communications biology

Article

https://doi.org/10.1038/s42003-024-07408-7

Genome-scale modeling identifies dynamic metabolic vulnerabilities during the epithelial to mesenchymal transition

Check for updates

Rupa Bhowmick[1,7], Scott Campit[2,7], Shiva Krishna Katkam[3], Venkateshwar G. Keshamouni[3,4,5] ✉ & Sriram Chandrasekaran[1,2,4,6] ✉

Epithelial-to-mesenchymal transition (EMT) is a conserved cellular process critical for embryogenesis, wound healing, and cancer metastasis. During EMT, cells undergo large-scale metabolic reprogramming that supports multiple functional phenotypes including migration, invasion, survival, chemo-resistance and stemness. However, the extent of metabolic network rewiring during EMT is unclear. In this work, using genome-scale metabolic modeling, we perform a meta-analysis of time-course transcriptomics, time-course proteomics, and single-cell transcriptomics EMT datasets from cell culture models stimulated with TGF-β. We uncovered temporal metabolic dependencies in glycolysis and glutamine metabolism, and experimentally validated isoform-specific dependency on Enolase3 for cell survival during EMT. We derived a prioritized list of metabolic dependencies based on model predictions, literature mining, and CRISPR-Cas9 essentiality screens. Notably, enolase and triose phosphate isomerase reaction fluxes significantly correlate with survival of lung adenocarcinoma patients. Our study illustrates how integration of heterogeneous datasets using a mechanistic computational model can uncover temporal and cell-state-specific metabolic dependencies.

Epithelial-to-mesenchymal transition (EMT) is a reversible developmental process that facilitates the transition from an epithelial (E) cell to a motile and invasive mesenchymal-like (M) cell, enabling tumor cells to initiate metastasis[1]. EMT is not a binary process but occurs through a spectrum of distinct intermediate states. In addition to metastasis, the process of EMT is implicated in several clinically relevant aspects including, tumor heterogeneity, stemness, modulation of immune surveillance, and drug resistance[2,3]. Understanding underlying regulatory mechanisms can help develop therapeutic strategies that prevent EMT and inhibit metastasis.

The cytokine-transforming growth factor β (TGF-β), is a potent inducer of EMT. Consistently, TGF-β levels are highly upregulated and directly correlate with tumor progression, enhanced invasion, metastasis, and poor survival in patients with non-small cell lung cancer (NSCLC)[4]. Cell culture models of TGF-β -induced EMT serves as robust in vitro model to investigate mechanisms of metastasis.

Extensive metabolic reprogramming is necessary to support various molecular and structural changes that occur during EMT. Studies have demonstrated enhanced glycolytic activity during EMT. While several metabolic rewiring strategies have been observed in diverse cancers[5], the causal relationship between metabolic changes and EMT is still unclear. A comprehensive systems-level characterization of metabolic reprogramming during EMT has not been carried out. A key challenge has been the heterogeneity in EMT responses at the transcriptional level[6]. While EMT has been extensively studied, no common metabolic programs have been identified. Which metabolic pathways predominate during different stages of EMT is unknown[7]. Methods that can infer systems-level metabolic dysregulation using omics data will be invaluable for understanding causal relationship between EMT and resulting functional consequences.

We hence analyzed transcriptomic and proteomic changes during EMT using metabolic network models. We hypothesized that integrating

[1]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA. [2]Program in Chemical Biology, University of Michigan, Ann Arbor, MI, USA. [3]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [4]Rogel Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA. [5]LTC Charles S. Kettles VA Medical Center, Research Service (151), Ann Arbor, MI, USA. [6]Center for Bioinformatics and Computational Medicine, Ann Arbor, MI, USA. [7]These authors contributed equally: Rupa Bhowmick, Scott Campit. ✉e-mail: vkeshamo@umich.edu; csriram@umich.edu

heterogenous omics data with metabolic network models harmonizes the discrepancies and missing information in the datasets and creates a holistic context-specific metabolic framework for analysis. Integrating heterogeneous data with genome-scale metabolic models can also reduce noise and increase consistency between datasets[8–10].

To further reduce heterogeneity due to model systems, we focus on a well-established and extensively studied model of EMT, namely, A549 lung adenocarcinoma cells undergoing EMT after stimulation with TGF-β. The focus on the human metabolic network and the analysis of large collection of A549 EMT datasets allowed us to gain statistical power and uncover many common metabolic alterations not reported previously.

Here we performed a meta-analysis of time-course bulk and single-cell omics studies using Constraint-Based Optimization and Reconstruction Analysis (COBRA), a widely used approach for simulating genome-scale metabolic models[11]. COBRA methods simulate metabolic fluxes by using the metabolic network architecture, nutrient availability, and omics data as constraints in an optimization problem. COBRA methods have inferred metabolic rewiring strategies in several cancer subtypes[8,12–14], for instance, incorporating metabolomics data to identify synthetically lethal metabolic genes in pancreatic cancer[15].

Till date, the use of COBRA methods to study EMT has been limited and focussed on bulk transcriptomics data[16]. Some studies have investigated metabolic regulation in EMT through experimental flux analysis and metabolomics[17,18]. However, COBRA and experimental methods have not yet been applied to characterize the metabolic properties of intermediate states during EMT. We provide an overview diagram of our approach in Fig. 1 and a more detailed flowchart of steps used for modeling and validation of predicted targets in Supplementary Fig. 1. From our analysis, we were able to identify known metabolic dependencies during EMT, such as uptake of glucose and glutamine. Notably, we observed metabolic reprogramming that indicates intermediate EM state cells switch to alternative energy metabolism fueled by glutamine and fatty acids. We also predicted new metabolic dependencies including the enolase (ENO) reactions and those related to alpha-ketoglutarate (AKG) metabolism. Surprisingly, many of these dependencies were time-specific, suggesting that there is a narrow temporal window during which the cells can be targeted with drugs that inhibit these pathways. We also predicted metabolic changes that showed consistent trends based on model predictions derived from both the bulk and single-cell studies, CRISPR gene knockout (KO) screens and patient survival analysis.

Results

Cells undergoing EMT exhibit enhanced glycolysis during early and late stages

We performed a meta-analysis of differentially expressed genes and proteins across four bulk EMT datasets. Two were transcriptomics datasets—GSE17708[19] and GSE17518[20], and two were proteomics-based datasets[21,22].

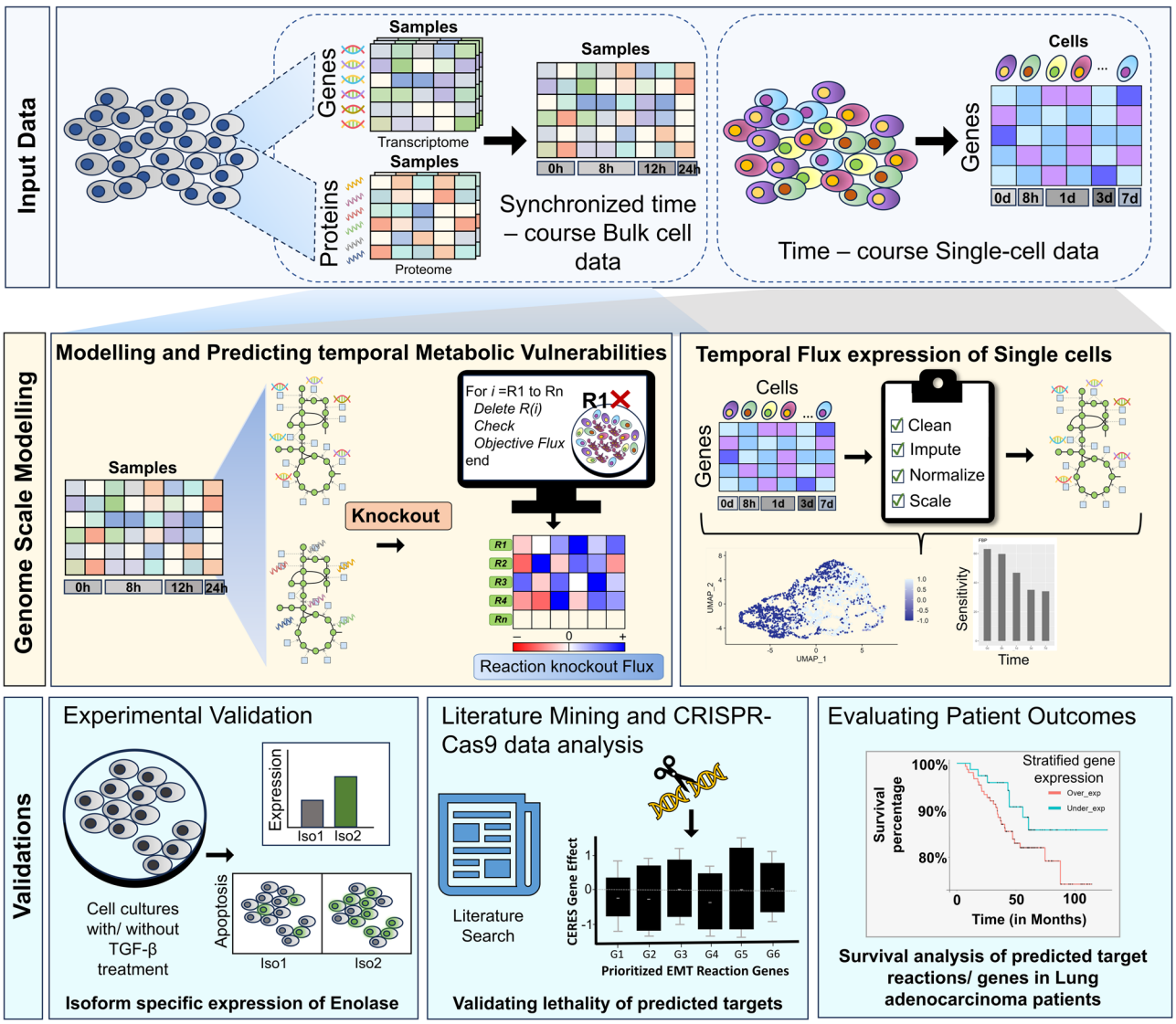


Fig. 1 | Schematic workflow of data processing and the methodologies used in this study.

We simulated the metabolic fluxes for each time-point using the transcriptomics and proteomics data to see how metabolic activity changes over time during EMT using a modified form of the iMAT algorithm[23,24] ("Methods"). This approach uses a linear optimization procedure with biological constraints, such as knowledge of the metabolic network structure (known as a stoichiometric (S) matrix) and expression levels as inputs to generate cell-state specific metabolic flux profiles. To account for redundancy in the output flux profiles due to the rank deficiency of the S matrix[11], Parsimonious Flux Balance Analysis or pFBA is further applied to provide a unique flux distribution by assuming optimal enzyme efficiency by minimizing the overall metabolic flux throughout the metabolic network[25]. Overall, this approach outputs the smallest (parsimonious) set of active reactions that best fit the transcriptomic or proteomic constraints while optimizing for a metabolic objective. We used the common optimization objective of maximizing biomass, as cancer cells typically optimize their metabolism for biomass production. Further, if cells reprogram their metabolism to optimize for other processes, this will result in a reduction of biomass production in our model and suggest alternative objectives.

Using genes and proteins that change over the time course of EMT as input, metabolic fluxes were determined. Our predictions reveal that there are more active reactions during the early and late phases of EMT. During the intermediate stages of EMT, metabolic activity goes down. As cells undergo dramatic structural rearrangements when transitioning to a mesenchymal cell, cells require energy substrates such as ATP to facilitate these processes. Our metabolic model assumed that these cancer cells were optimizing for increased biomass production, and the reduction of fluxes for biomass production during the intermediate EMT stages suggests that metabolic activity is being siphoned towards other processes such as motility. Our metabolic flux profile data suggests that cancer cells upregulate metabolism initially to build up metabolic substrate levels, and then divert metabolic resources towards other processes.

Samples within these time-points tend to have similar metabolic functions, as most active reactions are found within central carbon metabolism (glycolysis/gluconeogenesis, pentose phosphate pathway, folate metabolism) and nutrient exchange subsystems. These metabolic pathways contribute to biomass formation. We visualized the top 50 reactions sorted by prioritization scores ("Methods"; Supplementary Fig. 2; Supplementary Data 1, Eq. 2). The prioritization score takes into consideration the number of studies where a given metabolic gene(s) encoding a reaction was determined to be significant, and the absolute value of the gene effect size (log_2 fold change or Z-score). Developing a prioritization score enabled us to filter through 3744 reactions to provide a concise reaction list for downstream analyses.

Several glycolytic reactions were predicted to have increased metabolic activity and prioritization scores (Fig. 2A), which was expected given how cancer cells rewire glycolytic activity, as evidenced by the Warburg effect[5]. Several glycolytic substrates play a role in both cellular survival and cancer proliferation. It has been well established that TGF-β increases expression of several glycolytic enzymes[7]. We found that hexokinase, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and ENO were the top 3 glycolytic reactions that were highly active in both early and late EMT stages, supporting previous studies that suggest glycolysis is directly impacted by TGF-β induction. The timing of metabolic activity suggests that glycolysis is essential for initiating EMT and establishing metastasis at later stages. In contrast, flux through the citric acid cycle (Fig. 2B) and oxidative phosphorylation pathways (Fig. 2C) remained uniformly distributed throughout all the timepoints with lower prioritization scores compared to glycolysis. In order to better understand the differences in flux distribution patterns across these pathways, and the stage-specific metabolic dependence of EMT, we next performed a knockout sensitivity analysis.

