Supplementary Text: Mathematical formulation of the Subcellular Spatial Razor.

A. The subcellular spatial razor.

The spatial razor assumes that a given protein can be in the nucleus (n) and in the cytoplasm (c) for both unstimulated (u) cells and stimulated (s) cells. The corresponding abundances (Fig S1) are $A_{n,u}$, $A_{n,s}$, $A_{c,u}$, $A_{c,s}$. There are three SILAC ratios given by:

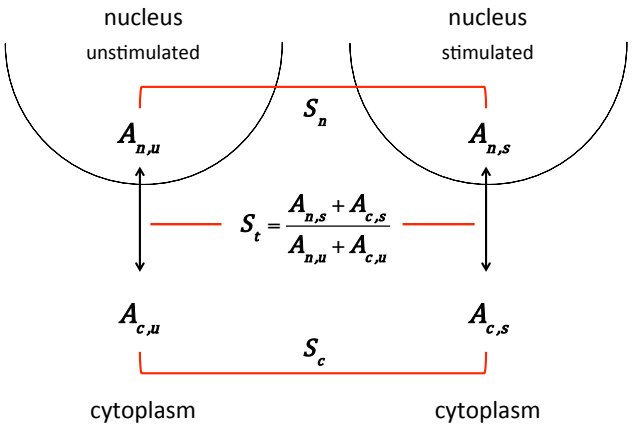


Fig. S1. The spatial razor model.

$$S_n = A_{n,s}/A_{n,u}, \qquad [1]$$
$$S_c = A_{c,s}/A_{c,u},$$
$$S_t = (A_{n,s} + A_{c,s})/(A_{n,u} + A_{c,u}).$$

The values for the fractions of the protein in the nucleus in unstimulated and stimulated cells, f_u and f_s, are closely related to the SILAC ratios:

$$f_u = A_{n,u}/(A_{n,u} + A_{c,u}) = (S_t - S_c)/(S_n - S_c), \qquad [2]$$
$$f_s = A_{n,s}/(A_{n,s} + A_{c,s}) = S_n(S_t - S_c)/S_t(S_n - S_c).$$

Furthermore, the fractions f_u and f_s are closely related to the parameters used in a three-dimensional orthogonal basis set for the experimental results (see below).

$$S_n/S_t = f_s/f_u, \qquad [3]$$
$$S_c/S_t = (1 - f_s)/(1 - f_u)$$

Experimental application of the spatial razor requires fractionating the total cellular proteins (total fraction) into two subcellular fractions such that: nucleus-enriched fraction + nucleus-depleted fraction (cytoplasm) = total fraction and measuring the set of SILAC ratios $\{S_n, S_c, S_t\}$. Although formulated here for the nucleus, a spatial razor could also be applied to other subcellular locations through the use of sample triplets such as {mitochondria-enriched, mitochondria-depleted, total-fraction}.

B. An orthogonal basis set.

Because changes in both the total abundance and the subcellular distribution of a protein influence its abundance in the nucleus/cytoplasm, the set of SILAC ratios $\{S_n, S_c, S_t\}$ are not an orthogonal basis set. As we have shown previously,[1] the 3D orthogonal basis set $\{S_n/S_t , S_c/S_t , S_t\}$ separates changes in total protein abundance (S_t) from changes in nucleus/cytoplasm distribution (the $\{S_n/S_t , S_c/S_t\}$ distribution plane). In this distribution plane, for any fraction of the protein f_u in the nucleus in the unstimulated cells, a unique curve is obtained as the fraction of the protein in the nucleus in the stimulated cells is varied over $0 < f_s < 1$ (Fig. S2).

The origin of the plot corresponds to f_u/f_s = 1, i.e. to no change in distribution upon stimulation of the cells. Conservation of mass requires that the data points lie in the two indicated quadrants of the distribution plane, which correspond to N ➔ C and C ➔ N redistribution of the subcellular location following cellular stimulation.

