

<https://doi.org/10.1038/s42003-021-02320-w>

OPEN

## A yeast-optimized single-cell transcriptomics platform elucidates how mycophenolic acid and guanine alter global mRNA levels

Guste Urbonaite<sup>1,2,6</sup>, Jimmy Tsz Hang Lee<sup>3,6</sup>, Ping Liu <sup>1,2</sup>, Guillermo E. Parada<sup>3</sup>, Martin Hemberg <sup>3,5</sup>  & Murat Acar <sup>1,2,4</sup> 

Stochastic gene expression leads to inherent variability in expression outcomes even in isogenic single-celled organisms grown in the same environment. The Drop-Seq technology facilitates transcriptomic studies of individual mammalian cells, and it has had transformative effects on the characterization of cell identity and function based on single-cell transcript counts. However, application of this technology to organisms with different cell size and morphology characteristics has been challenging. Here we present yeastDrop-Seq, a yeast-optimized platform for quantifying the number of distinct mRNA molecules in a cell-specific manner in individual yeast cells. Using yeastDrop-Seq, we measured the transcriptomic impact of the lifespan-extending compound mycophenolic acid and its epistatic agent guanine. Each treatment condition had a distinct transcriptomic footprint on isogenic yeast cells as indicated by distinct clustering with clear separations among the different groups. The yeastDrop-Seq platform facilitates transcriptomic profiling of yeast cells for basic science and biotechnology applications.

<sup>1</sup>Systems Biology Institute, Yale University, West Haven, CT, USA. <sup>2</sup>Department of Molecular Cellular and Developmental Biology, Yale University, New Haven, CT, USA. <sup>3</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. <sup>4</sup>Department of Physics, Yale University, New Haven, CT, USA. <sup>5</sup>Present address: Evergrande Center for Immunologic Disease, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA. <sup>6</sup>These authors contributed equally: Guste Urbonaite, Jimmy Tsz Hang Lee. ✉email: [mhemberg@bwh.harvard.edu](mailto:mhemberg@bwh.harvard.edu); [murat.acar@yale.edu](mailto:murat.acar@yale.edu)

In recent years, single-cell RNA sequencing (scRNA-seq) has become a widely used method for studying the transcriptomes of individual cells<sup>1</sup>. scRNA-seq has been used to identify populations of specific cell types within tissues<sup>2</sup>, to study particular cellular pathways<sup>3</sup>, and to better understand disease development<sup>3</sup>. One method to perform scRNA-seq is the Droplet Sequencing (Drop-Seq) platform<sup>2</sup>. Using a microfluidic device, this platform allows for the encapsulation of single cells, along with microbeads carrying barcoded primers and lysis buffer into oil droplets (Fig. 1). Upon cell lysis in the oil droplet, individual mRNA molecules hybridize to the primers on the microbeads. The barcoded primer sequences on each microbead allow for the distinction of mRNAs obtained from different cells and unique mRNAs obtained from the same cell. Next, all droplets are broken together and release their hybridized beads. Isolated beads are reverse-transcribed with template switching, generating cDNA strands. This is followed by the PCR-amplification of the cDNAs and the addition of the sequencing adapters. Finally, the barcoded mRNA samples are sent out for sequencing.

The original Drop-Seq platform has been specifically developed and optimized for mammalian cells with the goal of measuring mRNA counts at the single-cell level across genetic backgrounds and/or growth conditions. Size and structure/morphology differences between different cell types have made it difficult to directly apply the mammalian Drop-Seq platform to other organisms, including yeast which are much smaller than mammalian cells and have a cell wall. While there are well-based<sup>4</sup>, FISH-based<sup>5</sup> and droplet-based<sup>6</sup> (using commercial 10x Genomics platform) methods for measuring single-cell mRNA abundances in yeast cells, a table-top noncommercial Drop-Seq-based platform optimized for robust measurements of mRNA counts in yeast has not been available.

Here we present the yeastDrop-Seq platform, a yeast-optimized table-top scRNA-seq technology based on the original Drop-Seq platform<sup>2</sup>. As a proof-of-principle application of yeastDrop-Seq, we measure how Mycophenolic acid (MPA) and guanine impact mRNA counts globally at the single-cell level. MPA is a lifespan-extending compound that decreases de novo GMP synthesis<sup>7–9</sup>. Our work uncovers the global transcriptomic effects of MPA and

guanine in a pathway-specific manner at single-cell resolution, providing novel insights about how MPA extends the lifespan in yeast.

## Results

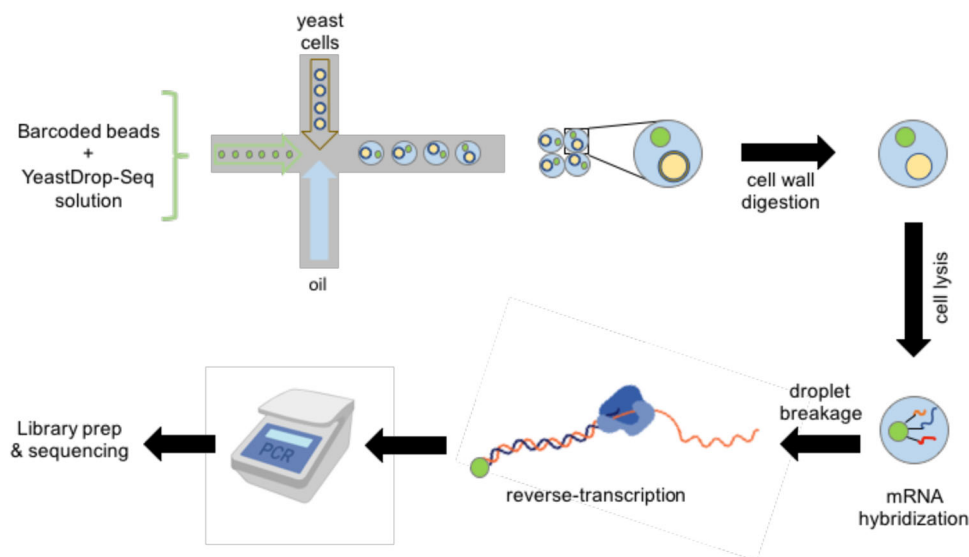
**yeastDrop-Seq as a yeast-optimized scRNA-Seq method.** We modified the original Drop-Seq protocol so that the differences in size and morphology between yeast and mammalian cells are factored in the relevant steps of the protocol.

To make sure that only one cell is present in each oil droplet, cell density at the time of cell-feeding into the microfluidic chip had to be sufficiently low. On the other hand, a density that is too low would lead to unfeasibly long cell-feeding times. We found that a density of 50 cells per microliter addressed these needs.

Unlike mammalian cells, yeast has a cell wall surrounding its outer membrane. To break the cell wall, we designed a solution containing Zymolyase in addition to a cell-lysis reagent and SDS (Methods). Zymolyase is a cell-wall digestion enzyme which does not alter RNA stability or folding<sup>10</sup>. After the initial formation of oil droplets by feeding into the microfluidic chip yeast cells (at the density of 50 cells per microliter) together with the microbeads resuspended in this solution, the oil droplets were incubated for 30 min at 30 °C to ensure that Zymolyase breaks the cell walls and then cell lysis occurs (Supplementary Fig. 1). After this incubation period, we evaluated oil-droplet quality and found that more than 95% of the droplets remained intact.

Next, oil droplets were broken, and the original Drop-Seq protocol<sup>2</sup> steps were applied to the pooled mRNA molecules hybridized to the primers on the microbeads, including reverse transcription, exonuclease I treatment, PCR-amplification of the cDNAs and the addition of the sequencing adapters. Sequencing was performed using the Illumina HiSeq2500 platform with 2 × 100 nts read pairs.

