Data-driven learning how oncogenic gene expression locally alters heterocellular networks

David J. Klinke II[1-3] ✉, **Audry Fernandez**[2,3], **Wentao Deng**[2,3], **and Anika C. Pirkey**[1]

[1]Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV
[2]Department of Microbiology, Immunology and Cell Biology, West Virginia University, Morgantown, WV
[3]WVU Cancer Institute, West Virginia University, Morgantown, WV

Discovering and developing pharmaceutical drugs increasingly relies on mechanistic mathematical modeling and simulation. In immuno-oncology, models that capture causal relations among genetic drivers of oncogenesis, functional plasticity, and host immunity provide an important complement to wet experiments, given the cellular complexity and dynamics within the tumor microenvironment. Unfortunately, formulating such mechanistic cell-level models currently relies on hand curation by experts, which can bias how data is interpreted or the priority of drug targets. In modeling molecular-level networks, rules and algorithms have been developed to limit a priori biases in formulating mechanistic models. To realize an equivalent approach for cell-level networks, we combined digital cytometry with Bayesian network inference to generate causal models that link an increase in gene expression associated with oncogenesis with alterations in stromal and immune cell subsets directly from bulk transcriptomic datasets. To illustrate, we predicted how an increase in expression of Cell Communication Network factor 4 (CCN4/WISP1) altered the tumor microenvironment using data from patients diagnosed with breast cancer and melanoma. Network predictions were then tested using two immunocompetent mouse models for melanoma. In contrast to hand-curated approaches, we posit that combining digital cytometry with Bayesian network inference provides a less biased approach for elaborating mechanistic cell-level models directly from data.

Heterocellular networks | digital cytometry | deconvolution | anti-tumor immunity | Bayesian network inference | functional plasticity

Correspondence: ***david.klinke@mail.wvu.edu***

Introduction

Tissues are dynamic structures where different cell types organize to maintain function in a changing environment. For instance, the mammary epithelium reorganizes during distinct stages of the ovarian cycle in preparation for lactation (Klinke, 2016). At the same time, immune cells clear dead cells and defend against pathogens present in the tissue microenvironment. Ultimately, the number and functional orientation of different cell types within a tissue interact to create a network, that is a heterocellular network. This heterocellular network is essential for creating and maintaining tissue homeostasis. While we know that tissue homeostasis is disrupted during oncogenesis, our understanding of how genetic alterations quantitatively and dynamically influence the heterocellular network within malignant tissues in humans is not well developed despite large efforts, like The Cancer Genome Atlas (TCGA), to characterize the genomic and transcriptomic landscape in human malignancy (Hoadley et al., 2018; Wells and Wiley, 2018). In parallel with these large scale data gathering efforts, two informatic developments, namely digital cytometry and Bayesian network inference, may be helpful in interrogating these datasets.

In cytometry, single-cell sequencing technology elicits a lot of excitement as it enables unbiased discovery of novel cell subsets in particular disease states (Papalexi and Satija, 2018; Singer and Anderson, 2019). Unfortunately, persistent challenges related to confounding of batch effects with biological replicates limit the statistical power of these datasets to link oncogenic transcriptional changes with reorganization of the cellular network (Grun et al., 2014; Stuart and Satija, 2019). Due to the high number of biological replicates, transcriptomic datasets, such as the Cancer Genome Atlas, provide a rich resource in characterizing the heterogeneity of oncogenic transformation. Yet, these data were obtained from homogenized tissue samples and reflect the expression of genes averaged across a heterogeneous cell population. Computationally, "Digital Cytometry" can deconvolute the prevalence of individual cell types present within a mixed cell population (Newman et al., 2019). The approach stems from the idea that the influx of a particular cell subset into a tissue corresponds to an increase in a gene signature uniquely associated with this particular cell subset (Shen-Orr et al., 2012; Yoshihara et al., 2013; Wang et al., 2018; Zaitsev et al., 2019). Gene signatures of immune cells have been developed in a number of studies, which increasingly leverage scRNAseq data and machine-learning methods (Shen-Orr et al., 2010; Becht et al., 2016; Schelker et al., 2017; Torang et al., 2019). Besides representing different cellular subsets, gene signatures can also represent intracellular processes associated with oncogenesis, like the epithelial-mesenchymal transition (Tan et al., 2014; Koplev et al., 2018; Malta et al., 2018; George et al., 2017; Klinke and Torang, 2020). Though, the predictive value of many of these tissue "features" in inferring how heterocellular networks are altered in diseased tissues remain unclear, as establishing correlations among features tends to be the end point of studies (e.g., (Tosolini et al., 2011; Malta et al., 2018; Thorsson et al., 2018)).

Increases in size and information content of transcrip-

tomic datasets enable using probabilistic inference methods to identify relationships within the data that could not be observed using simpler statistical techniques (Hill et al., 2016; Friedman, 2004). However to infer how heterocellular networks are altered in diseased tissues, we need to be able to identify the direction of information flow within the network, that is the causal relationships among interacting components. One method to identify the topology of a causal network in an unbiased way is to use algorithms that identify Bayesian networks (Scutari, 2010). Bayesian networks are a type of directed acyclic graphs (DAG), where each node represents a random variable, or "feature", and each edge represents a causal relationship between two nodes. As algorithms for reconstructing Bayesian networks emerged, they were used to model signaling pathways within cells (Sachs et al., 2002), to identify known DNA repair networks in *E. coli* using microarray data (Perrin et al., 2003) and to identify simple phosphorylation cascades in T lymphocytes using flow cytometry data (Sachs et al., 2005, 2009). While many more studies have been published since, a common conclusion is that the statistical confidence associated with an inferred network improves as the number of samples included in a dataset is greater than the number of random variables. However, transcriptomics data, like that obtained as part of the TCGA, typically have a large number of random variables (n_{genes}) and a small number of biological replicates ($n_{patients}$), which makes inferring gene-level networks computationally difficult (Zou and Conzen, 2005).

As summarized in Figure 1, we propose an approach that combines digital cytometry with Bayesian network inference to identify how heterocellular networks associated with functional plasticity and anti-tumor immunity change during oncogenesis in humans. Conceptually, digital cytometry improves the statistical power by projecting the transcriptomic space onto a smaller number of "features" that estimate the prevalence of stromal and immune cell types and the average differentiation state of malignant cells present within the tumor microenvironment, such that $n_{features} << n_{patients}$. The causal structure among these features can then be predicted using Bayesian network inference. While data unstructured in time, such as the TCGA datasets, are not ideal for inferring causality, we test the inferred networks using in vivo experiments using syngeneic murine tumor models.

To illustrate the approach, we focused on Cell Communication Network factor 4 (CCN4/WISP1), as it is upregulated in invasive breast cancer (Klinke, 2014) and correlates with a lower overall survival in patients diagnosed with primary melanoma (Deng et al., 2019). Functionally, CCN4 promotes metastasis in melanoma by promoting a process similar to the epithelial-mesenchymal transition (Deng et al., 2019, 2020). In developing state metrics that quantify functional plasticity in breast cancer and melanoma using an unsupervised approach, CCN4 was the only gene associated with both a mesenchymal state metric in breast cancer and a de-differentiated state metric in melanoma that results in a secreted protein (Klinke and Torang, 2020). The collective set of features, or simply nodes of a network, were quantified in three transcriptomic datasets obtained from bulk tissue samples from patients with breast cancer and melanoma and used to generate a casual network describing how expression of a "gene driver" associated with oncogenesis, such as CCN4, alters the heterocellular network within a tissue using Bayesian network inference.

Results

Generating causal graphs that link oncogenic changes in gene expression with changes in the heterocellular network. Bayesian network inference involves inferring the structure of the network, which captures the specific interactions or edges among the nodes of a network and represents them as a directed acyclic graph (DAG), and then estimating the parameters of the conditional probability distribution from the datasets. We used a two-step process to learn the causal structure associated the cell-level networks. First, we created a collection of edges that were consistently identified among the different structural learning algorithms, that is a consensus seed network. In the initial structure learning step, an overall flow of the network was specified by limiting the inclusion of edges into a proposed network. In particular, we considered only edges into the "CD8 T cells" node (i.e., a leaf node), only edges that originate from the "Cancer" node (i.e., a root node), mostly edges that originate from the "CCN4" node (with exception for the "Cancer" node), and only edges into the "CD4 T cells" and "Neutrophils" nodes. Specifying "CD4 T cells" and "Neutrophils" as leaf nodes follows from the high number of zero values for those features in the dataset, which were 350 and 439 samples in the BRCA dataset, respectively. This was implemented by assigning the corresponding edges to a "blacklist". Collectively, this represents a way to incorporate prior knowledge about causal relationships associated with oncogenesis and the roles that specific immune cells play in controlling tumor cell growth.

As algorithms for structural learning have different underlying assumptions, we used an ensemble approach to average across the different algorithms to identify an initial structure of the DAG. Specifically, we used nine different structural learning algorithms (Scutari, 2010), including a pairwise mutual information (ARACNE), constraint-based (Incremental association Markov Blanket - IAMB, Incremental association with false discovery rate control - IAMB.FDR, practical constraint - PC.STABLE), local discovery of undirected graphs (max-min parents and children - MMPC, Hiton parents and children - SI.HITON.PC), score-based (hill climbing - HC, Tabu search - Tabu), and hybrid learning (max-min hill-climbing - MMHC) algorithms. Bootstrap resampling was used in learning the network structure with each algorithm, which resulted in generating 10,000 network structures. For each algorithm, an averaged network structure was calculated from the collection of network structures, where the threshold for inclusion of a edge into the average network was automatically determined by each algorithm and was nominally 0.5. Whether a particular edge promotes or inhibits the target node was determined based on the sign of the cor-

relation coefficient between the two nodes. We applied the same approach to both the breast cancer (BRCA - Figures 2 and 3) and the two melanoma datasets (common melanocytic nevi and primary melanoma: GEO and primary melanoma from the TCGA: SKCM - Figures S4 and 5). To generate consensus seed networks, the BRCA dataset was analyzed alone (see Table S1) while results for the two melanoma datasets (see Tables S2 (GEO) and S3 (SKCM)) were used together. Including edges in the consensus seed network was based on consistency among algorithms, strength of the edge, and whether the edge provided a new connection linking the "Cancer" node with effector immune cell nodes, such as "CD4 T cells" or "CD8 T cells", or potential negative feedback mechanisms, which is illustrated graphically in Figures 2 and S4. For instance, in analyzing the BRCA dataset, edge numbers 32 ("Cancer" $\rightarrow$ "pM1"), 37 ("Cancer" $\rightarrow$ "B cells naive"), 45 ("CCN4" $\rightarrow$ "Macrophages"), 46 ("Cancer" $\rightarrow$ "resting NK cells"), and 53 ("CCN4" $\rightarrow$ "active NK cells") were included as they provided novel edges to the consensus seed network. The inferred direction of a number of edges varied among the algorithms (yellow bars in Figures 2 and S4) and were left out of the consensus seed network. The final network for each dataset was generated using a hybrid learning algorithm (mmhc) using a "blacklist" specified based on prior causal knowledge and a "whitelist" corresponding to the consensus seed network. Similar to the first step, bootstrap resampling (n_{boot} = 10,000) and network averaging were used to generate the DAGs shown in Figures 3 and 5. The averaged DAG was used to generate parameters for a linear Gaussian model estimated by maximum likelihood and conditioned on the network structure that approximates the joint probability distribution associated with the dataset. Values for the linear coefficients and the average node values were used to annotate the DAGs.

Oncogenesis in breast cancer was associated with a shift from epithelial to mesenchymal cell state accompanied by an increase in cell proliferation and a suppression of endothelial cells, which were inferred with high confidence. In turn, endothelial cells promote the infiltration of CD4 T cells. The local structure associated with "Cancer"'s influence on the "Mesenchymal" state via "CCN4" suggests an incoherent type-3 feed-forward motif to regulate the mesenchymal state, with CCN4 also inhibiting active NK cells. The high confidence edge between active NK and resting NK cells follows from these features being mutually exclusive in the dataset and very few samples having zero values for both features. The mesenchymal state increased cancer-associated fibroblasts ("s) with high confidence. Interestingly, oncogenesis was also associated with increasing the prevalence of a type 1 macrophage, which in turn promoted the recruitment of CD8 T cells. The prevalence of CD8 T cells are also connected to "Cancer" via a larger incoherent feed-forward motif involving "CCN4" and "CAFs" with high confidence.

As there was more data supporting the BRCA DAG, the resulting Bayesian network model was compared against the underlying experimental data and used to explore the impact of varying CCN4 expression in the context of normal and tumor tissue (Figure 4). To simulate "normal" and "tumor" tissue, we queried the conditional probability distribution by generating samples from the Bayesian network and filtered the values based on p("Cancer" < 0.15) and p("Cancer" $>$ 0.85), which are colored in orange and blue, respectively. Corresponding experimental data points and trendlines are overlaid upon the posterior distributions. The posterior distributions mirror the experimental data points, where there is an increase in CCN4 expression between "normal" and "tumor" tissue. The posterior distributions mirror the variability observed in the experimental data when comprised of non-zero values, such as CD8 T cells. In contrast, the prevalence of zero values increased the range of the posterior distribution, such as for CD4 T cells. In comparing "normal" to "tumor" tissue, CD8 T cells was the only feature, on average, increased in "tumor" tissue, while CD4 T cells, B cells, and cancer associated fibroblasts were decreased. Slopes of the trendlines highlight the influence of CCN4 gene expression on the prevalence of different immune cell populations. Increased CCN4 had the most pronounced inhibition on NK cells and also suppressed CD8 T cells. CCN4 also had a pronounced positive impact on the prevalence of cancer-associated fibroblasts, macrophages, and slightly promoted CD4 T cells. CCN4 seemed to have little to no impact on B cells in "normal" tissue while inhibited B cells in "tumor" tissue.

The breast cancer dataset contained 582 samples, of which 8.8% were from normal mammary tissue. In contrast, the two melanoma datasets contained 78 GEO samples, which includes 34.6% benign nevi, and 94 SKCM samples of primary melanoma only. While a lower number of samples limits the inferential power of a dataset, we decided not to combine them together as they had different distributions in transcript abundance as a function of transcript length. As the Bayesian network inference algorithm leverages differences in the magnitude of a feature within a population, approaches to harmonize these two datasets may introduce a systemic bias that is convoluted with oncogenic transformation, as the GEO dataset has many samples obtained from benign nevi while the SKCM dataset does not. We decided to analyze the melanoma datasets separately and combine the enriched edges in each dataset into a consensus seed network that reflects both datasets. In analyzing the melanoma datasets, edge numbers 26 ("CAF" $\rightarrow$ "CD8 T cells") and 30 ("CCN4" $\rightarrow$ "CAF") in the GEO analysis while edge numbers 17 ("CCN4" $\rightarrow$ "CAF"), 18 ("CAF" $\rightarrow$ "CD8 T cells"), and 22 ("Active NK cells" $\rightarrow$ "CD8 T cells") from the SKCM analysis were included in the consensus seed network. This consensus seed network was then included in the "whitelist" to learn the structure and parameters associated with Bayesian network inferred from the melanoma datasets (see Figure 5).

Given the high prevalence of samples from benign nevi in the GEO dataset, high confidence edges in the GEO network focus on changes associated with oncogenesis. Similar to the breast cancer analysis, oncogenesis was associated with a shift from an epithelial to a mesenchymal-like cell state and

the promotion of cell proliferation. Here, the mesenchymal cell state is promoted by both oncogenesis and CCN4 with a coherent feed-forward motif. Similar to the breast cancer analysis, oncogenesis promoted an increase in CD8 T cells, but indirectly by recruiting active NK cells. In contrast to the breast cancer analysis, CCN4 directly impacted CAFs and resting NK cells, although the "CCN4" $\rightarrow$ "resting NK cells" edge had both low confidence and low influence parameter. In analyzing the SKCM dataset, less emphasis is placed on the changes associated with oncogenesis but how expression of CCN4 influenced the network. Similarly to the GEO analysis, the SKCM dataset suggested that CCN4 directly impacted the mesenchymal state, CAFs, and resting NK cells, but the influence on resting NK cells changed from a slight inhibition in the GEO dataset (-0.11) to strong promotion in the SKCM dataset (0.75). In addition, the edge between CAFs and the mesenchymal state was inferred with high confidence but changing direction between GEO and SKCM datasets suggests that the algorithms were unable to discern edge direction from the data. In both melanoma datasets, CAFs influence CD8 T cells via an incoherent feed-forward motif involving the prevalence of macrophages. In addition, Neutrophils, macrophage polarization, and B cells were independent of oncogenesis and CCN4 expression. In all three analysis, there was high confidence associated with the edges among the nodes quantifying macrophage polarization, which is likely an artifact of formula used to calculate $p(M\Phi i)$'s. Queries of the conditional probability distribution based on the SKCM DAG for CD8 T cells, active NK cells, Macrophages, B cells, and CAFs were similar to the BRCA analysis (Fig. S5). Similar to the BRCA analysis, a high number of zero values for the CD4 T cell features in the SKCM dataset suggests caution in interpreting differences in CD4 T cell predictions.

Validating the impact of CCN4 on heterocellular networks using syngeneic mouse models. Syngeneic immunocompetent mouse models of cancer provide an important complement to retrospective studies of human data as they can aid in causally linking genetic alterations with cellular changes the tumor microenvironment. Here we used two syngeneic transplantable models for melanoma to test the predictions generated by the Bayesian network inference: the spontaneous B16F0 model and the YUMM1.7 model that displays Braf$^{V600E/WT}$ Pten$^{-/-}$ Cdkn2$^{-/-}$ genotype. As these cell lines basally produce CCN4, we generated CCN4 knock-out (KO) variants of these parental cell lines using a CRISPR/Cas9 approach and confirmed CCN4 KO by testing conditioned media for CCN4 by ELISA. Tumors were generated by injecting the cell variants subcutaneously in 6-8 week old female C57BL/6 mice and monitoring for tumor growth (n = 5 in each group). Once wt tumors reached between 1000 and 1500 mm^3 in size, tumors were surgically removed from all mice that were not considered outliers and processed into single cell suspensions (n = 3 for YUMM1.7 variants and n = 4 for B16F0 variants). The single cell suspensions were aliquoted among three antibody panels to characterize the tumor infiltrating lymphocytes by flow cytometry (see Supplementary Figures S8-S10 for gating strategies). While the B16F0 and YUMM1.7 KO variants were generated using a double nickase CRISPR/Cas9 approach, similar results were obtained using a homology directed repair strategy (Fernandez et al., 2020; Deng et al., 2020).

The percentage of CD45$^+$ cells among total live cells exhibited a semi-log dependence on tumor size (Fig. 6A - B16F0: R^2 = 0.607, F-test p-value = 7.27E-6; YUMM1.7: R^2 = 0.830, F-test p-value = 1.48E-7), where CCN4 KO resulted in smaller tumors in both cell models with greater CD45$^+$ cell infiltration. As illustrated in Figure 6A, YUMM1.7 variants had a much higher dependence on tumor size than B16F0 variants. Conventionally, flow cytometry data are normalized to tumor size to estimate the prevalence of a particular cell type per tumor volume. Yet, the dependence on tumor size could be a confounding factor in addition to CCN4 expression that could skew the results. Moreover, the Bayesian network analysis predicts the impact of CCN4 alone on the prevalence of specific immune cell subsets. Thus, we focused instead on the prevalence of a particular cell type within the live CD45$^+$ TIL compartment to compare against the Bayesian network predictions.

