A multi-tiered map of EMT defines major transition points and identifies vulnerabilities

Indranil Paul[1], Dante Bolzan[3], Heather Hook[4], Ahmed Youssef[1], Gopal Karemore[5], Michael UJ Oliphant[6], Weiwei Lin[1], Qian Liu[7], Sadhna Phanse[1], Carl White[1], Dzmitry Padhorny[10], Sergei Kotelnikov[10], Guillaume Andrieu[8], Pingzhao Hu[7], Gerald V. Denis[9], Dima Kozakov[10], Brian Raught[12], Trevor Siggers[4], Stefan Wuchty[3,11†], Senthil Muthuswamy[6†], Andrew Emili[1,2†*]

Affiliations

[1] Department of Biochemistry, Boston University School of Medicine, Boston University, 72 East Concord Street, Boston, MA 02118, USA

[2] Department of Biology, Charles River Campus, Boston University, Life Science & Engineering (LSEB-602), 24 Cummington Mall, Boston, MA 02215 USA

[3] Department of Computer Science, University of Miami, 1365 Memorial Drive, Coral Gables, FL 33146, USA

[4] (a) Department of Biology, Boston University, 24 Cummington Mall, Boston, MA 02115, USA; (b) Biological Design Center, Boston University, 610 Commonwealth Avenue, Boston, MA 02215

[5] Advanced Analytics, Novo Nordisk A/S, 2760 Måløv, Denmark

[6] Cancer Research Institute, Department of Medicine, Beth Israel Deaconnes Medical Center, Boston MA 02115, USA

[7] Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba R3E 0J9, Canada

[8] Team 2 - Institut Necker-Enfants Malades, INSERM U1151, Paris, France

[9] Boston Medical Center Cancer Center, Boston University, Boston University, 72 East Concord Street, Boston, MA 02118, USA

[10] (a) Department of Applied Mathematics and Statistics, Stony Brook University, 11794 Stony Brook, NY. (b) Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York, 11794, United States

[11] (a) Department of Biology, University of Miami, 1301 Memorial Drive, Coral Gables, FL 33146, USA; (b) Sylvester Comprehensive Cancer Center, 1475 NW 12th Ave, Miami, FL 33136, USA

[12] Discovery Tower (TMDT), 101 College St, Rm. 9-701A, University of Toronto, Toronto, Ontario, M5G 1L7, Canada

† Co-communicating authors

*Correspondence to: aemili@bu.edu

Summary

TGFβ mediated epithelial to mesenchymal transition (EMT) proceeds through hybrid 'E/M' states. A deeper understanding of these states and events which regulate entry to and exit from the E/M states is needed for therapeutic exploitation. We quantified >60,000 molecules across ten time points and twelve omic layers in mammary epithelial cells. Proteomes of whole cells, phosphoproteins, nucleus, extracellular vesicles, secretome and membrane resolved major shifts, E→E/M and E/M→M during EMT, and defined state-specific signatures. Metabolomics identified early activation of arachidonic acid pathway and an enzyme-mediated switch from Cytochrome P450 to Cyclooxygenase / Lipoxygenase branches during E→E/M. Single-cell transcriptomics identified GLIS2 as an early modulator of EMT. Integrative modeling-predicted combinatorial inhibition of AURKB, PP2A and SRC exposed vulnerabilities at E→E/M juncture. Covariance analysis revealed remarkable discordance between proteins and transcripts, and between proteomic layers, implying insufficiency of current approaches. Overall, this dataset provides an unprecedented resource on TGFβ signaling, EMT and cancer.

Short title

Multi-level time-resolved deconstruction of EMT program

Introduction

Epithelial to mesenchymal transition (EMT) regulates cell plasticity during embryonic development, tissue homeostasis and cancer, where polarized epithelial (E) cells dedifferentiate, transition through intermediate E/M hybrid states (E/M) and acquire mesenchymal (M) properties (Nieto et al., 2016). These E/M share several clinically-relevant attributes of circulating tumor cells (CTCs) and are responsible for EMT-associated stemness, chemoresistance, immune evasion and metastasis (Dongre and Weinberg, 2019). Complete molecular characterization of E/M states, and the mechanisms driving E→E/M and E/M→M transitions will enable development of refined mechanistic models and discovery of new therapeutic targets.

Approximately 150 genes are currently described as hallmarks of EMT (MSigDB) (http://www.gsea-msigdb.org/gsea/msigdb/) that were identified from studies measuring the expression of 'endpoint' markers (e.g. CDH1, MUC1, VIM, FN1) to track the process (Sha et al., 2019). The prevalent view on EMT as a transcriptionally-driven program is debatable (Yang et al., 2020). In fact, the poor correlation between genomic alterations or mRNA levels and (sub)cellular proteoforms during EMT has been noted (Califano and Alvarez, 2017; Liu et al., 2016). This is crucial because EMT is likely an emergent property where the shifts in cell physiology and phenotype are orchestrated by intra- and extra-cellular signaling (Sigston and Williams, 2017), extensive receptor-ligand crosstalk, protein relocalization (Hung and Link, 2011) and metabolic adaptations (Tabassum and Polyak, 2015; Thomson et al., 2019). Attempts made to model EMT primarily invoke gene-regulatory networks (GRNs) controlled by key transcription factors (TFs) and miRNAs. These models are usually based on a restricted number of factors and thus do not capture the multi-layered architecture of signaling during EMT (Hong et al., 2015; Zhang et al., 2014). To date, no EMT-focused studies have simultaneously measured metabolite and gene expression changes at different functional levels (e.g., mRNA, total protein, nucleus, secretome, etc.).

Consequentially, several aspects of EMT remain unclear. This includes dependencies between molecular layers, secreted molecules, and specific signatures at various stages of EMT, kinetics and scope of metabolic reprogramming, dynamics of subcellular protein localizations and ligand-receptor mediated intercellular crosstalk. To bridge these gaps, we employed multiple high-throughput platforms including microarray, scRNAseq and precision mass spectrometry (MS) to quantify molecules spanning 12 distinct layers of biological information. Our ability to integrate this information allowed us to define signatures of E/M states and identify molecular regulators of key transition points.

RESULTS

A comprehensive resource on TGFβ-induced EMT

The human mammary epithelial cell line MCF10A has been widely used to investigate TGFβ (transforming growth factor β) induced EMT (Liu et al., 2019b; Wei et al., 2015). To generate extensive temporally resolved quantitative expression maps of the evolving EMT landscape, cells were treated with TGFβ (TGF-β1; 10 ng/mL) over 12 days (0, 4 hrs, 1–6, 8 & 12 days), spanning multiple 'omics' layers and employing complementary technology platforms (**Fig. 1A-B**, **Fig. S1A-C**). Across all time points, after stringent quality control (see **STAR Methods**), we report nanoLC-MS/MS-based quantifications (**Fig. 1C**, **Table S1**) of 6,540 whole cell (*WC*), 4,198 nuclear (*Nuc*), 2,223 plasma membrane (*Mem*), 1,209 extracellular vesicle (*EV*) and 1,133 secreted (*Sec*) proteins. Using a serial enrichment workflow (see **STAR Methods**) we also quantified the total phosphoproteome (*Phos*; 8,741 high-confidence sites on 2,254 proteins; including 6975 Ser, 962 Thr and 140 Tyr residues), N-glycoproteome (*Glyco*; 549 proteins), acetylome (*Acet*; 349 sites on 165 proteins) and peptidome (*Pep*; 547 peptides from 202 proteins). For proteomics, samples for each point were multiplexed using isobaric tandem mass-tag (TMT-10), enabling higher throughput and robust comparisons. In addition to proteins, we tracked cellular metabolism in the same samples by nanoLC-MS/MS-based untargeted metabolomics, quantifying 8,325 HMDB-indexed endogenous small molecules (*Metabol*). Furthermore, we measured 23,787 gene transcripts (mRNA) and 2,578 microRNAs (miRNA) using microarrays. To assess cellular heterogeneity, we employed scRNAseq to quantify transcriptomes of 1,913 individual cells (>200 cells per time point) undergoing EMT.

In total, we provide temporal quantifications of >60,000 proteins, phosphosites, mRNAs, miRNAs, and metabolites combined, in addition to 9,785 mRNAs in our scRNAseq dataset (**Fig. 1C**). Subcellular enrichments were performed using previously established MS-compatible protocols (see **STAR Methods**), yielding high purity determined through keyword matching against a cellular compartment annotation database (**Fig. S1D**). Quantitative reproducibility across the 3 biological replicates was excellent (**Fig. S1E**). The expression profiles in **Fig. 1D** provide a snapshot of concurrent changes of a given gene over various layers during EMT. We reproduced expected expression behavior for many established markers of EMT, including an increase of M markers VIM, CDH2 and concomitant decrease of E markers SCRIB, MUC1. Using strict criteria for differential expression (Benjamini-Hochberg adj. *p-value* < 0.05; $r^2 \geq 0.6$ & $|\log2FC| \geq 1$; see **STAR Methods**) we identified >10,000 significantly regulated molecules (**Table S2**). As such, this study is the largest experimental description of EMT regulome induced by TGFβ, till date.

We found that molecular abundances were highly variable across layers and time-points (**Fig. S1F**) and all layers contributed significantly to the overall variation in the system (**Fig. 1E**). While each layer showed substantive alterations, the magnitudes (**Fig. S1G**), response profiles (**Fig. 1F**) and fraction of differentially expressed molecules (**Fig. 1G**) of each omic data set during the time course varied significantly. These findings provide direct evidence for global reorganization of the cellular architecture during EMT. For example, *Phos* and *Sec* showed the highest fractional change (36% & 40%, respectively) (**Fig. 1G**), suggesting extensive intra- and extra-cellular signaling during EMT (Scheel et al., 2011). These results add new depths to our current

133 understanding of TGFβ signaling and EMT (**Fig. 1H, I**).

134

135 Integrative multi-tiered topology of EMT reveals key transition points and 136 identifies molecular drivers

137 Although E→M progression is considered a continuum, distinct stages have been reported for
138 MCF10A cells (Zhang et al., 2014), and cancer tissues (Liu et al., 2019a; Pastushenko et al., 2018).
139 Using a phylogenetic clustering approach (Hughes and Friedman, 2009) we estimated 'distances'
140 between time steps to understand the transition kinetics (**Fig. 2A**). We observed that up to 24
141 hours cells maintained their parental E phenotype, while day 2 marked a switch-like exit from
142 the E state, suggesting that at day 2 the E/M began and continued through Day 6, after which cells
143 gradually entered the M state. Cells in day 2 to 6 can be further sub-divided into E/M-1 (Day 2/3,
144 late E) and E/M-2 (Day 4/5/6, early M) states. These cellular reconfigurations were consistent
145 with principal component analysis of individual layers, such as *Mem*, *Phos* and *WC* (**Fig. S2A**).

146 Our observations suggest that each molecular layer has a distinct and complementary role in
147 shaping the transition (**Fig 2B**, **Fig 1E-G**, **Fig S1F, G**). Although, the topology was generally
148 correlated (**Fig. 2C**), pairwise coefficients of determination (adj. R^2) revealed a strikingly low
149 concordance between total proteins and other proteomic layers (mean R^2 ranging from 0.015 for
150 *Acet* to 0.299 for *Glyco*) (**Fig. 2D, Upper panel**). The discordance was even stronger between
151 mRNA and various proteomic layers (mean R^2 ranging from 0.011 for *Pep* to just 0.109 for *WC*)
152 (**Fig. 2D, Lower panel**), which increased further with EMT progression. Collectively, these
153 observations illustrate that at a systems level mRNA quantity is a poor proxy of protein
154 abundance. Total protein or mRNA quantities are also unreliable predictors of a proteins PTMs
155 or subcellular trafficking, which ultimately determines signaling output.

156 Next, to model the regulatory patterns of molecules, we employed a 3-step computational
157 workflow. First, we used a regression strategy for time-course measurements (Conesa et al.,
158 2006) to remove residual noise (time dimension) and non-reliable (replicates) components.
159 Second, an unsupervised machine learning approach (self-organizing maps, SOMs) (Wirth et al.,
160 2012) was applied on the retained molecules to generate time-point specific SOM portraits (**Fig.
161 2E, Fig. S2B-D**, see **STAR Methods** for details). SOMs are a powerful integration tool for large
162 and diverse global datasets to extract fundamental underlying patterns of co-regulation (Lu et al.,
163 2009; Tamayo et al., 1999). Third, we performed pathway enrichments (Ulgen et al., 2019) of
164 these SOMs to elucidate the functional themes at each time-step of EMT (**Fig. 2F**).

165 SOMs traced the temporal unfolding of the transition and provided molecular fingerprints of each
166 time-step. Molecules ranking high (top 1% i.e., rank ≤ 250 & log2FC ≥ 1) with SOMs for control to
167 day 1 (430 unique molecules) primarily included proteins characteristic of an E state (**Fig. 2E,
168 Table S3**), such as, ACTG1 (*Mem*; establishes cell junctions and shape) and EPHB2 (*Glyco*;
169 regulates angiogenesis, contact dependent adhesion and migration; tumor suppressor). SOMs for
170 days 6–12 (252 unique molecules) contained established M markers (e.g., CALD1, CDH2, FLNA,
171 FN1, FSTL1, LGALS1, NT5E, TAGLN, TPM1, TPM2, TPM4, VIM). This group included molecules
172 with established roles in cancer and EMT, including miR-5189 (*miR*; targets ARF6 which

173 internalizes CDH1), CALR (*Acet*; Lys 62; calcium homeostasis, promotes metastasis, imparts
174 resistance to anoikis), ALCAM (*Nuc*; prognostic marker in multiple cancers linked to nuclear
175 translocation of β-catenin and stemness), ARHGAP33 (*Sec*; regulates intracellular trafficking),
176 RAPGEF5 (*Sec*; promotes nuclear translocation of β-catenin) and SOCS3 (*Phos*; Thr 3/13; E3
177 ligase, E3 ligase inhibiting TGFβ signaling). The onset of the major regulatory switch at day 2 is
178 likely driven by molecules peaking at its corresponding SOM (95 molecules) such as miR-675-
179 5p/H19 (*miR*; induces HIF1α, SNAIL activity), LTBP1 (*Sec*; master regulator of integrin-
180 dependent TGFβ activation) and CD44 (*Sec*; signal transduction). Notably, many of the
181 genes/molecules identified in the SOM, to our knowledge, have not been previously linked to
182 TGFβ signaling or EMT, presumably because they do not show clear changes in *mRNA* or *WC* but
183 in 'other' molecular layers, such as *EV* or *Sec* (**Fig. S2E**). This contrasts with most MSigDB
184 hallmarks of EMT for which we captured clear transcriptional profiles (**Fig. S2F**).

185 Enrichment analysis of active subnetworks (Ulgen et al., 2019) identified 237 significant
186 pathways (**Fig. 2F**, **Table S3**), discretized across sequential steps of EMT. For example, 'Beta
187 oxidation of hexanoyl-CoA to butanoyl-CoA' declined as cells leave E and enter E/M, indicating
188 reprogramming of mitochondrial fatty acid β-oxidation, consistent with a metastatic phenotype
189 (Ma et al., 2018). Conversely, 'RHO GTPases mediated activation of ROCKs/PAKs/IQGAPs'
190 increased as cells leave E/M and enter M, suggestive of their key role at this stage of EMT
191 (Ungefroren et al., 2018). The E/M specifically were associated with migration-associated
192 pathways such as 'anchoring fibril formation', 'ECM proteoglycans' and 'laminin interactions',
193 consistent with their shared property with CTCs.

194 Overall, we catalogued complex kinetics of thousands of molecules spanning multiple molecular
195 layers during EMT. Importantly, we predict signatures specific to each stage of EMT, e.g., the E/M,
196 with potential clinical value.

197

198 Metabolomics reveals kinetics and predicts novel enzyme-metabolite associations

199 TGFβ is known to regulate Warburg effect in cancer cells, but may also regulate other metabolic
200 pathways with implications for cancer management (Hua et al., 2019). Active subnetwork
201 analysis of *WC* and *Phos* datasets identified 13 enriched KEGG metabolic pathways during TGFβ-
202 induced EMT (**Fig. S3A**). Notably, significant enrichment (*p–value* ≤ 0.05) was only observed post
203 day 2 for pathways such as Steroid biosynthesis (SHB), Sphingolipid metabolism (SLM) and
204 Glycosaminoglycan degradation (GGM). To evaluate these gene-centric inferences of metabolic
205 phenotypes, we directly profiled intra-cellular small molecules by applying an optimized
206 untargeted metabolomics workflow to the same set of samples. Using stringent criteria (see **STAR**
207 **Methods**), we quantified >8,000 putative HMDB-compounds covering a wide range of chemical
208 classes (**Fig. 3A**).

209 Phylogenetic clustering (**Fig. 3B**) and SOM analysis (**Fig. 3C**) revealed kinetics of metabolic
210 remodeling prompted by TGFβ. In particular, a distinct temporal separation from parental E cells
211 was observed as early as 4 hours indicating rapid modulation of cellular metabolism by TGFβ,
212 preceding changes in most other layers. To glean further insights, we performed integrative

network analysis of metabolite SOMs with differential molecules in *WC* and *Phos* using MetaboAnalyst (Chong et al., 2019) (**Fig. 3D**). This analysis reiterated several enriched pathways predicted with protein expression alone, e.g., glycerophospholipid metabolism (GPLM) and lysine degradation (LD). However, integration of metabolite and protein data within the framework of metabolite SOMs revealed pathway activities representative of key transition steps of EMT driven primarily by corresponding metabolite signatures (**Fig. S3B-D**). Indeed, we observed that arachidonic acid metabolism (AAM), GPLM and LD pathways were activated within 4 hours of TGFβ stimulation, which was not captured by gene enrichment. Consistent with observations in **Fig 2F**, processes such as fatty acid metabolism, SHB and SLM appear after day 2, as cells prepare for a metastatic phenotype (Koundouros and Poulogiannis, 2020).

Pairwise-correlation based integration of metabolite profiles with proteomics measurements from the same samples could enable mechanistic predictions and aid discovery of novel players. To explore this, we chose AAM as an example. Arachidonic acid is an omega-6 fatty acid stored as membrane phosphoglycerolipid. Its cytosolic release enables stoichiometric chain reactions and results in >100 functionally diverse compounds (Hanna and Hafez, 2018) impacting processes such as redox state, proliferation, apoptosis and chemotaxis (Tallima and El Ridi, 2018). We computed Pearson's correlations between abundance of annotated enzymes, in *WC* and *Phos*, and metabolites of AAM pathway, in *Metabol*, quantified at SOM for 4 hours to day 1 (**Fig. 3E, F**). Overall, we found that metabolites mapping to AAM either increased linearly (CYP450 branch and thromboxane) or initially decreased before rising (Lipoxygenase branch and prostaglandin) (**Fig. 3G**), suggesting fine-tuned regulation of the different branches. Encouragingly, our enzyme-metabolite association map identified PLA2G4A, the rate-limiting phospholipase of AAM pathway (Hanna and Hafez, 2018), among the top candidates (**Fig. 3F**). Interestingly, correlation and expression profiles of PLA2G15, another phospholipase, (**Fig. 3E-G**) indicated a potential enzyme-mediated switch from Cytochrome P450 to Cyclooxygenase/Lipoxygenase branches during E→E/M.

