

Supplementary Information

A single-cell resolved genotype-phenotype map using genome-wide genetic and environmental perturbations

Mariona Nadal-Ribelles^{1,2,*}, Carme Solé^{1,2}, Anna Diez-Villanueva², Camille Stephan-Otto Attolini², Yaima Matas^{1,2}, Lars Steinmetz^{3,4}, Eulalia de Nadal^{1,2,*} and Francesc Posas^{1,2,*}

¹Department of Medicine and Life Sciences, Universitat Pompeu Fabra. Barcelona, Spain

²Institute for Research in Biomedicine (IRB Barcelona), the Barcelona Institute of Science and Technology. Barcelona, Spain

³Department of Genetics, Stanford University, School of Medicine, California, USA.

⁴European Molecular Biology Laboratory, Heidelberg, Germany.

*Corresponding authors; mariona.nadal@irbbarcelona.org (MNR); eulalia.nadal@irbbarcelona.org (EdN), francesc.posas@irbbarcelona.org (FP).

The file includes:

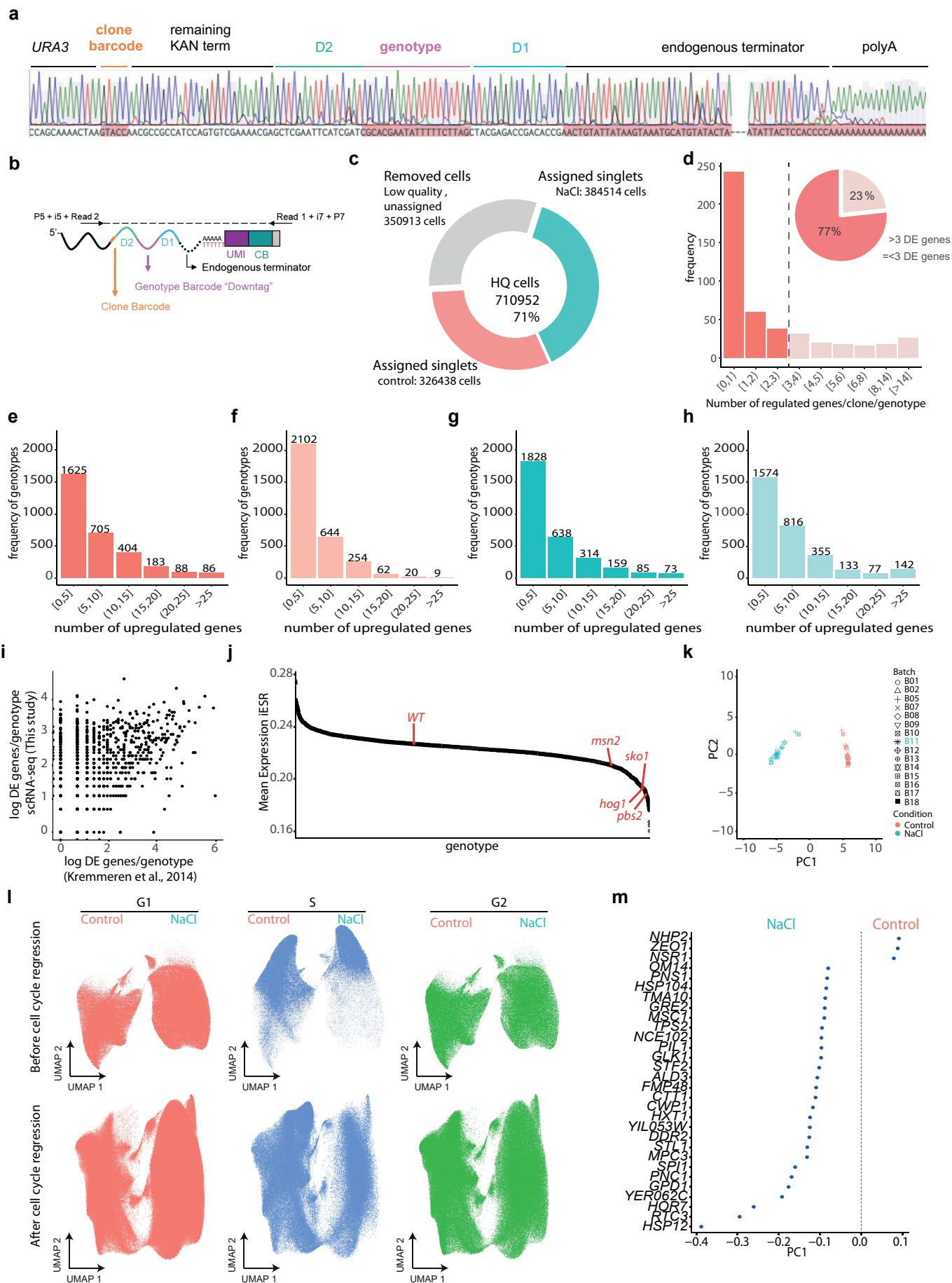
Supplementary Figures 1 to 5.