**Measuring the single-cell transcriptomic impact of MPA and guanine.** In our previous studies, we have elucidated that Mycophenolic acid (MPA) extended lifespan in yeast through GMP synthesis inhibition<sup>8,9</sup>. We further found that the longevity



**Fig. 1 A schematic of the single-cell mRNA sequencing workflow for transcriptomic analysis of yeast cells.** The yeastDrop-Seq solution containing chemicals for cell-wall digestion and cell lysis is mixed with barcoded microbeads, and the mixture is flown through one inlet of a microfluidic device. Yeast cells with intact cell walls and oil are flown through the second and third inlet of the microfluidic device, respectively. Encapsulated cells are then incubated to allow for cell-wall breakage and lysis. mRNA molecules hybridize to the barcoded microbeads via their polyA tails, droplets are broken, and reverse transcription occurs. cDNA is PCR-amplified and the cDNA libraries are prepared and submitted for sequencing.

effect of MPA was prevented by the supplementation of exogenous guanine, MPA's epistatic agent in GMP synthesis. MPA targets GMP synthesis by inhibiting IMP Dehydrogenase (IMD), decreasing de novo GMP synthesis<sup>7–9</sup>. IMD catalyzes the production of xanthosine 5'-phosphate (XMP) from inosine 5'-phosphate (IMP). This is followed by the production of GMP from XMP through the action of the GMP synthase Gua1. The guanylate kinase Guk1 converts GMP to GDP<sup>7</sup>. Despite the biochemical knowledge about MPA's involvement in the GMP synthesis pathway, how it affects single-cell transcriptomes downstream of GMP/GDP/GTP synthesis has been unknown.

To study the global impact of MPA and guanine on single-cell transcriptomes and to identify the specific genes that are up/downregulated upon MPA and/or guanine treatment, we grew yeast cells for 18 h in media containing DMSO (D), MPA (M), Guanine (G), or MPA + Guanine (MG) (Methods), subjected the cells to yeastDrop-Seq, and processed the raw mRNA sequencing data to obtain single-cell level mRNA abundances.

The processing of the raw data included trimming low quality reads in sequences (Supplementary Fig. 2), mapping the resulting reads to the *S. cerevisiae* genome, and the construction of [gene × cell] expression matrices for each treatment condition. We next performed quality control of the processed data to filter out low quality cells and features in order to improve the signal-to-noise ratio for downstream analyses. For this, thresholds were adjusted according to the total Unique Molecular Identifier (UMI) counts and total gene counts (Supplementary Fig. 3); cells whose total UMI or feature counts fell outside of the normal distribution were excluded.

**Analysis of gene expression datasets for the rate of cell-doublet formation.** An essential aspect of a droplet-based single-cell RNA measurement platform is its ability to encapsulate only one cell in each oil droplet or a low rate for doublet formation. To quantitatively characterize the doublet rate associated with yeastDrop-Seq, we applied five methods: DoubletFinder, Scrublet, DoubletDecon, scds, and solo (Supplementary Fig. 4). These methods first introduce artificial doublets based on the original dataset, followed by the training of a classifier to distinguish singlets from doublets. The methods then provide a ranking/score for the cells from the original dataset, and a threshold for doublet identification is defined.

While DoubletDecon predicted a surprisingly high doublet rate, which we attribute to false positives, the other four methods predicted low (~2–7%) rates for potential doublets for experiments run at 50 cells/uL (Methods). Also, the number of conserved doublets (i.e., doublets with the same barcode reported by more than one method) identified by the four methods was 12 out of 844 (Supplementary Fig. 4). We conclude that the cell-doublet rate of the yeastDrop-Seq platform is reasonably low. As a comparison, the original Drop-Seq platform developed for mammalian cells reported cell-doublet rates up to 11.3% at cell concentrations of 100 cells/uL and less than 5% at cell concentrations of 50 cells/uL.

**Characterization of sequencing reads, and transcript and gene counts.** We analyzed the filtered sequencing reads to obtain the condition-specific distributions of their GC contents and read lengths. We found that the distribution of GC contents of all valid reads in all valid cells peaks at around 23% for each of the four growth conditions, which matched the peak expected based on theoretical grounds (Supplementary Fig. 5a). From the distributions of read lengths, we found that the peak corresponded to 52–53 bp consistently for each condition (Supplementary Fig. 5b). The distribution of UMI counts showed that there was no bias toward transcript length (Supplementary Fig. 6a) and GC percentage (Supplementary Fig. 6b).

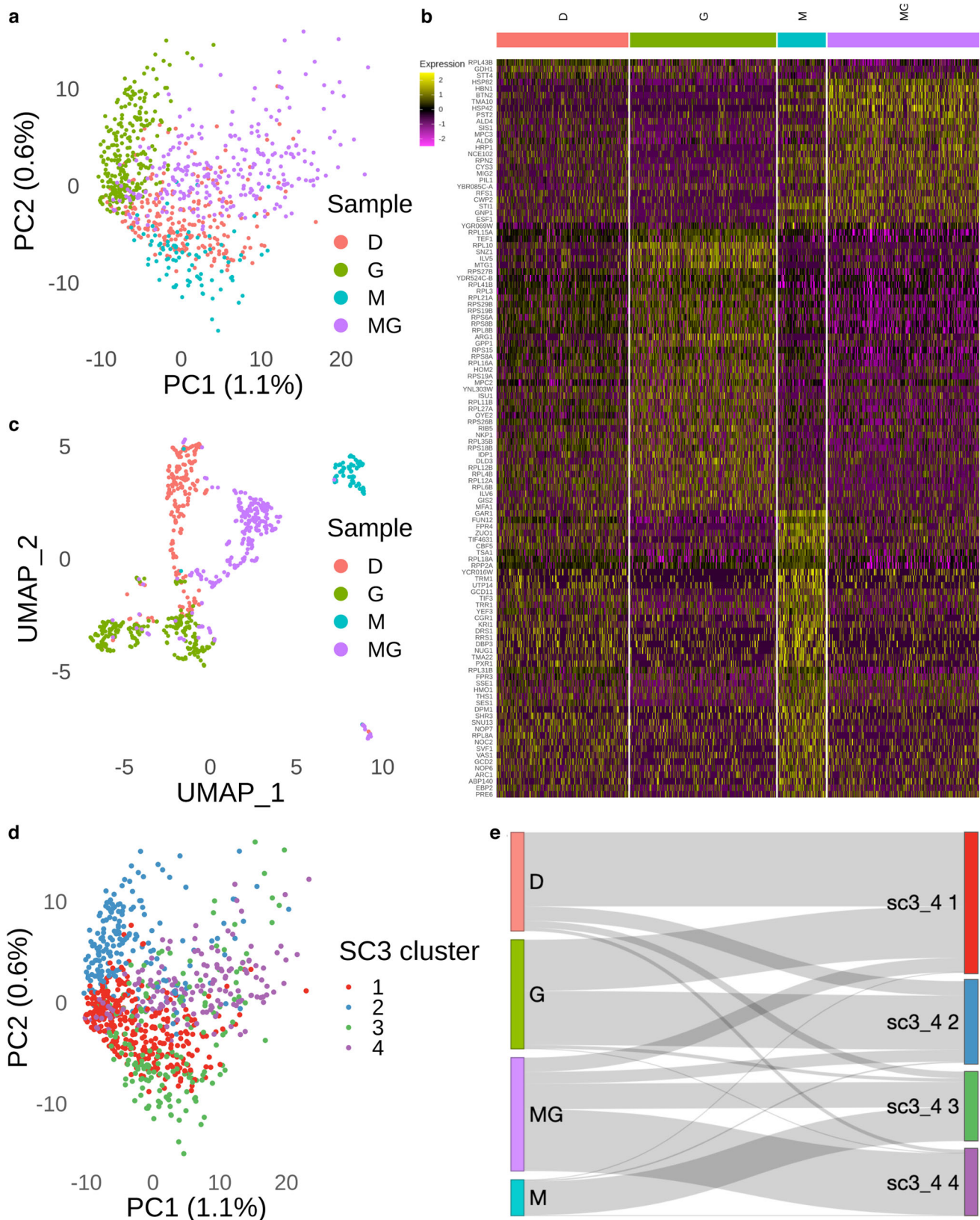
We next quantified the number of distinct mRNA transcripts, and identified the genes they corresponded to. First, analyzing our single-cell mRNA sequencing data for effective reads from valid barcodes (Supplementary Fig. 3) (Methods), we saw that the average number of effective reads from valid barcodes varied between 49.3% and 65.7% across the four growth conditions used. We then obtained the distribution of the number of distinct mRNA transcripts per cell for each treatment condition (Supplementary Fig. 3), and calculated the average number of distinct mRNA molecules (Supplementary Data 1). With the total number of cells identified in each condition ranging from 85 to 268, the mean transcript counts per cell ranged from 840.19 to 1335.62 across the four treatment conditions. Coefficients of variation (CV) were calculated across the cells in each treatment condition (Supplementary Data 2). We saw that the average number of distinct genes across cells of a sample, with the number of distinct genes quantifying the number of genes captured per cell in each sample, were 443 (G), 554 (M), 578 (MG), and 619 (D) (Supplementary Data 1). Since *S. cerevisiae* has 6275 genes, and since the set of distinct genes captured in each cell is different, the fraction of overall distinct genes captured by yeastDrop-Seq across all cells were 38.6% (M), 46.5% (G), 55.7% (D), and 62.2% (MG).

