A conserved expression signature predicts growth rate and reveals cell & lineage-specific differences

Zhisheng Jiang*[3], Serena Francesca Generoso*[1], Marta Badia[2], Bernhard Payer[†1,2], Lucas B. Carey[†2,3]

[1] Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003, Barcelona, Spain.
[2] Universitat Pompeu Fabra (UPF), Barcelona, Spain.
[3] Center for Quantitative Biology and Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[†]Correspondence: bernhard.payer@crg.eu and lucas.carey@pku.edu.cn
*These authors contributed equally to the work

Synopsis

By performing RNA-seq on cells FACS sorted by their proliferation rate, this study identifies a gene expression signature capable of predicting proliferation rates in diverse eukaryotic cell types and species. This signature, applied to scRNAseq data from *C.elegans*, reveals lineage-specific differences in proliferation during development. In contrast to the universality of the proliferation signature, mitochondria and metabolism related genes show a high degree of cell-type specificity; mouse pluripotent stem cells (mESCs) and differentiated cells (fibroblasts) exhibit opposite relations between mitochondria state and proliferation. Furthermore, we identified a slow proliferating subpopulation of mESCs with higher expression of pluripotency genes. Finally, we show that fast and slow proliferating subpopulations are differentially sensitive to mitochondria inhibitory drugs in different cell types.

Highlights

1. A FACS-based method to determine the transcriptomes of fast and slow proliferating subpopulations.

2. A universal proliferation-correlated transcriptional signature indicates high protein synthesis and degradation in fast proliferating cells across cell types and species.

3. Applied to scRNA-seq, the expression signature predicts correctly the global slowdown in proliferation during *C. elegans* development, with lineage-specific exceptions.

4. Mitochondria membrane potential predicts proliferation rate in a cell-type specific manner, with ETC complex III inhibitor having distinct effects on the proliferation of fibroblasts vs mESCs.

Abstract:

Isogenic cells cultured together show heterogeneity in their proliferation rate. To determine the differences between fast and slow-proliferating cells, we developed a method to sort cells by proliferation rate, and performed RNA-seq on slow and fast proliferating subpopulations of pluripotent mouse embryonic stem cells (mESCs) and mouse fibroblasts. We found that slowly proliferating mESCs have a more naïve pluripotent character. We identified an evolutionarily conserved proliferation-correlated transcriptomic signature that is common to all eukaryotes: fast cells have higher expression of genes for protein synthesis and protein degradation. This signature accurately predicted growth rate in yeast and cancer cells, and identified lineage-specific proliferation dynamics during development, using *C. elegans* scRNA-seq data. In contrast, sorting by mitochondria membrane potential revealed a highly cell-type specific mitochondria-state related transcriptome. mESCs with hyperpolarized mitochondria are fast proliferating, while the opposite is true for fibroblasts. The mitochondrial electron transport chain inhibitor antimycin affected slow and fast subpopulations differently. While a major transcriptional-signature associated with cell-to-cell heterogeneity in proliferation is conserved, the metabolic and energetic dependency of cell proliferation is cell-type specific.

Introduction

Rates of cell growth and division vary greatly, even among isogenic cells of a single cell-type, cultured in the same optimal environment[1]. Cell-to-cell heterogeneity in proliferation rate has important consequences for population survival in bacterial antibiotic resistance, stress resistance in budding yeast, and chemo-resistance in cancer[2-10]. Time-lapse fluorescence microscopy has shown that cell-to-cell variability in the expression of some genes, such as *p53* and *p21*, is associated with cell-to-cell variability in proliferation and survival[1, 11]. While microscopy can identify dynamic relationships between gene expression and cell fate, it is limited to measurements of one or two genes per cell. Single-cell RNA sequencing measures transcriptome-level heterogeneity but does not directly link this to cell-biological heterogeneity in organelle state, or dynamic heterogeneity in proliferation or drug resistance. Transcriptome-level approaches for understanding within-population cell-to-cell heterogeneity in proliferation and other dynamic processes are lacking. While the presence of intrapopulation variation in proliferation, transcriptome, and organelle-state in both steady-state and in differentiation populations is well established, the relationship among the three remains unclear.

One possibility is that the proliferation-correlated gene expression program is the same, regardless of if one looks at interpopulation variation due to genetic or environmental differences, or intrapopulation heterogeneity due to epigenetic differences and expression noise. However, in the budding yeast *Saccharomyces cerevisiae,* the expression program of intrapopulation heterogeneity in proliferation rate only partially resembles that of cells growing at different rates due to genetic or environmental perturbations[8]. The relation between gene expression and proliferation rate is much less well studied in mammalian cells.

In yeast, in tumors, and in organs, genetic, environmental and developmental changes cause changes in proliferation rate, and changes in the expression of hundreds or thousands of genes[12-16]. Unsurprisingly, many of the genes for which changes in expression are associated with changes in proliferation rate are associated with adverse clinical outcomes in cancer and with antibiotic and antifungal resistance[17, 18]. Within a population of microbes, and within a single multicellular organism, the correct balance of proliferation states and rates is essential. Yet measuring this heterogeneity is difficult, and without such data, understanding the consequences of this heterogeneity is impossible.

Gene expression is associated with phenotype, but mRNAs themselves do not often directly cause phenotypes. Instead, they serve as markers for cell-biological differences between cells. Phenotypes are mostly driven by larger cell-biological differences between cells, such as differences in metabolic state. Cell-to-cell heterogeneity in mitochondria state has been linked to differences in transcription rates, growth rates, proliferation and developmental trajectories[19-21]. Both cancer cells and pluripotent stem cells have atypical metabolisms and use glycolysis to produce much of their ATP, instead of the mitochondria-based oxidative phosphorylation, which is the predominant form of ATP-generation in differentiated cells[22]. It is unknown if this inter-population variation in proliferation, transcriptome, and mitochondria extends to intra-population variation among single cells within a single isogenic population.

Pluripotent stem cells exist in various states, such as naïve or primed, based on culture conditions and embryonic origin[23]. Mouse ESCs reflect the naïve pluripotency state of the blastocyst epiblast and can be cultured in either serum+LIF or 2i+LIF conditions, the latter involving inhibitors of FGF/ERK and GSK3 pathways. Culture in 2i+LIF conditions promotes a ground state more closely mirroring the *in vivo* situation with reduced heterogeneity in pluripotency gene expression and different cell cycle profile when compared to cells grown in serum+LIF[24-26]. Nevertheless, even in 2i+LIF conditions, mESCs display a certain amount of cell-to-cell heterogeneity[27, 28] and it is unclear, how this relates to heterogeneity in differentiated cell types when it comes to gene expression and its link to proliferation rate.

To understand the relation between intra-population transcriptome heterogeneity and heterogeneity in proliferation, we developed a FACS-based method to sort cells by proliferation rate. We applied this method to mouse immortalized fibroblasts and mESCs and performed RNA-seq on fast, medium and slow proliferating cell sub-populations. We identified a “proliferation signature”, mostly consisting of ribosome-biogenesis (protein synthesis) and proteasome-related (protein degradation) genes that are highly expressed in fast proliferating fibroblasts and ESCs. Moreover, the proliferation signature is conserved across cell-type and species, from yeast to cancer cells, allowing us to predict the relative proliferation rate from the transcriptome. We used this gene expression signature to predict proliferation rates in single cells from scRNA-seq data of *C. elegans* development. Unlike previous models to predict growth rate from gene expression[29], this model has no free parameters other than the set of genes, and does not suffer from overfitting – it can predict

differences in growth rate in yeast, cancer cells and *C. elegans*, in spite of no data from either species going into the initial model. When applied to scRNA-seq data from developing *C. elegans*, this model identified a global slowdown in proliferation rate during development, with lineage-specific exceptions where some lineages maintain constant proliferation scores, and others even increase proliferation rate. In contrast to the universality of this main transcriptional signature, many mitochondria-related genes were upregulated in fast proliferating fibroblasts, yet down-regulated in fast-proliferating mESCs. Consistent with this, we found that a high mitochondria membrane potential is indicative of slow proliferating fibroblasts, while in mESCs this is a property of fast proliferating cells. And the mitochondrial electron transport chain complex III inhibitor Antimycin treatment cause opposite effects on the proliferation of fibroblasts and ESCs. Fast, but not slow proliferating fibroblasts are particularly sensitive to the ATP synthase inhibitor oligomycin. Taken together, these results show the existence of a core protein-synthesis and protein-degradation expression program that is conserved across cell types and species, from yeast to mice, and a metabolic and energy-production program that is highly cell-type specific, with cell-type and proliferation-rate specific consequences on the effects of mitochondria inhibitors.

Results

A method to sort single mammalian cells by semi-heritable cell-to-cell heterogeneity in proliferation rate

To understand the causes and consequences of intrapopulation cell-to-cell heterogeneity in proliferation rate in mammalian cells we developed a method for sorting single mammalian cells by their proliferation rate **(Figure 1 and Figure S1)**. The cell-permeable dye carboxyfluorescein succinimidyl ester (CFSE) covalently binds free amines within cells, thus staining most intracellular proteins at lysine residues. In cell types that divide symmetrically, such as embryonic stem cells and immortalized fibroblasts[30], the equal dilution of CFSE into the two daughter cells enables counting of the number of divisions that each cell has undergone. This method is commonly used to differentiate proliferating from nonproliferating cells, and to count discrete numbers of cell division, such as in the study of T- and B-cell proliferation following antigen stimulation[31]. To eliminate confounding effects due to differences in initial staining we used fluorescence-activated cell sorting (FACS) to obtain an initially homogeneous cell population of cells with identical CFSE signals **(Figure 1A and Figure S1A, B)**. Thus, the initial CFSE signal is independent of initial cell-to-cell variation in dye uptake or protein content, as the initial distribution is determined by the FACS gate. $CFSE_{CFR2}$ conjugates are stable and unable to exit the cell[32]; the dye signal is stable for over eight weeks in non-dividing lymphocytes[33]. The measured CFSE signal should be relatively insensitive to cell-to-cell variation in protein degradation. We cultured this sorted starting cell population for several generations, during which time the CFSE signal decreases with each cell division **(Figure 1B)**. Consistent with the decrease in CFSE being mostly due to cell division, the population-level doubling time of each cell type can be calculated based on the decrease in CFSE signal over time **(Figure 1C, D)**, and these doubling times (19-21 hours for fibroblasts and 10-12 hours for mESCs) are consistent with those reported by other methods[34, 35].

To test if the intrapopulation heterogeneity in CFSE that develops after a few doublings corresponds to intrapopulation heterogeneity in proliferation rates, we stained cells with CFSE, isolated a homogenous population by FACS, grew ESCs or fibroblasts for 24h or 48h respectively, and used FACS to isolate the 20% of cells with the highest and lowest CFSE signal, and measured both viability and the fraction of cells in S phase **(Figure S1C, D)**. We found that fast proliferating (low CFSE) subpopulations maintain higher proliferation rates for at least three days **(Figure 1E, F and Figure S1E, F)**, and found no differences in viability between CFSE subpopulations **(Table S1)**. Thus, intrapopulation differences in CFSE correspond to semi-heritable differences in proliferation rates.

To identify genes whose expression is positively or negatively correlated with proliferation rate within a single population we grew fibroblasts MEF (mouse embryonic fibroblast) medium and mouse ESCs in pluripotent ground-state promoting 2i+LIF medium[36], stained cells with CFSE, performed the initial sort to isolate cells with the same CFSE signal, and then grew fibroblasts for five days, and ESCs for three days. We then used FACS to isolate cells with high, medium, and low CFSE signal, and performed RNA-seq on each sub-population **(Figure 1G)**.

Slow-proliferating ESCs are of more naïve pluripotent character than fast-proliferating ESCs

Embryonic stem cells exhibit cell-to-cell heterogeneity in the expression of naïve pluripotency marker genes such as *Nanog*, *Stella* (*Dppa3*) or *Rex1* (*Zfp42*)[37-39]. Although this heterogeneity is most apparent in ESCs cultured in serum+LIF, even when cultured in ground state-pluripotency-promoting 2i+LIF conditions, the sub-population of ESCs with low NANOG-levels displays a propensity for lineage-priming and differentiation[28, 40]. To determine if cell-to-cell variation in proliferation rate was caused by a sub-population of mESCs initiating a differentiation program, we determined the fold-change in expression between slow and fast proliferating sub-populations for a set of genes that are upregulated during lineage commitment (see “Differential expression of pluripotency…” in methods). We found no consistent enrichment of these differentiation genes in fast versus slow proliferating cells, as they could be found to be expressed in either population **(Figure 2A)**. However, the slow proliferating subpopulation did have higher expression of genes that are upregulated in naïve pluripotent cells, and in 2-cell(2C)-like state stem cells **(Figure 2B, C)**, suggesting that slow proliferating mESCs are in a more naïve pluripotent cell state than their fast proliferating counterparts.

Identification of biological processes correlated with proliferation rate that are conserved across cell-types and species, and within single populations

To identify functional groups of genes that are differentially expressed between fast and slow proliferating cells within a single population we performed gene set enrichment analysis (GSEA)[41, 42] **(Figure 3A, B and Figure S2A, B)** on mRNA-seq data from fast and slow proliferating subpopulations. We found that in both fibroblasts and ESCs, as well as for intrapopulation variability in budding yeast FACS-sorted by proliferation rate (data from van Dijk et al.[8]), genes involved in ribosome-biogenesis and the proteasome are more highly expressed in fast proliferating cells **(Figure 3C, D and Table S2)**. High expression of ribosomal

genes is a common signature for fast proliferating cells[12, 43], and cancer cells often exhibit high proteasome expression[44-46], but it is not clear if this is related to proliferation in-and-of-itself or due to aneuploidy and other genetic alterations[47]. Our results suggest that coordinated regulation of the ribosome and proteasome is an intrinsic signature of fast proliferating cells that is conserved across cell-types and species.

To test if the coupling between ribosome biogenesis and proteasome expression holds across species and in diverse cell types, we analyzed the bulk RNA-seq data across developmental stages, covering multiple organs in seven species[16]. Ribosome biogenesis and proteasome expression are highly correlated **(Figure 3E)**. The coordinated expression change with developmental stages between ribosome biogenesis genes and proteasome genes across organs and species suggests that the coordination between protein synthesis and degradation is likely to be a conserved feature across a large number of species and cell-types **(Figure S2D)**.

In addition to ribosome-biogenesis and the proteasome, several other gene sets are differentially expressed between fast and slow proliferating cells in both fibroblasts and ESCs **(Figure 3C)**. mTORC1 (mammalian Target Of Rapamycin Complex 1) functions as a nutrient sensor and regulator of protein synthesis, and is regulated by nutrient and cytokine conditions that cause differences in proliferation[48, 49]. We find that, even in the absence of genetic and environmental differences, mTORC1 is more active in fast proliferating cells. Activation of mTORC1 can promote ribosome-biogenesis[48, 50], however, there is still controversy about the regulation of proteasome activity by mTORC1[49, 51-55].

The transcription factor MYC **(Figure S2C)**, and MYC target genes **(Figure 3C and Table S2)** are more highly expressed in fast proliferating cells. MYC is frequently amplified in cancer, regulates the transcription ~15% of all genes[56] and is a master regulator of cell proliferation[57]. Overexpression of MYC promotes ribosome-biogenesis and cell growth rates[58, 59], and active mTORC1 can promote MYC activation[60, 61]. Our data suggest that increased expression of MYC and increased mTORC1 activity are general properties of fast-proliferating cells, and those genetic or environmental perturbations are not necessary to cause differential expression of these pathways.

Defining a proliferation signature to predict the growth rate across species

Expression of typical proliferation markers, such as PCNA and Ki67, did not correlate with intra-population heterogeneity in proliferation **(Figure 4A)**. The high degree of conservation of genes whose expression correlates with intra-proliferation heterogeneity in proliferation, from yeast to mouse, suggests that there should be a set of genes whose expression is predictive of growth rate across all eukaryotes. To build such a set we combined "proliferation correlated genes" – those with a Spearman correlation of rho = 1 in both fibroblasts and ESCs (243 genes) with genes from six ribosome biogenesis and proteasome related gene sets that are significantly enriched in both fibroblasts and ESCs, which result in a final gene set consisting of 370 genes **(Table S4)**, from whose expression we can calculate a proliferation signature score.

To test the ability of this proliferation signature to predict proliferation rates in new data we used out-of-species cross-validation. While several models have been developed to predict growth rates from gene expression[29], the performance of these models has been evaluated using within-experiment cross validation, in which a single sample (e.g., condition or genotype) was held-out (excluded) and used for testing model performance. Accurate prediction of growth rates in cells for which actual growth rates cannot be measured, such as tumors *in-vivo*, or from single-cell RNA-sequencing data, would be more useful. However, models tested using in-experiment cross-validation (also known as internal validation) are often over-fit, resulting in poor performance when the model is applied to new data from new experiments[62].

To overcome this problem, we used the proliferation signature from above, which was developed using data from mouse cells, to predict growth rate from gene expression in budding yeast. Our model has correlations of R = 0.82, 0.73 and 0.77 across three different datasets **(Figure 4B, S3B and S3C)**. In contrast, the model of Wyotck et al.[29], which was trained on these yeast data, has an out-of experiment predictive power of R < 0.15 (Figure S6 in Wyotck et al.[29]). The Wyotck et al. model is over-fit and cannot predict proliferation rates in new data. Similarly, our model has better performance (R = 0.65) than a cancer-specific model[63], which was trained on cancer cell-line data and cannot predict out-of-experiment **(Figure 4C)** (Figure 4 in Waldman et al.[63]). In contrast to most published models, our proliferation signature score model performs well on data on which it has not been trained.

When run on FACS-sorted 2C-like embryonic stem cells (2C::tdTomato+)[64], our proliferation signature model predicts that 2C-like mESCs proliferate slower **(Figure 4D)**. 2-cell embryos also have uniquely low proliferation scores (**Figure S4H**). These results are experimentally independent of, and biologically consistent with, our observation that expression of 2C-like cell state marker genes is higher in slow proliferating mESCs **(Figure 2C).** This provides further evidence that the proliferation signature we have identified can be universally applied to predict the proliferative state of many cell types.

