Chromosome-Specific Aneuploidy Engineering via dCas9-Induced Centromeric Chromatin Relaxation

Hugang Feng[1]*, Daqi Deng[1,18], Rashmi Dahiya[2,18], Libin Wang[3,18], Jingkun Zeng[4], Benjy Jek Yang Tan[1], Fiona Byrne[1], Scott T. C. Shepherd[1,5,6], Jiahao Wang[1,7], Sarah C. Johnson[8], Alison Harrod[9], Karen A. Lane[9], Annika Fendler[1,10], Anne-Laure Cattin[1], Zayd Tippu[1,5,6], Meilun Nie[3], Yiming Zhao[3], Ruijia Wang[3], Wei Ai[3], Omar Bouricha[1], Taja Barber[1], Yuliia Dovga[1], Yihan Xu[1], Yimeng Xu[11], Liani G. Devito[12], Lyn Healy[12], Zhen Sun[13], Weiming Shen[13], Hongye Wang[14], Eugénie S. Lim[15], Samuel M. O'Toole[15,16], Scott Akker[15], William M. Drake[15], Haojie Jin[13], Jessica A. Downs[9], Sarah E. McClelland[8], John F. X. Diffley[4], Peter Ly[2,17]& Samra Turajlic[1,5,6]*

Affiliations:

1. Cancer Dynamics Laboratory, The Francis Crick Institute, London, UK
2. Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA
3. Department of Cell Biology, School of Basic Medical Sciences, Peking University Stem Cell Research Center, Peking University Health Science Center, Peking University, Beijing, China
4. Chromosome Replication Laboratory, The Francis Crick Institute, London, UK
5. Skin and Renal Units, The Royal Marsden NHS Foundation Trust, London, UK
6. Melanoma and Kidney Cancer Team, The Institute of Cancer Research, London, UK
7. Department of Urology, Zhongshan Hospital, Fudan University, Shanghai, China
8. Centre for Cancer Evolution, Barts Cancer Institute, Queen Mary University of London, London, UK
9. Epigenetics and Genome Stability Team, The Institute of Cancer Research, London, UK
10. Department of Urology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany
11. PKUES, Peking University, Beijing, China
12. Human Embryo and Stem Cell Unit, The Francis Crick Institute, London, UK
13. State Key Laboratory of Systems Medicine for Cancer, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
14. Department of Interventional Oncology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
15. Department of Endocrinology, St. Bartholomew's Hospital, Barts Health NHS Trust, London, UK
16. Department of Endocrinology, Royal Hallamshire Hospital, Sheffield, UK
17. Department of Cell Biology, Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX, USA
18. These authors contributed equally: Daqi Deng, Rashmi Dahiya, Libin Wang

*Corresponding author. Email: Samra.Turajlic@crick.ac.uk; Geoffrey.Feng@crick.ac.uk

Abstract

Aneuploidy, characterised by an imbalanced chromosome copy number, is the cause of chromosomal disorders[1] as well as a hallmark of cancer[2]. Beyond pathological conditions, recent studies highlight the presence of lineage-specific recurrent aneuploidies in normal tissues[3-5]. Despite its prevalence, the functional significance of aneuploidies in cellular dysfunction and diseases remains elusive, hindered by experimental challenges in engineering relevant aneuploidy models for mechanistic investigation. Here, we discover that the centromeric recruitment of catalytically dead Cas9 (dCas9) induces efficient chromosome-specific mis-segregation, applicable to all 24 human chromosomes. We show that the mis-segregation phenotype is driven by dCas9-induced kinetochore chromatin relaxation, resulting in whole chromosome aneuploidy or arm-level aneuploidy with centromeric breakage. In induced pluripotent stem cells, we demonstrate aneuploidy manipulation for chromosome 13, 21, X, and Y. In primary renal epithelial cells, we engineered chromosome 3(p) loss, the tumour-initiating event in clear cell renal cell carcinoma (ccRCC)[6].. Overall, we describe a simple, efficient, and versatile approach for chromosome-specific aneuploidy generation that can facilitate preclinical aneuploid model development for functional interrogation of aneuploidy.

Introduction

Aneuploidy, the presence of chromosome-level copy number gains or losses, typically arises from errors in chromosome segregation during cell division. Aneuploidy in gametes or embryonic stem cells can lead to chromosomal disorders such as Down syndrome[1]. Somatic aneuploidy has long been recognised as a near-ubiquitous feature in almost all cancer types[2], while recent studies revealed the presence of somatic aneuploidy across various morphologically normal tissues[3-5].

Like genetic mutations, distinct patterns of aneuploidies are recurrently observed across different cancer types and normal tissues, highlighting context-dependent selection and functional significance[3-5,7]. However, in contrast to our ability to engineer gene-level alterations using techniques like CRISPR-Cas9, it has been challenging to efficiently and precisely engineer targeted aneuploidy in relevant preclinical models, impeding their functional interrogation[8]. Consequently, despite its prevalence, the exact role of aneuploidy in different contexts is largely unexplored.

To date, several approaches for generating aneuploidy have been developed, each with different limitations. Mitotic checkpoint inhibitors like reversine can efficiently generate near-random chromosome mis-segregation and aneuploidy but lack chromosome specificity[9]. More recently, chromosome-specific aneuploidy engineering methods were developed by inducing chromosome-specific mis-segregation, including ectopic kinetochore establishment[10], kinesin motor tethering[11], and manipulation of the histone H3 variant centromere protein A (CENPA)[12]; however, the application of these methods is restricted to a subset of chromosomes. In addition, approaches such as microcell-mediated chromosome transfer (MMCT)[13] and sequence-specific DNA breakage induction[14,15] can specifically target all chromosomes, but have limited efficiency and only generate gains or losses, respectively.

Recently, the complete human centromeric DNA sequence was resolved by the telomere-to-telomere (T2T) consortium, sparking increased interest in human centromere research[16-18]. The human centromere is mainly composed of α-satellite (αSat) DNA tandem repeats,

organised into higher-order repeat (HOR) structures spanning millions of bases[19]. In each human chromosome, a portion of αSat HOR region is epigenetically defined by the incorporation of CENPA, which is the key nucleating component of the kinetochore, a protein network essential to attach chromosomes to mitotic spindle microtubules for chromosome segregation[19]. Therefore, with the newly resolved centromeric sequence, CRISPR-based tools can now directly target the functionally important region for chromosome segregation. A recently described approach, KaryoCreate, achieved chromosome-specific mis-segregation and aneuploidy formation for 10 human chromosomes by targeting mutant kinetochore component Kinetochore Scaffold 1 ($KNL1^{mut}$) to centromeric αSat HOR regions using catalytically dead Cas9 (dCas9) for microtubule attachment disruption[20].

Despite these remarkable advances, we still lack a simple, efficient, specific and versatile aneuploidy engineering method that is applicable to all human chromosomes. Here, we report such an approach, which induces chromosome-specific mis-segregation by leveraging the newly characterised function of dCas9 in altering the chromatin organisation of the core centromeric region.

Centromeric dCas9 recruitment induces efficient chromosome-specific mis-segregation

Chromosome 3 (Chr3) aneuploidy is a near-universal tumour-initiating event in clear cell renal cell carcinoma (ccRCC)[6]. To obtain this aneuploid pre-tumoural population, we initially aimed to adapt CRISPR LiveFISH[21], a method that fluorescently labels repetitive genomic regions in live cells with dCas9 and fluorophore-tagged guide RNA (gRNA), to identify and isolate the rare pre-existing Chr3 aneuploid population from primary normal renal epithelial cells. While we successfully labelled Chr3 with dCas9 and a gRNA targeting Chr3-specific centromeric αSat HOR repetitive sequences, we unexpectedly observed that this approach also induced efficient Chr3-specific mis-segregation and micronucleus formation (Fig. 1a). As the CENPA-bound region, responsible for kinetochore assembly, constitutes a subset of the targeted αSat

HOR region (Extended Data Fig. 1a), we postulate that the observed mis-segregation phenomenon is likely attributed to dCas9-induced disruption of the CENPA-bound kinetochore region. To exploit this finding for the purpose of targeted aneuploidy engineering, we designed chromosome-specific gRNAs for all 24 human chromosomes, targeting their respective CENPA-bound regions as annotated by the T2T consortium[16-18] (Fig. 1b, Extended Data Fig. 1a, and Supplementary Table 1; Also see Methods).

We tested 454 gRNAs for their ability to induce micronucleus formation in induced pluripotent stem cell (iPSC) line iFCI013 (Extended Data Fig. 1b). We identified 56 functional gRNAs (defined as inducing ≥5% micronucleated cells in 48 hours) for all chromosomes except Chr13 and Chr22 (Fig. 1b and Extended Data Fig. 2). We validated the chromosome identity of the micronuclei induced by functional gRNAs using fluorescence *in situ* hybridisation (FISH). While most functional gRNAs demonstrated on-target specificity, three gRNAs exhibited significant off-target effects: two Chr14 gRNAs (Chr14-sg11 and sg14) induced micronuclei for both Chr14 and Chr22, and one Chr21 gRNA (Chr21-sg23) induced micronuclei for both Chr21 and Chr13 (Extended Data Fig. 2). Chr13/21 and Chr14/22 are two pairs of acrocentric chromosomes sharing high sequence homology within CENPA-bound regions[17], and the observed off-target activities are likely attributable to the identified dCas9 off-target sites of these three gRNAs (Extended Data Fig. 3).

Given this sequence homology, we designed 15 bi-chromosome-specific Chr13/21 gRNAs targeting the shared CENPA-bound sequences of Chr13 and Chr21 (Extended Data Fig. 1a, and Supplementary Table 1). We identified two functional Chr13/21 gRNAs, that bi-specifically induce micronucleus formation for the Chr13/21 pair with markedly improved efficiency compared to Chr21-sg23 (Fig. 1b and Extended Data Fig. 2).

Next, we selected a panel of 23 efficient functional gRNAs covering all 24 human chromosomes and further validated their performance in *TP53* proficient RPE1 retinal pigment epithelial cells (or K1088N primary renal epithelial cells for ChrY-sg3). We confirmed chromosome-specific micronucleus induction across different models for the selected 23

gRNAs (Fig. 1c,d, and Extended Data Fig. 4), however, we also noted seven gRNAs exhibiting consistent model-specific differences in efficiency (Fig. 1c, see Discussion).

In summary, using dCas9 and the validated functional gRNAs targeting the centromeric CENPA-bound region, we can induce efficient chromosome-specific mis-segregation for 20 human chromosomes and bi-chromosome-specific mis-segregation for Chr13/21 and Chr14/22. We referred to this approach as 'CRISPR-Taiji' (CRISPRt), inspired by the imperfect segregation depicted in the 'Taiji' symbol.

Kinetochore chromatin relaxation underlies dCas9-induced kinetochore rupture and chromosome mis-segregation

We next investigate the mechanism by which dCas9 compromises faithful chromosome segregation. After Chr3 CRISPRt in metaphase cells, the outer kinetochore component (HEC1) was present at Chr3 centromeres (Fig. 2a), suggesting dCas9 recruitment did not prevent mitotic kinetochore assembly. However, Chr3 already exhibited defects in properly aligning at the metaphase plate (Fig. 2a). Following anaphase progression, we observed that both the inner (CENPA/B/C indicated by CREST antibody staining) and outer (HEC1) kinetochore components were lost from the centromere of mis-segregating Chr3 (Fig. 2b), suggesting an acute loss of the entire kinetochore in anaphase. Consistently, following mitotic exit, only Chr3 in interphase micronuclei presented a near-complete loss of the inner kinetochore (Fig. 2c).

Interestingly, a similar set of mitotic defects has been described following Polo-like kinase 1 (*PLK1*) inhibition[22,23], including defective metaphase alignment despite normal kinetochore assembly, followed by entire kinetochore loss. Mechanistically, loss of *PLK1* function causes aberrant centromeric recruitment of DNA helicases, resulting in relaxation of kinetochore nucleosomes and chromatin[22,23]. This disruption compromises the interaction required to anchor the kinetochore at the centromere, leaving the kinetochore vulnerable to complete detachment under the mechanical force exerted by spindle microtubules, termed 'kinetochore rupture[22]. Given similar defects and the DNA unwinding activity of dCas9, we investigated

whether dCas9-induced mis-segregation also functions through kinetochore chromatin relaxation by measuring chromatin accessibility with ATAC-seq. Indeed, with functional Chr3-sg2 and Chr3-sg5 gRNAs, we observed a Chr3-specific increase in nucleosome accessibility across the CENPA-bound sequences (Fig. 2d,e), confirming kinetochore chromatin relaxation. Moreover, Chr3-sg2, which has a higher micronucleus induction efficiency, showed a greater Chr3-specific accessibility increase compared to Chr3-sg5; whereas Chr3-sg4, classified as a non-functional gRNA, induced minimal accessibility change (Fig. 2d,e). This suggests a positive correlation between kinetochore chromatin relaxation and chromosome mis-segregation. Finally, consistent with centromere-specific gRNA localisation (Fig. 2a), there was no obvious global chromatin accessibility change across chromosome arms (Extended Data Fig. 5a,b). Overall, our data suggest that CRISPRt induces kinetochore rupture and chromosome mis-segregation via dCas9-induced relaxation of kinetochore chromatin (Fig. 2f).

CRISPRt induces whole chromosome aneuploidy and arm-level aneuploidy with centromeric breakage

To assess the aneuploid events resulting from CRISPRt-induced chromosome mis-segregation, we used an immortalised renal epithelial cell line RPTEC-shP53 with stable *TP53* knockdown. We first confirmed efficient Chr3-specific mis-segregation and micronucleus formation in RPTEC-shP53 cells by CRISPRt (Extended Data Fig. 6a). Chromosome painting analysis of metaphase spreads revealed Chr3 aberrations in approximately 40% of RPTEC-shP53 cells 7 days post Chr3 CRISPRt (Extended Data Fig. 6b). We confirmed that CRISPRt-induced mis-segregation resulted in both whole chromosome gains and losses (Fig. 3a). In addition, we also observed frequent arm-level aneuploidy events (Fig. 3a), which are typically caused by centromeric breakage. Consistently, we identified a subset of CRISPRt-induced mis-segregated chromosomes that exhibited split chromosome arms connected by a centromeric DNA bridge across daughter cells (Fig. 3b, Extended Data Fig. 6c and Supplementary Video 1). Therefore, arm-level aneuploid events can arise from this subset of

chromosomes following the breakage of the centromeric bridge. Interestingly, the generation of split chromosome arms with a centromeric bridge was also observed following *PLK1* inhibition[23], suggesting its dependence on the kinetochore chromatin relaxation. Finally, we found shattered chromosome fragments indicative of chromothripsis following CRISPRt (Fig. 3c), consistent with the frequent occurrence of chromothripsis in micronucleus[24]. Overall, we demonstrate that CRISPRt-induced chromosome mis-segregation generates whole chromosome and arm-level gains and losses, as well as chromothripsis, highlighting its potential to model diverse aneuploidy types of interest.

CRISPRt generates Chr13, 21, X or Y aneuploidy in iPSCs

Patient-derived iPSCs are valuable models for studying chromosomal disorders, including Down syndrome (Chr21 Trisomy), Patau syndrome (Chr13 Trisomy), Klinefelter syndrome (47, XXY), Turner syndrome (45, X), and Jacob syndrome (47, XYY). Although, using isogenic iPSC lines with aneuploid and diploid karyotypes is desirable to precisely characterise aneuploidy-induced changes, chromosome-specific aneuploidy generation or correction in iPSCs remains experimentally challenging[25,26]. To test CRISPRt for targeted aneuploidy manipulation in iPSCs, we applied Chr13/21, X or Y CRISPRt in euploid iFCI013 iPSC line, followed by single-cell chromosome copy number characterisation using single-cell RNA sequencing (scRNA-seq) and InferCNV analyses (Fig. 4a, see Methods). Intended aneuploidies (both gains and losses) were successfully generated (Fig. 4b,c) without affecting the expression of key pluripotency genes (Extended Data Fig. 7a,b). ChrX-sg22 generated the most on-target aneuploid events (Fig. 4c), and Chr21-sg23 induced aneuploid events in both Chr13 and Chr21 (Fig. 4c), consistent with the observed micronucleus induction efficiency and specificity for the corresponding gRNAs. Notably, despite Chr13/21 bi-specific targeting by Chr21-sg23, the individual engineered aneuploid iFCI013 cells predominantly carry a single aneuploid event of either Chr13 or Chr21 (Extended Data Fig. 7c), suggesting that the occurrence of an undesired additional ‘off-target’ aneuploid event is infrequent at the

single-cell level when generating Chr13 or Chr21 aneuploidy with the bi-specific gRNA. Finally, we observed that ChrX CRISPRt generated both whole ChrX and arm-level aneuploidy events in iFCI013 cells (Extended Data Fig. 7d), consistent with the aneuploidy types identified in RPTEC-shP53 cells by Chr3 CRISPRt (Fig. 3a). Overall, our results demonstrate easy and efficient targeted aneuploidy manipulation in iPSCs using CRISPRt.

CRISPRt generates ccRCC initiating Chr3 aneuploid event in renal epithelial cells

Renal proximal tubule epithelial cells (PTECs) are considered the cell of origin of ccRCC[27]. In sporadic ccRCC, loss of Chr3 p-arm (Chr3p) is the near-universal tumour-initiating genetic alteration[6] (Extended Data Fig. 8a). The loss sometimes extends to include Chr3 q-arm or the entire Chr3, therefore, we refer to it as 'Chr3(p) loss' hereafter[6]. Chr3(p) loss results in the loss of heterozygosity (LOH) of multiple tumour suppressor genes, including *VHL*, which encodes the master regulator of the hypoxia-inducible factor (HIF) signalling pathway[28] (Extended Data Fig. 8b). In 90% sporadic ccRCC cases, the inactivation of the remaining *VHL* wild-type (WT) copy occurs after Chr3(p) loss through mutation or methylation leading to *VHL* loss, driving ccRCC tumorigenesis through aberrant HIF pathway activation (Extended Data Fig. 8a)[29].

Von Hippel-Lindau (VHL) disease is an inherited cancer predisposition syndrome caused by a pathogenic germline variant in *VHL* and manifests with multiple tumour types, including ccRCC[30]. In patients with VHL disease, ccRCC tumorigenesis is typically driven by the loss of Chr3(p) carrying the remaining *VHL* WT copy (Extended Data Fig. 8a), however, we also observed a minority of lesions driven by somatic inactivation of the remaining *VHL* WT copy without Chr3(p) loss[31]. To recapitulate and study these genetic stereotypes in relevant pre-clinical models, we established a primary PTEC organoid line (K1088N) from normal kidney tissue of a patient with VHL disease and applied Chr3 CRISPRt (Chr3-Ct) (Fig. 5a). In addition, we performed *VHL* WT-allele-specific knockout (*VHL*-ASK) to recapitulate the somatic inactivation of the remaining *VHL* WT copy (Fig. 5a and Extended Data Fig. 9a,b).