Supplementary Tables:

Supplementary Table 1: Plasmids generated in this study.

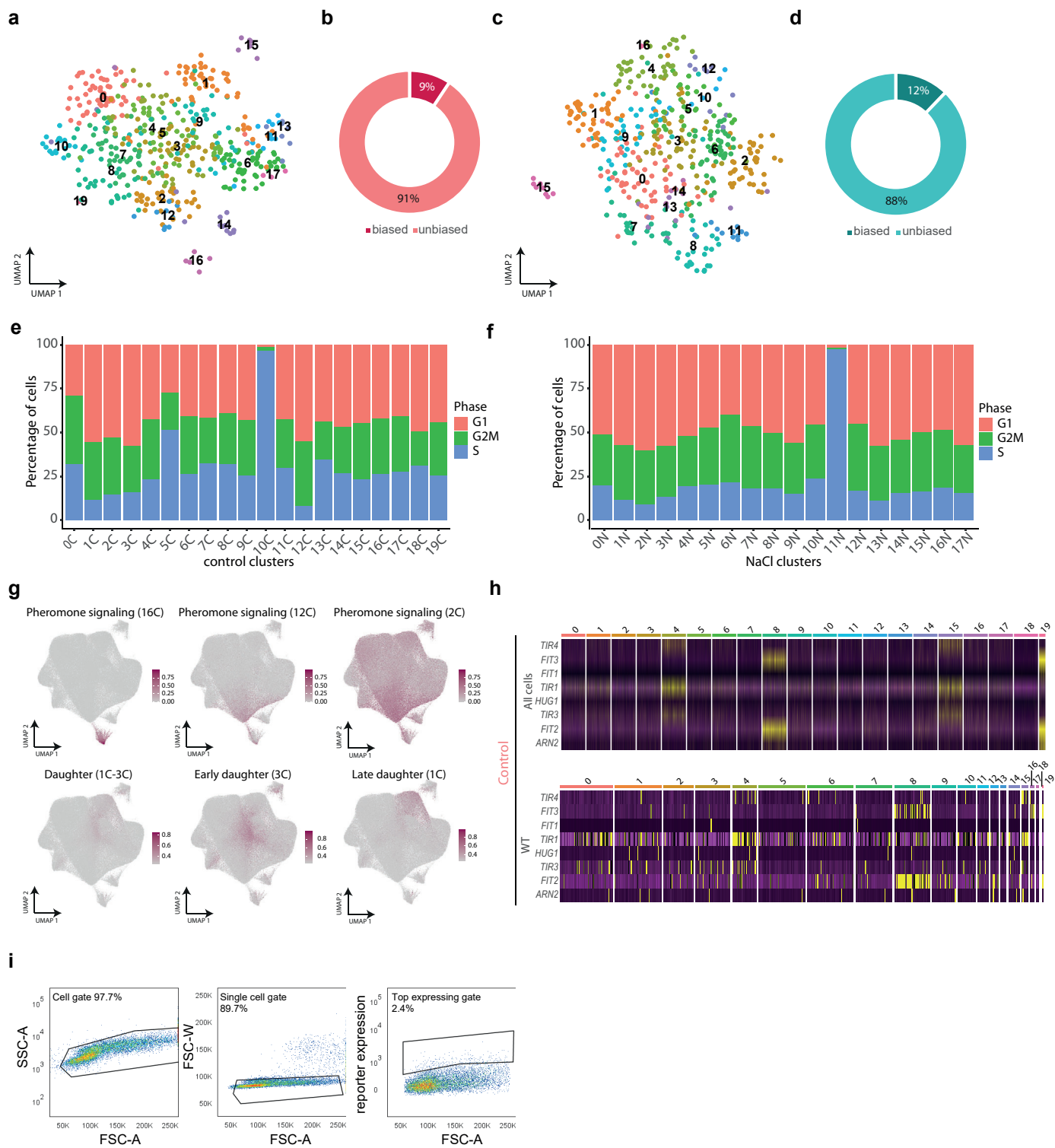
Supplementary Table 2: Yeast strains generated in this study.

