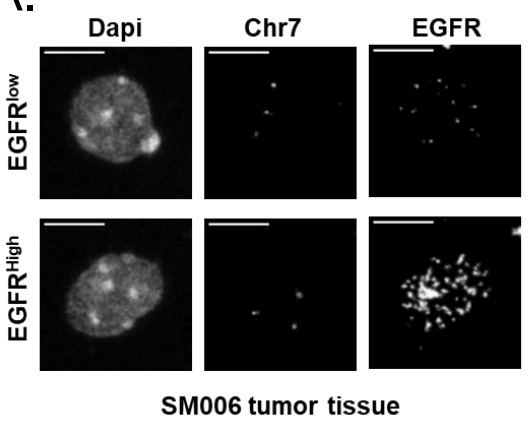
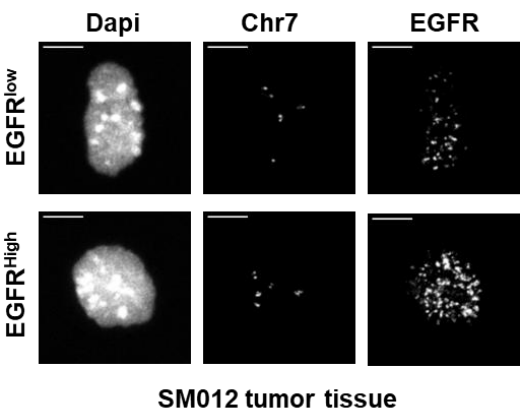
Supplementary Figure 1.

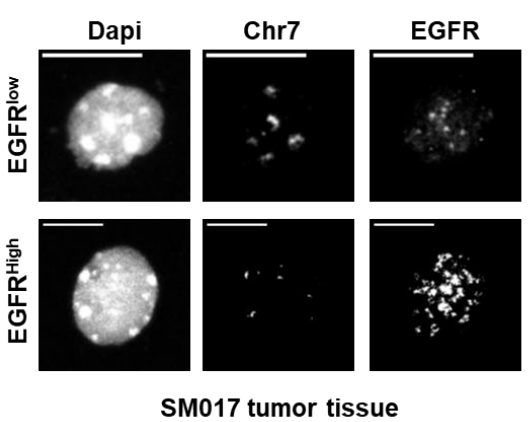
A.



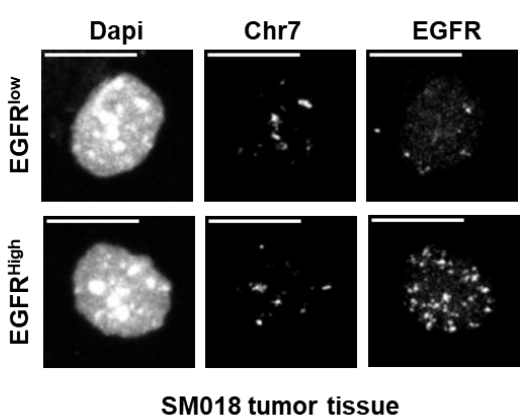
SM006 tumor tissue



SM012 tumor tissue

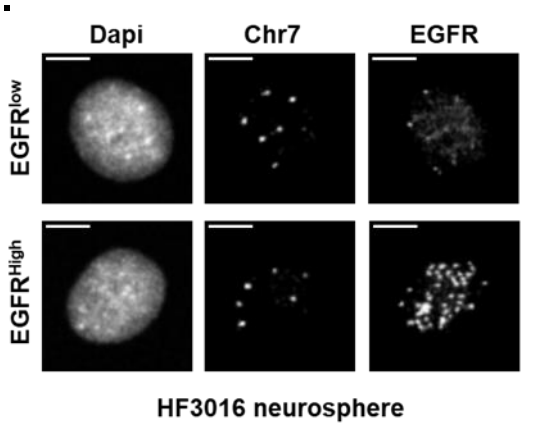


SM017 tumor tissue

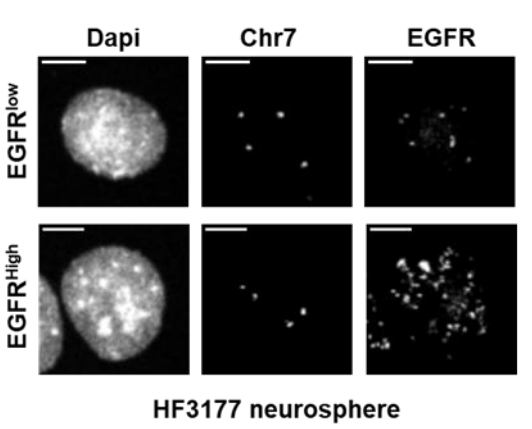


SM018 tumor tissue

B.



HF3016 neurosphere

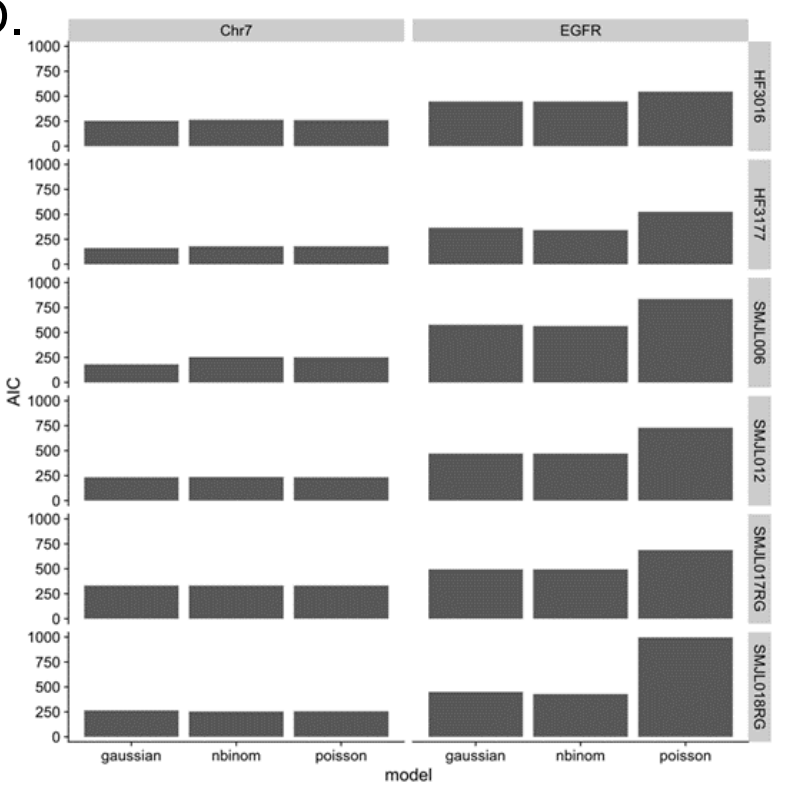


HF3177 neurosphere

C.

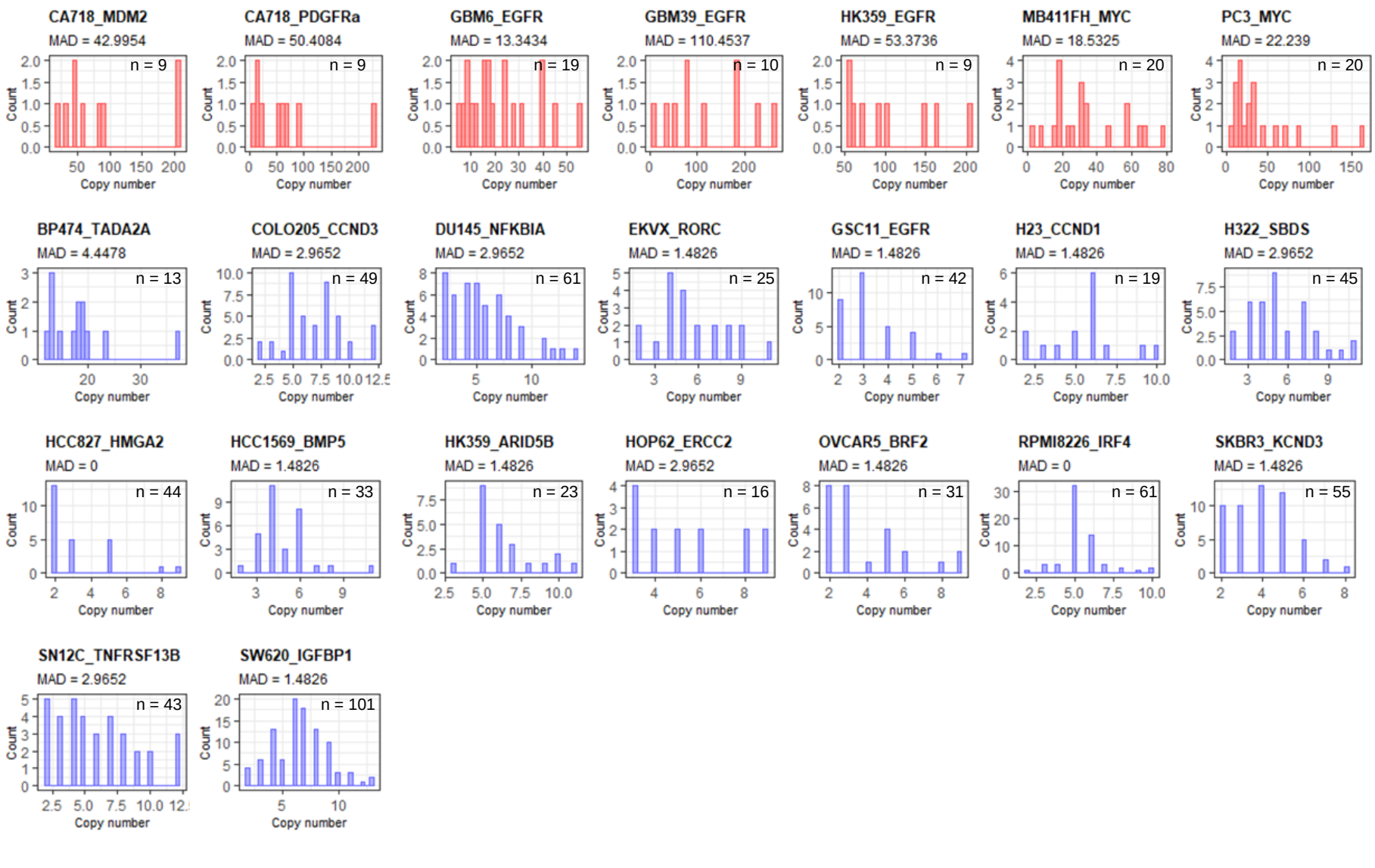
Cell line	Target	Relative fitness of negative binomial model	Relative fitness of Poisson model
HF3016	Chr7	0.01	0.02
HF3177	Chr7	0.00	0.00
SMJL006	Chr7	0.00	0.00
SMJL012	Chr7	0.23	0.63
SMJL017	Chr7	2.23	1.39
SMJL018	Chr7	119.87	57.72
HF3016	EGFR	1.00	0.00
HF3177	EGFR	87556.40	0.00
SMJL006	EGFR	1405.11	0.00
SMJL012	EGFR	0.74	0.00
SMJL017	EGFR	0.59	0.00
SMJL018	EGFR	22325.26	0.00

D.



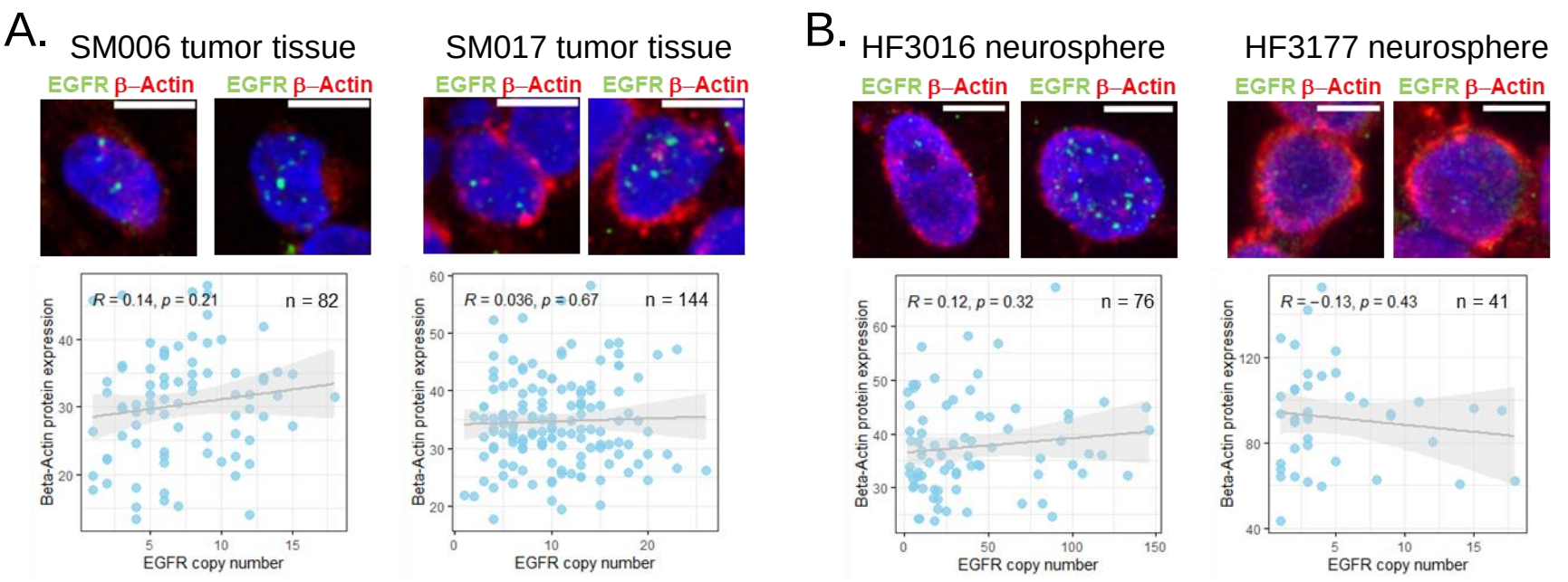
Supplementary Fig. 1 | *EGFR* copy number variation on tumor tissues and neurospheres . A. Grayscale images of *EGFR*/Chr7 FISH analysis on four GBM tumor tissues show *EGFR* copy number variation. **B.** Grayscale images of *EGFR*/Chr7 FISH analysis on two neurospheres show *EGFR* copy number variation. Scale bar, 10 μm. **C.** Relative fitness of non-Gaussian distributions compared to fitting data with Gaussian distribution. **D.** Akaike information criterion (AIC) values of each model. Lower values represent better fitness to a given probability distribution.

Supplementary Figure 2.



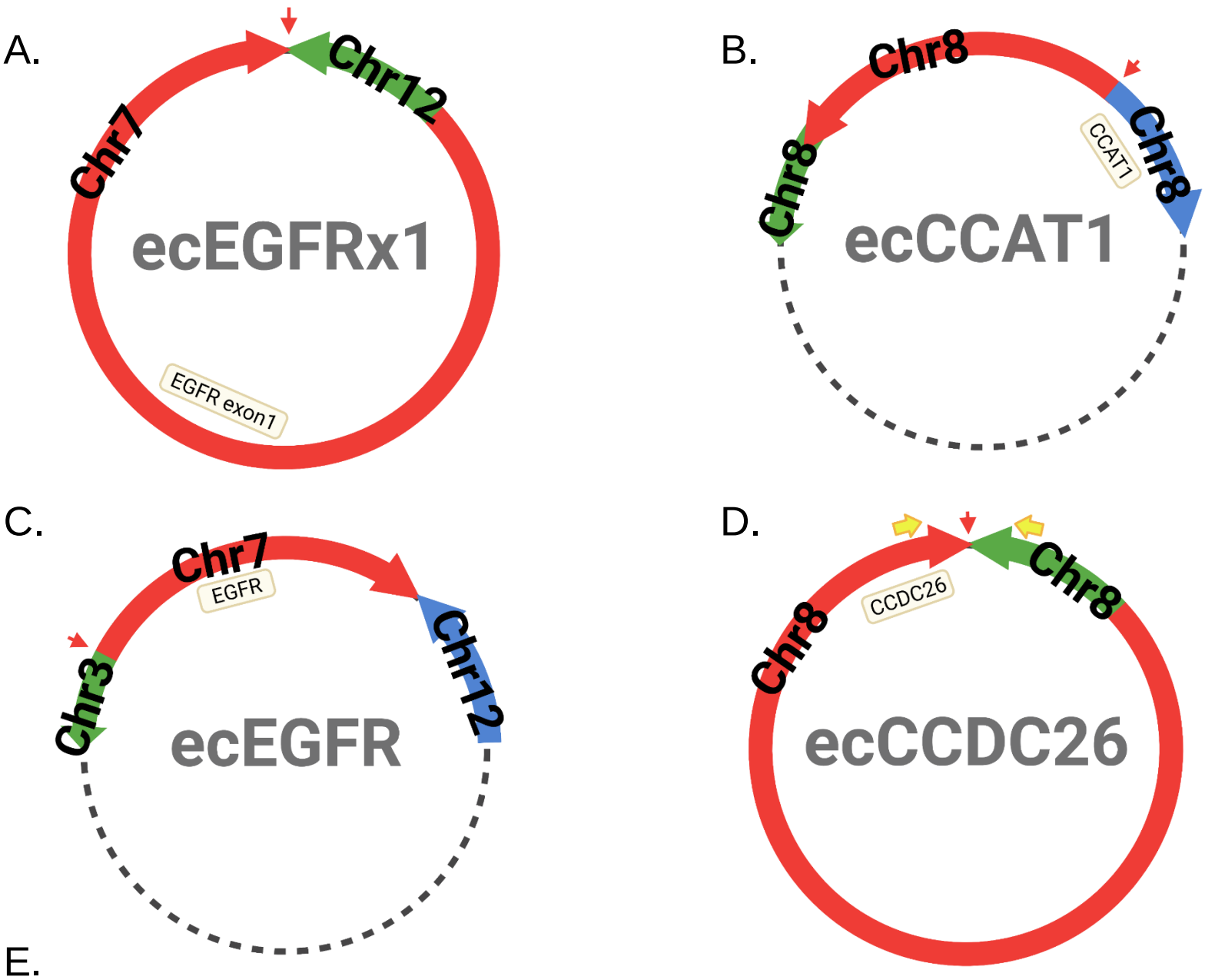
Supplementary Fig. 2 | Comparison of copy number distribution between ecDNA genes and linearly amplified genes. Individual copy number distribution on various types of cancer cell lines. Circularly amplified genes are considered as ecDNA gene and indicated by red. Linearly amplified genes were indicated by blue. The MADs are indicated at the top of each histogram.