Overall, these observations provide direct evidence for the kinetics of metabolic reprogramming during TGFβ induced EMT. We found metabolites and protein signatures coordinating processes such as AAM, GPLM and LD during key stages of the transition. We also demonstrate how our enzyme-metabolite correlation map could be used as a resource to predict novel enzymes for observed metabolic changes.

scRNAseq analysis reveals heterogeneous responses to TGFβ and novel transcriptional regulators of EMT

scRNAseq studies in murine epithelial Py2T cells treated with TGFβ (Krishnaswamy et al., 2018), MCF10A cells undergoing confluence-dependent EMT (McFaline-Figueroa et al., 2019) and LPS-induced EMT in alveolar epithelial cells (Riemondy et al., 2019), provided critical insights into EMT. However, a temporal analysis of EMT to understand the transition states has not been reported.

After quality control, we retained 1,913 single cells with a combined depth of 9,785 genes (**Fig.**

253 **S4A, Table S1**). As anticipated, many of the top expressing genes (TGFB1, TPT1, KRT6A, TMSB10,
254 MT2A) are key players in EMT (**Fig. S4B**). Interestingly, similar to many primary human tumors
255 (Aiello et al., 2018; Puram et al., 2018), we did not observe an explicit loss of several classical E
256 markers at the transcript level (**Fig. 4A, Fig. S4C**), indicating post-transcriptional regulation. This
257 may be energetically economical for tumor cells by conferring invasive properties without losing
258 capacity for metastatic seeding (Lambert et al., 2017).

259 To understand the stages of cell differentiation, we took advantage of the multiple time points in
260 our dataset (as opposed to a 'pseudo-time'), using Monocle3 (see **STAR Methods**) and identified
261 20 cell clusters in 3 disjoint partitions (**Fig. 4B, Fig. S4D**). Partition P2 (12 clusters) represents
262 the primary EMT axis while P1/P3 predominantly expressed genes related to cell cycle (**Fig. S4E**)
263 and were ignored for further analysis. We observed that C3 responded strongly to TGFβ (**Fig. 4C**),
264 C4/6 resisted EMT, C5/C8 are the 'transition' states and C13/14 represented terminal M cells (in
265 terms of hallmark M markers; **Fig. 4C**, **right panel**). Examining clusters C9/18/19 which were
266 composed of cells from nearly all time points, suggests presence of stable M-type cells in MCF10A
267 populations and appears to be transcriptional impasse to TGFβ signaling.

268 To explore the underlying gene expression program, we used hierarchical clustering to group the
269 individual clusters into 6 subtypes (**Fig. 4C**) and employed SCENIC (Aibar et al., 2017) to infer
270 TFs and GRNs underlying these subtypes (**Fig. D-F; Table S5**). For each subtype, we identified
271 several unique (**Fig. 4E**) or highly active TFs (**Fig. 4F**), including both established and novel
272 players. Several TFs implicated in EMT (TWIST2, FOXK2, ZEB1, ID2, MSX1, ING4) were over-
273 represented in S4-S6, corresponding to later stages of EMT. In contrast, direct evidence of
274 mechanistic links of TFs enriched in early stages of EMT (i.e., S3, S4) are lacking, indicating gaps
275 in current EMT models. Using human TF–binding arrays (see **STAR Methods, Fig. 4G**) we
276 confirmed elevated activity of three S3 TFs, GLIS2, SP1 and ZNF266 upon TGFβ induction (**Fig.
277 4H**). In addition to providing experimental evidence to our predictions, the TF-binding array also
278 identified several other TFs potentially playing important roles at the early stages of EMT (**Fig.
279 4I**).

280 Overall, we provide a high-resolution temporal map of gene expression programs of individual
281 cells as they respond to TGFβ signaling and undergo EMT. More generally, our data suggests that
282 most transcriptional changes occur at early time points (also **Fig. 2D, Lower panel**), followed by
283 further adaptations driven predominantly by secondary or post-transcriptional mechanisms.

284

285 Spatial regulation of proteins and inter-cellular communication during EMT

286 Regulation of protein distribution is a crucial signaling mechanism (Ferrell, 1998), but remains
287 insufficiently understood in EMT. The average pairwise Pearson's correlation between proteins
288 quantified in various cellular compartments (CCs) ranged between $r = 0.12$ to 0.58 indicating fine-
289 tuned regulation of protein distributions (**Fig. 5A**). We found 3,965 proteins localizing to ≥2 CCs
290 (**Fig. S5A**), which we categorized into 2 classes as follows: Class I proteins (1,424) displayed a
291 correlated trend ($r \geq 0.4$) consistently across all CC pairs suggesting regulation primarily at or
292 before the translation step (**Fig. 5B, C; Table S5**). Class II proteins (1,205) displayed anti-

293 correlated trends ($r \leq -0.4$) between any two CCs, implying active post-translational control of
294 their asymmetric distributions (**Fig. 5D, E; Table S5**).

295 The Bromodomain and ExtraTerminal (BET) cofactors, BRD2 and BRD4, were identified as Class
296 II proteins, with contrasting kinetic profiles (**Fig. 5F**). Using TF-binding arrays, we verified
297 enhanced global recruitment of BRD4, but not two other cofactors p300 and LSD1, (**Fig. 5G**)
298 indicating its importance during EMT. Indeed, treatment with a pan-BET inhibitor JQ-1
299 suppressed EMT (**Fig. 5H**). Another notable Class II protein was SCRIB (**Fig. 5I**), which regulates
300 apical-basal polarity and directional migration by acting as a molecular scaffold through protein-
301 protein interactions (Bonello and Peifer, 2019). Using an *in* vivo proximity ligation (BioID) screen
302 of SCRIB ((**Fig. S5B**, see **STAR Methods**), we identified multiple novel interactors (**Table S6**) of
303 which many have known roles in EMT (**Fig. 5J**). Using immunoprecipitation, we verified
304 interactions of SCRIB with SNAP23 and ARHGEF7 (**Fig. 5K**).

305 Regulation of extracellular signals is important for EMT (Scheel et al., 2011). Subtypes within a
306 cell population can differ in their capacity to send and receive signals, with implications for
307 metastasis and drug resistance (Kim et al., 2018; Tabassum and Polyak, 2015). To map the inter-
308 cellular communication between subgroups of cells during TGFβ-induced EMT, we integrate
309 proteomics and scRNAseq data to perform a systems-wide survey of ligand-receptor (L-R) pair
310 mediated crosstalk (**Fig. 5L**).

311 First, using a database of >2,500 curated binary L-R interactions (Ramilowski et al., 2015), we
312 searched for pairs of L and R in our *Sec* and *Mem/Glyco* datasets, respectively, assuming that co-
313 directional expression changes in L and/or R of a pair (FDR adj. *p-value* < 0.05 and combined L-R
314 $|\log2FC| \geq 1$) can indicate biological role. Currently, at least two L-R pairs are implicated in TGFβ-
315 induced EMT in several tissue models (Heldin et al., 2012). Our analysis detected 67 upregulated
316 and 12 downregulated L-R pairs at any given time point (**Fig. 5M, Table S7**) following TGFβ
317 treatment. Notably, none of these pairs have been directly implicated in TGFβ signaling or EMT,
318 although individually many of the identified L or R occur frequently in the context of EMT and/or
319 cancer. For instance, LAMC2 with its 7 receptors (CD151, COL17A1, ITGA2, ITGA3, ITGA6, ITGB1,
320 ITGB4) exhibit significant alteration during EMT. LAMC2 is overexpressed in cancers (Garg et al.,
321 2014) and its silencing can reverse EMT (Moon et al., 2015; Pei et al., 2019). The cognate
322 receptors, CD151, ITGA3, ITGA6 and ITGB1 synergize with TGFβ signaling to promote metastatic
323 behavior (Pellinen et al., 2018; Sadej et al., 2010; Shirakihara et al., 2013; Zhang et al., 2017).

324 Next, by systematically comparing the expression patterns of L & R (identified above) among the
325 16 clusters (identified in our scRNAseq dataset), we obtained cell-cell communication networks
326 (sender → receiver) (**Fig. 5N, Fig. S5C, Table S7**). For example, C13 cells, which appeared at day
327 3 and showed highest M genes expression (**Fig. 4C**), produced the receptor CD44 to its cognate
328 ligand MMP7 expressed by C17 (**Fig. S5D**). Together, this suggests that communication between
329 cell subgroups (C17:MMP7 → C13:CD44) may exist during EMT which can have potential
330 ramifications for tumor growth (Tabassum and Polyak, 2015). Encouragingly, several of these
331 identified L-R pairs showed strong correlation (**Fig. 5O**) in human breast invasive carcinoma
332 samples (Cancer Genome Atlas Network, 2012; Cerami et al., 2012).

333 Our study provides a comprehensive analysis of TGFβ-triggered subcellular trafficking as cells

334 undergo EMT. Such translocations, potentially driven by differential PPI, could mediate the tight
335 coordination between functional modules (e.g., SCRIB complex) and EMT phenotypes such as
336 cytoskeletal rearrangement (Huang et al., 2012). We also uncovered novel cell-cell
337 communication pathways via L-R interactions in driving EMT, representing an untapped clinical
338 opportunity.

339

340 Modeling phosphoproteome dynamics during EMT reveals kinase susceptibilities

341 Phosphoregulatory mechanisms are a key aspect of TGFβ signaling and EMT. We confidently
342 quantified 8,741 phosphosites (p-sites; 6,975 Ser, 962 Thr and 140 Tyr residues) (**Fig. S6A-C**)
343 over a dynamic range of 10^6 orders of magnitude (**Fig. S6D**), and phospho–STY frequencies (**Fig.**
344 **S6E**) in line with previous reports (D'Souza et al., 2014) mapping to 2,254 proteins (**Fig. S6F**). Of
345 all p-sites, 3,138 (35.8%) were differentially regulated in at least one time point (**Fig. S6G**). At
346 protein level, different patterns of regulation were noted; some proteins, such as DEK, VIM and
347 MISP, were regulated at ~90% of detected sites, some proteins, such as CAV1, CAMK2 and
348 GOLGA1, showed a ~50% mixture of regulated and unregulated sites, while others such as
349 AHNAK, PML and BCLAF1 showed ~2% differential sites (**Fig. 6A**). Interestingly, the fraction of
350 regulated p-sites in several proteins, e.g., VIM, increased with EMT progression (**Fig. 6B**).

351 We observed that in ~50%, p-sites dynamics were not explained ($r \leq 0.4$) by a corresponding
352 change at the protein level (**Fig. 6C, D**). Interestingly, for ~26%, a directionally opposite change
353 between phosphorylation and the corresponding protein abundance was noted ($r \leq -0.1$),
354 suggesting effects on protein stability. Phospho-regulated proteins were enriched for '*nucleus*',
355 '*cytoskeleton*' and '*focal adhesion*' annotations (**Fig. 6E**) reflecting the importance of CC
356 remodeling during EMT.

357 We also computed correlations between nuclear localization profiles (*Nuc*) and phosphorylation
358 kinetics (*Phos*) of individual proteins (**Fig. 6F, G**). For instance, phosphorylation of MICAL3 at
359 T684 and S685 regulates CSCs by promoting symmetric division (Tominaga et al., 2019). The
360 pattern of phosphorylation at residues T684, S685 and S687 (**Fig. 6D,** MICAL3) which are located
361 at the consensus NLS motif of MICAL3 suggests a role in regulating nuclear translocation and/or
362 retention at E/M→M transition. Indeed, the bipartite NLS motif of MICAL3 interacts with
363 Importin-α and the p-sites T684, S685, S687 are directly adjacent to the binding interface (**Fig.**
364 **6H**).

365 To analyze stage-specific kinase activities we generated a ternary model which distinguishes
366 active kinases into 3 broad stages (i.e., E, E/M, and M) of EMT (**Fig. 6I, Fig. S6H**). An example that
367 illustrates the utility of this model are AKT isoforms that have distinct and opposing roles during
368 cancer development (Hinz and Jücker, 2019). We predict AKT1 is strongly associated with the E-
369 stage, which is consistent with its role in maintaining the E phenotype (Li et al., 2016). In fact,
370 depletion of AKT1 in MCF10A cells promoted TGFβ-induced E→E/M transition (Iliopoulos et al.,
371 2009). Our model further predicts key roles for AKT2 and AKT3 at E/M→M transition. Indeed,
372 AKT2 and AKT3 were associated with tumor invasiveness, stemness and sensitivity to drug
373 treatment (Chin et al., 2014), key characteristics of the E/M populations (Dongre and Weinberg,

374 2019). Among several other kinases (**Fig. S6J**), our ternary model predicts critical roles for PRKCA
375 and AURKB at the junction of E→E/M and E/M→M transitions. PRKCA is reportedly a hub and
376 therapeutic target for EMT-induced breast CSCs (Tam et al., 2013). Similarly, inhibition of AURKB
377 was found to reverse EMT and reduce breast cancer metastasis *in vivo* (Zhang et al., 2020).

378 Overall, we reveal the rich intricacies of the phosphoregulome during EMT, identified functional
379 p-sites, predict novel kinase susceptibilities, and provide a mechanistic framework to enhance
380 understanding of the signaling mechanisms during EMT.

381

382 **Integrative systems causal model of EMT identifies mechanistic vulnerabilities**

383 Systems biology approaches that combine multiple molecular types (proteins, miRNAs,
384 metabolites) into a framework of established knowledge allow for a rich assessment of a
385 biological context (Hawe et al., 2019; Karczewski and Snyder, 2018). Using experimentally
386 validated functional priors (compiled from ENCODE, PhosphoSitePlus, SignaLink 2.0, SIGNOR 2.0,
387 HINT, miRTarBase and MetaBridge), we combined causal inference and PCSF (Prize Collecting
388 Steiner Forest) (Akhmedov et al., 2017) to construct hierarchical mechanistic models of the EMT
389 program (**Fig. 7A**, see **STAR Methods**). The final 'EMT network' comprised of 3,255 edges
390 connecting 2,217 molecules, including 723 kinase/phosphatase–substrate, 1,407 TF–target, 746
391 miRNA–target and 31 metabolite–gene interactions.

392 One of the potential applications of the EMT network is to discover signaling paths from
393 TGFBR1/2 to any gene(s) of interest within the network. As a demonstration, we queried several
394 genes (FN1, MMP7, CD44, SCRIB, TWISTNB, ZEB1, SNAI2) and recovered previously known and
395 unknown paths to them putatively active at multiple stages of EMT (**Fig. 7B**). Next, to identify key
396 factors driving EMT, we extracted 'controller' nodes (Liu et al., 2011; Vinayagam et al., 2016)
397 exerting a significant influence on EMT network topology (see **STAR Methods**). Survival analysis
398 against a large publicly available dataset of primary breast cancers with long-term patient
399 outcomes (Cancer Genome Atlas Network, 2012) showed a significant association between
400 tumors with altered expression of these controllers and shortened overall survival (**Fig. 7C**). Not
401 surprisingly, a few of them are established key regulators of metastasis (**Fig. 7D**), providing
402 independent support to our model. Unlike controllers, however, the non-controller nodes were
403 poorly represented in EMT literature (**Fig. 7E**), again highlighting gaps, and potentially
404 identifying new regulatory processes.

405 We queried the EMT network to identify the signaling context in which these controllers are
406 active at various key stages of EMT, which could also provide clues into mechanistic
407 vulnerabilities (**Fig. 7F**). As cells are stimulated with TGFβ, the TFs SMAD2 and SMAD3 are
408 activated, as expected. Another early responder was RHO GTPase RAC1, an effector of both KRAS
409 (Wu et al., 2014) and TGFβ signaling (Ungefroren et al., 2018), suggesting potential crosstalk. The
410 downstream effector of RAC1, MAPK14 (p38 MAPK), was also regulated early in EMT, suggesting
411 cooperation between RAC1 and MAPK pathways (Santibáñez et al., 2010). Our model suggests
412 SMAD3 regulates two other TF hubs, CEBPB (CCAAT/enhancer-binding protein β) and FOXA1.
413 Loss of CEBPB reportedly switches TGFβ signaling from growth-inhibiting to EMT-inducing

414 (Johansson et al., 2013), while FOXA1 is reportedly a key TF during EMT (Wang et al., 2013). We
415 observed that STAT3 is suppressed at later stages of EMT. A recent study in KRAS-driven lung
416 and pancreatic cancer found that STAT3 is required for maintaining the E state and is lost during
417 acquisition of M phenotypes (D'Amico et al., 2018).

418 The EMT network directly predicts novel avenues for blocking EMT. To assess this, we performed
419 a morphometry-based screening where we treated MCF10A cells with TGFβ in combination with
420 drugs which were predicted to inhibit several of the controller nodes active at E→E/M transition
421 (**Fig. 7G,** see **STAR Methods**). Our analysis using a custom-built and publicly accessible image
422 analysis software (Ochs et al., 2019) (**Fig. 7G**) showed noticeable efficacy of LB100+Barasertib,
423 LB100+PP1 and Sonidegib+Autocamtide in reverting the elongated (= 'eccentric') phenotype of
424 EMT-induced cells (**Fig. 7H**), thus providing direct experimental evidence to our predictions.

425 Overall, our EMT network recapitulates known signaling pathways, uncovers novel routes of
426 information flow to known regulators of EMT, identifies new signaling players and pathways,
427 makes substantial and provocative novel predictions and reveals cohesive time-resolved
428 regulatory patterns and mechanistic links between both controllers and non-controllers.

429

430 DISCUSSION

431 EMT and cancer are *emergent systems* (Sigston and Williams, 2017) wherein progression through
432 various stages is regulated by intricate networks of intra- and extra-cellular signaling within and
433 between cells. The key to understanding such complex biological phenomena are establishing
434 experimental workflows that integrate multiple tiers of biological information (Karczewski and
435 Snyder, 2018).

436 Theoretical discussions on EMT are often guided by the Waddington metaphor of a ball (=cell)
437 rolling over a phenotypic landscape, which is dynamically shaped by multiple parameters:
438 topologies of signaling networks, molecular stochasticity, extraneous cooperating and opposing
439 forces (e.g., EMT inducing and/or inhibitory ligands, interaction with other cells) (Li and Balazsi,
440 2018). By integrating several molecular layers, SOMs and 'neighbor joining' approaches revealed
441 the kinetics of cell-fate transitions and major phenotypic switch points driven by TGFβ. Several
442 studies have indicated molecular and phenotypic granularity in EMT continuum and suggested
443 existence of discrete metastable E/M (Sha et al., 2019). Many current EMT markers are biased
444 toward the later stages of EMT, when the process is approaching completion (Song et al., 2019).
445 Consequently, the molecular nature of E/M is still poorly understood. Our time-course integrative
446 SOM allowed us to trace the temporal unfolding and uncover the molecular nature of the E→M
447 transition.

448 While previous omics studies have measured either one or two molecular layers to describe EMT,
449 our multi-tiered datasets enabled the discovery of several new aspects of EMT which were not
450 captured by previous approaches. For example, the correlation between *mRNA* and proteins were
451 weak. Strikingly, correlations between the various *proteomics layers* were also found to be quite
452 poor, indicating that systems behavior cannot readily be extrapolated by any single layer (e.g.,
453 mRNA or total proteome), but instead needs an integrated analysis of several molecular layers.