Supplementary References.



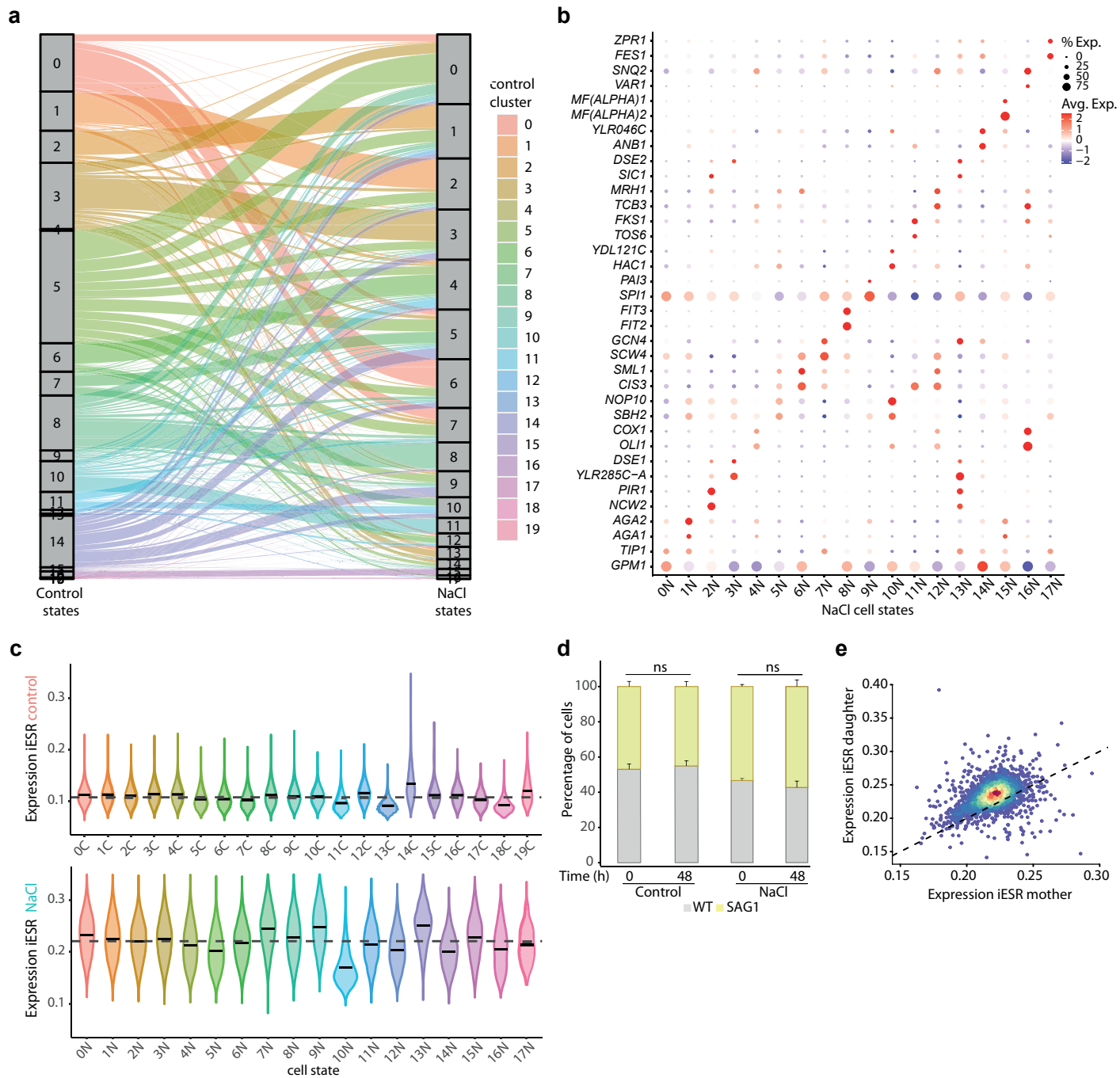
Supplementary Fig. 1: Redesign of the YKO collection with RNA-traceable deletions to enable genome-scale genetic and environmental perturbation screens.

