

Topological Tumor Graphs: a graph-based spatial model to infer stromal recruitment for immunosuppression in melanoma histology

Henrik Failmezger^{1,2}, Sathya Muralidhar^{1,2*}, Antonio Rullan^{3,4}, Carlos E. de Andrea⁵, Erik Sahai³, Yinyin Yuan^{1,2}*

¹ Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

² Division of Molecular Pathology, The Institute of Cancer Research, London, UK

³ Tumor Cell Biology Laboratory, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, UK

⁴ Targeted therapy Laboratory, The Royal Marsden NHS Foundation Trust, London, SW3 6JJ, United Kingdom.

⁵ Clinica Universidad de Navarra, Pamplona, Spain; CIBERONC, Madrid, Spain.

* Joint first authors

Running title

Stromal recruitment and immunosuppression in melanoma

Keywords

Melanoma, stromal interactions, computational pathology

Financial support

Y.Y. acknowledges funding from Cancer Research UK Career Establishment Award (C45982/A21808), Breast Cancer Now (2015NovPR638), Children's Cancer and Leukaemia Group (CCLGA201906), NIH U54 CA217376 and R01 CA185138, CDMRP Breast Cancer Research Program Award BC132057, European Commission ITN (H2020-MSCA-ITN-2019), Wellcome Trust (105104/Z/14/Z), and The Royal Marsden/ICR National Institute of Health Research Biomedical Research Centre. SM is funded by NIH U54 CA217376, R01 CA185138. CEA is funded by the Instituto de Salud Carlos III (AC14/00034) and co-financed by Fondos FEDER (Spain). A.R., and E.S. are supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001144), the UK Medical Research Council (FC001144) and the Wellcome Trust (FC001144). A.R. is additionally supported by the Spanish Society for Medical Oncology (Beca Fundación SEOM)

Correspondence

Dr Yinyin Yuan
Division of Molecular Pathology
Brookes Lawley Building, 2E12.9
15 Cotswold Road, Sutton, Surrey, UK, SM2 5NG
+44 20 8722 4111
yinyin.yuan@icr.ac.uk

Conflict of interest

The authors declare no potential conflicts of interest.

Abstract

Despite the advent of immunotherapy, metastatic melanoma represents an aggressive tumor type with a poor survival outcome. The successful application of immunotherapy requires in-depth understanding of the biological basis and immunosuppressive mechanisms within the tumor microenvironment. In this study, we conducted spatially explicit analysis of the stromal-immune interface across 400 melanoma H&E specimens from TCGA (The Cancer Genome Atlas). A computational pathology pipeline (CRImage) was used to classify cells in the H&E specimen into stromal, immune or cancer cells. The estimated proportions of these cell types were validated by independent measures of tumor purity, pathologists' estimate of lymphocyte density, imputed immune cell subtypes and pathway analyses. Spatial interactions between these cell types were computed using a graph-based algorithm (Topological Tumor Graphs: TTGs). This approach identified two stromal features, namely stromal clustering and stromal barrier, which represented the melanoma stromal microenvironment. Tumors with increased stromal clustering and barrier were associated with reduced intratumoral lymphocyte distribution and poor overall survival independent of existing prognostic factors. To explore the genomic basis of these TTG-derived stromal phenotypes, we used a deep learning approach integrating genomic (copy number) and transcriptomic data, thereby inferring a compressed representation of copy number-driven alterations in gene expression. This integrative analysis revealed that tumors with high stromal clustering and barrier had reduced expression of pathways involved in naïve CD4 signalling, MAPK and PI3K signalling. Taken together, our findings support the immunosuppressive role of stromal cells within metastatic melanoma via physical barrier and T cell exclusion within the vicinity of cancer cells.

Statement of significance

Computational histology-based stromal phenotypes within the tumor microenvironment are significantly associated with prognosis and immune exclusion in melanoma.

Introduction

Cutaneous melanoma remains one of the most aggressive cancer types with global increase in incidence and mortality, despite the advent of immunotherapy (1,2). Late stage diagnosis is associated with poor survival related to metastatic spread of disease to distant anatomical sites (3–5). Despite durable responses in patients treated with systemic anti-CTLA-4 and anti-PD-1 immunotherapy in the past decade, the fact remains that a significant proportion of patients do not respond to treatment (6). To this effect, diverse intrinsic mechanisms of tumor immune evasion have been described, including loss of tumor antigenicity (7), upregulation of co-inhibitory immune checkpoint molecules and recruitment of regulatory T cells. On the other hand, immunosuppression is modulated by the tumor microenvironment, which is composed of cell types including fibroblasts, endothelial cells, immune cells and tissue-resident interstitial cells. Fibroblasts can act as immunosuppressive agents through paracrine signalling, modulating the cytotoxic activity of natural killer cells(8), and inhibition of the antitumoral T cell response(9). For example, remodelling of the extra cellular matrix (ECM) by fibroblasts has been showed to result in increased ECM rigidity because of thickening and linearization of collagen fibres (10,11). This modified ECM could restrict access of immune cells to cancer cells, serving as a physical barrier (12,13). Understanding the mechanisms by which this stroma network is generated has the potential to aid the development of new therapeutics. However, to date there is no established method to provide systems characterisation of the stroma-mediated immunosuppression.

