Supplementary Information

Bin-free approach to computing relative variability

We sought to develop an alternative approach to compute local estimates of cell-to-cell variability without the need to define discrete bins for cells in each stage of the cell cycle. To do so, we now form the Conditional Variance Estimator (CVE), $\widehat{Var}[X_p](t)$, for a feature *X*, for each protein, *p*, at any possible time, *t*, based on the set of all cells identified, B_q for all other proteins, q.

$$\hat{E}[X_p](t) = \frac{\sum_{q \in C \setminus \{p\}} K_{h_p}(\vec{u}_p - \vec{u}_q) \sum_{c \in B_q} K'_{h'}(t - t_c) w_c x_c}{\sum_{q \in C \setminus \{p\}} K_{h_p}(\vec{u}_p - \vec{u}_q) \sum_{c \in B_q} K'_{h'}(t - t_c) w_c}$$

$$\widehat{Var}[X_p](t) = \frac{\sum_{q \in C \setminus \{p\}} K_{h_p}(\vec{u}_p - \vec{u}_q) \sum_{c \in B_q} K'_{h'}(t - t_c) w_c {x_c}^2}{\sum_{q \in C \setminus \{p\}} K_{h_p}(\vec{u}_p - \vec{u}_q) \sum_{c \in B_q} K'_{h'}(t - t_c) w_c} - \hat{E}[X_p](t)^2$$

In these formulas, we have used t_c to represent the cell stage estimate for cell, *c*, and we have introduced a second, univariate Gaussian kernel, *K'* and associated bandwidth parameter, *h'*. This second kernel weights all the cells by their distance in the cell cycle from the time of interest, *t*. This replaces the cell cycle bins in equation 5 in the main text, where only cells within the bin, *i*, were used to estimate the variance for that bin. We must now select this new bandwidth parameter, *h'*. We use a constant bandwidth

$$h' = \frac{1}{\sqrt{12}} \left(\frac{3}{4} N\right)^{-\frac{1}{5}}$$

Where *N* is the minimum expected number of cells $N = \sum_c w_c$, which we set to be *N*=25 to be consistent with the cutoff of at least 5 cells per bin that we used for the discrete bin method. This choice of bandwidth corresponds to the AMISE optimal bandwidth (Wand 1992) if the cells were Gaussian distributed with variance 1/12. We note that in this alternative (bin-free) approach, the 25 cells need not be distributed equally across the cell cycle stages. This allows us to include substantially more of the proteins in the analysis using this method.

Given the CVE, $\widehat{Var}[X_p](t)$, we compare it to the 'observed' variability at all times, $Var[X_p](t)$. Because we have only observed a finite sample of cells at specific times, we use another CVE to estimate this, but now we only consider the cells from protein *p*:

$$E[X_p](t) = \frac{\sum_{c \in B_q} K'_{h'}(t - t_c) w_c x_c}{\sum_{c \in B_q} K'_{h'}(t - t_c) w_c}$$

$$Var[X_p](t) = \frac{\sum_{c \in B_q} K'_{h'}(t - t_c) w_c {x_c}^2}{\sum_{c \in B_q} K'_{h'}(t - t_c) w_c} - E[X_p](t)^2$$

We define the new relative variability (RV') as an integral over all the cell cycle times, *t*,

$$RV'_p = \frac{1}{2}\int_0^1 \left[\log_2\left(Var[X_p](t)\right) - \log_2\left(\widehat{Var}[X_p](t)\right)\right]dt$$

Since we standardized the cell cycle to range between 0 and 1, these are the limits of this integral. To calculate the RV', we numerically evaluated the integral by summing 256 discrete points over the range of inferred cell stages, 0 to 1.

Overall, the RV in protein abundance or subcellular spread using the binning approach is highly correlated with the bin-free RV' ($r = 0.74$ and $r = 0.72$ respectively) and the extreme examples identified using the RV, such as Hsp78, Cbf1 and Rpc82 are still among the most highly spatially variable proteins identified using RV'. On the other hand, the ranks of many of the top variable proteins differ, indicating that the top hits identified by alternative methods might yield new candidate variable proteins. The complete list ranked according to RV' is available as supplementary data. Exploration of other kernel-based approaches to quantify cell-to-cell variability is an area for further research.