We first analysed K1088N Chr3-Ct cells 7 days (D7) post-treatment using scRNA-seq. We found efficient generation of Chr3-specific aneuploid events in K1088N cells with 17.6% cells carrying Chr3(p) loss (Fig. 5b,c), predominantly involving the loss of whole Chr3 (Extended Data Fig. 9c). To evaluate the fitness impact of Chr3(p) loss in PTECs, we further profiled K1088N Chr3-Ct cells 11 days (D7+4) post-treatment. The proportions of Chr3(p) loss decreased from 17.6% to 5.5% between D7 and D7+4 (Fig. 5b,c). This suggests that, despite Chr3(p) loss being a tumour-initiating event in ccRCC, it does not confer fitness advantages in primary PTECs at least under *in vitro* culture conditions. To preserve the engineered Chr3(p) loss population in diploid K1088N cells, we implemented a phenotypic selection strategy using flow sorting. Given that *VHL* loss leads to upregulation of the membrane protein CA9[32], we sorted CA9-positive (CA9+) cells from the K1088N Chr3-Ct and *VHL*-ASK D7+4 populations, aiming to enrich for cells with *VHL* loss due to Chr3(p) loss and *VHL* knockout, respectively (Extended Data Fig. 10a,b). Profiling the sorted CA9+ populations with scRNA-seq demonstrated successful enrichment of the Chr3(p) loss population from 5.5% to 27.6% (Fig. 5b,c) and also revealed significantly increased cell proportions with a high HIF pathway signature score (Extended Data Fig. 10c), indicative of *VHL* loss.

Together, we demonstrate efficient targeted aneuploidy engineering with CRISPRt in primary normal renal epithelial cells, followed by enrichment of the desired aneuploidy through phenotypic selection, for modelling of ccRCC initiation.

Discussion

We present CRISPRt, a simple and efficient approach to induce chromosome-specific mis-segregation by targeting dCas9 to the core centromeric region. CRISPRt is applicable to all 24 human chromosomes with the identified functional gRNAs. Currently, Chr13/21 and Chr14/22 can only be targeted bi-specifically in pairs due to their high CENPA-bound sequence homology, and improved specificity may be achievable by using dCas9 variants with higher fidelity or broader PAM sequence recognition. In addition, CRISPRt demonstrates

excellent efficiency in mis-segregation induction, achieving a median of 20% micronucleated cell formation within 48 hours across 24 human chromosomes in different *TP53*-proficient models. However, a subset of functional gRNAs showed varying efficiencies in inducing micronucleted cell formation across different models. This variability may be attributed to the recently reported sequence and location heterogeneities of the CENPA-bound centromeric regions across haplotypes [33,34].

Mechanistically, we demonstrate that dCas9 centromeric recruitment by functional CRISPRt gRNAs drives chromosome-specific kinetochore chromatin relaxation, which underpins CRISPRt-induced mitotic defects and chromosome mis-segregation. Another chromosome-specific aneuploidy engineering method, KaryoCreate, also explored dCas9 centromeric targeting to recruit mutant KNL1 for disrupting microtubule attachment[20]. KaryoCreate, employs a fusion protein of mutant kinetochore component KNL1^{Mut} with dCas9 (KNL1^{Mut} -dCas9), to leverage dCas9 αSat HOR targeting for centromeric KNL1^{Mut} recruitment, thereby disrupting microtubule-kinetochore attachment during mitosis[20]. To maximise KNL1^{Mut} recruitment, KaryoCreate screened for chromosome-specific αSat HOR targeting gRNAs capable of generating centromeric dCas9 foci, identifying nine functional gRNAs for targeting 10 chromosomes[20]. These nine KaryoCreate gRNAs are also in our CRISPRt gRNA design, as the CENPA-bound regions used in CRISPRt design are part of the αSat HOR regions used in KaryoCreate. Given the efficient dCas9 recruitment, we expected that the nine gRNAs would induce efficient kinetochore chromatin relaxation and mis-segregation with dCas9 alone, however, to our surprise, five of these nine gRNAs were non-functional with dCas9 alone (Supplementary Table 1). Consistently, KaryoCreate demonstrated that KNL1^{Mut}, in addition to dCas9, is indispensable for inducing significant mis-segregation and aneuploidy with KaryoCreate gRNAs[20]. Therefore, we postulate that dCas9-induced chromatin relaxation is not solely determined by dCas9 recruitment, and additional factors including the nucleosome binding position may be important. While KaryoCreate selected gRNAs based on efficient dCas9 recruitment, CRISPRt selected gRNAs for efficient mis-segregation that is dependent on dCas9-induced kinetochore chromatin relaxation. Given the different screening criteria, the

functional gRNA sets for KaryoCreate and CRISPRt likely represent distinct dCas9 functionalities at the centromeric region, warranting further characterisation.

Functionally, CRISPRt-induced mis-segregation results in gains and losses of whole chromosomes or chromosome arms. We also observed chromothripsis, which can lead to extensive genomic rearrangements and extra-chromosomal DNA formation[35], both of which are prevalent in cancers. In addition, although requiring further validation, our inferred copy number analyses suggested the formation of isochromosome (Fig. 5b), which is implicated in chromosomal disorders and cancers. These isochromosomes are likely arising from chromosome arm aneuploidy with centromeric breakage through breakage fusion of the sister chromatids (arms) after DNA replication.

We demonstrate efficient targeted aneuploidy engineering in iPSCs and primary normal epithelial cells using CRISPRt, highlighting its broad model compatibility. This is particularly important given the prevalence of aneuploidy across diverse biological contexts. By modelling Chr3(p) loss in PTECs, we revealed that the upregulated interferon response, a common transcriptomic feature of ccRCC cells, is attributed to Chr3(p) loss. This aligns with previous findings in lung squamous cell carcinoma, where Chr3p loss correlates to upregulated interferon responses in both patient tumour samples and engineered pre-clinical models[7]. Therefore, we demonstrate that the relevant aneuploidy model engineered using CRISPRt can uncover the underlying molecular and functional significance of a specific aneuploid event. Additional investigations are required to further understand how Chr3(p) loss elicits interferon response and contributes to ccRCC tumorigenesis, for which the engineered PTECs with Chr3(p) loss can serve as a useful model.

Overall, we believe that CRISPRt will be a valuable tool for advancing our understanding of the growing collection of recurrent aneuploid events identified in both physiological and pathological conditions.

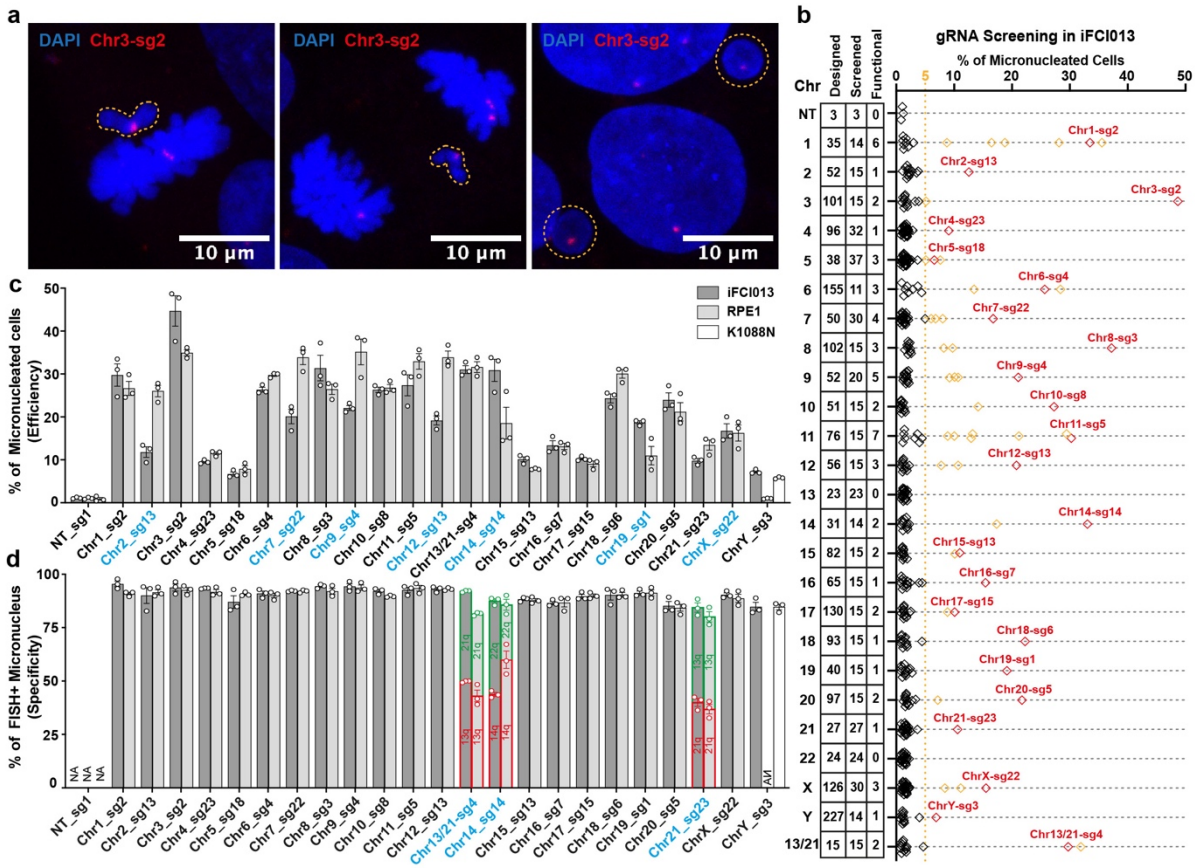


Fig. 1 Centromeric dCas9 recruitment induces efficient chromosome-specific mis-segregation

a, Representative images showing Chr3-specific mitotic mis-segregation and micronucleus formation induced by dCas9 recruitment to Chr3-specific centromeric αSat HOR sequences with Atto550-labeled Chr3-sg2 gRNA.

b, Percentage of micronucleated cells induced by dCas9 with chromosome-specific gRNAs in iFCI013 cells at 48 h (Supplementary Table 1). Rows represent chromosomes with the numbers of designed, screened, and functional gRNAs annotated. Diamonds represent gRNAs and the orange line indicates a 5% threshold above which a gRNA is deemed functional. Functional gRNAs selected for validation are shown in red with gRNA ID labeled and other functional gRNAs are shown in orange.

c, Percentage of micronucleated cells with the selected functional gRNAs in the indicated cell lines at 48 h (n = 3). Functional gRNAs with model-specific efficiency differences are marked in blue.

d, Percentage of chromosome-specific FISH positive (FISH+) micronucleus with the selected functional gRNAs in the indicated cell lines at 48 h (n = 3). Functional gRNAs that target two chromosomes due to off-target effects or bi-chromosome-specific design are marked in blue, and their corresponding bars are split by target chromosomes shown in red and green.

Scale bars,10 µm.

Error bars, mean ± SEM

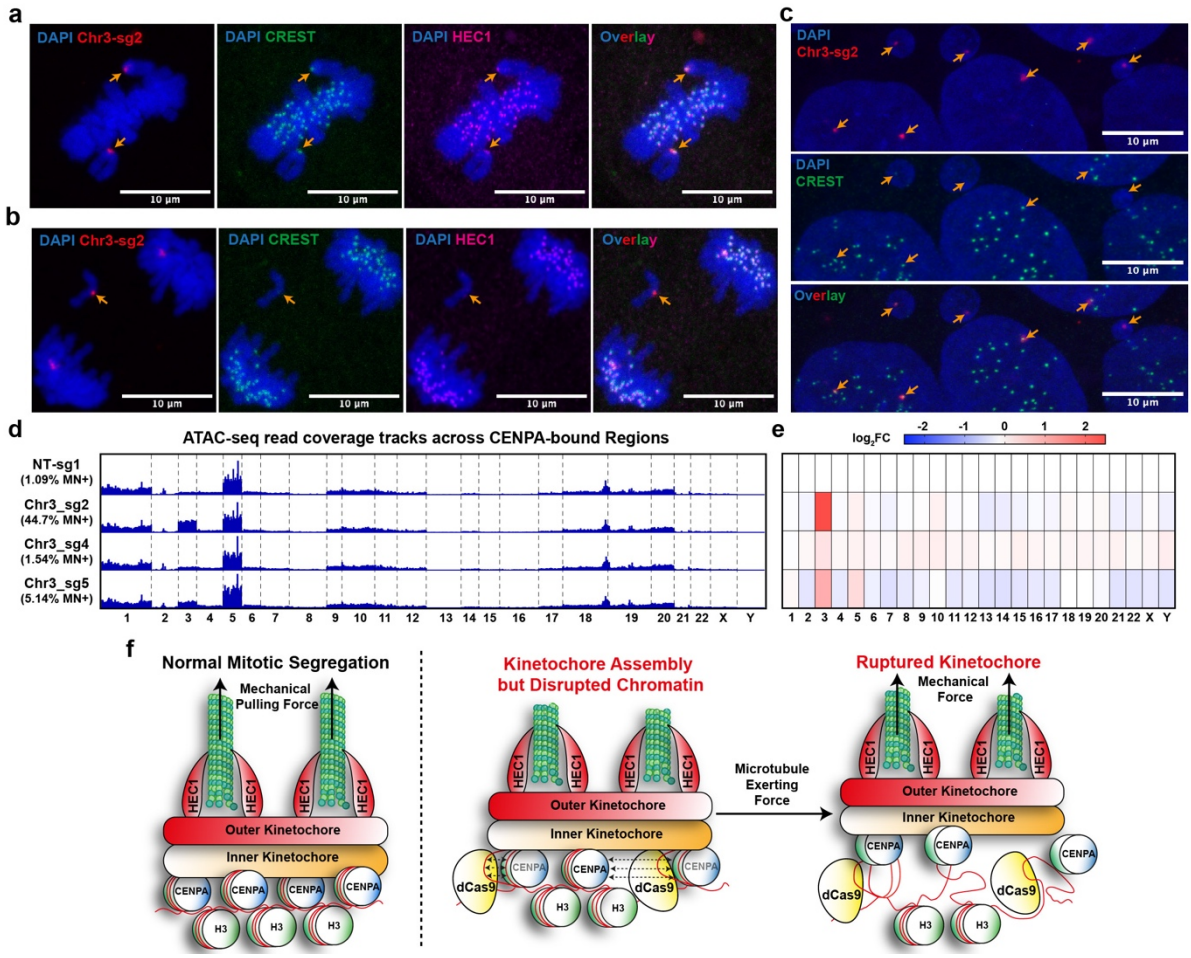


Fig. 2 CRISPRt induces kinetochore rupture through relaxation of kinetochore-associated chromatin

a-c, Representative images showing kinetochore components on Chr3 centromere in metaphase **(a)**, anaphase **(b)**, and interphase **(c)** 24 h after Chr3 CRISPRt with Atto550-labeled Chr3-sg2 gRNA in iFCI013 cells. Chr3-sg2 (red) marks Chr3 centromere, CREST antibody (green, stain for CENPA/B/C) marks the inner kinetochore, and HEC1 antibody (magenta) marks the outer kinetochore.

d,e, ATAC-seq analyses across the CENPA-bound regions 8 h after CRISPRt with the indicated gRNAs (with the corresponding micronucleated cell formation efficiency indicated) in iFCI013 cells (n = 2). Read coverage tracks **(d)** showing average normalised read distribution. Heatmap **(e)** showing average normalised $\log_2$ fold change ($\log_2$FC) of the read counts.

f, Schematics showing the proposed model that dCas9 recruitment disrupts the kinetochore chromatin organisation at the CENPA-bound region, leading to kinetochore rupture under the spindle microtubule force.

Scale bars,10 µm.

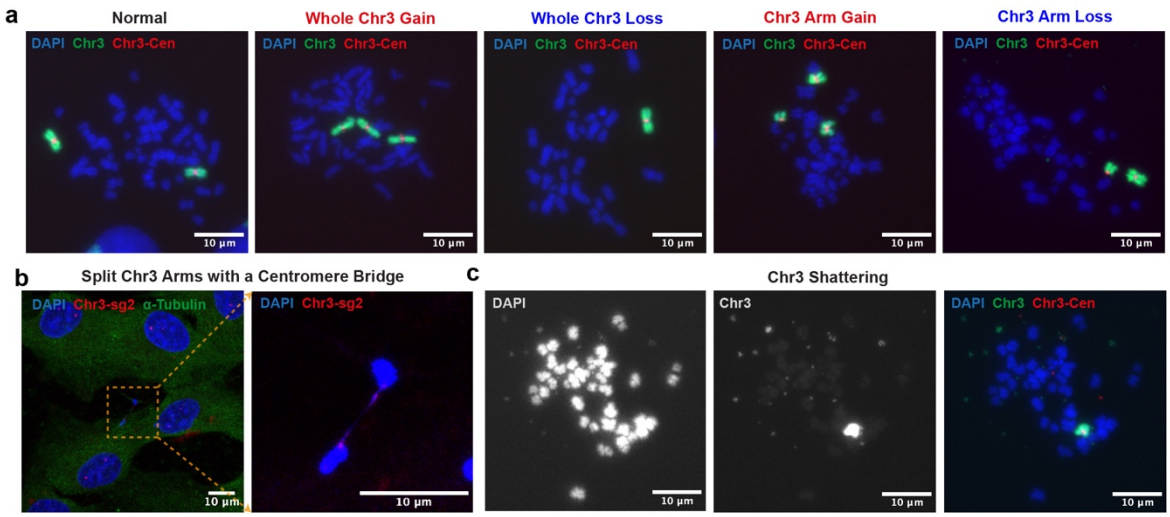


Fig. 3 CRISPRt generates a diverse spectrum of chromosomal alterations.

a, Representative images showing metaphase spreads with the indicated types of Chr3 aneuploid events 7 days after CRISPRt with Chr3-sg2 gRNA in RPTEC-shP53 cells. Chr3 arms and Chr3 centromere are shown in green and red, respectively.

b, Representative images showing the mis-segregated Chr3 with split chromosome arms connected by the centromere bridge 24 h after CRISPRt with Atto550-labeled Chr3-sg2 gRNA in RPE1 cells. Chr3-sg2 (red) marks Chr3 centromere and α-Tubulin (green) marks the cytoplasm.

c, Representative images showing shattered Chr3 fragments 7 days after CRISPRt with Chr3-sg2 gRNA in RPTEC-shP53 cells.