a Schematic representation of the RNA-barcoded structure and a representative Sanger sequencing chromatogram. **b** Overview of the targeted amplification PCR strategy, the *URA3* transcript and the contained barcodes are shown in colors. Rectangle represents the anchoring oligo dT. Arrows indicate the location of primers and dotted line the amplification product. **c** Distribution of cells in the entire dataset after quality check (QC). Plot represents a total of 1.061.865 cells. Colored regions show removed cells (low quality cells or doublet cells, grey) or cells that pass the QC. The total number per each section and kept for further analysis are shown. **d** Histogram the classification of clones based on the number of differentially expressed ($FC > 2$ $pvalue < 0.05$). Only genotypes with at least 10 cells per clone. Pie chart on top, depicts the percentage of clones that express > 3 differentially expressed genes ($n = 365$) (see online Methods). Bars and pie chart are colored according to the number of differential expressed genes and dashed line indicates the threshold. **e-h** Histograms depicting the frequency of number of differentially expressed genes across genotypes in control (red bars) and stress (blue bars). Histograms shows bins of 5 genes and an overflow bin for > 25 genes. Dark colored graphs represent the number of upregulated genes ($FC > 1$ $pvalue < 0.05$). Light graphs represent downregulated genes ($FC > -1$, $pvalue < 0.05$). **i** Scatter plot showing the median (log) number of differentially expressed genes in control conditions for the intersecting mutants from bulk microarray profiling (Kemmeren et al., 2014)¹. **j** Ranked mean iESR expression (y axis) across mutants (x axis) under stress condition. Wild type and reference mutants reported in the literature to display impaired induction of the response are highlighted in red. **k** Principal component analysis of the entire dataset. For each batch and condition the average value for PC1 and PC2 are represented. for each batch (shape) and colored by condition. Batch 11 contained only NaCl sample. **l** UMAP of the entire dataset colored by the indicated cell cycle phase. Cell cycle was determined using Seurat and canonical phase specific genes (see Methods). **m** Gene loadings of PC1 for Fig. 1j. Source data are provided as a Source Data file.



