Machine vision reveals micronucleus rupture as a potential driver of the transcriptomic response to aneuploidy

DiPeso, Lucian[1,4]; Pendyala, Sriram[3]; Huang, Heather Z[1]; Fowler, Douglas M[3]; Hatch, Emily M[1,2*]

1. Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
2. Human Biology Division, Fred Hutchinson Cancer Center, Seattle, WA, USA
3. Genome Sciences, University of Washington, Seattle, WA, USA
4. Molecular & Cellular Biology, University of Washington, Seattle, WA, USA

* Corresponding author
Emily Hatch
E-mail: ehatch@fredhutch.org

Keywords: Micronuclei, nuclear membrane rupture, neural net, image segmentation, aneuploidy, transcriptome

Abstract

Micronuclei are aberrant nuclear compartments that trap a portion of a cell's chromatin in a distinct organelle separate from the nucleus and are drivers of inflammation, DNA damage, chromosome instability, and chromothripsis. Many of the consequences of micronucleus formation stem from micronucleus rupture: the sudden loss of micronucleus compartmentalization, resulting in mislocalization of nuclear factors and the exposure of chromatin to the cytosol for the remainder of interphase. Micronuclei form primarily from segregation errors during mitosis, errors that also give rise to other, non-exclusive phenotypes, including aneuploidy and chromatin bridges. The stochastic formation of micronuclei and phenotypic overlap confounds the use of population-level assays or hypothesis discovery, requiring labor-intensive techniques to visually identify and follow micronucleated cells individually. In this study, we present a novel technique for automatically identifying and isolating micronucleated cells generally and cells with ruptured micronuclei specifically using a *de novo* neural net combined with Visual Cell Sorting. As a proof of concept, we compare the early transcriptomic responses to micronucleation and micronucleus rupture with previously published responses to aneuploidy, revealing micronucleus rupture to be a potential driver of the aneuploidy response.

Introduction

Micronuclei (MN) are aberrant nuclear compartments that trap some portion of a cell's chromatin in a spatially distinct organelle separate from the nucleus. They are produced principally via segregation errors or the presence of unrepaired DNA breaks during mitosis.[1] Micronuclei have been observed since the late 19th century, but in recent years, micronuclei were discovered to be key drivers of chromothripsis, chromosome instability, inflammation, and even embryonic mosaicism.[2–7] MN resemble nuclei in general structure and function—they possess a double-membrane envelope punctuated with nuclear pore complexes, undergirded with a nuclear lamina, and perform transcription and DNA replication. However, nuclear pore complexes are often at a significantly lower density than the nucleus, nuclear lamina may contain gaps, DNA replication is often delayed, and DNA damage repair inefficient.[1,8–10]

A major defect is their resistance to nuclear envelope repair during interphase.[11,12] MN frequently undergo nuclear membrane rupture during interphase and rarely regain compartmentalization. This results in mislocalization of soluble nuclear factors to the cytoplasm; the arrest of nuclear processes such as DNA damage recognition, replication, or transcription; and the exposure of contained chromatin to the cytoplasm for the rest of interphase.[5,10,11] Exposure to the cytoplasm persistently alters histone modifications, may activate the canonical DNA-sensing cGAS/STING pathway, and can result in damage from nucleases, leading to chromothripsis and kataegis.[5,8,13–16]

During chromothripsis, a micronucleated chromosome is fragmented during micronucleus rupture and subsequent mitosis and is then religated during the next interphase, leading to inversions, translocations, deletions, and the formation of circular double-minutes.[7,10,17,18]

Micronuclei have been the focus of intense research in the past two decades, work largely done with cultured cells treated to induce mitotic errors—*eg*, microtubule poisons or Mps1 inhibitors (Mps1i).[7,8,11,19,20] However, these mitotic errors only stochastically induce micronuclei, while also producing other, overlapping phenotypes, including aneuploidy, chromatin bridge formation, and misshapen nuclei.[19,21–23] This phenotypic overlap confounds our ability to use population-level assays or bulk screens to identify the specific consequences of micronucleation and micronucleus rupture. Prior research has instead largely relied on labor-intensive protocols dependent on finding individual micronucleated cells and tracking them over time. Because of their time- and labor-intensive nature, these protocols limit our ability to perform hypothesis-generating, rather than hypothesis-driven, experiments.

Current methods to circumvent these limitations include the specific missegregation of an artificial chromosome or the Y-chromosome in a cell population.[10,17,24] These methods allow bulk analyses, but are currently restricted to specific chromosomes in specific cell lines.

In this study, we sought to develop a general protocol to identify and isolate micronucleated cells, and to further distinguish cells with ruptured micronuclei versus those with intact micronuclei. We opted to develop a neural net that could segment micronuclei from micrographs taken with only a 20x widefield objective. We coupled this neural net with Visual Cell Sorting (VCS), a technique developed to isolate a specific population of cells identified by a visual phenotype using FACS in order to profile the early transcriptomic response to micronucleation and micronucleus rupture in an aneuploid cell population.[25] Finally, we are making available an updated micronucleus segmentation neural net, tested on a variety of cell lines and imaging conditions, to accelerate future research on micronucleation and micronucleus rupture.

Results

Machine vision identifies micronucleated RPE-1 cells within a mixed population

We developed an automated pipeline for identifying micronucleated cells in a mixed population of hTERT RPE-1 (RPE-1) cells. RPE-1 cells are a near-diploid, non-transformed human cell line not prone to micronucleation. In addition, the RPE-1 transcriptomic response to aneuploidy is already well studied, providing a convenient background to examine the specific effects of micronucleation and micronucleus rupture.[19,23,26] Treatment with a low dose of the Mps1 inhibitor (Mps1i) BAY-1217389 causes whole chromosome missegregation and micronucleation in approximately 50% of the cells (**Fig S1a**) with a modal MN number of 1 per cell in MN+ cells (**Fig S1b**). To identify intact and ruptured micronuclei, we transduced RPE-1 cells with H2B fused to emiRFP703—a far-red fluorophore—and a 3xDendra2, a photoconvertible fluorophore

repeat tagged with a nuclear localization signal (NLS) (**Fig 1a**).[11,27] This cell line will hereafter be referred to as RFP703/Dendra cells.

To maximize the efficiency of the VCS protocol, we imaged RFP703/Dendra cells on a widefield fluorescent microscope at 20x (**Fig 1b**) and developed an image analysis pipeline that privileges positive predictive value (PPV) over recall. In this pipeline, nuclei are segmented using the H2B signal with a pre-existing neural network trained to recognize nuclei (see Methods).[28] Image crops centered at each nucleus segment are generated and processed with Sobel edge detection, added as a second channel. Each 2-channel crop is used as input into a custom neural network that segments micronuclei within each crop. These micronucleus segments are reassembled onto the full, uncropped micrograph and assigned to “parent” nuclei by proximity. Manual analysis of micronucleus association with nuclei demonstrated that proximity is sufficient to correctly associate 97% of MN (**Fig S1c**). Nuclei with associated micronuclei are classed as micronucleus+ (MN+) and those without as micronucleus- (MN-). Comparison of micronucleus masks with manually identified MN demonstrated a recall value of approximately 70% and a PPV of 89% with a statistically significant reduction in recall among ruptured MN, determined manually by loss of NLS-3xDendra2 (**Fig 1c**). When the whole field is considered, including regions that are not captured by the crops or are excluded due to proximity to the field edge, MN recall drops to 50% (**Fig 1d**). Recall of MN- and MN+ cells is 86% and 65%, respectively, and PPV is 73% and 93%, respectively (**Fig 1e**). In summary, we can automatically identify the majority of

micronucleated and non-micronucleated cells in a low-resolution image with very high specificity.

Dendra2 intensity distinguishes intact micronuclei from ruptured micronuclei

Similar to previously used nuclear integrity markers, NLS-3xDendra2 is present in intact MN and absent in ruptured MN (**Fig 2a**).[11] To automatically identify cells with ruptured MN, the maximum Dendra2 intensity is measured in the nucleus and micronucleus segments and the ratio taken. Defining a threshold at 0.16 for MN rupture correctly classifies approximately 90% of micronuclei (**Fig 2b**). This information can be integrated into our pipeline to further classify MN+ cells into those with no ruptured micronuclei (rupture- cells) from those with 1 or more ruptured micronuclei (rupture+ cells) (**Fig S2a**), correctly identifying 60% of rupture- cells and 70% of rupture+ cells with a PPV near 75% in both cases (**Fig 2c**). 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 one ruptured micronucleus (**Figs S2b–c**).

Visual Cell Sorting specifically isolates multiple populations of photoconverted cells

VCS was previously shown to be able to specifically label and isolate multiple populations of cells based on differential UV pulse durations, which directly correlate to the proportion of photoconverted Dendra2.[25] To confirm that we could replicate previously demonstrated specificity with our cell line, we plated a 1:1 mixture of cells treated with CellTrace far red or with DMSO loading control (**Figs 3a–b**). We then segmented nuclei using the NLS-3xDendra2 channel and the same nuclear segmenter

used in our micronucleus classification pipeline. The masks generated were used to target a UV laser to individual nuclei. We thresholded on the nuclear far-red intensity, targeting CellTrace- cells with 200 ms UV pulses and CellTrace+ cells with 800 ms UV pulses. By thresholding on the Dendra2 red-emission/green-emission ratios during FACS, we could specifically separate a mixed population of cells either stained with CellTrace or left unstained (**Figs 3c–d**).

Cell yield from VCS recapitulates aneuploidy transcriptomic signatures

To confirm that our machine vision/VCS pipeline was suitable for generating input populations for RNAseq and to assess the effect of different UV pulses on cells treated with Mps1i, we randomly photoconverted asynchronous RFP703/Dendra cells treated with either Mps1i or DMSO loading control, with cells randomly assigned to either 800 ms or 200 ms of UV (**Fig 4a**). Given the duration of the complete experiment, we reimaged the first and last 5 positions at the end of an experiment to confirm that photoconverted cells retained differential levels of Dendra2 ratio throughout (**Fig S3a**). Further, to prevent cells from going through mitosis between the start and end of the experiment—and therefore diluting photoconverted Dendra2 signal or changing their MN status—we added the reversible Cdk1 inhibitor RO-3306 to inhibit mitosis.[29] This experimental design allowed us to look at differential gene expression between Mps1i and DMSO treatments and between UV pulse lengths. Conversion of 1,500–2,000 fields allowed us to capture 13k cells after FACS for each photoconverted population in each condition. Principle components analysis (PCA) revealed that, as expected, cells clustered first by treatment group and then by replicate (**Fig 4b**), confirming the effects

of Mps1 inhibition can be differentiated from loading control but also the need to control for batch effects. MA analysis identified 2,200 differentially expressed genes (DEGs), 63 of which had an absolute fold-changes > 1.5, suggesting a moderately strong response to Mps1 inhibition vs DMSO treatment (**Fig 4c**). Comparing DMSO-treated cells exposed to 800 ms vs 200 ms of UV radiation showed no difference in populations, confirming that these data were the result of Mps1i treatment and not the VCS pipeline itself (**Figs S4a–b**).To determine whether the VCS analysis preserved the previously reported aneuploidy response, we compared our data to RNAseq data generated by treating RPE-1 cells with a different Mps1 inhibitor or RPE-1 cells released from an 8-hour nocodazole block.[19,20,23] Both conditions identified a characteristic downregulation of many cell-cycle and metabolic pathways alongside upregulation of inflammatory pathways. To compare our results, we first reanalyzed these prior studies' raw counts using the same DESeq2 RNAseq pipeline we used for our data (see Methods) (**Figs S4c–f**).

Gene-set enrichment analysis (GSEA) using GO terms identified substantial overlap between all three data sets, though with many more terms identified in our data set than the other two (**Fig 4d**). GSEA identifies those genes that are most important for driving differential regulation in a pathway, termed "leading edge genes," and all three data sets also shared many leading edge genes. GSEA using Hallmark pathways identified a similar downregulation of cell-cycle pathways and upregulation of inflammatory pathways across the data sets (**Fig 4e**). Overall, we recapitulated a similar aneuploidy response as had been previously published. Our data set also resulted in many more

terms and pathways being identified, likely a result of greater paired-end coverage compared with the shallower, single-end reads used in the previously generated data sets (**Fig S4g**).

In summary, these data confirm that the VCS pipeline can reproduce data consistent with the previously published transcriptomic signature of aneuploid RPE-1 cells and validates this aneuploidy signature.

Micronucleus rupture, but not micronucleation, induces a limited general transcriptional response

To determine whether the formation of micronuclei induced a transcriptional response, we treated RFP703/Dendra cells with Mps1i and isolated MN+ and MN- cells using our machine-vision pipeline. Prior to processing samples for RNAseq, we performed several quality control checks: manually confirming in a subset of images that the classifier was performing as expected; replating an aliquot of sorted cells to check sorting accuracy between MN+ and MN- cells (**Fig S5a**); and checking the quality of extracted RNA by TapeStation analysis (**Fig S5b**). Running PCA on these data showed results clustered largely by replicate, rather than micronucleation status. This was confirmed by MA plot, which revealed nearly all DEGs to have absolute fold-changes less than 1.5. Only 2 genes were above this cutoff—the immediate-early gene EGR1, a transcription factor regulating cellular stress and growth responses, and A20, a ubiquitin-editing enzyme regulating inflammatory responses (**Fig 5b, c**).[30–33] These data suggest that

micronucleation alone produces a small initial transcriptional response in aneuploid cells.

Micronuclei continue to rupture throughout interphase. [8,11] Over the 4–5 hours necessary for photoconversion and isolation, we expected some intact micronuclei to rupture, decreasing the purity of the rupture- population (and conversely increasing the purity of the rupture+ population). To ensure minimal and predictable MN rupture during analysis and conversion of cells, we synchronized the cells in G1 using a Cdk4/6 inhibitor prior to release into Mps1i.[34] This causes a larger but more predictable change in rupture frequency compared with asynchronous cells (**Fig S5c**). In addition, a single technical replicate was performed per experiment to minimize the time between photoconversion and trypsinization (**Fig 5d**). Using this protocol, we determined the maximum change in rupture frequency from the start of an experiment vs the end, performed by re-imaging the first few fields of view at the end of the experiment. We found a significant increase in rupture frequency and used this change to estimate a reduction in rupture- PPV with a concomitant increase in rupture+ PPV (**Fig 5e**). These data suggest that RNAseq would still capture differential expression between these two populations, despite a loss in purity of the rupture- population.

To determine whether MN rupture induced a general transcriptional signature, we performed RNAseq analysis on isolated cells enriched for either intact-only MN (rupture-) or at least one ruptured MN (rupture+). PCA showed clustering based on condition, indicating a significant difference between the two cell populations. The MA

plot confirmed a transcriptomic change with 106 DEGs, 14 of which had absolute fold changes greater than 1.5 fold (**Figs 5f–g**). Both EGR1 and A20 are upregulated in cells with ruptured MN compared to cells with intact MN.