Scale bars,10 μm.

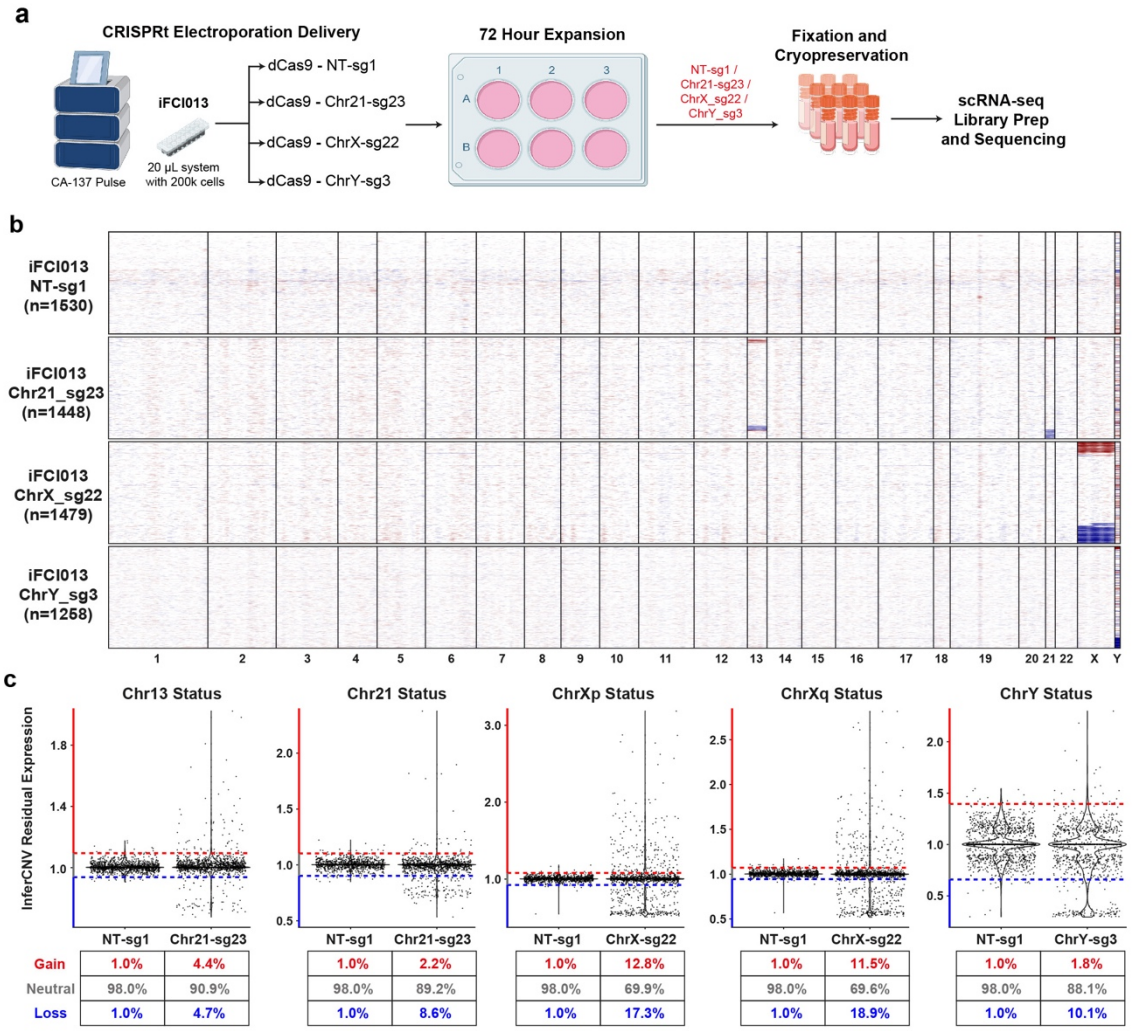
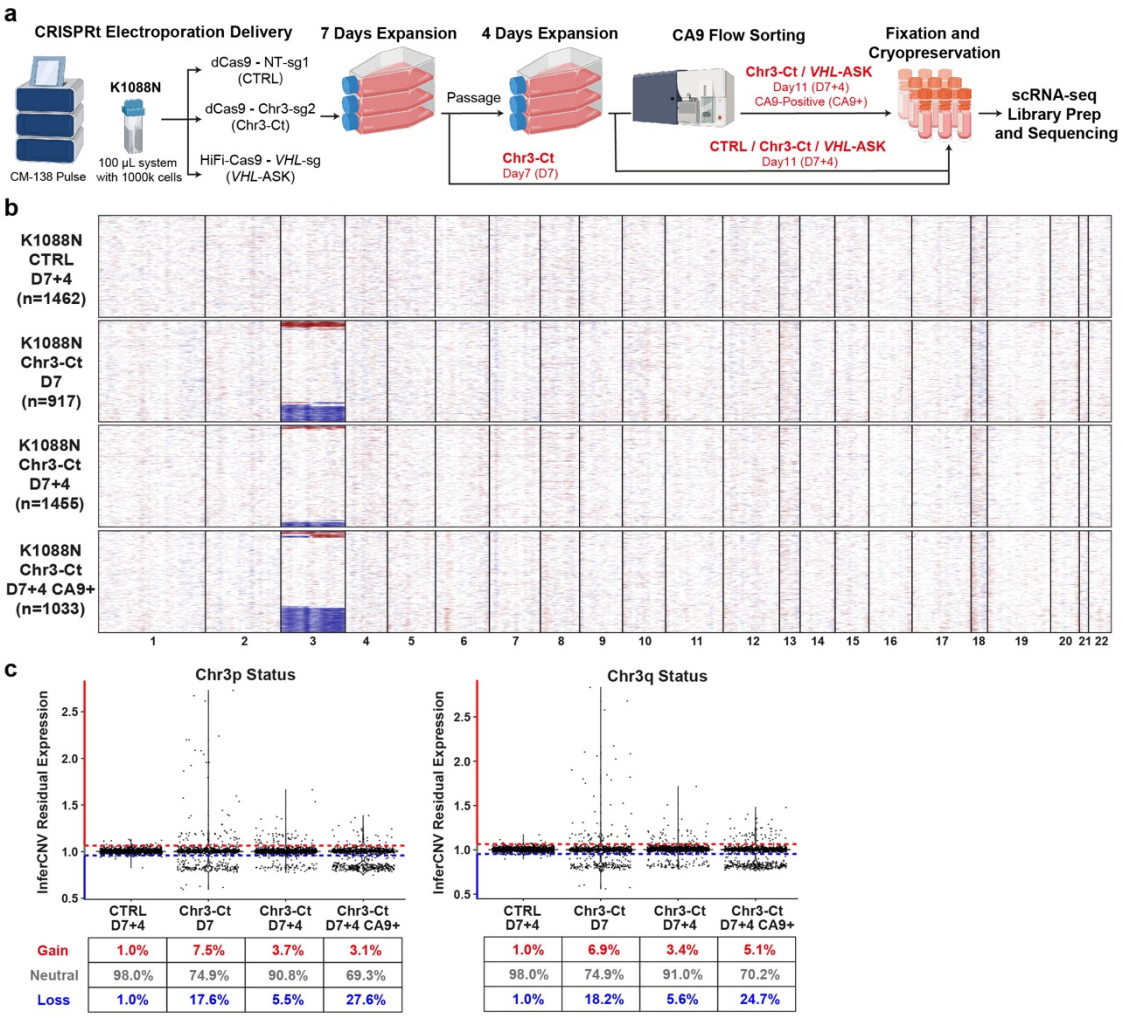


Fig. 4 CRISPRt generates Chr13, 21, X, or Y aneuploidy in iPSCs.

a, Schematics showing the experiment setup of Chr13/21, X, or Y aneuploidy engineering using CRISPRt with the indicated gRNAs in iFCI013 cells. The four samples fixed for scRNA-seq are marked in red.

b, Heatmaps showing single-cell chromosome copy number profiles inferred from scRNA-seq data of the indicated samples. The number of cells analysed (n) is shown. Rows represent individual cells and columns represent chromosomes. Copy number status is shown in blue (loss), red (gain) or white (neutral).

c, Violin plots showing InferCNV residual expression across the indicated chromosome or chromosome arm after the indicated CRISPRt. The dots represent individual cells. The blue and red dashed lines represent the thresholds (see Methods) for copy number loss and gain, respectively. The cell proportions with the indicated copy number status are shown in the tables.



	CTRL D7+4	Chr3-Ct D7	Chr3-Ct D7+4	Chr3-Ct D7+4 CA9+
Gain	1.0%	7.5%	3.7%	3.1%
Neutral	98.0%	74.9%	90.8%	69.3%
Loss	1.0%	17.6%	5.5%	27.6%

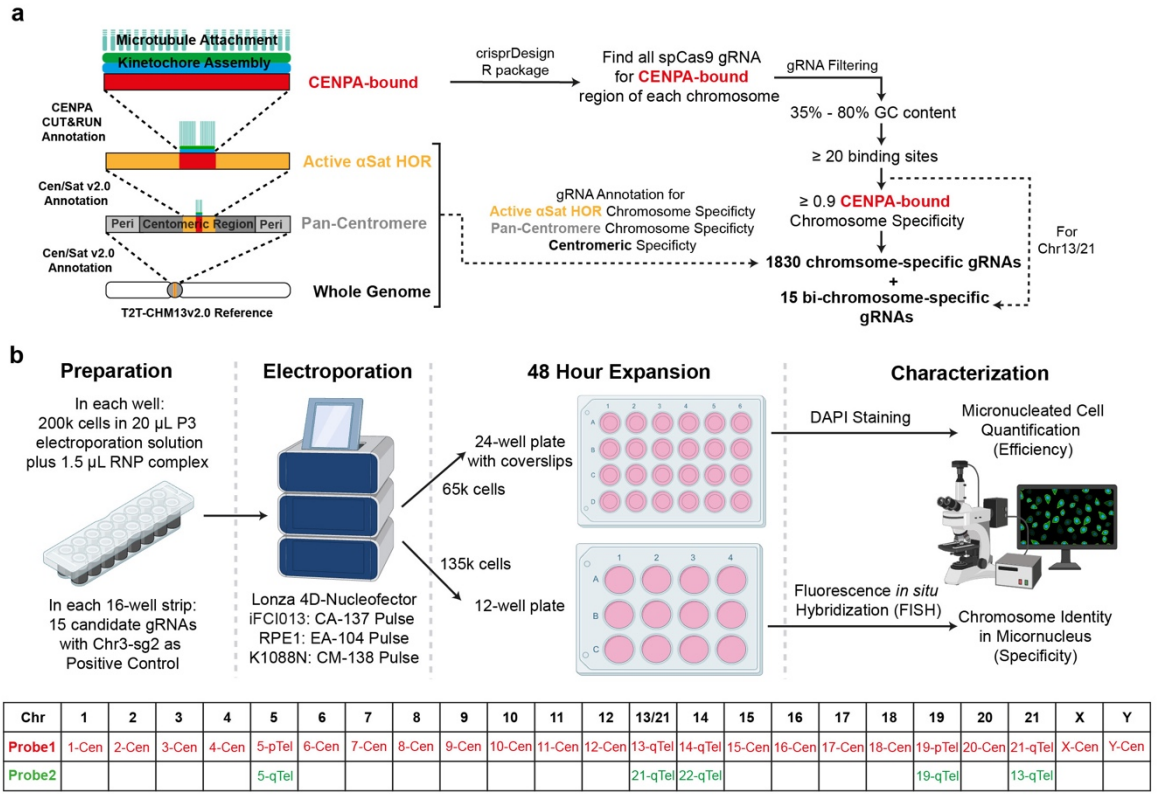
	CTRL D7+4	Chr3-Ct D7	Chr3-Ct D7+4	Chr3-Ct D7+4 CA9+
Gain	1.0%	6.9%	3.4%	5.1%
Neutral	98.0%	74.9%	91.0%	70.2%
Loss	1.0%	18.2%	5.6%	24.7%

Fig. 5 CRISPRt generates ccRCC initiating Chr3(p) loss in PTECs

a, Schematics showing the experiment setup of Chr3 aneuploidy engineering using CRISPRt with Chr3-sg2 (Chr3-Ct) or NT-sg1 (CTRL), and *VHL* allele-specific KO (*VHL*-ASK) in K1088N PTECs. Cells were cultured for 7 days (D7) or 11 days (D7+4) after treatments. Cells with Chr3-Ct or *VHL*-ASK on D7+4 were selected for CA9-positive (CA9+) cells with flow sorting. The six samples fixed for scRNA-seq are marked in red.

b, Heatmaps showing single-cell chromosome copy number profiles inferred from scRNA-seq data of the indicated samples. The number of cells analysed (n) is shown. Rows represent individual cells and columns represent chromosomes. Copy number status is shown in blue (loss), red (gain) or white (neutral).

c, Violin plots showing InferCNV residual expression for Chr3p and Chr3q of the indicated samples. The dots represent individual cells. The blue and red dashed lines represent the thresholds (see Methods) for copy number loss and gain, respectively. The cell proportions with the indicated copy number status are shown in the tables.

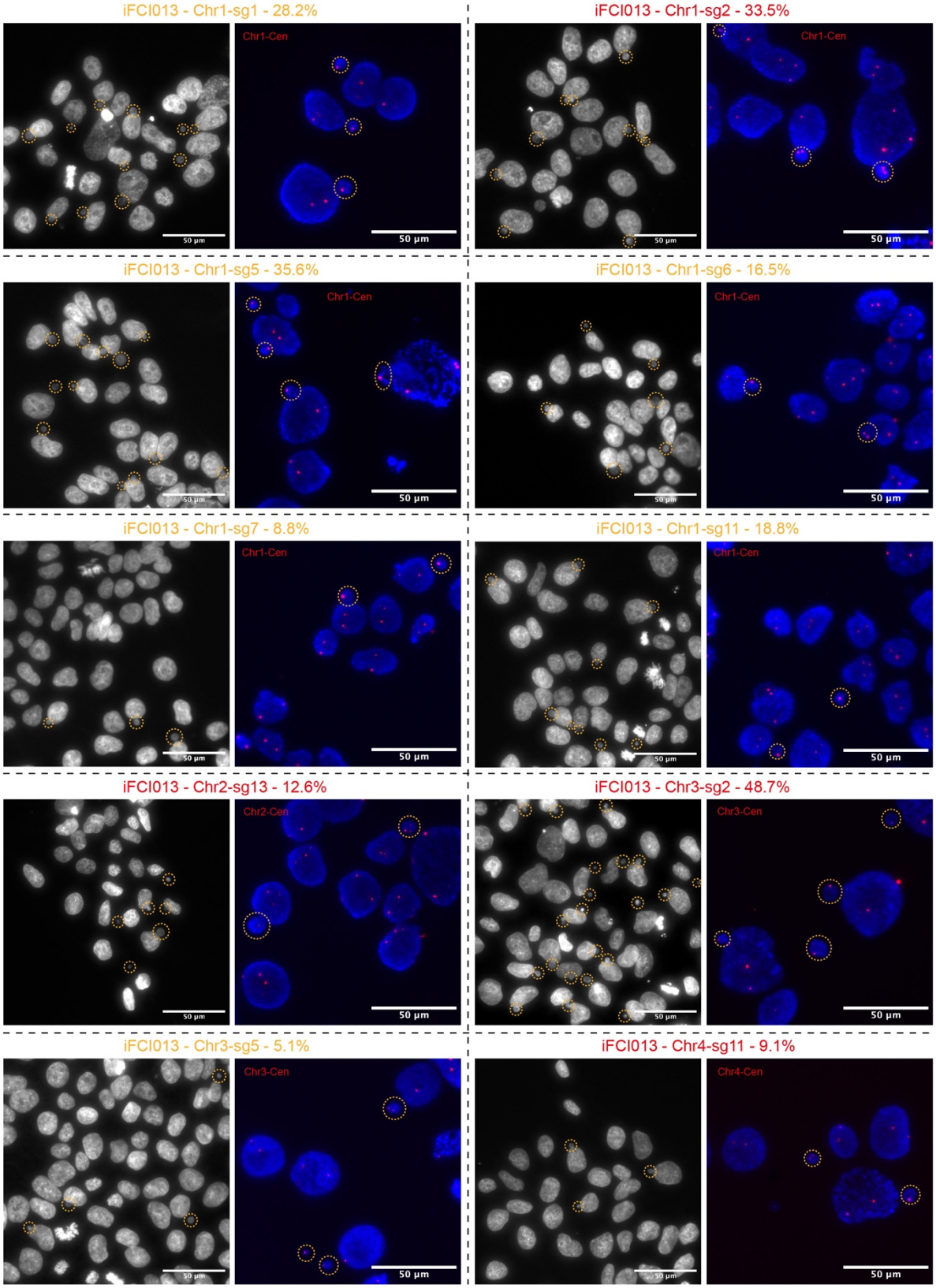


Chr	1	2	3	4	5	6	7	8	9	10	11	12	13/21	14	15	16	17	18	19	20	21	X	Y
Probe1	1-Cen	2-Cen	3-Cen	4-Cen	5-pTel	6-Cen	7-Cen	8-Cen	9-Cen	10-Cen	11-Cen	12-Cen	13-qTel	14-qTel	15-Cen	16-Cen	17-Cen	18-Cen	19-pTel	20-Cen	21-qTel	X-Cen	Y-Cen
Probe2					5-qTel								21-qTel	22-qTel					19-qTel		13-qTel		

