

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
<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. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P 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 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	No software was used for data collection.
Data analysis	Dino-ViT v1 open source code was used with modifications (detailed in Methods) to learn the self-supervised embedding from cell-based image tiles. Analysis code used to make evaluation and plots are available at https://github.com/insitro/cp-posh (accession ID: 10.5281/zenodo.17470307). Models trained in this study are available at https://huggingface.co/insitro/cp-posh . The models and code are provided under a CC BY-NC-SA 4.0 license.

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 single cell tile image data generated in this study are accessible at ([s3://insitro-research-2023-cellpaint-posh/](https://insitro-research-2023-cellpaint-posh/), Zenodo accession ID: 10.5281/zenodo.17470307). The NGS data generated in this study are accessible through NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA1359537. The

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](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	Data for the pooled screening experiments (POC, MOA, and druggable genome screens; Figs. 3-6) were acquired with high coverage. The mean number of individual cells analyzed per gene knockout (KO) was ~1300 (POC), ~4000 (MOA), and ~1500 (druggable genome). For the orthogonal, arrayed validation experiment (Figs. 6h-i), data were generated from two technical replicate wells per KO. The total number of images analyzed per condition was 567 (intergenic), 568 (AURKAIP1), 568 (HSD17B10), and 568 (CYP11B1).
Data exclusions	No data were excluded from the pooled screening analyses.
Replication	In pooled screening, each perturbation is represented by hundreds to thousands of single-cell measurements, providing a replicate depth equivalent to or greater than standard pooled CRISPR screens. In arrayed screening, data were generated from two technical replicate wells per KO.
Randomization	The pooled format inherently randomizes perturbations across the culture plates, mitigating spatial confounders (e.g., plate-position effects) common to arrayed screens. In arrayed format, the experiment was not randomized.
Blinding	In pooled screening, data collection was intrinsically blinded, as the perturbation identity at any given x-y coordinate was unknown to the experimenter until the in situ sequencing and decoding were complete. In arrayed format, investigators were not blinded during the experiment or outcome assessment.

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

Antibodies

Antibodies used	pS6 (Cell Signaling Technologies, cat. 2215S, lot 18), Tom20 (Novus, cat. NBP1-81556, lot. 12083), AlexaFluor 750-conjugated Secondary (cat. SA5-10043, lot. WL3448431)
Validation	Per manufacturer's validation

Eukaryotic cell lines

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

Cell line source(s)	A549 (ATCC)
Authentication	No authentication
Mycoplasma contamination	No mycoplasma testing
Commonly misidentified lines (See ICLAC register)	N/A

Plants

Seed stocks	N/A
Novel plant genotypes	N/A
Authentication	N/A