bioRχiv

Image-based identification and isolation of micronucleated cells to dissect cellular consequences

Lucian DiPeso,[a,d] Sriram Pendyala,[c] Heather Z. Huang,[a] Douglas M. Fowler,[c] Emily M. Hatch[a,b*]
[a]Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
[b]Human Biology Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
[c]Genome Sciences, University of Washington, Seattle, WA, USA
[d]Molecular & Cellular Biology, University of Washington, Seattle, WA, USA
*Correspondence should be addressed to E.M.H. (ehatch@fredhutch.org)

Recent advances in isolating cells based on visual phenotypes have transformed our ability to identify the mechanisms and consequences of complex traits. Micronucleus (MN) formation is a frequent outcome of genome instability, triggers extensive disease-associated changes in genome structure and signaling coincident with MN rupture, and is almost exclusively defined by visual analysis. Automated MN detection in microscopy images has proved extremely challenging, limiting unbiased discovery of the mechanisms and consequences of MN formation and rupture. In this study we describe two new MN segmentation modules: a rapid and precise model for classifying micronucleated cells and their rupture status (VCS MN), and a robust model for accurate MN segmentation (MNFinder) from a broad range of microscopy images. As a proof-of-concept, we define the transcriptome of non-transformed human cells with intact or ruptured MN after inducing chromosome missegregation by combining VCS MN with photoactivation-based cell isolation and RNASeq. Surprisingly, we find that neither MN formation nor rupture triggers a unique transcriptional response. Instead, transcriptional changes are correlated with increased aneuploidy in these cell classes. Our MN segmentation modules overcome a significant challenge to reproducible MN quantification, and, joined with visual cell sorting, enable the application of powerful functional genomics assays, including pooled CRISPR screens and time-resolved analyses of cellular and genetic consequences, to a wide-range of questions in MN biology.

Introduction

Recent advances in automated image analysis have led to the development of high-throughput platforms to isolate specific cell classes and match visual phenotypes to specific genetic and expression profiles. These platforms bring the power of pooled genetic screening and population-based analyses to a huge range of phenotypes that are defined solely by visual changes in subcellular features. One such feature are micronuclei (MN), nuclear compartments containing a few chromosomes or chromatin fragments that result from mitotic segregation errors and persistent DNA damage (Bona and Bakhoum, 2024; Guo et al., 2019). Increased MN frequency is a hallmark of carcinogen exposure, cancer development, and aging, and MN are potent drivers of massive genome structure changes, pro-inflammation and metastasis signaling, and senescence (Bakhoum et al., 2018; Dou et al., 2017; Harding et al., 2017; He et al., 2019; Mackenzie et al., 2017; Mohr et al., 2021; Soto et al., 2018; Zhang et al., 2015). These processes are linked to MN rupture, which exposes the chromatin to the cytosol for the duration of interphase (Hatch et al., 2013) and may contribute to tumorigenesis, metastasis, aging, and inflammatory disorders (Bona and Bakhoum, 2024; Guo et al., 2019).

Most studies on the biology and consequences of MN formation and rupture take advantage of the fact that MN can be induced with high frequency in cultured cells, for instance by inhibiting the spindle assembly checkpoint kinase Mps1 (Krupina et al., 2021). However, these interventions cause diverse "off-target" nuclear and cellular changes, including chromatin bridges, aneuploidy, and DNA damage. Several sophisticated techniques have been developed to overcome this challenge that enrich or isolate MN or micronucleated cells, including live-cell imaging of single cell arrays, inducing Y chromosome missegregation by disrupting the centromere, and purifying MN from lysed cells by flow cytometry, which have led to new insights into MN rupture, function, and consequences (Agustinus et al., 2023; Ly et al., 2016; Mohr et al., 2021; Papathanasiou et al., 2023; Zhang et al., 2015). However, all have significant limitations for unbiased analysis of the cellular consequences of MN formation and rupture and lack features necessary to perform high-throughput analyses on micronucleated cells. For these studies, what is needed is a way to visually identify micronucleated cells within a larger population rapidly and robustly and target them for downstream analysis.

Automated detection of MN in microscopy images using conventional morphological transformations is challenging due to the diversity of MN shapes and sizes, their similarity to nuclear features that co-occur at high rates, including nuclear blebs and chromatin bridges, and their frequent proximity to nuclei. To address this, we developed two image analysis pipelines that combine neural network-based pixel classification with pre- and post-processing steps to rapidly identify micronucleated cells from low resolution images (VCS MN) or segment MN with high recall across multiple cell lines, chromatin labels, and imaging conditions (MNFinder). We demonstrate the utility of this approach by combining VCS MN with a phenotype-based cell isolation method, called Visual Cell Sorting (VCS), to define the transcriptomic profile of hTERT-RPE1 cells with none, intact, or ruptured MN by RNAseq after inducing chromosome

missegregation. During VCS, single cells expressing nuclear-localized Dendra2 are photoconverted on-demand based on the results of the MN classifier. Specific cell classes are then isolated by gating on Dendra fluorescence ratios during FACS (Hasle et al., 2020). We show that we can recapitulate an established aneuploidy signature using this method and find that, surprisingly, neither micronucleation nor rupture is sufficient to induce substantial transcriptional changes in these conditions. We envision that the MN segmentation and cell isolation platforms described here will be widely applied to fundamental questions in cell division and nucleus biology and to cell-based models of human disease to enable new discoveries into the contribution of MN to cellular dysfunction.

Results

Machine vision identifies micronucleated cells within a mixed population

We initially developed an automated pipeline to identify micronucleated cells in a mixed population on hTERT RPE-1 (RPE1) cells. RPE1 cells are a near-diploid, non-transformed human cell line with very low frequencies of spontaneous MN that have been extensively used for studies of chromosome missegregation and micronucleation (He et al., 2019; Kneissig et al., 2019; Mammel et al., 2021; Santaguida et al., 2017; Zhang et al., 2015). RPE1 cells also do not activate the cGAS-STING innate immune pathway upon micronucleation (Chen et al., 2020), similar to many cancer cells (Kwon and Bakhoum, 2020; Stetson et al., 2008). We anticipated that this would limit inflammatory signaling and increase the sensitivity of downstream analyses of the consequences of MN formation and rupture (Bakhoum et al., 2018; He et al., 2018; Santaguida et al., 2017). To enable automated image analysis and live-cell marking, we co-expressed a fluorescent chromatin marker (H2B-emiRFP703) to identify nuclei and MN with 3xDendra2-NLS (nuclear localization signal) to identify ruptured MN and photoactivate selected cells (Hasle et al., 2020; Hatch et al., 2013; Matlashov et al., 2020) (Fig. 1A). We treated these cells, referred to as RFP703/Dendra, with a low dose of an Mps1 inhibitor (Mps1i) to induce MN, which increased MN frequency to 50% of cells with 1 MN per cell on average (Fig. S1A-B).

To distinguish MN from morphologically similar nuclear features also induced by Mps1i, including chromatin bridges and nuclear blebs (He et al., 2018; Maciejowski et al., 2015), we trained a neural net classifier on low resolution single channel images of RFP703/Dendra after Mps1i incubation. For this first model, called VCS MN, we opted for low resolution training images, increased speed of analysis, and positive predictive value (PPV) privileged over recall to optimize downstream classifier integration into a visual cell sorting platform. To train the model, H2B channel images were passed to a Deep Retina neural net (Caicedo et al., 2019a) to generate nuclear masks, which were then used to crop the field into single cell images, excluding cells on the image edges. On average, 75% of nuclei per field were correctly segmented and cropped. A U-Net classifier using pyTorch's ResNet18 pre-trained model as its base architecture (Ronneberger et al., 2015) was then trained on 2,000 single cell crops combining the H2B image with the results of Sobel edge detection. Classified MN pixels were converted to a mask that was mapped back onto the whole field image (Fig. 1B). MN segments were then assigned to "parent" nuclei by proximity, which correctly associated 97% of MN (Fig. S1C). Nuclei associated with at least one MN were labeled as MN+ cells and those associated with no MN were labeled as MN- cells. We validated the ability of VCS MN to classify cells on 6 whole field images from two experiments and calculated recall values of 86% and 65%, respectively, for MN- and MN+ labeled cells, and PPVs of 73% and 93% (Fig. 1C). Analysis of MN classification on cropped images found a recall value of approximately 70% and a PPV of 89% for MN identification (Fig. 1D), indicating that we successfully limited false positives in our MN+ cell class with the tradeoff of decreased purity of the MN- cell pool. We also observed a small, but statistically significant, reduction in recall for ruptured MN (Fig. 1D), likely driven by their smaller size.

To determine whether this pipeline could achieve similar accuracy in another cell line, we retrained the VCS MN classifier on images of multiple cell lines acquired at two magnifications (see Methods) and assessed prediction quality on low-resolution images of U2OS cells induced to form MN. We calculated recall values of 82% and 83% in MN- and MN+ cells, respectively, PPVs of 81% and 83%, and an MN PPV of 88% (Fig. 1E). In summary, VCS MN can automatically identify the majority of micronucleated and non-micronucleated cells in low-resolution images of cells from multiple sources containing a mix of contaminating objects with high precision.

RPE1 cells with ruptured MN were further classified from the MN+ population based on MN Dendra2 intensity: NLS-3xDendra2 is present in intact MN and absent in ruptured MN (Fig. 1F). We quantified the maximum Dendra2 intensity in the nucleus and corresponding MN segments and MN with a signal less than 0.16 of the nucleus was classified as "ruptured." This threshold correctly classified approximately 90% of MN (Fig. S1D). When appended to the VCS MN pipeline, this analysis correctly identified 60% of rupture- cells (cells with only intact MN) and 70% of rupture+ cells (cells with at least 1 ruptured MN) with a PPV near 75% in both cases (Fig. 1G). The difference in recall is likely due to the increased probability of a multi-micronucleated cell being correctly classified as MN+ and having at least 1 ruptured MN (Fig. S1E-F).

MNFinder accurately segments MN in images of attached cells

Due to the analysis constraints we imposed, MN classified by the VCS MN module were typically undersegmented (Fig. 2A) and performance diminished substantially on images taken using different magnifications or with different chromatin labeling agents. Therefore, we developed a new module, called MNFinder, that privileges accurate MN and nuclear segmentation across cell types, DNA labels, and image resolution over PPV. MNFinder takes a single channel chromatin image as input and integrates the results of two independent nucleus/MN (nuc/MN) and cell segmenter pipelines to generate three object group: 1) MN, 2) cells, which group nuclei with associated MN, and 3) nuclei. Both the nuc/MN and cell pipelines use a UNet classifier for initial

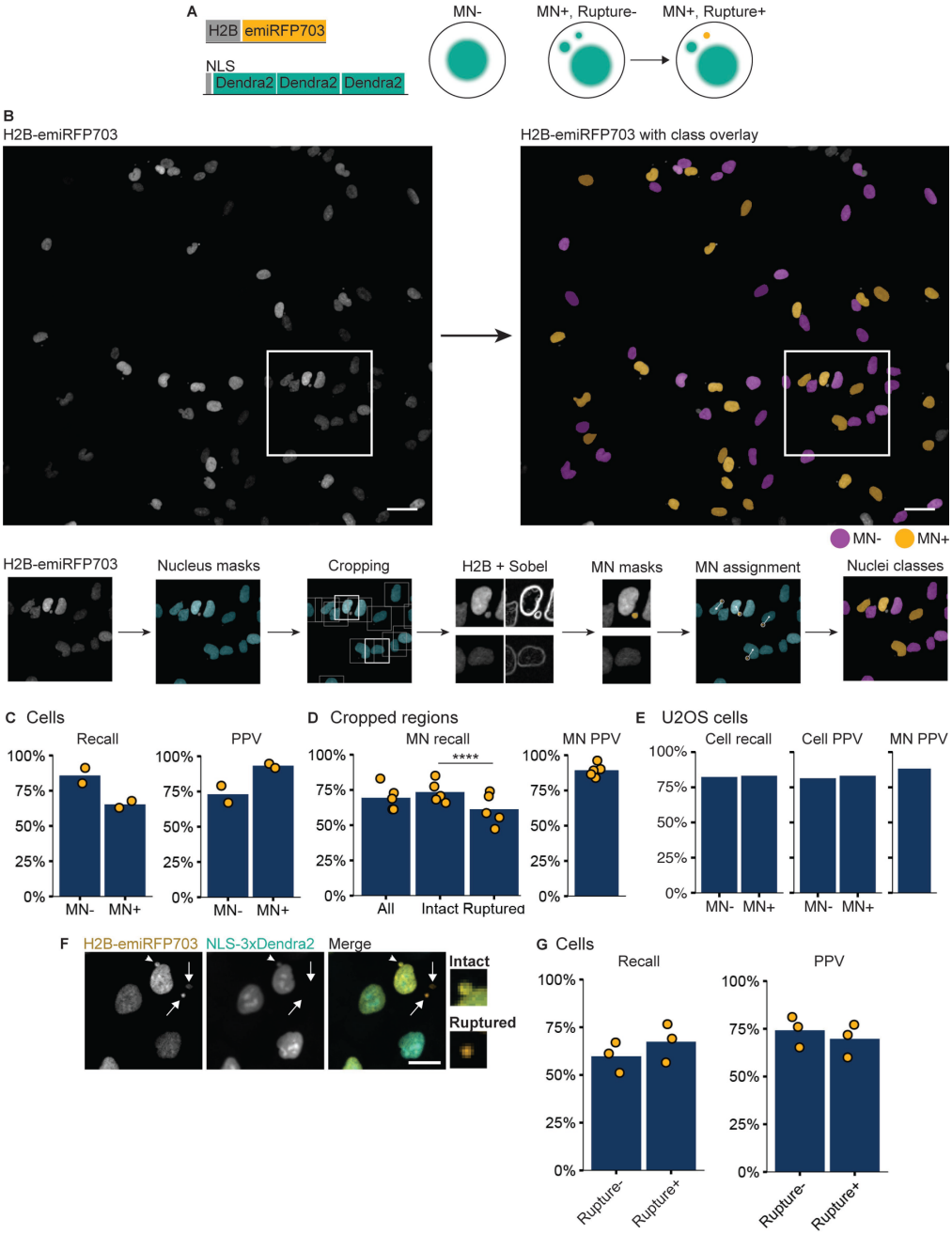

Figure 1: VCS MN neural net module identifies micronucleated cells. (A) Diagrams of constructs transduced into RFP703/Dendra cells and how these constructs localize in micronucleated cells before and following MN rupture. (B) Image of H2B-emiRFP703 in RFP703/Dendra cells at 20x and overlaid with MN+ and MN- cell classification results. Bottom: visual depiction of classification pipeline. (C) Recall and positive predictive value (PPV) of MN- and MN+ cells. N=2, n=328, 186. (D) Recall and PPV of MN classification within image crops. MN were manually scored as intact or ruptured by Dendra2 signal. N=5, n=264, 158, 365, 283, 249. (E) Same as for (d), except analyzed images are of U2OS cells expressing H2B-emiRFP703 and NLS-3xDendra2. N = 1, n = 85, 95. (F) RFP703/Dendra cells with ruptured (arrows) and intact (arrowhead) MN showing a loss of NLS-3xDendra signal in ruptured MN. Scale bar = 10 μm. (G) Recall and PPV for rupture-based cell classification. N=3, n=120, 91, 82.

segmentation followed by additional image processing steps to refine the results (Fig. 2B).