We also sorted our gene expression matrices for each gene represented across cells in each growth condition. Supplementary Data 2 shows the average single-cell mRNA transcript numbers averaged across all cells of each treatment condition, as well as the standard deviation (SD). To quantify transcriptional noise in the expression of each gene, we used CV as a metric and saw similar dependencies between CV and mean across the four treatment conditions. A second-degree polynomial function fitted the distribution of single-cell CV vs. mean transcript numbers pooled from all genes represented for each condition (Supplementary Fig. 7). As expected, genes with low transcript numbers were associated with high CV values.

**Isogenic yeast populations display distinct transcriptomic substructures.** To perform downstream analysis, we merged the data from the 4 samples, corresponding to the 4 growth conditions, using the 'merge' function in Seurat v3<sup>11</sup>, followed by normalization of the count matrices of the 4 samples. We performed Principal Component Analysis (PCA) on the merged gene expression data corresponding to the four treatment conditions (Fig. 2a). Each condition led to distinct clustering with clear separations among the different groups, which indicates that each treatment condition has a distinct transcriptomic footprint (Fig. 2b, Supplementary Data 3). The nonlinear dimension reduction performed through Uniform Manifold Approximation and Projection (UMAP) also showed consistent results (Fig. 2c). Despite the use of isogenic yeast cells across our experiments, we identified subclusters in the UMAP plot corresponding to transcriptomic heterogeneity within the DMSO, Guanine and MPA + Guanine-treated yeast populations; differently from these observations, the MPA treatment led to a relatively homogeneous population (Fig. 2c).

Although the PCA plot (Fig. 2a) clearly showed the presence of 4 clusters corresponding to the 4 treatment conditions, we wanted to obtain the best quantitative clustering outcome corresponding to our merged data. For this, we ran the single-cell consensus clustering (SC3) method which provides robust and accurate cell clustering as well as downstream analysis for single-cell RNA sequencing data<sup>12</sup>. We found that SC3 with  $k = 4$  provided the best outcome for our combined data (Fig. 2d).

Using a Sankey diagram, we traced how results from each treatment condition were mapped to the four SC3 clusters



**Fig. 2 Identification of DE genes in the 4 samples.** **a** PCA plot illustrating the distribution of the 4 samples (D  $n = 233$ , G  $n = 258$ , M  $n = 85$ , MG  $n = 268$  cells). **b** Heatmap illustrating the DE genes corresponding to the 4 samples (Supplementary Data 3). The color represents the normalized expression level in natural-log scale of the corresponding genes of a cell (yellow is high). **c** UMAP plot of the 4 samples (D  $n = 233$ , G  $n = 258$ , M  $n = 85$ , MG  $n = 268$  cells) based on the PC spaces in panel **a**. **d** Semi-supervised clustering identifies 4 clusters across the 4 samples. **e** Sankey plot illustrating the 4 samples mapping to the 4 SC3 clusters (cluster 1  $n = 331$ , cluster 2  $n = 193$ , cluster 3  $n = 164$ , cluster 4  $n = 156$  cells). DE genes means differentially expressed genes.

(Fig. 2e). As expected due to the low population heterogeneity observed, almost all MPA-treated cells were mapped to a single SC3 cluster (SC3\_4 3). On the other hand, most cells treated with DMSO or Guanine were mapped to three or four SC3 clusters.

To quantitatively evaluate the subpopulations associated with each treatment condition, we performed SC3 clustering analysis on each treatment-specific gene expression matrix separately. For the DMSO treatment, we found that  $k=2$  gave a reasonable resolution to distinguish two subclusters containing 177 and 56 cells (Fig. 3a). An area-under-the-receiver-operating-characteristic (AUROC) curve analysis identified the list of marker genes corresponding to each subcluster from the DMSO population (Supplementary Data 4). For the Guanine-treated cells, we again found two distinct subclusters represented by 135 and 123 cells (Fig. 3b). An AUROC curve analysis identified the list of marker genes corresponding to these subclusters (Supplementary Data 4). As expected for the MPA-treated cells, the SC3 analysis yielded a similar result: the apparent lack of heterogeneity prevented SC3 from resolving the population into more than one subcluster (Fig. 3c). Finally, for the cells treated with MPA and Guanine, there were 3 distinct subclusters (with 166, 77, 25 cells) and the result from consensus clustering showed that  $k=3$  was ideal (Fig. 3d), even though there were relatively small numbers of cells in the third subcluster. The small cluster was supported by the marker genes analysis which revealed a strong signal for the third subcluster (Supplementary Data 4).

Relabeling the cells with respect to the SC3 clustering results obtained from the analysis on each of the four treatment conditions, one can see the eight distinct subclusters under different plotting schemes (Fig. 3e, f, Supplementary Data 5). Finally, comparing these eight subclusters identified separately with the results obtained from a new SC3 clustering analysis (with  $k=8$ ) performed on the full data merged from all four treatment conditions, we found that a large fraction of cells could be consistently mapped between single subcluster counterparts (Fig. 3g).

**Identification of differentially expressed genes across treatment conditions and clusters.** We performed the Wilcoxon test to identify the differentially expressed genes. Using the thresholds of 50% of minimum percentage of cells in each condition/cluster and  $\pm 0.5$  average log-fold change (with adjusted  $p$ -value  $< 0.05$ ), we obtained a list of genes up- or downregulated for each treatment condition and cluster (Supplementary Data 6). Visual inspection of the heatmaps suggested that these genes could well represent each condition or distinct cluster (Fig. 3f). We also identified the gene candidates representing each subcluster in each treatment condition (Fig. 4, Supplementary Data 7). Finally, we obtained the set of differentially expressed genes for the clusters identified by SC3 with  $k=4$  (Fig. 5, Supplementary Data 8).