Genome-scale reaction knockout simulation predicts dependence on alternate energy metabolism during intermediate EM state

While our previous analysis focused on reaction fluxes, we next used Flux Balance Analysis (FBA) to simulate the impact of reaction knockout on cellular biomass production in each time-points across four independent A549 TGF-β induced EMT studies (Supplementary Fig. 3; Supplementary Data 2). We analyzed the distribution of knockouts across bulk experiments and by time using a sensitivity score (Eq. 3). The later stage of EMT were predicted to have more vulnerabilities (932 reactions) than in the early stage (874 reactions) and intermediate stage (660 reactions), suggesting mesenchymal cells are more vulnerable to metabolic perturbation (Fig. 3; *inset*).

We predicted 74 reactions that were selectively sensitive in specific EMT stages and studies (Supplementary Fig. 4). We show the top 25 of these reactions in Fig. 3 and discuss their relevance in context to EMT in this section. These also highlight the technical and biological variance that is observed in EMT studies across different omics modes. Notably, AKG transport between the cytosol and mitochondria was unique to the intermediate stage in two out of five independent studies. Mechanistically, AKG likely suppresses metastasis by counteracting the effects of other oncometabolites such as 2-hydroxyglutarate, succinate, and fumarate[26]. Our computational model suggests that knockout of AKG transport between the cytosol and mitochondria negatively affects EMT, providing clues about cellular compartment dynamics and their impact on cancer metastasis.

Different phases of glycolysis showed varying sensitivities at different timepoints. Reactions from the initial energy-investment phase, namely triose-phosphate isomerase (TPI), fructose-bisphosphate aldolase (FBA (enzyme)) and fructose bisphosphatase were sensitive 1 h after TGF-β induction in two of the studies. On contrary, reactions from the pay-off phase of glycolysis, namely ENO, phosphoglycerate mutase and lactate dehydrogenase (LDH_L), were predicted to be sensitive 72 h post TGF-β induction, according to EMT proteomics data[21]. Upregulated ENO levels are associated with promoting cell growth, migration, and invasion during EMT in various cancers[27,28]. Similarly, LDH is strongly associated with cancer metastasis, and has been shown to activate EMT during metastasis in lung cancer and other cancers[29,30]. Our modeling approach shows the dynamic regulation of glycolysis and essentiality of different glycolytic reactions during early and late stages of EMT, providing novel insights into the time-dependent sensitivity of these well-known targets.

At the intermediate EMT stage (~24 h), knockouts that impacted biomass involved key enzymes in the citric acid cycle and fatty acid metabolism. The escalated impact of several reactions in citric acid cycle and fatty acid metabolism reactions during late epithelial and intermediate EMT stages (Fig. 3) highlight a partial switching of metabolic dependencies from glucose to alternate energy sources. This switch likely supports the transformation from proliferative to invasive cellular objectives in the mesenchymal state. Despite this metabolic adjustment, the cells continued to show glucose dependence throughout all the timepoints. Glucose exchange reaction and hexokinase were projected to be sensitive throughout all time points and experiments (Supplementary Fig. 3), indicating that cells in all EMT stages are responsive to perturbations in glucose. These results are consistent with prior studies that describes how glucose, fatty acids, and glutamine support metastasizing lung cancer[31,32], and facilitate migration and invasion processes in EMT across various cancer types[33,34].

Fatty acyl-CoA synthase (FASN) has emerged as a promising therapeutic target for NSCLC, as its inhibition is shown to trigger apoptosis and suppress cell proliferation in cancer models[35,36]. Additionally, the heightened sensitivity to the L-glutamine exchange reaction observed at the intermediate stage of EMT (24 h) (Supplementary Fig. 3) emphasizes the shift towards alternative energy sources required for the progression of the EMT. Glutamine metabolism is essential for sustaining proliferation in multiple tumor lineages, including NSCLC, and its dysregulation is a hallmark of cancer metabolism[37]. It is also important in driving metastatic phenotypes in lung cancer[38]. To better understand the metabolic consequences of EMT cells' sensitivity to L-glutamine, we investigated flux patterns across glutamine metabolic reactions (Supplementary Fig. 5). We observed that at the 24-h mark, the enzyme glutamine synthetase, which converts glutamate to glutamine, showed no significant activity in any dataset (Supplementary

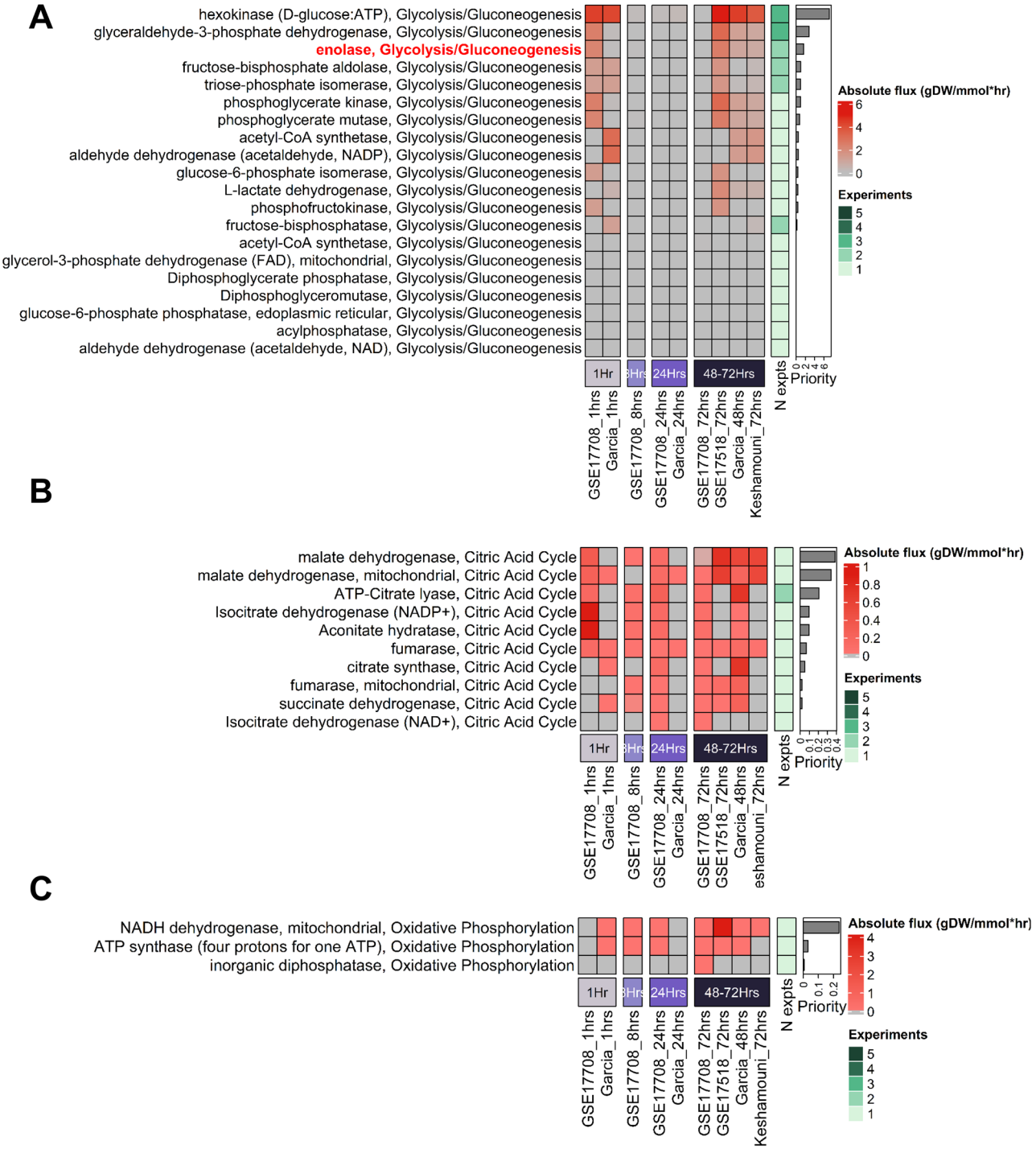


Fig. 2 | A549 metabolism is predicted to be overactive during the early and late phases of EMT. A Reactions are sorted based on the prioritization score, which is a function of the number of studies with significant metabolic genes that encode the reaction and the effect size (Z-score or log2 fold change). The metabolic fluxes were simulated using the RECON1 human metabolic reconstruction. Several metabolic reactions within the Glycolysis/Gluconeogenesis metabolic subsystem are overactive in the earlier stages (1 h) and late stages (48–72 h) of EMT, based on the absolute value of the metabolic fluxes predicted by constraint-based modeling. The top 5 reactions in the Glycolysis/Gluconeogenesis subsystem have at least two studies supporting the flux predictions. Enolase is bolded as it was prioritized for experimental validation. **B** In contrast to glycolysis, which shows high activity in early and late stages of EMT, reactions within the Citric Acid Cycle are predicted to have more uniform activity, i.e., it is not over or under active at the earlier or late time point of EMT, across all time points relative to control (unconstrained flux distribution) and have lower prioritization scores compared to glycolysis. **C** Metabolic reactions within the Oxidative Phosphorylation metabolic subsystem are also predicted to have more uniform activity relative to control across all time points.

Fig. 5). While glutaminase, that metabolizes glutamine to glutamate, and glutamate dehydrogenase has flux distributed through all the timepoints. This indicates that during the intermediate EM stage, the cells depend on external sources of glutamine to replenish the intracellular glutamine/glutamate pool and thereby maintain the reactions that support the metabolic objective of the cells.

A range of amino acids also show differential sensitivity during different phases of EMT. In the early phase, EMT cells are particularly sensitive

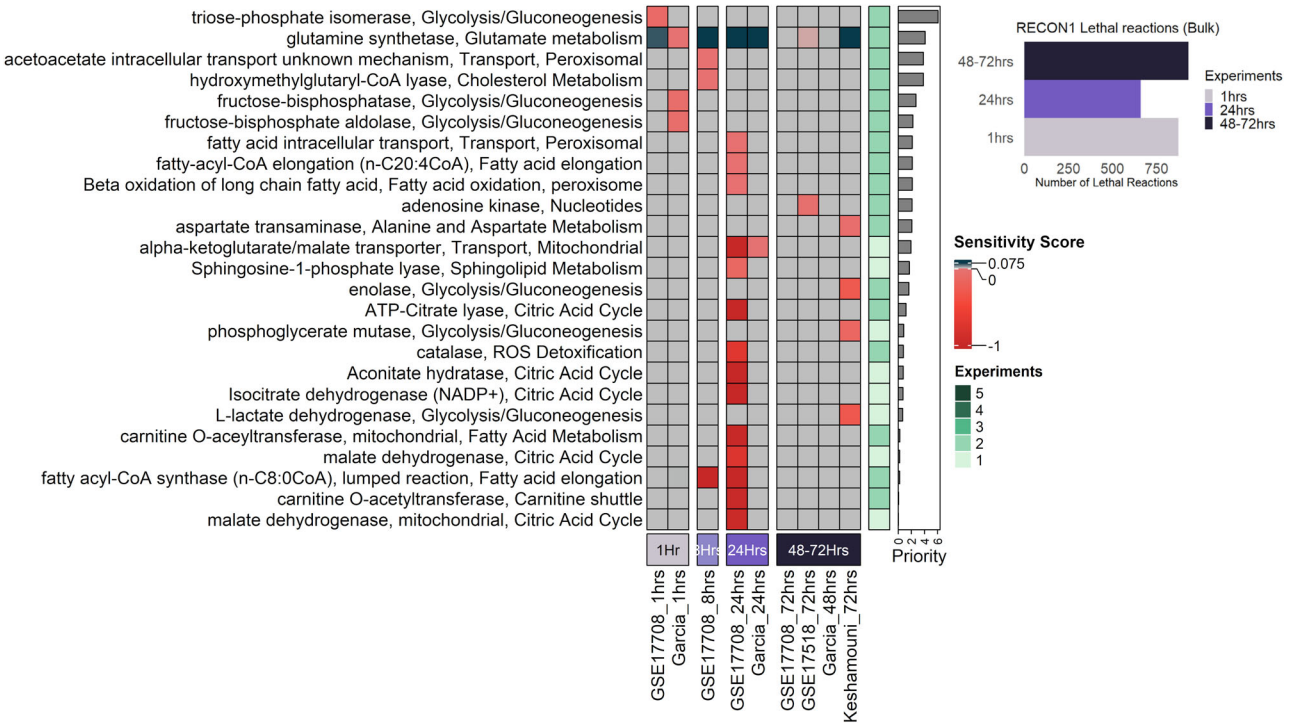


Fig. 3 | A heatmap visualization of the sensitivity of metabolic reactions to a specific EMT stage. The sensitivity of the reactions is determined by the reduction in the normalized KO Biomass flux, referred as Sensitivity Score. At any timepoint, the reactions showing a sensitivity score closer to a value of −1 have the highest sensitivity as compared to simulated reaction knockout in the control (an unconstrained metabolic reconstruction growth rate biomass flux upon reaction knockout in an unconstrained model, see "Methods"). Reactions are sorted based on prioritization scores. The bar plots (inset) show the distribution of sensitive reactions (sensitivity score < 0) for three timepoints 1 h, 24 h, and 48–72 h after TGF-**β** induction across all experiments in 3 time points. The red color in the heatmap represents a reduction in the Biomass flux upon reaction knockout, as compared to the control (sensitivity score < 0). The gray to dark gray colors represents an increase in the Biomass flux (sensitivity score > 0).

to changes in histidine and cysteine metabolism, whereas in the late phase, the metabolism of lysine, alanine, aspartate, as well as branched-chain amino acids like valine, leucine, and isoleucine, shows heightened sensitivity. The increased sensitivity of sphingosine-1-phosphate lyase is notable, given that, sphingosine-1-phosphate facilitates TGF-β induced EMT in A549 cells by helping acquire fibroblast-like features with upregulated mesenchymal and downregulated epithelial markers[39].