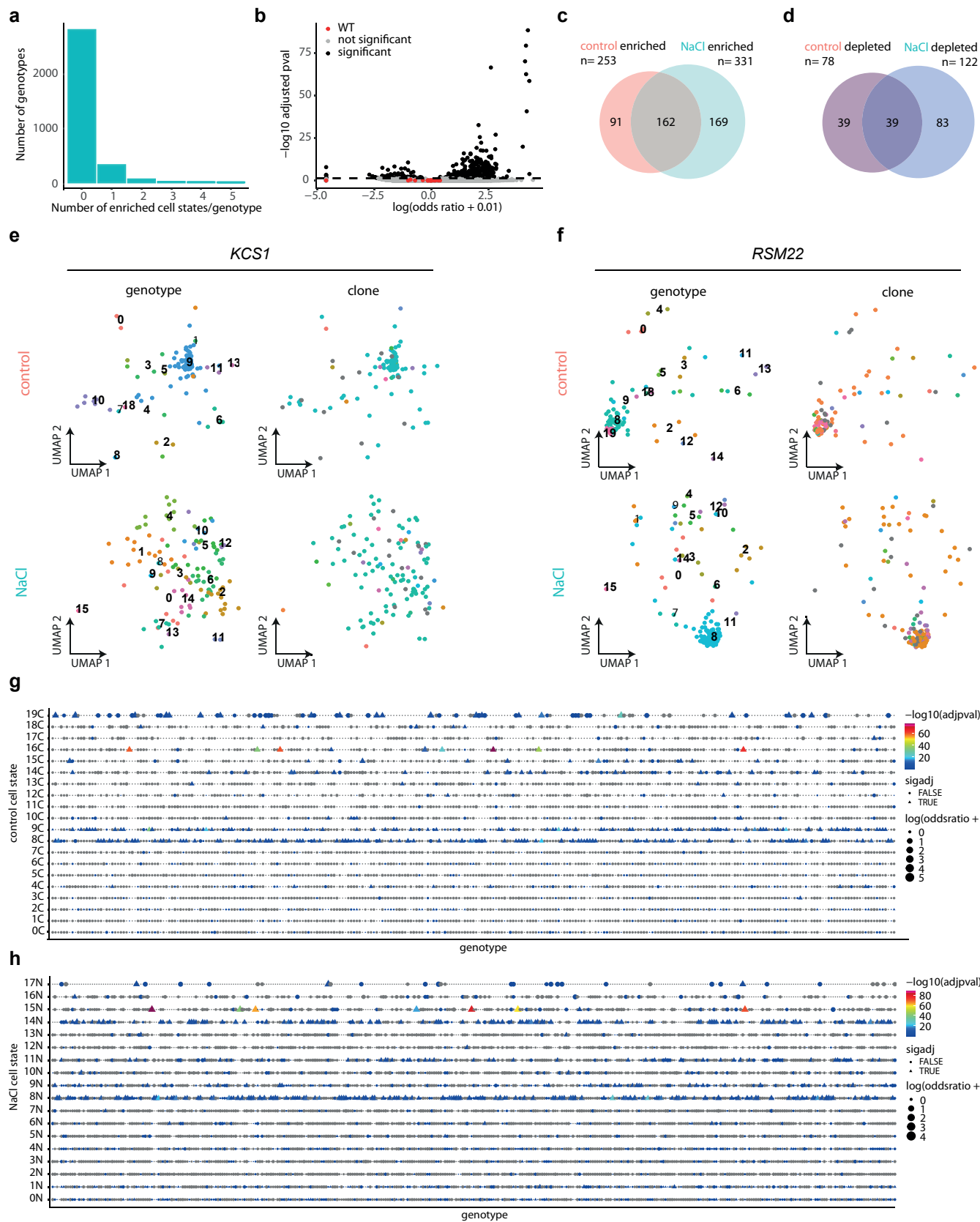
Supplementary Fig. 2: Cells in a population arrange in heterogeneous gene expression states associated with biological function.