Performing a GO term analysis using the R package clusterProfiler<sup>13</sup> we also found the pathways that are up- and downregulated in each treatment condition or treatment clusters. In Supplementary Data 9, we show the up- or downregulated pathways for each treatment condition. Delving deeper into the individual clusters for each treatment condition below we again identify up- or downregulated pathways or biological processes using the GO term analysis (Supplementary Data 10).

**Analysis of selected differentially expressed genes and biological processes in MPA-treated cells.** We saw an upregulation of mRNAs for proteins involved in rRNA pre-processing in cells treated with MPA (*GARI*, *UTP14*, *CGR1*, *DBP3*, and *PXR1*) (Fig. 4). However, it is known that MPA decreases ribosome biogenesis as well as rRNA levels, by decreasing RNA pol I and RNA pol III activity<sup>14,15</sup>. It has been suggested that this rRNA

decrease leads to the accumulation of ribosomal proteins (r-proteins) in the cell<sup>15</sup>. We hypothesize that cells upregulate the processome machinery, transcribed by RNA pol II, in an attempt to produce more mature rRNA, as there is a lack of such rRNA upon treatment with MPA.

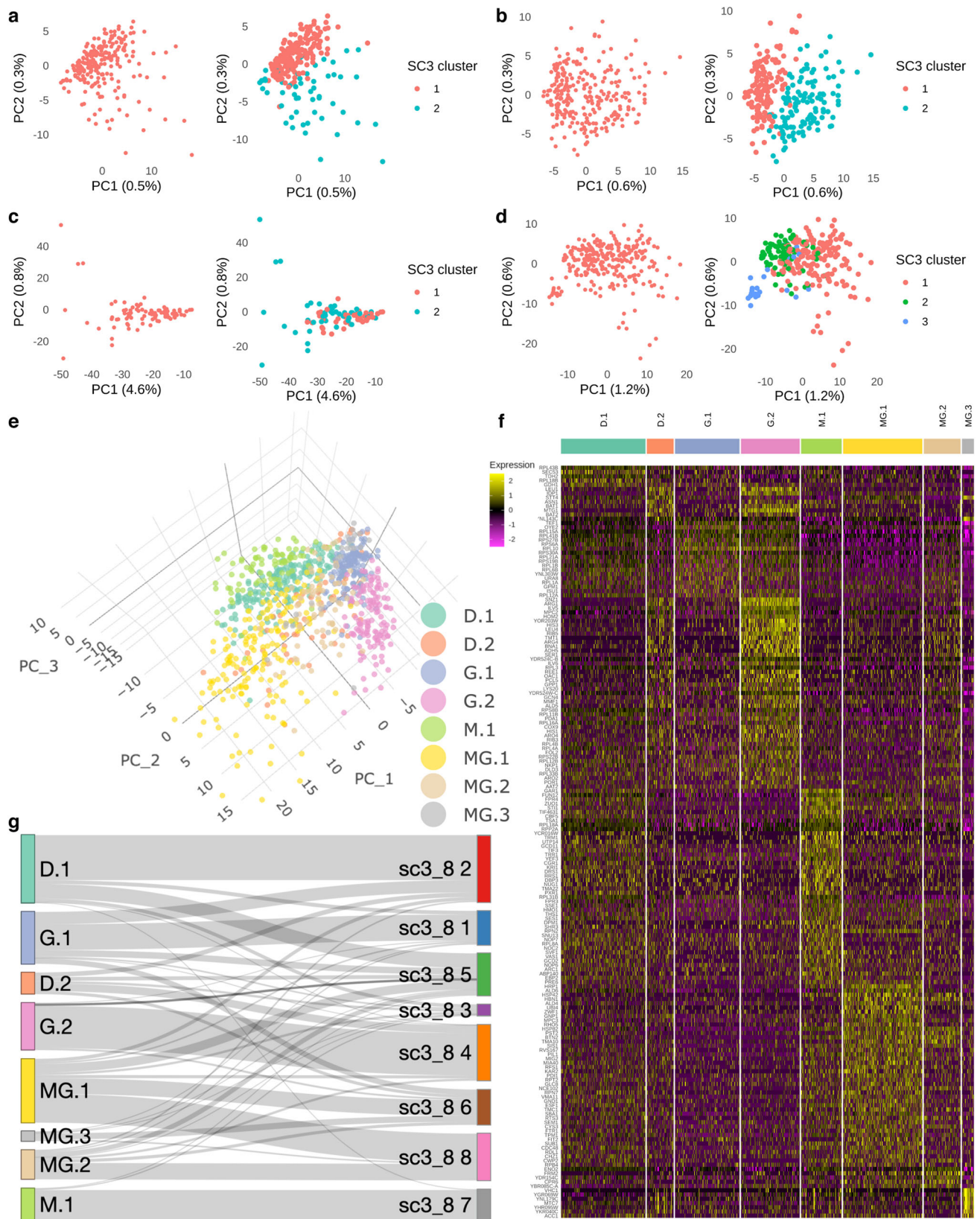
There is also an upregulation of tRNA synthetase and tRNA methyltransferase genes (*TRM1*, *SES1*, *ABP140*, *THS1*, *VAS1*) (Fig. 4). The decreased RNA PolI and PolIII activity due to MPA treatment has been shown to result in a decreased concentration of mature tRNAs<sup>16</sup>. We conjecture that MPA-treated cells attempt to replenish this population of tRNAs by upregulating genes, transcribed by RNA PolII, that are involved in their production. Histone and nucleosome assembly genes are also upregulated in cells treated with MPA (*FPR4*, *FPR3*) (Fig. 4). It is known that MPA acts as a transcription elongation repressor<sup>17</sup>, and we conjecture that MPA represses transcription by leading to a more packed chromatin structure. It is interesting to note that *FPR3* and *FPR4* are involved with rDNA, whose replication and recombination has been implicated in aging<sup>18</sup>. Additionally, translation initiation factor gene expression is upregulated, including *TIF3*, *GCD11*, *TIF4631*, *FUN12*, and *GCD2* (Fig. 4). We hypothesize that, because MPA treatment was previously reported to lead to a decrease in guanosine nucleotides<sup>19</sup>, protein production is decreased (rRNA and mRNA expression decreases) and cells attempt to rescue this phenotype by increasing translation initiation. Moreover, it was found that some components of the SAGA complex, which is involved in histone acetylation, translation initiation, and elongation, are also affected in cells treated with MPA, suggesting that MPA has an overall effect on the translation process<sup>20</sup>. MPA-treated cells also display an upregulation of antioxidant genes such as *TSA1* and *TRR1*. Cells treated with MPA appear to upregulate genes that provide protection from oxidative damage. This antioxidant effect of MPA has also been reported in mouse models previously<sup>21</sup>, and here we identify which yeast genes are involved in this process. Here we also observe that cells treated with MPA upregulate the expression of chaperone proteins (*STI1*, *SSE1*) (Fig. 4) that are involved in regulation of organization of amyloid-like proteins or unfolded proteins in general. It is interesting to note that cells treated with MPA have increased protein burden<sup>22</sup>. It is likely that upregulation of *STI1* and *SSE1* is a direct cause of this increased protein burden.

