

Structural variation drives enhancer hijacking via 3D genome disruption in ccRCC

Received: 12 August 2025

Accepted: 18 November 2025

Cite this article as: Dong, Y., Xia, W., Yang, Z. *et al.* Structural variation drives enhancer hijacking via 3D genome disruption in ccRCC. *npj Digit. Med.* (2025). <https://doi.org/10.1038/s41746-025-02186-w>

Yu Dong, Wenjiao Xia, Zitong Yang, Liangliang Ren, Hongru Wang, Yiyang Zhou, Qinchen Li, Zhi Chen, Zhinan Xia, Yichun Zheng, Feifan Wang, Ning He, Bing Cheng, Dongmei Ma, Wei Shao, Wei Guo, Shuwen Wang, Ziqiao Liu, Junxiao Shen, Yiming Qi, Xuke Gong, Juan Jin, Bo Xie, Guixin Zhu & Cheng Zhang

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Structural Variation Drives Enhancer Hijacking via 3D Genome Disruption in ccRCC

Yu Dong^{1#}, Wenjiao Xia^{1#}, Zitong Yang^{1#}, Liangliang Ren^{1#}, Hongru Wang^{1#}, Yiyang Zhou¹, Qinchen Li¹, Zhi Chen¹, Zhinan Xia², Yichun Zheng¹, Feifan Wang³, Ning He³, Bing Cheng⁴, Dongmei Ma⁵, Wei Shao⁶, Wei Guo¹, Shuwen Wang¹, Ziqiao Liu¹, Junxiao Shen¹, Yiming Qi¹, Xuke Gong¹, Juan Jin^{7,8*}, Bo Xie^{3*}, Guixin Zhu^{1*}, Cheng Zhang^{1*}

1. Department of Urology, Center for Oncology Medicine, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, China, 322000.
2. Department of Urology, the Fourth Affiliated Hospital of Harbin Medical University, Harbin, China, 150001.
3. Department of Urology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, 310003.
4. Department of Geriatric Medicine, Shulan (Hangzhou) Hospital, Zhejiang Shuren University Shulan International Medical College, Hangzhou, China, 310000.
5. Department of Anesthesiology, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, China, 322000.
6. State Key Laboratory of Green Chemistry Synthesis Technology, College of Chemical Engineering, Zhejiang University of Technology, Hangzhou, China, 310000.
7. Department of Nephrology, the First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, China, 310000.
8. Zhejiang Key Laboratory of Research and Translation for Kidney Deficiency-Stasis-Turbidity Disease, Hangzhou, China, 310000.

[#]*These authors contributed to the work equally.*

^{*}*Correspondence:*

Juan Jin, MD, 20233004@zcmu.edu.cn;

Bo Xie, MD, zdyxb@zju.edu.cn;

Guixin Zhu, PhD, zhuguixin@zju.edu.cn;

Cheng Zhang, PhD, zhangcheng13836182568@zju.edu.cn

Abstract

Clear cell renal cell carcinoma (ccRCC) is an aggressive malignancy associated with single-nucleotide variants in *VHL*, *PBRM1*, and *SETD2*. However, the role of structural variation (SV) —which have broader genomic impacts—and 3D genome architecture in ccRCC remains inadequately understood. Here, we reported a comprehensive molecular characterization. Through multi-omics analysis, we identify novel SV-associated oncogenic targets and reveal multidimensional 3D genome reorganization during ccRCC progression. We elucidate the dynamic interplay between SV and 3D chromatin architecture, demonstrating how structural rearrangements drive oncogenic dysregulation through 3D genome reorganization. Notably, an unrecognized pathogenic enhancer hijacking event is discovered and experimentally validated, leading to constitutive activation of the proto-oncogene *SEMA5B*. Furthermore, we developed a machine learning-based prognostic framework using enhancer hijacking signatures. Collectively, this work establishes a valuable resource for ccRCC research by elucidating how SVs and 3D genome reorganization collectively drive oncogenesis, and translates these findings into a clinically applicable prognostic tool.

Keywords

Structural variants, 3D genome structure, clear cell renal cell carcinoma, Hi-C, nanopore long-read sequencing, Machine learning, Prognostic model

1. Introduction

Renal cell carcinoma (RCC), one of the most common urological malignancies globally, was estimated to have caused approximately 430,000 new cases worldwide in 2020¹. Clear cell renal cell carcinoma (ccRCC), the predominant histological subtype, accounts for 70%-80% of all cases and is characterized by high heterogeneity and aggressiveness². Radical surgery (nephrectomy or partial nephrectomy) is the standard treatment; however, 20%-40% of patients experience recurrence and disease progression, and 25%-30% present with metastasis at diagnosis³. Unfortunately, surgical outcomes are poor for these patients, and ccRCC is inherently resistant to radiotherapy and chemotherapy, leading to a significantly reduced five-year survival rate. Therefore, improving the mortality rate of RCC depends on enhancing early detection sensitivity and developing new therapeutic targets, which necessitates a deeper understanding of the molecular biology and pathogenesis of this cancer.

With the rapid development of next-generation sequencing (NGS) technologies (short-read sequencing, SRS), our understanding of ccRCC has significantly deepened. Studies based on single nucleotide variants (SNVs) and small insertions/deletions (InDels) have identified mutations in key renal cancer genes such as *VHL*, *PBRM1*, *SETD2*, and *BAP1*, which have been proven to be critical initiating and driving factors in ccRCC³. However, the landscape and pathogenesis of larger-scale genomic alterations—structural variants (SVs)—remain underexplored in ccRCC. SVs, defined as insertions, deletions, duplications, inversions, and translocations of at least 50 base pairs, represent large-scale genomic alterations that account for a substantial proportion of genetic variation in the human genome⁴. SVs are ubiquitous in cancer⁵ and emerge during tumor development, progression, and therapeutic resistance⁶. They can dysregulate gene expression by amplifying, disrupting, or fusing cancer-related genes or by repurposing non-coding DNA regulatory elements⁷. While SVs were traditionally detected using next-generation sequencing, recent advances in long-read sequencing (LRS) technologies have made SV detection more precise and convenient⁸. Long-read sequencing can directly span SVs to identify breakpoint junctions, resolve high homology repetitive regions, and sequence complex genomic regions with high GC content⁹. Additionally, LRS employs single-molecule sequencing technology, which minimizes base mismatches¹⁰ and sequencing biases¹¹. Given these advancements, there is a compelling need to perform long-read sequencing on ccRCC genomes to precisely characterize SV features and uncover novel mechanisms and driver events.

Recent technological advances have made the detection of SVs more accessible; however, interpreting their pathogenicity and phenotypic consequences remains a formidable challenge. Notably, SVs not only exhibit dosage effects by altering gene coding but also exert position effects by influencing the location and/or function of cis-regulatory elements such as promoters and enhancers¹². Recent advances in three-dimensional genomic interaction technologies, such as high-throughput chromosome conformation capture (Hi-C) sequencing, have enabled the detection and characterization of the 3D genomic structure in tumors, thereby facilitating the evaluation of SV position effects¹³. Recent studies have revealed that dynamic changes in 3D chromatin organization are associated with the development of various solid tumors¹⁴⁻¹⁶. Importantly, SVs formation is influenced by 3D genome architecture¹⁷⁻¹⁹, while SVs can disrupt 3D genomic structures, thereby interfering with the precise regulatory networks of gene expression^{15,16}. This underscores the bidirectional interplay between SVs and 3D genomic organization, highlighting their dynamic and reciprocal relationship. Moreover, enhancer hijacking, as an important manifestation of the positional effects of structural variations (SVs), has been demonstrated to play a significant role in various cancers such as leukemia²⁰, high-grade glioma²¹, and glioblastoma²².

However, the interplay between SVs, 3D epigenomic organization, and dysregulated transcriptional programs in ccRCC remains underexplored, with critical gaps persisting in our understanding of its molecular pathogenesis. In this study, we obtained nanopore third-generation whole-genome sequencing (WGS), next-generation WGS, Hi-C, H3K27ac ChIP-seq, and RNA-seq data from two renal cancer cell lines (786-O and OS-RC-2) and an immortalized normal renal epithelial cell line (HEK293T) to comprehensively assess the structural variant landscape and 3D chromatin organization in renal cancer. We also performed RNA-seq on 12 clinical ccRCC samples to validate our findings. Overall, our study identifies novel variant targets in renal cancer from the perspective of structural variation, reveals multidimensional 3D genome reorganization during ccRCC oncogenesis, marked by compartment shifts, TAD boundary alterations, and chromatin loop reconfiguration. Our study further elucidates the dynamic interplay between SVs and the 3D genome, revealing how SVs drive aberrant oncogenic regulation through altering 3D chromatin organization.

Notably, we identify and validate a previously unrecognized pathological enhancer hijacking event that drives the constitutive activation of the proto-oncogene *SEMA5B*. Leveraging these mechanistic insights, we also developed a machine learning-based prognostic prediction model for ccRCC, founded on cancer-specific enhancer hijacking events. This model demonstrates robust predictive performance and stability. These findings not only expand our understanding of the genetic and molecular basis of ccRCC but also offer novel potential therapeutic targets and biomarkers for precision oncology, while providing a translational pathway for clinical application of the uncovered mechanisms.

2. Results

2.1 The Landscape of Structural Variation Signatures in ccRCC

In this study, we systematically characterized the spectrum of SVs associated with the tumorigenesis of normal renal epithelial cells using nanopore long-read whole-genome sequencing. Our nanopore sequencing data exhibited exceptional quality, with an N50 of approximately 50 kb and a mean sequencing quality of around 20 (Supplementary Figs.1a-1c, Supplementary Figs.1g-1h, Supplementary Tables 1 and 2). The data also achieved a mean map identity exceeding 95% and a mapping ratio above 99% (Supplementary Fig.1k, Supplementary Table 2), alongside sufficient sequencing depth ($\geq 25\times$) (Supplementary Figs. 1a-1f, Supplementary Table 2), and a throughput surpassing 43 Gb for reads longer than 50 kb (Supplementary Fig.1j).

Through alignment of the sequencing data to the GRCh38 (hg38) reference genome, we identified a substantial number of high-confidence SV events and constructed a detailed SV map (Fig. 1a, Supplementary Figs. 2d-2e). The total SV counts in 786-O, OS-RC-2, and HEK293T cells were 18,912, 18,792, and 21,571, respectively (Fig. 1b, Supplementary Tables 3 and 4). Consistent with patterns observed in other solid tumors^{23,24}, deletions and insertions were the predominant types, comprising approximately 45% and 51% of SVs in each tumor cell, respectively. However, unlike other tumors with a duplication rate of around 6%, ccRCC exhibited a markedly lower duplication rate, indicating that duplications may not be a defining feature of renal cancer²⁴. Our analysis of the length distribution of different SV types revealed that insertions and deletions shared similar patterns, with a prominent

peak near 300 bp and the majority of events occurring within 1 kb (Fig. 1c, Supplementary Fig. 2h). Furthermore, examination of the normalized counts of different SV types across each chromosome indicated that the distribution patterns of each specific SV type were largely consistent (Figs. 1e-1f, Supplementary Fig. 2j). In line with prior research^{25,26}, most SVs were located in intergenic and intronic regions (Fig. 1d, Supplementary Figs. 2f-2g).

We next investigated the prevalence of complex SVs, defined by overlapping breakpoint junctions. Comparative analysis revealed significantly higher frequencies of complex SVs in renal carcinoma cell lines (786-O and OS-RC-2) compared to normal renal epithelial cells (HEK293T) (Supplementary Table 5), suggesting enhanced genomic instability during carcinogenesis. To identify tumor-specific SVs potentially driving oncogenesis, we performed a systematic comparison between malignant and normal cell lines. Notably, approximately one-third demonstrated cell line specificity (Fig. 1b, Supplementary Table 6), implying functional relevance of lineage-restricted SVs in ccRCC pathogenesis. Subsequent analysis of SV-affected exonic regions identified 232, 234, and 255 genes disrupted in 786-O, OS-RC-2, and HEK293T, respectively (Fig. 1g, Supplementary Table 7). Importantly, 121 and 124 tumor-specific SV-associated genes in 786-O and OS-RC-2 were prioritized for functional characterization (Supplementary Table 8). Cross-referencing with the COSMIC database and TCGA mutation rank revealed 10 and 27 tumor-specific SV-associated genes corresponding to either canonical oncogenes or renal carcinoma-associated mutations (SNV-based top 30 candidates) (Supplementary Table 9). Of particular clinical relevance, both ccRCC models exhibited *HIF1A* alterations - a hypoxia-inducible oncogene regulated by *VHL* (a renal cancer gatekeeper gene) whose dysregulation correlates with poor prognosis²⁷⁻²⁹. In 786-O, we identified a 38-kb intragenic deletion corroborating previous cytogenetic and NGS reports of *HIF1A* loss (Figs. 1g-1h). Conversely, OS-RC-2 exhibited a novel duplication event, expanding the spectrum of *HIF1A* alterations in ccRCC. Furthermore, long-read sequencing uncovered a 2.6-kb insertion in *SETD2* (Supplementary Figs. 3f-3h), a chromatin modifier gene previously associated with renal oncogenesis through SNV/indel mutations^{3,30}. Collectively, these findings — including canonical ccRCC genomic hallmarks 3p loss, 5q gain³⁰ (Supplementary Figs. 2a-2b) — validated the molecular identity of our cell models. Both tumor lines shared mutations in *FLG* and *CAMKK2* — two genes ranked among the top 30 most frequently SNV-mutated loci in ccRCC (Supplementary Fig. 2k). Additionally, we detected a large ~258-kb deletion affecting the renal cancer-associated oncogene *NFIB* (Supplementary Fig. 2i). To independently

verify these findings, we performed multi-platform visualization of the relevant genomic regions across all three cell lines using Integrative Genomics Viewer (IGV) and Ribbon (Figs. 1h-1j).

We systematically compared the performance of third-generation and next-generation sequencing in detecting SVs in ccRCC. Long-read sequencing demonstrated superior sensitivity, identifying significantly more high-confidence SVs than next-generation sequencing (Supplementary Fig. 3a-3c). Genome-wide ideogram revealed a marked increase in breakpoint density detected by long-read sequencing, particularly in complex genomic regions such as centromeres and telomeres (Supplementary Figs. 3d-3e). Moreover, long-read sequencing resolved a large-scale insertion in *SETD2* (Supplementary Fig. 3f), a key renal cancer driver gene undetectable by next-generation sequencing. Consistent with prior studies²⁴, deletions and insertions dominated TGS-specific SVs, likely mediated by repetitive element activity³¹. Annotation of repeat-associated deletions revealed that TGS detected a higher burden of repetitive element-driven SVs compared to next-generation sequencing, with distinct compositional profiles: SINEs represented 58.8% of NGS-associated repeats but showed reduced proportions in long-read data, while other repeat classes (e.g., LINEs, LTRs) exhibited increased representation (Supplementary Fig. 3g). TGS-identified repeats displayed broader chromosomal distribution without positional bias (Supplementary Fig. 3j) and larger average sizes (Supplementary Fig. 3i). Subfamily-level classification and multi-repeat-overlapping SV analyses further refined these observations (Supplementary Fig. 4).

In summary, the SVs identified here directly impact oncogenes and tumor suppressor genes, potentially exerting significant influence on tumor malignancy and the maintenance of neoplastic phenotypes. Our long-read sequencing analysis delineates the landscape of SVs in human renal cancer, a previously underexplored area, and provides a valuable resource for future studies on the pathogenesis of ccRCC.

2.2 Extensive Remodeling of 3D Genome Organization Associates with Gene Expression Alterations in ccRCC

To investigate the global impact of SVs on higher-order chromatin structure, we analyzed the 3D genome of ccRCC cells and normal cells using Hi-C technology (Supplementary Figs. 5a-5c). Multi-scale comparative analysis revealed extensive reorganization of the 3D genome architecture in malignant cells compared to normal cell.

We first characterized A/B compartment distributions, observing compartment switching rates of 27.1% (14.0% A-to-B, 13.1% B-to-A) in 786-O and 27.1% (14.4% A-to-B, 12.8% B-to-A) in OS-RC-2 (Supplementary Table 11). Chromosomal-level profiling revealed a tendency toward compartment identity conservation, with stable A compartments representing the largest fraction in ccRCC cells (786-O: 39.5%; OS-RC-2: 38.8%; Fig. 2a). Consistent with the established role of compartments in gene regulation, compartment switches significantly correlate with gene expression changes, with stable A and B to A compartments showing higher expression, and A to B compartments exhibiting low transcription, suggesting silencing (Fig. 2b and Supplementary Fig. 5d, Supplementary Table 12). Further analysis indicated that the B-to-A compartment switch, despite involving fewer differentially expressed genes overall, was markedly enriched for transcriptionally activated oncogenes. This was statistically significant in both 786-O (30/322 vs. 6/388; $p = 2.745e-06$) and OS-RC-2 (22/299 vs. 4/342; $p = 6.99e-05$) cell lines, with the COSMIC-annotated oncogene *BTG1* serving as a representative example (Figs. 2c, 2e). Moreover, Pearson correlation heatmaps showed that cancer cell lines not only had extensive compartment remodeling but also significant enhancement of A/B compartment patterns (Fig. 2d).