We thus summarize our observations by underscoring key dynamic metabolic shifts during EMT. AKG transport gained prominence at the intermediate EMT stage. There was a transition from the energy-investment phase of glycolysis at early stages of EMT, to the pay-off phase in later stages including ENO, phosphoglucomutase, and lactate dehydrogenase—a set of enzymes linked to cancer cell invasion and proliferation. There was a notable shift from glucose to a reliance on citric acid cycle, fatty acid metabolism, and external glutamine sources at later stages, possibly to support the cellular transition toward invasion. Although, this metabolic shift still resulted in a consistent glucose metabolic sensitivity in EMT cells at all stages. L-glutamine uptake became critical, indicating external sources are essential for metabolic maintenance during the intermediate stage. Variability in amino acid metabolism sensitivity through different stages of EMT underscored the adaptability of nutrient dependencies during the transition. The knockout sensitivity analysis, thus, highlighted nutrient dependencies and metabolic vulnerabilities during EMT.

Isoform-specific role of enolase 3 in regulating cell survival during EMT

The glycolytic enzyme, ENO, was predicted by both our flux and gene knockout analysis to have high metabolic activity and sensitivity to knockout in EMT. During tumor progression, cancer cells must increase glucose metabolism. Owing to the hypoxic tumor microenvironments, cancer cells upregulate glycolytic enzymes, including Eno, to support anaerobic proliferation (Warburg effect). Eno is a key glycolytic enzyme that catalyzes the dehydration of 2-phosphoglycerate to phosphoenolpyruvate. It occurs as 3 isoforms, Eno1 (ubiquitously expressed in all cells), Eno2 (neuronal specific) and Eno3 (muscle specific)[40]. Our transcriptomic analysis shows that Eno3, the muscle specific isoform, which is catalytically more efficient, is 10-fold differentially expressed in cells undergoing EMT (Supplementary Fig. 7A). siRNA mediated inhibition achieved 73–97% knockdown of Eno3 and 87–93% knockdown of Eno1 at the protein level compared to untreated control, as assessed by western immunoblotting and quantifying western band intensities (Supplementary Fig. 6A). Similarly, at the mRNA level as assessed by qRT-PCR we achieved 96–98% knockdown for Eno3 and 86–91% knockdown for Eno1 compared to untreated controls (Supplementary Fig. 6B). siRNA mediated inhibition of Eno3 selectively induced apoptosis in cells undergoing EMT as assessed by Annexin-V/PI staining by flow cytometry (Fig. 4A and Supplementary Fig. 8) Eno3 siRNA at 50 nm and 100 nm concentrations induced 3-fold ($p = 0.007$) and 3.6-fold ($p = 0.001$) apoptosis over respective controls without TGF-β treatment. Interestingly, inhibition of the ubiquitously expressed isoform, Eno1, did not trigger significant apoptosis compared to untreated control (Fig. 4A). We further validated induction of apoptosis by assessing Caspase3/7 activation by fluorescence microscopy (Fig. 4B). Consistently, we also observed significant increases in Caspase3/7 activation in response to Eno3 knockdown, but not with Eno1 (Supplementary Fig. 6C). These observations suggest that EMT induces reprogramming of glycolysis to an Eno3 dependent pathway to meet the energy demands of migratory and invasive cells. In addition to migratory behavior, EMT confers multiple other functional phenotypes to cancer cells including invasion, stemness, survival, and chemoresistance. We observed that inhibition of Eno3 selectively induced apoptosis in cells undergoing EMT, demonstrating the Eno3

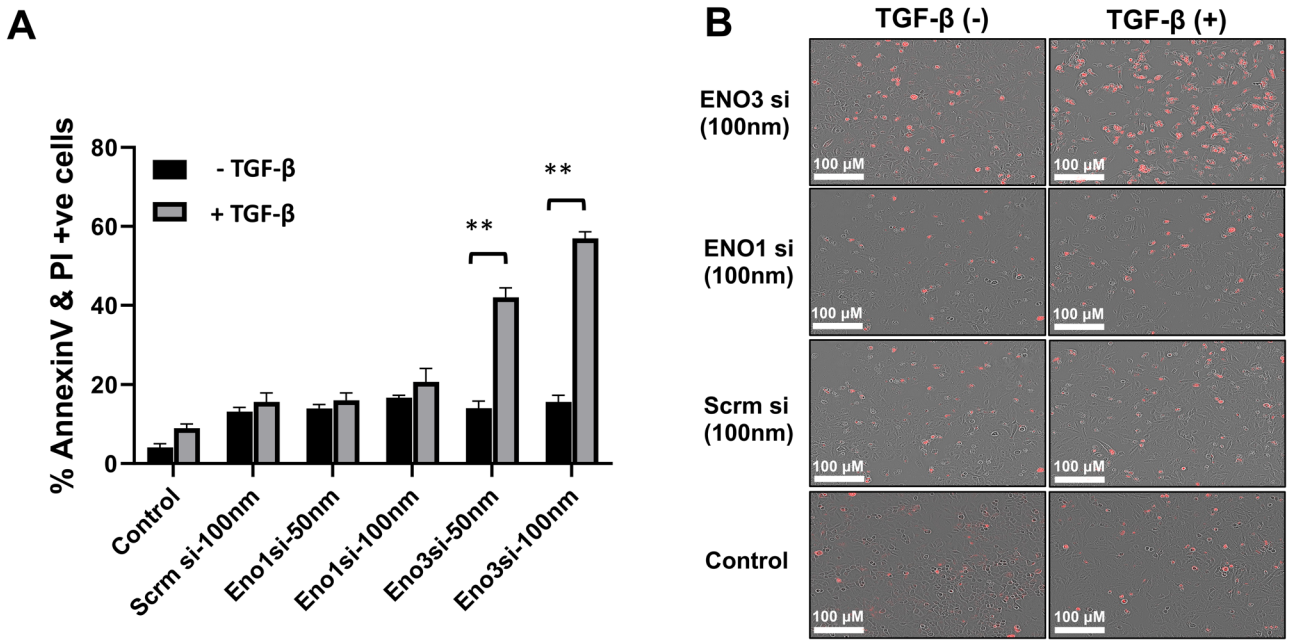


Fig. 4 | siRNA-mediated inhibition of Eno3, but not Eno1, triggers apoptosis in A549 cells undergoing EMT after 72 h TGF-β treatment. "Control" indicates no siRNA, and "Scrm" indicates scrambled siRNA. Efficiency of siRNA mediated knockdown of Eno1 and Eno3 was assessed by western immunoblotting and qRT-PCR (Supplementary Fig. 6A, B). **A** Apoptosis is assessed by staining for the percentage of AnnexinV/PI +ve cells by Flow cytometry (a representative gating strategy for 4 samples is presented in Supplementary Fig. 8). Histogram represent mean of three replicates with error bars representing standard error. Statistical significance assessed by paired *t* test * = <0.05, ** = <0.005. **B** Caspase3/7 activation is measured using a caspase3/7 specific substrate that fluoresces red after caspase3/7-mediated cleavage, cells were imaged and the number of red fluorescent cells per 96-well were counted (Supplementary Fig. 6C) using the CellCyte (Echo, San Diego, CA), a live cell imaging system. Data is representative of three independent biological replicates.

dependent survival. Given Eno3's potent effect on cell survival, any modulation of migratory behavior will be confounded by the cell death induced in the absence of Eno3, and therefore does not reflect a true effect on migration. We thus infer that, inhibition of Eno3 can selectively kill cells undergoing EMT and may prevent metastasis.

Single cell analysis highlights dynamic changes in carnitine and fatty acid metabolism across EMT

To determine the metabolic phenotypes of EMT at the single cell level, we analyzed single cell transcriptomics data of A549 cells induced with TGF-β (GSE147405[6]). To capture subtle metabolic differences as cells transition from E to M states, we reconstructed separate metabolic models for each cell based on its transcriptomic profile measured in the dataset. Single-cell fluxes were generated incorporating single-cell data in the model using the same linear iMAT algorithm as that used for bulk cell data ("Methods"). The data was preprocessed and imputed before incorporating in the model ("Methods"). We created a multimodal Seurat object using the transcriptome and the flux profiles of the single cells, which allowed paired measurements of metabolic gene expression with reaction fluxes. To ensure we were observing the transition between E to M in this dataset, we visualized the change in expression of the epithelial marker Keratinocyte 10 (KRT10) (Fig. 5B) and the mesenchymal marker vimentin (VIM) (Fig. 5C) in the UMAP embedding and found that the expression profiles are consistent with literature. Based on the UMAP distribution of TGF-β induced A549 cells at different time points (Fig. 5A), and the expression of different epithelial and mesenchymal markers, we classified the cells of Day 0 and 8 h as early stage EMT cells, Day 3 and 7 as late stage EMT, and Day 1 as intermediate EM stage cells.

Many of the reactions that were prioritized based on stage or study specific sensitivity in the bulk datasets also showed sensitivity across similar timepoints in the EMT trajectory in single cell data (Fig. 5D–G). We obtained single cell sensitivity scores for the knockout of 74 reactions that were prioritized from bulk data ("Methods"). Single-cell analysis revealed subtle changes in reaction knockout sensitivity not seen in bulk cell analysis. The bar plots along with the UMAPs show the percentage of sensitive cells for a given reaction knockout (Fig. 5, Supplementary Data 3). Reactions such as formimidoyltransferase cyclodeaminase (FTCD), FBA (enzyme), and acetoacetate intracellular transport (ACACtx) showed early-stage sensitivity in both bulk datasets (1 h and 8 h) and single cells (0 day, 8 h) (Fig. 5D–G).

We observed that the flux through the AKG transporter was higher in early stage EMT cells (Day 0 and 8 h) and highest in intermediate EM cells (Day 1) (Fig. 5E). Increased activity of this reaction may suppress metastatic transformation during the initial phases of EMT, as AKG can inhibit metastasis by counteracting the impact of other oncometabolites[26]. Furthermore, in line with our observations from bulk data analysis, we observe increased sensitivity of fatty acid metabolism reactions (ACACtx (Fig. 5D)), fatty acid intracellular transport (Fig. 5F), FASN (n-C8:0CoA) (Fig. 5E) around Day 1 in single cells. We observe similar patterns of sensitivity in reactions belonging to carnitine shuttle and metabolism (mitochondrial O-acetylcarnitine transport, mitochondrial Carnitine transferase and carnitine O-acetyltransferase) (Fig. 5E). Reactions belonging to fatty acid oxidation also show heightened sensitivity during this intermediate phase. Carnitine facilitates the intracellular transport of long chain fatty acids for subsequent β-oxidation[41]. We infer that the cells at intermediate EM state actively reprogram their energy metabolism towards alternate energy sources like fatty acids. Also, most of the citric acid cycle reactions showed increased sensitivity during this phase exemplified by isocitrate dehydrogenase (Fig. 5E). We infer that citric acid cycle intermediates serve as sources for the production of metabolites that might drive the EMT metabolism during the late stage EMT. Finally, Sphingosine-1-phosphate lyase shows increased sensitivity during Day 1 and verifies our observation made with bulk data analysis.