Supplementary Figure 3.



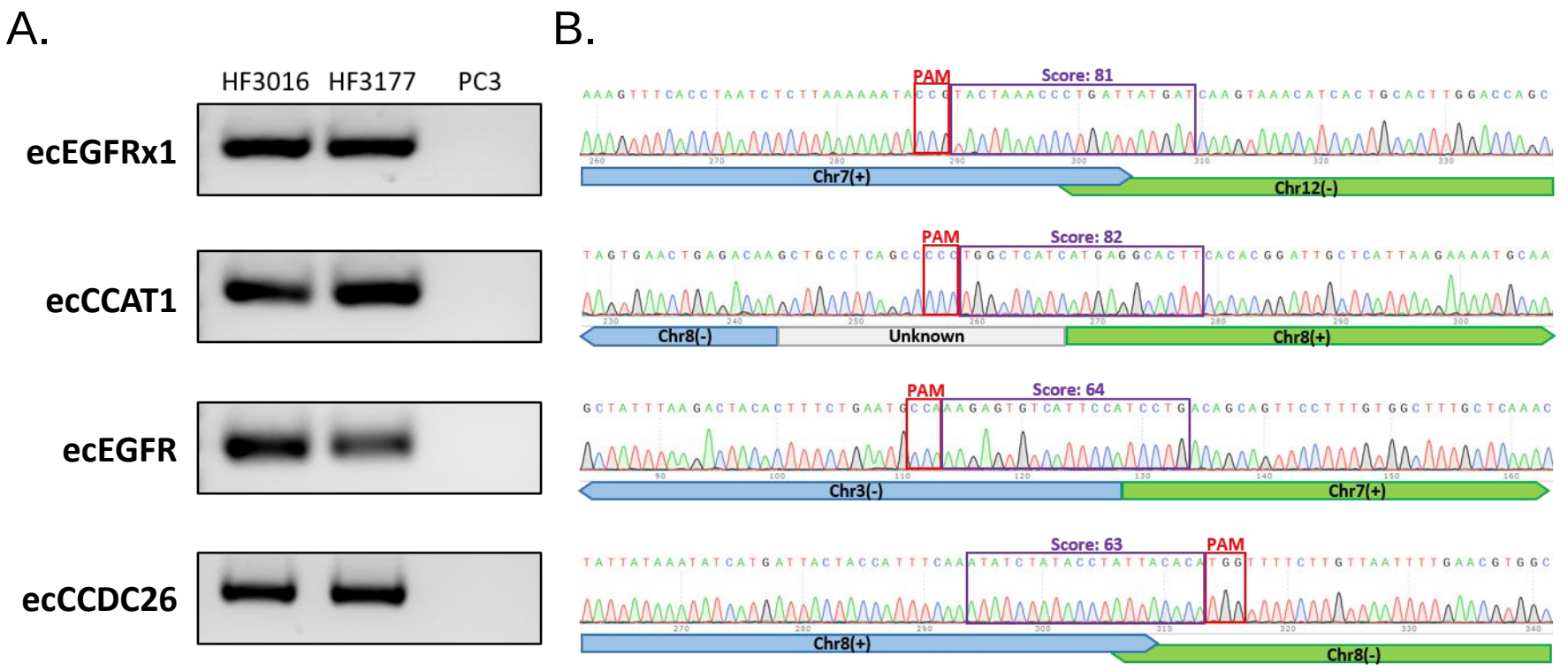
Supplementary Fig. 3 | ImmunoFISH of *EGFR* DNA FISH and β-Actin immunostaining. A-B. ImmunoFISH experiment on two GBM tumor tissues (A) and two neurosphere lines (B). Scale bar, 10 μm. Green signal indicates *EGFR* FISH signal. Red signal indicates β-Actin protein signal. Correlation between copy number of *EGFR* (number of *EGFR* DNA FISH signal foci) and β-Actin protein expression (quantified based on signal intensity) per cell and *p* values were determined by Pearson's correlation test (lower panel). β-Actin protein signals that appears to be derived from the nucleus is in fact cytoplasmic and on the cell surface, but appears nuclear as two-dimensional images were obtained from a three-dimensional cell image.

Supplementary Figure 4.



Supplementary Fig. 4 | EcDNA structures. A-D. ecDNA structures anticipated by AmpliconArchitect. Each ecDNA is named based on its cargo gene (A. ecEGFRx1 (exon 1); B. ecCCAT1; C. ecEGFR; D. ecCCDC26). **E.** Circular amplicons were predicted through analysis of whole-genome sequencing data using AmpliconArchitect. This showed that different ecDNAs were derived from the same genomic region. Unique breakpoint regions, which can be discretely labeled by ecTag, are shown labeled by a red circled number.

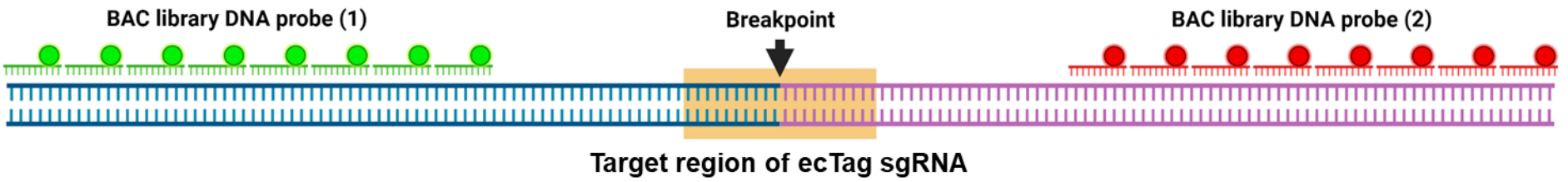
Supplementary Figure 5.



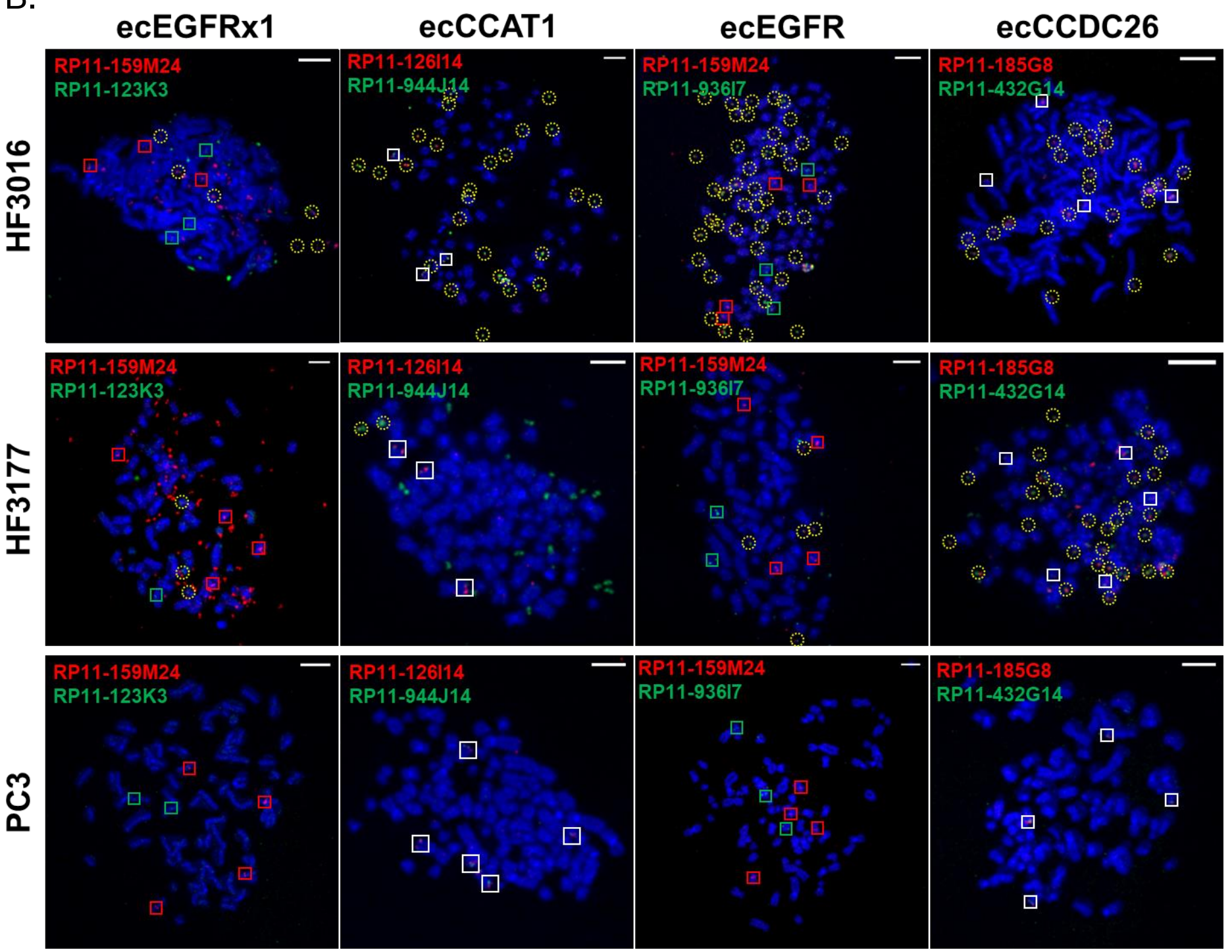
Supplementary Fig. 5 | Validation of breakpoint junction sequences. A. Gel-images of BP-PCR across breakpoint junctions. **B.** Chromatograms of Sanger sequencing results for each breakpoint. The target specificity score was determined by CRISPOR.

Supplementary Figure 6.

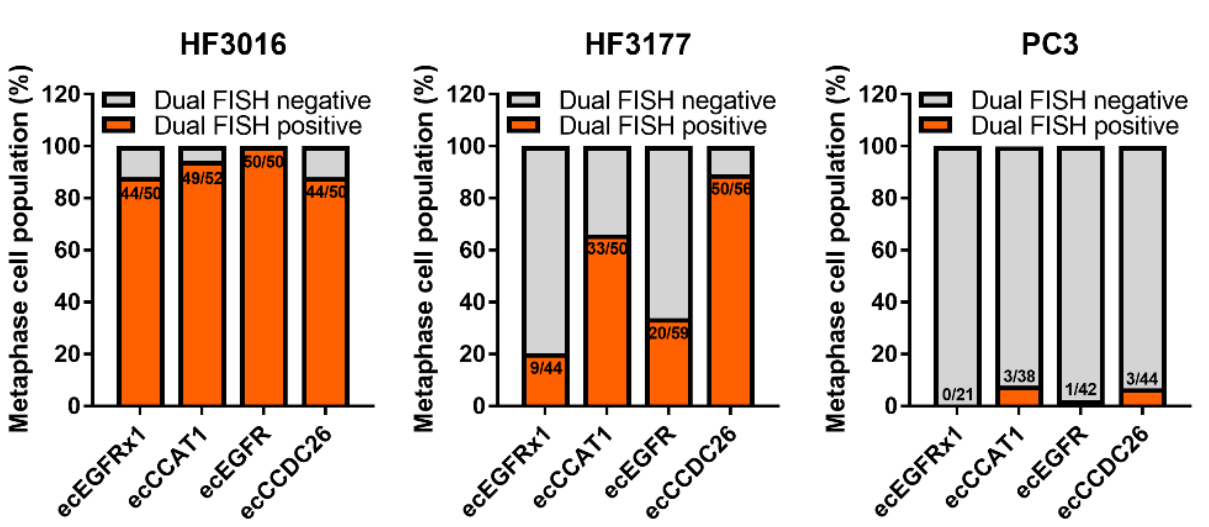
A.



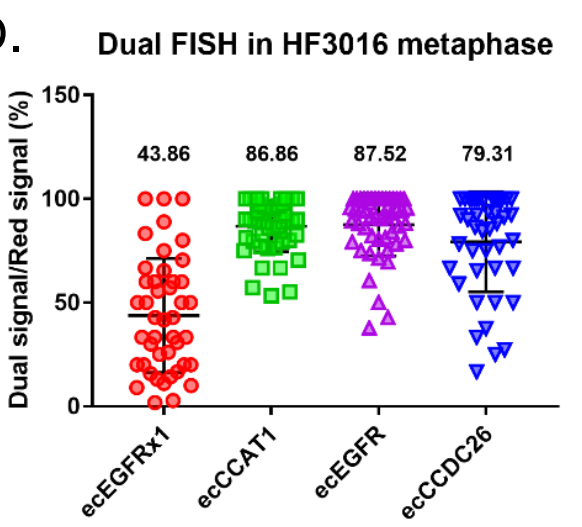
B.



C.

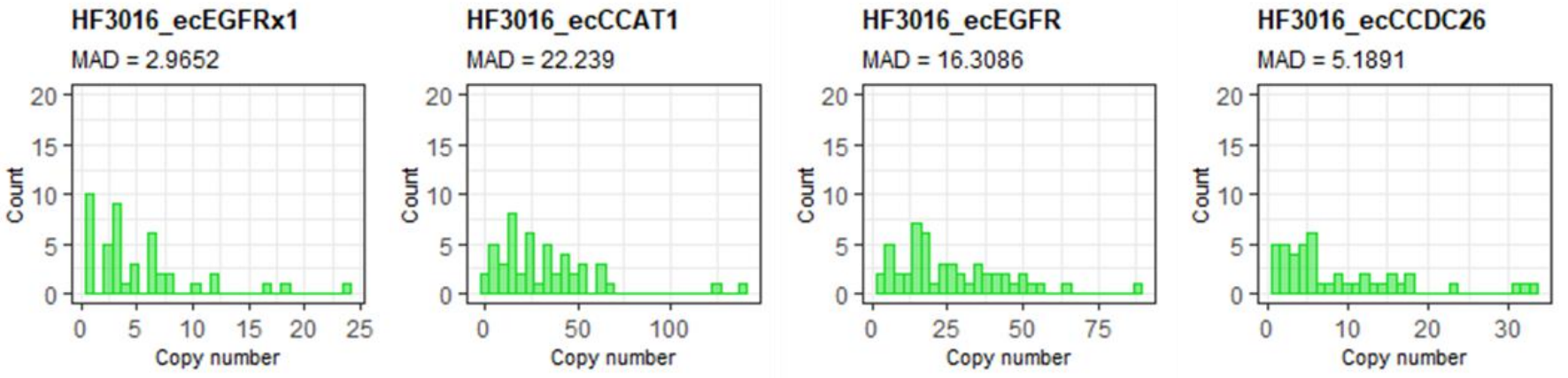


D.