At the finer scale of topologically associating domains (TADs), we identified 7,014, 7,823, and 7,580 TADs at 10 kb resolution in the 786-O, OS-RC-2, and HEK293T cell lines, respectively (Fig. 2f, Supplementary Table 13). Notably, 3,907 TADs were conserved across all three cell lines, representing 55.7%, 49.9%, and 51.5% of the total TADs in 786-O, OS-RC-2, and HEK293T, respectively. This high degree of conservation highlights the robustness of TADs in the human genome, which is consistent with previous studies^{15,32}. To characterize TAD alterations, we defined ccRCC-specific TADs (gained in tumor cells) and ccRCC-common TADs (shared between tumor cells but absent in normal cells). Our results showed that, compared to conserved TADs, ccRCC-specific and ccRCC-common TADs were significantly smaller in size (Fig. 2i), a finding similar to previous studies on solid tumors such as multiple myeloma, liver cancer, and prostate cancer¹⁴⁻¹⁶. We visualized interaction matrices using hicPlotTADs to

illustrate these characteristics, which clearly depicted the smaller TAD sizes on chromosomes 7 and 5 in the ccRCC cell lines (Figs. 2g-2h, Supplementary Figs. 5h-5i). Functionally, genes within ccRCC-specific TADs exhibited significantly higher expression levels than those in conserved TADs. (Fig. 2j, Supplementary Tables 15-16). Moreover, these genes were significantly enriched in pathways related to kidney cancer and other associated diseases (Fig. 2k, Supplementary Table 17), suggesting that these regions are hotspots for transcriptional dysregulation and oncogene activation.

In summary, our results demonstrate that ccRCC undergoes extensive cancer-specific remodeling of the 3D genome at both the compartment and TAD levels. These structural changes are closely associated with transcriptional dysregulation and the aberrant activation of oncogenes, implicating them as potential drivers of ccRCC development and progression.

2.3 Distribution of Structural Variants in 3D Chromatin Organization

Previous studies have established that somatic structural variant (SV) formation is influenced by and reciprocally modulates three-dimensional genome architecture¹⁷⁻¹⁹. To elucidate this relationship in ccRCC, we systematically investigated SV distribution patterns across hierarchical chromatin structures. Initial compartment-level analysis through permutation testing demonstrated significant enrichment of both ccRCC-specific and cell-type SVs in transcriptionally active A compartments (ccRCC-specific SVs: Z-score = 11.959, $P = 0.001$; 786-O SVs: Z-score = 19.221, $P = 0.001$; Fig. 3a). At local genomic scales, cancer cell lines (786-O and OS-RC-2) exhibited increased deletion and insertion densities in A compartments alongside decreased densities in B compartments, a pattern absent in normal cell (Fig. 3b, Supplementary Table 18). Further examination of compartment transition chromosomal regions revealed elevated deletion/insertion densities in stable A compartments and B to A transition regions (except OS-RC-2 insertion), while no significant changes were observed in A to B regions compared to genome-wide background levels (Supplementary Fig. 6a, Supplementary Table 20). These findings collectively suggest preferential SV accumulation in A compartments, potentially linked to the high transcriptional activity and consequent high DSB occurrence rate in A compartments, as well as to B to A compartment dynamics.

Analysis of tumor-specific SVs revealed distribution patterns mirroring those in parental cell lines, with deletion/insertion showing A compartment enrichment. Notably, 786-O-specific deletions displayed a significant reduction within B compartments (Fig. 3c, Supplementary Table 19). Compartment transition analysis showed ccRCC-specific deletions consistently enriched in B to A regions, while insertion and duplication distributions exhibited cell line-specific variations (Supplementary Fig. 6b, Supplementary Table 22). Despite density differences across compartment types, SV type composition remained comparable (Fig. 3d, Supplementary Fig. 6i).

TAD-level investigations demonstrated significant boundary enrichment for both ccRCC-specific and cell-type SVs (ccRCC-specific: $Z = 1.684$, $P = 0.043$; 786-O: $Z = 3.355$, $P = 0.001$; Fig. 3a). Moreover, boundary regions exhibited elevated densities for most SV types except 786-O duplication and OS-RC-2 insertion (Fig. 2b, Supplementary Table 18), consistent with prior researches^{15,19,33}. Comparative analysis revealed ccRCC-associated TAD alterations, with 786-O duplication and OS-RC-2 deletion/insertion preferentially localizing at lost TAD boundaries (Supplementary Fig. 6c). Tumor-specific SVs showed cell type-dependent associations with TAD remodeling, particularly OS-RC-2 insertion enrichment at disrupted boundaries (Supplementary Fig. 6d, Supplementary Table 23).

Classification of ccRCC-specific SVs by TAD positioning identified four distinct categories (Fig. 3e). The majority localized within TAD interiors (Inner TAD), suggesting minimal chromatin folding impact (Fig. 3f, Supplementary Fig. 6j). SVs with the potential to disrupt chromatin folding (i.e., those crossing TAD boundaries or partially overlapping with them) exhibited distinct SV type compositions, with deletions predominating in boundary-spanning events (Fig. 3g, Supplementary Fig. 6k).

These results collectively demonstrate that structural variant distributions in renal cell carcinoma exhibit non-random associations with three-dimensional chromatin architecture. Our analyses revealed preferential SV accumulation in transcriptionally active A compartments and their transition zones, accompanied by cell type-specific preferences in TAD boundary associations. Furthermore, the disruptive potential of SVs on topological organization appears contingent on variant type, with deletions predominantly affecting boundary-spanning regions.

The observed spatial coordination between somatic structural variation and chromatin architecture suggests a sophisticated interplay between genomic instability and 3D genome organization during renal carcinogenesis.

2.4 Interplay of ccRCC Specific SVs and TADs

Spatial genome organization is a critical determinant of chromosomal rearrangements and oncogenesis³⁴. Previous studies have demonstrated that SVs can induce alterations in chromatin architecture, particularly through disruption of TAD boundaries. Such perturbations may lead to fusion of adjacent TADs or aberrant spreading of active chromatin, thereby establishing oncogenic enhancer-promoter interactions^{12,18,25,35}. To investigate the impact of SVs on TAD disruption in renal cell carcinoma, we analyzed the correlation between distinct classes of TAD deletions and TAD fusion events. Our results revealed that deletions spanning TAD boundary regions exhibited significantly higher TAD fusion scores compared to other deletion types (Fig. 4a, Supplementary Table 24). These findings align with prior studies indicating that boundary-spanning deletions are strongly associated with TAD fusion³⁶⁻³⁸. Notably, SVs localized within inner TAD regions—even those of substantial size—showed minimal disruption to adjacent chromatin interactions, with neighboring TADs remaining largely unaffected (Fig. 4c, Supplementary Fig. 7d). Conversely, SVs traversing TAD boundaries induced a pronounced increase in chromatin interactions between adjacent TADs proximal to the deletion sites (Fig. 4d).

We further explored the relationship between TAD fusion and transcriptional dysregulation in two cancer cell lines. Differential gene expression analysis demonstrated that genomic regions harboring SVs with top 50% TAD fusion scores contained a significantly higher proportion of dysregulated genes compared to regions with lower fusion scores or genome-wide baselines (Fig. 4b, Supplementary Table 25). Strikingly, genes within high TAD fusion score regions exhibited elevated expression levels relative to the global genomic average (Fig. 4e).

These observations suggest that TADs maintain genome structural integrity, while their boundaries serve as vulnerable hubs susceptible to structural variation-driven remodeling of oncogenic circuits.

2.5 Remodeling of Enhancer and Focal Chromatin Interaction in ccRCC Oncogenesis

Having examined large-scale 3D genome architecture, we next focused on the enhancer landscape and focal chromatin interactions to elucidate the fine-scale regulatory logic underlying ccRCC pathogenesis. Using H3K27ac ChIP-seq data, we identified 27,340 ccRCC-specific and 9,752 normal-specific enhancer peaks across 786-O, OS-RC-2, and HEK293T cell lines (Supplementary Table 26). Comparative chromatin state analysis confirmed distinct enhancer activities: ccRCC-specific enhancers were active in tumor cell lines (786-O, OS-RC-2) but not in HEK293T cells, while normal-specific enhancers showed the opposite pattern (Fig. 5a). IGV plot of representative genomic loci further confirmed the robustness of our differential enhancer identification (Fig. 5b, Supplementary Fig. 9a). Functional annotation of tumor-specific enhancers through GO dataset based GREAT analysis revealed significant enrichment for pathways central to ccRCC biology, including angiogenesis³⁹ and hypoxia response⁴⁰ (Fig. 5d, Supplementary Table 28). In contrast, normal-specific enhancers were associated with urinary system developmental processes such as mesonephric tubule formation and ureteric bud morphogenesis (Fig. 5e). Mouse phenotype enrichment analysis demonstrated tumor-specific enhancer associations with renal morphological abnormalities (Supplementary Fig. 9c), reinforcing the biological relevance of our findings. Motif analysis identified Jun-AP1, ZNF416, NEUROD1, ZNF669, ZEB1, and MGA as key transcription factors potentially orchestrating enhancer reprogramming in ccRCC (Supplementary Fig. 9b).

We used Mustache at 5kb resolution to identify chromatin interactions genome-wide, detecting 3,272 loops in 786-O, 3,396 in OS-RC-2, and 2,225 in HEK293T cells (Supplementary Table 29). Differential analysis revealed 223 ccRCC common, 55 conserved, and 377 normal-specific loops (Supplementary Table 30). Aggregate peak analysis demonstrated pronounced clustering of ccRCC common interaction signals in tumor samples compared to normal counterparts (Fig. 5c). Notably, ccRCC-common loops exhibited significantly higher active enhancer signals compared to conserved loops and random control regions (Fig. 5g, Supplementary Table 32). Chromatin interaction heatmaps displayed long-range ccRCC common interactions, such as those involving *MAP4K4*, *RFX8*, and *CREG2* (Fig. 5f), with tumor common enhancer signals observed in these regions. To gain deeper insights into the impact of these remodeled interactions, we integrated chromatin interaction data with transcriptomic profiles. Our analysis

revealed that the anchors of ccRCC common loops were linked to markedly higher gene expression levels in comparison to all identified loops or conserved loops (Figs. 5h-5i, Supplementary Table 31).

In summary, our results demonstrate that ccRCC pathogenesis involves extensive enhancer reprogramming and a remodeling of focal chromatin interactions. The enrichment of active enhancers within ccRCC-specific loops and their strong association with elevated gene expression underscore the role of fine-scale 3D genome restructuring in driving transcriptional dysregulation during renal carcinogenesis.

2.6 Identification of Enhancer Hijacking Events Driving Oncogenic Dysregulation in ccRCC

Recent studies have revealed that structural variations (SVs) can aberrantly activate oncogenes through enhancer hijacking, a mechanism whereby SVs reposition enhancer elements into proximity with oncogenic promoters^{41,42}. To systematically identify such events in ccRCC, we applied Neoloopfinder. Based on comprehensive consensus SV data from TGS-WGS and Hi-C - corrected genomic CNVs, we identified ectopic chromatin interactions near SV breakpoints that activate oncogenes, termed neoloops (Fig. 6a, Supplementary Table 33). Aggregate peak analysis revealed significant tumor-specific enrichment of neoloop signals compared to normal tissue controls (Fig. 6d). Transcriptomic profiling demonstrated elevated expression of genes localized to neoloop anchor regions in tumor cell lines (Fig. 6e, Supplementary Table 34), supporting the functional relevance of these aberrant chromatin interactions in oncogene activation. Notably, systematic annotation identified multiple cancer drivers within these neoloop domains, including established oncogenes *THRAP3* and *DIRC1* (Supplementary Figs. 10a-10b). Among the identified neoloops, we characterized a complex SV involving a coordinated chromosomal inversion and translocation that repositioned a distal enhancer cluster (chr3:128,620,000-128,930,000) to a proximal position relative to the *SEMA5B* promoter (chr3:122,450,000-123,000,000) (Figs. 6c, 6g). Hi-C contact maps confirmed complete absence of promoter-enhancer interactions at this locus in normal cells, while tumor samples exhibited robust chromatin connectivity coinciding with *SEMA5B* transcriptional upregulation (Figs. 6c, 6f).

To validate the physical existence of these SVs, we performed experimental verification. Dual-color DNA-FISH using breakpoint-flanking probes demonstrated tumor-specific co-localization of chr2 and chr3 fragments (Fig. 6h, Supplementary Figs. 10c-10d), confirming the somatic origin of the translocation. Breakpoint junction PCR followed by Sanger sequencing (primer locations in Fig. 6j) successfully amplified and validated heterozygous inversion/translocation events in tumor cell lines (Figs. 6i, 6k). These findings were further corroborated by single-molecule long-read sequencing data showing continuous DNA reads spanning both rearrangement junctions exclusively in tumor samples. No evidence of these structural rearrangements was detected in matched normal cell lines across all experimental modalities.

In summary, our results demonstrate that SVs in ccRCC can disrupt local chromatin architecture to mediate long-range enhancer hijacking, thereby creating oncogenic enhancer-promoter loops. This 3D genome reorganization facilitates tumor-specific transcriptional activation by aberrantly engaging enhancers with new target genes. Furthermore, our multi-platform validation framework provides conclusive evidence for the existence of the complex SVs underlying these pathogenic rewiring events.

2.7 Enhancer Hijacking-Mediated *SEMA5B* Upregulation Promotes Tumorigenesis in ccRCC

Although semaphorins are emerging as clinical biomarkers and therapeutic targets in cancer⁴³, and *SEMA5B* has been implicated in various malignancies (47, 48) including as a potential target in renal cancer⁴⁴, its specific oncogenic role in ccRCC remains poorly characterized. To address this gap, we first evaluated the expression of *SEMA5B* across pan-cancer datasets using The Cancer Genome Atlas (TCGA). Results revealed that *SEMA5B* exhibits high expression predominantly in brain and renal tissues, with significantly elevated levels in renal cancer compared to normal renal tissue (Figs. 7a-7b), suggesting renal ccRCC-specific overexpression of *SEMA5B*. Further analysis demonstrated low transcriptomic and protein expression of *SEMA5B* in normal cell lines such as HK-2 and HEK293T, whereas OS-RC-2 renal cancer cells exhibited markedly higher expression (Figs. 7e-7f). These findings were validated in clinical samples using RNA-seq data and RT-qPCR, which confirmed tumor-specific overexpression of *SEMA5B* at the transcriptomic level (Figs. 7c-7d). Immunohistochemistry (IHC) further corroborated tumor-specific upregulation of *SEMA5B* at the protein level (Figs. 7g, 7i).

To investigate the regulatory mechanisms underlying *SEMA5B* overexpression, we identified enhancers (e4-e6) involved in ectopic interactions based on the coordinates of the neo-loop structures. We designed sgRNAs targeting these enhancers and the promoter region of *SEMA5B* for CRISPR interference (CRISPRi) experiments (Fig. 7h, Supplementary Table 36) and synthesized siRNAs targeting *SEMA5B* exons. The efficiency of knockdown (KD) and CRISPRi was validated via western blotting and RT-qPCR (Figs. 7k-7l, Supplementary Fig. 10e). CRISPRi targeting the ectopic enhancers significantly reduced *SEMA5B* expression (Figs. 7k-7l), validating the reliability of neo-loop identification and its transcriptional regulatory effects.

We further assessed the phenotypic consequences of disrupting the ectopic enhancers and promoter of *SEMA5B*. In vitro experiments revealed that ablation of these regulatory elements markedly attenuated cellular proliferation (Figs. 7j, 7p, and 7u) and invasive capacity (Figs. 7q and 7t), findings that were consistent with KD results (Figs. 7m, 7o, and 7s). In vivo experiments further confirmed that disruption of the ectopic enhancers and promoter of *SEMA5B* significantly slowed tumor proliferation (Figs. 7n, 7r, Supplementary Figs. 10f-10g).

In summary, our integrated analysis—spanning clinical cohorts, cellular models, and functional genomics—establishes *SEMA5B* as a ccRCC-specific oncogene activated by enhancer hijacking. The consistent phenotypic evidence from in vitro and in vivo models underscores the critical role of this neo-loop-mediated transcriptional rewiring in ccRCC pathogenesis, nominating *SEMA5B* as a promising therapeutic target for renal malignancies.

