bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Immunosuppressive traits of the hybrid epithelial/mesenchymal phenotype

Running title: Hybrid epithelial/mesenchymal cells have high PD-L1 levels

Sarthak Sahoo[1, 2, †], Sonali Priyadarshini Nayak[3, †], Kishore Hari[2], Susmita Mandal[2], Akash Kishore[4], Herbert Levine[5, 6], Mohit Kumar Jolly[2, *]

[1]Undergraduate program, Indian Institute of Science, Bangalore, Karnataka 560012, India

[2] Centre for BioSystems Science and Engineering, Indian Institute of Science, Bangalore, Karnataka 560012, India

[3] College for Integrated Studies, University of Hyderabad, Hyderabad, Telangana 500046, India

[4] Department of Computer Science & Engineering, SSN College of Engineering, Chennai, India

[5] Center for Theoretical Biological Physics, Northeastern University, Boston, MA

[6] Department of Physics, Northeastern University, Boston, USA

[*] Author to whom correspondence may be addressed (mkjolly@iisc.ac.in)

[†] These authors contributed equally

Abstract

Cancer metastasis remains a primary cause of cancer related mortality. Recent *in vitro* and *in vivo* data has indicated the high metastatic fitness of hybrid epithelial/mesenchymal (E/M) states, i.e. their enhanced abilities to initiate tumours at secondary tumour site. Mechanistic details about how such hybrid E/M cells survive the metastatic cascade remain unclear. Here, Univwe investigate immune-evasive strategies of hybrid E/M states, an issue that to date has been largely unexplored. We construct a minimalistic regulatory network that captures known associations between regulators of EMT (the epithelial mesenchymal transition) and levels of PD-L1, an established suppressor of immune response, and simulated the network's emergent dynamics. Our model recapitulates observations that cells undergoing EMT have increased PD-L1 levels, while reverting EMT can decrease these levels, indicative of a causal link between EMT drivers and PD-L1. Further, we show that hybrid E/M cells can have high levels of PD-L1, similar to those seen in cells with a full EMT phenotype, thus obviating the need for cancer cells to undergo a full EMT to evade the immune system. Finally, we identify various signalling pathways and cellular processes that can independently or in concert affect PD-L1 levels and EMT status. For instance, hybrid E/M cells can gain both immune-evasion and stemness through largely independent paths. Our results underscore another underlying reason for the high metastatic ability of hybrid E/M cells.

Keywords

Hybrid epithelial/mesenchymal; PD-L1; immune evasion; multistability

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The progression of cancer relies on a complex interplay of various cell autonomous and non-cell autonomous phenomena. Cancer cells can proactively create a microenvironment that aids their own survival. One of the strategies employed is to suppress various arms of immune system that can lead to cancer cell elimination (1). For instance, some tumor cells can inhibit the functions of effector T (T_{eff}) cells and/or induce a population of tolerogenic cells that ultimately result in the immune escape of the tumor. They can also facilitate accumulation of immune suppressive cells such as regulatory T (T_{reg}) cells, myeloid derived suppressor cells (MDSCs) and M2 macrophages/ tumor-associated macrophages (TAMs), leading to tumor growth (1). Understanding these strategies of tumor-driven reprogramming of the microenvironment would be a major step towards more effective guiding of various therapeutic interventions.

In addition to reprogramming the immune cells in the stroma, tumors employ cell autonomous mechanisms that help them directly evade cytotoxic CD8 T cells. A key mechanism via which tumor cells achieve this evasion is via the expression of programmed death-ligand 1 transmembrane protein (PD-L1) on their cell membranes (2). The binding of PD-L1 to PD-1 receptor on activated T cells drives the exhaustion of these T cells, reducing their cytotoxic abilities (3). In cancer cells, a multitude of molecular players modulate PD-L1 levels at various regulatory stages (2). Of interest here is the finding that PD-L1 levels can be increased as cells go through Epithelial-Mesenchymal Transition (EMT) and consequently gain the ability to migrate and invade (4–6). The process of EMT, however, is not typically a binary switch, as had been tacitly assumed in these earlier works. Instead, cells can stably maintain one or more hybrid epithelial/mesenchymal (E/M) phenotypes that can be much more metastatic than cells in a 'full EMT' or 'extremely mesenchymal' state (7). Besides, hybrid E/M phenotypes can be resilient to various therapies across cancers (7). However, the immune evasive properties of the hybrid E/M states are relatively poorly understood.

In this study, we identify a core regulatory network that helps us elucidate the immune evasive properties of different phenotypes along the epithelial-hybrid-mesenchymal spectrum. Our simulations indicate that hybrid E/M phenotypes are extremely likely to exhibit high PD-L1 levels, similar to mesenchymal cells, thus obviating the need to undergo a full EMT to display immunosuppression. Moreover, the switch from an epithelial/low-PDL1 state to hybrid/high-PDL1 or mesenchymal/high-PDL1 state is reversible, i.e., while EMT can induce PD-L1 levels, MET can reduce them. Our model predictions are consistent with extensive analysis of transcriptomic datasets across cancers.

Results

Hybrid E/M and Mesenchymal cell states are more likely to exhibit high PD-L1 levels

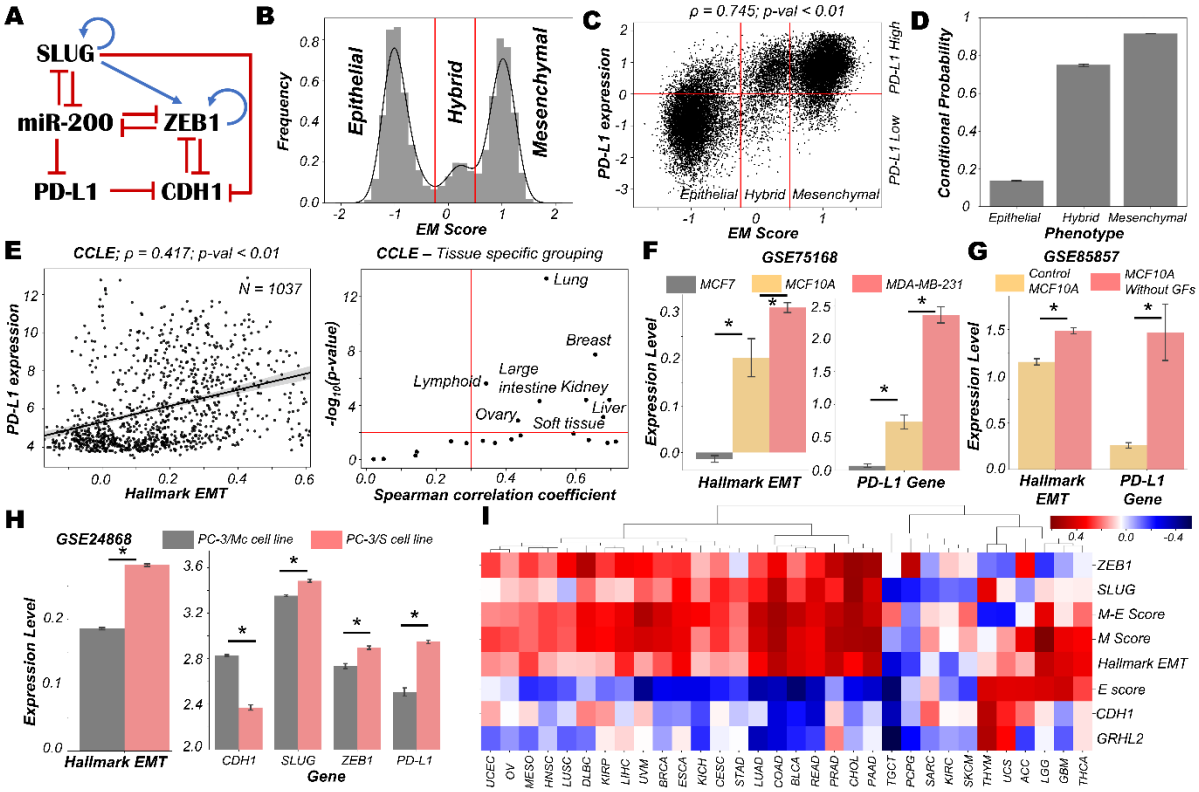
Capturing the essence of biological processes via mechanism-based mathematical modelling can be a daunting task given the vast complexity of biological systems. Identifying an appropriately sized gene regulatory network that incorporates the essential features of the underlying biological mechanism at hand in a minimalistic, yet informative manner is a key first step. To that extent, we started out with a core set of four well-reported biomolecules and their interactions that can capture the non-binary nature of the EMP spectrum and the ability of different phenotypes to modulate PD-L1 levels: ZEB1, miR-200, CDH1 (E-cadherin) and SLUG (**Fig. 1A**). Mutual inhibition between ZEB1 and miR-200, together with ZEB1 self-activation can enable epithelial, mesenchymal and hybrid epithelial/mesenchymal (E/M) phenotypes (8). SLUG has been reported to associate with hybrid E/M phenotype specifically both in experimental and computational analysis (9,10).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