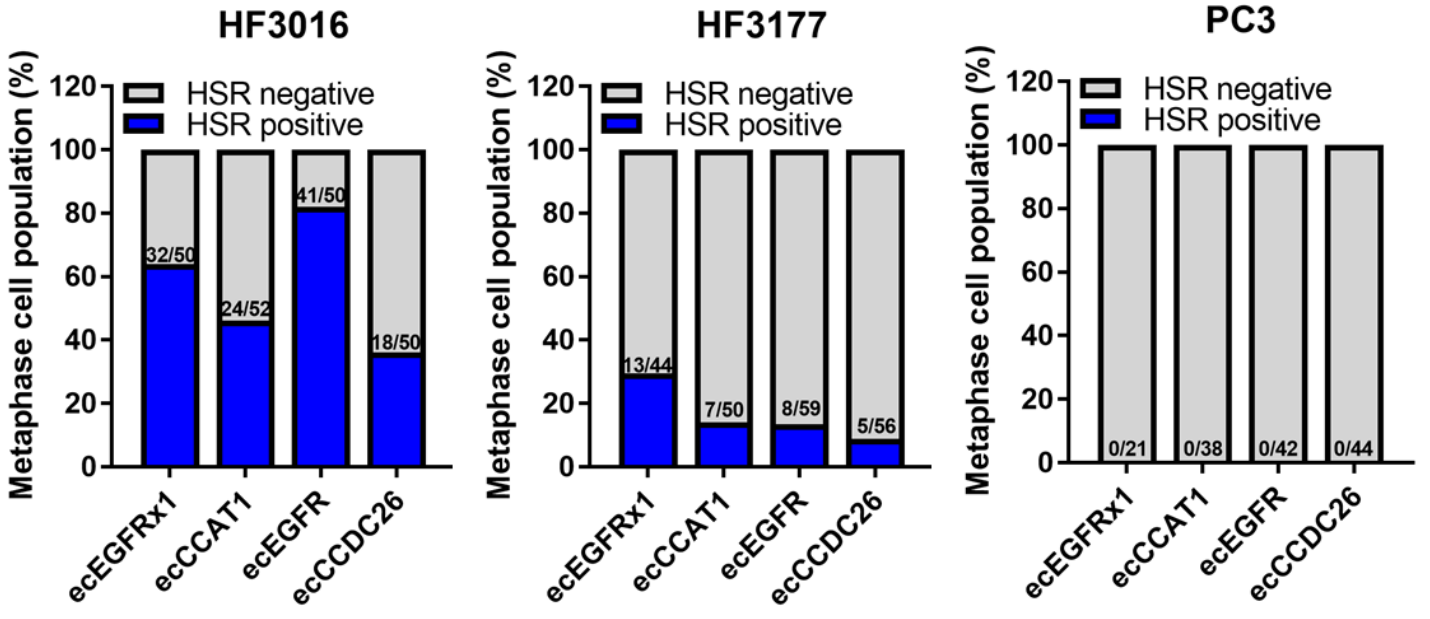


Supplementary Figure 6. -continued

E.

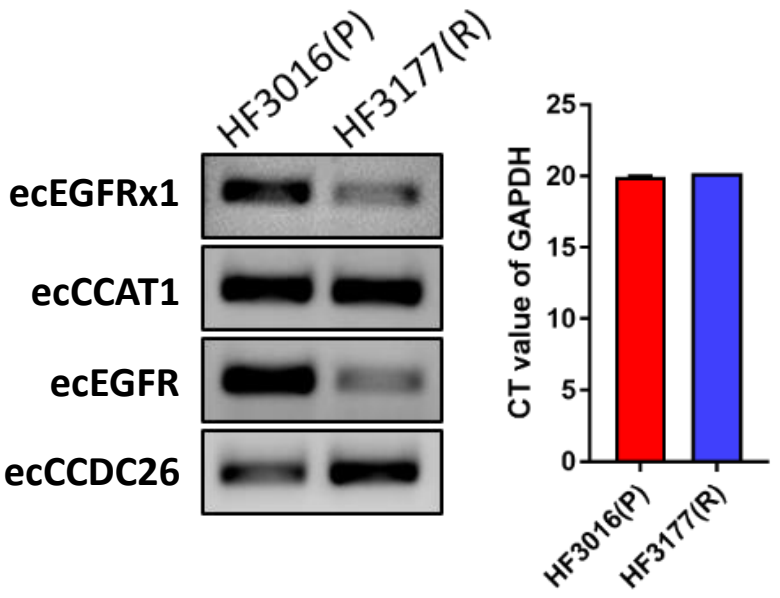


F.



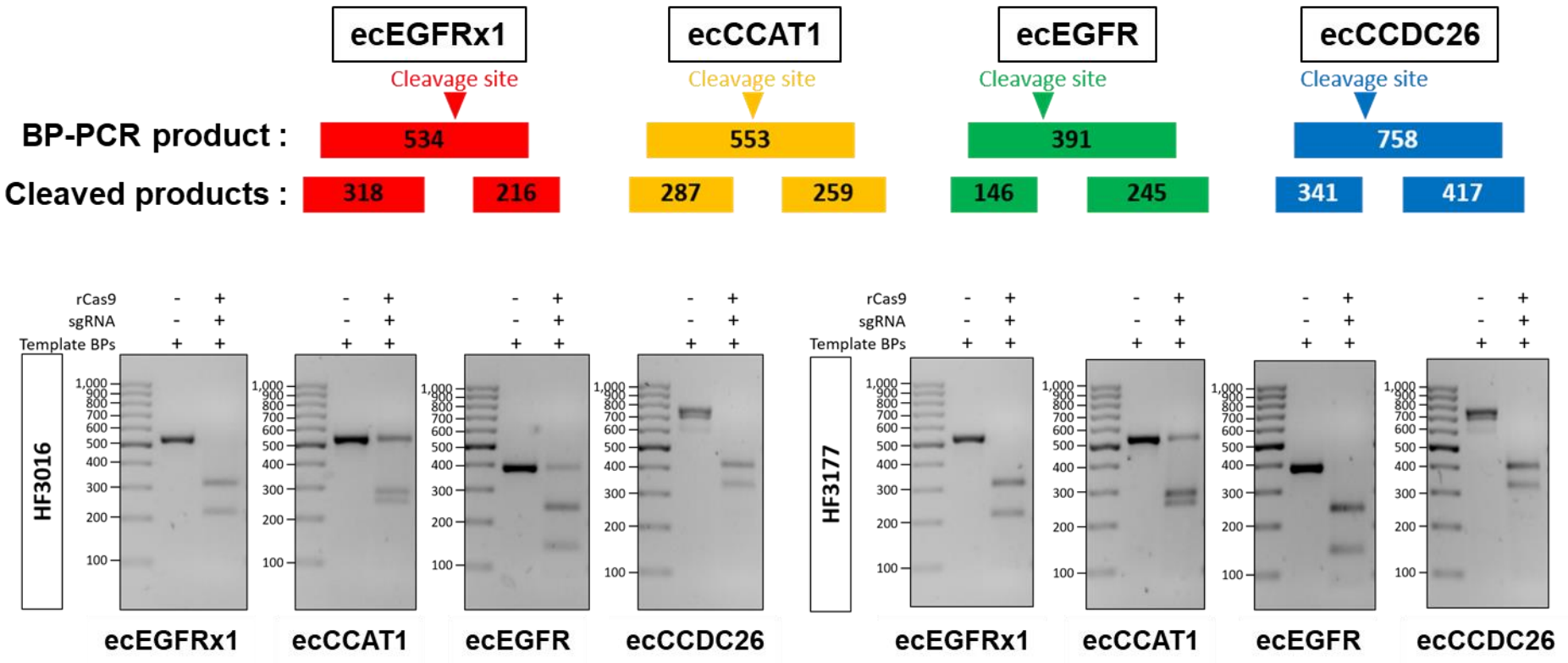
Supplementary Fig. 6 | Extrachromosomal breakpoint validation using Dual-FISH method. A. Schematic illustration of Dual-FISH method. **B.** Representative images of metaphase cells labeled with two BAC probes. Yellow dotted circle = ecDNA; Colored square = BAC probe signals from chromosomes. Scale bar, 10 μm. **C.** Dual-FISH signal-positive cell population in three cell lines showed that the breakpoints of these particular ecDNAs are specific to two neurosphere lines, but only detectable as background noise in PC3. The numbers shown are the count of Dual-FISH signal positive cells divided by the total number of cells. **D.** Proportion of dual-color signals out of red-color signals. The average proportion of dual-color signals in each target is indicated in the chart. The error bars represent S.D. Individual objective indicates a single metaphase cell. **E.** Distribution of copy-number of each target breakpoint labeled with both probes in HF3016. **F.** HSR-positive cell population in three cell lines.

Supplementary Figure 7.



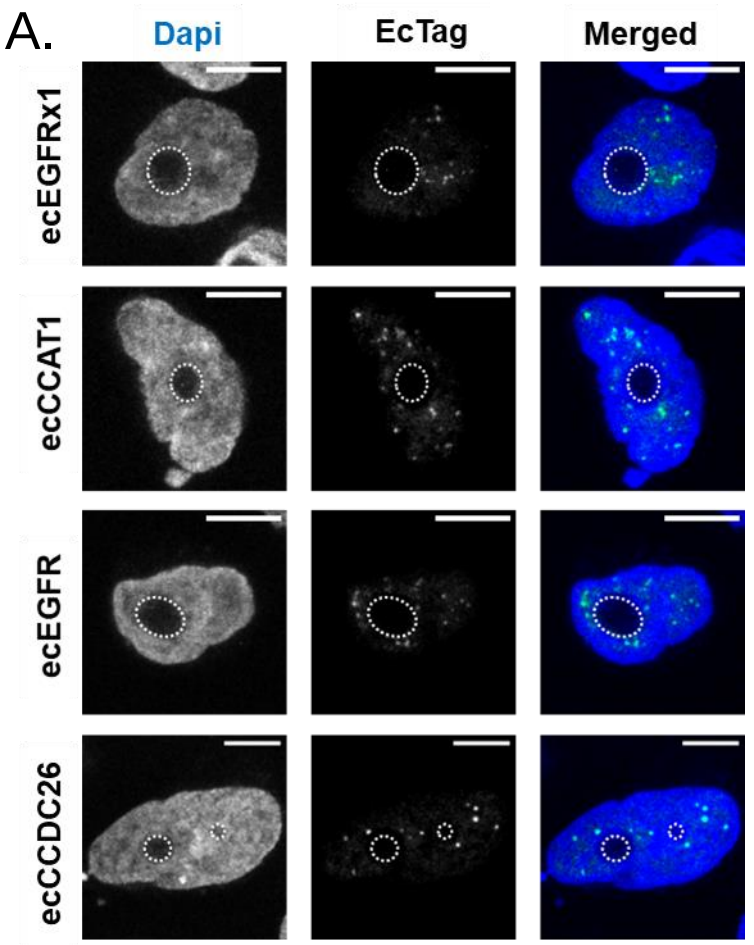
Supplementary Fig. 7 | Comparison of copy number of ecDNAs between primary and recurrent neurosphere. Gel-images of comparative BP-PCR performed on HF3016 and HF3177 (left panel). Input genomic DNAs were quantified by quantitative PCR (qPCR) normalized to GAPDH (right panel).

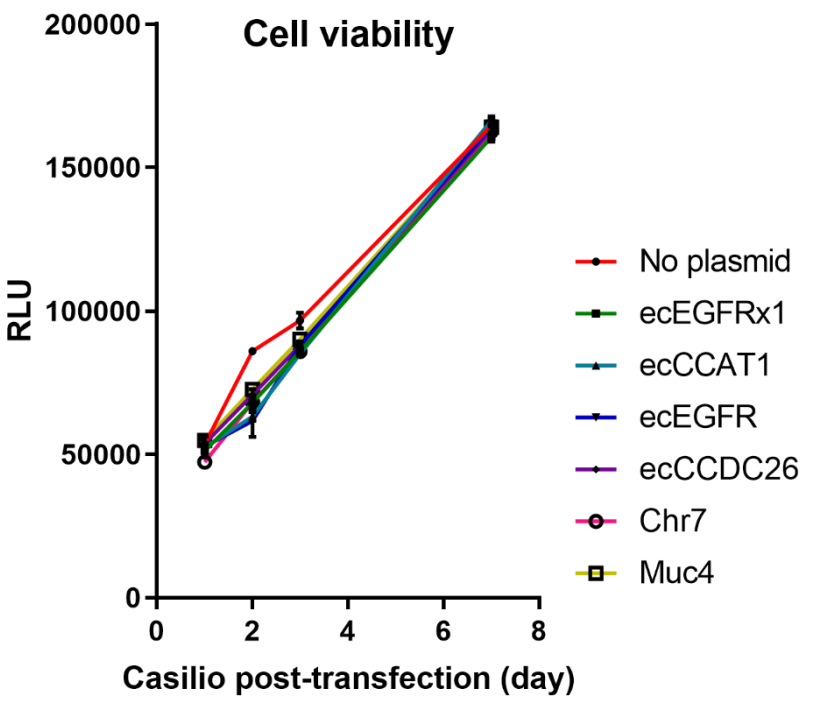
Supplementary Figure 8.

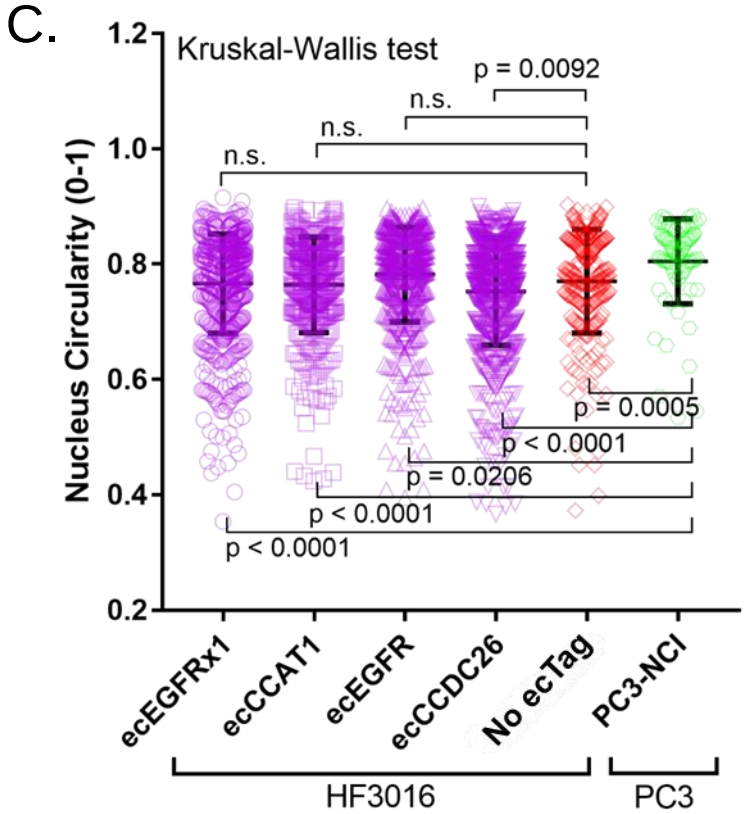


Supplementary Fig. 8 | Specificity test of breakpoint-targeting sgRNAs. Schematic illustration of the workflows of the specificity test (upper panel). Gel-images of BP-PCR amplicons sufficiently targeted by sgRNAs (lower panel).

Supplementary Figure 9.

