SUPPLEMENTARY TEXT

Title: **Global variability in gene expression and alternative splicing is modulated by mitochondrial content.**

Running title: **Non-genetic variability modulated by mitochondria**

Authors: Raul Guantes*[3], Alberto Rastrojo*[4], Ricardo Neves*[2,5], Ana Lima[5], Begoña Aguado[4], Francisco J Iborra[1,2]

Address: 1- Centro Nacional de Biotecnología, CSIC,
Darwin 3, Campus de Cantoblanco,
28049 Madrid, Spain
2- MRC Molecular Haematology Unit,
Weatherall Institute of Molecular Medicine,
John Radcliffe Hospital, University of Oxford
Headington, Oxford, OX3 9DS, UK
3- Department of Condensed Matter Physics, Materials Science Institute 'Nicolás Cabrera' and Institute of Condensed Matter Physics (IFIMAC)
Universidad Autónoma de Madrid; Campus de Cantoblanco,
28049 Madrid, Spain
4- Centro Biología Molecular "Severo Ochoa", CSIC-UAM,
C/ Nicolás Cabrera 1; Campus de Cantoblanco,
28049 Madrid, Spain
5- Present address: CNC - Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal; 2, Biocant - Center of Innovation in Biotechnology, 3060-197 Cantanhede, Portugal

*These authors contributed equally

Correspondence: FJ Iborra
Telephone : +0034915854845
E Mail: fjiborra@cnb.csic.es

Key words: RNA polymerase, transcription, splicing, translation, mitochondria, non-genetic variability

Supplementary Materials and Methods

Statistical analysis of mitochondrial contribution to variability

We used three different statistical methods to estimate the contribution of variability in mitochondrial content to total protein variability. The method described in the main text estimates the mitochondrial contribution to variability (MCV) as a ratio of two variabilities: the variability of the original protein distribution (quantified by the normalized interquartile-range of the distribution) and the variability of the protein distribution once the linear correlation with mitochondrial content has been removed (IQR ratio in Figure S2). The second method estimates the MCV simply as the fraction of variance explained by a linear regression model(Freedman 2009) where the mitochondrial content is the independent or explanatory variable (R^2 in Figure S2). The third method estimates the MCV as a ratio of the variance associated to mitochondrial variability and the total variance (Var. Ratio in Figure S2). We follow the procedure used by Snijder et al.(Snijder et al. 2009) and divide the CmxRos-Protein data in bins with equal number of points along the CmxRos axis. The MCV is then calculated as:

$$MCV = 100 \times \frac{\sigma^2_{between}}{\sigma^2_{within} + \sigma^2_{between}}$$

where $\sigma^2_{between}$ is the variance between the average values of each bin and σ^2_{within} gives the variance of each bin. The three methods yield very similar estimates (Figure S2).

Relative influence of cell cycle state and mitochodrial content on protein variability in single cells.

HeLa cells were stained with fluorescent probes for mitochondrial content (CmxRos), DNA content (DAPI, used as a reporter of cell cycle state), and different protein antibodies, and the levels of each marker were quantified by single-cell live microscopy. First we checked that the distribution of DAPI values in a cell population reflected the cell cycle stage (showing two peaks in G0/G1 and G2/M phases). We then analyzed the simultaneous influence of mitochondrial levels and DAPI content on six different proteins of our house-keeping genes set (Figure S3).

The correlation between protein amount and DAPI is much weaker than that between protein and mitochondria, Figure S3A-B. Moreover, there is also a positive correlation between mitochondrial content and cell cycle state (Figure S3C). If we consider cell cycle state and mitochondrial levels as variables (the predictors or independent variables) influencing protein variability (the outcome or dependent variable), associations or correlations between the independent variables must be removed when asking how much of the variability in outcome can be explained by each predictor(Shipley 2004).

Partial correlations provide a means of assessing the relative importance of each predictor in determining the outcome, and the relationships between variables can be visualized as undirected graphs where the nodes are the variables and the edges represent the partial correlations between them (a procedure also known as graphical Gaussian modelling(Opgen-Rhein and Strimmer 2007; Krumsiek et al. 2011), Figure S3D-I. All proteins analyzed show a much stronger dependence on mitochondrial content than on cell cycle state. Partial correlations represent the contribution of each predictor to a dependent variable after the contributions of the other predictors have been removed both from that predictor and the dependent variable, but it is not the only useful measure of dependence. Other useful measures are semipartial correlations (where the contributions of other predictors are removed from the predictor in question, but not from the dependent variable), and LMG coefficients (representing a natural decomposition of the total regression coefficient in the contributions from different independent variables(Lindeman et al. 1980)). Figure S3J displays the average values (for the six proteins) of these different statistical measures of causal association, showing the weak dependence of protein on cell cycle state, and the strong dependence on mitochondrial content.

FLIP experiments and quantification of RNA Pol II rate constants.

A clone stably expressing GFP–RNA Pol II (C23) (das Neves et al., 2010; Hieda et al., 2005; Sugaya et al., 2000) was cultured at 39°C, and images were collected with the microscope stage heated to 39°C. Fluorescence images were collected using a confocal microscope (Zeiss LSM 510 META), with an EC PlnN 40×/1.3 oil objective, with the pinhole completely open. We selected a rectangle at the bottom

half of each nuclei where we applied 100% laser power, in order to bleach all the fluorescent molecules in the rectangle. This operation was repeated every 5 s for a period of 1,200 s, and we analysed the decay of the fluorescence in the unbleached top half. Fluorescence intensity was analysed in MetaMorph 6.1 (Universal Imaging). Curves were analysed using Sigma Plot 11.0 for Windows. For the analysis we assumed that there were three populations: freely diffusible, bound to DNA and fully engaged in transcription (see model in Figure 4). FLIP analysis allows to estimate the dissociation constants of the different populations. In order to associate these constants to a functional process, we treat cells with inhibitors of specific steps of the transcription cycle, and then see which kinetic regime is affected. In this way, we can link a decay constant to a specific transcription step (Hieda et al. 2005, Sugaya et al. 2000).

For the fitting to the exponential decays in Figure S4b, we allowed the three components to optimize with no restriction (always r^2>0.99). These fittings allow us to estimate the fraction of the different populations and the rate constants K2 and K4 (Figure 4 and S4g). To obtain the other kinetic constants (K1 and K3) we fitted our data to the steady-state solution of the kinetic model in Figure 4a. We were concerned with the possible artifacts induced by FLIP. Therefore, transcription “run on” experiments were performed on photobleached cells, which demonstrated no alteration in the transcription pattern or intensity in the bleached area (data not shown).

Deoxy-Glucose experiments were performed by incubation for 60 min in 25 mM Deoxy-Glucose plus 10 mM Sodium Azide, previous to the FLIP. Under this condition ATP levels were depleted by 97%.

To assess the possible role of glycolysis in ATP production in our cell line, we performed experiments (Figure S4c) where cells were incubated in the absence of glucose and presence of galactose, showing no difference in ATP content with respect to control cells.

Estimation of relative mtDNA levels by qPCR

HeLa cells were sorted according to their mitochondrial content into High (H) and Low (L) subpopulations (inset in Figure S1a). Then, the DNA was extracted using the Illustra DNA Extraction Kit (BACC2) (RPN-8502) following manufacturer

instructions. DNA was quantified by qPCR using the primers and TaqMan fluorogenic probes specified in Table S2 (Ashley et al. 2005).

Table S2. Sequence of pairs of primers used for qPCR

Target name	Forward primer sequence	Reverse primer sequence
APP (nuclear)	TTTTTGTGTGCTCTCCCAGGTCT	TGGTCACTGGTTGGTTGGC
MitHu (mitoch.)	AGGACAAGAGAAATAAGGCC	AAGAAGAGGAATTGAACCTCTGACTG

cDNA synthesis and qRT-PCR

The isolation and purification of total RNA was performed with TRIzol® Reagent (Invitrogen) following the manufacturer's instructions. For cDNA synthesis 1µg total RNA was incubated with MultiScribe™ Reverse Transcriptase (Applied Biosystems®) according to the manufacturer's instructions. Quantitative RT-PCR (qRT-PCR) was performed using Power SYBR® Green PCR Master Mix (Applied Biosystems®) and data were collected on an Applied Biosystems® 7500 Real-Time PCR system. The sequences of the primers used for specific target amplification are provided in Table S3.

Table S3. Sequence of pairs of primers used for qRT-PCR.

Target name	Forward primer sequence	Reverse primer sequence
ALDOC	ACTCGTACCCAGCCCTTTCT	TGTGTTTTCCACCCCAATTT
FDFT1	ATAACCAATGCACTGCACCA	ATAACAGGCAGCCAAAGTGG
GTF2IRD1	CACTAAGGGAGCAGGTCCAG	CTCTTGATCCGCTTGTAGGG
ALDH3A1	CCTCGGAGCTGAGTGAGAAC	GCCCGTGTACAGGATATGGT
FDPS	CTGCAAGCTTTCTTCCTGGT	GCAGGCGGTAGATACATGCT
RMI2	GATGCAGGGCAGGGTAGT	TGAACCACTCCCATCACCA
NR2F1	GCCTCAAGAAGTGCCTCAAA	GATGTAGCCGGACAGGTAGC
MORC3	GTGCAGCTTGACGATGTGTT	CACACTGTGAACGGATTTGTG
RPS15A	GTGCTTATTAGGCCGTGCTC	TTTAGCCTGCCTGTGAGGTT
PCSK9	TCACCGACTTCGAGAATGTG	GGTGGGTGCCATGACTGT
PLD1	TCCACTTTGAGGGAGAGGAG	GACAGCCGGAGAGATACGTC
SLC2A3	TGAAGGTTTTGTTGGCTGAA	GCCGATTGTAGCAACTGTGA