Prediction of lineage-specific changes in proliferation rates during C. elegans development by the proliferation signature

Expression of the most commonly used markers (PCNA and Ki67) for measuring proliferation rates in bulk populations are cell-cycle regulated; what is really being measured is the fraction of the population that is in S-phase (PCNA) or is not in G0 (Ki67). Expression of these markers does not measure proliferation rates in single cells. Single-cell RNA sequencing is a powerful method for understanding development and differentiation *in vivo*, but it suffers from high levels of noise at the single-gene level[65]. We reasoned that our proliferation signature model would be ideal for measuring the proliferation rates of single cells from scRNA-sequencing data, as the model takes into account the expression of >300 genes, most of which are highly expressed and therefore have low levels of technical noise. To test the ability of the proliferation signature model on scRNAseq data we used a dataset

of 89,701 cells from *C. elegans* development[66]. We computed the proliferation signature score for each cell and divided the cells into terminal cell types vs preterminal cells. Non-terminally differentiated cells have a higher proliferation signature score (t-test, $p = 4.9\times10^{-41}$) **(Figure 5A)**. A visual comparison of the 89,701 cells in UMAP space, colored by either embryo age[66] or by proliferation signature score **(Figure 5B)**, suggested that, globally, proliferation rates in single cells decreases as development proceeds. However, three clusters of cells did not follow this pattern: germline, intestine and M cells. To quantify the relationship between proliferation rates of single cells and developmental time we binned all cells with same embryo time, and calculated the correlation between proliferation score and developmental time **(Figure 5C**). The proliferation signature score decreases as the embryo develops (rho = -0.65, $p = 9.3 \times10^{-19}$ and rho = -0.73, $p = 8.7\times10^{-24}$ when excluding the three outlier groups **(Figure S4B)**). This conclusion from our single-cell gene expression analysis using the proliferation signature score is therefore quantitatively consistent with lineage-tracing microscopy data from Sulston et al.[67], showing that the rate of cell division within the developing nematode decreases during development **(Figure S4D, S4E)**.

To our surprise, the predicted proliferation rates increased after 650 minutes **(Figure S4C**). To understand why, we grouped all cells in 650 minutes or older by lineage **(Figure 5D)**. The three outlier groups from UMAP space: germline, M cell and intestine, had the highest proliferation signature among all cell types late in development. Specifically, for these three cell types, the proliferation score did not decrease with the embryo time, but increased or maintained a high level **(Figure 5E)**. This can be explained by lineage-specific characteristics: the germline is the only cell type in *C. elegans* that is continuously proliferating, M cells are a highly proliferative single mesodermal blast cell, and the intestinal cells, although they do not proliferate, continue to increase in both biomass and DNA content through endoreplication. Other cell types with high proliferation scores, such as Z1/Z4, are also known to continue proliferation after 650 minutes[68]. The proliferation signature score decreases with embryo time for most cell types, including body wall cell, hypodermis and ciliated amphid neuron, which are the most prevalent in the single cell RNA-seq dataset **(Figure 5E)**.

Two cell types, hypodermis and seam cells, exhibited very low proliferation signature scores at late time points, while intestinal cells exhibited very high proliferation scores **(Figure 5D, S4F)**. Both these cell types contain multinucleated cells, but these multinucleated cells arise through very different mechanisms: hypodermis and seam cells through cell-fusion, and intestine through endoreplication[68]. Thus, two cell types with seemingly similar properties have highly divergent transcriptomes, and highly divergent mechanisms to reach their similar final state.

Upon development to an L1 larva, a nematode has more than half of the final number of cells present in the adult. In contrast, mammals continue to rapidly increase in cell number even after embryonic development is complete. This difference can be seen in the change in proliferation signature over time, which decreases in *C. elegans*, but increases in human (data from Petropoulos et al.[69]) and mouse (data from Deng et al.[70]) **(Figure 5F, G**

352 **and S4G, H)**. In conclusion, our proliferation signature genes obtained from mouse fibroblast
353 and ESC data can predict dynamic changes in proliferation rates during the development of
354 various cell types and species, thereby confirming its universal applicability.
355
356
357 *Cell-to-cell heterogeneity in mitochondria state predicts variation in proliferation both in ESCs*
358 *and fibroblasts, but in opposite directions.*
359 While the pattern of within-population proliferation-correlated expression in yeast,
360 mouse fibroblasts and ESCs was broadly similar with regard to genes involved in protein
361 synthesis and degradation, the behavior of metabolic and mitochondria-related genes in fast
362 and slow proliferating subpopulations was highly cell-type specific. Mitochondria membrane
363 and respiratory chain-related gene sets were more highly expressed in fast proliferating
364 fibroblasts, but not in fast proliferating ESCs **(Table 1)**. These results are consistent with
365 differential mitochondrial states in ESCs when compared to differentiated cells like
366 fibroblasts[22], which suggests the existence of different types of metabolism and
367 proliferation-related heterogeneity between pluripotent and differentiated cell-types. We
368 also observed cell-type specific differences in glycolysis, fatty acid metabolism, and other
369 metabolic processes, suggesting fundamental differences in the metabolic pathways required
370 for fast proliferation between pluripotent ESCs and differentiated cells like fibroblasts **(Table**
371 **1)**.
372
373 The mitochondrial membrane potential is a major predictor of cell-to-cell heterogeneity
374 in proliferation rate in budding yeast[9]. Mitochondria-related genes are more highly
375 expressed in the fast proliferating subpopulation of fibroblasts (**Table 1**). In contrast, these
376 genes are slightly more highly expressed in the slow proliferating subpopulation of ESCs. This
377 suggests that the relation between cell-to-cell heterogeneity in mitochondria state and
378 proliferation may be different in these two cell types. To test the ability of mitochondrial
379 membrane potential to predict proliferation rate in mammalian cells we used the
380 mitochondria membrane potential-specific dye tetramethylrhodamine ethyl ester (TMRE) to
381 stain fibroblasts and ESCs, and performed both RNA-seq and proliferation-rate assays on high
382 and low TMRE sub-populations **(Figure 6A)**.
383
384 Unlike the proliferation-based sort **(Figure 1)**, sorting ESCs and fibroblasts by
385 mitochondria-state **(Figure 6 and Figure S5A, B)** resulted in highly divergent expression
386 profiles. ESCs with high TMRE signal had high expression of ribosome-biogenesis,
387 proteasome, MYC-targets and mitochondrial-related genes, while in fibroblasts these gene
388 sets are more highly expressed in the low TMRE sub-population **(Figure 6B, C and Table S5)**.
389 This is consistent with the opposite behavior of mitochondria-related gene sets in
390 proliferation-rate sorted cells from the two cell types **(Table 1)**. We note that the differences
391 between high and low TMRE populations are smaller than the difference between high and
392 low CFSE population **(Figure S5C, D)**, either due to technical limitation of the dye, or because
393 there is less heterogeneity in mitochondria state than there is in proliferation rate.
394
395 These expression data make the following prediction: ESCs with high TMRE should have a

396 shorter doubling time, while fibroblasts with high TMRE should have a longer doubling time.
397 To test this, we sorted fibroblasts and ESCs by TMRE, and found that high TMRE fibroblasts
398 indeed do proliferate more slowly, while high TMRE ESCs proliferate more rapidly **(Figure 6D)**.
399 In addition, we tested the effect of ascorbic acid (vitamin C, an antioxidant) and O_2 levels
400 (ambient 21% atmospheric vs. low 5% physiological levels) on doubling time, but found no
401 significant effects **(Table S6)**.
402
403 To investigate additional cell types, we searched for RNA-seq data for cells sorted by
404 mitochondria state, and analyzed RNA-seq data of mouse CD8+ T-lymphocytes that have
405 been sorted by mitochondria membrane potential (TMRM)[21]. CD8+ T cells with high
406 TMRM signal (high ΔΨm) showed higher expression of ribosome-biogenesis and proteasome
407 related genes **(Table S7)**, and proliferate more rapidly[21], thereby behaving in a similar
408 fashion to ESCs.
409
410 Thus, across yeast, mouse ESCs, fibroblasts and CD8+ T cells, while mitochondria state
411 and proliferation rate co-vary within a single population, the direction of this correlation is
412 different, with yeast and fibroblasts behaving similarly with each other, and opposite to ESCs
413 and CD8+ T cells.
414
415
416 *Perturbation of mitochondria function affects fast and slow proliferating fibroblasts and ESCs*
417 *in different ways.*
418 In order to investigate the relation between proliferation rate, cell type, and
419 mitochondrial state, we performed perturbation experiments by directly inhibiting
420 mitochondria function. We stained both mouse ESCs and fibroblasts with CFSE and sorted 20%
421 of the viable cells on the peak of CFSE signal to have a homogeneous starting population.
422 After culturing them for 24h or 48h respectively, two bins were sorted: the lowest 20% (fast
423 proliferating cells) and the highest 20% CFSE (slow proliferating cells) **(Figure 7A)**. We then
424 cultured the sorted cells with either medium containing DMSO as mock control, the
425 mitochondrial electron transport chain complex III inhibitor Antimycin, the ATP synthase
426 inhibitor Oligomycin for 16h, washed out the drugs, and measured both viability and
427 proliferation rate.
428
429 Both fast fibroblasts and ESCs sorted by CFSE signal maintained a higher fraction of cells
430 in S-phase over two days in growth-media with DMSO, indicating that the proliferation status,
431 fast vs slow, is semi-heritable **(Figure 7B, S6A)**. Interestingly, we found cell-type and
432 proliferation-state specific effects of mitochondria perturbation. Antimycin strongly
433 decreased the fraction of slow-proliferating fibroblasts that were in S-phase but had a
434 weaker effect on fast-proliferating fibroblasts (t-test, $p = 0.0089$) **(Figure 7C)**. In ESCs, the
435 effect of antimycin appeared somewhat stronger on fast- than on slow-proliferating cells
436 (although it did not differ significantly).
437
438 In contrast, fast proliferating fibroblasts were highly sensitive to oligomycin treatment.
439 Specifically, we observed a change in cell morphology upon treatment **(Figure 7D)**. In

comparison with DMSO-treated cells, the cells lost their elongated shape and became more round and smaller. This led us to hypothesize if that morphology change might be explained by a mesenchymal to epithelial transition (MET) upon oligomycin treatment. In fact, during induced pluripotent stem cell (iPSC) reprogramming, MET of fibroblasts is an important early reprogramming step[71, 72]. In that context, oligomycin treatment has been recently shown not only to promote a metabolic switch from oxidative phosphorylation to glycolysis, but also to modulate mesenchymal markers during reprogramming[72, 73]. Therefore we measured the protein levels of the regulators N-cadherin (mesenchymal marker expressed in fibroblasts) and E-cadherin (epithelial marker expressed in ESCs) with and without treatment, by both immunostaining and Western Blot **(Figures 7E and S6A-C)**. We could not detect E-cadherin in fibroblasts, but we observed reduced expression of N-cadherin in comparison with DMSO treated-cells in particular in oligomycin-treated fast cycling fibroblasts **(Figure 7E and S6A-C)**. In addition to the change in morphology, oligomycin treatment reduced cell viability specifically in fast proliferating fibroblasts, but not in slow fibroblasts **(Figure 7F)**. In conclusion, although we observed both a change in cell viability, morphology and a reduction in N-cadherin levels, oligomycin treatment did not induce a complete mesenchymal to epithelial transition in the fast-proliferating fibroblasts as indicated by the lack of E-cadherin upregulation. Nevertheless, the distinct effects of antimycin treatment on the proliferation of fibroblasts and ESCs and the subpopulation-specific effect of oligomycin on fast fibroblasts are in line with a differential dependency on mitochondrial function between the different subpopulations of fibroblasts and ESCs.

Discussion

In summary, we have developed a method to sort cells by their proliferation rate and used these data to identify a pattern of proliferation-correlated gene expression that is conserved among eukaryotes. We used these data to develop a model that can predict proliferation rates from gene expression in multiple eukaryotic species and cell types, and for types of data, such as single-cell RNA sequencing in a developing organism, for which proliferation rates cannot be measured experimentally.

While the CFSE signal is not a measure of the instantaneous proliferation rate, but instead determined by the average proliferation rate integrated over several days, the fact that (A) the transcriptomes of the sorted cells are predictive of proliferation rates, and (B) the cells with low CFSE maintain faster proliferation rates over at least three days, suggests that there are not likely to be large differences in the instantaneous proliferation rate vs the average rate, at least for these cell types and experimental timescales.

We found that genes involved in protein synthesis (ribosome-biogenesis, translation initiation), and in protein degradation (the proteasome and proteasome-related protein degradation) are highly expressed in fast proliferating eukaryotic cells, including mammalian, nematode and yeast cells. Previous studies have reported that high expression of the proteasome in fast-growing cells may be necessary in order to degrade misfolded protein, because the fast protein synthesis in fast-growing cells produce more incorrectly folded

484 proteins[49, 74, 75]. Even with a constant translation and folding error rate, fast proliferating
485 cells will produce more protein, and therefore more misfolded protein that needs to be
486 degraded.
487
488 In all non-cancer mammalian cells we investigated, we also found the mTORC1 signaling
489 pathway enriched in fast proliferating cells and P53-targets enriched in slow proliferating
490 cells. Our results show both upregulations of the mTORC1 signaling pathway and
491 proteasome activity in fast proliferating cells, which is consistent with several previous
492 studies[9, 12-15].
493
494 Our analysis of fast versus slow proliferating ESCs cultured in 2i+LIF conditions indicated
495 at several levels that slow proliferating cells were of a more naïve ground state pluripotent
496 character than fast proliferating cells. First, this was supported by the fact that they displayed
497 a higher expression of naïve pluripotency marker genes and markers of 2C-like cells **(Figure**
498 **2B, C)**. Second, we observed enrichment of E2F targets and genes involved in G1 S cell cycle
499 phase transition **(Table 1)** in our fast cycling ESC population, indicative of a shortened G1
500 phase and a shorter doubling-time, as described for ESCs cultured in serum+LIF [26]. Finally,
501 although we could find differentiation genes to be expressed both in fast and slow
502 proliferating cells **(Figure 2A)**, we saw several differentiation pathways to be enriched
503 specifically in fast dividing ESCs **(Table 1)**. In summary, even when ESCs are cultured in
504 ground-state pluripotency promoting 2i+LIF conditions, they display heterogeneity in
505 proliferation rate, with the slow proliferating being of more naïve pluripotent character when
506 compared to fast dividing cells.
507
508 While we observed ESCs behave similarly to other cell types like fibroblasts or yeast
509 when it comes to gene expression signatures characteristic of fast proliferating cells related
510 to protein synthesis and turnover **(Figure 3C)**, we found a very different behavior when it
511 comes to regulation of metabolism. Although the growth rate can be predicted by
512 mitochondrial membrane potential in *Saccharomyces cerevisiae*[23], where it is negatively
513 correlated with proliferation rate like in fibroblasts as we show in this study, our results show
514 mitochondrial membrane potential to be positively correlated with proliferation rate in ESCs
515 **(Figure 6D)**. This suggests mitochondrial membrane potential has different functions in
516 pluripotent cells when compared to differentiated cell types or yeast. This is corroborated by
517 our gene expression analysis of cells with high vs. low mitochondrial membrane potential
518 **(Figure 6B, C)**, where we found pathways linked with fast proliferating cells to be enriched in
519 fibroblasts with low mitochondrial membrane potential but on the contrary, enriched in ESCs
520 with high mitochondrial membrane potential. Surprisingly, primed pluripotent stem cells
521 have been described to rely more on non-oxidative, glycolysis-based metabolism than naïve
522 pluripotent stem cells[76-78], which appears in contradiction with our result that our slow
523 proliferating, mitochondria activity low ESCs being more naïve-like. However, TMRE is not a
524 direct measure of ATP generation by mitochondria; yeast cells that are respiring and
525 producing all of their ATP using their mitochondria, and yeast cells unable to respire, both
526 have high TMRE signals[9]. Differentiated cells in general rely more on oxidative metabolism
527 than pluripotent cells, therefore our fast proliferating ESCs could potentially reflect a more

528 differentiation prone state.
529
530 *In vivo*, cells exhibit a great degree of variability in proliferation rates, from terminally
531 differentiated neurons, to slowly proliferating cancer stem cells, to rapidly proliferating
532 embryonic stem cells. Many cell types, such as hemopoietic stem cells, contain both
533 proliferating and non-proliferating populations. The proliferation signature model, because
534 of its applicability across all tested species and cell types, provides a useful tool for
535 understanding *in vivo* development for systems, in which precise measurements of
536 proliferation are impossible. Our model has been validated on scRNA-seq data, using
537 published time-lapse microscopy of cell lineages in *C. elegans* as the ground truth[67].
538 However, it is technically challenging to do microscopy or to otherwise measure proliferation
539 of individual cells inside of a developing mouse embryo, or in a tumor in a patient. Models
540 that can accurately predict difficult to measure properties, such as proliferation rate, from
541 easy to measure ones, such as gene expression, will therefore aid in our understanding of
542 complex biological processing during tumor formation, differentiation, and development.
543
544 ## MATERIALS AND METHODS
545
546 **Cell culture growth conditions**
547 Tail tip fibroblasts (TTFs) were isolated from a female newborn mouse from a *Mus musculus*
548 *x Mus Castaneus* cross and immortalized with SV40 large T antigen[79]. The clonal line
549 68-5-11[80] was established and maintained in DMEM supplemented with 10% serum
550 (LifeTech), HEPES (30mM, Life Tech), Sodium Pyruvate (1mM, Life Tech), non-essential amino
551 acids (NEAA) (Life Tech), penicillin-streptomycin (Ibian Tech), 2-mercaptoethanol (0.1mM,
552 Life Tech).
553 The mouse embryonic stem cell (ESC) line EL16.7 (40XX, Mus musculus/*M.castaneus* hybrid
554 background[81] was maintained on gelatin coated tissue culture dishes and passaged every 2
555 days by seeding around $2x10^6$ cells in 2i+LIF medium. Accutase (Merck Chemicals and Life
556 Science) was regularly used for cell detachment when passaging cells. 2i+LIF medium
557 contains a 1:1 mixture of DMEM/F12 supplemented with N2 (LifeTech) and neurobasal
558 media (LifeTech) supplemented with glutamine (LifeTech), B27 (LifeTech), insulin (Sigma),
559 penicillin-streptomycin (Ibian Tech), 2-mercaptoethanol (LifeTech), LIF (Orfgenetics),
560 PD0325901 (Sigma) and CHIR9021 (Sigma). Both TTFs and EL16.7 were cultured at 37C in 5%
561 CO_2.
562
563 **Proliferation and doubling time analysis**
564 ESCs and fibroblasts were plated on 10 cm plates at $5.3x10^6$ and $7.3x10^5$ concentrations,
565 respectively. Cells were expanded and counted for 7 days. To monitor distinct generations of
566 proliferating cells, carboxyfluorescein succinimidyl ester (CFSE, Thermo Fisher Scientific) was
567 used to stain the cells and the dilution of the dye was detected by flow cytometry every day.
568 CFSE was dissolved in dimethyl sulfoxide at a concentration of 5 mM as stock solution and
569 CFSE was added to a 1 ml cell suspension, to a final concentration of 5uM or 10uM. After the
570 addition of CFSE, cells were incubated at 37°C for 20 min. Then the cells were washed twice
571 with complete medium and maintained on ice until use in a buffer containing PBS, 2% serum