2.8 A Neoloop-Based Machine Learning Model for Prognostic Prediction in ccRCC

To evaluate the clinical potential of our findings, we developed a machine learning model based on ccRCC-specific neoloop-associated genes using the TCGA-KIRC cohort. This model demonstrated robust predictive performance, achieving area under the curve (AUC) values of 0.747 in the training cohort, 0.740 in the test cohort, and 0.743 in the entire patient cohort. These results surpassed the predictive capability of the traditional WHO grading system (AUCs: 0.745, 0.703, and 0.720, respectively; Figs. 8a-8c). Furthermore, the model exhibited stable and sustained

predictive accuracy over time, maintaining consistent performance across 1- to 5-year follow-up periods (Figs. 8d-8f). We also constructed a nomogram integrating multiple clinical parameters to predict overall survival (OS) in ccRCC patients (Fig. 8h). Calibration curve analysis confirmed the model's robust predictive performance (C-index = 0.767; Fig. 8g).

Based on the model-derived risk scores, patients were stratified into distinct high-risk and low-risk groups, which exhibited significantly different survival outcomes (Supplementary Figs. 11c-S11e). Kaplan-Meier survival analysis for Overall Survival (OS) and Progression-Free Survival (PFS) revealed significant divergence between the survival curves of the high-risk and low-risk groups in both the training cohort and the full TCGA-KIRC cohort ($p < 0.001$; Figs. 8i-8l), indicating the model's efficacy in discriminating patients with differential risk levels. Importantly, this significant separation of survival curves was consistently replicated in the testing cohort ($p < 0.001$; Figs. 8m-8n), confirming the model's robustness and predictive accuracy across independent patient subsets.

Notably, the model-derived risk scores showed significant positive correlations with established indicators of disease severity. Patients with more advanced disease—including those of older age, higher pathological TNM stage, elevated clinical stage, and advanced WHO grade—consistently exhibited increased risk scores (all $p < 0.05$; Supplementary Figs. 12a–12g). These associations further validate the clinical relevance and biological coherence of our neoloop-based prognostic model.

3. Discussion

RCC, particularly ccRCC, poses a significant clinical challenge due to its high metastatic potential and resistance to conventional therapies. Over the past decade, genomic studies have revealed recurrent somatic mutations in genes such as *VHL*, *PBRM1*, *SETD2*⁴⁵, and other epigenetic regulation related genes, underscoring the importance of genetic and epigenetic dysregulation in ccRCC pathogenesis. However, these efforts have primarily relied on next-generation sequencing, which focuses on single nucleotide variants (SNVs) and small insertions/deletions (indels),

leaving the contribution of structural variants (SVs)—large-scale genomic alterations including deletions, duplications, inversions, and translocations—poorly understood.

Recent advances in long-read sequencing and high-throughput chromatin conformation capture (Hi-C) have significantly deepened our understanding of the critical roles of SVs and three-dimensional (3D) genome architecture in tumorigenesis and cancer progression. However, in the context of ccRCC, the dynamic interplay between SVs, 3D chromatin organization, and their mechanistic contributions to malignant transformation remains poorly understood. In this study, we systematically integrated SV landscapes, 3D genome organization, epigenomic profiles, and transcriptomic data in ccRCC, to provide a comprehensive view of multidimensional alterations in chromosomal architecture and to elucidate how SVs interact with higher-order chromatin structure to drive oncogenic dysregulation. These findings expand our understanding of the molecular networks underlying ccRCC pathogenesis and highlight novel regulatory mechanisms that may inform therapeutic strategies.

To establish a detailed SV profile in ccRCC, we leveraged Nanopore long-read sequencing to generate a tumor-specific SV atlas, identifying over 18,000 high-confidence SVs. Our analysis offered substantially improved resolution over traditional next-generation sequencing, particularly in complex genomic regions such as repetitive sequences and structural rearrangements. Moreover, long-read sequencing identified a large-scale insertion in *SETD2* and duplication and deletion events affecting *HIF1A*. These mutations in key renal cancer driver genes were undetectable in previous research using next-generation sequencing. Previous studies have established that the histone methyltransferase *SETD2* maintains H3K36me3 to suppress chromatin accessibility. However, loss of *SETD2* function leads to a more open chromatin state and subsequent enhancer activation⁴⁶. A pertinent question for future research is whether this mutational background increases the proportion of active enhancers, thereby providing more opportunities for enhancer hijacking events. Furthermore, while the conventional view holds that *VHL* mutation upregulates *HIF1A* expression, thereby driving tumor-associated transcriptional changes²⁷, our study reveals that *HIF1A* itself can be subject to duplication variants. It remains to be explored whether such structural variations alone can lead to *HIF1A* transcriptional accumulation, and whether, in the context of concurrent *VHL* mutation—which impairs HIF degradation—this would result in a more pronounced oncogenic function of HIF1A.

The frequency and impact of these previously overlooked structural variants warrant further validation in large-scale cohorts utilizing third-generation sequencing data.

Notably, immortalized HEK293T cells exhibited an even higher SV burden (21,571 SVs), a phenomenon potentially linked to genomic instability induced by SV40 T antigen during the immortalization process. Furthermore, SVs showed a non-random genomic distribution, with the majority localized to intronic and intergenic regions rather than exonic sequences. This distribution pattern suggests that SVs may primarily contribute to tumor evolution by altering regulatory landscapes while maintaining coding sequence integrity—a balance that preserves genomic stability while enabling phenotypic diversity. At the 3D genome level, SVs were significantly enriched in transcriptionally active A compartments and at TADs boundaries, implying that chromatin architecture directly influences SV formation. Mechanistically, SV generation in these regions may arise from two major pathways: replication-based mechanisms, such as microhomology-mediated break-induced replication (MMBIR), and fusion-based mechanisms involving erroneous repair of double-strand breaks (DSBs)³⁵. The increased vulnerability of A compartments to DSBs is primarily attributed to their open chromatin configuration, active transcriptional processes, and the abundance of regulatory elements^{47,48}. These features collectively create a permissive environment for SV formation. For instance, the TOP2B enzyme, which relaxes DNA supercoils, can introduce DSBs in these active regions, thereby increasing SV risk³⁵.

The pathogenic mechanisms of structural variations (SVs) can be categorized into position effects and dosage effects. SVs that disrupt or encompass coding sequences are typically interpreted based on their impact on gene dosage⁴⁹. However, it is noteworthy that 98% of the human genome is non-coding⁵⁰, and a significant portion of this non-coding genome is involved in gene regulation⁵¹. In reality, SVs can cause diseases without altering the genomic sequence⁵². They achieve this by affecting the position and/or function of cis-regulatory elements such as promoters and enhancers¹², thereby influencing the expression of genes located far from the SV breakpoints and contributing to carcinogenesis and disease progression.

In this study, we first investigated SVs that directly impact the exons of oncogenes to explore the oncogenic dosage effects of SVs. Our findings revealed deletions and duplications in crucial oncogenes such as *HIF1A*, *NFIB*, and *CNTN3* that had previously gone unnoticed. Subsequently, we shifted our focus to the impact of local SVs on chromatin folding. We discovered that local SVs can induce the fusion of adjacent chromatin domains, leading to alterations in gene expression within neighboring TADs. Mechanistically, SVs spanning TAD boundaries remove the insulation effect of the original CCCTC-binding factor (CTCF)-associated boundary elements, triggering enhancer translocation and potentially disrupting enhancer-promoter communication, resulting in abnormal gene expression. Moreover, the TAD structural perturbations caused by TAD boundary-spanning deletions appear to be confined within adjacent TADs. This suggests that TADs tend to limit the effects of SVs to the maximum extent, thereby preserving the overall stability of the three-dimensional chromatin architecture. This observation is consistent with the evolutionary conservation of TADs, which may be a result of adaptive selection under natural evolutionary pressures.

To further investigate the oncogenic mechanisms of SV position effects in renal cell carcinoma (RCC), we focused on copy-number balanced distal SVs (including inversions and translocations) that are distant in the genome and disrupt chromatin loops. We first identified tumor-specific enhancer libraries closely associated with the biological functions and phenotypes of RCC. Then, by integrating genomic alterations, 3D genomic interactions, tumor-specific enhancers, and transcriptomic data, we identified enhancer hijacking as a novel oncogenic mechanism in ccRCC, exemplified by the activation of oncogenes such as *SEMA5B*, *DIRC1*, and *THRAP3*. We used long-read sequencing, DNA-FISH, PCR, and Sanger sequencing to validate the complex genomic breakpoints involved in enhancer hijacking. Hi-C analysis revealed ccRCC-specific translocation interaction signals, confirming the authenticity of the *SEMA5B* enhancer hijacking event. Further analysis of clinical samples and the TCGA database showed that *SEMA5B* is significantly overexpressed in ccRCC. We used CRISPRi to knock out the translocated enhancer clusters and promoter, and validated the oncogenic effects of *SEMA5B* both in vitro and in vivo. *SEMA5B* influences tumor progression through various mechanisms, including regulating tumor angiogenesis⁵³ and inducing tyrosine kinase phosphorylation to promote tumor metastasis⁵⁴. Recent studies indicate that the interaction between SEMA and its plexin receptor is targetable⁵⁵. Given the overexpression of *SEMA5B* in ccRCC and our phenotypic experimental results, there is an opportunity for intervention. Notably, *SEMA5B* has also been identified as an

optimal therapeutic target for antibody drug conjugates (ADCs) or small molecule drug conjugates (SMDCs) in renal cancer⁵⁶. Additionally, small-molecule inhibitors targeting tyrosine kinases have been approved for clinical treatment of ccRCC. Therefore, the downstream events of *SEMA5B* oncogene activation via the enhancer hijacking mechanism are worthy of further study and may provide new targets for clinical treatment.

Our study provides novel insights into the pathophysiological mechanisms of renal cell carcinoma, offering a foundation for advancing precision risk prevention strategies. Leveraging the comprehensive landscape of structural variation (SV)-mediated enhancer hijacking events, we constructed a risk assessment model using machine learning. This model demonstrated sensitive prognostic risk prediction and robust performance, effectively stratifying patients into distinct prognostic groups and establishing a new pathway for translating novel mechanistic discoveries into clinical applications.

In conclusion, our multi-omics approach reveals how SVs reshape 3D chromatin organization to activate oncogenic pathways in ccRCC, bridging the gap between structural genomic alterations and transcriptional dysregulation. While this study provides critical insights, limitations such as sample size and the absence of longitudinal data warrant further investigation. Future work should explore the temporal dynamics of SV accumulation during tumor evolution and assess the therapeutic efficacy of targeting enhancer hijacking events. By elucidating the interplay between SVs and 3D genome organization, this research not only advances our understanding of ccRCC biology but also opens new avenues for precision oncology. The resulting risk prediction model further translates these mechanistic insights into a clinically applicable tool for improved patient stratification.

4. Materials and Methods

Cell Culture

Human renal cell carcinoma cell lines 786-O and OS-RC-2 and human normal kidney cell lines HK-2 and HEK293T were obtained from the American Type Culture Collection (ATCC) (<https://www.atcc.org/>). All cell lines were cultured under recommended conditions, and 786-O, OS-RC-2, HK-2, and HEK293T was authenticated by

high-resolution small tandem repeat profiling and next-generation whole genome sequencing. The cell lines used in this study were identified by the following Research Resource Identifiers (RRIDs): 786-O (CVCL_1051), OS-RC-2 (CVCL_1626), and HEK293T (CVCL_0063). WGS data analysis revealed that the average alignment rate of the cell line reads to the human reference genome (hg38) exceeded 90% (Supplementary Table 2), confirming the absence of contamination in all cell lines.

Ultra-long genomic DNA extraction

Ultra-long genomic DNA was extracted from 5 million fresh ccRCC and normal kidney cells. The cell pellet was resuspended in 500 μ L of pre-cooled NSB and incubated on ice for 3 min. After centrifugation at 500 g for 5 min at 4 °C, the pellet was washed with 1000 μ L of wash buffer and centrifuged again under the same conditions. The pellet was then lysed in 5 mL of SDS extraction buffer (#3250GR500, BioFroxx) with 20 μ L of RNase A (100 mg/mL) (#9001-99-4, Solarbio) at 50 °C for 1 h, with gentle inversion every 20 min. The lysate was cooled, centrifuged at 5000 g for 10 min at 4 °C, and the supernatant was extracted with phenol/chloroform/isoamyl alcohol (25:24:1), followed by chloroform/isoamyl alcohol (24:1). DNA was precipitated with 0.8 volumes of chilled isopropanol, pelleted by centrifugation, washed twice with 75% ethanol, and air-dried. The DNA was resuspended in 200 μ L of Elution Buffer (#19086, Qiagen) and incubated overnight at 4 °C. DNA quality and concentration were measured using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) and Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). Fragment integrity and size distribution were preliminarily assessed by agarose gel electrophoresis prior to library construction.

Nanopore Long-read Whole Genome Sequencing

High-molecular-weight (HMW) genomic DNA (gDNA) was isolated from renal cancer cell lines 786-O and OS-RC-2, with detailed extraction procedures provided in the Supplementary information. The purified HMW gDNA was subsequently prepared for nanopore sequencing, comprising end-repair, adapter ligation with the SQK - LSK114 ligation sequencing kit (Oxford Nanopore Technologies, Oxford, UK), and purification via AMPure XP beads (#A63881, Beckman Coulter, Brea, USA). Sequencing was conducted on the PromethION platform (Oxford Nanopore Technologies, Oxford, UK) using R10.4.1 flow cells in accordance with the manufacturer's standardized

protocols. Base calling and FASTQ conversion were performed with MinKNOW (v23.07.12) (Oxford Nanopore Technologies, Oxford, UK) and Dorado (v0.9.0) (Oxford Nanopore Technologies, Oxford, UK). NanoPlot (v1.40.0) was employed for the quality control analysis of long-read sequencing data⁵⁷.

Next-generation Whole Genome Sequencing

Total genomic DNA from 786-O and OSRC-2 was extracted using TIANamp Genomic DNA kit (#DP304, TIANGEN, Beijing, China). RNase A (#19101, Qiagen, Hilden, Germany) was used to obtain RNA-free genomic DNA. Then gDNA was sheared using Covaris instruments to retrieve DNA fragment sizes from 400 to 500 bp. The DNA fragment sizes were determined using a Qsep100™ fully automated nucleic acid and protein analysis system (BiOptic Inc., New Taipei City, China). Then the DNA fragment was used for library construction with both NovaSeq X Plus, Novaseq 6000, and BGI DNBSEQ-T7. The library preparation was conducted according to the manufacturer's instructions (#ND607-01 and #NDM627-01, Vazyme, Nanjing, China).

Identification of Structural Variants

For long-read whole genome sequencing (WGS) data, Minimap2 (v2.24)⁵⁸ was applied to perform genomic mapping with default parameters. Sniffles (v2.0.7)²⁶, NanoSV (v1.2.4)⁵⁹, and cuteSV (2.0.0)⁶⁰ were used for long-read structural variants (SV) identification with default parameters. For next-generation WGS data, BWA (v0.7.17)⁶¹ was applied to perform genomic mapping with default parameters. Delly (v1.1.8)⁶², Lumpy (v0.3.1)⁶³, and Manta (v1.6.0)⁶⁴ were used for short-read SV identification with default parameters. Homo sapiens genome assembly GRCh38 (hg38) was used as reference genome. Consensus SVs were identified using SURVIVOR (v1.0.7)⁶⁵ when at least two SV callers detected them as the same type, in the same genomic orientation, with breakpoints within 1000 bp of each other for variants larger than 50 bp. ccRCC-specific SV was also identified by SURVIVOR with similar parameters. The SV classification was comprehensively defined with each SV caller. The SVs retrieved were annotated with AnnotSV (v3.4.4)⁶⁶. Complex SVs were characterized as local assemblies composed of multiple SV junctions originating from distinct genomic locations. In contrast, simple SVs were defined as those containing only a single junction event. Cancer gene intersects were identified using BEDTools (v2.30.0)⁶⁷ to analyze the intersection between consensus-variants and the GENCODE hg38 annotation, followed by

filtering against the COSMIC Cancer Gene Census (v98)⁶⁸ and the Bushman Lab's cancer-related gene list (http://www.bushmanlab.org/assets/doc/allOnco_May2018.tsv). The Circos plots depicted in Fig. 1 and Fig. 6 were generated using Circa (<http://omgenomics.com/circa>) with BED files encompassing diverse SVs. SV was visualized using IGV (v2.14.1)⁶⁹ and Ribbon (v1)⁷⁰.