High levels of ZEB1 and SLUG with low levels of CDH1 and miR-200 are frequently attributed to a mesenchymal phenotype, while high levels of CDH1 and miR-200 with a concurrent reduced level of ZEB1 and SLUG usually associates with an epithelial phenotype (11). Interactions between ZEB1, miR-200, CDH1, and SLUG have been extensively studied (10–13). Furthermore, miR-200 has been known to directly inhibit PD-L1 by binding to 3' untranslated region of its mRNA (4). PD-L1 can, in turn, repress the levels of CDH1 via indirect mechanisms (14) (Fig 1A).

We used RACIPE (15) to generate *in-silico* steady state gene expression values enabled by this gene regulatory network (**Methods**). RACIPE simulates a given gene regulatory network as a set of coupled ordinary differential equations (ODEs), with parameters being sampled from biologically relevant ranges. The ensemble of resultant steady states is indicative of the possibility space allowed by the network topology. To quantify the cellular E/M phenotype of given steady state solution, we defined an EM score from z-normalised expression values of ZEB1, SLUG, miR-200, and CDH1. The higher the EM score, the more mesenchymal is the corresponding phenotype. A histogram of these scores showed a clear trimodal distribution, which can be construed as epithelial, hybrid E/M, and mesenchymal phenotypes; these assignments can be confirmed by PCA plots (**Fig 1B, S1A-B**). Subsequently, we also observed a bimodal distribution of PD-L1 levels (**Fig S1C**) where high levels of PD-L1 can be viewed as an immune-evasive state while low PD-L1 denotes an immune-sensitive tumor cell state (16).

Next, we investigated the association between the EM scores and PD-L1 levels and observed a strong positive correlation between them (ρ = 0.745; p-value < 0.01) (**Fig 1C**). Conditional probability analysis shows that only a small percentage (~15%) of epithelial cells were PD-L1+. In contrast, a much larger percentage of hybrid E/M (~70%) and mesenchymal (~90%) cells were PD-L1+ (**Fig. 1D, S1D**). Consistently, PD-L1 was found to negatively correlate with CDH1 but positively with ZEB1 and SLUG (**Fig S1E**). Further, ZEB1, SLUG, and PD-L1 all had intermediate levels in hybrid E/M state compared to extreme states – epithelial and mesenchymal (**Fig. S1F**). Together, these results suggest that cells undergoing either a partial or full EMT can upregulate their levels of PD-L1 and consequently can exhibit immune evasion.



bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

***Fig. 1**. **Dynamics of regulatory network coupling EMT with PD-L1. A)** Regulatory network (GRN) interplay of EMT regulators coupled with PD-L1. Blue arrows stand for activation links, red hammers for inhibitory links. **B)** Density histogram of EM Score fitted with kernel density estimate showing a trimodal distribution. Red lines show the partition between phenotypes: Epithelial, Hybrid, and Mesenchymal. **C)** Scatter plot of PD-L1 expression and EM score. Horizontal red line shows the partition between PDL1 expression level being high vs. low. Vertical red lines show the partition between phenotypes: Epithelial, Hybrid, and Mesenchymal based on EM score. Spearman's correlation coefficient (ρ) and corresponding p-value (p-val) have been reported. **D)** Bar plot representing conditional probability of a phenotype being PD-L1 high given that it belongs to a given EMT phenotype. **E)** Scatter plot showing correlation between PD-L1 levels and the Hallmark EMT signature in cell lines from CCLE. Spearman's correlation coefficient (ρ) and corresponding p-value (p-val) are reported (left). Splitting CCLE cell lines reveals tissues that show a strong significant correlation ($\rho > 0.3$ and p-val < 0.01) (right) right panel. **F)** Activity levels of Hallmark EMT and PD-L1 expression levels in 3 breast cancer cell lines (GSE75168). **G)** Activity/Expression levels of Hallmark EMT and PD-L1 levels in MCF10A breast cancer cells treated with or without growth factors (GSE85857). **H)** Activity/Expression levels of Hallmark EMT and PD-L1 levels in two prostate cancer sub-lines of PC3 with different EMT status (GSE24868). **I)** Heatmap showing the Spearman's correlation coefficients between the different EM metrics and EMT associated genes and PD-L1 levels across 32 different cancer types in TCGA. * denotes a statistically significant difference (p-val < 0.05) between the represented groups assessed by a two-tailed Students t-test assuming unequal variances.*

To validate these model predictions, we analysed pan-cancer gene expression datasets such as CCLE (Cancer Cell Line Encyclopaedia), where we observed the ssGSEA scores of EMT to be positively correlated with PD-L1 levels (**Fig. 1E; left**). A tissue-specific analysis revealed a majority (16 out of 22) of cancers exhibited a strong correlation ($\rho > 0.3$) between EMT and PD-L1 expression (**Fig. 1E; right**). Next, we investigated more specific scenarios. For instance, three breast cancer cell lines along the EMP spectrum – MCF7 (epithelial), MCF10A (hybrid E/M) and MDA-MB-231 (mesenchymal) (17,18) – showed consistent trends with PD-L1 levels with MCF7 < MCF10A < MDA-MB-231 (**Fig 1F**). This pattern that was recapitulated in an analysis of TCGA luminal breast cancer samples (**Fig S1G**. Furthermore, MCF10A cells, when driven to a more mesenchymal phenotype upon growth factor depletion (19), showed concurrent increase in the levels of PD-L1 (**Fig 1G**). Similarly, comparing two sub-lines of prostate cancer cells PC-3 (20), the more mesenchymal one (PC-3/S) showed higher levels of PD-L1, ZEB1, and SLUG relative to the hybrid E/M PC-3/Mc cells (**Fig 1H**). A positive correlation between EMT signature and PD-L1 levels was also seen in A549 lung adenocarcinoma cells induced to undergo EMT (**Fig. S1H**), suggesting a pan-cancer association of EMT with PD-L1 levels.

Finally, we analysed TCGA patient cohort datasets for all the above-mentioned features. We calculated the Spearman's correlation coefficient between PD-L1 expression levels with those of CDH1, GRHL2, SLUG, and ZEB1 as well as with a Hallmark EMT signature, and epithelial and mesenchymal specific signatures (**Fig. 1I, S2A**). A majority of cancers (21 out of 32) showed a strong positive correlation of PD-L1 with the mesenchymal related metrics (ZEB1, SLUG, Hallmark EMT, mesenchymal score, and M-E score) while showing an intermediate to strong negative correlation with the epithelial ones (CDH1, GRHL2, and epithelial score), thereby endorsing our model predictions. Intriguingly, 6 cancer types (THYM, UCS, ACC, LGG, GBM, and THCA) showed a positive correlation of PD-L1 with both epithelial and mesenchymal signatures, highlighting a possible association of highest PD-L1 levels with the hybrid E/M phenotype. It should be noted that our model does not completely preclude the association of an epithelial state with high PD-L1 levels, although the likelihood of such association is relatively low (**Fig 1C**). This infrequent association may underlie context-specific behavior of epithelial tumours (such as Thymic epithelial tumours) also show high PD-L1 positivity (**Fig 1I**) (21).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Overall, *in silico* predictions, supported by analysis of *in vitro* and patient data, suggests that a change in the EMP status of the cell is positively associated with PD-L1 levels across various cancer types. These results clearly indicate the likelihood of the hybrid E/M phenotype being (almost) as immune evasive as the mesenchymal phenotype.