Mathematical model of variability in alternative splicing

The expression levels of different mature mRNAs out of the same genes are modelled taking into account transcription of pre-mRNA and AS/conversion to mature mRNA as two condensed steps depending on mitochondrial content (Figure 6d). For a single gene with two different mRNA isoforms, M_1 and M_2 , mRNA levels change in time according to the rate equations:

$$\frac{dP}{dt} = k - (\alpha_1 + \alpha_2)P \qquad (S1)$$

$$\frac{dM_1}{dt} = \alpha_1 P - \delta_1 M_1 \qquad (S2)$$

$$\frac{dM_2}{dt} = \alpha_2 P - \delta_2 M_2 \,, \qquad (S3)$$

where k is the transcription rate for the pre-mRNA (P), α_1 and α_2 the conversion rates of the pre-mRNA to isoforms M_1 and M_2 respectively, and δ_1/δ_2 the degradation rates of the mature isoforms M_1/ M_2. The steady-state levels of the two isoforms (quantified by RNA-seq) are given by

$$M_1^* = N_1 \frac{k}{\delta_1} \qquad (S4)$$

$$M_2^* = (1 - N_1) \frac{k}{\delta_2} \,. \qquad (S5)$$

where $N_1 = \alpha_1/(\alpha_1 + \alpha_2)$ gives the fraction of pre-mRNA converted to isoform M_1.
To simulate the experimental map of relative abundance for different isoform pairs (Figure 6c), first we need to assign values to the kinetic parameters and then consider all possible sources of variability caused by mitochondrial content. Since we do not have experimental access to the particular transcription and degradation rates of mRNA forms with altered AS, we statistically sample these parameters from a biologically reasonable distribution. We use a recent genome-wide quantification study of mRNA and protein synthesis and turnover in mouse fibroblasts (Schwanhausser et al. 2011) as a proxy for these distributions. In Figure S8 we show Schwanhausser et al. (Schwanhausser et al. 2011) experimental data for transcription rates (Figure S8a) and mRNA half-lives (Figure S8b) fitted to standard

statistical distributions. The experimental distribution of transcription rates k is well fitted by a log-normal function

$$f_{transc}(k;\mu,\sigma) = \frac{1}{k\sigma\sqrt{2\pi}} e^{\frac{-(\ln k-\mu)^2}{2\sigma^2}}$$

with parameters $\mu = 0.6, \sigma = 1$ (red line in Figure S5a), while the distribution of mRNA half-lives ($\tau_{1/2}$) is fitted with a Gamma probability function

$$f_{half-life}(\tau_{1/2};a,b) = \frac{1}{b^a\,\Gamma(a)} \tau_{1/2}^{a-1} e^{\frac{-\tau_{1/2}}{b}}$$

with parameters $a = 5.1, b = 2.1$ (red line in Figure S5b). The degradation rate δ_i of a particular mRNA form is related to the half-life by $\delta_i = \ln 2/\tau_{1/2}^i$. To get values of steady-state levels for pairs of isoforms, Eqs. (S4) and (S5), we generated independent random values of transcription rate k and degradation rates δ_1/δ_2 within these distributions, while the conversion fraction N_1 is a uniform random number between 0 and 1. We note that the mRNA synthesis and degradation rates given by Schwanhausser et al (Schwanhausser et al. 2011) are mean values averaged over time and cell population, and may differ from the steady-state values obtained from kinetic equations such as those in Eqs. (S1)-(S3) (Schwanhausser et al. 2011). Moreover, independent measurements of RNA synthesis and decay (Dolken et al. 2008; Rabani et al. 2011) yielded shorter RNA half-lives, probably due to differences in cell lines, culture conditions and/or experimental procedures. Nevertheless, since we are interested in reproducing, in a statistical sense, *relative* (fold-change) differences in expression levels due to different mitochondrial content, absolute values of the rate constants are not important provided they are estimated in a consistent way and reproduce the biologically relevant ranges, as is the case in Schwanhausser et al (Schwanhausser et al. 2011) data.

Mitochondrial content can modulate in several ways the different reaction rates. A global source of variation in transcription rate is due to coupling between ATP concentration and transcription elongation by RNA Pol II (das Neves et al. 2010). This induces a sigmoidal dependence between mitochondrial levels (quantified by CmxRos) and transcription elongation (measured as BrU) in a cell population (Figure S6B). To include this observed variability in our model, we first note that both BrU and CmxRos levels are normalized by their average values, so that at the response threshold BrU ~ 1. Thus we assume that the transcription rate k for a given gene is modulated by mitochondrial content as

$$k = k_r \times BrU(Cmx)$$

where k_r is the transcription rate value randomly chosen from the experimental distribution of Figure S8a and *BrU(Cmx)* the sigmoidal function fitted to the experimental data of Figure S6B. To mimic the Low/High cell sorting for RNAseq, and the spread observed in the experimental BrU levels at similar CmxRos values, we vary the transcription rate of the gene at the Low condition according to

$$k_{LOW} = k_r \times (BrU(Cmx_{LOW}) + \sigma_{LOW} \cdot r) \qquad (S6)$$

where Cmx_{LOW} is a random number in the interval $Cmx_{LOW} \in [0.2,0.5]$ (variability in cell sorting at Low), r a normally distributed random number and σ_{LOW} the standard deviation around Cmx_{LOW} (estimated from the error bars in Figure S6B). We use a similar expression to vary the transcription rate at the High condition (k_{HIGH}), with $Cmx_{HIGH} \in [1,2]$.

Mitochondrial levels can also have a global impact on mRNA degradation, affecting steady state mRNA levels. As deduced from Figure S7a, variations in mitochondrial content have a moderate effect on average mRNA half-life ($\tau_{1/2}$ ~46 min for High versus ~65 min for Low conditions). We nevertheless included this variability in the sampled values from the half-life distribution of Figure S8b, along the same lines as for the transcription rate above,

$$\delta_i^{LOW} = \delta_i^r \left(\frac{\delta_{av}^{LOW}}{\bar{\delta}_{av}} + \sigma_{LOW} \cdot r \right) \qquad (S7)$$

where δ_i^r is the random degradation rate for isoform i sampled from the experimental distribution, δ_{av}^{LOW} the population averaged degradation rate obtained from Figure S7a in Low condition (0.64 h^{-1}), $\bar{\delta}_{av} = \delta_{av}^{LOW} + (\delta_{av}^{HIGH} - \delta_{av}^{LOW})/2$ the average decay at mean mitochondrial levels, and σ_{LOW} the variability in average decay within the population at Low condition (which we take as an upper limit of 0.3 h^{-1} in view of Figure S7a). Variability in the High condition is included with a similar expression with δ_{av}^{HIGH}=0.91 h^{-1} and $\sigma_{HIGH} = \sigma_{LOW}$.

We first checked whether this simple model is able to reproduce the experimental alternative splicing map observed in Figure 6c, including only the global variability in transcription elongation and mRNA degradation given by Equations (S6) and (S7). We generated 8,400 random isoform pairs (equal to the experimental number of

genes with at least two isoforms expressed) and calculated steady state levels of these isoforms and fold-change expression,

$$FC_i = \ln\left(\frac{M_i^{HIGH}}{M_i^{LOW}}\right)$$

for each isoform i according to the procedure described above.

The result is shown in Figure S9a, where we plot with solid lines the thresholds delimiting the domains of altered alternative splicing in Figure 6c. Including only global variability in transcription elongation and degradation is not enough to produce strongly altered alternative splicing patterns as seen experimentally. We thus include in a heuristic way other possible sources of variation induced by mitochondria in our basic reaction steps. Apart from RNA Pol II elongation, gene transcription requires a sequence of promoter state transitions, including chromatin remodelling, assembly of the transcription machinery, and clearance of the promoter by RNA polymerase. These intermediate steps may be highly variable (Mao et al. 2010) and be affected by mitochondrial levels directly (by ATP content) or indirectly, for instance through differential expression of transcriptional activators/repressors. Use of alternative promoters may be also highly affected by mitochondrial levels (Figure S10a). These factors could originate effective transcription rates with strong differences even for the same gene. To take into account this within our modelling framework, we took the approach that the transcription rate k could vary in any way within our initial distribution from one condition to another (Low versus High), resampling the transcription rate value at each condition. This variation in transcription rate for the same gene between Low and High gives a dispersion in fold-change along the diagonal (yellow symbols in Figure S9b, where no variation in other kinetic rates is considered), and can be viewed as a *gene specific* contribution to fold-change variability. When adding independently variation in transcription elongation as in Figure S9a (*global* contribution, the same for all genes) fold-change values are scattered out of the diagonal (grey symbols in Figure S9b), giving a noticeable fraction of genes in the UP and DOWN regions of the plane (black and blue symbols respectively) as observed in the experimental data, Figure 6c. We see however that variation of transcription rate alone is not able to explain the inversions observed in alternative splicing (red symbols in Figure 6c). Mitochondria can also modify the conversion rates to AS isoforms by a variety of mechanisms, again direct, such as energetic

modulation of spliceosome assembly, AS regulators or RNA folding, or indirect through alteration of genes involved in splicing machinery (Figure S10b), or coupling between alternative splicing and different transcriptional events. Indeed, there is recent experimental evidence that both chromatin structure and histone modifications participate in the regulation of alternative splicing (Luco et al. 2011) and that transcription elongation is also coupled to alternative splicing (Schmidt et al. 2011; Dujardin et al. 2013; Kornblihtt et al. 2013) . To include these effects in our model, we assume that the conversion fraction parameter N_1 can be strongly modified by mitochondrial levels, and thus make a new random choice of this value in Low and High conditions. When only this parameter and global transcription elongation are allowed to vary, we see that fold-changes are highly spread along the opposite diagonal in the differential expression map, Figure S9c. This gives an appreciable fraction of genes with the AS pattern inverted (red symbols, INV region) as observed in the experimental situation (Figure 6c). These inversions are probably due either to coupling mechanisms between transcription elongation and/or initiation (chromatin remodelling) and alternative splicing, or alterations of specific splicing genes mediated by mitochondria.

Finally, we explored the possibility of gene specific variability induced by mitochondria in the individual degradation rates δ_1/δ_2 , re-sampling these rates at each Low/High condition. Changing these parameters alone (Figure S9d), although increasing variability, does not produce any appreciable number of genes in the above-threshold domains.

A combination of variability in conversion fraction N_1 and transcription rate k including gene specific and global factors as explained above is able to reproduce the experimental pattern, Figure 6e, along with the relative fractions of genes in each of the differential expression regions (inset of Figure 6e). We thus conclude that the observed alterations in AS can be only explained by considering both large changes in splicing site choice and large variability in pre-mRNA production from the same gene. Moreover, mitochondria is probably a key modulator of these processes.