In comparing the wt B16F0 and YUMM1.7 models, the relative prevalence of NK, CD4$^+$ T, and CD8$^+$ T cells were similar while B cells were almost 10-times more prevalent in the B16F0 tumors compared to YUMM1.7 tumors (Fig. 6B). The prevalence of these different cell types changed within the CD45$^+$ TIL compartment upon CCN4 KO (Fig. 6C and D). Figure 6C highlights the trends among the mouse models and compares against the Bayesian network predictions obtained from the BRCA and SKCM datasets. Predictions for the change in cell type prevalence by CCN4 expression were obtained by propagating a change in the "CCN4" node from 0 to 1 within the linear Gaussian model to the target node, which is also represented by the slope of the "Cancer" trendlines in Figures 4 and S5. Specifically, CD4 and CD8 T cells and B cells had analogous nodes in the Bayesian networks as assayed in the flow panel, while NK cells were mapped to "active NK cells" in the Bayesian network. The relative change in abundance was largely consistent among the four systems, with the YUMM1.7 model being the most different. NK cells were most reduced by CCN4, which was observed in the B16F0 model and BRCA and SKCM datasets. A reduction in CD8$^+$ T cells by increased CCN4 expression was consistent in both mouse models and both BRCA and SKCM datasets. The CD4$^+$ T cells seemed to vary where YUMM1.7 results seemed to be unaffected by CCN4 expression while the BRCA and B16F0 results showed an increase and SKCM showed a decrease. As stated previously, the BRCA and SKCM predictions for CD4 T cells should be interpreted with caution given the high frequency of zero values for the features. B cell response was mixed with both the BRCA and SKCM results suggesting no change and an increase in the B16F0 model and a decrease in the YUMM1.7 model, with the low number of B cells infiltrating YUMM1.7 tumors rendered the results more variable. Given the small sample size of the experimental mouse cohorts, only the extremes were

statistically significant, with NK cells significantly increased (p-value = 0.047) and B cells significantly decreased (p-value = 0.002) in B16F0 CCN4 KO tumors (Fig. 6D).

Concordance in CCN4-induced changes in the myeloid compartment are less clear. In addition to changes in T and NK cells within the live $CD45^+$ compartment, we also assayed myeloid subsets in tumors generated by wt and CCN4 KO variants of the B16F0 and YUMM1.7 cell lines. Using the gating strategy summarized in Figure S10, we focused on $CD11c^+$ and $CD11c^-$ macrophages and three different myeloid-derived suppressor cell (MDSC) subsets: granulocytic PMN-MDSC, monocytic Mo-MDSC, and $CD11c^+$ Mo-MDSC. In comparing tumors derived from wt cell lines, $CD11c^+$ macrophages were the most predominant infiltrating myeloid cell subset and the $CD11c^+$ subsets (macrophages and Mo-MDSC) were the most consistent (Fig. 7A). Upon CCN4 KO in the mouse models, the macrophage subsets tended to increase while the MDSC subsets decreased (Fig. 7B-E) within the $CD45^+$ compartment. The reduction in granulocytic PMN-MDSC in CCN4 KO variants were most pronounced and statistically significant (p=0.011 in B16F0 and p = 0.0011 in YUMM1.7). Conversely, neutrophils were increased within the $CD45^+$ compartment in CCN4 KO tumors derived from YUMM1.7 cells ($\log_{10}$ fraction in wt YUMM1.7: -1.49±0.03 versus CCN4 KO: -1.32±0.09, p = 0.041) but not statistically different in the B16F0 model (p = 0.10). Other myeloid subsets trended similarly but with differences that were not statistically significant, as expected given the small sample size. While Ly6G and Ly6C staining may have been a better staining strategy for distinguishing Mo-MDSC from PMN-MDSC, Rose et al. reported that there are likely overlapping epitopes between the Ly6G component of Gr1 and F4/80 (Rose et al., 2012). When stained simultaneously (as done here), the Gr1 antibody outcompetes the F4/80 mAb such that the $F4/80^-$ $Gr1^+$ cells are likely $Ly6G^+$ $Ly6C^-$ (PMN-MDSC) and the $F4/80^+$ $Gr1^+$ population is $Ly6G^-$ $Ly6C^+$ (Mo-MDSC) (see Fig. 7C and D). We observed a similar reduction in PMN-MDSCs in YUMM1.7 tumors upon CCN4 KO using Ly6G/Ly6C antibodies (Fernandez et al., 2020).

Comparing the trends in the myeloid compartment observed among the mouse models and the Bayesian network predictions obtained from the BRCA and SKCM datasets is less clear, given the uncertainty as to how the digital cytometry features map onto the quantified myeloid subsets in these mouse models. Key myeloid features in the Bayesian networks were macrophages oriented towards a M1 phenotype. Correspondingly, $CD11c^+$ macrophages, a subset that has been associated with pro-inflammatory M1 tumor-associated macrophages (Jeong et al., 2019), were the most predominant myeloid subset in wt B16F0 and YUMM1.7 tumors, which didn't change upon CCN4 KO. In the BRCA dataset, the prevalence of macrophages was influenced by CCN4 expression; yet, the functional orientation away from the M2 and towards the M1 phenotype depended solely on oncogenic transformation. Similarly, the prevalence of macrophages was influenced by both CCN4 expression and oncogenic transformation in both melanoma datasets. In contrast to the BRCA results, functional orientation of macrophages were independent of both oncogenic transformation and CCN4 expression. Neutrophils were predicted to be independent of CCN4 in the melanoma datasets, which is not surprising considering that the the majority of tumors had zero values for the Neutrophil feature (see Figs. S1-S3). Similarly, neutrophils were about 10 times less abundant than CD11c+ macrophages in the mouse models. Given the significant changes observed in MDSCs in the mouse models, challenging digital cytometry predictions in this way highlights features that can be improved, such as discriminating among terminally differentiated and immature subsets, like Mo-MDSC and PMN-MDSC.

CCN4 has no direct effect on T cell proliferation but alters $CD8^+$ T cell function. The local proliferation of $CD8^+$ T cells correlates with clinical response to immune checkpoint blockade (Huang et al., 2017; Twyman-Saint Victor et al., 2015). In addition, the DAGs inferred from both the breast cancer and melanoma datasets suggest that a decrease in $CD8^+$ T cells is driven indirectly through CCN4 via modulating cancer-associated fibroblasts or the activity of NK cells. While the structural learning algorithms rejected a direct edge between CCN4 and $CD8^+$ cells, we tested whether CCN4 directly inhibits T cell proliferation (see Fig. 8A and B) using a statistical analysis of Cell Trace distributions in $CD4^+$ and $CD8^+$ T cells stimulated in vitro (see Table S4). Specifically, splenocytes were stimulated in vitro with αCD3/αCD28-loaded beads in the presence of media conditioned by wt or CCN4 KO B16F0 cells or supplemented with 10 ng/ml recombinant mouse CCN4. In both the $CD4^+$ and $CD8^+$ T cell populations, the presence of tumor-conditioned media significantly inhibited the fraction of cells that divided at least once (Dil - CD4 p-value = 0.022, CD8 p-value = 0.018) and the probability that a cell will divide at least once (PF - CD4 p-value = 0.024, CD8 p-value = 0.013) while CCN4 exposure was not a statistically significant factor. For responding cells, the average number of divisions they undergo (PI) was not different among experimental conditions for $CD4^+$ T cells (p-value = 0.22) but reduced in $CD8^+$ T cells exposed to tumor-conditioned media (p-value = 0.0077). Overall, the presence of tumor-conditioned media and not CCN4 influenced T cell proliferation, which was consistent with the DAGs.

Another characteristic of $CD8^+$ T cells present within the tumor microenvironment is that they are dysfunctional (Li et al., 2019). As the digital cytometry approach used here doesn't estimate the functional state of $CD8^+$ T cells only their prevalence within a tissue sample, we decided to test whether CCN4 had a direct impact on $CD8^+$ T cell function, as quantified by target-specific ex vivo cytokine release as measured by ELISpot. First we generated YUMM1.7-reactive $CD8^+$ T cells by immunizing C57BL/6mice against YUMM1.7 cells and isolated $CD8a^+$ T cells from splenocytes three days after re-priming with live YUMM1.7 cells. We also created a variant of CCN4 KO YUMM1.7 cells with CCN4 expression induced by doxycycline and vector con-

trols that were used as target cells (see Fig. S11). IFNγ and TNFα ELISpots were used to quantify the CD8$^+$ T cell functional response to the different tumor targets in the presence or absence of tumor-produced CCN4. As expected for a re-call CD8a$^+$ effector T cell response, the most prominent IFNγ and lowest TNFα responses were against wt and CCN4 KO YUMM1.7 cells, with a slightly higher IFNγ response to wt YUMM1.7 targets (see Fig. 8C, p-value < 0.05). Interestingly, re-expression of CCN4 by CCN4 KO YUMM1.7 cells following doxycycline induction significantly reduced both IFNγ and TNFα production (p-value < 0.001), which suggests that CCN4 plays a direct role in inhibiting CD8a$^+$ T cell function. Of note, CCN4 was predicted to directly inhibit the activity of NK cells, which share cytokine release and cytotoxicity mechanisms with CD8$^+$ T cells. Overall, the changes observed between wt and CCN4 KO variants of the B16F0 and YUMM1.7 mouse models were consistent with the causal networks inferred from the breast cancer and melanoma datasets.

Discussion

Validating the role that a particular molecule plays in driving the disease state using targeted experiments is central for improving understanding of biological mechanisms or selecting among competing drug targets. Given the limited observability of the biological response in experimental models and patients, mechanistic modeling and simulation is playing an increasing role in helping answer many central questions in discovering, developing, and receiving federal approval of pharmaceutical drugs and also basic biology (Moore and Allen, 2019). In immuno-oncology, there is increasing interest in modeling the heterocellular network of relevance for a specific immunotherapy. The first step in creating mathematical models of cell-level networks is to create the topology of the network, which is expressed in terms of which nodes to include and how they influence each other. The structure of these cell-level models is created using a fully supervised approach, which means by hand using expert knowledge (Gadkar et al., 2016). For instance, systems of ordinary differential equations have been developed to capture multiple spatial compartments containing interacting malignant, antigen presenting, and T cells and to predict a general immune response (Palsson et al., 2013), a response to immune checkpoint blockade using CTLA-4, PD-1, and PD-L1 antibodies (Milberg et al., 2019) or adoptive cell transfer (Klinke and Wang, 2016).

While leveraging the knowledge of experts is a great starting point, hand-curated models can also implicitly impose bias on how data is interpreted. In the context of molecular-level networks, rules and algorithms have been developed to elaborate causal networks based on a limited set of rules (Chylek et al., 2014; Sekar and Faeder, 2012; Boutillier et al., 2018; Vernuccio and Broadbelt, 2019). The rules constrain the types of interactions, or edges, that are realistic between the nodes while the algorithms generate all possible edges that are consistent with the rules and collection of nodes. The resulting rule-based networks are then used to interpret data by filter the edges for the most consistent and, in the process, may reveal previously unappreciated pathways. For instance, a rule-based model was used to interpret single-molecule detection of multisite phosphorylation on intact EGFR to reveal new a role for the abundance of adaptor proteins to redirect signaling (Salazar-Cavazos et al., 2020). Given the challenges with representing the various activation states of a 12-subunit Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) holoenzyme that is essential for memory function, a rule-based model identified a molecular mechanism stabilizing protein activity that was obscured in prior reduced models (Pharris et al., 2019). Inspired by engineering better CAR T cells, Rohrs et al. developed a rule-based model to interpret site-specific phosphorylation dynamics associated with Chimeric Antigen Receptors (Rohrs et al., 2018).

To our knowledge, no equivalent approaches exist in the context of modeling cell-level networks.[1] We posit that coupling digital cytometry with Bayesian network inference is analogous to rule-based modeling in the context of modeling cell-level networks. Here, the rules comprise a limited set of constraints, or heuristics, related to the direction of information flow. Specifically, the rules limit how changes in gene expression within the malignant cell introduced during oncogenesis propagate to stromal and immune cells present within the tumor microenvironment and are implemented as a "black list". The algorithms that underpin Bayesian network inference search over all possible network topologies for edges that are consistent with the data. The resulting networks can be used in multiple ways. As an unsupervised approach, the network topology could complement existing workflows for creating mechanistic mathematical models fit for use in testing molecular targets (Gadkar et al., 2016; Ramanujan et al., 2019). In addition, DAGs represent explicit hypotheses generated from pre-existing human data that motivate new experiments to validate the predictions, as illustrated by the B16F0 and YUMM1.7 results.

While the focus here is in the context of breast cancer and melanoma due the pre-existing breadth of data, the approach could be generally applied to other biological contexts and motivate new experimental studies. For instance, one of the limitations of inferring the network topology in the form of directed acyclic graphs is that some direct and indirect causal relationships can be confounded, such as reciprocal feedback modes of communication between cells (Zhou et al., 2018). Discerning the difference between a direct and indirect causal relationship has practical importance, such as for selecting therapeutic targets (Pearl, 2005). Methods, like

[1] One might consider agent-based or cellular automata models to apply as the cellular interactions are specified by rules. In rule-based modeling of molecular networks, the rules and algorithms elaborate a network space that encompasses all possible topologies of the network and data is used to prune the network to the most relevant. Similarly, the edges included in the "blacklist" and "whitelists" can be considered as a Bayesian prior, where the strength of inclusion in the final DAG and the coefficient associated with a particular edge in the conditional probability function depend on the data. In contrast, agent-based or cellular automata models require specifying all interactions between cells as rules a priori and are validated qualitatively by comparing emergent behavior against experimental observations (Hwang et al., 2009; López et al., 2017; Mallet and De Pillis, 2006).

Granger causality and dynamic Bayesian networks (Finkle et al., 2018; Li et al., 2014; Zou and Conzen, 2005), do exist that could reveal direct and indirect causal relationships, but time-series data is required. Unfortunately, human tissue samples, like those in the TGCA, are very rarely sampled with time. Analysis of pre-existing human datasets can be complemented by a more focused experimental study of a pre-clinical model. Specifically, single-cell RNAseq to identify the cell types present and their associated gene signatures can be combined with bulk transcriptomic seqencing to capture the prevalence of all of the cell types within the tissue sample and provide a large number of biological replicates spanning the disease space - normal homeostasis; initiation; early, middle and late progression; and productive resolution or adverse outcomes. Similar network topologies would suggest similar biological mechanisms and help select relevant pre-clinical models for drug development. In short, we feel that combining digital cytometry with Bayesian network inference has the potential to become an indispensable unsupervised approach for discovering relevant heterocellular networks associated with disease.

Methods

Digital Cytometry. Transcriptomics profiling of bulk tissue samples using Illumina RNA sequencing for the breast cancer (BRCA) and cutaneous melanoma (SKCM) arms of the Cancer Genome Atlas was downloaded from TCGA data commons, where values for gene expression were expressed in counts using the "TCGAbiolinks" (V2.8.2) package in R (V3.6.1) and converted to TPM. RNA-seq data expressed in counts assayed in samples acquired from benign melanocytic nevi and untreated primary melanoma tissue and associated sample annotation were downloaded from GEO entry GSE98394 and converted to TPM. TCGA data and the benign nevi and melanoma data were filtered to remove sample outliers and normalized based on housekeeping gene expression (Eisenberg and Levanon, 2013). Digital cytometry features associated with the functional plasticity of tumor cells within an epithelial to mesenchymal-like state space were calculated based on state metrics developed separately for bulk breast cancer and melanoma tissue samples (Klinke and Torang, 2020). Cell proliferation features were calculated based on the median expression of genes associated with cell proliferation identified previously using human cell line data (Deng et al., 2020). Features corresponding to the prevalence of endothelial cells, cancer-associated fibroblasts, macrophages, and $CD4^+$ T cells were calculated using CIBERSORTx (https://cibersortx.stanford.edu) using the gene signatures derived from single cell RNAseq data (Tirosh et al., 2016) while the prevalence of B cells naïve, $CD8^+$ T cells, Macrophage M0 ($M\Phi 0$), Macrophage M1 ($M\Phi 1$), Macrophage M2 ($M\Phi 2$), activated NK cells, resting NK cells, and neutrophils were calculated using the LM22 immune cell gene signatures in CIBERSORTx run in absolute mode.

Given the potential lack of independence among the macrophage features, the LM22 macrophage features were combined to estimate the probability of the average functional orientation using the formula described previously(Kaiser et al., 2016):

$$p(M\Phi i) = \frac{M\Phi i}{M\Phi 0 + M\Phi 1 + M\Phi 2}, \qquad \textbf{(1)}$$

where $i = \{0, 1, 2\}$ and denotes the specific macrophage subtype. Additional cellular features were excluded from the analysis as they tended to have a large number of zero values across the datasets or were disconnected from the rest of the network in preliminary network inference studies. Sample attributes were transformed to numerical values, which were assumed to be extremes of a continuous variable (e.g., Normal = 0, Cancer = 1). The sample attributes, CCN4 gene expression, and estimated cellular features extracted from the bulk RNAseq data calculated for each sample are included in the GitHub repository.

Bayesian Network Inference. Prior to network inference, feature values were log transformed, normalized to values between 0 and 1, and discretized (BRCA: 15 intervals; GEO and SKCM: 6 intervals), as summarized in supplemental Figures S1-S3. The features were then assigned to nodes. The relationships among the nodes, or edges, were represented by directed acyclic graphs inferred from the datasets using a two-stage process, as detailed in the results section. Given the inferred structure, a Bayesian network in the form of a linear Gaussian model was fit to the datasets using maximum likelihood estimation of the model parameters. Conditional probability queries of the Bayesian networks were performed by logic sampling with 10^5 samples. Bayesian network inference was performed using the 'bnlearn' package (V4.5) in R (V3.6.1).

Reagents and Cell Culture. Cytokines and antibodies were obtained from commercial sources and used according to the suppliers' recommendations unless otherwise indicated. The mouse melanoma line B16F0 (purchased in 2008, RRID: CVCL_0604) was obtained from American Tissue Culture Collection (ATCC, Manassas, VA). The mouse melanoma line YUMM1.7 (received in September 2017, RRID: CVCL_JK16) was a gift from Drs. William E. Damsky and Marcus W. Bosenberg (Yale University) (Meeth et al., 2016). B16F0 and YUMM1.7 cells were cultured at 37^oC in 5% CO_2 in high-glucose DMEM (Cellgro/Corning) supplemented with L-glutamine (Lonza), penicillin-streptomycin (Gibco), and 10% heat-inactivated fetal bovine serum (Hyclone). All cell lines were revived from frozen stock, used within 10-15 passages that did not exceed a period of 6 months, and routinely tested for mycoplasma contamination by PCR. CCN4 knock-out variants of B16F0 and YUMM1.7 cells were generated using a double-nickase CRISPR/Cas9 editing strategy described previously (Deng et al., 2019). Briefly, two pairs of mouse CCN4 double nickase plasmids that target the mouse CCN4 gene at different locations were purchased from Santa Cruz Biotechnology, Inc. (Dallas, TX) and transfected into B16F0

and YUMM1.7 cells following the manufacturer's instructions. Following antibiotic selection, surviving single clones were isolated and expanded on 6-well plates. The concentration of CCN4 in the cell culture media from those wells was assayed using the Human WISP-1/CCN4 DuoSet ELISA Kit (R&D Systems, Minneapolis, MN) to confirm CCN4 knockout. CCN4-knockout cells were further expanded and aliquoted to create a low passage frozen stock.