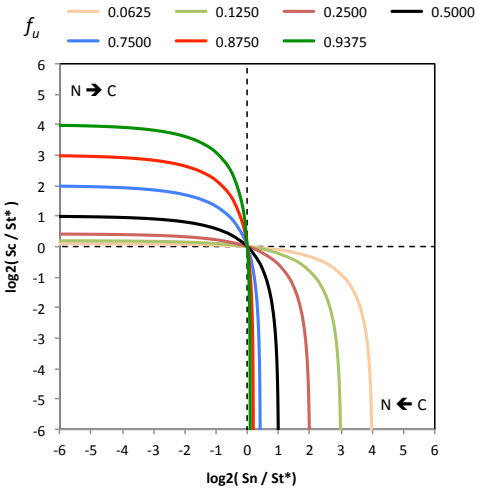


Fig. S2. For different values of f_u, the location in the distribution plane as f_s is varied over $0 < f_s < 1$.

C. Measurement and correction of S_t.

Fractionation of the sample into nuclear and cytoplasmic samples tends to increase coverage of the cellular proteome since the fractions are less complex than a total lysate sample.[1] An estimate of S_t for as many proteins as possible is essential to maximize the number of proteins that can be analysed for subcellular redistribution and can also be used to normalize the distribution plane (see below), but may be difficult to achieve experimentally for less abundant proteins in a total lysate sample.[1] In the present experiments we have optimized proteome coverage and estimated S_t by joint co-processing of the MS data recorded for the nuclear and cytoplasmic samples. Because 30 µg of nuclear/cytoplasmic proteins were used for the MS analyses of each biological replicate, and because the fractionation gave about 250/1450 µg of nuclear/cytoplasmic proteins, this involved an effective enrichment of the nuclear proteins by r = 5.8-fold during the MS analyses. Because the ratio S_n/S_c is independent of S_t (see eqn. [1]), this has no effect on the parameter $|\log_2(S_n/S_c)|$ that was used to select those proteins showing the strongest redistribution following stimulation of the cells (see main text). However, the experimental value for changes in abundance (S_t^*) is shifted from the real value (S_t) by: $S_t \rightarrow (rA_{n,s} + A_{c,s})/(rA_{n,u} + A_{c,u}) = S_t^*$. This can be written in terms of a correction factor that involves the enrichment factor r and the experimental SILAC ratios S_n, S_c and S_t^*:

$$S_t = S_t^* \left\{ \frac{\left(\frac{S_n}{S_t^*}-1\right)+\left(1-\frac{S_c}{S_t^*}\right)}{r\left(\frac{S_n}{S_t^*}-1\right)+\left(1-\frac{S_c}{S_t^*}\right)} \right\} \left\{ \frac{r\frac{S_c}{S_n}\left(\frac{S_n}{S_t^*}-1\right)+\left(1-\frac{S_c}{S_t^*}\right)}{\frac{S_c}{S_n}\left(\frac{S_n}{S_t^*}-1\right)+\left(1-\frac{S_c}{S_t^*}\right)} \right\}. \qquad [4]$$

The correction factor is symmetric with respect to enrichment/depletion of one fraction relative to the other. Note that for $f_s = f_u$ (no redistribution), $S_t^* = S_t$, i.e. there is no distortion of S_t for any value of f_u. This is because in the absence of subcellular redistribution, a conservation of mass equation $S_t = S_n = S_c$ applies. The magnitudes of the corrections can become substantial for large/small enrichment factors r coupled with substantial redistribution ($0.5 \gtrsim f_s/f_u \gtrsim 2$), but this is independent of whether there are large changes in total abundance with $1 \ll S_t$ or $1 \gg S_t$ (Fig. S3A). The correction does not affect S_n/S_c, but causes a shift in the distribution plane along a line parallel to the line $S_n/S_t = S_c/S_t$, with the magnitude of the shift dependent on the enrichment factor r and on f_s/f_u (Fig. S3B). This affects estimations of the basal distribution between the nuclear and

cytoplasmic compartments. However, if the enrichment factor r is known, the entire 3D space $\{S_n/S_t^*, S_c/S_t^*, S_t^*\}$can be corrected (Fig. S3C).