We made several changes to the neural net input and architecture to design MNFinder. To adjust for highly imbalanced data sets, we incorporated attention gates in the upsampling blocks and employed focal loss during training (Lin et al., 2018). We also modified the classifier to segment both nuclei and MN to better discriminate between nucleus-associated and MN-associated pixels, increased the size of input image tiles from 48x48 px to 128x128 px to increase contextual information, and oversampled input images by 25%. Final classification results integrate the predictions from all crops containing a given object.

For nucleus/MN segmentation, we tested a variety of UNet-derived architectures (Table 1) and found that a basic UNet with attention gates performed well across multiple cell lines (Oktay et al., 2018; Su et al., 2021). We also observed incorporating multiscale downsample blocks identified some MN otherwise missed, but produced an overall reduction in performance. Therefore we developed an ensemble classifier (Fig. S2A) that takes predicted MN weights from both UNet types as inputs to generate the final MN predictions. Nucleus predictions are retained from the basic attention gate UNet. This classifier was trained on images of live RPE1 and U2OS cells expressing H2B-emiRFP703 or fixed RPE1 cells, HeLa cells, and hTERT human fetal fibroblasts (HFF) labeled with DAPI using 128x128 random crops with image augmentation. To adjust for misclassification of large MN as nuclei, nuclei with an area below 250 px are automatically reclassified and undersegmentation of MN is further limited by expanding MN object boundaries to their convex hull (Fig. S2B).

Cell identification is not possible with standard UNets. To overcome this limitation and improve nucleus segmentation, we developed a UNet and image processing pipeline that outputs "cell" masks, defined as the concave hull of each nucleus and its associated MN, based on predicted distance and proximity maps (Fig. S2C). Map predictions are derived from single channel image crops using a multi-decoder UNet with a UNet3+ architecture in the two main decoders (Huang et al., 2020; Mahbod et al., 2022). To improve accuracy and decrease training time, we added a third arm to the UNet that classifies foreground pixels and feeds these predictions into the distance and proximity decoders (Fig. S2D). Given the complexity of this UNet, we used a constant feature depth at every level in the encoder and decoder pathways and

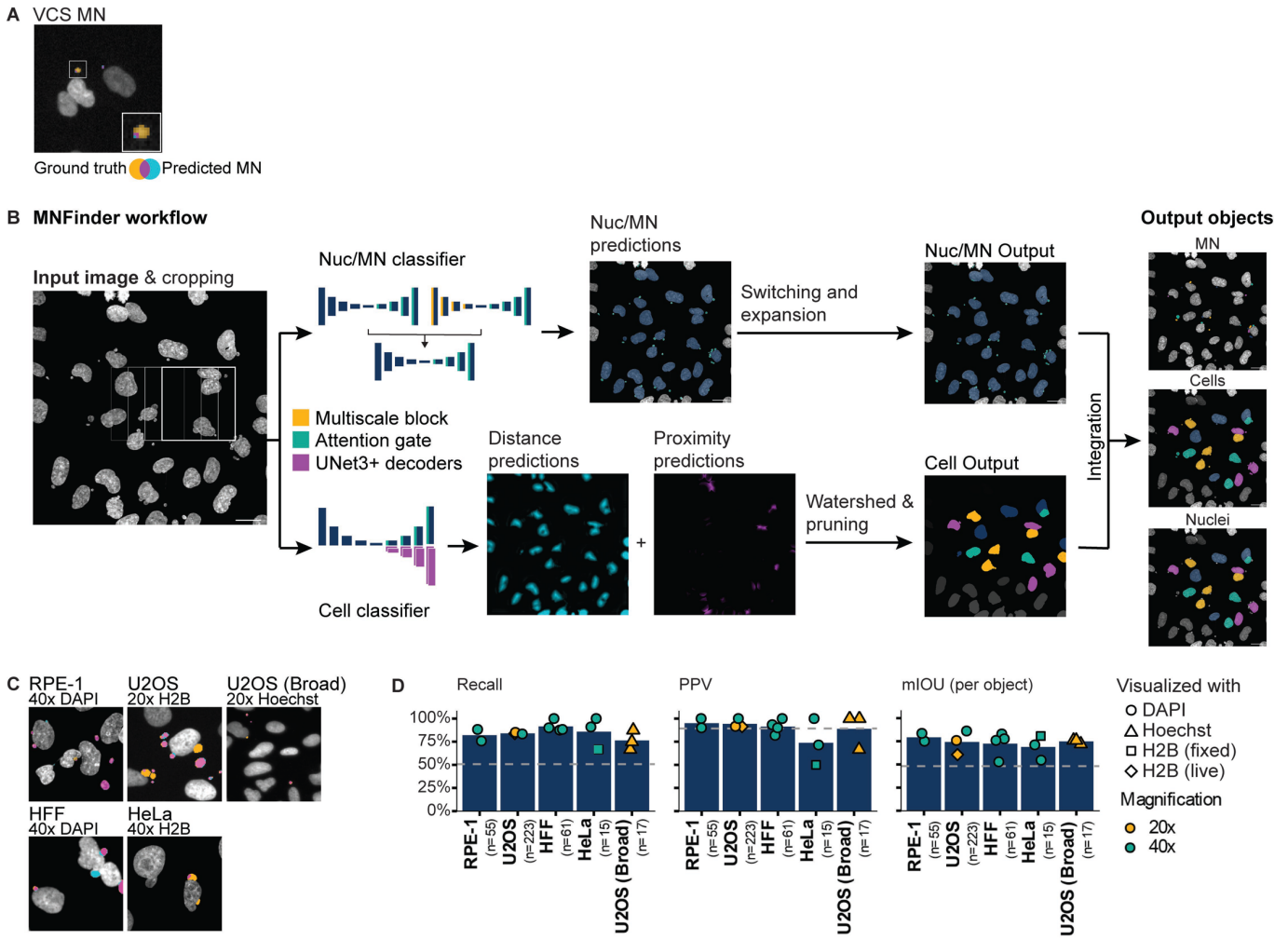

Figure 2: MNFinder module robustly segments MN across cell types and imaging conditions. (A) Representative image showing undersegmentation by the VCS MN neural net. (B) Overview of MNFinder module for classifying and segmenting MN and nuclei. Images are tiled by a sliding window and processed by 2 neural nets in parallel: one for classifying regions as nuclei or MN (Nuc/MN) and one for classifying cells. Nuc/MN classifier results are post-processed to correct MN that were misclassified as small nuclei and to expand MN masks. Cell classifier gradient map outputs are used to define cells through watershed-based post-processing. Nuc/MN results are then integrated with cell results to produce final labels of individual MN, cells, and MN. Image crops are reassembled onto the final image by linear blending. C-D) Example images and MN pixel predictions using MNFinder on multiple cell types (RPE1 H2B-emiRFP703, U2OS, HeLa H2B-GFP, and HFF), chromatin labels (DAPI, Hoechst, H2B-FP), and magnifications (20x, 40x). MN recall, PPV, and mean intersection-over-union (mIOU) per object were quantified across conditions. Dotted line = performance of the VCS neural net on RPE1 H2B-emiRFP703 live 20x images. Performance is similar across conditions except H2B-GFP in fixed HeLa cells (teal squares). N = 1. n (on graph) = cells.

replaced most concatenation operations with addition (Lu et al., 2022). This UNet was trained on the same images as the nucleus/MN segmenter after annotation with concave hulls automatically generated on annotated MN and nuclei (Fig. S2C). To generate cell masks from the predicted distance and proximity maps, the results are summed and used for watershed segmentation followed by elimination of false boundaries based on the proximity map predictions of true cell boundaries (Fig. S2E).

In the last step of the MNFinder module, the Nuc/MN and cell segmentation results are integrated to produce a final set of labels, identifying each unique MN, each cell with its nuclei and MN, and each unique nucleus (Fig 2B). We validated MNFinder on single channel images of RPE1 RFP703/Dendra, U2OS RFP703/Dendra, HeLa H2B-GFP, and HFF cells after incubation in Mps1i for 24 hours. Cells were imaged live and fixed, using a 20x widefield and 40x confocal microscopy, and using H2B and DAPI to visualize DNA. In these images, MN were present at a frequency between ~30-70% of cells due to induction of chromosome missegregation. These levels are elevated compared to some cancer cell lines and tumor samples (Jdey et al., 2017). Therefore, we also analyzed publicly available images of unperturbed U2OS cells from Broad Bioimage dataset BBBC039v1, which have an MN frequency of 8% (Table 2). MNFinder showed significant improvement in recall over VCS MN with an additional improvement in PPV for some conditions (Fig. 2D). Importantly, recall and PPV were largely insensitive to image resolution, DNA label, and cell type. HeLa H2B-GFP 40x images were an outlier in terms of performance for unclear reasons, potentially due to increased nuclear shape diversity. We also calculated the per object mIoU to determine the quality of the segmentation and found that most MN were accurately segmented with mIoU values between 69-79% (Fig. 2D, Table 1).

These metrics indicate that MNFinder provides accurate and robust MN segmentation across multiple cell lines and image acquisition settings. MNFinder identifies MN with similar sensitivity and substantially improved specificity compared to existing MN enumeration programs (Table 3)

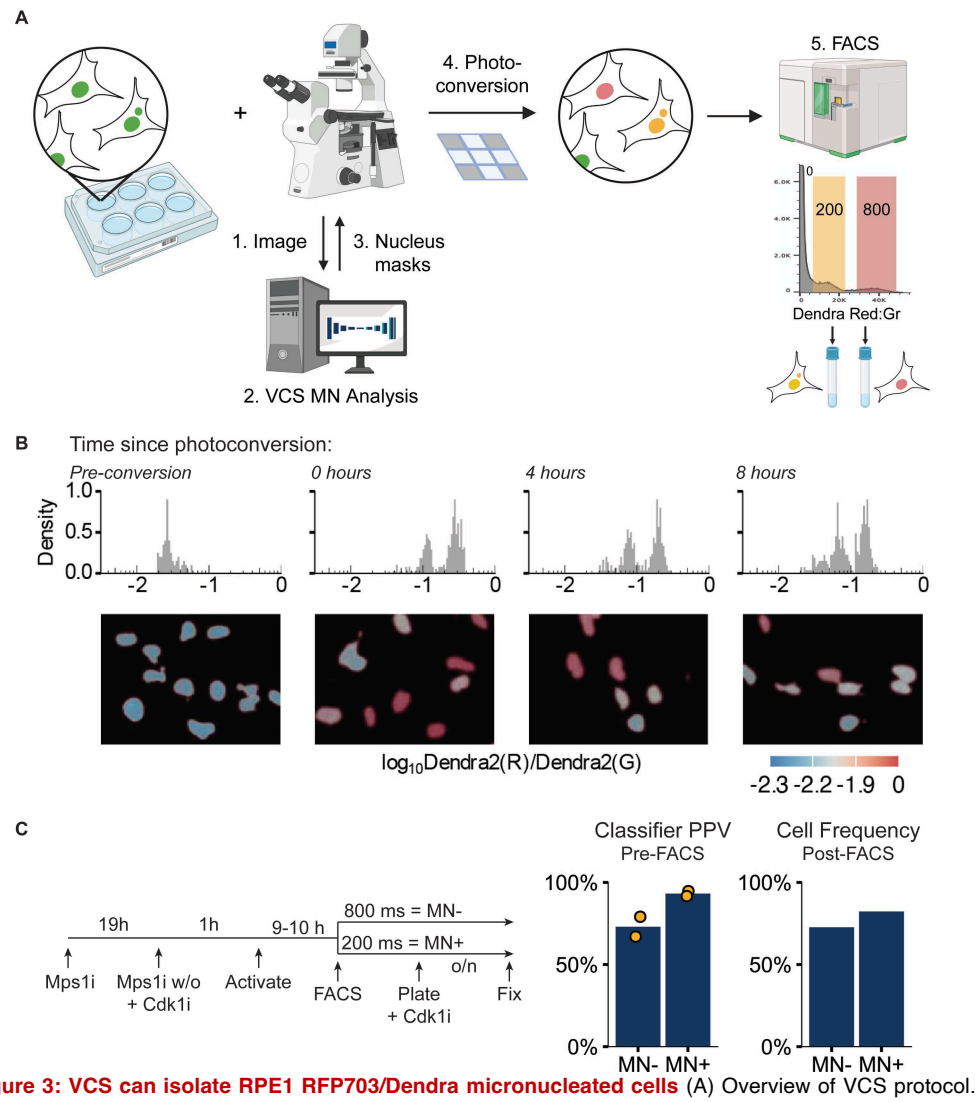

Figure 3: VCS can isolate RPE1 RFP703/Dendra micronucleated cells (A) Overview of VCS protocol. Cells are plated in multi-well plates, during imaging cellular phenotypes are quantified, VCS MN is deployed, and specified classes are photocoverted for either 200 or 800 ms, yielding two different ratios of red:green fluorescence. These differences are quantified by FACs and gated for cell sorting. Graphic created with BioRender.com. (B) Quantification of nuclear red:green ratios from images of the same field taken 0, 4, and 8 h after photoconversion displayed as histograms. Representative images from each time point pseudo-colored by log10 Dendra red:green ratio (below). N = 1, n = 82, 353, 285, 313. (C) Experimental design of MN cell isolation validation. Classifier PPV calculated on images acquired during activation and frequency of MN- or MN+ cells manually quantified in cells plated and fixed post-sorting. Pre-FACS: N = 2, n = 328, 186 post-FACS N=1. n = 338, 353.

(Ibarra-Arellano et al., 2024; Pons and Mauvezin, 2022) and is the only one to report a high mIoU, which is necessary for quantifying MN characteristics. This module is available as a Python package, MNFinder, via PyPI and on the Hatch Lab GitHub repository.

VCS MN suitability for analysis of micronucleated RPE1 RFP703/Dendra cell transcriptomes

To demonstrate the utility of our MN segmentation modules, we asked whether the VCS MN could be used for optical cell isolation to obtain cell populations substantially depleted and enriched for MN. Visual cell sorting (VCS) is a recently developed optical cell isolation pipeline that specifically labels and isolates multiple populations of adherent cells in a single experiment by combining on-demand image analysis with UV (405 nm) pulses of different durations targeted with single cell accuracy. It can generate up to four different proportions of converted Dendra2, which can be quantified and sorted by FACS (Hasle et al., 2020) (Fig. 3A).

We first validated the utility of micronucleated RPE1 RFP703/Dendra cells for VCS analysis. We confirmed that we could specifically activate and sort two populations of RFP703/Dendra cells by classifying cells in a mixed pool based on CellTrace far red labeling (Fig. S3A-C). We next confirmed that Dendra2 red:green ratios were stable for the duration of a VCS MN isolation experiment by randomly converting RFP703/Dendra cells using a short, long, or no UV pulse and analyzing nuclear fluorescence intensity 0, 4, and 8 hours after activation. Quantification of nuclear red:green ratios from the same fields over time showed the persistence of three distinct fluorescent populations and a minimal loss of red fluorescence (Fig. 3B), indicating that photoconversion persisted over the time required to activate multiple 6-well cell populations.

We also confirmed that VCS MN could accurately activate and isolate MN+ and MN- cells. RFP703/Dendra cells were incubated with Mps1i one day prior to imaging to generate MN and Cdk1i added prior to imaging to prevent mitosis, which dilutes the Dendra2(red) signal and frequently alters MN status (Hatch 2013). Cells were activated based on VCS MN analysis results and isolated by FACS. Isolated cells were replated in medium containing Cdk1i, fixed, and MN content quantified by manual fluorescent image analysis. Comparison of classifier PPV for MN+ and MN- cells during activation and MN+ and MN- cell frequency after FACs found a strong enrichment for the correct cell type in each group, with the increased purity of MN+ classified cells being retained during sorting (Fig. 3C).