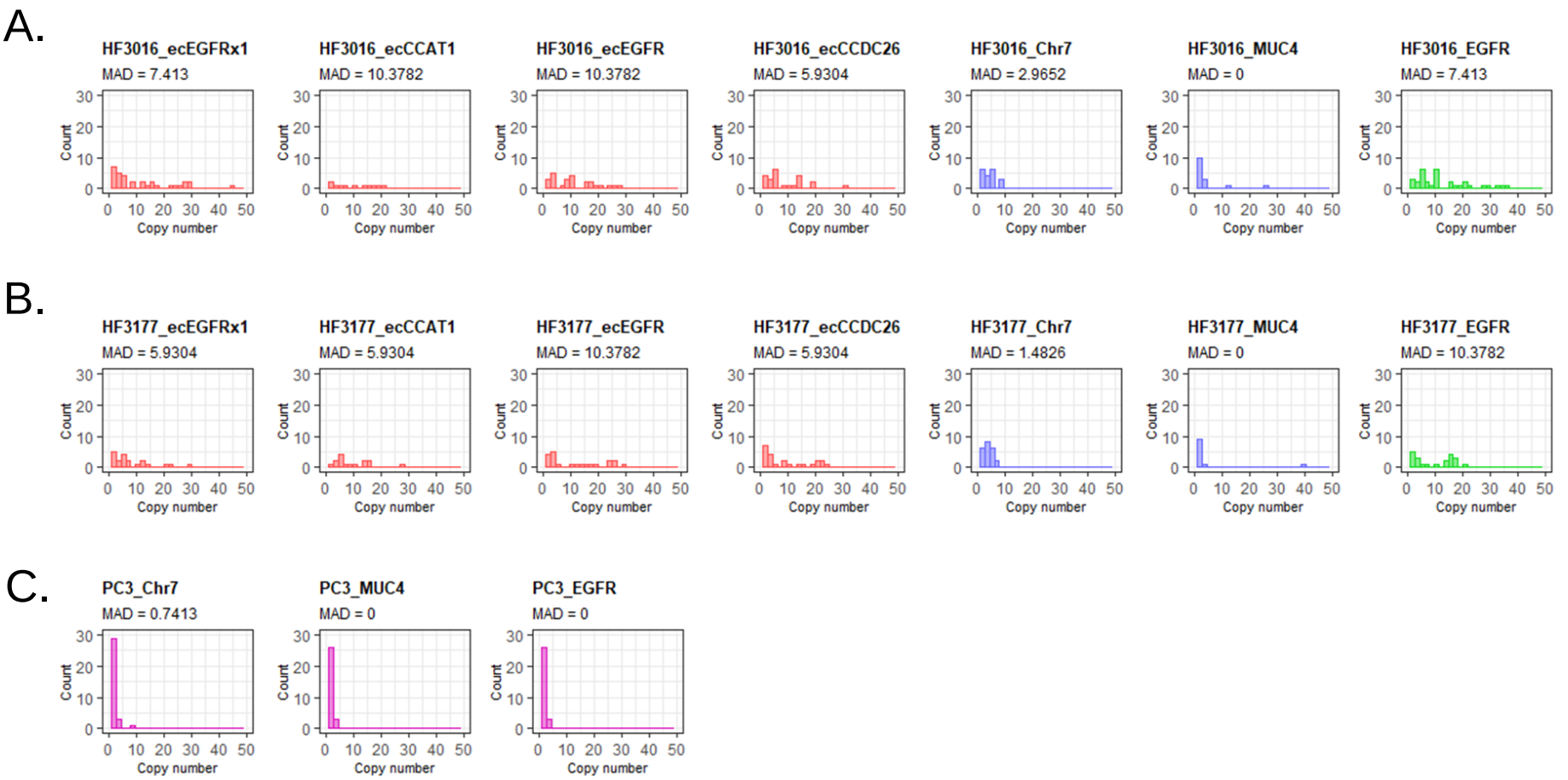






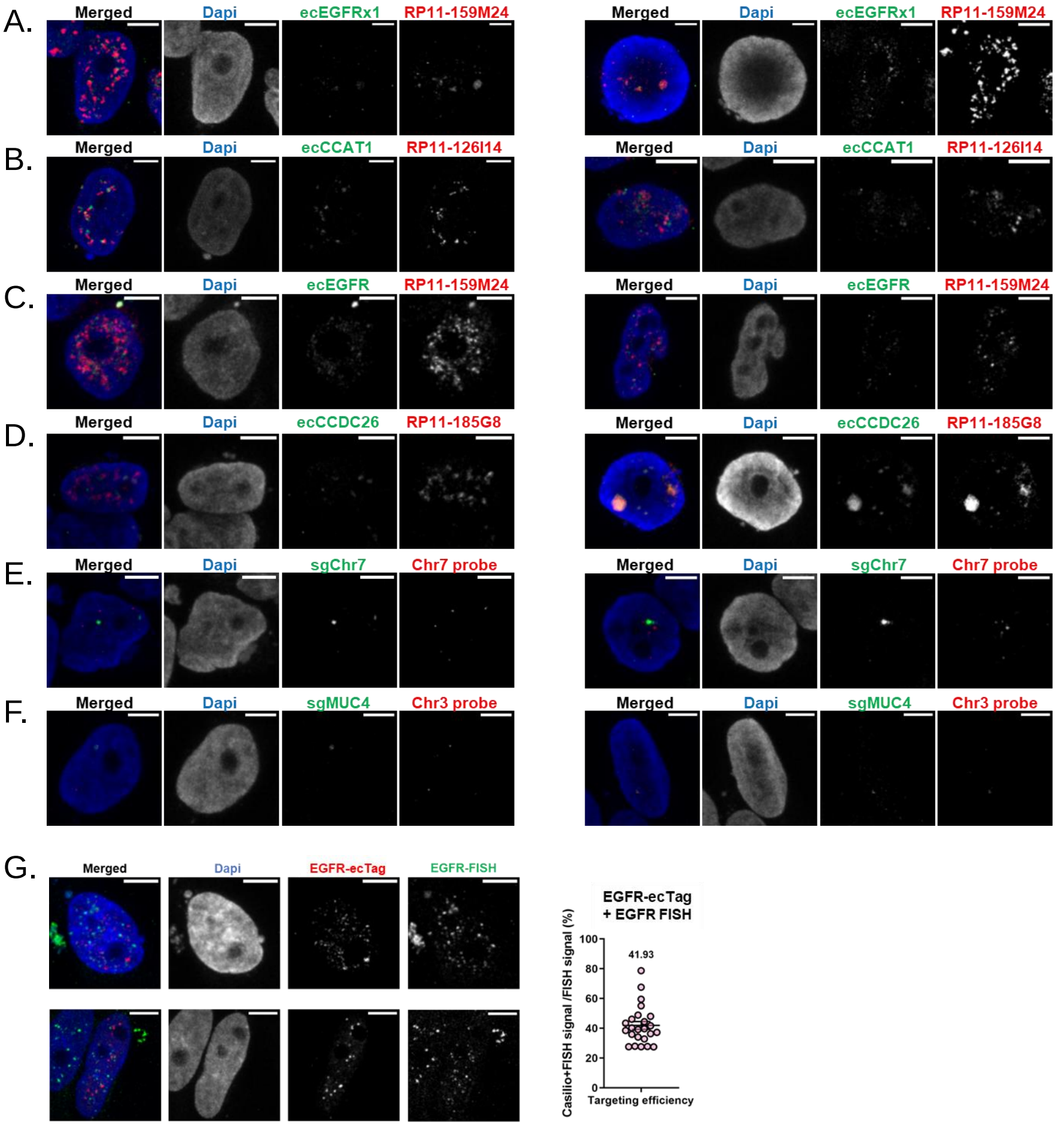
Supplementary Fig. 9 | EcTag transfection. A. Comprehensive single-channel images of ecTag-labeled cells. The nucleolus is indicated by white dotted circle. Scale bar, 10 μm. SgRNAs conjugated with 15 PUFBSs were used. **B.** EcTag-transfected HF3016 cells showed similar cell viability with HF3016 cells treated with transfection reagent (No plasmid). The error bars indicate S.E.M. **C.** Nucleus circularity was measured in ecTag-transfected HF3016 cells, original HF3016 neurosphere cells, and PC3 cells. Nucleus circularity was significantly decreased and more variable in ecTag-transfected HF3016 and original HF3016 in comparison with PC3, confirming that the abnormal nucleus morphology of glioblastoma cells is not the result of ecTag transfection. The error bars indicate S.D.

Supplementary Figure 10.



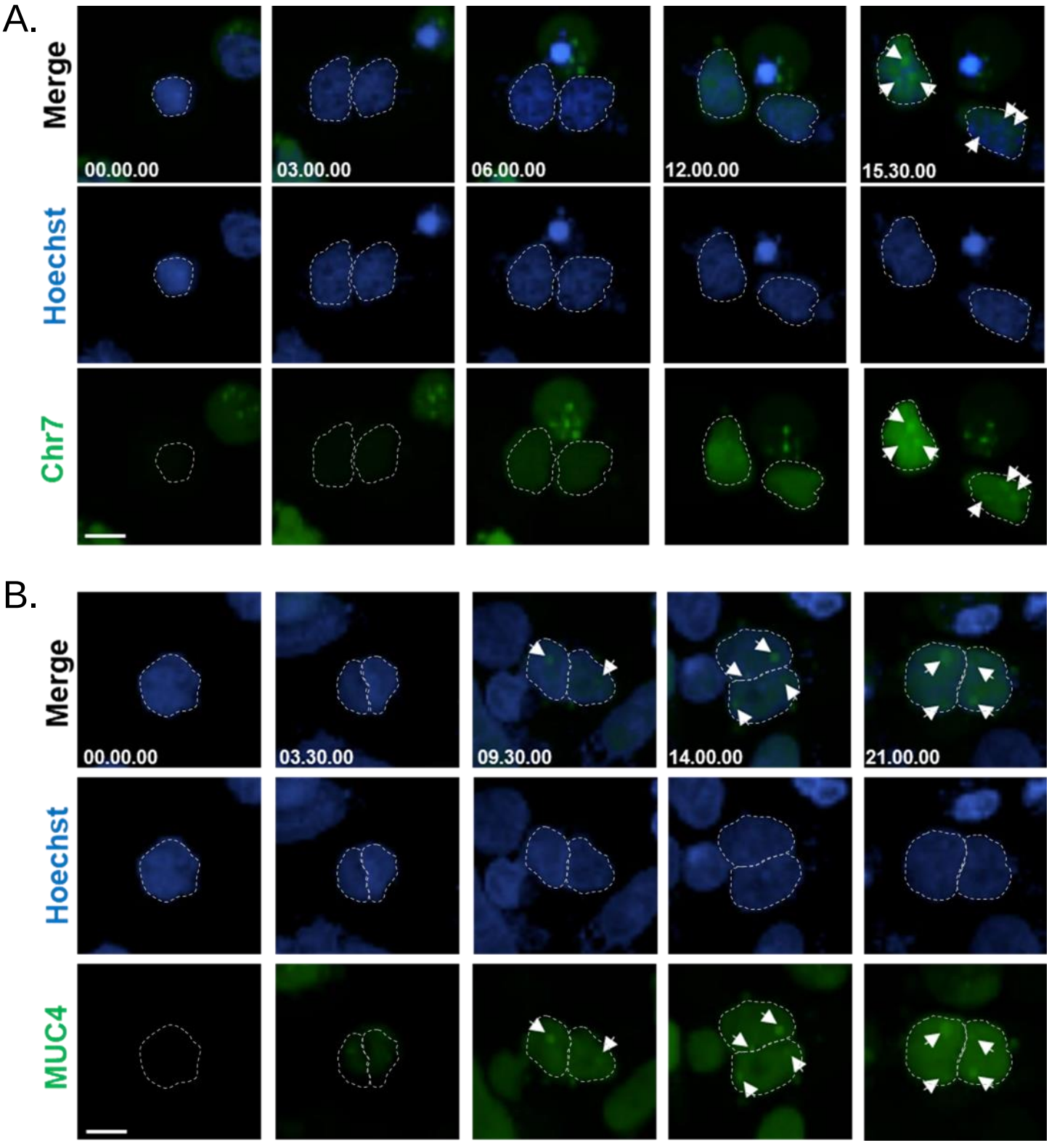
Supplementary Fig. 10 | EcTag recapitulates uneven distribution pattern of ecDNAs. A-B. Distribution of ecTag foci demonstrated patterns of ecDNA uneven segregation, reflected by high MAD score in the ecDNA-targeting sgRNA group (red bar plots) while the Chr7- and *MUC4*-targeting group (blue bar plots) showed low MAD scores or MAD scores equal to 0, reflecting even distribution in HF3016 (A) and HF3177 (B). *EGFR*-targeting sgRNA group (green bar plots) showed an uneven segregation pattern with a high MAD score. n = 11-32 cells per condition. SgRNAs conjugated with 15 PUFBSs were used. **C.** Chr7-, *MUC4*- and *EGFR*-targeting sgRNA in PC3 negative control cells (magenta bar plots) showed an even distribution pattern. n = 29-36 cells per condition. SgRNAs conjugated with 15 PUFBSs were used.

Supplementary Figure 11.



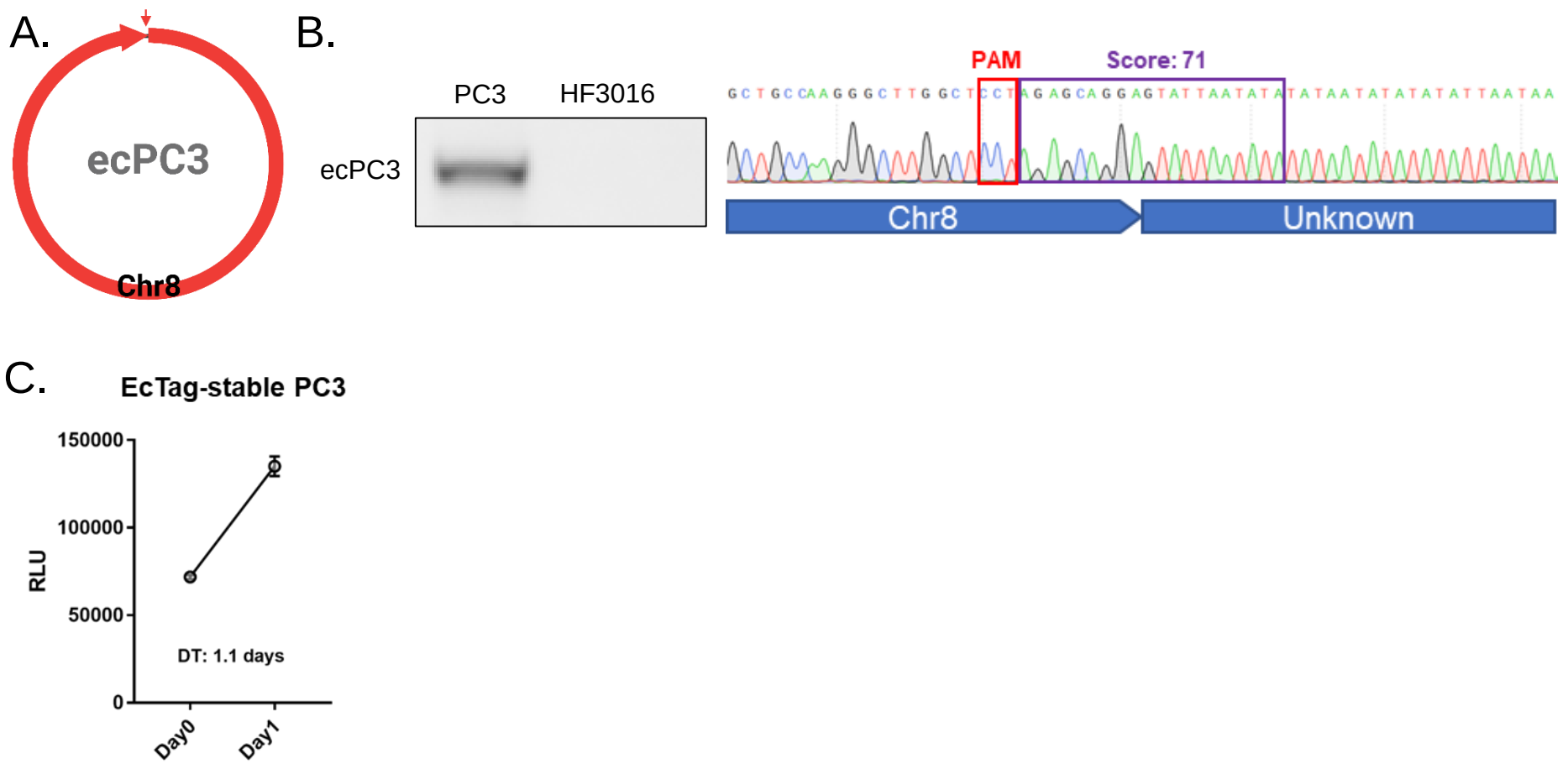
Supplementary Fig. 11 | EcTag targeting efficiency. A-F. Comprehensive single-channel images of two representative FISH validation images showing ecDNA labeled with ecTag (green) and FISH (red) in HF3016. Scale bar, 10μm. SgRNAs conjugated with 25 PUFBSs were used. **G.** Representative images of *EGFR* FISH validation (green BAC probe) performed on *EGFR*-ecTag (red)-transfected HF3016 cells (left panel). Scale bar, 10 μm. Proportion of ecTag and FISH double-positive signals out of the FISH (red) signals was calculated as targeting efficiency (right panel). SgRNAs conjugated with 15 PUFBSs were used. Individual dots in the plot represent the targeting efficiency of an individual cell. The mean of targeting efficiency was indicated in the plot. Error bar represents S.E.M.

Supplementary Figure 12.



