bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dynamics of single-cell protein covariation during epithelial–mesenchymal transition

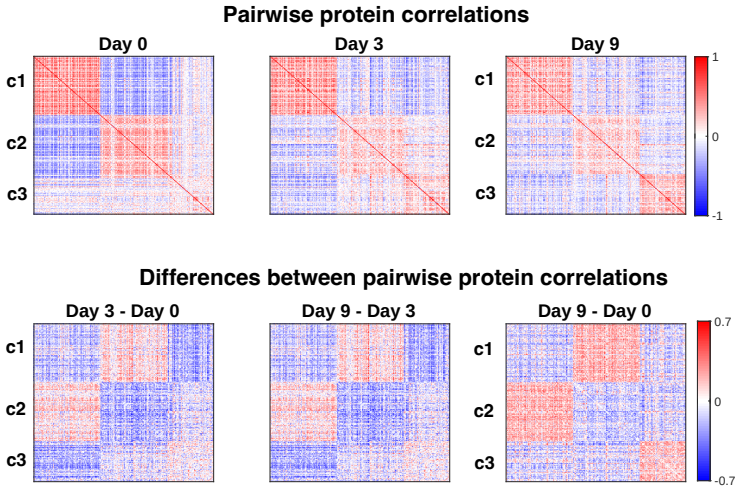
Saad Khan[1,2], Rachel Conover [1], Anand R. Asthagiri[1,2,3,✉], Nikolai Slavov[1,2,4,✉]

1 Department of Bioengineering, Northeastern University, Boston, MA, USA
2 Department of Biology, Northeastern University, Boston, MA, USA
3 Department of Chemical Engineering, Northeastern University, Boston, MA, USA
4 Parallel Squared Technology Institute, Watertown, MA 02472, USA

✉ Correspondence: a.asthagiri@northeastern.edu, nslavov@northeastern.edu
∈ Data, code & protocols: scp.slavovlab.net/Khan_et_al_2023

Abstract

Physiological processes, such as epithelial–mesenchymal transition (EMT), are mediated by changes in protein interactions. These changes may be better reflected in protein covariation within cellular cluster than in the temporal dynamics of cluster-average protein abundance. To explore this possibility, we quantified proteins in single human cells undergoing EMT. Covariation analysis of the data revealed that functionally coherent protein clusters dynamically changed their protein-protein correlations without concomitant changes in cluster-average protein abundance. These dynamics of protein-protein correlations were monotonic in time and delineated protein modules functioning in actin cytoskeleton organization, energy metabolism and protein transport. These protein modules are defined by protein covariation within the same time point and cluster and thus reflect biological regulation masked by the cluster-average protein dynamics. Thus, protein correlation dynamics across single cells offer a window into protein regulation during physiological transitions.



bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Epithelial-mesenchymal transition (EMT) plays a key role in embryonic and adult development, tissue repair and wound healing, and pathologies such as fibrosis and cancer[1–3]. EMT involves major changes in cell behavior. While most attention is given to cell morphology, adhesion, migration and invasiveness, EMT also impacts cell cycle activity, senescence, apoptosis, metabolism, genomic stability, stemness and drug resistance, among other cell behaviors. The pleiotropic effect of EMT on cell behaviors is mediated by complex regulatory networks, including transcription factors, post-transcriptional and post-translational signaling, intercellular communication and the microenvironment.

EMT regulation across single cells is nonuniform. Single-cell RNA and morphological analysis show significant variability among cells undergoing EMT[4,5]. In what ways variability in protein abundance contributes to single-cell heterogeneity during EMT remains unclear. Since RNA level is an unreliable predictor of protein abundance, protein level variation cannot be reliably inferred from single-cell RNA data[6–9]. Indeed, post-transcriptional processes, such as endocytosis, protein synthesis, modifications and degradation, play a significant role in EMT regulation[10–12].

While this variability poses challenges for population average measurements, it offers the potential to infer regulatory processes from protein covariation across single cells[13]. Indeed protein co-variation across individual cells may reflect protein interactions and can be detected by single-cell protein measurements with sufficient accuracy, precision and throughput[14,15]. Such analysis is becoming feasible due to the development of powerful single-cell mass spectrometry (MS) proteomic methods[16–24]. As a result, protein covariation across single cells can be quantified[25–28].

Here, we applied Single Cell ProtEomics[29–31] to quantify the proteomes of single cells undergoing EMT triggered by TGFβ, a prominent stimulus for EMT in physiological and pathophysiological processes. Analyzing the system across time, we observed within cluster variation across single cells and strong protein covariation across the single cells within each time point. We systematically quantified this covariation and its change in time. This revealed clusters of proteins whose covariation evolved concertedly in time. These concerted changes in protein covariation include singling proteins, cytoskeleton proteins and metabolic enzymes whose mean abundance per time point does not change. Thus, protein covariation across single cells provides information about cellular remodeling that is complementary to changes in protein abundance.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Dynamics of proteins abundance and variability during EMT

We aimed to quantify EMT-mediated changes in protein abundance over the time span of several days. To this end, we performed single-cell proteomics on non-transformed human mammary epithelial cells (MCF-10A) treated with transforming growth factor beta (TGFβ) for 0 (untreated), 3 and 9 days, Fig. 1. The chosen dosage of TGFβ has been shown by us and others to induce overt EMT in MCF-10A cells[32–34]. The time points were chosen to interrogate transient (3 days) and sustained (9 days) responses to TGFβ during which cells will be in intermediary and overt phases of EMT. Our MS measurements yielded data for 4,571 proteins across 420 individual cells. However, many of proteins were quantified in a small fraction of the cells, and we focused on a subset of 1,893 proteins that were quantified in over 30 single cells from the dataset and over 5 single cells from each time point.