We also observed differences in sensitivity for hydroxymethylglutaryl-CoA lyase (HMGLx) between the bulk and the single cell data. Bulk sensitivity analysis highlights increased sensitivity for HMGLx at an early time point (8 h). However, the single cell analysis predicted maximum sensitivity during Day 1. Cholesterol metabolism plays a multifaceted role in regulating

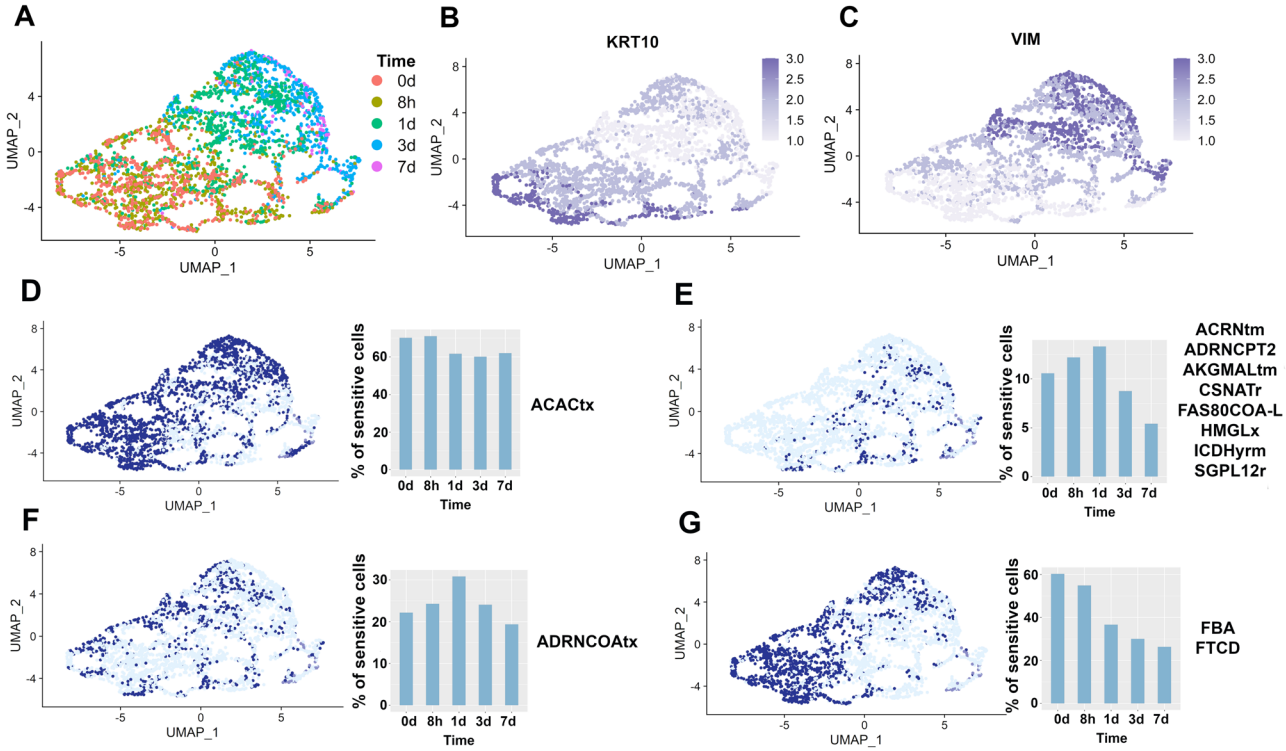


Fig. 5 | Single-cell RNASeq data of EMT biomarkers. A UMAP visualization of the temporal trajectory of A549 cells induced with TGF-β. **B** Expression profile of the epithelial marker Keranitocyte10 (KRT10). **C** Expression profile of the mesenchymal marker Vimentin (VIM). Panel figure D to O show the distribution of reaction knockout sensitive cells across the UMAP trajectory. The sensitivity scores for single cells were discretized into two categories. Sensitivity score < 0 was discretized to a value of −1 and >0–1. Dark blue depicts cells with value −1 representing that the cells had a reduced biomass compared to the control upon knockout of the reaction. The value 1 represents cells with higher biomass flux upon knockout as compared to the control. The bar plots along with the UMAPs show the percentage of sensitive cells at each timepoint. **D**, **G** Distribution of sensitive cells for the knockout of (**D**) Acetoacetate intracellular transport (ACACtx), (**E**) O-acetylcarnitine transport, mitochondrial (ACRNtm), Carnitine transferase, mitochondria (ADRNCPT2), alpha-ketoglutarate/malate transporter (AKGMALtm), carnitine O-acetyltransferase, carnitine shuttle (CSNATr), fatty acyl-CoA synthase (n-C8:0CoA), lumped reaction (FAS80COA_L), hydroxymethylglutaryl CoA lyase, peroxisome (HMGLx), Isocitrate dehydrogenase (ICDHyrm), and sphingosine-1-phosphate lyase (SGPL12r), (**F**) Fatty acid intracellular transport (ADRNCOAtx), (**G**) fructose bisphosphate aldolase (FBA (enzyme)), and Formimidoyltransferase cyclodeaminase (FTCD).

EMT across different cancers, although the precise nature of this association remains unconfirmed. Cholesterol-rich lipid rafts facilitate TGF-β-induced epithelial plasticity, while cholesterol depletion inhibits TGF-β-induced EMT[42]. Conversely, elevated plasma-membrane-associated cholesterol decreases membrane fluidity, thereby reducing cell motility and impeding mesenchymal transition[43]. Additionally, cholesterol may serve as a metabolic trigger for EMT[44]. The increased sensitivity towards cholesterol metabolism reaction during the early and intermediate EM stages may drive changes during the late phase.

We further computed the biomass fluxes of individual cells upon reaction knockout of the top 50 variable reactions in the single cells selected based on maximum standard deviation of fluxes across data (Supplementary Fig. 9). The sensitivity scores were calculated using the same approach as the bulk cell data (see "Methods"). The aggregated scores in the heatmap (Supplementary Fig. 9) reflect the average property of all the single cells at any particular timepoint. This contrasts with the UMAP visualization of the sensitive cells (Fig. 5), and hence, differences in the temporal sensitivity patterns of some reactions are expected. Most of the top variable reactions that showed sensitivity (sensitivity score < 0) belonged to glycolysis, citric acid cycle and pyruvate metabolism. Cancer cells undergoing EMT divert more glucose towards citric acid cycle as compared to epithelial cells, presumable to enhance amino acid production using citric acid cycle intermediates as backbone to enable mesenchymal transition[45]. An enhanced sensitivity was observed for the mitochondrial pyruvate transporter and pyruvate dehydrogenase, which bridge glycolysis and citric acid cycle, and citric acid cycle reactions (aconitate hydratase, isocitrate dehydrogenase (NADP+), malate dehydrogenase, and ATP citrate lyase) at Day 3 and Day 7 of EMT progression. Furthermore, hexokinase and pyruvate dehydrogenase knockouts had the highest impact on cellular biomass production at Day 3 and Day 7 (Supplementary Fig. 9). The sensitivity of acetoactyl-CoA: acetate-CoA transferase, a propanoate metabolism reaction also increases during the late phase EMT. A number of reactions from fatty acid metabolism and cholesterol metabolism also impacted the biomass production of cells during late phase EMT.

Single-cell analysis of TGF-β induced A549 EMT dataset thus revealed stage-specific metabolic reaction sensitivities, with some diverging from bulk cell observations. Stage-specific sensitivities in FTCD, FBA (enzyme), and ACACtx reactions were in early-stage EMT, while AKGMALtm flux peaked in the intermediate stage. Notably, fatty acid metabolism pathways involving acyl-CoA synthase and carnitine shuttle enzymes and the citric acid cycle were sensitive in the intermediate stage. Cholesterol metabolism, especially HMGLx, demonstrated varied sensitivity between early and intermediate phases, suggesting a potential role in EMT modulation. The analysis also highlighted a shifting dependency on glycolysis, pyruvate metabolism, and the citric acid cycle throughout EMT. Importantly, the citric acid cycle and fatty acid metabolism appear crucial for supporting the mesenchymal transition, consistent with literature.

Literature mining and CRISPR knockout screens support metabolic targets inferred using both bulk and single-cell simulations

We predicted the reactions that reduced biomass production upon knockout in bulk cell and single-cell analysis. 92 reactions intersected between both the bulk and single-cell data (Table 1, Supplementary Data 4).

Table 1 | Literature review of known and novel essential reactions predicted for NSCLC from both bulk and single-cell simulations

Reaction Name	Gene-Protein-Reaction Rule	Average NSCLC CERES Score	PMID
Strong evidence			
pyruvate carboxylase	(PC)	−0.5217	26070193
			30005601
			25607840
fructose-bisphosphate aldolase	(ALDOA) or (ALDOC) or (ALDOB)	−0.7012	32530543
			28444969
			31358528
enolase	(ENO1) or (ENO2) or (ENO3)	−0.0015	25887760
			29538538
			27996156
			28086938
			32521855
Medium evidence			
ribulose 5-phosphate 3-epimerase	(RPE)	−1.3731	32365991
triose-phosphate isomerase	(TPI1) or (TPIP2)	−1.1346	27908734
formimidoyltransferase cyclodeaminase	(FTCD)	−0.0908	30784016
			34819088
glyceraldehyde-3-phosphate dehydrogenase	(GAPDH) or (GAPDHS)	−0.7870	27878251
UTP-glucose-1-phosphate uridylyltransferase	(UGP2)	−0.2952	31243371
ATP-Citrate lyase	(ACLY)	−0.7807	23807225
acetone mitochondrial transport via proton symport	(SLC16A1)	−0.0556	31371390
aspartate transaminase	(GOT2)	−0.3341	23535601
Novel predictions			
Glutamate formimidoyltransferase	(FTCD)	−0.0908	
4-Hydroxyphenylpyruvate: oxygen oxidoreductase	(HPD)	−0.0390	
adenosylhomocysteinase	(AHCY) or (AHCYL1)	−0.6535	
Phosphatidylserine synthase	(PTDSS1)	−0.2514	

Note that glutamate formimidoyltransferase of histidine metabolism listed under the novel reactions is regulated by the gene FTCD, which also regulates formimidoyltransferase cyclodeaminase of folate metabolism, listed under reactions with medium evidence For full list of all predicted essential reactions, see Supplementary Data 4.

We performed extensive literature curation for these reactions to prioritize them for further analysis (see "Methods"). The reactions were prioritized using the prioritization score (Eq. 2) that was based on the number of studies found for each gene query (for the genes that encode these reactions) and its relevance to cancer and EMT. Three high confidence predictions that were found in both analyses included Pyruvate Carboxylase, FBA (enzyme), and ENO which were shown to contribute specifically to NSCLC progression and metastasis[46–48]. Pyruvate carboxylase (PCm) was one of the most important reactions, whose knockout significantly reduced biomass flux in EM transitioning A549 single cells (Supplementary Fig. 12). Eight reaction predictions had some evidence of being dysregulated in other cancer subtypes, but not NSCLC (Table 1). There are four reactions that have not been highlighted in the literature, and present opportunities for future experimental validation.

To further assess our model predictions against experimental data, we compared our bulk and single-cell knockout results against batch-corrected CRISPR-Cas9 essentiality knockout screens integrated from the Broad and Sanger Institute[49]. Given the limited availability of CRISPR-Cas9 screenings in EMT studies, we took NSCLC cancer cell lines from the DepMap dataset, which annotated them to be derived either from a primary tumor or a metastatic site. Further, we took EMT signatures from MSigDB and EMTome to classify cancer cell lines from the cancer cell line encyclopedia (CCLE) into epithelial-like or mesenchymal-like cell-lines (see "Methods"). When comparing the classification of CCLE cancer annotation with our EMT classification, we found that there was moderate agreement between cancer cell lines obtained from a primary site and the epithelial state and low agreement between the mesenchymal cell state and cell lines from a metastatic site (Supplementary Fig. 11).

The essentiality of metabolic enzymes derived from our model predictions were interpreted using the CRISPR gene knockout (CERES) Scores, where a lower score suggests that a given gene is essential for survival in a given cell line[50]. A score of 0 was used as the threshold to indicate the median effect of non-sensitive genes. We overlaid the CERES Scores of 14 NSCLC EMT specific metabolic genes (see "Methods") for CCLE cancer cell lines' primary and metastatic annotation and our EMT annotation (Fig. 6A, B). We found that the median of CERES score distribution for both primary vs metastatic and epithelial vs mesenchymal classification methods were similar for these genes. The AKG/malate transporter SLC25A11, and central carbon metabolism enzymes ENO1, ENO2, ENO3, IDH1, and LDH show lower median CERES scores than the threshold of 0. From this analysis, we were able to identify isoform-specific sensitivity in NSCLC, analogous to our validation of Eno3 dependency (Fig. 4). IDH1 gene depletion is associated with NSCLC essentiality, compared to IDH2 and IDH3 depletion, suggesting that targeting IDH1 expression in NSCLC may be an effective therapy to supplement existing therapeutics that target specific IDH1/2 mutations.

To evaluate how well our sensitivity score predictions performed against CRISPR-Cas9 experimental data, we took the ratio of the CERES scores for the Primary site derived NSCLC cell lines to the scores for the Metastatic site derived cell lines for each metabolic gene. We found that the sensitivity score and metabolic fluxes agreed significantly with the CERES ratios with $R = 0.31$ and 0.20 respectively, $P = 0.005$ and 0.036 (Supplementary Data 5). We determined the correlation between our predictions against all cell lines (i.e., including non-lung cell lines) as well but found that the correlations were not significant with the pan-cancer CERES ratios. These results match our expectations, as our COBRA models were constrained using A549 transcriptomics and proteomics data. We evaluated the quality of each dataset on COBRA predictions, and found that the single-cell RNASeq data best matched the CERES ratios (KO $R = 0.27$, KO $P = 0.01$; Flux $R = 0.20$, Flux $P = 0.03$; Supplementary Data 5) while bulk transcriptomics and proteomics data were weakly correlated ($R < 0.1$; $P > 0.05$; all bulk experiments).

The reduced magnitude of correlation, although statistically significant, between our predictions and experimental data can be attributed to several confounding variables. Firstly, we assumed metabolic similarities between primary tumor site cells and epithelial-like cells, and between metastatic cell lines and mesenchymal cells. To address this, we reclassified cells based on EMT gene markers, yielding similar correlations using CCLE annotation or our EMT classification (Supplementary Data 5). Secondly, our analysis used TGF-β-induced EM transitioning cell lines, contrasting with non-EMT-induced CRISPR screened cell lines. Furthermore, our simulations exclusively featured A549 cells, unlike the NSCLC CCLs utilized for evaluating statistical significance. As a result, the CERES scores failed to differentiate between ENO's three isoforms (Fig. 6B), in contrast to our siRNA knockdown experiment (Fig. 4B).