a UMAP of wild type cells extracted from the control dataset colored and labeled by cell state. **b** Percentage of mutants in control condition with unbiased cell distribution (light red) or biased distribution (dark red). **c** UMAP of wild type cells as in (A). **d** Percentage of mutants according to their distribution unbiased (light blue) or biased (dark blue). **e-f** Barplot represents the percentage of cells in each cell cycle phase for each cell state for control (e) or NaCl (f) dataset. **g** UMAP representing the expression signature defined by two genes of control clusters related to pheromone signaling (clusters 16C, 12C and 2C, upper panel). Projection of daughter cell state markers clusters (1C and 3C independently and combined). Darker color indicates higher expression. **h** Heatmap represents the expression of the aging signature for a downsampled subset of the entire control dataset (upper panel) and wild type cells only (lower panel). **i** Cells were gated sequentially using FSC-A/SSC-A, FSC-A/FSC-W (for single cell isolation), and reporter fluorescence (mCherry/iRFP) to isolate the top 2% expressing populations. For competition assays and expression analysis, cells were gated using FSC-A/SSC-A and analyzed for gene expression or population of interest (see Methods).



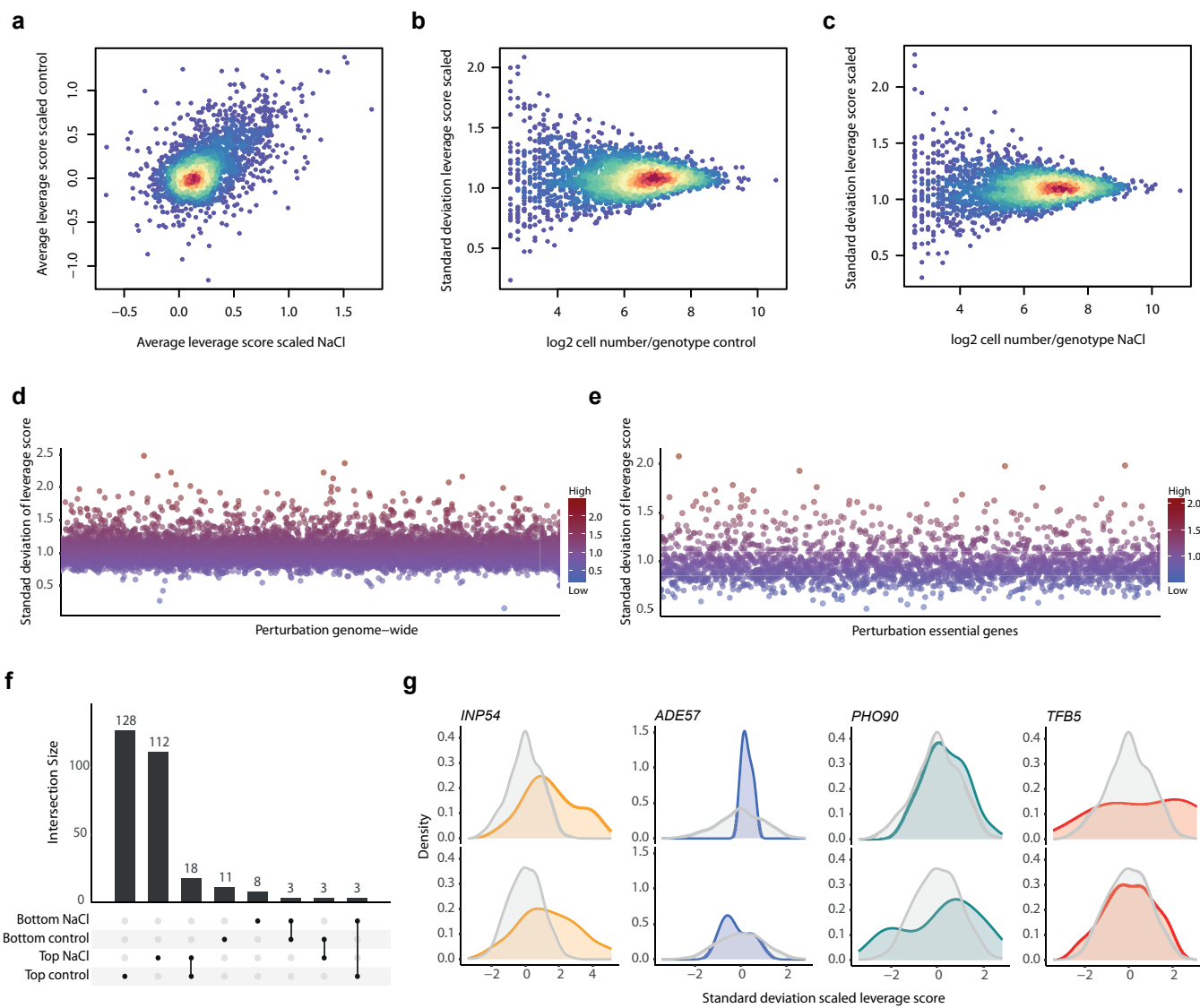
Supplementary Fig. 3: Transcriptome mapping unveils core and responsive cell states that determine fitness.