Table S1

Go description	Term_id	adj_p value
Catabolism		
cellular macromolecule catabolic process	GO:0044265	**2,15E-24**
proteolysis involved in cellular protein catabolic process	GO:0051603	**7,31E-22**
cellular protein catabolic process	GO:0044257	**7,31E-22**
modification-dependent protein catabolic process	GO:0019941	**2,30E-21**
RNA metabolism		
RNA processing	GO:0006396	**3,50E-15**
mRNA metabolic process	GO:0016071	**8,36E-11**
chromatin organization	GO:0006325	**3,24E-09**
mRNA processing	GO:0006397	**4,20E-09**
RNA splicing	GO:0008380	**9,42E-08**
transcription from RNA polymerase II promoter	GO:0006366	**1,81E-06**
chromatin modification	GO:0016568	**5,38E-07**
histone modification	GO:0016570	**6,74E-05**
Transport		
vesicle-mediated transport	GO:0016192	**7,89E-07**
nuclear transport	GO:0051169	**1,02E-09**
nucleocytoplasmic transport	GO:0006913	**1,33E-09**
intracellular protein transport	GO:0006886	**1,92E-14**
intracellular protein transmembrane transport	GO:0065002	**1,24E-05**
protein localization in organelle	GO:0033365	**1,38E-06**
intracellular transport	GO:0046907	**5,28E-18**
protein targeting	GO:0006605	**1,84E-09**
cellular protein localization	GO:0034613	**5,75E-15**
Protein metabolism		
translation	GO:0006412	**6,10E-09**
protein complex assembly	GO:0006461	**4,01E-08**
protein folding	GO:0006457	**3,94E-05**
Energy		
energy derivation by oxidation of organic compounds	GO:0015980	**7,35E-05**
carboxylic acid metabolic process	GO:0019752	**4,39E-08**
cellular lipid metabolic process	GO:0044255	**5,88E-10**
cellular carbohydrate metabolic process	GO:0044262	**1,33E-07**
electron transport chain	GO:0022900	**1,13E-05**
Cell cycle		
M phase of mitotic cell cycle	GO:0000087	**1,42E-06**
M phase	GO:0000279	**1,68E-06**
DNA metabolic process	GO:0006259	**5,88E-10**
mitosis	GO:0007067	**5,24E-07**
DNA damage		
cellular response to DNA damage stimulus	GO:0034984	**9,42E-08**
DNA repair	GO:0006281	**1,13E-06**
Others		
microtubule cytoskeleton organization	GO:0000226	**4,91E-06**
small GTPase mediated signal transduction	GO:0007264	**9,13E-06**
regulation of signal transduction	GO:0009966	**1,95E-05**

The gene ontology analysis of genes differentially expressed in cells with different mitochondria content showed 8 main families of genes with a p value < 10E-4.
As a global trend, the families of genes involved in macromolecule homeostasis are highlighted. Degradation processes seem to be the ones with higher representation, followed by RNA metabolism and protein metabolism. Another family of genes with high score are the genes involved in transport processes, which in part could be a consequence of the high biosynthetic activity of cells with high mitochondria content. It is not surprising to find also genes involved in energy metabolism because cells with more mitochondria require more intermediate energy precursors. We already have documented that mitochondria content impact in the length of the cell cycle (das Neves et al 2010), therefore it is not surprising to find that mitochondria affects cell cycle genes. It is well know the role played by the mitochondria as a source of free radicals, which can damage DNA. For such reason, it makes sense to find an effect of mitochondria content on genes involved in DNA damage response.

Supplementary Figures captions

Figure S1. CmxRos is a good reporter of mitochondrial mass.

a.In order to study the suitability of fluorescent probes for mitochondrial content quantification, we first sorted cells according to MitoTracker Green (MG) uptake in two subpopulations: High (H) and Low (L), inset in Figure S1a. Then, we quantified the relative mtDNA levels by qPCR as described in Supplementary Methods in both subpopulations. The ratios of MG staining and mtDNA levels between H and L populations show a perfect matching (Figure S1a). Error bars are the standard deviations of three biological replicates. b.To check whether CmxRos is a faithful reporter for mitochondrial content in individual cells, we first co-stained cells with CmxRos and MG, showing a high correlation (r^2=0.92, Pearson correlation). In order to further validate the use of CmxRos we co-stained cells with CmxRos and several mitochondrial markers. c-e. Three mitochondrial matrix proteins: Aconitase (ACO2), SOD2 and Pyruvate Dehydrogenase (PDP1) also correlate well with CmxRos: r^2=0.88 (Aconitase), r^2=0.82 (SOD2), r^2= 0.75 (PDP1). f. Complex I, which is located in the inner membrane, closely matches CmxRos levels (r^2=0.88). The antibodies used were ACO2 (AB110321), Complex-1 (AB109798), SOD2(AB110300) (Abcam) and PDH (2784) (Cell Signaling).

Figure S2. Mitochondrial contribution to variability using different statistical methods.

To assess the robustness of our estimate of mitochondrial contribution to global variability in protein expression, we tested other statistical measures of contribution to variance (see Supplementary Text). Black histograms: Method discussed in main text and used in Figures 1, 2 and 3. Gray histograms: Fraction of variance explained by the coefficient of determination (R^2) of linear regression. White histograms: ratio of variance along CmxRos axis to total variance(Snijder et al. 2009). Error bars are standard deviations of three replicates, each with 200-400 cells.

Figure S3. Relative influence of cell cycle state and mitochodrial content on protein variability in single cells.

A-B. Scatter plots of simultaneous values of DAPI (cell cycle state) and CmxRos (mitochondrial content) with levels of the protein LDHA. C. Variability of DAPI levels as a function of CmxRos. The corresponding pairwise correlation coefficients (Pearson) are shown in the inset. D-I. Partial correlation network of six different proteins taken from the set analyzed in Figure 1 (red nodes), with respect to mitochondrial and DNA (DAPI) content. The thickness of the edges is proportional to the partial correlation coefficient. For each protein, ensembles of 300-400 cells were analyzed to quantify CmxRos, DAPI and protein levels, and their values z-score normalized before calculating partial correlations. J. Average value of association (for the six proteins analyzed, bars representing standard errors), for the three possible associations between variables (Mitochondria-Protein, DAPI-Protein and Mitochondria-DAPI). Different measures of statistical and causal association were tested. Black histograms: pairwise Pearson correlation coefficients (r). Blue histograms: partial correlation coefficients (pr). Red: Semipartial correlation coefficients (sr). Gray: LMG coefficients (lr). See Supplementary Text. These results show that mitochondrial content appears consistently as a much stronger predictor of protein variability than cell cycle state.

Figure S4. Kinetics of RNA Pol II in vivo.

FLIP analysis of GFP–RNA Pol II. (a) Half of the nucleus was bleached continuously (white rectangle) as confocal images were collected approximately every 5 s. (b) Decay in fluorescence intensity of the unbleached area of the nucleus in the FLIP experiment (intensity, arbitrary units) (n = 60) in a natural log scale. The data can be fitted to three exponentials showing the existence of three different kinetic regimes, which according to our model can be assigned to three different populations of RNA Pol II: free polymerase, engaged RNA Pol II and elongating RNA Pol II. The intercept with the y axis was used to determine the steady state ratio of initiating and elongating RNA Pol II, and the slopes were used to estimate the kinetic rate constants K2 and K4. (c) ATP concentrations relative to control (blue bars, cells growing in 4.5g/l of glucose) in different conditions: Green bars, cells growing in media without glucose but supplemented with galactose for osmotic compensation (300 mOsm). Red bars: Cells treated with azide plus deoxyglucose (see main text). Error bars are standard deviations of three biological replicates. (d) The total amount of RNA Pol II is proportional to the amount of mitochondria in

individual cells. (e) Analysis in single cells of the fraction of initiating RNA Pol II as a function of the elongation constant K4. Red: Control cells. Blue: D-A treated cells. The fraction of initiating polymerase molecules is independent of the elongation constant.(f) Analysis in single cells of the fraction of elongating RNA Pol II molecules as a function of the elongation constant K4. Red: Control cells. Blue: D-A treated cells. The fraction of elongating molecules is proportional to the elongation constant. (g) Summary of kinetic model parameters and their estimation method.

Figure S5. qPCR quantification of mRNA abundance.

We selected genes (see Supplementary Text) with noticeable expression differences between Low and High conditions (FC > 5) in our RNAseq data and performed qPCR (SYBR Green) on sorted (with MitoTracker Green) populations with Low and High mitochondrial content. After quantification, their fold-changes were corrected by the same factor used to correct RNAseq data (since for the same amount of mRNA, Low samples approximately triplicate the number of cells of High samples as determined by polyA mRNA FISH, Methods). Left panel: Fold-changes (High/Low) taken from our RNAseq data (blue histograms) compared to fold-changes determined by qPCR quantification (maroon histograms). Bars are standard errors of three replicates. Right panel: Scatter plot of logarithmic fold-changes for RNAseq and qPCR experiments with the corresponding Pearson correlation coefficient.

Figure S6. Correlation between CmxRos and biosynthetic activities.

Quantitative microscopy was used to gather information of both mitochondria dosage (CmxRos) and gene expression products and activities. (A) Amount of double stranded RNA (highly enriched in ribosomal RNA, and thus an indicator of the cell ribosomal mass), stained with YOYO1, as a function of CmxRos. (B) HeLa cells stained with CmxRos and then pulsed for 30 min with BrU (a marker of nascent RNA). We binned the scattered data in CmxRos intervals and calculated average and standard deviation values, filled circles and error bars. The mean values are well fitted to a sigmoidal function (solid line) of the form $\frac{a_0+a_1(x/K)^n}{1+(x/K)^n}$, with parameters a_0=0.45, a_1=1.41, K=0.81, n=4.7. (C) Relationship between polyA

RNA(an indicator of total mRNA content) and CmxRos content. In all panels we show the value of the Spearman's correlation coefficient.

Figure S7. Global decay in mRNA and protein content in the two sorted subpopulations.

A.Red circles: mRNA decay of the Low subpopulation as measured by BrU intensity decay (Materials and Methods). Blue circles: mRNA decay of the High subpopulation. We fitted experimental data to exponential functions of the form $+be^{-ct}$, and estimated mRNA half-life with the expression $\tau_{1/2} = \frac{\ln b - \ln\left(\frac{1}{2} - a\right)}{c}$.

B.Red circles: protein decay of the Low subpopulation (Materials and Methods), Blue circles: protein decay of the High subpopulation. Experimental data are fitted to single exponential functions e^{-ct}. The time span of the x-axis is the average cell cycle period of HeLa cells (~23 h).

Figure S8.Transcription rates and half-lives distributions of mouse fibroblasts.

(a) Transcription rate distribution from the experimental data of (Schwanhausser et al. 2011). Synthesis rates were estimated using RNAseq, metabolic pulse labelling and a quantitative model of mRNA transcription/degradation in more than 5,000 genes. Red line is a fit to a log-normal distribution. (b) Distribution of half-lives from the data of (Schwanhausser et al. 2011) for the same set of genes. Red line is a fit to a Gamma distribution.

Figure S9. Differential expression map of alternative splicing using a two-step model.

Scatter plot of logarithmic fold-changes of pairs of alternatively spliced isoforms generated with our two-step model. Transcription and degradation rates are sampled from the experimental distributions in Figure S8, and the conversion fraction is a uniform random number in the unit interval (see *Supplementary Text*). Different panels show the effect of different sources of variability. (a) Only global

variability due to transcription elongation and global differences in decay (according to our experimental results in Figure S6B and Figure S8A). (b) Adding gene specific variability in transcription rate alters the AS pattern with an appreciable fraction of genes in the UP and DOWN domains (black and blue dots respectively). The yellow line corresponds to the suppression of extrinsic variability in transcription elongation and decay. (c) Adding gene specific variability in isoform conversion fraction changes the AS differential expression map favouring inversions of expression (INV domain, red dots). (d) Adding large gene specific variability only in RNA half-life does not explain the appearance of the different domains of altered AS. Black lines in all panels mark the threshold values in FC used in the analysis of Figures 6c and 6e.

Figure S10. Usage of alternative promoters in alternative splicing and expression of splicing genes.