In MPA-treated cells, we saw an upregulation of genes, including *NUG1*, *NOC2*, and *RRS1* (Fig. 4), involved in ribosome component transport. It is interesting to note that cells treated with MPA also exhibited altered expression of 40 S and 60 S ribosomal subunits. Some subunit components are downregulated (*RPLA1A*, *RPLA1B*, and *RPS29B*) (Fig. 4), while others are upregulated (*RPL18A*). There is a delicate balance among the r-proteins forming the various subunits of the ribosome in the cell and it has been suggested that during ribosomal stress there can be an accumulation of free r-proteins not assembling in a ribosome<sup>15</sup>.

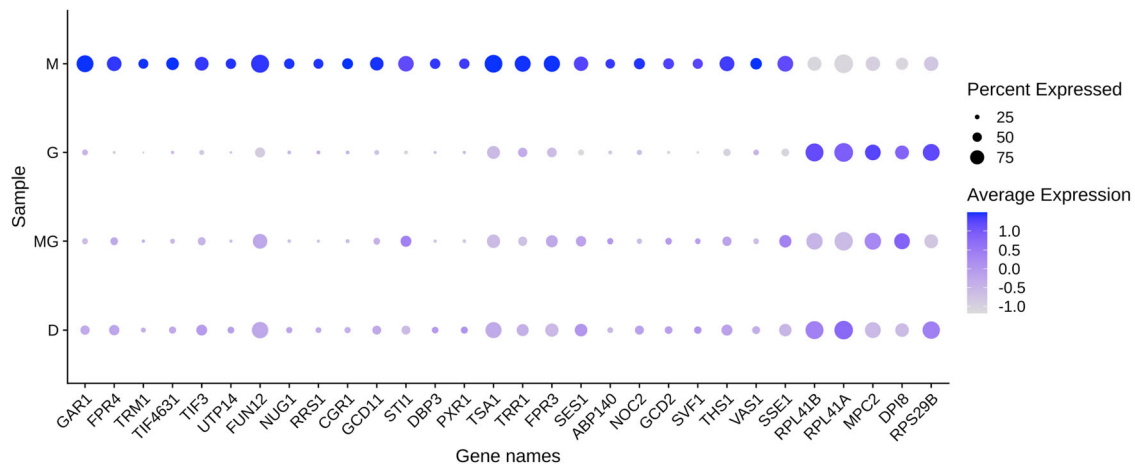
Additionally, cells treated with MPA downregulate mitochondrial components (*MPC2*, *DPI8*) (Fig. 4) and most likely mitochondrial function. It is known that mitochondrial membrane potential<sup>23</sup> and mitochondrial composition<sup>19</sup> are impacted by MPA treatment, so these downregulations upon MPA treatment, perhaps indicating a decrease in mitochondrial function, are not surprising.

Based on the GO terms analysis (Supplementary Data 9–10), there was an upregulation of ncRNA metabolic processes, involving tRNA aminoacylation for protein translation, as well as an upregulation of cellular component biogenesis involving proteasome components and ribosome assembly, such as amino acid activation, peptide biosynthetic process, translational initiation.

GO term analysis (Supplementary Data 9–10) also revealed several downregulated cellular processes including ribosomal



**Fig. 3** SC3 consensus clustering reveals the heterogeneity within yeast cells in the 4 samples. **a** PCA plot showing the 2 SC3 clusters in DMSO ( $n = 233$  cells). **b** PCA plot showing the 2 SC3 clusters in Guanine ( $n = 258$  cells). **c** PCA plot showing the 2 SC3 clusters in MPA ( $n = 85$  cells). **d** PCA plot showing the 3 SC3 clusters in MPA + Guanine (MG  $n = 268$  cells). **e** A 3-dimension PCA plot illustrating the distribution of the 8 subclusters (D.1  $n = 177$ , D.2  $n = 56$ , G.1  $n = 135$ , G.2  $n = 123$ , M.1  $n = 85$ , MG.1  $n = 166$ , MG.2  $n = 77$ , MG.3  $n = 25$  cells) in the 4 samples. **f** Heatmap illustrating the DE genes corresponding to the 8 subclusters identified by SC3 (Supplementary Data 5). The color represents the normalized expression level in natural-log scale of the corresponding genes of a cell (yellow is high). **g** Sankey plot illustrating the subclusters identified in separated analysis mapping to the 8 SC3 clusters identified in group analysis. DE genes means differentially expressed genes.



**Fig. 4 DE genes in the 4 samples.** Dotplot showing the key DE genes (Methods) identified in the 4 samples (D  $n = 233$ , G  $n = 258$ , M  $n = 85$ , MG  $n = 268$  cells). The size of the dot represents the percentage of cells within a sample expressing the corresponding genes, while the color represents the average expression level in natural-log scale of the corresponding genes across all cells within a sample (blue is high). DE genes means differentially expressed genes.

RNA transport, ribosome assembly and ribosomal RNA metabolic process. In addition, mitochondrial function is also signified by a downregulation in pyruvate metabolic process. Additionally, phosphatidylinositol-mediated signaling was downregulated in MPA-treated cells. This result also made sense based on the previous finding that MPA alters cholesterol and phosphatidylcholine concentrations which greatly impacts lipid-mediated signaling in intestinal cells<sup>19</sup>. Regulation of cell wall organization was also downregulated in MPA-treated cells. Finally, we observed a downregulation in the monocarboxylic acid metabolic process in cells treated with MPA. As MPA itself is a monocarboxylic acid and perturbs many cellular pathways, it is reasonable to expect that cells attempt to downregulate genes that promote its transport.

## Discussion

In this study, we introduce a yeast-optimized table-top scRNA-seq platform for measuring single-cell level mRNA counts in yeast. The yeastDrop-Seq platform is applied to cells treated with Mycophenolic acid (MPA) and/or guanine, and the resulting changes in global mRNA expression levels were quantified. We elucidate heterogeneous gene expression profiles in cells as well as how the expression profiles are clustered in isogenic populations (Fig. 3). We further uncover differentially up- or downregulated groups of genes and pathways as a result of the growth conditions we used (Figs. 3–5). Our table-top scRNA-seq technology is similar in its function to a recently-published<sup>6</sup> scRNA-seq technology that uses the commercial 10x Genomics platform; both technologies are droplet-based, involve microfluidics and adapted for yeast cells.