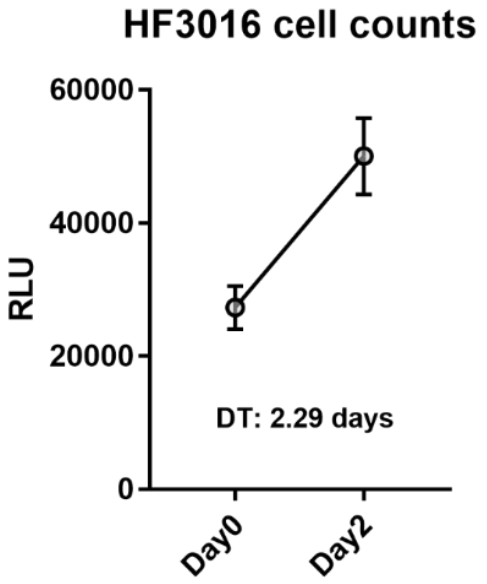
Supplementary Fig. 12 | Spatiotemporal tracking of Chr7 and *MUC4*. A-B. Captured time-lapse images of chromosome 7 (A) and *MUC4* (B) segregation during mitosis. SgRNAs conjugated with 25 PUFBSs were used. Scale bar, 10 μm.

Supplementary Figure 13.



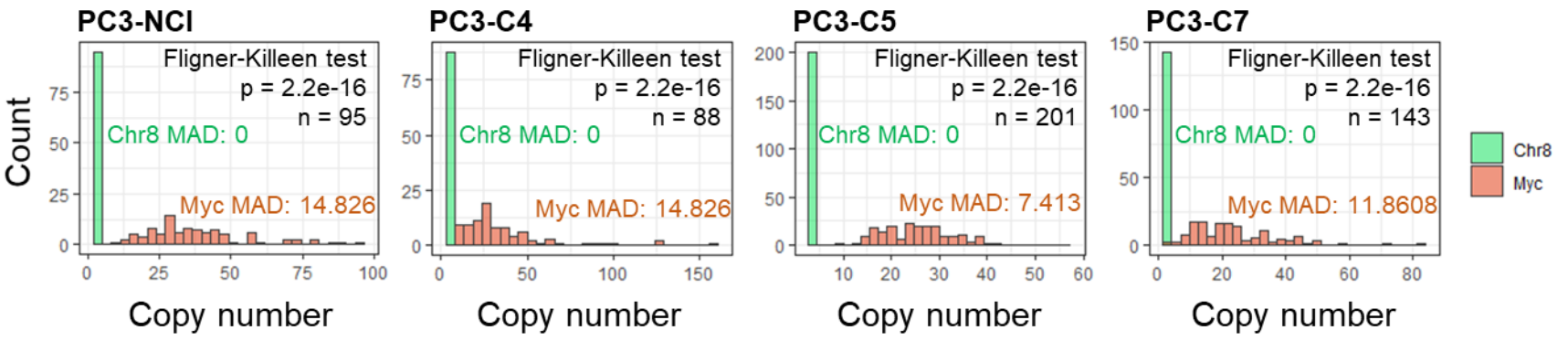
Supplementary Fig.13 | The validation of breakpoint junction sequence found in PC3. **A.** ecDNA structure anticipated by AmpliconArchitect generated from WGS data in PC3. **B.** Gel-images of BP-PCR across breakpoint junction in PC3 and HF3016 (Left). Chromatograms of Sanger sequencing result (Right). The target specificity score was determined by CRISPOR. **C.** Doubling time of ecTag-stable PC3 cells.

Supplementary Figure 14.



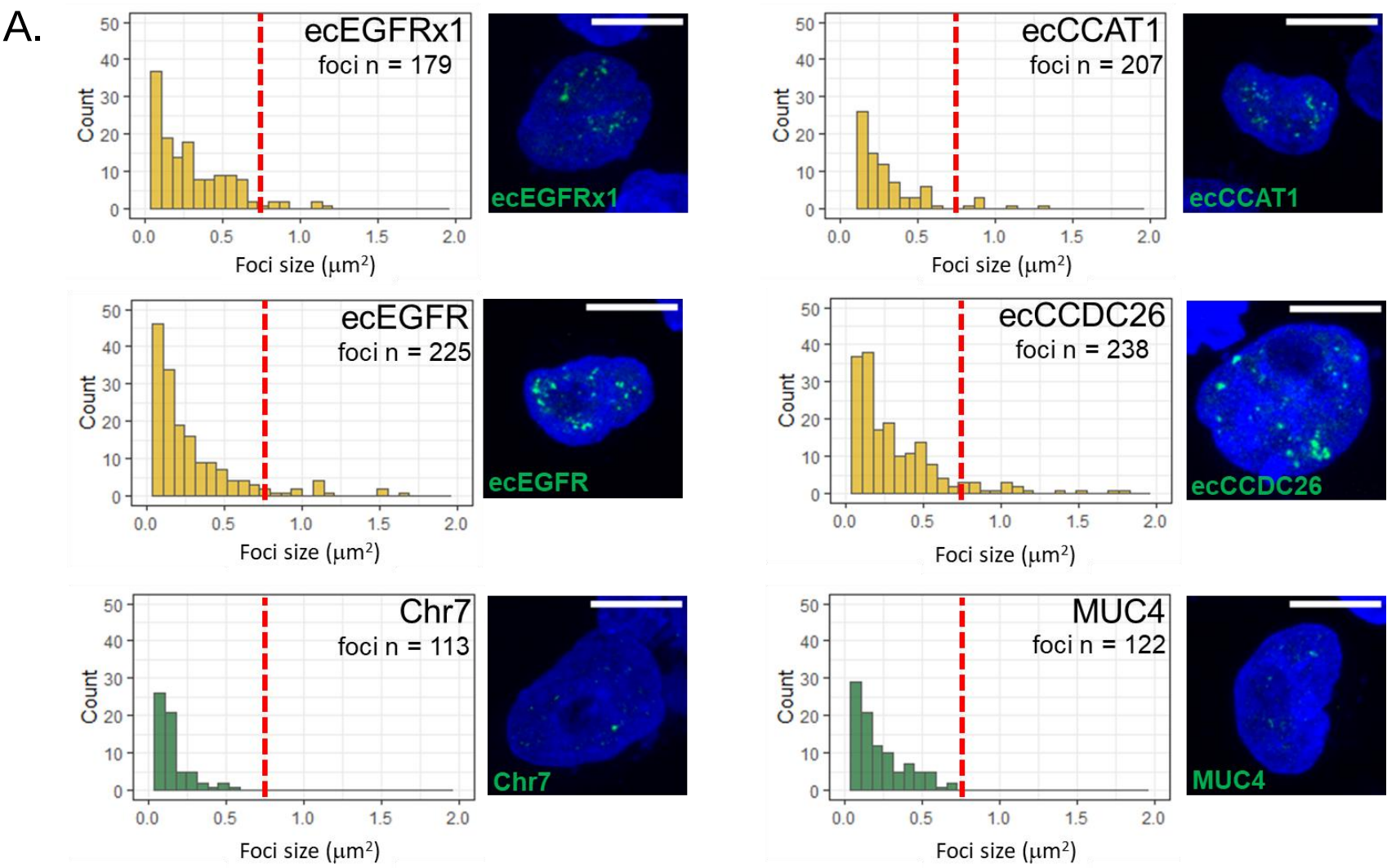
Supplementary Fig. 14 | HF3016 doubling time test. The doubling time (DT) of HF3016 cells were determined by cell viability assay.

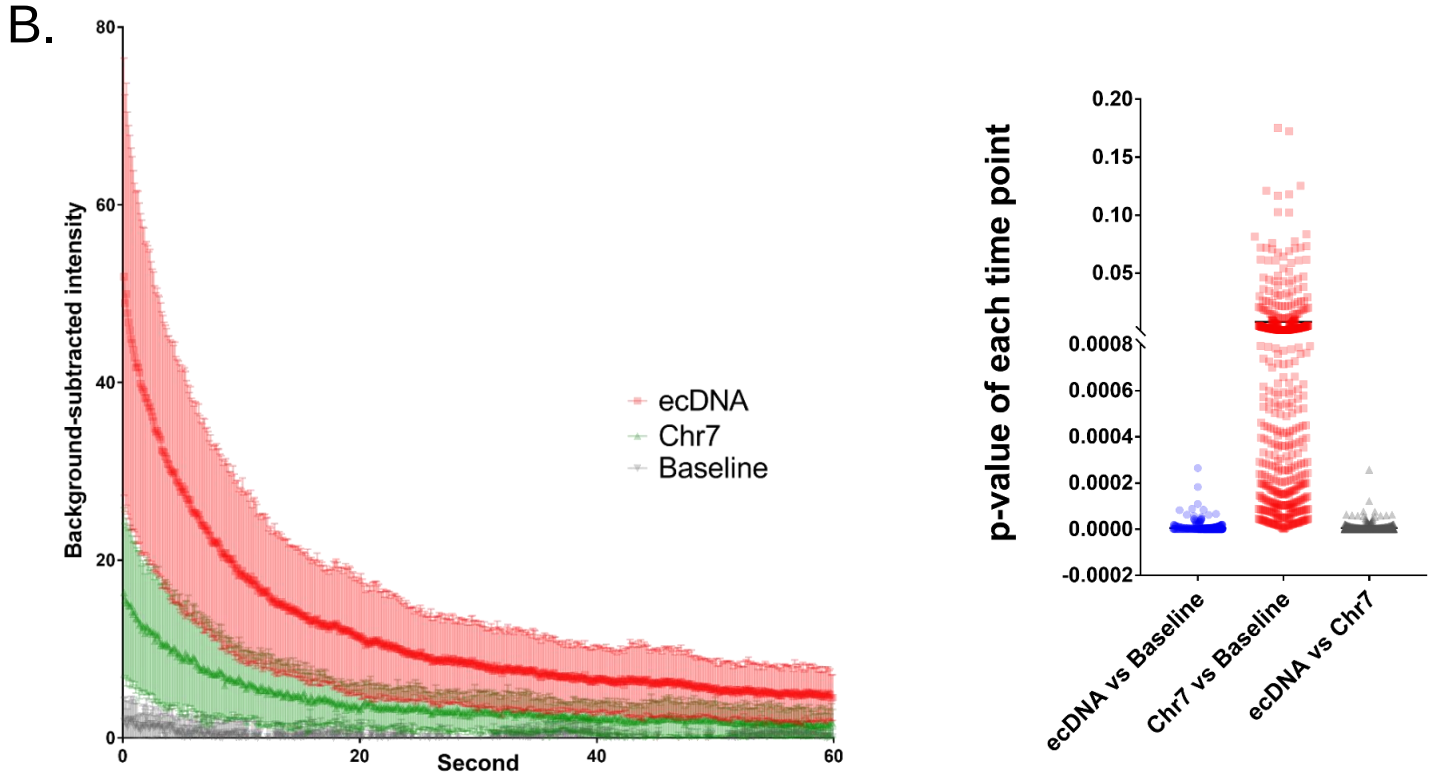
Supplementary Figure 15.



Supplementary Fig. 15 | *MYC* copy number distribution on PC3 and single cell-derived PC3 clones. Copy number was determined using FISH and *p* values indicate the homogeneity of variances between *MYC* and Chr8, determined using a Fligner-Killeen test.

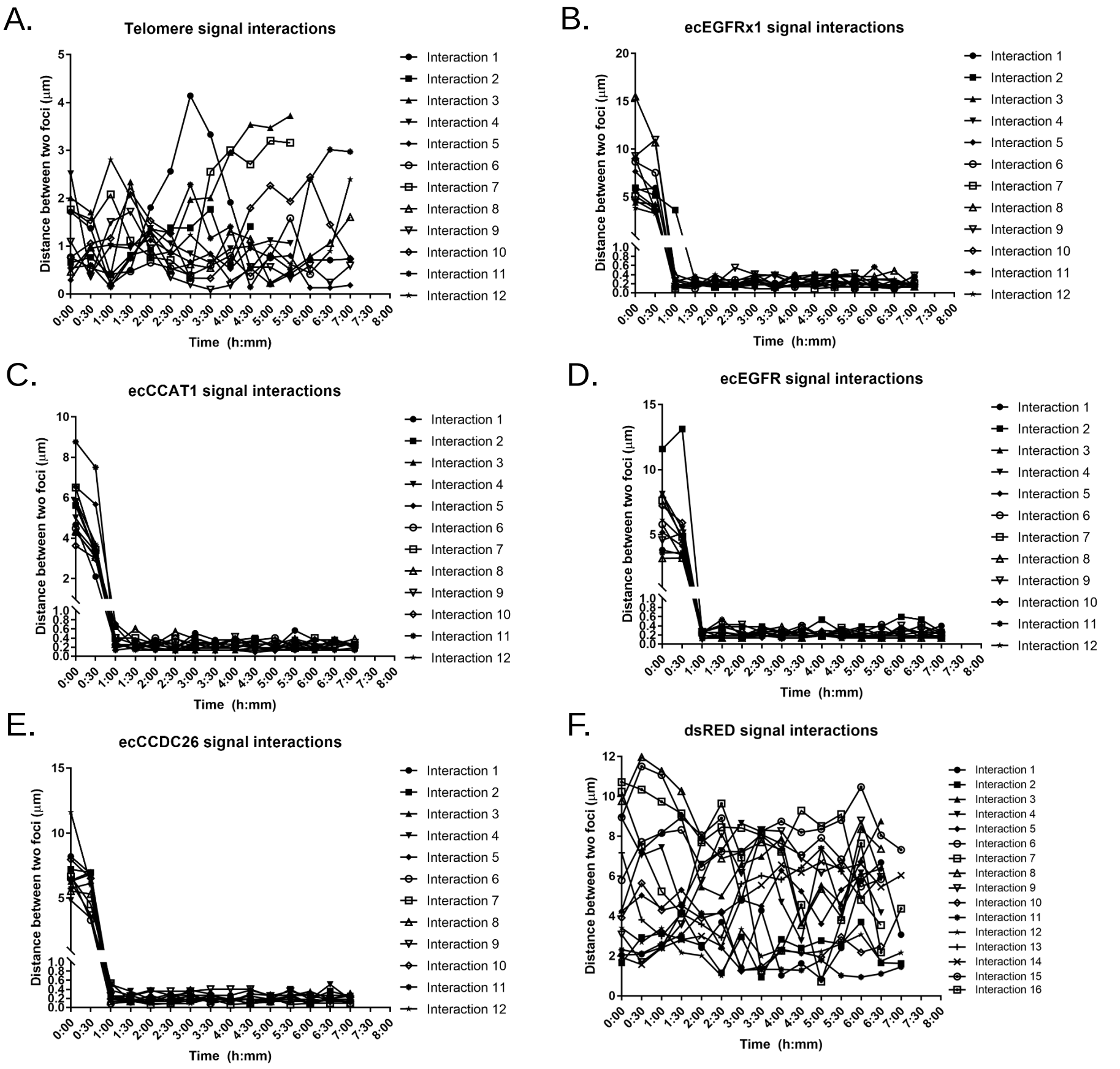
Supplementary Figure 16.





Supplementary Fig. 16 | Validation of ecDNA clustering. A. Size distribution of single ecTag signal foci. HF3016 cells transfected with ecTag including sgRNA tagged with less number of PUFBSs (x15) were imaged and the size of all signal foci were measured. Four ecDNAs showed lager signal size than the size threshold of controls (Chr7 threshold = 0.551 μm^2, MUC4 threshold = 0.680 μm^2). Representative image in right. Scale bar, 10 μm. **B.** Stepwise photobleaching analysis reveals ecDNA clusters consisting multiple ecDNA molecules. Background-subtracted spot intensity demonstrates significantly different decreases in spot intensity with photobleaching (left). P-values indicating intensity differences between groups were determined using t-test at each timepoint.

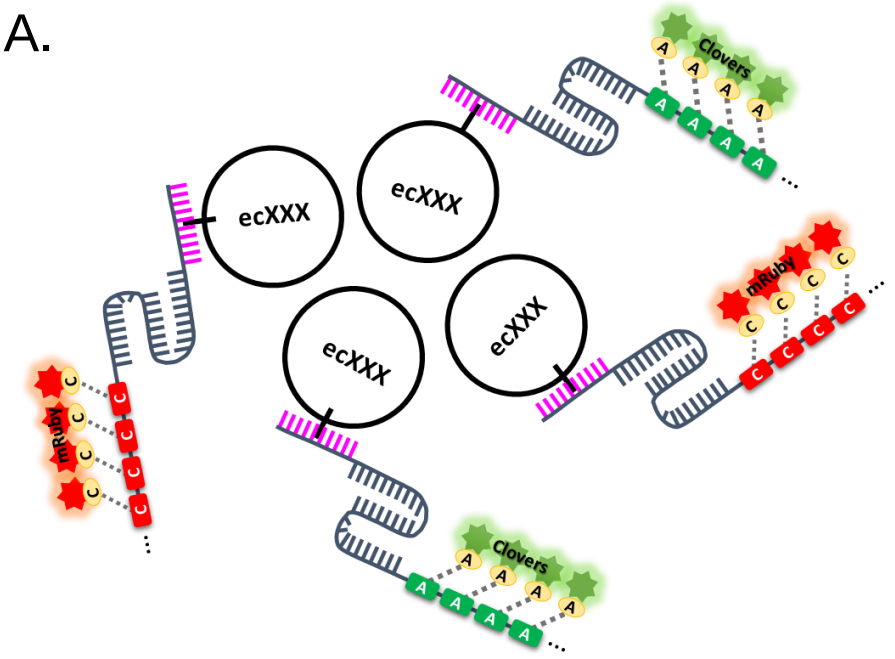
Supplementary Figure 17.