A. Distribution of logarithmic fold-changes induced by different mitochondrial content on RNA isoforms transcribed from different transcription start sites (TSS). The black line shows the distribution for isoforms in the UP domain (FC > 10) and the blue line in the DOWN domain (FC < 0.1). Inset: Usage of alternative promoters quantified as the ratio of isoforms transcribed from different TSS in each domain (relative to the total number of isoforms in the domain). As can be seen, ~70% of the total isoforms in UP and DOWN domains are expressed from alternative promoters. B. Cumulative distribution of logarithmic fold-changes (High/Low) in the expression of genes coding for splicing factors. The dashed red line marks the threshold for three-fold differences in expression between the two populations, which affect ~47% of the splicing genes (grey shaded area).

Figure S11. Influence of mitochondrial content on epigenetic modification of chromatin.

a.Cartoon showing the effect of high mitochondrial content on methylation and acetylation of chromatin. The steady state of histone methylation is the balance between methylation and demethylation reactions. Histone methylation is catalyzed by histone methyl transferases and the substrate S-adenosyl-Lmethionine (SAM). SAM is produced by the reaction L-methionine + ATP. Thus, SAM links energy production to histone methylation. H3K36me2 and H3K4me3 are demethylated by

the action of KDM2 which is activated by 2-oxoglutarate, whose concentration depends on ADP/ATP among others (Salminen et al. 2014). Then in cells with high mitochondria content, the ratio ADP/ATP must be low and so 2-oxoglutarate, keeping KDM2 activity low. Therefore, cells with high mitochondria content must show more methylated chromatin. Likewise, the balance between acetylation and deacetylation reactions determine the levels of acetylated histones. Histone acetylation is catalyzed by histone acetyl transferases (HATs) which use as a precursor Acetyl-CoA. This factor is produced by Pyruvate Dehydrogenase Complex (PDC). Cells with high mitochondrial abundance must contain a high concentration of Pyruvate because most glycolitic enzymes co-variate with mitochondria (Figure 1e). Furthermore, Pyruvate inhibits histone deacetylases HDAC1 and HDAC3 (Choudhary et al. 2014). b. In cells with low mitochondrial content the concentration of ATP is low and therefore the activity of the methylation reaction is reduced. Moreover, in those cells the ratio ADP/ATP must be higher, favouring the formation of 2-oxoglutarate activating histone demethylases (HDM) like KDM2. Likewise, acetylation reaction must have low activity due to reduced levels of Pyruvate. Moreover, the low concentration of Pyruvate release the repression of HDACs. Furthermore, cells with low mitochondrial content are expected to have a high ratio of NAD+/NADH (Jang et al. 2012) which is the regulator of Sirtuins activity, catalyzing histone deacetylation (Choudhary et al. 2014).

Figure S12. Influence of ROS on transcription and ATP content.

A.Contribution of oxidative stress to Br-RNA synthesis. Cells were grown for 2 hours in presence of 2 and 1 mM DTT(antioxidant), or 50 and 100 microM Diamide(prooxidant). Then, BrU was added to the medium for half an hour, cells were fixed and immunolabeled for Br-RNA. This panel shows a clear dependency of Br-RNA incorporation (RNA Pol II activity) on antioxidant capabilities of the cell. Bars show the average of 200 cells, together with the corresponding standard deviation. B. ATP is affected by antioxidants and prooxidants. Cells were grown in 125 microM DTT or 50 microM Diamide, for up to 12 h. Bars are the average of three independent experiments. Error bars are SD. This panel shows that DTT increases relative ATP production (presumably by improving mitochondrial function) and Diamide decreases relative ATP levels. Therefore, we cannot attribute the

increase of Br-RNA in antioxidant conditions only to the reduction of ROS, since ATP is also increased in these conditions. C. Cells were co-stained with MitotrackerGreen and Mitosox (a superoxide reporter) for 2 h. The blue dots are the average values of sampling in bins, and the solid line the best fit to these dots. The slope is close to one (0.93) which shows that the production of superoxide is completely matched by the amount of mitochondria. This means that cells have the same *relative* production of superoxide, as shown in panel D. Therefore retrograde signalling does not apply to our condition.

References

Ashley N, Harris D, Poulton J. 2005. Detection of mitochondrial DNA depletion in living human cells using PicoGreen staining. *Experimental cell research* **303**(2): 432-446.

Choudhary C, Weinert BT, Nishida Y, Verdin E, Mann M. 2014. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature reviews Molecular cell biology* **15**(8): 536-550.

das Neves RP, Jones NS, Andreu L, Gupta R, Enver T, Iborra FJ. 2010. Connecting variability in global transcription rate to mitochondrial variability. *PLoS biology* **8**(12): e1000560.

Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P et al. 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**(9): 1959-1972.

Dujardin G, Lafaille C, Petrillo E, Buggiano V, Gomez Acuna LI, Fiszbein A, Godoy Herz MA, Nieto Moreno N, Munoz MJ, Allo M et al. 2013. Transcriptional elongation and alternative splicing. *Biochim Biophys Acta* **1829**(1): 134-140.

Freedman DA. 2009. *Statistical Models: Theory and Practice*. Cambridge University Press, New York.

Jang SY, Kang HT, Hwang ES. 2012. Nicotinamide-induced mitophagy: event mediated by high NAD+/NADH ratio and SIRT1 protein activation. *The Journal of biological chemistry* **287**(23): 19304-19314.

Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nature reviews Molecular cell biology* **14**(3): 153-165.

Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. 2011. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC systems biology* **5**: 21.

Lindeman RH, Merenda PF, Gold RZ. 1980. *Introduction to Bivariate and Multivariate Analysis*. Scott Foresman, Glenview, IL.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**(1): 16-26.

Mao C, Brown CR, Falkovskaia E, Dong S, Hrabeta-Robinson E, Wenger L, Boeger H. 2010. Quantitative analysis of the transcription control mechanism. *Mol Syst Biol* **6**: 431.

Opgen-Rhein R, Strimmer K. 2007. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC systems biology* **1**: 37.

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N et al. 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**(5): 436-442.

Salminen A, Kauppinen A, Hiltunen M, Kaarniranta K. 2014. Krebs cycle intermediates regulate DNA and histone methylation: epigenetic impact on the aging process. *Ageing Res Rev* **16**: 45-65.

Schmidt U, Basyuk E, Robert MC, Yoshida M, Villemin JP, Auboeuf D, Aitken S, Bertrand E. 2011. Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *The Journal of cell biology* **193**(5): 819-829.

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473**(7347): 337-342.

Shipley B. 2004. *Cause and correlation in Biology*. Cambridge University Press, Cambridge, UK.

Snijder B, Sacher R, Ramo P, Damm EM, Liberali P, Pelkmans L. 2009. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature* **461**(7263): 520-523.

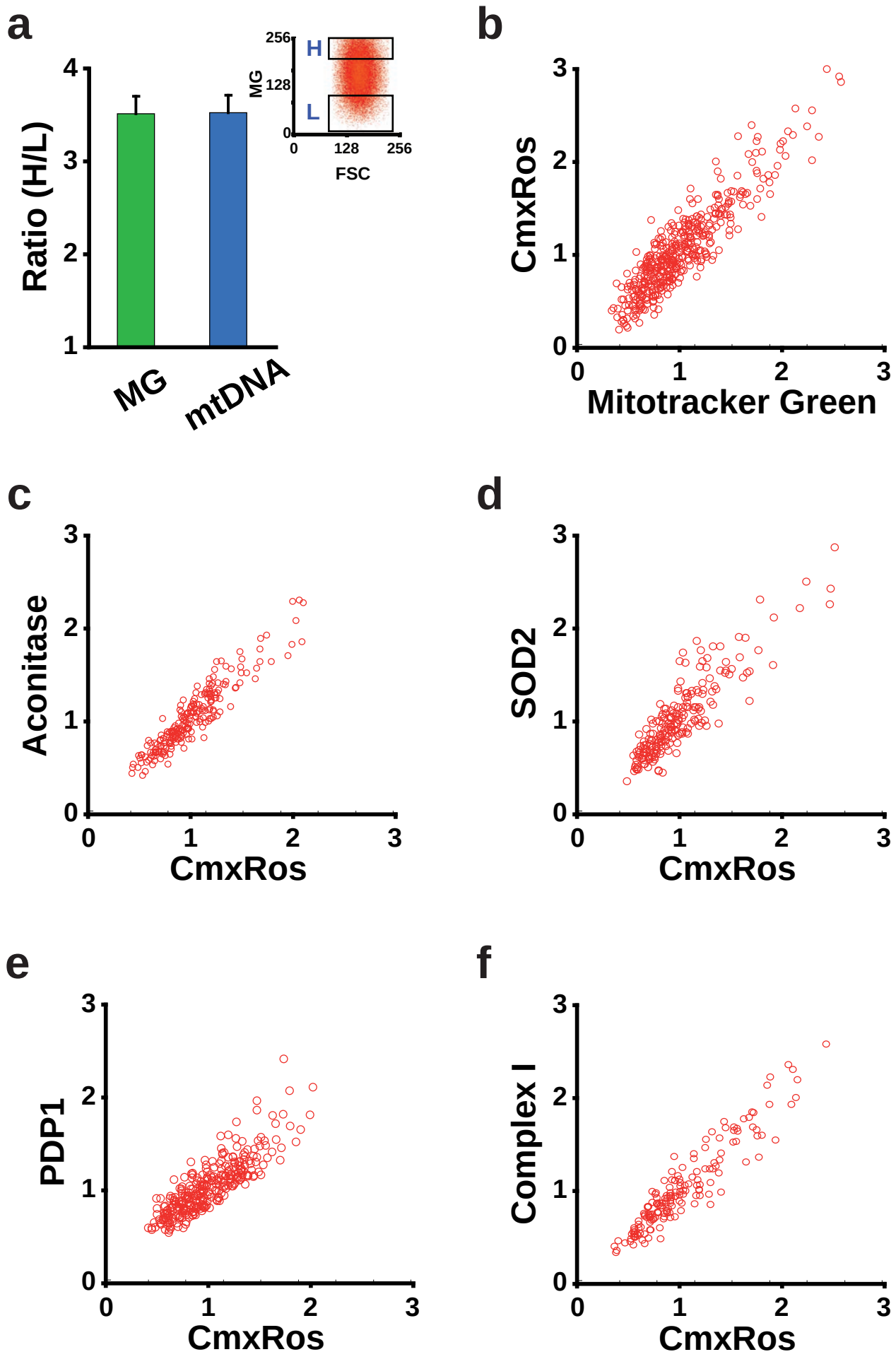

a
Ratio (H/L)
4
3
2
1
MG
mtDNA
256
128
0
MG
H
L
0
128
256
FSC
b
CmxRos
Mitotracker Green
c
Aconitase
CmxRos
d
SOD2
CmxRos
e
PDP1
CmxRos
f
Complex I
CmxRos