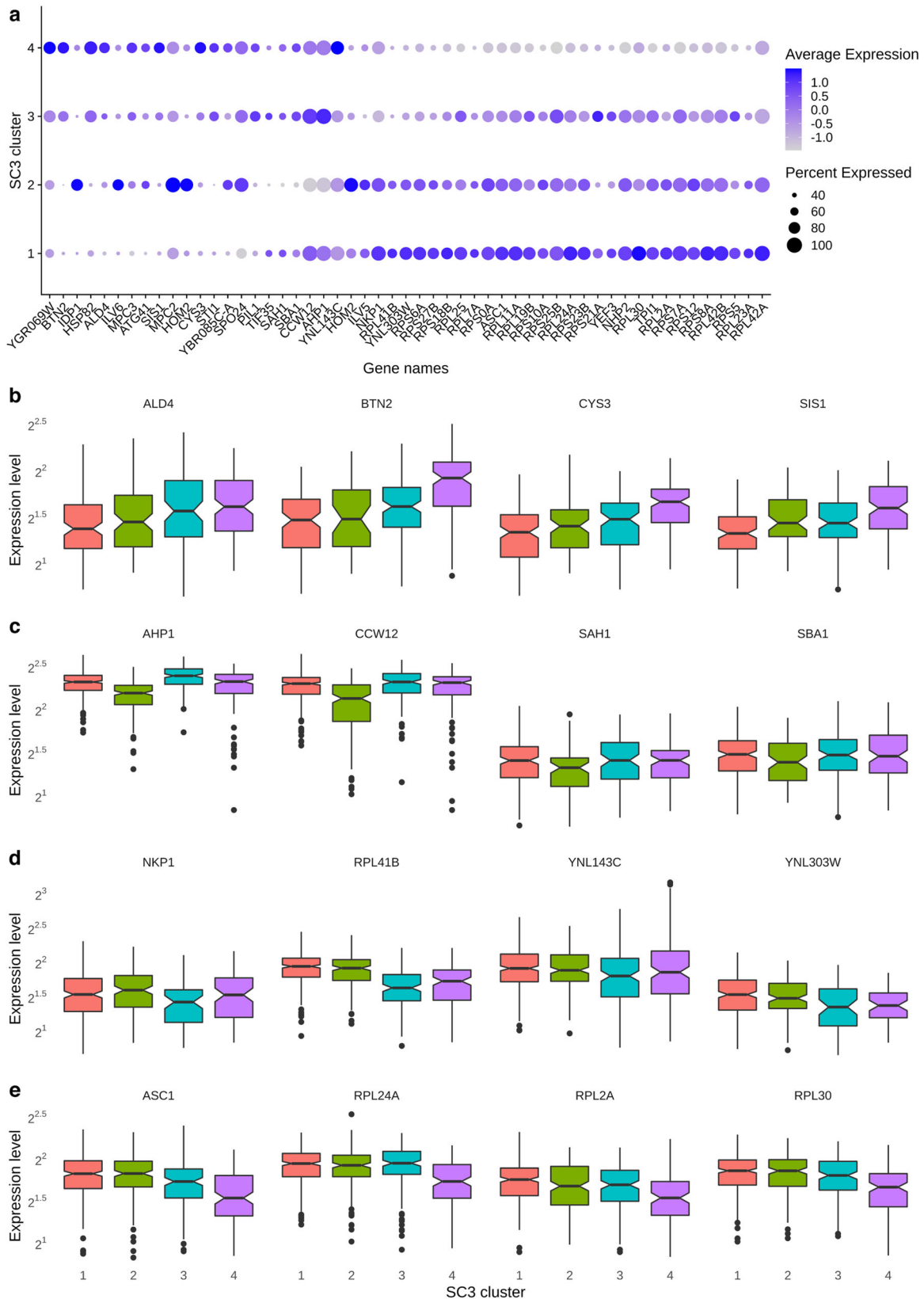
One major parameter to optimize for an effective yeastDrop-Seq is doublet formation rate, owing to the smaller size of *S. cerevisiae* cells (~4–5  $\mu\text{m}$  in diameter) compared to mammalian cells (~10–100  $\mu\text{m}$ ). Although doublet-detection tools tailored for yeast cells are yet not available, we assume transcription doublet in yeast cells share similar nature as that in mammalian cells. In this study, we applied five doublet-detection tools which are based on different algorithms to identify doublets in our dataset to estimate the doublet rate (Supplementary Fig. 4). Out of all tools, DoubletFinder<sup>24</sup>, scds<sup>25</sup>, Scrublet<sup>26</sup>, and solo<sup>27</sup> have detected an average of 3.5% doublet rate and 2% of conserved doublets, doublets identified by more than one method used. Surprisingly, DoubletDecon<sup>28</sup> has estimated an exceptionally high

doublet rate with only 0.01% of the detected doublets is in agreement with the conserved doublets. This result could be explained by the benchmark experiment against DoubletFinder, Scrublet and solo; DoubletDecon tends to have higher sensitivity and lower specificity<sup>28,29</sup>. Therefore, we considered this inconsistent result as an outlier.

In our computational analysis, we found that the expression data from the cells grown in the 4 conditions could be merged without batch correction, suggesting that our protocol is highly reproducible with consistency. The principal component analysis (Fig. 2a) showed that the 4 samples are separated into 4 distinct clusters suggesting that each treatment condition had a distinct transcriptomic footprint. This result is further supported by our clustering results (Fig. 2d) as we could observe a similar cluster distribution over the 4 samples. Although we saw 4 distinct clusters in our UMAP analysis, we also observed subclusters (Fig. 2c) among the cells treated with DMSO, Guanine and MPA + Guanine, suggesting transcriptomic heterogeneity in these samples. This was an interesting observation as we used isogenic yeast cells across our experiments. To identify the subclusters, we performed unsupervised clustering in each condition and successfully identified 8 subclusters and their corresponding DE genes over the 4 conditions (Fig. 3). Collectively, our result reveals the population heterogeneity in *S. cerevisiae* in multiple conditions.

While higher levels of gene expression heterogeneity and cellular switching between gene expression states can facilitate an adaptive strategy against changing environmental conditions, it can be detrimental to population fitness in stable external conditions as decreasing the fraction of population at the optimal gene expression level would hamper fitness. Surprisingly, our results indicate that cells treated with MPA form a single subcluster, which is indicative of a transcriptionally homogeneous population. Future studies are needed to show how MPA's lifespan-extension effect on yeast cells might be directly contributed by homogeneous expression of key lifespan-regulatory genes and how yeast cells aging in static environmental conditions could benefit more from the MPA treatment compared to cells aging in changing environments.

Although our current work focuses on the application of MPA on young yeast cells not enriched for old age, one can make predictions about how MPA extends lifespan based on the differential gene expression profiles we report here. For example, it is known that mitochondrial DNA accrues mutations as cells age,



**Fig. 5 DE genes in the SC3 clusters with  $k = 4$ .** **a** Dotplot showing the key DE genes (Methods) identified in the 4 SC3 clusters. The size of the dot represents the percentage of cells within a SC3 cluster expressing the corresponding genes, while the color represents the average expression level in natural-log scale of the corresponding genes across all cells within a sample (blue is high). **b** Boxplot showing the top 4 DE genes in SC3 cluster 1. **c** Boxplot showing the top 4 DE genes in SC3 cluster 2. **d** Boxplot showing the top 4 DE genes in SC3 cluster 3. **e** Boxplot showing the top 4 DE genes in SC3 cluster 4. Boxplots show the median (center line), interquartile range (hinges) and 1.5 times the interquartile range (whiskers); outlier data beyond this range are plotted as individual points. Expression levels are plotted in log 2 scale of normalized UMI (Methods). DE genes means differentially expressed genes. (cluster 1  $n = 331$ , cluster 2  $n = 193$ , cluster 3  $n = 164$ , cluster 4  $n = 156$  cells).



**Table 1 yeastDrop-Seq solution.**

Solution (total volume)	Reagent	Reagent volume	
yeastDrop-seq solution (2223.5 µL)	Zymolyase (Zymo Research E1004), 2 Units/µL	37.5 µL	
	Lysis buffer (no DTT)	1.5 mL	
	1 M DTT	75 µL	
	Zymolyase buffer	600 µL	
	10% SDS (final concentration at 0.05%)	11 µL	
	Lysis buffer (no DTT) <sup>2</sup> (950 µL)	H <sub>2</sub> O	500 µL
		20% Ficoll PM-100	300 µL
Sarkosyl		10 µL	
0.5 M EDTA		40 µL	
Zymolyase buffer <sup>37</sup> (CSH protocols Zymolyase Buffer) (100 mL)	2 M Tris pH 7.5	100 µL	
	2 M Sorbitol	50 mL	
	1 M K <sub>2</sub> HPO <sub>4</sub>	4.2 mL	
	1 M KH <sub>2</sub> PO <sub>4</sub>	0.8 mL	
	EDTA (0.5 M, pH 8)	1 mL	
	H <sub>2</sub> O	44 mL	

resulting in the accumulation of damaged mitochondria in the cell<sup>23</sup>. This leads to oxidative stress due to the mitochondrial malfunction and release of reactive oxygen species<sup>30</sup>. Forming a negative feedback regulation, the accumulation of reactive oxygen species can result in further damage in both mitochondrial and genomic DNA<sup>31–33</sup>. In this study, when yeast cells are treated with MPA, we see that, despite a downregulation of genes involved in mitochondrial function, there is an increase in the expression of genes coding for antioxidant proteins and response to reactive oxygen and decrease in the expression of genes coding for reactive oxygen species metabolic process (Supplementary Data 10). The upregulation of these genes could result in fewer reactive oxygen species and could therefore counter the effects of reactive oxygen species on cellular aging.