Aneuploidy is slightly enriched in cells with ruptured MN

To assess whether the transcriptional changes we detected in cells with ruptured MN were the result of micronucleation or aneuploidy, we analyzed single chromosome aneuploidies by FISH (**Fig 5h**). Asynchronous RPE-1 cells were treated with Mps1i for 24 h, fixed, and labeled with probes against chromosomes 1, 11, or 18. Quantification of the number of foci per cell demonstrated that aneuploidy was correlated with micronucleation for chromosomes 11 and 18 (**Fig 5i**). The contribution of micronucleation status to aneuploidy frequency was analyzed using a generalized estimating equation (GEE) with MN status, chromosome identity, and their interaction modeled as predictors against the frequency of single-chromosome aneuploidy. Micronucleation was found to be statistically significant, although with differing rates depending on the chromosome. While chromosome 1 was more likely to missegregate than the other two chromosomes (**Fig S5d**), when missegregated, it was less likely to correlate with the formation of MN or multiple micronucleation events (**Fig S5e**). The accompaniment of multiple micronuclei and micronucleus formation suggests missegregation of chromosomes 11 and 18 may be indicative of more severe mitotic defects, increasing the probability of aneuploidy and micronucleation. We performed the same aneuploidy analysis comparing cells with intact vs ruptured MN and found a strong correlation with chromosome identity, as before, but that rupture alone was not a

significant predictor of aneuploidy (**Figs 5j**). However, chromosome identity and rupture together were statistically significant, confirming an increase in the rates of aneuploidy among cells with missegregation in chromosomes 11 or 18. A correlation between MN rupture and aneuploidy may be explained by the increased frequency of multiple micronucleation events in cells with missegregated chromosomes 11 or 18 and an increase in the probability of at least one ruptured MN as the number of MN increase (**Figs S5e, S2b**). Chromosomes in a ruptured MN are no longer available to perform transcription or DNA replication. We reanalyzed the same FISH data, discounting any chromosomes within a ruptured MN to determine rates of "functional aneuploidy," finding nearly identical results (**Figs S5f–g**). Together, these data suggest that the transcriptional effects of aneuploidy are likely to be strong when comparing MN+ with MN- cells, but that the contribution of aneuploidy to the transcriptional response in rupture+ cells is less clear, given the weaker link between rupture and aneuploidy and the stronger transcriptional response produced.

ATF3 levels are increased among rupture+ cells

To validate our RNAseq results, we examined the expression of ATF3 in cells with and without ruptured MN by immunofluorescence. ATF3 is an immediate-early gene and transcription factor induced by a variety of stressors, and is transcriptionally upregulated in rupture+ cells (**Fig 5g**).[35,36]

As a positive control, we treated cells with doxorubicin, a known activator of ATF3 expression (**Fig 6a, S6a**).[37] Asynchronous RPE-1 cells expressing 2xRFP-NLS were

incubated with Mps1i or DMSO loading control for 24 hours, including 1 hour of treatment with doxorubicin in positive controls. Cells were then fixed and labeled with antibodies against γH2AX and ATF3, then manually identified as MN-, MN+ & rupture-, or MN+ & rupture+. [11,38–40] ATF3 and γH2AX intensity were quantified in automatically segmented nuclei. ATF3 intensity showed a strong increase in cells treated with doxorubicin and a correlation with γH2AX intensity (**Figs S6b–c**). Mps1i treatment induced an increase in ATF3 expression in a subset of cells (**Fig S6a**). Further classification of Mps1i cells into MN+/- and rupture+/- populations identified a small increase in cells with high ATF3 among MN- and rupture+ cells (**Fig 6b**). The increase in ATF3-high cells in MN- and rupture+ cells is consistent with our RNAseq results, which found ATF3 to be a DEG comparing Mps1i treatment against DMSO and comparing rupture+ with rupture- cells, but not between MN+ and MN- cells. (**Figs 4c, 5c, 5g**).

A potential confounding factor is that increased aneuploidy is associated with both an increase in micronucleation and DNA replication defects, DNA damage, and ATF3 activation. To account for this, we constructed a GEE to model the effects of rupture status and γH2AX intensity on ATF3 intensities among rupture+ and rupture- cells (**Fig 6c**). This model found that ATF3 intensity is predicted to increase as γH2AX intensity increases, but that this effect is slightly stronger in MN- cells, weak in rupture- cells, and pronounced among rupture+ cells. These data suggest that ATF3 expression is upregulated in response to DNA damage and this response is heightened in cells with ruptured MN, consistent with our RNAseq results.

Micronucleus rupture is a significant driver of transcriptomic changes in Mps1i-treated cells

To determine what pathways were affected by MN rupture, we compared changes to the Hallmark pathway ontology across our three analyses: Mps1i+/-, MN+/-, and rupture+/- conditions. There is broad overlap both in which pathways are differentially regulated and the direction and magnitude of those changes (**Fig 7a**). Many pathways share the same leading edge genes. Given the relatively small number of DEGs—of which some are regulatory transcription factors—we took the unique set of leading edge genes found in any differentially regulated pathway from each analysis, in order to determine if there was a core set of genes driving changes across many pathways. While there is some overlap in the leading edge genes from the Mps1i+/- experiment and the MN+/- experiment, the rupture+/- leading edge genes share significant overlap between both (**Fig 7b**). This suggests that rupture may be playing an important role in driving the changes in pathway regulation observed in both Mps1i-treated and micronucleated cell populations.

To determine the degree to which micronucleation and micronucleus rupture may be determining the GSEA results seen with Mps1i treatment, we plotted the cumulative proportion of differentially regulated pathways represented by a subset of leading edge genes (**Figs 7c–f**, in black). Along the x-axis, we ranked all leading edge genes by the number of pathways they belong to; on the y-axis are the cumulative proportion of differentially regulated pathways represented. All differentially regulated pathways in the

Mps1i+/- data set are fully represented by just the first 6% of leading edge genes (**Fig 7c**, dotted line). This top 6% of leading edge genes is also more likely to have a higher fold-change compared to the entire set (**Fig S7a**). Together, this suggests a core set of genes most responsible for differential pathway regulation.

To determine if these core genes were enriched in MN+/- leading edge genes, we plotted the contribution provided by leading edge genes found in both Mps1i+/- and MN+/- data sets (**Fig 7c**, in pink). These genes are 3-fold more likely to appear within the core set of genes (**Fig 7c**, odds ratio). The contribution is also somewhat greater than what would be expected by chance alone (**Fig 7c**, gray 66% and 95% confidence intervals). However, when we determine the contribution of leading edge genes found in both Mps1i+/- and MN+/- data sets, but not in the rupture+/- data set, the contribution of these genes to pathway enrichment in Mps1i+ cells drops to what would be expected by chance alone (**Fig 7d**). Leading edge genes found in both the Mps1i+/- and rupture+/- data sets are represented in 23 of the 24 differentially regulated pathways and are also significantly more likely to be present among the core Mps1i+ leading edge genes (**Fig 7e**). Conversely, Mps1i+/- leading edge genes not found in the rupture+/- data set are significantly more likely to be found outside of the core set, individually contribute to fewer pathways, and have a lower RNAseq fold change (**Figs 7f, S7a**). These data suggest that the genes most important for driving the changes identified by GSEA in Mps1i-treated cells are also leading edge genes seen in rupture+ cells when compared to rupture- cells, despite both populations having been treated with Mps1i. These trends also hold when considering the much larger set of GO terms (**Figs S7b–e**) and when

comparing the contribution of rupture to the transcriptional response to micronucleation (**Figs S8a–c**). Overrepresentation analysis of the Mps1i+/- Hallmark pathways identifies the specific pathways whose genes include a greater proportion of rupture+/- leading edge genes than would be expected by chance alone (**Fig 7a**, in bold).[41,42] These rupture-overrepresented pathways include inflammatory, stress and damage response, and cell proliferation pathways. In summary, these data suggest that cells with ruptured micronuclei, despite only representing ~25% of the total population of Mps1i-treated cells (**Fig S8d**), are disproportionately driving the transcriptomic response to Mps1i treatment, specifically those pathways related to stress, inflammation, and proliferation, suggesting that at least a portion of the aneuploidy response previously characterized may be a response to events correlated with MN rupture specifically.

A redesigned neural net improves overall performance in a wider range of conditions

The neural net used in the above experiments prioritized speed and PPV in specific imaging conditions. However, other experiments would benefit from an increase in recall or good performance in other imaging conditions. To generalize neural net performance to other imaging conditions, we redesigned the neural net architecture to use a simplified UNet model with no transfer learning employed. One difficulty in segmenting micronuclei is the imbalanced nature of the data set—there are many more pixels that are not micronuclei in an image than are. To improve training, we switched from a cross-entropy loss function to a focal loss function, which is designed to deal with highly imbalanced data sets.[43] A second challenge with micronucleus segmentation is the similarity between nucleus-associated and micronucleus-associated pixels. We modified

the training data and the neural net to segment both nuclei and micronuclei. While nucleus segmentation underperforms existing nucleus segmenters, providing this information to the micronucleus neural net improved its ability to distinguish the two. We also more than doubled the size of the cropped region to give the neural net greater contextual information, and then trained it on both RPE-1 and U2OS cells expressing H2B-emiRFP703, to broaden the neural net's ability to discern micronuclei in different cell types with different nuclear shapes. To improve processing speed, the original neural net relied on crops centered on already segmented nuclei. To improve coverage and predictions, at the cost of processing time, we oversample the entire field and predictions are averaged. Final micronucleus segments are compared with nucleus segments generated by an existing, purpose-built nucleus-segmenting neural net to trim down any micronucleus segments that overlap with nucleus predictions (**Fig 8a**).[44]

When tested against RFP703/Dendra cells, this new neural net has similar recall and PPV as the neural net used for VCS, though for the whole field rather than just cropped regions. However, it maintains a similar recall and PPV under different imaging conditions and with different cell lines, with a generally improved mIOU (**Figs 8b**). This neural net is being offered as a Python package installable with the Python package manager `pip`, described on the Hatch Lab GitHub repository.[45]

Discussion

In this study, we present a method for isolating subpopulations of micronucleated cells from a heterogeneous bulk population using a newly developed neural net combined

with Visual Cell Sorting, here used to separate MN+ and MN- cells or micronucleated cells that are rupture+ or rupture-. Further, we demonstrated that this technique can be used for hypothesis discovery, describing the early transcriptomic response to micronucleation and micronucleus rupture. Comparison of MN+/- and rupture+/- transcriptomes with the transcriptomic response to aneuploidy reveals micronucleus rupture to be strongly correlated with the previously described “aneuploidy response” in RPE-1 cells, suggesting that this aneuploidy response may largely be a response to micronucleus rupture rather than aneuploidy *per se*. Though, with some correlation between micronucleation, micronucleus rupture, and chromosome missegregation, further research is necessary to disentangle these related processes. Finally, we developed a modified neural net able to segment micronuclei in a wider variety of imaging contexts and with other cell lines, which we are making available as a Python package.

Materials and methods

Plasmid construction

pLVX-EF1a-NLS-3xDendra2-blast was created by PCR of NLS-3xDendra2 from pLenti-CMV-Dendra2x3-P2A-H2B-miRFP (a gift from Doug Fowler) with forward primer (with prepended NotI digest site) 5'-***ggaagg*gcggccgc**caagtttgtacaaaaaagttggcaccATGG-3' and reverse primer (with prepended EcoRI digest site and stop codon) 5'-***ggaagg*gaattcTTA**GGAAAAATTCGTTGCGCCGCTCCC-3', followed by insertion into pLVX-EF1a-blast by restriction digest cloning (NEB NotI-HF, #R3189; EcoRI-HF, #R3101). Bold signifies regions that were prepended with primer extension; bold-italic mark sites

added to facilitate restriction enzyme activity and are cleaved before insertion into the final plasmid.

pLVX-EF1a-H2B-emiRFP703-neo was created by PCR of H2B-emiRFP703 from pH2B-emiRP703 (a gift from Vladislav Verkhusha, AddGene #136567) with forward primer (with prepended NotI digest site) 5'-***gaaagg*****gcggccgc**cATGCCAGAGCCAGCGAAG-3' and reverse primer (with prepended MluI site) 5'-***gaaagg*****gaattc**cTTAGCTCTCAAGCGCGGTGATC-3', followed by insertion into pLVX-EF1a-neo by restriction digest cloning (NEB MluI-HF, #R3198).

All primers manufactured by IDT and resuspended to 100 µM stock solution in 1xTE (0.5 M EDTA [Sigma, #E5134] and 1 M Tris Base [Fisher, #BP152] in H_2O, adjusted to pH 8.0) and diluted to a 10 µM working solution in ultrapure H_2O (Genesee Scientific, #18-195).

Lentivirus production

For pLVXE-based plasmids, HEK293T cells were plated in 10 cm tissue-culture–treated dishes (Fisher Scientific, #664160) at 70% confluency in DMEM (Gibco, #11965092) supplemented with 10% FBS (Sigma, #F0926, lot: 18C539) and 1% pen/strep (Gibco, #15140122). The next day, we replaced with fresh media and transfected cells with a solution of 3 µg psPAX2, 1.5 µg pVSV-G (gifts from Didier Trono, AddGene #12260 and #12259, respectively), and 2 µg of pLVX-EF1a-NLS-3xDendra2-blast, pLVX-EF1a-H2B-emiRFP703-neo, or pLenti-CMV-Dendra2x3-P2A-H2B-miRFP after a 5 minute

incubation in OptiMEM (Gibco, #31985070) with 40 µg linear PEI (polsciences, #23966-2).

The following day, we refreshed media with media supplemented with a sterile 25 mM HEPES (Sigma, #H3375) solution in diH_2O.

The following day, we removed 5 mL of supernatant and filtered it through 0.45 µm PES syringe filters (Genesee Scientific, #25-246) and stored the filtrate at 4 ºC. The remaining media was removed and replaced with fresh, HEPES-supplemented media.

The next day, we repeated this procedure, adding the filtrate to the previous day's. We then added 2 U/mL DNaseI (Fisher, #FEREN0523) and $MgCl_2$ to a 2 mM final concentration. After incubating at room temperature for 30 min, we moved the filtrate to 4 ºC for another 4 hours, then aliquoted virus and snap froze in liquid N_2. Virus aliquots were thereafter stored at -80 ºC.

Cell lines

The hTERT RPE-1 NLS-3xDendra2/H2B-emiRFP703 and U2OS NLS-3xDendra2/H2B-emiRFP703 cell lines were produced through serial transduction of both plasmids. Immediately after splitting hTERT RPE-1 or U2OS cells to 40% confluency, a polybrene (Sigma, #H9268) solution in PBS (Gibco, #14190) was added to a final concentration of 6 µg/mL, followed by 1 mL of virus filtrate added dropwise. Cells were then split twice before selection was added: blasticidin (Invivogen, #ant-bl-1) to a final concentration of

10 µg/mL and Geneticin/G418 (Gibco, #11811031) to a final concentration of 500 µg/mL active G418. Cells were then sorted on by FACS on a BD Aria II for the top 20% brightest Dendra2 and emiRFP703 expressing cells.

hTERT RPE-1 NLS-3xDendra2-P2A-H2B-miRFP703 cells were created through a single transduction as described above, without selection applied, and sorted by Dendra2 and miRFP703 intensity.

Tissue culture

hTERT RPE-1 cell lines were kept in DMEM/F12 (Gibco, #11330057) supplemented with 10% FBS, 1% pen/strep, and 0.01 mg/mL hygromycin B (Sigma, #H3274) and stored in a cell incubator (HeraCell 150i) with a 5% CO_2 atmosphere at 37 ºC.

RPE-1 NLS-3xDendra2/H2B-emiRFP703 were supplemented with 10 µg/mL blasticidin and 500 µg/mL G418.

U2OS NLS-3xDendra2/H2B-emiRFP703 cells were kept in DMEM supplemented with 10% FBS and 1% pen/strep, and stored in a cell incubator with a 10% CO2 atmosphere at 37 ºC and likewise supplemented with blasticidin and G418.

Micronucleus segmentation and cell classifier

VCS neural net

The neural net used for the experiments in this paper was created using the FastAI 1.0 library in Python. It is a U-Net with Torchvision's ResNet18 pre-trained model as its base architecture.[46] Additional training was performed specifically for micronucleus recognition using ~2,000 48x48 crops of individual cells as training data, a further 164 for validation, and 177 for testing.

Training data was generated using the hTERT RPE-1 NLS-3xDendra2-P2A-H2B-miRFP703 cell line. Training data was generated by treating cells with either 0.5 μM reversine or DMSO loading control, then imaging the next day on a Leica DMi8 at the Fowler Lab (Fowler Lab microscope) using a 20x widefield objective in the NIR and GFP channels. Following nucleus segmentation using the Deep Retina neural net,[28] 48x48 px crops of the H2B channel were generated centered at each nucleus. We then annotated each crop for the presence of micronuclei using a custom program.