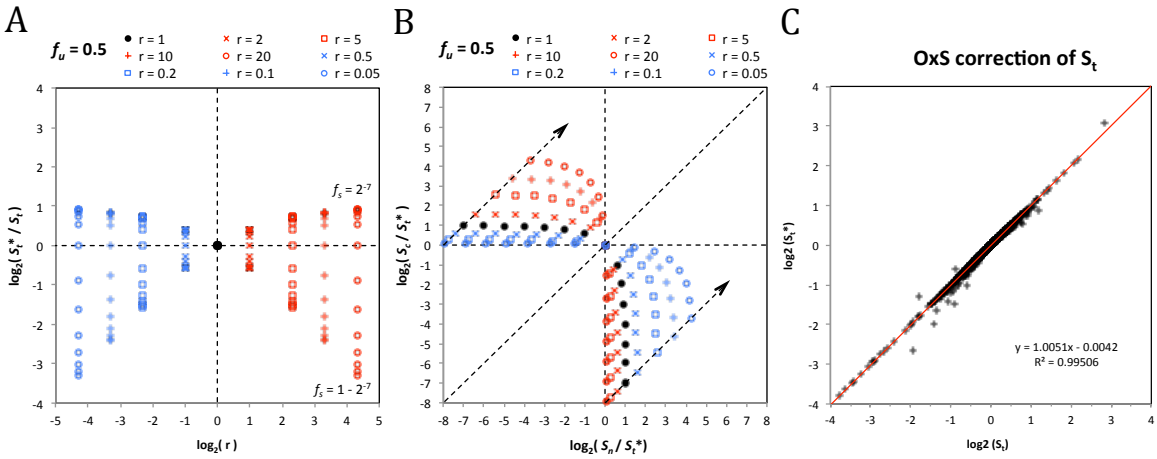


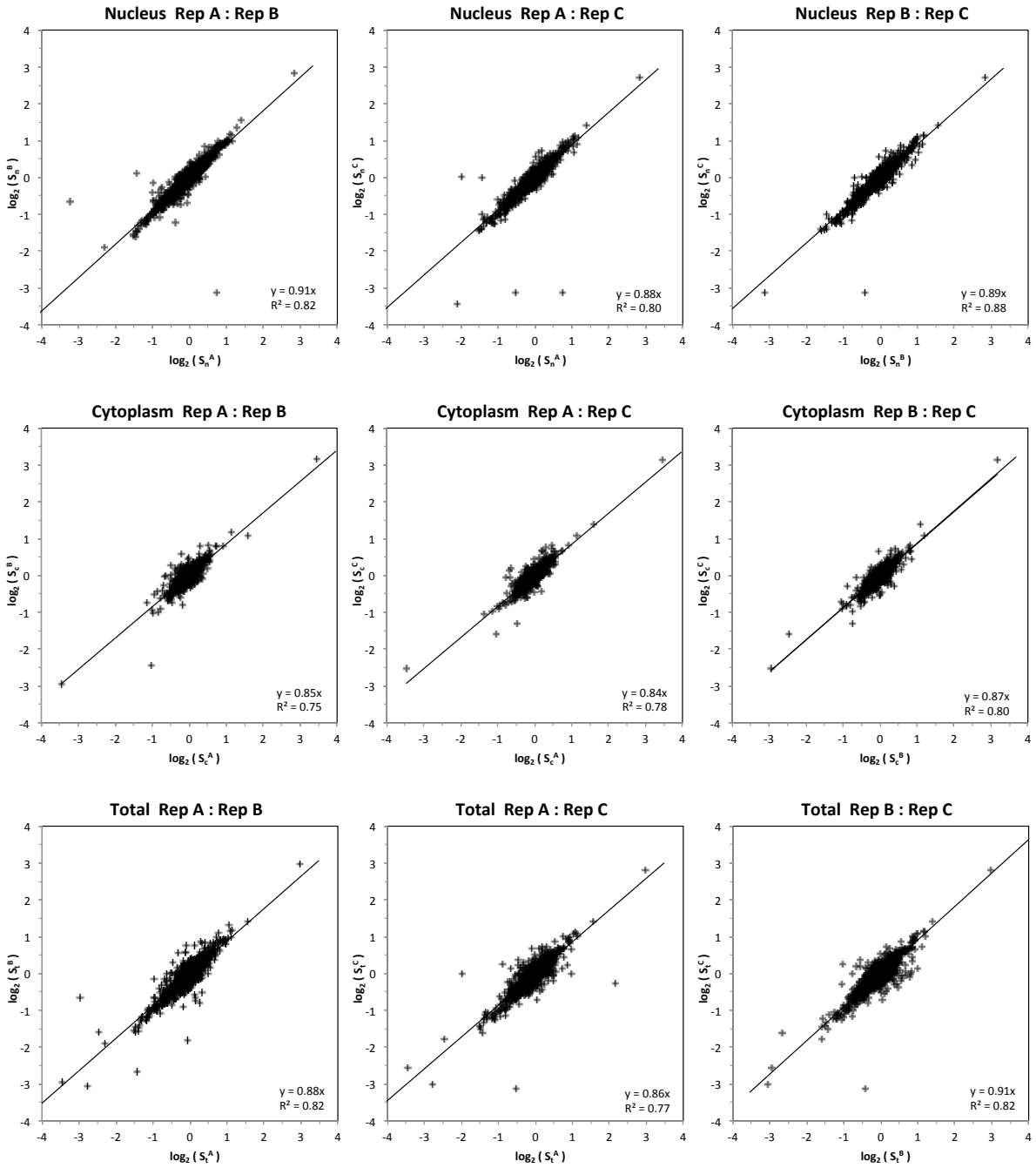
Fig. S3. Effects of nuclear enrichment during MS data collection on values of S_t and on estimation of basal nucleo-cytoplasmic distribution. (A) Variation of (S_t^*/S_t) for different values of the nuclear enrichment factor r with unstimulated nuclear fraction f_u = 0.5 and variation of the stimulated nuclear fraction f_s over the range $f_s = 2^{-7}$ -> 0.5 -> 1 – 2^{-7} in steps of 2^{-n}. (B) Variation in the subcellular distribution plane. (C) Correlation between corrected (S_t) and uncorrected (S_t^*) changes in total abundance for the experimental MS data for treatment of IMR90 cells with TBP.