We propose a new method to study the spatial structure of the tumor microenvironment, referred to as Topological Tumor Graphs (TTGs), which establishes spatial interactions among cell types as phenotypes of tumor-host interactions. We focus our analyses on specific cell types broadly defined by nuclear morphology, namely: tumor cells (large, round nuclei), lymphocytes (smaller, darker nuclei) and stromal cells (spindle-shaped cells likely to be fibroblasts, with possible inclusion of endothelial cells).

Based on TTGs inferred directly from pathological slide images, algorithms that are typically used for social network analysis can be directly applied to the investigation of architectural complexity of the tumor microenvironment. Building on the idea that the

TTGs transform seemingly chaotic cell distribution data in pathological slides, into highly structured microenvironmental phenotypes, integration with high-dimensional genomic data may inform cancer adaptive strategies that enable it to overcome external constraints. In this work, we: 1) developed a new graph-based method to characterise two distinct spatial features of the stromal microenvironment, stromal cell clustering and stromal barrier between cancer cells and lymphocytes; 2) developed a new unsupervised deep learning approach for simultaneous dimension reduction and integration of copy-number alterations (CNAs) and gene expression data, which further enables genotype-phenotype integration and 3) learned the clinical implications and biological processes underpinning stromal recruitment and their effects on immunosuppression.

Materials and Methods

Patients and samples

This study is based on 400 patients from TCGA SKCM cohort with no previous systemic therapy (except that adjuvant interferon- α ≥ 90 days prior was permitted) (14), wherein all specimens were obtained with written informed consent from the relevant institutional review board participated in TCGA. H&E-stained whole-tumor sections (obtained from the public TCGA Data Portal) from diagnostic samples were used owing to better preserved morphology compared to frozen samples. Clinical endpoint was overall survival (OS) which was censored at the date of death or the date of last contact (for living patients). Samples were randomly split into equal sized discovery and a validation cohorts.

A second cohort included 12 patients with melanoma who underwent serial biopsies were analyzed using the same methods (Table S1). Patients were treated with antibodies targeting CTLA4, PD-1 or a combination of both. Treatment was administered until disease progression or unacceptable toxicity. All specimens were obtained with consent from the patient in accordance to ethical approval 111/2010 granted by the institutional review board of the University of Navarre.

H&E image processing

An previously described computational pathology pipeline CRImage was used for image processing (15). In brief, watershed segmentation and Otsu thresholding for hematoxylin positive nuclei was performed. For every cell nucleus morphological features such as shape, intensity and texture features were calculated. Cell classification was based on a support vector (14–20) using 97 morphological and textural features. Each cell was classified into artefact, lymphocyte, cancer or stromal cell. The following measures were used to validate the automated cellular classifications for melanoma:

LScore: Pathologist-provided LScore was defined based on lymphocyte distribution and density within the tumors (21). LScore was calculated as the sum of two scores: lymphocyte distribution (0-3, 0 = no lymphocytes within the tissue, 1 = lymphocytes present involving <25% of the tissue cross sectional area, 2 = lymphocytes present in 25 to 50% of the tissue, 3 = lymphocytes present in >50% of tissue), and lymphocyte density (0-3; 0 = absent, 1 = mild, 2 = moderate, 3 = severe).

CPE (Consensus measure of Purity Estimation): Molecular measure of tumor purity value based on (22)

Imputed lymphocyte scores: From TIMER (23)

Immune phenotype clusters were derived from previously described (24) six immune phenotype groups as follows:

Immune phenotype 1 and 2 ('low cytotoxicity'): Cluster 1

Immune phenotype 3 and 4 ('intermediate cytotoxicity'): Cluster 2

Immune phenotype 5 and 6 ('high cytotoxicity'): Cluster 3

Differential Expression analyses: TCGA Biolinks

(<http://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html>) was used for differential gene expression, using the function TCGAanalyze_EAcomplete for enrichment analysis.

Topological Tumor Graphs

Based on cell spatial mapping provided by automated image analysis, each cell was treated as a node and edges between cells were drawn if they were in spatial proximity ($< 35\mu\text{m}$). This threshold was determined by calculating the distance of cancer cells to their closest stromal neighbour in regions of the tumor-stroma interface. The mean value of this distribution was used as threshold (Additional file 1: Fig. S3a). Cancer supernodes were created by deleting all edges of cancer cells to non-cancer cells. Afterwards the connected components in the remaining network were calculated. Connected components with more than 50 cells have been summarized in cancer supernodes and the network was connected again. For every network we kept only the largest connected component.

Network centrality measures

The node degree measures the number of neighbours (edges) of a node. The node degree for the whole network was calculated, regardless of the node's cell type. Cell type-specific node degrees were created by separating the nodes by their cell type (the neighbours were counted regardless of the cell type).

The clustering coefficient measures the degree in which the neighbourhood of a node is connected. For a particular node, clustering coefficient = $\frac{\text{Number of closed triplets}}{\text{Number of all triplets}}$, where a triplet consists of three nodes with edges connecting all of them.

Stromal clustering was defined as the average clustering coefficient of stromal cells within a tumor, whereas only the stromal neighbours of the (stromal) cell were considered in the calculation. In order to avoid biased results at the border of tumor clusters, we only calculated the clustering coefficient for stromal cells that did not directly border a cancer supernode.