To determine how MN frequency affects MN cell isolation, we used the PPV and recall values from the VCS MN analysis of untreated U2OS cells (Fig. 1E: U2OS Broad, MN frequency = 8%) to estimate MN+ and – cell population purity. As expected, the purity of the MN- population increases and the purity of the MN+ population decreases compared to populations with a higher MN frequency (Fig. S3D). However, this represents a nearly 7-fold enrichment of MN+ cells in the isolated population and the MN+ cell purity is comparable to the enrichment of tumor cells in patient

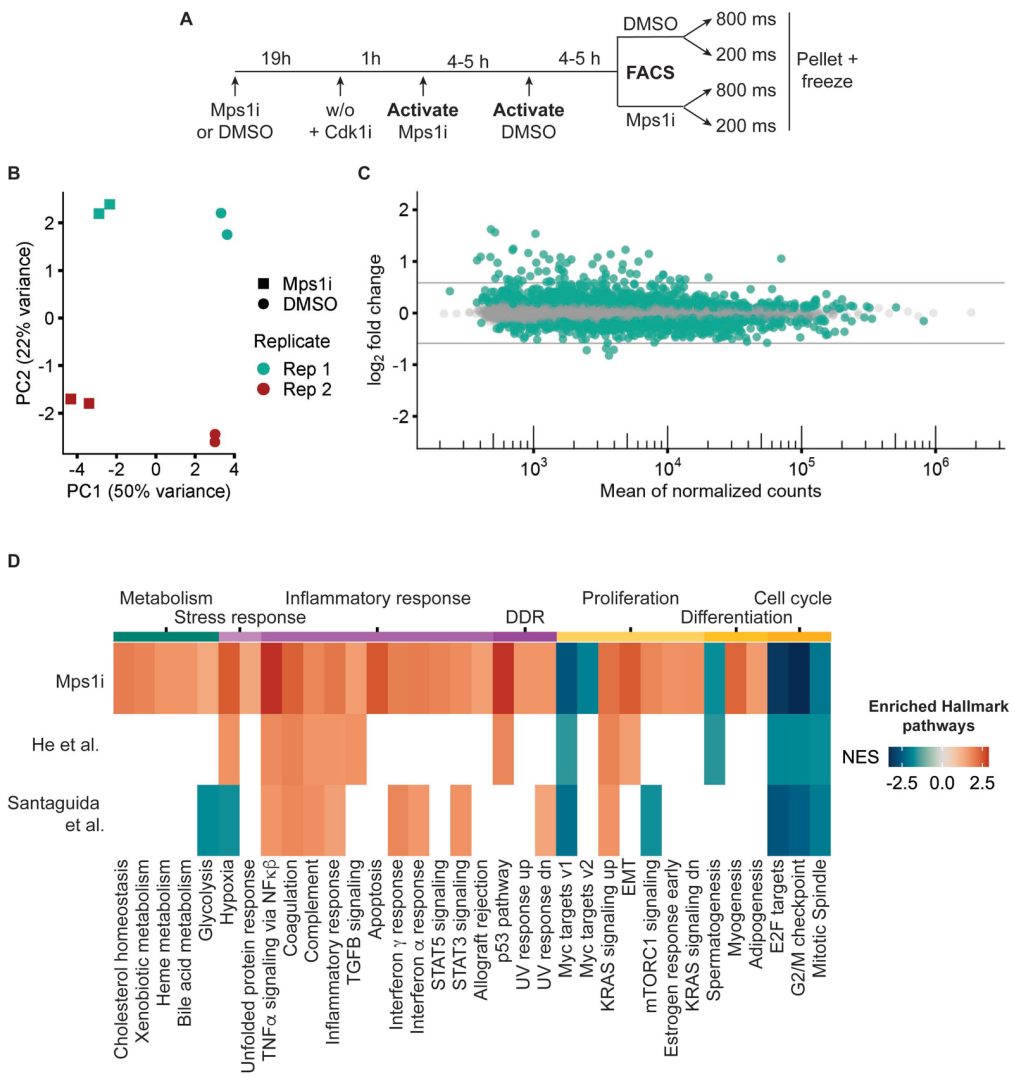

Figure 4: VCS pipeline identifies Mps1i transcriptional response. (A) Timeline of experiment. (B) PCA plot showing clustering of Mps1i-treated and DMSO-treated cells by treatment (major) and by replicate (minor). Each experimental replicate represents 2 technical replicates. (C) MA plot. Differentially expressed genes (FDR adjusted p-value < 0.05) are in green. Gray lines represent 1.5 fold-change in expression. (D) Heatmap of Hallmark pathway enrichment between VCS data and data from (Santaguida et al., 2017) and (He et al., 2019) analyses of RPE1 cells after induction of chromosome missegregation. Hallmark pathways (bottom) were clustered based on manually annotated categories (top).

biopsies (Wu et al., 2021). Thus, VCS can be combined with the VCS MN neural net to generate cell populations enriched for MN- and MN+ cells from a variety of conditions that are highly suitable for discovery, including genetic screening, bulk RNA and proteomic analyses, and single cell sequencing.

To validate this pipeline for transcriptome analysis, we used RNASeq to define gene expression changes in RPE1 cells after Mps1i incubation and VCS. Cells were incubated with DMSO or Mps1i and each population was randomly activated with short and long UV pulses (Fig. 4A). Conversion of 1,500–2,000 fields at 20x magnification allowed us to isolate 13k cells after FACS for each photoconverted population in each condition. Isolated populations were processed for RNASeq and principal component analysis (PCA) revealed that, as expected, cells clustered first by treatment group (Fig. 4B). Analysis of the DMSO samples found minimal differences in gene expression associated with UV pulse duration (Fig. S4A-B, Table 4), consistent with previous results (Hasle et al., 2020). Therefore, data from cells activated at both pulse lengths were pooled in subsequent analyses. MA analysis identified 2,200 differentially expressed genes (DEGs) in Mps1i versus DMSO treated cells, 63 of which had absolute fold-changes > 1.5 (Fig. 4C, Table 5-6). We used GSEA analysis to compare our results to previously identified changes in RPE1 cells after mitotic disruption to induce aneuploidy (Table 7-8) and found substantial overlap between enriched Hallmark pathways in Mpsi1i cells isolated by VCS and previous studies (He et al., 2018; Santaguida et al., 2017) (Table 9-11). These included increased expression of inflammation, EMT, and p53 associated genes (Fig. 4D). Additional changes observed in VCS-processed samples fell into similar function categories and are potentially due to differences in sequencing depth. These data confirm that the VCS MN pipeline can accurately and sensitively identifies biologically relevant transcriptional changes in aneuploid RPE1 cells.

MN rupture induces few unique transcriptional changes and does not contribute to the initial aneuploidy response.

To determine whether MN formation induces a transcriptional response, we treated RFP703/Dendra cells with Mps1i, activated MN+ and MN- cells based on VCS MN analysis results, and isolated differentially activated populations by FACS in duplicate (Fig. 5A). PCA revealed that results clustered largely by replicate and, consistent with this, only a few DEGs were identified with just two having absolute fold-changes greater than 1.5 (Fig. 5 B-C, Table 12). Both highly altered DEGs were also strongly upregulated by Mps1i treatment (Table 13). Although our analysis did identify some batch effects, our data strongly suggest that micronucleation does not induce a unique transcriptional response.

We next compared gene expression between micronucleated cells classified as rupture+ and rupture-. Because the overall MN rupture frequency increases over time (Hatch et al., 2013), we first synchronized cells in G1 using a Cdk4/6 inhibitor followed by release into Mps1i (Mammel et al., 2021) (Fig. 5D). This results in a more consistent rate of MN rupture (Fig. S5A). We modeled how the 4-5 hours required for analysis and activation of 1 well would alter

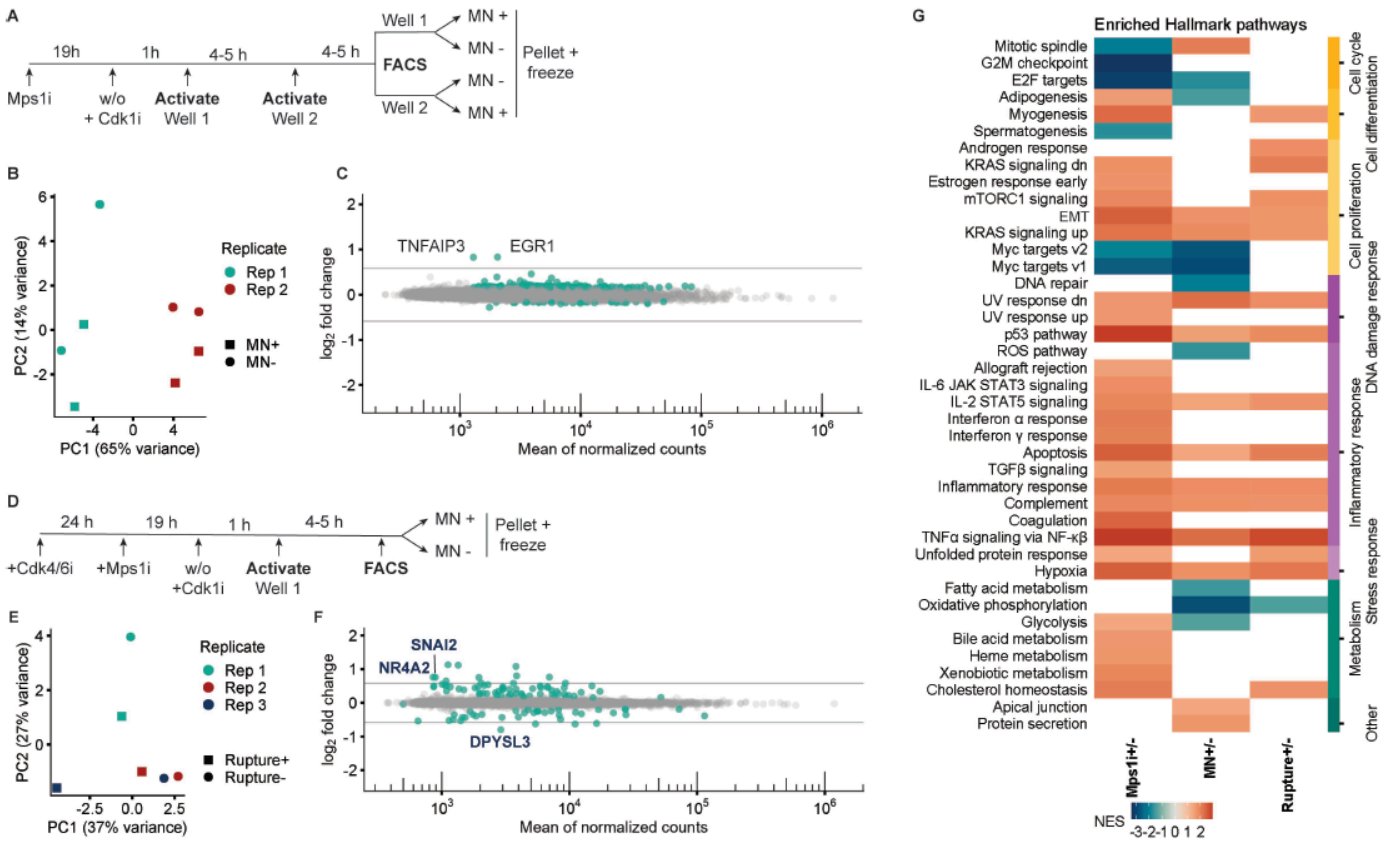

Figure 5: Micronucleation and rupture transcriptional changes largely overlap with aneuploidy response. (A) Timeline of experiment for MN+ and MN- cell isolation from RFP703/Dendra cells. (B) PCA plot showing clustering of MN+ and MN- cells by replicate (major) and condition (minor). (C) MA plot. Of identified DEGs, only 2 have fold-changes larger than 1.5. Both, TNFAIP3 and EGR1, are also significantly upregulated in Mps1i treated cells. (D) Timeline of experiment for rupture+ and rupture- cell isolation. (E) PCA plot showing clustering of intact MN and ruptured MN cells by condition and replicate. (F) MA plot. Three highly differently expressed genes unique to this dataset are indicated on plot. (G) Heatmap of Hallmark pathway enrichment in datasets of DMSO vs Mps1i, Mps1i-treated cells with and without MN, and synchronized, Mps1i-treated, MN+ cells with and without MN rupture. Pathways are grouped based on manual annotation (left) and show substantial overlap between categories enriched in Mps1i+ cells versus the MN+ and rupture+ subsets.

population purity by manually quantifying the frequency of rupture+ cells in images taken 5 hours apart. Based on the increase in rupture+ cells we observed, we estimated only a small decrease in purity of the rupture- population (Fig. S5B), with a sustained high level of enrichment for both populations. PCA revealed that results clustered first by condition, indicating a transcriptional difference between rupture+ versus rupture– cells (Fig. 5E), and the MA plot identified 106 DEGs, 14 of which had absolute fold changes greater than 1.5 fold (Fig. 5F, Table 14). Of these, 3 were unique to cells with ruptured MN (Table 15). GSEA analysis confirmed that most of the pathways altered in MN+ or rupture+ cells overlapped with those identified in the total aneuploid Mps1i population (Fig. 5G, Table 16-17).

We next asked whether micronucleation or MN rupture contributed to the transcriptional response to aneuploidy. We first quantified aneuploidy frequency in MN-, rupture-, and rupture+ cells to determine whether transcription changes could reflect underlying differences in ploidy. Cells were labeled with probes against chromosomes 1, 11, or 18, all of which frequently missegregate into MN (Fig. S6A), and ruptured MN were identified by loss of H3K27Ac (Mammel et al., 2021; Mohr et al., 2021) (Fig. 6A). Quantification of chromosome foci number found that aneuploidy frequency varied between chromosomes but was consistently higher for MN+ compared to MN- cells. This trend was also observed in rupture+ versus rupture- micronucleated cells (Fig. 6B). Similar results were obtained when transcription loss due to MN rupture was considered (functional aneuploidy) (Fig. S6B). We next compared the fold change of highly upregulated or downregulated genes in the Mps1i dataset to results from analysis of the subsetted populations of MN+ and rupture+ cells. All replicates were analyzed individually to reduce noise from batch effects in the MN+ results. This analysis identified one gene cluster that increased in expression in the subset of Mps1i cells with ruptured MN and included the genes FILIP1L, CREB5, TNFAIP3, ATF3, and EGR1 (Fig. 6C, Table 18). We attempted to validate increased protein expression of EGR1 and ATF3 in rupture+ cells by immunofluorescence. As a positive control, we quantified an increase EGR1 and ATF3 nuclear mean intensity after addition of hEGF and DNA damage by doxorubicin, respectively (Fig. S6C-D). Both genes were defined as upregulated by Mps1i and showed increased expression in Mps1i treated cells compared to controls by immunofluorescence (Fig. 6D-E). However, analysis of rupture+ versus other classes of Mps1i cells found no increase in EGR1 expression and only a small increase in ATF3 that was less than that observed between DMSO and Mps1i treated cells (Fig. 6D-E). Overall, our results strongly suggest that protein expression changes in MN+ and rupture+ cells are driven mainly by increased aneuploidy rather than cellular sensing of MN formation and rupture.

