PNAS

Supporting Information for
A survey of chromosomal instability measures across mechanistic models

Andrew R. Lynch[1,2], Shermineh Bradford[1,2], Amber S. Zhou[1,2], Kim Oxendine[3], Les Henderson[3], Vanessa L. Horner[3], Beth A. Weaver[1,2,5], Mark E. Burkard[1, 2, 4, 6]

Corresponding Author: Mark Burkard
Email: mburkard@wisc.edu

This PDF file includes:

Supplemental Methods

Cell line derivation

MCF10A-PLK4-WT-tetOn cells were kindly provided by the laboratory of Dr. David Pellman (12). We used lentiviral transduction to stably express H2B-mNeonGren and mScarlet-α-Tubulin. CAL51 cells were obtained from DSMZ-German Collection of Microorganisms and Cell Cultures and CAL51-TERF2-DN-tetOn was generated using retroviral transduction of pCMV Retro TetO into which we cloned the TERF2-DN mutant sequence (Addgene, 16069) and an mScarlet reporter. To make retrovirus, we transfected 293T cells then transduced CAL51 cells expressing TetR for ~18 hours with 4 ug/mL polybrene. We selected with puromycin (2 ug/mL) and sub-cloned in 96-wells to generate mon-oclonal lines. All cell lines were tested and free from mycoplasma contamination prior to study. Standard cultivation conditions are described in Supplemental Methods.

Cell line cultivation

Cells were maintained at 37ºC and 5% CO_2 in a humidified, water-jacked incubator and propagated in either Dulbecco's Modified Eagle's Medium (DMEM)/High Glucose (Cytiva Hyclone, SH3024301) for CAL51 cells or mammary epithelial basal media for MCF10A cells, consisting of DMEM/F12 (Cytiva Hyclone, SH3026101), 5% horse serum (Gibco, 16050122), 20 ng/mL EGF (Peptrotech, AF-100-15), 0.5 mg/ml hydrocortisone (MP Biomedicals, 0219456901), 100 ng/ml cholera toxin (Enzo Life Sciences, BMLG1170001), and 10 ug/ml insulin (Millipore Sigma, I9278). All growth media is supplemented with 10% fetal bovine serum (GeminiBio, 900-108),100 units/mL penicillin-streptomycin (Gibco, 15070063), and plasmocin prophylactic (Invivogen, ant-mpp) to a final concentration of 5 ug/mL.

Time lapse fluorescence microscopy

Cells were re-seeded in 4- or 8-well chamber slides (Ibidi, 80426) at 40% and grown to ~70-80% over 18-24 hours. At the time of re-plating, CAL51-TERF2-DN-mCherry-TetOn cells were transduced with adenovirus to express β-tubulin-mScarlet + P2A-H2B-NeonGreen to visualize mitotic spindles and DNA. They were then imaged on a Nikon Ti-E2 inverted fluorescence microscope outfitted with a Yokogawa CSU-W1 spinning disk confocal scanning unit. Images were collected every 2.5 minutes for 4-12 hours with a 20x/0.5NA (P Fluotar) objective and an ORCA Flash 4 digital sCMOS camera (Ha-mamatsu). Environmental control was maintained by a humidified, stage-top chamber (Tokei Hit) set to 37°C and 5% CO2. Micrograph montages were prepared in Affinity Designer.

For analysis, at least 20 cells for each replicate and each CIN model were tracked from nuclear en-velope breakdown through to cytokinesis and mitotic phenotypes were recorded for metaphase, ana-phase, and telophase. Exceptions are replicate 1 of CtrlC, with 17 cells, and replicate 5 of Pb, with 18 cells. Data from these replicates were retained and reported as the alternative measures of the replicate provided sufficient information. Standard procedures are described in Supplemental Methods.

Cytogenetics

Cells were re-seeded in T75 flasks at 40% and grown to ~70-80% density over 18-24 hours prior to harvest. For Po, AZ3146 was washed out at T-0 hours and harvest performed at T+24 hours for mitotic chromosome spreads and interphase fluorescence in situ hybridization (FISH) of centromeres of chro-mosomes 3, 4, 7, 9, 10, and 17. For analysis, at least 50 mitotic chromosome spreads were counted for each replicate of each model with the exception of the Mp model, which exhibited a relatively few mitot-ic cells in each biological replicate (18, 32, and 35 mitotic chromosome spreads). For FISH, at least 100 interphase cells were counted per 3-probe set. Standard procedures are described in Supplemental Methods.

Statistical Analyses

Statistical analysis was completed in R/Rstudio (v4.2.3/v1.2.5019) (1, 1). Unless otherwise specified, statistical significance between group means is determined using two-tailed, two-sample Student's *t*-tests and $\alpha = 0.05$ over at least 3 biological replicates and corrected for multiple comparisons via the Benjamini-Hochberg method.

Fixed immunofluorescence microscopy

Coverslips are rinsed in warmed (37°C) PBS (x1) before fixation in warmed 4% paraformaldehyde (PFA) in PHEM buffer (120 mM PIPES, 50 mM HEPES, 20 mM EGTA, 4 mM $MgSO_4$ in pure water, pH 7.0) for 10 minutes. Coverslips are rinsed of fixative (x3), extracted with 1% NP40 in PHEM buffer for 15 minutes, and blocked with 3% bovine surum albumin (BSA)(Millipore Sigma, A2153) and 0.1% Triton X-100 in PBS (BSA + PBSTx) for 30 minutes. Primary and secondary antibodies were pooled separately in BSA + PBSTx. Coverslips were incubated with primary antibodies for 1 hour at room temperature, rinsed (x3) in PBSTx, then incubated with secondary antibodies for 30 minutes, and rinsed (x3) again in PBSTx. Coverslips were counterstained with 10 ug/mL DAPI (Sigma-Aldrich, 62248), mounted on glass slides with Prolong Dimaond anti-fade medium (Molecular Probes, P36970), and cured for 48 hours. Cells were immunostained to visualize α-tubulin (Primary — 1:1000 mouse (γ1) anti-α-tubulin (DM1A)(Invitrogen, 14-4502-82, RRID: AB_1210456; Secondary — 1:350 goat anti-mouse (γ1) + Alexa Fluor 750 (Invitrogen, A-21037, RRID: AB_2535708) for CAL51 cells or goat anti-mouse (H+L) + Alexa Fluor 555 (Invitrogen, A-11001, RRID: AB_2534069) for MCF10A cells) and pericentrin (Primary — 1:1000 rabbit anti-pericentrin (Abcam, ab4448); Secondary — 1:350 chicken anti-rabbit (H+L) + Alexa Fluor 647 (Invitrogen, A-21443, RRID: AB_2535861) for CAL51 cells or goat anti-rabbit (H+L) + Alexa Fluor 488 (Invitrogen, A-11008, RRID: AB_143165) for MCF10A cells)

Image acquisition was performed on a Nikon Eclipse Ti inverted microscope equipped with motorized stage, LED epifluorescence light source (Spectra X), 60x/1.4NA (Plan Apo) DIC oil immersion objective, and ORCA Flash4.0 V2+ digital sCMOS camera (Hamamatsu). Optical sections were taken at 200-nm intervals and deconvolved using the LIM 3D Deconvolution module in Nikon Elements. Micrograph montages were prepared in Affinity Designer.