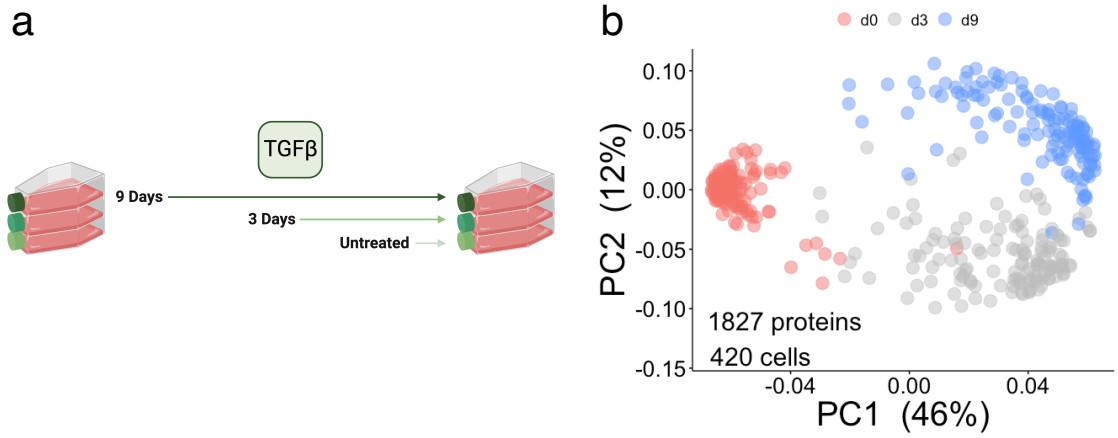


Figure 1 | Experimental design and dataset overview (**a**) MCF-10A cells were treated by TGFβ for the indicated duration and then collected for protein analysis. (**b**) Single cells plotted in the space of their principal components are colored by the day post TGFβ treatment.

Principal component analysis (PCA) of protein abundance data shows that cells from the three time points segregate into clusters in PC space. The separation suggests that magnitude of changes in protein abundance is large enough to readily distinguish cells at different phases of EMT. The first PC primarily separates untreated cells from TGFβ-treated cells regardless of how far into

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EMT they have progressed. Cells with transient (3 d) and sustained (9 d) exposure score similarly along PC1. Meanwhile, the second PC separates cells with transient vs. sustained exposure to TGFβ. Proteins loaded heavily along this axis distinguish intermediary vs. overt stages of EMT.

Dynamics of protein covariation during EMT

Next, we sought to analyze pairwise protein correlations for a set of proteins selected by correlation vector analysis (similar to previously uses in ref[29,35]) to be correlated at both days 0 and 9 but in different ways. An example of such pair is shown in Fig. 2a: RSPH10B and TUBB3 correlate negatively in epithelial cells (Day 0) and positively in mesenchymal cells (Day 9). This change in correlation is statistically significant (p val $< 10^{-13}$, q val $< 0.1\%$). Day 3 (whose data was not used for protein selection) exhibits and intermediate correlation pattern, suggesting concerted changes in correlation over EMT progression.

The change in pairwise correlations in Fig. 2a is not associated with significant changes in the mean abundance of these proteins in Days 0, 3 and 9, suggesting that the changes in correlation patterns may reveal molecular rearrangements inaccessible from measuring mean protein levels. This observation is bolstered by the data for other proteins pairs exhibiting similarly concerted temporal changes in covariation without significant changes in mean abundance, Fig. 2b.

Next, we sought to expand the correlation patterns suggested by individual pairs of proteins (Fig. 2) to a systematic exploration of changing global patterns of covariation, Fig. 3. To this end, we started by selecting the subset of proteins with multiple pairwise observations (i.e., all proteins for which we can compute pairwise correlations from measured protein abundances) and substantial changes in covariation. To identify proteins with changing correlations, we subtracted the correlations for Day 0 from the correlations for Day 9 and computed the norms of the vectors of correlation differences. Then we selected the proteins (50% percentile) having the largest magnitude difference between Day 0 and Day 9 (Fig. S2) or (20% percentile) having the largest magnitude difference between Day 0 and Day 9 (Fig. S3). The correlations between these proteins defined 3 main clusters when clustered hierarchically Fig. S2, and thus we used K-means clustering with k=3 to define 3 discrete clusters (c1, c2, and c3).

These 3 clusters are well defined, as shown by the matrices of pairwise correlations within each cluster (Fig. 3a) and their corresponding differences across time (Fig. 3b). The global change in correlations for each cluster is monotonic (Fig. 3), and this monotonicity is quantified by the dy-

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

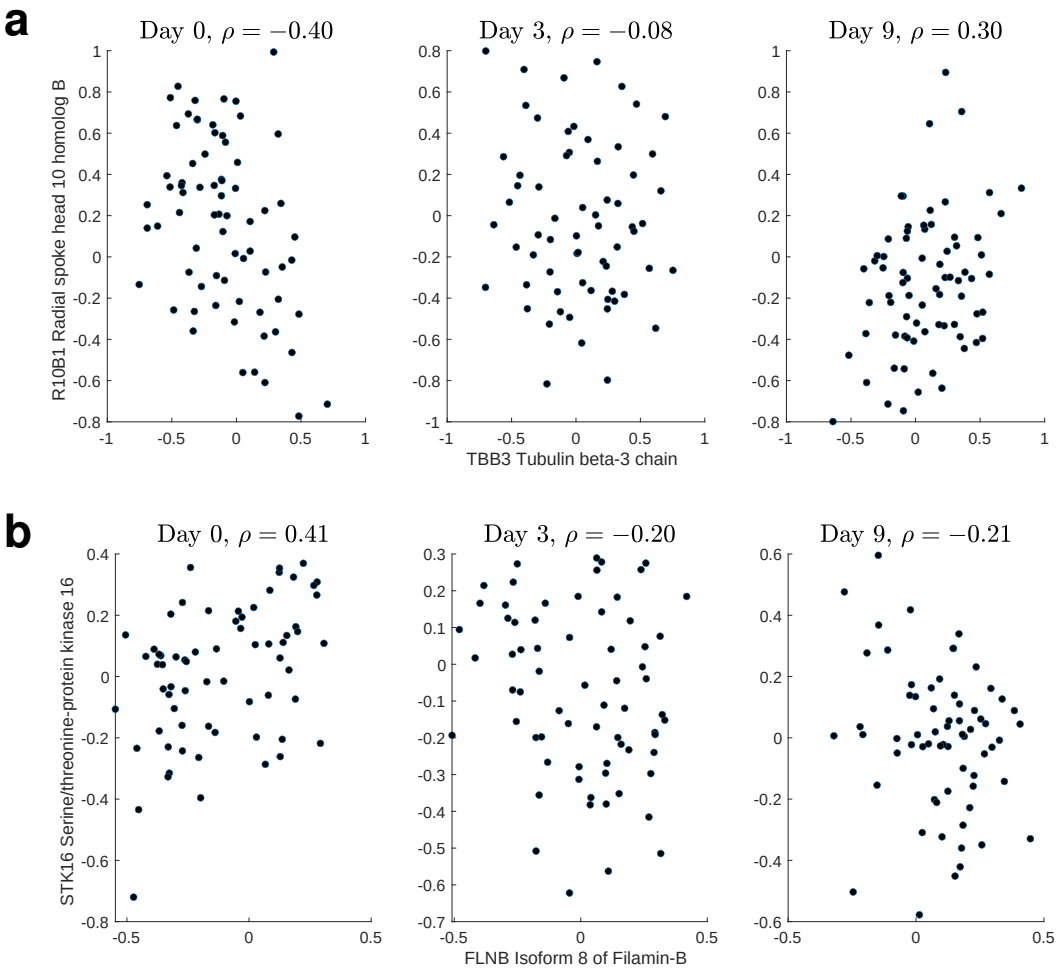


Figure 2 | Protein correlations change in the course of EMT. (**a**) Pairwise correlations between RSPH10B and TUBB3 at each of the 3 time points. (**b**) Pairwise correlations between STK16 and FLNB at each of the 3 time points. The probability of observing these changes in correlations in the randomized data is below low, $p < 10^{-8}$, $q < 10^{-2}$.