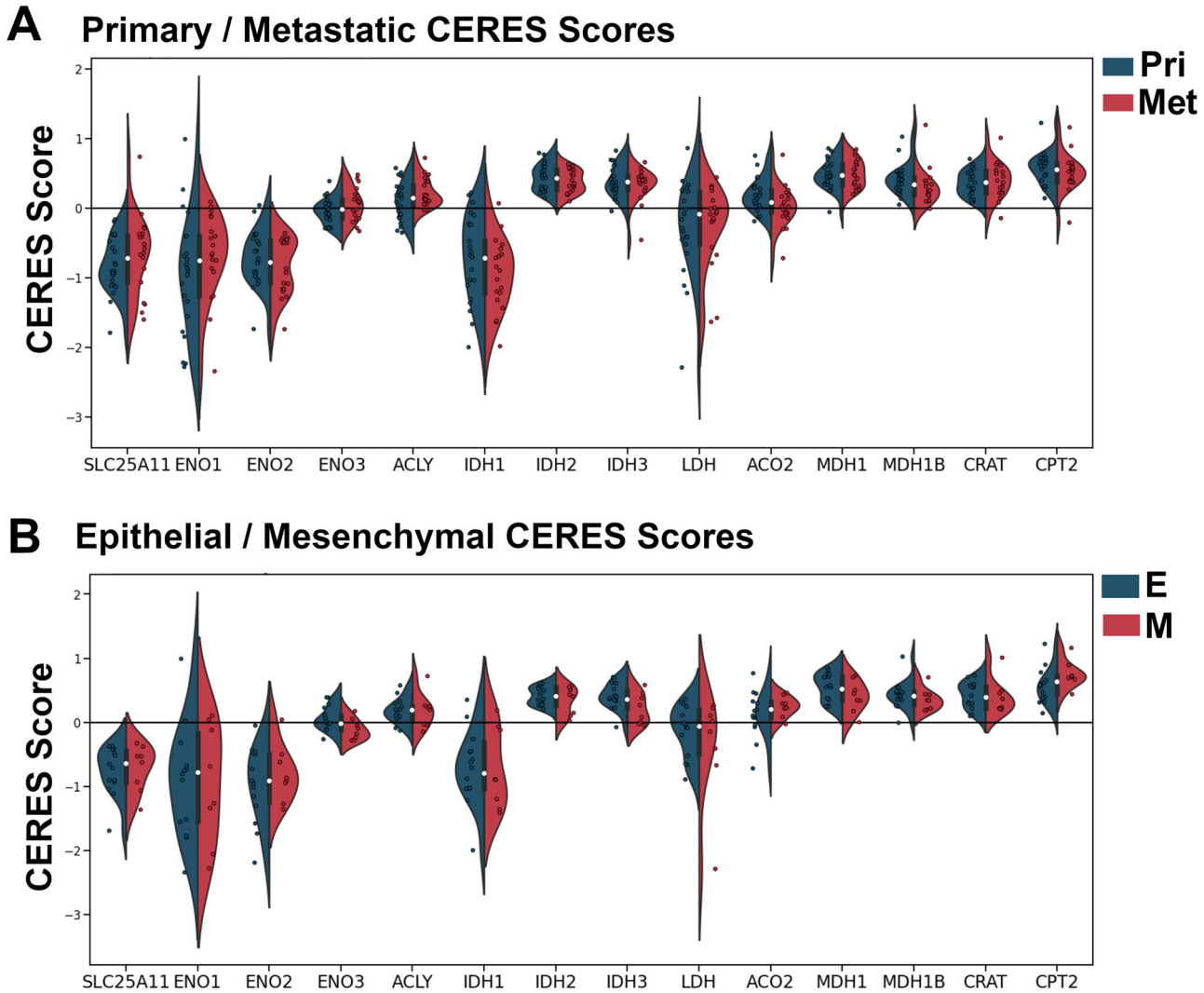


Fig. 6 | COBRA-prioritized NSCLC CERES Scores reveal metabolic vulnerabilities during EMT. A Reactions predicted by COBRA to be sensitive were compared against CERES Scores. Cell lines were classified as primary or metastatic, and their CERES score distributions are shown on the violin plots. The average predicted growth scores and fluxes through the reactions in hour 72 across all 5 experiments agreed with the Primary/Metastatic CERES Scores Ratios ($R = 0.31$ and 0.2; $P = 0.005$ and 0.035 respectively; Supplementary Data 5). These 10 metabolic genes were selected based on reactions of interest from our bulk COBRA knockout profiles and single-cell flux profiles (from Fig. 3). The colors dark blue (Pri) represents primary cell lines and red (Met) represents metastatic cell lines. The white dot inside the violin plots represent the median value of the two CERES score distribution. The analysis in (**B**) was repeated with the epithelial/mesenchymal annotations for sensitive reactions. The predicted growth scores and fluxes from the single-cell simulations (GSE147405) agreed with the Epithelial/Mesenchymal CERES Scores Ratios ($R = 0.28$, $p = 0.01$). The color dark blue (E) represents epithelial cells and red (M) represents mesenchymal cells.

Association of EMT states and metabolic reactions with patient survival outcomes

To test the relevance of metabolic reactions regulated during EMT with patient outcomes, we performed survival analysis. Gene expression profile from 226 lung adenocarcinoma (LUAD) patient samples were used to predict metabolic fluxes of each patient and compared against their corresponding survival information[51]. A set of 104 reactions prioritized through bulk and single cell EMT analysis were used to predict survival outcomes of patients. Seven of these reactions showed significant association with patient survival based on Cox regression, when their fluxes were used as variables (Supplementary Data 6). Of note, the calculated HR, CI and p values of the ENO, isocitrate dehydrogenase (ICDHxm) and triose phosphate isomerase (TPI) reactions were HR = 1.2 (1–1.4) with $p = 0.042$, HR = 1.2 (1.1–1.4) with p value 0.00073 and HR = 1.4 (1.1–1.7) with p value 0.0064 respectively. We then stratified patients based on the flux of the reactions into two groups using median z-score transformation. In line with our knockout and sensitivity analysis, we observed ENO (HR = 2.8, 95% CI = 1.4–5.5; $p = 0.0039$) and isocitrate dehydrogenase (ICDHxm) (HR = 2.3, 95% CI = 1.2–4.4; $p = 0.016$) to be significantly association with poor patient survival (Fig. 7A, B). Additionally, triose phosphate isomerase (TPI) was also weakly associated with poor survival (HR = 1.7, 95% CI = 0.87–3.4; $p = 0.12$) (Fig. 7C). To further demonstrate clinical significance, gene expression of ENO3, IDH3B and TPI1 were assessed in a primary human lung adenocarcinoma data set ($n = 2166$) available in the KM Plotter online tool[52]. Univariate Cox overall survival analysis revealed that increased gene expression of ENO3 (HR = 1.12 (1–1.26); logrank $p = 0.058$), IDH3B (HR = 1.19 (1.05–1.34); logrank $p = 0.0052$) and TPI1 (HR = 1.17 (1.04–1.32); logrank $p = 0.008$) correlated with poor patient survival (Supplementary Fig. 7). Other reactions with significant association with survival were related to pyruvate metabolism, glutamine metabolism, and aminosugar metabolism (Supplementary Data 6). Overall, this analysis indicates that prioritized EMT specific metabolic reactions are strongly associated with patient outcomes.

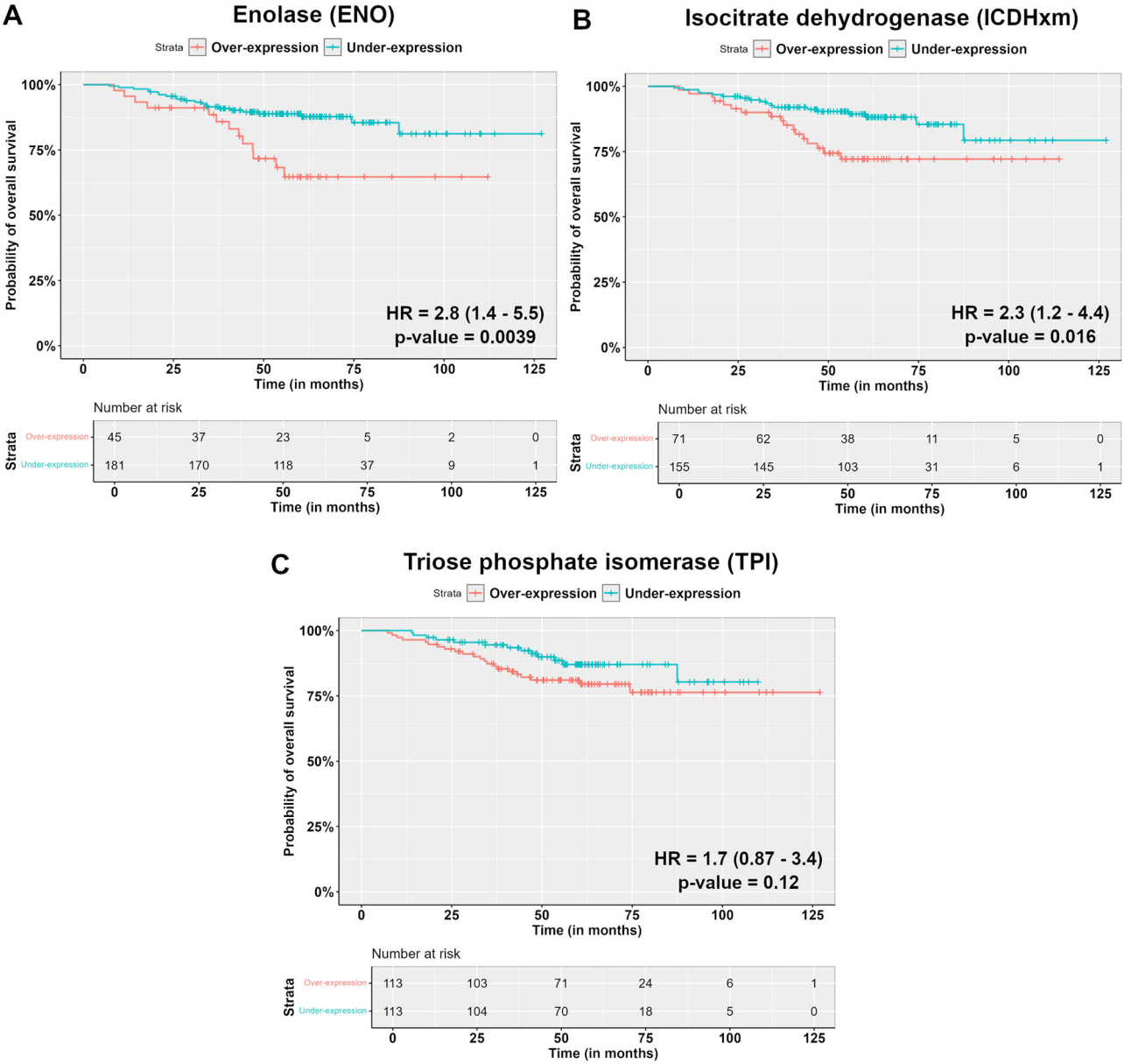


Fig. 7 | Kaplan Meier plots of metabolic reactions that show significant association with patient survival. Association of (**A**) Enolase (ENO), (**B**) Isocitrate dehydrogenase (ICDHxm), and (**C**) Triose phosphate isomerase (TPI) with patient survival. The stratification of patients was done using the median z-score of the flux values.

Discussion

We utilize constraint-based modeling informed by multiple omics data sources to predict metabolic activity and knockout sensitivity during EMT. Our predictions are supported from literature validation, siRNA knockout studies, and CRISPR-Cas9 essentiality screens. We further provide a list of high confidence metabolic reaction dependencies during EMT for future experimental validation. The therapeutic promise of targeting cellular metabolism lies in the pharmacological tractability of metabolic enzymes, as exemplified by existing small molecule inhibitors for ENO[53] and IDH[54]—two of the prioritized targets in our research. Furthermore, our approach also provides insights into metabolic activity at the single-cell level, which is not possible to infer with current experimental methodologies.

EMT leads to the transformation of cells from a proliferation-focused phenotype in the epithelial stage to a motile and invasive phenotype during mesenchymal transformation. Yet, the precise metabolic state of mesenchymal cells remains elusive. By utilizing time-course EMT data, we created distinct steady state models for each time point, enabling us to capture the dynamics of various EMT transition states. We specifically chose biomass as the objective function, assuming that, ideally the cancer cells would optimize for increasing biomass production, unless there is a requirement to reprogram the metabolism to other processes like motility, in which case, the biomass production would reduce. This is indeed what we observed during the mesenchymal stages of EMT, where biomass flux was lower compared to other time points. Through our knockout sensitivity analysis, we pinpointed reactions contributing to this biomass reduction. By integrating insights from both bulk and single-cell EMT analysis, we inferred reactions in glycolysis, pyruvate metabolism, citric acid cycle, and fatty acid metabolism, significantly impact biomass production in intermediate and mesenchymal states. Our modeling thus predicted reactions that could serve as objectives for the mesenchymal state.