572 and 1% pen-strep. Cell viability was determined by DAPI (Biogen Cientifica) staining. Dye
573 signals were measured on an LSRII flow cytometer.
574
575 **RNA-seq**
576 To collect cells with different growth rates, cells were isolated by sorting at room
577 temperature according to the CFSE signal (median and high CFSE signal). ESCs and fibroblasts
578 were sorted into 1.5 ml Eppendorf tubes containing medium and were cultured for 3 days
579 and 5 days respectively in specific culture conditions as described earlier. For each cell line
580 three bins were sorted: the lowest 10%, the median 10% and the highest 10% CFSE. Cells
581 were sorted into prechilled 1.5-ml Eppendorf tubes containing 200 µl medium each. Cells
582 were then centrifuged at 1000 rpm for 5 min, the media removed and the resulting cell pellet
583 was used for RNA extraction. All bins were treated identically throughout the process.
584 Cellular RNA was extracted using the Maxwell RNA Purification Kit and processed for RNA
585 sequencing. For biological replicates, all experiments were repeated on three or four
586 different days. Expression was quantified using Kallisto v0.42.3[82] from the raw reads (no
587 pre-processing) using the gencode.VM18.transcript annotations. We experimented with
588 multiple methods for batch effect removal using the R package SVA[83] and found that the
589 results of the GSEA, with regards to which gene sets were differentially expressed between
590 fast and slow, or high and low TMRE cell populations, did not change. We therefore used the
591 original data.
592 PCA on TMRE sorted biological replicates showed that one TMRE sort was an extreme outlier
593 **(Figure S5C)**; this pair was discarded from all analysis.
594
595 **BrdU Staining**
596 Cell Proliferation was measured by the incorporation of bromodeoxyuridine (BrdU). Every
597 24h BrdU was added at a final concentration of 10 µM to the cells. Incubation under the
598 appropriate growth conditions occurred for 30 minutes for ESCs and 45 minutes for
599 fibroblasts to pulse label the cells. Cells were trypsinized, spun down at 1050 rpm for 5
600 minutes. After washing them in ice-cold PBS, cells were fixed overnight in ice cold Ethanol
601 (70%) while maintaining a gentle vortex. The following day the Ethanol fixed cells were
602 centrifuged and the DNA denatured by adding 2N HCl - 0.5%Triton X-100 for 30min at room
603 temperature. Then cells were centrifuged and resuspended in 0.1 M $Na_2B_4O_7$, pH 8.5 for 10
604 minutes at room temperature. After spinning them down the cells were incubated overnight
605 at 4C with PerCP/Cy5.5 anti-BrdU (1:30 dilution) (BioLegend) in a buffer containing 0.5%
606 Tween 20 / 1% BSA/PBS and RNase (0.8 mg/ml). The following day cells were washed in PBS
607 and spun down at 1050rpm for 5min at room temperature. The pellet was resuspended in
608 PBS with DAPI (1:1000) and analyzed in an BD LSRII flow cytometer.
609
610 **Mitochondria inhibitor treatment assay**
611 For the assessment of chemical inhibitors on membrane potential changes, cells were
612 incubated with medium containing DMSO (0.1%, mock control), Antimycin A (500 nM),
613 Oligomycin (1uM) for 16h. Cells with or without treatment were washed with PBS and
614 trypsinized. After spinning the cells for 5 minutes at 1050 rpm at room temperature, the cell
615 pellet has been stained with 50nM TMRE for 20 min at 37C. After 2 times washes with PBS,

616 cells were resuspended in PBS containing DAPI and immediately analyzed by flow cytometer
617 BD LSRII.
618
619 **Western blot**
620 Cells were lysed in a lysis buffer (20 mM Tris-HCL, pH 8.0, 150 mM NaCl, 1% Triton X-100,
621 supplemented with protease inhibitors cocktail) and centrifuged for 30 minutes at 16000g.
622 The supernatant was boiled in SDS loading buffer. After SDS-PAGE, the samples were
623 transferred to a polyvinylidene difluoride membrane using a transfer apparatus according to
624 the manufacturer's protocols (Bio-Rad). After incubation with 5% nonfat milk in TBST (10
625 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween 20) for 1h, the membrane was washed once
626 with TBST and incubated with antibodies against N-Cadherin (BD Biosciences, 1:1000),
627 E-Cadherin (BD Biosciences, 1:1000), GAPDH (Santa Cruz, 1:5000), at 4 °C for 16 h.
628 Membranes were washed three times for 10 min and incubated with a 1:5000 dilution of
629 Rabbit Anti-Mouse Immunoglobulins/HRP for 1.5 h. Blots were washed with TBST three
630 times and developed with the ECL system (Amersham Biosciences) according to the
631 manufacturer's protocols.
632
633 **Immunofluorescence staining**
634 Cells were grown in 8-well Lab-Tek chamber slides (Thermo Fisher Scientific) and fixed in 4%
635 paraformaldehyde for 10min at room temperature. Then, washed three times in PBS. Fixed
636 cells were permeabilized in 0,5% Triton X-100 (Sigma-Aldrich) in PBS buffer for 10min at
637 room temperature. And then washed in PBST (PBS with 0.1% Tween (Sigma-Aldrich)) for
638 2min at RT. Then cells were incubated in a blocking solution containing 10% bovine serum
639 albumin (BSA, Sigma) and 0.01% Triton X-100 for 1h at room temperature. Cells were then
640 left at 4°C overnight in a blocking solution containing the primary antibody: mouse
641 E-Cadherin (BD Biosciences, 1:1000) and mouse N-Cadherin (BD Biosciences, 1:1000). The
642 next day, the cells were washed three times in PBS and then incubated with the secondary
643 antibody for 45min at room temperature. Goat anti-mouse IgG, (1:1000, Life Technologies)
644 conjugated to Alexa Fluor-555, was used as a secondary antibody. Nuclear staining was
645 performed with DAPI (1:1000, Biogen Cientifica). Images were taken with a Leica TCS SP8
646 confocal microscopy system and were analyzed with Fiji (ImageJ).
647
648 **Differential expression of pluripotency, 2C-like state and lineage commitment-related**
649 **genes in mESCs sorted by proliferation rate (CFSE)**
650 Pluripotent state gene markers are chosen from 4 different studies[28, 84-86], only genes
651 that are used as pluripotent state gene marker in at least 2 of these 4 papers are used in this
652 paper. Lineage commitment and 2C-like state gene markers are the same as genes in Figure
653 5A and key differentiation regulators in Figure 6 of Kolodziejczyk et al.[84]. To see the
654 corresponding pluripotent cell state of fast and slow proliferating mESCs, we calculated the
655 mean expression of naïve pluripotent markers in four fast-proliferating and four
656 slow-proliferating replicates and log2(fast/slow) was calculated to compare genes expression
657 in fast proliferating subpopulation and slow proliferating sub-population. The same method
658 was applied to lineage commitment gene markers and 2C-like state gene markers.
659

660 **Mitochondrial Membrane Potential Measurements.**
661 The relative mitochondrial transmembrane potential (ΔΨm) was measured using the
662 membrane-potential-dependent fluorescent dye TMRE (Tetramethylrhodamine, Ethyl Ester,
663 Perchlorate) (Molecular Probes, Thermo Fisher Scientific)[87]. For TMRE staining fibroblasts
664 and ESCs were grown, washed in PBS, trypsinized and resuspended in PBS with 0.1% BSA and
665 TMRE added at a final concentration of 50nM, from a 10uM stock dissolved in DMSO. Cells
666 were incubated for 20min at 37C, washed with PBS and were analyzed by flow cytometry or
667 sorted.
668
669 **Cell sorting**
670 Cells at 80% confluence in 10cm plates were trypsinized, centrifuged at 1000rpm for 5min
671 and stained with medium containing 10uM CFSE for 20min. Then cells were washed twice
672 with PBS and stained with DAPI as viability dye. To have a homogeneous starting population,
673 20% of the viable cells were sorted according to the proliferation rate on the peak of CFSE
674 signal and re-plated. ESCs and fibroblasts have been cultured for 24h or 48h respectively and
675 two bins were sorted: the lowest 20% (fast proliferating cells) and the highest 20% CFSE
676 (slow proliferating cells). Cells were sorted into prechilled 1.5-ml Eppendorf tubes containing
677 200 µl medium each. Cells were then centrifuged at 1000 rpm for 5 min, the media removed
678 and plated in their culture medium. To monitor their proliferation rate the dilution of the
679 CFSE dye was detected by flow cytometry every day up to 3 days for ESCs and 5 days for
680 fibroblasts. Dye signals were measured on an LSRII flow cytometer.
681
682 For the CFSE sort (no TMRE), cells were stained with CFSE and DAPI, and we used FACS to
683 obtain a population of viable cells the same CFSE signal. We then grew cells for 3 or 5 days,
684 and every 24 hours measured the CFSE signal using flow cytometry. Staining did not have a
685 strong effect on cell viability or proliferation **(Figure S7)**.
686
687 For the TMRE sort for proliferation rate, cells were stained with CFSE and TO-PRO-3, and we
688 used FACS to obtain a population of G1 cells with the same CFSE signal. We then grow cells
689 for 3 or 5 days, and every 24 hours measured the CFSE signal using flow cytometry.
690
691 In order to have a homogeneous starting population, both cell types were stained with
692 Hoechst (10 ug/ml, Life Technologies) to pick cells in G0/G1 phase. Within this population,
693 cells were selected according to the proliferation rate on the peak of CFSE signal prior to
694 staining them with the dye. Then cells were sorted by TMRE into three bins: low, medium
695 and high with a BD Influx cell sorter into prechilled 1.5 ml Eppendorf tubes containing 200 µl
696 medium each. Cells were then centrifuged at 1000 rpm for 5 min, the cell pellet was washed
697 with PBS and used for RNA extraction. All bins were treated identically throughout the
698 process. Cellular RNA was extracted using the Maxwell RNA Purification Kit and processed for
699 RNA sequencing. Cell viability was determined by TO-PRO-3 (Thermo Fisher Scientific)
700 staining.
701
702 To test the effect of O_2 levels and ascorbic acid/vitamin C in both cell types, sorted cells from
703 each bin were plated into each of the four different conditions (low O_2 (5%), normal oxygen

growing conditions, and with or without ascorbic acid/vitamin C (25 ug/ml, Sigma-Aldrich)) in duplicate. After one day of recovery from the sorting, the cells were washed in PBS, were trypsinized, and counted. After seeding the same initial number, the rest of the cells was analyzed on a BD Fortessa analyzer. Every day a sample from each condition and replicate was taken for counting, and stained with 50 nM TMRE, up to 3 days for ESCs and 5 days for fibroblasts, and both TMRE and CFSE were measured by flow cytometry.

Images of CFSE and TMRE stained cells are shown in **Figures S8** and **S9**.

Gene set enrichment analysis (GSEA)

GSEA was performed using the GSEA software and the MSigDB (Molecular Signature Database v6.2)[88, 89]. We use signal-to-noise (requires at least three replicates) or log2 ratio of classes (for experiments with less than three replicates) to calculate the rank of each gene. The maximum number of genes in each gene set size was set to 500, the minimum to 15, and GSEA was run with 1000 permutations. We provided all GSEA results in this study **(Table S8)**.

Enrichment map

Enrichment map of this study (Figure 3C, D and Figure 6B, C) are created using EnrichmentMap in Cytoscape[90, 91], we refer to Reimand et al's protocol[92] for using EnrichmentMap.

We imported the output file of GSEA to EnrichmentMap and set FDR threshold as 0.1, other parameters set as default.

Coordination of expression of ribosome biogenesis and proteasome related genes

We first calculate the mean expression (average of log2(TPM+1)) of ribosome biogenesis genes (genes in GO preribosome gene set) and proteasome genes (genes in GO proteasome complex gene set) across organ developmental time course, then we calculate the Pearson correlation of ribosome biogenesis and proteasome.

Calculation of proliferation signature scores

To obtain proliferation correlated genes, we first calculate, for each gene, the Spearman correlation with proliferation rate, as measured by the decrease in CFSE signal, in both fibroblasts and ESCs. We define "proliferation correlated genes" as genes that have a correlation of 1 in both fibroblasts and ESCs (243 genes). To this set we add genes from six ribosome biogenesis and proteasome related gene sets that are significantly enriched in both fibroblasts and ESCs, which result in a final gene set consisting of 370 genes **(Table S3)** and we called this gene set proliferation signature.

To apply proliferation signature in other species, the R package Biomart[93, 94] was used to obtain homologous genes of other species in this study and to map across different gene naming schemes (eg: transfer Ensemble gene id to Entrez gene id).

Prediction of growth rates using proliferation signature

747 Published expression profiling data for yeasts cultured in the chemostat with controlled
748 growth rate from Airoldi et al. (dataset1)[95], Slavov et al. (dataset2)[96], Regenberg et al.
749 (dataset3)[13] and cancer cell lines with corresponding growth rate[63] were downloaded.
750 For each dataset, we used ssGSEA to calculate the enrichment score of a gene set containing
751 all proliferation signature genes, and calculated the Pearson correlation of proliferation
752 signature score with growth rate. ssGSEA[97] is a rank-based method that computes an
753 overexpression measure for a gene set of interest relative to all other genes in the genome.
754 We use R package GSVA to apply ssGSEA with default settings[98].
755 We also used another method to calculate proliferation signature score for 3 yeast datasets.
756 We use the sum of genes expression for all genes in the proliferation gene set to represent
757 proliferation signature score **(Figure S3D-F)**, the result is slightly worse than the ssGSEA
758 method.
759
760 **Proliferation score of 2C-like mESCs and non-2C-like mESCs**
761 RNA-seq data (GSE33923) of 2C-like mESCs are from Macfarlan et al.[64], who FACS
762 separated 2C-like cells (high MuERVpromoter driven expression, 2C::tdTomato^{+}) from
763 non-2C-like mESCs (2C::tdTomato^{-}). We calculated the proliferation signature score for each
764 of the six samples, and used a paired t-test to control for differences between replicates.
765
766 **Brief description of experiments from other papers**
767 In van Dijk et al.[8] *cts1Δ* histone-GFP budding yeast undergo cytokinesis to separate mother
768 and daughter cells, but these cells remain physically attached to each other by their cell wall.
769 Thus, starting from an initial population of single cells in G1, variation in proliferation rate
770 can be measured by variability in histone-GFP signal in physically connected clusters of cells.
771 Each cluster contains cells descended from the same ancestor cell.
772 In Dhar et al.[9] wild-type yeast were stained with TMRE, and sorted into four bins with
773 varying TMRE signal.
774 In Sukumar et al.[21] pmel-1 T cell receptor (TCR) transgenic mice were injected with
775 recombinant vaccinia virus encoding hgp100 (gp100-VV). Five days after vaccination, they
776 isolated CD8+ T cells, stained them with the lipophilic cationic dye tetramethylrhodamine
777 methyl ester (TMRM) (25 nm for 30 min at 37°C) and FACS-sorted the highest and lowest
778 7-10% of cells for subsequent RNA-seq.
779
780 **Proliferation scores of preterminal cell lineages vs terminal cell types**
781 Preterminal cell lineage and terminal cell type pseudo-bulk RNAseq data of *C.elegans* were
782 downloaded from Murray et al.[66], specifically, gene expression profile for terminal cell
783 types and preterminal cell lineage is in Table S7 and Table S8, annotation file for terminal cell
784 types and preterminal cell lineage is in Table S2 and Table S4. As there are multiple time
785 points for one terminal cell type, we only use the sample with maximum time point to
786 represent corresponding terminal cell type, processed data provided in this study **(Table S9)**.
787 For each cell we calculate proliferation signature score, and a t-test was used to compare the
788 mean proliferation signature score of all cells in each of the two groups.
789
790 ***C. elegans* scRNA-seq data analysis**

C. elegans scRNA-seq data was provided in R package "VisCello.celegans". After loading the package, type cello() to load all data into the current environment. We calculated the proliferation signature score for all single cells, then color them by proliferation signature score in UMAP space. The calculation of proliferation signature score for single cell data is different from the calculation for bulk RNA-seq data. We just sum up the expression value of genes in proliferation signature gene set to get proliferation signature score, but not use ssGSEA consider ssGSEA is a rank-based method, however most of the genes have 0 expression in this scRNA-seq data set **(Figure S4A)**, which makes it not appropriate to use ssGSEA.

We binned all single cells by their embryo time. We first calculate the mean proliferation score for cells with same embryo time, then calculate Spearman correlation of this mean proliferation score with embryo time, the result is rho = -0.65 ($p = 9.3 \times 10^{-19}$), the correlation of unbinned data is -0.41 ($p < 2.2\times10^{-16}$). After excluding three special cell types germline, M cell and intestine, the result is rho = -0.73 ($p = 8.6 \times 10^{-24}$), the correlation of unbinned data is -0.45 ($p < 2.2\times10^{-16}$).