In vivo Tumor Assays and in vitro T cell proliferation assays. All animal experiments were approved by West Virginia University (WVU) Institutional Animal Care and Use Committee and performed on-site. C57BL/6Ncrl mice (6-8 week-old female) were from Charles River Laboratories. Mice were randomly assigned to treatment groups and co-housed following tumor initiation. Subcutaneous tumors were initiated by injecting mice subcutaneously with 3×10^5 of the indicated YUMM1.7 cells and 2.2×10^5 of the indicated B16F0 cells in 100 μL and, once palpable, tumor sizes were recorded every other day via caliper. Tumor volume was calculated using the formula: 0.5236 x width^2 x length, where the width is the smaller dimension of the tumor. Once WT tumors reached between 1000 and 1500 mm^3 in size, the tumors were surgically removed from mice in both arms of the study (WT and CCN4 KO) after euthanasia and processed into single cell suspensions. This normally occurred at Day 14 with the B16F0 model and at Day 27 with the YUMM1.7 model. Three tumors were processed separately for each YUMM1.7 variant while four tumors were processed for each B16F0 variant. Single-cell suspensions were obtained by enzymatically digesting the excised tumors using the Tumor Dissociation Kit and gentleMACS C system (Miltenyi Biotec, Auburn, CA). In addition to following the manufacturer's instructions, the gentleMACS program 37C_m_TDK_1 was used for B16F0 tumors and 37C_m_TDK_2 was used for YUMM1.7 tumors. Following lysing of the red blood cells, the remaining single-cell suspensions were washed and stained with Live/Dead Fixable Pacific Blue Dead Cell Stain Kit (ThermoFisher). Following blocking with Mouse BD Fc Block (BD Biosciences), the surface of the cells were stained with one of three different antibody mixes that focused on T cells (CD45, CD3, CD4, CD8, and PD1), NK and B cells (CD45, B220, NK11, DX5, and PD1), and myeloid cells (CD45, CD11b, CD11c, Gr-1, F4/80, and MHCII) and quantified by flow cytometry. The specific antibodies used are listed in Supplemental Table S1.

To assess the impact of CCN4 on T cell proliferation in vitro, splenocytes were obtained from naïve C57BL/6 mice and stained with CellTrace Pacific Blue Cell Proliferation Kit (ThermoFisher). Stained splenocytes (2.5×10^5) were stimulated for 3 days in 96 well plate with MACSiBeads loaded with anti-mouse CD3 and anti-mouse CD28 antibodies (AP beads, Miltenyi Biotec), at a 1:1 proportion. Fresh serum-free DMEM media conditioned for 24 hours by either confluent wild-type (WT TCM) or confluent CCN4 KO (CCN4 KO TCM) melanoma B16F0 cells were collected, centrifuged to remove cells and cell debris, and added at 50% final volume during T cell stimulation with AP beads. In addition, splenocytes were either left unstimulated or stimulated with AP beads alone, or stimulated in the presence of recombinant mouse CCN4 (rCCN4, R&D) at a final concentration of 10 ng/mL. After 72h, cells were washed and stained with Live/Dead Fixable Green Dead Cell Stain Kit (ThermoFisher). Surface staining with anti-mouse CD8/APC (Miltenyi Biotec), anti-mouse CD4/APC-Cy7 (BD Biosciences), anti-mouse CD62L/PE (eBioscience, ThermoFisher) and anti-mouse CD44/PerCPCy5.5 (eBioscience, ThermoFisher) was performed after incubating the cells with Mouse BD Fc Block (BD Biosciences). The proliferation of both CD4 and CD8 T cells were quantified by flow cytometry.

In vitro suppression of CD8$^+$ T cell function. Inducible mouse CCN4 expression lentiviral vector (IDmCCN4) was constructed with Gateway cloning using Tet-on destination lentiviral vector pCW57.1 (Addgene Plasmid #41393, a gift from David Root) and pShuttle Gateway PLUS ORF Clone for mouse CCN4 (GC-Mm21303, GeneCopoeia). Lentiviruses were packaged as described (Deng et al., 2019) to transduce YUMM1.7 cell with *Ccn4* CRISPR knockout (Ym1.7-KO1) (Deng et al., 2019). After puromycin selection, two pools of cells with inducible mCCN4 (Ym1.7-KO1-IDmCCN4) or vector control (Ym1.7-KO1-IDvector) were obtained. ELISA tests with doxycycline (Dox, final 0.5μg/ml) induction revealed the mCCN4 expression was under stringent control and the secreted protein was in the similar level as compared with wild-type YUMM1.7 cells (data not shown).

To generate YUMM1.7-reactive CD8+ T cells, healthy C57BL/6Ncrl mice were inoculated subcutaneously with irradiated YUMM1.7 cells (10^5/mouse), followed by live YUMM1.7 cells (3×10^5/mouse) 3 weeks later. The mice without tumor growth in the next five weeks were maintained. Three days before the assay, the mice were injected again with live YUMM1.7 cells (10^5/mouse). On the day of assay, these mice were euthanized and the YUMM1.7-reactive cells were isolated from mouse splenocytes using mouse CD8a+ T Cell Isolation Kit (130-104-075, Miltenyi Biotec), resuspended in a concentration of 10^6/ml. 50μl (5×10^4) of the YUMM1.7-reactive CD8+ T cells were aliquoted into each well on a 96-well plate for ELISpot assay using Mouse IFNγ/TNFα Double-Color ELISpot kit (Cellular Technology Limited, CTL) following manufacturer's instructions. Briefly, target tumor cells were stimulated with IFNγ (200U/ml, or, 20ng/ml) for 24 hours, harvested and resuspended in a concentration of 2×10^6/ml. 50μl (10^5) of indicated tumor cells in triplicates were aliquoted into each well, with or without doxycycline (Dox, final 0.5μg/ml). The reactions were incubated at 37°C for 24 hours and colored spots were developed (Red for IFNγ and blue for TNFα). The spots were counted and imaged using an Olympus MVX10 Microscope and the result was plotted and analyzed by GraphPad Prism (version 5).

Flow Cytometry. Single cell suspensions described above were stained with specific antibodies or isotype controls

using conventional protocols. Fluorescence-activated cell counting was performed using a BD LSRFortessa and FACS-Diva software (BD Biosciences) as where the fluorescence intensity for each parameter was reported as a pulse area with 18-bit resolution. Unstained samples were used as negative flow cytometry controls. Single-stain controls were used to establish fluorescence compensation parameters. For TIL analysis, greater than 5×10^5 events were acquired in each antibody panel in each biological replicate. In analyzing enriched cell populations, 2×10^4 events were acquired in each biological replicate. Flow cytometric data were exported as FCS3.0 files and analyzed with using R/Bioconductor (V3.5.1), as described previously (Klinke and Brundage, 2009). The typical gating strategies for T cells, NK and B cells, and myeloid cells are shown in supplementary Figures S4-S6, respectively. The statistical difference in tumor infiltrating lymphocytes between wt and CCN4 KO variants was assessed using log-transformed values and a two-tailed homoscedastic Student's t test. Cell proliferation was quantified using metrics: fraction diluted (Dil), Precursor frequency, %dividing cells (PF), Proliferation index (PI), and proliferation variance (SD^D) (Roederer, 2011). Statistical differences among these proliferation parameters were assessed using type III repeated measures ANOVA in the "car" (V3.0-7) package in R. A p-value < 0.05 was considered statistically significant.

Data and Code Availability. The code used in the analysis can be obtained from the following GitHub repository:

- https://github.com/KlinkeLab/CellNetwork_2020

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation (NSF CBET-1644932 to DJK) and National Cancer Institute (NCI 1R01CA193473 to DJK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or NCI. We also used equipment from the WVU Flow Cytometry & Single Cell core, which was supported by the National Institutes of Health Grants GM103488/RR032138, GM104942, GM103434, and OD016165.

AUTHOR CONTRIBUTIONS

These contributions follow the International Committee of Medical Journal Editors guidelines: http://www.icmje.org/recommendations/. Conceptualization: DJK; Study Design: DJK; Data Acquisition: DJK, AF, and WD; Data Analysis: DJK and WD; Data Interpretation: DJK; Funding acquisition: DJK; Methodology: DJK; Project administration: DJK; Software: DJK and ACP; Supervision: DJK; Writing - original draft: DJK; Writing - review & editing: all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Bibliography

Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautes-Fridman, C., Fridman, W.H., de Reynies, A., 2016. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. Genome Biol. 17, 218.

Boutillier, P., Maasha, M., Li, X., Medina-Abarca, H.F., Krivine, J., Feret, J., Cristescu, I., Forbes, A.G., Fontana, W., 2018. The Kappa platform for rule-based modeling. Bioinformatics 34, i583–i592.

Chylek, L.A., Harris, L.A., Tung, C.S., Faeder, J.R., Lopez, C.F., Hlavacek, W.S., 2014. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. Wiley Interdiscip Rev Syst Biol Med 6, 13–36.

Deng, W., Fernandez, A., McLaughlin, S.L., Klinke, D.J., 2019. WNT1-inducible signaling pathway protein 1 (WISP1/CCN4) stimulates melanoma invasion and metastasis by promoting the epithelial-mesenchymal transition. J. Biol. Chem. 294, 5261–5280.

Deng, W., Fernandez, A., McLaughlin, S.L., Klinke, D.J., 2020. Cell Communication Network Factor 4 (CCN4/WISP1) Shifts Melanoma Cells from a Fragile Proliferative State to a Resilient Metastatic State. Cell Mol Bioeng 13, 45–60.

Eisenberg, E., Levanon, E.Y., 2013. Human housekeeping genes, revisited. Trends Genet. 29, 569–574.

Fernandez, A., Deng, W., McLaughlin, S., Pirkey, A., Rellick, S., Klinke, D.J., 2020. Cell communication network factor 4 (ccn4/wisp1) promotes tumor-induced immunosuppression in melanoma. bioRxiv , (submitted).

Finkle, J.D., Wu, J.J., Bagheri, N., 2018. Windowed Granger causal inference strategy improves discovery of gene regulatory networks. Proc. Natl. Acad. Sci. U.S.A. 115, 2252–2257.

Friedman, N., 2004. Inferring cellular networks using probabilistic graphical models. Science 303, 799–805.

Gadkar, K., Kirouac, D.C., Mager, D.E., van der Graaf, P.H., Ramanujan, S., 2016. A Six-Stage Workflow for Robust Application of Systems Pharmacology. CPT Pharmacometrics Syst Pharmacol 5, 235–249.

George, J.T., Jolly, M.K., Xu, S., Somarelli, J.A., Levine, H., 2017. Survival Outcomes in Cancer Patients Predicted by a Partial EMT Gene Expression Scoring Metric. Cancer Res. 77, 6415–6428.

Grun, D., Kester, L., van Oudenaarden, A., 2014. Validation of noise models for single-cell transcriptomics. Nat. Methods 11, 637–640.

Hill, S.M., Heiser, L.M., Cokelaer, T., Unger, M., Nesser, N.K., Carlin, D.E., Zhang, Y., Sokolov, A., Paull, E.O., Wong, C.K., Graim, K., Bivol, A., Wang, H., Zhu, F., Afsari, B., Danilova, L.V., Favorov, A.V., Lee, W.S., Taylor, D., Hu, C.W., Long, B.L., Noren, D.P., Bisberg, A.J., Mills, G.B., Gray, J.W., Kellen, M., Norman, T., Friend, S., Qutub, A.A., Fertig, E.J., Guan, Y., Song, M., Stuart, J.M., Spellman, P.T., Koeppl, H., Stolovitzky, G., Saez-Rodriguez, J., Mukherjee, S., Afsari, B., Al-Ouran, R., Anton, B., Arodz, T., Sichani, O.A., Bagheri, N., Berlow, N., Bisberg, A.J., Bivol, A., Bohler, A., Bonet, J., Bonneau, R., Budak, G., Bunescu, R., Caglar, M., Cai, B., Cai, C., Carlin, D.E., Carlon, A., Chen, L., Ciaccio, M.F., Cokelaer, T., Cooper, G., Coort, S., Creighton, C.J., Daneshmand, S.M., de la Fuente, A., Di Camillo, B., Danilova, L.V., Dutta-Moscato, J., Emmett, K., Evelo, C., Fassia, M.K., Favorov, A.V., Fertig, E.J., Finkle, J.D., Finotello, F., Friend, S., Gao, X., Gao, J., Garcia-Garcia, J., Ghosh, S., Giaretta, A., Graim, K., Gray, J.W., Grosseholz, R., Guan, Y., Guinney, J., Hafemeister, C., Hahn, O., Haider, S., Hase, T., Heiser, L.M., Hill, S.M., Hodgson, J., Hoff, B., Hsu, C.H., Hu, C.W., Hu, Y., Huang, X., Jalili, M., Jiang, X., Kacprowski, T., Kaderali, L., Kang, M., Kannan, V., Kellen, M., Kikuchi, K., Kim, D.C., Kitano, H., Knapp, B., Komatsoulis, G., Koeppl, H., Kramer, A., Kursa, M.B., Kutmon, M., Lee, W.S., Li, Y., Liang, X., Liu, Z., Liu, Y., Long, B.L., Lu, S., Lu, X., Manfrini, M., Matos, M.R., Meerzaman, D., Mills, G.B., Min, W., Mukherjee, S., Muller, C.L., Neapolitan, R.E., Nesser, N.K., Noren, D.P., Norman, T., Oliva, B., Opiyo, S.O., Pal, R., Palinkas, A., Paull, E.O., Planas-Iglesias, J., Poglayen, D., Qutub, A.A., Saez-Rodriguez, J., Sambo, F., Sanavia, T., Sharifi-Zarchi, A., Slawek, J., Sokolov, A., Song, M., Spellman, P.T., Streck, A., Stolovitzky, G., Strunz, S., Stuart, J.M., Taylor, D., Tegner, J., Thobe, K., Toffolo, G.M., Trifoglio, E., Unger, M., Wan, Q., Wang, H., Welch, L., Wong, C.K., Wu, J.J., Xue, A.Y., Yamanaka, R., Yan, C., Zairis, S., Zengerling, M., Zenil, H., Zhang, S., Zhang, Y., Zhu, F., Zi, Z., 2016. Inferring causal molecular networks: empirical assessment through a community-based effort. Nat. Methods 13, 310–318.

Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C.K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A.G., Schneider, B.G., Lawrence, M.S., Noushmehr, H., Malta, T.M., Cancer Genome Atlas Network, Stuart, J.M., Benz, C.C., Laird, P.W., 2018. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 173, 291–304.

Huang, A.C., Postow, M.A., Orlowski, R.J., Mick, R., Bengsch, B., Manne, S., Xu, W., Harmon, S., Giles, J.R., Wenz, B., Adamow, M., Kuk, D., Panageas, K.S., Carrera, C., Wong, P., Quagliarello, F., Wubbenhorst, B., D'Andrea, K., Pauken, K.E., Herati, R.S., Staupe, R.P., Schenkel, J.M., McGettigan, S., Kothari, S., George, S.M., Vonderheide, R.H., Amaravadi, R.K., Karakousis, G.C., Schuchter, L.M., Xu, X., Nathanson, K.L., Wolchok, J.D., Gangadhar, T.C., Wherry, E.J., 2017. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. Nature 545, 60–65.

Hwang, M., Garbey, M., Berceli, S.A., Tran-Son-Tay, R., 2009. Rule-Based Simulation of Multi-Cellular Biological Systems-A Review of Modeling Techniques. Cell Mol Bioeng 2, 285–294.

Jeong, H., Hwang, I., Kang, S.H., Shin, H.C., Kwon, S.Y., 2019. Tumor-Associated Macrophages as Potential Prognostic Biomarkers of Invasive Breast Cancer. J Breast Cancer 22, 38–51.

Kaiser, J.L., Bland, C.L., Klinke, D.J., 2016. Identifying causal networks linking cancer processes and anti-tumor immunity using Bayesian network inference and metagene constructs. Biotechnol. Prog. 32, 470–479.

Klinke, D.J., 2014. Induction of Wnt-inducible signaling protein-1 correlates with invasive breast cancer oncogenesis and reduced type 1 cell-mediated cytotoxic immunity: a retrospective study. PLoS Comput. Biol. 10, e1003409.

Klinke, D.J., 2016. Eavesdropping on altered cell-to-cell signaling in cancer by secretome profiling. Mol Cell Oncol 3, e1029061.

Klinke, D.J., Brundage, K.M., 2009. Scalable analysis of flow cytometry data using R/Bioconductor. Cytometry A 75, 699–706.

Klinke, D.J., Torang, A., 2020. An unsupervised feature extraction and selection strategy for identifying epithelial-mesenchymal transition state metrics in breast cancer and melanoma. iScience , (in press).

Klinke, D.J., Wang, Q., 2016. Inferring the Impact of Regulatory Mechanisms that Underpin CD8+ T Cell Control of B16 Tumor Growth In vivo Using Mechanistic Models and Simulation. Front Pharmacol 7, 515.

Koplev, S., Lin, K., Dohlman, A.B., Ma'ayan, A., 2018. Integration of pan-cancer transcriptomics with RPPA proteomics reveals mechanisms of epithelial-mesenchymal transition. PLoS Comput. Biol. 14, e1005911.

Li, H., van der Leun, A.M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., van Akkooi, A.C.J., van den Braber, M., Rozeman, E.A., Haanen, J.B.A.G., Blank, C.U., Horlings, H.M., David, E., Baran, Y., Bercovich, A., Lifshitz, A., Schumacher, T.N., Tanay, A., Amit, I., 2019. Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. Cell 176, 775–789.

Li, P., Gong, P., Li, H., Perkins, E.J., Wang, N., Zhang, C., 2014. Gene regulatory network inference and validation using relative change ratio analysis and time-delayed dynamic Bayesian network. EURASIP J Bioinform Syst Biol 2014, 12.

López, A.G., Seoane, J.M., Sanjuán, M.A.F., 2017. Dynamics of the cell-mediated immune response to tumour growth. Philos Trans A Math Phys Eng Sci 375.

Mallet, D.G., De Pillis, L.G., 2006. A cellular automata model of tumor-immune system interactions. J. Theor. Biol. 239, 334–350.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kami/'nska, B., Huelsken, J., Omberg, L., Gevaert, O., Colaprico, A., Czerwi/'nska, P., Mazurek, S., Mishra, L., Heyn, H., Krasnitz, A., Godwin, A.K., Lazar, A.J., Cancer Genome Atlas Research Network, Stuart, J. M. Hoadley, K.A., Laird, P.W., Noushmehr, H., Wiznerowicz, M., 2018. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell 173, 338–354.

Meeth, K., Wang, J.X., Micevic, G., Damsky, W., Bosenberg, M.W., 2016. The YUMM lines: a series of congenic mouse melanoma cell lines with defined genetic alterations. Pigment Cell Melanoma Res 29, 590–597.

Milberg, O., Gong, C., Jafarnejad, M., Bartelink, I.H., Wang, B., Vicini, P., Narwal, R., Roskos, L., Popel, A.S., 2019. A QSP Model for Predicting Clinical Responses to Monotherapy, Combination and Sequential Therapy Following CTLA-4, PD-1, and PD-L1 Checkpoint Blockade. Sci Rep 9, 11286.

Moore, H., Allen, R., 2019. What Can Mathematics Do for Drug Development? Bull. Math. Biol. 81, 3421–3424.

Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., Diehn, M., Alizadeh, A.A., 2019. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. 37, 773–782.

Palsson, S., Hickling, T.P., Bradshaw-Pierce, E.L., Zager, M., Jooss, K., O'Brien, P.J., Spilker, M.E., Palsson, B.O., Vicini, P., 2013. The development of a fully-integrated immune response model (FIRM) simulator of the immune response through integration of multiple subset models. BMC Syst Biol 7, 95.