In practice, for the enrichment factor of r = 5.8 in the present experiments, only very few proteins were modified appreciably by correction for nuclear protein enrichment. For example, for the present data sets, only 22 OxS proteins showed > 10% variation between S_t and S_t^* (Fig. S3C). The slightly altered locations of these proteins in the $\{S_n/S_t, S_c/S_t, S_t\}$ space did not change the overall interpretation of the nature of their response to the stimulations and mainly affects estimations of the basal skewing of abundance between the two compartments. For this reason, the supplementary data tables and the figures in the main text have been prepared on the basis of uncorrected values of S_t. For other applications of the spatial razor, where enrichment factors might be larger or more extreme compartmental redistribution is encountered, a correction for subcellular sample enrichment could become essential.

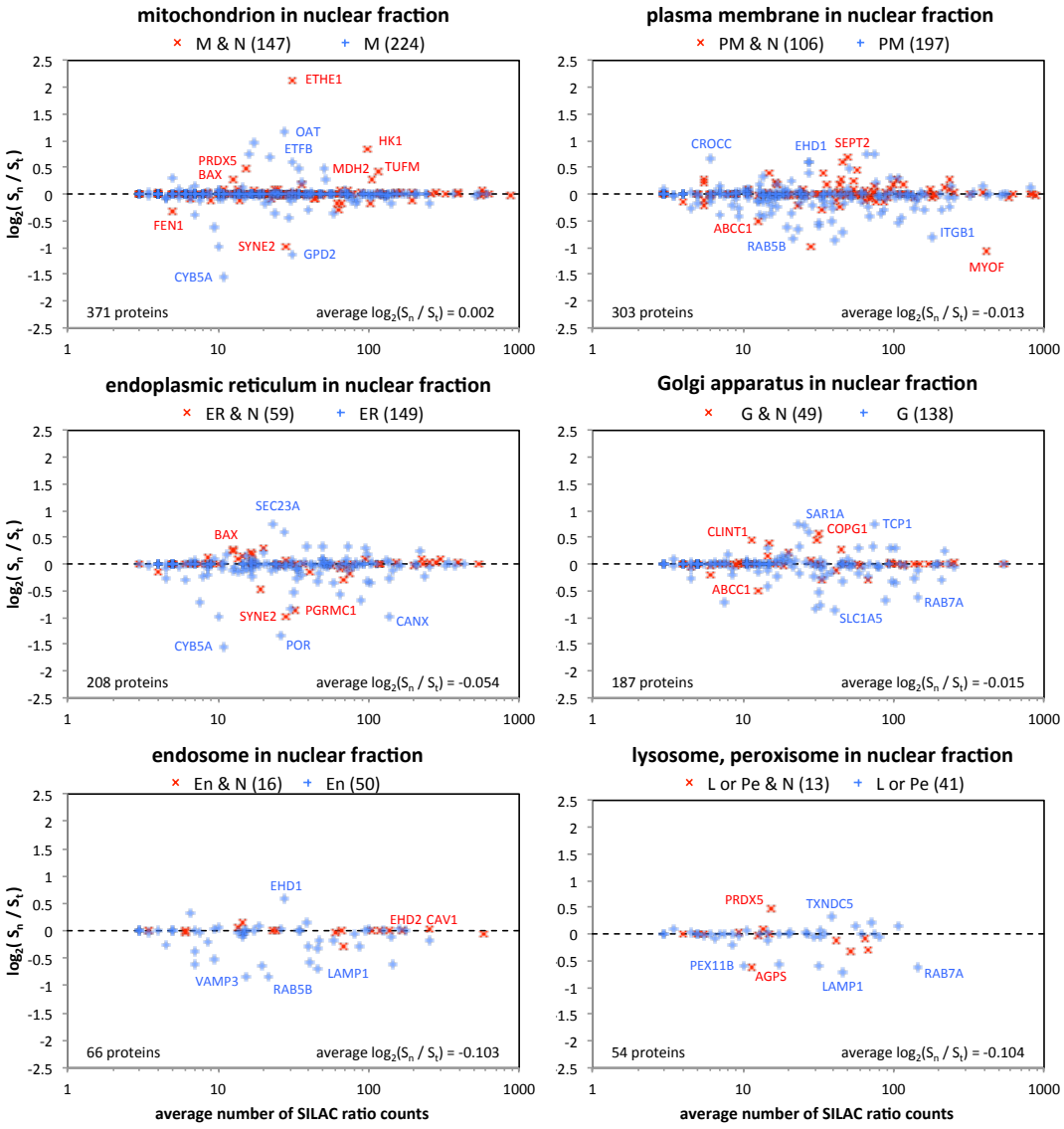
In principle it is possible to use the subcellular spatial razor determine the fraction of a protein in the nuclear compartments for both unstimulated (f_u) and stimulated (f_s) cells (eqns. [2]). As we have described elsewhere,[2] this is a robust method for dealing with MS sampling of different sample types (nucleus and cytoplasm), but requires an independent measurement of S_t to allow checks for each protein of conservation of mass during cellular fractionation, protein extraction and MS sampling.

[1] Mulvey, C. M.; Tudzarova, S.; Crawford, M.; Williams, G. H.; Stoeber, K.; Godovac-Zimmermann, J.: Subcellular proteomics reveals a role for nucleo-cytoplasmic trafficking at the DNA replication origin activation checkpoint. *J Proteome Res* **2013**, *12*, 1436-53.