Stromal barrier was calculated by counting the number of stromal cells that a lymphocyte has to cross in order to reach a cancer cluster. Lymphocytes that directly bordered a cancer cluster were excluded. The overall stromal barrier of a sample was calculated as the average of the individual stromal barriers of the lymphocytes in the sample.

Betweenness centrality measures the network flow or traffic in a communication network or graph based on shortest paths. The betweenness centrality for a node in the network is the number of the shortest paths between every pair of nodes that pass through this node. For each type of cells betweenness was computed and averaged to represent the betweenness centrality of this cell type within the cell network.

Survival analysis

Univariate Cox proportional hazards model was used to test the differential OS between patients whose tumors had low or high stromal clustering and barrier. Patients stratification into low (n=97) or high (n=97) stromal barrier was defined using the upper/lower quartile of stromal barrier, since this produced the best group separation. In the case of stromal clustering, since OS in the upper/lower quartile did not vary significantly, an optimal cut-off was identified. This cut-off (stromal clustering ≤ 0.590253) was identified in the discovery subset and found to be significant in the validation subset, stratifying the tumors into low (n=156) or high (n=230) stromal clustering. Multivariate Cox regression was performed using the same patient stratification, after adjusting for clinical parameters. The above described stratification of low/high stromal barrier and clustering were used to create the combination groups. Kaplan Meier curves were used for visual representation.

Genomic and transcriptomic data

RNAseq-derived transcriptomic data was downloaded from TCGA biolinks (<http://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html>)(25).

Thresholded copy number data was downloaded from the Broad Institute TCGA Genome Data Analysis Centre (Broad Institute TCGA Genome Data Analysis Centre (2016): SNP6 Copy number analysis (GISTIC2). Copy number data and the gene expression data were scaled between zero and one using min-max scaling.

For differential gene expression analysis, we used RNASeq data that was aligned against the human reference genome hg19. The data was normalized using within-lane and between-between lane normalization using (25) and filtered using quantile filtering with a threshold of 0.25 (25).

CNx

Integration of transcriptomic and copy number data was done using CNx, which was inspired by the Stacked Denoising Autoencoders (SDA) previously applied to gene expression data (26,27). Transcriptomic data were binarized using Gaussian mixture clustering and used in the loss function of the unsupervised deep learning network. In contrast to an SDA, the input for CNx was copy number data and the loss function was the cross-entropy between copy number data and binarized gene expression data. Genes that contributed to the nodes were determined by multiplying the weight matrices (27) of the individual layers and selecting the genes that have a weight that is more than two standard deviations from the mean.

The bottleneck layer of CNx was visualised as a network, by binarizing the node activities (28). We identified each node's highest and lowest activity values and defined 10 equally spaced activation thresholds between these values. We evaluated the agreement in the binarized activation values between two nodes at each threshold. Two nodes were connected if they had an agreement larger than 55% of the binarized activation values for one of the thresholds.

Software and Data

Image processing was performed using the Bioconductor packages CRImage (15) and EBImage (29). Topological Tumor Graphs were created using the Python package NetworkX (30). The encoder-decoder network was implemented using Python and Tensorflow (4). Genomic analysis was performed in R using the packages TCGAbiolinks (25).

Results

Mapping cell spatial distribution in whole-tumor histology melanoma sections

To spatially map the locations and types of cells, 400 full face, H&E-stained sections from FFPE (formalin-fixed paraffin-embedded) diagnostic blocks of melanoma patients from TCGA were subjected to fully automated image analysis (Fig. 1a, Fig. S1a, Methods). Samples consisted of primary tumors (n=91), regional lymph node metastases (n=191), metastases from regional skin (n=61) and from distant sites

(n=51), with 6 samples being unclassified. Our quantitative morphological classifier identified the following cell types from whole-section images: cancer cells (an average of 662,276 cells with standard deviation $\pm 489,852$), lymphocytes ($\pm 213,832$) and 123,081 stromal cells that included fibroblasts and endothelial cells ($\pm 121,035$). Individual cellular percentages were computed as a proportion of all cell types (Stromal%, Cancer% and Lymphocyte% henceforth).

The cellular classifications derived from image analyses were validated using single cell annotations on the H&E images, pathologist scores, imputed immune subtypes, tumor purity measure and pathway analysis. Using a set of 3230 single-cell annotations, the balanced accuracy of our classifier was found to be 84.9% (81.9% recall, 90.9% precision). Lymphocyte% was significantly higher in tumors with higher pathologist-measured LScore ($p < 0.001$, Fig. S1b) while Cancer% was significantly higher in tumors with increased tumor purity (as measured by CPE score, $p < 0.001$, Fig. S1c). Lymphocyte% was significantly higher in tumors with high imputed lymphocyte score (Fig. S1d, $p < 0.001$) and in imputed immune subtypes which were described as 'high-immune' subtype (24) ("Cluster 3", $p < 0.014$, Fig. S1e). Differential gene expression analysis revealed that samples with high lymphocyte ratio were enriched for immune-related pathways, including B Cell and T Cell Receptor Signalling (Fig. S2, Methods). Collectively, these orthogonal analyses supported the validity of automated H&E image analysis when applied to the TCGA melanoma dataset.