Papalexi, E., Satija, R., 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat. Rev. Immunol. 18, 35–45.

Pearl, J., 2005. Direct and indirect effects, in: Proc Am Stat Assoc Joint Stat Meetings (Minneapolis, MN), MIRA Digital Publishing. pp. 1572–1581.

Perrin, B.E., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., dlchBuc, F., 2003. Gene networks inference using dynamic bayesian networks. Bioinformatics 19, ii138–ii148.

Pharris, M.C., Patel, N.M., VanDyk, T.G., Bartol, T.M., Sejnowski, T.J., Kennedy, M.B., Stefan, M.I., Kinzer-Ursem, T.L., 2019. A multi-state model of the CaMKII dodecamer suggests a role for calmodulin in maintenance of autophosphorylation. PLoS Comput. Biol. 15, e1006941.

Ramanujan, S., Chan, J.R., Friedrich, C.M., Thalhauser, C.J., 2019. A Flexible Approach for Context-Dependent Assessment of Quantitative Systems Pharmacology Models. CPT Pharmacometrics Syst Pharmacol 8, 340–343.

Roederer, M., 2011. Interpretation of cellular proliferation data: avoid the panglossian. Cytometry A 79, 95–101.

Rohrs, J.A., Zheng, D., Graham, N.A., Wang, P., Finley, S.D., 2018. Computational Model of Chimeric Antigen Receptors Explains Site-Specific Phosphorylation Kinetics. Biophys. J. 115, 1116–1129.

Rose, S., Misharin, A., Perlman, H., 2012. A novel Ly6C/Ly6G-based strategy to analyze the mouse splenic myeloid compartment. Cytometry A 81, 343–350.

Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., Lauffenburger, D.A., 2002. Bayesian network approach to cell signaling pathway modeling. Science Signaling 2002, pe38.

Sachs, K., Itani, S., Carlisle, J., Nolan, G.P., Pe'er, D., Lauffenburger, D.A., 2009. Learning signaling network structures with sparsely distributed data. J Comp Biol 16, 201–212.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., Nolan, G.P., 2005. Causal protein-signaling networks derived from multiparameter single-cell data. Science 308, 523–529.

Salazar-Cavazos, E., Nitta, C.F., Mitra, E.D., Wilson, B.S., Lidke, K.A., Hlavacek, W.S., Lidke, D.S., 2020. Multisite EGFR phosphorylation is regulated by adaptor protein abundances and dimer lifetimes. Mol. Biol. Cell 31, 695–708.

Schelker, M., Feau, S., Du, J., Ranu, N., Klipp, E., MacBeath, G., Schoeberl, B., Raue, A., 2017. Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. Nat Commun 8, 2032.

Scutari, M., 2010. Learning Bayesian Networks with the bnlearn R Package. J Stat Software 35, 1–22.

Sekar, J.A., Faeder, J.R., 2012. Rule-based modeling of signal transduction: a primer. Methods Mol. Biol. 880, 139–218.

Shen-Orr, S., Tibshirani, R., Butte, A., 2012. Gene expression deconvolution in linear space. Nature Methods 9, 9.

Shen-Orr, S.S., Tibshirani, R., Khatri, P., Bodian, D.L., Staedtler, F., Perry, N.M., Hastie, T., Sarwal, M.M., Davis, M.M., Butte, A.J., 2010. Cell type-specific gene expression differences in complex tissues. Nat. Methods 7, 287–289.

Singer, M., Anderson, A.C., 2019. Revolutionizing Cancer Immunology: The Power of Next-Generation Sequencing Technologies. Cancer Immunol Res 7, 168–173.

Stuart, T., Satija, R., 2019. Integrative single-cell analysis. Nat. Rev. Genet. 20, 257–272.

Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y., Thiery, J.P., 2014. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO Mol Med 6, 1279–1293.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., Ziv, E., Culhane, A.C., Paull, E.O., Sivakumar, I.K.A., Gentles, A.J., Malhotra, R., Farshidfar, F., Colaprico, A., Parker, J.S., Mose, L.E., Vo, N.S., Liu, J., Liu, Y., Rader, J., Dhankani, V., Reynolds, S.M., Bowlby, R., Califano, A., Cherniack, A.D., Anastassiou, D., Bedognetti, D., Rao, A., Chen, K., Krasnitz, A., Hu, H., Malta, T.M., Noushmehr, H., Pedamallu, C.S., Bullman, S., Ojesina, A.I., Lamb, A., Zhou, W., Shen, H., Choueiri, T.K., Weinstein, J.N., Guinney, J., Saltz, J., et al., 2018. The Immune Landscape of Cancer. Immunity 48, 812–830.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.R., Cohen, O., Shah, P., Lu, D., Genshaft, A.S., Hughes, T.K., Ziegler, C.G., Kazer, S.W., Gaillard, A., Kolb, K.E., Villani, A.C., Johannessen, C.M., Andreev, A.Y., Van Allen, E.M., Bertagnolli, M., Sorger, P.K., Sullivan, R.J., Flaherty, K.T., Frederick, D.T., Jane-Valbuena, J., Yoon, C.H., Rozenblatt-Rosen, O., Shalek, A.K., Regev, A., Garraway, L.A., 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196.

Torang, A., Gupta, P., Klinke, D.J., 2019. An elastic-net logistic regression approach to generate classifiers and gene signatures for types of immune cells and T helper cell subsets. BMC Bioinformatics 20, 433.

Tosolini, M., Kirilovsky, A., Mlecnik, B., Fredriksen, T., Mauger, S., Bindea, G., Berger, A., Bruneval, P., Fridman, W., Pagos, F., Galon, J., 2011. Clinical impact of different classes of infiltrating t cytotoxic and helper cells (th1, th2, treg, th17) in patients with colorectal cancer. Cancer Research 71, 1263–1271.

Twyman-Saint Victor, C., Rech, A.J., Maity, A., Rengan, R., Pauken, K.E., Stelekati, E., Benci, J.L., Xu, B., Dada, H., Odorizzi, P.M., Herati, R.S., Mansfield, K.D., Patsch, D., Amaravadi, R.K., Schuchter, L.M., Ishwaran, H., Mick, R., Pryma, D.A., Xu, X., Feldman, M.D., Gangadhar, T.C., Hahn, S.M., Wherry, E.J., Vonderheide, R.H., Minn, A.J., 2015. Radiation and dual checkpoint blockade activate non-redundant immune mechanisms in cancer. Nature 520, 373–377.

Vernuccio, S., Broadbelt, L.J., 2019. Discerning complex reaction networks using automated generators. AIChE Journal 65, e16663.

Wang, Z., Cao, S., Morris, J.S., Ahn, J., Liu, R., Tyekucheva, S., Gao, F., Li, B., Lu, W., Tang, X., Wistuba, I.I., Bowden, M., Mucci, L., Loda, M., Parmigiani, G., Holmes, C.C., Wang, W., 2018. Transcriptome Deconvolution of Heterogeneous Tumor Samples with Immune Infiltration. iScience 9, 451–460.

Wells, A., Wiley, H.S., 2018. A systems perspective of heterocellular signaling. Essays Biochem. 62, 607–617.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P., Levine, D., Carter, S.L., Getz, G., Stemke-hale, K., Mills, G.B., Verhaak, R., 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications 4, 2612.

Zaitsev, K., Bambouskova, M., Swain, A., Artyomov, M., 2019. Complete deconvolution of cellular mixtures based on linearity of transcriptional signatures. Nature communications 10, 2209.

Zhou, X., Franklin, R.A., Adler, M., Jacox, J.B., Bailis, W., Shyer, J.A., Flavell, R.A., Mayo, A., Alon, U., Medzhitov, R., 2018. Circuit Design Features of a Stable Two-Cell System. Cell 172, 744–757.

Zou, M., Conzen, S.D., 2005. A new dynamic bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics 21, 71–79.

Figure Legends

Figure 1 - A computational workflow combines digital cytometry with Bayesian network inference to estimate how a genetic driver impacts the heterocellular network within a tissue. Digital cytometry deconvolutes a bulk transcriptomic profile using gene signatures that correspond to different stromal, malignant, and immune cell types. The results estimate the prevalence of the different cell types within the tissue sample, that is digital cytometry features. By using bulk transcriptomic profiles of defined patient populations, underlying variation in the inferred cellular composition coupled with changes in expression of a putative gene driver can be used to estimate how the heterocellular network is impacted by a gene driver using Bayesian Network inference. To illustrate the approach, we focused on CCN4 as a gene driver. The resulting directed acyclic graphs represent the collective conditional independence among modeled nodes of the network.

Figure 2 - Summary of the evidence obtained from the TCGA breast cancer dataset supporting the consensus edges in the seed network. Edges ordered based on the number of algorithms that detected that an edge was enriched (bar graph - left axis) and the strength of enrichment (dotted lines - right axis). The lines associated with the strength of enrichment represent the minimum (dashed line) and maximum (dotted line) values obtained by the different algorithms. Coloring of bar graph indicates whether a clear direction was associated with an edge (green), an edge was significantly enriched but without a clear direction (yellow), or that an edge was excluded from the consensus seed network list (tan).

Figure 3 - A directed acyclic graph (DAG) representing the conditional probability distribution inferred using the digital cytometry features extracted from the breast cancer arm of the TCGA. The nodes of the graph represent features, such as CCN4 gene expression (rectangle), sample attribute (hexagon), or the prevalence of a particular cell type/state (oval). The edges represent inferred causal relationships among the nodes. The black lines with arrow heads represent a positive causal relation while red lines with horizontal bars represent a negative or inhibitory causal relation, where the extent of influence of the parental node is annotated by the number beside the edge. The number included within the node symbol represents the average normalized value of the digital cytometry feature within the dataset with values of all of the parental nodes set to zero. The width of the edge is proportional to the posterior probability of inclusion into the DAG.

Figure 4 - Conditional probability query of the BRCA DAG compared against digital cytometry estimates obtained from experimental data. Experimental samples obtained from normal mammary and tumor tissue are shown as filled versus open circles, respectively. Samples of the conditional probability model for p(Cancer < 0.15) (orange) and p(Cancer > 0.85) (blue) for CD8 T cells (A), CD4 T cells (B), active NK cells (C), B cells (D), Macrophages (E) and Cancer Associated Fibroblasts (F). Linear trendlines are superimposed on the conditional probability samples.

Figure 5 - Two DAGs representing the conditional probability distributions inferred using the digital cytometry features extracted from the two melanoma-related datasets. (A) Analysis of a bulk RNAseq dataset obtained from patients with common pigmented nevi and primary melanoma ($n_{samples}$ = 78). (B) Analysis of primary melanoma samples extracted from the SKCM arm of the TCGA ($n_{samples}$ = 94). The DAGs are summarized using similar notation as described in Figure 3. Dotted lines indicate edges that were included in the consensus seed network but, as the samples were all from patients with cancer, had no evidence in the TCGA dataset.

Figure 6 - CCN4 knock-out in two syngeneic mouse models of melanoma induces a similar shift in NK cells and T and B lymphocytes as observed in human breast cancer and melanoma. (A) The percentage of live CD45+ cells isolated from tumors generated by inoculating s.c. with wt (red) and CCN4 KO (blue) variants of B16F0 (o and x's) and YUMM1.7 (□ and +'s) cells, where the log-linear trends are highlighted by dotted lines. CD45+ values were obtained from three different antibody panels that quantified T cells, B/NK cells, and myeloid cells in TIL isolates from each mouse (n = 3 for YUMM1.7 and n = 4 for B16F0 variants). (B) A comparison of the ratio of NK cells (black), CD8+ T cells (red), CD4+ T cells (blue), and B cells (green) to live CD45+ TILs in s.c. tumors generated using wt B16F0 and YUMM1.7 cells. (C) Comparing the log ratio in prevalence of the different cell types when CCN4 is present (WT) versus absent (CCN4 KO) predicted by the BRCA (1st column) and SKCM (4th column) DAGs and observed experimentally using the B16F0 (2nd column) and YUMM1.7 (3rd column) mouse models. Median results for NK cells (black), CD8+ T cells (red), CD4+ T cells (blue), and B cells (green) in the different settings are connected by lines. (D) TIL comparison upon CCN4 KO in B16F0 and YUMM1.7 mouse models stratified by NK cells, CD8+ T cells, CD4+ T cells, and B cells (top to bottom). p-values calculated between wt and CCN4 KO pairs using Student's t-test.

Figure 7 - Myeloid immune cell subsets differentially infiltrate tumors derived from wt B16F0 and YUMM1.7 cells but shift in similar ways upon CCN4 knock-out. (A) A comparison of the ratio of CD11c- (black) and CD11c+ (gray) macrophages, CD11c+ Mo-MDSC (green), Mo-MDSC (blue), and PMN-MDSC (red) to live CD45+ TILs in s.c. tumors generated using wt B16F0 and YUMM1.7 cells. (B) Comparing the log ratio in prevalence of the different myeloid cell types when CCN4 is present (WT) versus absent (CCN4 KO) predicted by the BRCA (1st column) and SKCM (4th column) DAGs and observed experimentally using the B16F0 (2nd column) and YUMM1.7 (3rd column) mouse models. Macrophages are the only myeloid cell subset inferred from the BRCA and SKCM datasets and are assumed to be related to CD11c- macrophages in mouse models. Median results in the different settings are connected by lines. (C and D) A representative scatter plot of GR1 versus CD11c expression in gated live CD45+ CD11b+ TILs obtained from wt (C) and CCN4 KO (D) YUMM1.7 tumors. The right panel shows MHCII versus F4/80 expression in the $GR1^{+}$ $CD11c^{-}$ subset. (E) TIL comparison upon CCN4 KO in B16F0 and YUMM1.7 mouse models stratified by myeloid cell subsets (top to bottom). p-values calculated between wt and CCN4 KO pairs using Student's t-test.

Figure 8 - CCN4 has no direct effect on T cell proliferation but impairs CD8$^+$ T cell function. The distribution in cell trace staining among live CD4$^+$ (A) and CD8$^+$ (B) T cells stimulated with αCD3/αCD28 (AP beads) alone or in the presence of media conditioned by wt B16F0 cells (AP beads + WT TCM), media conditioned by CCN4 KO B16F0 cells (AP beads + CCN4 KO TCM), or with 10 ng/ml of recombinant mouse CCN4 (AP beads + rCCN4). The distribution in the corresponding unstimulated cells (gray) are shown at the bottom. The colored vertical lines indicate the predicted dilution of cell trace staining in each generation based on the unstimulated controls. (C) CD8+ T cells isolated from the spleens of C57BL/6 mice that rejected YUMM1.7 tumors were cultured in an in vitro ELISPOT assay using variants of the YUMM1.7 cell line as targets (wt YUMM1.7 - yellow, CCN4 KO YUMM1.7 - light green, CCN4 KO YUMM1.7 with a blank inducible expression vector - dark green and blue, CCN4 KO YUMM1.7 with a CCN4 inducible expression vector - purple and red). Variants containing the inducible expression vector were also cultured in the absence (dark green and purple) or presence of doxycycline (blue and red). CD8+ T cells expressing IFNγ and TNFα were quantified following 24 hour co-culture (bar graph). Statistical significant between pairs was assessed using a Student's t-test, where * = p-value < 0.05 and *** = p-value < 0.001.

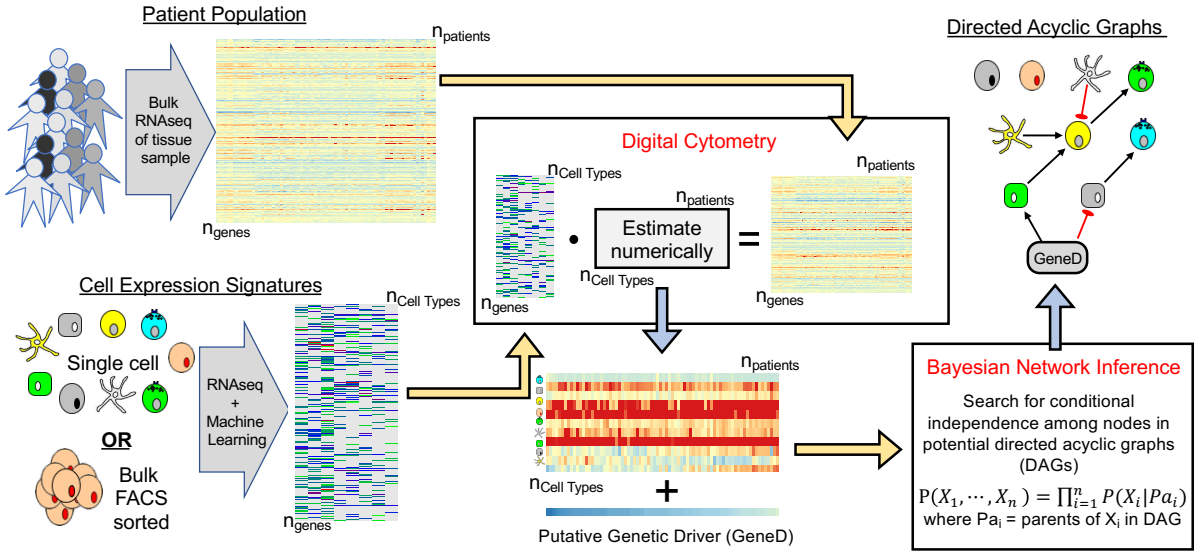

Fig. 1. A computational workflow combines digital cytometry with Bayesian network inference to estimate how a genetic driver impacts the heterocellular network within a tissue.

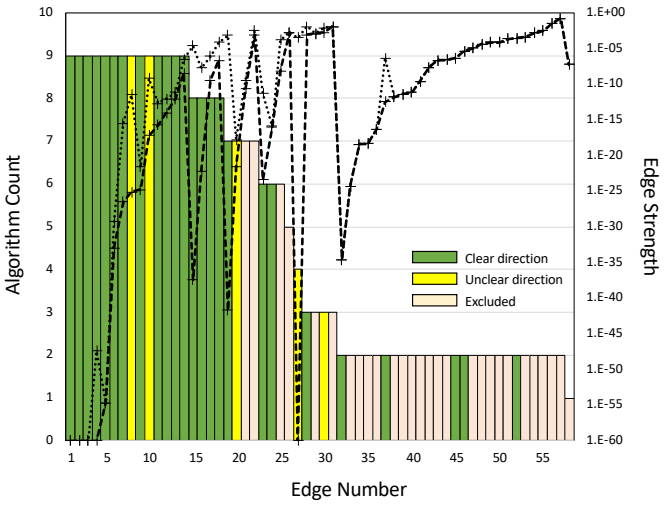

Fig. 2. Summary of the evidence obtained from the TCGA breast cancer dataset supporting the consensus edges in the seed network.

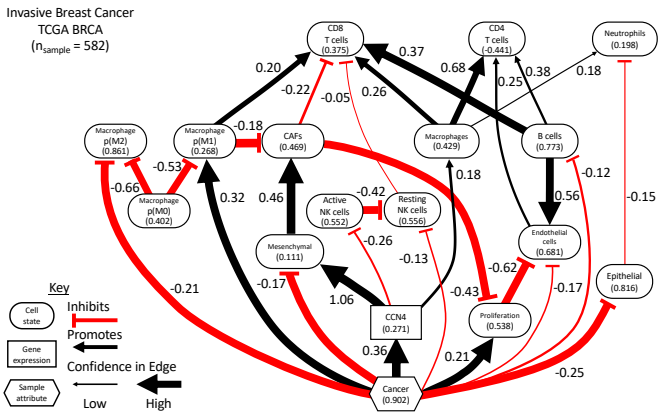

Fig. 3. A directed acyclic graph (DAG) representing the conditional probability distribution inferred using the digital cytometry features extracted from the breast cancer arm of the TCGA.