Traversal of cells on the EMP spectrum can alter the PD-L1 status of the cells

After establishing a pan-cancer correlation of more mesenchymal status and higher PD-L1 levels, we examined a causal connection between them. We simulated the set of coupled ODEs for a representative tristable parameter set with diverse initial conditions and observed convergence to three distinct EM states (**Fig. 2A**). Corresponding PD-L1 levels followed the previously observed patterns with epithelial state showing the least PD-L1 levels, while both the hybrid E/M and mesenchymal showing nearly equal levels of PD-L1 higher than those of the epithelial state (**Fig 2A**). Stochastic simulations for this parameter set created a landscape indicative of the steady states under the influence of biological noise. This landscape revealed the co-existence of three states (shown by valleys) – (Epithelial, PD-L1 low), (Hybrid E/M, PD-L1 high) and (Mesenchymal, PD-L1 high) (**Fig. 2B**), depicting that as cells change their EMP status, their corresponding PD-L1 levels are also altered. In another parameter set to study the stochastic dynamics of this network (22), we observed spontaneous switches between epithelial and mesenchymal states with a concurrent change in the levels of PD-L1 (**Fig. 2C**). Together, this analysis points towards the possibility of a switch like behaviour in acquisition of an immune evasive phenotype as the cells undergo EMT.

To characterize the impact of perturbations on our core regulatory network we simulated the scenarios of EMT induction and MET induction. EMT was induced by down expression (DE) of miR-200 or over expression (OE) of SLUG; conversely, MET was induced by OE of miR-200 or DE of SLUG (13). SLUG-OE or miR-200 DE increased the proportion of mesenchymal cell states with a concurrent decrease in epithelial cases (**Fig 2D-E, Fig S3A-B, D-E**). This change resulted in a significant increase in PD-L1 levels (**Fig S3C, F**). Opposite trends were observed in the cases of MET induction via miR-200 or SLUG-DE, with resultant changes in PD-L1 levels (**Fig 2D-E**).

Next, we probed whether our model prediction about a concurrent switch in EMP status and PD-L1 levels is supported by experimental data through analysing corresponding transcriptomic data. HeLa cells treated with TGFβ and EGF led to an induction of EMT, evident by increases in SLUG and ZEB1 levels, as well in the activity (as measured via ssGSEA; see Methods) of the Hallmark EMT gene set (**Fig 2E**). In treated cells, CDH1 levels were significantly decreased while PD-L1 levels were increased (**Fig 2E**). This phenomenon of EMT-driven increase in PD-L1 was also seen in non-cancerous cells where TGFβ treatment of primary airway epithelial cells led to upregulation of EMT and PD-L1 (**Fig S3G**), indicating that this association between EMT and PD-L1 levels need not be restricted to cancer cells. Furthermore, we compared the profiles of triple negative breast cancer cells DKAT when grown in a medium supporting epithelial growth (MEGM) vs when grown in a medium containing stromal factors (SCGM). These have been shown to differ in their EM status: while culturing in SCGM facilitated a mesenchymal phenotype, that in MEGM drove an epithelial one (GSE33146). Consistently, SLUG, ZEB1 and PD-L1 levels were significantly higher in cells grown in SCGM rather than in MEGM (**Fig S3G**). The reciprocal nature of expression between ERα and EMT programme have been studied in the context of the acquisition of reversible drug resistance in ER+ breast cancer (23). Thus, the loss of ERα that drives a more mesenchymal phenotype in luminal breast cancer can also lead to an increase in PD-L1 levels (**Fig S3I**). Next, we analysed a set of experiments in which MCF10A cells were induced to undergo EMT either via TGFβ application or by induced overexpression of SNAIL. This time course experiment resulted in

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

an increase of activity of Hallmark EMT genes and PD-L1, irrespective of cells' initial EMP status (**Fig 2F**).

Finally, we asked whether induction of MET can decrease the levels of PD-L1 in cancer cells. HMLE cells, upon overexpression of MET-inducing factor GRHL2, displayed a more epithelial state (increased CDH1, decreased ZEB1 and hallmark EMT signature) with a substantial drop seen in PD-L1 levels (**Fig 2H**), pinpointing that EMT-driven changes in PD-L1 levels are reversible. Similar observations were made for PD-L1 levels in LNCaP prostate cancer cells which were first induced to undergo EMT and subsequently MET was induced. A two-dimensional plot of EM score and PD-L1 levels revealed an increase in PD-L1 as EMT was induced, and a decrease when MET was induced. Intriguingly, the cell population did not retrace its original path during MET induction, indicative of hysteresis in the system (24). The overall levels of PD-L1 were lower at the end of 20 days of MET than for the uninduced cells themselves, suggesting that MET induction can reset the baseline PD-L1 levels upon a cycle of EMT and MET. Collectively, these results underscore that induction of EMT or MET in cancer cells can regulate their immune evasion status through altered levels of PD-L1.