Hi-C Library Construction and Sequencing

Ten million crosslinked cells were lysed in 100 μ L of 0.55% SDS at room temperature, 62°C and 37°C sequentially for 10 min each, followed by the addition of 50 μ L 10% Triton X-100 to quench the SDS and incubation at 37°C for 30 min. To fragmentize the chromatin, nuclei were digested with DpnII (#R0543, NEB) at 37°C overnight. The digested chromatin was filled-in with dNTP (#N0447, NEB) and DNA Polymerase I, Large (Klenow) Fragment (#M0209, NEB) at 37°C for 1.5 hours, and then subjected to dA-tailing in a solution containing 1 $\times$ CutSmart buffer, BSA (#B9000S, NEB), dATP (#N0440, NEB), and Klenow Fragment (3' $\rightarrow$ 5' exo-) (#M0212, NEB) at room temperature for 1 hour. Proximity ligation was performed by adding biotinylated bridge linker, T4 DNA ligase (#M0202, NEB) and incubating at 16°C overnight.

The ligated DNA was then decrosslinked with Proteinase K and purified using a ChIP DNA Clean & Concentrator kit (#D5205, Zymo Research). The purified DNA was sheared by sonication to 400 bp using covaris machine. Biotinylated fragments were captured with M-280 streptavidin Dynabeads (#11206D, Thermo Fisher Scientific). The end repair, A-tailing, and adaptor ligation were performed on-beads using a VAHTS Universal DNA Library Prep Kit for Illumina V3 (#ND607, Vazyme). The final library was amplified by PCR, purified with Ampure beads (#A63881, Beckman Coulter), and quality was controlled using Qubit high-sensitive DNA (Thermo Fisher Scientific) and Qsep100™ (BiOptic Inc.) analysis. Hi-C library sequencing was performed using the Illumina HiSeq X Ten platform with 150 bp paired-end reads.

Hi-C Contact Matrix Building and Data Normalization

For Hi-C data, read quality was evaluated using FastQC (v0.12.1) (<https://github.com/s-andrews/FastQC>) and processed with Trim Galore (v0.6.7) (<https://github.com/FelixKrueger/TrimGalore>) for quality control. Trimmed reads were aligned against the human reference genome (GRCh38/hg38) utilizing Bowtie2 (version 2.5.1)⁷¹ with default mapping parameters. Subsequent data processing implemented the Hi-C Pro software (v2.10.0)⁷² to systematically eliminate non-informative read pairs through multistage filtering. This rigorous quality control process excluded unmapped reads, multimapped reads, and invalid paired-end reads including dangling-end pairs, re-ligation pairs, self-cycle pairs, single-end pairs, and dumped pairs, retaining only uniquely mapped valid interactions for downstream analysis. The interaction matrices at various resolutions (bin sizes of 1 Mb, 500 kb, 100 kb, 40 kb, 20 kb, and 5 kb at the genome-wide level) were constructed using HiC-Pro software (version 3.1.0) with default parameters⁷². To remove potential Hi-C interaction bias, an improved computational efficiency ICE (Iterative Correction) method was utilized⁷².

For bridge linker Hi-C data, raw sequencing reads underwent quality assessment using FastQC (v0.12.1) (<https://github.com/s-andrews/FastQC>) and processed with Trim Galore (v0.6.7) (<https://github.com/FelixKrueger/TrimGalore>) for quality control. Trimmed reads were aligned to the human reference genome (GRCh38/hg38) using the Burrows-Wheeler Aligner (BWA) (v0.7.17)⁷³, employing a hybrid of BWA-ALN and BWA-MEM algorithms. Only uniquely mapped reads with a mapping quality score (MAPQ) ≥ 30 were retained. The pipeline categorizes read pairs based on bridge-linker sequences: (i) no linker (none-tags), (ii) a linker and one genomic tag (mono-tags), and (iii) a linker and paired-end tags (di-tags). For di-tags, each tag was extended by 500 bp in the 5' direction to identify chromatin interaction anchors. Interaction matrices at multiple resolutions (1 Mb, 500 kb, 100 kb, 40 kb, 20 kb, and 5 kb) were constructed using processed PETs. Matrices were normalized using the ICE (Iterative Correction) method to correct for biases in Hi-C data⁷².

Identification of A/B Compartments

A/B compartments were determined using Juicer tools⁷⁴ with Hi-C interaction matrices normalized by the Knight-Ruiz (KR) algorithm. Eigenvectors were computed for each chromosome at 100-kb resolution through principal component analysis of Pearson's correlation matrices derived from intrachromosomal interaction frequencies.

Compartment assignments were established by correlating eigenvectors with genome-wide gene density profiles (wigCorrelate). Eigenvector signs were inverted for chromosomes exhibiting negative correlations to align transcriptional activity (A compartments) with gene-rich regions. Final compartment annotations were merged across all chromosomes and formatted into a genome-wide bedGraph file. A/B compartment switches were called when genomic coordinates showed discordant eigenvector signs (positive/negative) between tumor and normal samples after eigenvector sign normalization.

Identification of TAD Boundaries and Domains

TADs were identified using the hicFindTADs algorithm in HiCExplorer (v2.2.3)⁷⁵ with normalized Hi-C interaction matrices at 10-kb resolution. Key parameters included a threshold comparison value (--thresholdComparisons) of 0.05 and a boundary detection sensitivity (--delta) of 0.01. TAD domains were defined as genomic regions flanked by two consecutive statistically validated boundaries. Differential TAD analysis was performed using HiCExplorer's hicDifferentialTAD tool (v2.2.3)⁷⁵ with normalized 10-kb-resolution matrices. Control sample TAD architectures served as reference frameworks to quantify structural divergence in experimental samples (e.g., ccRCC vs. normal). Interaction frequency differences within corresponding TAD regions were statistically evaluated using Wilcoxon rank-sum tests, followed by false discovery rate (FDR) correction. Genomic regions exhibiting significant divergence (FDR-adjusted $P < 0.01$) were classified as differential TADs, while regions below this threshold were categorized as conserved TADs.

Identification of Chromatin Loops

Chromatin loops were detected using Mustache (v1.2.0)⁷⁶ with its default parameters, analyzing 5-kb-resolution Hi-C interaction matrices. Statistically significant loops were retained following filtration based on interaction frequency thresholds that surpassed genome-wide background levels ($P < 0.05$). Differential chromatin loops were further analyzed using the differential analysis module (diff_mustache.py) implemented in Mustache (v1.2.0)⁷⁶, with statistical significance thresholds applied ($P < 0.05$).

Quantification of TAD Structural Disruption

To assess potential structural damage caused by deletion-induced TAD disruption, we employed the TAD fusion score⁷⁷ with default parameters to quantitatively evaluate the degree of TAD boundary disintegration.

Bulk RNA-seq and Data Analysis

Total RNA was extracted using the RNAeasy Mini Kit (#74104; Qiagen; Hilden; Germany) following the manufacturer's protocol. Poly(A)-enriched strand-specific libraries were constructed according to the manufacturer's instructions (#NRM605 and #NR60, Vazyme, Nanjing, China) and sequenced on the Illumina NovaSeq 6000 or DNBSEQ-T7 platform. Three biological replicates per sample were generated, yielding 20 million reads per replicate. Read quality was evaluated using FastQC (v0.12.1) (<https://github.com/s-andrews/FastQC>) and processed with Trim Galore (v0.6.7) (<https://github.com/FelixKrueger/TrimGalore>) for quality control. Reads were aligned to the human reference genome (hg38) using STAR (v2.7.11)⁷⁸, with gene expression quantification performed using HTSeq-count (v2.0.4)⁷⁹. Downstream statistical analyses were conducted in RStudio, with differential gene expression analysis performed using DESeq2 (v1.38.3)⁸⁰. Differentially expressed genes were identified based on a Benjamini-Hochberg adjusted q-value threshold of < 0.05 coupled with a minimum absolute log2-fold change of ≥ 1 .

ChIP-seq Data Analysis

Similar quality control procedures were implemented using FastQC (v0.12.1) and Trim Galore (v0.6.7). Reads were aligned to the hg38 genome using Bowtie 2 (v2.5.1)⁷¹, with duplicate removal performed using SAMtools (v1.18)⁸¹. For ChIP-seq data, peak calling was conducted using MACS2 (v2.2.9.1)⁸² with specific parameters: -f BAMPE, -ghs, --nomodel, -B, and -q 1e-2. BigWig tracks were generated using deepTools bamCoverage (v3.5.1)⁸³ with RPGC normalization. Differential analysis between cancer and normal samples was performed using BEDTools (v2.30.0)⁶⁷. Data quality was assessed according to ENCODE standards for histone ChIP-seq. Track visualization of H3K27ac ChIP-seq profiles at specific regions was performed using IGV (v2.14.1)⁶⁹, and heatmap of H3K27ac

ChIP-seq signals across ccRCC-specific and normal-specific regions were generated by the computeMatrix and plotHeatmap algorithms from deepTools (v3.5.1)⁸³.

Analysis of SV Distribution Across the 3D genome Architecture

Dynamic chromosomal regions were defined by contrasting cancer cell lines with normal kidney cell lines. Specifically, comparing 786-O (or OS-RC-2) with HEK293T cell lines enables classification of compartment regions into stable A, stable B, A-B, and B-A compartments; and chromosomal regions into ccRCC-gained TAD boundaries, ccRCC-lost TAD boundaries, and stable TAD boundaries. ccRCC-specific SVs, determined by SURVIVOR, were those present in 786-O (or OS-RC-2) but absent in HEK293T. Dynamic chromosomal regions were identified using BEDTools. To investigate SV distribution, we compared SV densities in different chromosomal regions (A/B compartments, TAD domains, and TAD boundaries). SV density was calculated as the number of SVs divided by the length of each chromosomal region, while background density was the number of SVs divided by the whole-genome length. SV enrichment was assessed by comparing the regional SV proportion to the background using R's prop.test function. Genomic region associations were assessed through permutation testing via regioneR (v1.38.0)⁸⁴ with default parameters.

Identification of Repetitive Elements

RepeatMasker (version 3.7) was employed for the annotation of repetitive elements. Specifically, mutation sequence information corresponding to TGS and NGS - identified DELs and INSs was obtained from the vcf files and subsequently utilized as input for RepeatMasker (<https://github.com/Dfam-consortium/RepeatMasker/tree/master>).

Enrichment Analysis

The list of genes located within ccRCC-common TADs was analyzed in Enrichr (v23.6.8)⁸⁵ using the Jensen_DISEASES term (<https://diseases.jensenlab.org/Search>). The bed files for tumor-specific and normal-specific enhancers were uploaded to the GREAT website (v4.0.4) (<http://great.stanford.edu/public/html/>). GREAT analysis

was performed with hg38 as the reference genome and the whole genome as the background region, and the results were visualized using the R package ggplot2 (v3.4.2) (<https://github.com/tidyverse/ggplot2>).

Neoloop Analysis

In each sample, CNV profiles were estimated from Hi-C matrices using a generalized additive model. Large SV breakpoints were identified through nanopore long-read WGS analysis and Hi-C breakfinder²⁵. These breakpoints were then used as inputs for NeoLoopFinder⁸⁶, a computational framework employed to construct and identify local genome structures and neo-loops.

Cell Transfection

The short interference RNAs that targeted *SEMA5B* (si *SEMA5B*) and corresponding siRNA negative controls (siNC) were purchased from Gene Pharma (Shanghai, China). Transfections were performed using the Lipofectamine 3000 kit (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA) according to the manufacturer's protocol. The knockdown efficiency was assessed by reverse transcription-quantitative polymerase chain reaction (RT- qPCR) 48 h after transfection.

CRISPR-based Epigenetic Perturbation

The promoter- or enhancer-targeting sgRNAs and non-targeting sgRNAs with no genome recognition sites were cloned into pLV-hU6-sgRNA-hUbc-dCas9-KRAB-T2a-Puro (Addgene: 71236). The cells were infected with these vectors and then selected with puromycin (5 µg/ml) for at least 3 days before extracting RNA. All sgRNA sequences used are listed in Supplementary Table 36.

DNA FISH Combined with Immunofluorescence

OS-RC-2 and HEK293T cells were cultivated in a chamber slide (ibidi) and subjected to doxycycline (25 ng/ml) for 24 hours. After fixation with 4% paraformaldehyde for 10 min, the cells were rinsed three times with ice-cold PBS.

For DNA FISH, cells were permeabilized by sequential incubation in 70%, 80%, 90%, and 100% ethanol (5 min each) at room temperature. The probe hybridization mixture was formulated by combining FISH Hybridization Buffer (39 μ l) (#Bes1011, BersinBio) with FISH probes (1 μ l). Ten microliters of this mixture were deposited on each slide, and a coverslip was affixed on top, sealed with rubber cement. Upon solidification of the rubber cement, genomic DNA and probes were denatured at 75°C for 8 min. The slides were then incubated at 37°C overnight in darkness. Subsequently, the coverslip was removed, and the slides were sequentially washed with prewarmed 2 $\times$ SSC at 53°C for 5 min, prewarmed 0.1% NP-40/2 $\times$ SSC at 42°C for 5 min, and 2 $\times$ SSC at room temperature for 5 min. Post-washing, the slides were stained with 4',6-diamidino-2-phenylindole and examined on a confocal laser scanning microscope (Zeiss). The DNA FISH probes used in this study were custom-synthesized by BersinBio, with their sequences listed in Supplementary Table 35.

PCR across the Breakpoint Junctions

We designed specific primers on each side of the breakpoint and verified the breakpoint junction by PCR followed by sequencing to obtain nucleotide resolution sequence information. All primers used were listed in Supplementary Table 37.

RNA Isolation and Quantitative Real-time PCR (qPCR)

Total RNA was extracted from the indicated cells and frozen tissue specimens using TRIzol reagent (Invitrogen, USA) and reverse transcribed into cDNA using a Prime Script RT reagent kit (Takara, Japan) according to the manufacturers' instructions. Real-time PCR was performed using SYBR reagent (Applied Biosystems). The sequences of the primers used for quantitative real-time PCR are provided in Supplementary Table 38. For quantitative data analysis, gene expression was calculated using the $2^{-\Delta\Delta C_t}$ method.

Western Blot Assay

Proteins for western blotting analysis were extracted using RIPA buffer supplemented with protease inhibitors, and their concentrations were quantified with a BCA assay kit. Equal amounts of proteins were separated on 4-20%

SDS-PAGE gels and subsequently transferred onto nitrocellulose membranes. The membranes were then blocked with 5% skimmed milk for 1 h at room temperature. Following blocking, the membranes were incubated overnight at 4°C with the following primary antibodies: rabbit anti-human *SEMA5B* (#HPA066548, Sigma-Aldrich; Merck KGaA), anti- β -tubulin (#abs131994, Absin, Shanghai, China), and the following antibodies from Cell Signaling Technology (CST, Danvers, USA): anti-Jak1 (6G4) (#3344S), anti-phospho-Jak1 (Tyr1034/1035) (#3331S), anti-Stat3 (D3Z2G) (#12640S), anti-phospho-Stat3 (Tyr705) (D3A7) XP® (#9145S), anti-mTOR (#2972S), anti-phospho-mTOR (Ser2481) (#2974S), anti-Akt (pan) (C67E7) (#4691S), and anti-phospho-Akt (Thr308) (#9275S). β -Tubulin was used as an internal loading control.

After extensive washing, the membranes were incubated for 1 h at room temperature with horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG (Beijing Zhongshan Golden Bridge Biotechnology, Co. Ltd., Beijing, China) or HRP-conjugated goat anti-mouse IgG (Beijing Zhong-shan Golden Bridge Biotechnology, Co. Ltd., Beijing, China). Protein bands were visualized using the SuperSignal™ West Pico Plus Chemiluminescent Substrate kit (Thermo Scientific Inc., IL, USA) and detected with a Molecular Imager ChemiDoc XRS system (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Quantitative analysis of the band intensities for *SEMA5B* and signal-related proteins was performed using ImageJ software, with the values normalized to β -tubulin.

HE and Immunohistochemical Staining

All tissues were fixed in 10% neutral-buffered formalin. Lungs were inflation-fixed and all other tissues were drop-fixed for 24 h before paraffin embedding. Histological assessment was performed on hematoxylin and eosin (H&E)-stained sections. For immunohistochemical analysis, sections were dewaxed, rehydrated and subjected to microwave-based antigen retrieval with 20 min of boiling in citrate buffer (10 mM citrate and 0.05% Tween-20, pH 6.0) under pressure. Sections were probed with primary antibodies to *SEMA5B* (Sigma-Aldrich; Merck KGaA). Secondary antibodies (ImmPRESS HRP reagent kit, anti-rabbit, Beijing Zhongshan Golden Bridge Biotechnology, Co. Ltd., Beijing, China) were incubated at room temperature for 45 min and finally visualized with the peroxidase substrate DAB (Sigma-Aldrich; Merck KGaA) for the same amount of time (maximum 5 min) at room temperature. The stained tissue sections were counterstained with Meyer's hematoxylin.