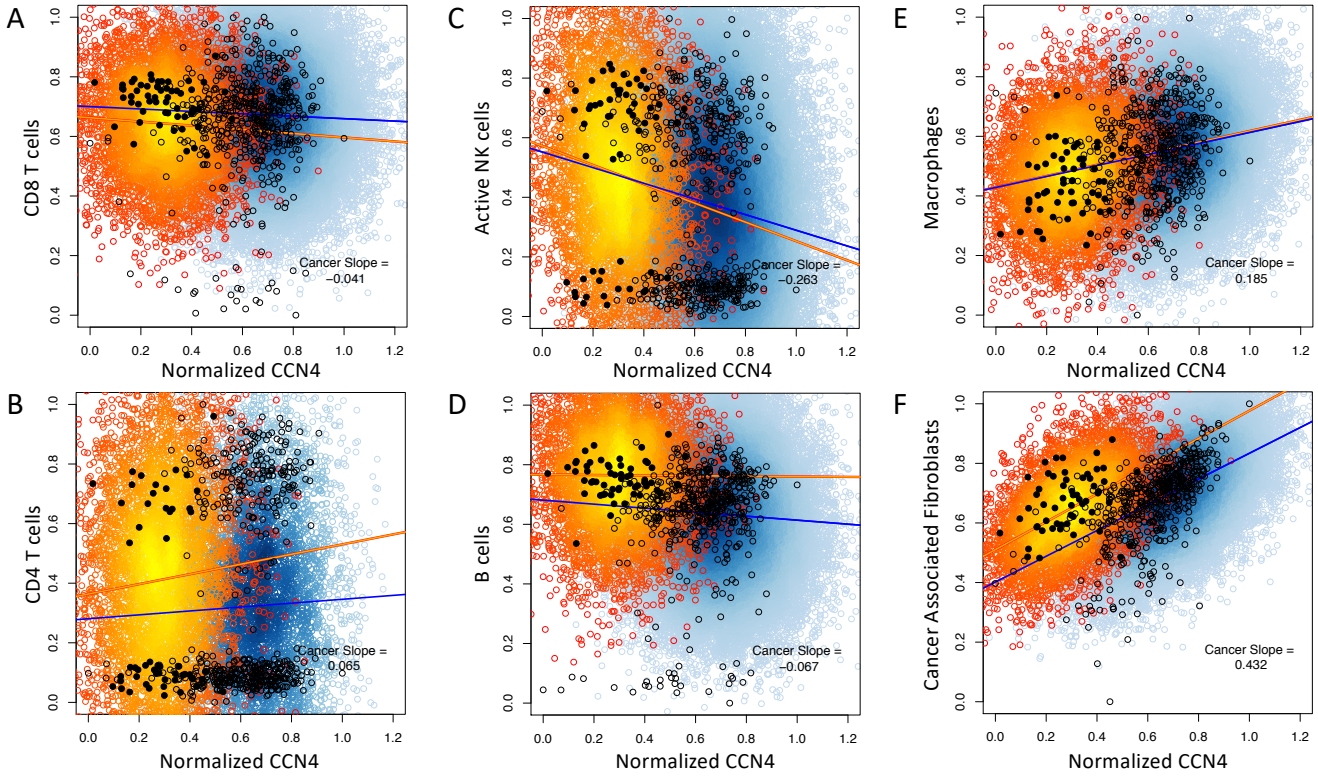

Fig. 4. Conditional probability query of the BRCA DAG compared against digital cytometry estimates obtained from experimental data.

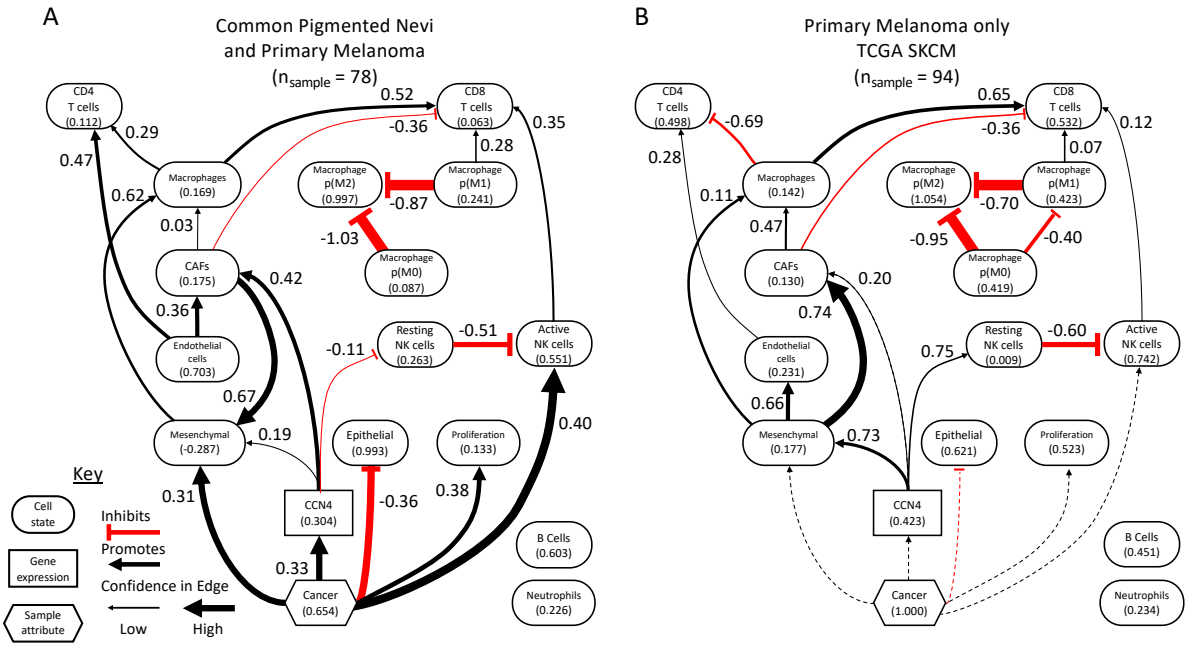

Fig. 5. A DAG representing the conditional probability distribution inferred using the digital cytometry features extracted from the two melanoma-related datasets.

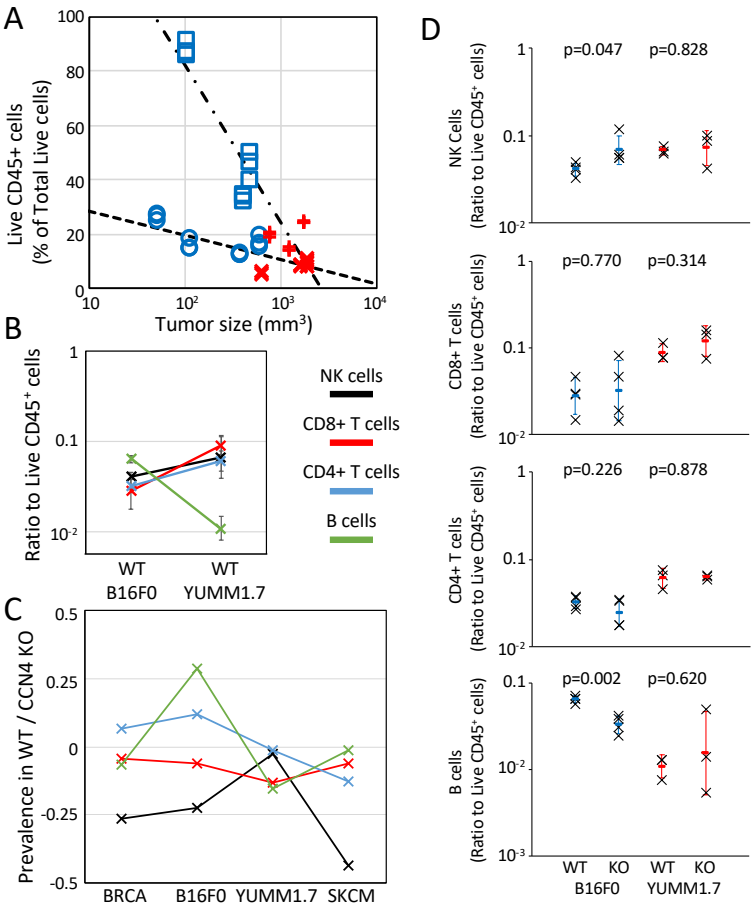

Fig. 6. CCN4 knock-out in two syngeneic mouse models of melanoma induces a similar shift in NK cells and T and B lymphocytes as observed in human breast cancer and melanoma.

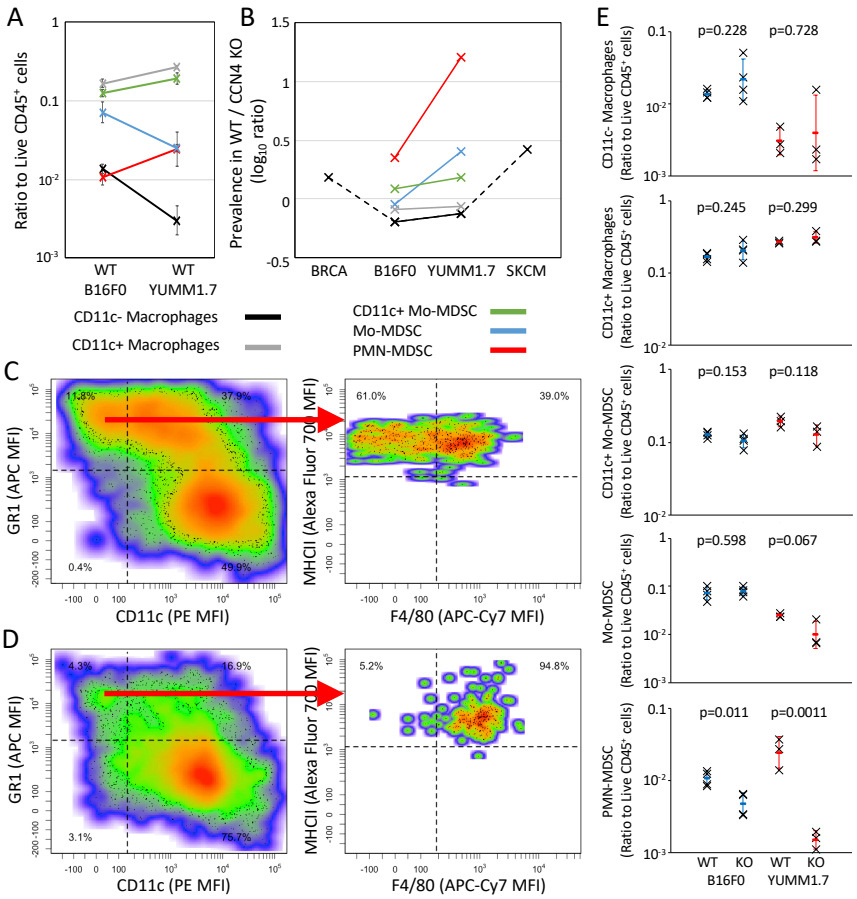

Fig. 7. Myeloid immune cell subsets differentially infiltrate tumors derived from wt B16F0 and YUMM1.7 cells but shift in similar ways upon CCN4 knock-out.

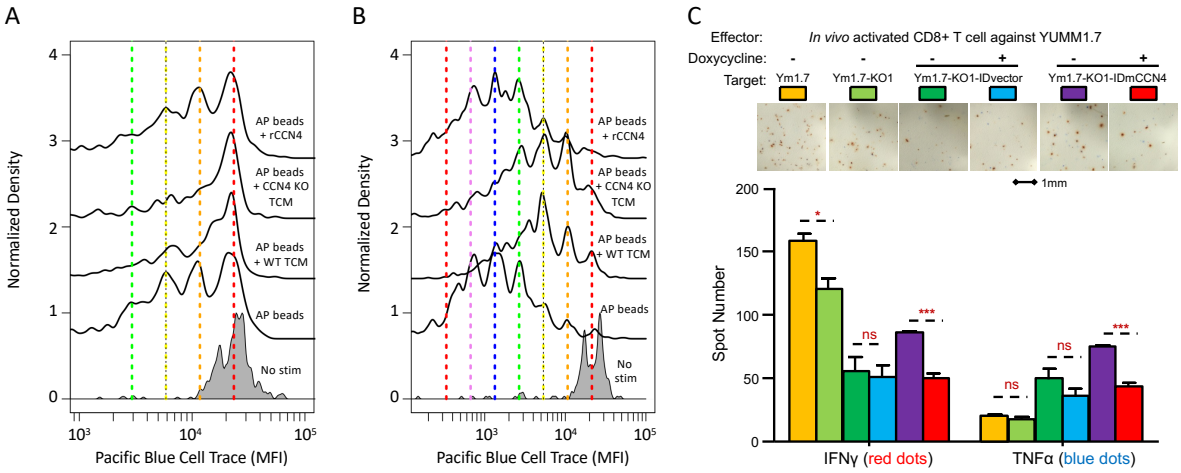

Fig. 8. CCN4 has no direct effect on T cell proliferation but impairs CD8$^+$ T cell function.

Table S1. List of edges, whether an edge was predicted to promote or inhibit the target node, and the strength inferred using the different structure learning algorithms in analyzing the features present in TCGA breast cancer dataset. Rows highlighted in green were included in the consensus seed network, yellow indicate that the directionality was unclear, and red indicate edges included in the "blacklist". The edge numbers correspond to the x-axis in Figure 2.