Experimental data for *C. elegans* development

Developmental data of *C. elegans* was extracted from figure 4 of Sulston et al.[67]. This figure is cell number (live nuclei number) change over embryo time and we use WebPlotDigitizer[99] to extract data. We use the data to plot log2 cell number change over embryo time. The difference of log2 cell number for two adjacent time points divided by the difference of time is the proliferation rate of mean of two time points.

DATA AND CODE AVAILABILITY

The code for performing analysis is available at https://github.com/carey-lab/Proliferation_Signature_Public. Raw and processed RNA-seq data created in this study are available on GEO (Gene Expression Omnibus) with the accession code GSE139594. Information of all published data used in this study is provided **(Table S10)**.

AUTHOR CONTRIBUTIONS

Z.J. and S.F.G. made the figures. Z.J. and M.B. analyzed the data. S.F.G. and M.B. did the experiments. L.B.C. and B.P. supervised the project. L.B.C., Z.J., S.F.G. and B.P. wrote the manuscript. All authors read and approve of the final manuscript.

ACKNOWLEDGMENTS

We thank the CRG/UPF flow-cytometry core and CRG Genomics facilities for help with experiments, and Yang Zhao for comments on the manuscript.

This work has been funded by the Spanish Ministry of Science, Innovation and Universities (BFU2014-55275-P and BFU2017-88407-P to B.P. and BFU2015-68351-P to L.B.C.), the AXA Research Fund and the Agencia de Gestio d'Ajuts Universitaris i de Recerca (AGAUR, 2017 SGR 346 to B.P. and 2014 SGR 0974 & 2017 SGR 1054 to L.B.C.). We would like to thank the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership,

to the 'Centro de Excelencia Severo Ochoa', and the Unidad de Excelencia María de Maeztu, funded by the MINECO (MDM-2014-0370). We also acknowledge the support of the CERCA Programme of the Generalitat de Catalunya. L.B.C. was supported by funding from Peking University and from the Peking-Tsinghua Center for Life Sciences.

FIGURE LEGENDS AND TABLE

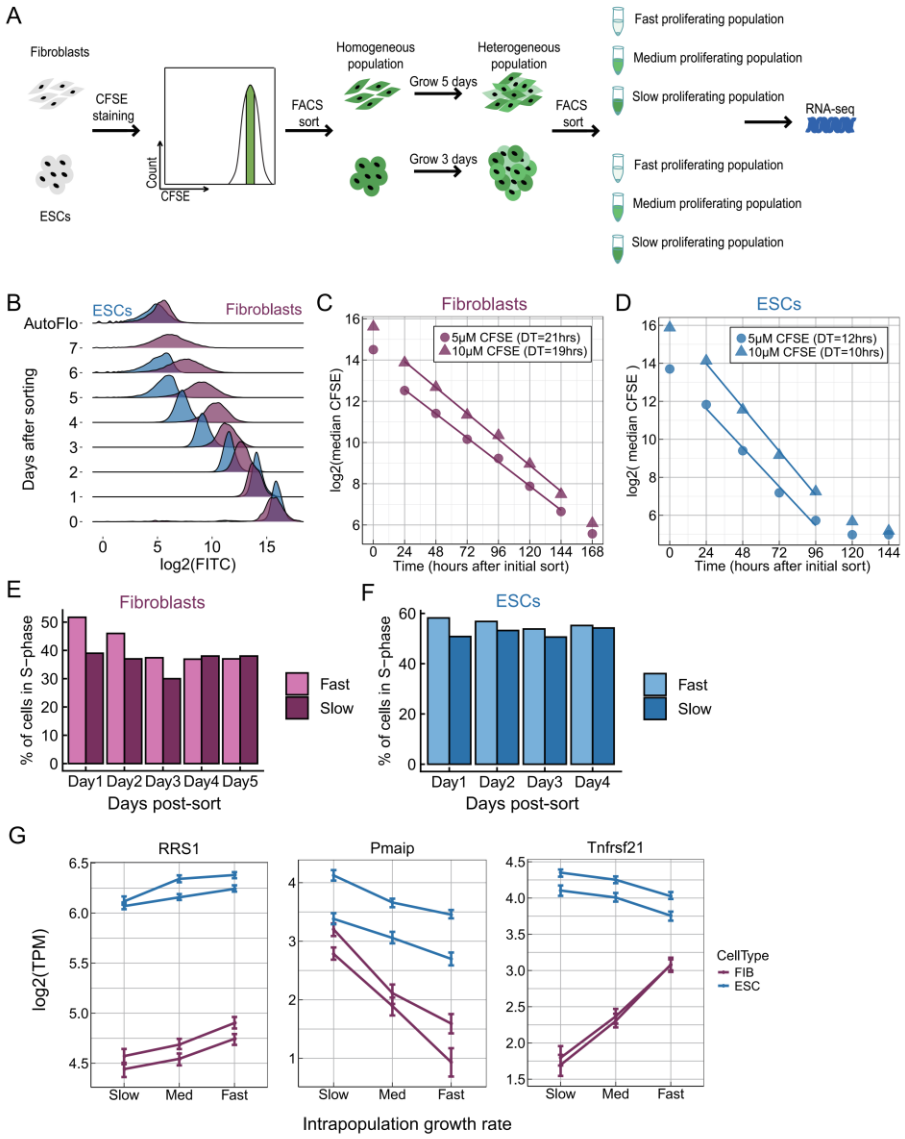


Figure 1. A CFSE-based method to sort mammalian cells by proliferation rate.
(A) Cells were stained with CFSE and a subpopulation of cells with identical CFSE levels was collected by FACS. Growth for several generations resulted in a heterogeneous cell population with a broad CFSE distribution, and cells with high, medium, and low CFSE signal (slow, medium and fast proliferation, respectively) were sorted by FACS for RNA-sequencing. **(B)** The change in the CFSE distribution over time, for fibroblasts and ESCs. **(C, D)** The

population-level doubling time can be calculated by fitting a line to the median of the log2(CFSE) signal. We discard data from time 0, cells immediately after the sort, because the CFSE signal decreases in the initial hours, even in the absence of cell division, likely due to efflux pumps. **(E, F)** BrdU was used to measure the % of cells in S-phase for FACS-sorted fast and slow proliferating subpopulations. Fibroblasts: 4 replicates, p = 0.0002441. ESCs: 3 replicates for ESCs, p = 0.001953. p-values are for binomial tests across all biological replicates that the two populations have the same percentage of cells in S-phase. **(G)** Examples of genes whose expression positively or negatively correlated with proliferation rate. Each line is one biological replicate, and the error bars are 95% confidence intervals for each expression value.

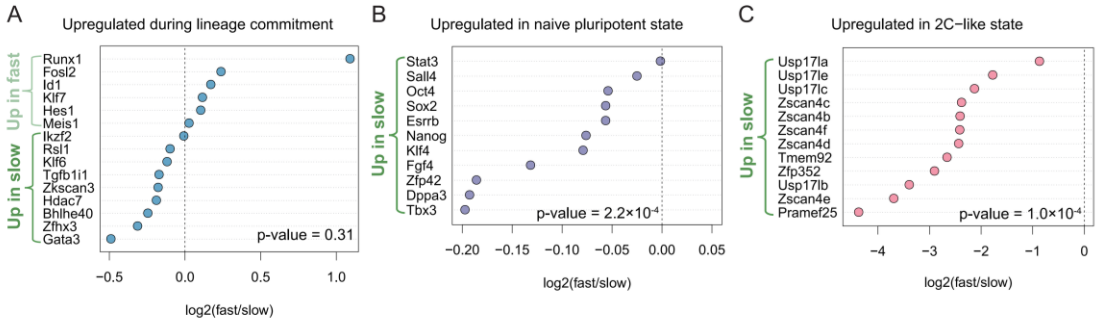


Figure 2. Slow-proliferating ESCs display a more naive pluripotent stemness character than fast-proliferating ESCs.

(A) Comparison of lineage commitment-related gene expression between fast and slow proliferating subpopulations. **(B)** Comparison of pluripotency-associated gene expression between fast and slow proliferating subpopulations. **(C)** Comparison of 2C-like state markers expression between fast proliferating subpopulation and slow proliferating sub-population. Dashed lines separate genes expressed preferentially in slow- (left of dashed line) or in fast-proliferating (right of dashed line) ESCs. P-values are from binomial tests, testing if genes are more often highly expressed in slow cells than would be expected by chance (53.5% of all genes are more highly expressed in slow cells).

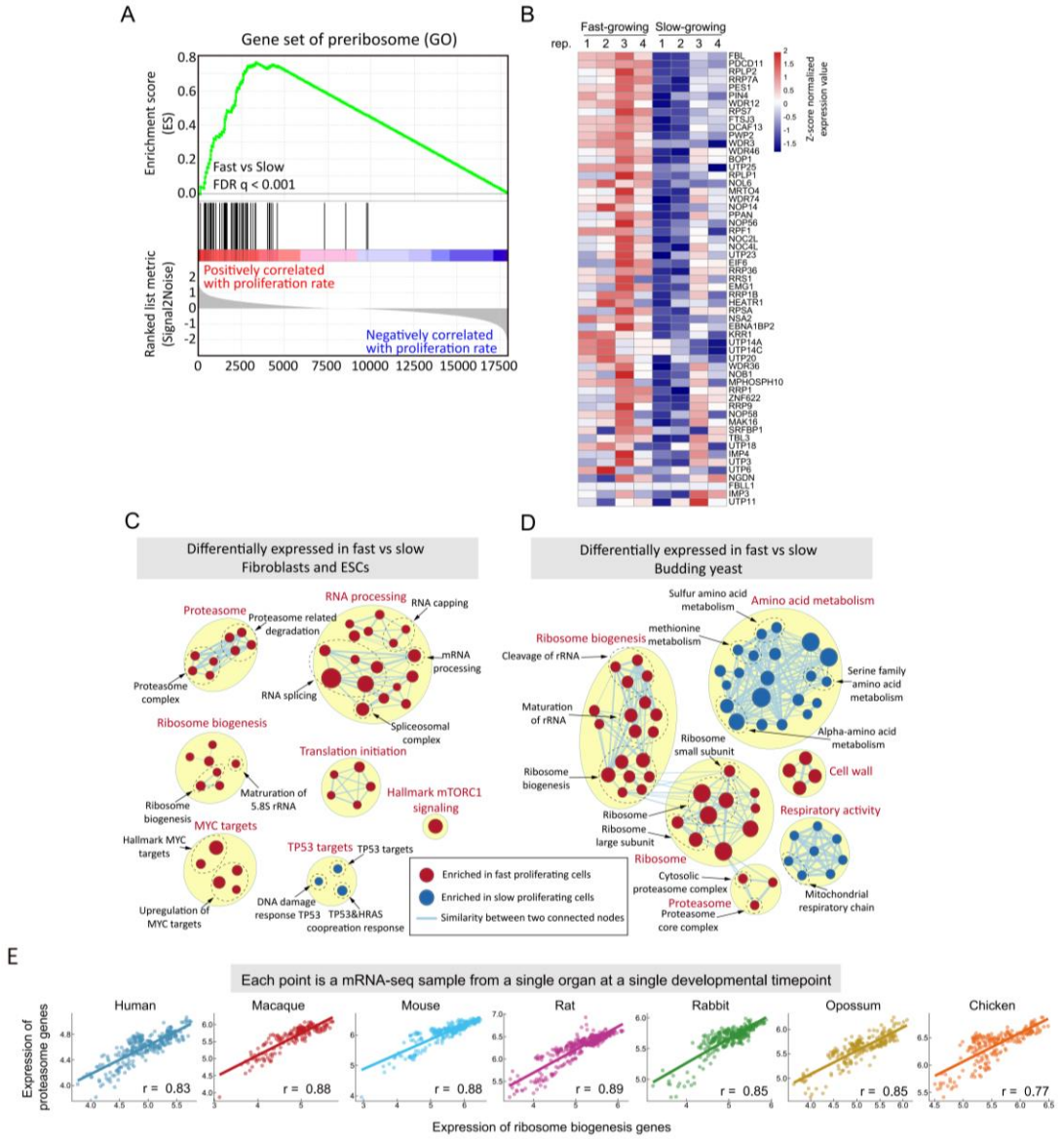


Figure 3. Functional pathways for which cell-to-cell heterogeneity in expression correlates with proliferation rate across cell types and species.

(A) In Gene Set Enrichment Analysis, genes are sorted by their fast/slow expression value (left panel, bottom), and each gene is represented by a single black line (left panel middle). The enrichment score is calculated as follows: for each gene not in the GO preribosome gene set, the value of the green line decreases, and for each gene in the gene set, the value of the green line increases. The ES score will be near zero if the genes in a gene set are randomly distributed across the sorted list of genes, positive if most genes are to the left, and negative if most genes are to the right. **(B)** The heatmap (right panel) shows the expression (z-scored read counts) of preribosome genes in fibroblasts across four biological replicates of the CFSE sorting experiment. **(C)** Gene sets enriched (FDR<0.1) in both fibroblasts and ESCs were mapped as a network of gene sets (nodes) related by mutual overlap (edges), where the color (red or blue) indicates if the gene set is more highly expressed in fast (red) or slow (blue) proliferating cells. Node size is proportional to the total number of genes in each set and edge thickness represents the number of overlapping genes between sets. **(D)** GSEA results

(FDR<0.1) of *S. cerevisiae*[8] that sorted by cell-to-cell heterogeneity in proliferation rate. **(E)** Pearson correlations of mean expression (average of log2(TPM+1)) of ribosome biogenesis genes vs proteasome genes across organ developmental time courses in seven species (see also Figure S1).

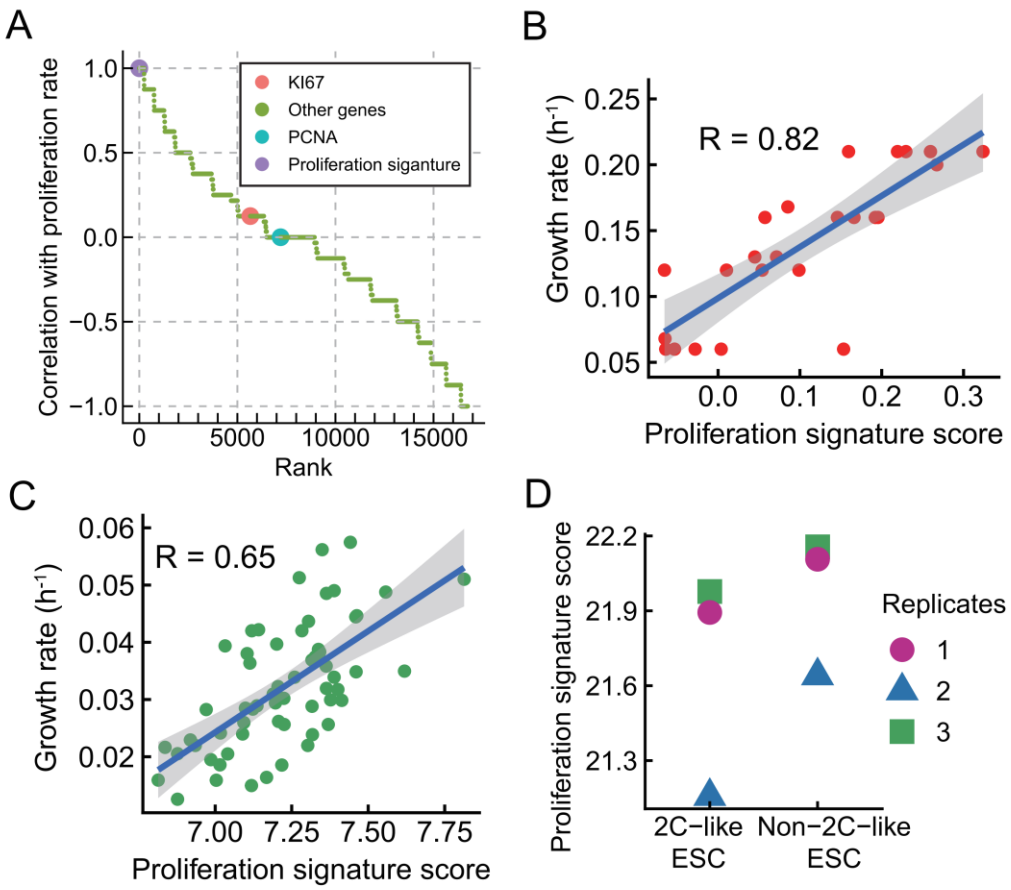


Figure 4. A proliferation signature model can predict relative growth rates from gene expression for species and cell-types on which it was not trained.

(A) Genes and proliferation signature spearman correlation with proliferation rate (sorted by CFSE). Compare with Ki67 or PCNA, proliferation signature has a better correlation with proliferation rate. **(B)** Using the proliferation signature to predict growth rate in budding yeast, we apply ssGSEA to calculate the enrichment score of proliferation signature for each sample. The Pearson correlation of proliferation signature score with growth rate is 0.82 ($p = 8.9\times10^{-7}$). **(C)** Using the proliferation signature to predict growth rate in cancer cell lines, the Pearson correlation is 0.65 ($p = 1.9\times10^{-8}$). **(D)** Comparison of proliferation signature score between 2C-like ESC and non-2C-like ESC (paired t-test, p = 0.04669).

