nature portfolio

Corresponding author(s): Young Seok Ju, Ji Won Oh

Last updated by author(s): Jun 19, 2021

# 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 | All the sequences were generated by Illumina sequencer (Hiseq X Ten or Novaseq 6000) |
| Data analysis | Sequenced reads were mapped to the human reference genome (GRCh37) using the BWA-MEM (v0.7.17-r1188) algorithm. The duplicated reads were removed by Picard (v2.1.0), and indel realignment and base quality score recalibration were performed by GATK (v4.0). Initially, we identified base substitution and short indels by using HaplotypeCaller (v4.0) and VarScan2 (v2.3.9). We identified somatic genomic rearrangements of the whole genome sequenced samples using Delly (v0.7.6). Segmented copy number profiles were estimated for the whole genome sequenced samples by Sequenza (v3.0.0). Detected variants were inspected using Integrative Genomics Viewer (IGV v2.8.7). Custom codes were written by Python (v3.7.6) and R (v3.6.0). Following packages were used in R: ggtree (v1.16.6), abc (v2.1), Rtsne (v0.15), and pracma (v2.2.9). All custom scripts are available at GitHub (https://github.com/seongyeol-park/Human_Lineage_Tracing and https://github.com/chrono0707/Human_Lineage_Tracing). |

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

Whole-genome and targeted sequencing data have been deposited in the European Genome-phenome Archive (EGA) with accession id: EGAS00001004824.

# 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 | No statistical methods were used to predetermine the sample size. We acquired various number of single-cell clones from individuals, and we estimated the latest developmental time which can be revealed by our dataset. |
| Data exclusions | Among a total of 374 clonally expanded cells, 18 with low depth of coverage (mean depth < 10) were excluded. Based on the VAFs of somatic mutations and established phylogenetic trees, we removed additional 19 samples thought to be multiclonal: they have variants in mutually exclusive lineages simultaneously, and/or low peak VAFs. In addition, three samples which show unexplainable high peak VAFs were excluded for data integrity. Finally, we analyzed the remaining 334 whole-genome sequencing data from single-cell derived clones to reconstruct phylogenetic trees. In targeted-deep sequencing, we only used the mutations with the sufficient read depth for the accurate analysis. The exclusion criteria were not pre-established. We determined the criteria based on the mutation sharing pattern and certainty of lineage assignment. |
| Replication | Five biological replicates (culturing the sample twice from three samples, and separating cultured cell pool into two from two samples) and two technical replicates (sequencing the DNA pool twice) were made. All attempts at replication were successful, and all the detected early mutations were validated in the replicates. |
| Randomization | Not applicable since there was no predetermined group of samples. All possible samples were used. |
| Blinding | Not applicable since this is a descriptive study. Based on the assumption that all the sample information (e.g. anatomical position) is accurate, we traced early embryogenesis using both genomic data and sample information. |

# 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 |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☐ | ☒ Human research participants |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Human research participants

Policy information about studies involving human research participants

| | |
|---|---|
| Population characteristics | We analyzed cells and tissues from seven donors. Age, gender, and cause of death were described at Supplementary Table 1. |

| | |
|---|---|
| Recruitment | From June 2016 to January 2019, ten immediate autopsies had been performed at the Kyungpook National University Hospital and Department of Anatomy at Kyungpook National University, School of Medicine after informed consent. In seven of them, primary culture of tissues was successful and the donors were enrolled in this study. Since most donors were hospitalized before death, cause of death in our cases could be biased. |
| Ethics oversight | All the procedures related to warm-autopsy and tissue sampling were approved by Institutional Review Board of Kyungpook National University (KNU-2018-0088 and KNU-2019-0151) and KAIST (KH2020-029). |

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