The classifier takes as input a 20x image with the H2B-emiRFP703 and NLS-3xDendra2 channels. It uses the Deep Retina neural net to segment nuclei, discarding any touching the edge of the image. It then generates 48x48 px crops centered at each nucleus. Each crop is processed with Sobel edge detection. The VCS neural net architecture is based on Torchvision's ResNet18, which expects a 3-channel image. To accommodate this architecture, we transform each crop into a 3-channel image: the first

2 channels are duplicates of the H2B channel and the final channel the results of Sobel edge detection.

Crops are processed by the VCS neural net. Segmented micronuclei are localized on the original, full 20x image, then assigned to the closest nucleus segment. Any micronucleus segments whose nearest segmented nucleus is more than 40 px away are discarded.

Once micronuclei are assigned, the classifier takes the maximum Dendra2 intensity from each micronucleus segment and its associated nucleus segment to generate micronucleus/nucleus Dendra2 ratios. Micronuclei with a ratio below 0.16 are classified as being ruptured micronuclei. This threshold was chosen by manually identifying micronuclei as ruptured or intact from experiments isolating MN+ and MN- cells, then using the `JRip` classifier in Weka 3.8.6 to identify an optimal threshold to separate intact and ruptured micronuclei by this Dendra2 ratio.[47,48]

Nuclear segments are classified as MN+ or MN- based on the presence or absence of an associated micronucleus segment. MN+ segments are then further classified into those with only intact micronuclei (rupture-) or those with at least one ruptured micronucleus (rupture+).

Images taken and the classification results from each experiment were saved. After an experiment finished, we compared classifier predictions against ground truth by manual

inspection. Micronuclei with attached chromatin bridges were ignored to ensure that micronuclei were discrete nuclear compartments.

Individual micronuclei were segmented by hand using PixelStudio on an Apple iPad. Recall was calculated as the proportion of all micronuclei that overlapped with a predicted segment. Positive predictive value was calculated as the proportion of all predicted segments that overlapped with a micronucleus.

Improved neural net

The improved neural net was created using TensorFlow 2.0 also using a U-Net architecture, but without transfer learning. Training was performed on 128x128 px crops generated at time of training from images of hTERT RPE-1 NLS-3xDendra2/H2B-emiRFP703 and U2OS NLS-3xDendra2/H2B-emiRFP703 cells taken using a 20x widefield objective on the Fowler Lab microscope in the NIR channel. Annotations of nuclei and micronuclei were performed by hand using PixelStudio version 4.5 on an Apple iPad.

Input is expected to be an image of nuclei taken at 20x; images taken at 40x were first scaled down by a factor of 2. A 128x128 px sliding window is used to generate crops, advancing the window by 64 px horizontally and vertically to oversample the image. Crops are expanded into 2-channel images, with the second channel the result of Sobel edge detection. Crops are processed by the improved neural net to segment micronuclei and the average value taken from oversampled regions.

The same image is used as input for the Stardist nucleus segmenter to identify nuclei. Any regions where micronucleus segments overlap nucleus segments are removed to produce a final set of micronucleus labels.[44]

The improved neural net was validated against images of hTERT RPE-1, U2OS, hTERT HFF, and HeLa cells. Live RPE-1 and U2OS cells expressing NLS-3xDendra2/H2B-emiRFP703 were taken on the Fowler Lab microscope with a 20x widefield objective in phenol-free media. Fixed cells were taken on either a Leica DMi8 (Hatch Lab microscope) or Leica spinning disk microscope with a 40x confocal objective after being fixed in 4% PFA (Electron Microscopy Sciences, #15710) in PBS for 5 min at room temperature with coverslips mounted on glass slides with VectaShield anti-fade medium (Vector Laboratories, #H-1000) and sealed with nail polish (Wet *n* Wild, wildshine base coat).

HeLa cell images were taken by Molly Zych and HFF cell images were taken Emily Hatch, both at the Hatch Lab.

Recall and positive predictive values were calculated using the same protocol used for VCS classifier validation.

VCS CellTrace sorting accuracy validation

We trypsinized hTERT RPE-1 NLS-3xDendra2/H2B-emiRFP703 and hTERT RPE-1 cells with 0.25% trypsin in EDTA (Gibco, #25200-056) after rinsing cells in PBS. After pelleting cells at 300 rcf for 5 min, we resuspended each population of cells in 1 mL PBS. We then made 2 aliquots of 1 million cells for each population. Cells were then re-pelleted and resuspended in 1 mL PBS. To one aliquot from each population, we added 1 µL CellTrace far red (ThermoFisher, #C34572). To the other aliquots, we added 1 µL DMSO (US Biological, #D8050-01).

Aliquots were incubated at 37 ºC for 10 min in the dark, gently mixed by pipette, then incubated for a further 10 min. We then added 200 µL PBS to each aliquot, mixing gently by pipette and incubated for another 5 min.

Aliquots were pelleted as before, then resuspended in media. Cells were then plated on a 6-well glass-bottom dish (Cellvis, #P06-1.5H-N), 175k cells/well, as follows: A1: RPE-1, -CellTrace; A2: RPE-1, +CellTrace; B1: Dendra2-expressing, -CellTrace; B2: Dendra2-expressing, +CellTrace; A3: 1:1 mix of Dendra2-expressing +/- CellTrace.

Cells were then transported from the Hatch Lab at the Fred Hutchinson Cancer Center to the Fowler Lab at the University of Washington (~20 min transport) in the dark and placed in a cell incubator (HeraCell Vios 160i) with a 5% CO_2 atmosphere at 37 ºC.

The following day, we rinsed cells 3x with PBS then replaced media with phenol-free DMEM/F12 (Gibco, #21041025) supplemented with 10% FBS, 1% pen/strep, 0.01 mg/mL hygromycin B, and 10 μM Cdk1i (RO-3306, Sigma, #SML0569) that had been filtered through a 0.22 μm PES syringe filter (Genesee Scientific, #25-244) to remove residual RO-3306 precipitate. Cells were then placed in a temperature-controlled, CO_2-supplemented microscope incubation chamber.

1 hour following treatment with Cdk1i, the mixture well was imaged by scanning across the well using a 20x widefield objective on the Fowler Lab microscope, imaging in NIR and GFP wavelengths. The NIR channel captures both H2B-emiRFP703 and CellTrace far red, though the latter is significantly brighter. Custom journals transmitted each image to another computer equipped with a CellTrace classifier.

This classifier segmented nuclei with the Deep Retina neural net using the GFP channel as input, then measured the mean NIR intensity in the nucleus. Cells above a given threshold, determined on the day by capturing a test image and measuring the CellTrace intensity of CellTrace+ and CellTrace- cells by hand, were classified as CellTrace+ and those below this threshold were classified as CellTrace-. 1-bit masks of each population were transmitted back to MetaMorph, which used a digital micromirror device to target CellTrace+ cells with 800 ms UV and CellTrace- cells with 200 ms UV.

Following photoconversion of the mixture well, cells were trypsinized and resuspended in media. We then pelleted cells at 300 rcf for 5 min, then resuspended cells in FACS

buffer (1 % BSA [Sigma, #A4503, lot: SLCG9729], 0.2 mM EDTA, 25 mM HEPES in PBS, sterile filtered with 0.22 µm PES syringe filter) and placed on ice.

We then sorted cells with a BD Aria III, using Alexa 700 to gate on emiRFP703+ cells and FITC to gate on Dendra2-hi cells. Final sorting was performed by gating on a ratio of PE-blue/FITC. Compensation for PE-blue exciting unphotoconverted Dendra2 was performed using Dendra2-expressing cells without CellTrace, as was compensation for CellTrace far red and emiRFP703 spectral overlap. The sorted populations were then re-analyzed on the same sorter to validate the purity of each sorted population.

Mps1i+/- experiment and VCS validation

We plated 175k hTERT RPE-1 NLS-3xDendra2/H2B-emiRFP703 cells in 3 wells of a glass-bottom 6-well dish. The following day, media was replaced and supplemented with 100 nM Mps1i (BAY-1217389, Fisher, #501872752) or DMSO loading control. Cells were then transported in the dark from labs at the Fred Hutchinson Cancer Center to the Fowler Lab at the University of Washington, as above. 19 hours later, we rinsed cells 3x with PBS and then replaced media with filtered phenol-free media supplemented with 10 µM Cdk1i. Cells were then placed in a temperature-controlled, CO_2-supplemented microscope incubation chamber.

1 hour following treatment with Cdk1i, two wells were imaged sequentially with a 20x widefield objective in the NIR and GFP channels. As previously described, MetaMorph scanned over each well, and custom journals transmitted each image to another

computer equipped with a random classifier. 1-bit masks of nuclei randomly assigned to receive 800 ms or 200 ms UV pulses were transmitted back to MetaMorph, which directed UV pulses at the segmented nuclei.

At the end of each well, the first and last few positions were re-imaged in NIR, GFP, and RFP channels to ensure cells were being properly photoconverted. At the end of the experiment, several positions in the third, un-photoconverted well was also imaged.

Following Dendra2 photoconversion of two wells, cells were trypsinized, resuspended in media, and pelleted. Cells were then resuspended in FACS buffer and placed on ice.

We sorted cells with a BD Aria III, using Alexa 700 to gate on emiRFP703+ cells and FITC to gate on Dendra2-hi cells. Final sorting was performed by gating on a ratio of PE-blue/FITC. Compensation for PE-blue exciting unphotoconverted Dendra2 was performed using the final well not treated with UV.

After collecting ~13k cells in a given condition, sorting was paused and collection tubes swapped out in order to collect further samples of ~13k cells. Samples were then pelleted, media aspirated, and pellets flash frozen in dry ice before being placed at -80 ºC for RNA extraction.

To validate the maintenance of a Dendra ratio over the course of the experiment, we thresholded images taken at the end of the experiment using the GFP channel to

segment nuclei. We used these segments to measure the mean GFP and RFP channel intensities. Images taken of the last few positions represented cells that had just been photoconverted, while images taken of the first few positions represented cells that had been photoconverted 4 or 8 hours prior. Plotting the ratio of RFP/GFP intensities revealed three distinct populations in all three image sets, representing unphotoconverted cells, cells that received 200 ms UV, and cells that received 800 ms UV.

Micronucleus+/- isolation

This followed the same protocol used for the Mps1i+/- experiment, but with all cells receiving 100 nM Mps1i. As above, two wells were imaged sequentially with a 20x widefield objective in the NIR and GFP channels. 1-bit masks of MN+ nuclei and MN- nuclei were transmitted back to MetaMorph, which directed UV pulses at the segmented nuclei. In one well, MN+ nuclei received 800 ms UV and MN- nuclei 200 mn UV; in the other well, this was reversed. Whether the MN+:800 ms/MN-:200 ms well or the MN+:200 ms/MN-: 800 ms was imaged first was flipped from one experiment to the next to ensure UV exposure could be controlled for during downstream analyses.

Sorting was performed as described above. During one experiment, one aliquot was retained for replating instead of being flash frozen, as described below.

MN+/- sorting accuracy determination

To assess sorting accuracy, one aliquot was transported back to the Hatch Lab on ice in the dark (travel time ~20 min) where cells were pelleted and resuspended in media supplemented with 10 μM Cdk1i to prevent mitosis, then replated into 24-well plates (ThermoFisher, #930186) on lysine-coated coverslips. The following day, we rinsed cells in PBS then fixed in 4% PFA in PBS at room temperature for 5 min. Cells were permeabilized in PBS-BT (PBS supplemented with 0.4% TritonX-100 [Sigma, #T9284], 3% BSA, and 0.02% sodium azide [Sigma, #S2002]) for 15 min, then incubated in 1 μg/mL DAPI [Invitrogen, #D1306] in PBS for 5 min. After a final rinse in PBS, we mounted coverslips on glass slides with VectaShield and sealed with nail polish.

We then imaged coverslips on the Hatch Lab microscope with a 40x confocal objective, taking z-stacks of multiple positions until ~100 cells of each population were imaged. We then randomized the names of images, to obscure which image came from which population, then counted the number of MN+ cells in each image. Once counting was finished, we re-identified which images belonged to which population.

Rupture frequency pulse-chase

RPE-1 NLS-3xDendra2/H2B-emiRFP703 cells were plated into 2 wells of a 6-well glass-bottom dish. One well was plated at 225k cells/well supplemented with 1 μM Cdk4/6i (PD-0332991, Sigma, #PZ0199) to synchronize cells at the G1/S transition. The other well was plated at 175k cells with DMSO loading control.

24 hours later, we rinsed wells 3x with PBS before replacing media supplemented with 100 nM Mps1i. Cells were transported to the Fowler Lab as described above. 19 hours later, we rinsed cells 3x with PBS and replaced media with phenol-free media supplemented with 10 μM Cdk1i that had been filtered to remove precipitate.

After incubating cells in the microscope incubation chamber for 1 hour, we imaged 5 positions of each well, running the micronucleus classifier as described above. However, we restricted photoconversion to cells that were classified as rupture-, exposing them to 800 ms UV pulses. No cells were treated with 200 ms UV.

After photoconversion, every hour for 11 hours, we imaged a 25-position window centered at the original 5 photoconverted positions in the NIR, GFP, and RFP channels to track those cells that had been photoconverted. Images were then analyzed to validate by eye the proportion of cells that were rupture-, rupture+, or MN-.

Rupture+/- isolation

Cells were plated as in the micronucleus+/- section, but only 2 wells were plated at 225k cells/well. In addition, cells were originally plated in media supplemented with 1 μM Cdk4/6i (PD-0332991, Sigma, #PZ0199) to synchronize cells.

24 hours later, we rinsed cells 3x with PBS and replaced with media supplemented with 100 nM Mps1i and then transported to the Fowler Lab.

The following day, cells were prepared as described above with Cdk1i, phenol-free media and one well was photoconverted, with 1-bit masks for rupture- and rupture+ cells being returned by the classifier. Rupture- cells received 800 ms UV and rupture+ cells 200 ms. Cells were then sorted as described in the micronucleus+/- isolation section.

RNA isolation and sequencing

We extracted RNA from frozen pellets using ThermoFishers's RNAqueous micro kit (ThermoFisher, #AM1931), starting with immediately resuspending frozen pellets in lysis solution and vortexing vigorously. All other steps were according to the manufacturer's protocol. Residual DNA was removed with the optional DNase I treatment that is included with the kit, resulting in ~20 µL of RNA solution.

To remove residual guanidium salts, RNA was further purified by glycogen precipitation. 20 µL pH 7 TE was added, then 4 µL of 3 M sodium acetate (Fisher, #S210) and 0.4 µL 15 mg/mL GlycoBlue (ThermoFisher, #AM9515) or 0.3 µL 20 mg/mL RNA-grade glycogen (ThermoFisher, #R0551). 133 µL of -20 ºC 100% ethanol was then added and the mixture placed at -20 ºC overnight. The next day, the sample was centrifuged at 14,000 rcf for 10 min at 4 ºC in a microcentrifuge and then supernatant removed by pipette. The glycogen/RNA pellet was then washed with 500 µL of -20 ºC 70% ethanol in ultra-pure H_2O and pelleted again at 21,000 rcf for 10 min at 4 ºC. The supernatant was removed by pipette and the tube allowed to dry with the cap open at room temperature. Finally, the RNA was resuspended 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.

cDNA synthesis and library preparations were performed by the Genomics Core at the Fred Hutchinson Cancer Center using the SMARTv4 for ultra-low RNA input and Nextera XT kits, respectively (Takara). Sequencing was also performed by the Genomics Core using an Illumina NextSeq 2000 sequencing system with paired-end, 50 bp reads.

RNAseq analysis

We quantified transcripts with Salmon to map reads against the UCSC hg38 assembly provided by http://refgenomes.databio.org (digest: `2230c535660fb4774114bfa966a62f823fdb6d21acf138d4`), using bootstrapped abundance estimates and corrections for GC bias.[49]

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.[19,23] No GC-bias correction was applied, as this is still experimental at time of writing for single-end reads.