Fig S1

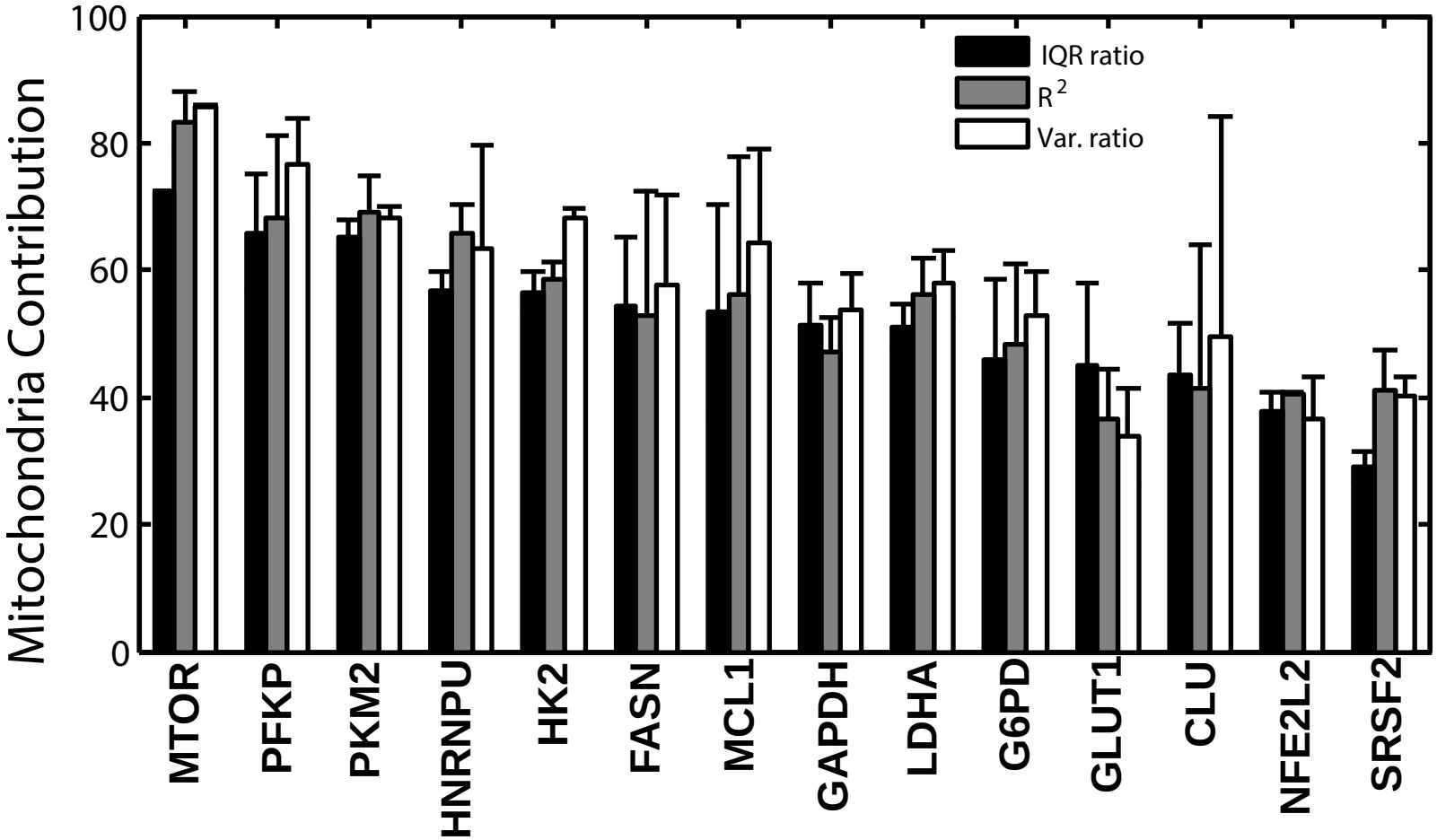

Mitochondria Contribution
100
80
60
40
20
0
IQR ratio
R 2
Var. ratio
MTOR
PFKP
PKM2
HNRNPU
HK2
FASN
MCL1
GAPDH
LDHA
G6PD
GLUT1
CLU
NFE2L2
SRSF2

Fig. S2

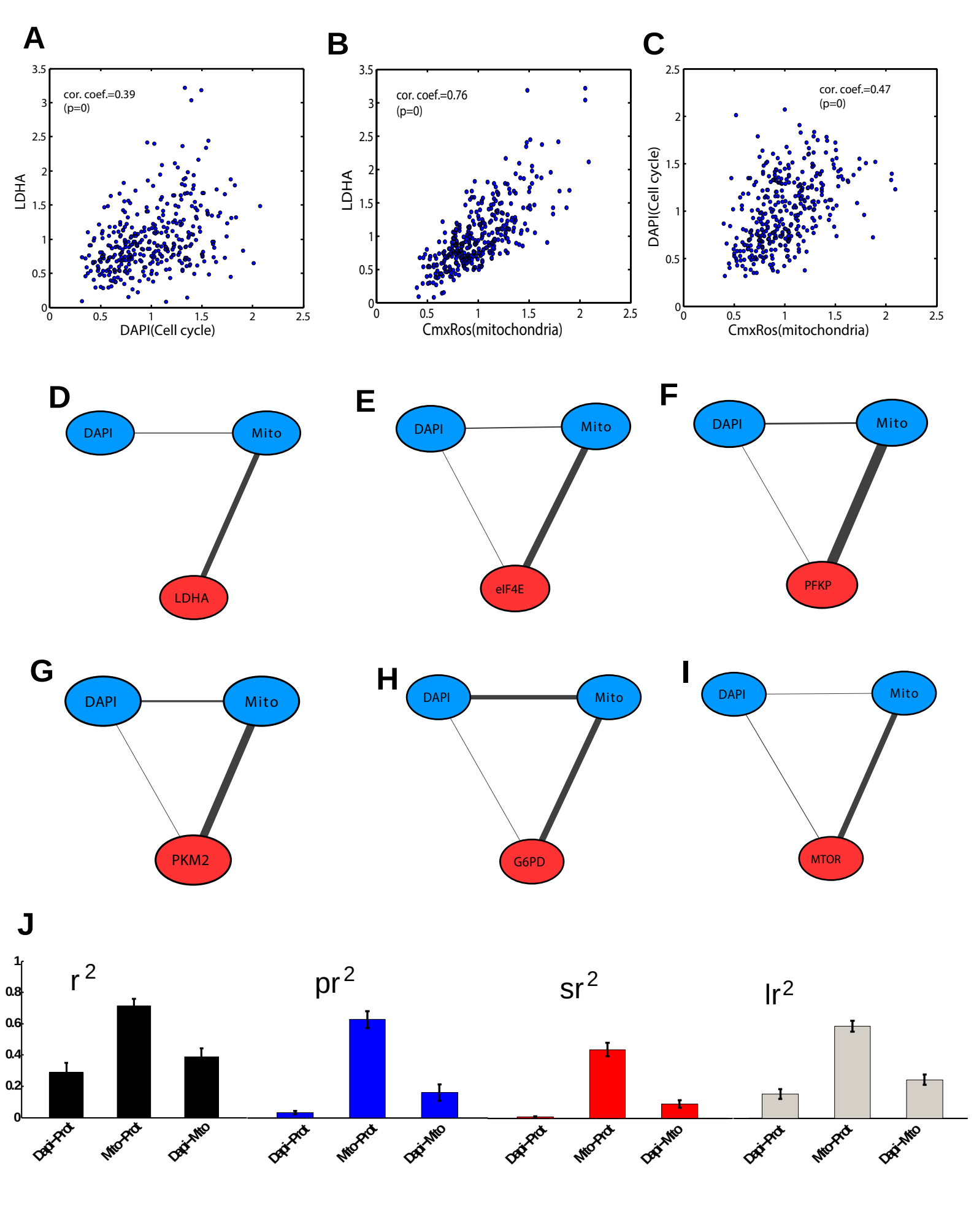
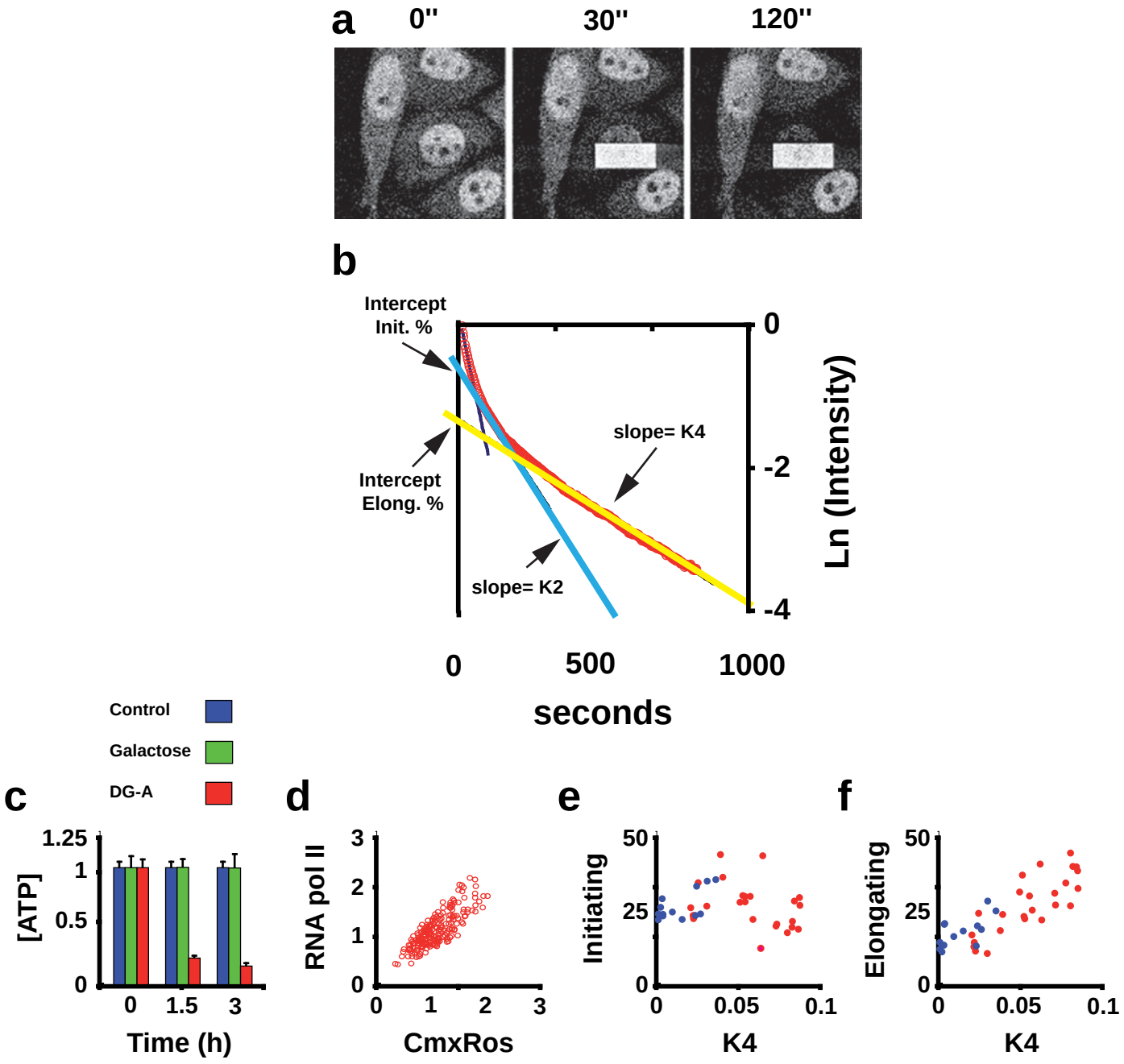