Cell Proliferation and Viability Assays

Cell proliferation was quantified using the Cell Counting Kit-8 (CCK-8; Beyo time Institute of Biotechnology, Shanghai, China) following the manual. Briefly, 3×10^3 /well cells were seeded and transfected in a 96-well plate (Corning Incorporation). After indicated time, 10 μ l of CCK-8 solution was added into each well, and the cells were incubated for 4 h at 37 °C. The absorbance was measured using a microplate reader (Bio-Tek Instruments, VT, USA) at 450 nm.

Colony Formation Assay

A colony formation assay was also performed. Cells were added to 6-well plates (1×10^3 cells/well) after transfection for 2 weeks. Colonies were fixed with 100% methanol at room temperature for 20 min and stained with 0.1% crystal violet (Sigma-Aldrich; Merck KGaA) at 25°C for 30 min. The total number of visible colonies were imaged and counted using a light microscope. All experiments were repeated three times.

Trans-well Invasion Assay

Trans-well membranes with 8 μ m pore size (Corning Costar, MA, USA) coated with Matrigel (BD Biosciences, NJ, USA) were used for cancer cell invasion assay as previously described. After indicated transfection, cells were resuspended into fresh serum-free medium, and re-plated into the upper chamber. Fresh medium was directly added to the lower chamber. After additional 24 h incubation, the invasive cells penetrated to the lower surface, and were fixed with 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS) and stained with crystal violet (Sigma-Aldrich, MO, USA). The number of invasive cells was counted with a microscope (BX51, Olympus).

Human Samples

Tissue samples of ccRCC patients were from the Fourth Affiliated Hospital of School of Medicine, Zhejiang University. Informed consent was obtained from all study subjects, and sample collection was approved by the Ethics Committee of the Fourth Affiliated Hospital of School of Medicine, Zhejiang University (K20242224).

Animal Studies

All experiments using animals were performed in accordance with protocols approved by the local animal experimentation committee of the Zhejiang University (license no. ZJU20241020). Balb/C-nude mice were used at 8 weeks of age. Luciferase-expressing OS-RC-2 cells were injected subcutaneously on both flanks and monitored twice per week for 4 weeks by electronic caliper or bioluminescence imaging using Xenogen IVIS Imaging System 100 (STTARR facility, UHN). Investigators who measured and analyzed the samples were blind to the group information.

Risk Prediction Model Construction

ccRCC patients were randomly partitioned 1:1 into training and testing cohorts using the createDataPartition R function. Potential prognostic ccRCC-specific enhancer hijacking genes were initially identified via univariate Cox regression analysis. Subsequently, the R package glmnet (v=4.1.8) (<https://cran.r-project.org/web/packages/glmnet/index.html>) facilitated LASSO Cox regression to select the most significant genes. A multivariate stepwise Cox regression model was then constructed to further refine genes selection and optimize the prognostic model. Finally, a risk score equation was derived by combining the multivariate Cox regression coefficients with the expression levels of the selected enhancer hijacking genes, as follows:

$$\text{Risk} = \sum (\text{Gene expression} * \text{coef})$$

Survival Analysis and Patient Stratification

Cohorts (TCGA, internal training, and testing datasets) were stratified into high- and low-risk groups based on the median risk score. Kaplan-Meier survival curves, generated using the R package survminer (v=0.4.9) (<https://github.com/kassambara/survminer>), and log-rank tests ($p < 0.05$) assessed significant differences in OS and

PFS between risk groups. The R package timeROC (v=0.4) (<https://cran.r-project.org/web/packages/timeROC/index.html>) facilitated ROC analysis to evaluate the model's sensitivity and specificity for predicting OS, comparing its AUC against other clinical features. Associations between the risk score and clinical factors (age, gender, stage, T, M, N, grade) were then examined. To enhance prognostic precision and predictive capacity, a nomogram integrating the risk score with clinical parameters was developed, providing quantitative survival probability estimates.

Quantification and Statistical Analysis

The associations between Hi-C interactions, transcriptome dynamics, and SV were examined by Fisher's test and rank sum test. Gene expression data were presented as median with interquartile range (IQR), with statistical significance evaluated by Wilcoxon rank-sum test. Constituent ratios were analyzed using χ^2 test or Fisher's exact test. Survival curves and pan-cancer expression patterns were acquired from GEPIA2 using standard analytical parameters⁸⁷. All statistical analyses of experiments were performed using either Prism 9 (GraphPad Software), with appropriate statistical tests applied based on data distribution and experimental design.

SUPPLEMENTARY INFORMATION

Supplementary information is available on journal website or from the author.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary information. The raw sequence data for nanopore long-read WGS, NGS-WGS, Hi-C, and RNA-seq reported in this paper have been deposited in the Genome Sequence Archive⁸⁸ in National Genomics Data Center⁸⁹, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA012404) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human>. The H3K27ac ChIP-seq data were obtained from NCBI Gene Expression Omnibus (GEO) under accession GSE147279. Hi-C data were sourced from

GEO accessions GSE147279 and GSE216399. HEK293T nanopore long-read WGS data were retrieved from European Nucleotide Archive (ENA) project PRJEB44816. Transcriptional data and clinical information of ccRCC patients for model construction were downloaded from The Cancer Genome Atlas (TCGA) database.

Code availability

The source code supporting this study is not publicly available. However, it can be made available to bona fide researchers upon reasonable request to the corresponding author.

Acknowledgements

This work was supported by the Natural Science Foundation of China (no.32370749, no.82403511, 22075249, and 82372731), Key Project of Natural Science Foundation of Zhejiang Province (no. LZ22H160008), Natural Science Foundation of Zhejiang Province (no.LQ24H160005), Hangzhou Medical and Health Technology Project (no.B20241639), and the Natural Science Foundation of Zhejiang Province, Exploration Project/Exploration Public welfare/Medical health (no. LTGY24H160020). We gratefully acknowledge Zhen You, Zhimin Zhou, and Yangchen Ye for their valuable support in data analysis. We kindly appreciate the insightful contributions from Shengcai Chen, Qi Lv, and Pei Wang regarding the validation of breakpoint junctions. Additionally, we thank Yanfen Zhu and Liang Gong for their support for Hi-C library construction.

Author contributions

C.Z., G.Z., B.X., and J.J. conceived the project. Y.D., W.X., and Z.Y. conceptualized, designed, and wrote the manuscript. Y.D., L.R., Q.L., Z.C., and Z.X. assembled and processed the data. Y.D., H.W., and L.R. performed formal analyses. W.X., Z.Y., and H.W. developed the CRISPRi and knockdown samples for SEMA5B expression. W.X., Z.Y., H.W. and Y.Z. conducted in vitro experiments. Y.Z. and Q.L. assisted with cell culture. W.X. and Z.Y. carried out the in vivo experiments with help from Y.Z., F.W., H.N., and D.M., B.C. and W.S. provided intellectual advice. W.G., S.W., Z.L., J.S., Y.Q., and X.G. assisted in collecting clinical specimens.

Competing interests

The authors declare no competing interests.

Figure legends

ARTICLE IN PRESS

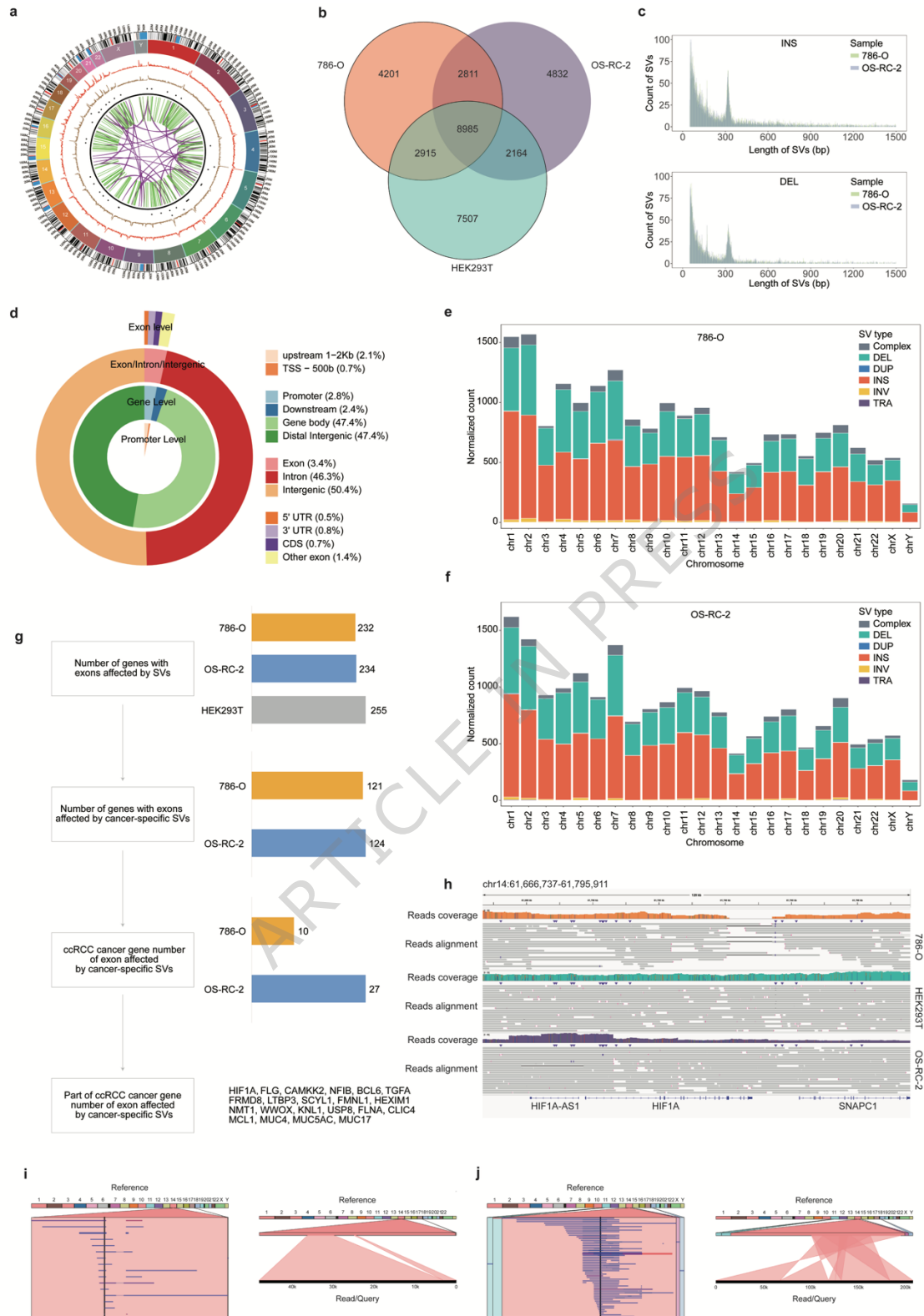


Fig. 1. SV landscapes in 786-O, OS-RC-2, and HEK293T cell lines.

- a) Circos plot illustrating the genome-wide distribution of 786-O high-confidence SVs of identified by Sniffles, cuteSV, and NanoSV, and integrated by SURVIVOR. The tracks from the outermost to innermost circles represent deletions, insertions, duplications, with the innermost green curve indicating inversions and the purple curve translocations.
- b) Venn plot depicting the intersection of structural variants in two ccRCC cell lines (786-O and OS-RC-2) and one normal renal cell line (HEK293T), with counts shown.
- c) Histograms displaying the length distribution of different types of ccRCC SVs.
- d) Distribution of SVs in various genomic regions in 786-O.
- e-f) Distribution of standardized total SV burden (dark blue: complex SV, green: deletion, light blue: duplication, orange: insertion, yellow: inversion, purple: translocation) across chromosomes in 786-O (e) and OS-RC-2 (f).
- g) Pipeline for identifying cancer-related genes with exons directly impacted by SVs in 786-O and OS-RC-2. ccRCC cancer genes were defined as those annotated by COSMIC or ranked among the top 30 ccRCC single nucleotide variation genes in TCGA.
- h) IGV image showing a heterozygous deletion on chromosome 3 covering *HIF1A* in 786-O and a heterozygous duplication on chromosome 3 covering *HIF1A* in OS-RC-2.
- i) Ribbon image showing a heterozygous deletion on chromosome 3 covering *HIF1A* in 786-O.
- j) Ribbon image showing a heterozygous duplication on chromosome 3 covering *HIF1A* in OS-RC-2.

ARTICLE IN PRESS

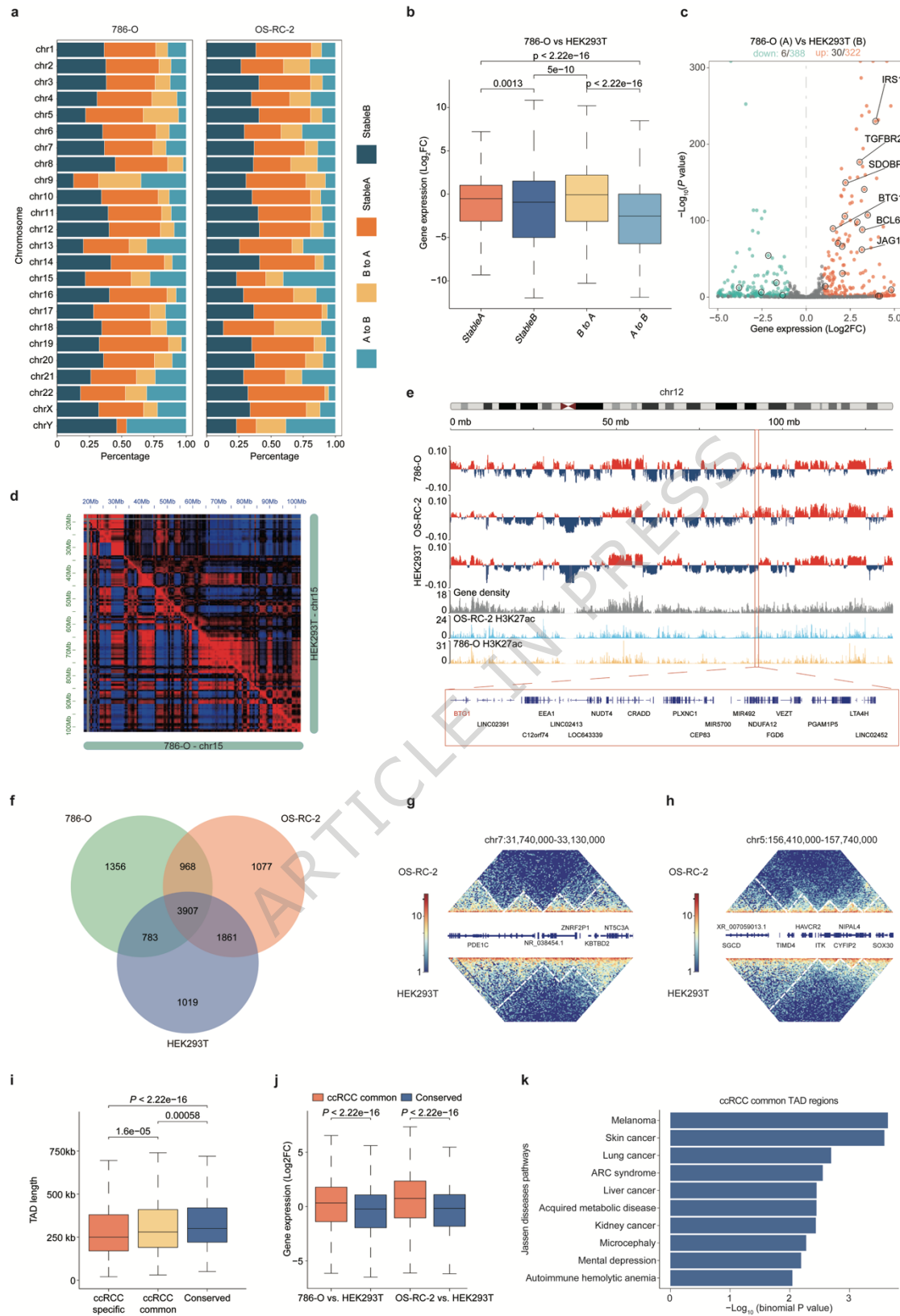


Fig. 2. 3D genome architectural reorganization correlates with gene expression alterations in ccRCC.