Resulting transcript abundances were then processed to find differentially expressed genes (DE genes) with the R package `DESeq2` version 3.16 in R 4.2.1, RStudio 2022.07.2 build 576, and Sublime Text build 4143. Briefly, files were imported into

`DESeq2` with the R package `tximeta`.[50,51] 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.[52]

Gene-set enrichment analysis

GSEA was performed using the R package `fgsea` version 1.25.1, comparing log-fold changes of genes against either the full *Homo sapiens* Gene Ontology database (GO) version 2022.1 or the Hallmark Gene Sets version 2022.1, both part of the MSigDB resource published by UC San Diego and the Broad Institute.[53–60] Santaguida, *et al.*, provides data from cells that underwent cell-cycle arrest and cells that continued to cycle; our GSEA found no significantly enriched terms from still-cycling cells.

We generated a unique list of leading edge genes for each data set and the number of pathways or terms that leading edge gene was a member of. These genes were then divided into core genes and non-core genes, with the former being the unique list of genes that were members of at least 99% of pathways (for the Hallmark ontology) or 98% of terms (for the GO ontology).

To compare the contributions of different subsets of leading edge genes, we generated an odds ratio by comparing the ratio of core/non-core genes in the subset vs the total leading edge set. To test if the subset's contribution was more or less than what was expected by chance, we generated a confidence interval from a null distribution created by 200 bootstrapped subsets whose overlap with the total leading edge gene list was the same proportion as the subset of interest.

To identify Mps1i+/- pathways whose genes include more rupture+/- leading edge genes than would be expected, we used the R package `clusterProfiler` version 4.6.2

XCE FISH and analysis

We plated 175k hTERT RPE-1 cells/well in a 6-well dish (ThermoFisher, #130184) onto lysine-coated coverslips. The following day, we replaced media with fresh media supplemented with 100 nM Mps1i. 20 hours following, we rinsed cells in PBS then fixed in -20 ºC 100% methanol for 10 min.

To determine the rupture status of micronuclei, we performed immunofluorescence for acetylated lysine 27 on histone 3 (H3K27-acetyl).[8,11,13] Briefly, we rehydrated cells with a 2x wash in PBS for 5 min each at room temperature, then blocked in PBS-BT for 30 min at room temperature. We then incubated coverslips in polyclonal rb α H3K27-acetyl antibody (Abcam, #ab4729) at 2 µg/mL in PBS for 30 min at room temperature followed by a 3x PBS wash for 5 min each at room temperature. We then incubated coverslips in

a secondary gt α rb antibody conjugated to AlexaFluor-488 (Invitrogen, #A11034) at 1 µg/mL in PBS-BT for 30 min at room temperature. After a final 3x PBS wash for 5 min each at room temperature, we re-fixed cells in 4% PFA in PBS for 5 min at room temperature.

We then rinsed coverslips in 2xSSC (20x SSC, Sigma, #S6639 diluted in ultrapure H_2O), then permeabilized cells in a solution of 0.2 M HCl (Sigma, #258148) and 0.7% TritonX-100 in ultrapure H_2O for 15 min at room temperature. After 2, 5-min washes in 2xSSC at room temperature, we incubated coverslips in a 1:1 solution of formamide (Millipore, #4650) and 2xSSC for 1 hour at room temperature. After another 2, 5-min washes in 2xSSC at room temperature, we applied 3 µL of chr 1, 11, or 18 XCE orange probe (MetaSystems, #D-0801-050-OR, D-0818-050-OR, and D-0811-050-OR) directly to the coverslip and mounted it on a glass slide, sealed with rubber cement. We immediately placed the slide on a 74 ºC hot plate for 3 min, then moved the slide to a pre-warmed, humidified incubation chamber at 37 ºC. We incubated each slide at 37 ºC for either 4 hours (chr 18) or 14.5 hours (chrs 1 and 11).

Following incubation, we removed rubber cement and placed the coverslip into a 0.4xSSC solution heated to 74 ºC, and incubated each coverslip at 74 ºC for 5 min. We then performed 2, 5-min washes in 2xSSC supplemented with 0.1% Tween20 (Fisher, #BP337). Finally, we incubated the coverslip in 1 µg DAPI in PBS supplemented with 0.4% TritonX-100 for 5 min at room temperature, rinsed once in PBS, then mounted

each coverslip onto a glass slide using VectaShield anti-fade medium, sealed with nail polish.

Images were taken on the Hatch Lab microscope using a 40x confocal objective. For each cell, the number of centromeres in the cell, the number of centromeres in micronuclei, and the number of centromeres in ruptured micronuclei were counted. For determining missegregation frequencies, cells without 2 centromeres for a given chromosome were classified as missegregated. For determining functional aneuploidy, the cells without 2 centromeres in any intact nuclear compartment (ie, centromeres in ruptured micronuclei did not count toward the total) were classified as aneuploid.

Micronucleation status was performed by hand. Rupture status of any given micronucleus was determined by hand through examining the presence or absence of H3K27-acetyl.

In total, 3 replicates of chromosomes 11 and 18 and 2 replicates of chromosome 1 were performed.

Immunofluorescence

For ATF3 validation, we plated 175k RPE-1 2xRFP-NLS cells in 4 wells of a 6-well dish along with lysine-coated coverslips. The following day, we changed the media in all wells, replacing one well's media with that supplemented with 100 nM Mps1i and the others with DMSO loading control. 23 hours later, we replaced media in all wells

following a 3x rinse with PBS at room temperature with fresh media supplemented with 2 µg/mL doxorubicin hydrochloride (Fisher, #BP25165) or DMSO loading control. The resulting treatment conditions were: DMSO ⇒ DMSO, Mps1i ⇒ DMSO, or DMSO ⇒ doxorubicin. Following a 1 hour incubation, we rinsed cells 1x in PBS and fixed them in a 4% PFA solution in PBS at room temperature in the dark for 5 min.

Immunofluorescence proceeded in the dark at room temperature, starting with a 30 min block in PBS-BT. Coverslips were then incubated in mouse-anti-γH2A.X (BioLegend, #613401) at a dilution of 1:499 in PBS for 30 min, followed by 3x 5-min washes in PBS. Coverslips were incubated in a secondary antibody, goat-anti-mouse conjugated to Alexa647 (Life Technologies, #A21236) diluted 1:999 in PBS-BT for 30 min. After 3x 5-min PBS washes, we incubated coverslips in rabbit-anti-ATF3 (Cell Signaling Technology, #18665S) at a dilution of 1:399 in PBS for 30 min, followed by 3x 5-min washes in PBS and incubation in goat-anti-rabbit conjugated to Alexa488 (Life Technologies, #A11034) diluted 1:1999 in PBS-BT for 30 min. After a final set of 3x 5-min PBS washes, coverslips were briefly incubated in 1 µg/mL DAPI in PBS followed by a final rinse in PBS and mounting to slides using VectaShield and nail polish.

ATF3 quantification

ATF3-stained coverslips were imaged as z-stacks on the Hatch Lab microscope using a 40x confocal objective. Following flat-field correction, images were subject to rolling-ball background subtraction in FIJI 2.9.0 with a pixel radius of 60 px. Using FIJI, in the middle of each z-stack, we thresholded on the RFP-NLS signal to create ROIs. For

each ROI, we identified the middle of the nucleus using the RFP-NLS channel, then measured the mean intensities in both the ATF3 and γH2AX channels using the ROI as a mask. In addition, to determine the number of γH2A.X foci, we ran the `Find maxima…` command in FIJI, with a threshold of 20. Measuring the raw integrated density and dividing by 255 rendered an approximate number of foci.

To aid in the visual comparison of ATF3 intensities across replicates, the values displayed in **Fig 6b** were normalized by first scaling all ATF3 mean intensities to the minimum and maximum values within that replicate, then translating the values such that the 90th percentile in the DMSO-treated condition was at 1. These normalized values were only to aid in visual inspection and were not used in any statistical analysis.

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

The correlation between MN rupture status and segmenter recall in **Figs 1c–d** was determined using a generalized estimating equation (GEE) using a binomial distribution and a logit link function:[61]

(# recalled, # missed) ~ MN status

where *MN status* is whether an MN was ruptured or not. The change in rupture frequency in **Fig 5e** was likewise analyzed with a GEE using a binomial distribution and a logit link function.

Chromosome missegregation and functional aneuploidy results were tested with a GEE using a binomial distribution and a logit link function:

(# 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 was calculated using the `drop1` function in R.

The predicted change to classifier PPV in **Fig 5e** was determined by taking the difference in mean rupture frequencies between the beginning and end of the experiments and adjusting the PPV of rupture- and rupture+ populations by reducing the true positive rate in the rupture- population by this difference and increasing the true positive rate in the rupture+ population by the same:

To determine the effect of rupture on the measured intensity of ATF3 in **Fig 6c**, we constructed a GEE using a Gamma distribution and log link function:

ATF3 intensity ~ Population x γH2AX

p-values were calculated as above.

Odds ratios in **Figs 7 & S7** were tested by Barnard's exact test.

Acknowledgements

This work supported by a National Institutes of Health grant (R35GM124766-02, awarded to E.M. Hatch), a training grant (T32CA009657, awarded to L. DiPeso), and the Rita Allen Foundation Scholars program.

1. Guo, X. *et al.* The molecular origins and pathophysiological consequences of micronuclei: New insights into an age-old problem. *Mutat. Res. - Rev. Mutat. Res.* **779**, 1–35 (2019).
2. Howell, W. The life history of the formed elements of the blood, especially the red blood corpuscles. *J. Morphol.* **4**, 57–116 (1890).
3. Kirsch-Volders, M., Bolognesi, C., Ceppi, M., Bruzzone, M. & Fenech, M. Micronuclei, inflammation and auto-immune disease. *Mutat. Res. Rev. Mutat. Res.* **786**, 108335 (2020).
4. MacDonald, K. M., Benguerfi, S. & Harding, S. M. Alerting the immune system to DNA damage: micronuclei as mediators. *Essays Biochem.* **64**, 753–764 (2020).
5. Mackenzie, K. J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461–465 (2017).
6. Vázquez-Diez, C., Yamagata, K., Trivedi, S., Haverfield, J. & FitzHarris, G. Micronucleus formation causes perpetual unilateral chromosome inheritance in mouse embryos. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 626–631 (2016).
7. Zhang, C.-Z. *et al.* Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
8. Mammel, A. E., Huang, H. Z., Gunn, A. L., Choo, E. & Hatch, E. M. Chromosome length and gene density contribute to micronuclear membrane stability. *Life Sci. Alliance* **5**, (2022).
9. Liu, S. *et al.* Nuclear envelope assembly defects link mitotic errors to chromothripsis. *Nature* **561**, 551–555 (2018).

10. Crasta, K. *et al.* DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**, 53–58 (2012).
11. Hatch, E. M., Fischer, A. H., Deerinck, T. J. & Hetzer, M. W. Catastrophic nuclear envelope collapse in cancer cell micronuclei. *Cell* **154**, 47–60 (2013).
12. Vietri, M. *et al.* Unrestrained ESCRT-III drives micronuclear catastrophe and chromosome fragmentation. *Nat. Cell Biol.* **22**, 856–867 (2020).
13. Mohr, L. *et al.* ER-directed TREX1 limits cGAS activation at micronuclei. *Mol. Cell* **81**, 724-738.e9 (2021).
14. Nader, G. P. de F. *et al.* Compromised nuclear envelope integrity drives TREX1-dependent DNA damage and tumor cell invasion. *Cell* **184**, 5230-5246.e22 (2021).
15. MacDonald, K. M. *et al.* Antecedent chromatin organization determines cGAS recruitment to ruptured micronuclei. *Nat. Commun.* **14**, 556 (2023).
16. Papathanasiou, S. *et al. Transgenerational transcriptional heterogeneity from cytoplasmic chromatin*. http://biorxiv.org/lookup/doi/10.1101/2022.01.12.475869 (2022) doi:10.1101/2022.01.12.475869.
17. 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).
18. Shoshani, O. *et al.* Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**, 137–141 (2021).
19. He, Q. *et al.* Chromosomal instability-induced senescence potentiates cell non-autonomous tumourigenic effects. *Oncogenesis* **7**, 1–18 (2018).

20. Santaguida, S., Tighe, A., D'Alise, A. M., Taylor, S. S. & Musacchio, A. Dissecting the role of MPS1 in chromosome biorientation and the spindle checkpoint through the small molecule inhibitor reversine. *J. Cell Biol.* **190**, 73–87 (2010).
21. Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J. & de Lange, T. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell* **163**, 1641–1654 (2015).
22. Sadaie, M. *et al.* Cell-based screen for altered nuclear phenotypes reveals senescence progression in polyploid cells after Aurora kinase B inhibition. *Mol. Biol. Cell* **26**, 2971–2985 (2015).
23. Santaguida, S. *et al.* Chromosome Mis-segregation Generates Cell-Cycle-Arrested Cells with Complex Karyotypes that Are Eliminated by the Immune System. *Dev. Cell* **41**, 638-651.e5 (2017).
24. Dertinger, S. D., Torous, D. K., Hayashi, M. & MacGregor, J. T. Flow cytometric scoring of micronucleated erythrocytes: an efficient platform for assessing in vivo cytogenetic damage. *Mutagenesis* **26**, 139–145 (2011).
25. Hasle, N. *et al.* High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol. Syst. Biol.* **16**, e9442 (2020).
26. Dürrbaum, M. *et al.* Unique features of the transcriptional response to model aneuploidy in human cells. *BMC Genomics* **15**, 139 (2014).
27. Matlashov, M. E. *et al.* A set of monomeric near-infrared fluorescent proteins for multicolor imaging across scales. *Nat. Commun.* **11**, 239 (2020).
28. Lopez-Urrutia, A. Deep Retina 3th place solution to Kaggle's 2018 Data Science Bowl. (2022).

29. Vassilev, L. T. Cell Cycle Synchronization at the G2/M Phase Border by Reversible Inhibition of CDK1. *Cell Cycle* **5**, 2555–2556 (2006).
30. Vereecke, L., Beyaert, R. & van Loo, G. The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. *Trends Immunol.* **30**, 383–391 (2009).
31. Verstrepen, L. *et al.* Expression, biological activities and mechanisms of action of A20 (TNFAIP3). *Biochem. Pharmacol.* **80**, 2009–2020 (2010).
32. Baron, V., Adamson, E. D., Calogero, A., Ragona, G. & Mercola, D. The transcription factor Egr1 is a direct regulator of multiple tumor suppressors including TGFβ1, PTEN, p53, and fibronectin. *Cancer Gene Ther.* **13**, 115–124 (2006).
33. Wang, B. *et al.* The Role of the Transcription Factor EGR1 in Cancer. *Front. Oncol.* **11**, (2021).
34. Rocca, A., Farolfi, A., Bravaccini, S., Schirone, A. & Amadori, D. Palbociclib (PD 0332991): targeting the cell cycle machinery in breast cancer. *Expert Opin. Pharmacother.* **15**, 407–420 (2014).
35. Hai, T., Wolford, C. C. & Chang, Y.-S. ATF3, a Hub of the Cellular Adaptive-Response Network, in the Pathogenesis of Diseases: Is Modulation of Inflammation a Unifying Component? *Gene Expr.* **15**, 1–11 (2018).
36. Ku, H.-C. & Cheng, C.-F. Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Front. Endocrinol.* **11**, (2020).
37. Hasim, M. S., Nessim, C., Villeneuve, P. J., Vanderhyden, B. C. & Dimitroulakos, J. Activating Transcription Factor 3 as a Novel Regulator of Chemotherapy Response in Breast Cancer. *Transl. Oncol.* **11**, 988–998 (2018).