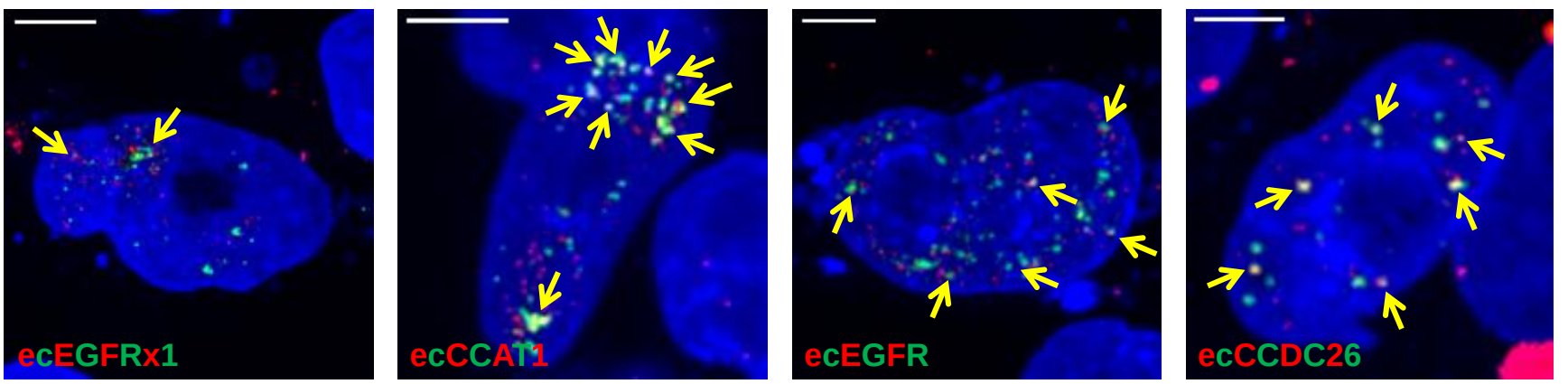
Supplementary Fig. 17 | ecTag signal interactions. A. Telomere signal interactions. SgRNA targeting telomere repeats (target sequence: GTTAGGGTTAGGGTTAGGGTTA) was used. **B-E.** EcDNA signal interactions. SgRNA targeting each ecDNA was used in HF3016. EcDNA clustering demonstrated by stable contacts between signals. **F.** DsRED-targeting ecTag signal interactions. SgRNA targeting dsRED sequence (target sequence: TGCATTACGGGGCCGTCGGA) was used in dsRED-transfected HF3016. The distance between two interacting signal foci (from the core of one foci to another) were measured every 30 minutes, for 7 hours. SgRNAs conjugated with 25 PUFBSs were used. Telomere-imaging and dsRED-plasmid imaging were acquired using Dragonfly confocal microscopy. EcDNAs-imaging were acquired using Phenix confocal microscopy. Twelve interactions for telomeres and each ecDNAs were assessed. Sixteen interactions for dsRED-plasmids were assessed.

Supplementary Figure 18.

A.

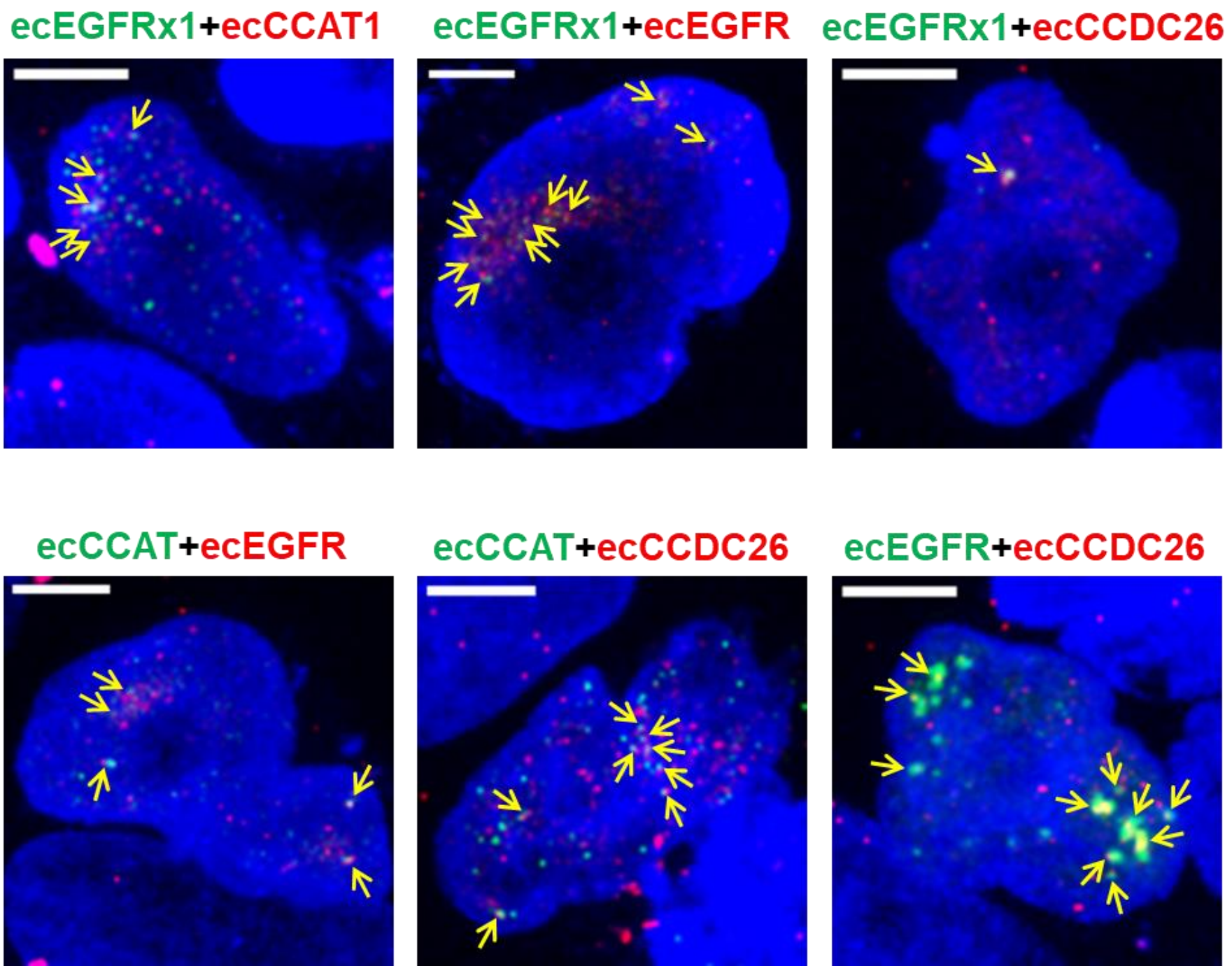


B.



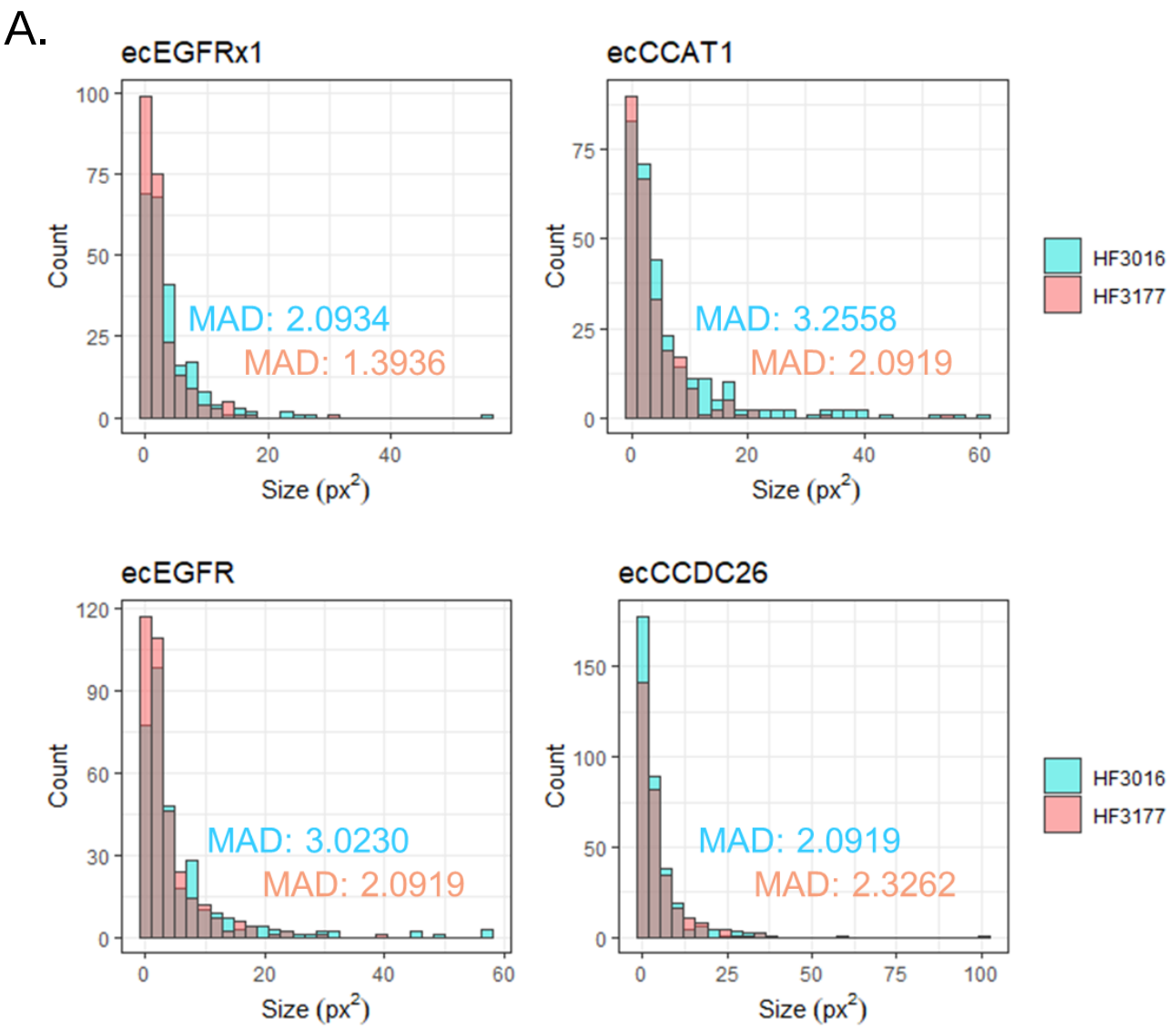
Supplementary Fig. 18 | Dual-color ecDNA labeling system. A. Schematic illustration of dual-color ecDNA labeling system. **B.** Representative dual-color labeling experiments. The ecTag signals were derived from sgRNA containing 15 repeats of PUFBSs. The yellow arrow indicates ecDNA clustering. Scale bar, 10 μm

Supplementary Figure 19.



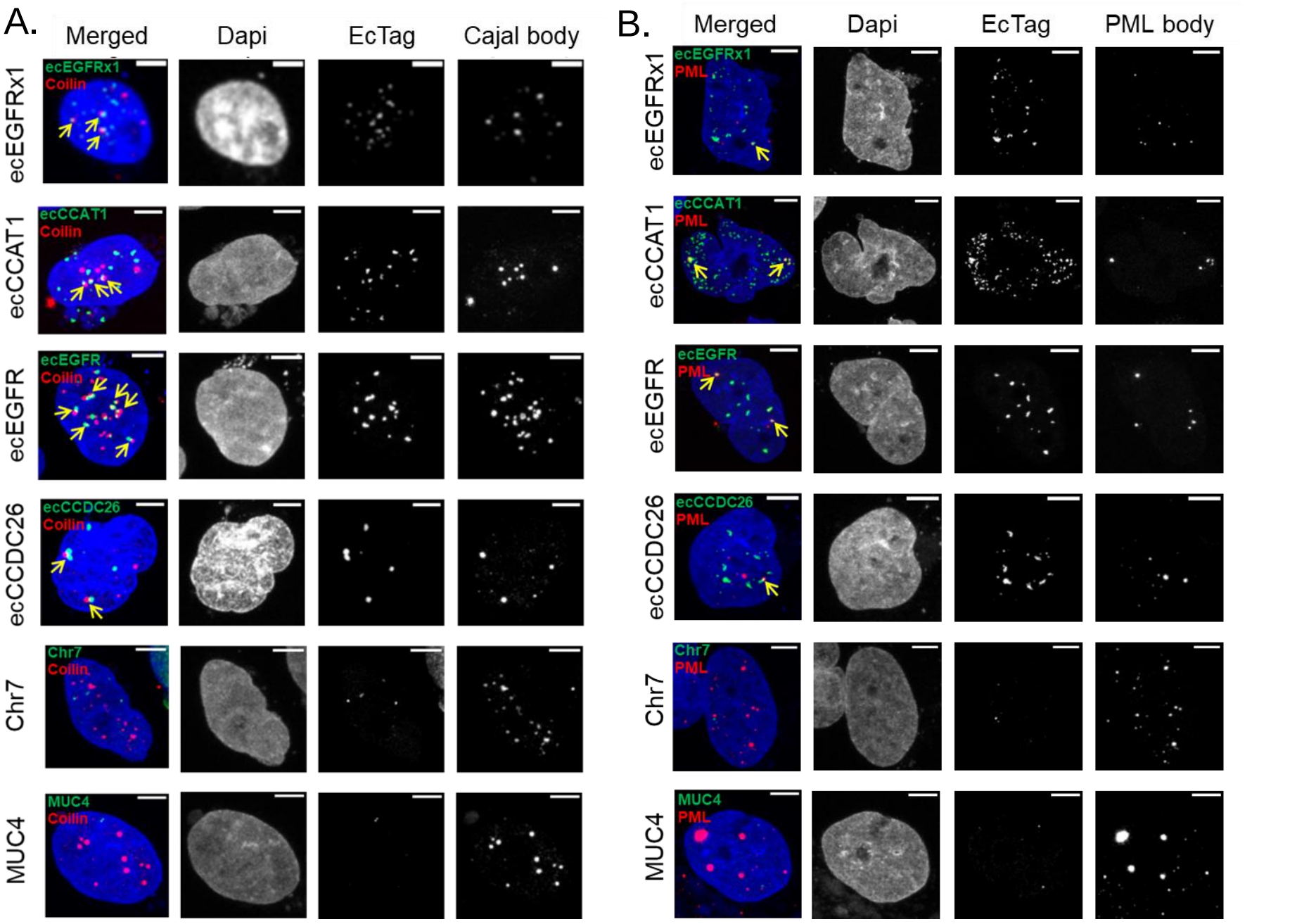
Supplementary Fig. 19 | Multiple ecDNA clustering. Two ecDNAs labeled with two different colors using ecTag. EcTag signals were derived from sgRNA containing 15 repeats of PUFBSs. The yellow arrow indicates multiple ecDNA clustering. Scale bar, 10 μm.

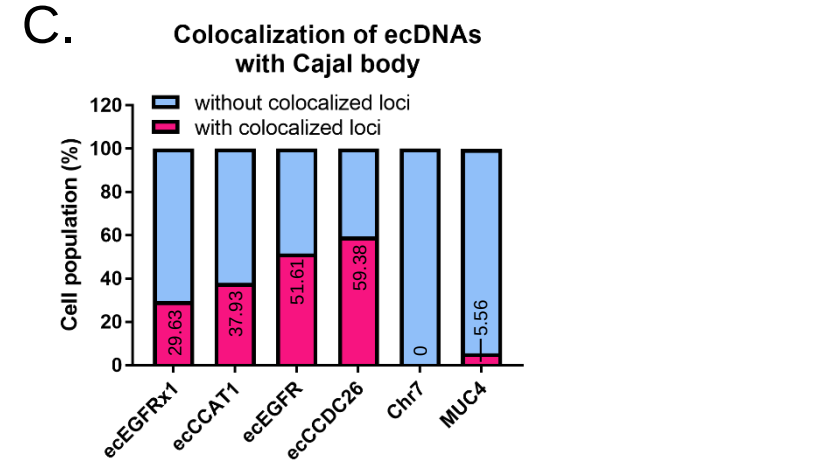
Supplementary Figure 20.