namics of mean cluster correlations displayed in Fig. 4a. Since the data from Day 3 have not been used for selecting or clustering proteins, their consistency with the monotonic trends bolsters our confidence in the results.

Remarkably, the dynamics of the mean protein correlations for clusters 1,2, and 3 are not reflected in the corresponding dynamics of mean protein abundances, Fig. 4. For a particular cluster, such as c3, the correlations scale in time while the proteins from the cluster remain positively correlated to each other, as shown in Fig. 4a. Based on this observation, we estimated the mean correlation for each cluster and plotted the estimates over time, Fig. 4b. Similarly we estimates, the mean protein abundance of each cluster over time (Fig. 4c), and evaluated the dependence between these mean cluster estimates as shown in Fig. 4d. The result indicates no dependence between the dynamics

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

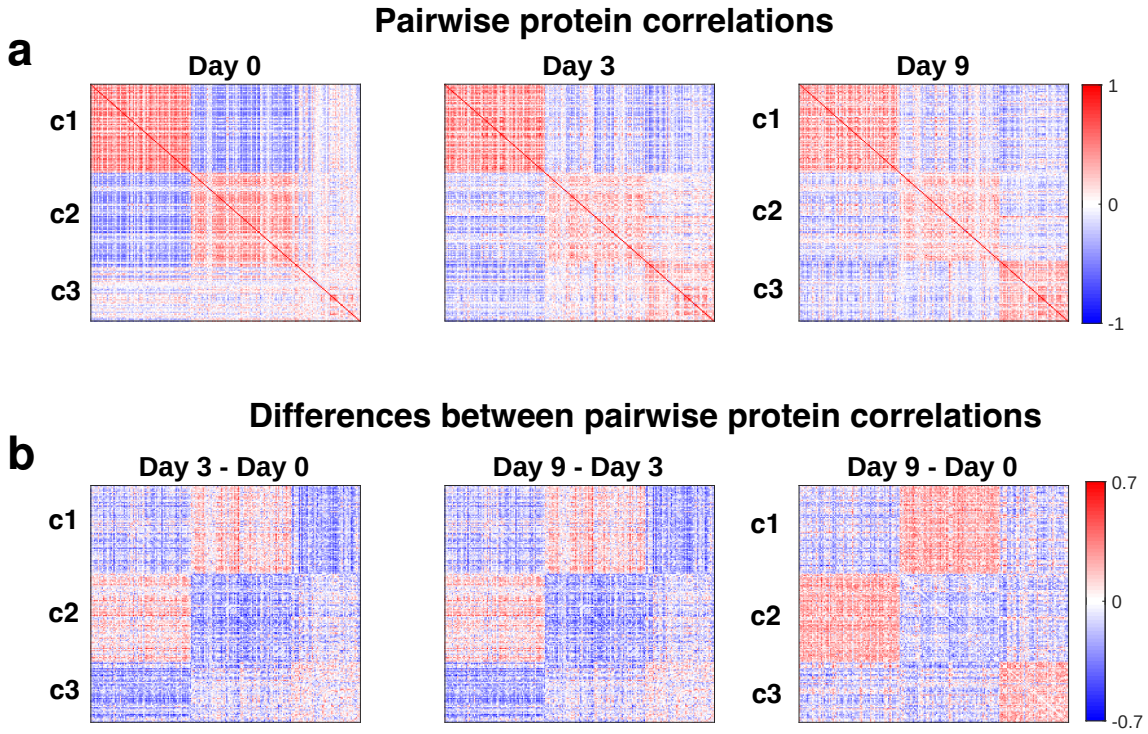


Figure 3 | Dynamics of protein covariation during EMT (**a**) Matrices of pairwise protein correlations at days 0, 3 and 9 clustered using k-means clustering with $k = 3$. The 3 clusters are denoted by c1, c2, and c3. (**b**) Matrices of differences between pairwise protein correlations for the indicated time points. The rows and columns for all days correspond to the same proteins ordered in the same way, namely based on clustering the matrix of correlation differences between Day 9 and 0.

of protein correlations and mean protein abundance, consistent with the observation for the protein pairs shown in Fig. 2. There results reveal biological signals reflected in the protein covariation across single cells from the same time point but not from the corresponding cluster-average protein abundance, Fig. 4a-c.

To identify biological functions corresponding to each cluster, we performed gene ontology (GO) term analysis[36] for terms enriched in each cluster relative to all analyzed proteins. We found a many statistically significant protein groups in these clusters, which are provided as Supplemental Tables and a few characteristic groups are highlighted in Table 1. The first cluster comprises proteins involved in cytsokeletal regulation, including actin, vimentin, tubulin subunits, vinculin, filamins and contractility regulators, such as RhoA, myosin and tropomyosin. The proteins in this cluster broadly span cytoskeletal regulation and show statistically significant enrichment when the analysis is extended to more proteins, Fig. S3. Meanwhile, the proteins in the second and third clusters showed statistically-significant enrichment for metabolism and protein synthesis and transport, respectively (Supplemental Tables). The second cluster was enriched for proteins involved