38. Dickey, J. S. *et al.* H2AX: functional roles and potential applications. *Chromosoma* **118**, 683–692 (2009).
39. Nikolova, T. *et al.* The γH2AX Assay for Genotoxic and Nongenotoxic Agents: Comparison of H2AX Phosphorylation with Cell Death Response. *Toxicol. Sci.* **140**, 103–117 (2014).
40. Turinetto, V. & Giachino, C. Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic Acids Res.* **43**, 2489–2498 (2015).
41. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, (2021).
42. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).
43. Lin, T.-Y., Goyal, P., Girshick, R., He, K. & Dollar, P. Focal Loss for Dense Object Detection. in 2980–2988 (2017).
44. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds. Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) 265–273 (Springer International Publishing, 2018). doi:10.1007/978-3-030-00934-2_30.
45. hatch-lab/mnfinder: MN segmentation neural net. https://github.com/hatch-lab/mnfinder.

46. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, 2015). doi:10.1007/978-3-319-24574-4_28.
47. Frank, Hall, M. A., Witten, I. H. & Pal, C. J. The WEKA Workbench. in *Data Mining: Practical Machine Learning Tools and Techniques* (Morgan Kaufmann, 2016).
48. Cohen, W. W. Fast Effective Rule Induction. in 115–123 (1995).
49. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
51. Love, M. I. *et al.* Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLOS Comput. Biol.* **16**, e1007664 (2020).
52. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).
53. Korotkevich, G. *et al.* Fast gene set enrichment analysis. 060012 Preprint at https://doi.org/10.1101/060012 (2021).
54. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
55. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

56. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

57. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

58. The Gene Ontology Consortium *et al.* The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).

59. Greene, D., Richardson, S. & Turro, E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics* **33**, 1104–1106 (2017).

60. Crameri, F. Geodynamic diagnostics, scientific visualisation and StagLab 3.0. *Geosci. Model Dev.* **11**, 2541–2562 (2018).

61. Højsgaard, S., Halekoh, U. & Yan, J. The R Package geepack for Generalized Estimating Equations. *J. Stat. Softw.* **15**, 1–11 (2006).

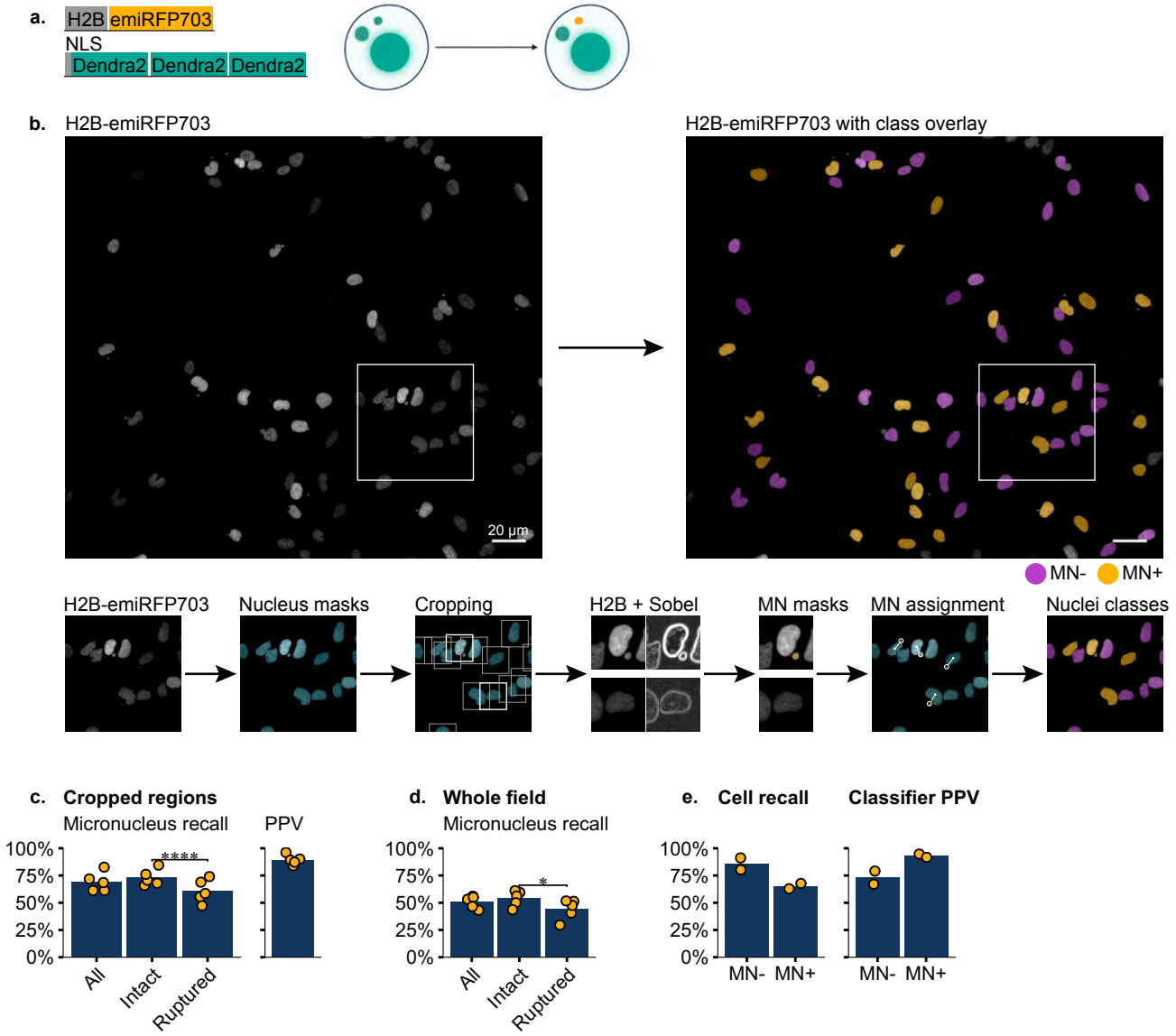


Figure 1: Machine vision identifies micronucleated cells
a) Diagrams of constructs transduced into RFP703/Dendra cells and how these constructs localize in micronucleated cells before and following micronucleus rupture. Graphic created with BioRender.com. **b)** Image of H2B-emiRFP703 in RFP703/Dendra cells at 20x and overlaid with MN+ and MN- classification results cells, with a visual depiction of the entire classification. **c,d)** Recall and positive predictive value (PPV) of micronuclei within regions analyzed by the neural net and recall for the entire field of view. Micronuclei were manually classified as intact or ruptured by Dendra2 intensity. N=5, n = 264, 158, 365, 283, 249. **e)** Recall and PPV of MN- and MN+ cells. N=2, n=328, 186.

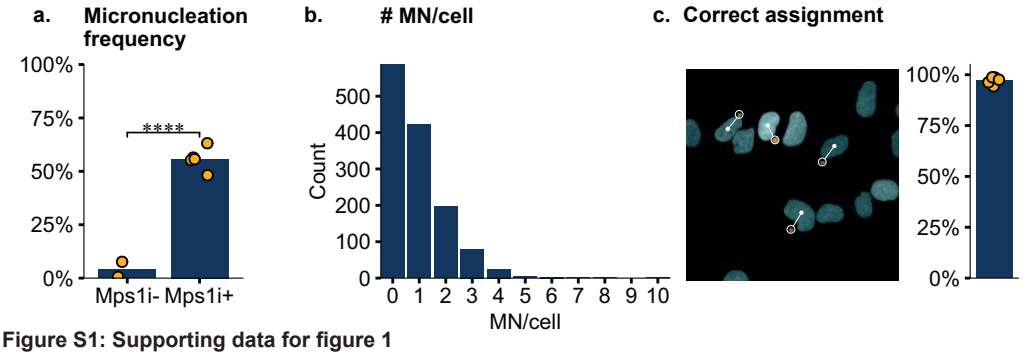


Figure S1: Supporting data for figure 1
a) Micronucleation rates in hTERT RPE-1s treated with either DMSO (Mps1-) or the Mps1i-inhibitor BAY-1217389 (Mps1i+) for 20 hrs. **b)** Histogram of the number of MN/cell in Mps1i+ RFP703/Dendra cells after treatment with Mps1i+ for 20 hrs. **c)** The proportion of MN assigned to the correct nucleus by proximity alone.

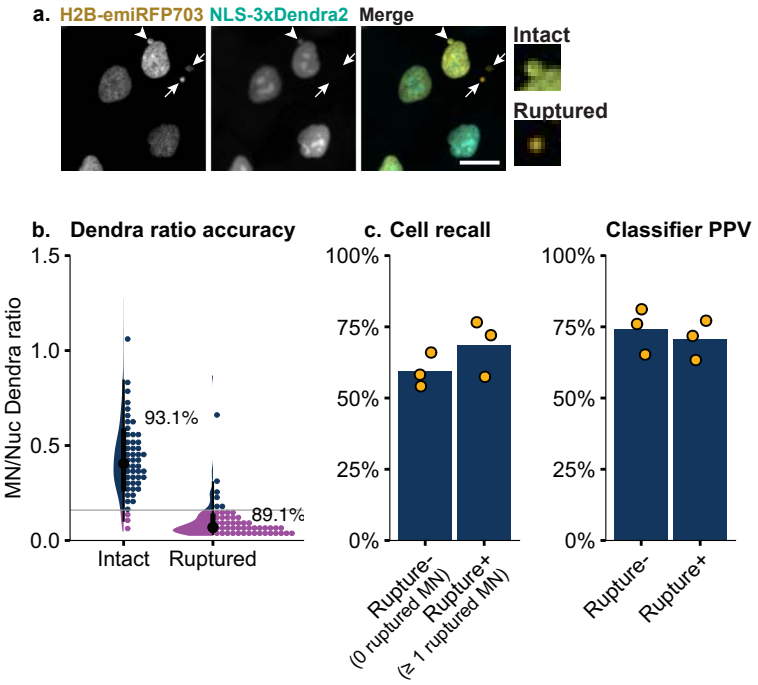


Figure 2: Micronucleated cells can be further classified into rupture+ and rupture- cells
a) RFP703/Dendra cells with ruptured (arrows) and intact (arrowhead) micronuclei showing a loss of NLS-3xDendra signal in ruptured micronuclei. **b)** Distributions of micronucleus/nucleus Dendra2 intensity ratio between manually classified intact o ruptured MN. Solid gray line = threshold used to automatically classify intact and ruptured MN. N=3, n=179, 113, 105. **c)** Recall and PPV for rupture-based cell classification. N=3, n=120, 91, 82.

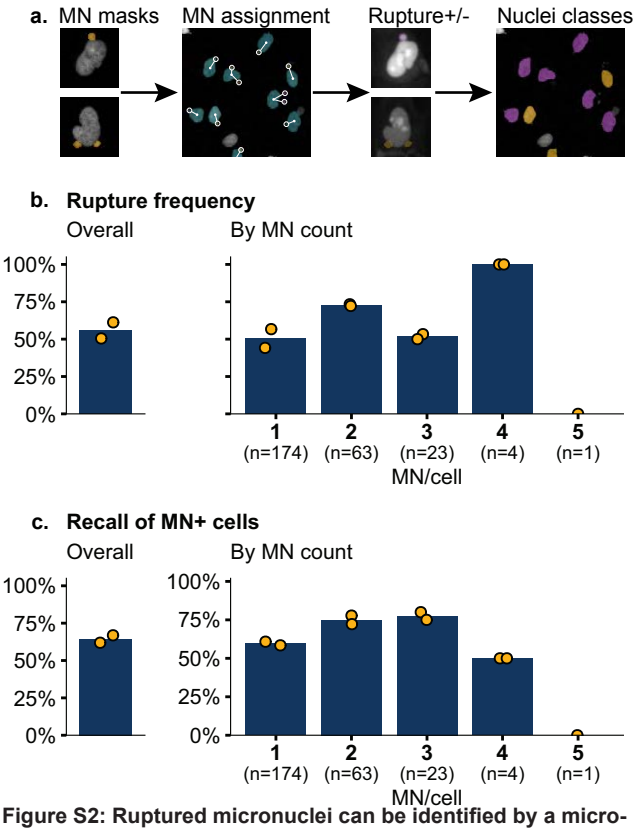


Figure S2: Ruptured micronuclei can be identified by a micronucleus/nucleus Dendra2 ratio
a) Modification to classification pipeline to include MN rupture status. **b,c)** Rupture frequency and recall of MN+ cells by # MN. Cells manually classified. N=2, n=328, 186.

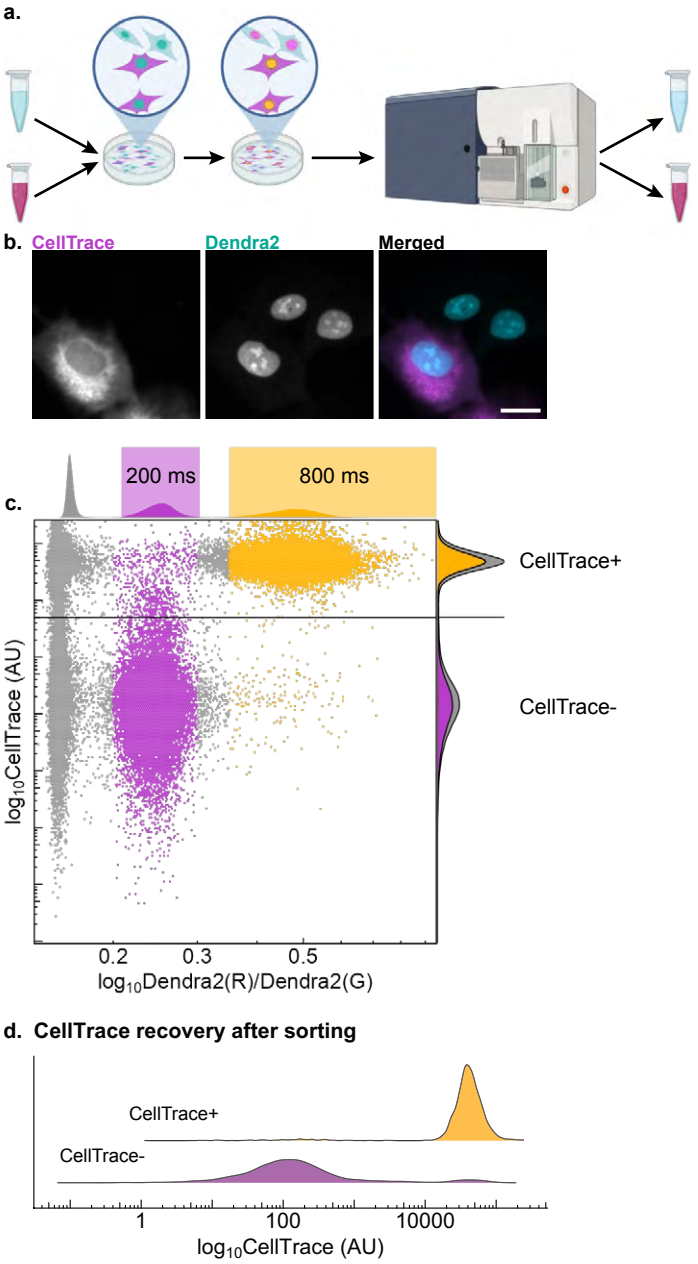


Figure 3: VCS can specifically isolate multiple cell populations from RFP703/Dendra cells
a) Overview of CellTrace mixture experiment: 1:1 mix of CellTrace+:CellTrace- RFP703/Dendra cells differentially photoconverted and sorted by Dendra ratio to recover the original unmixed populations. Graphic created BioRender.com. **b)** Representative image of unphotoconverted cells from CellTrace mixture experiment. **c)** FACS plot of cell population after differential photoconversion of CellTrace+ and CellTrace- cells with gates used to sort populations by UV exposure along the top margin. **d)** Re-analysis of CellTrace intensity of the two recovered populations after FACS.