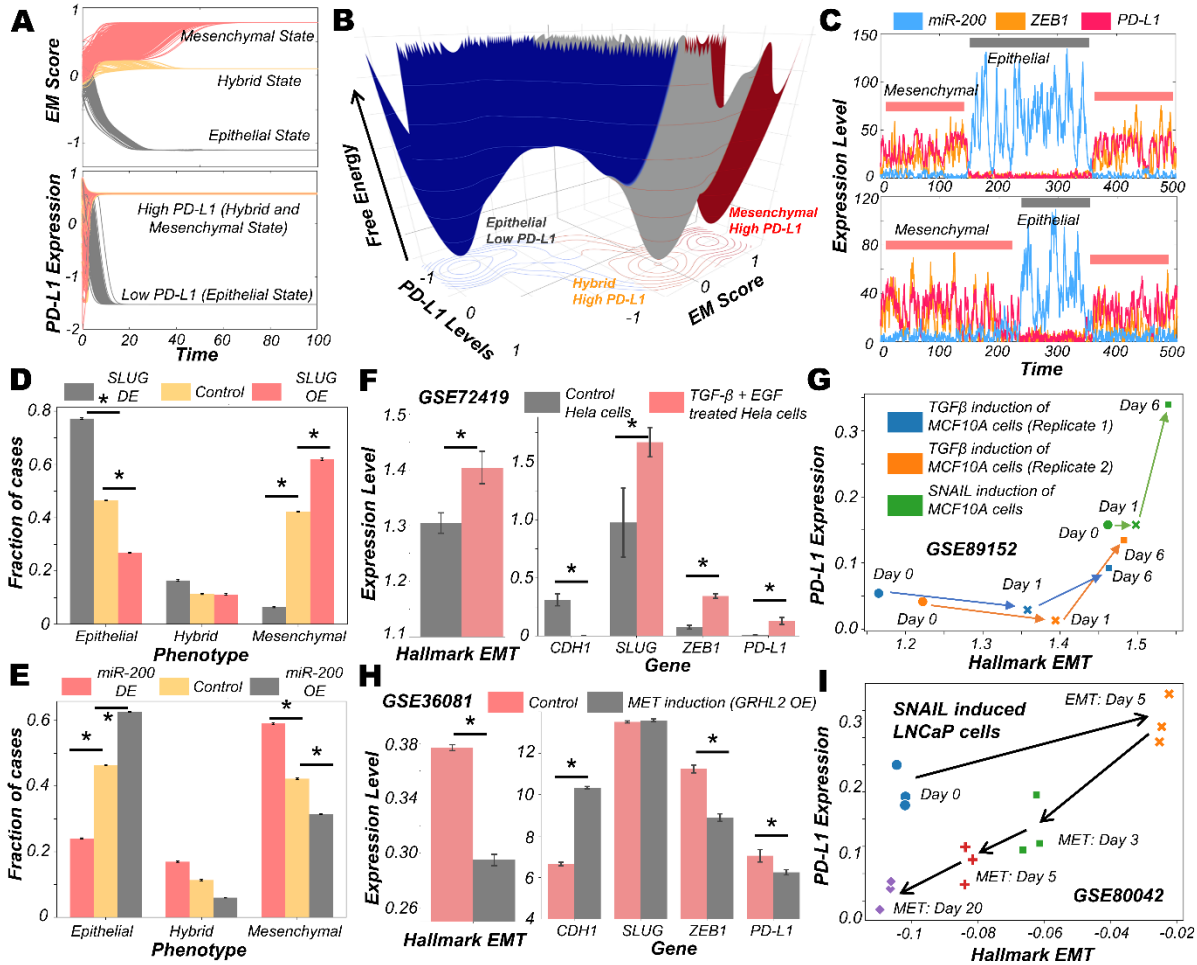


Fig. 2. Evidence for causal links between EMT associated genes and PD-L1 levels. A) *Dynamics of EM score and PD-L1 showing presence of epithelial, hybrid, and mesenchymal phenotypes and their corresponding PD-L1 levels, when simulated from multiple initial conditions.* ***B)*** *Free energy landscape on the PD-L1 and EM score plane, with the valleys representing the stable states possible in the system. Three distinct states – Epithelial/ PD-L1 low, Hybrid E-M/ PD-L1 high, and Mesenchymal/ PD-L1 high – are observed.* ***C)*** *Stochastic simulations of gene regulatory network via sRACIPE showing spontaneous switching between different states.* ***D)*** *Simulation results showing the fraction of cases of epithelial, hybrid, and mesenchymal phenotypes*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

under control (yellow), SLUG DE (grey) and SLUG OE (orange) conditions. ***E)*** *Same as D but for miR-200 OE (grey) and miR-200 OE (orange) conditions.* ***F)*** *Activity/Expression levels of Hallmark EMT and PD-L1 levels in Hela cells induced to undergo EMT (GSE72419).* ***G)*** *Two-dimensional Hallmark EMT and PD-L1 plot showing trajectory of MCF10A cells induced with TGFβ or SNAIL (GSE89152).* ***H)*** *Activity/Expression levels of Hallmark EMT and PD-L1 levels in HMLE cells where MET has been induced via overexpression of GRHL2 (GSE36081).* ***I)*** *Two-dimensional Hallmark EMT and PD-L1 plot showing trajectory of LNCaP prostate cancer cells that have been induced with SNAIL to undergo EMT followed by removal of signal to induce MET (GSE80042). * denotes a statistically significant difference (p-val < 0.05) between the represented groups assessed by a two-tailed Students t-test assuming unequal variances.*

Various signalling pathways can either independently or in concert modulate the immune evasive properties of cancer cells on the EMP spectrum

The above-mentioned interconnections among the EMT regulators and PD-L1 levels seldom work in isolation. Multiple signalling pathways can independently or in concert affect the expression of EMT status of the cells and/or the PD-L1 expression. To investigate such effects, we calculated the degree of correlation of 15 well-defined signalling pathways with EMT and with PD-L1 levels across different cancers in the TCGA cohort (**Fig S4A**). A scatter plot of corresponding correlation coefficients revealed pan-cancer consistency in signaling pathways associated with EMT and with PD-L1 levels (ρ = 0.37; p-value < 0.01) (**Fig S4B**). Next, we ranked which pathways correlate most strongly with EMT signature or PD-L1 levels (**Fig 3A**). TGFβ, IL2-STAT5, TNFα-NFκB, IL6-STAT3 and NOTCH signalling were found to correlate strongly with EMT, consistent with their expected roles (25–29). Similarly, PD-L1 levels are most correlated with IL6-STAT3, IFNγ, IL2-STAT5, IFNα and TNFα-NFκB, all of which have been previously implicated (30). Plotting these pathways through their normalised ranks allows identifying the pan-cancer independent regulators of PD-L1 levels and EMT; for instance, TGFβ and NOTCH can be considered as more EMT-specific, while IFNγ and IFNα are PD-L1 specific modulators. IL6-JAK-STAT3, IL2-STAT5 and TNFα-NFκB pathways correlated both with PD-L1 and EMT (**Fig 3B**). IL1β is known to act partially by the NFκB pathway (31). Thus, it is not surprising to see that treatment of cancer cells with IL1β caused a concerted increase in EMT as well as PD-L1 levels; the consistency in these trends was also visible upon withdrawal of the signal (**Fig 3C**).

The interplay between stemness and EMT has been extensively investigated (32,33). Thus, we asked whether EMT, stemness and PD-L1 levels all vary together. To investigate this crosstalk, we simulated an extended regulatory network including stemness regulators (OCT4, miR-145, LIN28, let-7) via RACIPE (**Fig S4C). A** stemness window was defined based on the distribution of stemness (SN) score (**Fig S4D**). This network showed conserved trends between EM score and PD-L1 expression level and found that most hybrid E/M solutions lay within the stemness window (**Fig 3D-E**). Quantifying these trends among EMT status, stemness status and PD-L1 levels revealed that while hybrid E/M cells were very likely to exhibit both PD-L1 and enhanced stemness; the stemness status by itself (irrespective of EMT status) could not predict any association with PD-L1 (**Fig 3F**). The non-monotonic nature of association between EMT states and stemness was confirmed by pan-cancer data analysis, where the stemness signature was most enriched in cells with hybrid E/M status (**Fig 3G; Fig S4E**) while no such trend was seen for a direct association of stemness with PD-L1 levels (**Fig 3H**). Together, we conclude that while hybrid E/M cells are more stem-like and immune-evasive, these two features are likely acquired independent of one another.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