454 We show that TGFβ-induced EMT is only partially driven by transcription, where several E genes
455 are repressed only post-transcriptionally, while transcripts of M genes are upregulated but
456 mostly during earlier time points. At later stages, post-translational mechanisms become more
457 prominent in driving the process, suggesting that the regulatory control of EMT may be more
458 flexible than previously appreciated. Similarly, our results also predict the mechanistic
459 importance of protein subcellular localization during EMT. A comparison of proteins detected in
460 *EV*, *Sec*, *Glyco*, *Mem* and *Nuc* indicated extensive regulation of protein localizations. Further,
461 previous studies on EMT have largely focused on cell-autonomous signaling, whereas multiple
462 inter-cellular signaling mechanisms are evident from our integrative analysis. Indeed, we found
463 both *EV* and *Sec* to be extensively regulated during EMT. Once considered as cellular 'garbage
464 bins', their active participation in signaling and crosstalk is increasingly recognized (H. Rashed et
465 al., 2017). Our results demonstrating limited qualitative overlap between *EV* and *Sec* hint at a
466 potentially overlooked mechanistic distinction and provides opportunities for new biological
467 insights. We utilized our *Mem*, *Glyco* and *Sec* datasets to provide a repertoire of 79 putative new
468 L-R pairs with potential roles during EMT, which should be validated in future studies. By
469 combining these results with the scRNAseq profiles, we were able to define an extensive network
470 of inter-cellular communications.

471 In conclusion, we have established a comprehensive multi-tiered molecular landscape of TGFβ-
472 induced EMT. This study aims to provide a valuable resource which is accessible through an
473 interactive website (https://www.bu.edu/dbin/cnsb/emtapp/) (**Fig. S7**) and will strongly
474 complement hypothesis-driven research with direct implications for epithelial cancers.

475

476

477

478 # REFERENCES

479 Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G.,
480 Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: Single-cell regulatory
481 network inference and clustering. Nat. Methods *14*, 1083–1086.

482 Aiello, N.M., Maddipati, R., Norgard, R.J., Balli, D., Li, J., Yuan, S., Yamazoe, T., Black, T., Sahmoud,
483 A., Furth, E.E., et al. (2018). EMT Subtype Influences Epithelial Plasticity and Mode of Cell
484 Migration. Dev. Cell *45*, 681-695.e4.

485 Akhmedov, M., Kedaigle, A., Chong, R.E., Montemanni, R., Bertoni, F., Fraenkel, E., and Kwee, I.
486 (2017). PCSF: An R-package for network-based interpretation of high-throughput data. PLOS
487 Comput. Biol. *13*, e1005694.

488 Bonello, T.T., and Peifer, M. (2019). Scribble: A master scaffold in polarity, adhesion,
489 synaptogenesis, and proliferation. J. Cell Biol. *218*, 742–756.

490 Califano, A., and Alvarez, M.J. (2017). The recurrent architecture of tumour initiation,
491 progression and drug sensitivity. Nat. Rev. Cancer *17*, 116–130.

492 Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast
493 tumours. Nature *490*, 61–70.

494 Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J.,
495 Heuer, M.L., Larsson, E., et al. (2012). The cBio Cancer Genomics Portal: An Open Platform for
496 Exploring Multidimensional Cancer Genomics Data: Figure 1. Cancer Discov. *2*, 401–404.

497 Chin, Y.R., Yoshida, T., Marusyk, A., Beck, A.H., Polyak, K., and Toker, A. (2014). Targeting Akt3
498 Signaling in Triple-Negative Breast Cancer. Cancer Res. *74*, 964–973.

499 Chong, J., Wishart, D.S., and Xia, J. (2019). Using MetaboAnalyst 4.0 for Comprehensive and
500 Integrative Metabolomics Data Analysis. Curr. Protoc. Bioinforma. *68*, e86.

501 Conesa, A., Nueda, M.J., Ferrer, A., and Talón, M. (2006). maSigPro: a method to identify
502 significantly differential expression profiles in time-course microarray experiments.
503 Bioinforma. Oxf. Engl. *22*, 1096–1102.

504 D'Amico, S., Shi, J., Martin, B.L., Crawford, H.C., Petrenko, O., and Reich, N.C. (2018). STAT3 is a
505 master regulator of epithelial identity and KRAS-driven tumorigenesis. Genes Dev. *32*, 1175–
506 1187.

507 Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial–
508 mesenchymal transition and implications for cancer. Nat. Rev. Mol. Cell Biol. *20*, 69–84.

509 D'Souza, R.C.J., Knittle, A.M., Nagaraj, N., Dinther, M. van, Choudhary, C., Dijke, P. ten, Mann, M.,
510 and Sharma, K. (2014). Time-resolved dissection of early phosphoproteome and ensuing
511 proteome changes in response to TGF-β. Sci. Signal. *7*, rs5–rs5.

512 Ferrell, J.E. (1998). How regulated protein translocation can produce switch-like responses.
513 Trends Biochem. Sci. *23*, 461–465.

514 Garg, M., Braunstein, G., and Koeffler, H.P. (2014). LAMC2 as a therapeutic target for cancers.
515 Expert Opin. Ther. Targets *18*, 979–982.

516 H. Rashed, M., Bayraktar, E., K. Helal, G., Abd-Ellah, M.F., Amero, P., Chavez-Reyes, A., and
517 Rodriguez-Aguayo, C. (2017). Exosomes: From Garbage Bins to Promising Therapeutic Targets.
518 Int. J. Mol. Sci. *18*.

519 Hanna, V.S., and Hafez, E.A.A. (2018). Synopsis of arachidonic acid metabolism: A review. J. Adv.
520 Res. *11*, 23–32.

521 Hawe, J.S., Theis, F.J., and Heinig, M. (2019). Inferring Interaction Networks From Multi-Omics
522 Data. Front. Genet. *10*.

523 Heldin, C.-H., Vanlandewijck, M., and Moustakas, A. (2012). Regulation of EMT by TGFβ in
524 cancer. FEBS Lett. *586*, 1959–1970.

525 Hinz, N., and Jücker, M. (2019). Distinct functions of AKT isoforms in breast cancer: a
526 comprehensive review. Cell Commun. Signal. CCS *17*.

527 Hong, T., Watanabe, K., Ta, C.H., Villarreal-Ponce, A., Nie, Q., and Dai, X. (2015). An Ovol2-Zeb1
528 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial
529 and Mesenchymal States. PLOS Comput. Biol. *11*, e1004569.

530 Hua, W., ten Dijke, P., Kostidis, S., Giera, M., and Hornsveld, M. (2019). TGFβ-induced metabolic
531 reprogramming during epithelial-to-mesenchymal transition in cancer. Cell. Mol. Life Sci.

532 Huang, R.Y.-J., Guilford, P., and Thiery, J.P. (2012). Early events in cell adhesion and polarity
533 during epithelial-mesenchymal transition. J. Cell Sci. *125*, 4417–4422.

534 Hughes, A.L., and Friedman, R. (2009). A phylogenetic approach to gene expression data:
535 evidence for the evolutionary origin of mammalian leukocyte phenotypes. Evol. Dev. *11*, 382–
536 390.

537 Hung, M.-C., and Link, W. (2011). Protein localization in disease and therapy. J. Cell Sci. *124*,
538 3381–3392.

539 Iliopoulos, D., Polytarchou, C., Hatziapostolou, M., Kottakis, F., Maroulakou, I.G., Struhl, K., and
540 Tsichlis, P.N. (2009). MicroRNAs differentially regulated by Akt isoforms control EMT and stem
541 cell renewal in cancer cells. Sci. Signal. *2*, ra62.

542 Johansson, J., Berg, T., Kurzejamska, E., Pang, M.-F., Tabor, V., Jansson, M., Roswall, P., Pietras, K.,
543 Sund, M., Religa, P., et al. (2013). MiR-155-mediated loss of C/EBPβ shifts the TGF-β response
544 from growth inhibition to epithelial-mesenchymal transition, invasion and metastasis in breast
545 cancer. Oncogene *32*, 5614–5624.

546 Karczewski, K.J., and Snyder, M.P. (2018). Integrative omics for health and disease. Nat. Rev.
547 Genet. *19*, 299–310.

548 Kim, E., Kim, J.-Y., Smith, M.A., Haura, E.B., and Anderson, A.R.A. (2018). Cell signaling
549 heterogeneity is modulated by both cell-intrinsic and -extrinsic mechanisms: An integrated
550 approach to understanding targeted therapy. PLoS Biol. *16*.

551 Koundouros, N., and Poulogiannis, G. (2020). Reprogramming of fatty acid metabolism in
552 cancer. Br. J. Cancer *122*, 4–22.

553 Krishnaswamy, S., Zivanovic, N., Sharma, R., Pe'er, D., and Bodenmiller, B. (2018). Learning time-
554 varying information flow from single-cell epithelial to mesenchymal transition data. PLoS ONE
555 *13*.

556 Lambert, A.W., Pattabiraman, D.R., and Weinberg, R.A. (2017). Emerging Biological Principles of
557 Metastasis. Cell *168*, 670–691.

558 Li, C., and Balazsi, G. (2018). A landscape view on the interplay between EMT and cancer
559 metastasis. Npj Syst. Biol. Appl. *4*, 1–9.

560 Li, C.-W., Xia, W., Lim, S.-O., Hsu, J.L., Huo, L., Wu, Y., Li, L.-Y., Lai, C.-C., Chang, S.-S., Hsu, Y.-H., et
561 al. (2016). AKT1 Inhibits Epithelial-to-Mesenchymal Transition in Breast Cancer through
562 Phosphorylation-Dependent Twist1 Degradation. Cancer Res. *76*, 1451–1462.

563 Liu, X., Li, J., Cadilha, B.L., Markota, A., Voigt, C., Huang, Z., Lin, P.P., Wang, D.D., Dai, J., Kranz, G., et
564 al. (2019a). Epithelial-type systemic breast carcinoma cells with a restricted mesenchymal
565 transition are a major source of metastasis. Sci. Adv. *5*, eaav4275.

566 Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on
567 mRNA Abundance. Cell *165*, 535–550.

568 Liu, Y., Xue, M., Du, S., Feng, W., Zhang, K., Zhang, L., Liu, H., Jia, G., Wu, L., Hu, X., et al. (2019b).
569 Competitive endogenous RNA is an intrinsic component of EMT regulatory circuits and
570 modulates EMT. Nat. Commun. *10*, 1637.

571 Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of complex networks. Nature
572 *473*, 167–173.

573 Lu, R., Markowetz, F., Unwin, R.D., Leek, J.T., Airoldi, E.M., MacArthur, B.D., Lachmann, A., Rozov,
574 R., Ma'ayan, A., Boyer, L.A., et al. (2009). Systems-level dynamic analyses of fate change in
575 murine embryonic stem cells. Nature *462*, 358–362.

576 Ma, Y., Temkin, S.M., Hawkridge, A.M., Guo, C., Wang, W., Wang, X.-Y., and Fang, X. (2018). Fatty
577 acid oxidation: An emerging facet of metabolic transformation in cancer. Cancer Lett. *435*, 92–
578 100.

579 McFaline-Figueroa, J.L., Hill, A.J., Qiu, X., Jackson, D., Shendure, J., and Trapnell, C. (2019). A
580 pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the
581 epithelial-to-mesenchymal transition. Nat. Genet. *51*, 1389–1398.

582 Moon, Y.W., Rao, G., Kim, J.J., Shim, H.-S., Park, K.-S., An, S.S., Kim, B., Steeg, P.S., Sarfaraz, S.,
583 Changwoo Lee, L., et al. (2015). LAMC2 enhances the metastatic potential of lung
584 adenocarcinoma. Cell Death Differ. *22*, 1341–1352.

585 Nieto, M.A., Huang, R.Y.-J., Jackson, R.A., and Thiery, J.P. (2016). EMT: 2016. Cell *166*, 21–45.

586 Ochs, F., Karemore, G., Miron, E., Brown, J., Sedlackova, H., Rask, M.-B., Lampe, M., Buckle, V.,
587 Schermelleh, L., Lukas, J., et al. (2019). Stabilization of chromatin topology safeguards genome
588 integrity. Nature *574*, 571–574.

589 Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van
590 Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour
591 transition states occurring during EMT. Nature *556*, 463–468.

592 Pei, Y.-F., Liu, J., Cheng, J., Wu, W.-D., and Liu, X.-Q. (2019). Silencing of LAMC2 Reverses
593 Epithelial-Mesenchymal Transition and Inhibits Angiogenesis in Cholangiocarcinoma via
594 Inactivation of the Epidermal Growth Factor Receptor Signaling Pathway. Am. J. Pathol. *189*,
595 1637–1653.

596 Pellinen, T., Blom, S., Sánchez, S., Välimäki, K., Mpindi, J.-P., Azegrouz, H., Strippoli, R., Nieto, R.,
597 Vitón, M., Palacios, I., et al. (2018). ITGB1-dependent upregulation of Caveolin-1 switches TGFβ
598 signalling from tumour-suppressive to oncogenic in prostate cancer. Sci. Rep. *8*, 1–14.

599 Puram, S.V., Parikh, A.S., and Tirosh, I. (2018). Single cell RNA-seq highlights a role for a partial
600 EMT in head and neck cancer. Mol. Cell. Oncol. *5*, e1448244.

601 Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M.,
602 Kawaji, H., Carninci, P., Rost, B., et al. (2015). A draft network of ligand–receptor-mediated
603 multicellular signalling in human. Nat. Commun. *6*, 1–12.

604 Riemondy, K.A., Jansing, N.L., Jiang, P., Redente, E.F., Gillen, A.E., Fu, R., Miller, A.J., Spence, J.R.,
605 Gerber, A.N., Hesselberth, J.R., et al. (2019). Single-cell RNA sequencing identifies TGF-β as a key
606 regenerative cue following LPS-induced lung injury. JCI Insight *4*, e123637.

607 Sadej, R., Romanska, H., Kavanagh, D., Baldwin, G., Takahashi, T., Kalia, N., and Berditchevski, F.
608 (2010). Tetraspanin CD151 Regulates Transforming Growth Factor β Signaling: Implication in
609 Tumor Metastasis. Cancer Res. *70*, 6059–6070.

610 Santibáñez, J.F., Kocić, J., Fabra, A., Cano, A., and Quintanilla, M. (2010). Rac1 modulates TGF-β1-
611 mediated epithelial cell plasticity and MMP9 production in transformed keratinocytes. FEBS
612 Lett. *584*, 2305–2310.

613 Scheel, C., Eaton, E.N., Li, S.H.-J., Chaffer, C.L., Reinhardt, F., Kah, K.-J., Bell, G., Guo, W., Rubin, J.,
614 Richardson, A.L., et al. (2011). Paracrine and Autocrine Signals Induce and Maintain
615 Mesenchymal and Stem Cell States in the Breast. Cell *145*, 926–940.

616 Sha, Y., Haensel, D., Gutierrez, G., Du, H., Dai, X., and Nie, Q. (2019). Intermediate cell states in
617 epithelial-to-mesenchymal transition. Phys. Biol. *16*, 021001.

618 Shirakihara, T., Kawasaki, T., Fukagawa, A., Semba, K., Sakai, R., Miyazono, K., Miyazawa, K., and
619 Saitoh, M. (2013). Identification of integrin α3 as a molecular marker of cells undergoing
620 epithelial-mesenchymal transition and of cancer cells with aggressive phenotypes. Cancer Sci.
621 *104*, 1189–1197.

622 Sigston, E.A.W., and Williams, B.R.G. (2017). An Emergence Framework of Carcinogenesis. Front.
623 Oncol. *7*.

624 Song, J., Wang, W., Wang, Y., Qin, Y., Wang, Y., Zhou, J., Wang, X., Zhang, Y., and Wang, Q. (2019).
625 Epithelial-mesenchymal transition markers screened in a cell-based model and validated in lung
626 adenocarcinoma. BMC Cancer *19*, 680.

627 Tabassum, D.P., and Polyak, K. (2015). Tumorigenesis: it takes a village. Nat. Rev. Cancer *15*,
628 473–483.

629 Tallima, H., and El Ridi, R. (2018). Arachidonic acid: Physiological roles and potential health
630 benefits – A review. J. Adv. Res. *11*, 33–41.

631 Tam, W.L., Lu, H., Buikhuisen, J., Soh, B.S., Lim, E., Reinhardt, F., Wu, Z.J., Krall, J.A., Bierie, B., Guo,
632 W., et al. (2013). Protein Kinase C α Is a Central Signaling Node and Therapeutic Target for
633 Breast Cancer Stem Cells. Cancer Cell *24*, 347–364.

634 Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub,
635 T.R. (1999). Interpreting patterns of gene expression with self-organizing maps: Methods and
636 application to hematopoietic differentiation. Proc. Natl. Acad. Sci. *96*, 2907–2912.

637 Thomson, T.M., Balcells, C., and Cascante, M. (2019). Metabolic Plasticity and Epithelial-
638 Mesenchymal Transition. J. Clin. Med. *8*, 967.

639 Tominaga, K., Minato, H., Murayama, T., Sasahara, A., Nishimura, T., Kiyokawa, E., Kanauchi, H.,
640 Shimizu, S., Sato, A., Nishioka, K., et al. (2019). Semaphorin signaling via MICAL3 induces
641 symmetric cell division to expand breast cancer stem-like cells. Proc. Natl. Acad. Sci. *116*, 625–
642 630.

643 Ulgen, E., Ozisik, O., and Sezerman, O.U. (2019). pathfindR: An R Package for Comprehensive
644 Identification of Enriched Pathways in Omics Data Through Active Subnetworks. Front. Genet.
645 *10*.

646 Ungefroren, H., Witte, D., and Lehnert, H. (2018). The role of small GTPases of the Rho/Rac
647 family in TGF-β-induced EMT and cell motility in cancer. Dev. Dyn. *247*, 451–461.

648 Vinayagam, A., Gibson, T.E., Lee, H.-J., Yilmazel, B., Roesel, C., Hu, Y., Kwon, Y., Sharma, A., Liu, Y.-
649 Y., Perrimon, N., et al. (2016). Controllability analysis of the directed human protein interaction
650 network identifies disease genes and drug targets. Proc. Natl. Acad. Sci. U. S. A. *113*, 4976–4981.

651 Wang, H., Meyer, C.A., Fei, T., Wang, G., Zhang, F., and Liu, X.S. (2013). A systematic approach
652 identifies FOXA1 as a key factor in the loss of epithelial traits during the epithelial-to-
653 mesenchymal transition in lung cancer. BMC Genomics *14*, 680.

654 Wei, S.C., Fattet, L., Tsai, J.H., Guo, Y., Pai, V.H., Majeski, H.E., Chen, A.C., Sah, R.L., Taylor, S.S.,
655 Engler, A.J., et al. (2015). Matrix stiffness drives Epithelial-Mesenchymal Transition and tumour
656 metastasis through a TWIST1-G3BP2 mechanotransduction pathway. Nat. Cell Biol. *17*, 678–
657 688.