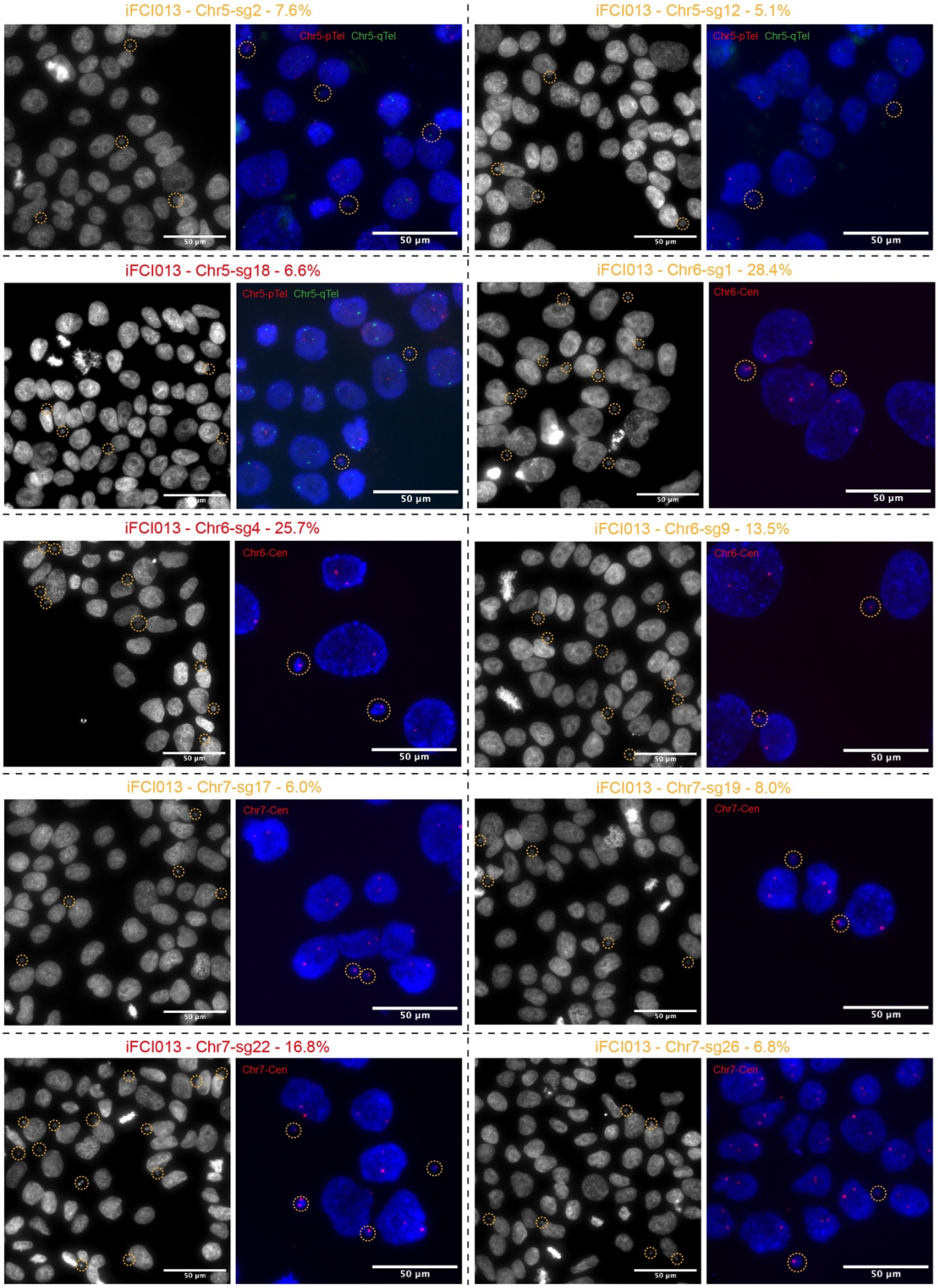
Extended Data Fig. 1 Design and screening of chromosome-specific gRNAs targeting CENPA-bound region.

a, Schematic showing the gRNA design, filtering, and annotation strategy (see Methods). Each human chromosome harbours the (peri)centromeric region (Pan-centromere), active alpha-satellite higher order repeat region (Active αSat HOR), and CENPA-bound region (CENPA-bound) in the hierarchical structure shown. The gRNA design uses the CENPA-bound sequences from T2T-CHM13v2.0 reference resolved and annotated by the T2T consortium. The gRNA filtering considers gRNA GC-content (35% - 80%), the number of binding sites (≥20), and CENPA-bound chromosome specificity (≥ 0.9). Three additional specificity scores are calculated and annotated for the 1830 filtered chromosome-specific gRNAs. 15 bi-chromosome-specific gRNAs targeting Chr13/21 pair are designed. The gRNA information is summarized in Supplementary Table 1.

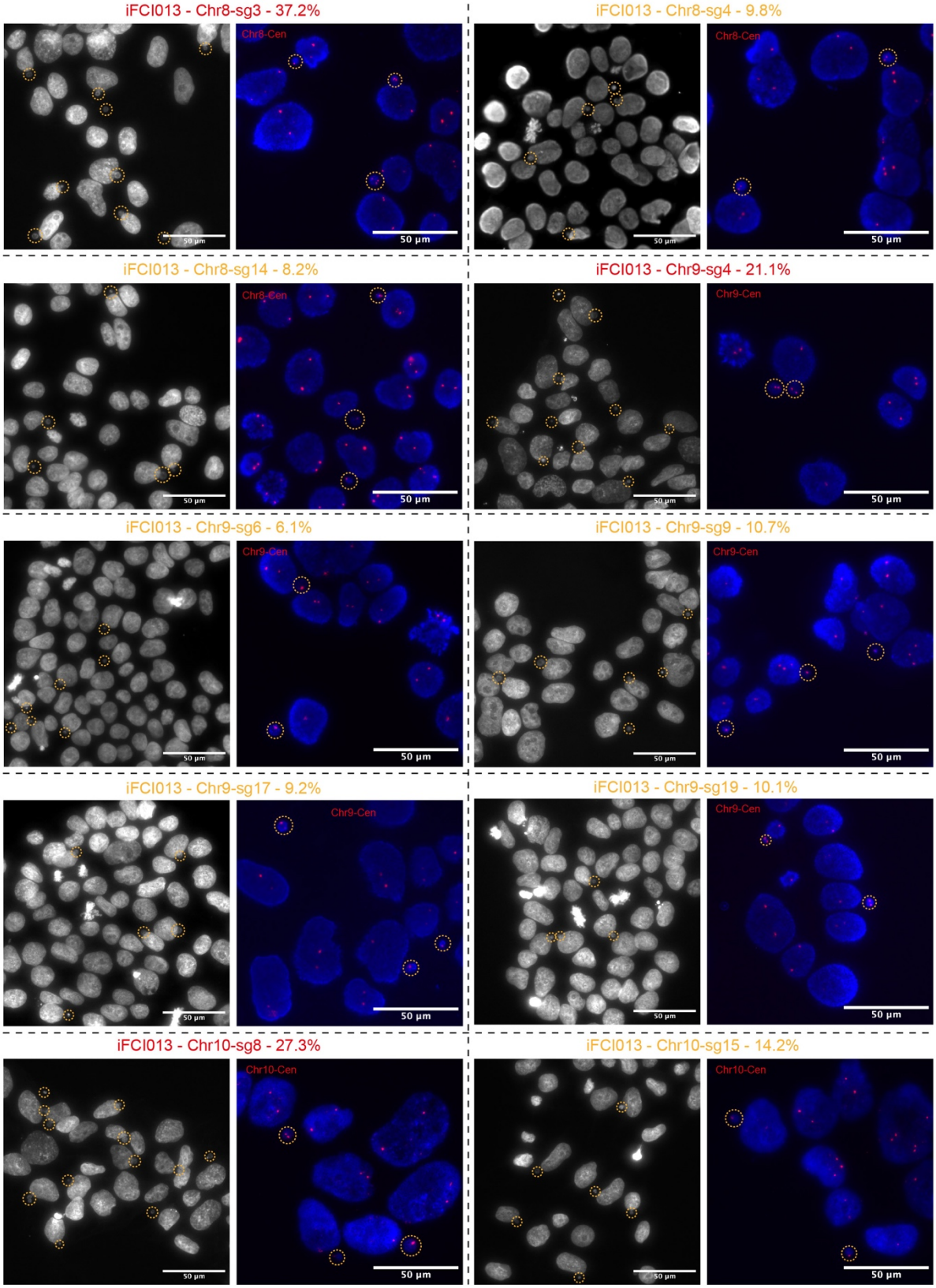
b, Schematics showing the screening setup for identifying functional gRNAs that induce chromosome-specific micronucleated cell formation. Cells are electroporated with dCas9-gRNA RNP complex and cultured for 48 h. DAPI staining is performed to quantify micronucleated cell formation (efficiency). Fluorescence *in situ* hybridisation (FISH) is performed to validate the identity of the chromosome within the micronucleus. The bottom table indicates the FISH probes used for each chromosome – 'Cen' represents the centromeric probe and 'pTel' or 'qTel' represents the telomeric probe specific to chromosome p-arm or q-arm, respectively. The gRNA screening results are Supplementary Table 1.


iFCI013 - Chr1-sg1 - 28.2%
Chr1-Cen
50 µm
iFCI013 - Chr1-sg2 - 33.5%
Chr1-Cen
50 µm
iFCI013 - Chr1-sg5 - 35.6%
Chr1-Cen
50 µm
iFCI013 - Chr1-sg6 - 16.5%
Chr1-Cen
50 µm
iFCI013 - Chr1-sg7 - 8.8%
Chr1-Cen
50 µm
iFCI013 - Chr1-sg11 - 18.8%
Chr1-Cen
50 µm
iFCI013 - Chr2-sg13 - 12.6%
Chr2-Cen
50 µm
iFCI013 - Chr3-sg2 - 48.7%
Chr3-Cen
50 µm
iFCI013 - Chr3-sg5 - 5.1%
Chr3-Cen
50 µm
iFCI013 - Chr4-sg11 - 9.1%
Chr4-Cen
50 µm

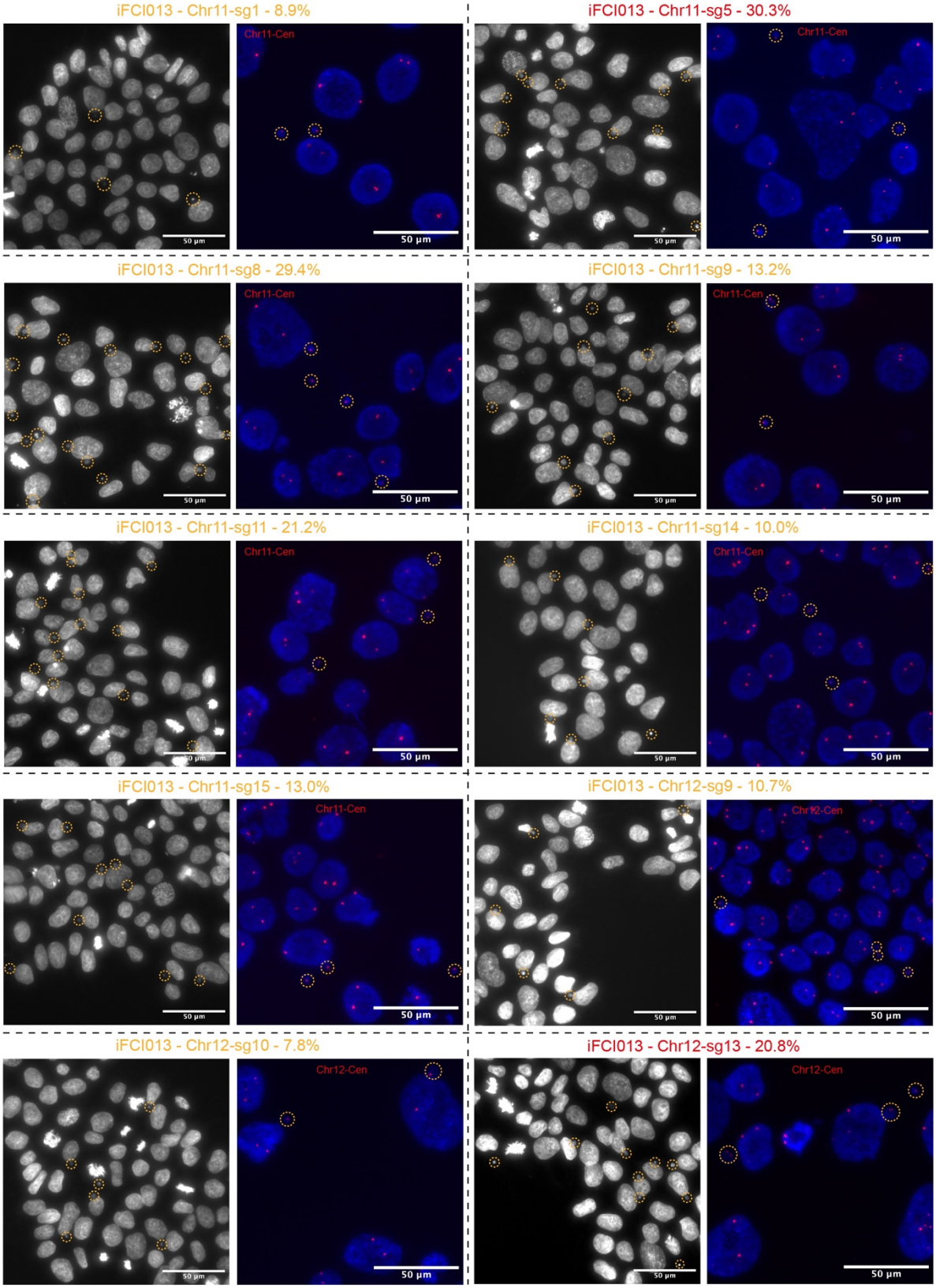
Continued


iFCI013 - Chr5-sg2 - 7.6%
Chr5-pTel Chr5-qTel
iFCI013 - Chr5-sg12 - 5.1%
Chr5-pTel Chr5-qTel
iFCI013 - Chr5-sg18 - 6.6%
Chr5-pTel Chr5-qTel
iFCI013 - Chr6-sg1 - 28.4%
Chr6-Cen
iFCI013 - Chr6-sg4 - 25.7%
Chr6-Cen
iFCI013 - Chr6-sg9 - 13.5%
Chr6-Cen
iFCI013 - Chr7-sg17 - 6.0%
Chr7-Cen
iFCI013 - Chr7-sg19 - 8.0%
Chr7-Cen
iFCI013 - Chr7-sg22 - 16.8%
Chr7-Cen
iFCI013 - Chr7-sg26 - 6.8%
Chr7-Cen
50 µm

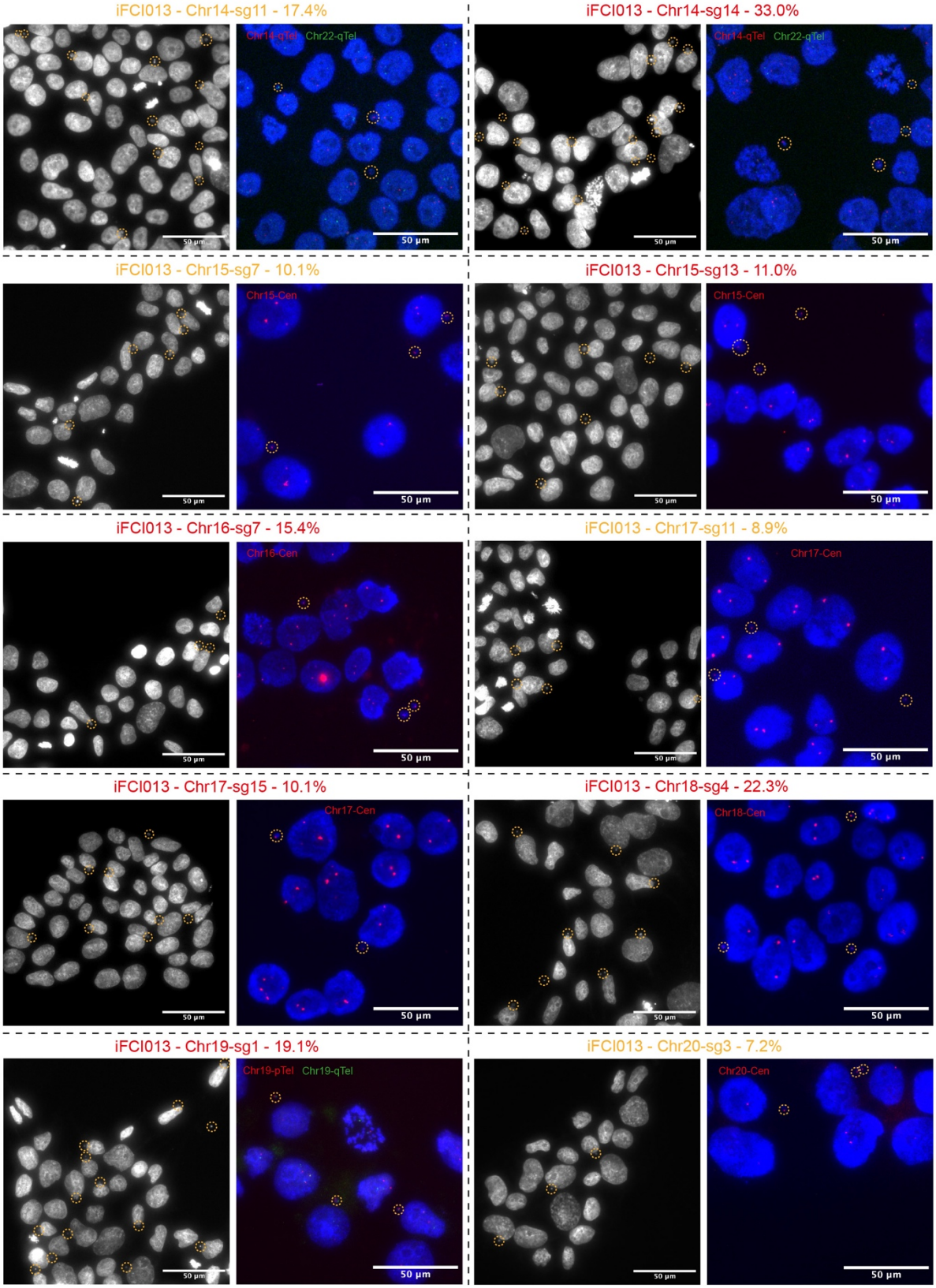
Continued


iFCI013 - Chr8-sg3 - 37.2%
Chr8-Cen
50 µm
iFCI013 - Chr8-sg4 - 9.8%
Chr8-Cen
50 µm
iFCI013 - Chr8-sg14 - 8.2%
Chr8-Cen
50 µm
iFCI013 - Chr9-sg4 - 21.1%
Chr9-Cen
50 µm
iFCI013 - Chr9-sg6 - 6.1%
Chr9-Cen
50 µm
iFCI013 - Chr9-sg9 - 10.7%
Chr9-Cen
50 µm
iFCI013 - Chr9-sg17 - 9.2%
Chr9-Cen
50 µm
iFCI013 - Chr9-sg19 - 10.1%
Chr9-Cen
50 µm
iFCI013 - Chr10-sg8 - 27.3%
Chr10-Cen
50 µm
iFCI013 - Chr10-sg15 - 14.2%
Chr10-Cen
50 µm