a The alluvial plot represents percentage of corresponding cells pairs across the indicated clusters and conditions obtained from performing a label transfer. Clusters identified in the control condition (left) were used as a reference to predict cell type labels in the stress (right). The streams are colored based on the control clusters and indicate the connectivity across the indicated clusters (see Supplementary Data 5). **b** Expression of two representative cell state marker genes of the stress dataset. Dot size represent the percentage of expressing cells and are colored from high (red) to low (blue) expression. **c** Expression of the iESR signature for all cell states in each condition compared to the median of population (grey line). **d** Cells with the top 2% *SAG1* expression (p*SAG1*-UbiM-mCherry-t*SAG1*) were sorted and grown in combination with wild type cells (labelled with GFP) in rich media (control) or in the presence of stress (NaCl). Growth of the two populations was assessed at time 0 or after 48 hours by flow cytometry. Data represents mean and standard deviation of three independent experiments. Paired t-test comparing time 0h and 48h of each condition is shown on top (n=3). *p* value (control; 0,47 and NaCl 0.14). **e** Scatter plot shows the expression of the iESR signature (n=175 genes) per each genotype as a function of generation (mother cells, x axis and daughter cells y axis). Each point represents a genotype and points are colored by density (warmer colors indicate higher density). Source data are provided as a Source Data file.



Supplementary Fig. 4: Single-cell genotype-to-phenotype analysis highlights regulators of yeast cell states.

a Distribution of the number of significant cell states per genotype in stress conditions. **b** Volcano plot shows the cell state enrichment of mutants (wild type is shown in red) in control conditions. Black line and dots show the threshold for statistical significance (two-sided Fisher test Benjamini-Hochberg adjusted *pvalue* <0.05). **c-d** Overlap between cell state enriched (c) or depleted (d) genotypes across conditions. **e-f** UMAP representation of genetically determined cell state for the indicated mutants across conditions (top/bottom panels). Points are colored and labeled based on the Seurat cluster (left panel) or based on their clone identity (grey points represent unassigned clones). **g-h** Phenomap shows the top500 genotypes with biased cell state enrichment (x axis) as a function of cell state (y axis) in control (g) and stress (h). Shape indicates statistical significance and are colors indicate the odds ratio value. Source data are provided as a Source Data file.



Supplementary Fig. 5: Different cellular functions drive transcriptional heterogeneity under control and stress conditions.

a Correlation of the average leverage score per each genotype (blue dots) in control (x axis) and stress (y axis) conditions. **b-c** Distribution of standard deviation of scaled leverage score (y axis) against the number of cells per genotype in control (b) and NaCl (c) datasets. Point density is shown in warmer colors (high) and colder colors (low). **d-e** Distribution of the standard deviation of scaled leverage score (y axis) in human Perturb-seq in K562 genome-scale (d) and essential gene screens (e) (Replogle et al., 2022)². **f** Upset plot shows the overlap between negative and positive regulators identified in each condition. Bars are ordered in descending order in each indicated category and the total number is shown above. **g** Representative distributions of transcriptional heterogeneity (leverage score) patterns for the indicated mutants (colored distributions). The background filled distribution in grey is shown for the wild type strain in each condition. Source data are provided as a Source Data file.