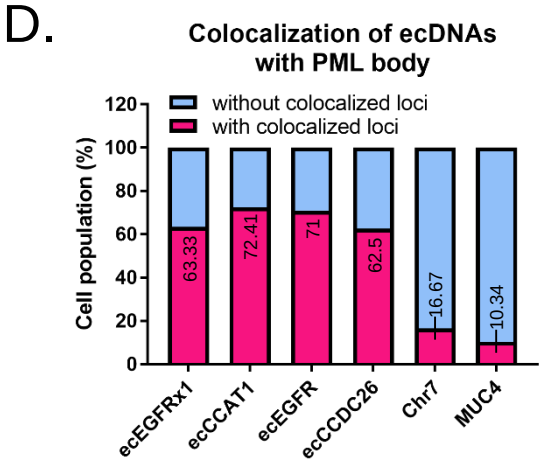


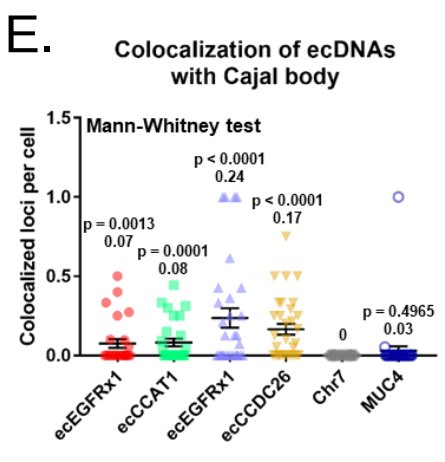
Supplementary Fig. 20 | Comparison of ecTag signal size between primary and recurrent neuroshpere. A. EcTag signals generated using sgRNA tagged with 25 PUFBS repeats were imaged and the size of each signal foci were measured. Size distribution of the four ecDNAs was comparable between primary and recurrent neurosphere models. **B.** No significant differences of the median MAD of signal size distribution between primary and recurrent neurosphere found (Mann-Whitney U test). The error bars represent S.D.

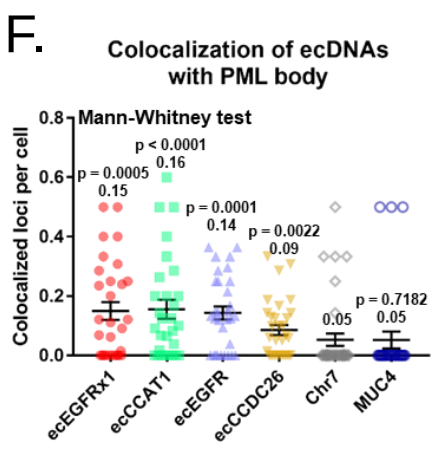
Supplementary Figure 21.

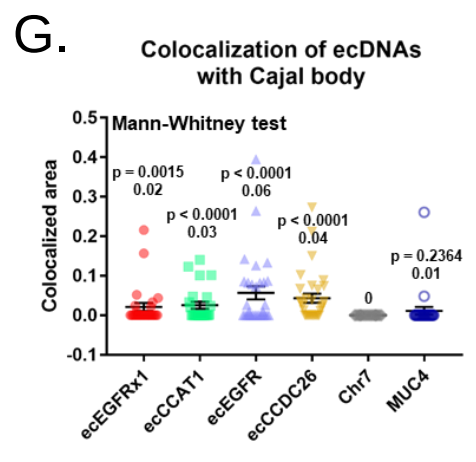


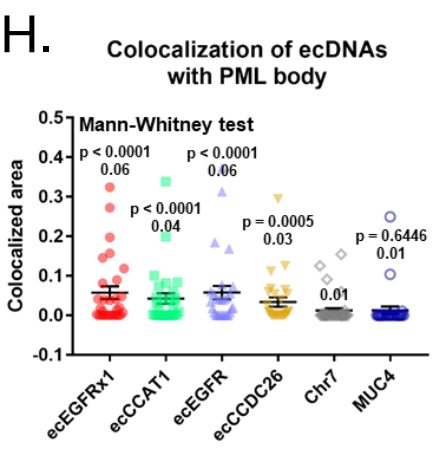








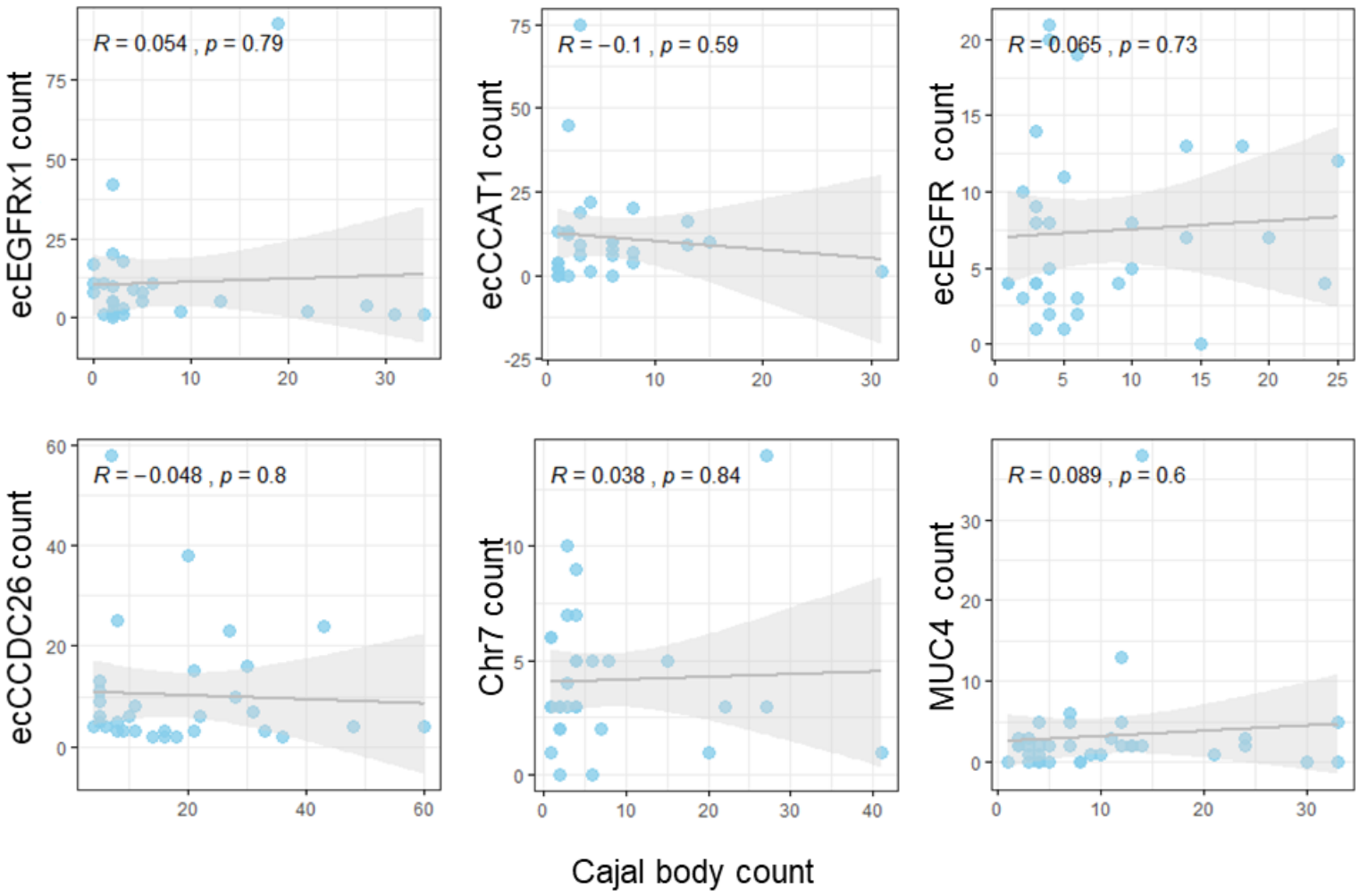




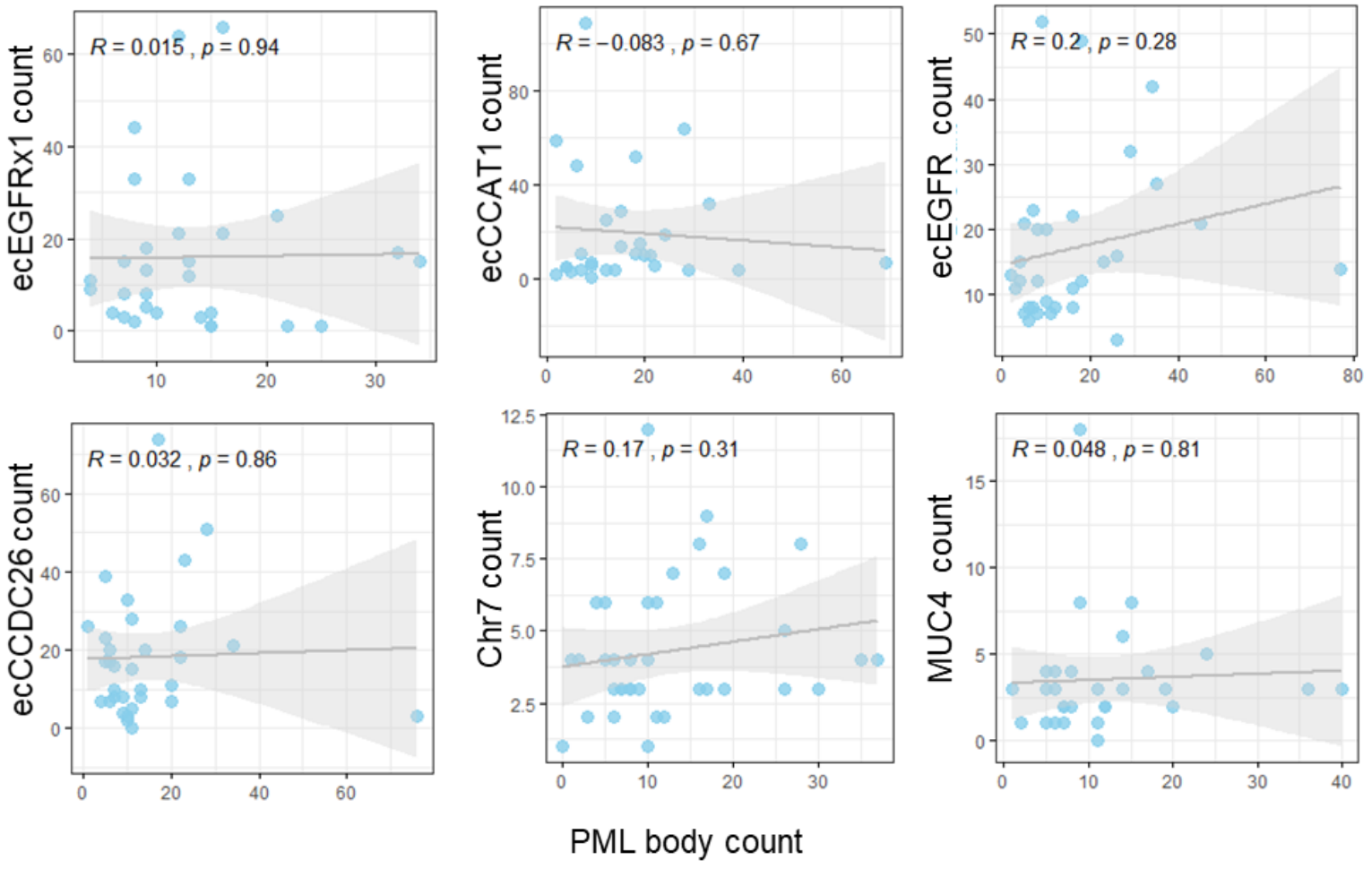
Supplementary Fig. 21 | Colocalization of ecDNAs with nuclear bodies. A. Representative images of Cajal body immunofluorescent staining, Scale bar, 10 μm. SgRNAs conjugated with 25 PUFBSs were used. **B.** Representative images of PML body immunofluorescent staining, Scale bar, 10 μm SgRNAs conjugated with 25 PUFBSs were used. **C.** Proportion of cells with or without the loci colocalized with Cajal body. **D.** Proportion of cells with or without the loci colocalized with PML body. **E.** Colocalized loci with Cajal body per cell. All value was normalized by each ecTag signal. The values of ecDNAs and *MUC4* were compared with Chr7. *p* values were determined by Mann-Whitney U test. Average values are indicated under each *p* value. At least 25 single-cell images per group were analyzed. The error bars represent S.E.M. **F.** Colocalized loci with PML body per cell. All value was normalized by each ecTag signal. The values of ecDNAs and *MUC4* were compared with Chr7. *p* values were determined by Mann-Whitney U test. Average values are indicated under each *p* value. At least 30 single-cell images per group were analyzed. The error bars represent S.E.M. **G.** EcTag signal area merged with Cajal body marker was normalized by each ecTag signal area. The values of ecDNAs and *MUC4* were compared with Chr7. *p* values were determined by Mann-Whitney U test. Average values are indicated under each *p* value. The error bars represent S.E.M. **H.** EcTag signal area merged with PML body marker was normalized by each ecTag signal area. The values of ecDNAs and *MUC4* were compared with Chr7. *p* values were determined by Mann-Whitney U test. Average values are indicated under each *p* value. The error bars represent S.E.M.

Supplementary Figure 22.

A.



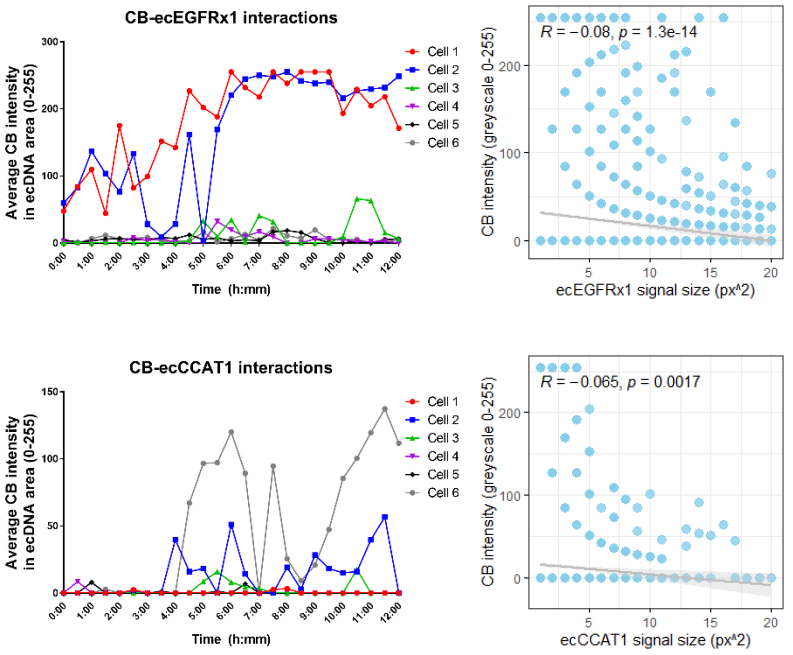
B.



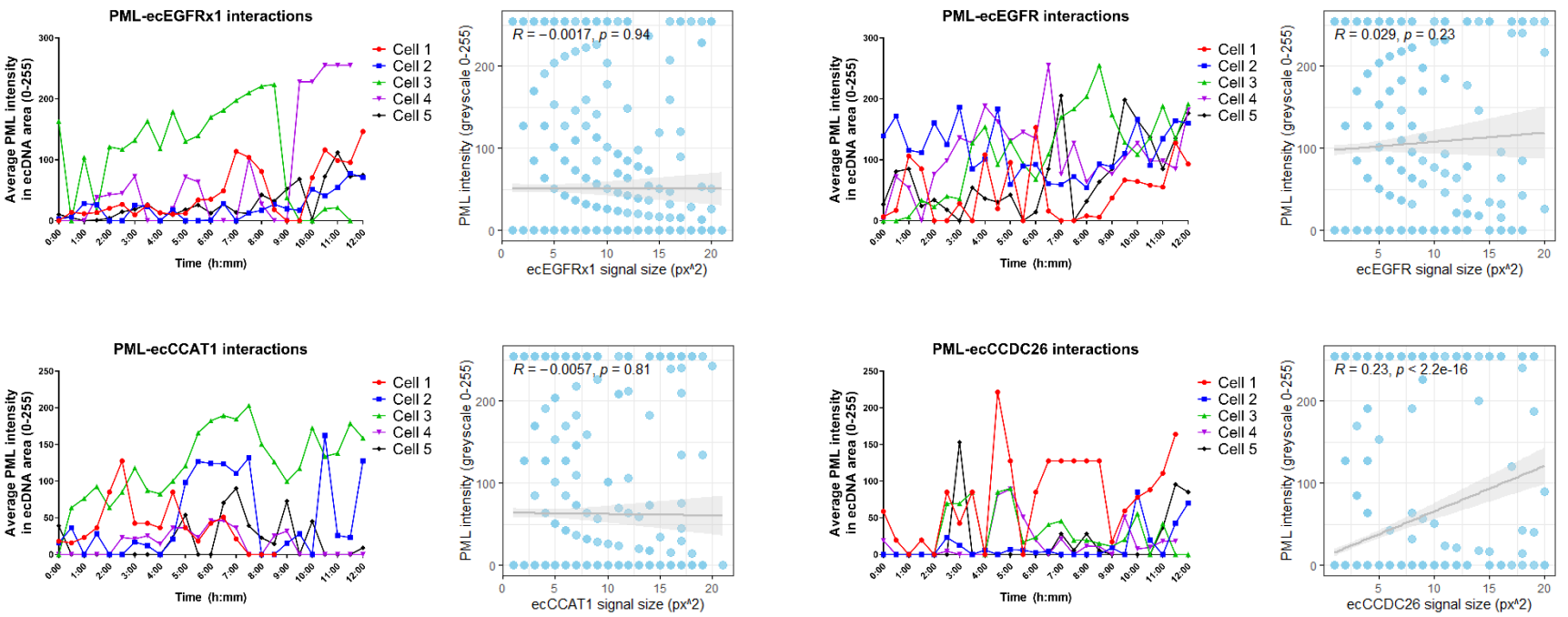
Supplementary Fig. 22 | Correlation of ecDNAs with nuclear bodies. A. Correlation between copy number of ecDNAs and Cajal body count. Correlation score and *p* values were determined by Pearson's correlation test. At least 25 single-cell images per group were analyzed. **B.** Correlation between copy number of ecDNA and PML body count. Correlation score and *p* values were determined by Pearson's correlation test. At least 30 single-cell images per group were analyzed.