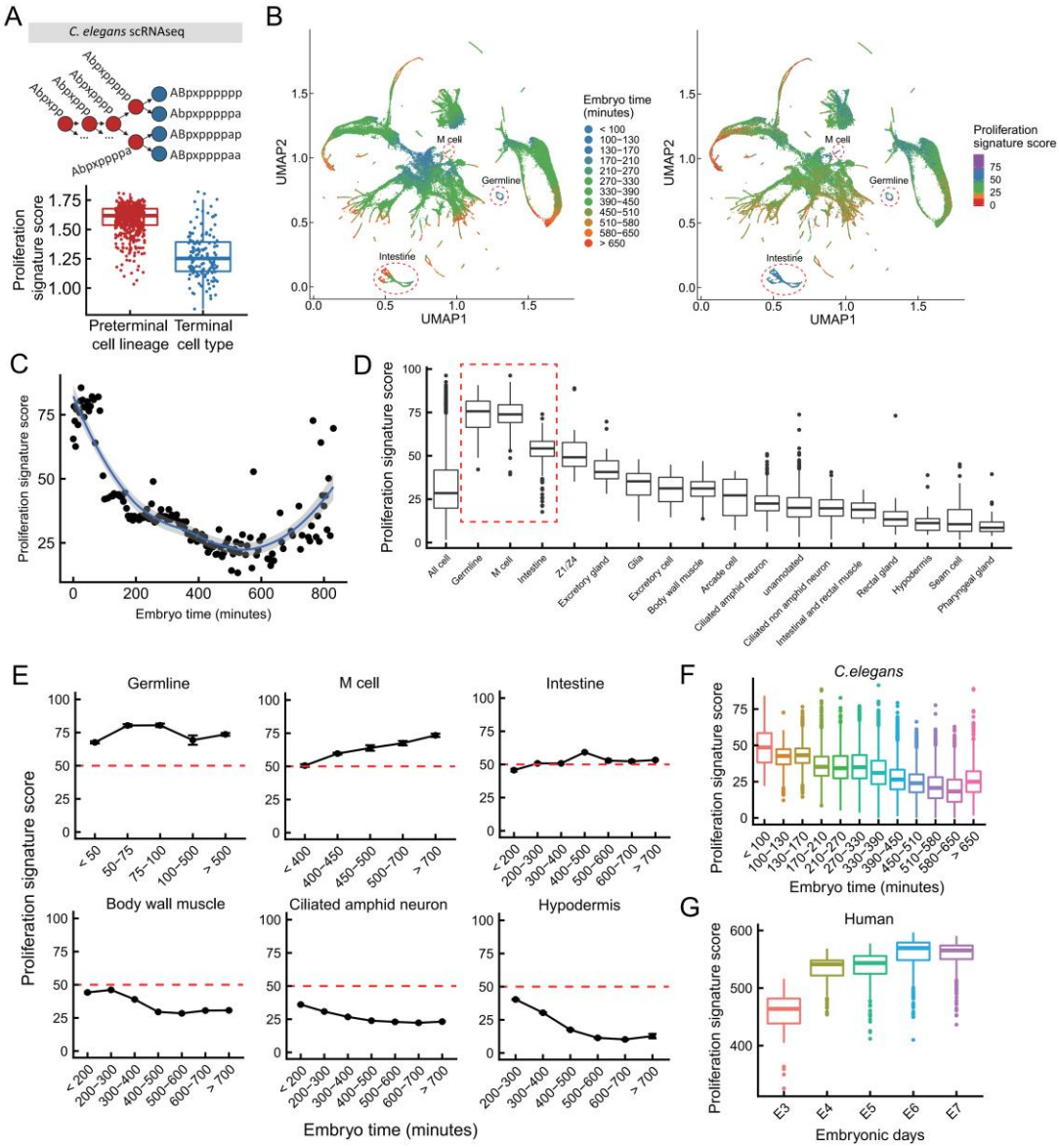


Figure 5. Proliferation signature of cell development.

(A) A cartoon showing four terminal cells, and a partial linage showing the final four generations of preterminal cells. Comparison of single-cell proliferation signatures between preterminal cell lineage and terminal cell types (t-test, $p = 4.9\times10^{-41}$). **(B)** UMAP projection of 89,701 cells. Cells in the left panel are colored by estimated embryo times; in the right panel by proliferation signature score. **(C)** To calculate the proliferation signature score (y-axis) at each time point (x-axis) cells are binned by embryo time, and the mean proliferation signature score for all cells in the same bin is calculated. The spearman correlations are -0.65 ($p = 9.3\times10^{-19}$) for binned data and -0.42 ($p < 2.2e-16$) for unbinned data. **(D)** Boxplots (line shows median, boxes interquartile range) of proliferation signature score for all cells with embryo time > 650min. **(E)** Temporal dynamics of proliferation scores of select cell lineages, showing the average proliferation score for all single cells in that lineage, at each time point. **(F-G)** Boxplot of *C. elegans* **(F)** and human **(G)** proliferation signatures as a function of developmental time, from scRNAseq data.

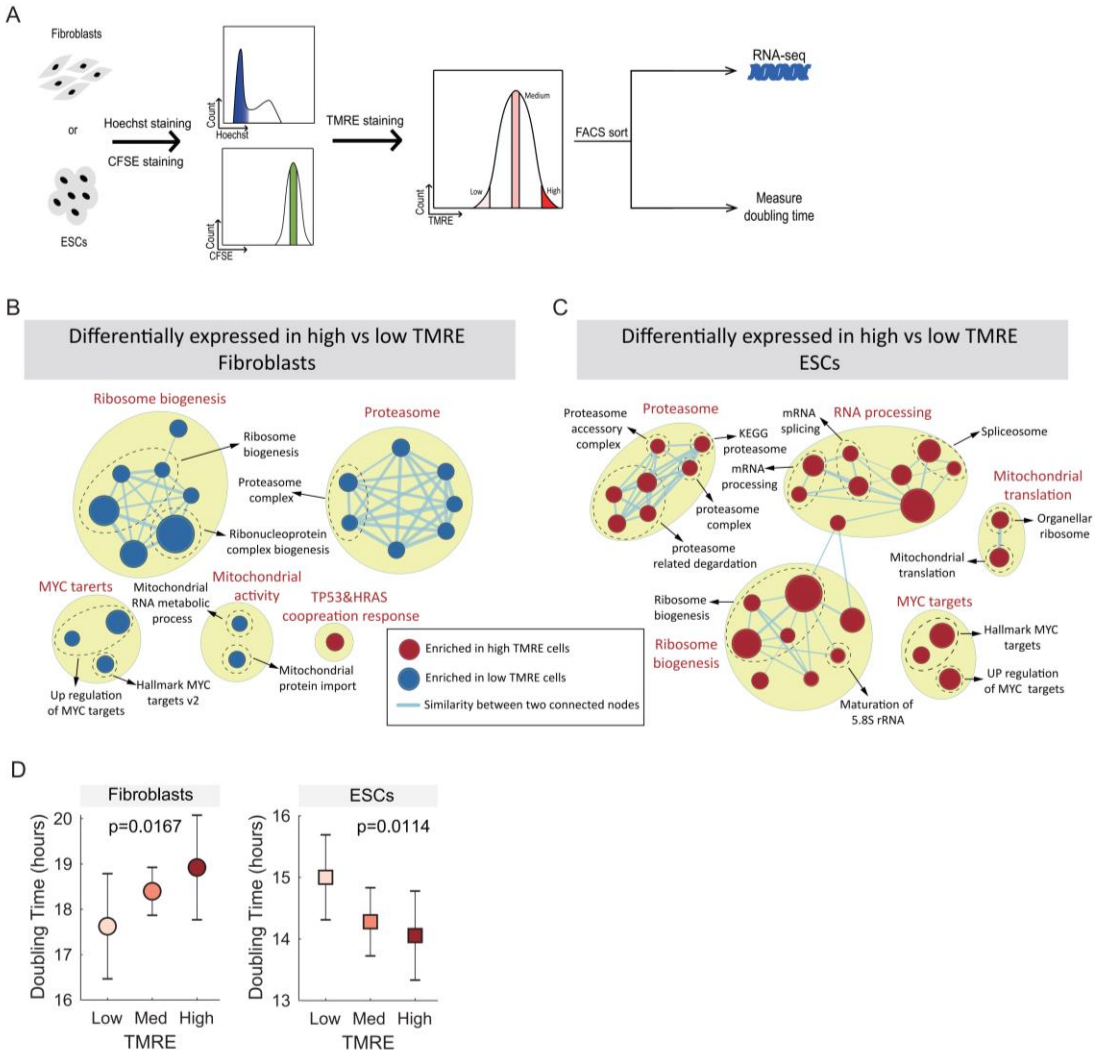


Figure 6. Expression of proliferation-related gene sets in cells sorted by intra-population heterogeneity in mitochondria membrane potential.

(A) Cells were stained with Hoechst and CFSE and a homogenous population of equally sized cells in G1 with equal CFSE was obtained by FACS. These cells were stained with TMRE sorted by TMRE, and then used for RNA-seq, or allowed to proliferate to measure the doubling time of each TMRE sub-population. **(B, C)** Enrichment maps of fibroblasts and ESCs sorted by TMRE. **(D)** Doublings times, as estimated by the measured by the decrease in CFSE signal over time, for high, medium and low TMRE sorted cells. P-values are from ANOVA, testing if TMRE is predictive of doubling time (see methods).

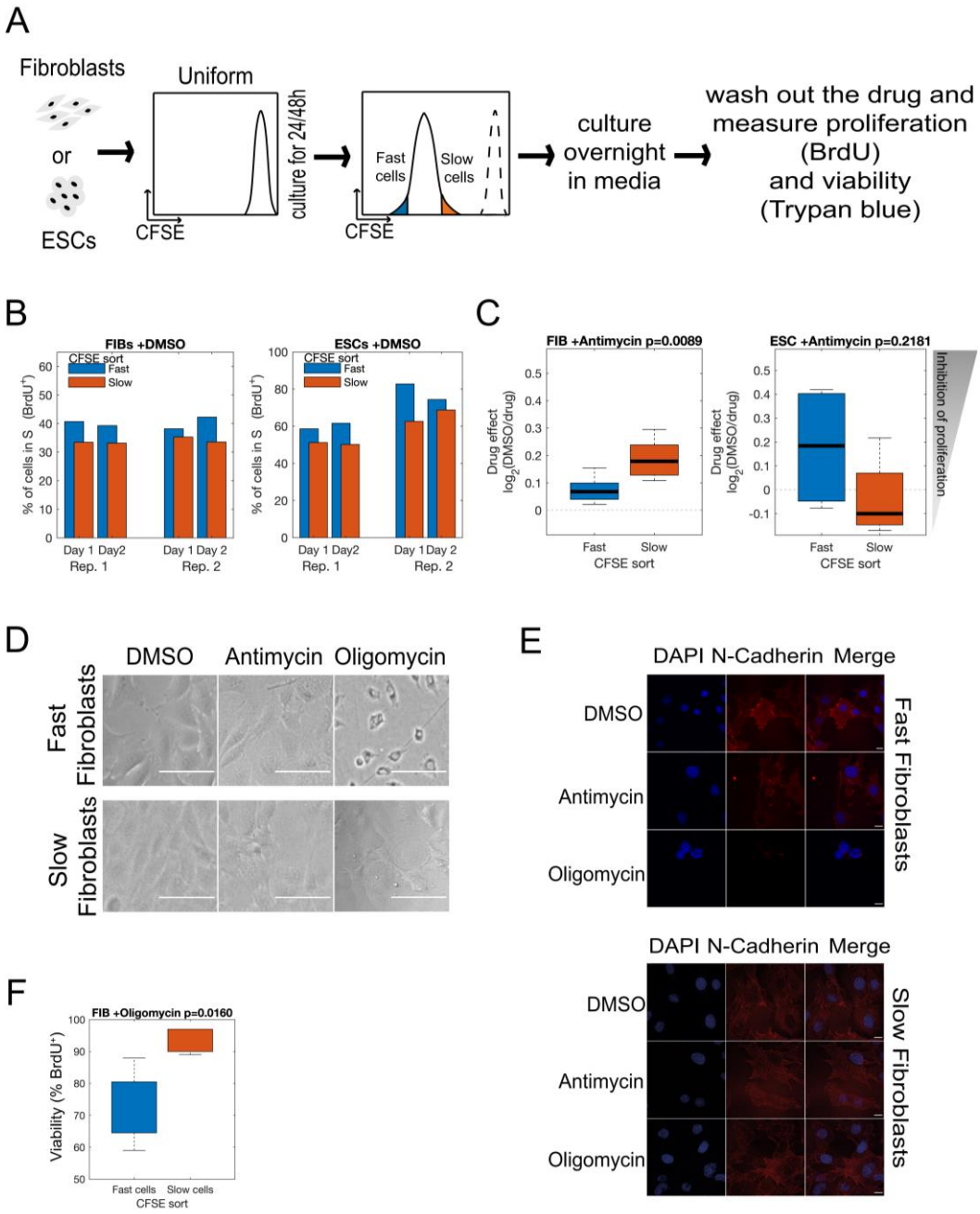
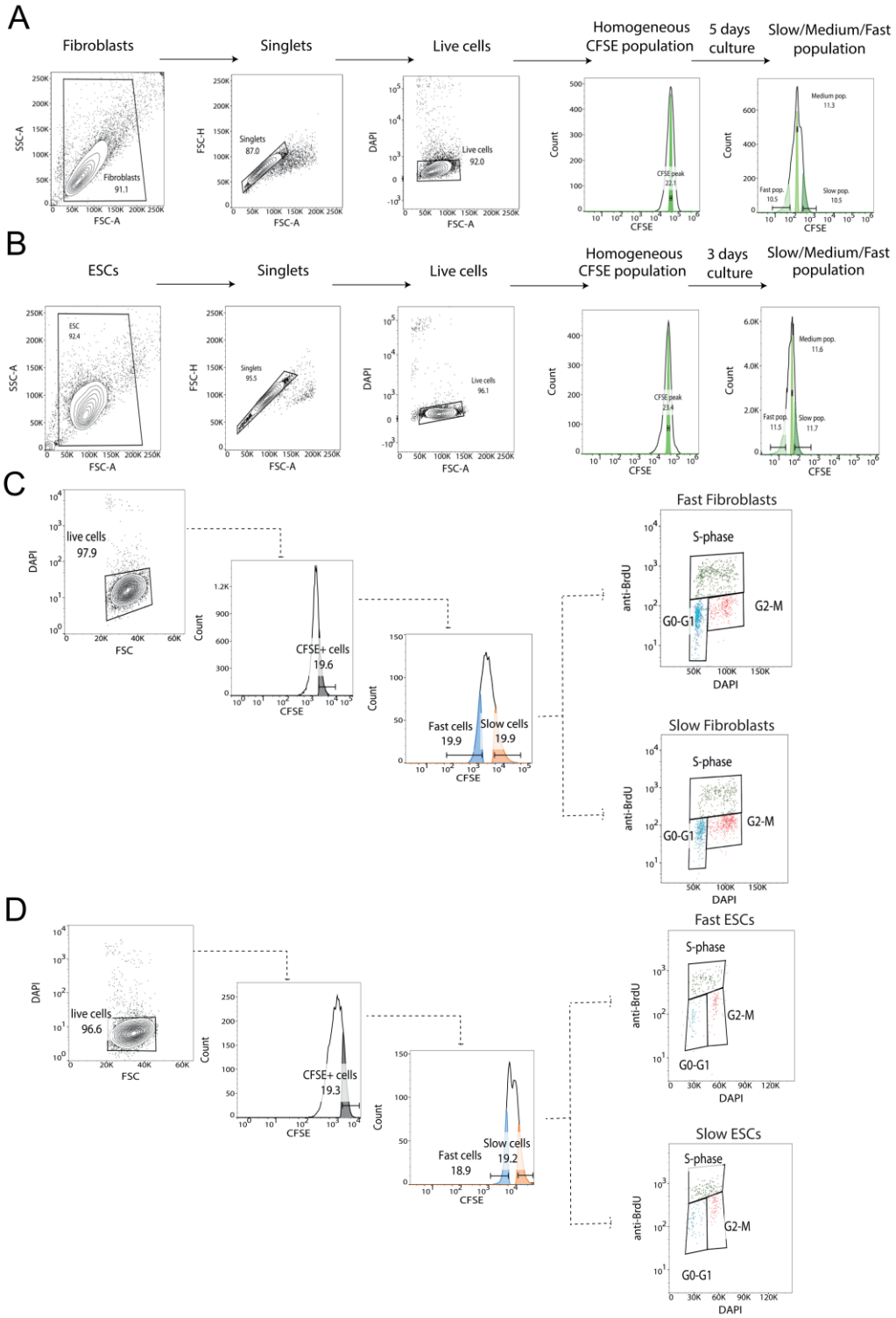
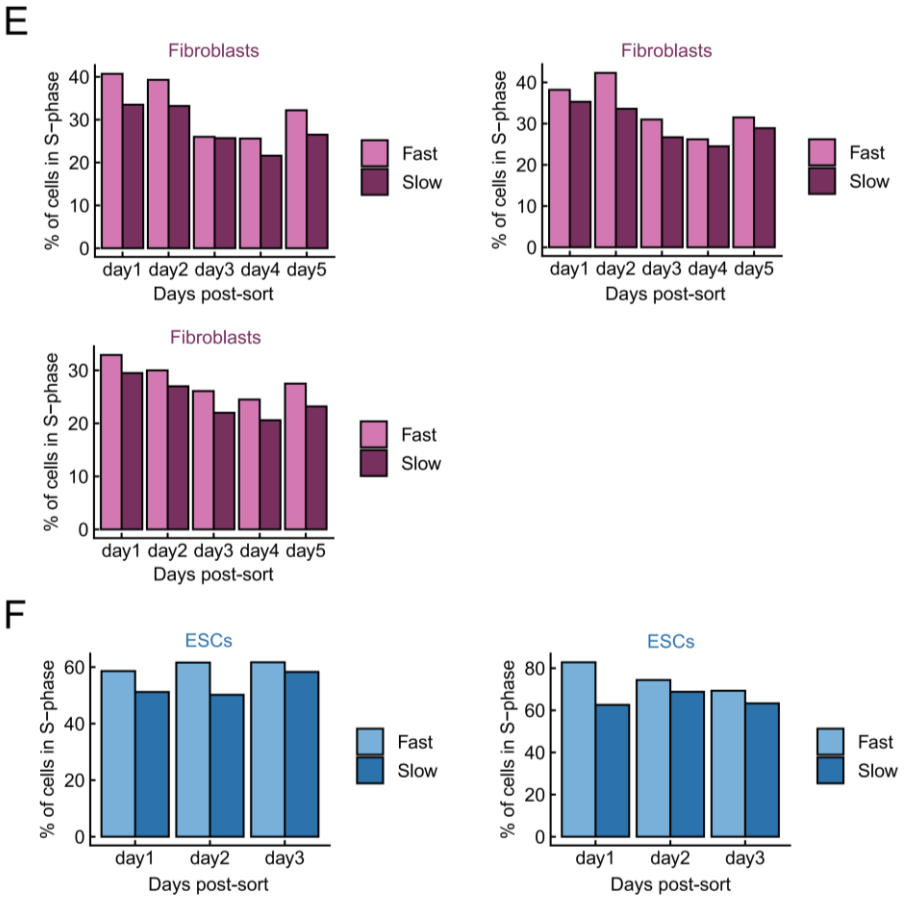


Figure 7. Growth-rate and cell-type specific effects of mitochondria inhibitors on proliferation rate, cell viability and cell state. (A) Schematic of the experimental setup for measuring the effects of mitochondria inhibitors on slow and fast proliferating cells. **(B)** Fast proliferating Fibroblast and ESCs sorted by CFSE signal maintained a higher fraction of cells in S phase over two days of growth in media+DMSO. **(C)** Effect of antimycin treatment on fast and slow proliferating fibroblasts and ESCs. **(D)** Fast fibroblasts changed morphology after the treatment with oligomycin. Scale bars = 80µm. **(E)** Immunostaining of fibroblasts for N-Cadherin and DAPI after drug treatment. Fast fibroblasts lose N-Cadherin staining specifically after oligomycin treatment. Scale bars = 15µm. **(F)** Effect of oligomycin and antimycin treatment on fibroblast viability. Oligomycin has a specific effect on the viability of fast-proliferating fibroblasts.