Time lapse fluorescence microscopy

They were then imaged on a Nikon Ti-E2 inverted fluorescence microscope outfitted with a Yokogawa CSU-W1 spinning disk confocal scanning unit. Images were collected every 2.5 minutes for 4-12 hours with a 20x/0.5NA (P Fluotar) objective and an ORCA Flash 4 digital sCMOS camera (Hamamatsu). Environmental control was maintained by a humidified, stage-top chamber (Tokei Hit) set to 37°C and 5% CO_2. Micrograph montages were prepared in Affinity Designer.

Imaging-based approximation of mis-segregations per diploid division (MDD)

MDD was approximated using fixed immunofluorescence and time lapse fluorescence imaging, as previously described (36), using the formula

$$\mathrm{MDD} = \sum_{\varphi_i} \frac{\pi \times \mu \times (1-\rho) \times 46}{\theta}$$

where φ is the mitotic phenotype of CIN (i.e. lagging, bridging etc.), π is the penetrance or rate of that mitotic phenotype per division, μ is the magnitude or number of chromosomes involved in that mitotic phenotype, ρ is the resolution rate or rate at which the a chromosome that is involved in a mitotic defect is segregated into the correct daughter cell and θ is the ploidy or total number of modal chromosomes in the sample. In long-from, this is

$$\mathrm{MDD} = \sum_{\varphi_i} \frac{\text{Mitotic Error Penetrance} \times \text{Missegregation } \mathit{Magnitude} \times (1 - \text{Resolution Rate}) \times 46}{\text{Number of Modal Chromosomes}}.$$

To derive the values for this formula, we made the following reasonable assumptions. Bridging chromosomes are assumed to produce 1 mis-segregation (magnitude of 1, 0% resolution). Lagging chromosomes are assumed to produce less than 1 mis-segregation (magnitude of 1, 90% resolution, an approximation based on Thompson & Compton 2011 (13). Polar chromosomes are assumed to produce 7.8 mis-segregations as we observed cells treated under the Po model (prior to anaphase onset) had an average of 15.6 polar ACA foci **(Supplemental Figure 1A)** (7.8 chromosomes, 0% resolution as polar segregated polar chromosomes will always affect both sister chromatids). Pseudobipolar spindles are assumed to not produce their own mis-segregation events in and of themselves as, aside from presentation through other phenotypes (lagging, polar etc.), a number of chromosomes cannot be discerned. Multipolar spindles are assumed to produce 18 mis-segregations based on Lynch et al. 2022 (18 chromosomes, 0% resolution. CAL51 cells have 46 chromosomes and MCF10A cells have 47 chromosomes based on our sequencing data. The penetrance of any phenotype is taken at the phase of the cell cycle where it is most readily detectable. Thus the fraction of cells with polar chromosomes represents metaphase cells wherein they are most readily detectable. All other CIN phenotypes are taken from anaphase or telophase cells.

To assess the sensitivity of these assumptions we performed a standard univariate sensitivity analysis where we individually changed the magnitude by each assumed phenotype (lagging, bridging, etc.) by ± 50%. We then computed a sensitivity index (14) for each phenotype as the percentage difference in MDD between the minimum and maximum values as follows:

$$SI_{\varphi} = \frac{MDD_{\varphi,max} - mdd_{\varphi,min}}{MDD_{\varphi,max}}$$

where MDD_{max} and MDD_{min} are the maximum and minimum values of MDD when the magnitude of phenotype φ is changed and the others are kept the same.

Cytogenetics

Colcemid (Millipore Sigma, 234109) was added to a final concentration of 50 ng/mL and incubated for 2 hours to enrich mitotic cells. We retained the media and trypsinized cells, rinsing once in warm PBS. Cells were resuspended and swelled for 5 minutes in 75 mM KCl with 10 drops of Carnoy fixative (3:1 methanol:glacial acetic acid). This was followed by three successive resuspensions in Carnoy fixative, after which, samples were stored at -30°C then cells were dropped onto slides and dried in a drying chamber.

Fluorescence in situ hybridization (FISH) enumeration of chromosomes 3, 4, 7, 9, 10, and 17 was completed using the following 2 probe mixes: Vysis CEP 3 (D3Z1) labeled SpectrumOrange (Vysis, 06J3613) localizing to 3p11.1-q11.1, Vysis CEP 7 (D7Z1) labeled SpectrumAqua (Vysis, 06J5427) localizing to 7p11.1-q11.1, Vysis CEP 9 labeled SpectrumGreen (Vysis, 06J3719) localizing to 9p11-q11 in IntelliFISH hybridization buffer (Vysis, 08N8701), and Vysis CEP 4 labeled SpectrumGreen (Vysis, 06J3714) localizing to 4p11-q11, Vysis CEP 10 labeled SpectrumAqua (Vysis, 06J5420) localizing to 10p11.1-q11.1 and Vysis CEP 17 (D17Z1) labeled SpectrumOrange (Vysis, 06J3697) localizing to 17p11.1-q11.1 in IntelliFISH hybridization buffer (Abbott Molecular, Des Plaines, IL). Slides were aged for 2 minutes in 2xSSC at 74°C and put through a dehydration ethanol series (70%, 85%, and 95%). Slides and probe were codenatured by heating at 80°C for 2 minutes using a ThermoBrite instrument (Abbott Molecular). Hybridization was performed overnight at 37°C. Finally, the slides were mounted with Vectashield containing DAPI (Vector Laboratories). Localization of the probes was confirmed on pooled cytogenetically normal blood controls.

Single-cell DNA sequencing and analysis

Single cell DNA sample preparation

Cells were re-seeded in 6-well plates at 40% and grown to ~70-80% density over 18-24 hours prior to harvest. Cells were washed with HBSS, trypsinized, pelleted, and washed by resuspension in wash buffer (2% FBS in 1x PBS). Pelleting once more, cells were resuspended in 500 µL wash buffer and fixed by dropwise addition to 4.5 mL 80% ethanol. Samples were stored at -30°C.