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

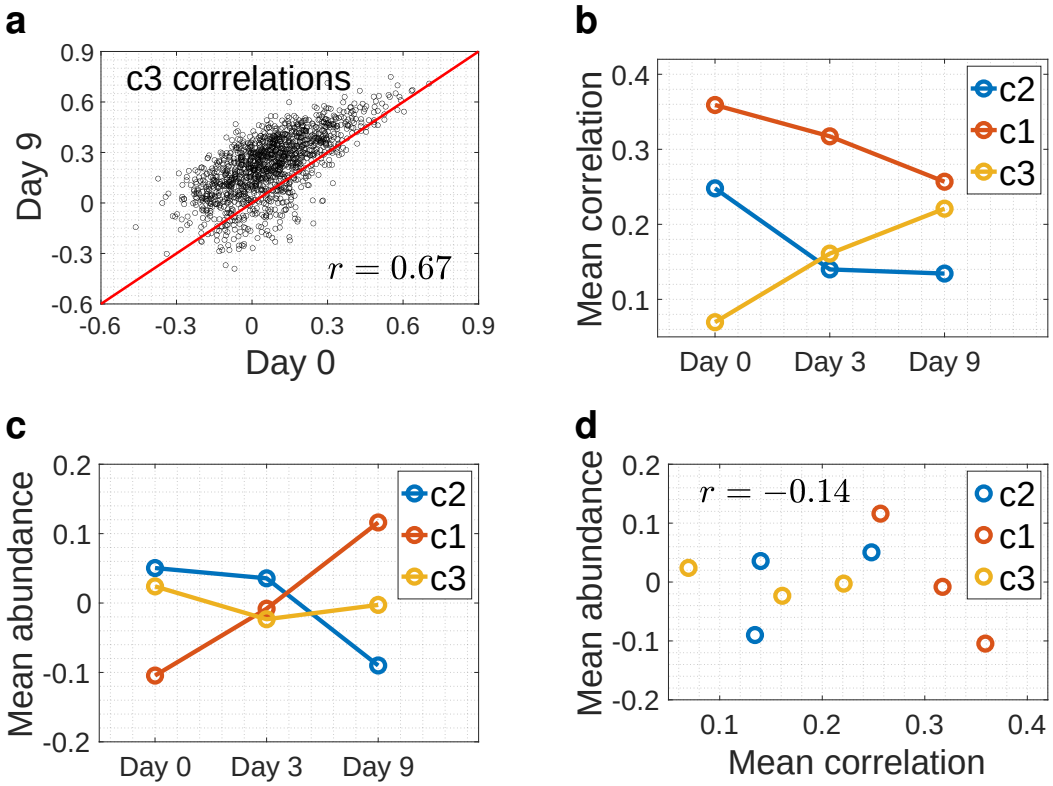


Figure 4 | Differences between the dynamics of within-cluster protein correlations and cluster-average abundance. (**a**) Comparison of pairwise correlations for cluster 3 (c3) between days 0 and 9. (**b**) Mean correlations among the proteins from each cluster across time. (**c**) Mean abundance of the proteins from each cluster across time. (**d**) Correspondence between mean cluster correlations and mean cluster protein abundances. The correlation is weak ($r = -0.14$) and not statistically significant ($p = 0.7$).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cluster	Enriched function	Proteins
c1	regulation of actin filament-based process	fascin, RhoA, IQGAP1, Arp3
c2	glucose and pyruvate metabolic processes	pyruvate kinase, aldolase, enolase, lactate dehydrogenase
	oxidative phosphorylation	ATP synthase, cytochrome C oxidase
c3	protein synthesis and transport	ribosomal proteins
	telomere maintenance, cell response to DNA damage	XRCC5, XRCC6, PARP1, PCNA

Table 1 | Enriched biological functions in the correlation clusters. The table summarizes protein sets and associated representative proteins form performing GO Gorilla analysis on cluster c1, c2, and c3. All terms are significant at 1% FDR. The full results from the analysis are available as Supplemental Tables.

in glycolysis and oxidative phosphorylation, consistent with the role of EMT in regulating aerobic and anaerobic utilization of glucose[37]. Many of the enriched functions found in the third cluster involved protein synthesis and transport, including core ribosomal proteins. The third cluster is enriched for other functions associated with EMT, including telomere regulation, and senescence and cell response to DNA damage. Taken together, the protein functions found across the three clusters correspond to the multi-faceted effect of EMT on cellular functions.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Protein covariation across single cells may identify regulatory interactions[13], and here we demonstrate its potential to delineate dynamic remodeling of protein networks during EMT. Our work builds upon previous observations of RNA covariation[38] and protein covariation[25–28] and extends the anaysis and interpretation towards temporal dynamics during cellular transitions.

Proteins covariation has to important benefits relative to RNA covariation for inferring biological regulation. First, proteins quantification is based on sampling sufficient number of copies from most proteins (often 100s of copies)[29,39,40] to support reliable quantification of correlations across single cells (not clusters of cells), as shown in Fig. 2. Second, much of the regulation may be driven by protein synthesis and degradation of protein subunits of complexes[41], and this component is detectable only at the protein level. For these two reasons, protein covariation offers an informative perspective towards biological regulation[13].

We based our analysis on correlations between proteins with many pairwise observations. Many of the proteins quantified in our dataset did not have sufficient number of pairwise observations to be included in this analysis due to the stochastic approach of shotgun data acquisition. This limitations can be overcome in future studies by using prioritised data acquisition (pSCoPE)[28,42] or multiplexed data independent acquisition (plexDIA)[43–45]. Thus, using the latest generation of single-cell proteomic technology and future innovations[46] will further empower the approach that we used in this study.

Materials and methods

Cell culture

Non-transformed human mammary epithelial cells (MCF-10A, ATCC) were maintained in growth medium consisting of DME medium/Ham's F-12 (ThermoFisher) containing HEPES and L-glutamine supplemented with 5% (v/v) horse serum (ThermoFisher), 20 ng/ml EGF (Peprotech), 0.5 μg/ml hydrocortisone, 0.1 μg/ml cholera toxin, 10 μg/ml insulin (Sigma), and 1% penicillin-streptomycin (ThermoFisher), as described previously[47]. To induce EMT, cells were treated with 20 ng/ml TGFβ (Peprotech) in growth medium for 0, 3 and 9 days[32]. TGFβ-containing medium was refreshed every three days.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sample preparation

Cells were harvested as single-cell suspension and prepared for MS analysis using Nano-ProteOmic sample Preparation (nPOP) as described by Leduc *et al.*[48,49]. The automated collection of prepared samples had not been developed yet[26] and so samples were manually collected using a pipette (using $5\mu l$ of mass spectrometry grade Acetonitrile then water respectively) and transferred into a 384-well plate (Thermo AB1384). The samples were then dried down in a SpeedVac vacuum evaporator and resuspended in $1.07\mu l$ of 0.1% Formic Acid (buffer A) and tightly sealed using an aluminium foil cover (Thermo Fisher AB0626).