A
cor. coef.=0.39
(p=0)
LDHA
DAPI(Cell cycle)
B
cor. coef.=0.76
(p=0)
LDHA
CmxRos(mitochondria)
C
cor. coef.=0.47
(p=0)
DAPI(Cell cycle)
CmxRos(mitochondria)
D
DAPI
Mito
LDHA
E
DAPI
Mito
eIF4E
F
DAPI
Mito
PFKP
G
DAPI
Mito
PKM2
H
DAPI
Mito
G6PD
I
DAPI
Mito
MTOR
J
r 2
pr 2
sr 2
lr 2
Dapi-Prot
Mito-Prot
Dapi-Mito

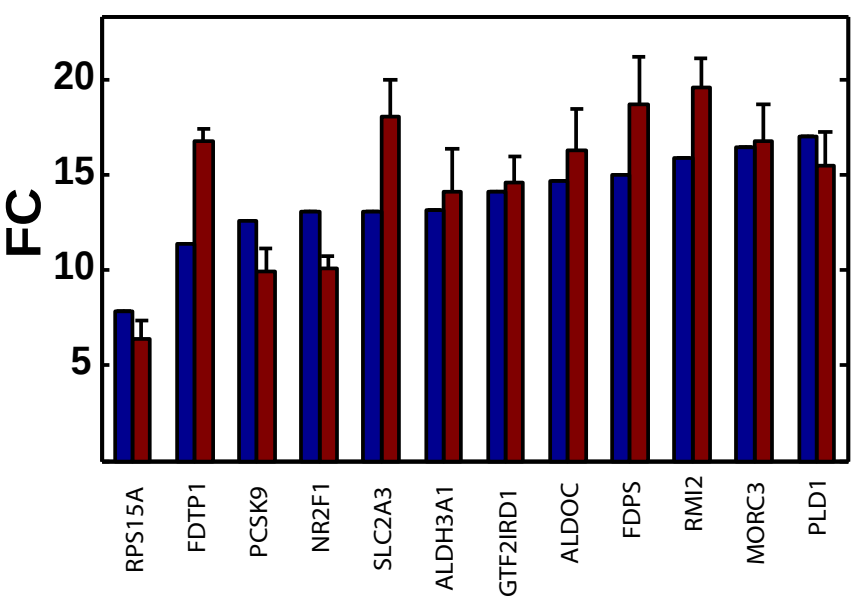
Fig S3



g

Symbol	Meaning	Method for determination
K1	DNA entry constant	Model solution
K2	Abortive constant	Best fit experiment
K3	Initiation constant	Model solution
K4	Elongation constant	Best fit experiment
Pol II_{Free}	Free fraction	Best fit experiment
Pol $II_{Init.}$	Initiating fraction	Best fit experiment
Pol $II_{Elong.}$	Elongating fraction	Best fit experiment

Fig S4


FC
20
15
10
5
RPS15A
FDTP1
PCSK9
NR2F1
SLC2A3
ALDH3A1
GTF2IRD1
ALDOC
FDPS
RMI2
MORC3
PLD1

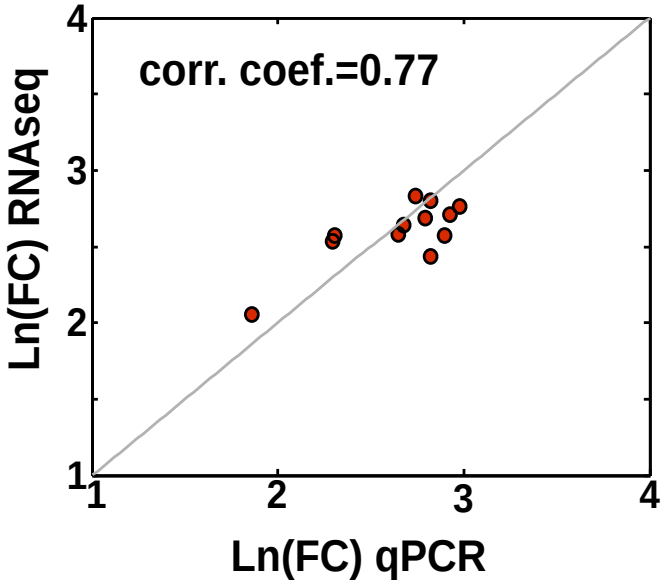

corr. coef.=0.77
Ln(FC) RNAseq
Ln(FC) qPCR
1
2
3
4

Fig S5

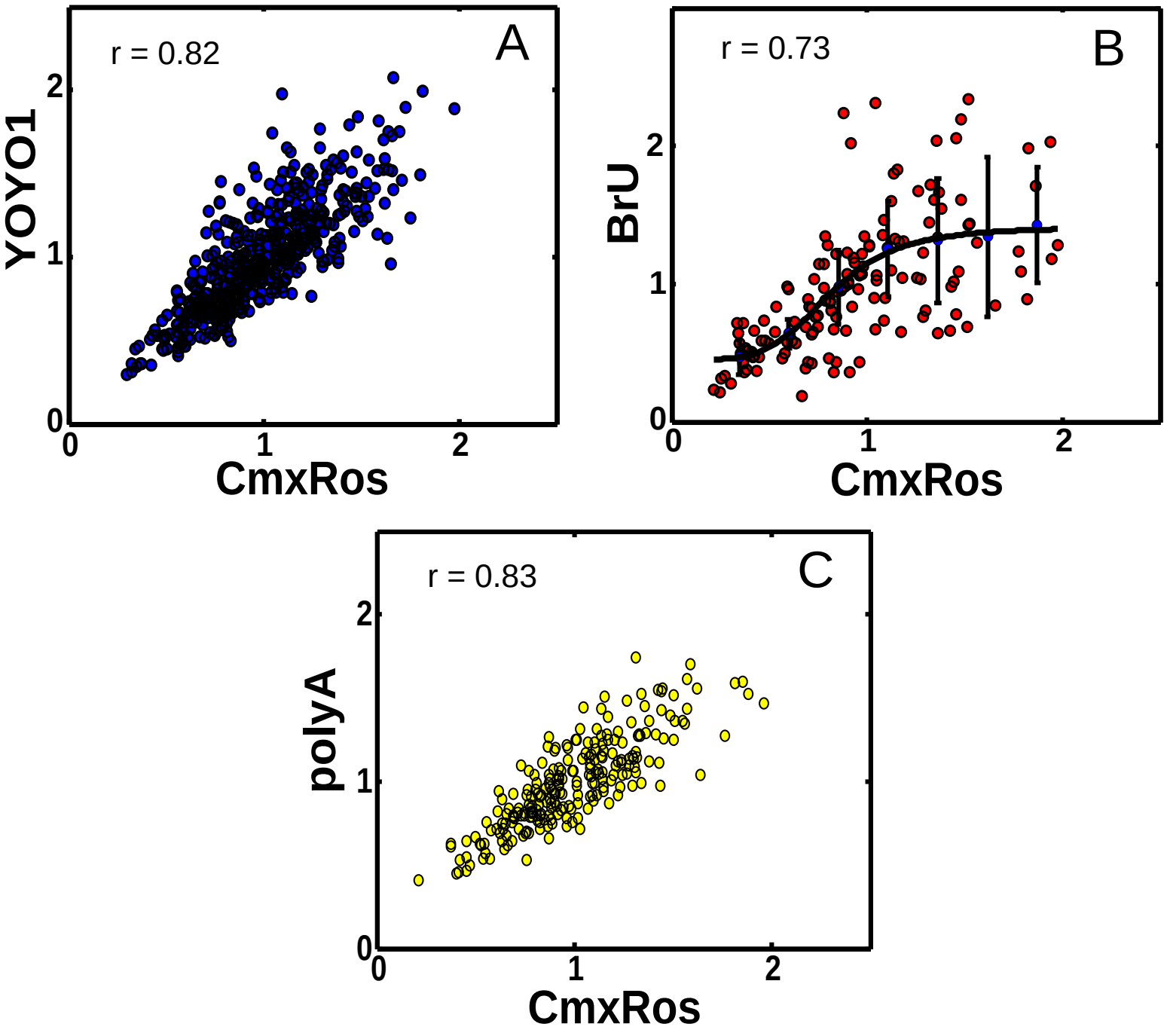


Fig S6

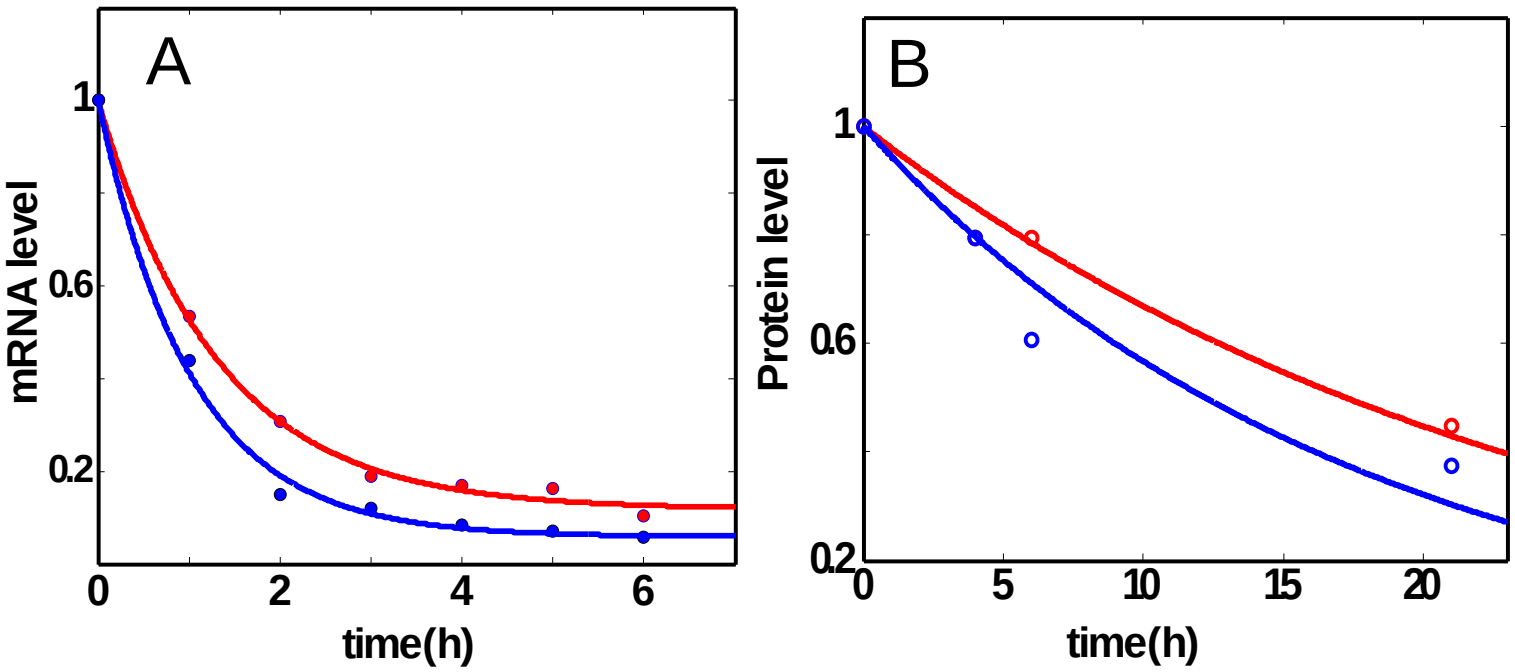

A
mRNA level
1
0.6
0.2
0
2
4
6
time(h)
B
Protein level
1
0.6
0.2
0
5
10
15
20
time(h)