Biological annotation enrichment analysis

Several studies have reported various correlations between cell-cell variability in protein abundance and biological function (e.g., Newman et al. 2006). In this section, we investigate whether proteins with high or low RV share biological function or not. To do so, we evaluated the significance of enrichments of occurrences of biological annotations (GO Annotations, Ashburner et al. 2000) within groups of proteins with extreme RV using the hypergeometric distribution with a Bonferroni correction for multiple testing.

We found that the enrichments in extreme RV are mostly insignificant (Table S1). Only proteins associated with chromosome segregation are over-represented among those with high spatial variability. The absence of strongly significant functional enrichments might be because proteins with similar functions tend to have similar profiles of feature measurements (Handfield et al. 2013), so that the CVEs are likely to be based on variability estimates from proteins of similar function.

RV in	Protein group	# in group (# in 2860 proteins)	Annotation category (most significantly enriched)	Log_{10} P-value (corrected)
Abundance	Top 10%	9 (20)	GO:0031225 Anchored to membrane	-1.00
	Bottom 10%	51 (333)	GO:0005524 ATP binding	+0.26
Subcellular Spread	Top 10%	13 (30)	GO:0007059 Chromosome segregation	-2.45
	Bottom 10%	12 (50)	GO:0000324 fungal type vacuole	+0.72

Table S1 Enrichment analysis for proteins with high or low RV

The Bonferroni correction for 1805 annotation enrichment tests renders all but one P-value insignificant (P-value > 0.05)

