Branching process deconvolution algorithm reveals a detailed cell-cycle transcription program

Xin Guo[a], Allister Bernard[a], David A. Orlando[b,c], Steven B. Haase[b,c], and Alexander J. Hartemink[a,b,c,1]

Departments of [a]Computer Science and [b]Biology, and [c]Program in Computational Biology and Bioinformatics, Duke University, Durham, NC 27708

AUTHOR SUMMARY

Recent advances in imaging have made it possible to measure the dynamics of a small number of molecules in single cells. However, accurate, genome-wide quantification of many molecular species is still only possible in populations of cells. For a population of cells to provide insight into the cell-cycle dynamics of molecules in individual cells, it is necessary to synchronize the cells by arresting them at one stage of the cell cycle and then "releasing" them, allowing them to resume progress through subsequent division cycles. Molecular species can be monitored in the population at various time points after release. However, perfect cell synchrony is neither attainable at synchronization nor maintainable after release. Equally important, cell division is an asymmetric procedure in many kinds of cells, such as budding yeast: After cell division, the newborn daughter cells are smaller than their mothers and have a significantly longer cell cycle. For these reasons, time-series measurements of a cell population do not accurately reflect the dynamics of individual cells as they traverse the cell cycle, but instead represent the convolved dynamics of all cells in the imperfectly synchronized population.

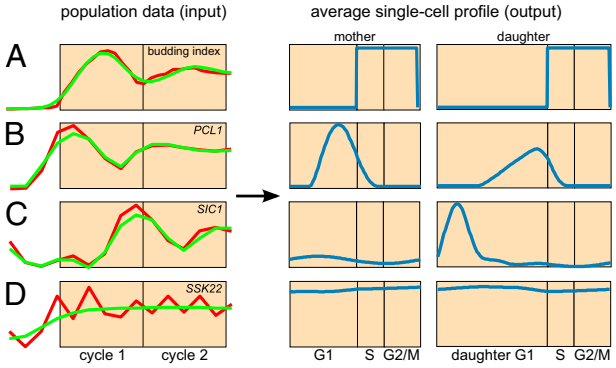

Fig. P1. Deconvolution recovers average single-cell profiles from population-level data for both mother and daughter cells. (*A–D*) (*Left*) Input population-level time-series data (*red*) and the fit to these data under our algorithm (*green*). (*Right*) Deconvolved transcription profiles for both mother and daughter cells (*blue*). (*A*) Budding index measurements: Cells become budded just before the start of S phase. (*B*) *PCL1*: Transcription profiles are fairly similar between mother and daughter cells. (*C*) *SIC1*: Transcription profiles show a sharp burst of expression that is specific to daughter cells. (*D*) *SSK22*: Transcription profiles are denoised by deconvolution.

Here, we present a branching process deconvolution algorithm, a method for efficiently removing the effects of "synchrony loss" from population-level measurements. We use this algorithm to reveal a detailed cell-cycle transcription program at the level of an average single cell. Our algorithm is built upon the CLOCCS (characterizing loss of cell-cycle synchrony) (1, 2) framework for quantitatively determining cell-cycle distributions in population synchrony experiments. The CLOCCS framework models a population of cells during a synchronized time-series experiment, using a branching process to explicitly account for cell division. Using morphological markers, CLOCCS precisely estimates how cells in a population are distributed throughout the cell cycle at any time following synchrony release. Our algorithm then uses these CLOCCS parameter estimates to construct a convolution function, quantifying how individual cell-level measurements will be transformed when collected instead at the population level. The task of deconvolution can therefore be viewed as an inverse problem, seeking to reverse this transformation. We use a wavelet-basis regularization approach to address the ill-posed nature of the problem and to simultaneously tackle the issue of noise in the input data.

Our approach can be applied to population-level measurements (e.g., of transcript levels, protein levels, or genomic occupancy levels of nucleosomes or transcription factors) of any cell type undergoing any dynamic cell-cycle process. To demonstrate the usefulness of our approach, we applied the algorithm to a recent cell-cycle transcription time course in the eukaryote *Saccharomyces cerevisiae*. The algorithm inputs included replicate profiles of transcript levels for 5,670 genes at a temporal resolution of 16 min, along with accurate CLOCCS parameter estimates characterizing the synchrony loss of each replicate. The outputs consisted of 5,670 jointly learned deconvolved transcription profiles at a nominal temporal resolution of less than 1 min, with distinct transcription programs for mother and daughter cells (Fig. P1).

We analyzed the resulting transcription profiles and found that our algorithm reveals subtle timing differences in transcript levels that were obscured in previous population-level measurements. Our method also allowed us to rank genes on the basis of the degree of transcript fluctuation throughout the cell cycle; because of the increased dynamic range afforded by the algorithm, even the 1,500th gene in the ranking exhibited evident fluctuation, suggesting that many more genes may exhibit a greater degree of transcriptional regulation during the cell cycle than has previously been recognized (3–5).

A unique feature of our deconvolution algorithm is that, when applied to population-level measurements across the yeast cell cycle, it can learn distinct cell-cycle programs for mother and daughter cells, because we model them as distinct within the branching process. For example, our algorithm identified 82 genes that appear to be transcribed specifically during G1 in daughter cells. We anticipate that this finding will prove useful for studying late mitotic and early G1 cell-cycle events, as well as the process of cell differentiation in yeast.

Author contributions: X.G., A.B., D.A.O., S.B.H., and A.J.H. designed research; X.G., A.B., and A.J.H. performed research; X.G., A.B., S.B.H., and A.J.H. analyzed data; and X.G. and A.J.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: amink@cs.duke.edu.

See full research article on page E968 of www.pnas.org.

Cite this Author Summary as: PNAS 10.1073/pnas.1120991110.

The ability to distinguish dynamic programs of biologically relevant subpopulations is not limited to mother and daughter cells in budding yeast. By modifying the underlying branching process model, this feature of our deconvolution algorithm could be extended to other systems, thereby leading to the identification of other transcription programs that occur only in distinct subpopulations of cells.

1. Orlando DA, Iversen ES, Jr., Hartemink AJ, Haase SB (2009) A branching process model for flow cytometry and budding index measurements in cell synchrony experiments. *Ann Appl Stat* 3(4):1521–1541.
2. Mayhew MB, Robinson JW, Jung B, Haase SB, Hartemink AJ (2011) A generalized model for multi-marker analysis of cell cycle progression in synchrony experiments. *Bioinformatics* 27(13):i295–i303.
3. Orlando DA, et al. (2008) Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature* 453(7197):944–947.
4. Pramila T, Wu W, Miles S, Noble WS, Breeden LL (2006) The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. *Genes Dev* 20(16): 2266–2278.
5. Spellman PT, et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 9(12): 3273–3297.