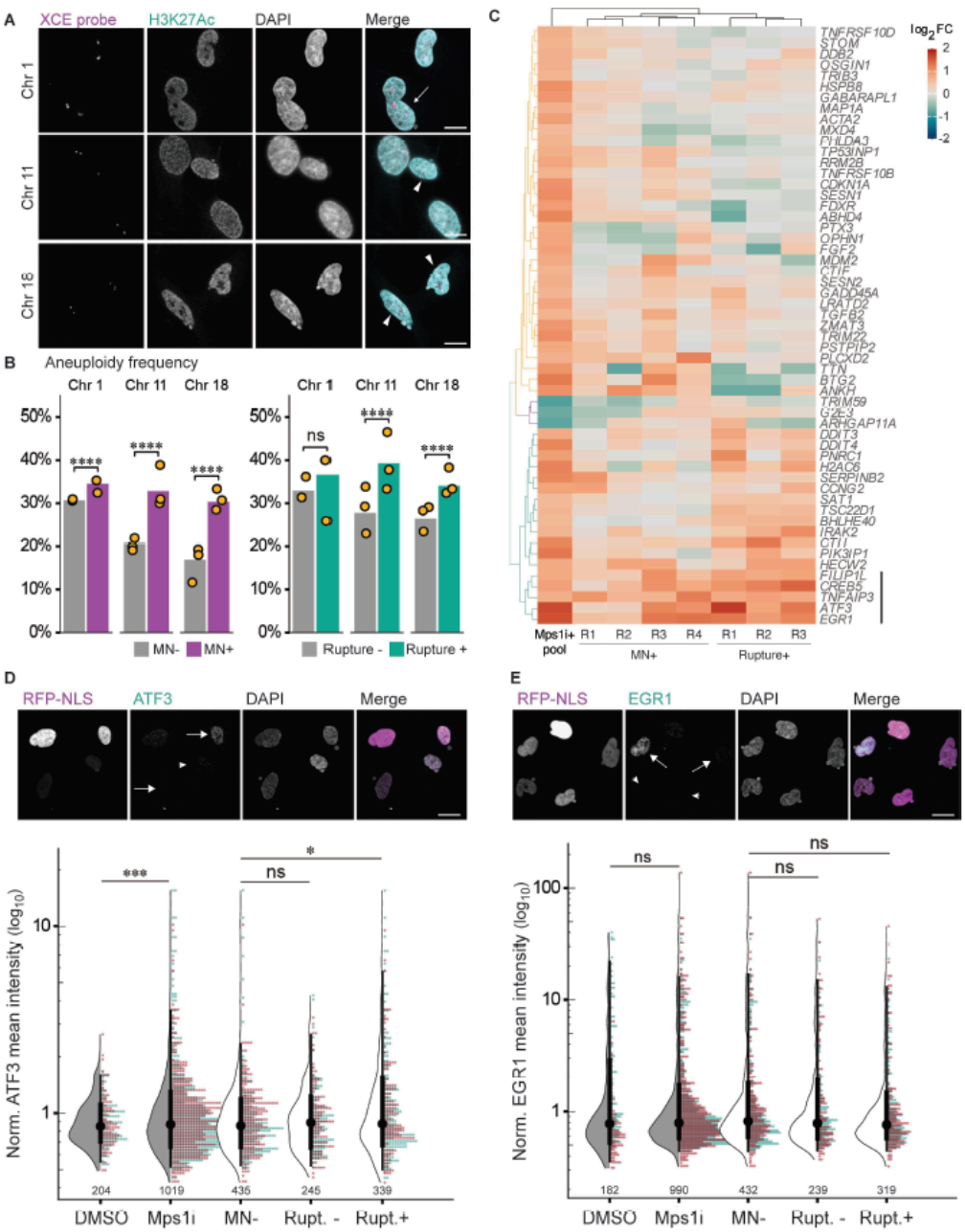

Figure 6: Micronucleation and rupture do not significantly contribute to the aneuploidy transcription response. (A) Examples of DNA FISH for chromosomes 1, 11, and 18 and H3K27Ac identification of intact MN. Arrows = ruptured MN, arrowheads = intact MN. (B) Quantification of aneuploidy frequency (foci ≠ 2) per chromosome. Cells manually classified as MN- or MN+, and rupture- or rupture+. MN: Chr 1: N=2, n=429, 158; Chr 11: N=3, n=406, 313, 160; Chr 18: N=3, n=425, 202, 230. Rupture: Chr 1: N=2, n=187, 74; Chr 11: N=3, n=190, 108, 71; Chr 18: N=3, n=186, 102, 101. (C) Heatmap of highly different Mps1i+ DEGs (cutoff = absolute FC 1.5) compared to MN+ and rupture+ replicates. Euclidean distances calculated for features and samples and clustered by complete-linkage. Genes with lacking values for at least one class were excluded. Line = gene cluster upregulated in rupture+ cells. (D-E) Representative images of ATF3 and EGR1 labeling in RPE1 2xRFP-NLS cells after Mps1i incubation. Arrows = ruptured MN cell, arrowheads = intact MN cell (top). Quantification of normalized ATF3 and EGR1 mean nuclear intensity in manually classified cells (bottom). N = 2 (graph colors), n = on graph, p: ns > 0.05, * ≤ 0.05, *** < 0.001 by GEE. Scale bar = 20 μm.

Discussion

In this study, we present two machine-learning based modules to identify MN and micronucleated cells based on single channel fluorescence images and combine one with visual cell sorting to profile transcriptional responses to MN formation and rupture. We demonstrate that our MN segmentation pipeline, MNFinder, can robustly classify and segment MN from DNA labeled images across multiple cell types and fluorescent imaging conditions. Further, we demonstrate that a separate MN cell classifier, VCS MN, rapidly and robustly identifies micronucleated cells from low resolution images and can be combined with single-cell photoconversion to accurately isolate live cells with none, intact, or ruptured MN from a mixed population. Using this platform, we find that, unexpectedly, neither micronucleation nor rupture triggers gene expression changes beyond those associated with increased aneuploidy. Overall, our study brings a powerful high-throughput optical isolation strategy to MN biology and we anticipate that it will enable a wide range of new investigations.

VCS MN isolation has several advantages over current methods to identify the mechanisms and consequences of MN formation and rupture. First, it can be used on any adherent cell line in the absence of genetic perturbations. This overcomes challenges involved with using lamin B2 overexpression to inhibit MN rupture (Hatch et al., 2013), which is limited to specific cell lines and MN types (Mammel et al., 2021; Xia et al., 2019) and is complicated by additional changes in mitosis and gene expression (Agustinus et al., 2023; Han et al., 2020; Kuga et al., 2014; Liwag et al., 2024). In addition, it overcomes cell line and MN content restrictions imposed by systems that induce missegregation of single chromosomes or chromosome arms (Lin et al., 2023; Ly et al., 2019, 2016; Shoshani et al., 2021; Trivedi et al., 2023) by enabling analysis of all missegregation events in any genetic background. Unlike live single cell assays, it is highly scalable and eliminates selection pressures and restrictions added by clonal expansion (Mohr et al., 2021; Papathanasiou et al., 2023; Zhang et al., 2015). Importantly, VCS MN isolation captures whole live cells, overcoming limitations associated MN purification (Agustinus et al., 2023; Klaasen et al., 2022; Mohr et al., 2021; Papathanasiou et al., 2023; Tang et al., 2022) and permits time-resolved analyses of cellular changes and MN

chromatin by population-level analyses. VCS has several advantages over similar optical isolation or *in situ* sequencing techniques as it can be adapted to any wide-field microscope by adding a digital micromirror to existing equipment, and can be performed on attached cells, which are critical to achieve the nuclear and cytoplasm spreading required for accurate MN identification (Li et al., 2015).

VCS MN isolation does have limitations. Due to Dendra2 signal decay and ongoing MN rupture, only about 200,000 cells can be analyzed and targeted per experiment. For optical pooled screening or analysis of rare cells, this limits the number of genes or depth of analysis that can be achieved. Cell fixation would overcome this issue and efforts to improve sample extraction in these conditions are ongoing (Kanfer et al., 2021; Yan et al., 2021). VCS MN isolation also requires introduction of at least one photoconvertible or activatable protein to mark the cells and a second fluorescent protein to discriminate ruptured MN. This limits the channels available for additional phenotype identification. However, recent advances in cell structure prediction (Johnson et al., 2023) may vastly expand the phenotypic information available from limited cell labels. VCS MN segmentation and MNFinder precision vary across cell types, and widely divergent nuclear morphologies from the training set could significantly impair performance. Additional training of the neural net should improve this metric, but different algorithm architectures will likely be required to identify MN in signal-rich environments like organoids or tissue samples.

We observed upregulation of several pathways, including inflammation, endothelial-to-mesenchymal transition, and p53, in Mps1i-treated RPE1 cells that were previously identified as enriched in similar studies (He et al., 2019; Santaguida et al., 2017). These results demonstrate the suitability of our platform for detecting biologically relevant transcript changes in aneuploid cells. However, our analysis of micronucleated cells and cells with ruptured MN found only a handful of genes that were uniquely upregulated by MN rupture and no changes that indicated a contribution of either condition to the aneuploidy response. Thus, in line with previous results (Santaguida et al., 2017), our findings suggest that MN and MN rupture are not sensed by the cell outside of their contribution to aneuploidy through dysregulated transcription and limited replication of the sequestered chromatin (Hatch et al., 2013; Papathanasiou et al., 2023; Zhang et al., 2015). Of significant interest is whether similar results will be obtained in cells with more robust cGAS/STING signaling. There is a discrepancy about whether cGAS binding to ruptured MN is sufficient to initiate signaling, and how MN chromatin content may mediate this (Bakhoum et al., 2018; Chen et al., 2020; Dou et al., 2017; Harding et al., 2017; MacDonald et al., 2023; Mackenzie et al., 2017; Mohr et al., 2021; Willan et al., 2019), that VCS MN isolation is ideal for resolving.

VCS MN isolation is a highly flexible platform that enables powerful new approaches to address fundamental questions in MN biology. VCS MN isolation can be used for optical pooled screening, an unbiased method that would be ideal to identify mechanisms of MN rupture, genetic changes that promote proliferation of micronucleated cells, and, in combination with dCas9-based chromosome labeling (Chen et al., 2013; Maass et al., 2018; Tanenbaum et al., 2014), mechanisms that enrich specific chromosomes in MN and could drive cancer-specific aneuploidies (Ben-David and Amon, 2020). Recovering live cell populations of cells with intact and ruptured MN will also enable precise analysis of post-mitotic genetic and functional changes caused by these conditions. For instance, these cell populations can be analyzed for acquisition of disease-associated behaviors, including proliferation, migration, and used in *in vivo* tumorigenesis and metastasis assays to directly assess their contribution to cancer development. In summary, automated MN segmentation and VCS MN isolation are poised to provide critical insights into a wide-range of questions about how MN form, rupture, and cause disease pathologies.

References

Agustinus AS, Al-Rawi D, Dameracharla B, Raviram R, Jones BSCL, Stransky S, Scipioni L, Luebeck J, Bona MD, Norkunaite D, Myers RM, Duran M, Choi S, Weigelt B, Yomtoubian S, McPherson A, Toufektchan E, Keuper K, Mischel PS, Mittal V, Shah SP, Maciejowski J, Storchova Z, Gratton E, Ly P, Landau D, Bakhoum MF, Koche RP, Sidoli S, Bafna V, David Y, Bakhoum SF. 2023. Epigenetic dysregulation from chromosomal transit in micronuclei. *Nature* 1–8. doi:10.1038/s41586-023-06084-7

Bakhoum SF, Ngo B, Laughney AM, Cavallo J-A, Murphy CJ, Ly P, Shah P, Sriram RK, Watkins TBK, Taunk NK, Duran M, Pauli C, Shaw C, Chadalavada K, Rajasekhar VK, Genovese G, Venkatesan S, Birkbak NJ, McGranahan N, Lundquist M, LaPlant Q, Healey JH, Elemento O, Chung CH, Lee NY, Imielenski M, Nanjangud G, Pe'er D, Cleveland DW, Powell SN, Lammerding J, Swanton C, Cantley LC. 2018. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* **553**:467–472. doi:10.1038/nature25432

Ben-David U, Amon A. 2020. Context is everything: aneuploidy in cancer. *Nature reviews Genetics* **21**:44–62. doi:10.1038/s41576-019-0171-x

Bona MD, Bakhoum SF. 2024. Micronuclei and Cancer. *Cancer Discov* **14**:214–226. doi:10.1158/2159-8290.cd-23-1073

Bray M-A, Singh S, Han H, Davis CT, Borgeson B, Hartland C, Kost-Alimova M, Gustafsdottir SM, Gibson CC, Carpenter AE. 2016. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc 11:1757–1774. doi:10.1038/nprot.2016.105

Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, Heng C, Becker T, Doan M, McQuin C, Rohban M, Singh S, Carpenter AE. 2019. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat Methods* **16**:1247–1253. doi:10.1038/s41592-019-0612-7

Caicedo JC, Roth J, Goodman A, Becker T, Karhohs KW, Broisin M, Molnar C, McQuin C, Singh S, Theis FJ, Carpenter AE. 2019b. Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images. Cytom Part A 95:952–965. doi:10.1002/cyto.a.23863

Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li G-W, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. 2013. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155**:1479–1491. doi:10.1016/j.cell.2013.12.001

Chen J, Harding SM, Natesan R, Tian L, Benci JL, Li W, Minn AJ, Asangani IA, Greenberg RA. 2020. Cell Cycle Checkpoints Cooperate to Suppress DNA- and RNA-Associated Molecular Pattern Recognition and Anti-Tumor Immune Responses. *Cell Rep* **32**:108080. doi:10.1016/j.celrep.2020.108080

Cohen WW. 1995. Fast Effective Rule Induction In: Prieditis A, Russell S, editors. Machine Learning Proceedings 1995. Morgan Kaufmann. pp. 115–123. doi:10.1016/b978-1-55860-377-6.50023-2

Crameri F. 2018. Geodynamic diagnostics, scientific visualisation and StagLab 3.0. *Geosci Model Dev* **11**:2541–2562. doi:10.5194/gmd-11-2541-2018

Dou Z, Ghosh K, Vizioli MG, Zhu J, Sen P, Wangensteen KJ, Simithy J, Lan Y, Lin Y, Zhou Z, Capell BC, Xu C, Xu M, Kieckhaefer JE, Jiang T, Shoshkes-Carmel M, Tanim KMAA, Barber GN, Seykora JT, Millar SE, Kaestner KH, Garcia BA, Adams PD, Berger SL. 2017. Cytoplasmic chromatin triggers inflammation in senescence and cancer. *Nature* **550**:402–406. doi:10.1038/nature24050