Table 1. Gene sets whose expression exhibits opposite correlations with growth between fibroblasts and ESCs. Shown are representative gene sets whose expression is significantly correlated with proliferation in either fibroblasts or ESCs, but whose expression changes with proliferation in opposing directions.

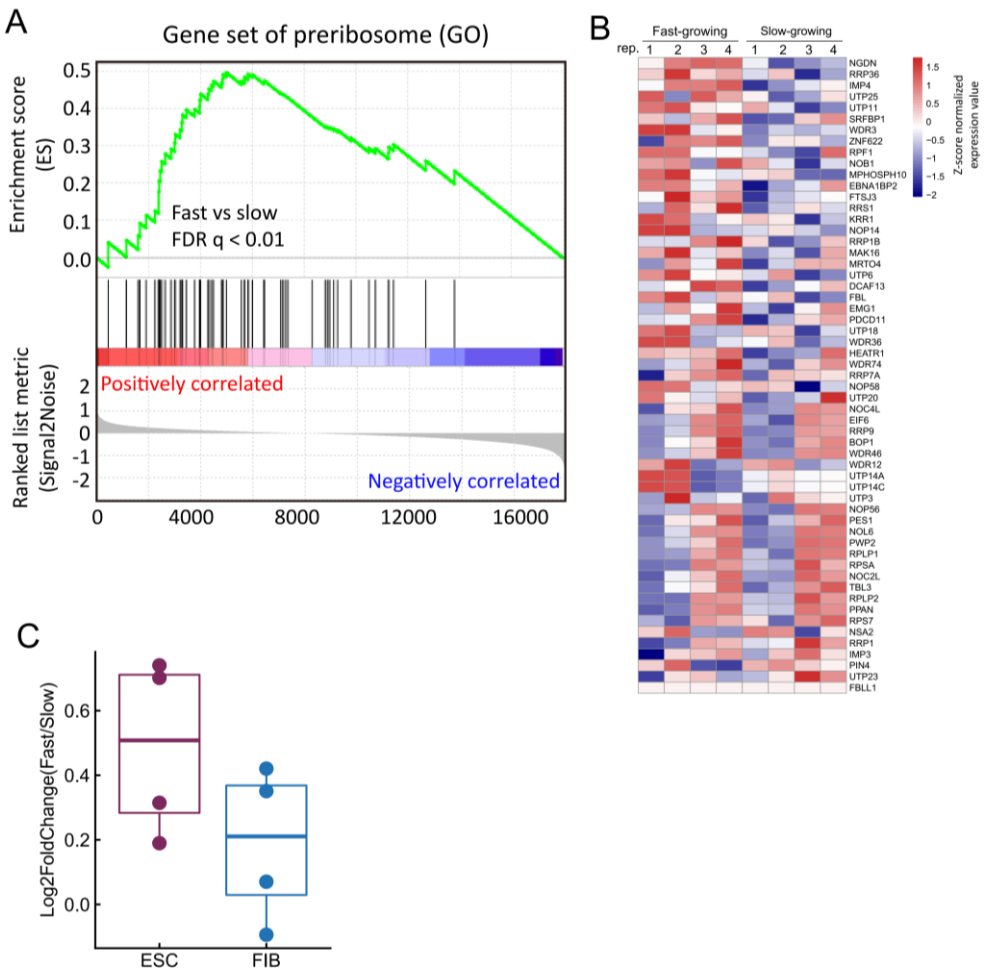
SUPPLEMENTAL FIGURE LEGENDS





957

958 **Figure S1. Fast proliferating subpopulations maintain higher proliferation rates than slow**
959 **proliferating subpopulations for at least 3 days.**
960 **FACS gating strategy for CFSE staining to get cell subpopulations with different proliferation**
961 **rates.**
962 **(A, B)** The gating strategy for CFSE staining to get cell subpopulations with different
963 proliferation rates in fibroblasts **(A)** and in ESCs **(B)**. Slow, medium and fast proliferating cell
964 subpopulations were sorted by FACS according to their CFSE signal. Then RNA-seq was
965 performed on each of the three subpopulations. **(C, D)** FACS gating strategy for measuring
966 the heritability of proliferation rates. The gating strategy for measuring the heritability of
967 proliferation rates in fibroblasts **(C)** and in ESCs **(D)**. In all experiments, the laser voltage was
968 increased so that, when sorting high and low CFSE cells, the modal CFSE signal was at least
969 10^3; the voltage is not the same for the first and second CFSE sorts. **(E)** 3 Replicates of
970 fibroblasts that similar to figure 1E. **(F)** 2 Replicates of ESCs that similar to figure 1F.
971


A
Gene set of preribosome (GO)
Enrichment score (ES)
Fast vs slow
FDR q < 0.01
Ranked list metric (Signal2Noise)
Positively correlated
Negatively correlated
B
Fast-growing
Slow-growing
rep.
Z-score normalized expression value
C
Log2FoldChange(Fast/Slow)
ESC
FIB

972

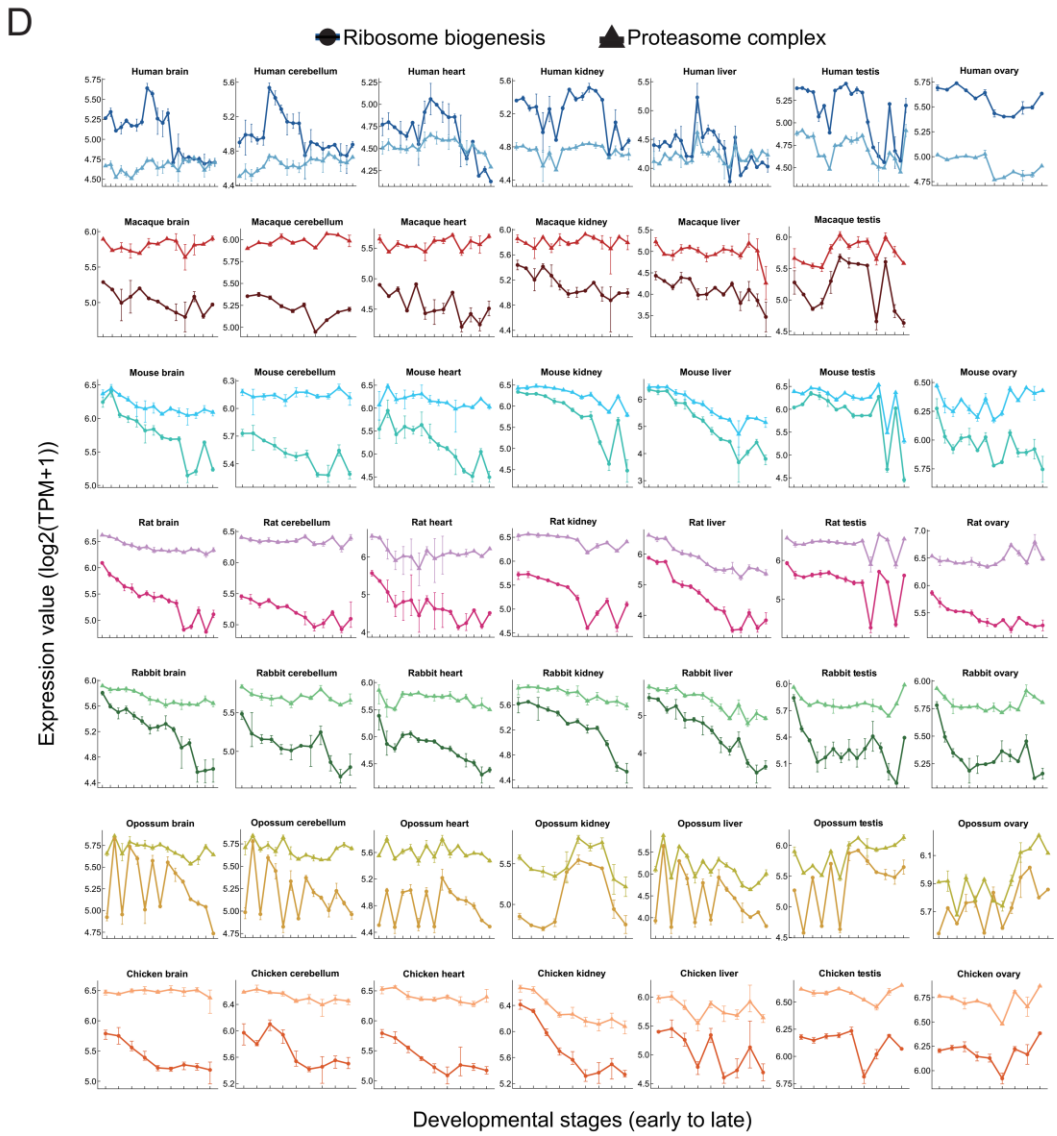


Figure S2. Functional pathways for which cell-to-cell heterogeneity in expression correlates with proliferation rate across cell types and species.

(A) GSEA result plot of Go preribosome genes set for ESC. **(B)** The heatmap (right panel) shows the expression (z-scored read counts) of preribosome genes in ESCs across four biological replicates of the CFSE sorting experiment. **(C)** Higher expression of Myc in both fast proliferating ESCs and fibroblasts. log2 fold change of Myc expression between fast and slow proliferating subpopulation in both ESCs and fibroblasts, each cell type 4 replicates. **(D)** Correlated changes in the expression of ribosome biogenesis and proteasome related genes during organ development. Change of average expression of log2(TPM+1) of genes in ribosome biogenesis (Go preribosome) gene set and proteasome complex (Go proteasome complex) gene set with developmental stages across different organs in seven species[16]. Points (circle and triangle) are the mean expression of replicates, error bars represent the maximum and minimum value in the replicates.

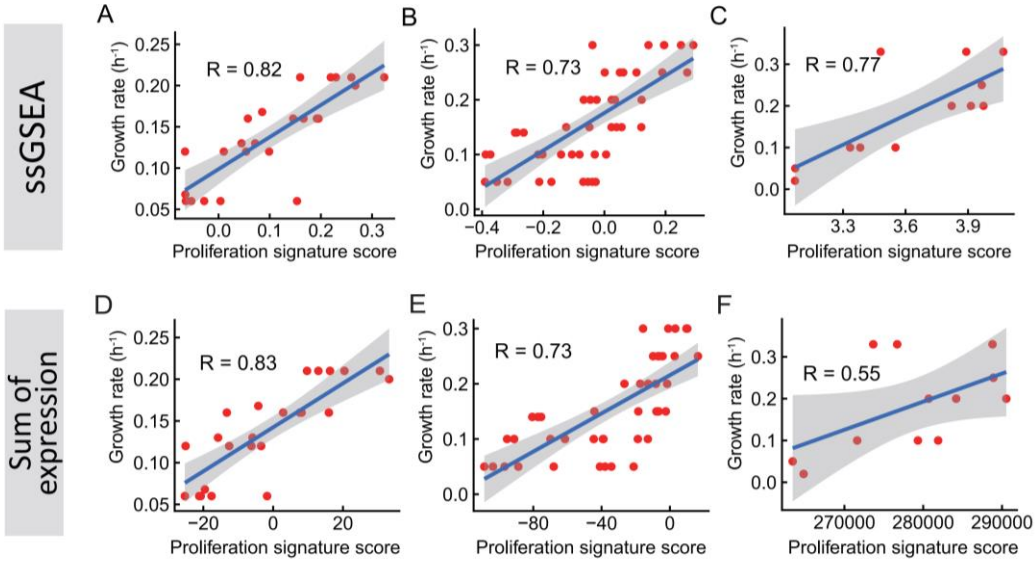


Figure S3. Proliferation signature scores predict growth rate, using different methods of calculation, and different species.

(A-C) Using the Normalized Enrichment Score from ssGSEA to predict growth rate in three different data sets. The Pearson correlation of proliferation signature score with growth rate in are R = 0.82 ($p = 8.9\times10^{-7}$), R = 0.73 ($p = 1.3\times10^{-8}$) and R = 0.77 ($p = 3.7\times10^{-3}$). **(D-F)** Similar to figure A-C, but using the sum of expression values for all genes in the proliferation signature gene set to calculate proliferation signature score. The Pearson correlation of proliferation signature score with growth rate are R = 0.83 ($p = 7.0\times10^{-7}$), R = 0.73 ($p = 1.6\times10^{-8}$) and R = 0.55 ($p = 0.65\times10^{-2}$).

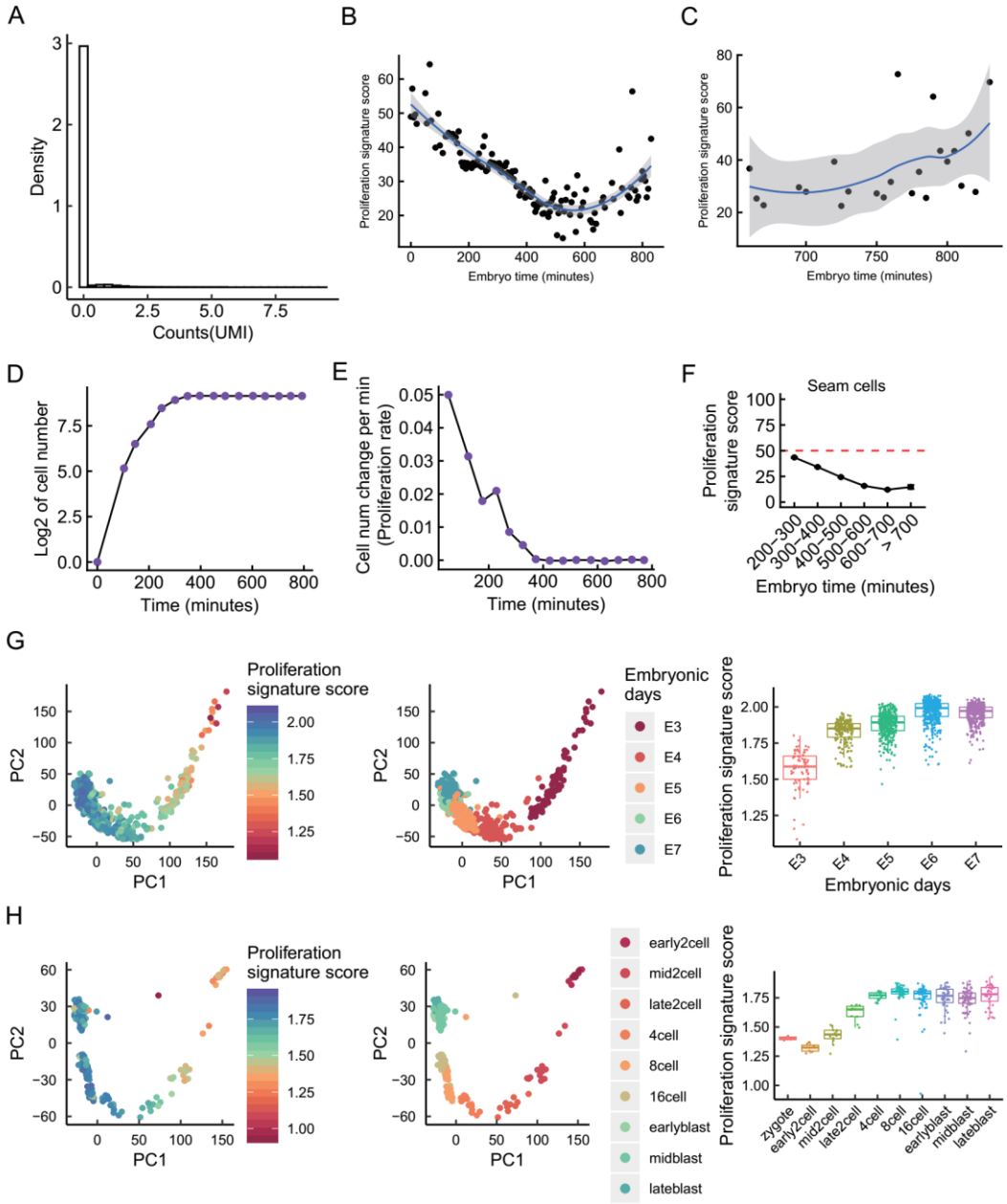


Figure S4. Lineage-specific proliferation signature scores during *C. elegans* development. **(A)** A density histogram of counts (UMI) across 1000 randomly sampled cells; 95.6% of genes have zero reads. This causes ssGSEA to give unreliable results, so the sum of expression values method is used for calculating the proliferation signature score for single cells. **(B)** Cells are binned by embryo time, and the mean proliferation signature score for all cells not the three outlier cell types (germline, intestine and M cells). The Spearman correlation is rho = -0.73 ($p = 8.7\times10^{-24}$) for binned data, and rho = -0.45 ($p < 2.2\times10^{-16}$) for unbinned data. **(C)** Similar to figure 5C, but only showing cells with an age higher than 650 minutes, rho = 0.5 ($p = 1.5\times10^{-2}$). **(D, E)** The change in cell number, and the rate of change in cell number, during development, as measured by microscopy[67]. **(F)** Change in proliferation signature score for

seam cells, which form multinucleated cells through cell-fusion. **(G)** Single-cell gene expression data from Petropoulos et al.[69] projected onto the first two principal components and colored by proliferation signature score or developmental stages. And boxplot shows the change of proliferation signature score with developmental stages. **(H)** Similar to figure G, but use scRNA-seq data from Deng et al.[70].