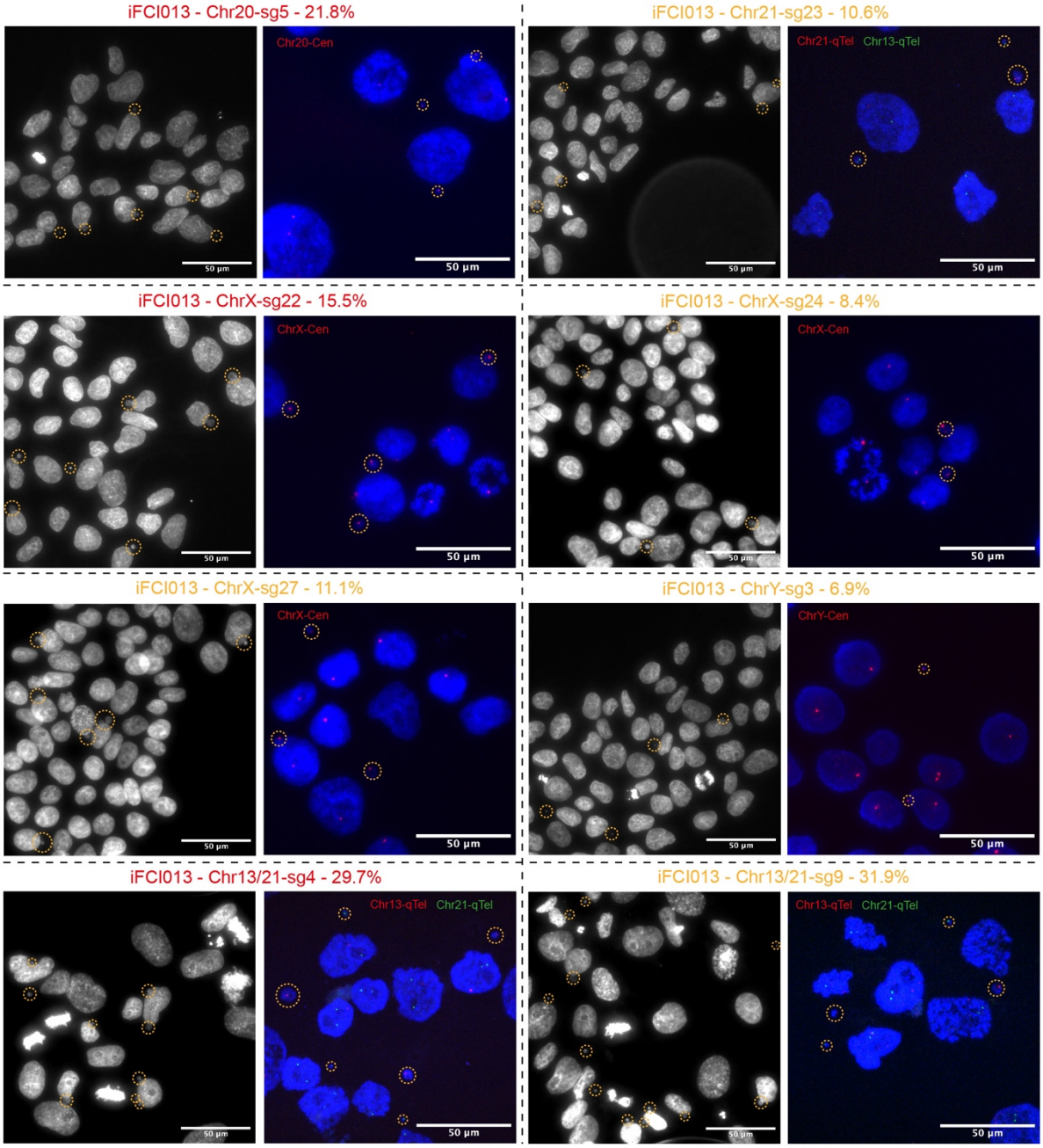
Continued


iFCI013 - Chr11-sg1 - 8.9%
Chr11-Cen
50 µm
iFCI013 - Chr11-sg5 - 30.3%
Chr11-Cen
50 µm
iFCI013 - Chr11-sg8 - 29.4%
Chr11-Cen
50 µm
iFCI013 - Chr11-sg9 - 13.2%
Chr11-Cen
50 µm
iFCI013 - Chr11-sg11 - 21.2%
Chr11-Cen
50 µm
iFCI013 - Chr11-sg14 - 10.0%
Chr11-Cen
50 µm
iFCI013 - Chr11-sg15 - 13.0%
Chr11-Cen
50 µm
iFCI013 - Chr12-sg9 - 10.7%
Chr12-Cen
50 µm
iFCI013 - Chr12-sg10 - 7.8%
Chr12-Cen
50 µm
iFCI013 - Chr12-sg13 - 20.8%
Chr12-Cen
50 µm

Continued


iFCI013 - Chr14-sg11 - 17.4%
Chr14-qTel Chr22-qTel
50 µm
iFCI013 - Chr14-sg14 - 33.0%
Chr14-qTel Chr22-qTel
50 µm
iFCI013 - Chr15-sg7 - 10.1%
Chr15-Cen
50 µm
iFCI013 - Chr15-sg13 - 11.0%
Chr15-Cen
50 µm
iFCI013 - Chr16-sg7 - 15.4%
Chr16-Cen
50 µm
iFCI013 - Chr17-sg11 - 8.9%
Chr17-Cen
50 µm
iFCI013 - Chr17-sg15 - 10.1%
Chr17-Cen
50 µm
iFCI013 - Chr18-sg4 - 22.3%
Chr18-Cen
50 µm
iFCI013 - Chr19-sg1 - 19.1%
Chr19-pTel Chr19-qTel
50 µm
iFCI013 - Chr20-sg3 - 7.2%
Chr20-Cen
50 µm

Continued



Extended Data Fig.2 Functional CRISPR-Taiji gRNAs induce chromosome-specific micronucleus formation in iFCI013 iPSCs.
Representative images showing the micronucleus induction results of the 58 identified functional gRNAs in iFCI013 iPSCs. For each gRNA, the percentage of micronucleated cells at 48 h is annotated next to the gRNA ID and the selected gRNAs are shown in red; the left image shows the micronucleated cell formation with DAPI staining (micronuclei circled in dashed line), and the right image shows FISH validation of chromosome identity within micronucleus (circled in dashed line) with the indicated FISH probe(s).
Scale bars, 50 µm.

Chr14-sg11

	GGAGCGCCCTGAGGCCTACG GGG
Chr22 - 130 sites	GGAGCGCTCTGAGGCCTACG GTG

Chr21-sg23

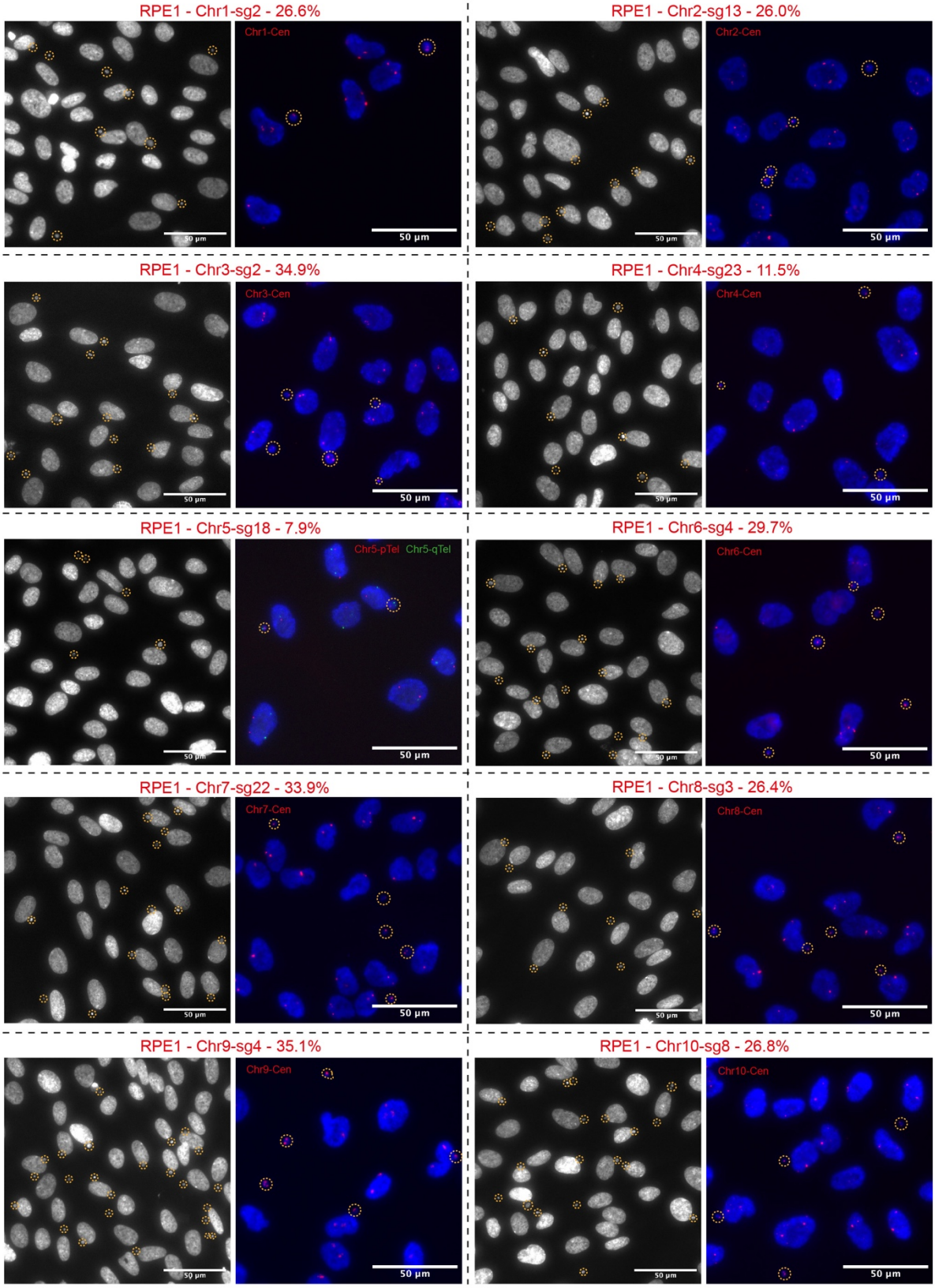
	TGTGTACTCAGCTAACAGAG TGG
Chr13 - 397 sites	TGTGTACTCGGCTAACAGAG TGG

Chr14-sg14

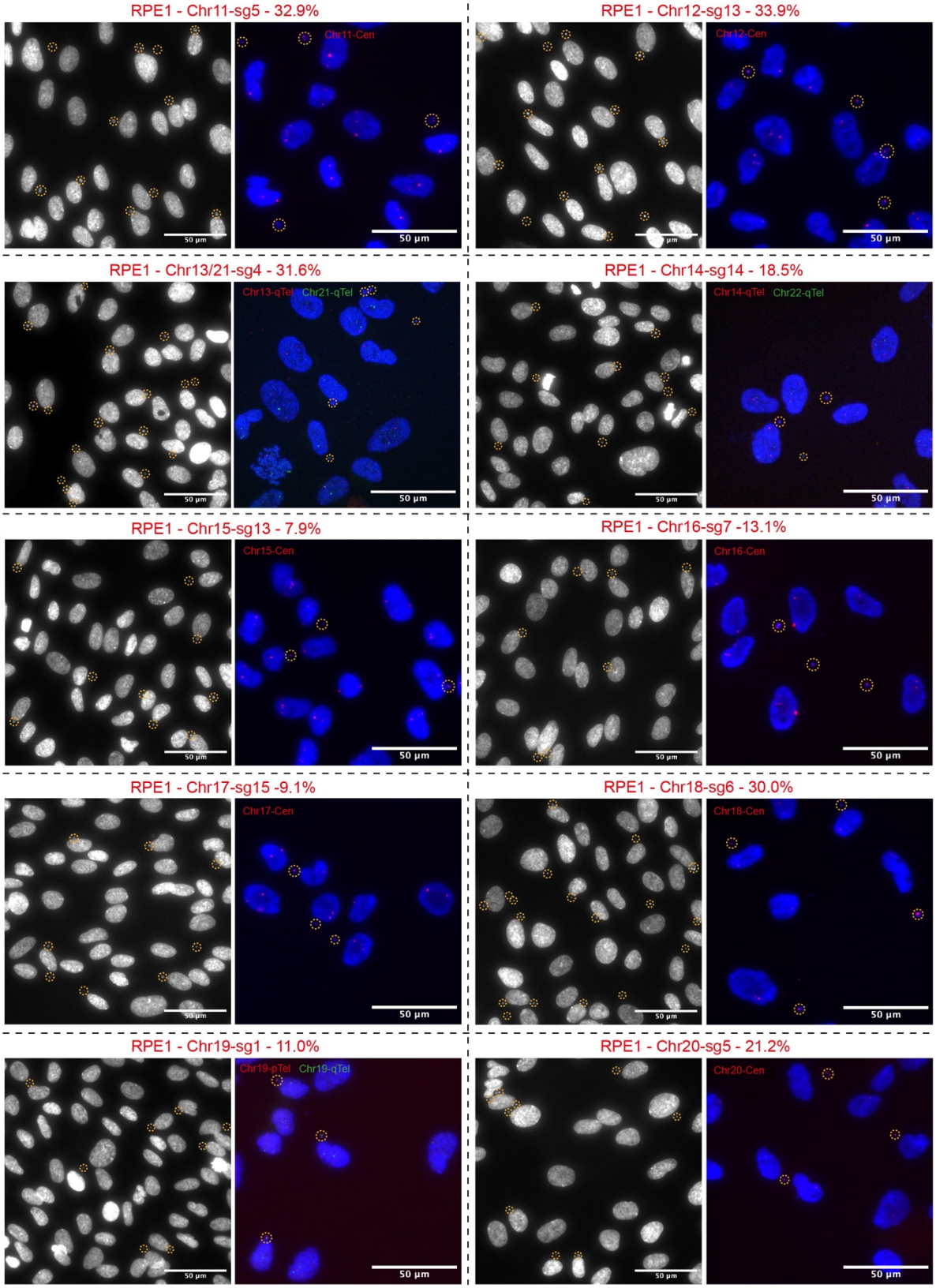
	TTGGAGCGCCCTGAGGCCTA CGG
Chr22 - 130 sites	TTGGAGCGCTCTGAGGCCTA CGG
Chr22 - 119 sites	TTGGAGCGCCTTGACGCCTA CGG

Extended Data Fig.3 Off-target sites for Chr14 and Chr21 functional gRNAs showing off-target effects.

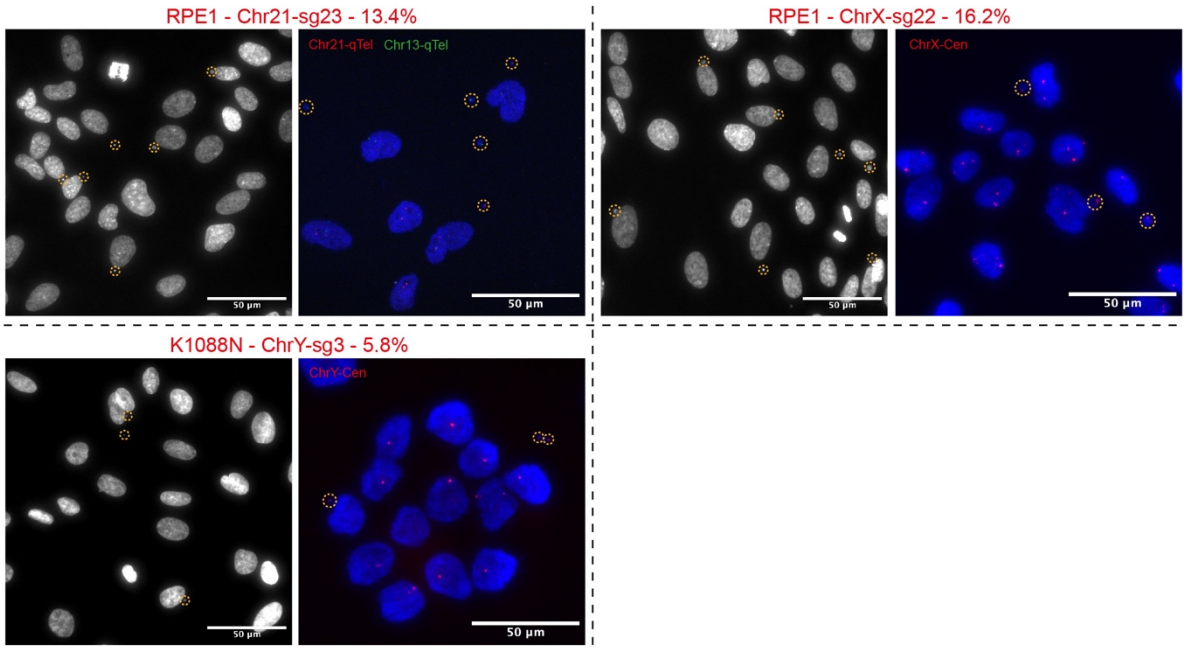
DNA sequence alignments showing the predicted off-target gRNA binding sequences with more than twenty binding sites and no more than two mismatches (shown in red) at the CENPA-bound region of the observed off-target chromosomes for the indicated functional gRNAs. The off-target chromosome and the number of predicted off-target sites are shown.