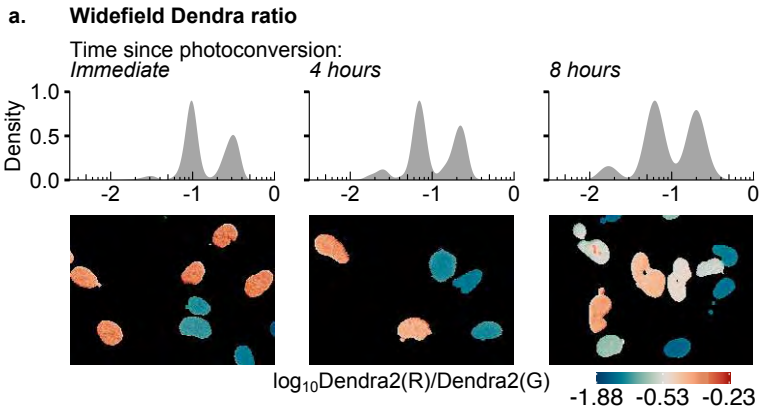


Figure S3: Dendra ratios persist for many hours
a) The Dendra-red/Dendra-green ratios from images taken post-photoconversion, pre-FACS during the experiment in Fig 4 showing 3 discernible populations even after 8 hours, including representative images from each time point color-coded by Dendra ratio.

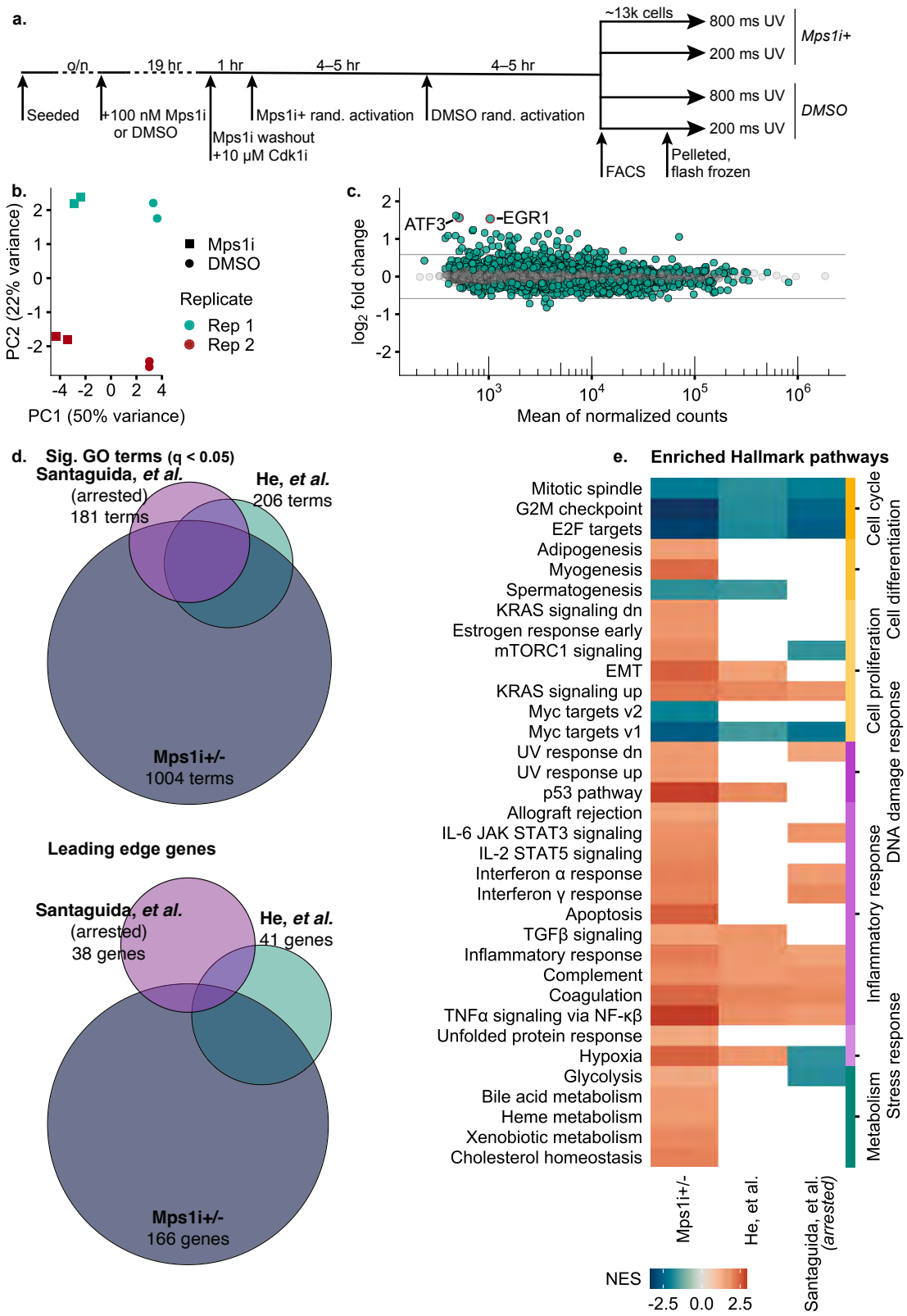


Figure 4: Spindle assembly checkpoint inhibition recapitulates characteristic aneuploidy transcriptional response

a) Timeline of experiment. **b)** PCA plot showing clustering of Mps1i-treated and DMSO-treated cells, along with minor clustering along a replicate axis. Each experimental replicate includes 2 technical replicates. **c)** MA plot. Genes with a differential expression q-value < 0.05 are highlighted in green. Gray lines represent 1.5 fold-change in expression. **d)** Venn diagram of the overlap in GO terms and leading edge genes between our data and the reanalyses of data from Santaguida, *et al.*[23] and He, *et al.*[19] **e)** Heatmap of Hallmark pathway enrichment between the 3 data sets in d). Hallmark pathways were grouped into broad categories on the right.

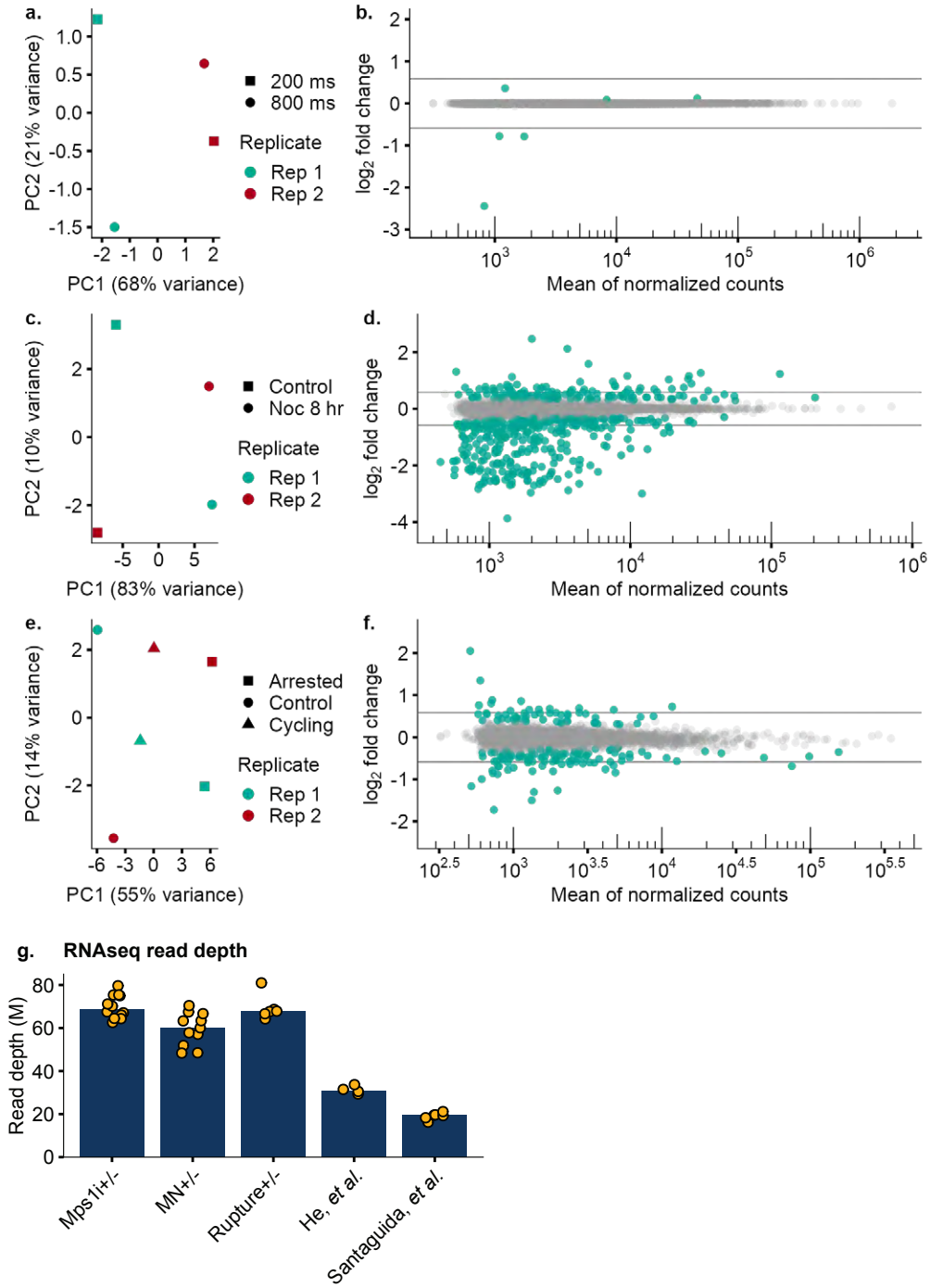
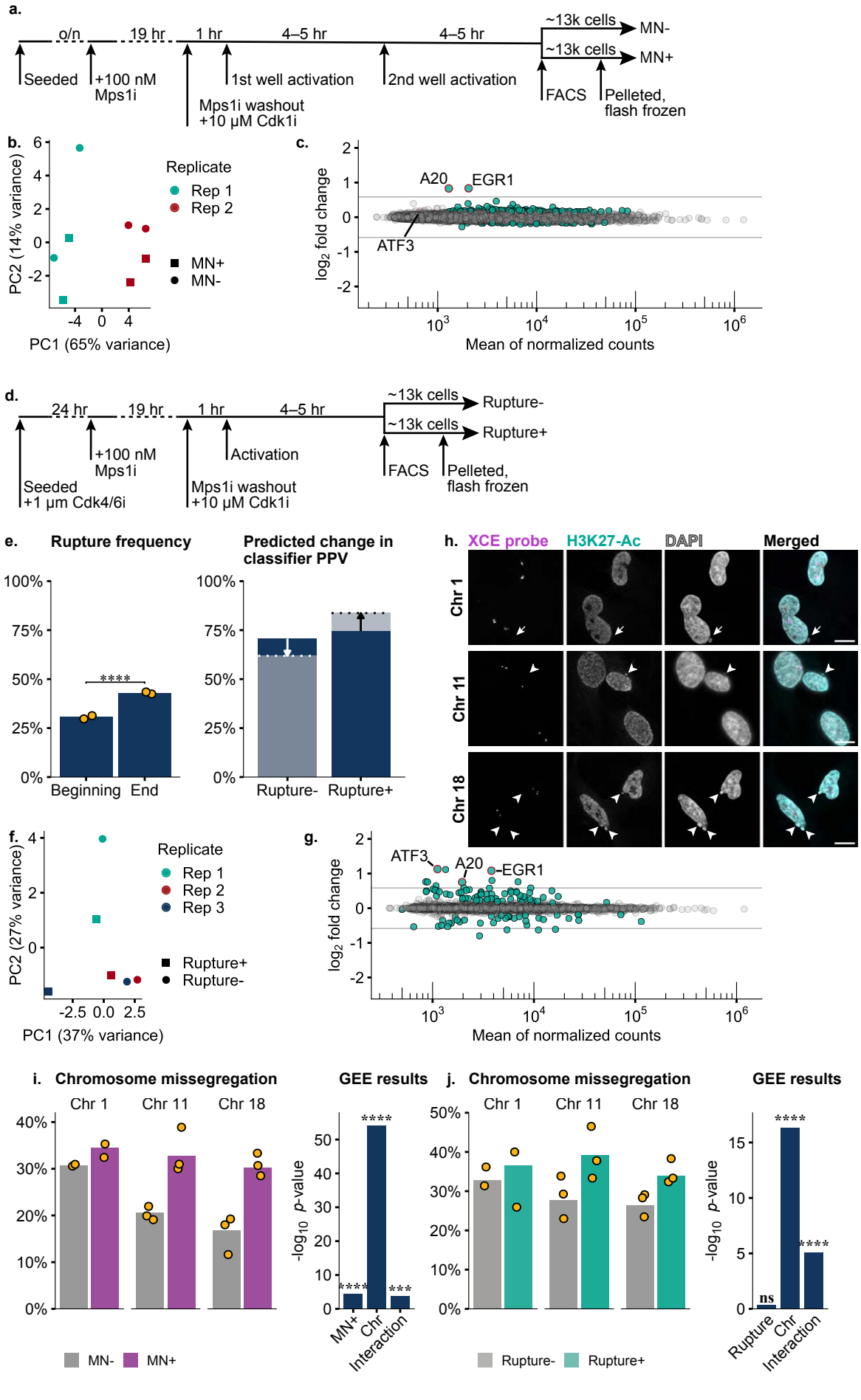


Figure S4: Controls and comparisons of differential expression analysis in Fig 4
a) PCA plot of cells treated with DMSO comparing those exposed to 800 ms UV vs 200 ms. Clustering is limited to replicate. **b)** MA plot of the data in a). Only 6 differentially-expressed genes were identified in cells exposed to 800 ms vs 200 ms UV. **c,d)** PCA plot and MA plot of the data reanalyzed from He, *et al.*[19] **e)** PCA plot of the data reanalyzed from Santaguida, *et al.*,[23] recapitulating the three clusters of cells they identified: cells treated with DMSO, cells treated with the Mps1 inhibitor reversine and which were still cycling, and cells treated with Mps1i but which arrested. **f)** MA plot of the data in e) comparing Mps1i+ cells with control. **g)** RNAseq read depth of our 3 analyses vs the data from Santaguida, *et al.* and He, *et al.* Dots represent each technical replicate. n=12, 11, 6, 4, 6.


a.
o/n
19 hr
1 hr
4–5 hr
4–5 hr
~13k cells
MN-
~13k cells
MN+
Seeded
+100 nM Mps1i
Mps1i washout +10 μM Cdk1i
1st well activation
2nd well activation
FACS
Pelleted, flash frozen
b.
PC2 (14% variance)
PC1 (65% variance)
Replicate
Rep 1
Rep 2
MN+
MN-
c.
A20
EGR1
ATF3
log2 fold change
Mean of normalized counts
d.
24 hr
19 hr
1 hr
4–5 hr
~13k cells
Rupture-
~13k cells
Rupture+
+100 nM Mps1i
Activation
Seeded +1 μm Cdk4/6i
Mps1i washout +10 μM Cdk1i
FACS
Pelleted, flash frozen
e.
Rupture frequency
Predicted change in classifier PPV
****
Beginning
End
Rupture-
Rupture+
h.
XCE probe
H3K27-Ac
DAPI
Merged
Chr 1
Chr 11
Chr 18
f.
PC2 (27% variance)
PC1 (37% variance)
Replicate
Rep 1
Rep 2
Rep 3
Rupture+
Rupture-
g.
ATF3
A20
EGR1
log2 fold change
Mean of normalized counts
i.
Chromosome missegregation
Chr 1
Chr 11
Chr 18
MN-
MN+
GEE results
-log10 p-value
****
****
***
MN+
Chr
Interaction
j.
Chromosome missegregation
Chr 1
Chr 11
Chr 18
Rupture-
Rupture+
GEE results
-log10 p-value
****
****
ns
Rupture
Chr
Interaction