658 Wirth, H., von Bergen, M., and Binder, H. (2012). Mining SOM expression portraits: feature
659 selection and integrating concepts of molecular function. BioData Min. *5*, 18.

660 Wu, C.-Y.C., Carpenter, E.S., Takeuchi, K.K., Halbrook, C.J., Peverley, L.V., Bien, H., Hall, J.C.,
661 DelGiorno, K.E., Pal, D., Song, Y., et al. (2014). PI3K regulation of RAC1 is required for KRAS-
662 induced pancreatic tumorigenesis in mice. Gastroenterology *147*, 1405-1416.e7.

663 Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A.,
664 Casanova, J., Christofori, G., et al. (2020). Guidelines and definitions for research on epithelial–
665 mesenchymal transition. Nat. Rev. Mol. Cell Biol. 1–12.

666 Zhang, J., Tian, X.-J., Zhang, H., Teng, Y., Li, R., Bai, F., Elankumaran, S., and Xing, J. (2014). TGF-β–
667 induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple
668 feedback loops. Sci Signal *7*, ra91–ra91.

669 Zhang, J., Lin, X., Wu, L., Huang, J.-J., Jiang, W.-Q., Kipps, T.J., and Zhang, S. (2020). Aurora B
670 induces epithelial–mesenchymal transition by stabilizing Snail1 to promote basal-like breast
671 cancer metastasis. Oncogene *39*, 2550–2567.

672 Zhang, K., Myllymäki, S.-M., Gao, P., Devarajan, R., Kytölä, V., Nykter, M., Wei, G.-H., and
673 Manninen, A. (2017). Oncogenic K-Ras upregulates ITGA6 expression via FOSL1 to induce
674 anoikis resistance and synergizes with αV-Class integrins to promote EMT. Oncogene *36*, 5681–
675 5694.

676

677

Author contributions

Indranil Paul: Conceptualization, Methodology, Software, Data curation, Writing – Original draft, Visualization, Investigation, Formal analysis
Dante Bolzan: Software, Formal analysis
Heather Joan Hook: Investigation
Ahmed Youssef: Software
Gopal Karemore: Software
Michael Uretz John Oliphant: Investigation
Weiwei Lin: Investigation
Qian Liu: Formal analysis
Sadhna Phanse: Software
Dzmitry Padhorny: Software
Sergei Kotelnikov: Software
Carl White: Software
Andrieu Guillaume: Investigation
Pingzhao Hu: Formal analysis
Gerald Denis: Investigation
Dima Kozakov: Software
Brian Raught: Investigation
Trevor Siggers: Investigation
Stefan Wuchty: Software, Formal analysis
Senthil Muthuswamy: Conceptualization, Investigation
Andrew Emili: Conceptualization, Supervision, Writing – Original draft, Resources, Funding acquisition, Formal analysis

Acknowledgements

We dedicate our sincere gratitude to Stefano Monti, Xarelabos Varelas, Valentina Perissi and Matthew Layne for providing critical feedback on the manuscript. We thank the Microarray and Sequencing Core Facility at Boston University School of Medicine for acquisition of mRNA, miRNA and scRNAseq data.

Funding

S.M., G.D., and A.E. acknowledge joint funding (UO1CA243004) from the NCI program Research Projects in Cancer Systems Biology. The CNSB has generous ongoing support from Boston University.

Competing interests

The authors declare no competing financial interest.

Main figure legends

Figure 1. A multi-dimensional resource on TGFβ-induced EMT

(A) MCF10A cells were exposed to TGFβ for the indicated time points. A major premise of the study was to characterize various cell-states and the transition points through which EMT occurs.
(B) Samples from 3 biological replicates were aliquoted and multiple technology platforms were employed to quantify various molecular layers.
(C) An overview of numbers of molecules quantified in various layers.
(D) Expression snapshots of some well-known EMT markers. Heatmaps show log2FC values (relative to Control, FDR = 0.05).
(E) Variance explained by each layer over 4 principal components.
(F) Ridge plot showing differential molecules, as % of total quantified, for each layer.
(G) Pie chart showing the overall fraction (in %) of differential molecules (yellow portion) relative to all molecules quantified in each layer.
(H) Overlap between established EMT databases (MSigDB, www.gsea-msigdb.org; dbEMT2.0, http://dbemt.bioinfo-minzhao.org) and differential proteins and miRNAs (adj. *p-value* < 0.05; $r^2 \geq 0.6$ & $|\log2FC| \geq 1$) from this study.
(I) Differential molecules (proteins, miRNAs, metabolites) from this study were used to assess the number of coherent functional modules, i.e., biological interactions or network edges, by employing the Prize-Collecting Steiner Forest algorithm on an interaction network compiled from PathwayCommons, miRTarBase and STITCH.

Figure 2. The topological architecture of EMT

(A) Neighbor-joining was applied as a distance metric to reveal the similarities between timepoints, considering all bulk measurements, excluding *mRNA* (see text).
(B) Combined pseudo-eigenvalues space of all datasets, indicating the contribution of each dataset to the eigenvalue (variance).
(C) Matrix correlations between each pair of datasets.
(D) Line plots show the distribution of adjusted coefficient of determination (R^2) values between layers as a function of time points.
(E) *Left panel.* SOM-portraits characterizing expression landscape of the samples. Color gradient refers to over- or under-expression of metagenes (=cluster of genes) in each time-point compared to the mean expression level of the metagene in the pool of all time-points: red = high, yellow/green = intermediate levels and blue = low (see **STAR Methods** for details). *Middle panel*. Representative examples that appeared among the highest-ranking features (top 1%) in each time-point. *Right panel*. Barplot shows the number of features that were contributed by each layer among the highest-ranking features (top 1%) in time-point.
(F) Heatmap depicts sample-wise pathway scores, which were derived from subnetwork-based pathway enrichment analysis using the highest-ranking molecules (top 1%) for each SOM.

756 **Figure 3. Dynamics of TGFβ-induced metabolic adaptations**

757 (A) Barplot showing the 'class' distribution of quantified metabolite features as defined in
758 HMBD. Also shown are the relative proportions of differentially expressed features for
759 each class (*p-value* ≤ 0.01, absolute log2FC ≥ 1).
760 (B) Neighbor-joining tree for the *Metabol* dataset.
761 (C) SOM analysis of the *Metabol* dataset.
762 (D) Top 10% of metabolite features of each SOM were grouped based on clusters in `B` and
763 used for "Network analysis" using MetaboAnalyst (https://www.metaboanalyst.ca).
764 (E) Identified metabolites of AAM pathway in SOM for 4 hours to day 1 were taken and
765 Pearson's correlation computed with known metabolic enzymes (KEGG) quantified in the
766 *Phos* & *Prot* datasets.
767 (F) The plot shows enzymes ranked according to their Pearson's correlation with metabolites
768 of AAM pathway, as detected in this study.
769 (G) Schematics of information flow from TGFβ signaling to AAM pathway, mediated by known
770 enzymes. Heatmaps and line plots display 'standardized' expression values over the time
771 points.

772 **Figure 4. scRNAseq analysis reveals cellular dynamics and novel TFs for EMT**

773 (A) Violin plots showing expression of well-known EMT hallmarks in each time point.
774 (B) UMAP of scRNAseq by Monocle3. Dots represent single cells and are colored by inferred
775 clusters, while trajectories depict cells during EMT.
776 (C) Heatmap showing the number of cells (in log2 scale) in each cluster of partition P2. Mean
777 expression of some well-known E and M markers in each cluster are also shown.
778 (D) The plot shows all TFs ranked according to their SCENIC score. TF names are shown for
779 the 5 TFs with the highest scores and some well-known EMT associated TFs.
780 (E) Tree displays unique TFs identified by SCENIC for each subtype. TFs highlighted in 'bold'
781 are known players in EMT as of dbEMT2.0. The genes in the outer circle are
782 representative examples used by SCENIC to infer TF activity.
783 (F) Barplot shows the inferred activities of top 15 TFs based on SCENIC scores in each
784 subtype.
785 (G) Schematics of the human TF-binding array workflow (see **STAR Methods** for details).
786 (H) & (I) Density plots of ∂B-scores of indicated TFs. Two-sided Kolmogorov-Smirnov test
787 was performed to evaluate the significance of distribution differences between Control
788 and TGFβ treated conditions.

789 **Figure 5. Spatial regulation of proteins and intercellular communication**

790 (A) The schematic summarizes correlations based on Pearson's coefficients between the
791 indicated layers. Each pie chart depicts the fraction of differential proteins (orange slice)
792 with respect to all proteins quantified in the layer. The dotted lines indicate correlation
793 of overlapping proteins between layers.
794 (B) The plot displays top 25 Class I proteins. Each pair is represented by a different shape, as
795 shown.
796 (C) Expression profiles of top 3 Class I proteins. Each colored line represents a molecular
797 layer, as shown.
798 (D) The plot displays top 25 Class II proteins. Each pair is represented by a different shape,
799 as shown.

800 (E) Expression profiles of top 3 Class II proteins. Each colored line represents a molecular
801 layer, as shown.
802 (F) Expression profiles of BRD2 and BRD4 in various layers, as indicated. Legend as in E
803 (G) Density plots of ∂B-scores of indicated co-factors. Two-sided Kolmogorov-Smirnov test
804 was performed to evaluate the significance of difference in distributions.
805 (H) Phase-contrast images after 6 days of TGFβ treatment in presence or absence of active JQ-
806 1 (100 nM) or its inactive analogue.
807 (I) Expression profile of SCRIB in various layers, as indicated. Legend as in E
808 (J) PPI network of SCRIB interactors identified using BioID.
809 (K) Immunoblots showing interactions of a few SCRIB partners identified using BioID.
810 (L) Schematics of analysis pipeline for discovering active L-R pairs (see **STAR Methods**).
811 (M) Heatmap showing combined log2FCs of L-R pairs in the *Sec* and *Mem/Glyco* datasets.
812 (N) Network plot showing L-R interactions detected between different P2 cell clusters.
813 (O) Scatter plot of correlations between indicated gene-pairs in Breast invasive carcinoma
814 samples. Regression line is shown in red.

815 **Figure 6. Phosphoproteome dynamics during EMT**

816 (A) Fraction of detected p-sites on a protein that are regulated during EMT.
817 (B) Detected p-sites on VIM and their expression during EMT. The gray lines indicate all p-
818 sites that are catalogued in PhosphoSitePlus database.
819 (C) Distribution of Pearson's correlation between expression of proteins and the p-sites
820 detected on them.
821 (D) Schematic showing two examples each where expression of proteins and p-sites showed
822 either low (≤–0.1; CDS2, CBX1) or high (≥0.4; MISP, MICAL3) correlation.
823 (E) Gene ontology enrichment of genes with at least a single regulated p-site at any time point.
824 (F) Distribution of Pearson's correlation between expression of proteins detected in *Nuc*
825 layer and the p-sites detected on them in the *Phos* layer. A few EMT hallmarks are
826 highlighted.
827 (G) Heatmaps of expression profiles of indicated molecules in *Nuc* and *Phos* layers.
828 (H) Structural model of MICAL3 p-sites, NLS and Importin-α.
829 (I) Ternary plot of kinase activity scores binned into 3 broad stages of EMT, i.e., E, ICS, and
830 M.
831 (J) *Top*. Expression of PRKCA as detected in various layers. *Bottom*. Pathway enrichment of
832 all differential substrates of PRKCA detected in our dataset.
833 (K) *Top*. Expression of AURKB as detected in various layers. *Bottom*. Pathway enrichment of
834 all differential substrates of AURKB detected in our dataset.

835 **Figure 7. Integrative systems causal model of EMT**

836 (A) Schematics of network and causal inference workflow. Non-redundant genes with most
837 significant expression (Hotelling's T^2 statistic) profiles were used for the analysis.
838 CausalPath-estimated logical networks were used to augment a custom-built confidence-
839 weighted scaffold interactome which was then used to solve the Steiner forest problem
840 using the OmicsIntegrator software. Only differential molecules (relative to Control, FDR
841 adj.*p-value* ≤ 0.05, |log2FC| ≥ 1) were considered as 'prizes'.
842 (B) A subgraph showing the signaling context of well-known EMT players 'queried' on the
843 EMT network.

(C) Kaplan-Meier plots comparing prognostic performance of MSigDB hallmarks and 'controllers' identified in this study, as indicated.
(D) Overlap between established EMT databases (as in Figure 1) and the 'controllers' identified in this study.
(E) Overlap between the EMT databases and the 'non-controllers' identified in this study.
(F) A simplified schematic showing the hierarchical relationships between several bottleneck genes at different stages of EMT, as indicated.
(G) Workflow for morphometric screening of drug combinations. The tool GENIMASEG is available here: https://figshare.com/articles/software/GenImaSeg_Tool_zip/14371778
(H) The barplot of results of the above experiment indicates varying degrees of synergy or antagonism between inhibitors, in influencing EMT-associated changes in cell shape (eccentricity). Error bars indicate spread of datapoints across all quantified cells in each condition.

Supplementary figure legends

Figure S1. Related to Figure 1. A multi-dimensional resource on TGFβ-induced EMT

(A) TGFβ treatments were staggered, at defined time periods, such that all plates were harvested at the same time. Cells were serum starved for 16 hours before harvesting.
(B) Phase-contrast images of MCF10A cells at different time points. Scale bar = 100 μm.
(C) For maximizing efficiency while maintaining compatibility with technology-specific protocols, after harvesting cells were collected in 4 aliquots per replicate, as shown. Conditioned media were also collected.
(D) Upper panel – Barplots showing percentage of quantified proteins with annotation in the 'Cellular Component' category. Lower panel – Boxplots showing intensity values (log2) of proteins (= markers) commonly used to assess sub-cellular fractionation purity. Box edges correspond to 25th and 75th percentiles, whiskers include extreme data points.
(E) Cumulative distribution (%) of Pearson's coefficients across the samples (= time points). Significant overlap between the 3 biological replicates, shown as 3 colors, indicates high reproducibility.
(F) Heatmaps of molecular expression profiles of indicated layers. The mean of all quantified molecules across the 3 replicates was used.
(G) Violin plots showing the spread of log2FCs of molecules at each time point (relative to Control) for each layer.

Figure S2. Related to Figure 2. The topological architecture of EMT

(A) PCA of the various molecular layers. Time points are shown with different shapes and colors. Points with similar shape/color indicate biological replicates.
(B) The population map presents the number of genes mapped to each individual metagene.
(C) The plot summarizes co-variance structure of datasets at the metagene level.
(D) The plot shows correlation of expression patterns of individual genes and the metagenes in which they are contained.
(E) Temporal expression profiles of a few example genes identified in SOM analysis. Each colored line represents a molecular layer, as shown.
(F) Temporal expression profiles of a few example genes enlisted as 'EMT hallmarks' in MSigDB database. Each colored line represents a molecular layer, as shown.

887

888 **Figure S3. Related to Figure 3. Dynamics of TGFβ-induced metabolic adaptations**

889 (A) The heatmap displays the results of active subnetwork analysis using differential
890 molecules in *Prot* and *Phos* datasets.
891 (B) – (D) Metabolite-metabolite interaction networks, using MetaboAnalyst, for indicated
892 clusters. High connectivity indicates co-regulation of functionally related metabolites.

893 **Figure S4. Related to Figure 4. scRNAseq analysis reveals cellular dynamics and novel TFs for**
894 **EMT**

895 (A) An outline of the QC pipeline employed for scRNAseq data analysis.
896 (B) The plot shows top 25 most expressed genes. Each row corresponds to a gene and each
897 bar corresponds to the expression of the gene in single cells.
898 (C) Expression of indicated genes in *mRNA* layer.
899 (D) Developmental trajectories of MCF10A cells in response to TGFβ, inferred by Monocle3.
900 Clusters are indicated by colors.
901 (E) Heatmap showing aggregate expression of groups of genes (=Modules) with similar
902 expression pattern across the *partitions*, by Monocle3. Modules 9/12 were highly
903 expressed in P1/P3 and were enriched for 'Cell cycle' related GO annotations.

904 **Figure S5. Related to Figure 5. Spatial regulation of proteins and intercellular communication**

905 (A) The plot shows number of common genes (intersection size, *y* axis) between layers as
906 indicated. Only differential genes in each layer ('Set size', FDR ≤ 0.05; adj. *p-value* ≤ 0.05;
907 $r^2 \geq 0.6$ & |log2FC|, relative to Control, of ≥ 1) were considered for the analysis.
908 (B) Schematics of the BioID experiment performed to discover novel SCRIB partners induced
909 by TGFβ signaling.
910 (C) Circos plot showing L-R interactions between the P2 clusters.
911 (D) UMAP plots of scRNAseq data highlighting the co-expression patterns of MMP7 and CD44.

912 **Figure S6. Related to Figure 6. Phosphoproteome dynamics during EMT**

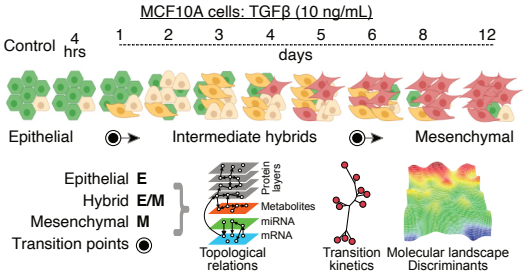
913 (A) An outline of QC pipeline for *Phos* data analysis.
914 (B) Enrichment efficiency for TiO_2 workflow employed for the study.
915 (C) About 74% of all detected p-sites were reliably localized (=Class I) by MaxQuant.
916 (D) Quantifications of p-sites were achieved with a dynamic range of 10^6 orders of magnitude.
917 (E) Proportions of number of phosphate moieties in each detected p-site.
918 (F) Number of phosphoproteins regulated over the time course.
919 (G) Magnitudes of log2FC values for p-sites over the time course.
920 (H) The plot shows all kinases ranked according to their KSEA enrichment. Mean of |z-score|
921 values over all time points were taken.