RPE1 - Chr1-sg2 - 26.6%
Chr1-Cen
50 µm
RPE1 - Chr2-sg13 - 26.0%
Chr2-Cen
50 µm
RPE1 - Chr3-sg2 - 34.9%
Chr3-Cen
50 µm
RPE1 - Chr4-sg23 - 11.5%
Chr4-Cen
50 µm
RPE1 - Chr5-sg18 - 7.9%
Chr5-pTel Chr5-qTel
50 µm
RPE1 - Chr6-sg4 - 29.7%
Chr6-Cen
50 µm
RPE1 - Chr7-sg22 - 33.9%
Chr7-Cen
50 µm
RPE1 - Chr8-sg3 - 26.4%
Chr8-Cen
50 µm
RPE1 - Chr9-sg4 - 35.1%
Chr9-Cen
50 µm
RPE1 - Chr10-sg8 - 26.8%
Chr10-Cen
50 µm

Continued

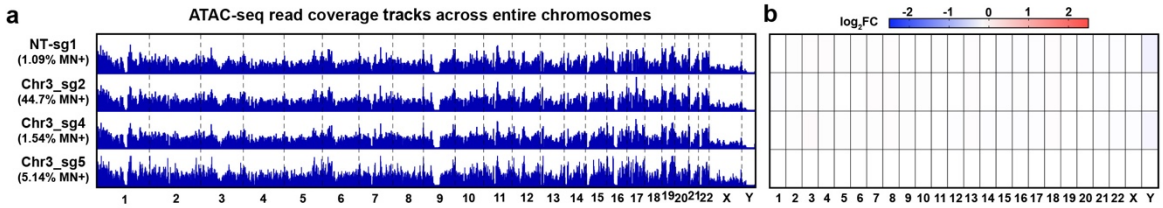

RPE1 - Chr11-sg5 - 32.9%
Chr11-Cen
50 µm
RPE1 - Chr12-sg13 - 33.9%
Chr12-Cen
50 µm
RPE1 - Chr13/21-sg4 - 31.6%
Chr13-qTel Chr21-qTel
50 µm
RPE1 - Chr14-sg14 - 18.5%
Chr14-qTel Chr22-qTel
50 µm
RPE1 - Chr15-sg13 - 7.9%
Chr15-Cen
50 µm
RPE1 - Chr16-sg7 -13.1%
Chr16-Cen
50 µm
RPE1 - Chr17-sg15 -9.1%
Chr17-Cen
50 µm
RPE1 - Chr18-sg6 - 30.0%
Chr18-Cen
50 µm
RPE1 - Chr19-sg1 - 11.0%
Chr19-pTel Chr19-qTel
50 µm
RPE1 - Chr20-sg5 - 21.2%
Chr20-Cen
50 µm

Continued



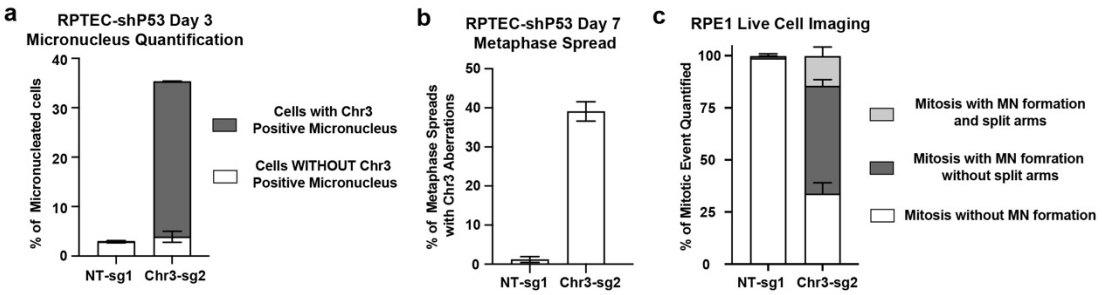
Extended Data Fig.4 Validation of the selected functional gRNAs in RPE1 or K1088N cells.

Representative images showing the characterisation results of the 23 selected functional gRNAs in RPE1 or K1088N cells. For each gRNA, the cell line used and the average percentage of micronucleated cells at 48 h of three independent repeats is annotated next to the gRNA ID; the left image shows the micronucleated cell formation with DAPI staining (micronuclei circled in dashed line), and the right image shows FISH validation of chromosome identity within micronucleus (circled in dashed line) with the indicated FISH probe(s).
Scale bars, 50 µm.

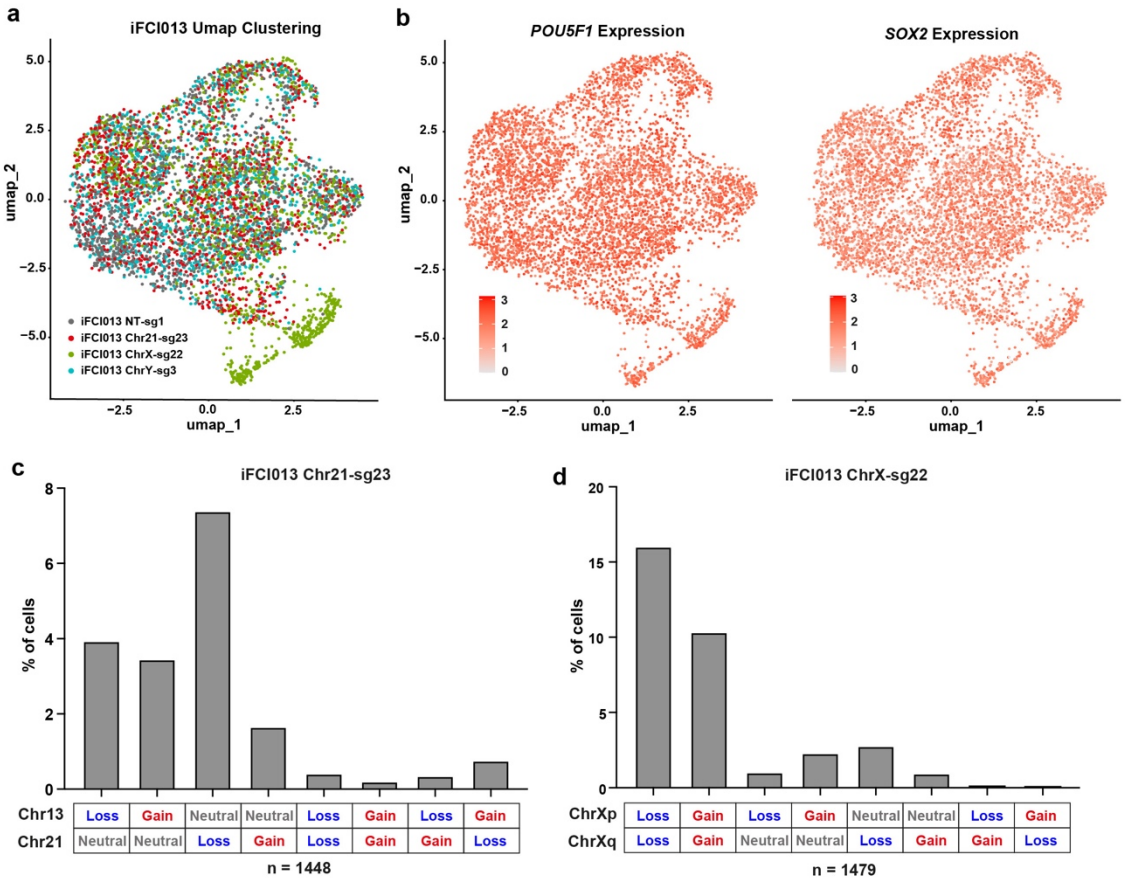


Extended Data Fig.5 CRISPRt does not affect whole chromosome-level chromatin accessibility.

a,b, ATAC-seq analyses across the entire chromosomes 8 h after CRISPRt with the indicated gRNAs (with the corresponding micronucleated cell formation efficiency indicated) in iFCI013 cells (n = 2). **(a)** Read coverage tracks showing average normalised read distribution. **(b)** Heatmap showing average normalised $\log_2$ fold change ($\log_2$FC) of the read counts.



Extended Data Fig.6 CRISPRt generates efficient and diverse types of aneuploid events.
a, Percentage of micronucleated cells (efficiency) with or without Chr3 chromosome painting signal positive micronucleus (specificity) 3 days after CRISPRt (lipofectamine RNP delivery, see Methods) with the indicated gRNAs in RPTEC-shP53 cells (n = 2).
b, Percentage of metaphase spreads with Chr3 aberrations (numerical aneuploidy, chromosome shattering, and/or simple rearrangement) 7 days after CRISPRt with the indicated gRNAs in RPTEC-shP53 cells (n = 2).
c, Percentage of mitotic events with the classified outcomes (Supplementary Video 1, see Methods) after CRISPRt with the indicated gRNAs in RPE1 cells (n=3) with time-lapse live cell imaging.
Error bars, mean ± SEM



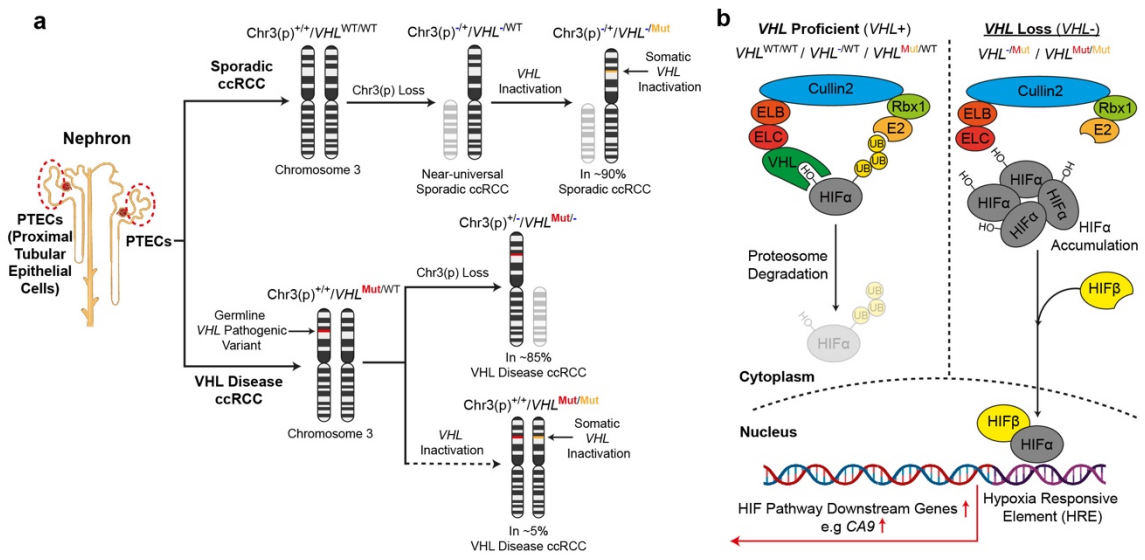
Extended Data Fig.7 Chr13, 21, X or, Y CRISPRt aneuploidy engineering in iFCI013 iPSCs.

a, UMAP clustering plot of the scRNA-seq analysis in iFCI013 cells.

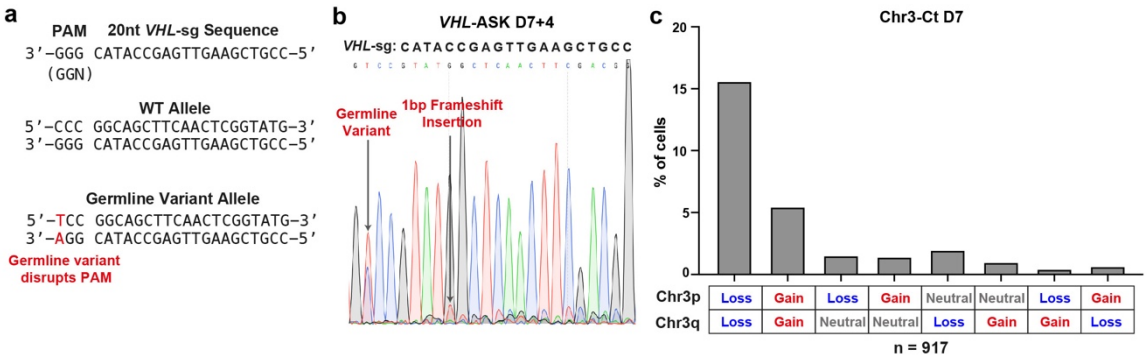
b, Feature plots showing the expression of pluripotency genes *POU5F1* (OCT4) and *SOX2* in individual iFCI013 cells. Colour gradient represents the normalized expression levels.

c, Bar plot showing the percentage of cells with the indicated copy number status of Chr13 and Chr21 in iFCI013 cells 72 h after CRISPRt with Chr21-sg23 gRNA.

d, Bar plot showing the percentage of cells with the indicated copy number status of ChrXp and ChrXq arms in iFCI013 cells 72 h after CRISPRt with ChrX-sg22 gRNA.



Extended Data Fig.8 Tumour initiation in sporadic or VHL disease ccRCC.
a, Schematics showing the tumour initiation from normal renal proximal tubule epithelial cells (PTECs) for sporadic or VHL disease ccRCC through combinations of Chr3(p) loss and *VHL* somatic or germline mutation. Chr3(p) allele status is annotated as '+' (unchanged) or '-' (loss). *VHL* gene status is annotated as 'WT' (wild-type), 'Mut' (mutated), or '-' (deleted).
b, Schematics showing *VHL* protein function in repressing the hypoxia-inducible factor (HIF) signalling pathway when *VHL* is proficient (*VHL*+, at least one WT *VHL* copy) under normal oxygen condition, and the aberrant constitutive HIF pathway activation when *VHL* is lost (*VHL-*, both *VHL* copies inactivated), driving expression of HIF pathway downstream genes.

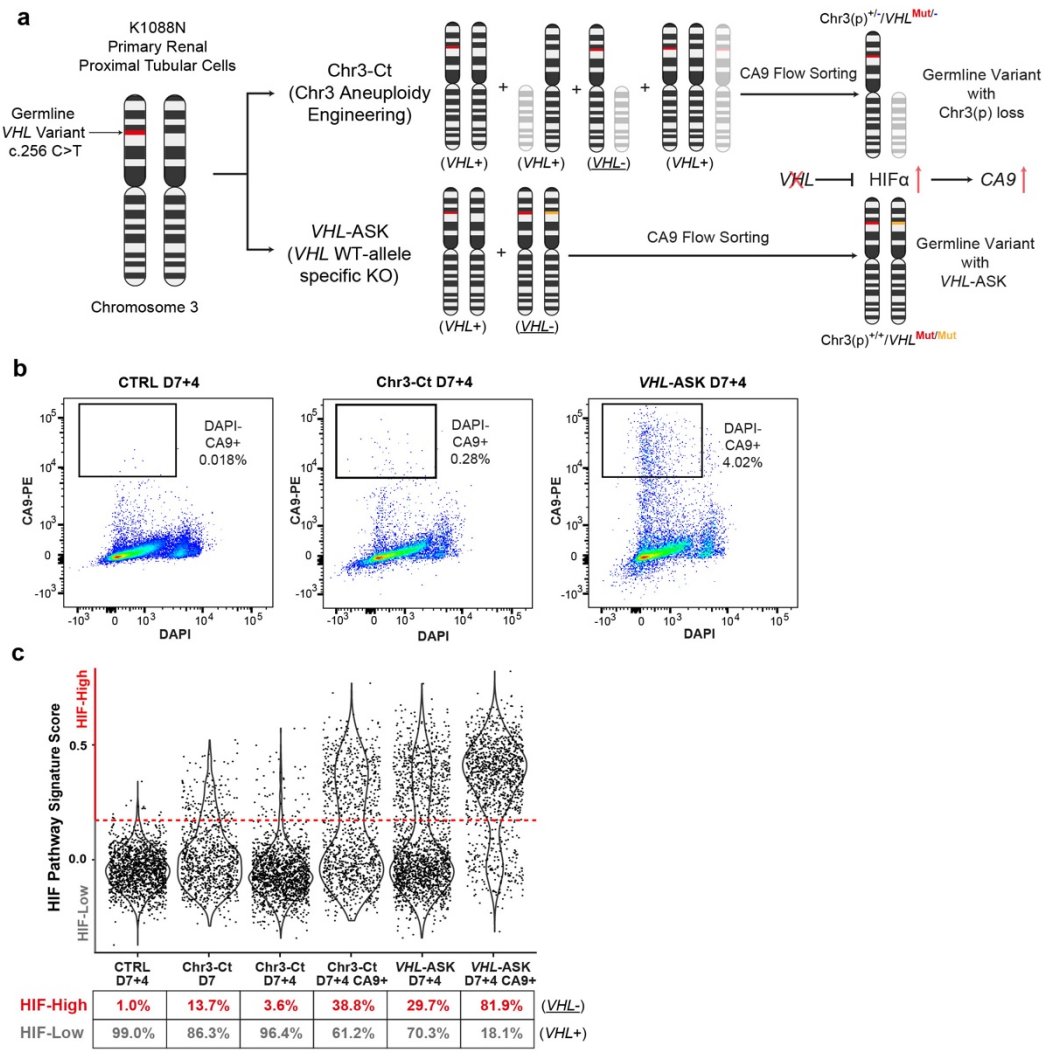


Extended Data Fig. 9 Genetic engineering results of *VHL*-ASK and Chr3-Ct in K1088N PTECs.

a, The design of gRNA (*VHL*-sg) for *VHL* allele-specific KO (*VHL*-ASK) in K1088N PTECs. The *VHL* germline variant in K1088 patient disrupts the protospacer adjacent motif (PAM) of *VHL*-sg gRNA.

b, Sanger sequencing chromatogram showing the 1bp frameshift insertion generated by *VHL*-ASK in K1088N PTECs.

c, Bar plot showing the percentage of cells with the indicated copy number status of Chr3p and Chr3q arms in K1088N PTECs 7 days (D7) after Chr3-Ct.



Extended Data Fig. 10 Enrichment of *VHL* loss population with CA9 flow sorting after Chr3-Ct or *VHL*-ASK in K1088N PTECs.

a, Schematics showing the expected genotypes generated after Chr3-Ct or *VHL*-ASK in K1088N PTECs with the corresponding *VHL* status annotated (*VHL*+ for *VHL* proficient, and *VHL-* for *VHL* loss) and the target genotypes with *VHL* loss enriched by CA9-positive (CA9+) flow sorting.

b, Scatter plots showing the CA9 and DAPI staining intensities of K1088N PTECs after the indicated treatment. The rectangular box shows the gating strategy to flow sort for live (DAPI-) CA9-positive (CA9+) cells.

c, Violin plots showing the signature scores of a curated HIF pathway activation gene set (see Methods) in K1088N PTECs after the indicated treatment. The dots represent individual cells. The dashed line represents the calculated threshold (see Methods) for defining cells with a high (HIF-High) or a low (HIF-Low) HIF pathway signature score, used for assigning *VHL* proficient (*VHL*+) or *VHL* loss (*VHL-*) status.

Supplementary Table 1 CRISPR-Taiji gRNA information

Supplementary Table 2 ATAC-seq quality check information

Supplementary Video 1 Split chromosome arms induced by CRISPRt.

Representative video clip of time-lapse live cell imaging showing the induced chromosome mis-segregation in an RPE1 cell after CRISPRt with Chr3-sg2 gRNA. One mis-segregated chromosome presents the phenotype of split chromosome arms into the two daughter cells with a putative centromere bridge, and the other mis-segregated chromosome simply forms a micronucleus in one daughter cell. The left view shows the merged 'Trans' channel (cell) in grey and 'mTurquoise-H2B' channel (DNA) in blue. The right view shows only 'mTurquoise-H2B' channel (DNA) in grey. Scale Bars, 30 μm.