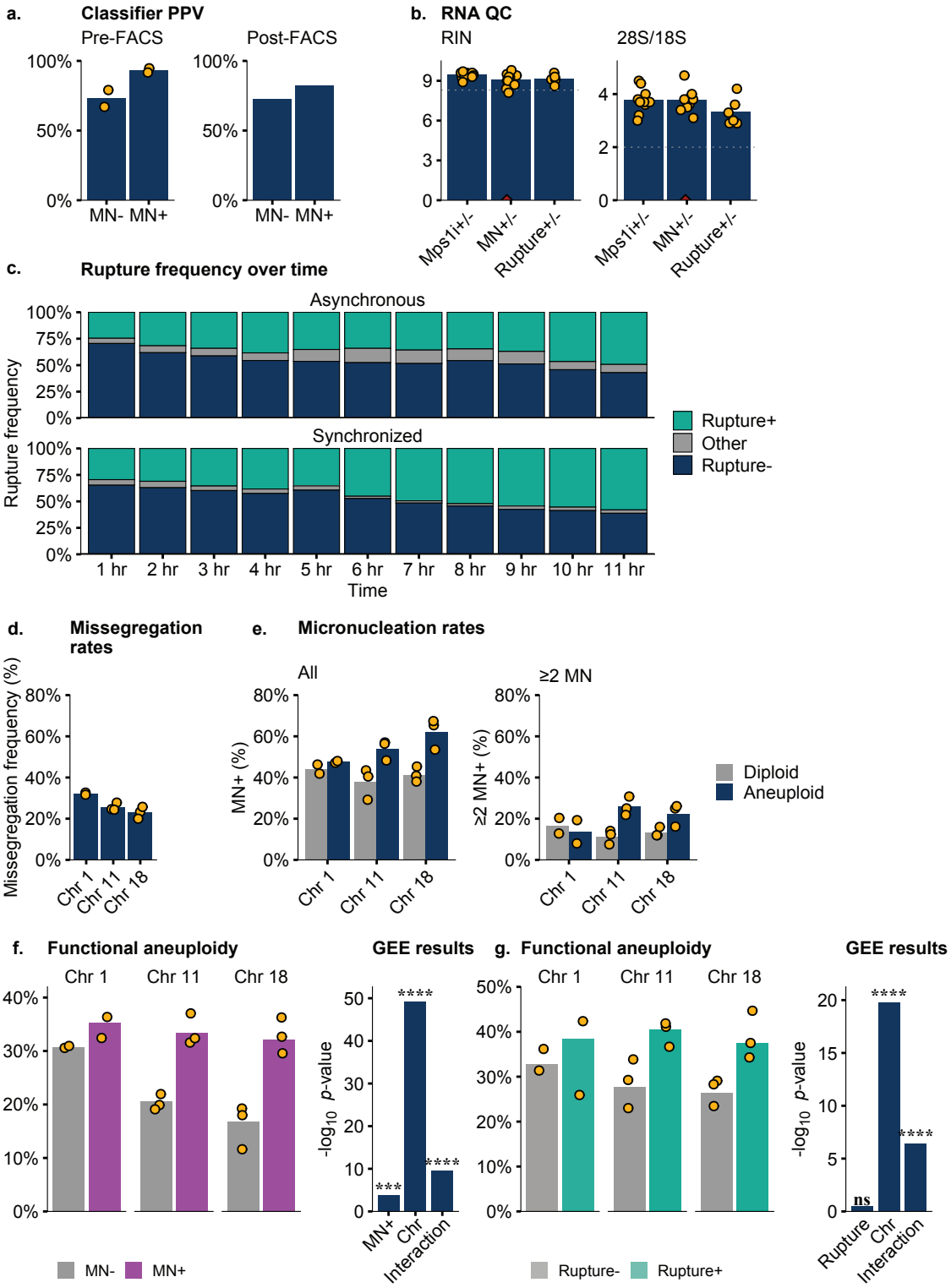


Figure S5: Supporting data related to figure 5

a) Accuracy after sorting cells and replating in Cdk1i to inhibit mitosis. Pre-FACS N=2; post-FACS N=1. **b)** RNA quality control measures from each experiment showing RIN scores and 28S/18S area measurements. Dots represent each technical replicate. Mps1i+/-: n=12, 11, 6. One technical replicate had insufficient RNA concentration to measure RIN or 28S/18S area, represented in red. **c)** Change in rupture frequency over time in asynchronous and synchronized cells treated with Cdk1i. Other represents mitotic or MN- cells, or cells without Dendra2 expression. N=1, n=~200 cells. **d,e)** Quantification of data in Fig 5h, showing overall missegregation frequency in chromosomes 1, 11, or 18, and micronucleation rates in cells with and without missegregated chromosomes 1, 11, or 18. N=2,3,3. **f,g)** Same as Figs 5i,j but with a different measure of aneuploidy. Chromosomes in ruptured micronuclei were considered to be functionally lost to the cell, so were ignored when tallying chromosomes/cell.

Figure 5: Micronucleus rupture produces a moderate early transcriptional response, but micronucleation *per se* does not

a) Timeline of experiment to isolate cells with and without MN. **b,c)** PCA and MA plots comparing MN+ vs MN- cells. **d)** Timeline of experiment to isolate micronucleated cells with only intact or at least 1 ruptured MN. **e)** Change in micronucleus rupture frequency between the start and end of an experiment and the predicted change in classifier PPV because of previously intact micronuclei rupturing before the end of an experiment. **f,g)** PCA and MA plots of rupture+ cells vs rupture- cells. h) Example of centromere FISH on chromosomes 1, 11, and 18 and the use of H3 lysine 27 acetylation (H3K27-Ac) to mark intact micronuclei. Ruptured micronuclei are annotated with arrows and intact with arrowheads. i) Quantification of the data in d). Chromosome missegregation examines the total number of chromosomes in each cell, regardless of its placement within a micronucleus. Chr 1: N=2, n=429, 158; Chr 11: N=3, n=406, 313, 160; Chr 18: N=3, n=425, 202, 230. j) Same as i), but comparing rupture- and rupture+ cells. Chr 1: N=2, n=187, 74; Chr 11: N=3, n=190, 108, 71; Chr 18: N=3, n=186, 102, 101.

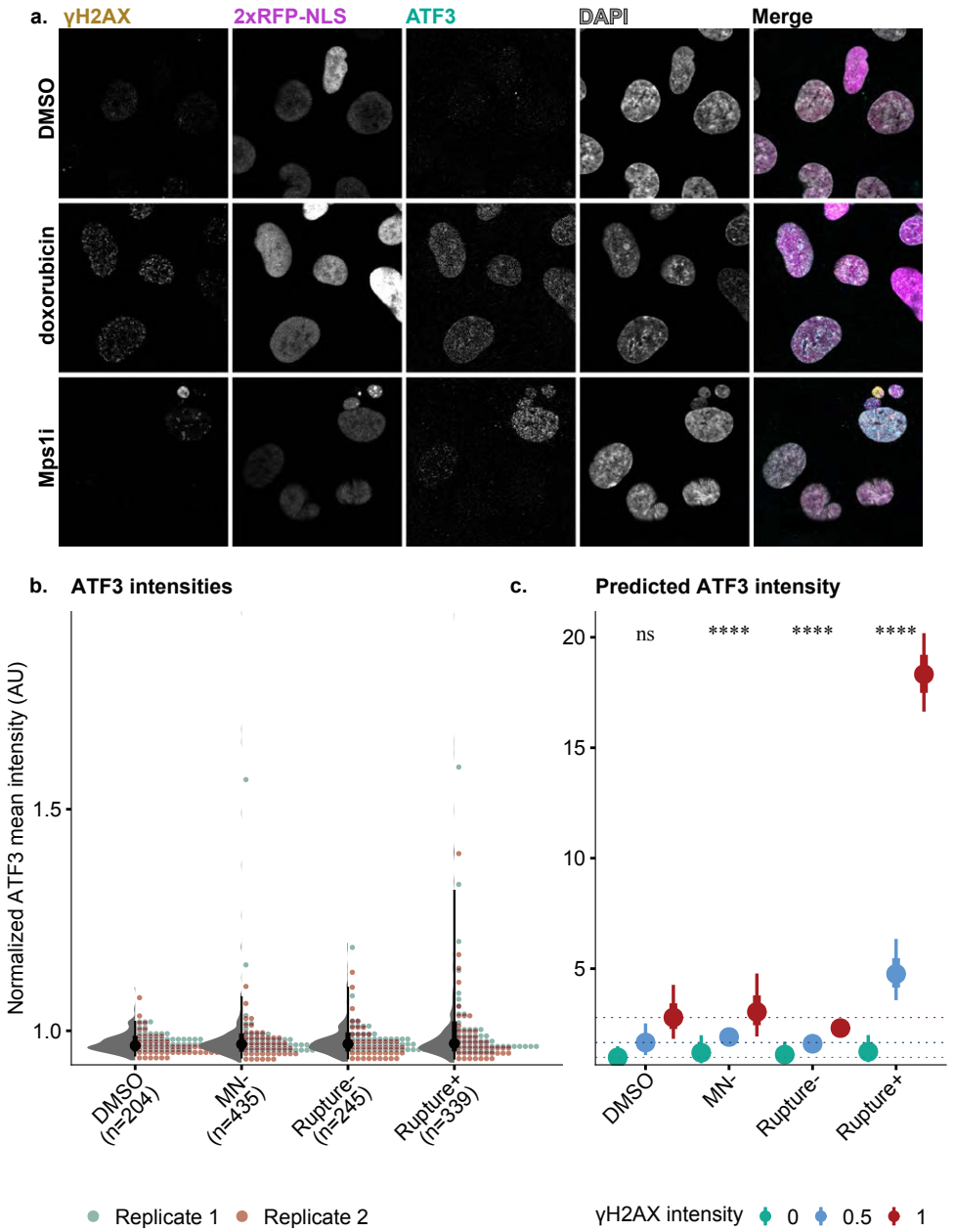


Figure 6: ATF3 is moderately increased in MN- and rupture+ populations
a) Example immunofluorescence images for quantifying γH2AX and ATF3 expression in cells treated with DMSO (neg. control), 2 µg/mL doxorubicin (pos. control), or with Mps1i. **b)** Violin & dot plots of ATF3 intensities from Mps1i- and DMSO-treated cells. ATF3 intensity was normalized to max and min ATF3 intensities from all conditions per replicate then translated such that the 90th percentile of DMSO data was 1. **c)** GEE predicted ATF3 values showing interactions between population subgroups and γH2AX intensity, with GEE significance results for each population. Intervals show mean prediction and 95% and 66% CIs. Dotted horizontal lines show the mean prediction from the DMSO group.

Figure 5: Micronucleus rupture produces a moderate early transcriptional response, but micronucleation *per se* does not
h) Example of centromere FISH on chromosomes 1, 11, and 18 and the use of H3 lysine 27 acetylation (H3K27-Ac) to mark intact micronuclei. Ruptured micronuclei are annotated with arrows and intact with arrowheads. **i)** Quantification of the data in h). Chromosome missegregation examines the total number of chromosomes in each cell, regardless of its placement within a micronucleus. Chr 1: N=2, n=429, 158; Chr 11: N=3, n=406, 313, 160; Chr 18: N=3, n=425, 202, 230. **j)** Same as i), but comparing rupture- and rupture+ cells. Chr 1: N=2, n=187, 74; Chr 11: N=3, n=190, 108, 71; Chr 18: N=3, n=186, 102, 101.

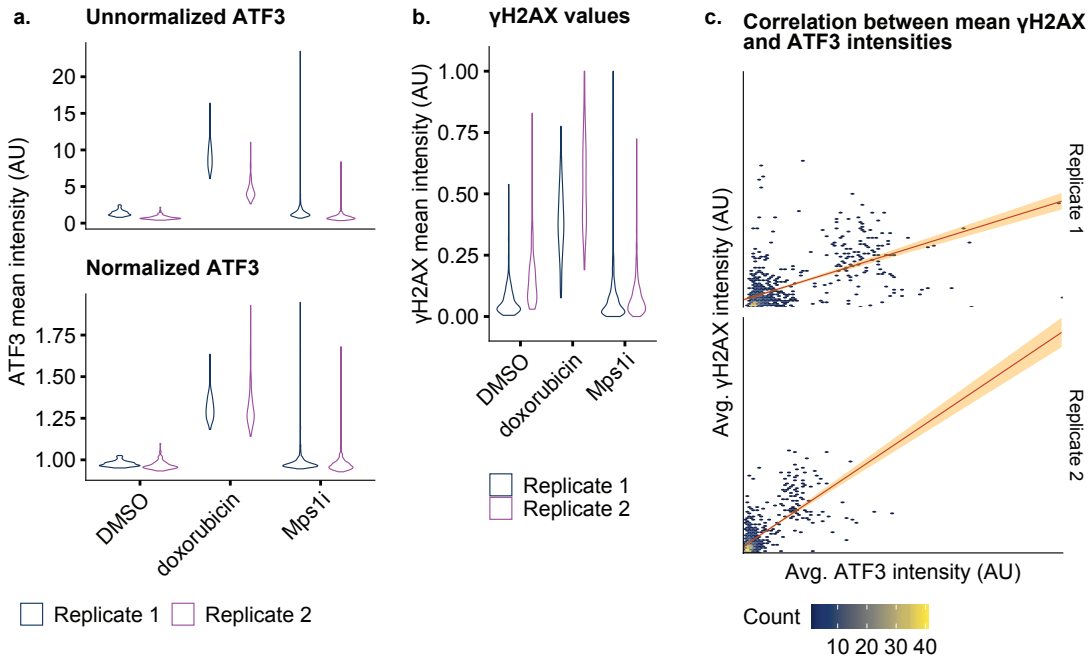


Figure S6: Supporting data for Figure 6

a) Violin plots of ATF3 intensities with and without rescaling or translation applied in all treatment conditions. **b)** Violin plots of γH2AX intensities in all treatment conditions. **c)** Plots of correlation between γH2AX and ATF3 intensities per replicate, justifying the inclusion of γH2AX intensity in the GEE.