Duckham M, Kulik L, Worboys M, Galton A. 2008. Efficient generation of simple polygons for characterizing the shape of a set of points in the

plane. *Pattern Recognit* **41**:3224–3236. doi:10.1016/j.patcog.2008.03.023
Greene D, Richardson S, Turro E. 2017. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics* **33**:1104–1106. doi:10.1093/bioinformatics/btw763
Guo X, Ni J, Liang Z, Xue J, Fenech MF, Wang X. 2019. The molecular origins and pathophysiological consequences of micronuclei: New insights into an age-old problem. *Mutation Research* **779**:1–35. doi:10.1016/j.mrrev.2018.11.001
Halekoh U, Højsgaard S, Yan J. 2006. The R Package geepack for Generalized Estimating Equations. *J Stat Softw* **15**. doi:10.18637/jss.v015.i02
Han L, Choudhury S, Mich-Basso JD, Ammanamanchi N, Ganapathy B, Suresh S, Khaladkar M, Singh J, Maehr R, Zuppo DA, Kim J, Eberwine JH, Wyman SK, Wu YL, Kühn B. 2020. Lamin B2 Levels Regulate Polyploidization of Cardiomyocyte Nuclei and Myocardial Regeneration. *Dev Cell* **53**:42-59.e11. doi:10.1016/j.devcel.2020.01.030
Harding SM, Benci JL, Irianto J, Discher DE, Minn AJ, Greenberg RA. 2017. Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**:466–470. doi:10.1038/nature23470
Hasle N, Cooke A, Srivatsan S, Huang H, Stephany JJ, Krieger Z, Jackson D, Tang W, Pendyala S, Monnat RJ, Trapnell C, Hatch EM, Fowler DM. 2020. High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol Syst Biol* **16**:e9442. doi:10.15252/msb.20209442
Hatch EM, Fischer AH, Deerinck TJ, Hetzer MW. 2013. Catastrophic Nuclear Envelope Collapse in Cancer Cell Micronuclei. *Cell* **154**:47–60. doi:10.1016/j.cell.2013.06.007
He B, Gnawali N, Hinman AW, Mattingly AJ, Osimani A, Cimini D. 2019. Chromosomes missegregated into micronuclei contribute to chromosomal instability by missegregating at the next division. *Oncotarget* **10**:2660–2674. doi:10.18632/oncotarget.26853
He Q, Au B, Kulkarni M, Shen Y, Lim KahJ, Maimaiti J, Wong ChengKit, Luijten MoniqueNH, Chong HC, Lim EH, Rancati G, Sinha I, Fu Z, Wang X, Connolly JohnE, Crasta KC. 2018. Chromosomal instability-induced senescence potentiates cell non-autonomous tumourigenic effects. *Oncogenesis* **7**:62. doi:10.1038/s41389-018-0072-4
Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, Han X, Chen Y-W, Wu J. 2020. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. *ICASSP 2020 - 2020 IEEE Int Conf Acoust, Speech Signal Process (ICASSP)* **00**:1055–1059. doi:10.1109/icassp40776.2020.9053405
Ibarra-Arellano MA, Caprio LA, Hada A, Stotzem N, Cai L, Shah S, Melms JC, Wünneman F, Izar B, Schapiro D. 2024. micronuclAI: Automated quantification of micronuclei for assessment of chromosomal instability. bioRxiv 2024.05.24.595722. doi:10.1101/2024.05.24.595722
Jdey W, Thierry S, Popova T, Stern M-H, Dutreix M. 2017. Micronuclei Frequency in Tumors Is a Predictive Biomarker for Genetic Instability and Sensitivity to the DNA Repair Inhibitor AsiDNA. Cancer Res 77:4207–4216. doi:10.1158/0008-5472.can-16-2693
Johnson GT, Agmon E, Akamatsu M, Lundberg E, Lyons B, Ouyang W, Quintero-Carmona OA, Riel-Mehan M, Rafelski S, Horwitz R. 2023. Building the next generation of virtual cells to understand cellular biology. *Biophys J* **122**:3560–3569. doi:10.1016/j.bpj.2023.04.006
Kanfer G, Sarraf SA, Maman Y, Baldwin H, Dominguez-Martin E, Johnson KR, Ward ME, Kampmann M, Lippincott-Schwartz J, Youle RJ. 2021. Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. *J Cell Biol* **220**:e202006180. doi:10.1083/jcb.202006180
Kiyono T, Foster SA, Koop JI, McDougall JK, Galloway DA, Klingelhutz AJ. 1998. Both Rb/p16INK4a inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* **396**:84–88. doi:10.1038/23962
Klaasen SJ, Truong MA, Jaarsveld RH van, Koprivec I, Štimac V, Vries SG de, Risteski P, Kodba S, Vukušić K, Luca KL de, Marques JF, Gerrits EM, Bakker B, Foijer F, Kind J, Tolić IM, Lens SMA, Kops GJPL. 2022. Nuclear chromosome locations dictate segregation error frequencies. *Nature* **607**:604–609. doi:10.1038/s41586-022-04938-0
Kneissig M, Keuper K, Pagter MS de, Roosmalen MJ van, Martin J, Otto H, Passerini V, Sparr AC, Renkens I, Kropveld F, Vasudevan A, Sheltzer JM, Kloosterman WP, Storchova Z. 2019. Micronuclei-based model system reveals functional consequences of chromothripsis in human cells. *Elife* **8**:e50292. doi:10.7554/elife.50292
Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. 2021. Fast gene set enrichment analysis. *bioRxiv* 060012. doi:10.1101/060012
Krupina K, Goginashvili A, Cleveland DW. 2021. Causes and consequences of micronuclei. *Curr Opin Cell Biol* **70**:91–99. doi:10.1016/j.ceb.2021.01.004
Kuga T, Nie H, Kazami T, Satoh M, Matsushita K, Nomura F, Maeshima K, Nakayama Y, Tomonaga T. 2014. Lamin B2 prevents chromosome instability by ensuring proper mitotic chromosome segregation. *Oncogenesis* **3**:e94–e94. doi:10.1038/oncsis.2014.6
Kwon J, Bakhoum SF. 2020. The Cytosolic DNA-Sensing cGAS–STING Pathway in Cancer. *Cancer Discov* **10**:26–39. doi:10.1158/2159-8290.cd-19-0761
Li Y, Lovett D, Zhang Q, Neelam S, Kuchibhotla RA, Zhu R, Gundersen GG, Lele TP, Dickinson RB. 2015. Moving Cell Boundaries Drive Nuclear Shaping during Cell Spreading. *Biophysical Journal* **109**:670–686. doi:10.1016/j.bpj.2015.07.006
Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**:417–425. doi:10.1016/j.cels.2015.12.004
Lin T-Y, Goyal P, Girshick R, He K, Dollár P. 2018. Focal Loss for Dense Object Detection. *IEEE Trans Pattern Anal Mach Intell* **42**:318–327. doi:10.1109/tpami.2018.2858826
Lin Y-F, Hu Q, Mazzagatti A, Valle-Inclán JE, Maurais EG, Dahiya R, Guyer A, Sanders JT, Engel JL, Nguyen G, Bronder D, Bakhoum SF, Cortés-Ciriano I, Ly P. 2023. Mitotic clustering of pulverized chromosomes from micronuclei. *Nature* **618**:1041–1048. doi:10.1038/s41586-023-05974-0
Liwag EMP, Wei X, Acosta N, Carter LM, Yang J, Almassalha LM, Jain S, Daneshkhah A, Rao SSP, Seker-Polat F, MacQuarrie KL, Ibarra J, Agrawal V, Aiden EL, Kanemaki MT, Backman V, Adli M. 2024. Depletion of lamins B1 and B2 promotes chromatin mobility and induces differential gene expression by a mesoscale-motion-dependent mechanism. *Genome Biol* **25**:77. doi:10.1186/s13059-024-03212-y
Ljosa V, Sokolnicki KL, Carpenter AE. 2012. Annotated high-throughput microscopy image sets for validation. Nat Methods 9:637–637. doi:10.1038/nmeth.2083
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8
Love MI, Soneson C, Hickey PF, Johnson LK, Pierce NT, Shepherd L, Morgan M, Patro R. 2020. Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLoS Comput Biol* **16**:e1007664. doi:10.1371/journal.pcbi.1007664
Lu H, She Y, Tie J, Xu S. 2022. Half-UNet: A Simplified U-Net Architecture for Medical Image Segmentation. *Front Neuroinformatics* **16**:911679. doi:10.3389/fninf.2022.911679
Ly P, Brunner SF, Shoshani O, Kim DH, Lan W, Pyntikova T, Flanagan AM, Behjati S, Page DC, Campbell PJ, Cleveland DW. 2019. Chromosome segregation errors generate a diverse spectrum of simple and complex genomic rearrangements. *Nature genetics* **51**:705–715. doi:10.1038/s41588-019-0360-8
Ly P, Teitz LS, Kim DH, Shoshani O, Skaletsky H, Fachinetti D, Page DC, Cleveland DW. 2016. Selective Y centromere inactivation triggers chromosome shattering in micronuclei and repair by non-homologous end joining. *Nature cell biology* **19**:68–75. doi:10.1038/ncb3450
Maass PG, Barutcu AR, Shechner DM, Weiner CL, Melé M, Rinn JL. 2018. Spatiotemporal allele organization by allele-specific CRISPR live-cell imaging (SNP-CLING). *Nature Structural & Molecular Biology* **25**:176–184. doi:10.1038/s41594-017-0015-3
MacDonald KM, Nicholson-Puthenveedu S, Tageldein MM, Khasnis S, Arrowsmith CH, Harding SM. 2023. Antecedent chromatin organization determines cGAS recruitment to ruptured micronuclei. *Nat Commun* **14**:556. doi:10.1038/s41467-023-36195-8
Maciejowski J, Li Y, Bosco N, Campbell PJ, Lange T de. 2015. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163**:1641–1654. doi:10.1016/j.cell.2015.11.054
Mackenzie KJ, Carroll P, Martin C-A, Murina O, Fluteau A, Simpson DJ, Olova N, Sutcliffe H, Rainger JK, Leitch A, Osborn RT, Wheeler AP, Nowotny M, Gilbert N, Chandra T, Reijns MAM, Jackson AP. 2017. cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **2**:1199–465. doi:10.1038/nature23449
Mahbod A, Schaefer G, Dorffner G, Hatamikia S, Ecker R, Ellinger I. 2022. A dual decoder U-Net-based model for nuclei instance segmentation in hematoxylin and eosin-stained histological images. *Front Med* **9**:978146. doi:10.3389/fmed.2022.978146
Mammel AE, Huang HZ, Gunn AL, Choo E, Hatch EM. 2021. Chromosome length and gene density contribute to micronuclear membrane stability. *Life Sci Alliance* **5**:e202101210. doi:10.26508/lsa.202101210

Matlashov ME, Shcherbakova DM, Alvelid J, Baloban M, Pennacchietti F, Shemetov AA, Testa I, Verkhusha VV. 2020. A set of monomeric near-infrared fluorescent proteins for multicolor imaging across scales. *Nat Commun* **11**:239. doi:10.1038/s41467-019-13897-6

Mohr L, Toufektchan E, Morgen P von, Chu K, Kapoor A, Maciejowski J. 2021. ER-directed TREX1 limits cGAS activation at micronuclei. *Mol Cell*. doi:10.1016/j.molcel.2020.12.037

Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla NY, Kainz B, Glocker B, Rueckert D. 2018. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv*. doi:10.48550/arxiv.1804.03999

Papathanasiou S, Mynhier NA, Liu S, Brunette G, Stokasimov E, Jacob E, Li L, Comenho C, Steensel B van, Buenrostro JD, Zhang C-Z, Pellman D. 2023. Heritable transcriptional defects from aberrations of nuclear architecture. *Nature* 1–9. doi:10.1038/s41586-023-06157-7

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**:417–419. doi:10.1038/nmeth.4197

Pons C, Mauvezin C. 2022. PIQUE: An ImageJ plugin for the quantification of toroidal nuclei in biological images. bioRxiv 2022.11.07.515391. doi:10.1101/2022.11.07.515391

Ronneberger O, Fischer P, Brox T. 2015. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III. *Lect Notes Comput Sci* 234–241. doi:10.1007/978-3-319-24574-4_28

Santaguida S, Richardson A, Iyer DR, M'Saad O, Zasadil L, Knouse KA, Wong YL, Rhind N, Desai A, Amon A. 2017. Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Developmental Cell* **41**:638-651.e5. doi:10.1016/j.devcel.2017.05.022

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**:676–682. doi:10.1038/nmeth.2019

Shoshani O, Brunner SF, Yaeger R, Ly P, Nechemia-Arbely Y, Kim DH, Fang R, Castillon GA, Yu M, Li JSZ, Sun Y, Ellisman MH, Ren B, Campbell PJ, Cleveland DW. 2021. Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**:137–141. doi:10.1038/s41586-020-03064-z

Soto M, García-Santisteban I, Krenning L, Medema RH, Raaijmakers JA. 2018. Chromosomes trapped in micronuclei are liable to segregation errors. *Journal of cell science* **131**:jcs.214742. doi:10.1242/jcs.214742

Stephens M. 2017. False discovery rates: a new deal. *Biostatistics* **18**:275–294. doi:10.1093/biostatistics/kxw041

Stetson DB, Ko JS, Heidmann T, Medzhitov R. 2008. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**:587–598. doi:10.1016/j.cell.2008.06.032

Su R, Zhang D, Liu J, Cheng C. 2021. MSU-Net: Multi-Scale U-Net for 2D Medical Image Segmentation. *Front Genet* **12**:639930. doi:10.3389/fgene.2021.639930

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102**:15545–15550. doi:10.1073/pnas.0506580102

Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD. 2014. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**:635–646. doi:10.1016/j.cell.2014.09.039

Tang S, Stokasimov E, Cui Y, Pellman D. 2022. Breakage of cytoplasmic chromosomes by pathological DNA base excision repair. *Nature* **606**:930–936. doi:10.1038/s41586-022-04767-1

Trivedi P, Steele CD, Au FKC, Alexandrov LB, Cleveland DW. 2023. Mitotic tethering enables inheritance of shattered micronuclear chromosomes. *Nature* **618**:1049–1056. doi:10.1038/s41586-023-06216-z

Willan J, Cleasby AJ, Flores-Rodriguez N, Stefani F, Rinaldo C, Pisciottani A, Grant E, Woodman P, Bryant HE, Ciani B. 2019. ESCRT-III is necessary for the integrity of the nuclear envelope in micronuclei but is aberrant at ruptured micronuclear envelopes generating damage. *Oncogenesis* **8**:29. doi:10.1038/s41389-019-0136-0

Witten IH, Frank E, Hall MA, Pal CJ, editors. 2017. Appendix B - The WEKA workbenchData Mining (Fourth Edition). Morgan Kaufmann. pp. 553–571. doi:10.1016/b978-0-12-804291-5.00024-6

Wu F, Fan J, He Y, Xiong A, Yu J, Li Y, Zhang Y, Zhao W, Zhou F, Li W, Zhang J, Zhang X, Qiao M, Gao G, Chen S, Chen X, Li X, Hou L, Wu C, Su C, Ren S, Odenthal M, Buettner R, Fang N, Zhou C. 2021. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. Nat Commun 12:2540. doi:10.1038/s41467-021-22801-0

Xia Y, Pfeifer CR, Zhu K, Irianto J, Liu D, Pannell K, Chen EJ, Dooling LJ, Tobin MP, Wang M, Ivanovska IL, Smith LR, Greenberg RA, Discher DE. 2019. Rescue of DNA damage after constricted migration reveals a mechano-regulated threshold for cell cycle. *Journal of Cell Biology* **218**:2545–2563. doi:10.1083/jcb.201811100

Yan X, Stuurman N, Ribeiro SA, Tanenbaum ME, Horlbeck MA, Liem CR, Jost M, Weissman JS, Vale RD. 2021. High-content imaging-based pooled CRISPR screens in mammalian cells. *J Cell Biol* **220**:e202008158. doi:10.1083/jcb.202008158

Zhang C-Z, Spektor A, Cornils H, Francis JM, Jackson EK, Liu S, Meyerson M, Pellman D. 2015. Chromothripsis from DNA damage in micronuclei. *Nature* **522**:179–184. doi:10.1038/nature14493

Acknowledgements

This work was supported by a National Institutes of Health grant (R35GM124766, awarded to E.M. Hatch), a National Human Genome Research Institute grant (RM1HG010461, awarded to D.M. Fowler), a training grant (T32CA009657, awarded to L. DiPeso), the Rita Allen Foundation Scholars program (awarded to E.M. Hatch), and the Fred Hutchinson Cancer Center Bioinformatics and Genomics cores (funded by NIH grant P30CA015704).