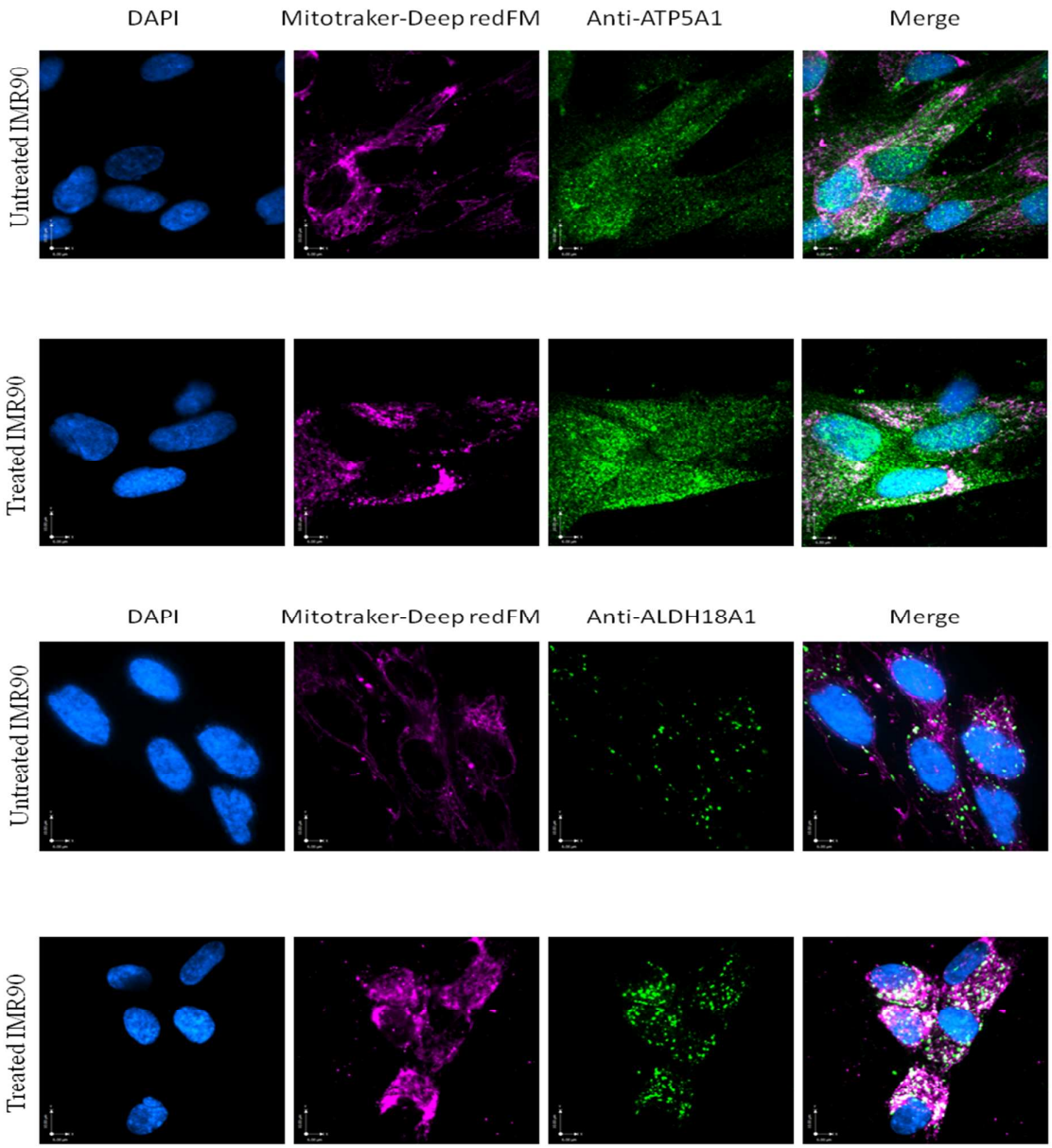
[2] Pinto, G.; Alhaiek, A. A.; Amadi, S.; Qattan, A.; Crawford, M.; Radulovic, M.; Godovac-Zimmermann, J.: Systematic Nucleo-Cytoplasmic Trafficking of Proteins Following Exposure of MCF7 Breast Cancer Cells to Estradiol. *J Proteome Res* **2014**.



Supplementary Figure 1. Correlation of SILAC ratios between replicates for the nucleus, cytoplasm and total (C&N) samples. Proteins with ≥ 3 ratio counts in both replicates are included. With the exception of a handful of outliers with low ratio counts and low recorded MS intensity, correlation between replicates is high.



Supplementary Figure 2. Analysis of the enrichment of the nuclear fraction. Plots of $\log_2(f_s/f_u)$ as a function of the average number of SILAC ratio counts over the nucleus and total data sets for proteins with GO annotation to the indicated subcellular organelles. For each of the subcellular organelles, the set of proteins was divided into two groups: (a) those annotated to the location and the nucleus (red data points); and, (b) those annotated to the location, but not to the nucleus (blue data points). The number of proteins in each set is indicated in the legend at the top of each plot. For each organelle only a minority of proteins show appreciable changes in the fraction of the protein in the nucleus for oxidative stress (f_s) compared to control cells (f_u). Both increases and decreases in nuclear fraction are seen for cells subjected to oxidative stress, but the average over all proteins is $\log_2(f_s/f_u) \sim 0$ for all organelles. There is no discernible difference between proteins that are/are not also annotated to the nucleus. The pattern for all organelles is consistent with appreciable nuclear redistribution for a small set of specific proteins from each organelle. The data also suggests that current GO annotations underestimate the number of different proteins in the nucleus of IMR90 cells.



Supplementary Figure 3. Confocal immunofluorescence imaging of ATP5A1 (top panels) and ALDH18A1 (bottom panels) for IMR90 cells with/without treatment with tert-butyl hydrogen peroxide. Nuclei were stained with DAPI and mitochondria with Mitotracker Deep red FM. The proteins were visualized with the same fluorescent secondary FITC antibody (see Materials and Methods).