Our analysis predicted metabolic enzymes that are novel as well as those with experimental evidence in literature supporting their role in tumor progression, such as glucose and glutamine transport and fatty acid metabolism. We found that glycolytic reactions were differentially overactive in the early and late stages of EMT. We further experimentally validate the essential role of the ENO reaction in EMT. The ENO enzyme is implicated

in cancer progression for various tissue lineages, but so far has not been identified as a crucial player in NSCLC metastasis. Our modeling predicted reaction catalyzed by ENO as highly active during the late stages of EMT and ENO knockout to have a negative impact on cellular biomass production. ENO has three isoforms. Eno1 is ubiquitously expressed in all cells, Eno2 is neuronal specific, and Eno3 is a muscle specific isoform. In our transcriptomic data sets, we observed expression of Eno1 and Eno3, but not Eno2. Even though our modeling did not distinguish between isoforms, we were able to experimentally demonstrate an isoform specific function for Eno3 in cell survival during EMT. This is consistent with the kinetically more active muscle specific Eno3 regulating energy-intensive migratory behavior of cancer cells. This observation also fits with the broader trend between catalytic activity and enzyme over-expression observed across various cancers[8]. The modeling observation related to ENO was further supported by the association of high ENO reaction flux levels with low LUAD patient survival (Fig. 7A). Using LUAD datasets available in KM Plotter, the hazard ratio (HR) for overexpression of both Eno1 and Eno3 genes were statistically significant with Eno1 showing higher HR (1.35, $p = 9 \times 10^{-7}$) compared to Eno3 (1.12, $p = 0.058$) (Supplementary Fig. 7). However, the data does not specifically differentiate metastatic patients, which would provide a more accurate context for assessing the effects of overexpressing these two isoforms.

Single-cell transcriptome and flux analyses of TGF-β induced A549 cells showed variation in fatty acid metabolism, citric acid cycle, and pyruvate metabolism. The knockout analysis showed the importance of pyruvate dehydrogenase during the late phase of EMT (Day 3 and 7). The serine-threonine kinase, pyruvate dehydrogenase kinase (PDK4) actively regulates EMT via phosphorylation and inhibition of pyruvate dehydrogenase (PDH)[55]. In breast cancer, TGF-β induced mesenchymal cells show reduced expression of PDK4 that results in enhanced PDH activity diverting the glucose flux towards Citric acid cycle[45]. Due to unavailable gene expression profile, we were unable to trace the expression of PDK4 across the single cell UMAP embedding. However, by visualizing the gene expression of Estrogen–related receptor α (ESRRA)—a key activator of PDK4[56], we could infer the possibility of higher PDK4 expression during early stage of EMT (Supplementary Fig. 10). We further inferred from our analyses that the reactions belonging to fatty acid synthesis and elongation are prominently expressed during the early stages of EMT, while the reactions driving propanoate metabolism heighted heightened flux during the later stages. This aligns with the finding that TGF-β treatment of cancer cells induce propanoate metabolism which induces EMT and acquisition of pro-aggressive traits via methylmalonic acid[57].

Our model predictions exhibited a significant correlation with CRISPR knockout gene essentiality scores from cancer cell lines. Remarkably, single-cell knockout simulations displayed a stronger correlation with CRISPR-Cas9 gene knockout essentiality data than models derived from bulk omics data, revealing the single-cell models' improved capacity to capture transient vulnerabilities during EMT. However, some confounding factors influenced the correlation between knockout analysis and CRISPR-Cas9 CERES scores. Notably, our metabolic models used EMT datasets induced by TGF-**β**, while the CRISPR screen used non-EMT-induced cancer cell lines. Our analysis focused only on A549 cells, differing from diverse NSCLC lines in the CRISPR-Cas9 data. Despite this limitation, A549 cells serve as a robust in-vitro model for in-vivo metastasis for EMT. Multiple groups including ours utilized A549 model system to gain therapeutic insights[58] and gene signatures that are prognostic to lung adenocarcinoma[59].

The reactions prioritized by our model are strong candidates for drug development because they reduce cellular biomass production in cells from later timepoints (mesenchymal/ metastatic-like) relative to earlier ones (epithelial/benign-like). Reactions, such as glutaminase (FTCD), 4-hydroxyphenylpyruvate oxidoreductase (HPD), adenosyl homocysteinase (AHCY), and phosphatidylserine synthase (PTDSS1), predicted by our model lacked literature support but yielded negative CERES scores, suggesting that they may be novel drug targets. In addition, our model predicted ATP-Citrate lyase (ACLY) to be sensitive in intermediate EM stage cells and essential in the CERES score analysis. ACLY has been implicated as a crucial metabolic enzyme that facilitates cancer progression and its upregulation is associated with poor patient prognosis[60]. Survival analysis of the LUAD patients with respect to differentially expressed reaction fluxes showed that the glycolysis/gluconeogenesis reactions ENO, triose phosphate isomerase, and lactaldehyde dehydrogenase, and the citric acid cycle reaction isocitrate dehydrogenase were significantly associated with patient survival. In summary, our computational model accurately captures metabolic activity and gene essentiality during EMT. Overall, our computational modeling approach holds the potential for broader application across a diverse array of diseases.

Methods

Differential expression (bulk studies)

We analyzed two transcriptomics[19,20] and two proteomics[21,22] EMT time-course studies with A549 as the cell model undergoing TGF-β induction. We looked for the modulation of core EMT markers across the datasets and calculated their EMT scores using methods described in section '*Patient stratification using EMT scores and reaction fluxes*'. EMT scores consistently increased across the time points, indicating progressive EMT in the datasets (Supplementary Fig. 13A–D). All studies compared later time point after TGF-β induction over day 0 to obtain differentially expressed genes and proteins. When possible, authors' methods and provided datasets were used to obtain a list of up- and downregulated gene sets. If no preprocessed data was provided (as in the case of GSE17518), *limma-voom*[61] was performed to determine differentially expressed genes between conditions. Additionally, a GAM-LOESS model was used to determine differentially expressed genes in GSE147405[6], aggregating single-cells at the time-course level. The regression coefficients from the GAM-LOESS model were used to determine the sign of regulation (up/down). *P* values from *limma-voom* and a GAM-LOESS model were adjusted using the Benjamini-Hochberg method and the significance threshold used was $P < 0.05$. The expression matrix containing statistically significant normalized scores for all metabolic genes across all 5 experiments can be found in Supplementary Data 7.

Single cell differential expression

We computed differentially expressed genes for individual cells without TGF-β removal in GSE147405 to simulate individual cell fluxes and reaction knockout biomass fluxes. Data preprocessing included data scaling, removing contaminant artifacts such as cells with more that 10% of mitochondrial genes, and removing cells with total gene counts less than 500. This was performed on the raw data object. Further, we used the data imputation algorithm Markov Affinity-based Graph Imputation of Cells (MAGIC)[62] to fill in drop out values. The MAGIC-imputed data was transformed to a Z-score using a Z-score method that subtracts out the median and centers the data based on the median absolute deviation (MAD). The formula for the robust Z-score for a specific gene *i* in a given cell *j* is shown in Eq. 1:

$$Z_{i,j} = \frac{x_{i,j} - median(x_i)}{MAD} \quad (1)$$

For any given cell, a gene was determined to be upregulated if the robust z-score >1 and downregulated if z-score <1.

Prioritizing metabolic gene targets across multiple studies

We evaluated the robustness metabolic gene dysregulation across five EMT studies using the following prioritization score η. The method to compute η shown in Eq. 2:

$$\eta_i = M \sum_{i=1}^{N} abs(E_i) \quad (2)$$

Where M is the number of studies where the gene was determined to be significant and E_i is the gene effect size ($\log_2$ fold change or Z-score) for gene

i. The prioritization scores were ranked in descending order and used to prioritize reactions for further investigation.

COnstraint-Based Reconstruction and Analysis (COBRA)

We use the RECON1 model with 3744 reactions, 1905 genes and 2771 metabolites to build EMT specific metabolic models[63]. Similar results were obtained in terms of reactions and subsystems when we compared results from RECON1 with a more recent version of the Human GEM (Human1 v1.4.0)[64]. Specifically, the 200 most variable reactions in each model across all the bulk cell datasets were compared. Subsystem-wise classification highlighted common (68%) and unique (32%) pathways in both the models (Supplementary Data 8). Further analysis of the datasets using both the models revealed dynamic changes in common pathways. As such, we used the RECON1 model for further analysis as it is more widely used in literature.

To incorporate the differentially expressed genes and proteins from the time-course bulk and single-cell datasets[6,19–22] into the metabolic model, a linear version of the iMAT algorithm was used[23,24,65]. In this approach, the z-score transformed transcriptome was fed into the human metabolic model, RECON1. Using the z-score cut-off of 1 and −1, reactions associated with upregulated genes are assigned higher flux bounds while reactions associated with downregulated genes are assigned lower bounds. Downregulated reactions were not removed from the model but rather the flux through those reactions were minimized[24,66]. Note that, differential expression, rather than absolute expression, was used as we were interested in variation between stages of EMT. Furthermore, the use of diverse datatypes, including microarrays, RNA-sequencing and relative proteomics, prevents meaningful comparison of absolute levels of transcripts and proteins. To capture the dynamics of EMT, we create steady state metabolic models corresponding to specific time points measured in the omics data. Each model consisted of 3744 reactions, 1905 genes and 2771 metabolites. The number of upregulated and downregulated genes changed for each model, depending on the number of genes that qualified the z-score threshold. Hence, the flux profile of each sample/single-cell was unique and determined by the differential gene expressed at each time-point in each study. The model can be used to include addition omics data, and the future availability of dynamic metabolomics data can enable the construction of a fully dynamic FBA model[67].

Single gene and reaction knockout simulations were obtained by systematically shutting off (upper and lower bounds were set to 0) each metabolic reaction to simulate a "knockout", while the biomass objective was optimized using Flux balance analysis (FBA)[11] (Supplementary Data 4). Biomass is a comprehensive measure of cellular proliferation and growth, encompassing various macromolecular components such as carbohydrates, amino acids, nucleotides, lipids, and cofactors. It quantifies the overall production of cellular material necessary for cell division and proliferation. We further applied parsimonious enzyme usage (pFBA)[25] to obtain a unique metabolic flux distribution and to minimize fluxes that did not contribute to biomass formation. Since numerous flux solutions are feasible in FBA, the use of both reaction knockouts and pFBA allowed us to prioritize the most significant reactions that show both differential activity and vulnerability during EMT as drug targets. Although it is unclear if cells undergoing EMT are metabolically efficient and solely relying on the pFBA assumption might lead to imprecise results. Hence, we also performed reaction knockout analysis, which does not depend on this assumption. Intersecting reactions that appear in both knockout and flux analysis helped us prioritize the most promising leads.

Differential metabolic activity and knockout sensitivity analysis

To determine differentially active metabolic reactions, we used the prioritization score described in Eq. 2 on the absolute value of the metabolic fluxes. Most reactions show zero flux, and so reactions that showed metabolic activity were considered to be overactive metabolic reactions.

To determine the impact metabolic genes and reactions on the biomass production during different stages of EMT, we computed a sensitivity score θ comparing the reduction in EMT versus control biomass flux for each gene/reaction knockout. The biomass flux generated upon knockout in an unconstrained model was used as the control. The equation to compute the bulk sensitivity score is shown in Eq. 3:

$$\theta_i = \text{mean}\left(\frac{\text{cancer}_{i,j}/\text{mean}(\text{cancer}_j)}{\text{control}_i/\text{mean}(\text{control})}\right) - 1 \quad (3)$$

where i is the gene/reaction and j is the sample.

The KO biomass flux was mean-normalized for the control, and the TGF-β treated time-course samples. The final score was taken as the ratio of the mean normalized TGF-β treated KO biomass flux over the mean normalized control KO biomass flux for a given reaction knockout. This score was calculated across all cells to identify differentially sensitive metabolic genes/reactions. The score was centered at 0. The rationale behind a zero centered θ is as follows: if the sensitivity score is 0, the gene/reaction knockout has no difference between cancer and control. If the score is less than 0, the knockout impacts the cancer cell more than control, and is considered to be essential for biomass production. To rank and prioritize metabolic targets for experimental validation, we used the same prioritization score as we did to rank differentially expressed genes (Eq. 2).

Classifying cancer cell lines based on EMT

Cancer cell lines from the CCLE were annotated by their source from a primary tumor or metastatic tumor. To also classify whether a cell was in the epithelial versus mesenchymal state, we calculated the Z-score and *p* value for all genes in the CCLE and mapped them to known EMT markers. Upregulated genes were classified as having a positive Z-score and a significance threshold of $p < 0.05$, while downregulated genes were classified as having a negative Z-score with the same significance threshold.

EMT markers (with up- and downregulated signatures) were taken from MSigDB[68] across three studies from different tissues of origin induced using TGF-β. We further filtered this list with NSCLC markers without up/downregulation annotations from EMTome[69]. The final number of markers obtained for EMT was 14 genes, which were used to classify cancer cell lines (CCLs). We chose to classify cell lines using upper and lower quantiles of gene markers. Up- and downregulated genes from NSCLC CCLs were cross referenced to the list of EMT signatures and labeled as E if the number of signatures was less than 5 genes or M if the number of signatures was greater than 8.