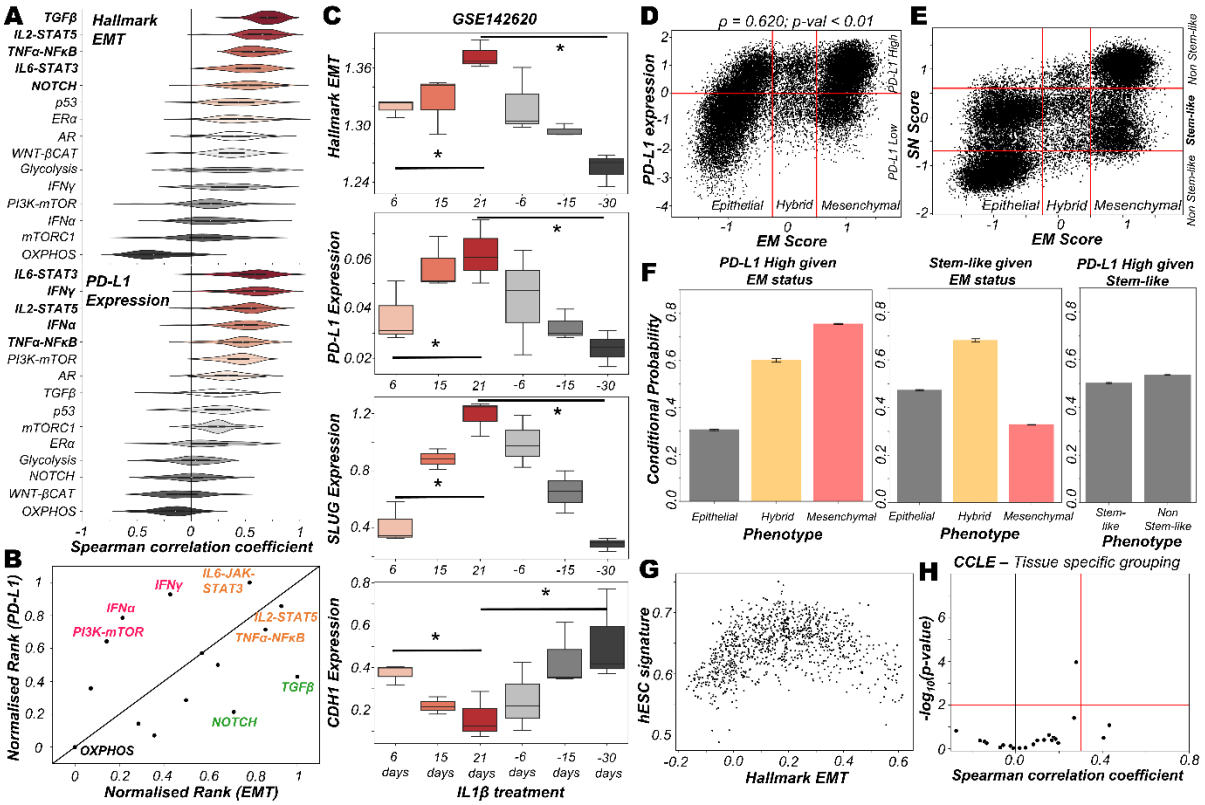


Fig. 3. Signalling pathways and biological processes that can affect PD-L1 and/or EMT. A) *Violin plots of Spearman's correlation values of different signalling pathways with Hallmark EMT programme (top) or with PD-L1 levels (bottom) ordered by corresponding median values across 27 cancer types in TCGA.* ***(B)*** *Scattered plot between normalised ranks of signalling pathways with the EMT programme and with PD-L1 expression levels. Signalling pathways hypothesised to be specific to EMT programme are labelled in green, those specific for PD-L1 highlighted in pink and those with both in orange.* ***C)*** *Activity/expression levels of Hallmark EMT, PD-L1, SLUG, and CDH1 levels in lung cancer cells treated with IL-1β and subsequent removal of signal (GSE142620).* ***D)*** *Scatter plot of PD-L1 expression and EM score. Horizontal red line shows the partition between PDL1 expression level being high vs low for the circuit in Fig S4C. Vertical red lines show the partition between phenotypes: Epithelial, Hybrid, and Mesenchymal based on EM score. Spearman's correlation coefficient (ρ) and corresponding p-value (p-val) are reported.* ***E)*** *Scatter plot of SN score and EM score showing the presence of clusters having predominantly stem-like hybrid and the presence of both stem like and non-stem like epithelial and mesenchymal cells scattered in the plane. Horizontal red lines show the partition between stem-like and non-stem-like based on SN score and EM phenotypes.* ***F)*** *Bar plot representing conditional probability of PD-L1 being high given EM status, stem-like phenotype given EM status, and PD-L1 high given stemness status respectively.* ***G)*** *Scatter plot showing the non-monotonic association between the hESC signature and the Hallmark EMT signature in CCLE dataset.* ***H)*** *No tissue in CCLE show a strong significant Spearman's correlation (ρ > 0.3 and p-val < 0.01) between hESC signature and PD-L1 levels. * denotes a statistically significant difference (p-val < 0.05) between the represented groups assessed by a two-tailed Students t-test assuming unequal variances.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Cellular processes are full of regulatory feedback loops and mechanisms that maintain a dynamic equilibrium, thus enabling cells to adapt to various internal and external fluctuations. The expression of PD-L1 on the surface of cells is one such mechanism that keeps the inflammatory responses from uncontrolled activation, by providing necessary brakes. Tumour cells exploit this checkpoint to escape from both immunological detection and elimination. PD-L1 on cancer cells' surface enables them to prevent the activation of T cells while simultaneously causing them to be exhausted, to prevent T cells from targeting of cancer cells. High PD-L1 has been exhibited in circulating tumor cells as well across cancers, and EMT has been associated with higher PD-L1 levels. Here, we have investigated PD-L1 levels in hybrid E/M phenotypes, given their higher fitness for metastasis and evasion of various other treatment options.

Through *in silico* simulations for underlying networks incorporating crosstalk between PD-L1 and EMT regulators, we observed that hybrid E/M states can show high PD-L1 similar to those seen in a 'full EMT' (mesenchymal) phenotype. This model prediction is substantiated by analysis of gene expression pan-cancer datasets. We further show that EMT/MET can alter PD-L1 status reversibly in cancer cells, a trend validated in multiple *in vitro* datasets. Intriguingly, while hybrid E/M phenotype was found to be associated with both enhanced PD-L1 and higher and stemness, a direct association between PD-L1 levels and stemness was not found. Our results thus highlight another dimension of the high metastatic fitness of hybrid E/M cells: their immune-suppressive traits.

We acknowledge, as with all models, the limitations of our model. We considered here a minimalistic regulatory network that captures key features of interest, and, thus is far from being comprehensive. On one hand, our model recapitulates key observations especially including previously reported associations between EMT and PD-L1 levels; on the other, it provides testable predictions, regarding hybrid E/M states being likely to be PD-L1 positive). The overarching positive correlations between EMT and PD-L1 levels across a majority of the cancers in TCGA shows the broad applicability of our conclusions; these were supported in by the analysis of the CCLE and of specific datasets dealing with perturbations. However, our generic model cannot explain some specific exceptions, specifically why certain cancer types such as TGCT, PCPG, SARC and SKCM do not show strong correlation of PD-L1 with either epithelial or mesenchymal programmes. Interestingly, some cancers of mesenchymal origins (LGG, GBM) show positive correlations of PD-L1 with epithelial signatures, suggesting the association of a hybrid E/M state with maximal PD-L1 levels. Finally, our extended model, obtained by considering the additional players OCT4, LIN28, miR-145 and let-7, finds no significant association between stemness and PD-L1 levels. Various additional nodes in the network not considered here may alter this trend, which may explain previously reported correlations between immune evasion and stemness. Thus, future efforts are needed to understand these context-specific scenarios in terms of interplay between EMT, stemness and immune evasion.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Materials and methods

Stochastic simulations and Landscape construction

We simulated the gene regulatory using the Euler-Maruyama method for a representative parameter set (**Table S4**) that showed the co-existence of 3 phenotypes: epithelial with low levels of PD-L1; hybrid E/M with high levels of PD-L1 and mesenchymal with high levels of PD-L1. The corresponding equation is as follows:

$$X_i(t+1) = X_i(t) + \Delta t * g_{X_i} * \prod_j H^s\left(X_j(t), X_{ji}0, n_{ji}, \lambda_{ji}\right) - k_{X_i} * X_i(t) * \Delta t + \sqrt{\Delta t} * N(0,1)$$

The equation is just a discrete form of the ODE presented before, with an addition of the noise term $\sqrt{\Delta t} * N(0,1)$, where Δt is the time step and N(0,1) is a normal random variable with mean 0 and standard deviation 1. For the parameter set, we simulated the network for 100 different initial conditions sampled uniformly from the range $\left[0, 1.5 * \frac{g_{X_i}}{k_{X_i}}\right]$. We then normalized the trajectories using the mean and standard deviation of each node expression obtained from RACIPE and converted the trajectories to EM scores and PD-L1 levels in order to classify them into the observed phenotypes.

Using these trajectories, we constructed obtained a probability density (P) of the EM-PD-L1 score pairs and constructed a potential landscape by calculating the pseudo potential as – log (P) (34).

Gene expression data analysis

Publicly available microarray datasets and RNA-Seq datasets were obtained from GEO. Single-sample gene set enrichment analysis (ssGSEA) (35) was performed on the Hallmark signalling pathways gene signatures from MSigDB (Molecular Signatures Database) (36) to estimate the activity of pathway.

sRACIPE simulations

sRACIPE simulations has been performed to generate random set of parameters and to simulate the system with a fixed amount of noise. We have used the webserver of Gene Circuit Explorer (GeneEx) to simulate stochastic time evolution dynamics of our core gene regulatory circuits (22). Parameter values used for the simulation are presented in **Table S5**.

Stemness circuit and stemness score calculation

We have considered a gene regulatory network shown in **Fig S4C** in which 5 nodes of our core regulatory network are present along with 4 other nodes (OCT4, miR-145, LIN28, and let7) which represents the key players of stemness signature (**Table S1)**. The stemness scores (SN) were calculated by difference in normalized expression values of node representing stem-like and non-stem-like signatures: (LIN28 + OCT4 – let7 – miR145)/4. Subsequently based on the distribution, cells were categorised into non-stem-like (SN score <-0.5 and SN score >0.5), and stem-like (SN score = -0.5 to 0.5) represented in **Fig 3C**.

EM score calculation

The scores were calculated by difference in normalized expression values of node representing mesenchymal (M) and epithelial (E) signatures. EM score = (ZEB1 + SLUG - miR200 - CDH1)/4. Subsequently based on the distribution, cells were categorised into epithelial (<-0.25), hybrid (-0.25 to 0.5), and mesenchymal (>0.5) represented in **Fig 1B**.

