nature portfolio

Corresponding author(s): Eunjung Alice Lee and Kwangseog Ahn

Last updated by author(s): Aug 4, 2024

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Immunoblot assay images were collected using FusionCapt Advance Solo 4 16.15 (VILBER).<br>Fluorescence microscopic images were obtained using LSM700 Confocal microscopy (Zeiss). |
| Data analysis | Total RNA-seq data for MRC5 fibroblasts (GSE120890) data were analyzed using Trim Galore! (v0.6.7), STAR (v2.7.1a), GENCODE (v37), TEtranscripts (v2.2.1), Sambamba (v0.8.1), and DESeq2 (v1.34.0). ATAC-seq data were analyzed using Trim Galore! (v0.6.7), BWA (v0.7.17-r1188), Sambamba (v0.8.1), deepTools (v3.5.1), and MACS2 (v2.2.7.1). L1 promoter motifs were searched using HOMER (v4.11).<br>HAT-seq data were analyzed using PEAR (v0.9.6) and BWA-MEM algorithm (v0.7.17-r1194-dirty). TE insertion in HCMV DNA were analyzed using BWA-MEM (v0.7.17-r1194-dirty) and SCRAMble (https://github.com/GeneDx/scramble).<br>WGS data for HCMV DNA were processed using fastp. FastQC, Fastq Screen, BBTools Repair, BWA-MEM, SAMtools, Picard, Qualimap, and Mosdepth. Short variant analyses were performed using Mutect, VarScan2, and BCFtools. Genome-side maps of variants were generated using Circos.<br>Statistical analysis was performed using GraphPad Prism 7 software. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Genomic sequencing and HAT-seq data have been deposited at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) under the BioProject PRJNA1144403. ATAC-seq has been deposit at NCBI under accession number GSE217152. All data supporting the findings of this study are provided in the accompanying Supplementary Information and Source Data files. Source data are provided with this paper.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | n/a |
| Reporting on race, ethnicity, or other socially relevant groupings | n/a |
| Population characteristics | n/a |
| Recruitment | n/a |
| Ethics oversight | n/a |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample size was equivalent to that of routinely used for any particular assay. All experiments were executed in independent biological replicates and individual replicate numbers are given in the figure legends. All values from the independent experiments were indicated in dot plots through individual data points. Immunoblot images are representative of at least two independent experiments. |
| Data exclusions | No data were excluded from the analysis. |
| Replication | Most experiments are reliably reproduced in two or more independent experiments. The number of replications for each experiment are indicated in figure legend respectively. |
| Randomization | No randomization was required because the study was based on molecular and cellular biology techniques. |
| Blinding | The investigators who carried out the experiments were not blinded to the groups and we did not consider blinding necessary because all groups were treated in unbiased manner by algorithms and computer programs. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| [ ] | [x] Antibodies |
| [ ] | [x] Eukaryotic cell lines |
| [x] | [ ] Palaeontology and archaeology |
| [x] | [ ] Animals and other organisms |
| [x] | [ ] Clinical data |
| [x] | [ ] Dual use research of concern |
| [x] | [ ] Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| [x] | [ ] ChIP-seq |
| [ ] | [x] Flow cytometry |
| [x] | [ ] MRI-based neuroimaging |

# Antibodies

Antibodies used

Anti-LINE-1 ORF1p Antibody, clone 4H1 (Merck, MABC1152)
Anti-LINE-1 Retrotransposable Element ORF2 Protein [1G4E11] mouse Antibody (Kerafast, ETL005)
Anti-Cytomegalovirus (IE1/2) Antibody, clone 8B1.2 (Merck, MAB810R)
Anti-HCMV UL99 (pp28) Monoclonal Antibody (Virusys, CA004)
Anti-CMV ICP36 (UL44) Monoclonal Antibody (Virusys, CA006)
Anti-Vinculin Antibody (Sigma, V9131)
Recombinant Anti-YY1 Antibody [EPR4651] (Abcam, ab109228)
Anti-RUNX3 Antibody [2B3] (Abcam, ab135248)
Anti-FLAG® M2 mouse Antibody (Sigma, F3165)
Anti-DYKDDDDK Tag (D6W5B) Rabbit monoclonal Antibody (Cell Signaling, #14793)
Anti-HA tag Mouse Antibody [HA.C5] (Abcam, ab18181)
Anti-HA tag Rabbit Antibody (Abcam, ab9110)
Anti-Histone H2A.X antibody - ChIP Grade (Abcam, ab20669)
Anti-Phospho-Histone H2A.X (Ser139) (20E3) Rabbit monoclonal Antibody (Cell Signaling, #9718)
Anti-alpha Tubulin Antibody (BioVender, LF-PA0146)
Anti-BrdU (Bu20a) Mouse monoclonal Antibody (Cell Signaling, #5292)
Anti-PCNA (PC10) Mouse monoclonal Antibody (Cell Signaling, #2586)
Peroxidase AffiniPure Goat Anti-Mouse IgG (Jackson ImmunoResearch Laboratories, 115-035-062)
Peroxidase AffiniPure Goat Anti-Rabbit IgG (Jackson ImmunoResearch Laboratories, 111-035-003)
Fluorescein (FITC) AffiniPure Rabbit Anti-Goat IgG (Jackson ImmunoResearch Laboratories, 305-095-047)
Donkey anti-Mouse IgG (H+L) Highly Cross-absorbed secondary anbitody, Alexa Fluor 488 (Invitrogen, A21202)
Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568 (Invitrogen, A11004)
Goat anti-Rabbit IgG (H+L) Cross-absorbed secondary antibody, Alexa Fluor 568 (Invitrogen, A11011)
Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ Plus 647 (Invitrogen, A32728)

Validation

All antibodies are commercially available. Antibodies were validated to react with the human cell or virus-produced proteins by the manufacturer. Additionally, L1 ORF1p, YY1, RUNX3 antibodies were validated with in our own experimental model using RNAi-mediated knockdown.

# Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

Human Foreskin Fibroblast (HFF), HeLa, and U373MG cells were purchased in ATCC (HFF-1, SCRC-1041; HeLa, CCL-2; U373MG, ATCC HTB-17).

Authentication

All cells were authenticated by ATCC cell authentication using the short tandem repeat (STR) method.

Mycoplasma contamination

All cells were confirmed as mycoplasma-free by COSMOGENETECH.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

## Plants

| | |
|---|---|
| Seed stocks | n/a |
| Novel plant genotypes | n/a |
| Authentication | n/a |

## Flow Cytometry

### Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

| | |
|---|---|
| Sample preparation | Cells were trypsinized, washed with PBS, fixed with 3.7% formaldehyde and ice-cold 70% ethanol at 4°C. Cells were then incubated in 0.5% Tween-20 and 1% BSA in PBS, and antibody solution. The detailed procedures is in the method section. |
| Instrument | Flow cytometry data were collected using a BD FACS Canto II flow cytometer (BD Biosciences) |
| Software | Flow cytometry data were analyzed using Flowjo software (Flowjo). |
| Cell population abundance | N/A. No sorting procedure included. |
| Gating strategy | Initial gate were set on the live and singlet cells were based on FSC/SSC parameters. The HCMV-infected cell population was determined by gating on cells with UL44 staining, and DNA replicainting cells were determined on the EdU staining. The total number of cells analyzed per sample ranged from 10,000 to 50,000. The gating strategy is provided in Supplementary figure. |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.