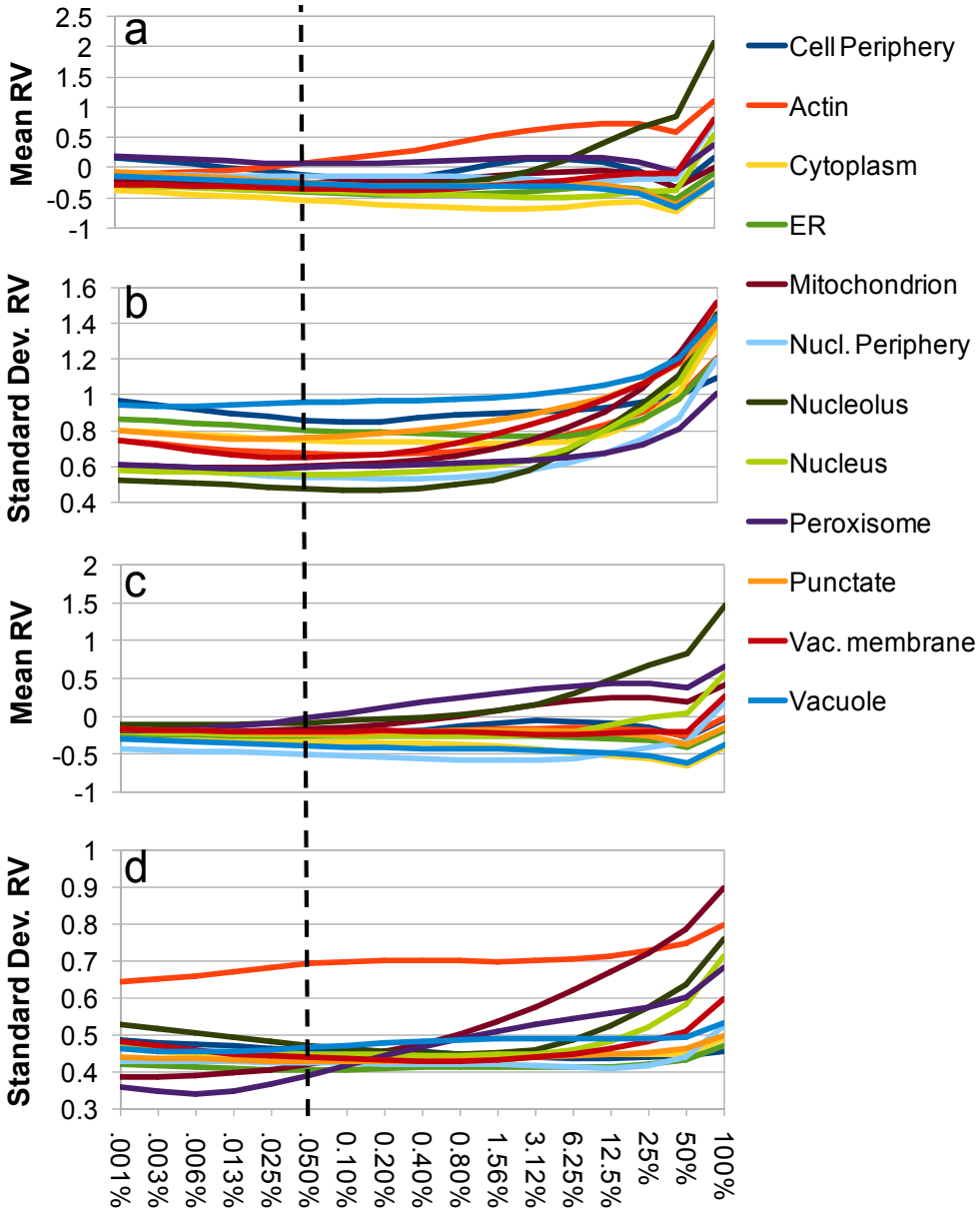
References

Ashburner M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25, 2000.

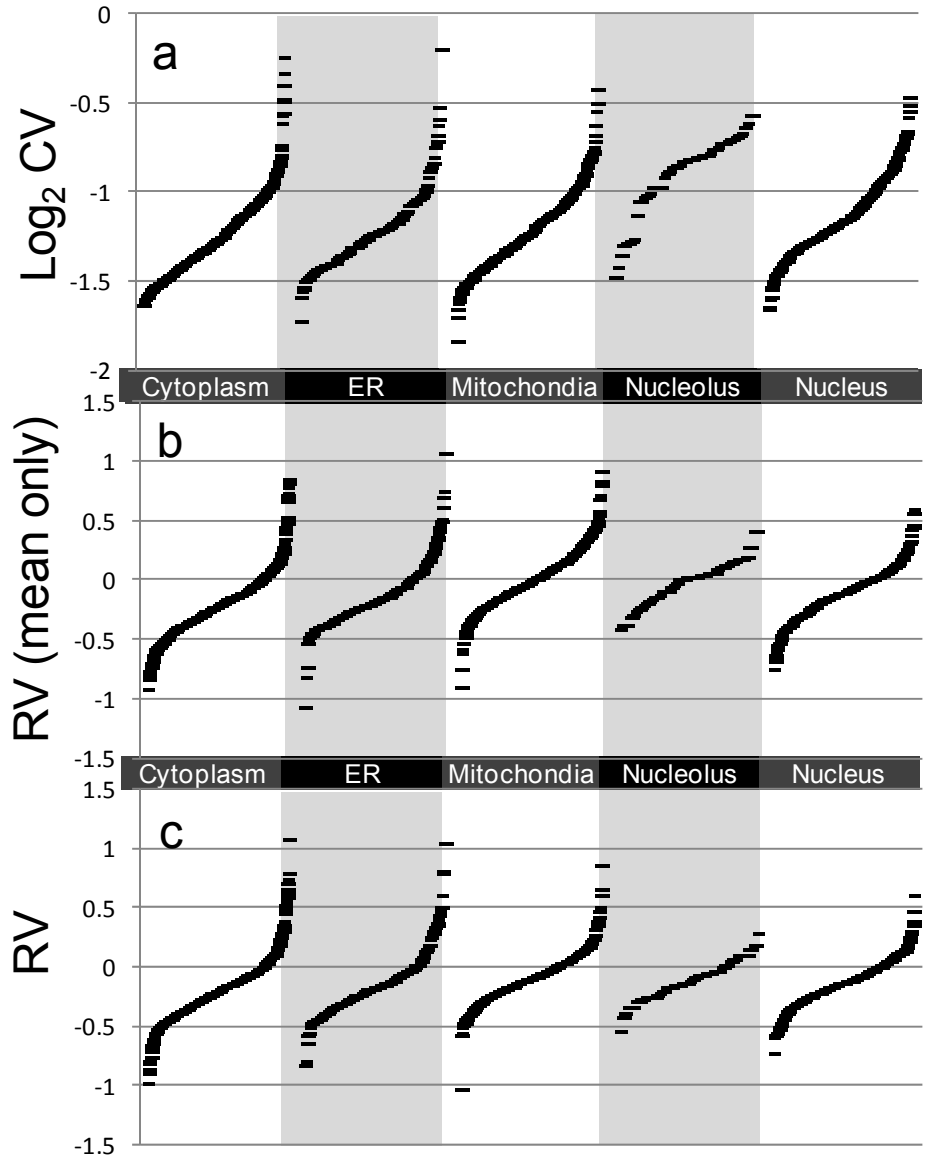
Handfield, L.-F., Chong, Y. T., Simmons, J., Andrews, B. J., and Moses, A. M. (2013). Unsupervised clustering of subcellular protein expression patterns in high-throughput microscopy images reveals protein complexes and functional relationships between proteins. *PLoS computational biology*, 9(6), e1003085.