Edge No	from	to	mmpc CorSign	aracne CorSign	hiton CorSign	iamb strength	iamb CorSign	iamb.fdr strength	iamb.fdr CorSign	tabu strength	tabu CorSign	mmhc strength	mmhc CorSign	hc strength	hc CorSign	pc_stable strength	pc_stable CorSign	Hit Number	Min strength	Max strength
1	CCN4	Mesenchymal	+	+	+	2.02E-138	+	2.02E-138	+	1.95E-148	+	1.79E-139	+	1.95E-148	+	1.79E-139	+	9	1.95E-148	2.02E-138
1	Mesenchymal	CCN4	+	+	+													9		
2	pM0	pM1	-	-	-					3.77E-48	-			3.77E-48	-			9	0.00E+00	3.77E-48
2	pM1	pM0	-	-	-	0.00E+00	-	0.00E+00	-			0.00E+00	-			0.00E+00	-	9		
3	pM0	pM2	-	-	-					0.00E+00	-			0.00E+00	-			9	0.00E+00	0.00E+00
3	pM2	pM0	-	-	-	0.00E+00	-	0.00E+00	-			0.00E+00	-			0.00E+00	-	9		
4	CAF_lg	Mesenchymal	+	+	+													9	2.94E-97	9.62E-73
4	Mesenchymal	CAF_lg	+	+	+	9.62E-73	+	9.62E-73	+	2.94E-97	+	2.35E-86	+	2.94E-97	+	8.08E-86	+	9		
5	Cancer	CCN4	+	+	+	2.06E-55	+	2.06E-55	+	2.06E-55	+	2.06E-55	+	2.06E-55	+	2.06E-55	+	9	2.06E-55	2.06E-55
5	CCN4	Cancer	+	+	+													9		
6	NK.cells.active_lg	NK.cells.rest_lg	-	-	-	4.80E-30	-	4.80E-30	-	1.15E-33	-	4.80E-30	-	1.15E-33	-	4.80E-30	-	9	1.15E-33	4.80E-30
6	NK.cells.rest_lg	NK.cells.active_lg	-	-	-													9		
7	B.cells.ive_lg	T.cells.CD8_lg	+	+	+	9.25E-27	+	2.92E-27	+	3.10E-16	+	2.92E-27	+	3.10E-16	+	2.92E-27	+	9	2.92E-27	3.10E-16
7	T.cells.CD8_lg	B.cells.ive_lg	+	+	+													9		
8	Cancer	Epithelial	-	-	-	1.66E-25	-	1.66E-25	-	1.66E-25	-	1.66E-25	-	1.66E-25	-	2.23E-22	-	9	1.66E-25	2.23E-22
8	Epithelial	Cancer	-	-	-													9		
9	CAF_lg	proliferation	-	-	-							7.61E-26	-					9	7.61E-26	3.37E-12
9	proliferation	CAF_lg	-	-	-	6.37E-21	-	6.37E-21	-	3.37E-12	-			3.37E-12	-	1.80E-18	-	9		
10	Endothelial.cells_lg	proliferation	-	-	-					5.75E-10	-			5.75E-10	-			9	6.23E-18	5.75E-10
10	proliferation	Endothelial.cells_lg	-	-	-	2.88E-15	-	2.88E-15	-			2.88E-15	-			6.23E-18	-	9		
11	CD4Tcell_sc_lg	Macrophages_sc_lg	+	+	+													9	1.91E-16	2.01E-13
11	Macrophages_sc_lg	CD4Tcell_sc_lg	+	+	+	1.91E-16	+	5.09E-16	+	3.07E-14	+	5.09E-16	+	3.07E-14	+	2.01E-13	+	9		
12	pM1	T.cells.CD8_lg	+	+	+	8.18E-13	+	3.53E-14	+	8.98E-15	+	3.53E-14	+	8.98E-15	+	3.53E-14	+	9	8.98E-15	8.18E-13
12	T.cells.CD8_lg	pM1	+	+	+													9		
13	Macrophages_sc_lg	T.cells.CD8_lg	+	+	+	9.15E-13	+	2.38E-12	+	6.56E-12	+	2.38E-12	+	6.56E-12	+	2.38E-12	+	9	9.15E-13	6.56E-12
13	T.cells.CD8_lg	Macrophages_sc_lg	+	+	+													9		
14	CAF_lg	T.cells.CD8_lg	-	-	-	2.84E-07	-	3.06E-07	-	3.42E-09	-	3.06E-07	-	3.42E-09	-	3.06E-07	-	9	3.42E-09	3.06E-07
14	T.cells.CD8_lg	CAF_lg	-	-	-													9		
15	Cancer	proliferation	+		+	6.68E-23	+	6.68E-23	+	4.70E-12	+	3.72E-10	+	4.70E-12	+	1.93E-08	+	8	6.68E-23	1.93E-08
15	proliferation	Cancer	+		+													8		
16	Cancer	Endothelial.cells_lg	-		-	7.40E-09	-	7.40E-09	-	3.67E-38	-	7.40E-09	-	3.67E-38	-	2.10E-05	-	8	3.67E-38	2.10E-05
16	Endothelial.cells_lg	Cancer	-		-													8		
17	CD4Tcell_sc_lg	Endothelial.cells_lg	+		+													8	3.64E-10	7.84E-07
17	Endothelial.cells_lg	CD4Tcell_sc_lg	+		+	5.65E-08	+	2.93E-07	+	3.64E-10	+	2.93E-07	+	3.64E-10	+	7.84E-07	+	8		
18	B.cells.ive_lg	CD4Tcell_sc_lg	+		+	2.95E-06	+	1.95E-06	+	2.17E-07	+	1.95E-06	+	2.17E-07	+	7.31E-05	+	8	2.17E-07	7.31E-05
18	CD4Tcell_sc_lg	B.cells.ive_lg	+		+													8		
19	Cancer	pM2	-	-	-					7.00E-04	-	1.91E-42	-	7.00E-04	-	1.91E-42	-	7	1.91E-42	7.00E-04
19	pM2	Cancer	-	-	-													7		
20	B.cells.ive_lg	Endothelial.cells_lg	+	+	+											4.01E-20	+	7	2.71E-22	1.52E-18
20	Endothelial.cells_lg	B.cells.ive_lg	+	+	+	2.71E-22	+	2.71E-22	+			1.52E-18	+					7		
21	pM2	proliferation	-	-	-					1.02E-10	-	2.09E-11	-	1.02E-10	-	3.30E-10	-	7	2.09E-11	3.30E-10
21	proliferation	pM2	-	-	-													7		
22	Epithelial	Neutrophils_lg	-		-	2.82E-03	-			1.62E-03	-	6.79E-04	-	1.62E-03	-	6.79E-04	-	7	6.79E-04	2.82E-03
22	Neutrophils_lg	Epithelial	-		-													7		
23	CAF_lg	pM1	-		-													6	3.74E-24	4.37E-12
23	pM1	CAF_lg	-		-					4.37E-12	-	3.74E-24	-	4.37E-12	-	1.06E-21	-	6		
24	Cancer	Mesenchymal	+		+					8.89E-17	+	1.16E-16	+	8.89E-17	+	1.16E-16	+	6	8.89E-17	1.16E-16
24	Mesenchymal	Cancer	+		+													6		
25	Neutrophils_lg	pM2	+		+													6	7.51E-09	1.57E-04
25	pM2	Neutrophils_lg	+		+					7.51E-09	+	1.57E-04	+	7.51E-09	+	1.57E-04	+	6		
26	Macrophages_sc_lg	Neutrophils_lg	+		+	1.59E-03	+					1.86E-03	+			1.86E-03	+	5	1.59E-03	1.86E-03
26	Neutrophils_lg	Macrophages_sc_lg	+		+													5		
27	pM1	pM2								0.00E+00	-			0.00E+00	-			4	0.00E+00	3.43E-04
27	pM2	pM1				3.43E-04	-	3.43E-04	-									4		
28	NK.cells.rest_lg	T.cells.CD8_lg				1.24E-02	-			7.88E-04	-			7.88E-04	-			3	7.88E-04	1.24E-02
28	T.cells.CD8_lg	NK.cells.rest_lg																3		
29	NK.cells.active_lg	CD4Tcell_sc_lg				1.26E-03	-			1.08E-03	-			1.08E-03	-			3	1.08E-03	1.26E-03
29	CD4Tcell_sc_lg	NK.cells.active_lg																3		
30	B.cells.ive_lg	Epithelial	+													1.73E-03	+	3	1.73E-03	8.19E-03
30	Epithelial	B.cells.ive_lg	+									8.19E-03	+					3		
31	CD4Tcell_sc_lg	pM1	+		+													3	1.23E-02	1.23E-02
31	pM1	CD4Tcell_sc_lg	+		+											1.23E-02	+	3		
32	Cancer	pM1								2.35E-35	+			2.35E-35	+			2	2.35E-35	2.35E-35
32	pM1	Cancer																2		
33	Cancer	pM0								5.03E-25	+			5.03E-25	+			2	5.03E-25	5.03E-25
33	pM0	Cancer																2		
34	Mesenchymal	Endothelial.cells_lg								4.02E-19	+			4.02E-19	+			2	4.02E-19	4.02E-19
34	Endothelial.cells_lg	Mesenchymal																2		
35	B.cells.ive_lg	pM1								4.18E-19	+			4.18E-19	+			2	4.18E-19	4.18E-19
35	pM1	B.cells.ive_lg																2		
36	Macrophages_sc_lg	CAF_lg								5.33E-17	-			5.33E-17	-			2	5.33E-17	5.33E-17
36	CAF_lg	Macrophages_sc_lg																2		
37	Cancer	B.cells.ive_lg								3.36E-13	-			4.79E-07	-			2	3.36E-13	4.79E-07
37	B.cells.ive_lg	Cancer																2		
38	pM0	Mesenchymal								1.51E-12	+			1.51E-12	+			2	1.51E-12	1.51E-12
38	Mesenchymal	pM0																2		
39	pM1	Endothelial.cells_lg								4.23E-12	+			4.23E-12	+			2	4.23E-12	4.23E-12
39	Endothelial.cells_lg	pM1																2		
40	CCN4	proliferation								7.29E-12	-			7.29E-12	-			2	7.29E-12	7.29E-12
40	proliferation	CCN4																2		
41	Cancer	T.cells.CD8_lg								2.30E-10	-			2.30E-10	-			2	2.30E-10	2.30E-10
41	T.cells.CD8_lg	Cancer																2		
42	Macrophages_sc_lg	pM0								2.14E-08	-			2.14E-08	-			2	2.14E-08	2.14E-08
42	pM0	Macrophages_sc_lg																2		
43	NK.cells.rest_lg	pM0								2.12E-07	+			2.12E-07	+			2	2.12E-07	2.12E-07
43	pM0	NK.cells.rest_lg																2		
44	Mesenchymal	Neutrophils_lg								2.39E-07	+			2.39E-07	+			2	2.39E-07	2.39E-07
44	Neutrophils_lg	Mesenchymal																2		
45	CCN4	Macrophages_sc_lg								4.84E-07	+			4.84E-07	+			2	4.84E-07	4.84E-07
45	Macrophages_sc_lg	CCN4																2		
46	Cancer	NK.cells.rest_lg								4.25E-06	-			4.25E-06	-			2	4.25E-06	4.25E-06
46	NK.cells.rest_lg	Cancer																2		
47	Epithelial	CD4Tcell_sc_lg								9.98E-06	-			9.98E-06	-			2	9.98E-06	9.98E-06
47	CD4Tcell_sc_lg	Epithelial																2		
48	B.cells.ive_lg	NK.cells.rest_lg								3.72E-05	-			3.72E-05	-			2	3.72E-05	3.72E-05
48	NK.cells.rest_lg	B.cells.ive_lg																2		
49	Epithelial	Mesenchymal								7.35E-05	+			7.35E-05	+			2	7.35E-05	7.35E-05
49	Mesenchymal	Epithelial																2		
50	Endothelial.cells_lg	Neutrophils_lg								7.61E-05	-			7.61E-05	-			2	7.61E-05	7.61E-05
50	Neutrophils_lg	Endothelial.cells_lg																2		
51	pM2	CAF_lg								3.71E-04	+			3.71E-04	+			2	3.71E-04	3.71E-04
51	CAF_lg	pM2																2		
52	Mesenchymal	T.cells.CD8_lg								2.04E-04	+			2.04E-04	+			2	2.04E-04	2.04E-04
52	T.cells.CD8_lg	Mesenchymal																2		
53	CCN4	NK.cells.active_lg								2.64E-04	-			2.64E-04	-			2	2.64E-04	2.64E-04
53	NK.cells.active_lg	CCN4																2		
54	B.cells.ive_lg	CAF_lg								1.48E-03	-			1.48E-03	-			2	1.48E-03	1.48E-03
54	CAF_lg	B.cells.ive_lg																2		
55	Epithelial	CAF_lg								2.94E-03	+			2.94E-03	+			2	2.94E-03	2.94E-03
55	CAF_lg	Epithelial																2		
56	NK.cells.rest_lg	proliferation								2.80E-02	+			2.80E-02	+			2	2.80E-02	2.80E-02
56	proliferation	NK.cells.rest_lg																2		
57	Macrophages_sc_lg	proliferation								1.53E-01	-			1.53E-01	-			2	1.53E-01	1.53E-01
57	proliferation	Macrophages_sc_lg																2		
58	CCN4	B.cells.ive_lg								5.87E-08	+							1	5.87E-08	5.87E-08
58	B.cells.ive_lg	CCN4																1		

Table S2. List of edges, whether an edge was predicted to promote or inhibit the target node, and the strength inferred using the different structure learning algorithms in analyzing the features present in dataset comprised of common melanocytic nevi and primary melanoma tissue samples (GEO). Rows highlighted in green were included in the consensus seed network, yellow indicate that the directionality was unclear, and red indicate edges included in the "blacklist". The edge numbers correspond to the x-axis in Figure 4A.

Edge No	from	to	mmpc CorSign	aracne CorSign	hiton CorSign	iamb strength	iamb CorSign	iamb.fdr strength	iamb.fdr CorSign	tabu strength	tabu CorSign	mmhc strength	mmhc CorSign	hc strength	hc CorSign	pc_stable strength	pc_stable CorSign	Hit Number	Min strength	Max strength
1	pM1	pM2	-	-	-	2.40E-41	-	2.40E-41	-	2.40E-41	-	2.40E-41	-	2.40E-41	-	2.40E-41	-	9	2.40E-41	2.40E-41
1	pM2	pM1	-	-	-													9		
2	pM0	pM2	-	-	-	3.31E-38	-	3.31E-38	-	3.31E-38	-	3.31E-38	-	3.31E-38	-	3.31E-38	-	9	3.31E-38	3.31E-38
2	pM2	pM0	-	-	-													9		
3	Cancer	Mesenchymal	+	+	+	6.29E-09	+	2.09E-18	+	8.02E-13	+	6.29E-09	+	8.02E-13	+	2.46E-12	+	9	2.09E-18	6.29E-09
3	Mesenchymal	Cancer	+	+	+													9		
4	Cancer	Epithelial	-	-	-	4.24E-15	-	4.24E-15	-	4.24E-15	-	4.24E-15	-	2.79E-16	-	4.24E-15	-	9	2.79E-16	4.24E-15
4	Epithelial	Cancer	-	-	-													9		
5	CAF_lg	Mesenchymal	+	+	+			6.70E-15	+							2.20E-12	+	9	6.70E-15	5.98E-05
5	Mesenchymal	CAF_lg	+	+	+	2.75E-14	+			7.46E-12	+	2.75E-14	+	5.98E-05	+			9		
6	Macrophages_sc_lg	T.cells.CD8_lg	+	+	+	9.32E-06	+	1.82E-11	+	7.68E-06	+	1.79E-04	+	7.68E-06	+	1.79E-04	+	9	1.82E-11	1.79E-04
6	T.cells.CD8_lg	Macrophages_sc_lg	+	+	+													9		
7	NK.cells.active_lg	NK.cells.rest_lg	-	-	-							1.02E-06	-					9	1.01E-09	1.68E-05
7	NK.cells.rest_lg	NK.cells.active_lg	-	-	-	1.01E-09	-	1.68E-05	-	1.01E-09	-			1.01E-09	-	1.68E-05	-	9		
8	CD4Tcell_sc_lg	Endothelial.cells_lg	+	+	+													9	8.60E-08	4.15E-07
8	Endothelial.cells_lg	CD4Tcell_sc_lg	+	+	+	4.15E-07	+	8.60E-08	+	4.15E-07	+	4.15E-07	+	4.15E-07	+	4.15E-07	+	9		
9	CCN4	Mesenchymal	+	+	+	6.48E-05	+			8.38E-07	+	6.48E-05	+	8.38E-07	+	2.29E-02	+	8	8.38E-07	2.29E-02
9	Mesenchymal	CCN4	+	+	+													8		
10	NK.cells.active_lg	T.cells.CD8_lg	+	+	+	1.05E-04	+			1.19E-03	+	3.93E-03	+	1.19E-03	+	3.93E-03	+	8	1.05E-04	3.93E-03
10	T.cells.CD8_lg	NK.cells.active_lg	+	+	+													8		
11	CD4Tcell_sc_lg	Macrophages_sc_lg	+	+	+													8	4.58E-04	4.58E-04
11	Macrophages_sc_lg	CD4Tcell_sc_lg	+	+	+	4.58E-04	+			4.58E-04	+	4.58E-04	+	4.58E-04	+	4.58E-04	+	8		
12	NK.cells.rest_lg	pM0	+	+	+					2.22E-04	+			2.22E-04	+	1.17E-03	+	7	6.30E-05	1.17E-03
12	pM0	NK.cells.rest_lg	+	+	+							6.30E-05	+					7		
13	pM1	T.cells.CD8_lg	+	+	+					5.96E-03	+	3.70E-03	+	5.96E-03	+	3.70E-03	+	7	3.70E-03	5.96E-03
13	T.cells.CD8_lg	pM1	+	+	+													7		
14	Cancer	NK.cells.active_lg		+		8.85E-16	+			8.85E-16	+			8.85E-16	+			4	8.85E-16	8.85E-16
14	NK.cells.active_lg	Cancer		+														4		
15	CAF_lg	Endothelial.cells_lg		+												3.11E-06	+	4	3.11E-06	8.04E-04
15	Endothelial.cells_lg	CAF_lg		+						8.04E-04	+			4.19E-04	+			4		
16	Cancer	proliferation				1.47E-09	+			6.39E-11	+			6.39E-11	+			3	6.39E-11	1.47E-09
16	proliferation	Cancer																3		
17	Macrophages_sc_lg	Mesenchymal		+														3	7.24E-09	7.24E-09
17	Mesenchymal	Macrophages_sc_lg		+						7.24E-09	+			7.24E-09	+			3		
18	Endothelial.cells_lg	pM0		-	-													3	1.27E-04	1.27E-04
18	pM0	Endothelial.cells_lg		-	-					1.27E-04	-							3		
19	B.cells.naive_lg	Endothelial.cells_lg		+														3	3.41E-03	3.41E-03
19	Endothelial.cells_lg	B.cells.naive_lg		+						3.41E-03	+			3.41E-03	+			3		
20	Cancer	CCN4								5.06E-11	+			5.06E-11	+			2	5.06E-11	5.06E-11
20	CCN4	Cancer																2		
21	Endothelial.cells_lg	Mesenchymal								2.75E-10	+			2.75E-10	+			2	2.75E-10	2.75E-10
21	Mesenchymal	Endothelial.cells_lg																2		
22	Macrophages_sc_lg	pM1								5.29E-07	+			5.29E-07	+			2	5.29E-07	5.29E-07
22	pM1	Macrophages_sc_lg																2		
23	Epithelial	pM0								3.92E-05	-			3.92E-05	-			2	3.92E-05	3.92E-05
23	pM0	Epithelial																2		
24	Endothelial.cells_lg	proliferation								5.19E-05	-			5.19E-05	-			2	5.19E-05	5.19E-05
24	proliferation	Endothelial.cells_lg																2		
25	CCN4	Neutrophils_lg								7.22E-04	+			7.22E-04	+			2	7.22E-04	7.22E-04
25	Neutrophils_lg	CCN4																2		
26	CAF_lg	T.cells.CD8_lg								8.87E-03	+			8.87E-03	+			2	8.87E-03	8.87E-03
26	T.cells.CD8_lg	CAF_lg																2		
27	Macrophages_sc_lg	NK.cells.rest_lg																2	2.08E-02	2.08E-02
27	NK.cells.rest_lg	Macrophages_sc_lg								2.08E-02	-			2.08E-02	-			2		
28	pM0	pM1								2.28E-02	-			2.28E-02	-			2	2.28E-02	2.28E-02
28	pM1	pM0																2		
29	Endothelial.cells_lg	Epithelial												9.69E-04	+			1	9.69E-04	9.69E-04
29	Epithelial	Endothelial.cells_lg																1		
30	CAF_lg	CCN4																1	2.35E-01	2.35E-01
30	CCN4	CAF_lg												2.35E-01	+			1		
31	Cancer	pM2		-														1	1.00E+00	1.00E+00
31	pM2	Cancer		-														1		

Table S3. List of edges, whether an edge was predicted to promote or inhibit the target node, and the strength inferred using the different structure learning algorithms in analyzing the features present in primary melanoma tissue samples in the TCGA SKCM dataset. Rows highlighted in green were included in the consensus seed network, yellow indicate that the directionality was unclear, and red indicate edges included in the "blacklist". The edge numbers correspond to the x-axis in Figure 4B.

Edge No	from	to	mmpc CorSign	aracne CorSign	hiton CorSign	iamb strength	iamb CorSign	iamb.fdr strength	iamb.fdr CorSign	tabu strength	tabu CorSign	mmhc strength	mmhc CorSign	hc strength	hc CorSign	pc_stable strength	pc_stable CorSign	Hit Number	Min strength	Max strength
1	pM0	pM2	-	-	-					6.36E-46	-	2.04E-18	-	6.36E-46	-			9	6.36E-46	2.04E-18
1	pM2	pM0	-	-	-	6.36E-46	-	6.36E-46	-							6.36E-46	-	9		
2	pM0	pM1	-	-	-					9.96E-07	-	9.96E-07	-	9.96E-07	-			9	1.59E-34	9.96E-07
2	pM1	pM0	-	-	-	1.59E-34	-	1.59E-34	-							1.59E-34	-	9		
3	CAF_lg	Mesenchymal	+	+	+					5.38E-14	+			3.26E-18	+			9	3.26E-18	5.38E-14
3	Mesenchymal	CAF_lg	+	+	+	3.26E-18	+	3.26E-18	+			3.26E-18	+			3.26E-18	+	9		
4	CAF_lg	Macrophages_sc_lg	+	+	+	1.45E-08	+	1.45E-08	+	8.38E-11	+	1.45E-08	+	1.02E-11	+	1.45E-08	+	9	1.02E-11	1.45E-08
4	Macrophages_sc_lg	CAF_lg	+	+	+													9		
5	Macrophages_sc_lg	T.cells.CD8_lg	+	+	+	1.43E-07	+	1.43E-07	+	1.23E-11	+	1.43E-07	+	1.23E-11	+	1.43E-07	+	9	1.23E-11	1.43E-07
5	T.cells.CD8_lg	Macrophages_sc_lg	+	+	+													9		
6	NK.cells.active_lg	NK.cells.rest_lg	-	-	-					2.37E-09	-			2.37E-09	-	2.37E-09	-	9	2.01E-10	2.37E-09
6	NK.cells.rest_lg	NK.cells.active_lg	-	-	-	2.01E-10	-	2.01E-10	-			2.01E-10	-					9		
7	Endothelial.cells_lg	Mesenchymal	+	+	+											1.28E-07	+	8	3.73E-08	1.28E-07
7	Mesenchymal	Endothelial.cells_lg	+	+	+	3.73E-08	+			3.73E-08	+	3.73E-08	+	3.73E-08	+			8		
8	CD4Tcell_sc_lg	Macrophages_sc_lg	-	-	-													8	1.96E-06	2.61E-04
8	Macrophages_sc_lg	CD4Tcell_sc_lg	-	-	-	2.61E-04	-			1.96E-06	-	2.61E-04	-	1.96E-06	-	2.61E-04	-	8		
9	CCN4	NK.cells.rest_lg	+	+	+	1.32E-03	+			1.57E-02	+	1.32E-03	+	1.57E-02	+	1.57E-02	+	8	1.32E-03	1.57E-02
9	NK.cells.rest_lg	CCN4	+	+	+													8		
10	CCN4	Mesenchymal	+	+	+	1.14E-06	+			2.38E-02	+					3.83E-06	+	6	1.14E-06	2.38E-02
10	Mesenchymal	CCN4	+	+	+													6		
11	NK.cells.active_lg	pM0		-						3.16E-05	-			3.16E-05	-			3	3.16E-05	3.16E-05
11	pM0	NK.cells.active_lg		-														3		
12	Macrophages_sc_lg	proliferation		-						1.72E-04	-			1.72E-04	-			3	1.72E-04	1.72E-04
12	proliferation	Macrophages_sc_lg		-														3		
13	Endothelial.cells_lg	Epithelial		+						1.01E-02	+			1.01E-02	+			3	1.01E-02	1.01E-02
13	Epithelial	Endothelial.cells_lg		+														3		
14	CD4Tcell_sc_lg	Epithelial		+														3	1.36E-02	1.36E-02
14	Epithelial	CD4Tcell_sc_lg		+						1.36E-02	+			1.36E-02	+			3		
15	pM1	pM2								9.70E-30	-			9.70E-30	-			2	9.70E-30	9.70E-30
15	pM2	pM1																2		
16	Macrophages_sc_lg	NK.cells.active_lg																2	2.51E-06	1.29E-04
16	NK.cells.active_lg	Macrophages_sc_lg								2.51E-06	+			1.29E-04	+			2		
17	CAF_lg	CCN4																2	4.75E-06	4.75E-06
17	CCN4	CAF_lg								4.75E-06	+			4.75E-06	+			2		
18	CAF_lg	T.cells.CD8_lg								1.58E-05	-			1.58E-05	-			2	1.58E-05	1.58E-05
18	T.cells.CD8_lg	CAF_lg																2		
19	CAF_lg	CD4Tcell_sc_lg								5.54E-03	+			5.54E-03	+			2	5.54E-03	5.54E-03
19	CD4Tcell_sc_lg	CAF_lg																2		
20	Macrophages_sc_lg	pM1																1	1.64E-02	1.64E-02
20	pM1	Macrophages_sc_lg												1.64E-02	+			1		
21	B.cells.naive_lg	T.cells.CD8_lg		+														1	1.00E+00	1.00E+00
21	T.cells.CD8_lg	B.cells.naive_lg		+														1		
22	NK.cells.active_lg	T.cells.CD8_lg		+														1	1.00E+00	1.00E+00
22	T.cells.CD8_lg	NK.cells.active_lg		+														1		

Table S4. Proliferation metrics associated CD4$^+$ and CD8$^+$ T cells stimulated in vitro in different conditions. DiI: fraction diluted; PF: Precursor frequency, %dividing cells; PI: Proliferation index; and SD^D: proliferation variance. Summary statistics were calculated from three biological replicates and represented as mean (standard deviation). Statistical significance was assessed using type III repeated measures ANOVA, where * indicates a p-value < 0.05.