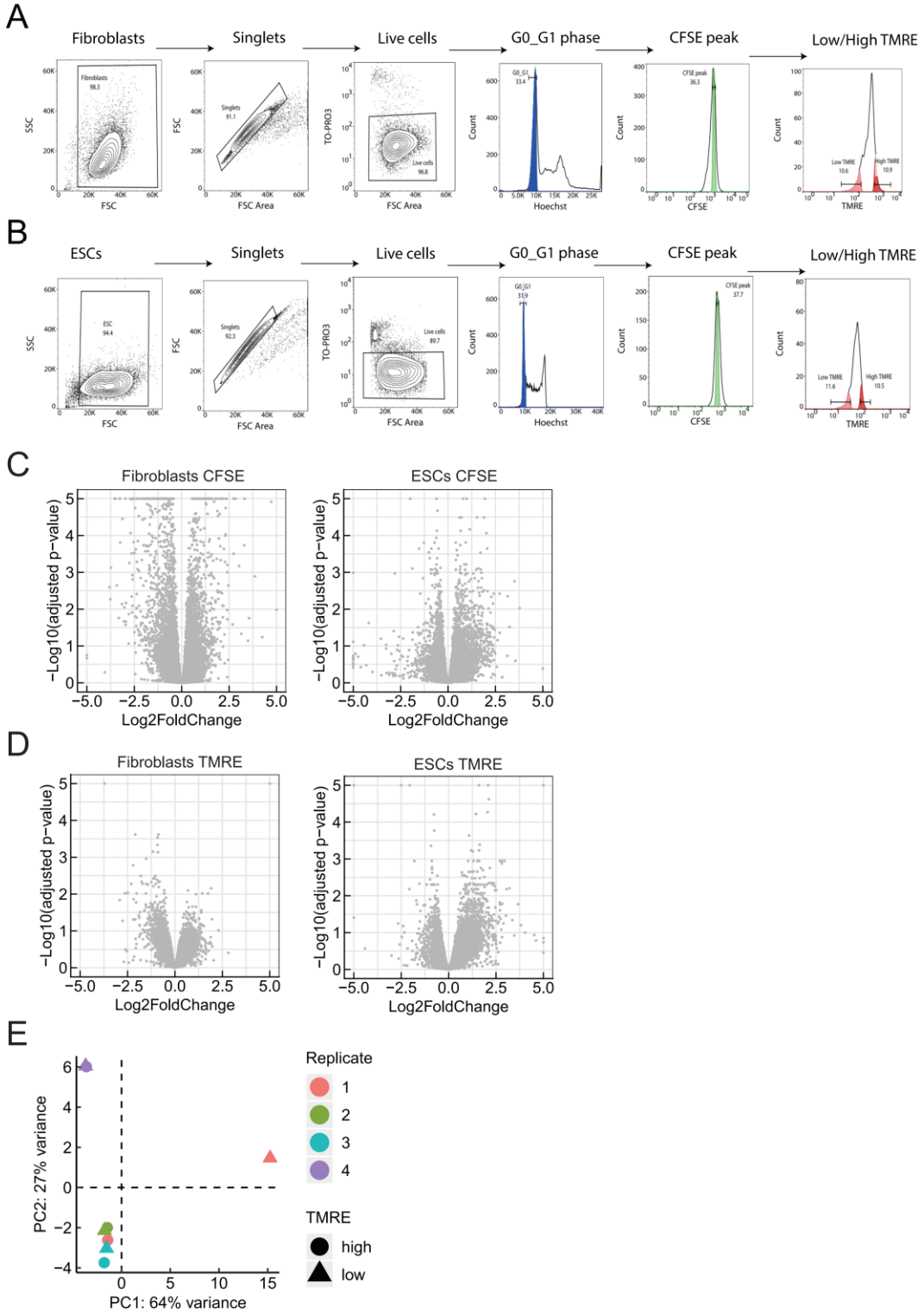


Figure S5. FACS gating strategy for TMRE staining and volcano plots for fibroblasts and ESCs sorted by CFSE or TMRE.

(A, B) The gating strategy for TMRE staining to get cell subpopulations with different mitochondrial states in fibroblasts **(A)** and in ESCs **(B)**. We use Hoechst to get cells in G0/G1, gate by CFSE to get a more uniform cell population, and separate populations with high and low TMRE signal, then do RNA-seq on each of the two subpopulations. **(C, D)** Deseq2 was

used to calculate log2 fold change and adjusted p-values for CFSE sorting **(C)** and TMRE sorting **(D)**, combining biological replicates. To set the axes to be maximally informative, genes with $p < 10^{-5}$ had p set to 10^{-5}, and those abs(log2 fold change) > 5 were truncated at -5 or +5. **(E)** PCA for RNA-seq data of ESCs sorted by TMRE. Low TMRE ESCs of replicate 1 is an outlier, so we remove replicate 1 for all analysis.

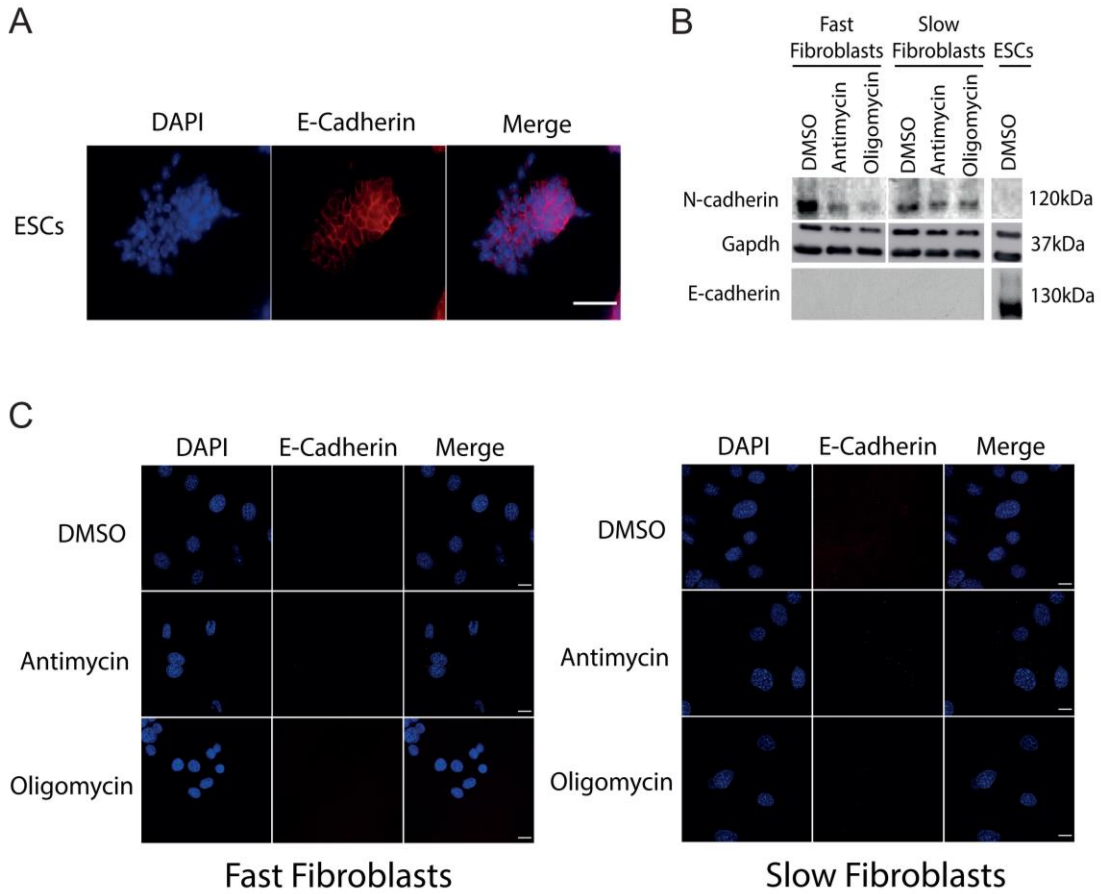
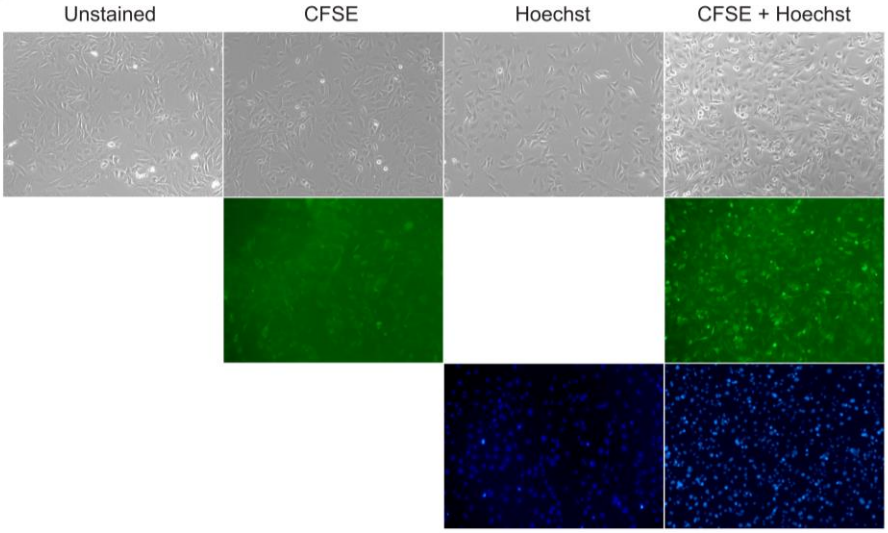


Figure S6. N-cadherin and E-cadherin levels in oligomycin and antimycin treated cells. **(A)** E-cadherin staining of ESCs as positive control for E-cadherin detection, Scale bar = 80µm. **(B)** Western blot for N-cadherin and E-cadherin (Gapdh = loading control) in DMSO- and drug-treated fibroblasts and ESCs. **(C)** Immunostaining for E-cadherin does not show detectable levels in fibroblasts, Scale bars = 15µm.

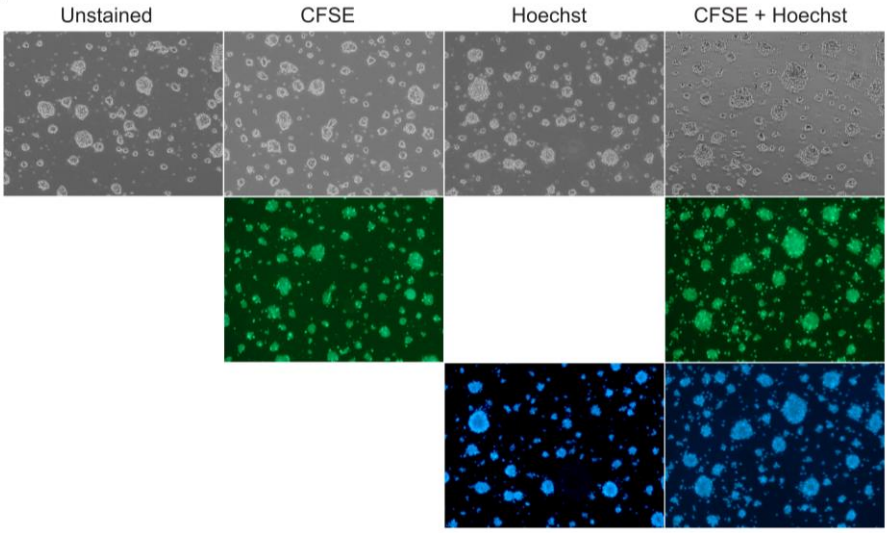
A



	Seeded cell number	Harvested cell number (after 24h)	Viability	Doubling time
Unstained	150000	$3.4x10^5$	97%	≈ 20.3
CFSE	150000	$3.52x10^5$	98%	≈ 19.5
Hoechst	150000	$3.34x10^5$	98%	≈ 20.7
CFSE+Hoechst	150000	$3.46x10^5$	96%	≈ 19.9

1041

B

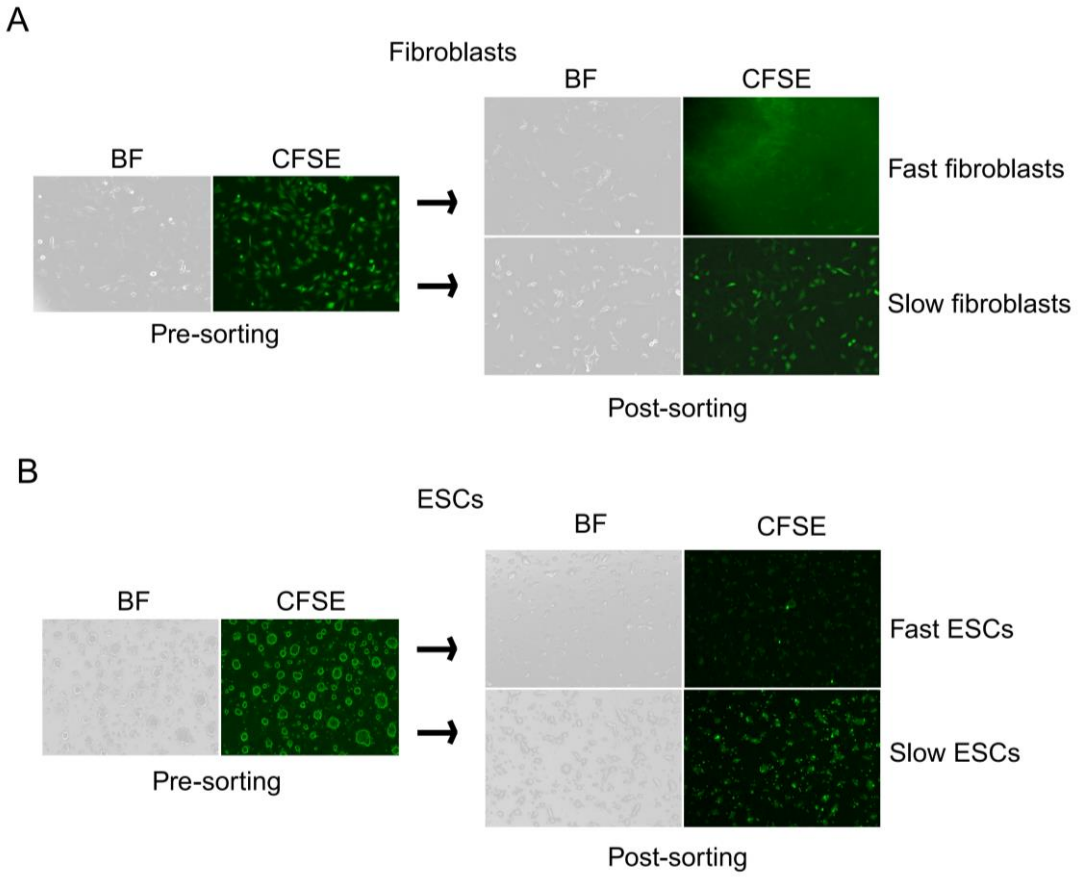


	Seeded cell number	Harvested cell number (after 24h)	Viability	Doubling time
Unstained	750000	$2.8x10^6$	90%	≈ 12.6
CFSE	750000	$2.60x10^6$	89%	≈ 13
Hoechst	750000	$2.72x10^6$	88%	≈ 12.9
CFSE + Hoechst	750000	$2.65x10^6$	90%	≈ 13.1

1042

1043 **Figure S7. The effects of Hoechst and CFSE staining on cell viability and proliferation rates.**
1044 Shown are the estimated doubling times (based on the increased in cell number after 24hrs
1045 growth) and measured viability (trypan blue) for fibroblasts **(A)** and ESCs **(B)**. The microscopy

1046 shows that stained cells maintain the correct morphology.
1047
1048



1049
1050 **Figure S8. The images of CFSE staining for both fast and slow proliferating fibroblasts and**
1051 **ESCs.**
1052 Fibroblasts **(A)** and ESCs **(B)** were stained by CFSE and sorted into two bins: fast proliferating
1053 cells (low CFSE) and slow proliferating cells (high CFSE).
1054

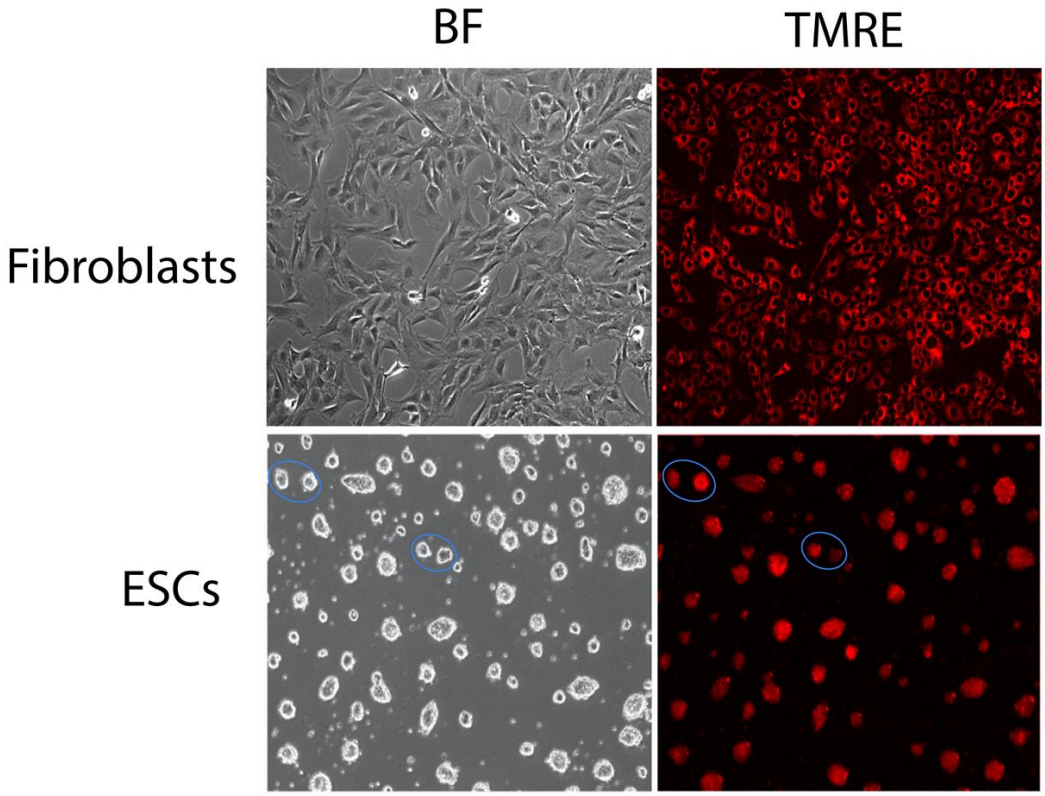


Figure S9. Images of TMRE staining for fibroblasts and ESCs, showing heterogeneity. Bright-field and TMRE staining images for both Fibroblasts and ESCs. Two pairs of ESC colonies of similar size but showing staining heterogeneity is circled in blue.