Materials and Methods

Patient Information

Two male von Hippel-Lindau (VHL) disease patients, designated K985 and K1088, with VHL germline variant c.407T>C (p.F136S) and c.256 C>T (p.P86S), respectively, were recruited to the TRACERx Renal study with patient's consent at the Barts Health NHS Trust, London, UK. The TRACERx Renal study is an ethically approved prospective cohort study (NCT03226886, National Health Service Research Ethics Committee approval 11/LO/1996) sponsored by the Royal Marsden NHS Foundation Trust and coordinated by its Renal Unit. Peripheral blood mononuclear cells (PBMCs) and renal surgery specimens, including normal kidney specimens, were collected from the recruited patients. Specimens were handled and stored according to standard protocols for downstream workflows.

Cell lines

The induced pluripotent stem cell (iPSC) line iFCI013 (also known as CRICKi010-A) was reprogrammed from the PBMCs of patient K985 at the Francis Crick Institute. The reprogramming steps and quality validation of iFCI013, including karyotyping and pluripotency marker expression, have been published[36]. iFCI013 cells were maintained in mTeSR Plus medium (STEMCELL Technologies) on cultureware coated with Cultrex ReadyBME (R&D Systems), at 37°C in a humidified incubator with 5% CO2.

The primary renal proximal tubule epithelial cells (PTECs) K1088N were derived from macroscopically normal kidney tissue obtained from patient K1088. The kidney tissue was processed into small fragments and digested with 2 mg/mL collagenase P (Roche, 249002001) at 37°C for 1 h on a shaking platform to release single cells. The digested suspension was filtered through a 40 µm cell strainer to remove undigested tissue fragments, followed by centrifugation to pellet the cells. The cell pellet was resuspended in cold growth factor-reduced Matrigel (Corning, 356231) and plated on pre-warmed cultureware as a 3D organoid culture. Organoids were maintained in kidney tubuloid medium (Advanced

DMEM/F12 supplemented with 1x GlutaMAX, 1x B27 supplement, 10 mM HEPES, 1 mM N-acetyl-L-cysteine, 5 μM A83-01, 50 ng/mL EGF, 100 ng/mL FGF-10, 0.1 mg/mL primocin, and 10% R-Spondin 1 conditioned medium) at 37°C in a humidified incubator with 5% CO2[37].

The RPE1-hTERT (RPE1) cell line, obtained from ATCC, was maintained in DMEM (Sigma, D6429) supplemented with 10% fetal bovine serum (FBS) and 100 U/mL Penicillin-Streptomycin, at 37°C in a humidified incubator with 5% CO2. The RPE1 cell line expressing mTurquoise-H2B for live cell imaging was generated and characterised in the previous publication[38].

The RPTEC-shP53 cell line (*TERT1* immortalised renal proximal tubule epithelial cells expressing a short hairpin RNA against TP53), was generated from the parental RPTEC/TERT1 cell line, obtained from ATCC, by retroviral transduction of pSRZ.p53shRNA retroviral construct (a gift from Dr. Jerry Shay's lab, UTSW) expressing a short hairpin RNA against TP53. The RPTEC-shP53 cells were maintained in renal epithelial cell growth basal medium (Lonza) supplemented with 0.5% tetracycline-free FBS, 100 U/mL Penicillin-Streptomycin, 10 ng/mL hEGF, 1x insulin-transferrin-selenium solution, 1 μg/mL hydrocortisone, 10 μM epinephrine, 50 ng/mL triiodo-l-thyronine, 30 μg/mL gentamicin and 15 ng/mL amphotericin B at 37°C in a humidified incubator with 5% CO2.

CRISPR-Taiji gRNA Design

The design and annotation steps of CRISPR-Taiji (CRISPRt) guide RNA (gRNA) spacer sequences are outlined in Extended Data Fig. 1a with gRNA spacer details summarised in Supplementary Table 1.

The complete "Telomere-to-Telomere" (T2T) reconstruction of a human genome with Y (T2T-CHM13v2.0) and centromere/satellite repeat annotations (Cen/Sat v2.0) from the T2T Consortium, were downloaded from github.com/marbl/CHM13[16-18]. For this study, Cen/Sat v2.0 annotation was used to select the active α-satellite higher-order-repeat region (active αSat HOR) and the entire (peri)centromere regions (pan-centromere). Additionally, CENPA

Cut&Run characterisation data for the T2T-CHM13v2.0 assembly, generated by the T2T Consortium, was used to select the CENPA-bound regions[17,18].

The crisprDesign R package was used to identify all Streptococcus pyogenes Cas9 (SpCas9) gRNA spacers within the CENPA-bound sequences for each chromosome[39]. All gRNA spacers were subsequently filtered based on the following criteria: GC-content between 35% and 80%, and a minimum of 20 canonical binding sites within the CENPA-bound region of at least one human chromosome. The GC-content range was selected based on previously published findings[40]. The CENPA-bound chromosome specificity was calculated as the number of binding sites in the CENPA-bound region of a specific chromosome divided by the total number of binding sites in the CENPA-bound regions across all chromosomes. A specificity threshold of ≥0.9 was applied, resulting in 1,830 chromosome-specific gRNAs across 24 human chromosomes. In addition, 15 bi-chromosome-specific gRNAs with binding sites on both Chr13 and Chr21 (i.e. not satisfying ≥0.9 CENPA-bound chromosome specificity) were included for efficiency improvement. Finally, 3 non-targeting gRNAs were included as negative control.

Further analyses were conducted to calculate the chromosome specificity of the CRISPRt gRNA set across the active α-sat HOR region and the pan-centromere region (similar to CENPA-bound chromosome specificity but across different centromeric regions). gRNAs with active αSat HOR or pan-centromere chromosome specificity <0.9 were flagged in red but not excluded (Supplementary Table 1). Additionally, the centromere specificity of the gRNAs was calculated as the total number of binding sites across the pan-centromere region divided by the total number of binding sites across the entire genome (T2T-CHM13v2.0). All selected CRISPRt gRNAs demonstrated a centromere specificity score >0.98, indicating limited on-target binding sites of CRISPRt gRNAs outside the centromeric region.

dCas9/HiFi-Cas9-gRNA RNP Electroporation

The single-guide RNAs (sgRNAs) used for CRISPRt screening were purchased from Genscript Biotech. The CRISPR RNAs (crRNAs, XT version) and trans-activating crRNA

(tracrRNA)-Atto550 conjugate used for immunofluorescence imaging were purchased from IDT. All synthetic RNAs are synthesised with chemical modifications at the first three 5' and 3' terminal RNA residues, which protect them from nucleases. All synthetic RNAs were resuspended in low-EDTA TE buffer to a stock concentration of 100 μM. High-fidelity Cas9 (Alt-R S.p.HiFi-Cas9 V3, 10 μg/μL) and catalytically dead Cas9 (Alt-R S.p. dCas9 V3, 10 μg/μL) were purchased from IDT.

For preparing CRISPR ribonucleoprotein (RNP) with sgRNAs for a 20 μL (or 100 μL) electroporation reaction, mix 1 μL (or 3 μL) of 100 μM sgRNA with 0.5 μL (or 1.5 μL) of dCas9 or HiFi-Cas9 in a PCR tube. For preparing RNP with crRNAs for a 20 μL electroporation reaction, mix 0.75 μL of 100 μM crRNA with 0.75 μL of 100 μM tracrRNA-Atto550 in a PCR tube; then anneal the crRNA and tracrRNA by incubating at 95°C for 5 min, followed by a gradual cool down to 25°C with a ramp rate of 0.2°C/s in a thermocycler; then mix the crRNA/tracrRNA complex with 0.5 μL dCas9. Incubate the mixed dCas9/HiFi-Cas9 and gRNA at room temperature for 30 min (in the dark if using crRNA-tracrRNA-Atto550).

The electroporation was performed with the Lonza P3 Primary Cell 4D-Nucleofector X kit S (20 μL reaction) or L (100 μL reaction), with Lonza 4D-Nucleofector X Unit according to the manufacturer's instructions. Briefly, single-cell suspension was prepared using Accutase treatment for iFCI013 cells, TrypLE treatment for RPE1 cells, and Dispase followed by Accutase treatment for K1088N organoids. For each 20 μL (or 100 μL) electroporation reaction, 200k (or 1000k) cells were transferred to a sterile tube and centrifuged. All the supernatant was carefully removed to avoid negative impacts on electroporation efficiency. The cell pellet was then resuspended in 20 μL (or 100 μL) of complete P3 electroporation buffer, prepared by mixing 16.4 μL (or 82 μL) of P3 Solution with 3.6 μL (or 18 μL) of Supplement 1. The cell suspension was gently mixed with the corresponding RNP complex by pipetting without introducing bubbles and then transferred to the electroporation well or cuvette. Electroporation pulses were immediately applied using the Lonza 4D-Nucleofector X Unit with pulse codes CA-137 (iFCI013 cells), EA-104 (RPE1 cells), or CM-138 (K1088N cells).

After electroporation, 135 μL (or 675 μL) of the corresponding culture medium described below was added to the electroporation well or cuvette with gentle mixing, and the cells were incubated for 5 min. If using sgRNA, the cell mixture was directly seeded into suitable cultureware after gentle mixing. If using rRNA/tracrRNA-Atto550, cells were centrifuged to remove all the supernatant containing the excess fluorescent RNP complex, before resuspending and seeding into the appropriate cultureware. Post-electroporation, iFCI013 cells were cultured on Cultrex ReadyBME-coated cultureware or glass coverslip with mTeSR Plus medium supplemented with CloneR2 (STEMCELL Technologies) to support single-cell survival; K1088N cells were cultured on Matrigel-coated cultureware or glass coverslip with the kidney tubuloid medium supplemented with 1% Matrigel as Matrigel-sandwiched 2D culture (reduced environmental hypoxia compared to 3D culture); RPE1 cells were cultured with DMEM medium with 10% FBS.

dCas9-gRNA RNP Lipofectamine Transfection

The lipofectamine CRISPRMAX transfection reagent (Invitrogen) was used to reverse-transfect the dCas9-sgRNA RNP complex into RPTEC-shp53 cells. Briefly, mix 1 μL of 100 μM sgRNA, 1.2 μL of dCas9, and 10 μL of Cas9 Plus reagent in 250 μL Opti-MEM in one tube, and mix 15 μL of CRISPRMAX reagent in 250 μL Opti-MEM in another tube. Mix the contents of the two tubes and incubate for 10 min at room temperature. Mix the lipofectamine-RNP mixture with 1.2 million RPTEC-shp53 cells resuspended in 2.5 mL medium before seeding into a 6 cm dish. Replace the medium after 24 h.

CRISPR-Taiji gRNA Screening and Validation

For each chromosome, the designed gRNAs with more canonical binding sites within the corresponding CENPA-bound region were prioritised for screening in batches until functional gRNAs were identified or all designed gRNAs were tested. The screened gRNAs were assigned with a CRISPRt ID, for example, ‘Chr1-sg1’ representing Chr1-specific sgRNA1.

For gRNA screening and validation, the dCas9-sgRNA RNP complex was introduced into iFCI013, RPE1, or K1088N cells using the 20 µL electroporation reaction system with the 16-well electroporation strips as described above. For each batch, the 3 non-targeting (NT) sgRNAs were included as the negative control. For each 16-well electroporation strip, a maximum of 15 candidate sgRNAs were tested and Chr3-sg2 was included as the positive control to check electroporation efficiency. Following electroporation, for each gRNA, 65k cells were seeded into 24-well plate with 13mm coverslips, while 135k cells were seeded into 12-well plate with the corresponding medium described above. Cells were cultured for 48 h before characterisation.

For micronucleated cell quantification, coverslips were fixed in 4% paraformaldehyde (PFA).. The fixed cells were then permeabilised and stained with PBS containing 0.2% Triton X-100 and 1 µg/mL DAPI for 5 min. DAPI images were captured using the Echo Revolve epifluorescence microscope with a 20x objective. The number of nuclei per image was quantified using CellProfiler software and the number of micronucleated cells was manually counted in a blinded manner to ensure unbiased assessment.

Fluorescence *in situ* Hybridisation (FISH)

For validating chromosome identity within the micronucleus, single-cell suspensions were generated using TrypLE treatment (important to make sure that the Cultrex/Matrigel coating is fully digested). The cell pellet was gently dissociated by tapping the tube, then fixed by slowly adding ice-cold Carnoy's fixative (methanol and acetic acid in 3:1 volume-to-volume ratio) dropwise while gently vortexing. Cells were incubated at room temperature for 20 min. After centrifugation to remove the fixative, the cells were subjected to a second round of fixation with the same procedure. The fixed cell pellet was resuspended in 50 µL of ice-cold Carnoy's fixative. Cells were subsequently dropped onto slides and air-dried for further processing.

For metaphase spread and chromosome painting, RPTEC-shP53 cells were treated with 120 ng/mL KaryoMAX colcemid for 6 h before collection. Cell pellets were gently resuspended

in 5 mL of 75 mM KCl solution dropwise while gently vortexing. Cells were incubated for 6 min in 37° C water bath and then fixed by slowly adding ice-cold Carnoy's fixative, followed by centrifugation and resuspension in Carnoy's fixative. Cells were subsequently dropped onto slides to form metaphase spread and air-dried for further processing.

For staining, the slides were then rehydrated in 2x Saline-Sodium Citrate (SSC) buffer for 2 min, followed by sequential dehydration in an ethanol series (70%, 85%, and 100%) for 2 min each. Next, the cell spots were hybridised with diluted centromere probes (OGT) or telomere probes (OGT), corresponding to the target chromosomes (Extended Data Fig. 1b), or with Chr3 chromosome painting probes (MetaSystems) for RPTEC-shP53 metaphase spread staining. Briefly, the probes were applied to the cell spot on the slide and covered with a 13mm coverslip, which was then sealed with rubber solution. Slides were incubated at 75°C for 2 min for denaturation, followed by overnight hybridisation in a humidified dark chamber at 37°C. The following day, coverslips were removed, and the slides were immersed without agitation in 0.4x SSC buffer at 72°C for 2 min, and then in 2x SSC buffer with 0.05% Tween-20 at room temperature for 30 s. The slides were counterstained with DAPI, air-dried, and mounted in antifade mounting solution.

Micronucleus FISH images were captured using the Echo Revolve epifluorescence microscope with a 20x objective. Micronuclei with or without the FISH signals of the corresponding target chromosome were manually quantified. Metaphase chromosome painting images were acquired on the Metafer Scanning and Imaging Platform microscope (MetaSystems) with a 63x objective. Metaphase images were analysed using the Isis Fluorescence Imaging Platform (MetaSystems, v.5.8.15).

Immunofluorescence

The dCas9-crRNA-tracrRNA-Atto550 RNP complex was introduced into iFCI013 cells by electroporation as described above. At 24 h post-electroporation, iFCI013 cells seeded on coverslips were fixed using ice-cold methanol for 30 min at -20 °C. After fixation, the cells were rehydrated in PBS for 5 min at room temperature and washed twice with PBS to remove

residual methanol. Blocking was performed with 5% milk-PBS for 1 h at room temperature, followed by two PBS washes to remove any remaining blocking solution. Primary antibodies were diluted in 2% BSA-PBS, and the cells were incubated with the primary antibody solution overnight at 4 °C. The primary antibodies used are Anti-Centromere Protein (CREST, 1:300, AntibodiesInc, 15-234), anti-HEC1 (1:300, ThermoFisher, MA1-23308), and anti-α-Tubulin (1:300, ThermoFisher, A11126). After primary antibody incubation, cells were washed three times in PBS on a shaking platform for 5 min each. Secondary antibody incubation was performed in 2% BSA-PBS for 1 h at room temperature. The secondary antibodies used are Anti-Mouse-IgG-Alexa-488 (1:500, Invitrogen, A332766) and Anti-Human-IgG-Alexa-647 (1:500, Invitrogen, A21445). Following secondary antibody incubation, the cells were washed three times in PBS for 5 min each. Coverslips were then mounted with antifade mounting medium containing DAPI. Z-stack fluorescence images were captured using a Zeiss Invert880 confocal system with Zeiss ZEN software with a 63x oil immersion lens.

ATAC-seq library preparation and sequencing

The dCas9-sgRNA RNP complex was introduced into iFCI013 cells by electroporation as described above, and experiments were performed with two independent biological replicates. Single cells were harvested using TrypLE treatment 8 h post-electroporation, allowing sufficient time for dCas9 binding with minimal mitotic progression. ATAC-seq library preparation followed the published ATAC-seq protocol[41], with all reagents and buffer recipes following the protocol unless otherwise specified. Briefly, 50k cells were used as input for each sample. Nuclei were prepared by incubating cells in ATAC-seq Lysis Buffer for exactly 3 min on ice. The nuclei were then incubated with the Transposition mix at 37 °C for 30 min in a thermomixer shaking at 1,000 rpm, generating transposed fragments. These fragments were purified and eluted in nuclease-free water using the DNA Clean & Concentrator-5 kit (Zymo). The purified fragments were then barcoded and amplified by PCR and the exact number of amplification cycles for each sample was determined using the NEBNext Library Quant Kit for Illumina, according to the manufacturer's protocol. The quality and fragment size of the final

libraries were assessed using the TapeStation 4200 (Agilent Technologies). Libraries were sequenced to 25 million PE-100bp target read pairs per sample using the Illumina NovaSeq 6000 at the Genomics Science Technology Platform at the Francis Crick Institute.

ATAC-seq analysis

The ATAC-seq library was first processed using the nf-core/atacseq pipeline (https://nf-co.re/atacseq/2.1.2) for quality checks (Supplementary Table 2)[42]. To assess the centromeric ATAC-seq signals, we adapted the previously published method for CENPA CUT&RUN data analysis with minor modifications[17]. Briefly, ATAC-seq raw reads were processed with TrimGalore (v0.6.1) to remove the adapter sequence. Trimmed reads were aligned to the T2T-CHM13v2.0 reference or a customised reference consisting of the CENPA-bound sequences used for gRNA design (T2T-CENPA) using BWA (v0.7.17) with the parameters: bwa mem -k 50 -c 1000000 [reference] [R1] [R2][17,43]. The resulting sam files were filtered using SAMtools (v1.13) with the command: samtools view -b -F 2308 [sam], which randomly assigns reads with multi-mapping sites to a single mapping location to prevent mapping biases in highly repetitive regions. PCR duplicates were marked and removed using Picard (v2.25.1). For visualisation, the processed bam files were converted to bigwig files using deepTools (v3.5.1) and normalised by the number of mapped reads.