	Live CD4+ T cells				Live CD8+ T cells			
Experimental Conditions	DiI	PF	PI	SD^D	DiI	PF	PI	SD^D
AP beads + rCCN4	0.670 (0.012)	0.392 (0.014)	1.407 (0.033)	0.274	0.983 (0.003)	0.851 (0.014)	2.655 (0.045)	0.103
AP beads + CCN4 KO TCM	0.472* (0.008)	0.221* (0.003)	1.404 (0.023)	0.274	0.914* (0.015)	0.715* (0.032)	1.729* (0.044)	0.205
AP beads + WT TCM	0.552* (0.047)	0.282* (0.038)	1.403 (0.031)	0.272	0.920* (0.020)	0.704* (0.044)	1.923* (0.052)	0.189
AP beads	0.655 (0.043)	0.366 (0.035)	1.473 (0.060)	0.255	0.982 (0.002)	0.841 (0.026)	2.756 (0.187)	0.106
No stimulation	0.046 (0.025)	0.016 (0.013)	1.763 (1.056)	0.371	0.062 (0.018)	0.008 (0.004)	2.520 (0.441)	0.229

Table S5. List of fluorophore-conjugated antibodies using to quantify cell subsets by flow cytometry.

Marker	Clone	Fluorophore	Manufacturer
LIVE/DEAD Fix	--	Violet/Pacific Blue	Invitrogen
CD45	30-F11	BB515	BD Biosciences #564590
CD3e	500A2	Alexa Fluor 700	BioLegend #152316
CD4	GK1.5	APC-Cy7	BD Biosciences #552051
CD8a	REA601	APC	Miltenyi 130-109-248
CD161 (NK-1.1)	PK136	APC-Cy7	BioLegend #108723
CD45R/B220	RA3-6B2	APC	BioLegend #103212
CD49b	DX5	PerCP/Cy5.5	Biolegend #108915
CD11b	M1/70	PerCP/Cy5.5	eBioscience #45-0112-80
CD11c	N418	PE	eBioscience #12-0114-81
F4/80	BM8	APC-Cy7	BioLegend #123117
Ly-6G/Ly-6C (Gr-1)	RB6-8C5	APC	BioLegend #108412
CD279 (PD-1)	REA802	PE	BioLegend #135205
I-A/I-E (MHC-II)	M5/114.15.2	Alexa Fluor 700	BioLegend #107622

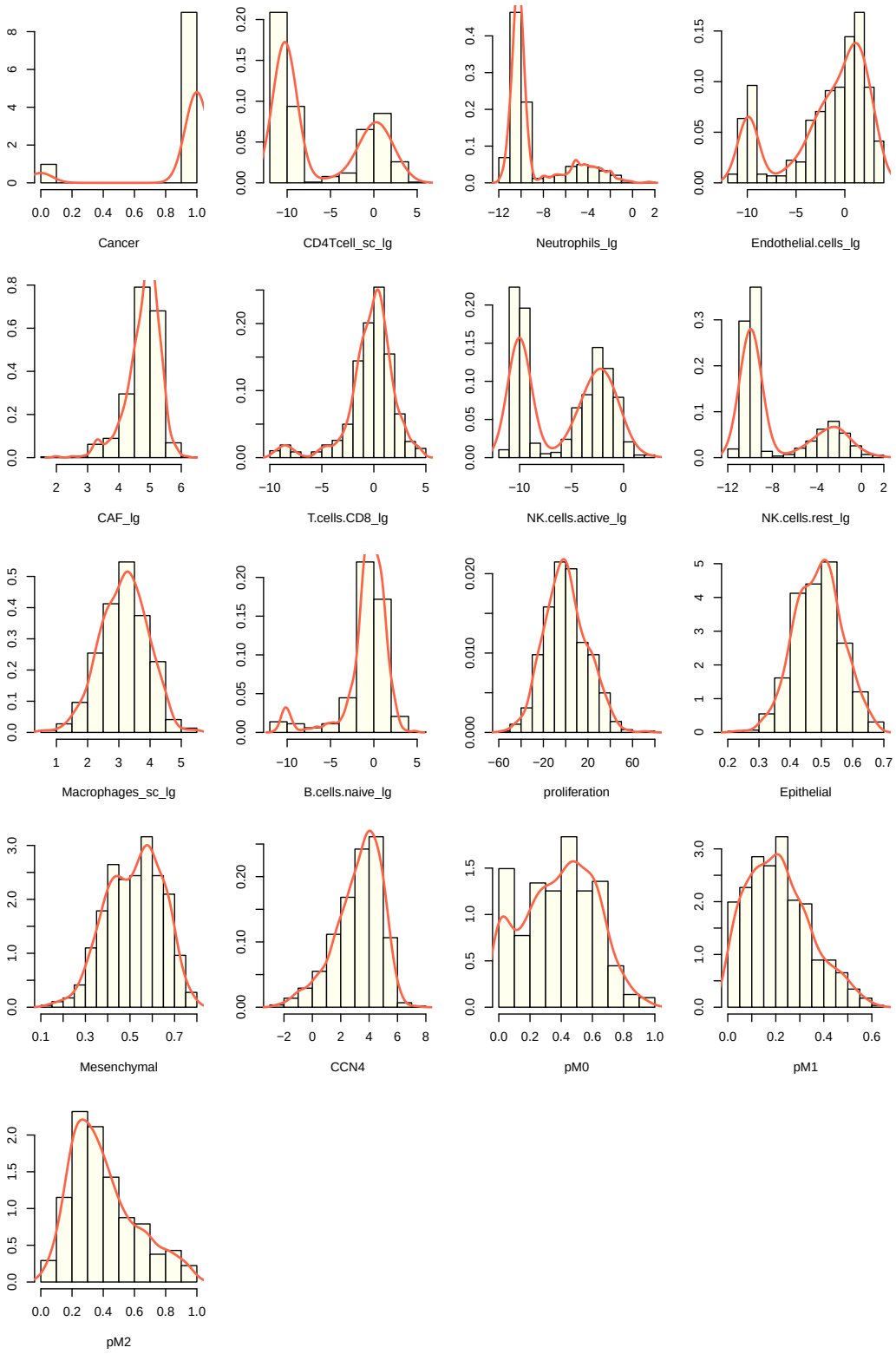

Fig. S1. Distribution of extracted features associated with the BRCA TCGA dataset. Figure represents a normalized histogram (bar graph) and distribution (red line) in log-transformed feature values. The panels from left to right, top to bottom are Cancer attribute, CD4 T cells, Neutrophils, Endothelial cells, Cancer associated fibroblasts (CAFs), CD8 T cells, active NK cells, resting NK cells, Macrophages, näive B cells, proliferation, epithelial cell state, mesenchymal cell state, CCN4 gene expression, p($M\Phi 0$), p($M\Phi 1$), and p($M\Phi 2$).

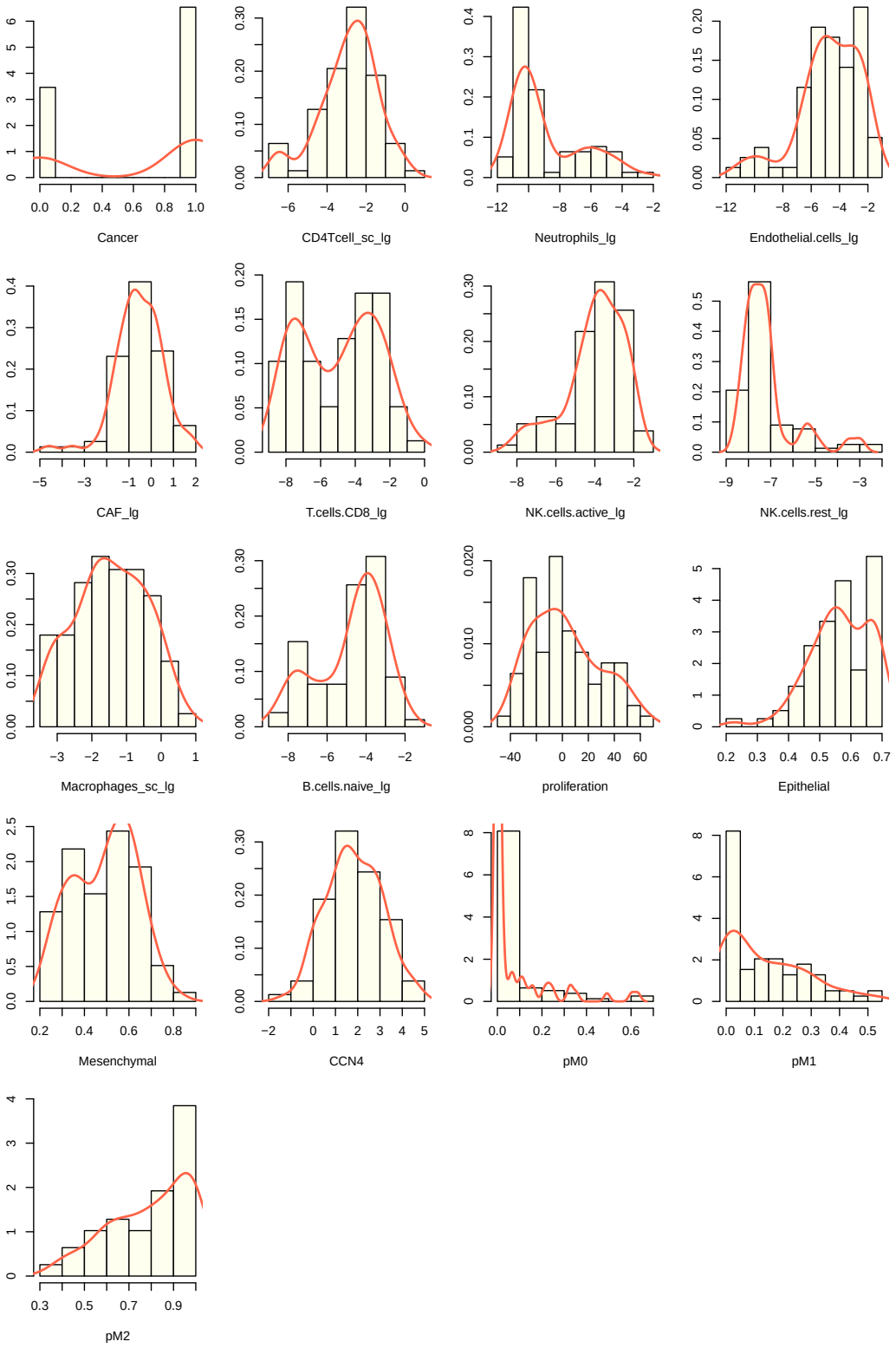

Fig. S2. Distribution of extracted features associated with the dataset containing common melanocytic nevi and primary melanoma tissue samples (GEO). Figure represents a normalized histogram (bar graph) and distribution (red line) in log-transformed feature values. The panels from left to right, top to bottom are Cancer attribute, CD4 T cells, Neutrophils, Endothelial cells, Cancer associated fibroblasts, CD8 T cells, active NK cells, resting NK cells, Macrophages, naïve B cells, proliferation, epithelial cell state, mesenchymal cell state, CCN4 gene expression, p($M\Phi 0$), p($M\Phi 1$), and p($M\Phi 2$).

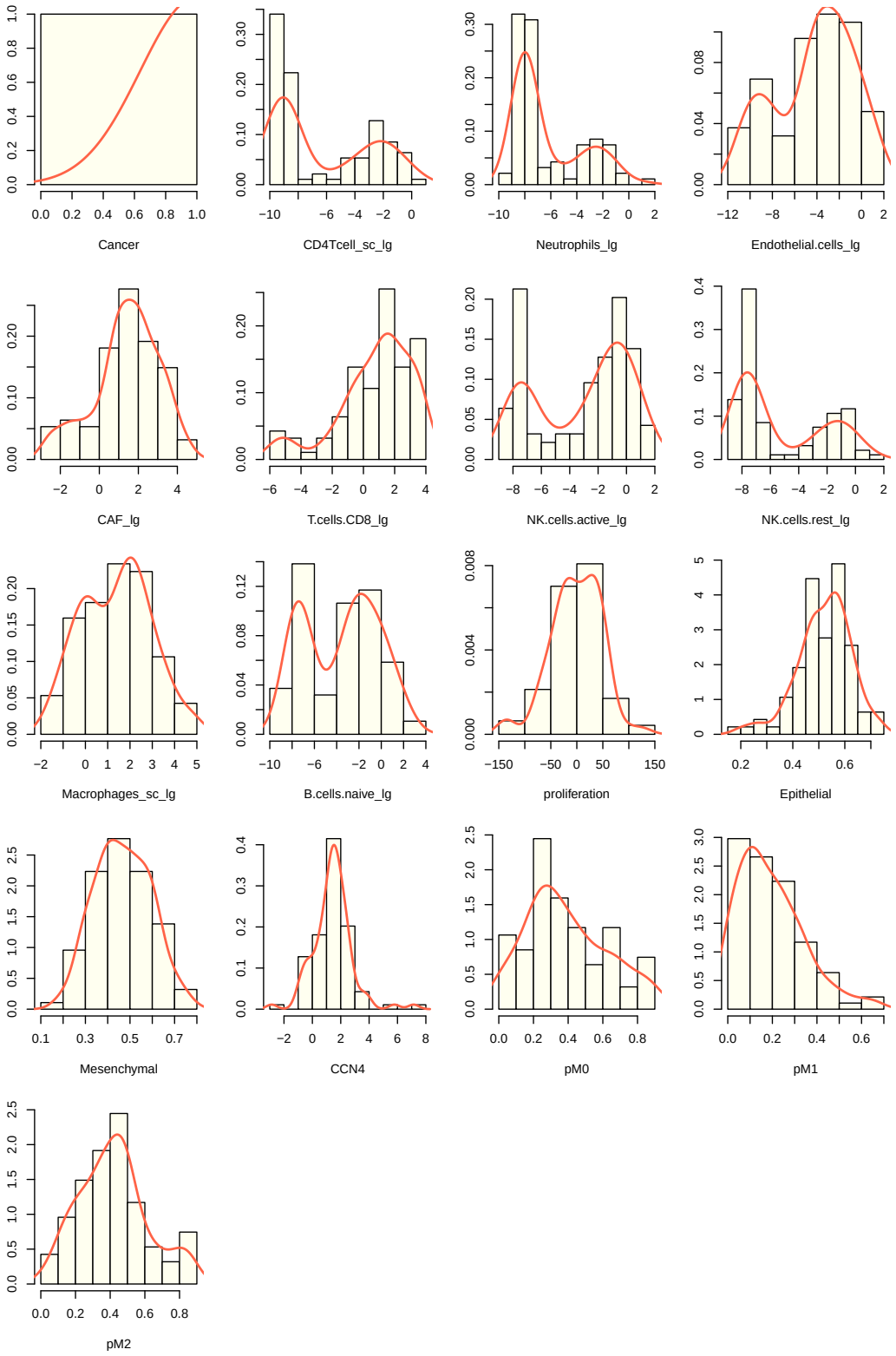

Fig. S3. Distribution of extracted features associated with primary melanoma samples in the TCGA SKCM dataset. Figure represents a normalized histogram (bar graph) and distribution (red line) in log-transformed feature values. The panels from left to right, top to bottom are Cancer attribute, CD4 T cells, Neutrophils, Endothelial cells, Cancer associated fibroblasts, CD8 T cells, active NK cells, resting NK cells, Macrophages, naïve B cells, proliferation, epithelial cell state, mesenchymal cell state, CCN4 gene expression, p($M\Phi 0$), p($M\Phi 1$), and p($M\Phi 2$).

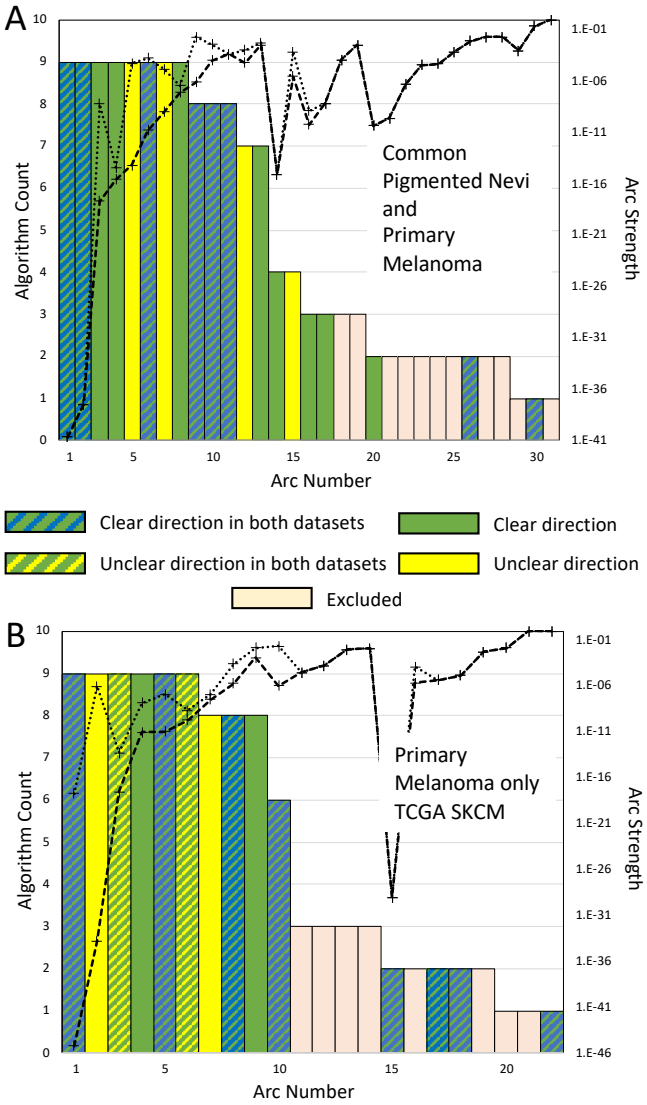

Fig. S4. Summary of the evidence obtained from two melanoma datasets supporting the consensus edges in the seed network. Analysis of datasets containing samples from both common pigmented nevi and primary melanoma (A) and from only primary melanoma (B). Edges ordered based on the number of algorithms that detected that an edge was enriched (bar graph - left axis) and the strength of enrichment (dotted lines - right axis). The lines associated with the strength of enrichment represent the minimum (dashed line) and maximum (dotted line) values obtained by the different algorithms. Coloring of bar graph indicates whether a clear direction was associated with an edge in one dataset (green) and in both datasets (green/blue), an edge was significantly enriched but without a clear direction in one dataset (yellow) and in both datasets (green/yellow), or that an edge was excluded from the consensus seed network list (tan). .