- a) Stacked bar chart showing frequencies of whole-chromosome A/B compartment shifts in 786-O and OS-RC-2 compared to HEK293T.
- b) Box plot showing gene expression distribution: the box spans the interquartile range (IQR), the center line marks the median, and the whiskers extend to $1.5 \times \text{IQR}$ (or to the maximum/minimum values if within range). Gene expression comparisons are presented as Log_2FC (786-O vs HEK293T), with p-values calculated using the Wilcoxon rank-sum test.
- c) Volcano plot highlighting differentially expressed genes (orange: significantly upregulated; green: significantly downregulated) and cancer-related genes in regions with B-to-A compartment shifts. Differentially expressed genes are defined by $|\text{Log}_2\text{FoldChange}| > 1$ and adjusted p-value < 0.05 , with examples of cancer-related genes circled in black.
- d) Pearson correlation heatmap of chr15 in OS-RC-2 and HEK293T.
- e) IGV images display an example of A/B compartment shifts on chromosome 12 in 786-O and OS-RC-2 versus HEK293T. Assignment of A (red) and B (blue) compartments is based on eigenvector values > 0 and < 0 . Gene density in the genome is shown as histograms, and enhancer activity is marked by H3K27ac ChIP-seq peaks, presented as histograms in 786-O (yellow track) and OS-RC-2 (blue track). The red box indicates a common B-to-A event covering the BTG1 gene in both cancer cell lines.
- f) Venn plot depicting the overlap of TADs derived from 10-kb resolution interaction matrices across the three cell types, with the number of TADs in each category shown.
- g-h) Examples of TAD alterations in regions of interest (G: chr7:31,740,000-33,130,000, H: chr5:156,410,000-157,740,000) in OS-RC-2 compared with HEK293T. Boxes in the interaction heatmaps delineate TADs, with involved genes displayed (RefSeqGene).
- i) Box plot showing the distribution of TAD length: the box spans the interquartile range (IQR), the center line marks the median, and the whiskers extend to $1.5 \times \text{IQR}$ (or to the maximum/minimum values if within range). P-values were calculated using the Wilcoxon rank-sum test.
- j) Box plot showing gene expression distribution: the box spans the interquartile range (IQR), the center line marks the median, and the whiskers extend to $1.5 \times \text{IQR}$ (or to the maximum/minimum values if within range). Gene

expression comparisons are presented as Log₂FoldChange (ccRCC-common TAD region vs conserved TAD region), with *P*-values calculated using the Wilcoxon rank-sum test.

k) Jensen- diseases pathway enrichment of differentially expressed genes located in ccRCC-common TADs. *P*-values were obtained by Fisher's exact test using EnrichR.

ARTICLE IN PRESS

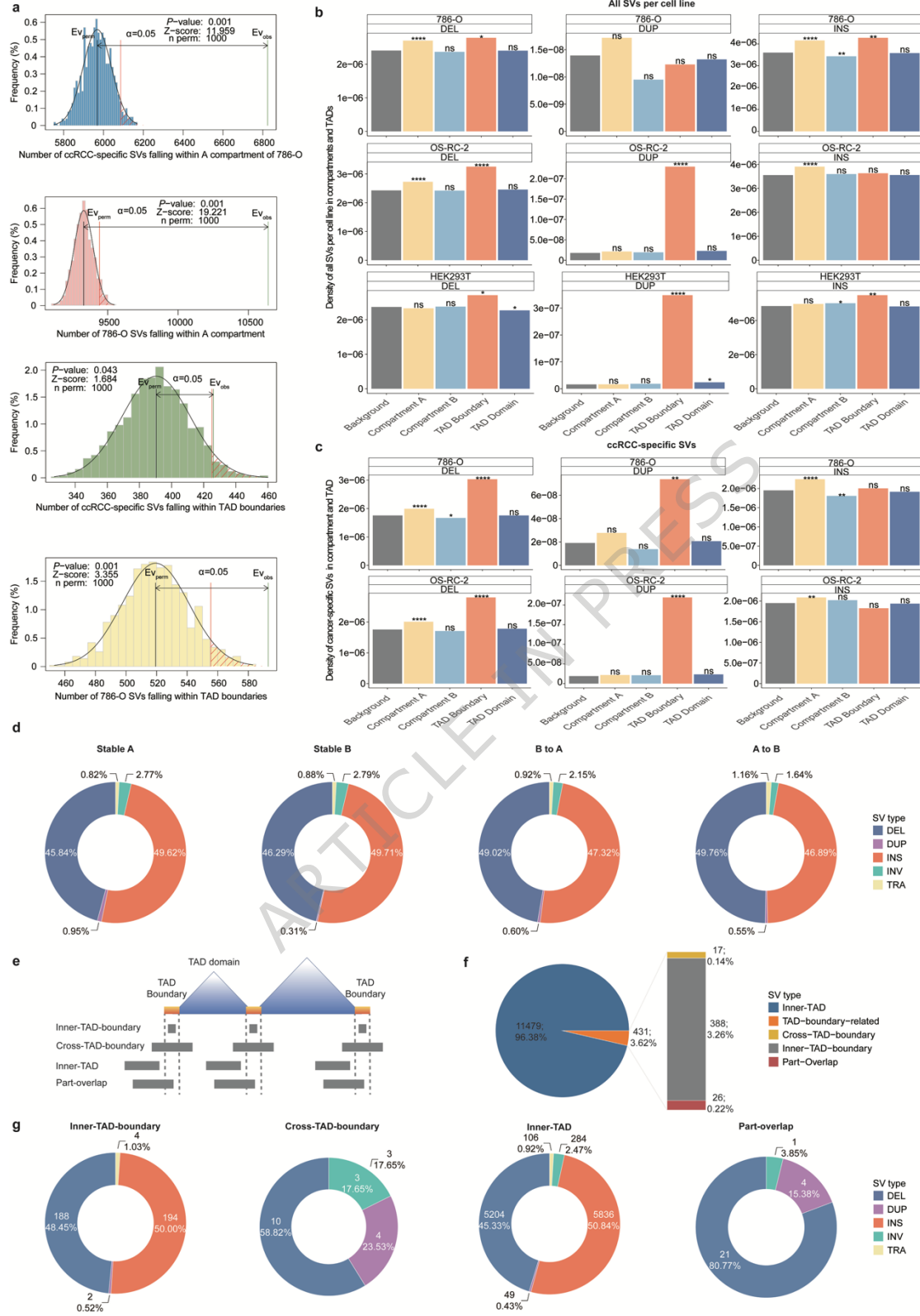


Fig. 3. Distributions of structural variants across 3D genome architectures.

- a) Permutation testing was employed to assess the spatial distribution patterns of both pan-cellular and ccRCC specific structural variants across 3D chromatin organizations. The Expected distribution was derived from 1000 iterations of random shuffling of chromosomal regions.
- b) Density of SVs (insertions, deletions, and duplications) across A/B compartments and TADs. SV density, calculated as the number of SVs normalized by the length of corresponding chromosomal regions, is represented as follows: gray bars indicate the background SV density across entire chromosomes; yellow bars show SV density in A compartments; light blue bars represent SV density in B compartments; orange bars depict SV density at TAD boundaries; and dark blue bars illustrate SV density within TAD domains. Enrichment analyses were conducted with proportionality test using R's proportion test, comparing the proportion of SVs in each region of interest to the proportion of that region's length in the whole genome. Significance levels are denoted as follows: **** $p \leq 0.0001$, *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$.
- c) Density of ccRCC-specific SVs (insertions, deletions, and duplications) across A/B compartments and TADs. ccRCC-specific SVs are defined as those occurring in 786-O or OS-RC-2 but absent in HEK293T. SV density was calculated using the same method as in the legend of (B). Enrichment analyses for these ccRCC-specific SVs were also performed using R's proportion test, following the same approach as described in (B).
- d) Proportions of ccRCC specific SV subtypes across distinct A/B compartment switching regions in 786-O cell.
- e) Schematic of SV classification based on the positional relationship between ccRCC specific SV breakpoints and TAD boundaries.
- f) Number and proportion of ccRCC specific SV types in 786-O according to (E).
- g) Number and proportion of ccRCC specific SV types in 786-O according to (E).

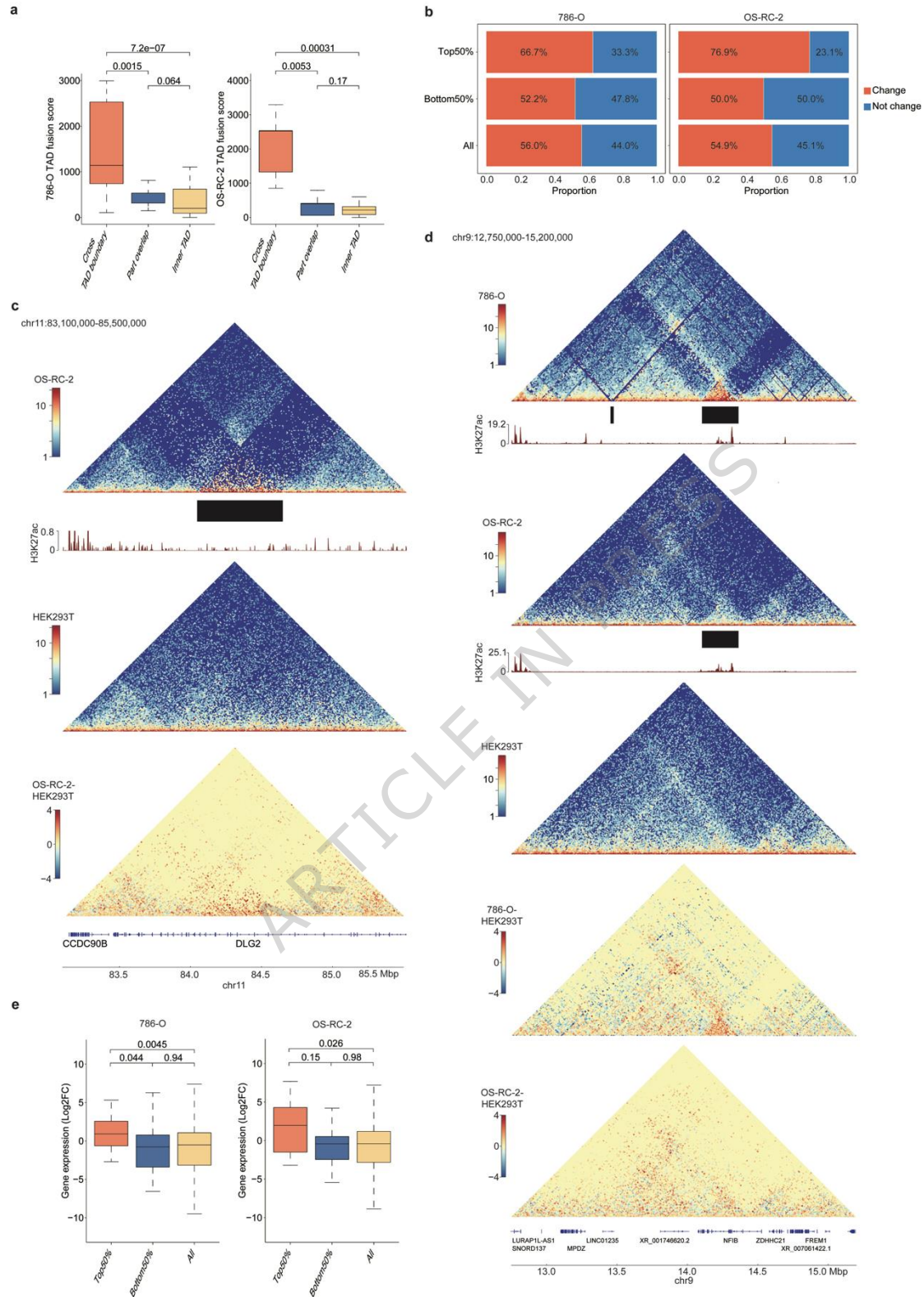


Fig. 4. Oncogenic structural variations drive transcriptional reprogramming through TAD remodeling in ccRCC.

- a) Comparative analysis of TAD fusion scores across distinct deletion classes. Boxplots depict distributions stratified by TAD deletion classes (boundary-spanning vs. internal deletions), with box ranges representing interquartile distances (IQR), central lines indicating medians, and whiskers extending to $1.5 \times \text{IQR}$ or extreme values within range. Statistical significance was determined by Wilcoxon rank-sum test.
- b) Differential gene expression burden in SV hotspots. Stacked bars quantify proportions of differentially expressed genes (DEGs) in genomic regions stratified by TAD fusion score percentiles (top/bottom 50th percentile) versus genome-wide baselines across 786-O and OS-RC-2 cell lines.
- c) Examples of the impact of inner TAD deletion frequency on the chromatin folding domain in OS-RC-2. Triangle heatmaps represent chromatin contact frequency, with the top showing OS-RC-2, middle showing HEK293T, and bottom showing the subtractive results. Histogram representing roadmap epigenome enhancer activity, marked by H3K27ac, in OS-RC-2 (red).
- d) Examples of the impact of Cross TAD boundary deletion frequency on the chromatin folding domain in ccRCC. Triangle heatmaps represent chromatin contact frequency for 786-O (top), OS-RC-2 (second), HEK293T controls (third), with differential contact maps (fourth: 786-O vs control; fifth: OS-RC-2 vs control). Histograms show roadmap epigenome enhancer activity marked by H3K27ac in 786-O and OS-RC-2 (red).
- e) Box plot of gene expression levels in regions with TAD fusion scores in the top 50%, bottom 50%, and genome-wide levels. The box spans the interquartile range (IQR), with the center line indicating the median and whiskers extending to $1.5 \times \text{IQR}$ (or the maximum/minimum values if within range). P-values were calculated using the Wilcoxon rank-sum test.

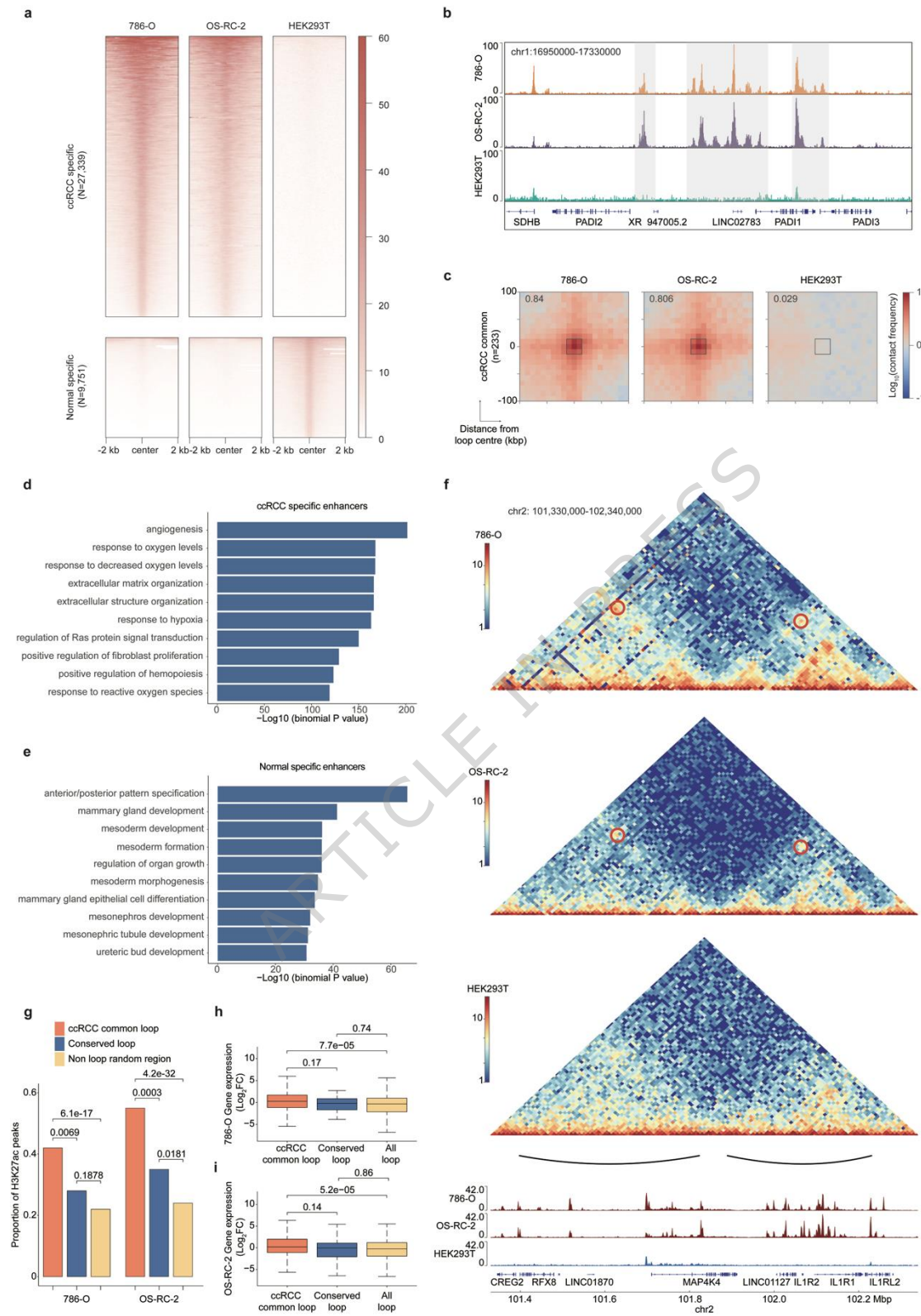


Fig. 5. Enhancer and chromatin interaction remodeling drives transcriptomic alterations during oncogenesis.