RACIPE output of core circuit and its z-normalization

RACIPE generates the steady-state values in the log2 scale, which we have further converted into z-scores by using the equation represented in **Eq. (1)** for a better comparative study of the expression of each gene node.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$Z_i = \frac{S_i - \overline{S_m}}{\sigma_m} \quad (1)$$

Z_i = Z normalised expression value,

S_i = each steady state value of a given node,

$\overline{S_m}$ = combined mean of untransformed expression values.

σ_m = combined standard deviation of untransformed expression values.

We have used all z normalised steady-state solutions up to penta-stable parameter sets for principal component analysis (PCA) plots. PCA was performed on all the z-normalised steady-state solutions. We identified 3 optimal clusters by performing hierarchical clustering on the z-normalised RACIPE data. Contributions of the various node to the principal component axes PC-1 and PC-2 presented in **Table S3**.

For RACIPE, we have used 10,000 parameter sets and 100 initial conditions for each mathematical model to integrate the dynamical equations using Euler's Method numerically. RACIPE takes a topology file as an input (**Table S1**) and samples the parameters for dynamical simulations from a biologically relevant range (**Table S2**). Depending on the particular parameters, a single model has the potential to give rise to one or more stable steady-state solutions, dependent upon the initial conditions. However, in our current analysis, we have considered up to 5 stable steady-state solutions. As from our initial analysis, we have found a minuscule contribution for >5. We have also performed both overexpression and downexpression of miR-200 and SLUG by 20 fold respectively using RACIPE on the same core circuit.

References

1. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24:541–50.

2. Sun C, Mezzadra R, Schumacher TN. Regulation and Function of the PD-L1 Checkpoint. Immunity. Elsevier Inc.; 2018;48:434–52.

3. Hashimoto M, Kamphorst AO, Im SJ, Kissick HT, Pillai RN, Ramalingam SS, et al. CD8 T Cell Exhaustion in Chronic Infection and Cancer: Opportunities for Interventions. Annu Rev Med. 2018;69:301–18.

4. Chen L, Gibbons DL, Goswami S, Cortez MA, Ahn Y, Byers L a, et al. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. Nat Commun. 2014;5:5241.

5. Dongre A, Rashidian M, Reinhardt F, Bagnato A, Keckesova Z, Ploegh HL, et al. Epithelial-to-mesenchymal transition contributes to immunosuppression in breast carcinomas. Cancer Res. 2017;77:3982–9.

6. Guo Y, Lu X, Chen Y, Rendon B, Mitchell RA, Cuatrecasas M, et al. Zeb1 induces immune checkpoints to form an immunosuppressive envelope around invading cancer cells. Sci Adv. 2021;7:eabd7455.

7. Jolly MK, Somarelli JA, Sheth M, Biddle A, Tripathi SC, Armstrong AJ, et al. Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. Pharmacol Ther. 2019;194:161–84.

8. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. Proc Natl Acad Sci. 2013;110:18144–9.

9. Subbalakshmi AR, Sahoo S, Biswas K, Jolly MK. A computational systems biology approach identifies SLUG as a mediator of partial Epithelial-Mesenchymal Transition (EMT). Cells Tissues Organs. 2021;in press.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10. Sterneck E, Poria DK, Balamurugan K. Slug and E-Cadherin: Stealth Accomplices? Front Mol Biosci. 2020;7:138.

11. Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, et al. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep. 2008;9:582–9.

12. Wels C, Joshi S, Koefinger P, Bergler H, Schaider H. Transcriptional activation of ZEB1 by Slug leads to cooperative regulation of the epithelial-mesenchymal transition-like phenotype in melanoma. J Invest Dermatol. 2011;131:1877–85.

13. Liu YN, Yin JJ, Abou-Kheir W, Hynes PG, Casey OM, Fang L, et al. MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. Oncogene. 2013;32:296–306.

14. Chen L, Xiong Y, Li J, Zheng X, Zhou Q, Turner A, et al. PD-L1 Expression Promotes Epithelial to Mesenchymal Transition in Human Esophageal Cancer. Cell Physiol Biochem. 2017;42:2267–80.

15. Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, Onuchic JN. Interrogating the topological robustness of gene regulatory circuits by randomization. PLoS Comput Biol. 2017;13:e1005456.

16. Juneja VR, McGuire KA, Manguso RT, LaFleur MW, Collins N, Nicholas Haining W, et al. PD-L1 on tumor cells is sufficient for immune evasion in immunogenic tumors and inhibits CD8 T cell cytotoxicity. J Exp Med. 2017;214:895–904.

17. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X. An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. PLOS Comput Biol. 2015;11:e1004569.

18. George JT, Jolly MK, Xu S, Somarelli JA, Levine H. Survival outcomes in cancer patients predicted by a partial EMT gene expression scoring metric. Cancer Res. 2017;77:6415–28.

19. Hong D, Messier TL, Tye CE, Dobson JR, Fritz AJ, Sikora KR, et al. Runx1 stabilizes the mammary epithelial cell phenotype and prevents epithelial to mesenchymal transition. Oncotarget. 2017;8:17610–27.

20. Celià-Terrassa T, Meca-Cortés Ó, Mateo F, De Paz AM, Rubio N, Arnal-Estapé A, et al. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. J Clin Invest. 2012;122:1849–68.

21. Padda SK, Riess JW, Schwartz EJ, Tian L, Kohrt HE, Neal JW, et al. Diffuse high intensity PD-L1 staining in thymic epithelial tumors. J Thorac Oncol. 2015;10:500–8.

22. Kohar V, Lu M. Role of noise and parametric variation in the dynamics of gene regulatory circuits. npj Syst Biol Appl. 2018;4:40.

23. Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. bioRxiv. 2021;435359.

24. Tripathi S, Levine H, Jolly MK. The Physics of Cellular Decision-Making during Epithelial-Mesenchymal Transition. Annu Rev Biophys. 2020;49:1–18.

25. Xu J, Lamouille S, Derynck R. TGF-B-induced epithelial to mesenchymal transition. Cell Res. 2009;19:156–72.

26. Ranganathan P, Weaver KL, Capobianco AJ. Notch signalling in solid tumours: A little bit of everything but not all the time. Nat Rev Cancer. 2011;11:338–51.

27. Koppikar P, Lui VWY, Man D, Xi S, Chai RL, Nelson E, et al. Constitutive Activation of STAT5 Contributes to Tumor Growth, Epithelial-Mesenchymal Transition, and Resistance to EGFR Targeting. Clin Cancer Res. 2008;14:7682–90.

28. Wu Y, Zhou BP. TNF-α/NFκ-B/Snail pathway in cancer cell migration and invasion. Br. J.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cancer. 2010.

29. Sullivan NJ, Sasser AK, Axel AE, Vesuna F, Raman V, Ramirez N, et al. Interleukin-6 induces an epithelial-mesenchymal transition phenotype in human breast cancer cells. Oncogene. 2009;28:2940–7.

30. Yi M, Niu M, Xu L, Luo S, Wu K. Regulation of PD-L1 expression in the tumor microenvironment. J Hematol Oncol. 2021;14:10.

31. Hai Ping P, Feng Bo T, Li L, Nan Hui Y, Hong Z. IL-1β/NF-kb signaling promotes colorectal cancer cell growth through miR-181a/PTEN axis. Arch Biochem Biophys. 2016;604:20–6.

32. Morel A-P, Lièvre M, Thomas C, Hinkal G, Ansieau S, Puisieux A. Generation of breast cancer stem cells through epithelial-mesenchymal transition. PLoS One. 2008;3:e2888.

33. Jolly MK, Huang B, Lu M, Mani SA, Levine H, Ben-Jacob E. Towards elucidating the connection between epithelial − mesenchymal transitions and stemness. J R Soc Interface. 2014;11:20140962.

34. Wang J, Zhang K, Xu L, Wang E. Quantifying the Waddington landscape and biological paths for development and differentiation. Proc Natl Acad Sci U S A. 2011;108:8257–62.

35. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.

36. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739–40.