Peptide separation and MS data acquisition

The separation was performed at a constant flow rate of 200nL/min using a Dionex UltiMate 3000 UHPLC. From the $1.07\mu l$ of sample in each well, 1 μl was loaded onto a 25cm $\times$ 75 μM IonOpticks Odyssey Series column (ODY3-25075C18). The separation gradient was 4% buffer B (80% acetonitrile in 0.1% Formic Acid) for 11.5 minutes, a 30 second ramp up to 8%B followed by a 63 minute linear gradient up to 35%B. Subsequently, buffer B was ramped up to 95% over 2 minutes and maintained as such for 3 additional minutes. Finally, buffer B was dropped to 4% in 0.1 minutes and maintained for 19.9 additional minutes.

The mass spectra were acquired using a Thermo Scientific Q-Exactive mass spectrometer from minutes 20 to 95 of the LC method. The electrospray voltage of 1700 V was applied at the liquid liquid junction of the analytical column and transfer line. The temperature of the ion transfer tube was 250°C, and the S-lens RF level was set to 80.

After a precursor scan from 450 to 1600 m/z at 70,000 resolving power, the top 7 most intense precursor ions with charges 2 to 4 and above the AGC min threshold of 20,000 were isolated for MS2 analysis via a 0.7 Th isolation window with a 0.3 Th offset. These ions were accumulated for at most 300 ms before being fragmented via HCD at a normalized collision energy of 33 eV (normalized to m/z 500, z = 1). The fragments were analyzed at 70,000 resolving power. Dynamic exclusion was used with a duration of 30 s with a mass tolerance of 10 ppm.

MS data searching

The raw data was searched by MaxQuant, a software package for proteomics data analysis, against a protein sequence database that included all entries from the human SwissProt database and known contaminants. The MaxQuant search was performed using the standard workflow, which includes trypsin digestion and allows for up to two missed cleavages for peptides with 7 to 25 amino acids. Tandem mass tags (TMTPro 16plex) were specified as fixed modifications, while

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

methionine oxidation and protein N-terminal acetylation were set as variable modifications. Carbamidomethylation was disabled as a fixed modification because it was not performed. Second peptide identification was also disabled. The calculation of peak properties was enabled. All peptide-spectrum-matches (PSMs) and peptides found by MaxQuant were exported to the evidence.txt files. The confidence in the PSMs was further updated using Dart-ID, which is a Bayesian framework for increasing the confidence of PSMs that were consistently identified at the same retention time with high-confidence PSMs for the same amino acid sequences[50]. The updated data were filtered at 1% FDR for both peptides and proteins as described by Petelski *et al.*[30].

Data filtering and analysis

The peptide by cells matrices were processed by the SCoPE2 analysis pipeline[29,30,51], which resulted in 4,571 proteins quantified across 420 single cells. However, each single cell had only about 1,000 proteins quantified and many proteins were quantified in relatively few single cells. Thus, we subset the proteins to the 1,893 proteins quantified in at least 30 single cells from the dataset and at least 3 single cells from each time point.

We computed the pairwise Pearson correlations for each time point between these 1,893 proteins using only measured abundances, without imputation, which resulted in three $1,893 \times 1,893$ correlation matrices, R_1, R_2, and R_3 for days 0, 3 and 9 respectively. We further computed the Pearson correlations among correlation vectors of R_1 and R_3 corresponding to the same protein as previously described[25,29,35], resulting in a vector $\mathbf{v}$. Each element of $\mathbf{v}$ corresponds to one protein and quantifies the similarity of its correlations to the remaining 1,892 proteins. The elements of $\mathbf{v}$ were used to explore pairwise combination of proteins whose correlations change significantly between days 0 and 9 and the examples shown in Fig. 2. To calculate the statistical significance for the change in the correlations, we computed the same correlation difference for 10^8 randomized samples and estimated the fraction of randomized samples whose correlation difference exceeds the difference observed in the data.

For the systematic analysis of correlations in Fig. 3, we computed the matrix of correlation differences $\Delta R = R_3 - R_1$ and preserved only its rows and columns corresponding to a set ϕ of 418 proteins with no missing data, i.e, the corresponding correlations could be computed only from quantified proteins both for day 0 and 9. Then, we quantified the average magnitude of correlation change for each of these 418 proteins by computing the norm of its corresponding column in ΔR_ϕ, resulting in a vector $\mathbf{m}$. To select the proteins whose correlations change the most, we selected the subset ω of 209 proteins having norms in $\mathbf{m}$ larger than the 50% percentile of $\mathbf{m}$ (the median of $\mathbf{m}$). Then we performed means clustering with $k = 3$ on ΔR_ω and used the 3 resulting clusters to display $R_{1\omega}$, $R_{2\omega}$, and $R_{3\omega}$ and their differences in Fig. 3. As an alternative approach, we clustered

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ΔR_ω hierarchically and used the resulting permutation to order the rows and columns of $R_{1\omega}$, $R_{2\omega}$, and $R_{3\omega}$ and their differences, as displayed in Fig. S2.

To quantitatively display the dynamics of correlations for the 3 clusters derived by K-means clustering, we computed the mean correlation for all pairwise correlations between proteins assigned to a cluster for each of the 3 time points, Fig. 4. Similarly, we computed the average protein abundance for all proteins (log2 fold change relative to the mean) assigned to a cluster for each of the 3 time points. Only measured protein values were used for computing the average-cluster protein abundance; no imputation values were used.

Availability

Data, metadata, code, and protocols are organized according to community recommendations[24] and available at supplemental information and at scp.slavovlab.net/Khan_et_al_2023. The Raw MS data are available at MassIVE: MSV000092872 and ProteomeXchange: PXD045423.

Acknowledgments

We thank Morgan Benson, Audrey Kidd and Andrew Leduc for help with sample preparation and members of the Asthagiri and Slavov labs for feedback and helpful discussions. The work was funded by an NCI award R21CA246150-01A1 to A.R.A. and N.S., and Allen Distinguished Investigator award through The Paul G. Allen Frontiers Group to N.S., an NIGMS award R01GM144967 to N.S., and a MIRA award from the NIGMS of the NIH to N.S. (R35GM148218) to N.S.

Competing interests

Nikolai Slavov is a founding director and CEO of Parallel Squared Technology Institute, which is a non-profit research institute.