Constructing Topological Tumor Graphs for structural mapping of the Melanoma microenvironment

We quantitatively dissected the stromal architecture using graph-based algorithms. This enabled transformation of the three cell types (lymphocytes, stromal and cancer cells) and their spatial relationships into a new representation, namely TTGs (see Methods). In the TTGs, each cell (irrespective of the type) was represented as a node, while the spatial proximity between cells were represented as edges (Fig 1b). Since cancer cells were frequently found to be clustered, those with low edge length ('connected' cancer cells) were merged into 'cancer supernodes'. The TTG produced the following network centrality measures, which gauged the spatial relationship between cell types: i) Node degrees which quantified the number of edges connected

to a node ii) clustering coefficient which measured the degree of clustering among cells and iii) betweenness centrality which quantified the amount of control over communications among cells in a network. The mean node degree of stromal cells was significantly lower than that of cancer cells ($p < 2.2e-16$), indicating increased spatial interactions of cancer cells with neighbouring cells than did the stromal cells. Lymphocytes had a wider spread of node degree, indicating a mixture of connection patterns (Fig. 1c). Importantly, the clustering coefficient of stromal cells had higher variability compared to cancer cells and lymphocytes, which share similar distribution of clustering coefficient (Fig. 1d). In contrast, there was no significant difference in betweenness centrality among the cell types, indicating little difference in cellular position relative to the network centre (Fig. 1e). Thus, we further investigated the clustering coefficient of stromal cells ('stromal clustering' henceforth), from among the aforementioned network features. Additionally, in an effort to gauge lymphocytes' accessibility to cancer cells (a frequent histopathological observation in melanomas), we measured the 'stromal barrier', which is a measure of the barrier presented by stromal cells to cancer infiltrating lymphocytes (Fig. 2a-b, see Methods).

In comparing stromal features with clinicopathological features in the TCGA melanoma cohort (Table 1), stromal clustering and barrier did not vary significantly with AJCC stage or NRAS/BRAF mutation subtype, but stromal clustering was significantly inversely associated with Breslow thickness and ulceration status. In comparing with tumor type, stromal clustering was significantly lower in primary melanomas compared to metastases, with no significant difference in stromal barrier between primary and metastatic melanomas. Notably, stromal clustering and barrier were positively associated with stromal% and inversely associated with lymphocyte% respectively. The inverse correlation of lymphocyte% with stromal features persisted in a stratified analysis of tumor type (primary tumors, regional lymph node metastases and distant regional cutaneous metastases), suggesting that stromal features are inversely correlated with lymphocytes, irrespective of the tumor type.

Quantitative measures of tumor microenvironmental architecture are independent predictors of patient survival

Using univariate survival analyses, high stromal clustering (Fig.3a) and barrier (Fig.3b) were both associated with poor 10-year OS (Table 2) while stromal% was not significantly associated with survival (Fig. S5a). Multivariable survival analyses revealed that the poor prognosis associated with high stromal clustering and stromal barrier was independent of ulceration status and Breslow depth (Table 2). Notably, the prognostic significance of both stromal clustering and barrier were also independent of stromal% and lymphocyte% (Table 2). Even among tumors with high lymphocyte%, high stromal clustering and stromal barrier were associated with poor 10-year OS (Fig. S3a). In tumor site-specific analyses, high stromal clustering was associated with poor prognosis within regional lymph nodes and distant metastasis (Fig. S3b). High stromal barrier was associated with poor prognosis within regional lymph nodes but not in distant metastases (Fig. S3c).

We then interrogated the combination of stromal clustering and barrier by defining four combination groups: low-clustering/low-barrier, low-clustering/high-barrier, high-clustering/high-barrier and high-clustering/low-barrier (Fig. 3c). Tumors with high-clustering/high-barrier had significantly worse 10-year OS compared to low-clustering/low-barrier tumors (Fig. 3d, Table 2). Lymphocyte% was significantly lower in tumors with high-clustering/high-barrier compared to those with low-clustering/low-barrier (Fig 3e).

Taken together, histology-based measures of stromal architecture were significant predictors of melanoma prognosis, independent of the common clinicopathological predictors of OS as well as the content of stroma and lymphocytes.

Immunosuppressive potential of stromal barrier and clustering

Given that high stromal clustering and barrier were independently predictive of poor prognosis, we tested the hypothesis that stromal barrier and clustering could promote immunosuppression by impeding cancer-lymphocyte interactions. To this effect, we observed a significant negative correlation of lymphocyte-to-tumor area ratio with stromal barrier and clustering (Fig. 4a). Concordantly, stromal barrier was significantly higher in tumors belonging to the 'low-cytotoxicity' immune phenotype ("Cluster 1") compared to 'high-cytotoxicity' immune subtypes ("Cluster 3") (Fig. 4b), which also had the highest lymphocyte%. However, stromal clustering and barrier did not vary

significantly across the previously reported melanoma molecular subtypes (Fig. S4a). In order to delineate the immune contexture associated with stromal features, we compared 16 previously-reported (24) immune cell signatures between tumors with low or high stromal features. While T-regulatory cells (T-regs) and T-effector memory cells were significantly higher in tumors with low stromal barrier, CD8 T cells scores were significantly higher in tumors with low stromal clustering (Fig. 4c). Taken together, this suggested that increased stromal cell barrier and spatial clustering are associated with reduced distribution of lymphocytes over the entire tumor tissue.