Acknowledgement

MKJ was supported by Ramanujan Fellowship (SB/S2/RJN-049/2018) awarded by Science and Engineering Research Board (SERB), Department of Science & Technology, Government of India. KH acknowledges support from Prime Ministers' Research Fellowship (PMRF). SS acknowledges support by KVPY fellowship.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental figures

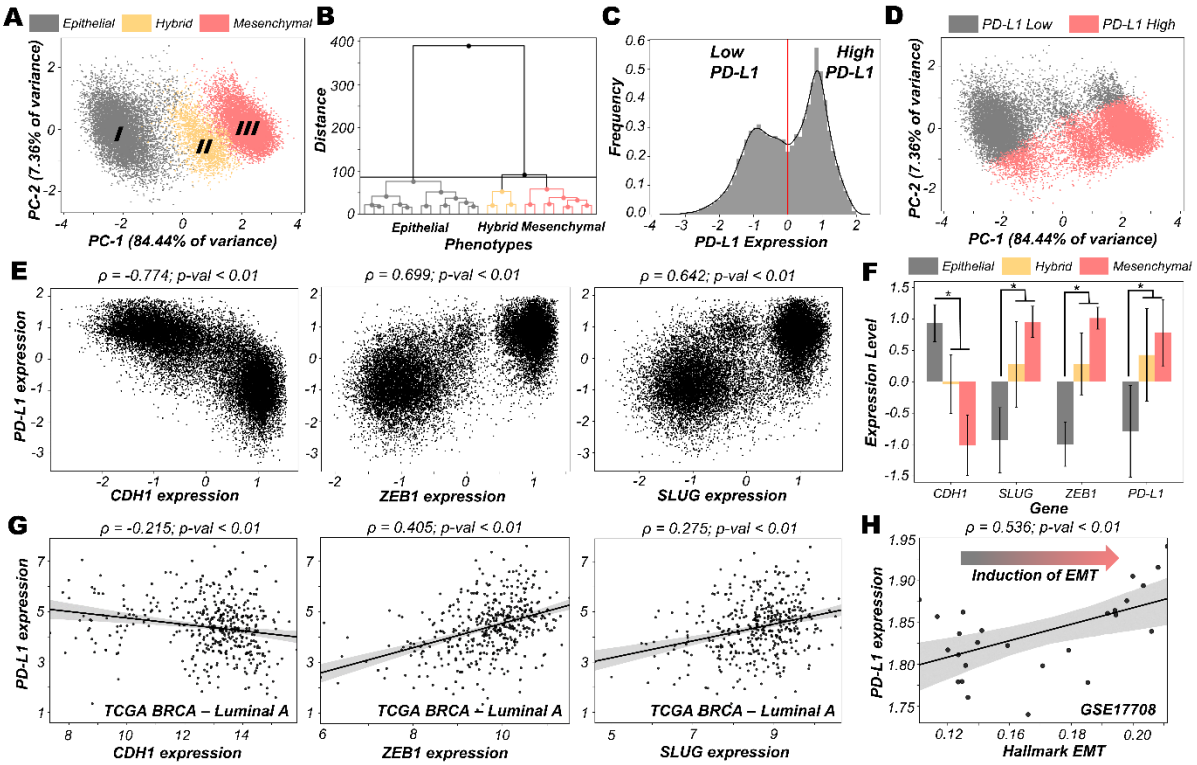


Fig S1. EMT regulatory network coupled with PD-L1. A) *PCA (Principal Component Analysis) plot showing the presence of different clusters emerging from z-normalised scores from RACIPE analysis. Composition of PC1 and PC2 are listed in Table S3.* ***B)*** *Hierarchical clustering for z-normalised RACIPE output.* ***C)*** *Density histogram of PD-L1 expression fitted with kernel density estimate, showing bimodality. Red lines show the partition between PD-L1 high and PD-L1 low.* ***D)*** *PCA plot colored by PD-L1 high vs. PD-L1 low levels, showing the enrichment of high PD-L1 levels in hybrid E/M and mesenchymal phenotypes, and that of low PD-L1 levels in an epithelial phenotype.* ***E)*** *Simulation results showing scatter plot of PD-L1 expression with CDH1, ZEB1, and SLUG, as obtained from RACIPE simulations. Spearman's correlation coefficient (ρ) and corresponding p-value (p-val) are reported.* ***F)*** *Bar graph showing expression of CDH1, SLUG, ZEB1 and PD-L1 in corresponding phenotypes (defined based on EM scores) respectively.* ***G)*** *Scatter plot showing experimental validation from TCGA BRCA – Luminal A cohort of patients of correlations between expression of PD-L1 with CDH1, ZEB1, and SLUG, which was earlier represented in E).* ***H)*** *Scatter plot showing positive correlation between PD-L1 expression and Hallmark EMT signature in A549 lung adenocarcinoma cells treated with TGFb to induce EMT over 3 days (GSE17708). * denotes a statistically significant difference (p-val < 0.05) between the represented groups assessed by a two-tailed Students t-test assuming unequal variances.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

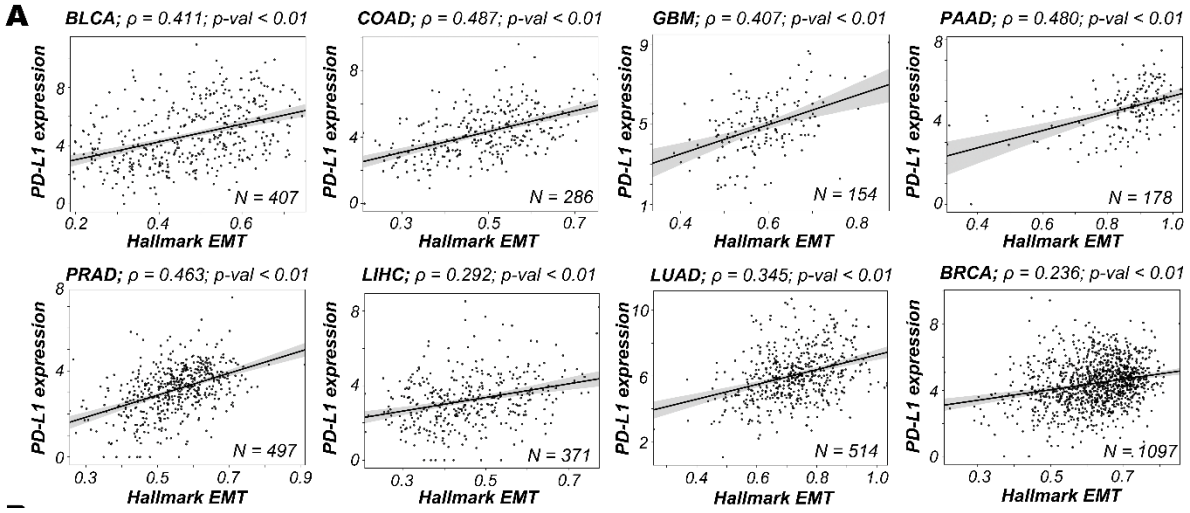


Fig S2. Clinical evidence supporting mathematical model predictions. A) *Scatter plots between expression levels of PD-L1 and Hallmark EMT in representative TCGA cancer types. Spearman's correlation coefficient (ρ) and corresponding p-value (p-val) are reported.*