References

1. Min, M. and S.L. Spencer, *Spontaneously slow-cycling subpopulations of human cells originate from activation of stress-response pathways.* PLOS Biology, 2019. **17**(3): p. e3000178.
2. Wakamoto, Y., et al., *Dynamic persistence of antibiotic-stressed mycobacteria.* Science, 2013. **339**(6115): p. 91-5.
3. Fridman, O., et al., *Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations.* Nature, 2014. **513**(7518): p. 418-21.
4. Balaban, N.Q., et al., *A problem of persistence: still more questions than answers?* Nature Reviews Microbiology, 2013. **11**(8): p. 587-591.
5. Gupta, P.B., et al., *Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells.* Cell, 2011. **146**(4): p. 633-44.
6. Brown, R., et al., *Poised epigenetic states and acquired drug resistance in cancer.* Nat Rev Cancer, 2014. **14**(11): p. 747-53.
7. Marusyk, A., V. Almendro, and K. Polyak, *Intra-tumour heterogeneity: a looking glass for cancer?* Nat Rev Cancer, 2012. **12**(5): p. 323-34.
8. van Dijk, D., et al., *Slow-growing cells within isogenic populations have increased RNA*

1079 *polymerase error rates and DNA damage.* Nat Commun, 2015. **6**: p. 7972.
1080 9. Dhar, R., et al., *Single cell functional genomics reveals the importance of mitochondria in*
1081 *cell-to-cell phenotypic variation.* eLife, 2019. **8**: p. e38904.
1082 10. Yaakov, G., et al., *Coupling phenotypic persistence to DNA damage increases genetic*
1083 *diversity in severe stress.* Nat Ecol Evol, 2017. **1**(1): p. 16.
1084 11. Paek, A.L., et al., *Cell-to-Cell Variation in p53 Dynamics Leads to Fractional Killing.* Cell,
1085 2016. **165**(3): p. 631-42.
1086 12. Brauer, M.J., et al., *Coordination of Growth Rate, Cell Cycle, Stress Response, and*
1087 *Metabolic Activity in Yeast.* Molecular Biology of the Cell, 2007. **19**(1): p. 352-367.
1088 13. Regenberg, B., et al., *Growth-rate regulated genes have profound impact on*
1089 *interpretation of transcriptome profiling in Saccharomyces cerevisiae.* Genome biology,
1090 2006. **7**(11): p. R107-R107.
1091 14. Im, H.K., et al., *Mixed effects modeling of proliferation rates in cell-based models:*
1092 *consequence for pharmacogenomics and cancer.* PLoS genetics, 2012. **8**(2): p.
1093 e1002525-e1002525.
1094 15. Choy, E., et al., *Genetic Analysis of Human Traits In Vitro: Drug Response and Gene*
1095 *Expression in Lymphoblastoid Cell Lines.* PLOS Genetics, 2008. **4**(11): p. e1000287.
1096 16. Cardoso-Moreira, M., et al., *Gene expression across mammalian organ development.*
1097 Nature, 2019.
1098 17. Venet, D., J.E. Dumont, and V. Detours, *Most random gene expression signatures are*
1099 *significantly associated with breast cancer outcome.* PLoS computational biology, 2011.
1100 **7**(10): p. e1002240-e1002240.
1101 18. Levy, S.F., N. Ziv, and M.L. Siegal, *Bet hedging in yeast by heterogeneous, age-correlated*
1102 *expression of a stress protectant.* PLoS Biol, 2012. **10**(5): p. e1001325.
1103 19. Johnston, I.G., et al., *Mitochondrial Variability as a Source of Extrinsic Cellular Noise.* PLOS
1104 Computational Biology, 2012. **8**(3): p. e1002416.
1105 20. das Neves, R.P., et al., *Connecting variability in global transcription rate to mitochondrial*
1106 *variability.* PLoS biology, 2010. **8**(12): p. e1000560-e1000560.
1107 21. Sukumar, M., et al., *Mitochondrial Membrane Potential Identifies Cells with Enhanced*
1108 *Stemness for Cellular Therapy.* Cell Metabolism, 2016. **23**(1): p. 63-76.
1109 22. Mathieu, J. and H. Ruohola-Baker, *Metabolic remodeling during the loss and acquisition*
1110 *of pluripotency.* Development, 2017. **144**(4): p. 541.
1111 23. Nichols, J. and A. Smith, *Naive and Primed Pluripotent States.* Cell Stem Cell, 2009. **4**(6): p.
1112 487-492.
1113 24. Wray, J., T. Kalkan, and Austin G. Smith, *The ground state of pluripotency.* Biochemical
1114 Society Transactions, 2010. **38**(4): p. 1027.
1115 25. Kolodziejczyk, Aleksandra A., et al., *Single Cell RNA-Sequencing of Pluripotent States*
1116 *Unlocks Modular Transcriptional Variation.* Cell Stem Cell, 2015. **17**(4): p. 471-485.
1117 26. ter Huurne, M., et al., *Distinct Cell-Cycle Control in Two Different States of Mouse*
1118 *Pluripotency.* Cell Stem Cell, 2017. **21**(4): p. 449-455.e4.
1119 27. Nair, G., et al., *Heterogeneous lineage marker expression in naive embryonic stem cells is*
1120 *mostly due to spontaneous differentiation.* Scientific Reports, 2015. **5**: p. 13339.
1121 28. Abranches, E., et al., *Stochastic NANOG fluctuations allow mouse embryonic stem cells to*
1122 *explore pluripotency.* Development (Cambridge, England), 2014. **141**(14): p. 2770-2779.

29. Wytock, T.P. and A.E. Motter, *Predicting growth rate from gene expression.* Proc Natl Acad Sci U S A, 2019. **116**(2): p. 367-372.
30. Smith, A., *10 Embryonic Stem Cells.* Cold Spring Harbor Monograph Archive; Volume 40 (2001): Stem Cell Biology, 2001: p. 205-230.
31. Cho, B.K., et al., *Homeostasis-stimulated proliferation drives naive T cells to differentiate directly into memory T cells.* J Exp Med, 2000. **192**(4): p. 549-56.
32. Parish, C.R., *Fluorescent dyes for lymphocyte migration and proliferation studies.* Immunology & Cell Biology, 1999. **77**(6): p. 499-508.
33. Weston, S.A. and C.R. Parish, *New fluorescent dyes for lymphocyte migration studies: Analysis by flow cytometry and fluorescence microscopy.* Journal of Immunological Methods, 1990. **133**(1): p. 87-97.
34. Romano, P., et al., *Cell Line Data Base: structure and recent improvements towards molecular authentication of human cell lines.* Nucleic acids research, 2009. **37**(Database issue): p. D925-D932.
35. Tamm, C., S. Pijuan Galitó, and C. Annerén, *A Comparative Study of Protocols for Mouse Embryonic Stem Cell Culturing.* PLOS ONE, 2013. **8**(12): p. e81156.
36. Ying, Q.-L., et al., *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**: p. 519.
37. Hayashi, K., et al., *Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states.* Cell stem cell, 2008. **3**(4): p. 391-401.
38. Toyooka, Y., et al., *Identification and characterization of subpopulations in undifferentiated ES cell culture.* Development, 2008. **135**(5): p. 909.
39. Chambers, I., et al., *Nanog safeguards pluripotency and mediates germline development.* Nature, 2007. **450**: p. 1230.
40. Nair, G., et al., *Heterogeneous lineage marker expression in naive embryonic stem cells is mostly due to spontaneous differentiation.* Scientific reports, 2015. **5**: p. 13339-13339.
41. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545.
42. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics (Oxford, England), 2011. **27**(12): p. 1739-1740.
43. Athanasiadou, R., et al., *Growth Rate-Dependent Global Amplification of Gene Expression.* bioRxiv, 2016: p. 044735.
44. Kumatori, A., et al., *Abnormally high expression of proteasomes in human leukemic cells.* Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(18): p. 7071-7075.
45. Chen, L. and K. Madura, *Increased Proteasome Activity, Ubiquitin-Conjugating Enzymes, and eEF1A Translation Factor Detected in Breast Cancer Tissue.* Cancer Research, 2005. **65**(13): p. 5599.
46. Arlt, A., et al., *Increased proteasome subunit protein expression and proteasome activity in colon cancer relate to an enhanced activation of nuclear factor E2-related factor 2 (Nrf2).* Oncogene, 2009. **28**: p. 3983.
47. Cetin, B. and D.W. Cleveland, *How to survive aneuploidy.* Cell, 2010. **143**(1): p. 27-29.
48. Iadevaia, V., R. Liu, and C.G. Proud, *mTORC1 signaling controls multiple steps in*

ribosome biogenesis. Semin Cell Dev Biol, 2014. **36**: p. 113-20.

49. Zhang, Y., et al., *Coordinated regulation of protein synthesis and degradation by mTORC1.* Nature, 2014. **513**(7518): p. 440-3.
50. Lempiainen, H. and D. Shore, *Growth control and ribosome biogenesis.* Curr Opin Cell Biol, 2009. **21**(6): p. 855-63.
51. Choi, J.-H., et al., *mTORC1 accelerates retinal development via the immunoproteasome.* Nature Communications, 2018. **9**(1): p. 2502.
52. Uprety, B., A. Kaja, and S.R. Bhaumik, *TOR Facilitates the Targeting of the 19S Proteasome Subcomplex To Enhance Transcription Complex Assembly at the Promoters of the Ribosomal Protein Genes.* Mol Cell Biol, 2018. **38**(14).
53. Yun, Y.S., et al., *mTORC1 Coordinates Protein Synthesis and Immunoproteasome Formation via PRAS40 to Prevent Accumulation of Protein Stress.* Mol Cell, 2016. **61**(4): p. 625-639.
54. Zhang, Y., et al., *Rapamycin extends life and health in C57BL/6 mice.* The journals of gerontology. Series A, Biological sciences and medical sciences, 2014. **69**(2): p. 119-130.
55. Zhao, J., G.A. Garcia, and A.L. Goldberg, *Control of proteasomal proteolysis by mTOR.* Nature, 2016. **529**: p. E1.
56. Gearhart, J., E.E. Pashos, and M.K. Prasad, *Pluripotency redux--advances in stem-cell research.* N Engl J Med, 2007. **357**(15): p. 1469-72.
57. Dang, C.V., *MYC, metabolism, cell growth, and tumorigenesis.* Cold Spring Harbor perspectives in medicine. **3**(8): p. a014217.
58. *Drosophila myc Regulates Cellular Growth during Development.* Cell, 1999. **98**(6): p. 779 - 790.
59. van Riggelen, J., A. Yetil, and D.W. Felsher, *MYC as a regulator of ribosome biogenesis and protein synthesis.* Nat Rev Cancer, 2010. **10**(4): p. 301-9.
60. Csibi, A., et al., *The mTORC1/S6K1 pathway regulates glutamine metabolism through the eIF4B-dependent control of c-Myc translation.* Curr Biol, 2014. **24**(19): p. 2274-80.
61. Liu, P., et al., *A functional mammalian target of rapamycin complex 1 signaling is indispensable for c-Myc-driven hepatocarcinogenesis.* Hepatology, 2017. **66**(1): p. 167-181.
62. Riley, R.D., et al., *External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges.* BMJ, 2016. **353**: p. i3140.
63. Waldman, Y.Y., T. Geiger, and E. Ruppin, *A genome-wide systematic analysis reveals different and predictive proliferation expression signatures of cancerous vs. non-cancerous cells.* PLoS Genet, 2013. **9**(9): p. e1003806.
64. Macfarlan, T.S., et al., *Embryonic stem cell potency fluctuates with endogenous retrovirus activity.* Nature, 2012. **487**(7405): p. 57-63.
65. Hicks, S.C., et al., *Missing data and technical variability in single-cell RNA-sequencing experiments.* Biostatistics, 2017. **19**(4): p. 562-578.
66. Packer, J.S., et al., *A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution.* Science, 2019: p. eaax1971.
67. Sulston, J.E., et al., *The embryonic cell lineage of the nematode Caenorhabditis elegans.* Dev Biol, 1983. **100**(1): p. 64-119.

68. Altun, Z.F., Herndon, L.A., Wolkow, C.A., Crocker, C., Lints, R. and Hall, D.H. *WormAtlas*. (ed.s) 2002-2020.
69. Petropoulos, S., et al., *Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos.* Cell, 2016. **167**(1): p. 285.
70. Deng, Q., et al., *Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells.* Science, 2014. **343**(6167): p. 193-6.
71. Samavarchi-Tehrani, P., et al., *Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming.* Cell Stem Cell, 2010. **7**(1): p. 64-77.
72. Li, R., et al., *A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts.* Cell Stem Cell, 2010. **7**(1): p. 51-63.
73. Sun, H., et al., *Metabolic switch and epithelial-mesenchymal transition cooperate to regulate pluripotency.* EMBO J, 2020. **39**(8): p. e102961.
74. Conn, C.S. and S.B. Qian, *Nutrient signaling in protein homeostasis: an increase in quantity at the expense of quality.* Sci Signal, 2013. **6**(271): p. ra24.
75. Valvezan, A.J. and B.D. Manning, *Molecular logic of mTORC1 signalling as a metabolic rheostat.* Nature Metabolism, 2019. **1**(3): p. 321-333.
76. Teslaa, T. and M.A. Teitell, *Pluripotent stem cell energy metabolism: an update.* The EMBO journal, 2015. **34**(2): p. 138-153.
77. Lu, V. and M.A. Teitell, *Alpha-ketoglutarate: a "magic" metabolite in early germ cell development.* 2019. **38**(1): p. e100615.
78. Tischler, J., et al., *Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate.* 2019. **38**(1): p. e99518.
79. Brown, M., et al., *A recombinant murine retrovirus for simian virus 40 large T cDNA transforms mouse fibroblasts to anchorage-independent growth.* Journal of virology, 1986. **60**(1): p. 290-293.
80. Minajigi, A., et al., *A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation.* Science, 2015: p. aab2276.
81. Lee, J.T. and N. Lu, *Targeted mutagenesis of Tsix leads to nonrandom X inactivation.* Cell, 1999. **99**(1): p. 47-57.
82. Bray, N.L., et al., *Erratum: Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(8): p. 888.
83. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
84. Kolodziejczyk, A.A., et al., *Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation.* Cell Stem Cell, 2015. **17**(4): p. 471-85.
85. Boroviak, T., et al., *The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification.* Nat Cell Biol, 2014. **16**(6): p. 516-28.
86. Young, R.A., *Control of the embryonic stem cell state.* Cell, 2011. **144**(6): p. 940-954.
87. Crowley, L.C., M.E. Christensen, and N.J. Waterhouse, *Measuring Mitochondrial Transmembrane Potential by TMRE Staining.* Cold Spring Harb Protoc, 2016. **2016**(12).
88. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-40.
89. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for*

1255 *interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p.
1256 15545-50.
1257 90. Shannon, P., et al., *Cytoscape: a software environment for integrated models of*
1258 *biomolecular interaction networks.* Genome Res, 2003. **13**(11): p. 2498-504.
1259 91. Merico, D., et al., *Enrichment map: a network-based method for gene-set enrichment*
1260 *visualization and interpretation.* PLoS One, 2010. **5**(11): p. e13984.
1261 92. Reimand, J., et al., *Pathway enrichment analysis and visualization of omics data using*
1262 *g:Profiler, GSEA, Cytoscape and EnrichmentMap.* Nat Protoc, 2019. **14**(2): p. 482-517.
1263 93. Durinck, S., et al., *BioMart and Bioconductor: a powerful link between biological*
1264 *databases and microarray data analysis.* Bioinformatics, 2005. **21**(16): p. 3439-3440.
1265 94. Durinck, S., et al., *Mapping identifiers for the integration of genomic datasets with the*
1266 *R/Bioconductor package biomaRt.* Nat Protoc, 2009. **4**(8): p. 1184-91.
1267 95. Airoldi, E.M., et al., *Steady-state and dynamic gene expression programs in*
1268 *Saccharomyces cerevisiae in response to variation in environmental nitrogen.* Mol Biol
1269 Cell, 2016. **27**(8): p. 1383-96.
1270 96. Slavov, N. and D. Botstein, *Coupling among growth rate response, metabolic cycle, and*
1271 *cell division cycle in yeast.* Mol Biol Cell, 2011. **22**(12): p. 1997-2009.
1272 97. Barbie, D.A., et al., *Systematic RNA interference reveals that oncogenic KRAS-driven*
1273 *cancers require TBK1.* Nature, 2009. **462**(7269): p. 108-12.
1274 98. Hanzelmann, S., R. Castelo, and J. Guinney, *GSVA: gene set variation analysis for*
1275 *microarray and RNA-seq data.* BMC Bioinformatics, 2013. **14**: p. 7.
1276 99. Rohatgi, A., *WebPlotDigitizer*. 2017, Austin, Texas, USA.
1277