Flow cytometry and fluorescence activated cell sorting

Fixed cell suspensions were pelleted and resuspended in wash buffer containing 10 µg/mL DAPI. Single cells or bulk samples (500 cells) were sorted by FACS (BD FACS AriaII), gating on 0.5-1.5x the DAPI signal intensity of the G1 peak, through a 130 µm low-pressure deposition nozzle into 96 well PCR plates containing 10 µL 1x prepared Cell Lysis and Fragmentation Buffer Mix (Millipore Sigma, WGA4). Doublets were excluded from analysis via standard FSC/SSC gating procedures. DNA content was analyzed via DAPI excitation at 355

nm and 450/50 emission using a 410 nm long pass dichroic filter. Gating for cell sorting was defined by 0.5x/1.5x (lower/upper) the DAPI intensity of the G1 peak. After sorting, the PCR plates were immediately centrifuged at 100 x g for 1 minute before library preparation. Plates were kept at 4°C when not on the cell sorter.

Single cell DNA library preparation

Initial cell lysis, genomic fragmentation, and genomic amplification reactions were done with the GenomePlex® Single Cell Whole Genome Amplification Kit (Millipore Sigma, WGA4). Initial genomic library distributions were assessed on a 1.5% agarose gel and purified using a ZR-96 Genomic DNA Clean & Concentrator-5 Kit (Zymo Research, D4067). Library concentrations were quantified using the Quant-iT™ dsDNA Broad Range Assay Kit (Invitrogen, Q33130) and normalized before additional preparation. Genomic libraries were enzymatically fragmented to ~250 bp, 5'-phosphorylated, 3'-dA-tailed, and index adaptor-ligated with the sparQ DNA Frag & Library Prep Kit (Quantabio, 95194). Ligated adaptors were standard P5 and custom uniquely indexed P7 illumina adaptors described previously (2). Indexed libraries underwent eight additional amplification cycles, purification with Axygen® AxyPrep MAG PCR Clean-Up beads (Corning, MAG-PCR-CL), and DNA concentration quantification as above. Up to 96 libraries with unique indices were pooled in equimolar concentrations. Library quality was validated on an Agilent TapeStation and concentration was measured via Qubit 2.0 fluorometer and qPCR.

Single cell sequencing and data pre-processing

Paired-end bulk RNA sequencing (2x150bp) was performed on an Illumina HiSeq sequencer and demultiplexed using Illumina bcl2fastq (v2.20). Reads were trimmed of adaptor sequences, quality filtered, and de-duplicated in fastp (v0.23.2) and aligned to reference sequence hg19 with bowtie2 (v2.3.5). Format conversions were performed with samtools (v1.9) and bedtools (v2.29.2).

Inference of mis-segregation rates

Agent based simulation. We initiated simulated populations with 100 diploid cells and evolved exponentially growing populations using a pseudo-Moran process to reduce computational demands (a random 50% of cells are culled when the population surpasses 3000 cells). Euploid cells had a 50% chance to divide at every step, a probability that is modified according to a cell's fitness level. We used a karyotype selection scheme that emulates stabilizing selection by negatively selecting genetically unbalanced karyotypes. The contribution of each chromosome to karyotype selection is dependent on the abundance of genes it encodes. We simulated populations using the following parameters: MDD = [0, 0.046, 0.092, …, 46], S = [0, 1, 2, …, 200], Time Steps = [0, 1, 2, …, 100]. We assumed whole-chromosome mis-segregation and that chromosome copy numbers below 1 and above 7 would result in cell death. Accordingly, as cells divide and mis-segregate chromosomes, more aneuploid cells with more unbalanced karyotypic stoichiometries are less likely to continue division. A random selection of 300 karyotypes is exported to measure summary statistics at each time step. Each combination was repeated 3 times for a total of 1,006,005 independent simulations. See Lynch et al. 2022 for additional details.

Population summary statistics. To summarize the characteristics of karyotypes from simulated populations and scDNAseq data, we quantified three features: aneuploidy, mean karyotype variance (MKV), and Colless index. Aneuploidy was quantified as the mean variance of copy numbers within each cell's sub-clonal karyotype (normalized to the modal karyotype). MKV was quantified as the mean variance of copy numbers for each chromosome across the population. Colless index was measured using the R package phyloTop (v2.1.1)(3) and was normalized to the number of leaf tips. Phylogenetic trees for measuring the Colless index were reconstructed from chromosome copy numbers by computing Euclidean distance matrices and hierarchical clustering using complete linkage, both in the R stats package (v.4.2.3)(1). Hierarchical clustering may not produce the same results every time, and rare, highly different dis-similar observations can drastically skew Colless indices, both of which result in measurement noise. To reduce this noise we, we permute the copy numbers of individual homologous chromosomes (i.e., a population's copy numbers for chromosome 1 are permuted separately from those of chromosome 2) across the population and repeat this Colless index measurement, taking the average of 200 iterations. This preserves phylogenetic imbalance for populations with sub-clonal alterations but reduces imbalance for those with a rare highly dissimilar karyotype, resulting in a stable and reproducible measure. Also, because hierarchical clustering requires at least 3 observations, we removed all simulated datasets that failed to maintain at least 3 cells, which could occur when high mis-segregation rates force cells past the pre-defined lethal chromosome copy number thresholds (1 to 7), promoting population collapse.

Approximate Bayesian computation. Parameter inference of mis-segregation rates and selection pressure from scDNAseq data was performed with approximate Bayesian computation using our simulated dataset as a prior dataset. We used the R package abc (v2.1) (4) to complete the analysis with rejection sampling and a tolerance threshold of 0.01. We chose independent prior datasets that best matched each control and experimental group. For example, we assigned control groups a prior dataset that encompassed 30-50 time steps (i.e., 15-25 generations) to reflect the number of passages after mono-clonal culture. Groups with doxycycline-induced CIN were assigned a prior dataset with ≤ 4 time steps (2 generations) to encompass about 48 hours of growth with fully penetrant CIN induction (this assumes induction of the tetOn constructs takes about 24 hours to become fully induced). The polar chromosome model (Po), which was induced using sequential chemical treatments and enrichment of about 50% of the population, was assigned a prior dataset with ≤ 2 time steps (1 generation).

Bulk DNA sequencing and analysis

Bulk sample and library preparation

Sample preparation, sorting, and bulk DNA library preparation were prepared in parallel with and in the same manner as single-cell DNA samples. 500 cells were sorted for each bulk DNA sample.

Bulk DNA sequencing and data pre-processing

Paired-end bulk DNA sequencing (2x150bp) was performed on an Illumina HiSeq to ~10x coverage and demultiplexed using Illumina bcl2fastq (v2.20). Sequencing reads were adaptor-trimmed, filtered, de-duplicated, and aligned to hg19 with the Illumina DRAGEN Bio-IT Platform using default settings.