Competing interest statement

The authors declare they have no competing interests.

Materials and Methods

Plasmid construction

pLVX-EF1a-NLS-3xDendra2-blast was created by PCR of NLS-3xDendra2 from pLenti-CMV-Dendra2x3-P2A-H2B-miRFP using primer sequences 5'-caagtttgtacaaaaaagttggcaccATGG-3' and 5'-TTAGGAAAAATTCGTT-GCGCCGCTCCC-3', followed by ligation into pLVX-EF1a-blast. pLVX-EF1a-H2B-emiRFP703-neo was created by PCR of H2B-emiRFP703 from pH2B-emiRP703 (a gift from Vladislav Verkhusha, AddGene #136567) with primers 5'-ATGCCAGAGCCAGCGAAG-3' and 5'-TTAGCTCTCAA-GCGCGGTGATC-3', followed by ligation into pLVX-EF1a-neo.

Cell culture and construction of cell lines

hTERT RPE-1 cell lines were cultured in DMEM/F12 (Gibco) supplemented with 10% FBS (Sigma), 1% penicillin-streptomycin (Sigma), and 0.01 mg/mL hygromycin B (Sigma) at 5% CO2 and 37 ºC. U2OS, hTERT-human fetal fibroblasts (HFF), and HeLa H2B-GFP cells were cultured in DMEM (Gibco) supplemented with 10% FBS and 1% pen/strep at 10% CO2 at 37 ºC. For ATF3 validation, cells were incubated in 2 µg/mL doxorubicin hydrochloride (Fisher Sci) for 1 hour prior to fixation. For EGR1 validation, cells were incubated in 5 ng/mL hEGF (Peprotech) for 1 hour prior to fixation. For MN induction, cells were incubated 100 nM BAY1217389 (Msp1i, Fisher Sci) for the indicated times.

hTERT RPE-1 NLS-3xDendra2/H2B-emiRFP703 and U2OS NLS-3xDendra2/H2B-emiRFP703 cell lines were produced through serial transduction of lentiviruses. RPE1 and U2OS cells were validated by STR sequencing. Lentivirus was produced in HEK293T cells using standard protocols and filtered medium was added with polybrene (Sigma, #H9268) for transduction. Cells were selected with 10 µg/mL blasticidin (Invivogen) and 500 µg/mL active G418 (Gibco) and FACS sorted on an Aria II sorter (BD Biosciences) for the top 20% brightest double positive cells. hTERT RPE-1 NLS-3xDendra2-P2A-H2B-miRFP703 cells were created through viral transduction and FACs sorting for the brightest double positive population. HeLa-H2B cells were a gift from Dr. Daphne Avgousti (Fred Hutchinson Cancer Center) and were originally acquired from Millipore (SCC117). HFF cells were a gift from Dr. Denise Galloway (Fred Hutchinson Cancer Center) (Kiyono et al., 1998).

Microscopy

VCS experiments were performed on a Leica DMi8 widefield fluorescence microscope with Adaptive Focus outfitted with an i8 incubation chamber (Leica) with temperature (PeCon: TempController 2000-1) and gas control (Oko) and a Mosaic 3 Digital Micromirror (Andor). Images were acquired with a 20x 0.8 NA apochromatic objective (Leica) using an iXon Ultra 888 EMCCD camera and MetaMorph v7.1.0.1.161 (Molecular Devices).

Fixed cell training and validation images were acquired with a Leica DMi8 laser scanning confocal microscope using the Leica Application Suite (LAS X) software and a 40x/1.15 NA Oil APO CS objective (Leica) or on a Leica DMi8 microscope outfitted with a Yokogawa CSU spinning disk unit, Andor Borealis illumination, ASI automated Stage with Piezo Z, with an environmental chamber and Automatic Focus using a 40x/1.3 NA Oil PLAN APO objective. Images on the spinning disk microscope were captured using an iXon Ultra 888 EMCCD camera and MetaMorph software (v7.10.4).

Micronucleus segmentation and cell classifiers

VCS MN: The neural net was created using the FastAI 1.0 library in Python, a UNet with Torchvision's ResNet18 pre-trained model as its base architecture (Ronneberger et al., 2015). Training for MN recognition was performed using ~2,000 images of individual cells as training data, a further 164 for validation, and 177 for testing. Training images were of RPE-1 NLS-3xDendra2-P2A-H2B-miRFP703 cells after incubation in 0.5 μM reversine (an Mps1 inhibitor, EMD Millipore) or DMSO for 24 h and taken with a 20x widefield objective on the VCS microscope. Nuclei were segmented on H2B channel images using the Deep Retina neural net (Caicedo et al., 2019) and 48x48 px image crops were generated centered on each nucleus. For training, MN pixels in cropped images were manually annotated. MN associated with chromatin bridges were ignored to ensure that labeled MN were discrete nuclear compartments.

The VCS MN classifier takes as input a 2-channel 20x image. It applies the Deep Retina neural net to the H2B channel to segment nuclei, discards any touching the edge of the image, and generates a 48x48 px crops centered on each nucleus. Each crop is processed with Sobel edge detection and linearly enlarged to 96x96 px. To accommodate the ResNet18 3-channel architecture, each crop is expanded to the H2B channel, a duplicate of the H2B channel, and the results of Sobel edge detection. Identified MN are mapped back to the full image and assigned to the closest segmented nucleus. MN more than 40 px away from a nucleus are discarded.

Once MN are assigned to cells, the classifier calculates the maximum Dendra2 MN/nucleus intensity ratio for each MN. MN with a ratio below 0.16 are classified as ruptured. This threshold was identified using the JRip classifier in Weka 3.8.6 to define the optimal threshold to separate manually annotated intact and ruptured MN (Cohen, 1995; Witten et al., 2017). Nuclear segments are classified as MN+ or MN- cells based on the presence or absence of an associated MN segment. MN+ cells are then further classified into those with only intact MN (rupture-) or those with at least one ruptured MN (rupture+).

For analysis of MN recall and PPV, MN were segmented using PixelStudio 4.5 on an iPad (Apple). Recall was calculated as the proportion of all MN that overlapped with a predicted segment. Positive predictive value was calculated as the proportion of all predicted segments that overlapped with a MN. Mean Intersection over Union (mIoU) was calculated per object by quantifying the overlap between groups of true positive pixels and their respective ground truths.

For analysis of U2OS cells, the VCS MN segmentation module was retrained on a collection of images of RPE1, U2OS, HFF, and HeLa cells after incubation in 100 nM BAY1217389 or 0.5 μM reversine for 24 h. Live images of RPE1 and U2OS NLS-3xDendra2/H2B-emiRFP703 cells were acquired on the VCS microscope at 20x. Images of fixed cells were taken on either the LSM or spinning disk confocal microscopes at 40x after fixation in 4% paraformaldehyde (Electron Microscopy Sciences, #15710) for 5 min at room temperature. Cells were labeled with DAPI as indicated. ~2,300 crops of U2OS NLS-3xDendra2/H2B-emiRFP703 cells taken on the VCS microscope at 20x were used for training with another 233 held back for validation and 910 for testing. Three images of Hoechst labeled U2OS cells taken at 20x on a widefield microscope at 16 bit depth were downloaded from the Broad Bioimage Benchmark Collection (BBBC039v1, (Bray et al., 2016; Caicedo et al., 2019b; Ljosa et al., 2012) and linearly scaled by 0.5. Crops were generated centered on manually-annotated cell nuclei and fed to VCS MN to determine PPV, recall, and mIoU for this data set.

MNFinder: The MNFinder neural nets were created using TensorFlow 2.0 without transfer learning. Training was performed using 128x128 px crops generated from the same training and validation data used for retraining the VCS MN.

For nucleus/MN segmentation (semantic segmentation) predictions are taken from two UNet-based neural nets, with MN predictions fed into a third ensembling UNet. All UNets are trained independently but are otherwise identical, save for the incorporation of multiscale downsampling into one of the input UNets. For cell segmentation (instance segmentation) a UNet architecture incorporating 3 decoder pathways is used to predict distance maps, proximity maps, and foreground pixels. The distance and proximity map decoders incorporate features from a UNet3+ design: specifically, additional skip connections from multiple layers of the encoder and decoder pathways and deep supervision during training (Huang et al., 2020). Training data were generated from annotated nuclei and MN images by generating a concave hull grouping a nucleus and associated MN using the cdBoundary package in Python (Duckham et al., 2008). This hull was transformed into a distance map by calculating the Euclidean distance transform (EDT) with each pixel value encoding the shortest distance between that pixel and the background. Proximity maps were generated by setting all pixels as foreground pixels except for those belonging to other hulls and applying an EDT, masked by the cell's boundaries, and raising this to the 4th power to sharpen edges. Both maps are scaled from 0–1 for each cell.

MNFinder input images taken at 20x are cropped using a 128x128 px sliding window, advancing the window by 96 px horizontally and vertically to oversample the image. 40x images are scaled down by a factor of 2 prior to input. Crops are expanded into 2-channel images, with the second channel the result of Sobel edge detection. These images are processed by the neural nets, post-processed as described, and reassembled by linear blending into a complete field. Recall and positive predictive values were calculated using the same as for VCS classifier validation.

MNFinder was validated on images of RPE1, U2OS, HFF, and HeLa cells after incubation in 100 nM BAY1217389 or 0.5 μM reversine for 24 h. Live images of RPE1 and U2OS NLS-3xDendra2/H2B-emiRFP703 cells were acquired on the VCS microscope at 20x. Images of fixed cells were taken on either the LSM or spinning disk confocal microscopes at 40x after incubation in 4% paraformaldehyde (Electron Microscopy Sciences, #15710) for 5 min at room temperature. PPV and recall for MN segmentation were calculated for individual input UNets and the ensemble UNet.

Outline of VCS MN cell isolation experiments

Cells for VCS were plated onto 6-well glass-bottom, black-walled plates at a density of 50,000–225,000 cells per well 1-2 days before activation. An extra unactivated well was plated as a control. One day before imaging, 100 nM Mps1i was added to the medium. One hour prior to imaging, cells were washed 1x in PBS and medium changed to phenol red free (GIBCO) containing 10 μM RO-3306 (Sigma). The plate was transferred to the microscope, the plate center and micromirror device were aligned, and the appropriate journals (see (Hasle et al., 2020)) were initiated for VCS activation. Imaging conditions were optimized for each experiment. Images were acquired using MetaMorph and analyzed on a dedicated linked computer. 1-bit masks of MN+ nuclei and MN- nuclei were transmitted back to MetaMorph, which directed UV pulses at the segmented nuclei. Activation occurred using either a 200 ms or 800 ms pulse of the 405 nm laser. After imaging, the initial 5 positions and last 5 positions were reimaged for quality control as well as 5 random positions in the unactivated well. Classifier predictions were compared to the first 3 and last 3 images from each VCS experiment, each manually annotated prior to downstream analysis, including RNA extraction.

Activated and unactivated cells were trypsinized, suspended in 2% FBS, and sorted using a FACS Aria II (BD Biosciences). Compensation for PE-blue excitation of unconverted Dendra2 was performed on the unactivated cells. Dendra2 activation-based sorting gates were defined on single cells positive for both Dendra2 and emiRFP703 using the PE-Blue-A/FITC-A ratio. Cells were sorted into 2% FBS then pelleted and either flash frozen on dry ice or replated onto poly-L-lysine coated coverslips.

CellTrace: Activation and sorting accuracy were analyzed for RPE1 RFP703/Dendra2 cells by incubating cells in CellTrace far red (ThermoFisher) for 10 min at 37 ºC, trypsinizing and pelleting cells, mixing 1:1 with unlabeled cells, and plating. A classifier segmented nuclei with the Deep Retina neural net on the GFP channel and measured the mean far-red intensity in the nucleus (Hasle et al., 2020). Threshold intensity for activation was experimentally determined. Cells were sorted by FACs for Dendra2 ratio and CellTrace intensity using compensation to eliminate emiRFP703 spectral overlap and then reanalyzed on the same machine.

Mps1i+/- isolation: Cells incubated in Msp1i or DMSO were imaged and activated using a random classifier. 1-bit masks of nuclei generated using the Deep Retina neural net were randomly assigned to receive 800 ms or 200 ms pulses. At least 13k cells were collected per sorting bin and samples were pelleted, flash frozen, and stored at -80°C.

Micronucleus+/- isolation: Two wells were imaged sequentially per experiment with the activation time for MN+ and MN- nuclei reversed between wells.

Rupture+/- isolation: Cells were plated 2 days before imaging in medium containing 1 μM Cdk4/6i (PD-0332991, Sigma). Twenty-four hours later, cells were rinsed 3x with PBS and the medium replaced with 100 nM BAY.

Only 1 well was imaged per experiment with rupture- cells receiving 800 ms and rupture+ cells receiving 200 ms pulses.

RNA isolation and sequencing

We extracted RNA from frozen cell pellets using the RNAqueous micro kit (ThermoFisher), according to the manufacturer's protocol. Residual DNA was removed by DNase I treatment and RNA was further purified by glycogen precipitation (RNA-grade glycogen; ThermoFisher) and resuspension in ultra-pure H_2O heated to 65 ºC. RNA quality and concentration was checked by the Genomics Core at the Fred Hutchinson Cancer Center with an Agilent 4200 Tapestation HighSense RNA assay and only samples with RIN scores above 8 and 28S/18S values above 2 were further processed. cDNA synthesis and library preparations were performed by the Genomics Core using the SMARTv4 for ultra-low RNA input and Nextera XT kits (Takara). Sequencing was also performed on an Illumina NextSeq 2000 sequencing system with paired-end, 50 bp reads.

RNAseq and gene-set enrichment analysis

We quantified transcripts with Salmon to map reads against the UCSC hg38 assembly at http://refgenomes.databio.org (digest: 2230c535660fb4774114bfa966a62f823fdb6d21acf138d4), using bootstrapped abundance estimates and corrections for GC bias(Patro et al., 2017). For comparisons with data from He, et al. and Santaguida, et al., the original FASTA files deposited at the Sequence Read Archive were downloaded with NCBI's SRA Toolkit and quantified with Salmon (He et al., 2019; Santaguida et al., 2017). No GC-bias correction was applied as only single-end reads were available.

Transcript abundances were processed to find differentially expressed genes (DEGs) with the R package DESeq2 version 3.16 in R 4.2.1, RStudio 2022.07.2 build 576, and Sublime Text build 4143. Files were imported into DESeq2 with the R package tximeta (Love et al., 2020, 2014), estimated transcript counts were summarized to gene-level, and low-abundance genes were filtered by keeping only those genes with estimated counts ≥ 700 in at least 2 samples. DEGs were identified using a likelihood ratio test comparing the full model with one with the condition of interest dropped and an FDR of 0.05. Log-fold changes were corrected using empirical Bayes adaptive shrinkage (Stephens, 2017). Operations were performed before pseudogenes were filtered from dataset.

GSEA was performed using the R package fgsea version 1.25.1, comparing log-fold changes of all DEGs against the full Homo sapiens Hallmark Gene Sets version 2022.1, part of the MSigDB resource (UC San Diego, Broad Institute) (Crameri, 2018; Greene et al., 2017; Korotkevich et al., 2021; Liberzon et al., 2015; Subramanian et al., 2005).

Live-cell imaging for MN rupture frequency analysis

RPE1 NLS-3xDendra2/H2B-emiRFP703 cells were plated 2 days before imaging and treated for 24 hours with either 1 µM Cdk4/6i or DMSO. One day before imaging, cells were rinsed and incubated in 100 nM BAY1217389. Nineteen hours later, the media was exchanged for Cdk1i medium, 5 positions were imaged in each well and rupture- cells were activated. These positions and the surrounding area were imaged every hour for 11 hours and the status of photoconverted cells manually recorded.