Newman, J., Ghaemmaghami, S., Ihmels, J., Breslow, D., Noble, M., DeRisi, J., and Weissman, J. (2006). Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095), 840–846.

Wand M.P. Error analysis for general multivariate kernel estimators. *Journal of Nonparametric* Statistics, 2(1):1-15, 1992.

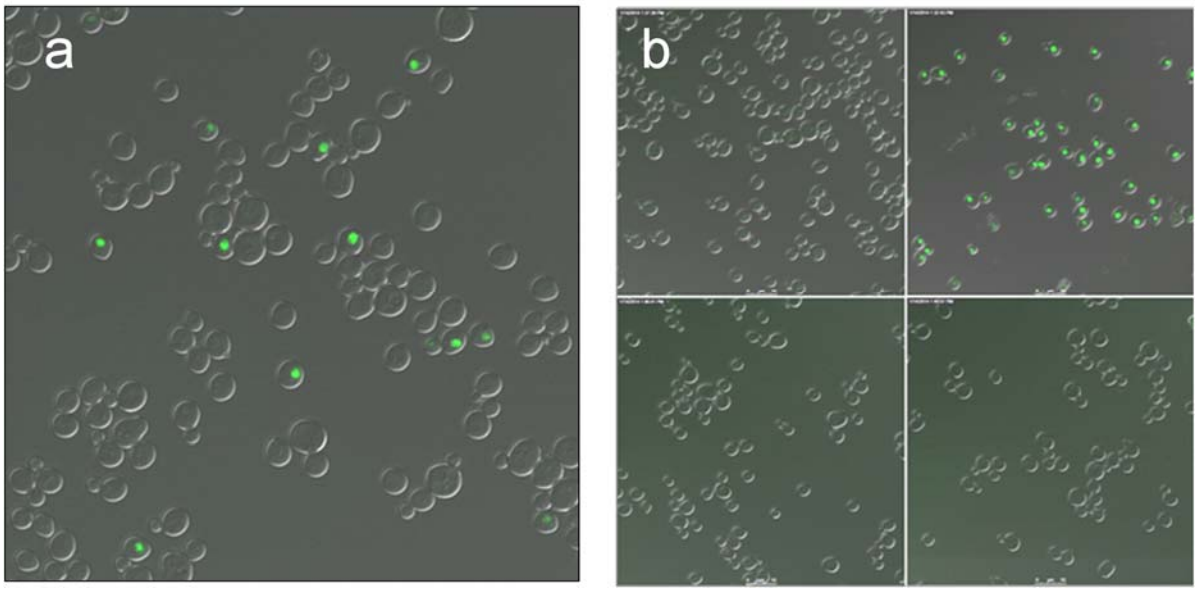


Supplementary Figure 1. Effect of bandwidth (h_p) on relative variability estimates
Plotted is the mean or standard deviation of the relative variability (RV) for various localization classes (indicated with coloured traces) vs. percent of data set when estimating the bandwidth (b_p, equation (6) in the text). Panels a) and b) are the mean and standard deviation for RVs of protein abundance, while c) and d) are the mean and standard deviation for RVs of spatial variability (as measured by the spatial spread feature). The distribution of RV estimates is not very sensitive to changes in the bandwidth. The dotted line indicates the value used in the text.