A) Heatmap of H3K27ac ChIP-seq signals showing the intensity of ccRCC-specific and normal-specific enhancers across the genome.

B) IGV screenshot showing an example region around chr1:16,950,000-17,330,000 corresponding to the PADI1 gene, highlighting ccRCC-specific enhancers.

C) Aggregated peak analysis for ccRCC-common chromatin interactions in ccRCC (786-O and OS-RC-2) and normal HEK293T cell lines (n=233).

D) GREAT pathway enrichment analysis based on the GO dataset for ccRCC-specific enhancers.

E) GREAT pathway enrichment analysis based on the GO dataset for normal-specific enhancers.

F) ccRCC-common chromatin interactions in the MAP4K4 genomic region. Red circles on the chromatin interaction heatmap and black arcs below highlight the positions of ccRCC-specific interactions. The top two panels of the chromatin interaction heatmap are from 786-O and OS-RC-2 tumor cells, and the bottom panel shows HEK293T normal cells. The tracks below show H3K27Ac ChIP-seq profiles from renal cancer cell lines (red) and normal cell lines (blue).

G) Histograms showing the proportion of H3K27ac peaks in different types of loops in the genomes of 786-O and OS-RC-2. P-values were calculated using R's proportionality test.

H-I) Box plots illustrating the expression levels of genes overlapping with double anchors of ccRCC-common loops, conserved loops, and all identified loops in 786-O (H) and OS-RC-2 (I). The box denotes the interquartile range (IQR), with the center line representing the median and whiskers extending to 1.5×IQR (or the maximum/minimum values if within range). P-values were determined using the Wilcoxon rank-sum test.

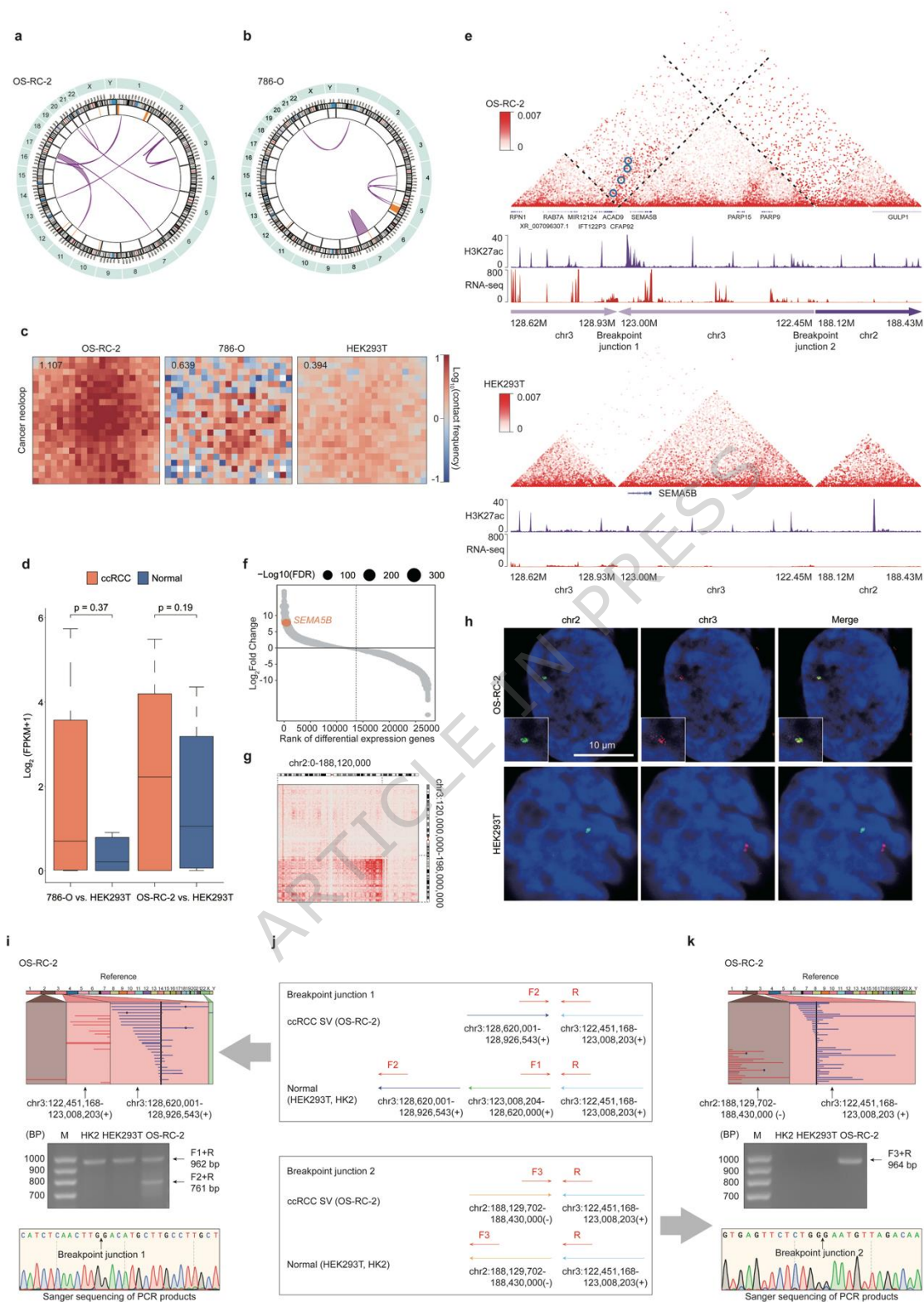


Fig. 6. Identification and validation of enhancer hijacking events driving oncogenic dysregulation in ccRCC.

- a-b) Circos plots illustrating enhancer hijacking and ectopic chromatin interactions mediated by SVs detected by TGS-WGS in OS-RC-2 (a) and 786-O (b) cancer cells. The purple curve represents neoloop spanning different chromosomes, while the orange curve represents neoloop within the same chromosome.
- c) Aggregated peak analysis of ectopic chromatin interactions (neoloops) in ccRCC (786-O and OS-RC-2) and normal HEK293T cells.
- d) Box plot of gene expression levels for genes at neoloop anchors in tumor and normal cells. The box spans the interquartile range (IQR), with the center line indicating the median and whiskers extending to 1.5×IQR. P-values were calculated using the Wilcoxon rank-sum test.
- e) Top panel: The reconstructed Hi-C map and genomic ChIP-seq, RNA-seq tracks for the enhancer hijacking event involving *SEMA5B* dysregulation in OS-RC-2 cells, showing intrachromosomal inversion on chr3 and interchromosomal translocation between chr2 and chr3. Blue circles indicate neoloops' genomic position and interaction intensity. Bottom panel: The same regions in normal kidney cells.
- f) Gene ranking dotplot showing DEGs between OS-RC-2 and HEK293T cells. Orange circles highlight *SEMA5B* dysregulation from enhancer hijacking.
- g) KR balanced Hi-C map showing translocation-related chromatin interactions in renal cancer at chr2: 0-188,120,000 and chr3: 120,000,000-198,000,000.
- h) DNA-FISH confirms co-localization of chr3:188,120,000 and chr2:122,450,000 DNA fragments in OS-RC-2 cells, absent in HEK293T cells.
- i) Validation of breakpoint junction 1 of inversion event involving enhancer hijacking. Top panel: Ribbon plot of nanopore long-read WGS reads capturing breakpoint junction 1 on chr3. Middle panel: PCR validation of breakpoint junction 1, amplifying a 761bp inversion sequence only in renal cancer cells. Bottom panel: Sanger sequencing validation of the 761bp PCR product.
- j) Schematic of PCR validation strategy for complex SV breakpoint junctions.
- k) Validation of breakpoint junction 2 of translocation event involving enhancer hijacking. Top panel: Ribbon plot of nanopore long-read WGS reads capturing breakpoint junction 2 between chr2 and chr3. Middle panel: PCR

validation of breakpoint junction 2, amplifying a 964bp translocation sequence only in renal cancer cells. Bottom panel: Sanger sequencing validation of the 964bp PCR product.

ARTICLE IN PRESS

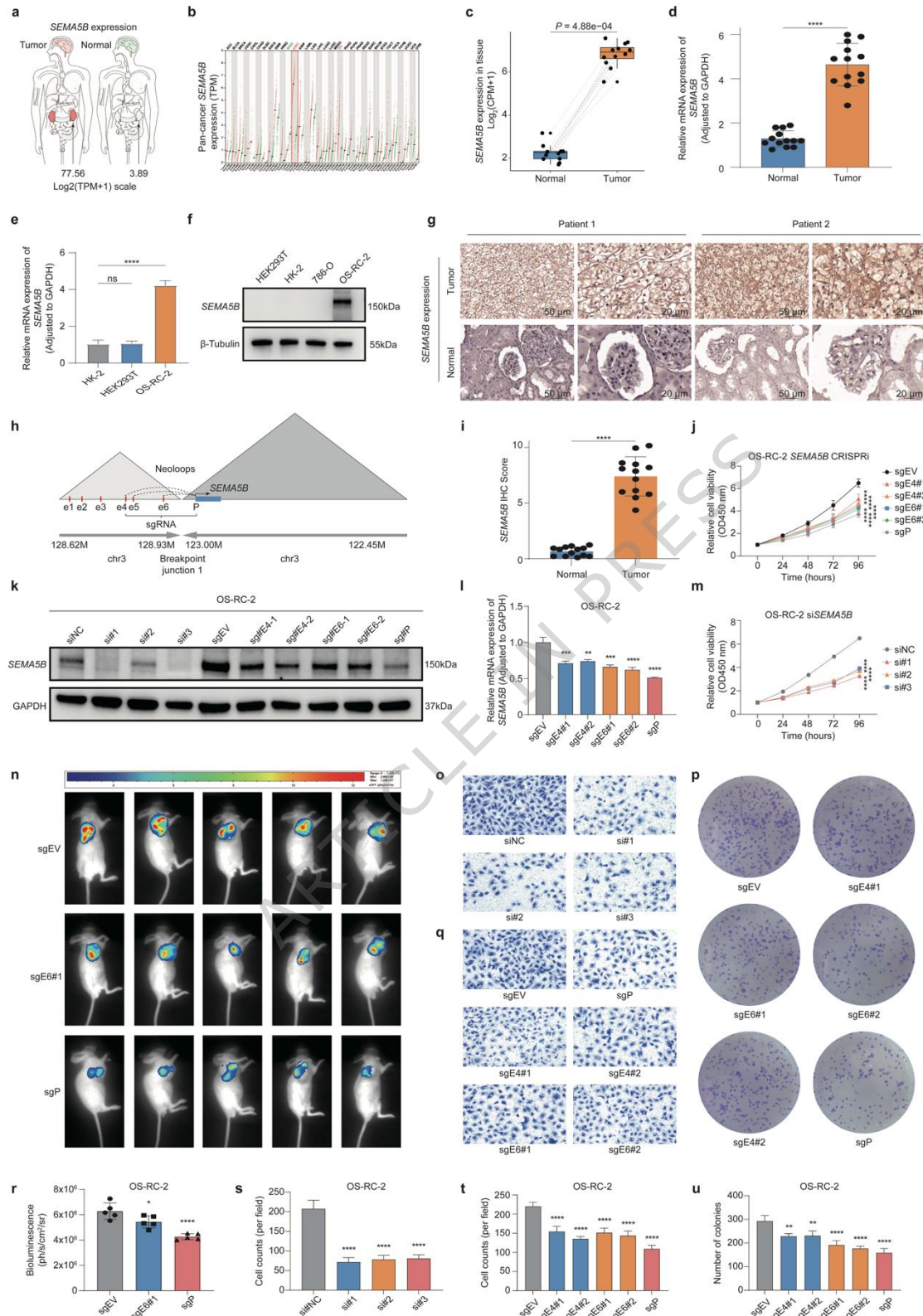


Fig. 7. Enhancer Hijacking Mediates *SEMA5B* Overexpression in ccRCC and Promotes Renal Carcinogenesis

- a) Bodymap analysis from GEPIA revealed significantly elevated *SEMA5B* expression in ccRCC compared to normal tissues.
- b) Pan-cancer analysis of *SEMA5B* expression across 33 tumor types in the GEPIA database demonstrated markedly higher expression in ccRCC.
- c) Box plot and scatter plot of RNA-seq data showing *SEMA5B* expression levels in paired tumor and normal tissues (N = 13). The box represents the interquartile range (IQR), with the median indicated by the center line and whiskers extending to 1.5×IQR. P-values were calculated using the Wilcoxon rank-sum test. Each scatter point represents an individual sample, and lines connecting tumor and normal samples indicate paired renal tissues from the same patient.
- d) RT-qPCR quantification of *SEMA5B* expression in paired tumor and normal tissues (N = 13). P-values were derived from unpaired two-sided Student's t-test. **** $p \leq 0.0001$.
- e) RT-qPCR analysis of *SEMA5B* expression in HK2, HEK293T, 786-O, and OS-RC-2 cell lines. N = 3 biological replicates. P-values were derived from unpaired two-sided Student's t-test. **** $p \leq 0.0001$.
- f) Western blot validation of *SEMA5B* protein levels in HK2, HEK293T, 786-O, and OS-RC-2 cell lines.
- g) Representative immunohistochemical (IHC) staining images of *SEMA5B* in paired tumor and normal tissues. Scale bars: 50 μm and 20 μm .
- h) Schematic illustration of *SEMA5B* neoloops and CRISPRi design targeting enhancers e4-e6 or the promoter region in OS-RC-2 cells.
- i) Quantitative IHC score of *SEMA5B* expression in paired tumor and normal tissues (N=13). P-values were derived from unpaired two-sided Student's t-test. **** $p \leq 0.0001$.
- j and m) Short-term cell proliferation assays demonstrating that *SEMA5B* promotes ccRCC cell proliferation. **** $p \leq 0.0001$. siNC indicates a negative control transfected with a non-specific sequence, and sgEV represents non-targeting sgRNAs with no genome recognition sites.
- k) Western blot analysis of *SEMA5B* expression following CRISPRi targeting of enhancers e4 and e6 or the promoter region of *SEMA5B*, or siRNA-mediated *SEMA5B* knockdown in ccRCC cells.

l) RT-qPCR quantification of *SEMA5B* expression after CRISPRi targeting of enhancers e4 and e6 or the promoter region in OS-RC-2 cells. N = 3 biological replicates. P-values were derived from one-way ANOVA. **** $p \leq 0.0001$, *** $p \leq 0.001$, ** $p \leq 0.01$.

n and r) bioluminescence measurement on the indicated days of mice with different treatment OS-RC-2 cell engraftment. N = 5 biological replicates. P-values were derived from one-way ANOVA.

o and s) Transwell migration assays demonstrated that knockdown of *SEMA5B* significantly reduced cell invasion compared to normal control cells. P-values were derived from one-way ANOVA. **** $p \leq 0.0001$, ** $p \leq 0.01$.

q and t) Transwell migration assays demonstrated that CRISPRi of *SEMA5B* significantly reduced cell invasion compared to empty vector control. N = 3 biological replicates. P-values were derived from one-way ANOVA. **** $p \leq 0.0001$, ** $p \leq 0.01$.

p and u) Long-term colony formation assays showing that *SEMA5B* enhances ccRCC cell proliferation. N = 3 biological replicates. P-values were derived from one-way ANOVA. **** $p \leq 0.0001$, ** $p \leq 0.01$.