To further explore the evolution of the stromal features over time, we used a small melanoma cohort with serial biopsies (Table S1). There was a significant decrease in stromal clustering but not in stromal barrier (Fig. S5a). Since stromal clustering was significantly higher in metastasis tissues compared to primary tumors (Fig. S5b) in the TCGA melanoma cohort, it is unlikely that this decrease can be attributed to the difference in the type of samples. Therefore, in this very limited, sequential cohort of melanoma patients, our preliminary analysis on melanoma evolution over time suggest that stromal clustering decreases with disease progression, possibly indicating a higher diversity in immune evasion strategies in progression compared with the initial stage of the disease.

Taken together, these data support the immunosuppressive potential and prognostic value of stromal organization, identified by the proposed stromal graph-based measures, to suppress anti-tumoral activity of lymphocytes.

Deciphering the molecular basis of stromal features from omics-based deep learning approaches

To identify biological processes underpinning histologically-defined stromal recruitment, matching copy-number and transcriptomic data were integrated and analysed using an unsupervised deep-learning framework (named CNx). This enabled the identification of gene expression changes driven by copy-number alterations in cancer cells, within the context of the microenvironmental measures described thus far. CNx consisted of an input layer (15667 nodes), an encoding layer (7000 nodes), a bottleneck layer (200 nodes), a decoding layer (7000 nodes) and an output layer (15667 nodes). To integrate the histology-based stromal architecture measures into

this model, we identified the nodes in the bottleneck layer that had the highest Spearman correlation between their activation and stromal clustering or barrier. These nodes and their relationship with other nodes were visualised as a network, along with the most weighted genes as nested nodes (Fig. 4d). Since this compressed layer captures copy-number-driven gene expression (both *cis* and *trans*), the list of genes contributing to the activation of nodes associated with stromal feature therefore constitute a catalogue of putative genes implicated in cancer-stromal interactions. To this effect, CNx identified 578 genes (Table S2) and 633 genes (Table S3) associated with stromal barrier and clustering respectively.

The genes identified by CNx had significantly higher correlation between copy number and gene expression data, compared to the genes identified by differential expression analysis as being correlated with stromal clustering and barrier (Fig. 4e).

Stroma-mediated down-regulation of lymphocyte cytotoxic program

The genes downregulated in tumors with high stromal barrier (268 genes at $FC < 0$, red-labelled, Table S2) and high stromal clustering (324 genes at $FC < 0$, red-labelled, Table S3) were enriched for pathways such as TCR signalling in naïve CD4 T cells, Wnt, Interleukin, PI3K-Akt and MAPK signalling (Table S4, S5). Conversely, genes upregulated in tumors with high stromal barrier (310 genes at $FC > 0$, green-labelled, Table S2) and clustering (309 genes at $FC > 0$, green-labelled, Table S3) were enriched for Mitochondrial translation, non-alcoholic fatty liver disease and processing of Capped intron-containing pre-mRNA (Table S6, S7).

We compared these CNx-derived findings with a lasso regression-based approach (detailed in supplementary materials), which identified copy number of 396 cytobands associated with gene expression events, analogous to the bottleneck nodes identified by CNx. These 'nodal cytobands' pertained to 692 and 723 constituent genes correlated significantly with stromal barrier and clustering respectively. In comparing the pathways enriched for these genes with those identified by CNx, we report that CNx identified genes enriched for immune-related pathways such as IL12-, IL2-, CXCR4- and TCR-mediated signalling, to be downregulated in tumors with high stromal clustering and barrier, as opposed to the lasso regression-based approach (Table S8, S9).

Since chemokines and cytokines direct cell trafficking and mediate cell recruitment in the tumor microenvironment (31), we investigated them and related

simulator/regulator in the CNx gene lists (Fig. 4f). There were only three genes shared between the cytokine/chemokine networks for stromal clustering and barrier: *GAB2*, *RASGRP1*, *XCL1*. Exclusive to stromal clustering were genes involved in JAK/STAT signalling (*JAK2*, *STAT2*, *STAT3*), chemokines and interleukin receptors (*CCL3*, *IL2RA*, *IL7R*, *IL15*), adhesion proteins (*VCAM1*, *ICAM1*), key hypoxia and differentiation markers (*HIF1A*, *TWIST1*, *FOXO1*, *JUN*). Among genes exclusive to the stromal barrier network were *CD80*, *CD38* and *TNFRSF1A*. The apparent lack of overlap between these two gene networks suggests functional mutual exclusivity, which was also consistent with the independent prognostic value of these two features.

Thus, CNx enabled the curation of a catalog of genomic alterations in cancer cells significantly over-represented in specific stromal phenotypes, thereby linking genes with microenvironmental context and providing a catalog of putative genomic features of stromal recruitment.