Bulk copy number calling

Bulk DNA copy numbers were called in R using QDNAseq (v1.34.0) (5) and a bin size of 30 Kb. Segment copy numbers were called using bin copy numbers smoothed over 2 bins and Anscombe transformed (transformFun = "sqrt" in the 'segmentBins' function).

Quantification of CIN signatures

Putative CIN signatures were derived from bulk copy number profiles and reported in the same manner as for single cell DNA copy number profiles as described above.

Bulk RNA sequencing

Bulk RNA sample preparation

Cells were re-seeded in 6-well plates at 40% and grown to ~70-80% density over 18-24 hours prior to harvest. Media was removed from cells and immediately lysed with TRI Reagent (ThermoFisher Scientific, AM9738) to preserve RNA, which was stored at -80°C. Total RNA was isolated using Zymo Direct-zol RNA MiniPrep kit (Zymo Research, R2050) and the concentration and quality were assessed with a Qubit 2.0 fluorometer and Agilent TapeStation respectively.

Bulk RNA library preparation

RNA libraries were prepared using the NEBNext Ultra II RNA Illumina Library Prep Kit (New England Biolabs, E7775) and mRNAs were enriched with oligo-d(T) beads then fragmented for 15 minutes at 94°C prior to first and second strand cDNA synthesis. cDNAs were end-repaired, 3'-adenylated, ligated with universal Illumina adapters and unique index sequences, then enriched by PCR. Library quality was validated on an Agilent TapeStation and concentration was measured via Qubit 2.0 fluorometer and qPCR.

Bulk RNA sequencing and data pre-processing

Paired-end bulk RNA sequencing (2x150bp) was performed on an Illumina HiSeq sequencer and demultiplexed using Illumina bcl2fastq (v2.20). Demultiplexed reads were trimmed of universal Illumina index sequences with BBMap's bbduk (v38.95) (6). We aligned trimmed reads to reference sequence hg19 and generated fragment count matrices in R/Rstudio (v4.2.3/v1.2.5019) (1, 7) using Rsubread (v2.4.3) (8). Library size normalization and dispersion estimation was performed in edgeR (v.3.32.1) (9) while linear modeling and differential expression analysis was performed in limma (v3.46) (10, 11).

Supplemental References

1. R. C. Team, R: A language and environment for statistical computing. (2021).

2. M. L. Leung, *et al.*, Highly multiplexed targeted DNA sequencing from single nuclei. *Nat Protoc* **11**, 214--235 (2016).

3. M. Kendall, M. Boyd, C. Colijn, phyloTop: Tools for calculating and viewing topological properties of phylogenetic trees (2018).

4. K. Csillery, O. Francois, M. G. B. Blum, Abc: An R package for approximate Bayesian computation (ABC). *Methods Ecol Evol* (2012) https:/doi.org/10.1111/j.2041-210x.2011.00179.x.

5. I. Scheinin, *et al.*, DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res* **24**, 2022–2032 (2014).

6. B. Bushnell, BBMap (2022).

7. T. RStudio, RStudio: Integrated Development for R. (2020).

8. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* (2019) https:/doi.org/10.1093/nar/gkz114.

9. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

10. M. E. Ritchie, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47–e47 (2015).

11. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014)

12. S. A. Godinho et al., Oncogene-like induction of cellular invasion from centrosome amplification. Nature 510, 167–171 (2014).

13. S. L. Thompson, D. A. Compton, Chromosome missegregation in human cells arises through specific types of kinetochore–microtubule attachment errors. Proc. Natl. Acad. Sci. U.S.A. 108, 17974–17978 (2011).

14. G. Qian, A. Mahdi, Sensitivity analysis methods in the biomedical sciences. Math. Biosci. 323, 108306 (2020).

Supplemental Table 1 – Variables for calculating MDD

CIN Phenotype (ϕ_i)	Phenotype Penetrance (π)	Mis-segregation Magnitude (μ)	Resolution Rate (ρ)	Modal Chromosomes (θ)	Measured Phase	Assumption Support
Bridging	See Supplemental Table 2	1	0	46 (CAL51) or 47 (MCF10A)	Anaphase or Telophase	
Lagging	See Supplemental Table 2	1	0.9*	46 (CAL51) or 47 (MCF10A)	Anaphase or Telophase	(69)
Pseudobipolar	See Supplemental Table 2	0	0	46 (CAL51) or 47 (MCF10A)	Anaphase or Telophase	
Multipolar	See Supplemental Table 2	18*	0	46 (CAL51) or 47 (MCF10A)	Anaphase or Telophase	(27)
Polar	See Supplemental Table 2	7.8*	0	46 (CAL51) or 47 (MCF10A)	Metaphase	Supplemental Figure 1

Values for each variable required for approximating MDD from imaging experiments. Columns marked with asterisks represent variables requiring assumptions which are backed by data or previous reports.

Supplemental Table 2 – MDD by phenotype approximated by imaging

Group	Phenotype	Phenotype Penetrance (π)	± SE	Approximate MDD	± SE
Fixed Imaging					
CtrlC	Multipolar	0	0	0.09	0.09
	Polar	0.02	0.01	0.13	0.08
	Bridging	0.03	0.01	0.03	0.01
	Lagging	0	0	0	0
Br	Multipolar	0.01	0	0.24	0.06
	Polar	0.01	0.01	0.07	0.07
	Bridging	0.32	0.02	0.32	0.02
	Lagging	0.01	0.01	0	0
CtrlM	Multipolar	0.01	0	0.21	0.04
	Polar	0.02	0.01	0.14	0.05
	Bridging	0.03	0.01	0.03	0.01
	Lagging	0.01	0.01	0	0
Pb	Multipolar	0.05	0.02	0.9	0.31
	Polar	0.03	0.01	0.25	0.06
	Bridging	0.03	0.02	0.03	0.02
	Lagging	0.04	0.01	0	0
Mp	Multipolar	0.23	0.04	4.12	0.69
	Polar	0.05	0.02	0.35	0.13
	Bridging	0.03	0	0.03	0
	Lagging	0.11	0.03	0.01	0
Po	Multipolar	0.01	0.01	0.13	0.13
	Polar	0.99	0	7.56	0.04
	Bridging	0.07	0.04	0.07	0.04
	Lagging	0.06	0.03	0.01	0
Time Lapse Imaging					
CtrlC	Multipolar	0.08	0.05	1.51	0.84
	Polar	0.02	0.02	0.19	0.19
	Bridging	0.08	0.03	0.08	0.03
	Lagging	0.06	0.06	0.01	0.01
Br	Multipolar	0.04	0.02	0.66	0.37
	Polar	0.06	0.06	0.49	0.49
	Bridging	0.4	0.02	0.4	0.02
	Lagging	0.07	0.03	0.01	0
CtrlM	Multipolar	0.04	0.04	0.7	0.7
	Polar	0	0	0	0
	Bridging	0.01	0.01	0.01	0.01
	Lagging	0.07	0.04	0.01	0
Pb	Multipolar	0.14	0.03	2.55	0.53
	Polar	0.03	0.02	0.22	0.12
	Bridging	0.05	0.02	0.05	0.02
	Lagging	0.12	0.02	0.01	0
Mp	Multipolar	0.24	0.05	4.28	0.81
	Polar	0.01	0.01	0.1	0.1
	Bridging	0.08	0.08	0.08	0.08
	Lagging	0.2	0.04	0.02	0
Po	Multipolar	0.02	0.02	0.42	0.42
	Polar	0.98	0.02	7.51	0.13
	Bridging	0.08	0.02	0.08	0.02
	Lagging	0.18	0.02	0.02	0