Supplementary Figure 23.

A.

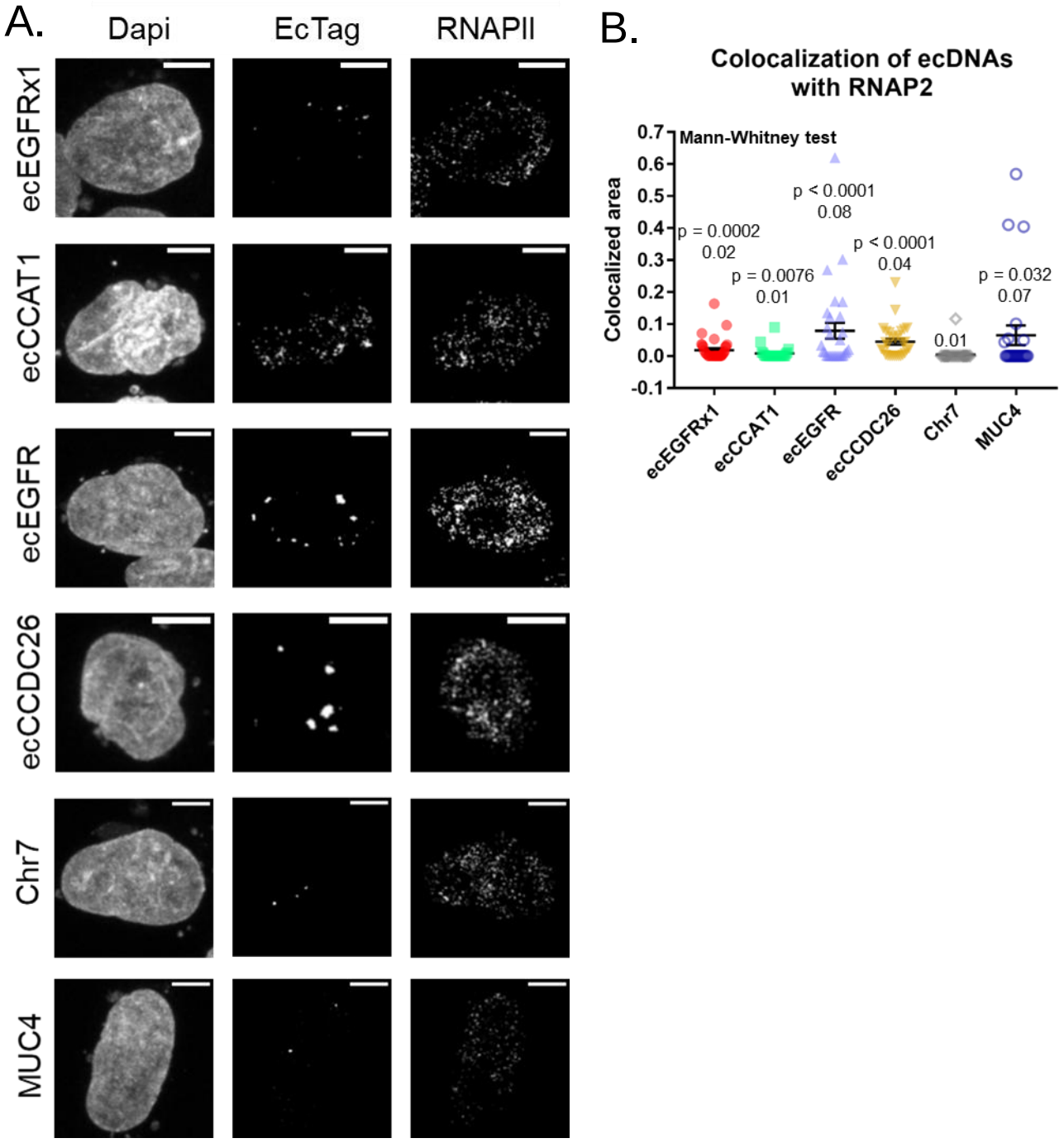


B.



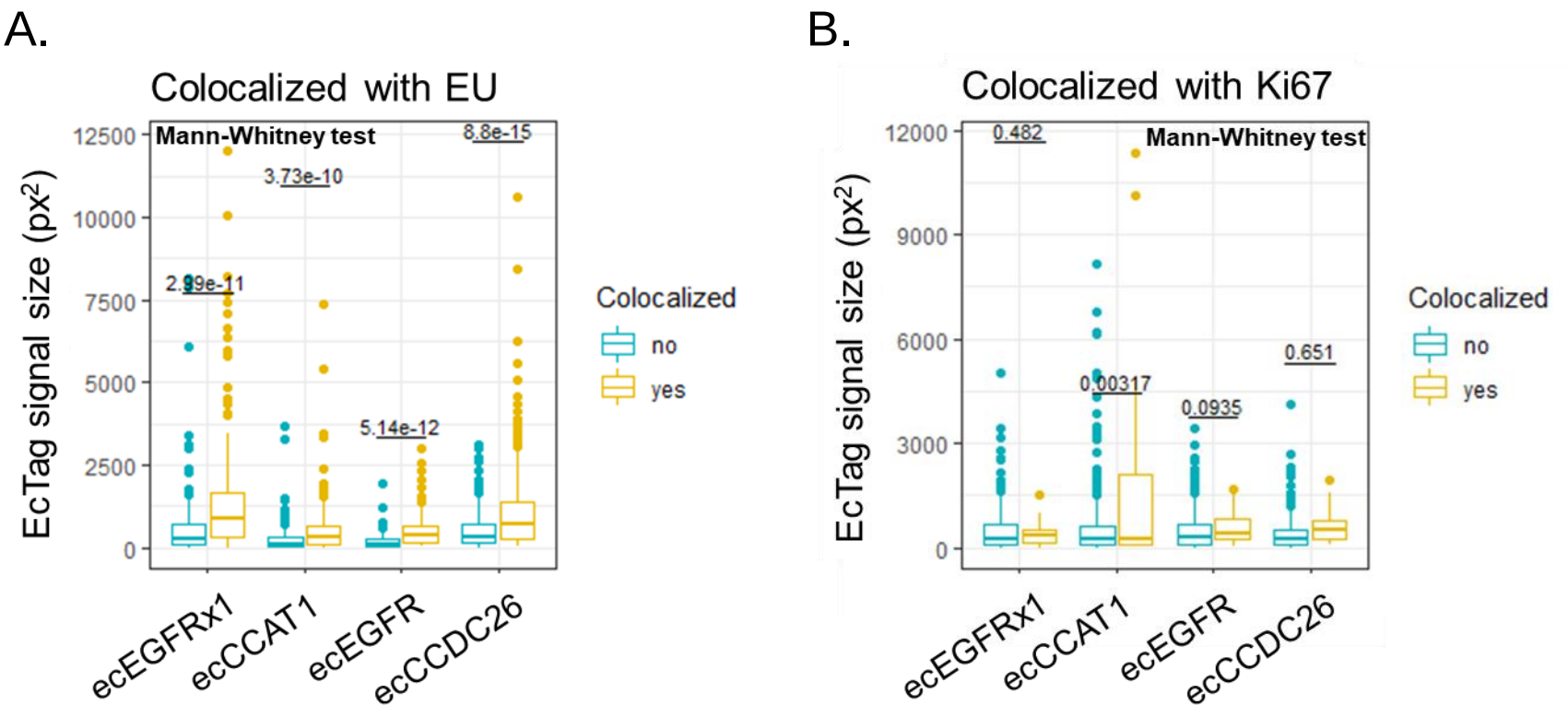
Supplementary Fig. 23 | Live-cell tracking of interaction between ecDNA and nuclear bodies. Cajal bodies- and PML bodies-reporter plasmids conjugated with GFP were used for live-cell tracking. Each ecDNA was labeled by ecTag system with mRuby (red fluorophore). SgRNAs conjugated with 25 PUFBSs were used. Average greyscale intensity (0-255) of the reporter GFP signals captured in the area of ecTag signal foci every 30 minutes for 12 hours. Correlation score and *p* values were determined by Pearson's correlation test. **A.** EcDNA interactions with Cajal bodies. **B.** EcDNA interactions with PML bodies. n = 5-6 cells.

Supplementary Figure 24.



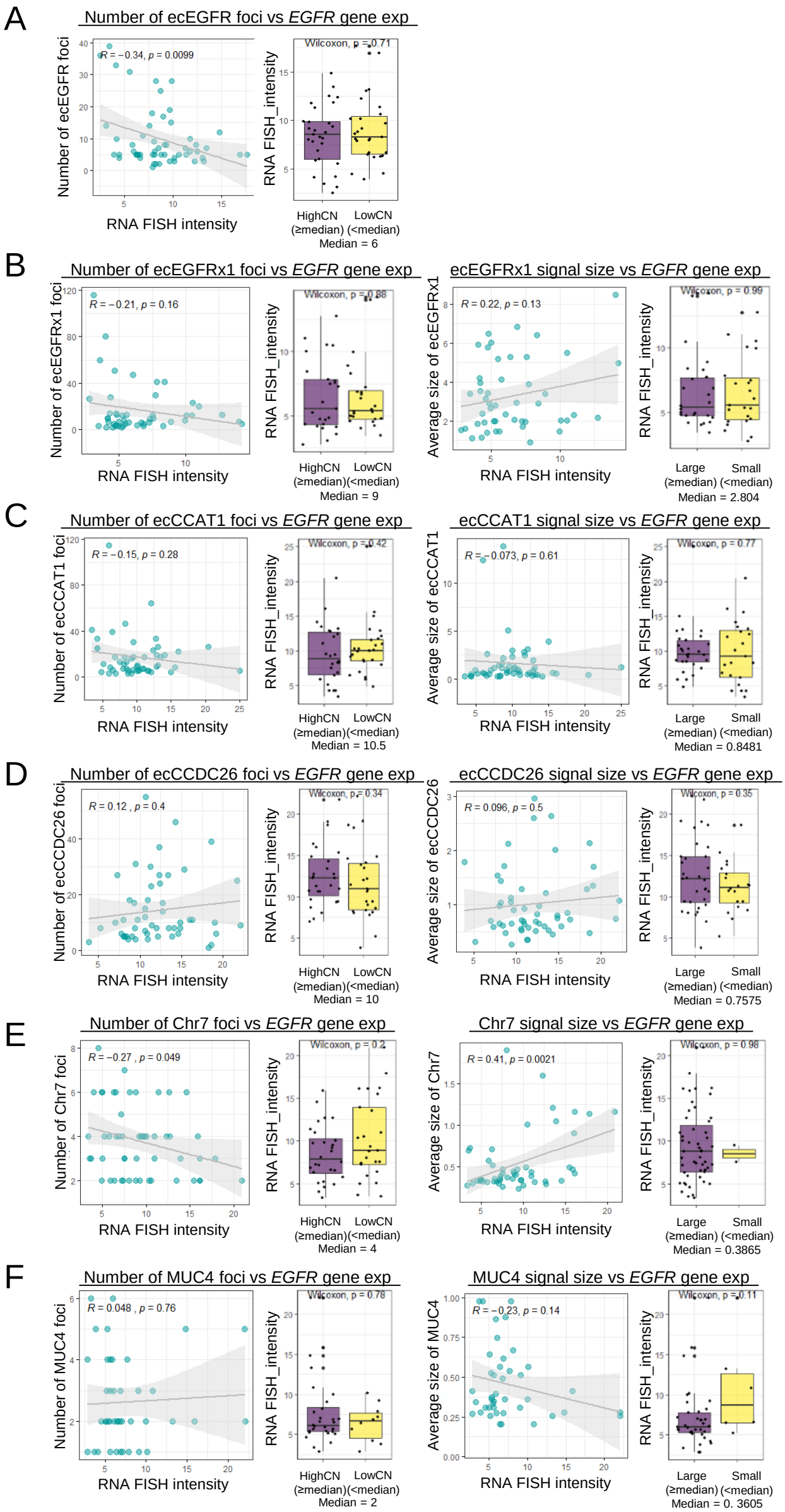
Supplementary Fig. 24 | Colocalization of ecDNAs with RNAPII. A. Grayscale images of Fig 4Aii. Scale bar, 10 μm. SgRNAs conjugated with 25 PUFBSs were used. **B.** EcTag signal area merged with RNAPII marker was normalized by each ecTag signal area. The values of ecDNAs and *MUC4* were compared with Chr7. *p* values were determined by Mann-Whitney U test. Average values are indicated under each *p* value. The error bars represent SE.

Supplementary Figure 25.



Supplementary Fig. 25 | EcDNA clustering colocalized with EU or Ki67. Comparison of ecDNA signal size co-localizing with EU (A) and Ki67 (B). *p* values were determined by Mann-Whitney U test. SgRNAs conjugated with 25 PUFBSs were used.

Supplementary Figure 26.



Supplementary Fig. 26 | Correlation between ecDNA clustering and *EGFR* gene expression. A. Correlation between number of ecEGFR signal foci and *EGFR* gene expression. **B-F.** Correlation between number of each ecTag signal foci and *EGFR* gene expression (left panels). Correlation between the average signal size of each ecTag signals and *EGFR* gene expression (right panels). The correlation was determined by Pearson's correlation test. The bar plots represented the comparison of average *EGFR* gene expression by number of signal foci and signal size. The category of copy number and signal size was determined by its median value. At least 40 single-cell images per group were analyzed. The unit of signal size is μm.

Target	Target cell	Target BP sequence (5' -> 3')	BAC probe (Coordinates, Hg19)	
ecEGFRx1	HF3016 HF3177	ATCATAATCAGGGTTTAGTA	RP11-159M24 (R)	CHR7: 54991664-55140696
			RP11-123K3 (G)	CHR12: 57560622-57740406
ecCCAT1	HF3016 HF3177	AAGTGCCTCATGATGAGCCA	RP11-126I14 (R)	CHR8: 135297772-135460675
			RP11-944J14 (G)	CHR8: 128272851-128459567
ecEGFR	HF3016 HF3177	CAGGATGGAATGACACTCTT	RP11-159M24 (R)	CHR7: 54991664-55140696
			RP11-936I7 (G)	CHR12: 58010761-58186776
ecCCDC26	HF3016 HF3177	ATATCTATACCTATTACACA	RP11-185G8 (R)	CHR8: 130157621-130311270
			RP11-432G14 (G)	CHR8: 135488210-135670526
Chr7	All	AGCTGGGCCAGGAGAGGAGA	Control Chr7 probe	7q11.1
MUC4	All	TTCGGTCAACTACGCAGCCA	Control Chr3 probe	3p11.1
GAL4	All	GAACGACTAGTTAGGCGTGTA		
EGFR	All	GGCAGTACTACAAAGCGGAC	*EGFR* DNA probe	Entire *EGFR* region in Chr7(p11.2)

Supplementary Table 1 | Target sequences of sgRNAs and genomic coordinate of BAC probes. (R) BAC probe conjugated with red fluorophores. (G) BAC probe conjugated with gold fluorophores. The gold probe signals were presented as a green color in all images. All DNA FISH probe was synthesized by Empire Genomics.