Discussion

In this study, we present Topological Tumor Graphs (TTGs): a new approach to study the spatial architecture of the tumor microenvironment in primary and metastatic melanoma. Although cell-cell network analysis has previously been used for tasks such as aiding nucleus identification (32), it has not been explicitly used to study tumor-host interactions. To construct TTGs, the melanoma tumor microenvironment learned from whole-tumor pathological section images were transformed into a network of cells using computational pathology. An understanding of the spatial interplay of heterogeneous cell types in the melanoma tumor microenvironment was derived network centrality measures such as node degree (the number of neighbours), clustering coefficient (a measure for the connectedness of a node), and betweenness (the number of shortest path in the network that cross this node). The node distribution of a melanoma TTG was similar to a random graph, suggesting that most cells have similar numbers of neighbours and there are not many hubs in the network. However, there were strong differences in node degree distribution between the cell types, particularly stromal cells having smaller node degrees than epithelial cells and more widespread clustering coefficients. The variable distribution patterns for different cell

types suggested that spatial interaction patterns among cancer, lymphocyte and stromal cells are dependent on cell identity.

The roles of stromal cells, in particular cancer-associated fibroblasts (CAFs), in regulating T cell activity and function are well known. As an important source of TGF- β , CAFs could promote cell death of effector CD8⁺ T cells by inhibiting expression of the pro-survival protein Bcl-2 (33), or alter cytotoxic CD8⁺T cell function by inhibiting the expression of perforin, granzymes A and B, Fas ligand, and IFN- γ (34,35). Alternatively, CAFs could restrict CD4⁺ and CD8⁺ T cells motility in dense matrix areas. Aligned fibres around tumor epithelial cell regions could restrict T cells from entering tumor islets (36). These studies and others, highlight the importance of stromal spatial positioning in the study of tumor immune response.

To this effect, we defined spatially explicit measurements based on TTGs to define two phenotypes of stromal recruitment, spatial clustering of stromal cells within the tumor, and stromal barrier which quantified the average number of stromal cells that a lymphocyte has to cross along a shortest path to infiltrate a cancer cluster. High degree of both stroma clustering and barrier significantly and independently correlated with decreased lymphocytic infiltration and poor survival in melanoma. These support the immunosuppressive potential of stromal cells in melanoma and their clinical relevance.

To identify intrinsic factors in cancer cells that drive stromal recruitment and enhance cell fitness, we investigated genomic alterations and gene expression in the context of stromal architecture. We applied a new deep learning approach, CNx, to identify copy number alterations that most influence the transcriptome and therefore best predict gene expression data even when reduced in dimension. There is a surge of interest in applying deep learning to cancer omics (16,17). Approaches which assess the association between gene expression and copy number changes, while being informative, do not always account for indirect effects of copy number change such as alterations in transcriptional factors. To this effect, we have adopted an unsupervised deep learning-based approach to directly integrate genome-wide copy number and tumor transcriptomics data. Furthermore, this study describes an integration of stromal phenotypes derived from whole-tumor section digital pathology, with omics data, in contrast to studies which focus on characterising microenvironmental phenotypes

using transcriptional immune/stromal signatures (24,37). Therefore, we propose that CNx could be a novel solution to infer independent, compressed representation of copy-number driven features from high-dimensional and highly inter-correlated genomic data in search of molecular features underpinning specific morphological phenotype.

Limitations of this study include the limit of access to more extensive, independent clinical cohorts, and the availability of data on response to immunotherapy, and the lack of specific marker staining for immune cell subsets. Nevertheless, we have been able to identify novel features of the genomic imprint of tumors that may potentially enable specific stromal features. New phenotypes of stroma-mediated immunosuppression identified using spatial histology analysis of stroma-cancer interface could help illuminate potential molecular machineries that allow tumors to engage with this interface. Furthermore, the concept of TTGs could easily be adapted for the analysis of other cell-cell interactions in the tumor microenvironment, e.g. tumor-lymphocyte interactions or lymphocyte-lymphocyte clusters.

Acknowledgements

The funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Data availability

Pathological images and clinicopathological information of the TCGA samples are available in a public repository from the TCGA Data Portal (<https://tcga-data.nci.nih.gov/tcga/>). All other data supporting the findings of this study are available as part of the reproducible Sweave package.

Code availability

CRIImage is available from Bioconductor. Code used to generate Topological Tumour Graphs and R code for performing relevant analyses are provided (<https://github.com/Henrik86/Topological-tumour-graphs>) for reproducibility.