922 **Figure S7. Related to Figure 7. Integrative systems causal model of EMT**

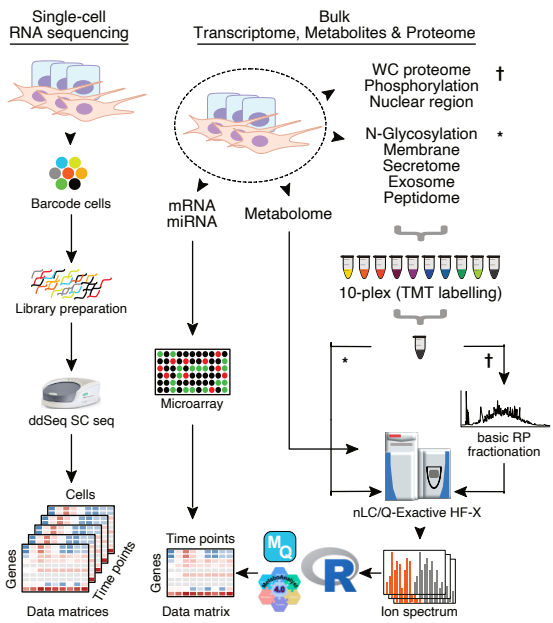
923 (A) A snapshot of the companion website which can be interactively and freely accessed at
924 (https://srphanse.shinyapps.io/EMTAPP_2/).
925

Figure 1

A Experimental design and premise



B Data acquisition, technologies and platforms



C Datasets (this study, 10 time points)

Proteomics

Component	Peptides	Genes (signif. up/dn) **	PTM sites * (signif. up/dn) **
Whole cell proteome (WC)	552,716	6,540 (403/470)	--
Nuclear region (Nuc)	286,162	4,198 (298/259)	--
Membrane (Mem)	35,870	2,223 (149/131)	--
Peptidome (Pep)	735	202 (31/31)	--
Secretome (Sec)	27,136	1,133 (247/200)	--
Extracellular vesicles (EV)	20,438	1,209 (149/123)	--
Glycosylated proteins (Glyco)	6696	590 (17/26)	--
Phosphorylation (Phos)	149,469	2,254	8,741 (1,536/1,602)
Total unique proteins	--	7,632	--

Transcriptomics

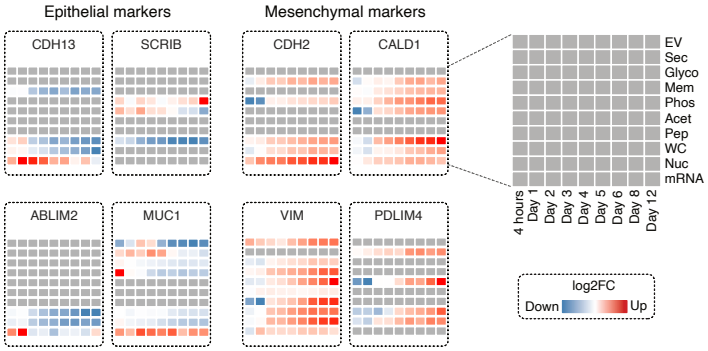
Single cell	Matrix	Total RNA	Genes (signif. up/dn)
Number of cells ***	1,913	mRNA	23,787 (2,441/1,993)
mRNA	9,785	mi-RNA	2,574 (178/118)

Metabolomics

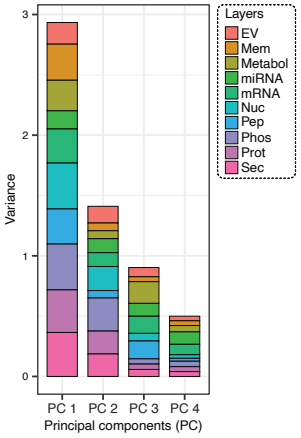
	Features (MS1 peaks)	HMDB indexed	KEGG indexed
Features (Level 2) ****	27,759	8,325 (610/468)	1,118 (65/59)

* Unique PTM sites remaining after QC steps; ** relative to `Control`; absolute log2FC ≥ 1, FDR ≤ 0.05, r^2 ≥ 0.6;
*** Number of cells & genes remaining after QC steps **** Level 2 identifications (Salek et al., 2013)

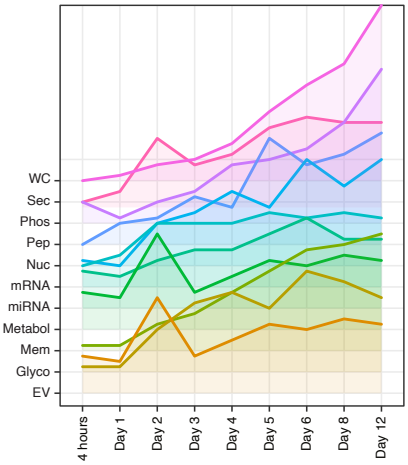
D MultiOmic profiles of EMT markers



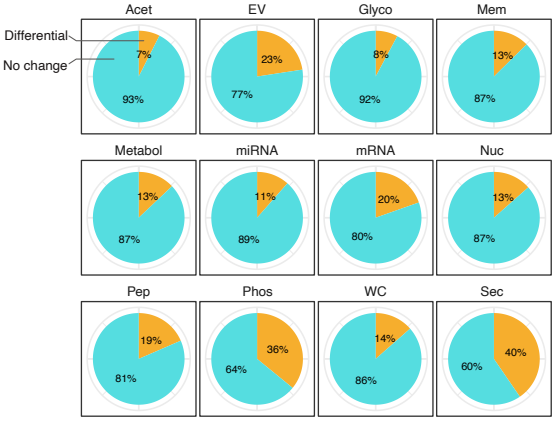
E Variance contribution



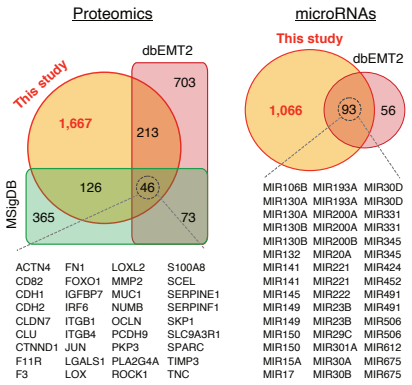
F Differential molecules over time points



G Differential molecules as % of total detected



H Comparison to current knowledge



I Enhanced discovery of coherent functional modules

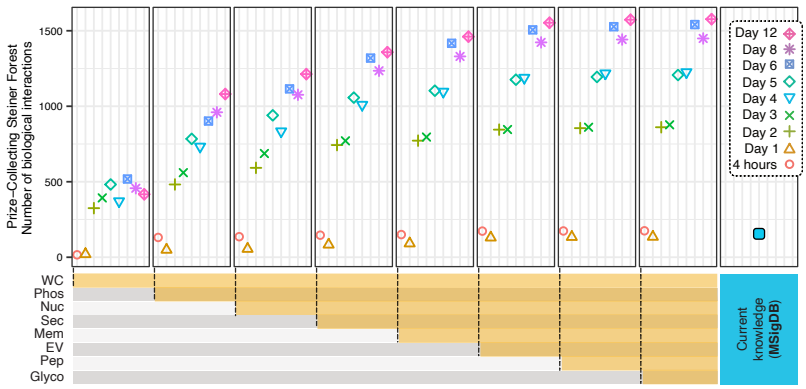


Figure 2

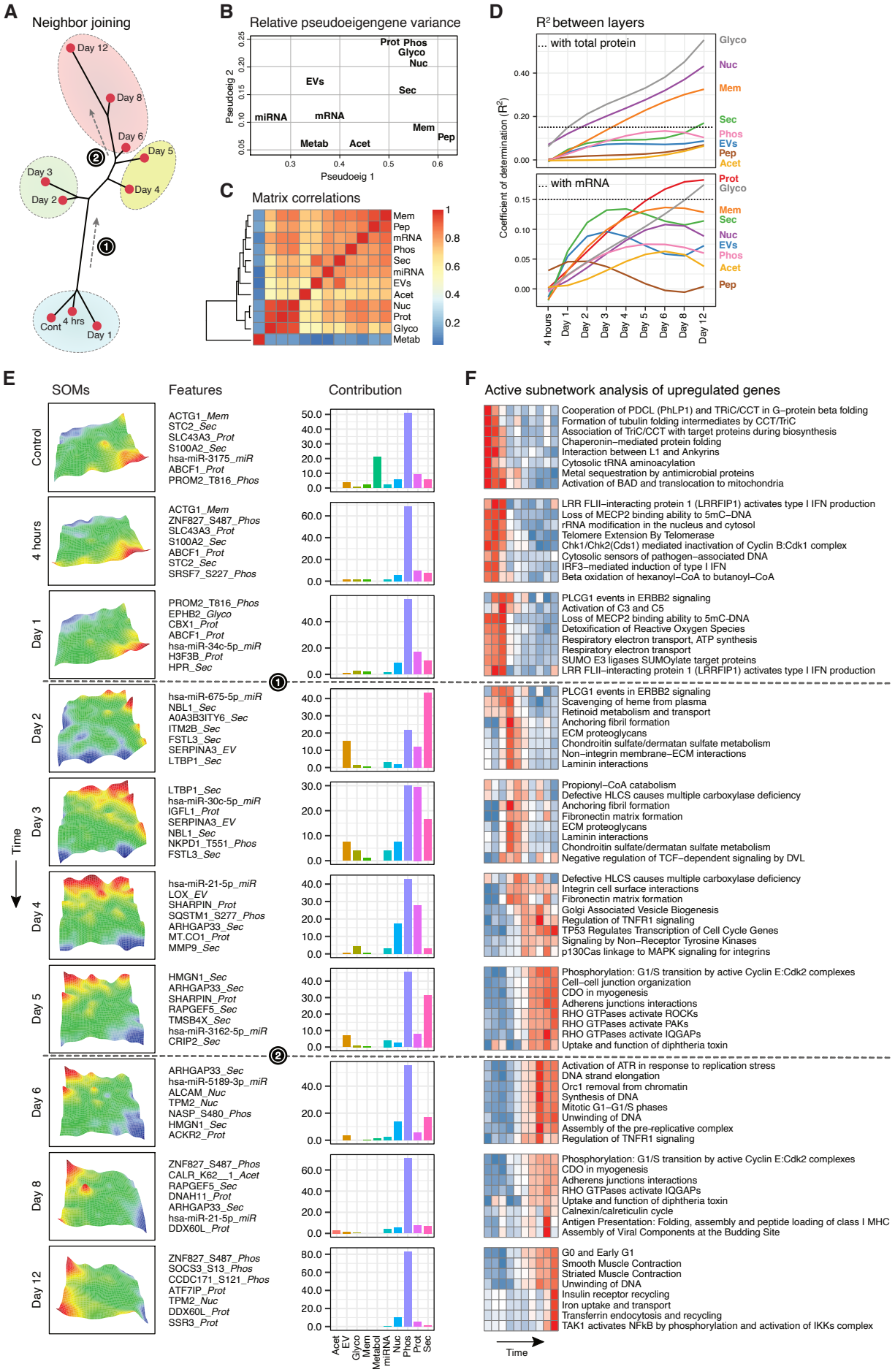

A
Neighbor joining
Day 12
Day 8
Day 6
Day 5
Day 4
Day 3
Day 2
Cont
4 hrs
Day 1
B
Relative pseudoeigengene variance
Pseudoeig 2
Pseudoeig 1
Prot Phos
Glyco
Nuc
EVs
Sec
miRNA
mRNA
Mem
Pep
Metab
Acet
C
Matrix correlations
Mem
Pep
mRNA
Phos
Sec
miRNA
EVs
Acet
Nuc
Prot
Glyco
Metab
D
R² between layers
... with total protein
... with mRNA
Coefficient of determination (R²)
4 hours
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 8
Day 12
E
SOMs
Features
Contribution
Control
4 hours
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 8
Day 12
Time
ACTG1_Mem
STC2_Sec
SLC43A3_Prot
S100A2_Sec
hsa-miR-3175_miR
ABCF1_Prot
PROM2_T816_Phos
ACTG1_Mem
ZNF827_S487_Phos
SLC43A3_Prot
S100A2_Sec
ABCF1_Prot
STC2_Sec
SRSF7_S227_Phos
PROM2_T816_Phos
EPHB2_Glyco
CBX1_Prot
ABCF1_Prot
hsa-miR-34c-5p_miR
H3F3B_Prot
HPR_Sec
hsa-miR-675-5p_miR
NBL1_Sec
A0A3B3ITY6_Sec
ITM2B_Sec
FSTL3_Sec
SERPINA3_EV
LTBP1_Sec
LTBP1_Sec
hsa-miR-30c-5p_miR
IGFL1_Prot
SERPINA3_EV
NBL1_Sec
NKPD1_T551_Phos
FSTL3_Sec
hsa-miR-21-5p_miR
LOX_EV
SHARPIN_Prot
SQSTM1_S277_Phos
ARHGAP33_Sec
MT.CO1_Prot
MMP9_Sec
HMGN1_Sec
ARHGAP33_Sec
SHARPIN_Prot
RAPGEF5_Sec
TMSB4X_Sec
hsa-miR-3162-5p_miR
CRIP2_Sec
ARHGAP33_Sec
hsa-miR-5189-3p_miR
ALCAM_Nuc
TPM2_Nuc
NASP_S480_Phos
HMGN1_Sec
ACKR2_Prot
ZNF827_S487_Phos
CALR_K62__1_Acet
RAPGEF5_Sec
DNAH11_Prot
ARHGAP33_Sec
hsa-miR-21-5p_miR
DDX60L_Prot
ZNF827_S487_Phos
SOCS3_S13_Phos
CCDC171_S121_Phos
ATF7IP_Prot
TPM2_Nuc
DDX60L_Prot
SSR3_Prot
Acet
EV
Glyco
Mem
Metabol
miRNA
Nuc
Phos
Prot
Sec
F
Active subnetwork analysis of upregulated genes
Cooperation of PDCL (PhLP1) and TRiC/CCT in G–protein beta folding
Formation of tubulin folding intermediates by CCT/TriC
Association of TriC/CCT with target proteins during biosynthesis
Chaperonin–mediated protein folding
Interaction between L1 and Ankyrins
Cytosolic tRNA aminoacylation
Metal sequestration by antimicrobial proteins
Activation of BAD and translocation to mitochondria
LRR FLII–interacting protein 1 (LRRFIP1) activates type I IFN production
Loss of MECP2 binding ability to 5mC–DNA
rRNA modification in the nucleus and cytosol
Telomere Extension By Telomerase
Chk1/Chk2(Cds1) mediated inactivation of Cyclin B:Cdk1 complex
Cytosolic sensors of pathogen–associated DNA
IRF3–mediated induction of type I IFN
Beta oxidation of hexanoyl–CoA to butanoyl–CoA
PLCG1 events in ERBB2 signaling
Activation of C3 and C5
Loss of MECP2 binding ability to 5mC-DNA
Detoxification of Reactive Oxygen Species
Respiratory electron transport, ATP synthesis
Respiratory electron transport
SUMO E3 ligases SUMOylate target proteins
LRR FLII–interacting protein 1 (LRRFIP1) activates type I IFN production
PLCG1 events in ERBB2 signaling
Scavenging of heme from plasma
Retinoid metabolism and transport
Anchoring fibril formation
ECM proteoglycans
Chondroitin sulfate/dermatan sulfate metabolism
Non–integrin membrane–ECM interactions
Laminin interactions
Propionyl–CoA catabolism
Defective HLCS causes multiple carboxylase deficiency
Anchoring fibril formation
Fibronectin matrix formation
ECM proteoglycans
Laminin interactions
Chondroitin sulfate/dermatan sulfate metabolism
Negative regulation of TCF–dependent signaling by DVL
Defective HLCS causes multiple carboxylase deficiency
Integrin cell surface interactions
Fibronectin matrix formation
Golgi Associated Vesicle Biogenesis
Regulation of TNFR1 signaling
TP53 Regulates Transcription of Cell Cycle Genes
Signaling by Non–Receptor Tyrosine Kinases
p130Cas linkage to MAPK signaling for integrins
Phosphorylation: G1/S transition by active Cyclin E:Cdk2 complexes
Cell–cell junction organization
CDO in myogenesis
Adherens junctions interactions
RHO GTPases activate ROCKs
RHO GTPases activate PAKs
RHO GTPases activate IQGAPs
Uptake and function of diphtheria toxin
Activation of ATR in response to replication stress
DNA strand elongation
Orc1 removal from chromatin
Synthesis of DNA
Mitotic G1–G1/S phases
Unwinding of DNA
Assembly of the pre-replicative complex
Regulation of TNFR1 signaling
Phosphorylation: G1/S transition by active Cyclin E:Cdk2 complexes
CDO in myogenesis
Adherens junctions interactions
RHO GTPases activate IQGAPs
Uptake and function of diphtheria toxin
Calnexin/calreticulin cycle
Antigen Presentation: Folding, assembly and peptide loading of class I MHC
Assembly of Viral Components at the Budding Site
G0 and Early G1
Smooth Muscle Contraction
Striated Muscle Contraction
Unwinding of DNA
Insulin receptor recycling
Iron uptake and transport
Transferrin endocytosis and recycling
TAK1 activates NFkB by phosphorylation and activation of IKKs complex
Time