Approximated MDD for each CIN phenotype in each model. MDD was calculated using $MDD = \left(\frac{Errors\ per\ Defect \times Defect\ Rate \times Penetrance}{Modal\ Chromosomes}\right) \times 46$ and assumptions of the number of chromosomes mis-segregated for each defect (see Materials and Methods). The fraction of cells with polar chromosomes represents metaphase cells wherein they are most readily detectable. All other CIN phenotypes are taken from anaphase or telophase cells.

Supplemental Table 3 – CIN70 and HET70 genes

CIN70		HET70	
TPX2	MSH6	AHCYL1	LPP
PRC1	EZH2	AKT3	MED8
FOXM1	CTPS1	ANO10	MMP2
CDK1	DKC1	ANTXR1	MUL1
TGIF2	OIP5	ATP6V0E1	MYO10
MCM2	CDCA8	ATXN1	NAGK
H2AZ1	PTTG1	B4GALT2	NR1D2
TOP2A	CEP55	BASP1	NRIP3
PCNA	H2AX	BHLHE40	P4HA2
UBE2C	CMAS	BLVRA	PKIG
MELK	NCAPH	CALU	PLOD2
TRIP13	MCM10	CAP1	PMP22
NCAPD2	LSM4	CAST	POFUT2
MCM7	NCAPG2	CAV1	POMGNT1
RNASEH2A	ASF1B	CLIC4	PRKAR2A
RAD51AP1	ZWINT	CTSL	MOK
KIF20A	PBK	CYB5R3	RHOC
CDC45	ZWILCH	ELOVL1	RRAGC
MAD2L1	CDCA3	EMP3	SEC22B
ESPL1	ECT2	FKBP14	SERPINB8
CCNB2	CDC6	FN1	SPAG9
FEN1	UNG	FST	SQSTM1
TTK	MTCH2	GNA12	TIMP2
CCT5	RAD21	GOLT1B	EMC3
RFC4	ACTL6A	HECTD3	TRIM16
ATAD2	GPI	HEG1	TRIO
CKAP5	SRSF2	HOMER3	TUBB2A
NUP205	HDGF	IGFBP3	VEGFC
CDC20	NXT1	IL6ST	VIM
CKS2	NEK2	ITCH	WASL
RRM2	DHCR7	P3H1	YIPF5
ELAVL1	AURKA	P3H2	YKT6
CCNB1	NDUFAB1	LEPROT	ZBTB38
RRM1	NEMP1	LGALS1	ZCCHC24
AURKB	KIF4A	LIMA1	ZMPSTE24

Genes included in the CIN70 and HET70 gene expression panels. Genes are sorted in their original, published order, CIN70 by correlational rank and HET70 in alphabetical order.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure 1 – Additional imaging data

(A) Number of anticentromere antibody (ACA) foci at or behind the spindle pole in late prometaphase (i.e., discernable metaphase plates) MCF10A cells treated under the Po condition but prior to anaphase induction. **(B)** Differences in observed frequencies of mitotic defects between fixed immunofluorescence and time lapse fluorescence imaging in metaphase or anaphase/telophase. Significance values beneath data are from two-tailed, one-sample Student's *t*-tests where H_o: $\mu = 0$. Colors of individual points indicate the biological replicate. Bars and error bars indicate mean and standard error. Significance values above data are from a two-tailed, two-sample Student's *t*-tests. Both are corrected for multiple comparisons via the Benjamini-Hochberg method.

Supplemental Figure 2 – Sensitivity analysis for approximating MDD via imaging

Approximated MDD values for fixed and time lapse imaging of each phenotypic model and control (columns) while changing the approximate magnitude (μ) of each mitotic defect (rows) by ± 50% (blue and red bars). Sensitivity index (SI) of each mitotic index is calculated and shown for each phenotypic model and control.

Supplemental Figure 3 – Gene expression in inducible phenotypic models of CIN

(A) Expression levels (log2 counts per million (CPM)) of *TERF2* and *PLK4* in CAL51-TERF2-DN-tetOn and MCF10A-PLK4-WT-tetOn cells respectively. Two-tailed, two-sample Student's *t*-tests are shown above data. $N \geq 3$ biological replicates. **(B)** Volcano plots of differential gene expression in each CIN-induced model compared to its uninduced control and a pooled analysis of all CIN-induced groups compared to all non-induced groups. Dashed lines indicate log2 fold-change thresholds of -1/1 and unadjusted P value of 0.05.

Supplemental Figure 4 – DNA content analysis and gating for pre-scDNAseq FACS

Cell cycle profiles (measured by flow cytometric analysis of DAPI intensity) for each replicate of each phenotypic model of CIN. Dashed lines indicate gating strategy for FACS of cells for scDNAseq and were determined by 50% and 150% of the intensity of the G1 peak. Red ticks indicate the DAPI intensity of cells sorted for scDNAseq. Note the broader G1 peaks of CIN-induced models Pb, Mp, and Po, indicative of extensive aneuploidy.

Supplemental Figure 5 – Breakpoint analysis in scDNAseq data

Breakpoint analysis of discordant and concordant telomere-proximal copy number alterations. Copy numbers are normalized relative to those of the modal karyotype of each phenotypic CIN model to uncover subclonal or relatively recent alterations. Only segments with copy number alteration consensus across the 4 telomere-proximal genomic bins (10 Mb), but which terminate at or before the centromere, are considered. Discordant subclonal telomere-proximal alterations are defined as those that are not matched by an alteration on the opposite chromosome arm. Concordant alterations are those that are matched on the opposite chromosome arm. The quantification shows the average number of each alteration per cell for each model.

Supplemental Figure 6 – Putative CIN signatures in inducible phenotypic models of CIN at single cell resolution

(A) Heatmap of all normalized putative CIN signature measurements in single cells showing with rows clustered by similarity. Colorized annotation on the left indicates phenotypic CIN model and shade indicates biological replicate. **(B)** Normalized putative CIN signature measurements in single cells grouped by model. Results from two-tailed, two-sample Student's *t*-tests are shown above data using all cells across 3 biological replicates and are corrected for multiple comparisons via the Benjamini-Hochberg method. Shape and shade of data points indicate biological replicate.