CRISPR-Cas9 analysis

We analyzed batch corrected CERES Scores[49] for metabolic genes that were predicted by COBRA to have increased metabolic activity or resulted in a reduction of biomass flux from knockout. CERES Scores were separated based on their association with metastatic (Met) and primary (Prim) cell lines, which was determined based on the CCLE metadata[70]. We also compared the CCLE annotations against our own Epithelial (E) and Mesenchymal (M) annotations, methods described above.

To evaluate how well our model predictions related to CERES Scores, we calculated the Pearson correlation coefficient between our predicted sensitivity scores and ratios of M / E and Met / Prim CERES Scores. Further, we compared the sensitivity scores against different subsets of the CERES Score data, including NSCLC only cell lines and all cell lines. Given the large number of observations, the correlations obtained were highly significant statistically, despite being of low magnitude. The *p* values were calculated by MATLAB based on a t statistic:

$$t = r\sqrt{\frac{n-2}{1-r^2}} \quad (4)$$

where r is the correlation and n is the number of observations.

Survival analysis

Gene expression profile of 226 LUAD patients with survival status was used to generate whole genome reaction fluxes and EMT scores[51]. Reactions fluxes for each of the samples were obtained by incorporating differential metabolic gene expression profiles in RECON1 model, as described in the section 'COnstraint-Based Reconstruction and Analysis (COBRA)'. The upregulation or downregulation of reaction fluxes across samples was calculated using the Z-score distribution of the flux data (Fig. 7A–C). We used a median z-score cut-off to define differential flux expression. Patient samples with z-scores of a reaction greater than the median z-score represented over-expressed category and those with z-scores less than or equal to the median value represented under-expressed category. Kaplan Meier survival analysis was performed using the R package "survival-3.5.5" and plotted using "ggplot2". Hazard ratios of top differential metabolic reactions were calculated using a univariate Cox proportional hazards regression model, fitted with the survival time and survival status of the patient as the dependent variable. The beta coefficients of regression, hazard ratios within 95% confidence interval (CI), Wald test statistics, and their level of significance were calculated. Reactions with a Wald test statistics > 3 and $p < 0.05$ were considered to significantly affect patient survival. We performed the Cox regression analysis for the flux data with (Fig. 7) and without stratification (i.e., regression) (Supplementary Data 6)

To demonstrate the clinical significance of these prioritized reactions with respect to patient survival, we further validated the association of high vs low expression of genes regulating these metabolic reactions using a more comprehensive data available at KM Plotter[52]. We used the lung adenocarcinoma patient cohort with 2166 patient data available in KM Plotter as of May 28, 2024 and used "median" as a measure to categorize high vs low gene expression.

Calculation of EMT scores

EMT scores were calculated using method described by Tagliazucchi et al. [71], and Chae et al. [72]. The EMT scores were generated to trace EMT progression in the lung cancer datasets GSE17518 (transcriptome), GSE17708 (transcriptome), Garcia data (proteome), and Keshamouni data (proteome) used for the bulk studies (Supplementary Fig. 13). We checked the distribution EMT scores across the different timepoints in these datasets. The following set of epithelial and mesenchymal markers were used:

Epithelial markers: CDH1, DSP, CRB3, OCLN.

Mesenchymal markers: VIM, CDH2, FOXC2, SNAI1, SNAI2, TWIST1, FN1, ITGB6, MMP2, MMP3, MMP9, SOX10, GSC, ZEB1, ZEB2, TWIST2.

To compute the EMT scores of the samples, the average z-score transformed expression of mesenchymal marker genes were subtracted from the average z-score transformed epithelial genes' expression. The higher the score, the more mesenchymal the sample is. We analyzed the distribution of EMT scores for samples belonging to each timepoint in the datasets to verify that the datasets indicated progressive epithelial-mesenchymal transition.

Identifying metabolic enzymes and EMT studies for systematic literature validation

To be considered for our systematic literature validation, we pooled a list of metabolic enzymes predicted from COBRA from bulk and single-cell reaction knockouts that had lethal reactions in at least 2 studies (sensitivity score < 0). The query was performed using PubMed and Google with the following keywords using AND filtering: "EMT", "Metabolism", "A549", "metastasis", "cancer", "cancer metabolism", and the individual gene of interest. The list of the manually curated results can be found in Supplementary Data 4.

We expanded the scope of our literature search to encompass all cancer cell lines. The query was manually curated to either support or refute the COBRA predictions. None of the model predictions contradicted the literature. Reaction predictions and their confidence were scored (1–3), where 1 has no evidence based on literature and 3 has strong A549 or lung adenocarcinoma specific evidence. The rules to assign each score for each reaction prediction are shown below:

- **1:** Prediction has no literature support.
- **2:** Prediction has literature evidence with general cancer lineages.
- **3:** Prediction has literature evidence either with specific experiments from A549 or related lung adenocarcinoma tissue/cell lines.

Cell culture, siRNA transfection, and EMT induction

A549 human lung adenocarcinoma cell line was obtained from the American Type Culture Collection (Manassas, VA) and maintained in RPMI-1640 medium with glutamine (Cat# 11875127, Gibco, ThermoFisher, Waltham, MA) supplemented with 10% FBS (Cat# SH30396.03, Hyclone, Marlborough, MA), penicillin-streptomycin (Cat# SV3007901, Hyclone) at 37 °C in 5% CO2. For inducing EMT, cells at 40–50% confluency in complete medium were serum starved for 24 h and treated with TGF-β (5 ng/ml) (Cat#100-21, PeproTech, Cranbury, NJ) for 72 h. We used 40–50% confluent cells as optimum to be able to obtain enough protein necessary for downstream molecular analysis versus inducing complete EMT. We have previously established that at 50% confluence, A549 cells undergo complete EMT as determined by observing <90% downregulation of epithelial marker, E-cadherin, and robust induction of mesenchymal markers N-cadherin and vimentin in response to 5 ng/ml TGF-β treatment for 72 h[21].

Isoform specific siRNA for ENOs includes a pool of 4 SMART selection-designed synthetic duplexes (Eno1 (Cat# M-004034-02-0005)), Eno3 (Cat# M-009718-02-0005, siGenome SMARTpool, Horizon Discovery/Dharmacon, Lafayette, CO). Cells at 40–50% confluency were transfected with siRNA using Lipofectamine 2000 (Cat No: 18324-012, Invitrogen, ThermoFisher, Waltham, MA) and optiMEM medium (Cat# 31985-070, Gibco, ThermoFisher, Waltham, MA) following the manufacturer's instructions. After 6 h of transfection cells were washed and allowed to recover from transfection in RPMI 1640 medium with 10% FBS before inducing EMT as described above.

To assess the knockdown efficiency of siRNA, first, Eno1 and Eno3 mRNA expression was assessed by qRT-PCR (Supplementary Fig. 6B). Briefly, Total RNA was isolated from the samples using the RNeasy Mini Kit (Cat# 74104, Qiagen, German Town, MD) following the manufacturer's detailed protocol. To generate complementary DNA (cDNA) from the isolated RNA, 1 µg of total RNA was used as the template. The cDNA synthesis was carried out using the High-Capacity cDNA Reverse Transcription Kit (Cat# 4368814 Applied Biosystems, ThermoFisher, Waltham, MA). RT-qPCR was performed using the Power SYBR Green Master Mix (Cat# 4385612, Applied Biosystems, ThermoFisher, Waltham, MA). The qPCR conditions were as follows: Initial denaturation: 95 °C for 20 s. Amplification: 40 cycles of 95 °C for 3 s and 60 °C for 30 s. GAPDH expression is assessed as internal control. Eno1 and En03 gene expression is normalized to GAPDH and fold change was calculated by comparing to respective untreated controls. qRT-PCR primers for ENO1 were Forward Primer: 5′ GCGTTCAATGTCATCAATGG 3′, and Reverse Primer:5′ GTTGACTTTGAGCAGGAGGC3′, and for ENO3 were Forward primer: 5′CGCAATGGGAAGTACGATCT3′, and Reverse primer:5′ ACTTGCGTCCAGCAAAGATT 3′.

Eno1 and Eno3 protein expression was assessed by Western Immunoblotting (Supplementary Fig. 6A). Briefly, cells were lysed in RIPA buffer (Cat# 20-188), containing Sodium Fluoride (Cat# 450022-5G), Sodium Orthovanadate (Cat# 567540-5GM) and protease inhibitor cocktail (Cat# 4693116001), from MilliporSigma, St. Louis, MO. Samples containing 20 µg of total protein were electrophoresed on SDS-polyacrylamide gels and transferred onto a polyvinyl difluoride membrane by electro blotting. Membranes were probed with indicated primary antibodies (Eno1, Rabbit Monoclonal Antibody (JM63-53), [Cat# MA5-32756], Eno3, Rabbit Monoclonal Antibody (003), [Cat# MA5-29239], Invitrogen, ThermoFisher, Waltham, MA), GAPDH (Rabbit polycolonal, [Cat# SAB4300645], Sigma-Aldrich, St.Louis, MO) with overnight incubation at 4 °C, followed by horseradish peroxidase-conjugated secondary antibodies (Cat# 7074,

Cell Signaling Technologies, Danvers, MA). Finally, the immunoblots were visualized using enhanced chemiluminescence (ECL) reagent (Cat# 170-5061, Bio-Rad, Hercules, CA), band intensities were quantified using Amersham Imager 600 (GE Healthcare Life Sciences) and normalized against corresponding GAPDH levels.

ENO1 and ENO3 mRNA expression analysis

Data in Supplementary Fig. 7A is derived from our previously published data set (GSE 17708). A549 cells were stimulated with 5 ng/ml of TGF-β, then were harvested at 0, 0.5, 1, 2, 4, 8, 16, 24, and 72 h after TGF-β stimulation: total RNA was isolated. From three biological replicates, RNA transcripts were assayed using Affymetrix HG-U133_plus_2 arrays containing 54675 probe-sets, representing ~20,000 distinct genes. Fold-changes were estimated as anti-log of the average differences for the log-transformed data as described before[19]. Two-way ANOVA models with effects for 3 experiments at indicated time points were fit to the data for each probe-set. We compared each time point with the 0 h time point, and counted probe-sets for which the *p* value for the comparison was smaller than 0.001. In cases where a gene was represented by more than one probe-set, we report data from the probe-set that had smallest *p* value for the overall F-test.

Apoptosis assays

Apoptosis was assessed by two independent methods; (1) *AnnexinV/Propodium Iodide (PI) staining:* At the end of the EMT experiment described above, all cells (including floating cells) were collected and stained for AnnexinV/PI using a kit (Cat# 640914, Biolegend, San Diego, CA). Briefly, collected cells were washed and resuspended in Annexin V binding buffer. 100 µl of cell suspension was stained with 5 µl of FITC-Annexin V, followed by 10 µl of PI staining solution. After 15 min incubation at room temperature in dark, 400 µl of Annexin V binding buffer is added and assessed for FITC-Annexin V and PI staining by flow cytometry. Annexin V and PI double positive cells are considered for assessing apoptosis. We show the sequential gating strategy for the flowcytometric analysis of AnnexinV /PI staining for four samples in (Supplementary Fig. 8) *Assessing Caspase 3/7 activation:* To assess casapase3/7 activation, an artificial caspase3/7 substrate (DVED) coupled to a red fluorescent DNA-binding dye (CellEvent™Cat# C10430, Invitrogen, ThermoFisher, Waltham, MA) is added to the cell culture (Fig. 4B and Supplementary Fig. 6C). When caspase3/7 is activated, it cleaves the DNA-binding dye which enters the nuclei and labels an apoptotic cell with red fluorescence allowing its imaging. Red fluorescent apoptotic cells were imaged and the number of fluorescent cells were counted per 96-well using CellCyte (Echo, San Diego, CA), a live cell imaging system.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All datasets used for analysis are publicly available. Processed data and composite analysis outputs are provided in the supplementary materials. Bulk cell transcriptome data is publicly available at Gene Expression omnibus with accession codes GSE17708 and GSE17518. Single cell transcriptome data can be found using the accession code GSE147405. Bulk proteome data were obtained from Keshamouni et al. [21] and Lu et al. [22]. CRISPR CERES scores can be downloaded from the DepMap portal (https://figshare.com/articles/dataset/DepMap_20Q2_Public/12280541/4).

Code availability

All codes used in this work are publicly available at https://github.com/sriram-lab/emt-cobra. The codes can be also accessed on Synapse at https://doi.org/10.7303/syn64332665[73].