Figure 3

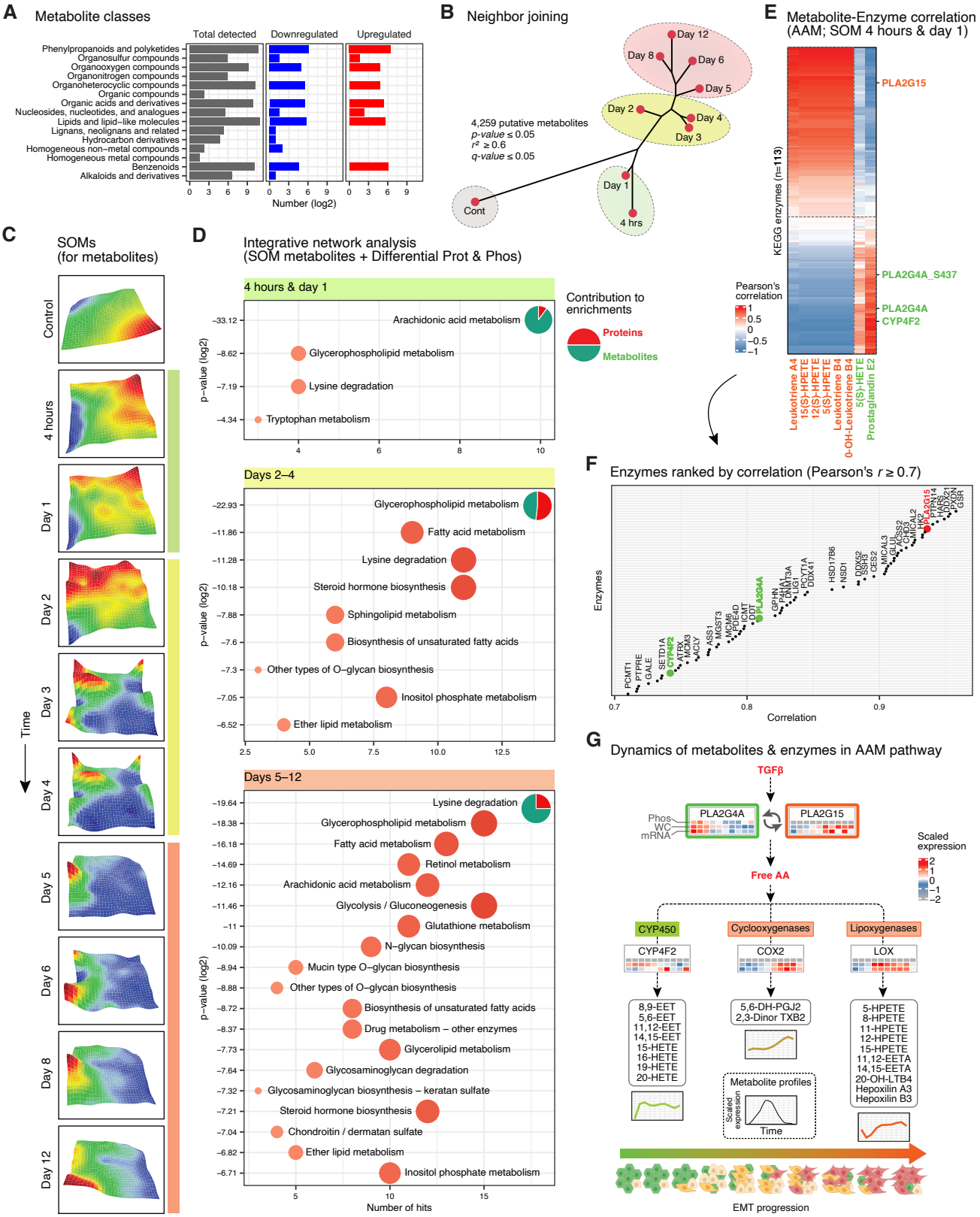

A
Metabolite classes
Total detected
Downregulated
Upregulated
Phenylpropanoids and polyketides
Organosulfur compounds
Organooxygen compounds
Organonitrogen compounds
Organoheterocyclic compounds
Organic compounds
Organic acids and derivatives
Nucleosides, nucleotides, and analogues
Lipids and lipid–like molecules
Lignans, neolignans and related
Hydrocarbon derivatives
Homogeneous non–metal compounds
Homogeneous metal compounds
Benzenoids
Alkaloids and derivatives
Number (log2)
B
Neighbor joining
4,259 putative metabolites
p-value ≤ 0.05
r² ≥ 0.6
q-value ≤ 0.05
Cont
4 hrs
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 8
Day 12
E
Metabolite-Enzyme correlation
(AAM; SOM 4 hours & day 1)
KEGG enzymes (n=113)
PLA2G15
PLA2G4A_S437
PLA2G4A
CYP4F2
Pearson's correlation
Leukotriene A4
15(S)-HPETE
12(S)-HPETE
5(S)-HPETE
Leukotriene B4
0-OH-Leukotriene B4
5(S)-HETE
Prostaglandin E2
C
SOMs
(for metabolites)
Control
4 hours
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 8
Day 12
Time
D
Integrative network analysis
(SOM metabolites + Differential Prot & Phos)
4 hours & day 1
Arachidonic acid metabolism
Glycerophospholipid metabolism
Lysine degradation
Tryptophan metabolism
Contribution to enrichments
Proteins
Metabolites
Days 2–4
Glycerophospholipid metabolism
Fatty acid metabolism
Lysine degradation
Steroid hormone biosynthesis
Sphingolipid metabolism
Biosynthesis of unsaturated fatty acids
Other types of O–glycan biosynthesis
Inositol phosphate metabolism
Ether lipid metabolism
Days 5–12
Lysine degradation
Glycerophospholipid metabolism
Fatty acid metabolism
Retinol metabolism
Arachidonic acid metabolism
Glycolysis / Gluconeogenesis
Glutathione metabolism
N–glycan biosynthesis
Mucin type O–glycan biosynthesis
Other types of O–glycan biosynthesis
Biosynthesis of unsaturated fatty acids
Drug metabolism – other enzymes
Glycerolipid metabolism
Glycosaminoglycan degradation
Glycosaminoglycan biosynthesis – keratan sulfate
Steroid hormone biosynthesis
Chondroitin / dermatan sulfate
Ether lipid metabolism
Inositol phosphate metabolism
p–value (log2)
Number of hits
F
Enzymes ranked by correlation (Pearson's r ≥ 0.7)
Enzymes
Correlation
G
Dynamics of metabolites & enzymes in AAM pathway
TGFβ
Phos
WC
mRNA
PLA2G4A
PLA2G15
Scaled expression
Free AA
CYP450
Cyclooxygenases
Lipoxygenases
CYP4F2
COX2
LOX
8,9-EET
5,6-EET
11,12-EET
14,15-EET
15-HETE
16-HETE
19-HETE
20-HETE
5,6-DH-PGJ2
2,3-Dinor TXB2
5-HPETE
8-HPETE
11-HPETE
12-HPETE
15-HPETE
11,12-EETA
14,15-EETA
20-OH-LTB4
Hepoxilin A3
Hepoxilin B3
Metabolite profiles
Time
EMT progression

Figure 4

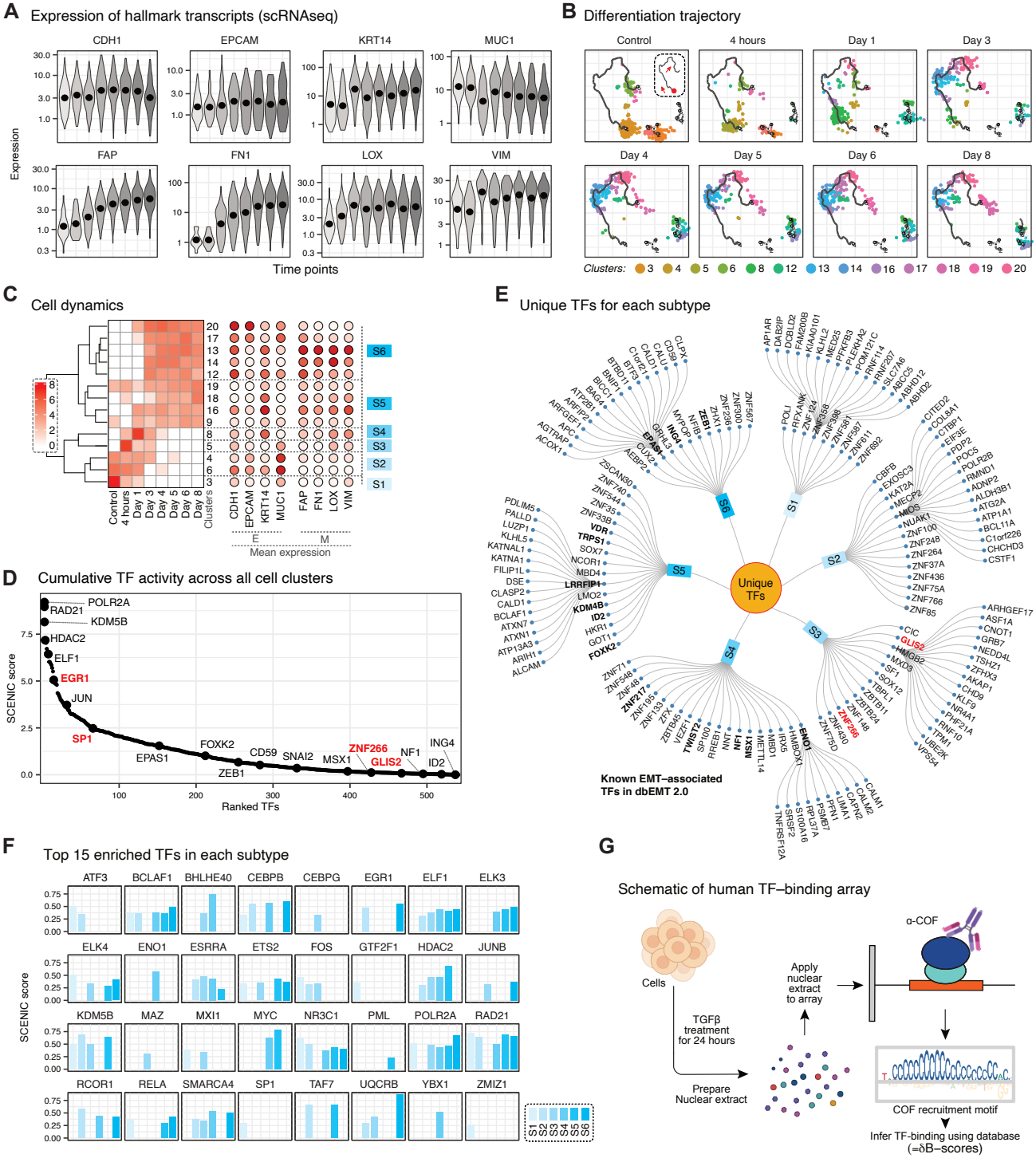

A
Expression of hallmark transcripts (scRNAseq)
CDH1
EPCAM
KRT14
MUC1
FAP
FN1
LOX
VIM
Expression
Time points
B
Differentiation trajectory
Control
4 hours
Day 1
Day 3
Day 4
Day 5
Day 6
Day 8
Clusters: 3 4 5 6 8 12 13 14 16 17 18 19 20
C
Cell dynamics
Control
4 hours
Day 1
Day 3
Day 4
Day 5
Day 6
Day 8
Clusters
CDH1
EPCAM
KRT14
MUC1
FAP
FN1
LOX
VIM
E
M
Mean expression
S6
S5
S4
S3
S2
S1
D
Cumulative TF activity across all cell clusters
SCENIC score
Ranked TFs
POLR2A
RAD21
KDM5B
HDAC2
ELF1
EGR1
JUN
SP1
EPAS1
FOXK2
ZEB1
CD59
SNAI2
MSX1
ZNF266
GLIS2
NF1
ID2
ING4
E
Unique TFs for each subtype
Unique TFs
S1
S2
S3
S4
S5
S6
Known EMT–associated TFs in dbEMT 2.0
F
Top 15 enriched TFs in each subtype
SCENIC score
ATF3
BCLAF1
BHLHE40
CEBPB
CEBPG
EGR1
ELF1
ELK3
ELK4
ENO1
ESRRA
ETS2
FOS
GTF2F1
HDAC2
JUNB
KDM5B
MAZ
MXI1
MYC
NR3C1
PML
POLR2A
RAD21
RCOR1
RELA
SMARCA4
SP1
TAF7
UQCRB
YBX1
ZMIZ1
S1 S2 S3 S4 S5 S6
G
Schematic of human TF–binding array
Cells
TGFβ treatment for 24 hours
Prepare Nuclear extract
Apply nuclear extract to array
α-COF
COF recruitment motif
Infer TF-binding using database (=δB–scores)
H
δB–score distributions
Transcription factors
ZNF263
SP1
GLIS2
p-val ≤ 0.006
D ≥ 0.27
p-val ≤ 0.007
D ≥ 0.34
p-val ≤ 0.000
D ≥ 0.45
δB–scores
Recovered
Consensus
I
δB–score distributions
Transcription factors
PAX1
GCM2
EGR1
p-val ≤ 0.001
D ≥ 0.32
p-val ≤ 0.120
D ≥ 0.25
p-val ≤ 0.092
D ≥ 0.23
δB–scores
Recovered
Consensus

Figure 5

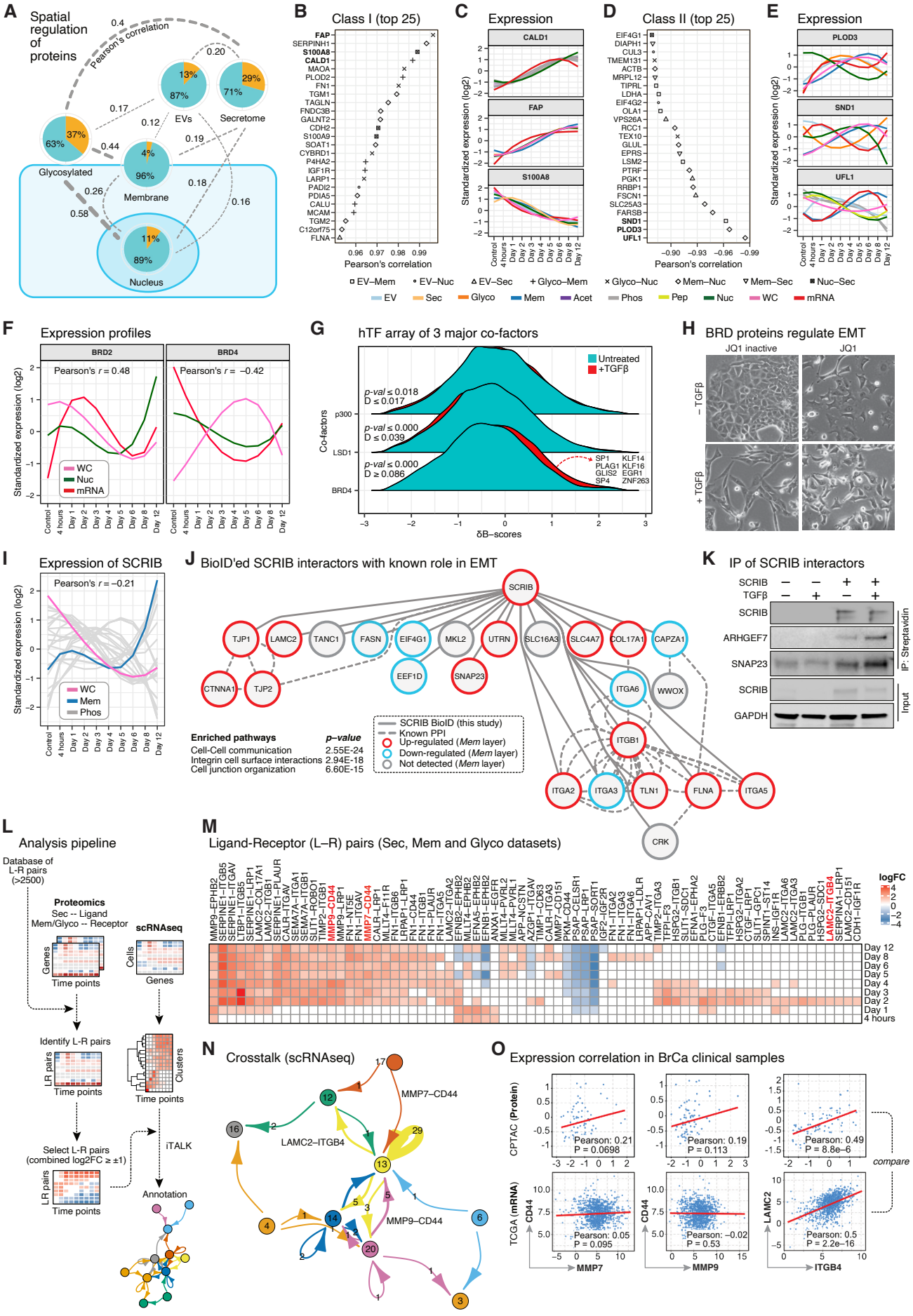

A
Spatial regulation of proteins
Pearson's correlation
EVs
Secretome
Glycosylated
Membrane
Nucleus
B
Class I (top 25)
Pearson's correlation
C
Expression
CALD1
FAP
S100A8
Standardized expression (log2)
D
Class II (top 25)
Pearson's correlation
E
Expression
PLOD3
SND1
UFL1
EV–Mem
EV–Nuc
EV–Sec
Glyco–Mem
Glyco–Nuc
Mem–Nuc
Mem–Sec
Nuc–Sec
EV
Sec
Glyco
Mem
Acet
Phos
Pep
Nuc
WC
mRNA
F
Expression profiles
BRD2
BRD4
Pearson's r = 0.48
Pearson's r = −0.42
WC
Nuc
mRNA
G
hTF array of 3 major co-factors
Untreated
+TGFβ
Co-factors
p300
LSD1
BRD4
δB–scores
H
BRD proteins regulate EMT
JQ1 inactive
JQ1
− TGFβ
+ TGFβ
I
Expression of SCRIB
Pearson's r = −0.21
WC
Mem
Phos
J
BioID'ed SCRIB interactors with known role in EMT
SCRIB
Enriched pathways
p–value
Cell-Cell communication 2.55E-24
Integrin cell surface interactions 2.94E-18
Cell junction organization 6.60E-15
SCRIB BioID (this study)
Known PPI
Up-regulated (Mem layer)
Down-regulated (Mem layer)
Not detected (Mem layer)
K
IP of SCRIB interactors
SCRIB
TGFβ
ARHGEF7
SNAP23
GAPDH
IP: Streptavidin
Input
L
Analysis pipeline
Database of L-R pairs (>2500)
Proteomics
Sec -- Ligand
Mem/Glyco -- Receptor
scRNAseq
Genes
Time points
Cells
Identify L-R pairs
LR pairs
Clusters
Select L-R pairs (combined log2FC ≥ ±1)
iTALK
Annotation
M
Ligand-Receptor (L–R) pairs (Sec, Mem and Glyco datasets)
logFC
Day 12
Day 8
Day 6
Day 5
Day 4
Day 3
Day 2
Day 1
4 hours
N
Crosstalk (scRNAseq)
MMP7–CD44
LAMC2–ITGB4
MMP9–CD44
O
Expression correlation in BrCa clinical samples
CPTAC (Protein)
TCGA (mRNA)
Pearson: 0.21
P = 0.0698
Pearson: 0.19
P = 0.113
Pearson: 0.49
P = 8.8e−6
compare
Pearson: 0.05
P = 0.095
Pearson: −0.02
P = 0.53
Pearson: 0.5
P = 2.2e−16
MMP7
MMP9
ITGB4

