

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
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
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 python 3.9.7, scipy 1.3.7, numpy 1.20.3, pandas 1.3.5

Data analysis python 3.9.7, scipy 1.3.7, numpy 1.20.3, pandas 1.3.5, scikit-learn 1.0.2, umap-learn 0.5.3, lifelines 0.27.4, R 4.2.2, survival 3.5-3

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](#)

The register-based data that support the findings of this study are available from The Danish Health Data Agency but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. All other data are available from the authors upon reasonable request. The browsable biomarker dataset of the human pathome is available at pathoage.com.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	Some of the findings in this manuscript are sex-specific. Sex is included in the provided registry-based data. Sex specific analyses were performed.
Population characteristics	Covariates-relevant population characteristics in this study include age, sex, mortality and structured clinical pathology codes.
Recruitment	Recruitment is not relevant in this register-based study.
Ethics oversight	Use of The Danish National Health Registry of Pathology in this study was approved by The Danish Data Protection Agency (Datatilsynet).

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 calculation was not relevant to the registry-based portion of this study.
Data exclusions	Pathology records with five or fewer clinical text features were excluded from this study.
Replication	Experiments were performed in at least three separate biological replicates.
Randomization	Randomization was not relevant to this study.
Blinding	Blinding was not relevant to this study.

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

Methods

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

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Human male primary fibroblasts (Coriell, NJ, USA) AG08498 (AG), GM22159 (159) and GM22222 (222)
Authentication	None of the cell lines used were authenticated. Cells were obtained directly from Coriell.

Mycoplasma contamination

All cell lines were tested negative for mycoplasma contamination. Mycoplasma testing was performed with Eurofins Genomics

Commonly misidentified lines
(See [ICLAC](#) register)

Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

Drosophila melanogaster, wild-type w1118.

Wild animals

The study did not involve wild animals.

Reporting on sex

Findings apply to male fruit flies.

Field-collected samples

The study did not involved samples collected from the field.

Ethics oversight

No ethical approval or guidance was required for Drosophila melanogaster work.

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