## Methods

**Cell growth conditions.** Yeast cells with BY4741 genetic background (MATA his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) were grown in 10 mL CSM minimal media containing 2% glucose. Four different treatment/growth conditions were used: minimal media containing (i) 10 µM DMSO (American Bio AB00435), (ii) 10 µM DMSO and 10 µM mycophenolic acid (Sigma M5255), (iii) 10 µM DMSO and 10 µM guanine (Sigma–Aldrich G11950) and 10 µM mycophenolic acid, (iv) 10 µM guanine. Cells were grown for 18 h to a final density (OD<sub>600</sub>) between 0.1 and 0.25. Cells were then diluted to 50 cells/µL for the final yeastDrop-Seq cell-collection using the microfluidic chip.

Doubling time of cells in each treatment condition was calculated as follows:

$$OD_f = OD_i \times 2^{\frac{\Delta t}{t_d}} \rightarrow \log_2 \frac{OD_f}{OD_i} = \frac{\Delta t}{t_d} \rightarrow t_d = \frac{\Delta t}{\log_2 \frac{OD_f}{OD_i}}$$

where: OD<sub>f</sub> = OD<sub>600</sub> final

OD<sub>i</sub> = OD<sub>600</sub> initial

Δt = elapsed culture time (18 h)

t<sub>d</sub> = calculated doubling time

Based on two-point cell density measurements, we found that the doubling time for MPA-treated cells was 187.8 min, for DMSO-treated cells was 130 min, for guanine-treated cells was 130 min, and for GM-treated cells was 158.9 min.

**yeastDrop-Seq experiments.** Cells were subjected to Drop-Seq following the original protocol<sup>2</sup> with the following modifications. Barcoded beads were resuspended in the yeastDrop-Seq solution containing the reagents as shown in Table 1. Cells were diluted to 50 cells/µL. After initial droplet formation, droplets were incubated for 30 min at 30 °C to ensure zymolyase breakdown of cell walls followed by cell bursting (Supplementary Fig. 1). After incubation, droplet quality was evaluated and >95% droplet-uniformity remained intact.

Downstream breakage of oil droplets followed by reverse transcription, exonuclease I treatment, and PCR for cDNA amplification were carried out as per the Drop-Seq protocol<sup>2</sup>. cDNA was then tagged using the NEBNext Ultra II DNA Library Prep with Sample Purification Beads kit (NEB #E7103S).

Sequencing was done using the Illumina HiSeq2500 platform with 2 × 100 read pairs.

**Single-cell RNA-seq data processing.** Sequencing reads were trimmed using Trimmomatic (version 0.39) for adapters and any base with less than 30 quality score was also removed. Reads were aligned to the *S. cerevisiae* reference genome (Genome assembly: R64-1-1 (GCA\_000146045.2), ensembl genomes) using STAR (version 2.7.5a). The expression count matrix was generated using UMI-tools (version 1.0.1) on *Saccharomyces cerevisiae*.R64-1-1.102 transcript annotation. Quality control was performed using the scater R package (version 1.12.2). Number of transcripts counts (UMI) in cells and number of features (genes) of each sample were fitted to the Gaussian distribution and the outliers were filtered. Refer to the code availability section for the code and parameters.

**SC3 clustering.** The cells of the 4 samples were combined and normalized using the command ‘NormalizeData()’ from the Seurat R package (version 3.2.0). The PCA plot was performed using the top 500 variable features. Sample heterogeneity was analysed using the SC3 R package (version 1.15.1). The number of clusters in each sample was evaluated by assessing the heatmap of consensus clustering and the silhouette plot of clusters. The map of cells from samples to the SC3 determined clusters was made as a Sankey plot from the network3D R package (version 0.4).

**Identification of differentially expressed genes.** Differentially expressed genes in each sample and consensus clusters k = 4, k = 8 were analysed using the Seurat R package (version 3.2.0). All samples are combined and log normalized with a scale factor of 10<sup>4</sup>. The Wilcoxon test was used for statistical tests, and the Bonferroni correction method was used to adjust the p-values of each gene. The key DE genes in Figs. 4 and 5 are defined by a threshold of 50% minimum percentage of cells with the gene detected and 0.5 log<sub>2</sub> fold change was used to filter the gene list.

**Doublet rate analysis.** Doublet analyses were performed over 5 tools, including DoubletFinder (version 2.0.3), DoubletDecon (version 1.1.6), scds (version 1.1.2) R packages, Scrublet (version 0.2.1) and solo (version 0.6) python packages. For DoubletFinder, the value of pK in each sample was determined by the mean-variance normalized bimodality coefficient. For DoubletDecon, the rhop value was set to 0.8. For scds, the threshold for doublet was determined by the co-expression based doublet scores, binary classification based doublet scores, and the hybrid scores. A threshold of top 4% cells was set for each score. For Scrublet, the transcriptomic doublet was determined by comparing to the doublet scores of simulated doublets in each sample. For solo, the default parameters were used. For the cells that were flagged as doublets in more than one of the tools (excluding DoubletDecon due to inconsistency), were considered as high-confidence doublets.

**GO term analysis.** Each cluster had genes that were classified as up- or down-regulated based on fold change in expression and statistical significance with p-value less than 0.05, using the Seurat (version 3). The R package clusterProfiler [ref] was then used to perform the GO term enrichment and pathway identifications. In brief, the ‘enrichGO()’ function was used with the org.Sc.gd.db organism database and the Biological Process of the GO term database. The reported p-value is further adjusted using the Benjamini–Hochberg procedure to correct for multiple testing<sup>34</sup>.

**Statistics and reproducibility.** The information on how statistical tests were conducted is provided in the relevant subsections of the Methods section. For example, the Wilcoxon test was used as the statistical test associated with the identification of differentially expressed genes, with the use of the Bonferroni correction method to adjust the p-values of each gene. The sample size was targeted based on the original (mammalian cell optimized) Drop-Seq technique<sup>2</sup>. Corresponding n number of each sample or cluster is indicated where appropriate. Due to the inherent nature of stochasticity associated with gene expression, there were cell-to-cell differences in the analyzed RNA transcript numbers even for cells grown in the same growth condition.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The accession number for the raw sequencing data corresponding to the four treatment conditions is GEO: GSE165686. Source data used as inputs to generate the figure panels, with each Excel file containing the data shown in specific figure panel(s) are available at the DOI-minting repository Zenodo at <https://doi.org/10.5281/zenodo.4767298><sup>35</sup>. All other data are available from the corresponding authors on reasonable request.

## Code availability

The code for analyses and parameters are available at <https://github.com/hemberg-lab/yeastDrop-Seq> and the DOI-minting repository Zenodo at <https://doi.org/10.5281/zenodo.4762526><sup>36</sup>.