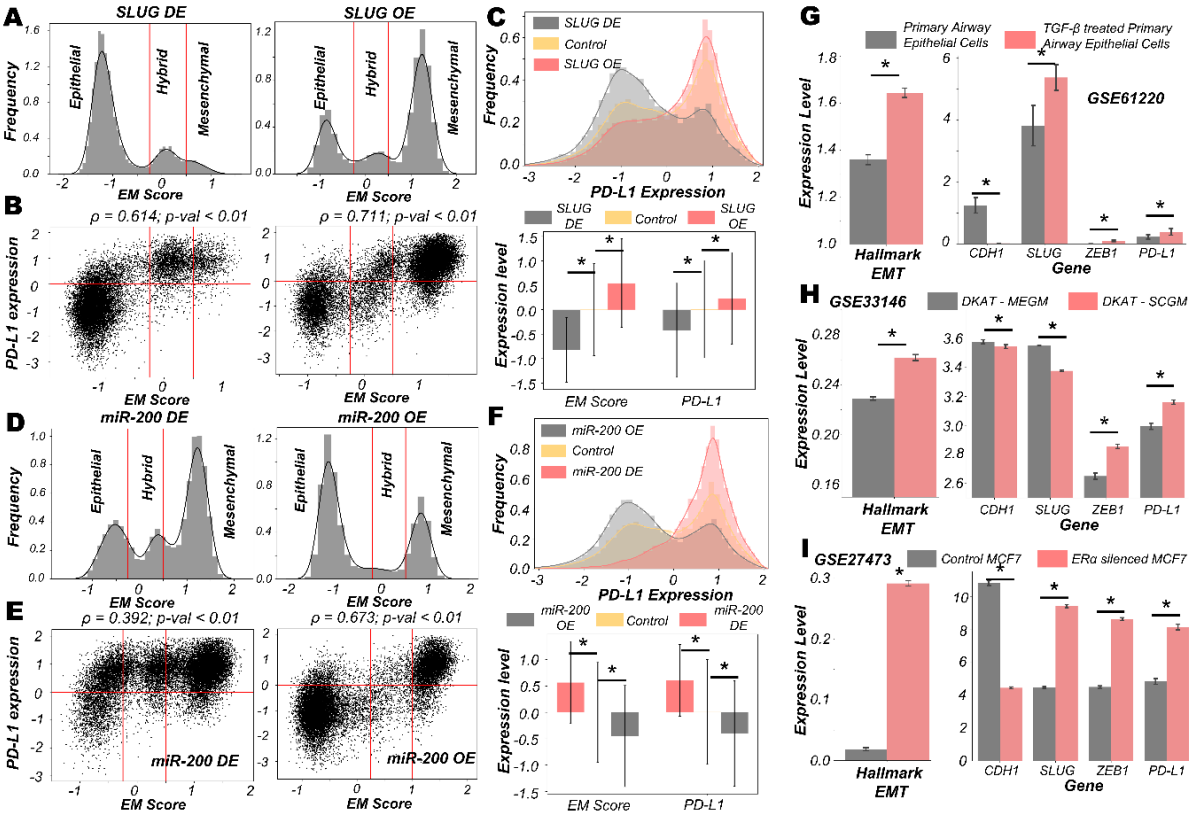


Fig S3. Dynamics upon perturbation of core regulatory network. A-C) *Upon miR-200 down expression (DE) and miR-200 over expression (OE):* ***A)*** *density histogram of EM Score fitted with kernel density estimate;* ***B)*** *Scatter plot of PD-L1 expression and EM score;* ***C)*** *Density histogram of PD-L1 expression fitted with kernel density estimate and Bar graph showing change in expression of EM score and PD-L1.* ***D-F)*** *Same as A-C but for SLUG DE and SLUG OE. Horizontal red line shows the partition between PD-L1 expression level being high and low. Vertical red lines*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.21.449285; this version posted June 22, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

show a partition between phenotypes: Epithelial, Hybrid E/M, and Mesenchymal based on EM score. Spearman's correlation coefficient (ρ) and corresponding p-value (p-val) are given. ***G)*** *Activity/Expression levels of Hallmark EMT and PD-L1 levels in non-cancerous airway epithelial cells where EMT has been induced (GSE61220).* ***H)*** *Activity/Expression levels of Hallmark EMT and PD-L1 levels in triple negative breast cancer (DKAT) cells grown in either epithelial growth medium (MEGM) or stromal growth medium (SCGM) (GSE33146).* ***I)*** *Activity/Expression levels of Hallmark EMT and PD-L1 levels in ER+ breast cancer cells (MCF7) cells in control or ERα silenced conditions (GSE27473). * denotes a statistically significant difference* (p-val < 0.05) *between* the represented groups *assessed by a two-tailed Students t-test assuming unequal variances.*

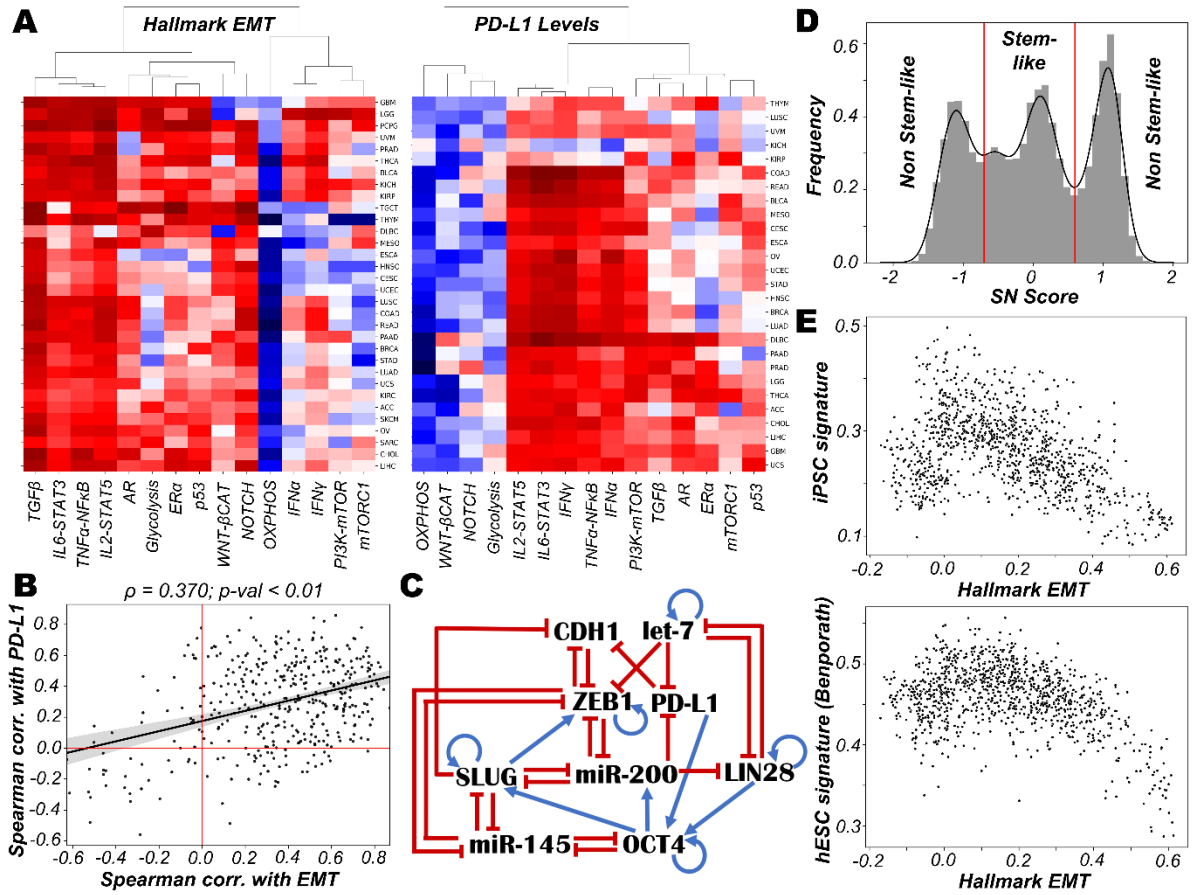


Fig S4. Different pathways that may influence the EMT/PD-L1 association. A) *Heatmap showing Spearman's correlation between various signalling pathways and Hallmark EMT/PD-L1 levels respectively. Spearman's correlation coefficient (ρ) and e corresponding p-value (p-val) are reported.* ***B)*** *Scatter plots between the Spearman's correlation of expression levels of PD-L1 and spearman correlation of Hallmark EMT showing the concordance between two heatmaps in (A).* ***C)*** *Schematic representation of stemness circuit diagram with nodes representing various EMT, immune evasion, and stemness signature players.* ***(D)*** *Density histogram of Stemness Score (SN score) (LIN28 + OCT4 – let7 – miR145)/4 fitted with kernel density estimate showing predominantly a trimodal distribution. Vertical red lines show the partition between stem-like and non-stem-like based on SN score; where intermediate levels of SN score lie within the 'stemness window'.* ***E)*** *Scatter plots between expression levels of PD-L1 with iPSC signature and hESC signature (Benporath et al. 2008) respectively in CCLE datasets.*