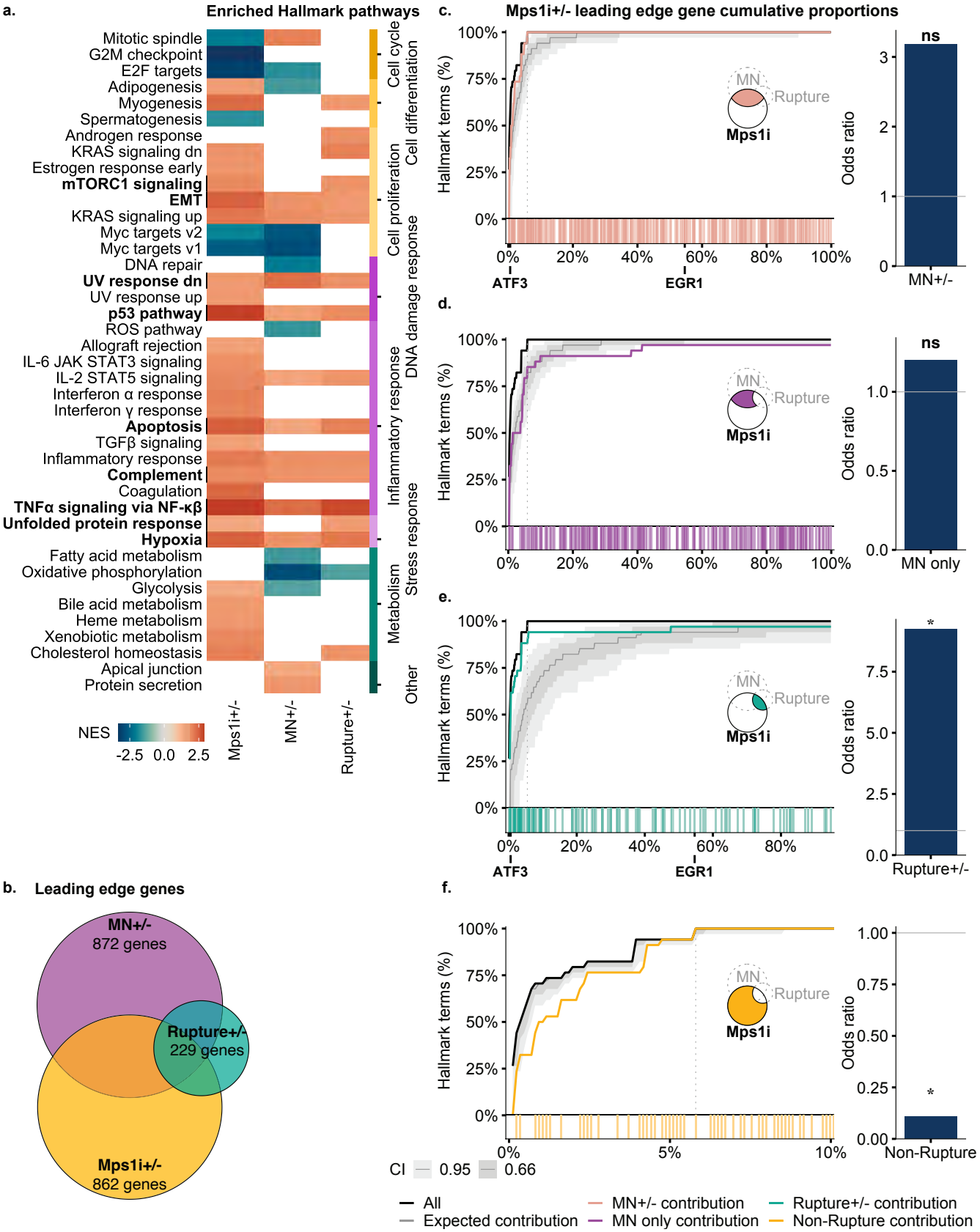
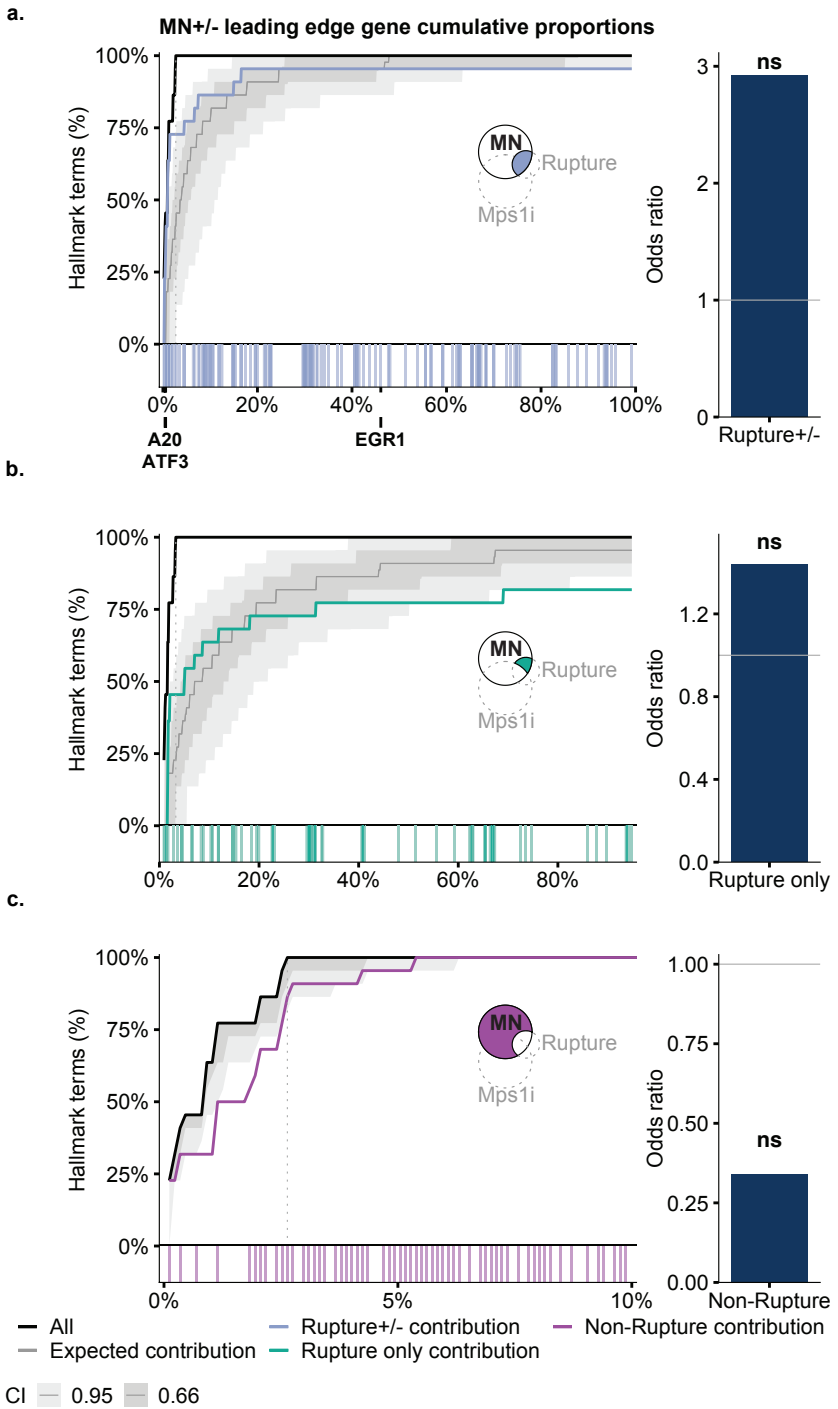


Figure 7: MN rupture is a strong contributor to the transcriptomic response to Mps1i treatment
a) Heatmap of Hallmark pathways identified in RFP703/Dendra cells treated with DMSO vs Mps1i; Mps1i-treated cells with and without MN; and synchronized, Mps1i-treated, MN+ cells with and without MN rupture. Pathways with an overrepresentation of rupture+/- leading edge genes are marked in bold. **b)** Venn diagram of overlap in leading edge genes between the three data sets. **c)** Analysis of the contribution of micronucleation leading edge genes to the Mps1i+/- transcriptomic response. Genes that are sufficient to represent all Hallmark pathways identified in these experiments (core genes) are to the left of the dotted line, and the placement of shared MN+/- genes displayed along the x-axis. The number of Hallmark pathways represented only by the MN+/- subset are represented in pink, with the expected contribution if these genes were randomly distributed in gray. An odds ratio comparing the odds of shared genes being among core genes is represented on the right. **d–f)** Same as c, but with different subsets of genes identified from overlap as indicated. f) displays core gene only.



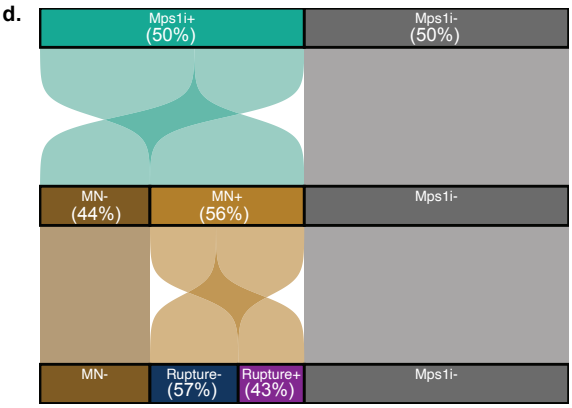


Figure S7: MN rupture is a contributer to the transcriptomic response to micronucleation
a–c) Same as Figs 7b–f, but examining contributions to the MN+/- transcriptomic response. **d)** Sankey diagram showing the breakdown of populations in the Mps1i+/- treatment condition.

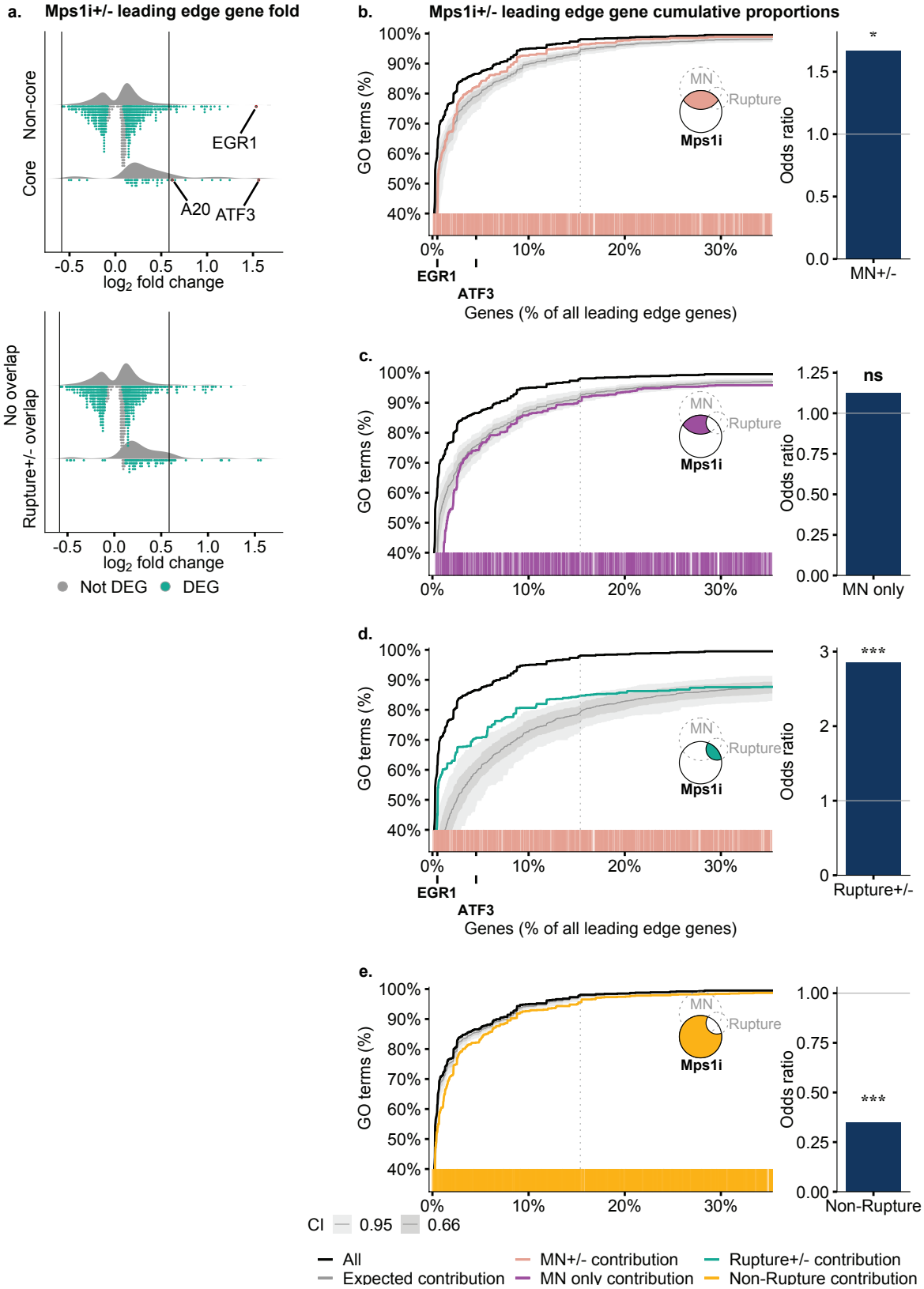


Figure S8: MN rupture predicts changes to GO terms
a) Rain cloud plots of the fold-change distribution of leading edge genes from the transcriptomic response to Mps1i treatment, by core and non-core genes, and by those genes that are leading edge genes in rupture+ cells vs not. DEGs are highlighted in green. Vertical lines mark a 1.5 fold change. **b–e)** Same as Figs 7c–f but looking at the GO ontology.

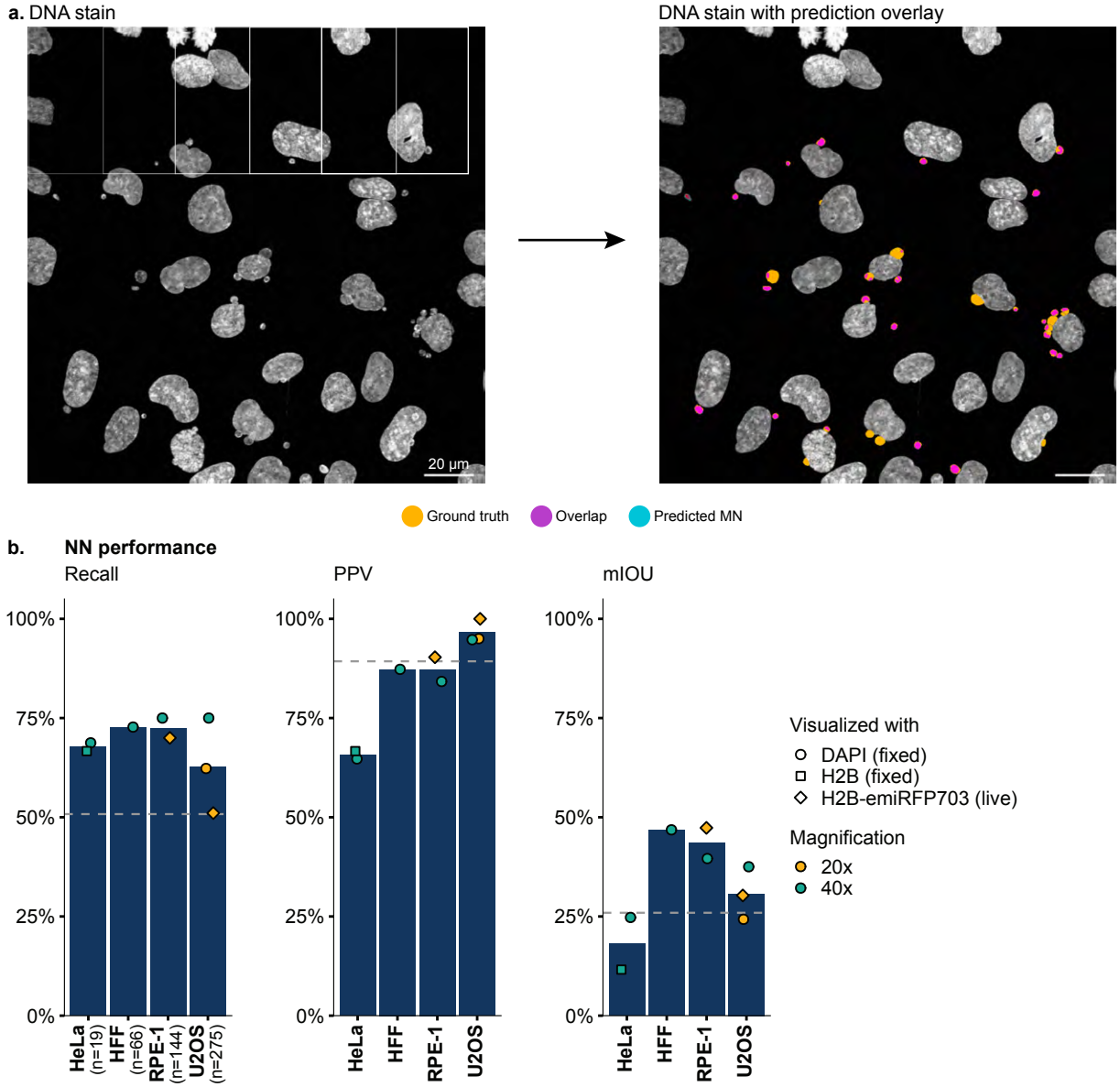


Figure 8: An updated neural net improves micronucleus segmentation for several different cell lines and imaging conditions
a) Overview of the neural net process showing oversampling by sliding window, with overlays showing ground truth, MN predictions, and overlap. **b)** Recall, PPV, and mean intersection over union (mIOU) by cell line. Bars indicate mean performance and dashed line the baseline performance from the VCS neural net.