Fig S7

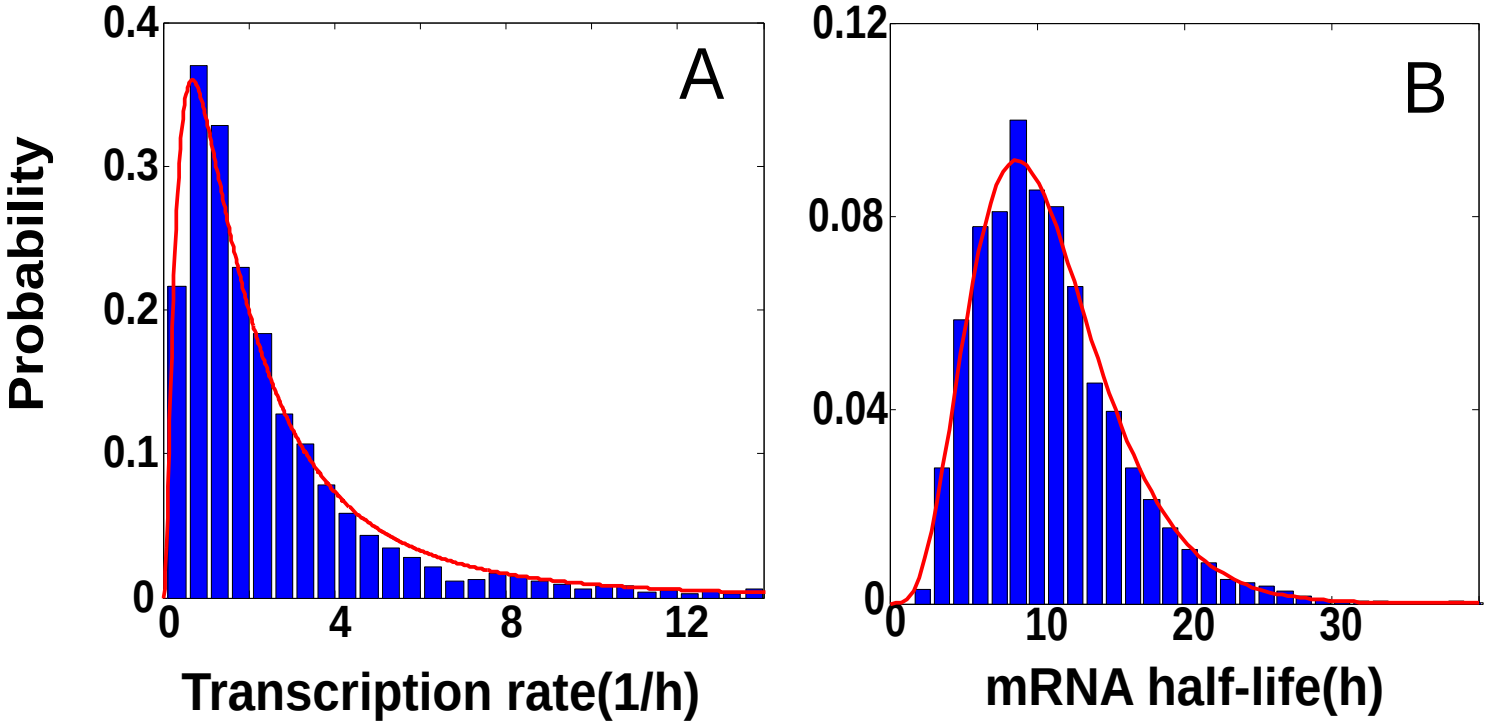

Probability
0.4
0.3
0.2
0.1
0
A
0
4
8
12
Transcription rate(1/h)
0.12
0.08
0.04
0
B
0
10
20
30
mRNA half-life(h)