Figure 6

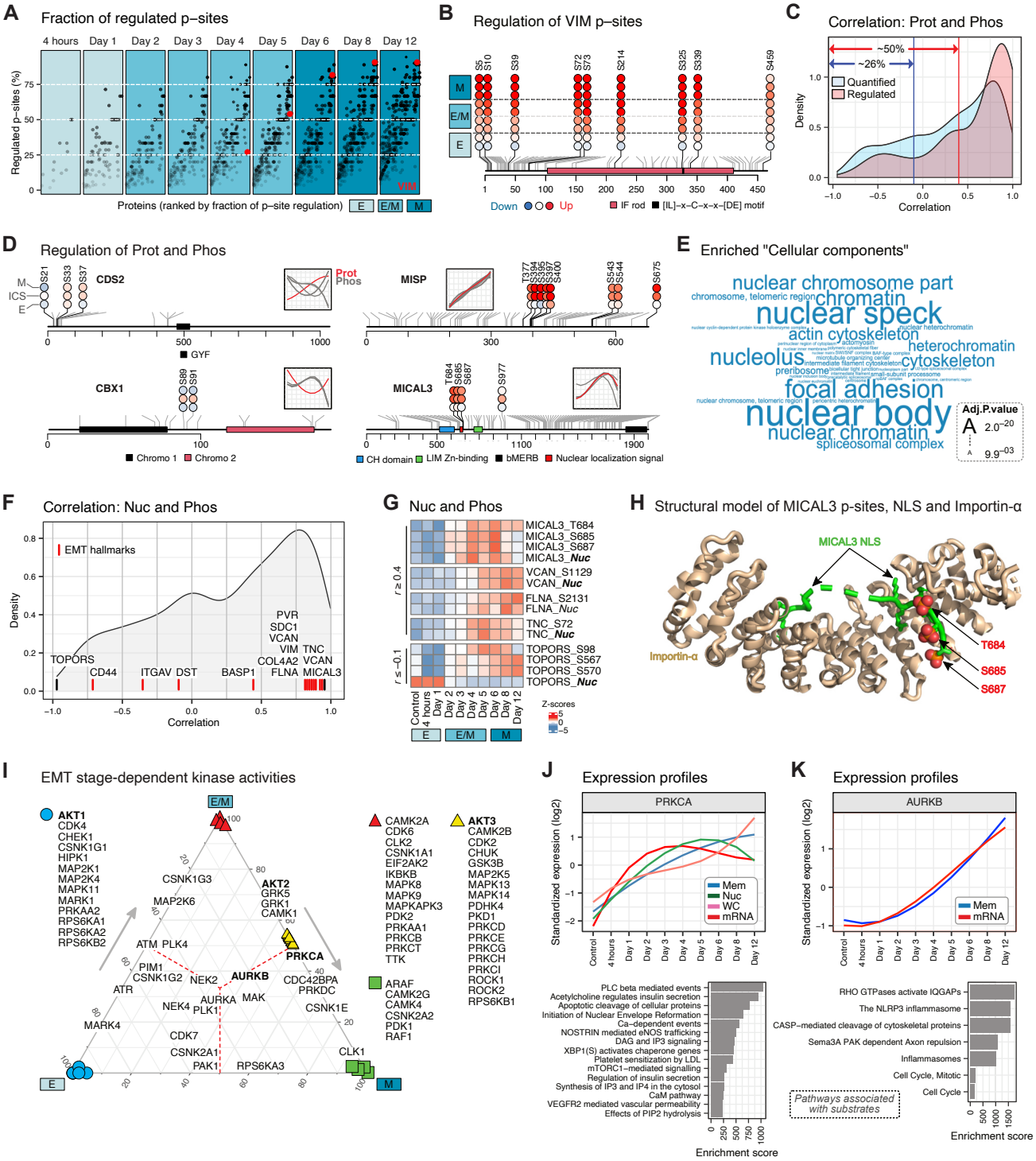

A
Fraction of regulated p–sites
4 hours
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 8
Day 12
Regulated p–sites (%)
VIM
Proteins (ranked by fraction of p–site regulation)
E
E/M
M
B
Regulation of VIM p–sites
S5
S10
S39
S72
S73
S214
S325
S339
S459
Down
Up
IF rod
[IL]–x–C–x–x–[DE] motif
C
Correlation: Prot and Phos
~50%
~26%
Quantified
Regulated
Density
Correlation
D
Regulation of Prot and Phos
CDS2
MISP
CBX1
MICAL3
Prot
Phos
GYF
Chromo 1
Chromo 2
CH domain
LIM Zn-binding
bMERB
Nuclear localization signal
E
Enriched "Cellular components"
nuclear chromosome part
chromosome, telomeric region
chromatin
nuclear speck
actin cytoskeleton
heterochromatin
nucleolus
cytoskeleton
focal adhesion
nuclear body
nuclear chromatin
spliceosomal complex
Adj.P.value
F
Correlation: Nuc and Phos
EMT hallmarks
Density
Correlation
TOPORS
CD44
ITGAV
DST
BASP1
PVR
SDC1
VCAN
VIM
TNC
COL4A2
FLNA
MICAL3
G
Nuc and Phos
MICAL3_T684
MICAL3_S685
MICAL3_S687
MICAL3_Nuc
VCAN_S1129
VCAN_Nuc
FLNA_S2131
FLNA_Nuc
TNC_S72
TNC_Nuc
TOPORS_S98
TOPORS_S567
TOPORS_S570
TOPORS_Nuc
Z-scores
H
Structural model of MICAL3 p-sites, NLS and Importin-α
MICAL3 NLS
Importin-α
T684
S685
S687
I
EMT stage-dependent kinase activities
AKT1
CDK4
CHEK1
CSNK1G1
HIPK1
MAP2K1
MAP2K4
MAPK11
MARK1
PRKAA2
RPS6KA1
RPS6KA2
RPS6KB2
CAMK2A
CDK6
CLK2
CSNK1A1
EIF2AK2
IKBKB
MAPK8
MAPK9
MAPKAPK3
PDK2
PRKAA1
PRKCB
PRKCT
TTK
AKT3
CAMK2B
CDK2
CHUK
GSK3B
MAP2K5
MAPK13
MAPK14
PDHK4
PKD1
PRKCD
PRKCE
PRKCG
PRKCH
PRKCI
ROCK1
ROCK2
RPS6KB1
ARAF
CAMK2G
CAMK4
CSNK2A2
PDK1
RAF1
AKT2
GRK5
GRK1
CAMK1
PRKCA
AURKB
CDC42BPA
PRKDC
CSNK1E
CLK1
CSNK1G3
MAP2K6
ATM
PLK4
PIM1
CSNK1G2
NEK2
ATR
AURKA
MAK
NEK4
PLK1
MARK4
CDK7
CSNK2A1
PAK1
RPS6KA3
J
Expression profiles
PRKCA
Standardized expression (log2)
Mem
Nuc
WC
mRNA
PLC beta mediated events
Acetylcholine regulates insulin secretion
Apoptotic cleavage of cellular proteins
Initiation of Nuclear Envelope Reformation
Ca–dependent events
NOSTRIN mediated eNOS trafficking
DAG and IP3 signaling
XBP1(S) activates chaperone genes
Platelet sensitization by LDL
mTORC1–mediated signalling
Regulation of insulin secretion
Synthesis of IP3 and IP4 in the cytosol
CaM pathway
VEGFR2 mediated vascular permeability
Effects of PIP2 hydrolysis
Enrichment score
K
Expression profiles
AURKB
Standardized expression (log2)
Mem
mRNA
RHO GTPases activate IQGAPs
The NLRP3 inflammasome
CASP–mediated cleavage of cytoskeletal proteins
Sema3A PAK dependent Axon repulsion
Inflammasomes
Cell Cycle, Mitotic
Cell Cycle
Pathways associated with substrates
Enrichment score

Figure 7

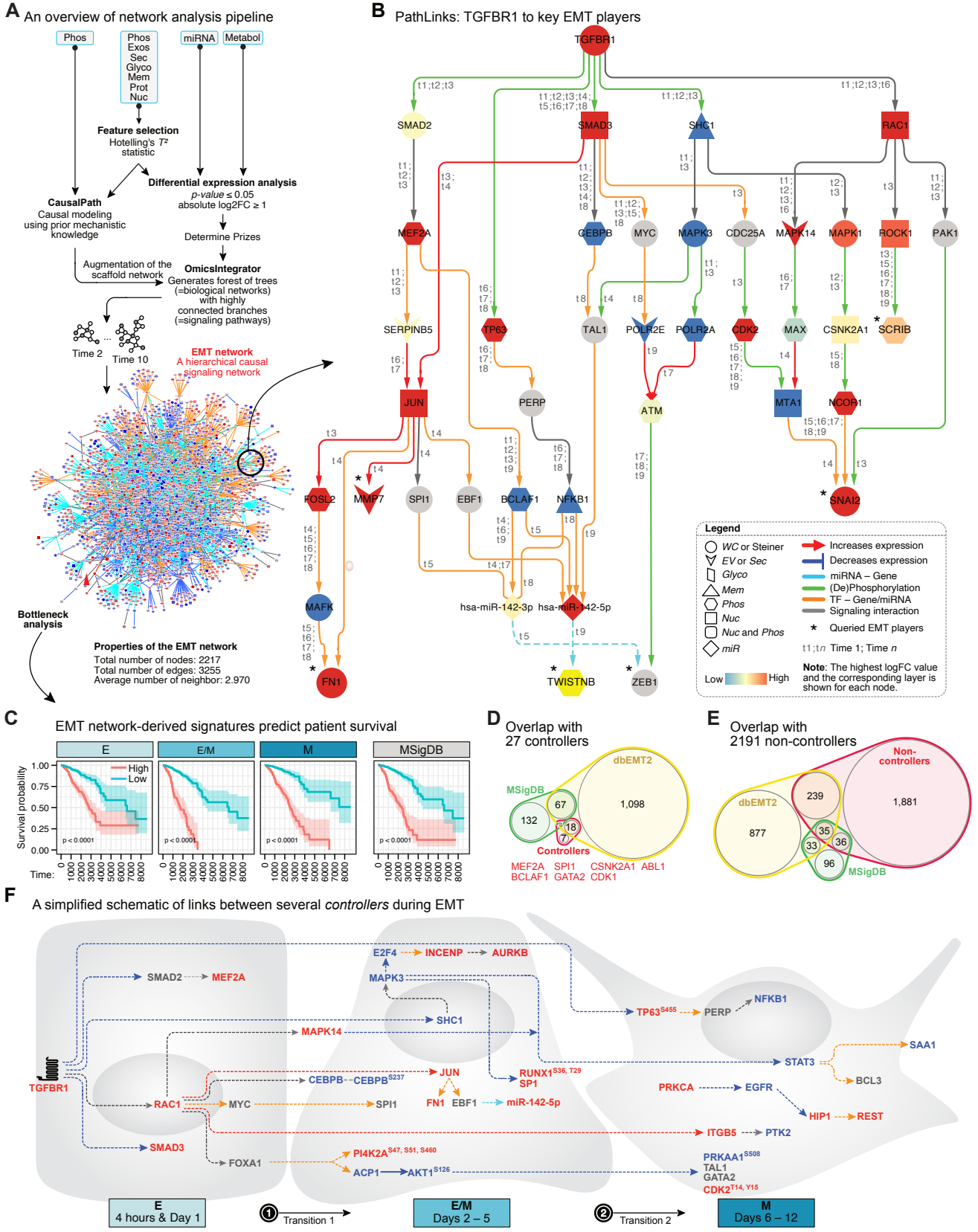

A An overview of network analysis pipeline
Phos
Phos Exos Sec Glyco Mem Prot Nuc
miRNA
Metabol
Feature selection
Hotelling's T² statistic
Differential expression analysis
p-value ≤ 0.05
absolute log2FC ≥ 1
CausalPath
Causal modeling using prior mechanistic knowledge
Determine Prizes
Augmentation of the scaffold network
OmicsIntegrator
Generates forest of trees (=biological networks) with highly connected branches (=signaling pathways)
Time 2
Time 10
EMT network
A hierarchical causal signaling network
Bottleneck analysis
Properties of the EMT network
Total number of nodes: 2217
Total number of edges: 3255
Average number of neighbor: 2.970
B PathLinks: TGFBR1 to key EMT players
TGFBR1
SMAD2
SMAD3
SHC1
RAC1
MEF2A
CEBPB
MYC
MAPK3
CDC25A
MAPK14
MAPK1
ROCK1
PAK1
SERPINB5
TP63
TAL1
POLR2E
POLR2A
CDK2
MAX
CSNK2A1
SCRIB
JUN
PERP
ATM
MTA1
NCOR1
FOSL2
MMP7
SPI1
EBF1
BCLAF1
NFKB1
SNAI2
MAFK
hsa-miR-142-3p
hsa-miR-142-5p
FN1
TWISTNB
ZEB1
Legend
WC or Steiner
EV or Sec
Glyco
Mem
Phos
Nuc
Nuc and Phos
miR
Low
High
Increases expression
Decreases expression
miRNA – Gene
(De)Phosphorylation
TF – Gene/miRNA
Signaling interaction
Queried EMT players
t1;tn Time 1; Time n
Note: The highest logFC value and the corresponding layer is shown for each node.
C EMT network-derived signatures predict patient survival
E
E/M
M
MSigDB
Survival probability
High
Low
p < 0.0001
Time:
D Overlap with 27 controllers
dbEMT2
MSigDB
1,098
132
67
18
7
Controllers
MEF2A SPI1 CSNK2A1 ABL1
BCLAF1 GATA2 CDK1
E Overlap with 2191 non-controllers
Non-controllers
dbEMT2
239
1,881
877
35
33
36
96
MSigDB
F A simplified schematic of links between several controllers during EMT
E2F4
INCENP
AURKB
SMAD2
MEF2A
MAPK3
TP63S455
PERP
NFKB1
SHC1
MAPK14
STAT3
SAA1
BCL3
TGFBR1
CEBPB
CEBPBS237
JUN
RUNX1S36, T29
SP1
PRKCA
EGFR
RAC1
MYC
SPI1
FN1
EBF1
miR-142-5p
HIP1
REST
ITGB5
PTK2
SMAD3
FOXA1
PI4K2AS47, S51, S460
ACP1
AKT1S126
PRKAA1S508
TAL1
GATA2
CDK2T14, Y15
E
4 hours & Day 1
Transition 1
E/M
Days 2 – 5
Transition 2
M
Days 6 – 12
G Workflow for morphometric analysis of inhibitor screen
Controllers
Selected Inhibitors
Experiment
Analysis
Inhibitor
Target
EMT inhibitor 1
Known EMT blocker
Sonidegib
Hh inhibitor
LB-100
PP2A inhibitor
Barasertib
AURKB inhibitor
Autocamtide
CaMK-II inhibitor
PP1
SRC inhibitor
APCP
NT5E inhibitor
TGFβ
+
Inhibitor permutations
Control
Day 3
GEN IMA SEG
Generic Image Segmentation Tool
Original
Segmented
Mask
Measured: Circularity of cells
H Exposed combinatorial vulnerabilities
TGFβ/Control
EMT inhibitor 1
Sonidegib
LB-100
Barasertib
Autocamtide
PP1
APCP

