

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	No special software or code was used to collect the data.
Data analysis	We used Python for all data analysis. Custom code was at GitHub (https://github.com/peterzheng98/mochiagent) and the Macau University of Science and Technology AI-X code repository (https://must.digitwinmed.com/must/MochiAgent).

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](#)

Restrictions apply to the availability of the developmental and validation datasets, which were used with permission of the participants for the current study. Deidentified data may be available for research purposes from the corresponding authors on reasonable request.

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	We used "sex" in our manuscript because we considered the biological attributes.
Reporting on race, ethnicity, or other socially relevant groupings	Information on race, ethnicity or other socially relevant groupings was not available for all included datasets.
Population characteristics	Data for this study were drawn from the China Health Assessment Investigation 2 (CHAI 2), a project of the International Consortium of Digital Twins in Medicine that prospectively assembles longitudinal EHR data to model current and future health risks and intervention outcomes. The maternal and infant arm of CHAI 2 focuses on establishing maternal pregnancy and infant developmental clocks and on quantifying individual disease risks. The development cohort comprised Cohort #1 (The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, China) and Cohort #2 (The First Affiliated Hospital of Chongqing Medical University, Chongqing, China), jointly contributing 3,682,209 longitudinal clinical visits from 1,064,733 mothers and 719,390 longitudinal clinical visits from 393,375 infants. The external validation cohort was drawn from Dazhou Central Hospital (Sichuan, China), a regional referral centre for women's and children's health, and contributed 263,452 visits from 75,816 mothers and 23,192 visits from 14,418 infants. Biological sex was recorded for all participants. For model training, "healthy" individuals were defined as mothers or infants without any disease diagnosis recorded in the EHR at the time of the corresponding clinical visit. Cohort assembly followed explicit inclusion and exclusion criteria for eligible mother–infant pairs and EHR-derived biomarker availability, and records with insufficient, inconsistent, or duplicate clinical information were excluded. The composition of the study cohort by hospital site, demographic distribution, sex, age, and visit counts is summarized in Extended Data Table 1 and Extended Data Fig. 1.
Recruitment	Participants have consented to use their EHR data for biomedical research. We used de-identified data in our study.
Ethics oversight	The study was performed in compliance with the tenets of the Declaration of Helsinki and was approved by Institutional Review Board/Ethics Committee of the participated hospitals.

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](https://www.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	The size of the cohort was indicated in the tables.
Data exclusions	No data were excluded after passing the initial quality-control step.
Replication	Replication was not relevant. We used independent validation cohorts.
Randomization	Samples were randomly allocated to the training, tuning and testing sets.
Blinding	All data were de-identified to remove any patient-related 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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.