References

1. Brabletz, S., Schuhwerk, H., Brabletz, T. & Stemmler, M. Dynamic EMT: a multi-tool for tumor progression. *EMBO J* **40,** e108647 (2021).
2. Nieto, M., Huang, R., Jackson, R. & Thiery, J. EMT: 2016. *Cell* **166,** 21–45 (2016).
3. Lambert, A. W. & Weinberg, R. A. Linking EMT programmes to normal and neoplastic epithelial stem cells. *Nat. Rev. Cancer* **21,** 325–338 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Deshmukh, A. *et al.* Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. *Proc Natl Acad Sci U S A* **118,** e2102050118 (2021).

5. Wang, W., Poe, D., Yang, Y., Hyatt, T. & Xing, J. Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. *Elife* **11,** e74866 (2022).

6. Franks, A., Airoldi, E. & Slavov, N. Post-transcriptional regulation across human tissues. *PLoS computational biology* **13,** e1005535. https://doi.org/10.1371/journal.pcbi.1007082 (2017).

7. Eraslan, B. *et al.* Quantification and discovery of sequence determinants of protein-per-mRNA amount in 29 human tissues. *Mol. Syst. Biol.* **15,** e8513 (Feb. 2019).

8. Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165,** 535–550 (2016).

9. Van den Berg, P. R., Bérenger-Currias, N. M. L. P., Budnik, B., Slavov, N. & Semrau, S. Integration of a multi-omics stem cell differentiation dataset using a dynamical model. *PLOS Genetics* **19,** 1–23. https://doi.org/10.1371/journal.pgen.1010744 (May 2023).

10. Díaz, V., Viñas-Castells, R. & García de Herreros, A. Regulation of the protein stability of EMT transcription factors. *Cell Adh Migr* **8,** 418–428 (2014).

11. Aiello, N. *et al.* EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev Cell* **45,** 681–695.e4 (2018).

12. Janda, E. *et al.* Raf plus TGFbeta-dependent EMT is initiated by endocytosis and lysosomal degradation of E-cadherin. *Oncogene* **25,** 7117–7130 (2006).

13. Slavov, N. Learning from natural variation across the proteomes of single cells. *PLOS Biology* **20,** 1–4. https://doi.org/10.1371/journal.pbio.3001512 (Jan. 2022).

14. Slavov, N. Unpicking the proteome in single cells. *Science* **367,** 512–513. https://doi.org/10.1126/science.aaz6695 (2020).

15. Slavov, N. Scaling Up Single-Cell Proteomics. *Molecular & Cellular Proteomics* **21,** 100179. ISSN: 1535-9476. https://doi.org/10.1016/j.mcpro.2021.100179 (2022).

16. Budnik, B., Levy, E., Harmange, G. & Slavov, N. Mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *bioRxiv 10.1101/102681.* https://doi.org/10.1101/102681 (2017).

17. Singh, A. Towards resolving proteomes in single cells. en. *Nat. Methods* **18,** 856 (Aug. 2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Furtwängler, B. *et al.* Real-Time Search Assisted Acquisition on a Tribrid Mass Spectrometer Improves Coverage in Multiplexed Single-Cell Proteomics. en. *Mol. Cell. Proteomics* (2022).

19. Slavov, N. Single-cell protein analysis by mass spectrometry. *Current Opinion in Chemical Biology* **60,** 1–9. ISSN: 1367-5931. https://doi.org/10.1016/j.cbpa.2020.04.018 (2020).

20. Cong, Y. *et al.* Ultrasensitive single-cell proteomics workflow identifies > 1000 protein groups per mammalian cell. *Chemical Science* **12,** 1001–1006 (2021).

21. Johnston, S. M. *et al.* Rapid, One-Step Sample Processing for Label-Free Single-Cell Proteomics. en. *J. Am. Soc. Mass Spectrom.* ISSN: 1044-0305, 1879-1123. https://pubs.acs.org/doi/10.1021/jasms.3c00159 (July 2023).

22. Matzinger, M., Müller, E., Dürnberger, G., Pichler, P. & Mechtler, K. Robust and Easy-to-Use One-Pot Workflow for Label-Free Single-Cell Proteomics. en. *Anal. Chem.* ISSN: 0003-2700, 1520-6882. http://dx.doi.org/10.1021/acs.analchem.2c05022 (Feb. 2023).

23. Orsburn, B. C. Metabolomic, Proteomic, and Single-Cell Proteomic Analysis of Cancer Cells Treated with the KRASG12D Inhibitor MRTX1133. *Journal of Proteome Research* **22,** 3703–3713 (2023).

24. Gatto, L. *et al.* Initial recommendations for performing, benchmarking, and reporting single-cell proteomics experiments. *Nat. Methods* **20,** 375–386. https://doi.org/10.1038/s41592-023-01785-3 (2023).

25. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass-spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biology* **19,** 161. https://doi.org/10.1186/s13059-018-1547-5 (2018).

26. Leduc, A., Huffman, R. G., Cantlon, J., Khan, S. & Slavov, N. Exploring functional protein covariation across single cells using nPOP. *Genome Biology* **23,** 261. https://doi.org/10.1186/s13059-022-02817-5 (2022).

27. Hu, M. *et al.* Correlated Protein Modules Revealing Functional Coordination of Interacting Proteins Are Detected by Single-Cell Proteomics. en. *J. Phys. Chem. B* **127,** 6006–6014. ISSN: 1520-6106, 1520-5207. http://dx.doi.org/10.1021/acs.jpcb.3c00014 (July 2023).

28. Huffman, R. G. *et al.* Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. *Nat. Methods.* http://dx.doi.org/10.1038/s41592-023-01830-1 (2023).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biology* **22** (2021).

30. Petelski, A. A. *et al.* Multiplexed single-cell proteomics using SCoPE2. *Nature Protocols* **16,** 5398–5425. https://doi.org/10.1038/s41596-021-00616-z (2021).

31. Leduc, A., Koury, L., Cantlon, J. & Slavov, N. Massively parallel sample preparation for multiplexed single-cell proteomics using nPOP. *bioRxiv* (2023).

32. Milano, D. *et al.* Positive Quantitative Relationship between EMT and Contact-Initiated Sliding on Fiber-like Tracks. *Biophys J* **111,** 1569–1574 (2016).

33. Natividad, R., Lalli, M., Muthuswamy, S. & Asthagiri, A. Golgi Stabilization, Not Its Front-Rear Bias, Is Associated with EMT-Enhanced Fibrillar Migration. *Biophys J* **115,** 2067–2077 (2018).

34. Brown, K. *et al.* Induction by transforming growth factor-beta1 of epithelial to mesenchymal transition is a rare event in vitro. *Breast Cancer Res* **6,** R215–31 (2004).

