Combined effects of protein expression variance and correlation on multicomponent systems

Kyle M. Kovary and Mary N. Teruel

Updated: 2020-02-22 16:12:03

Contents

Abstract

Protein expression variation leads to phenotypic variance between cells. This has been demonstrated in cell signaling and differentiation decisions. Additionally coordinated expression of proteins between cells can tune signaling pathways either towards a more binary or analog modality. Though there is some evidence in bulk cell measurements and in bacteria that certain heteromeric subunits or metabolic pathways may be expressed in a coordinated fashion (i.e. operons in bacteria), there has not been a direct measurement of coordinated expression of proteins (independent of TFs) in metazoans (vertebrates?). Here, we measure cell-to-cell variability of relative protein abundance using quantitative proteomics of individual Xenopus laevis eggs and show that proteins involved in metabolic pathways or members of heteromeric complexes tend to have high correlations with other members of those pathways/complexes. Our previous work highlighted the fact that correlated expression increases the total variation of a pathway, so one would reason that certain pathways or complexes would need to compensate of this extra source of by reducing the variation of expression of these pathways or complexes. To test this we computed total variance score that took into account both the coefficient of variance and correlations between proteins in a pathway and found that the lower 10% of GO terms were highly enriched for metabolic pathways. When we looked at the relationship between CV and R between these GO terms we found a negative relationship between them, demonstrating that increased correlation needs to come at the expense of decreased variance. Simple molecular models of heteromeric complexes and metabolic pathways demonstrate that this tradeoff can result in higher efficiencies in both function and reduced energy waste. Together, our study argues for a control principle whereby coordinated expression of proteins in a pathway can require lower variance in order to reduce the overall pathway variance, enabling accurate control of active complexes or metabolic pathway activity.

Introduction

The coordinated expression of proteins is vital across cellular function, from maintaining a dynamic steady state to differentiation of cells into specialized types in different tissues. The ability to regulate coordinated expression has been demonstrated to occur via transcription (transcription factors, chromatin regulation), translation (specialized ribosomes, mRNA structure/modifications), and degradation (E3 ligases). These processes, as well as the molecules they target, are all subject to noise, which has been shown to lead to

differences in cell behavior and decisions in otherwise identical cells. Variability and coordinated expression (correlation) of proteins leads to increased population level control in binary decisions, and a decreased ability to execute analog signaling.

Despite these important insights, a global understanding of the variability of protein expression between single cells and the coordinated expression of groups of proteins is largely an nonexistent. To systematically assess these properties of single cells, we carried out a proteomics experiment on exceeding large single cells, Xenopus laevis eggs. An advantage of this approach is that we can get around much of the signal to noise issues that accompany studying single cells due to low sample amounts. Additionally, at this stage of development transcription is restricted so we are able to gain insights into non-transcriptional control of protein expression. We also utilized isobaric tagging in order to measure 25 individual eggs in 5 mass spectrometry runs. In this study, we were able to measure the relative abundance >1000 proteins across single cells in order to better understand the properties of stochastic and coordinated expression of proteins.

With this dataset we have been able to, for the first time, measured the relationship between protein expression variance and coordinated protein expression on a proteome scale in single cells. We have observed that certain classes of proteins, including protein complexes and metabolic pathways, are expressed in such a way that increased coordinated expression is balanced by decreased variation. By doing this, cells are able to decrease the total variance of a given complex or pathway, an elegant balancing act that allows for finer control of metabolic throughput and controlling the number of potentially formed complexes though stoichiometric control. Though this kind of coordinated expression leads to an increase in variance in a population of cells, it can reduce variation within a cell.

Results

Single cell proteomics reveals global protein expression variability and coordinated expression between protein pairs.

Activated Xenopus laevis eggs were collected at 5 time points across the first cell cycle (0, 20, 40, 60, and 80 minutes), with 5 eggs at each time point. Using TMT multiplexing and mass spectrometry, we were able to determine the relative abundance of more than 1300 proteins. Expression of these proteins were largely invariant across the cell cycle, revealing that these highly expressed genes are likely not regulated by cell cycle processes. Additionally, a PCA analysis of these eggs showed no discernible clustering on cell cycle time (Fig S1).

To determine the variance across these proteins, we calculated the coefficient of variation (CV) using all time points collected. This showed a wide range of variation containing multiple distributions (Fig 1B), many of which are consistent with our previous study of variation using targeted mass spectrometry (Fig S2). In order to see if variation was a regulated process, we grouped the protein variation by gene ontology terms (GO terms) and plotted them in ranked order. We saw that in general, processes xxx were...

Since we were able to measure all of these proteins within single cells, we are able to calculate coordinated expression of protein pairs at single cell resolution. Using Pearson correlation coefficient, we could determine coordinated expression of nearly 2 million protein pairs (Fig 1C) The distribution of correlation coefficients fit a normal distribution, with the majority of protein pairs appearing to not show significant coregulation. However, there appear to be a significant number of protein pairs containing high correlation coefficients, and the heatmap showed a lot of clustering of these highly coregulated pairs (Fig 1D).