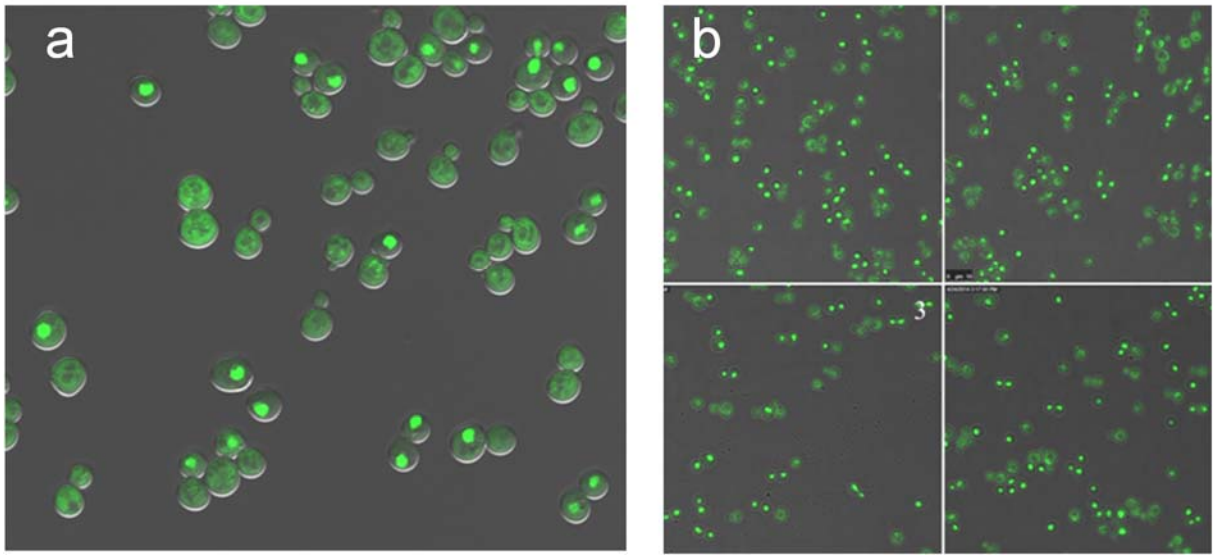


Supplementary Figure 2. Effect of a lower dimensional distance metric on relative variability (RV)