35. Slavov, N. & Dawson, K. A. Correlation signature of the macroscopic states of the gene regulatory network in cancer. *Proceedings of the National Academy of Sciences* **106,** 4079–4084. eprint: http://www.pnas.org/content/106/11/4079.full.pdf+html (2009).

36. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinformatics* **10,** 1–7 (2009).

37. Jia, D. *et al.* Towards decoding the coupled decision-making of metabolism and epithelial-to-mesenchymal transition in cancer. *Br J Cancer* **124,** 1902–1911. https://pubmed.ncbi.nlm.nih.gov/33859341 (2021).

38. Chapman, A. R. *et al.* Correlated gene modules uncovered by high-precision single-cell transcriptomics. en. *Proc. Natl. Acad. Sci. U. S. A.* **119,** e2206938119. ISSN: 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.2206938119 (Dec. 2022).

39. Specht, H. & Slavov, N. Transformative opportunities for single-cell proteomics. *Journal of Proteome Research* **17,** 2563–2916. https://doi.org/10.1021/acs.jproteome.8b00257 (8 June 2018).

40. MacCoss, M. J. *et al.* Sampling the proteome by emerging single-molecule and mass spectrometry methods. en. *Nat. Methods* **20,** 339–346. https://www.nature.com/articles/s41592-023-01802-5 (Mar. 2023).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41. Gonçalves, E. *et al.* Widespread Post-transcriptional Attenuation of Genomic Copy-Number Variation in Cancer. en. *Cell Syst* **5,** 386–398.e4. ISSN: 2405-4712. http://dx.doi.org/10.1016/j.cels.2017.08.013 (Oct. 2017).

42. Extending the sensitivity, consistency and depth of single-cell proteomics. en. *Nat. Methods.* ISSN: 1548-7091, 1548-7105. http://dx.doi.org/10.1038/s41592-023-01786-2 (Apr. 2023).

43. Derks, J. *et al.* Increasing the throughput of sensitive proteomics by plexDIA. *Nature Biotechnology.* https://doi.org/10.1038/s41587-022-01389-w (2022).

44. Framework for multiplicative scaling of single-cell proteomics. en. *Nat. Biotechnol.,* 1–2. https://www.nature.com/articles/s41587-022-01411-1 (July 2022).

45. Derks, J. & Slavov, N. Strategies for increasing the depth and throughput of protein analysis by plexDIA. *Journal of Proteome Research* **22,** 697–705. https://doi.org/10.1021/acs.jproteome.2c00721 (2023).

46. Slavov, N. Driving Single Cell Proteomics Forward with Innovation. *Journal of Proteome Research* **20,** 4915–4918. https://doi.org/10.1021/acs.jproteome.1c00639 (2021).

47. Kim, J. & Asthagiri, A. Matrix stiffening sensitizes epithelial cells to EGF and enables the loss of contact inhibition of proliferation. *J Cell Sci* **124,** 1280–1287 (2011).

48. Leduc, A., Huffman, R. G. & Slavov, N. Droplet sample preparation for single-cell proteomics applied to the cell cycle. *bioRxiv 2021.04.24.441211* (2021).

49. Leduc, A., Huffman, R., Cantlon, J., Khan, S. & Slavov, N. *Highly Parallel Droplet Sample Preparation for Single Cell Proteomics* en. https://www.protocols.io/view/highly-parallel-droplet-sample-preparation-for-sin-4r3l24r7qg1y/v3. Accessed: 2022-9-3. Apr. 2022.

50. Chen, A. T., Franks, A. & Slavov, N. DART-ID increases single-cell proteome coverage. *PLOS Computational Biology* **15,** 1–30. https://doi.org/10.1371/journal.pcbi.1007082 (July 2019).

51. Specht, H. *et al.* Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Zenodo,* 10.5281/zenodo.4339954. 10.5281/zenodo.4339954 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Figures

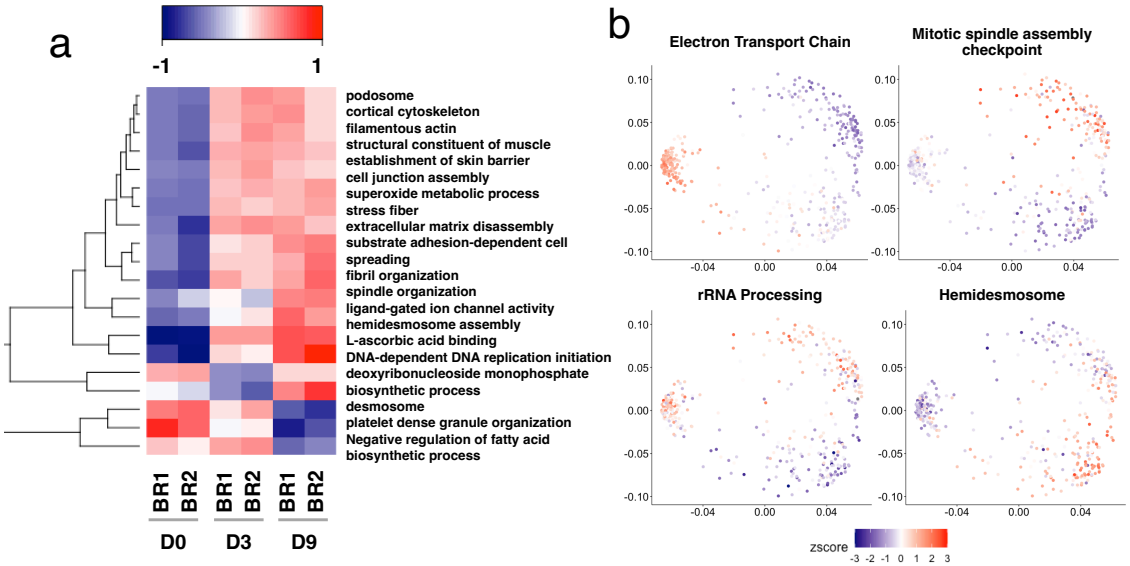


Figure S1 | Gradients of expression in single cells extend mean levels observed in orthogonal bulk data. (**a**) Statistically significant enrichment of related proteins in bulk data over the course of TGFβ treatment observed across two biological replicates. (**b**) Z scored abundances of select, significantly differential bulk GO terms have been used to color each single cell in PC space.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

