A data assimilation approach to predict cell population epithelial-mesenchymal transition dynamics

Mario J. Mendez, Matthew J. Hoffman, Elizabeth M. Cherry, Christopher A. Lemmon, Seth H. Weinberg

Epithelial-mesenchymal transition (EMT) is a biological process that plays a central role in embryonic development, tissue regeneration, and cancer metastasis. Transforming growth factor-β (TGFβ) is a potent inducer of this cellular transition, comprised of transitions from an epithelial state to partial EMT state(s), then to a mesenchymal state. Recent experimental studies have shown that within a population of cells, different phenotypical profiles were correlated with different time- and dose-dependent responses to TGFβ1 during EMT. This offers a challenge for computational models since most model parameters are averages from a range of experimental conditions. Thus, model simulations generally represent typical cell responses, not necessarily specific responses. In this study, we applied a data assimilation approach, which combines limited noisy observations with predictions from a computational model, paired with a parameter estimation to minimize the effects of parameter uncertainty and model error. We mimic the biological heterogeneity in cell states that is observed on epithelial cell populations, by generating a large population of model parameter sets. We performed a series of *in silico* experiments where a forecasting system was tasked with reconstructing the EMT dynamics of a phenotypically different virtual epithelial cell population. The cell population was exposed to a time-dependent exogenous TGFβ doses and, at a selected time threshold, an EMT-suppressing or EMT-promoting perturbations. We find that a data assimilation approach can successfully reconstruct and predict the dynamics of a heterogeneous virtual epithelial cell population in the presence of physiological model error and parameter uncertainty. Furthermore, we observed that when estimating single critical parameters, we can accurately predict the phenotypical responses to specific EMT suppressing or EMT promoting parametric perturbations.