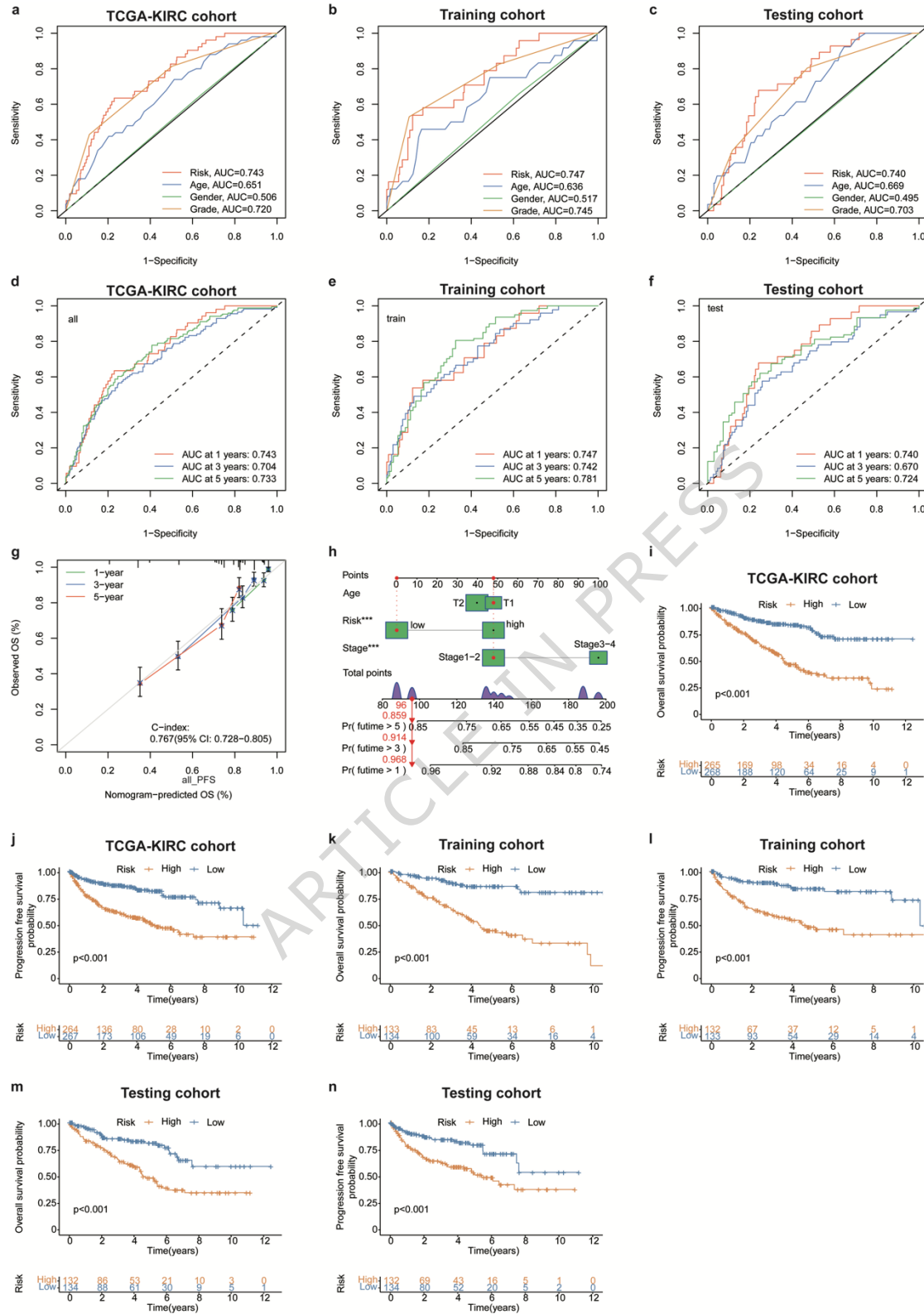


Fig. 8. Construction of a ccRCC Prognostic Risk Model Based on ccRCC-specific Enhancer-Hijacking Events Using Machine Learning

a-c) Receiver Operating Characteristic (ROC) curves demonstrating the robust prognostic predictive performance of the risk model in the TCGA-KIRC cohort (a), training cohort (b), and testing cohort (c).

d-f) Time-dependent Receiver Operating Characteristic (ROC) curves evaluating the risk model's performance at 1-, 3-, and 5-year intervals in the TCGA-KIRC cohort (d), training cohort (e), and testing cohort (f).

g) Calibration curve analysis validating the stability and reliability of model predictions.

h) Nomogram for predicting 1-, 3-, and 5-year overall survival (OS) in ccRCC patients within the TCGA-KIRC cohort.

i-j) Kaplan-Meier survival curves depicting significant divergence in overall survival (OS) (i) and progression-free survival (PFS) (j) between high- and low-risk groups in the TCGA-KIRC cohort.

k-l) Kaplan-Meier survival analysis showing distinct overall survival (OS) (k) and progression-free survival (PFS) (l) outcomes for high- versus low-risk groups in the training cohort.

m-n) Kaplan-Meier survival curves confirming differential overall survival (OS) (m) and progression-free survival (PFS) (n) between risk strata in the testing cohort.

References

- 1 Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249, doi:10.3322/caac.21660 (2021).
- 2 Hsieh, J. J. *et al.* Renal cell carcinoma. *Nat Rev Dis Primers* **3**, 17009, doi:10.1038/nrdp.2017.9 (2017).
- 3 Walton, J. *et al.* PBRM1, SETD2 and BAP1 - the trinity of 3p in clear cell renal cell carcinoma. *Nat Rev Urol* **20**, 96-115, doi:10.1038/s41585-022-00659-1 (2023).
- 4 Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* **14**, 125-138, doi:10.1038/nrg3373 (2013).
- 5 Pan-cancer analysis of whole genomes. *Nature* **578**, 82-93, doi:10.1038/s41586-020-1969-6 (2020).
- 6 Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210-216, doi:10.1038/s41586-019-1689-y (2019).
- 7 Beroukhi, R., Zhang, X. & Meyerson, M. Copy number alterations unmasked as enhancer hijackers. *Nat Genet* **49**, 5-6, doi:10.1038/ng.3754 (2016).
- 8 Method of the Year 2022: long-read sequencing. *Nat Methods* **20**, 1, doi:10.1038/s41592-022-01759-x (2023).
- 9 van Dijk, E. L. *et al.* Genomics in the long-read sequencing era. *Trends Genet*, doi:10.1016/j.tig.2023.04.006 (2023).
- 10 Warburton, P. E. & Sebra, R. P. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annu Rev Genomics Hum Genet* **24**, 109-132, doi:10.1146/annurev-genom-101722-103045 (2023).
- 11 Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat Rev Genet* **21**, 597-614, doi:10.1038/s41576-020-0236-x (2020).
- 12 Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat Rev Genet* **19**, 453-467, doi:10.1038/s41576-018-0007-0 (2018).

- 13 Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661-678, doi:10.1038/nrg.2016.112 (2016).
- 14 Taberlay, P. C. *et al.* Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res* **26**, 719-731, doi:10.1101/gr.201517.115 (2016).
- 15 Wu, P. *et al.* 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nat Commun* **8**, 1937, doi:10.1038/s41467-017-01793-w (2017).
- 16 Kim, Y., Yang, H. & Lee, D. Cell line-specific features of 3D chromatin organization in hepatocellular carcinoma. *Genomics Inform* **21**, e19, doi:10.5808/gi.23015 (2023).
- 17 Sidiropoulos, N. *et al.* Somatic structural variant formation is guided by and influences genome architecture. *Genome Res* **32**, 643-655, doi:10.1101/gr.275790.121 (2022).
- 18 Akdemir, K. C. *et al.* Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* **52**, 294-305, doi:10.1038/s41588-019-0564-y (2020).
- 19 Akdemir, K. C. *et al.* Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nat Genet* **52**, 1178-1188, doi:10.1038/s41588-020-0708-0 (2020).
- 20 Botten, G. *et al.* Structural Variation Cooperates with Permissive Chromatin to Control Enhancer Hijacking-Mediated Oncogenic Transcription. *Blood*, doi:10.1182/blood.2022017555 (2023).
- 21 Wang, J. *et al.* Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. *Sci Adv* **7**, doi:10.1126/sciadv.abg4126 (2021).
- 22 Xie, T. *et al.* Pervasive structural heterogeneity rewires glioblastoma chromosomes to sustain patient-specific transcriptional programs. *Nat Commun* **15**, 3905, doi:10.1038/s41467-024-48053-2 (2024).
- 23 Xu, L. *et al.* Long-read sequencing identifies novel structural variations in colorectal cancer. *PLoS Genet* **19**, e1010514, doi:10.1371/journal.pgen.1010514 (2023).
- 24 Aganezov, S. *et al.* Comprehensive analysis of structural variants in breast cancer genomes using single-molecule sequencing. *Genome Res* **30**, 1258-1273, doi:10.1101/gr.260497.119 (2020).
- 25 Dixon, J. R. *et al.* Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet* **50**, 1388-1398, doi:10.1038/s41588-018-0195-8 (2018).
- 26 Sedlazeck, F. J. *et al.* Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**, 461-468, doi:10.1038/s41592-018-0001-7 (2018).
- 27 Hoefflin, R. *et al.* HIF-1 α and HIF-2 α differently regulate tumour development and inflammation of clear cell renal cell carcinoma in mice. *Nat Commun* **11**, 4111, doi:10.1038/s41467-020-17873-3 (2020).
- 28 Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173**, 581-594 e512, doi:10.1016/j.cell.2018.03.057 (2018).
- 29 Kroeger, N. *et al.* Deletions of chromosomes 3p and 14q molecularly subclassify clear cell renal cell carcinoma. *Cancer* **119**, 1547-1554, doi:10.1002/cncr.27947 (2013).
- 30 Jonasch, E., Walker, C. L. & Rathmell, W. K. Clear cell renal cell carcinoma ontogeny and mechanisms of lethality. *Nat Rev Nephrol* **17**, 245-261, doi:10.1038/s41581-020-00359-2 (2021).
- 31 Fujimoto, A. *et al.* Whole-genome sequencing with long reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer. *Genome Med* **13**, 65, doi:10.1186/s13073-021-00883-1 (2021).
- 32 Okhovat, M. *et al.* TAD evolutionary and functional characterization reveals diversity in mammalian TAD boundary properties and function. *Nat Commun* **14**, 8111, doi:10.1038/s41467-023-43841-8 (2023).
- 33 Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269, doi:10.1038/nature19800 (2016).
- 34 Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908-921, doi:10.1016/j.cell.2012.02.002 (2012).
- 35 Dubois, F., Sidiropoulos, N., Weischenfeldt, J. & Beroukhim, R. Structural variations in cancer and the 3D genome. *Nat Rev Cancer* **22**, 533-546, doi:10.1038/s41568-022-00488-9 (2022).
- 36 Sarni, D. *et al.* 3D genome organization contributes to genome instability at fragile sites. *Nat Commun* **11**, 3613, doi:10.1038/s41467-020-17448-2 (2020).
- 37 Grzeda, K. R. *et al.* Functional chromatin features are associated with structural mutations in cancer. *BMC Genomics* **15**, 1013, doi:10.1186/1471-2164-15-1013 (2014).
- 38 Currall, B. B., Chiang, C., Talkowski, M. E. & Morton, C. C. Mechanisms for Structural Variation in the Human Genome. *Curr Genet Med Rep* **1**, 81-90, doi:10.1007/s40142-013-0012-8 (2013).
- 39 Qian, C. N., Huang, D., Wondergem, B. & Teh, B. T. Complexity of tumor vasculature in clear cell renal cell carcinoma. *Cancer* **115**, 2282-2289, doi:10.1002/cncr.24238 (2009).

- 40 Liao, C., Hu, L. & Zhang, Q. Von Hippel-Lindau protein signalling in clear cell renal cell carcinoma. *Nat Rev Urol* **21**, 662-675, doi:10.1038/s41585-024-00876-w (2024).
- 41 Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* **49**, 65-74, doi:10.1038/ng.3722 (2017).
- 42 Liu, T. *et al.* Enhancer Coamplification and Hijacking Promote Oncogene Expression in Liposarcoma. *Cancer Res* **83**, 1517-1530, doi:10.1158/0008-5472.Can-22-1858 (2023).
- 43 Mastrantonio, R., You, H. & Tamagnone, L. Semaphorins as emerging clinical biomarkers and therapeutic targets in cancer. *Theranostics* **11**, 3262-3277, doi:10.7150/thno.54023 (2021).
- 44 Hirota, E. *et al.* Genome-wide gene expression profiles of clear cell renal cell carcinoma: identification of molecular targets for treatment of renal cell carcinoma. *Int J Oncol* **29**, 799-827 (2006).
- 45 Dizman, N., Philip, E. J. & Pal, S. K. Genomic profiling in renal cell carcinoma. *Nat Rev Nephrol* **16**, 435-451, doi:10.1038/s41581-020-0301-x (2020).
- 46 Xie, Y. *et al.* SETD2 loss perturbs the kidney cancer epigenetic landscape to promote metastasis and engenders actionable dependencies on histone chaperone complexes. *Nat Cancer* **3**, 188-202, doi:10.1038/s43018-021-00316-3 (2022).
- 47 Haffner, M. C., De Marzo, A. M., Meeker, A. K., Nelson, W. G. & Yegnasubramanian, S. Transcription-induced DNA double strand breaks: both oncogenic force and potential therapeutic target? *Clin Cancer Res* **17**, 3858-3864, doi:10.1158/1078-0432.Ccr-10-2044 (2011).
- 48 Gerhauser, C. *et al.* Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell* **34**, 996-1011.e1018, doi:10.1016/j.ccell.2018.10.016 (2018).
- 49 Lupski, J. R. *et al.* Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. *Nat Genet* **1**, 29-33, doi:10.1038/ng0492-29 (1992).
- 50 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, doi:10.1126/science.1058040 (2001).
- 51 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 52 Hecht, F. & Kaiser-McCaw, B. Position effect in 8;14 translocation in Burkitt's lymphoma. *N Engl J Med* **304**, 174-175, doi:10.1056/nejm198101153040317 (1981).
- 53 Zhou, H., Yang, Y. H. & Basile, J. R. Characterization of the Effects of Semaphorin 4D Signaling on Angiogenesis. *Methods Mol Biol* **1493**, 429-441, doi:10.1007/978-1-4939-6448-2_31 (2017).
- 54 Conrotto, P. *et al.* Sema4D induces angiogenesis through Met recruitment by Plexin B1. *Blood* **105**, 4321-4329, doi:10.1182/blood-2004-07-2885 (2005).
- 55 Matsunaga, Y., Bashiruddin, N. K., Kitago, Y., Takagi, J. & Suga, H. Allosteric Inhibition of a Semaphorin 4D Receptor Plexin B1 by a High-Affinity Macrocyclic Peptide. *Cell Chem Biol* **23**, 1341-1350, doi:10.1016/j.chembiol.2016.09.015 (2016).
- 56 Musch, A., Peille, A.-L., Hasani, H. A., Fiebig, H.-H. & Vuaroqueaux, V. Abstract LB107: Identification of a promising renal cell carcinoma target using a unique in silico approach. *Cancer Research* **81**, LB107-LB107, doi:10.1158/1538-7445.Am2021-lb107 (2021).
- 57 De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, doi:10.1093/bioinformatics/btad311 (2023).
- 58 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).
- 59 Cretu Stancu, M. *et al.* Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun* **8**, 1326, doi:10.1038/s41467-017-01343-4 (2017).
- 60 Jiang, T. *et al.* Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol* **21**, 189, doi:10.1186/s13059-020-02107-y (2020).
- 61 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 62 Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333-i339, doi:10.1093/bioinformatics/bts378 (2012).
- 63 Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* **15**, R84, doi:10.1186/gb-2014-15-6-r84 (2014).
- 64 Chen, X. *et al.* Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220-1222, doi:10.1093/bioinformatics/btv710 (2016).
- 65 Jeffares, D. C. *et al.* Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* **8**, 14061, doi:10.1038/ncomms14061 (2017).

- 66 Geoffroy, V. *et al.* AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* **34**, 3572-3574, doi:10.1093/bioinformatics/bty304 (2018).
- 67 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
- 68 Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* **45**, D777-d783, doi:10.1093/nar/gkw1121 (2017).
- 69 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26, doi:10.1038/nbt.1754 (2011).
- 70 Nattestad, M., Aboukhalil, R., Chin, C. S. & Schatz, M. C. Ribbon: intuitive visualization for complex genomic variation. *Bioinformatics* **37**, 413-415, doi:10.1093/bioinformatics/btaa680 (2021).
- 71 Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421-432, doi:10.1093/bioinformatics/bty648 (2019).
- 72 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).
- 73 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
- 74 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).
- 75 Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189, doi:10.1038/s41467-017-02525-w (2018).
- 76 Roayaei Ardakany, A., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol* **21**, 256, doi:10.1186/s13059-020-02167-0 (2020).
- 77 Huynh, L. & Hormozdiari, F. TAD fusion score: discovery and ranking the contribution of deletions to genome structure. *Genome Biol* **20**, 60, doi:10.1186/s13059-019-1666-7 (2019).
- 78 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 79 Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943-2945, doi:10.1093/bioinformatics/btac166 (2022).
- 80 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 81 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
- 82 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 83 Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).
- 84 Gel, B. *et al.* regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289-291, doi:10.1093/bioinformatics/btv562 (2016).
- 85 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
- 86 Wang, X. *et al.* Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. *Nat Methods* **18**, 661-668, doi:10.1038/s41592-021-01164-w (2021).
- 87 Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* **47**, W556-w560, doi:10.1093/nar/gkz430 (2019).
- 88 Chen, T. *et al.* The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Genomics Proteomics Bioinformatics* **19**, 578-583, doi:10.1016/j.gpb.2021.08.001 (2021).
- 89 Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2022. *Nucleic Acids Res* **50**, D27-d38, doi:10.1093/nar/gkab951 (2022).