Supplementary Table 1. List of plasmids used for experimental validation.

Plasmid num	Plasmid Name	Description	Source
pDC198	pYTK-Spect- <i>LEU2</i> _int- <i>HIS3</i> (pYTK147)	backbone for integrative plasmids	Canadaell et al. ³
pNO82	pYTK-Nat-HO_int-Nat (pYTK168)	backbone for integrative plasmids	This study
pRP093	pYTK147- <i>pFIT3</i> -UbiM-mCherry 3b- <i>tFIT3</i>	Integrating plasmid carrying the <i>FIT3</i> reporter consisting of the promoter of <i>FIT3</i> , an N-terminal degradation signal fused to mCherry and terminator <i>FIT3</i> .	This study
pRP108	pYTK147-pHUG1 dBsaI-UbiM-mCherry 3b-tHUG1	Integrating plasmid carrying the <i>HUG1</i> reporter consisting of the promoter of <i>HUG1</i> , an N-terminal degradation signal fused to mCherry and terminator <i>HUG1</i> . dBsaI indicates the internal BsaI site was mutated to enable correct assembly.	This study
pRP128	pYTK168-pHSP12-UbiM-iRFP 3b-tHSP12	Integrating plasmid carrying the <i>HSP12</i> reporter consisting of the promoter of <i>HSP12</i> , an N-terminal degradation signal fused to iRFP and terminator <i>HSP12</i> .	This study
pDC448	pYTK144-TEF1i-YmukG1-tTDH1	Integrating plasmid carrying the <i>TEF1</i> reporter consisting of the promoter of <i>TEF1</i> , an N-terminal degradation signal fused to YmukG1 and terminator <i>TDH1</i> .	This study
pRP102	pYTK147-pSAG1-UbiM-mCherry 3b-tSAG1	Integrating plasmid carrying the <i>SAG1</i> reporter consisting of the promoter of <i>SAG1</i> , an N-terminal degradation signal fused to mCherry and terminator <i>SAG1</i> .	This study

Supplementary Table 2. List of strains generated in this study for experimental validation.

Strain name	Genotype	Source
yRP148	BY4741-pYTK147-pFIT3-UbiM-mCherry 3b- tFIT3:HIS	This study
yRP171	BY4741-pYTK147-pHUG1 dBsaI-UbiM-mCherry 3b- tHUG1:HIS	This study
yRP178	BY4741-pYTK144-TEF1i-YmukG1-tTDH1:LEU	This study
yRP193	BY4741-pYTK168-pHSP12-UbiM-IRFP 3b-tHSP12: NAT	This study
yRP465	BY4741-pYTK147-pSAG1-UbiM-mCherry 3b- tSAG1:HIS	This study
yRP211	MRP7::URA3-pYTK147-pFIT3-UbiM-mCherry-tFIT3:HIS	This study
yRP225	MSS116::URA3-pYTK147-pFIT3-UbiM-mCherry-tFIT3:HIS	This study
yRP231	RIM8::URA3 (4_G)-pYTK147-pFIT3-UbiM-mCherry-tFIT3:HIS	This study
yRP241	SNF7::URA3-pYTK147-pFIT3-UbiM-mCherry-tFIT3:HIS	This study
yRP264	YMN479-pYTK147-pFIT3-UbiM-mCherry-tFIT3 (WT)	This study
NA	SAS4::URA3	This study
NA	SAS5::URA3	This study
NA	YPL216W::URA3	This study

Supplementary References

1. Kemmeren, P. *et al.* Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. *Cell* **157**, 740–752 (2014).
2. Replogle, J. M. *et al.* Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559-2575.e28 (2022).
3. Canadell, D. *et al.* Implementing re-configurable biological computation with distributed multicellular consortia. *Nucleic Acids Res* **1**, 1–18 (2022).