References

1. Karlsson AK, Saleh SN. Checkpoint inhibitors for malignant melanoma: a systematic review and meta-analysis. *Clin Cosmet Investig Dermatol*. Dove Press; 2017;10:325–39.
2. Ali Z, Yousaf N, Larkin J. Melanoma epidemiology, biology and prognosis. *EJC Suppl EJC Off J EORTC, Eur Organ Res Treat Cancer* . [et al]. Elsevier; 2013;11:81–91.
3. Pasquali S, Hadjinicolaou A V, Chiarion Sileni V, Rossi CR, Mocellin S. Systemic treatments for metastatic cutaneous melanoma. *Cochrane Database Syst Rev*. John Wiley & Sons, Ltd; 2018;
4. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016;
5. Rodríguez-Cerdeira C, Carnero Gregorio M, López-Barcenas A, Sánchez-Blanco E, Sánchez-Blanco B, Fabbrocini G, et al. Advances in Immunotherapy for Melanoma: A Comprehensive Review. *Mediators Inflamm*. 2017;2017:1–14.
6. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*. NIH Public Access; 2012;12:252–64.
7. Beatty GL, Gladney WL. Immune escape mechanisms as a guide for cancer immunotherapy. *Clin Cancer Res*. NIH Public Access; 2015;21:687–92.
8. Ziani L, Safta-Saadoun T Ben, Gourbeix J, Cavalcanti A, Robert C, Favre G, et al. Melanoma-associated fibroblasts decrease tumor cell susceptibility to NK cell-mediated killing through matrix-metalloproteinases secretion. *Oncotarget*. Impact Journals, LLC; 2017;8:19780–94.
9. Gajewski TF, Schreiber H, Fu Y-X. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol*. NIH Public Access; 2013;14:1014–22.
10. Lu P, Weaver VM, Werb Z. The extracellular matrix: A dynamic niche in cancer progression. *J Cell Biol*. Rockefeller University Press; 2012;196:395–406.
11. Sangaletti S, Chiodoni C, Tripodo C, Colombo MP. The good and bad of targeting cancer-associated extracellular matrix. *Curr Opin Pharmacol*. 2017;35:75–82.
12. Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. *Science (80-)*. 2015;348:74–80.
13. Sorokin L. The impact of the extracellular matrix on inflammation. *Nat Rev Immunol*. 2010;10:712–23.
14. Verhaak RGW, Tamayo P, Yang JY, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J Clin Invest*. 2013;123:517–25.
15. Yuan Y, Failmezger H, Rueda OM, Ali HR, Graf S, Chin SF, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med*. 2012/10/27. 2012;4:157ra143.

16. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning–based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res*. 2018;
17. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A*. 2018;
18. Chakravarthy A, Furness A, Joshi K, Ghorani E, Ford K, Ward MJ, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun* [Internet]. Nature Publishing Group; 2018 [cited 2018 Sep 19];9:3220. Available from: <http://www.nature.com/articles/s41467-018-05570-1>
19. Beach JR, Hussey GS, Miller TE, Chaudhury A, Patel P, Monslow J, et al. Myosin II isoform switching mediates invasiveness after TGF- β -induced epithelial-mesenchymal transition. *Proc Natl Acad Sci U S A* [Internet]. National Academy of Sciences; 2011 [cited 2018 Sep 26];108:17991–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22025714>
20. Heindl A, Khan AM, Rodrigues DN, Eason K, Sadanandam A, Orbegoso C, et al. Microenvironmental niche divergence shapes BRCA1-dysregulated ovarian cancer morphological plasticity. *Nat Commun*. 2018;
21. Akbani R, Akdemir KC, Aksoy BA, Albert M, Ally A, Amin SB, et al. Genomic Classification of Cutaneous Melanoma. *Cell*. 2015;161:1681–96.
22. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. Nature Publishing Group; 2015;6:8971.
23. Li B, Severson E, Pignon J-C, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. BioMed Central; 2016;17:174.
24. Tamborero D, Rubio-Perez C, Mui~ Nos F, Sabarinathan R, Piulats JM, Muntasell A, et al. Biology of Human Tumors A Pan-cancer Landscape of Interactions between Solid Tumors and Infiltrating Immune Cell Populations. 2018;
25. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. *TCGAbiolinks*: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. Oxford University Press; 2016;44:e71–e71.
26. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning–Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer. *Clin Cancer Res*. 2017;
27. Danaee P, Ghaeini R, Hendrix DA. A DEEP LEARNING APPROACH FOR CANCER DETECTION AND RELEVANT GENE IDENTIFICATION. *Pac Symp Biocomput*. NIH Public Access; 2017;22:219–29.
28. Tan J, Ung M, Cheng C, Greene CS. Unsupervised feature construction and knowledge extraction from genome-wide assays of breast cancer with denoising autoencoders. *Pac Symp Biocomput*. NIH Public Access; 2015;20:132–43.
29. Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBImage--an R package for

- image processing with applications to cellular phenotypes. *Bioinformatics*. 2010/03/27. 2010;26:979–81.
30. Schult DA, Schult DA. Exploring network structure, dynamics, and function using NetworkX. *Proc 7TH PYTHON Sci Conf (SCIPY)*. 2008;11--15.
 31. Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat Rev Immunol*. NIH Public Access; 2017;17:559–72.
 32. Wang X, Janowczyk A, Zhou Y, Thawani R, Fu P, Schalper K, et al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H&E images. *Sci Rep*. Nature Publishing Group; 2017;7:13543.
 33. Sanjabi S, Mosaheb MM, Flavell RA. Opposing Effects of TGF- β and IL-15 Cytokines Control the Number of Short-Lived Effector CD8+ T Cells. *Immunity*. 2009;31:131–44.
 34. Thomas DA, Massagué J. TGF- β directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell*. 2005;8:369–80.
 35. Ahmadzadeh M, Rosenberg SA. TGF-beta 1 attenuates the acquisition and expression of effector function by tumor antigen-specific human memory CD8 T cells. *J Immunol*. 2005;174:5215–23.
 36. Salmon H, Franciszkiwicz K, Damotte D, Dieu-Nosjean M-C, Validire P, Trautmann A, et al. Matrix architecture defines the preferential localization and migration of T cells into the stroma of human lung tumors. *J Clin Invest*. 2012;122:899–910.
 37. Chakravarthy A, Furness A, Joshi K, Ghorani E, Ford K, Ward MJ, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun*. Nature Publishing Group; 2018;9:3220.