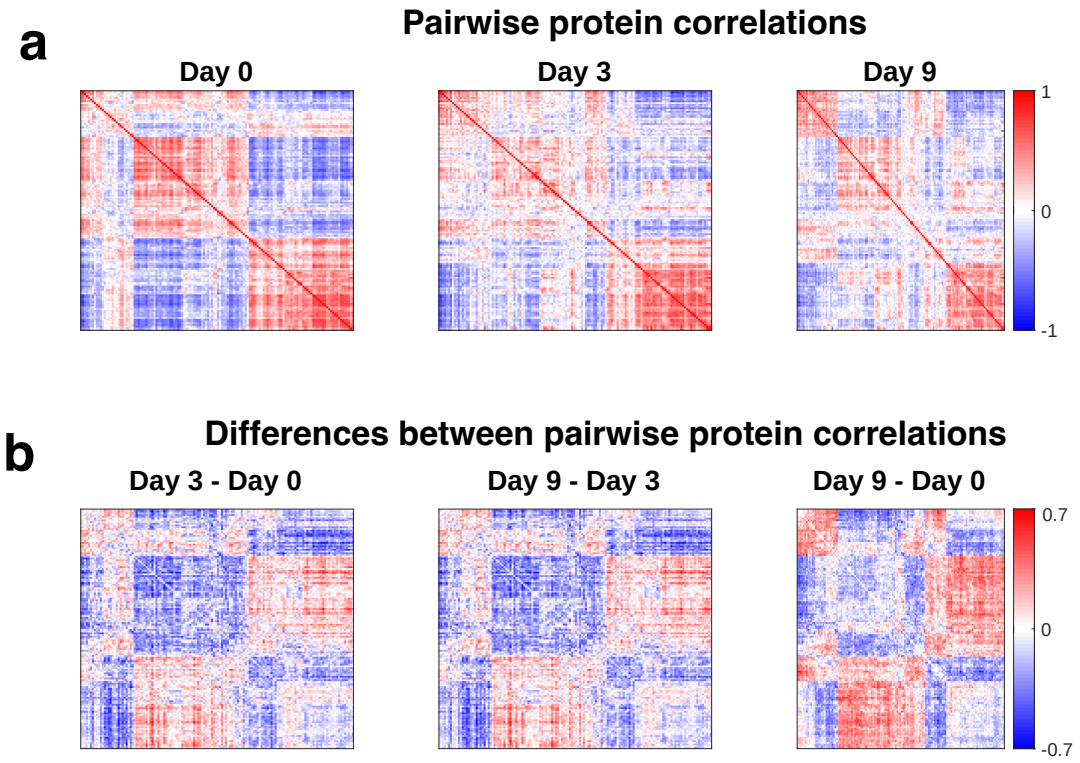


Figure S2 | Dynamics of protein covariation during EMT The the proteins whose correlations have the largest magnitude (exceeding the 50% percentile) difference between Day 0 and Day 9 were selected and their correlation matrices clustered hierarchically. (**a**) Matrices of pairwise protein correlations at days 0, 3 and 9. (**b**) Matrices of differences between pairwise protein correlations for the indicated time points. The rows and columns for all days correspond to the same proteins ordered in the same way, namely based on clustering the matrix of correlation differences between Day 9 and 0.

bioRxiv preprint doi: https://doi.org/10.1101/2023.12.21.572913; this version posted December 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

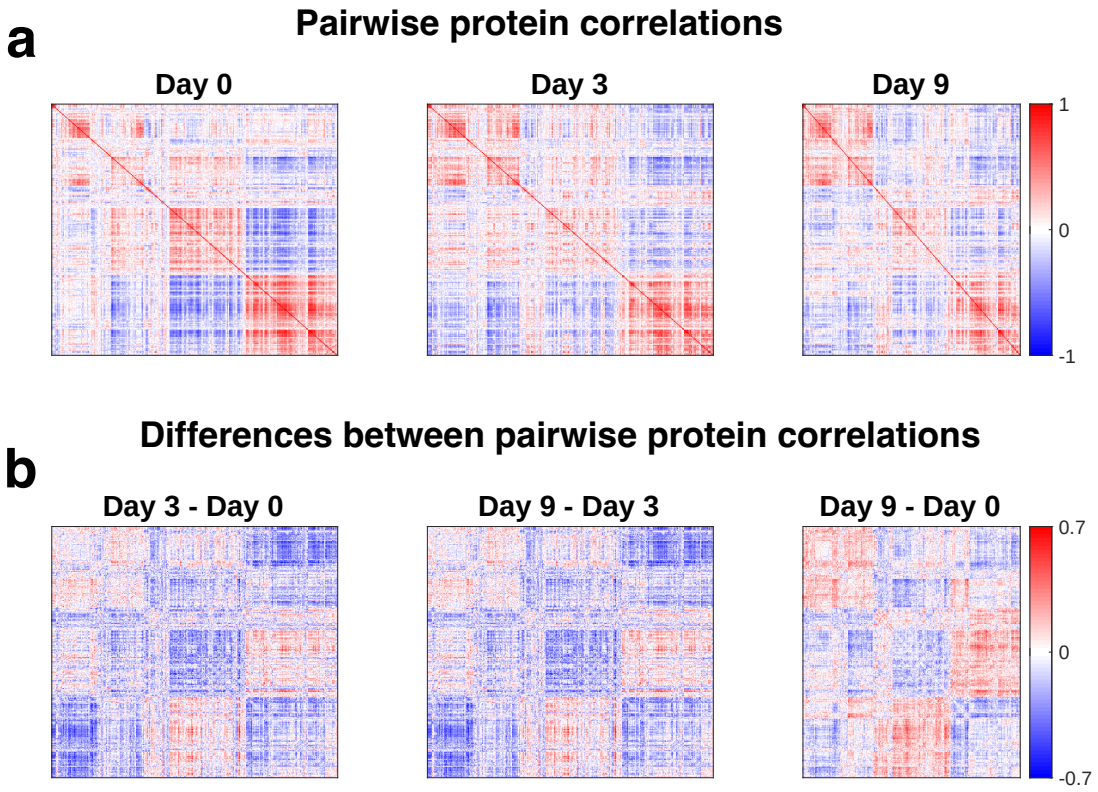


Figure S3 | Dynamics of protein covariation during EMT. The the proteins whose correlations have the largest magnitude (exceeding the 20% percentile) difference between Day 0 and Day 9 were selected and their correlation matrices clustered hierarchically. (**a**) Matrices of pairwise protein correlations at days 0, 3 and 9. (**b**) Matrices of differences between pairwise protein correlations for the indicated time points. The rows and columns for all days correspond to the same proteins ordered in the same way, namely based on clustering the matrix of correlation differences between Day 9 and 0.