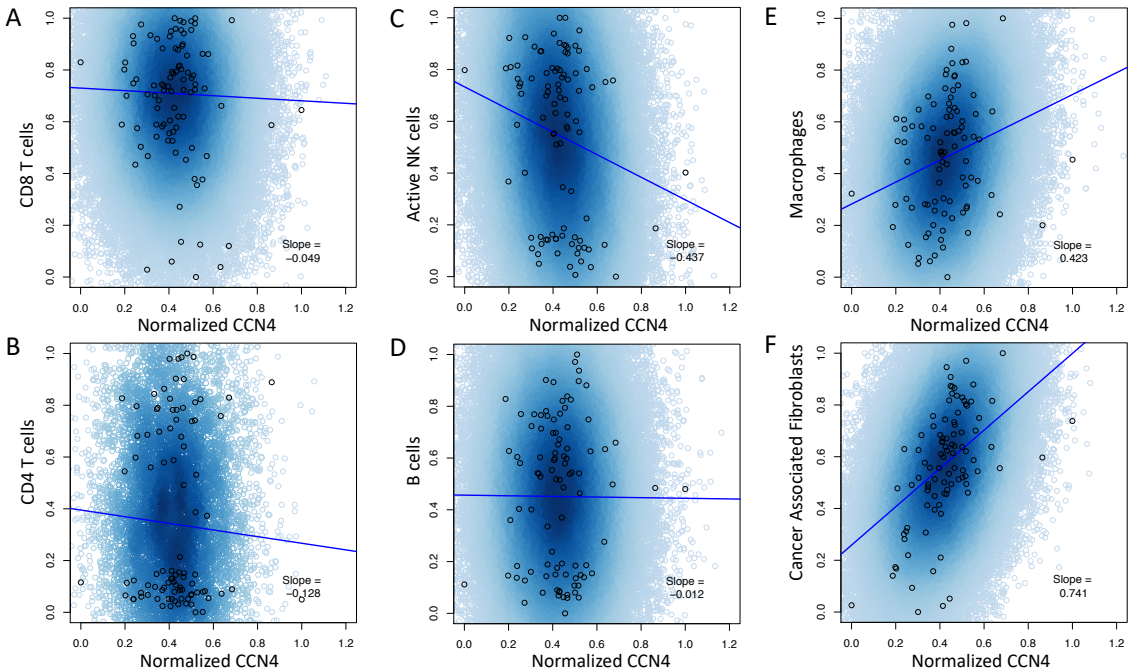

Figure S2B. Conditional probability query of the SKCM DAG compared against digital cytometry estimates obtained from experimental data. Experimental samples obtained from primary melanoma tissue are shown as open circles. Samples of the conditional probability model for p(Cancer > 0.85) (blue) for CD8 T cells (A), CD4 T cells (B), Active NK cells (C), B cells (D), Macrophages (E) and Cancer Associated Fibroblasts (F). Linear trendlines are superimposed on the conditional probability samples.

Fig. S5. Conditional probability query of the SKCM DAG compared against digital cytometry estimates obtained from experimental data. Experimental samples obtained from primary melanoma tissue are shown as open circles. Samples of the conditional probability model for p(Cancer $>$ 0.85) (blue) for CD8 T cells (A), CD4 T cells (B), active NK cells (C), B cells (D), Macrophages (E) and Cancer Associated Fibroblasts (F). Linear trendlines are superimposed on the conditional probability samples.

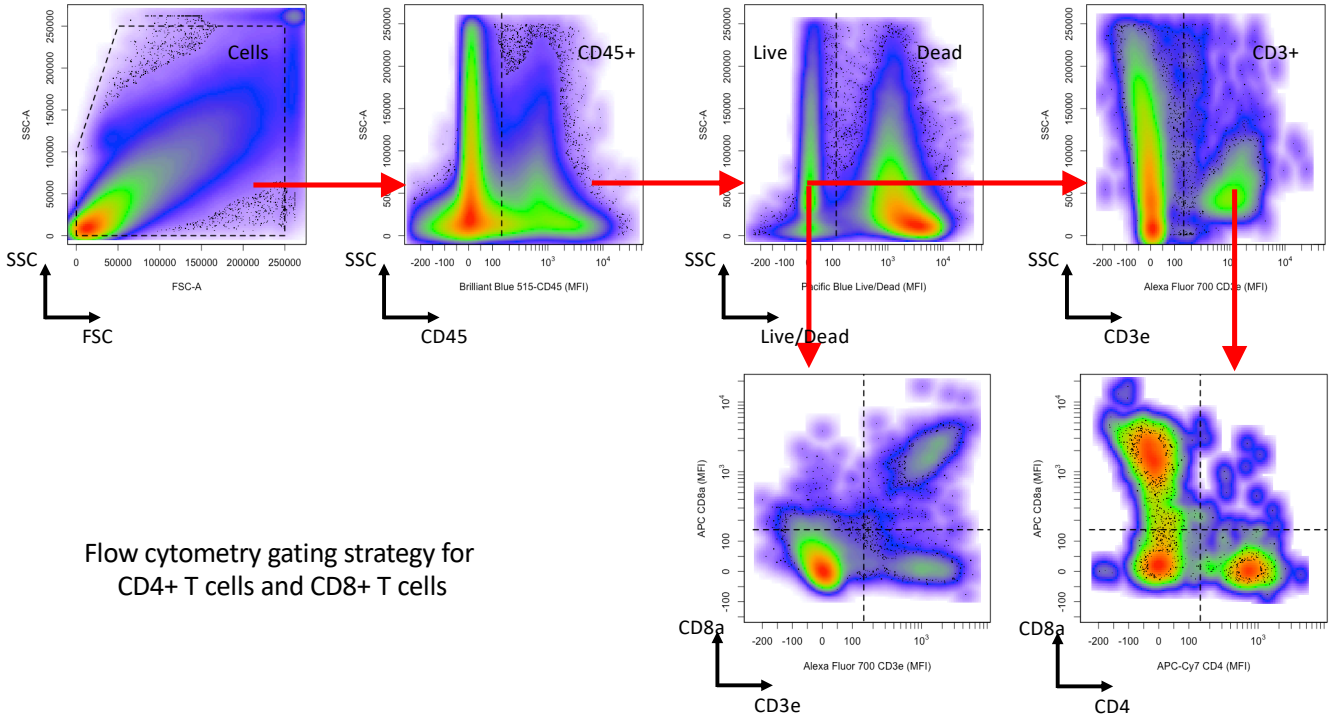

Fig. S6. Flow cytometry gating strategy for T cells. CD45 staining versus side scatter area was used to gate for CD45+ cells. Live Dead Pacific Blue staining versus side scatter area was used to then gate for Live CD45$^+$ cells, which were then gated based on CD3e$^+$ expression. Live CD45$^+$ CD3e$^+$ cells were further subdivided into CD8$^+$ T cells (live CD8$^+$ CD3e$^+$ CD45$^+$ cells), CD4 T cells (live CD4$^+$ CD3e$^+$ CD45$^+$ cells), and double negative T cells (live CD8$^-$ CD4$^-$ CD3e$^+$ CD45$^+$ cells).

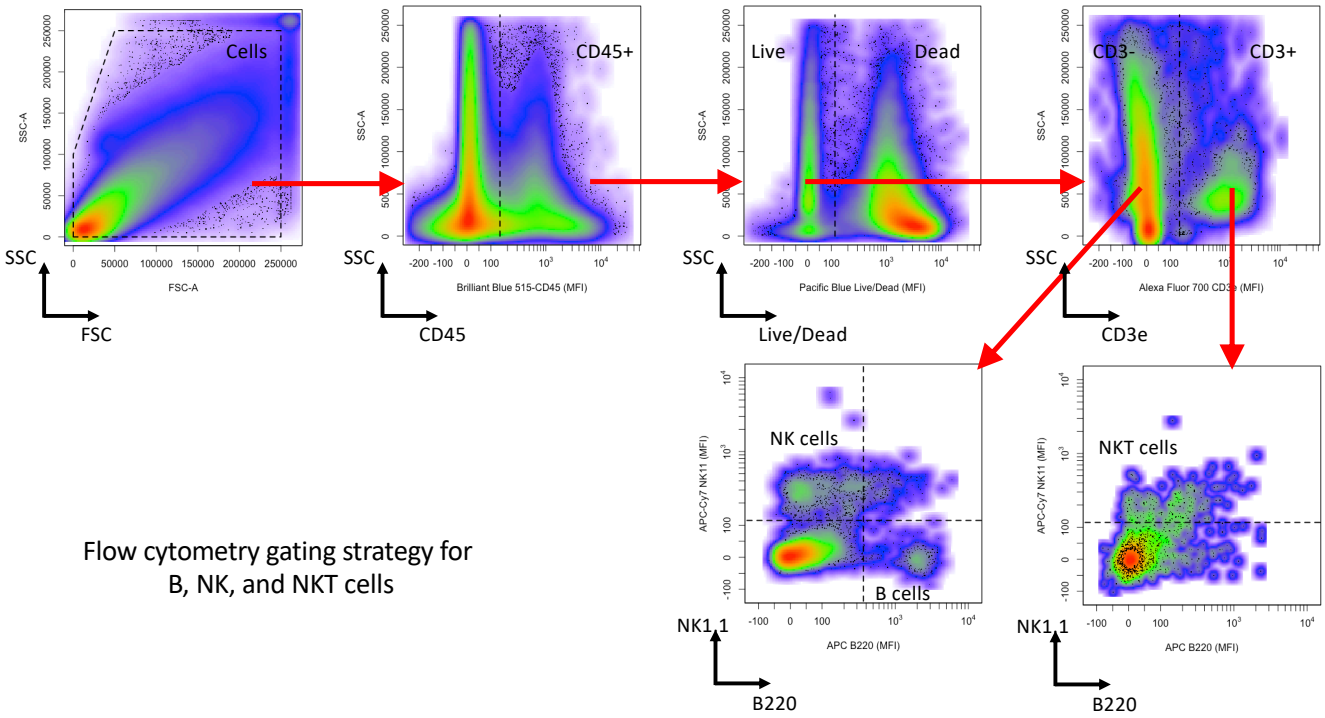

Fig. S7. Flow cytometry gating strategy for B, NK, and NKT cells. CD45 staining versus side scatter area was used to gate for CD45+ cells. Live Dead Pacific Blue staining versus side scatter area was used to gate for Live CD45^{+} cells, which were then subdivided into B cells (live NK1.1^{-} B220^{+} CD3^{-} CD45^{+} cells), NK cells (live NK1.1^{+} B220^{-} CD3^{-} CD45^{+} cells), and NKT cells (live NK1.1^{+} CD3e^{+} CD45^{+} cells).

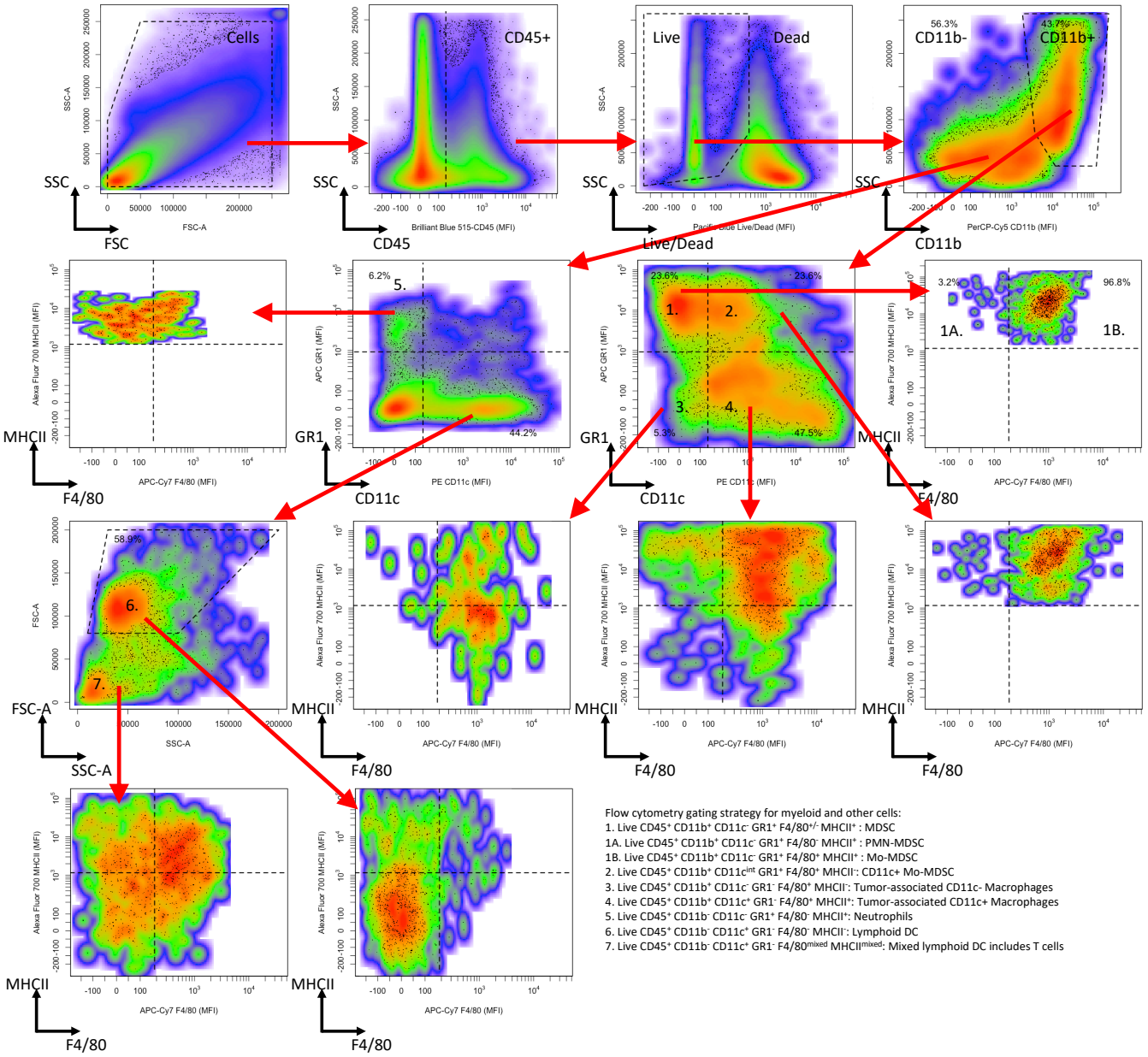

Fig. S8. Flow cytometry gating strategy for Tumor associated neutrophils and myeloid cell subsets. CD45 staining versus side scatter area was used to gate for CD45$^+$ cells. Live Dead Pacific Blue staining versus side scatter area was used to gate for Live CD45$^+$ cells, which were then subdivided into subsets based on CD11b staining followed by Gr1 versus CD11c staining. From the CD11b$^+$ gate, myeloid-derived suppressor cells (MDSC) (live Gr1$^+$ CD11c$^-$ CD11b$^+$ CD45$^+$ cells) were subdivided into PMN-MDSC (F4/80$^-$ MHCII$^+$) and Mo-MDSC (F4/80$^+$ MHCII$^+$). CD11c$^+$ Mo-MDSC were also quantified as (live F4/80$^+$ MHCII$^+$ Gr1$^+$ CD11c^{int} CD11b$^+$ CD45$^+$ cells). Also from the CD11b$^+$ gate, macrophages (live Gr1$^-$ F4/80$^+$ CD11b$^+$ CD45$^+$ cells) were subdivided into tumor-associated CD11c$^+$ (CD11c$^{int/+}$ MHCII^{hi}) and CD11c$^-$ (CD11c$^-$ MHCII^{lo})subsets. The CD11b$^-$ subset included tumor-associated neutrophils (TAN) (Gr1$^+$ CD11c$^-$ MHCII^{hi} F4/80$^-$) and lymphoid dendritic cells (Gr1$^-$ CD11c$^+$ FSC-A^{hi} MHCII^{lo} F4/80$^-$).

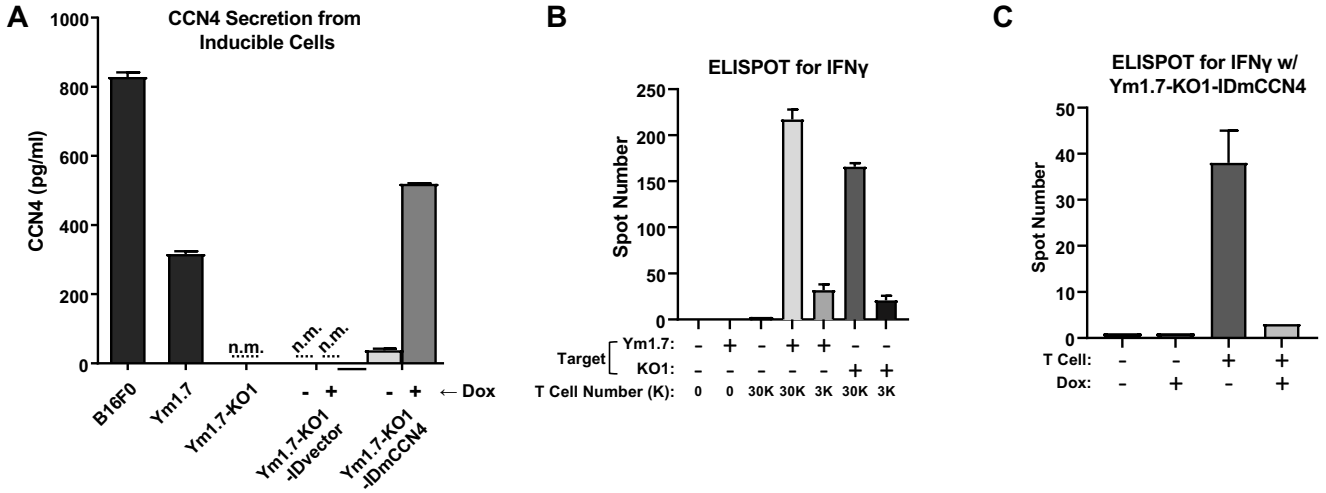

Fig. S9. Control experiments related to ELISPOT assay using an inducible CCN4 YUMM1.7 cell line. (A) CCN4 secretion, measured with ELISA, from CCN4-inducible cells in conditioned media in the presence of absence of doxycycline. (B) ELISPOT for IFNγ release with different target cells and different amount of effector CD8$^+$ T cells (In vivo activated CD8$^+$ T cells against YUMM1.7 (Ym1.7)). (C) ELISPOT for IFNγ with CCN4-inducible cells as targets using in vivo activated CD8+ T cell against YUMM1.7. Results shown as mean $\pm$ S.D. for three biological replicates.