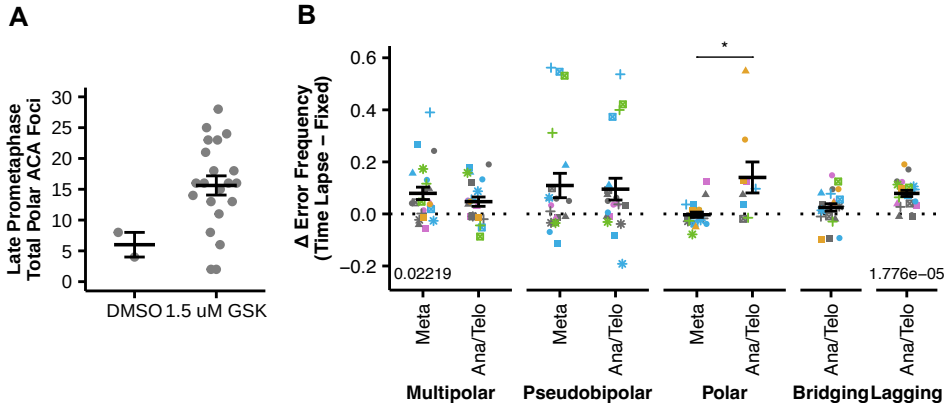
Supplemental Figure 7 – Extended pairwise correlations of CIN measurements

Pairwise correlations between each CIN measurement. Colors of data points indicate the phenotypic model of CIN. Pearson correlation coefficients and p-values of regression are shown.

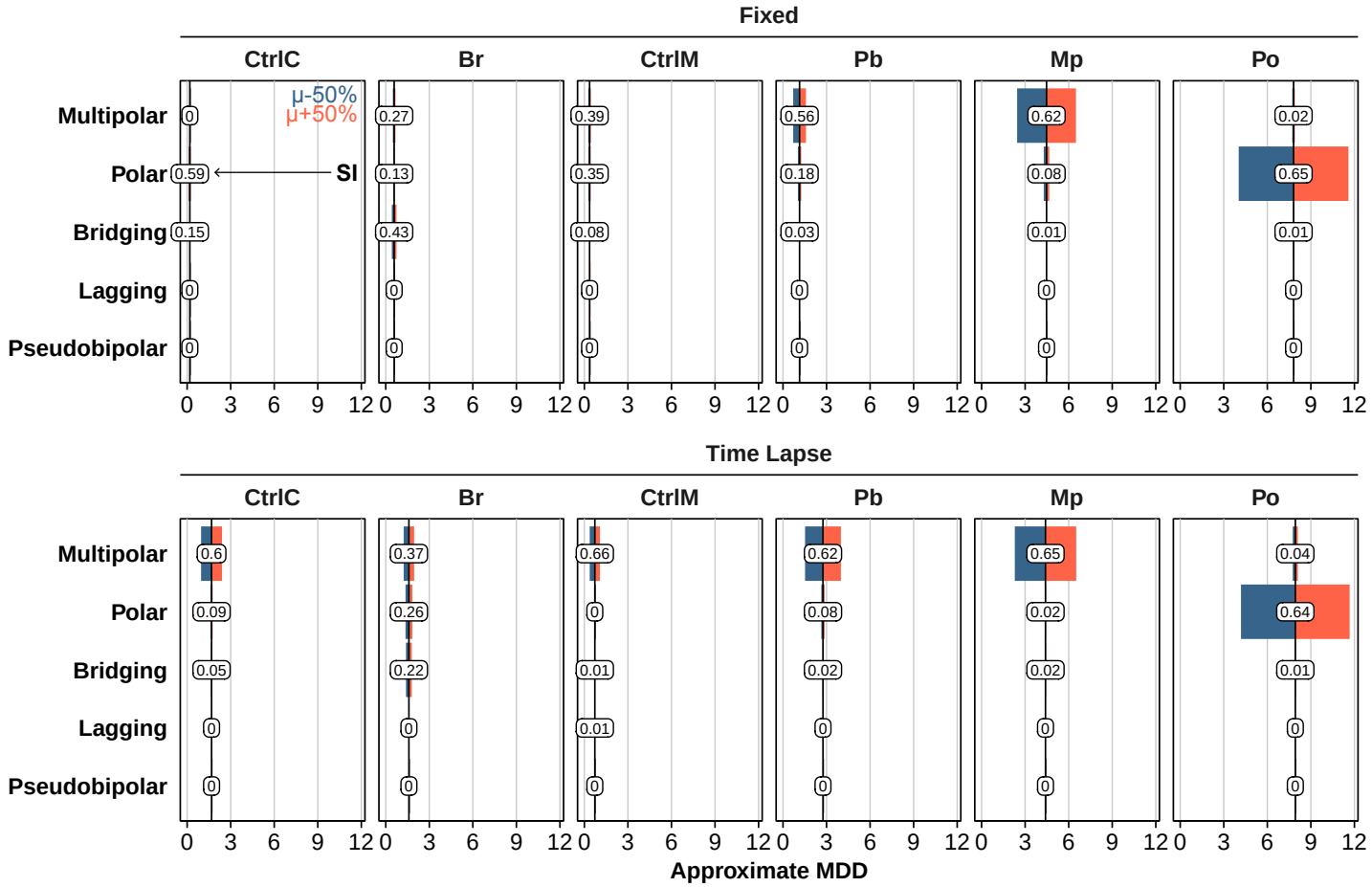
Supplemental Figure 8 – Sensitivity power analysis of CIN measures

Sensitivity power analysis of critical effect sizes (Hedge's *g*) of CIN measures for each phenotypic model (compared to its respective control). Shaded grey boxes indicate the critical effect size *g* at power (1-β) levels 0.2, 0.4, 0.6, and 0.8 (darkest to lightest). Effect sizes were calculated using the means for each available biological replicate (rather than individual cells or genes for some measures, for instance) and do not account for multiple comparisons correction. Asterisks on the right indicate whether the comparison is statistically significant at α=0.05 with Benjamini-Hochberg correction for multiple comparisons. Individual data points are the observed standard effect size (Hedge's *g*) ± the 95% confidence interval. Note that the Hedge's *g* approximates the effect size in units of pooled standard deviation (where 1 $g = 1\ SD_{pooled}$). Also note that statistical comparisons for scDNAseq and bRNAseq CIN signatures were calculated differently (using all cells and all genes, respectively, rather than mean values of biological replicates) in their respective sections.

