A123 · DTWarp: Dynamic Time Warping Alignment for RNA and Protein Identifies Protein-Level Effectors of Epithelial-to-Mesenchymal Transition

Presenter: Ruohong Wang — Boston University

Authors: Ruohong Wang, Pawel F Przytycki

Session: Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

Epithelial-to-mesenchymal transition (EMT) is a fundamental process in development and cancer metastasis, yet the protein-level effectors that execute it remain poorly characterized. Transcriptomic profiling has identified gene expression programs associated with EMT, but mRNA abundance is a limited predictor of protein levels, reflecting variation in translation rates. Prior multi-omic analyses of EMT have also not accounted for the temporal lag between transcript and protein accumulation, making it difficult to distinguish transcriptional responders from downstream protein effectors. To address this gap, we developed DTWarp, which uses soft-DTW to model the mRNA-to-protein temporal lag by aligning single-cell transcriptomic data to bulk proteomic time series, then applies quadratic programming to estimate per-cluster proteomic contributions. Soft-DTW considers multiple alignment paths rather than committing to a single optimal warp, accommodating timing uncertainty between modalities. We applied DTWarp to TGFβ-stimulated MCF-10A cells profiled with bulk transcriptomics, proteomics, and single-cell RNA-seq collected at shared timepoints but not paired at the cellular level. Across ten transcriptionally defined clusters, DTWarp nominated cluster-specific genes as candidate protein actuators. Because the lag correction shifts the frame of reference from transcription to protein accumulation, these candidates capture protein-level behavior rather than transcriptional timing. We validated them against an independent single-cell proteomics dataset from the same EMT system, confirming enrichment of candidate proteins in their predicted transition stages. These results show that modeling the mRNA-to-protein temporal lag reveals protein-level regulators of EMT otherwise invisible to temporally naive multi-omic analyses.

DTWarp: Dynamic Time Warping Alignment for RNA and Protein Identifies Protein Level Effectors of Epithelial-to-Mesenchymal Transition

Ruohong Wang[1*] and Pawel F Przytycki[1,2#]

1. Graduate Program in Bioinformatics, Boston University, Boston, MA, USA
2. Faculty of Computing & Data Sciences, Boston University, Boston, MA, USA

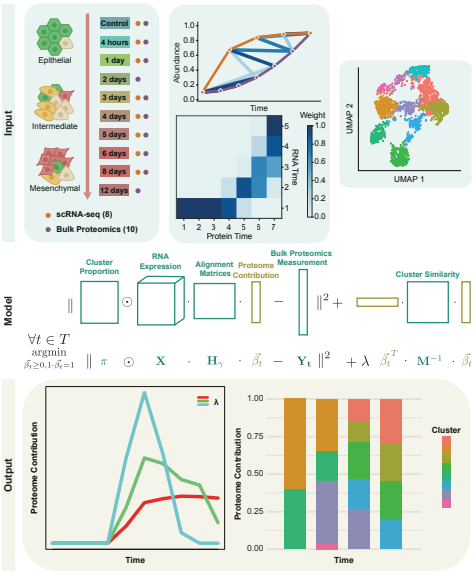
[#] Corresponding author: pawel@bu.edu

Background. Epithelial-to-mesenchymal transition (EMT) is a fundamental process in development and cancer metastasis, yet the protein-level effectors that execute it remain poorly characterized. Transcriptomic profiling has identified gene expression programs associated with EMT, but mRNA abundance is a limited predictor of protein levels, reflecting variation in translation rates. Single cell RNA sequencing resolves the subpopulations that define hybrid EMT states but reports only transcription, whereas bulk proteomics measures protein while averaging across cells. No approach links these layers while accounting for the gene specific timing offset between transcription and translation.

Methods. We used a TGFβ induced MCF10A EMT dataset[1] with single cell RNA sequencing and bulk proteomics across time. DTWarp aligns each gene's mRNA and protein trajectories independently with soft-DTW[2], an alignment method that accommodates gene specific, nonuniform lags and produces a per gene alignment matrix remapping transcript dynamics onto protein time. A quadratic program then combines these alignments with single cell cluster proportions to decompose the bulk proteome into subpopulation specific contributions at each time point, regularized by a cross-cluster similarity term that ties contributions across transcriptomically related clusters and constrained, so contributions sum to one per time point. Finally, a Leave-One-Gene-Out (LOGO) procedure refits the model with each gene removed and scores it by how strongly its removal shifts the predicted contribution across time points, clusters, and regularization strengths, yielding time sensitive markers.

Results. The decomposition recovers a per cluster proteome contribution that follows cluster proportion over time while correcting for the transcription to translation lag, with the consistent protein lagging RNA pattern emerging from the data itself. LOGO markers accumulated in an order concordant with each cluster's position along the epithelial-to-mesenchymal axis, with epithelial clusters accumulating markers early and mesenchymal clusters later, recovering the expected sequence of EMT without any time information built into selection. In independent single-cell proteomics data, marker sets assigned to the epithelial, intermediate, and mesenchymal stages each peaked at their matching stage, including an epithelial adhesion protein (DST), an intermediate cytoskeletal linker (EZR), and a mesenchymal TGFβ SMAD target (TAGLN).

Conclusion. DTWarp links transcription and translation with gene specific timing to resolve the bulk proteome into subpopulation specific contributions, recovering EMT ordering from data alone and validating stage specific markers at the protein level.



Overview of DTWarp. Inputs and alignment: scRNA-seq and bulk proteomics across time are aligned with soft dynamic time warping that remaps transcript dynamics onto protein time, and cluster similarity is calculated across scRNA-seq clusters. **Decomposition Model:** combining gene level alignments with cluster proportions and similarities to decompose the bulk proteome into subpopulation specific contributions per time point. **Output:** the estimated per cluster proteome contribution over time.

References. 1. Paul, I. et al. Parallelized multidimensional analytic framework applied to mammary epithelial cells uncovers regulatory principles in EMT. Nat Commun 14, 688 (2023). 2. Cuturi, M. & Blondel, M. Soft-DTW: a Differentiable Loss Function for Time-Series. Proceedings of the 34th International Conference on Machine Learning 70, 894-903 (2017).