Abbreviations

EMT	Epithelial-to-mesenchymal transition
TGF-β	Transforming Growth Factor β
NSCLC	Non-Small Cell Lung Cancer
COBRA	Constraint-Based Optimization and Reconstruction Analysis
FBA	Flux Balance Analysis
pFBA	Parsimonious FBA
CCLE	Cancer Cell Line Encyclopedia
CCL	Cancer Cell Line
GSMM	Genome-scale metabolic modeling

Received: 19 October 2023; Accepted: 17 December 2024;
Published online: 27 December 2024

References

1. Pastushenko, I. & Blanpain, C. EMT transition states during tumor progression and metastasis. *Trends Cell Biol.* **29**, 212–226 (2019).
2. Ramesh, V., Brabletz, T. & Ceppi, P. Targeting EMT in cancer with repurposed metabolic inhibitors. *Trends Cancer* **6**, 942–950 (2020).
3. Chockley, P. J. et al. Epithelial-mesenchymal transition leads to NK cell-mediated metastasis-specific immunosurveillance in lung cancer. *J. Clin. Investig.* **128**, 1384–1396 (2018).
4. Padua, D. & Massagué, J. Roles of TGFβ in metastasis. *Cell Res.* **19**, 89–102 (2009).
5. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).
6. Cook, D. P. & Vanderhyden, B. C. Context specificity of the EMT transcriptional response. *Nat. Commun.* **11**, 2142 (2020).
7. Jia, D. et al. Towards decoding the coupled decision-making of metabolism and epithelial-to-mesenchymal transition in cancer. *Br. J. Cancer* **124**, 1902–1911 (2021).
8. Oruganty, K., Campit, S. E., Mamde, S., Lyssiotis, C. A. & Chandrasekaran, S. Common biochemical properties of metabolic genes recurrently dysregulated in tumors. *Cancer Metab.* **8**, 5 (2020).
9. Jerby, L. & Ruppin, E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin. Cancer Res.* **18**, 5572–5584 (2012).
10. Sung, J., Wang, Y., Chandrasekaran, S., Witten, D. M. & Price, N. D. Molecular signatures from omics data: from chaos to consensus. *Biotechnol. J.* **7**, 946–957 (2012).
11. Orth, J. D., Thiele, I., Palsson & B, Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
12. Yizhak, K. et al. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *Elife* **3**, e03641 (2014).
13. Nilsson, A. et al. Quantitative analysis of amino acid metabolism in liver cancer links glutamate excretion to nucleotide synthesis. *Proc. Natl. Acad. Sci. USA* **117**, 10294–10304 (2020).
14. Bidkhori, G. et al. Metabolic network-based stratification of hepatocellular carcinoma reveals three distinct tumor subtypes. *Proc. Natl. Acad. Sci. USA* **115**, E11874–E11883 (2018).
15. Nelson, B. S. et al. Tissue of origin dictates GOT1 dependence and confers synthetic lethality to radiotherapy. *Cancer Metab.* **8**, 1 (2020).
16. Choudhary, K. S. et al. EGFR signal-network reconstruction demonstrates metabolic crosstalk in EMT. *PLoS Comput. Biol.* **12**, e1004924 (2016).
17. Matadamas-Guzman, M., Zazueta, C., Rojas, E. & Resendis-Antonio, O. Analysis of epithelial-mesenchymal transition metabolism identifies possible cancer biomarkers useful in diverse genetic backgrounds. *Front. Oncol.* **10**, 1309 (2020).
18. Wang, Q. et al. UDP-glucose dehydrogenase expression is upregulated following EMT and differentially affects intracellular glycerophosphocholine and acetylaspartate levels in breast mesenchymal cell lines. *Mol. Oncol.* **16**, 1816–1840 (2022).
19. Keshamouni, V. G. et al. Temporal quantitative proteomics by iTRAQ 2D-LC-MS/MS and corresponding mRNA expression analysis

identify post-transcriptional modulation of actin-cytoskeleton regulators during TGF-beta-Induced epithelial-mesenchymal transition. *J. Proteome Res.* **8**, 35–47 (2009).
20. Hecker, L. et al. NADPH oxidase-4 mediates myofibroblast activation and fibrogenic responses to lung injury. *Nat. Med.* **15**, 1077–1081 (2009).
21. Keshamouni, V. G. et al. Differential protein expression profiling by iTRAQ-2DLC-MS/MS of lung cancer cells undergoing epithelial-mesenchymal transition reveals a migratory/invasive phenotype. *J. Proteome Res.* **5**, 1143–1154 (2006).
22. Lu, C. et al. Coordination between TGF-β cellular signaling and epigenetic regulation during epithelial to mesenchymal transition. *Epigenetics Chromatin* **12**, 11 (2019).
23. Zur, H., Ruppin, E. & Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142 (2010).
24. Shen, F., Boccuto, L., Pauly, R., Srikanth, S. & Chandrasekaran, S. Genome-scale network model of metabolism and histone acetylation reveals metabolic dependencies of histone deacetylase inhibitors. *Genome Biol.* **20**, 49 (2019).
25. Lewis, N. E. et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* **6**, 390 (2010).
26. Wei, Q., Qian, Y., Yu, J. & Wong, C. C. Metabolic rewiring in the promotion of cancer metastasis: mechanisms and therapeutic implications. *Oncogene* **39**, 6139–6156 (2020).
27. Song, Y. et al. Alpha-enolase as a potential cancer prognostic marker promotes cell growth, migration, and invasion in glioma. *Mol. Cancer* **13**, 65 (2014).
28. Zhao, M. et al. Enolase-1 is a therapeutic target in endometrial carcinoma. *Oncotarget* **6**, 15610–15627 (2015).
29. Hou, X.-M., Yuan, S.-Q., Zhao, D., Liu, X.-J. & Wu, X.-A. LDH-A promotes malignant behavior via activation of epithelial-to-mesenchymal transition in lung adenocarcinoma. *Biosci. Rep*. **39**, BSR20181476 (2019).
30. Zhang, Y., Lin, S., Chen, Y., Yang, F. & Liu, S. LDH-Apromotes epithelial-mesenchymal transition by upregulating ZEB2 in intestinal-type gastric cancer. *Onco. Targets Ther.* **11**, 2363–2373 (2018).
31. Bergers, G. & Fendt, S. M. The metabolism of cancer cells during metastasis. *Nat. Rev. Cancer* **21**, 162–180 (2021).
32. Lu, J. The Warburg metabolism fuels tumor metastasis. *Cancer Metastasis Rev.* **38**, 157–164 (2019).
33. Xu, X. et al. Hyperglycemia promotes Snail-induced epithelial-mesenchymal transition of gastric cancer via activating ENO1 expression. *Cancer Cell Int* **19**, 344 (2019).
34. Liu, M., Quek, L.-E., Sultani, G. & Turner, N. Epithelial-mesenchymal transition induction is associated with augmented glucose uptake and lactate production in pancreatic ductal adenocarcinoma. *Cancer Metab.* **4**, 19 (2016).
35. Menendez, J. A. et al. Inhibition of fatty acid synthase (FAS) suppresses HER2/neu (erbB-2) oncogene overexpression in cancer cells. *Proc. Natl. Acad. Sci. USA* **101**, 10715–10720 (2004).
36. Pizer, E. S. et al. Inhibition of fatty acid synthesis delays disease progression in a xenograft model of ovarian cancer. *Cancer Res.* **56**, 1189–1193 (1996).
37. Yang, L., Venneti, S. & Nagrath, D. Glutaminolysis: a hallmark of cancer metabolism. *Annu. Rev. Biomed. Eng.* **19**, 163–194 (2017).
38. Vanhove, K. et al. Glutamine addiction and therapeutic strategies in lung cancer. *Int. J. Mol. Sci.* **20**, 252 (2019).
39. Riemma, M. A. et al. Sphingosine-1-phosphate/TGF-β axis drives epithelial mesenchymal transition in asthma-like disease. *Br. J. Pharmacol.* **179**, 1753–1768 (2022).
40. Chang, G.-C. et al. Identification of α-enolase as an autoantigen in lung cancer: its overexpression is associated with clinical outcomes. *Clin. Cancer Res.* **12**, 5746–5754 (2006).
41. Longo, N., Frigeni, M. & Pasquali, M. Carnitine transport and fatty acid oxidation. *Biochim. Biophys. Acta* **1863**, 2422–2435 (2016).
42. Zuo, W. & Chen, Y. G. Specific activation of mitogen-activated protein kinase by transforming growth factor-beta receptors in lipid rafts is required for epithelial cell plasticity. *Mol. Biol. Cell* **20**, 1020–1029 (2009).
43. Zhao, W. et al. Candidate antimetastasis drugs suppress the metastatic capacity of breast cancer cells by reducing membrane fluidity. *Cancer Res.* **76**, 2037–2049 (2016).
44. Gabitova-Cornell, L. et al. Cholesterol pathway inhibition induces TGF-β signaling to promote basal differentiation in pancreatic cancer. *Cancer Cell* **38**, 567–583.e11 (2020).
45. Sun, Y. et al. Metabolic and transcriptional profiling reveals pyruvate dehydrogenase kinase 4 as a mediator of epithelial-mesenchymal transition and drug resistance in tumor cells. *Cancer Metab.* **2**, 20 (2014).
46. Sellers, K. et al. Pyruvate carboxylase is critical for non-small-cell lung cancer proliferation. *J. Clin. Investig.* **125**, 687–698 (2015).
47. Fu, X.-G. et al. Histidine decarboxylase-expressing PMN-MDSC-derived TGF-β1 promotes the epithelial-mesenchymal transition of metastatic lung adenocarcinoma. *Int. J. Clin. Exp. Pathol.* **13**, 1361–1371 (2020).
48. Fu, Q.-F. et al. Alpha-enolase promotes cell glycolysis, growth, migration, and invasion in non-small cell lung cancer through FAK-mediated PI3K/AKT pathway. *J. Hematol. Oncol.* **8**, 22 (2015).
49. Pacini, C. et al. Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 (2021).
50. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
51. Okayama, H. et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* **72**, 100–111 (2012).
52. Győrffy, B., Surowiak, P., Budczies, J. & Lánczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS ONE* **8**, e82241 (2013).
53. Lin, Y.-H. et al. Author Correction: An enolase inhibitor for the targeted treatment of ENO1-deleted cancers. *Nat. Metab.* **3**, 122 (2021).
54. Tian, W. et al. Recent advances of IDH1 mutant inhibitor in cancer therapy. *Front. Pharmacol.* **13**, 982424 (2022).
55. Wang, X., Shen, X., Yan, Y. & Li, H. Pyruvate dehydrogenase kinases (PDKs): an overview toward clinical applications. *Bioscience Reports* **41**, BSR20204402 (2021).
56. Wende, A. R., Huss, J. M., Schaeffer, P. J., Giguère, V. & Kelly, D. P. PGC-1alpha coactivates PDK4 gene expression via the orphan nuclear receptor ERRalpha: a mechanism for transcriptional control of muscle glucose metabolism. *Mol. Cell Biol.* **25**, 10684–10694 (2005).
57. Gomes, A. P. et al. Altered propionate metabolism contributes to tumour progression and aggressiveness. *Nat. Metab.* **4**, 435–443 (2022).
58. Reka, A. K. et al. Identifying inhibitors of epithelial-mesenchymal transition by connectivity map-based systems approach. *J. Thorac. Oncol.* **6**, 1784–1792 (2011).
59. Reka, A. K. et al. Epithelial-mesenchymal transition-associated secretory phenotype predicts survival in lung cancer patients. *Carcinogenesis* **35**, 1292–1300 (2014).
60. Migita, T. et al. ATP citrate lyase: activation and therapeutic implications in non-small cell lung cancer. *Cancer Res.* **68**, 8547–8554 (2008).
61. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
62. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e27 (2018).

63. Duarte, N. C. et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **104**, 1777–1782 (2007).
64. Robinson, J. L. et al. An atlas of human metabolism. *Sci. Signal.* **13**, eaaz1482 (2020).
65. Campit, S. emt-cobra. *Github* https://github.com/ScottCampit/emt-cobra (2022).
66. Chandrasekaran, S. & Price, N. D. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. USA* **107**, 17845–17850 (2010).
67. Shen, F., Cheek, C. & Chandrasekaran, S. Dynamic network modeling of stem cell metabolism. *Methods Mol. Biol.* **1975**, 305–320 (2019).
68. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
69. Vasaikar, S. V. et al. EMTome: a resource for pan-cancer analysis of epithelial-mesenchymal transition genes and signatures. *Br. J. Cancer* **124**, 259–269 (2021).
70. Barretina, J. et al. 22 the cancer cell line encyclopedia - using preclinical models to predict anticancer drug sensitivity. *Eur. J. Cancer* **48**, S5–S6 (2012).
71. Malagoli Tagliazucchi, G., Wiecek, A. J., Withnell, E. & Secrier, M. Genomic and microenvironmental heterogeneity shaping epithelial-to-mesenchymal trajectories in cancer. *Nat. Commun.* **14**, 789 (2023).
72. Chae, Y. K. et al. Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci. Rep.* **8**, 2918 (2018).
73. Chandrasekaran, S. Metabolic model for epithelial mesenchymal transition. Software code available at https://doi.org/10.7303/SYN64332665 (2024).

Acknowledgements

This work was supported by faculty start-up funds from the University of Michigan, the Rogel Cancer Center, Camille and Henry Dreyfus Foundation, and R35 GM13779501 from NIH to S.C., Proteome Informatics of Cancer Training Program T32 CA140044 to S.E.C., American Lung Association Lung cancer Discovery Award (#563971), Michigan Medicine-Peking University Joint Institute for Translational & Clinical Research Discovery Award, NIH R01-HL-156998-01A1 and VA merit CX002233-01A2 to V.G.K.

Author contributions

S.C., V.G.K. conceived the study and designed research, S.E.C. and R.B. designed and performed research, S.K.K. performed experimental analysis at revision, and S.C., V.G.K., S.E.C., and R.B. wrote the manuscript.

Competing interests

S.C. is a consultant for Axcella Health Inc. Authors declare no other competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-07408-7.

Correspondence and requests for materials should be addressed to Venkateshwar G. Keshamouni or Sriram Chandrasekaran.

Peer review information *Communications Biology* thanks Eytan Ruppin, Tian-Gen Chang, Neel Sanghvi, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Toril Holien and Joao Valente.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024