Figure S1. Related to Figure 1

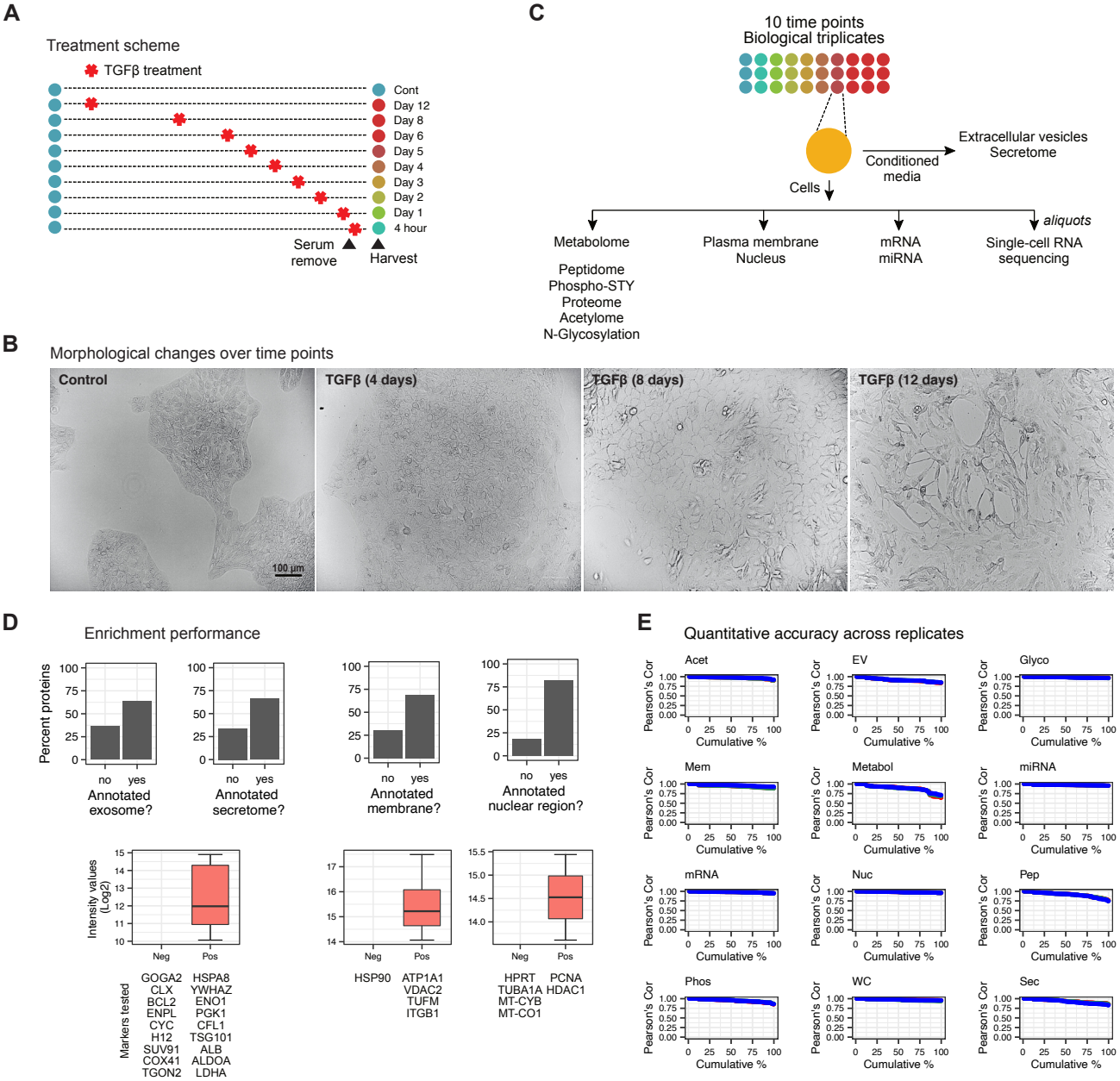


Figure S2. Related to Figure 2

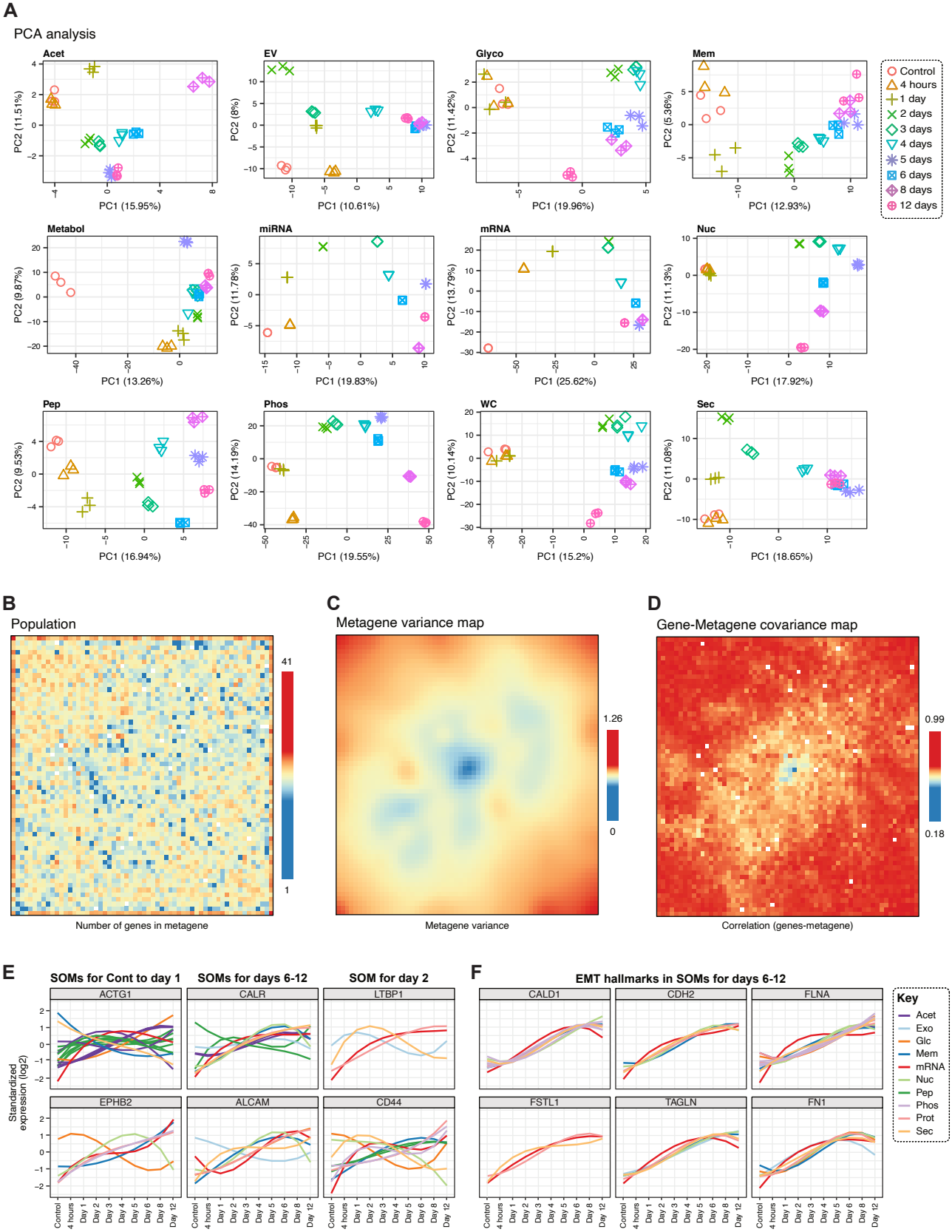
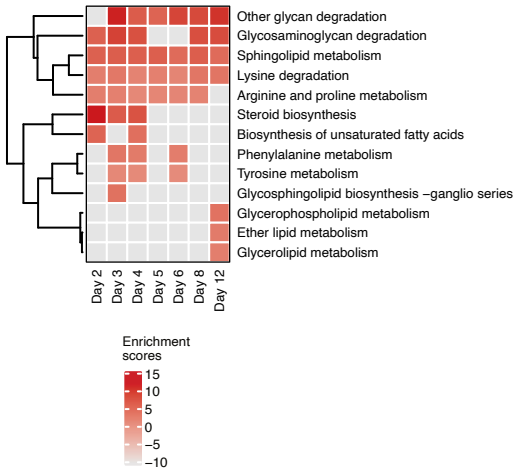
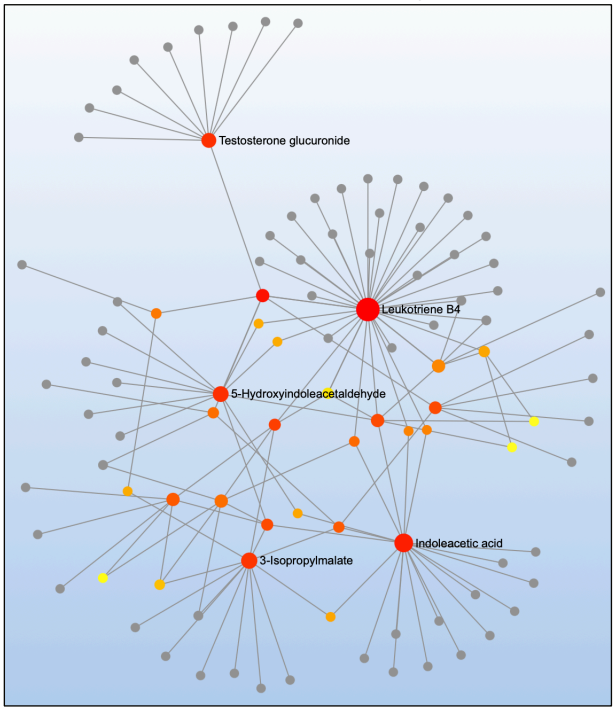


Figure S3. Related to Figure 3

A GeneSet enrichment of metabolic pathways using active subnetworks



B Metabolite-metabolite interaction network of metabolites identified in SOM for **4 hours to day 1**



C Metabolite-metabolite interaction network of metabolites identified in SOM for **day 2 to day 4**

D Metabolite-metabolite interaction network of metabolites identified in SOM for **day 5 to day 12**

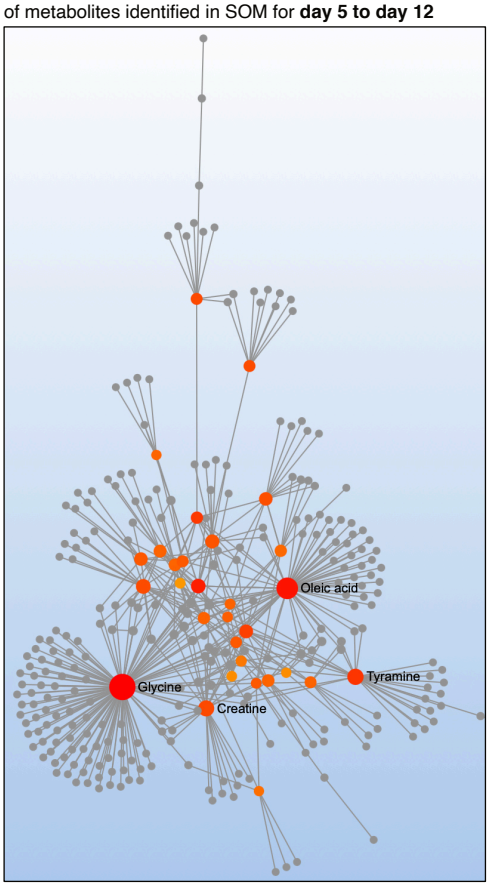


Figure S4. Related to Figure 4

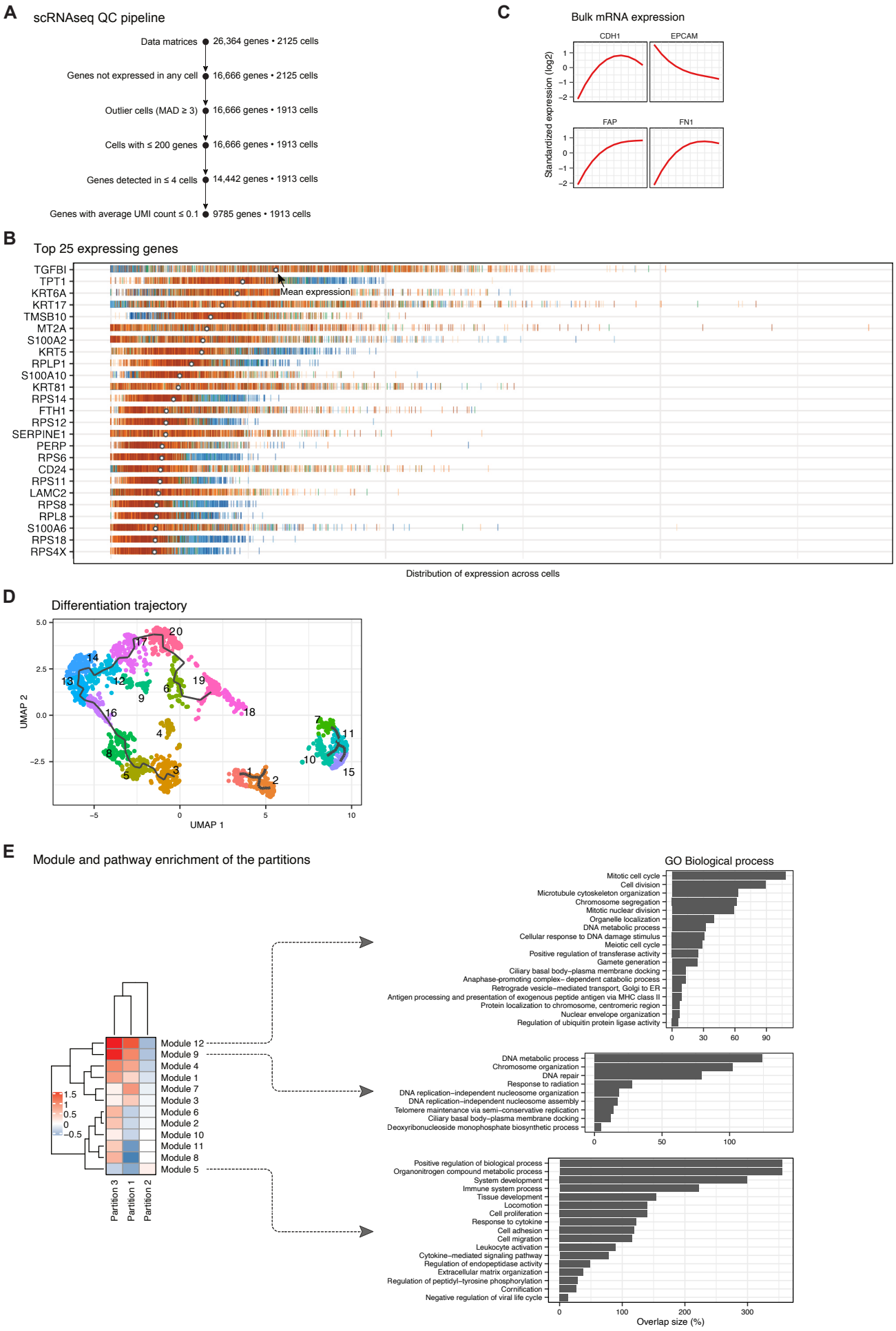


Figure S5. Related to Figure 5

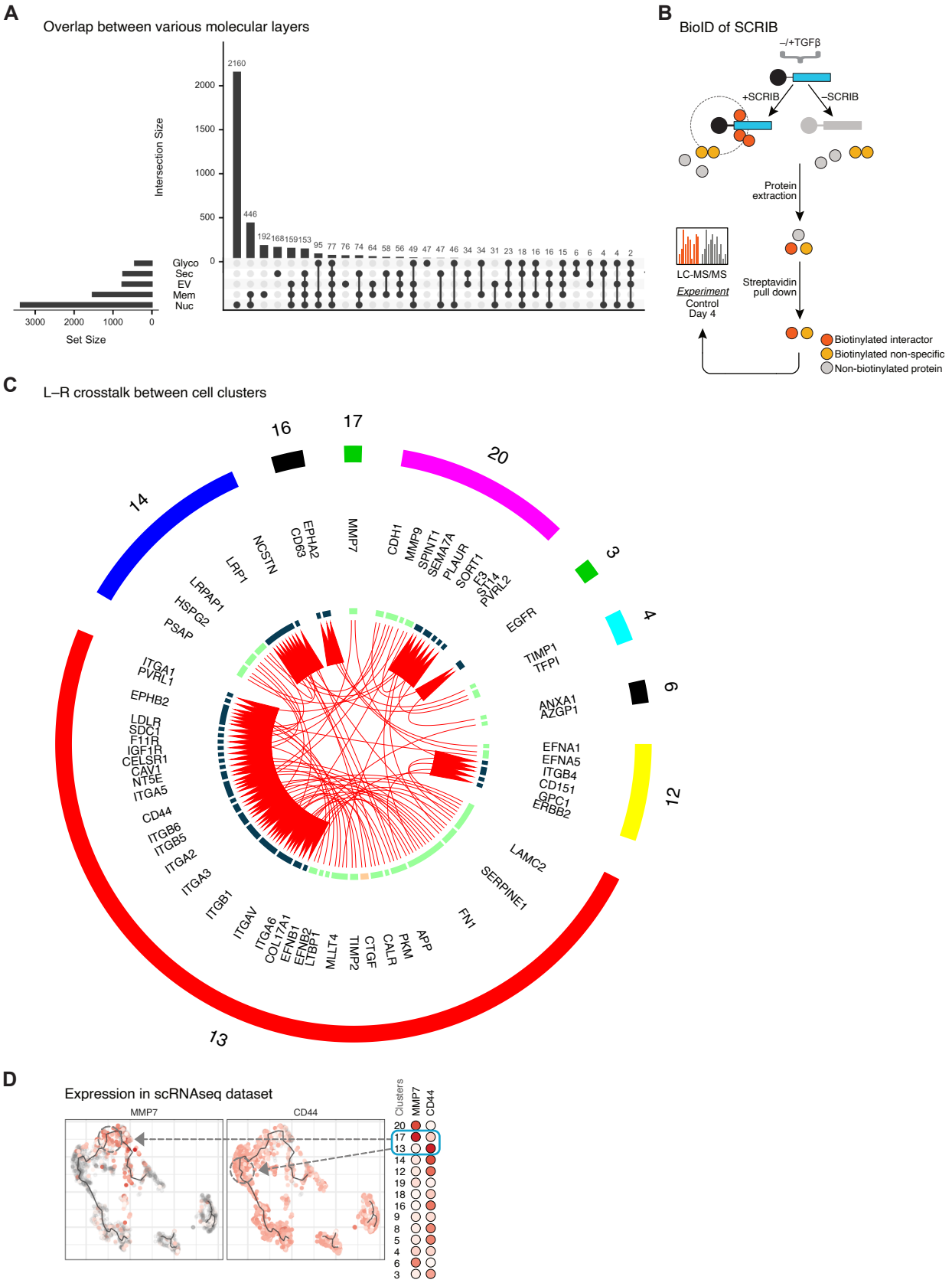


Figure S6. Related to Figure 6

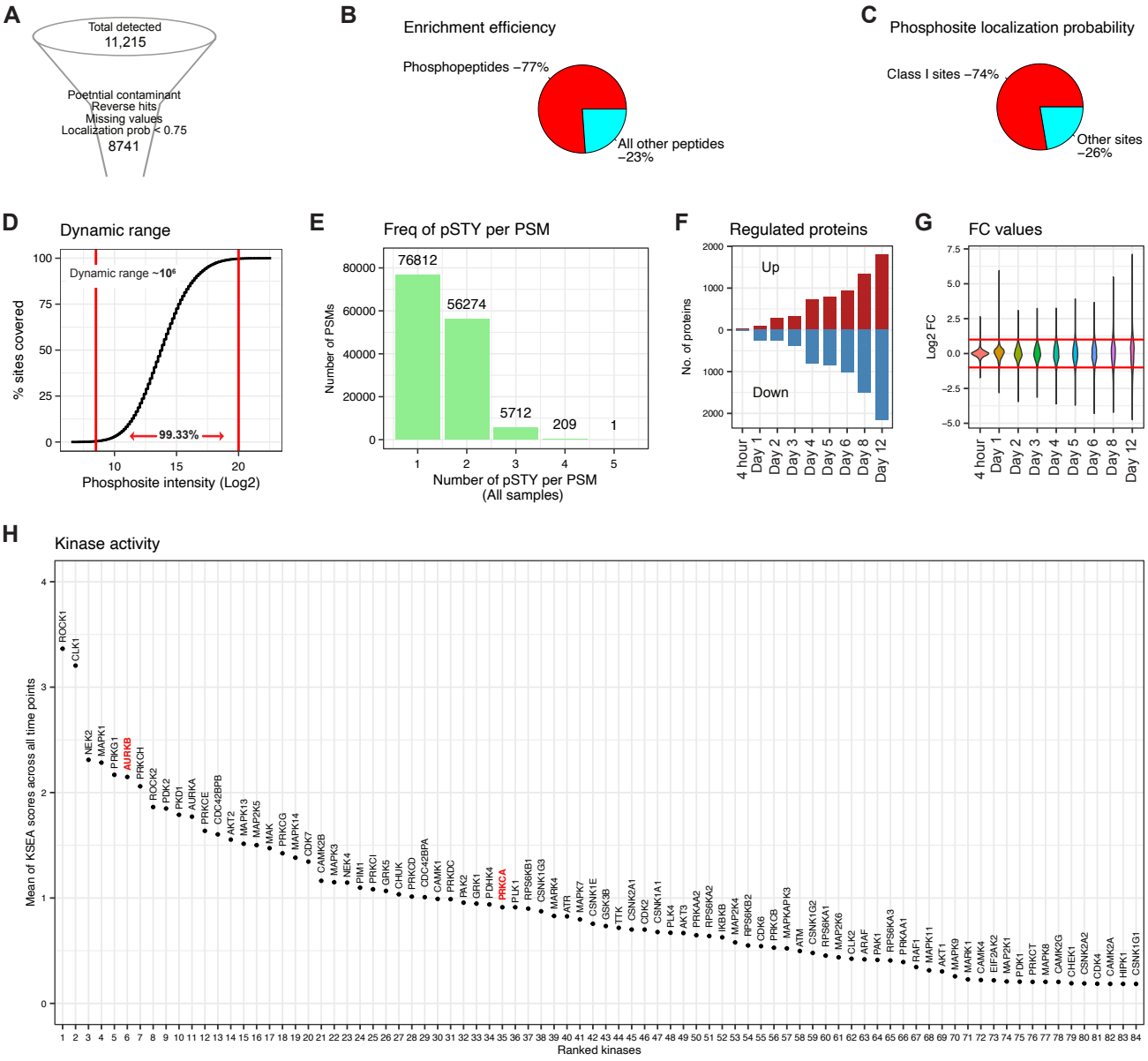
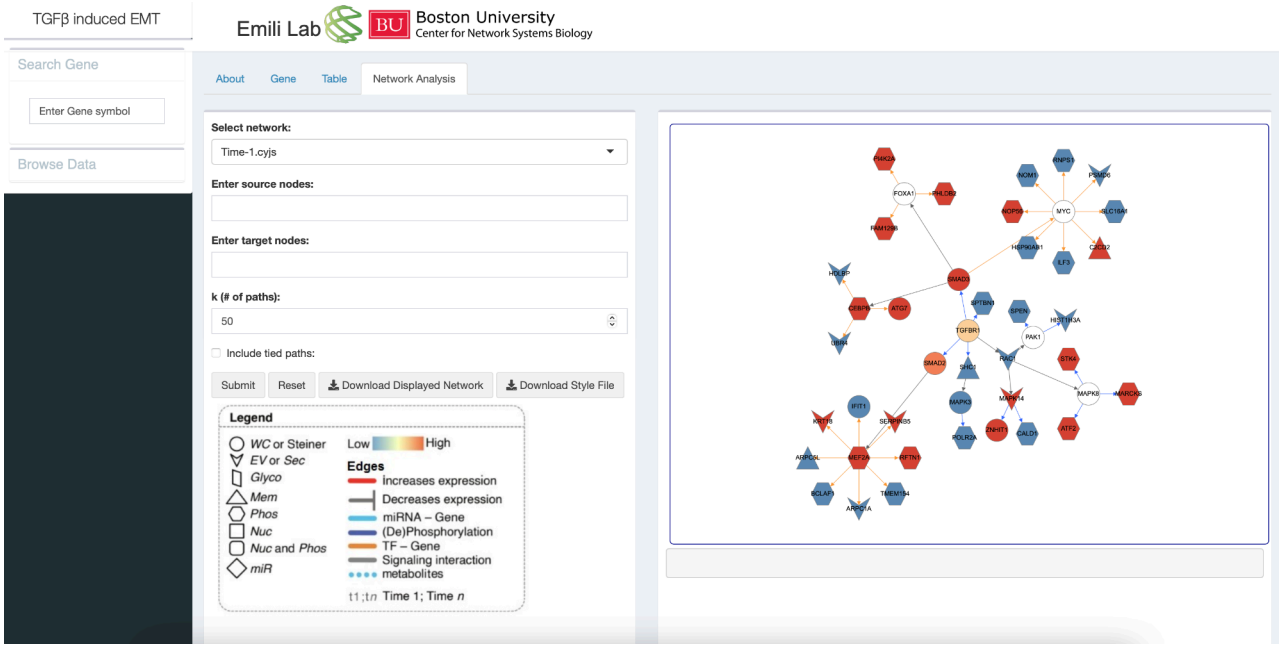


Figure S7. Related to Figure 7

Interface for querying the EMT network