Immunofluorescence (IF)

Cells plated on poly-L-lysine coated coverslips or glass bottomed plates were fixed for IF in 4% paraformaldehyde for 5 min unless otherwise indicated. Cells were permeabilized for 30 min at RT in PBSBT (1xPBS (GIBCO), 3% BSA, 0.4% Triton X-100, 0.02% sodium azide (all Sigma)), followed by incubation in primary antibodies diluted in PBSBT for 30 min, secondary antibodies diluted in same for 30 min, and 5 min in 1 µg/mL DAPI (Invitrogen). Coverslips were mounted in VectaShield (VectorLabs) and sealed with nail polish before imaging. Primary antibodies used were: mouse-a-γH2AX (1:500; BioLegend, 613401), rabbit-a-ATF3 (1:400; Cell Signaling Technology, 18665), rabbit-a-EGR1 (1:1600; Cell Signaling Technology, 4154) and rabbit α H3K27Ac (2 µg/mL; Abcam, ab4729). Secondary antibodies used were: AF647 goat-a-mouse (1:1000; Life Technologies, A21236) and AF488 goat-a-rabbit (1:2000; Life Technologies, A11034).

DNA FISH

RPE1 cells plated onto poly-L lysine coverslips were fixed in -20ºC 100% methanol for 10 min, rehydrated for 10 min in 1xPBS and processed for IF. Cells were then refixed in 4% PFA for 5 min at RT then incubated in 2xSSC (Sigma) for 2 x 5 min RT. Cells were permeabilized in 0.2 M HCl (Sigma), 0.7% TritonX-100 in H2O for 15 min at RT, washed in 2xSSC, and incubated for 1 h at RT in 50% formamide (Millipore). Cells were rewashed in 2xSSC, inverted onto chr 1, 11, or 18 XCE probes (MetaSystems), and the coverslips sealed with rubber cement. Probes were hybridized at 74ºC for 3 min and then incubated for 4 hours (chr 18) or overnight (chrs 1 and 11) at 37 ºC. After hybridization, coverslips were washed in 0.4xSSC at 74 ºC for 5 min, then 2xSCC 0.1% Tween20 (Fisher) for 2 x 5 min at RT. DNA was labeled by incubation in 1 µg/mL DAPI for 5 min at RT, and coverslips mounted in VectaShield. Images were acquired as 0.45 µm step z-stacks through the cell on the confocal LSM with a 40x objective. Cells that had more or less than two FISH foci were classified as aneuploid for that chromosome.

Image analysis

Dendra2 ratio stability: Nuclei were segmented on images taken at the start and end of an Mps1i +/- VCS experiment by thresholding on the GFP channel, measuring the mean intensity of GFP and RFP, and calculating the RFP:GFP ratio per nucleus for each image group.

MN+/- sorting accuracy: Cells replated and fixed after sorting were imaged on the LSM confocal at 40x with 0.45 µm z-stacks through the cell. Image names were randomized prior to quantification of MN+ cells.

ATF3 and EGR1 intensity: Images were acquired as 0.45 µm step z-stacks through the cell on the confocal LSM with a 40x objective. Images were corrected for illumination inhomogeneity by dividing by a dark image and background subtracted using a 60 px radius rolling-ball in FIJI (Schindelin et al., 2012) (v2.9.0). Single in focus sections of each nucleus was selected and nuclei masks generated by thresholding on RFP-NLS. Mean intensity of ATF3 or EGR1 were calculated for each nucleus and normalized for each replicate by scaling to the median value for the DMSO control. Statistics were calculated on the raw values.

Statistical analyses

Shorthand p-values are as follows:
ns: p-value >= 0.05
*: p-value < 0.05
**: p-value < 0.01
***: p-value < 0.001
****: p-value < 0.0001

Generalized estimating equations (GEE) were used to determine statistical differences for nominal data with multiple variables using binomial distributions and a logit link function (Halekoh et al., 2006). For Fig. 1C–D, data were assessed using the formula: (# recalled, # missed) ~ MN status where MN status is whether ruptured or intact. For Fig. 5E, we also used a binomial distribution and a logit link function. For Fig. 5I–J and S5F–G, we used the formula: (# aneuploid, # normal) ~ Status × Chr where Status is whether the cell was MN+/- (Figs 5I, S5F) or Rupture+/- (Figs 5J, S5G) and Chr is chromosome identity. p-values for each individual property were calculated using the drop1 function in R. In Fig. 6D–E and S6 C–D, we used a gamma distribution and the formula mean intensity ~ Population. Statistical significance for differences between single nominal variables in other figures were by Barnard's exact test.

The predicted change to classifier PPV in Fig. 5E was determined by reducing the true positive rate in the rupture- population by the difference in mean rupture frequencies between the beginning and end of the experiments and increasing the true positive rate in the rupture+ population by the same.

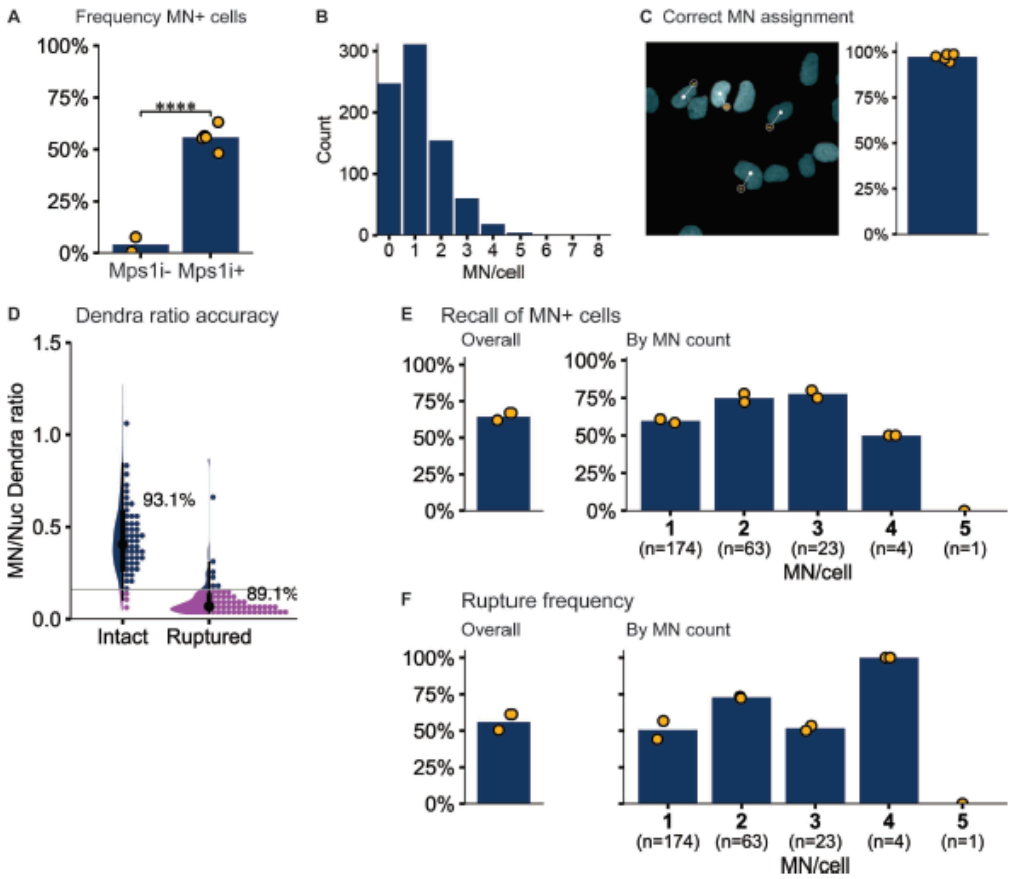

Figure S1: Supporting data for figure 1. (A) Micronucleation frequencies in hTERT RPE-1s treated with either DMSO (Mps1-) or the Mps1i-inhibitor BAY-1217389 (Mps1i+) for 20 hrs. N = 7, n = 328, 186, 175, 344, 228, 237, 262. **** = $p < 0.0001$ by GEE. (B) Histogram of number of MN/cell in Mps1i+ RFP703/Dendra cells. N = 5, n = 1323. (C) Quantification of proportion of MN assigned to the correct nucleus by proximity alone. N = 5, n = 264, 158, 365, 283, 249. (D)Distributions of MN/nucleus Dendra2 intensity ratio for intact and ruptured MN. Solid gray line = calculated threshold. N=3, n=179, 113, 105. E, F) Recall and rupture frequency in MN+ cells by # MN. Cells manually classified. N=2, n=328, 186.

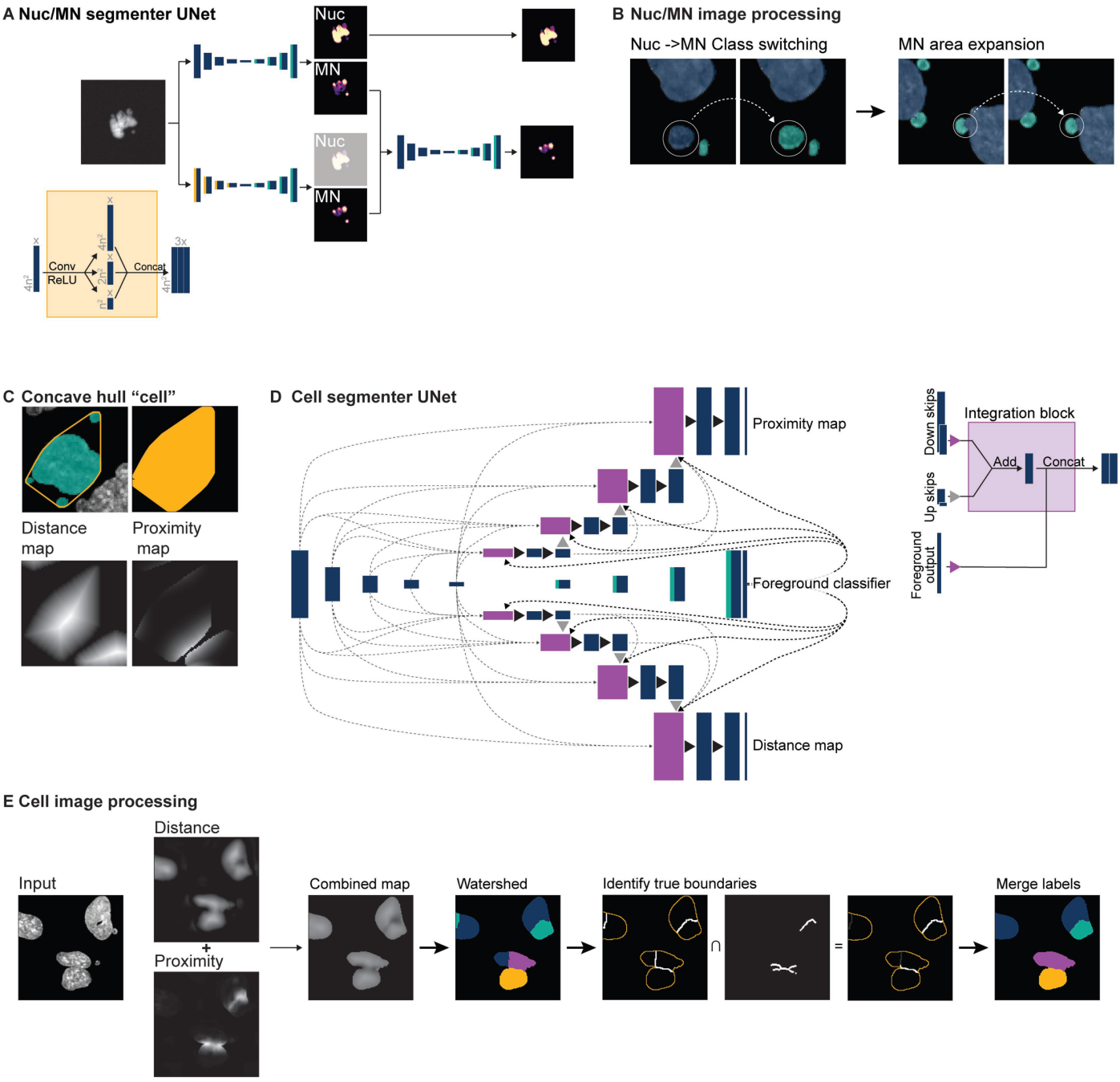

Figure S2: Details of UNet architectures and output post-processing in MNFinder module.
(A) The Nuc/MN ensemble classifier takes a single channel input image of chromatin and feeds it into two parallel, attention-gated UNets, one of which also has multiscale downsamplers (yellow). In these blocks input is fed into three parallel, differently-sized convolution operations that are then concatenated. The nucleus weights from the basic UNet are retained and both sets of MN weights are fed to a third UNet for ensembling to produce the final predictions. (B) Results from the Nuc/MN UNet are further processed to improve accuracy. To limit misclassification of large MN as small nuclei, nuclei under a user defined area threshold are reclassified as MN. To limit MN undersegmentation, MN pixel groups are expanded by transforming each into their convex hulls. (C) Example of how a "cell" is generated from existing training data by defining a concave hull that groups a nucleus and any associated MN. Distance and proximity maps are used to define cell boundaries and are derived from convex hulls for training as described in Methods. (D) Diagram of the triple decoder cell segmenter UNet. Two of the decoders have a UNet3+-like architecture with multiple skip connections and deep supervision during training. Feature depths are kept constant and most concatenation/max-pooling operations are replaced with addition to reduce training overhead. One decoder generates distance maps of a concave hull containing each nucleus and any associated MN (a "cell") and the other generates a proximity map of each cell's distance to any other. The output of a third decoder that uses a standard UNet with attention gates to segment foreground pixels (nuclei or MN) is used as input into every level of the distance- and proximity-map decoders via an integration block (magenta). (E) Resulting distance and proximity maps from the cell segmenter UNet are combined to generate seeds for watershed segmentation. To correct for oversegmentation, only labels with boundaries that intersect a skeletonized proximity map or border background pixels are retained.