Received: 19 February 2020; Accepted: 3 June 2021;

Published online: 30 June 2021

## References

- Kulkarni, A., Anderson, A. G., Merullo, D. P. & Konopka, G. Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Curr. Opin. Biotechnol.* **58**, 129–136 (2019).
- Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Ranzoni, A. M., Strzelecka, P. & Cvejic, A. Application of single-cell RNA sequencing methodologies in understanding haematopoiesis and immunology. *Essays Biochem.* **63**, 217–225 (2019).
- Nadal-Ribelles, M., Islam, S., Wei, W., Latorre, P. & Steinmetz, L. M. Sensitive high-throughput single-cell RNA-Seq reveals within-clonal transcript-correlations in yeast populations. *Nat. Microbiol.* **4**, 683–692 (2019).
- Li, G., Neuert, G. Multiplex RNA single molecule FISH of inducible mRNAs in single yeast cells. *Sci. Data* **6**, 94 (2019).
- Jariani, A. et al. A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. *eLife* **9**, e55320 (2020).
- Ransom, J. T. Mechanism of action of mycophenolate mofetil. *Ther. Drug Monit.* **6**, 681–684 (1995).
- Liu, P., Sarnoski, E. A., Olmez, T. T., Young, T. Z. & Acar, M. Characterization of the impact of GMP/GDP synthesis inhibition on replicative lifespan extension in yeast. *Curr. Genet.* **66**, 813–822 (2020).
- Sarnoski, E., Liu, P. & Acar, M. A high-throughput screen for Yeast replicative lifespan identifies lifespan-extending compounds. *Cell Rep.* **21**, 2639–2646 (2017).
- Herrero, E., Sanz, P. & Sentandreu, R. Cell wall proteins liberated by zymolyase from several ascomycetous and imperfect Yeasts. *J. Microbiol.* **133**, 2895–2903 (1987).
- Stuart, T., Butler, A., Hoffman, P. & Satija, R. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
- Kielev, V., Kirschner, K., Schaub, M., Andrews, T. & Hemberg, M. SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic* **16**, 284–287 (2012).
- Deisenroth, C. & Zhang, Y. Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. *Oncogene* **29**, 4253–4260 (2010).
- Gomez-Herreros, F., Rodrigues-Galan, O., Morillo-Huesca, M., Douglas, M. & Munoz-Centeno, M. C. Balanced production of ribosome components is required for proper G1/S transition in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **288**, 31689–31700 (2013).
- Jurkiewicz, A. et al. Inhibition of tRNA gene transcription by the immunosuppressant mycophenolic acid. *Mol. Cell Biol.* **40**, e00294–19 (2020).
- Gaillard, H. et al. Genome-wide analysis of factors affecting transcription elongation and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair. *PLoS Genet.* **5**, e1000364 (2009).
- Sinclair, D. A. & Guarente, L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* **91**, 1033–1042 (1997).
- Heischmann, S., Dzieciatkowska, M., Hansen, K., Liebfritz, D. & Christians, U. The immunosuppressant mycophenolic acid alters nucleotide and lipid metabolism in an intestinal cell model. *Sci. Rep.* **7**, 45088 (2017).
- Morillo-Huesca, M. & Vanti, M. A simple in vivo assay for measuring the efficiency of gene length-dependent processes in yeast mRNA biogenesis. *FEBS J.* **273**, 756–769 (2006).
- Dalmarco, E. M., Budni, P., Parisotto, E. B., Filho, D. W. & Fröde, T. S. Antioxidant effects of mycophenolate mofetil in a murine pleurisy model. *Transpl. Immunol.* **1**, 12–17 (2009).
- Farkas, Z. et al. Hsp70-associated chaperones have a critical role in buffering protein production costs. *eLife* **7**, e29845 (2018).
- Choi, S. J. N., Lee, H. K., Kim, N. H. & Chung, S. Y. Mycophenolic acid mediated mitochondrial membrane potential transition change lead to T lymphocyte apoptosis. *J. Korean Surg. Soc.* **81**, 235–241 (2011).
- McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).
- Bais, A. B. & Kostka, D. scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinformatics* **36**, 1150–1158 (2020).
- Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
- Bernstein, N. J. et al. Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.* **11**, 95–101 (2020).
- DePasquale, E. A. et al. DoubletDecon: deconvoluting doublets from single-cell RNA-sequencing data. *Cell Rep.* **29**, 1718–1727 (2019).
- Xi, N. M. & Li, J. J. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. *Cell Syst.* **12**, 176–194.e6 (2021).
- Jang, J. Y., Blum, A., Liu, J. & Finkel, T. The role of mitochondria in aging. *J. Clin. Invest.* **128**, 3662–3670 (2018).
- Young, T. Z., Liu, P., Urbonaite, G. & Acar, M. Quantitative insights into age-associated DNA repair inefficiency in single cells. *Cell Rep.* **28**, 2220–2230 (2019).
- Song, R., Sarnoski, E. A. & Acar, M. The systems biology of single-cell aging. *iScience* **7**, 154–169 (2018).
- Sarnoski, E. A., Song, R., Ertekin, E., Koonce, N. & Acar, M. Fundamental characteristics of single-cell aging in diploid Yeast. *iScience* **7**, 96–109 (2018).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1994).
- Lee, J. T. H. yeastDrop-Seq\_ProcessedData (Version v1.0.0) [Data set]. *Zenodo* <https://doi.org/10.5281/zenodo.4767298> (2021).
- Lee, J. T. H. Computational pipeline for the yeastDrop-seq—a yeast-optimized single-cell transcriptomics platform elucidates how mycophenolic acid and guanine alter global mRNA levels (Version v1.0.0). *Zenodo* <https://doi.org/10.5281/zenodo.4762526> (2021).
- Zymolyase Buffer. *Cold Spring Harbor Protocols* [http://cshprotocols.cshlp.org/content/2015/10/pdb.rec088542.full?text\\_only=true](http://cshprotocols.cshlp.org/content/2015/10/pdb.rec088542.full?text_only=true) (2015).

## Acknowledgements

We thank the Acar and Hemberg Laboratory members for useful discussions; Kaya Bilguvar, Karla Neugebauer, Patrick Sung, and Josien van Wolfswinkel for comments and discussions at various stages of this work; the Yale Center for Genome Analysis for running samples through the Illumina sequencer as well as for demultiplexing the samples. M.A. acknowledges funding from the US National Institutes of Health (1U54CA20992). G.U. was partly funded by a Gruber Science Fellowship and by an NIH training grant (T32 GM007499). J.L., G.P., and M.H. were supported by a core grant from the Wellcome Trust.

## Author contributions

G.U. and M.A. designed the experiments and development/optimization aspects of the yeastDrop-Seq setup/platform. G.U. performed the experiments and collected the data, with contributions by P.L., J.T.H.L. and M.H. designed data-analysis methods with contributions by G.U., G.P., and M.A. J.T.H.L. analyzed the data and prepared the figures, with data-analysis contributions by G.U. and G.P. M.A. conceived the project. M.A. and M.H. supervised the project. G.U., J.T.H.L., M.H., and M.A. interpreted the results and wrote the manuscript. All authors read and approved the manuscript.

## Competing interests

The authors declare the following competing interests: M.A. is an inventor on a patent application filed after the discovery of Mycophenolic acid as a lifespan-extending chemical compound. All other authors declare no competing interests.

## Additional information


**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02320-w>.

**Correspondence** and requests for materials should be addressed to M.H. or M.A.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Brooke LaFlamme.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2021