Fig S8

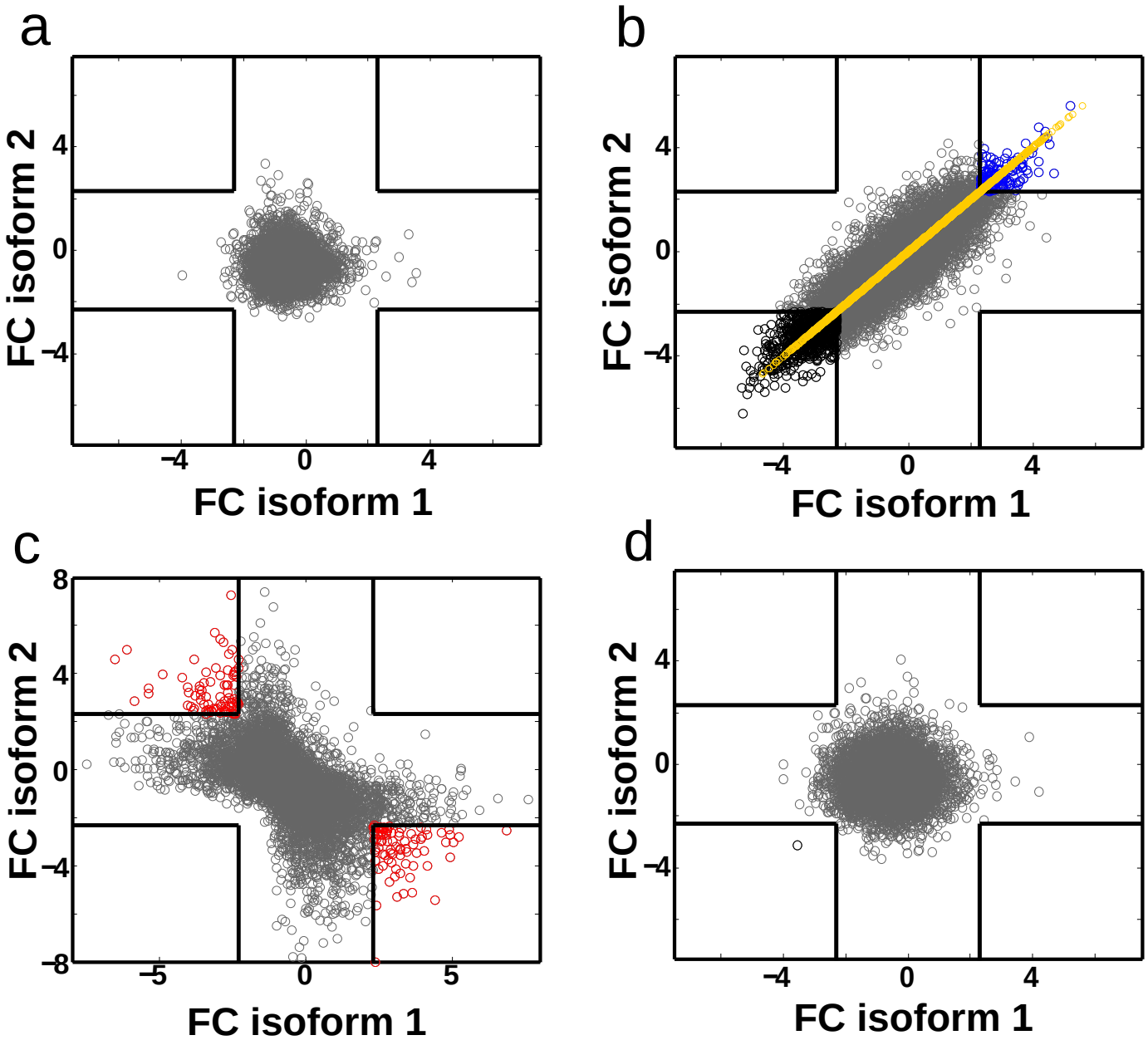


Fig S9

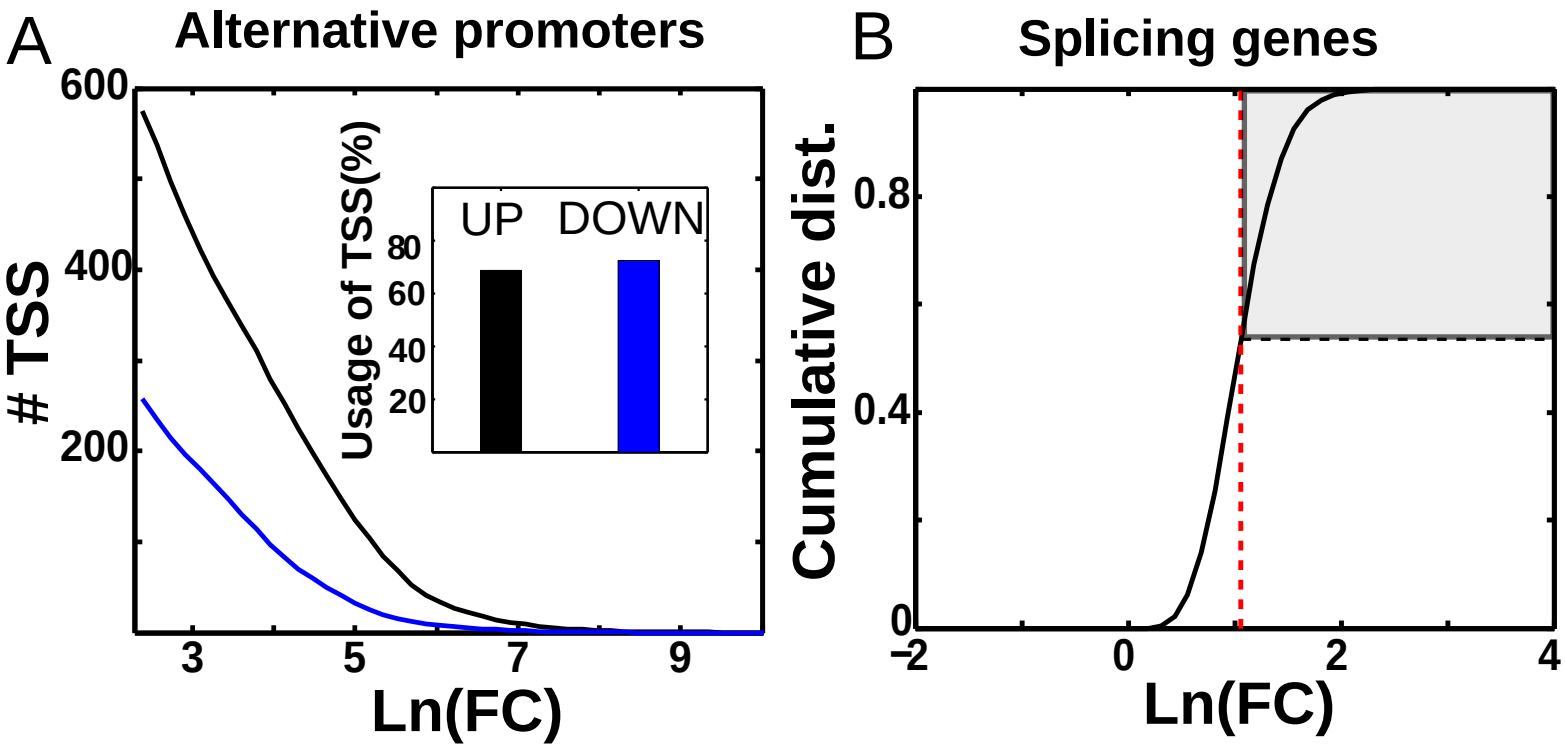

A
Alternative promoters
600
400
200
TSS
3
5
7
9
Ln(FC)
Usage of TSS(%)
80
60
40
20
UP
DOWN
B
Splicing genes
Cumulative dist.
0.8
0.4
0
–2
0
2
4
Ln(FC)

Fig S10

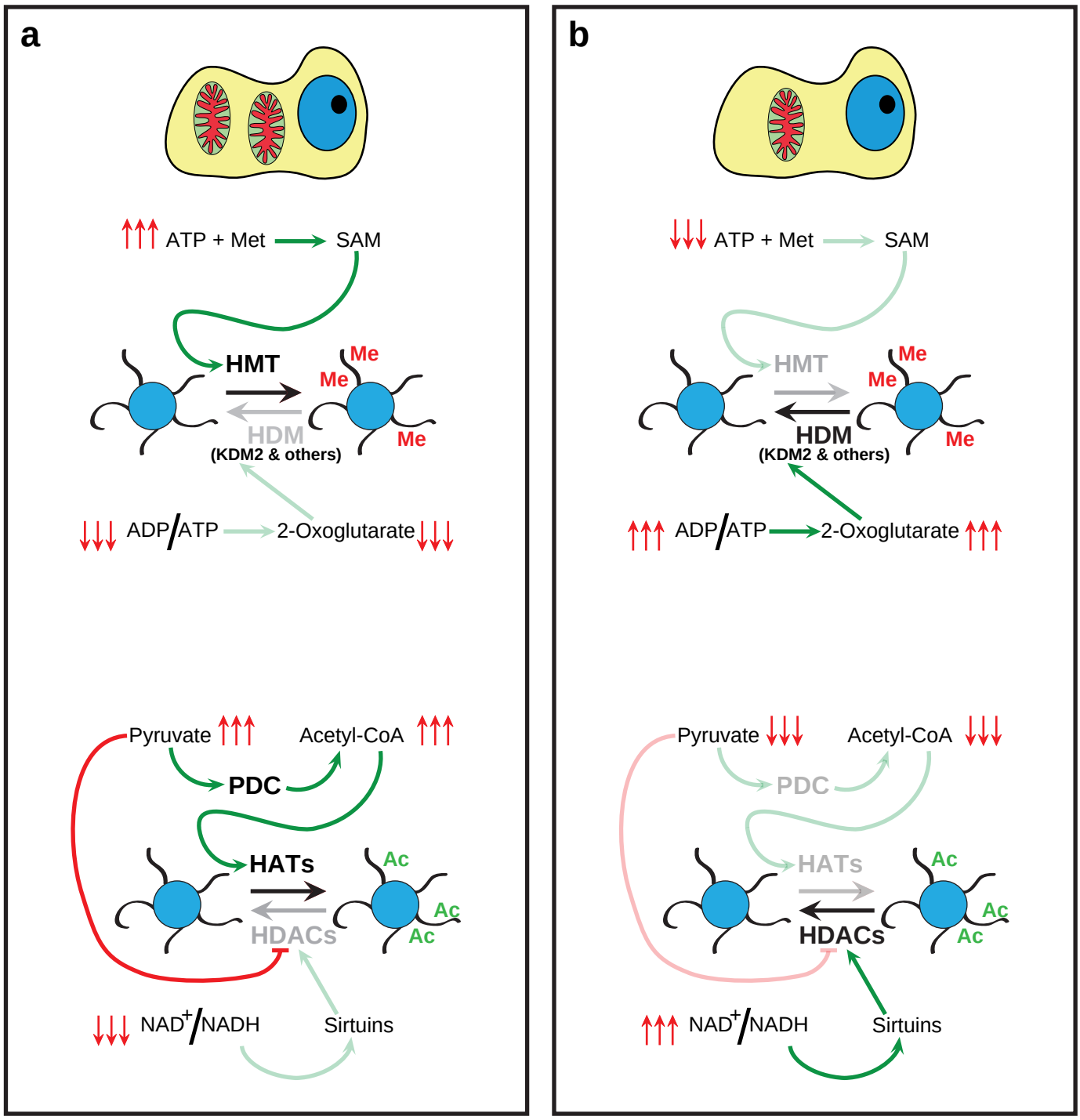

a
ATP + Met
SAM
HMT
HDM
(KDM2 & others)
Me
Me
Me
ADP/ATP
2-Oxoglutarate
Pyruvate
Acetyl-CoA
PDC
HATs
HDACs
Ac
Ac
Ac
NAD+/NADH
Sirtuins
b
ATP + Met
SAM
HMT
HDM
(KDM2 & others)
Me
Me
Me
ADP/ATP
2-Oxoglutarate
Pyruvate
Acetyl-CoA
PDC
HATs
HDACs
Ac
Ac
Ac
NAD+/NADH
Sirtuins

Fig S11

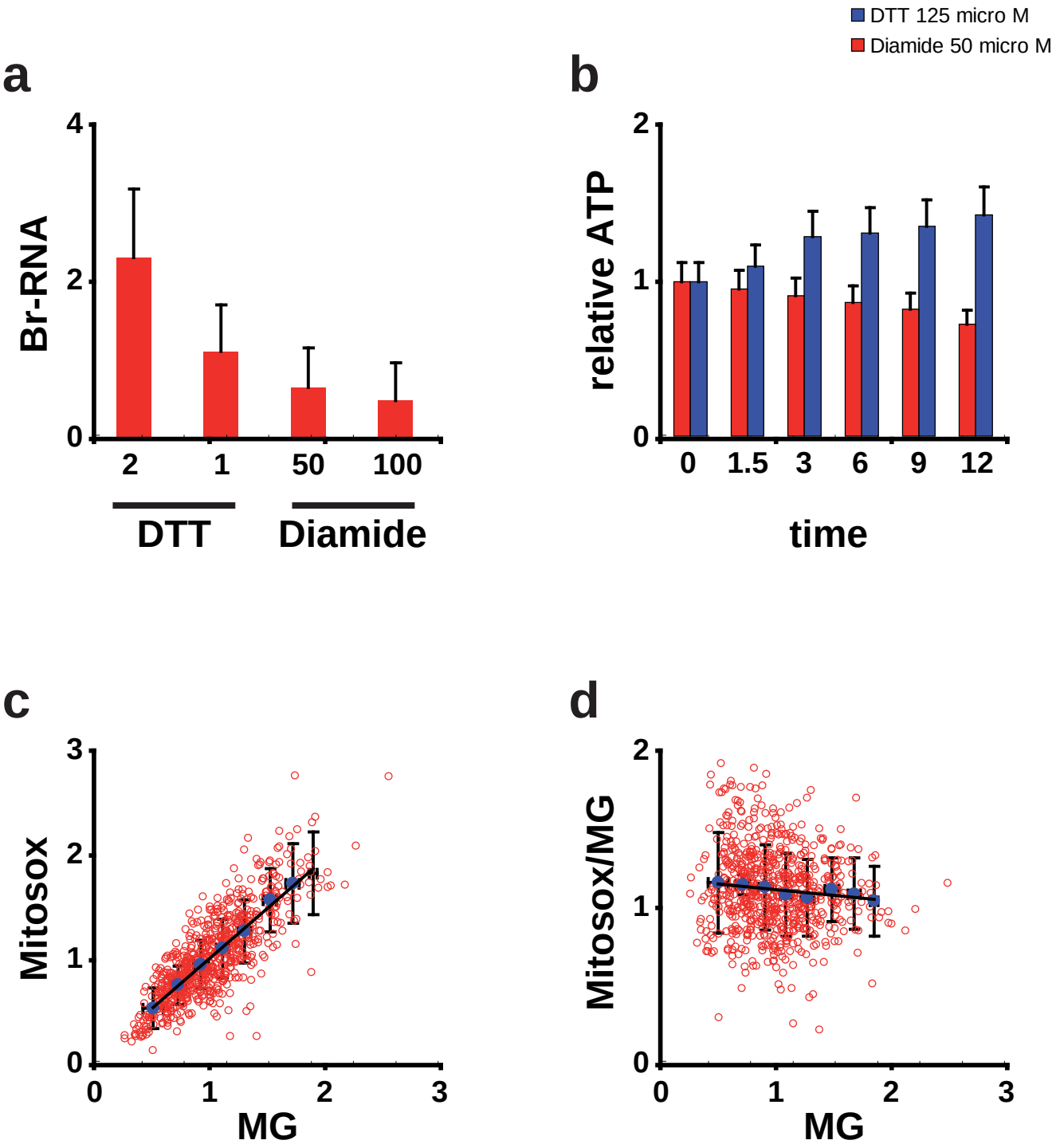

a
Br-RNA
4
2
0
2
1
50
100
DTT
Diamide
b
DTT 125 micro M
Diamide 50 micro M
relative ATP
2
1
0
0
1.5
3
6
9
12
time
c
Mitosox
3
2
1
0
0
1
2
3
MG
d
Mitosox/MG
2
1
0
0
1
2
3
MG

Fig S12