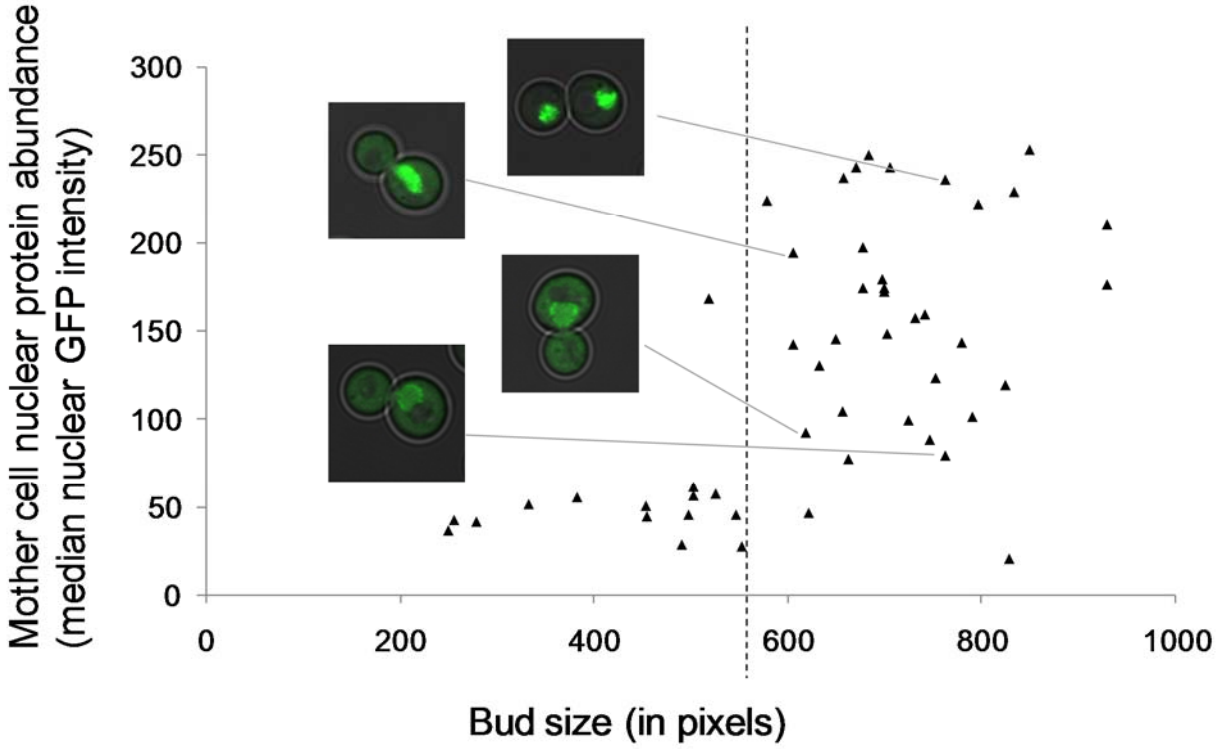
Panels a) and c) show the log CV and RV for protein abundance (as measured by GFP intensity) as shown in Figure 2 of the manuscript. b) shows the RV where only the mean protein abundance features (mean GFP intensity in 5 cell stages for both mother and bud) are used to compute distances between proteins. While this also yields comparable variability estimates, there is still a tendency for mitochondrial proteins to show high variability.



Supplementary Figure 3. Reimaging of Rpc82 indicates that it is likely a mixed culture
a) Image of the GFP-collection strain shows cell-cell variability. Only a small fraction of cells (appear as contours in DIC image) show bright nuclear GFP (shown in green overlay). b) individual colonies from the strain do not recapitulate the mixed pattern. Images show either no cells with nuclear GFP or all cells show nuclear GFP (top right panel). Similar results were obtained for Cbf1 (data not shown).



Supplementary Figure 4. Reimaging of Rnr4 confirms variability
a) Image of the GFP-collection strain shows cell-cell variability. Some cells show bright nuclear GFP (shown in green overlay), while other cells show only cytoplasmic GFP. b) individual colonies from the strain recapitulate the mixed pattern.



Supplementary Figure 5. Manual staging of yeast cells indicates variability in Rnr4 protein localization

The scatter plot shows the GFP intensity in mother cell nucleii as a function of the bud cell size. Each symbol represents a single mother-bud pair. Dashed line indicates the bud size above which variability is observed. Inlaid are four example mother-bud images illustrating the diversity of protein expression patterns.

Supplementary Methods. Microscopy and manual image analysis protocol

GFP-tagged strains of Cbf1 and Rpc82 from the GFP collection were initially grown overnight in liquid media (YPD) and streaked onto plates (synthetic drop-out media supplemented with 2% glucose and 2% amino acids, minus histidine; SD-His). For panel b), clearly isolated colonies were selected and inoculated into SD-His and grown overnight @ 30˚C. Prior to imaging, stationary cultures were diluted 1/10 in fresh SD-His and grown at 30˚C for 4 hours to ensure log-phase growth and proper expression of GFP fusion products. Cells were imaged using a Leica TCS SP8 confocal microscope with a 63X oil-immersion objective and a 10X eye-piece objective. GFP excitation was at 488nm.