To quantify chromosome-specific changes in ATAC-seq signals, we calculate the percentage of reads mapped to the entire chromosome (with T2T-CHM13v2.0 reference) or the CENPA-bound region (with T2T-CENPA reference) for each individual chromosome. Average log2 fold changes (log2FC) were then calculated by comparing to NT-sg1.

Time-lapse Live cell imaging

The dCas9-sgRNA RNP complex was introduced into RPE1 cells expressing mTurquoise-H2B by electroporation as described above. After electroporation, 30k cells per well were seeded into an 8-well chambered polymer coverslip (Ibidi, 80806) and allowed to settle for 16 h. Time-lapse live cell imaging was then performed using a Nikon Eclipse Ti inverted

microscope, equipped with a custom humidified enclosure (Okolabs) that maintains 37°C with 5% CO2. The Nikon Perfect Focus System (PFS) was used for auto-focus during image acquisition. Phase-contrast and fluorescent images were captured every 3 min for 48 h using ImageJ-mManager software with a 40x objective. Laser intensity and exposure time were optimised to prevent phototoxicity or photobleaching in the cells. Image processing was performed using FIJI software, and chromosome mis-segregation events were analysed and quantified manually. The mitotic event was first classified by the presence of micronucleus formation, and then mitotic events with micronucleus formation were further classified by the presence of chromosome arm split (Supplementary Video 1).

VHL Allele-specific Knockout

The gRNA spacer VHL-sg (5'- CATACCGAGTTGAAGCTGCC -3') was designed to specifically target the wild-type VHL allele while sparing the germline variant allele of patient K1088. This selective targeting was achieved by designing VHL-sg adjacent to the VHL germline variant of patient K1088, therefore, the SpCas9 protospacer-adjacent motif (PAM) site (5'-NGG-3'), essential for Cas9 binding, is altered from 5'-GGG-3' to 5'-GGA-3' by the germline variant, which prevents Cas9 from targeting the germline variant allele. The high-fidelity Cas9 (HiFi-Cas9) was used to further ensure the allele specificity. The HiFi-Cas9-sgRNA RNP complex was introduced into K1088N cells via electroporation as described. To confirm the successful introduction of frameshift mutations, Sanger sequencing was performed 11 days post-electroporation. The primer sequences used for PCR amplification of VHL exon 1 are VHLe1-FW (5'- TGGTCTGGATCGCGGAG -3') and VHLe1-RV (5'-CTTCAGACCGTGCTATCGTC -3').

Flow Cytometry and Cell Sorting

The dCas9-sgRNA RNP complex was introduced into K1088N cells by electroporation as described above. Cells were initially cultured for 7 days to confluence before splitting with 1:6 ratio, and the cells were further cultured for 4 days before harvesting for flow sorting. Single-

cell suspension was generated with Accutase treatment, and the cell pellet was washed with PBS once. Cells were stained with 1:50 diluted anti-CA9-PE antibody (Miltenyi, 130-131-321) in flow sorting buffer (1% bovine serum albumin (BSA) and 1μM EDTA in PBS) for 30 min at 4 °C in the dark. The cell pellet was washed with flow sorting buffer twice before resuspended in 250 μl flow sorting buffer for every 1 million cells supplemented with 1 μg/mL DAPI. Cell sorting was performed on BD FACSAria Fusion cell sorter with BD FACSDiva software. CA9 and DAPI gate thresholds were determined with the non-targeting dCas9-NT-sg1 (CTRL) sample. Singlet cells that are CA9 positive and DAPI negative for HiFi-Cas9-VHL-sg (*VHL*-ASK) or dCas9-Chr3-sg2 (Chr3-Ct) samples were sorted with 'Yield' precision mode scRNA-seq characterisation.

scRNA-seq Library Preparation and Sequencing

Up to 1 million singlet cells from each sample were fixed using the Parse Bioscience Evercode Fixation v3 kit, then cryopreserved as aliquots at -80 °C following the manufacturer's protocol. After fixation, single-cell barcoded cDNA libraries were prepared using the Parse Bioscience Evercode WT Mini v3 kit, targeting 1,650 cells per sample according to the manufacturer's instructions. The quality and fragment size of the final library (consisting of two sub-libraries) were assessed using the TapeStation 4200 (Agilent Technologies). The library was sequenced to an average depth of 60k PE-100bp read pairs per cell using the Illumina NovaSeq 6000 at the Genomics Science Technology Platform at the Francis Crick Institute.

Parse scRNA-seq Analysis

For Parse scRNA-seq, raw FASTQ reads were processed using the proprietary Parse Bioscience pipeline (v1.2.1) with default parameters to generate the filtered raw cell-gene count matrix. The filtered matrix was then used to generate the Seurat objects excluding features in fewer than 3 cells and cells with fewer than 200 features using the Seurat package (v5.0.0)[44]. For the filtering steps, cells with more than 20% mitochondrial reads were excluded as dead cells, and cells with feature counts more than the mean plus 2x standard deviation

were excluded as potential doublets. In addition, "MALAT1" and mitochondrial genes were removed for the downstream analyses. Normalization and SCTransform were applied, followed by UMAP dimensionality reduction and clustering using the Seurat package[44].

To calculate the single-cell copy number profiles, scRNA-seq gene expressions were used to infer the chromosome copy numbers using the InferCNV package (v1.21.0)[45] with the NT-sg1 (CTRL) as the reference. Average residual gene expressions for the chromosomes (or chromosome arms) of interest were calculated. Thresholds for defining copy number loss or gain were set at the average residual gene expressions where 1% of CTRL cells were classified as loss or gain, respectively. Cells between these two thresholds were classified as copy number neutral.

To assign K1088N cells with *VHL* loss status after Chr3-Ct or *VHL*-ASK, a hypoxia-inducible factor (HIF) pathway signature score was calculated with a curated HIF pathway activation gene set[46] using the Seurat package[44]. A threshold was set at the score where 1% of CTRL cells were classified as HIF pathway signature high (HIF-High). Cells with scores above this threshold were classified as HIF-High and *VHL* loss (*VHL-*). Cells with scores below this threshold were classified as HIF pathway signature low (HIF-Low) and *VHL* proficient (*VHL*+, with functional *VHL* gene copy).

For gene set enrichment analysis (GSEA) in K1088N samples, cells were classified by engineering targets (Chr3-Ct or *VHL*-ASK), Chr3(p) status (Neutral, Loss or Gain based on the inferred copy number) and inferred *VHL* status (*VHL*+ or *VHL-* based on the HIF pathway signature scores). Differentially expressed genes (DEG) for the classified cell groups compared to the corresponding unaltered group were identified using the FindMarkers function of the Seurat package[44]. GSEA was performed with the DEG analysis output for Human MSigDB hallmark gene sets using the fgsea package (v1.24.0)[47-49]. Gene sets with an adjusted p-value < 0.01 in at least one classified group were selected for visualisation. The MHC Class I (*B2M*, *HLA-A*, *HLA-B*, *HLA-C*, *HLA-E*, and *HLA-G*) gene signature score was calculated using the Seurat package[44].

Statistics and reproducibility

The number of biological replicates is indicated in the corresponding legends and the standard error of mean (SEM) is plotted as the error bar. The p-values for GSEA analysis were calculated and adjusted for multiple testing using the Benjamini-Hochberg procedure as the default setting of the fgsea package. The p-values for comparing signature scores were calculated with a two-tailed Mann-Whitney test.

Materials availability

All the reagents used in this article are publicly available. The cell models used in the article are available from the corresponding authors upon request, subject to a legal agreement.

Code availability

The codes with software versions used for CRISPRt gRNA design, ATAC-seq analyses, and scRNA-seq analyses are available at https://github.com/FrancisCrickInstitute/CRISPR-Taiji.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data Availability

The scRNA-seq and ATAC-seq data have been deposited in the Gene Expression Omnibus (GEO) with accession GSE278125. The published ccRCC scRNA-seq datasets used in this article are available according to the original publication.

Reference

1 Nagaoka, S. I., Hassold, T. J. & Hunt, P. A. Human aneuploidy: mechanisms and new insights into an age-old problem. *Nat Rev Genet* **13**, 493-504 (2012). https://doi.org:10.1038/nrg3245

2 Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899-905 (2010). https://doi.org:10.1038/nature08822
3 Gao, T. *et al.* A pan-tissue survey of mosaic chromosomal alterations in 948 individuals. *Nature Genetics* **55**, 1901-1911 (2023). https://doi.org:10.1038/s41588-023-01537-1
4 Li, R. *et al.* A body map of somatic mutagenesis in morphologically normal human tissues. *Nature* **597**, 398-403 (2021). https://doi.org:10.1038/s41586-021-03836-1
5 Lin, Y. *et al.* Normal breast tissues harbour rare populations of aneuploid epithelial cells. *Nature* (2024). https://doi.org:10.1038/s41586-024-08129-x
6 Creighton, C. J. *et al.* Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-49 (2013). https://doi.org:10.1038/nature12222
7 Taylor, A. M. *et al.* Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **33**, 676-689.e673 (2018). https://doi.org:10.1016/j.ccell.2018.03.007
8 Lakhani, A. A., Thompson, S. L. & Sheltzer, J. M. Aneuploidy in human cancer: new tools and perspectives. *Trends Genet* **39**, 968-980 (2023). https://doi.org:10.1016/j.tig.2023.09.002
9 Watson, E. V. *et al.* Chromosome evolution screens recapitulate tissue-specific tumor aneuploidy patterns. *Nat Genet* **56**, 900-912 (2024). https://doi.org:10.1038/s41588-024-01665-2
10 Tovini, L. *et al.* Targeted assembly of ectopic kinetochores to induce chromosome-specific segmental aneuploidies. *Embo j* **42**, e111587 (2023). https://doi.org:10.15252/embj.2022111587
11 Truong, M. A. *et al.* A kinesin-based approach for inducing chromosome-specific mis-segregation in human cells. *Embo j* **42**, e111559 (2023). https://doi.org:10.15252/embj.2022111559
12 Ly, P. *et al.* Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nat Cell Biol* **19**, 68-75 (2017). https://doi.org:10.1038/ncb3450
13 Fournier, R. E. & Ruddle, F. H. Microcell-mediated transfer of murine chromosomes into mouse, Chinese hamster, and human somatic cells. *Proc Natl Acad Sci U S A* **74**, 319-323 (1977). https://doi.org:10.1073/pnas.74.1.319
14 Tsuchida, C. A. *et al.* Mitigation of chromosome loss in clinical CRISPR-Cas9-engineered T cells. *Cell* **186**, 4567-4582.e4520 (2023). https://doi.org:10.1016/j.cell.2023.08.041
15 Girish, V. *et al.* Oncogene-like addiction to aneuploidy in human cancers. *Science* **381**, eadg4521 (2023). https://doi.org:10.1126/science.adg4521
16 Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022). https://doi.org:10.1126/science.abj6987
17 Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022). https://doi.org:10.1126/science.abl4178
18 Rhie, A. *et al.* The complete sequence of a human Y chromosome. *Nature* **621**, 344-354 (2023). https://doi.org:10.1038/s41586-023-06457-y
19 McKinley, K. L. & Cheeseman, I. M. The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* **17**, 16-29 (2016). https://doi.org:10.1038/nrm.2015.5

20 Bosco, N. *et al.* KaryoCreate: A CRISPR-based technology to study chromosome-specific aneuploidy by targeting human centromeres. *Cell* **186**, 1985-2001.e1919 (2023). https://doi.org:10.1016/j.cell.2023.03.029
21 Wang, H. *et al.* CRISPR-mediated live imaging of genome editing and transcription. *Science* **365**, 1301-1305 (2019). https://doi.org:10.1126/science.aax7852
22 Lera, R. F. *et al.* Plk1 protects kinetochore-centromere architecture against microtubule pulling forces. *EMBO Rep* **20**, e48711 (2019). https://doi.org:10.15252/embr.201948711
23 Addis Jones, O., Tiwari, A., Olukoga, T., Herbert, A. & Chan, K. L. PLK1 facilitates chromosome biorientation by suppressing centromere disintegration driven by BLM-mediated unwinding and spindle pulling. *Nat Commun* **10**, 2861 (2019). https://doi.org:10.1038/s41467-019-10938-y
24 Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53-58 (2012). https://doi.org:10.1038/nature10802
25 Omori, S. *et al.* A Pair of Maternal Chromosomes Derived from Meiotic Nondisjunction in Trisomy 21 Affects Nuclear Architecture and Transcriptional Regulation. *Sci Rep* **7**, 764 (2017). https://doi.org:10.1038/s41598-017-00714-7
26 Hirota, T. *et al.* Fertile offspring from sterile sex chromosome trisomic mice. *Science* **357**, 932-935 (2017). https://doi.org:10.1126/science.aam9046
27 Young, M. D. *et al.* Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599 (2018). https://doi.org:10.1126/science.aat1699
28 Jaakkola, P. *et al.* Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292**, 468-472 (2001). https://doi.org:10.1126/science.1059796
29 Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* **173**, 595-610.e511 (2018). https://doi.org:10.1016/j.cell.2018.03.043
30 Latif, F. *et al.* Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* **260**, 1317-1320 (1993). https://doi.org:10.1126/science.8493574
31 Shepherd, S. *et al.* Abstract 133: Rewinding the tape of life: somatic profiling in Von Hippel-Lindau (VHL) disease reveals evolutionary contingency and constraints. *Cancer Research* **83**, 133-133 (2023). https://doi.org:10.1158/1538-7445.AM2023-133
32 Wykoff, C. C. *et al.* Hypoxia-inducible expression of tumor-associated carbonic anhydrases. *Cancer Res* **60**, 7075-7083 (2000). https://doi.org:Not-Available
33 Logsdon, G. A. *et al.* The variation and evolution of complete human centromeres. *Nature* **629**, 136-145 (2024). https://doi.org:10.1038/s41586-024-07278-3
34 Volpe, E. *et al.* The complete diploid reference genome of RPE-1 identifies human phased epigenetic landscapes. *bioRxiv* (2023). https://doi.org:10.1101/2023.11.01.565049
35 Ly, P. *et al.* Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nat Genet* **51**, 705-715 (2019). https://doi.org:10.1038/s41588-019-0360-8
36 Devito, L. G. *et al.* Generation of TWO iPSC lines (CRICKi009-A; CRICKi010-A) from patients with type 1 von Hippel-Lindau (VHL) and histopathologically

confirmed renal cell carcinoma (RCC). *Stem Cell Research* **81**, 103611 (2024). https://doi.org:10.1016/j.scr.2024.103611
37 Schutgens, F. *et al.* Tubuloids derived from human adult kidney and urine for personalized disease modeling. *Nat Biotechnol* **37**, 303-313 (2019). https://doi.org:10.1038/s41587-019-0048-8
38 Zeng, J., Hills, S. A., Ozono, E. & Diffley, J. F. X. Cyclin E-induced replicative stress drives p53-dependent whole-genome duplication. *Cell* **186**, 528-542.e514 (2023). https://doi.org:10.1016/j.cell.2022.12.036
39 Hoberecht, L., Perampalam, P., Lun, A. & Fortin, J. P. A comprehensive Bioconductor ecosystem for the design of CRISPR guide RNAs across nucleases and technologies. *Nat Commun* **13**, 6568 (2022). https://doi.org:10.1038/s41467-022-34320-7
40 Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nature Biotechnology* **32**, 1262-1267 (2014). https://doi.org:10.1038/nbt.3026
41 Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-seq. *Nat Protoc* **17**, 1518-1552 (2022). https://doi.org:10.1038/s41596-022-00692-9
42 Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**, 276-278 (2020). https://doi.org:10.1038/s41587-020-0439-x
43 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* **1303** (2013). https://doi.org:10.48550/arXiv.1303.3997
44 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015). https://doi.org:10.1038/nbt.3192
45 Tickle T, T. I., Georgescu C, Brown M, Haas B (Klarman Cell Observatory, Broad Institute of MIT and Harvard, 2019).
46 Lombardi, O. *et al.* Pan-cancer analysis of tissue and single-cell HIF-pathway activation using a conserved gene signature. *Cell Rep* **41**, 111652 (2022). https://doi.org:10.1016/j.celrep.2022.111652
47 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org:10.1073/pnas.0506580102
48 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org:10.1016/j.cels.2015.12.004
49 Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*, 060012 (2019). https://doi.org:10.1101/060012

Acknowledgement

We are grateful to the patients and their families who generously participated in the TRACERx Renal Study. We thank Dr Frank Uhlmann, Dr Greg Findlay, Dr James Turner, Dr Michael Howell, and Dr Maxim Molodtsov from the Francis Crick Institute, London, UK, as well as Dr Francisco Barriga from the Vall d'Hebron Institute of Oncology for the fruitful discussion and

valuable feedback to drive the project forward. We thank We thank Dr Jerry Shay from UTSW for providing pSRZ.p53shRNA construct. We thank Ms. Laila Parvanta and Ms. Chetna Varsani from Barts Health NHS Trust, London for their support in recruiting *VHL* patients and providing surgical specimens. We thank support from the Science Technology Platforms at the Francis Crick Institute, London, UK, especially Genomics Science Technology Platform, Flow Cytometry Facility, and Light Microscopy Facility. BioRender icons were used in figure preparation under the Francis Crick Institute's subscription license. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK, the UK Medical Research Council and the Wellcome Trust (CC2044, S.T.), by Cancer Research UK (C50947/A29911, S.T.), by VHL Alliance (Research Grant 2023, S.T., H.F., D.D., and S.T.C.S.), by US Department of Defense (W81XWH2210764, P.L.), and by Beijing Natural Science Foundation (7242083, L.W.).

Author contributions

Conceptualisation: S.T. and H.F.

Methodology: H.F., L.W., H.J., J.A.D., S.E.M., J.F.X.D., P.L. and S.T.

Investigation: H.F., D.D., R.D., L.W., J.Z., F.B., S.T.C.S., J.W., S.C.J., A.F., M.N., Y.Z., R.W., W.A., O.B., Y.D., Yih. X., Yim. X., Z.S., W.S. and H.W.

Resources: S.T.C.S., A.L.C., Z.T., T.B., L.G.D., L.H., E.S.L., S.M.O., S.A. and W.M.D.

Data Curation: H.F., D.D., R.D., L.W., J.Z., B.J.Y.T. and Yim. X.

Formal Analysis: H.F., R.D., B.J.Y.T., A.H. and K.A.L.

Manuscript Preparation: H.F. and S.T. with contributions from all the authors

Competing interests

S.T. has received speaking fees from Roche, Astra Zeneca, Novartis, and Ipsen. S.T. has the following patents filed: Indel mutations as a therapeutic target and predictive biomarker PCTGB2018/051892, PCTGB2018/051893, and P113326GB.