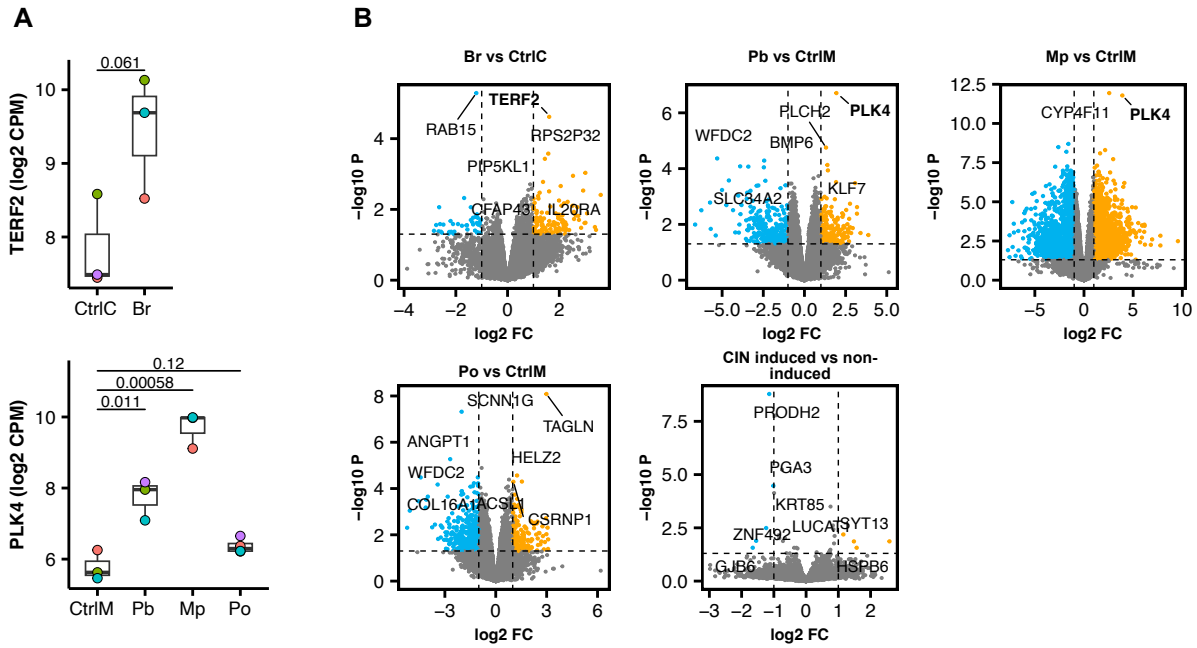
Supplemental Figure 1



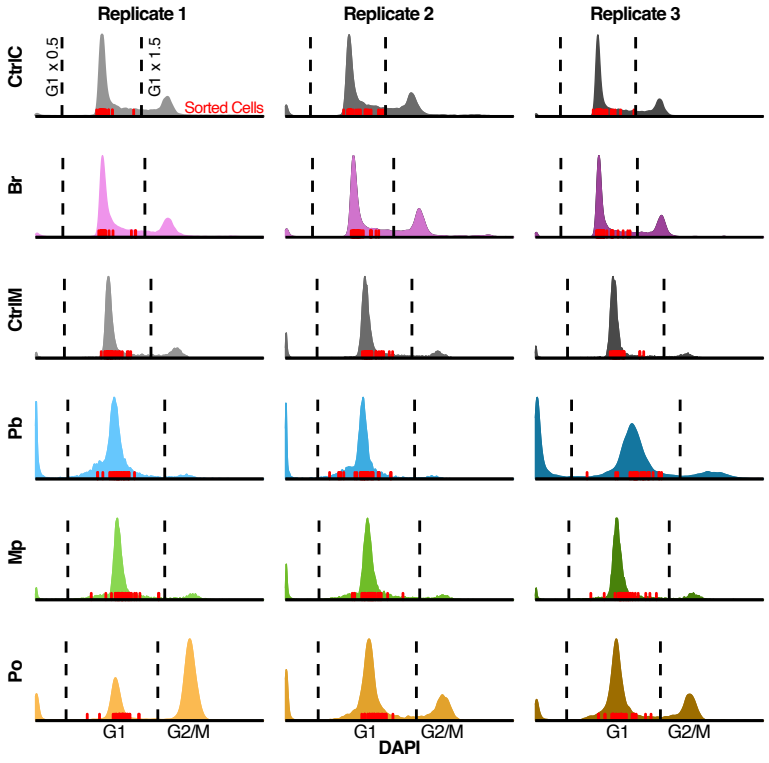
Supplemental Figure 2



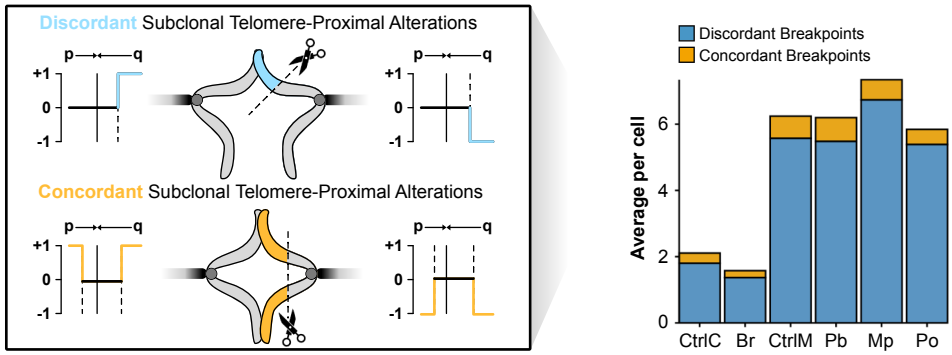