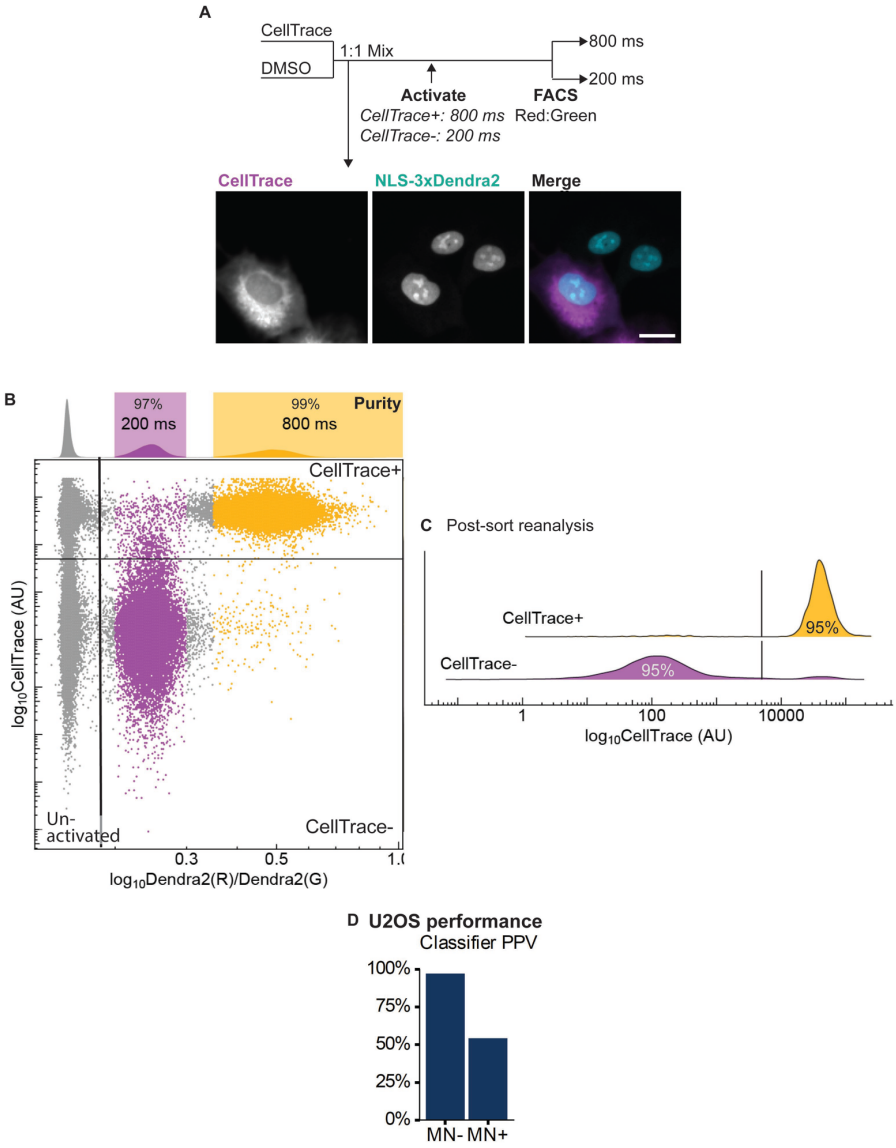

Figure S3. Controls for VCS MN isolation experiments. (A) Outline of RFP703/Dendra VCS validation experiment using CellTrace labeling as the activation trigger. Cells were incubated with CellTrace far-red and mixed with unlabeled cells at a 1:1 ratio. Nuclei were classified based on CellTrace fluorescence intensity and converted with either an 800 ms (CellTrace+) or 200 ms (CellTrace-) UV pulse. The well was only partially converted prior to FACs analysis and sorting. Representative image of the mixed population prior to photoconversion is shown. Scale bar = 10 μm. (B) FACS plot of Dendra2 red:green ratio versus CellTrace fluorescence. Colored bars represent gates. Values are percentage of negative and positive CellTrace cells present in 200 ms and 800 ms gate, respectively. (C) Histogram of CellTrace fluorescence in cells sorted by Dendra2 ratio after re-analysis by FACs. (D) Predicted classifier PPV (population purity) for untreated low MN frequency U2OS cells (U2OS Broad). We observe a lower but still substantial enrichment of micronucleated cells in the MN+ population compared in a high MN frequency population (Fig. 3C). N = 1, n = 17 cells.

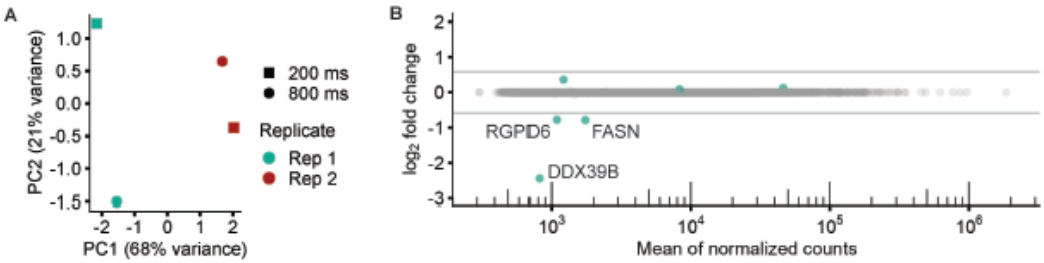

Figure S4: Differential UV pulses do not induce substantial transcriptional changes. (A) PCA plot of cells treated with DMSO and exposed to 800 ms or 200 ms UV. (B) MA plot of the data in A). Only 6 differentially expressed genes were identified in cells exposed to 800 ms vs 200 ms UV and only 3 were downregulated over 1.5 fold: DDX39B, FASN, RGPD6.

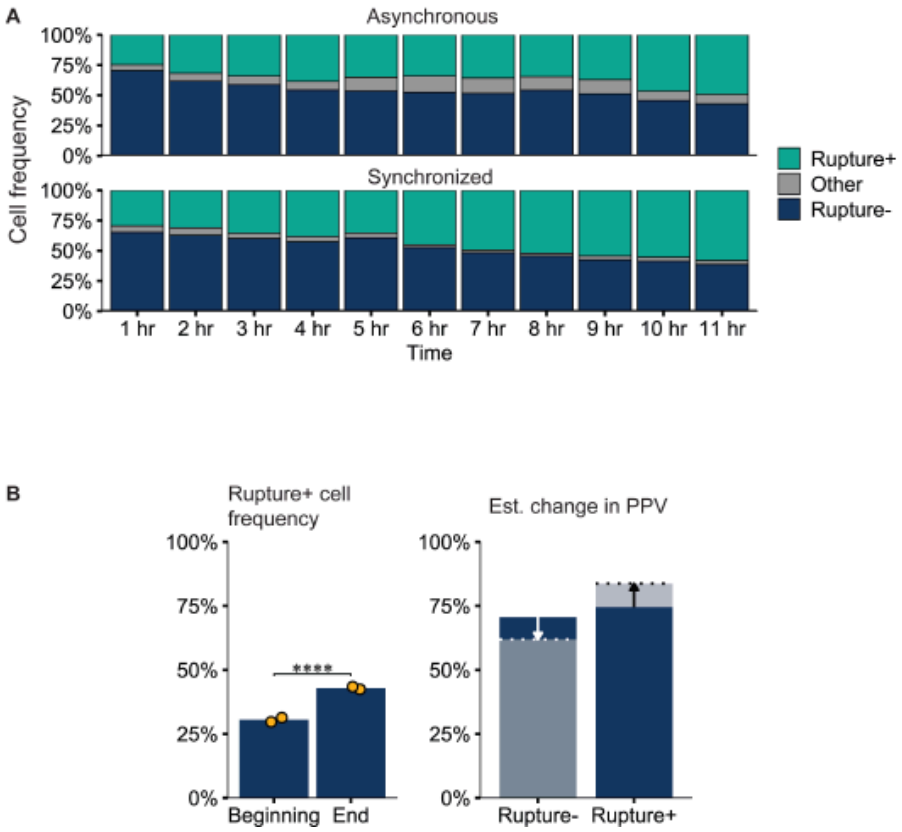

Figure S5: Cell synchronization reduces loss of intact MN cell population purity. (A) Change in rupture frequency over time in asynchronous and synchronized cells treated with Cdk1i. Other = mitotic, MN-, or Dendra2- cells. N=1, n=~200 cells per time point. (B) Change in MN rupture frequency between the start and end of a VCS experiment (4h) and predicted change in classifier PPV due to ongoing rupture of intact MN based on values in A).

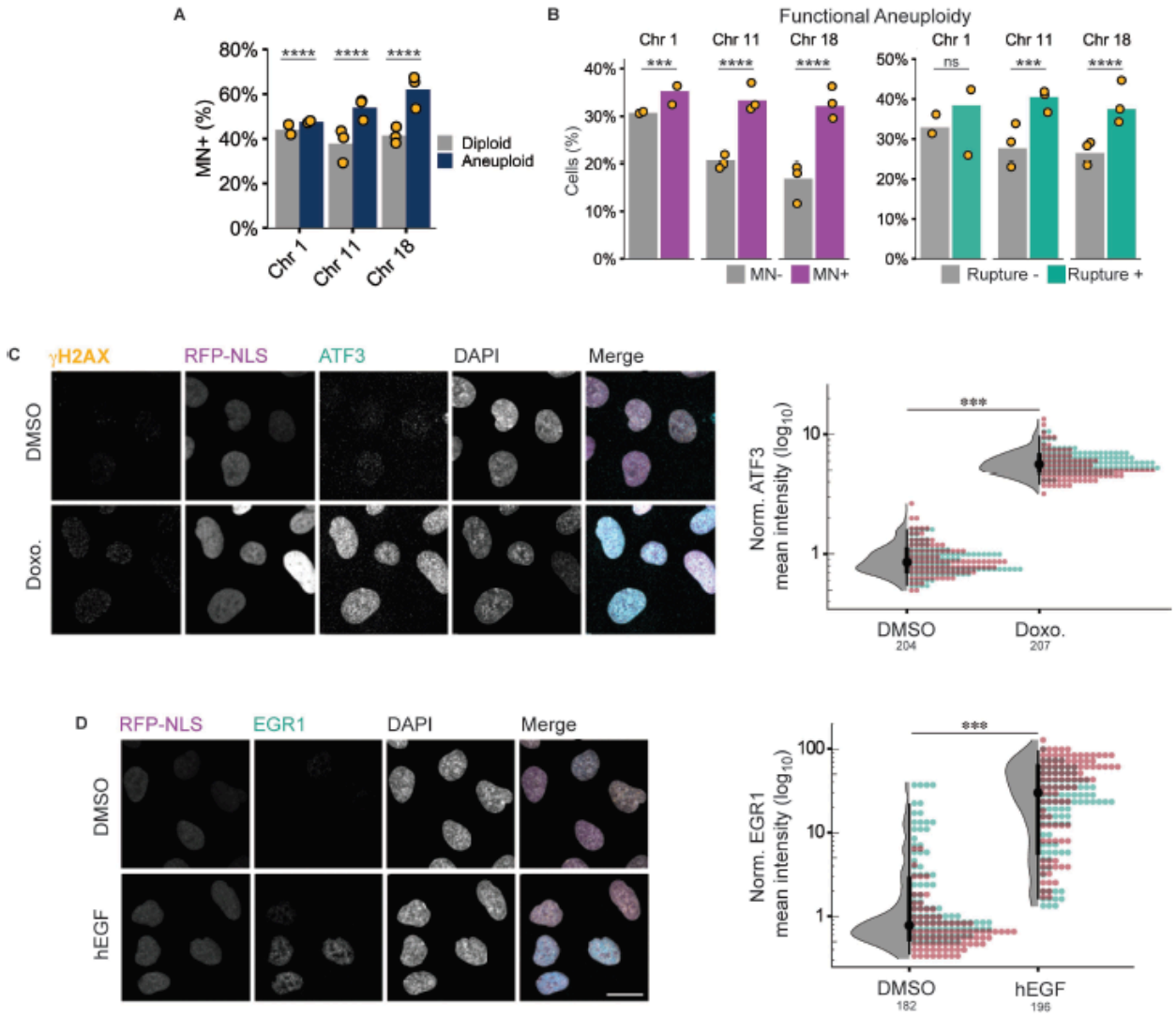

Figure S6: Controls related to Figure 6. (A) Quantification of chromosome 1, 11, and 18 micronucleation rates in RPE1 Mps1i cells, grouped by chromosome ploidy. N=2, 3, 3. n = 587, 879, 857. p: **** ≤ 0.0001. (B) Same analysis as Fig. 6b, but with chromosomes in ruptured MN excluded from the foci count. Similar levels of aneuploidy were observed between groups as in Fig. 6b. (C) Representative images and quantification of ATF3 nuclear mean fluorescence intensity in cells treated with DMSO or doxorubicin (Doxo.). N = 2 (colors on graph), n = on graph. (D) Representative images and quantification of EGR1 nuclear mean fluorescence intensity in cells treated with DMSO or hEGF. N = 2 (colors on graph), n = on graph. p: *** ≤ 0.001, by GEE. Scale bar = 20 μm.

List of Supplementary Tables

Table 1: Model testing. Comparison of NN model metrics for MN segmentation across cell lines

Table 2: MN frequency. Quantification of MN frequency (percentage of cells with at least 1 MN) in images used to evaluate MNFinder.

Table 3: MN analysis comparison. Describes the flexibility, sensitivity, specificity, and segmentation metrics of MNFinder compared to existing automated programs that quantify MN frequency in adherent cell images.

Table 4: DMSO 800 vs 200 DEG. List of differentially expressed genes (FDR ≤ 0.05) in DMSO treated RPE1 cells isolated after an 800 ms or 200 ms UV pulse. For all DEG analyses: padj = false-discovery rate adjusted p-values.

Table 5: Msp1i DEG. List of differentially expressed genes (FDR ≤ 0.05) in Mps1i versus DMSO RPE1 cells.

Table 6: Msp1i DEG log2FC. List of differentially expressed genes in Mps1i versus DMSO RPE1 cells filtered for $\log_2$ fold change above 0.58 or below -0.58.

Table 7: He et al DEG. List of differentially expressed genes (FDR ≤ 0.05) in RPE1 cells treated with nocodazole for 8 h versus control, initially reported in (He et al., 2018). Overlap with Mps1i DEG list noted in last column.

Table 8: Santaguida et al DEG. List of differentially expressed genes (FDR ≤ 0.05) in RPE1 cells that were treated with the Mps1i molecule reversine for 12 h, released, and ceased to divide, compared to control cells. Results initially reported in (Santaguida et al., 2017). Overlap with Mps1i DEG list noted in last column.

Table 9: Hallmark Mps1i. List of MSigDB Hallmark gene sets that are significantly enriched (FDR ≤ 0.05) in Mps1i treated RPE1 cells. For all Hallmark lists, NES = normalized enrichment score, padj = false discovery rate adjusted p-values, size = number of genes in gene set, leading_edge = , sig_genes_in_geneset = the gene names of set genes that were significantly upregulated.

Table 10: Hallmark He et al. List of MSigDB Hallmark gene sets that are significantly enriched in nocodazole treated RPE1 cells from (He et al., 2018).

Table 11: Hallmark Santaguida et al. List of MSigDB Hallmark gene sets that are significantly enriched in reversine treated RPE1 cells from (Santaguida et al., 2017).

Table 12: MN+ DEG. List of differentially expressed genes (FDR ≤ 0.05) in Mps1i treated RPE1 cells with MN versus without MN.

Table 13: MN+ DEG log2FC. List of differentially expressed genes in Mps1i treated RPE1 cells with MN versus without MN filtered for $\log_2$ fold change above 0.58 or below -0.58. Overlap with Mps1i DEG list noted in last column.

Table 14: Ruptured+ DEG. List of differentially expressed genes (FDR ≤ 0.05) in Mps1i treated micronucleated RPE1 cells with only intact MN versus at least 1 ruptured MN.

Table 15: Ruptured DEG+ log2FC. List of differentially expressed genes in Mps1i treated micronucleated RPE1 cells with only intact MN versus at least 1 ruptured MN MN filtered for absolute $\log_2$ fold change ≥ 0.58. Overlap with Mps1i DEG list noted in last column.

Table 16: Hallmark MN+. List of MSigDB Hallmark gene sets that are significantly enriched in Mps1i treated RPE1 cells with MN.

Table 17: Hallmark ruptured+. List of MSigDB Hallmark gene sets that are significantly enriched in Mps1i treated RPE1 cells with MN and at least 1 ruptured MN versus no ruptured MN.

Table 18: Log_2FC per replicate. List of $\log_2$ fold change values for Mps1i DEGs with absolute $\log_2$FC ≥ 0.58 broken out by MN+ and rupture+ replicate. Genes with all NA values for either all MN+ or all rupture+ replicates were excluded from analysis. enriched_in_rupture+_cluster = genes present in cluster enriched in increased expression over Mps1i DEGs.