Supplemental Figure 3



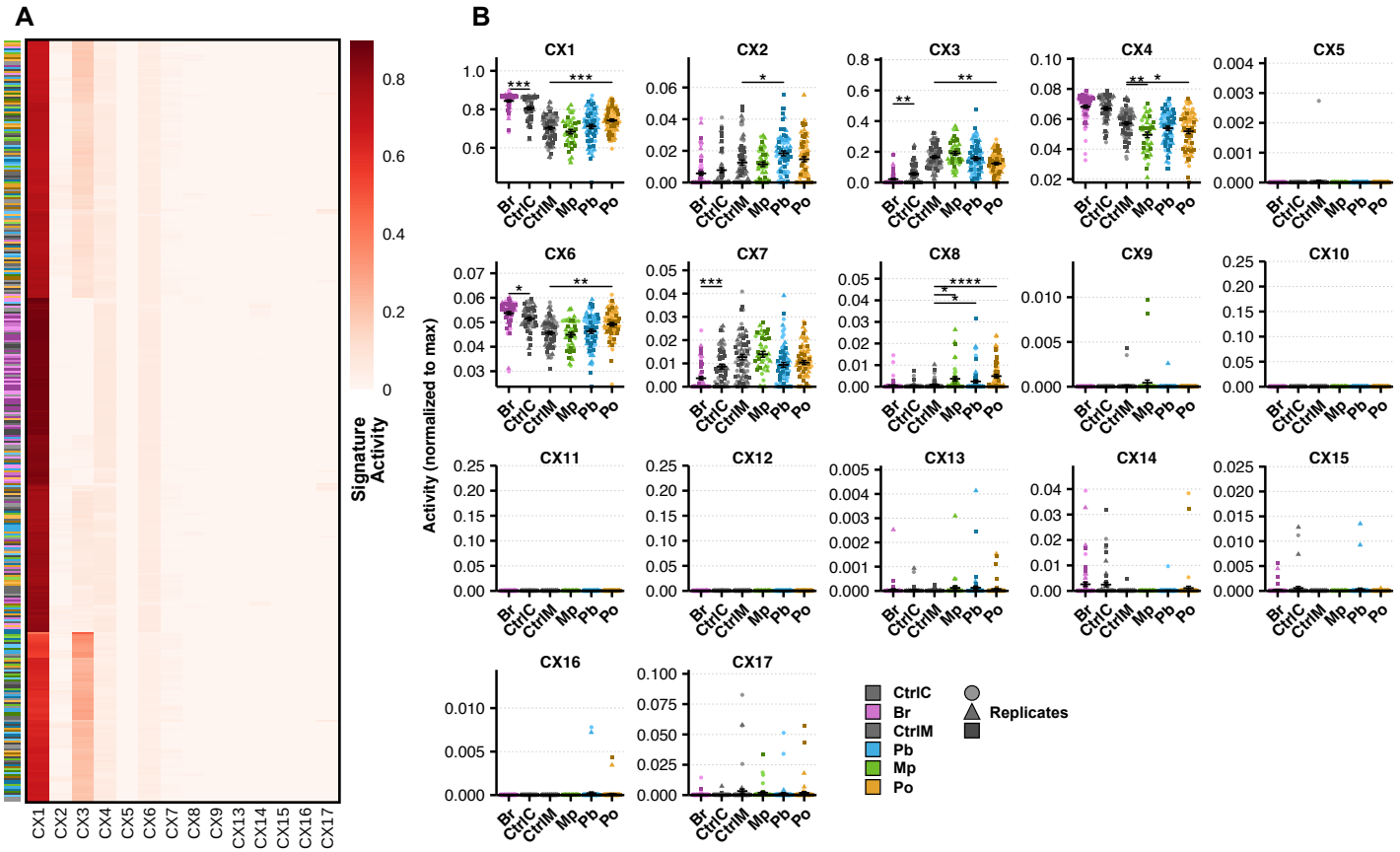
Supplemental Figure 4



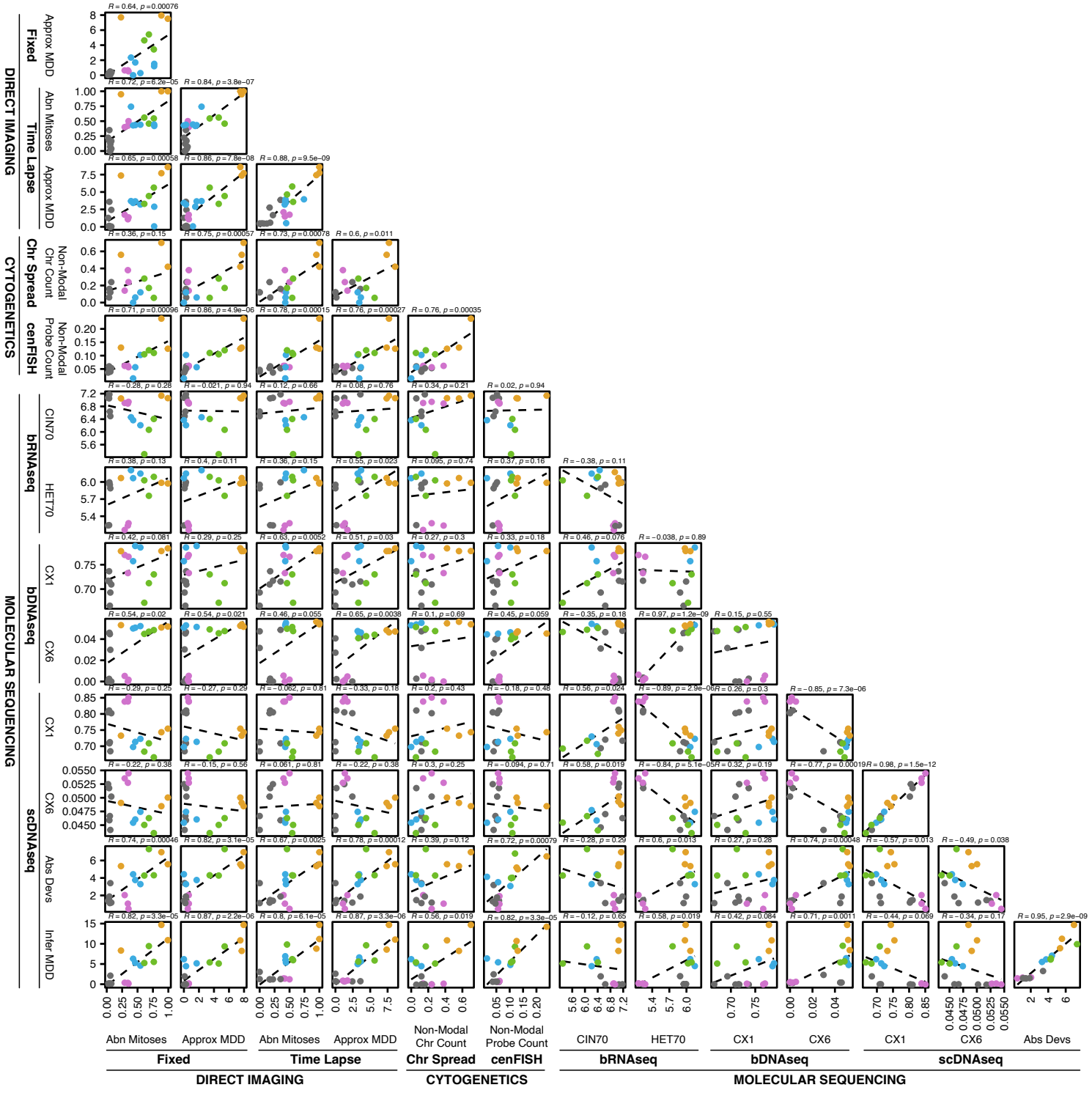
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 8