Table 1: Association of stromal features with clinicopathological measures in the TCGA melanoma cohort

Association of stromal phenotype with	Stromal clustering	Stromal barrier
AJCC stage	p=0.78	p=0.98
NRAS/BRAF status	p=0.64	p=0.88
Breslow thickness	R= -0.167, p=0.003	R= 0.01, p=0.80
Ulceration	p=0.02	p= 0.23
Tumour type (primary vs metastases)	p=0.0037	p=0.237
Stromal%	R=0.33, p=1.7e-11	R=0.58, 2.2e-16
Lymphocyte%		
- Whole data set	R=-0.31, p=2.2e-10	R=-0.18, p=0.00018
- Primary tumours only	R=-0.32, p=0.0012	R=-0.31, p=0.0016
- Distant regional cutaneous metastases	R=-0.51, p=4.6e-09	R=-0.25, p=0.0058
- Regional lymph node metastases	R=-0.34, p=1.2e-06	R=-0.10, p=0.15

Table 2: Association of stromal features with patient survival using univariate and multivariate analyses. Low and high stromal clustering are denoted by optimal cut-off (stromal clustering ≤ 0.590253), low and high stromal barrier are denoted by the first and fourth quartiles

	P-value	HR	HR 95% CI
Stromal clustering			
Univariate (low vs high)	0.002	1.97	1.26-3.07
Adjusted for Breslow depth and ulceration	0.003	2.50	1.35-4.62
Adjusted for Lymphocyte%	0.006	1.99	1.21-3.27
Adjusted for Stromal %	0.005	1.93	1.22-3.06
Stromal barrier			
Univariate (low vs high)	0.017	1.92	1.12-3.29
Adjusted for Breslow depth and ulceration	0.017	1.78	1.10-2.86
Adjusted for Lymphocyte%	0.03	1.82	1.05-3.166
Adjusted for Stromal%	0.04	1.87	1.02-3.43
Mixed model: Stromal barrier + Stromal clustering			
high clustering/high barrier vs			
high clustering/low barrier	0.62	0.158	0.31-1.20
low clustering/high barrier	0.53	0.308	0.16-1.77
low clustering/low barrier	0.35	0.007	0.16-0.76

Figure legends

Figure 1. Computational pipeline schema. (a) Schema of automated computational pathology pipeline to analyse H&E slides classified at single-cell resolution, and converted to topological graphs (TTGs). TTGs-derived tumor phenotypes were integrated with omic data using a convolutional neural network with an hourglass shaped architecture. This network was designed to learn a sparse representation of CNAs that drive gene expression changes. (b) Illustration of a part of a TTG from a whole-tumor section, with cancer cluster summarized into a supernode. (c-e) Cell type specific distribution of node degree, clustering coefficient and betweenness, P-value from t-test.

Figure 2. Conceptual illustration of stromal barrier for lymphocytic infiltration. (a) Example of the stromal barrier in a region of an H&E slide, where a fibrous stroma can be seen between a lymphoid aggregate and a tumor islet. (b) The corresponding TTG, in which cancer cell nodes have been summarized to a cancer supernode, whereas, each of the lymphocyte/stromal node represent a single lymphocyte/stromal cell. The number of stromal nodes on the shortest path from a lymphocyte to the nearest cancer supernode was calculated and later averaged to derive the quantitative measure of stromal barrier.

Figure 3. Stromal barrier and clustering are associated with patient survival. Effect of Stromal barrier (a) and stromal clustering (b) on 10-year OS (Overall Survival), P-values from univariate Cox proportional hazards model. (c) Scatter plot of stromal barrier versus clustering, with colours corresponding to the four combination groups. Comparison of (d) 10-year OS and (e) lymphocyte% between 4 combination groups. P-values in (e) are from Tukey multiple comparisons of means test

Figure 4. Immunosuppressive potential of stromal barrier and clustering. (a) Boxplot showing the relationship between the ratio of lymphocyte to tissue area and stromal barrier, Right panel: the same for stromal clustering. (b) difference in stromal barrier between clusters derived from previously described immune phenotypes (24): Cluster 1 ('Low-cytotoxicity'), Cluster 2 ('intermediate-cytotoxicity'), Cluster 3 ('High-cytotoxicity') (c) Visualisation of the bottleneck layer of CNx. Nodes in the bottleneck/compressed layer were connected if they shared similar activity values (Methods) and visualised as a network. The nodes with highest correlation with different image phenotypic features were coloured. Genes with highest activation for these nodes were shown as nested network. (d) Correlation between CNA and gene expression of genes identified by CNx and by Differential Gene expression analysis. (e) Overrepresented pathways for genes identified by Differential Gene Expression and by CNx. (f) Cytokine genes from each gene list illustrated as protein-protein interaction networks (String interaction score >0.4).

Figure 1
 a

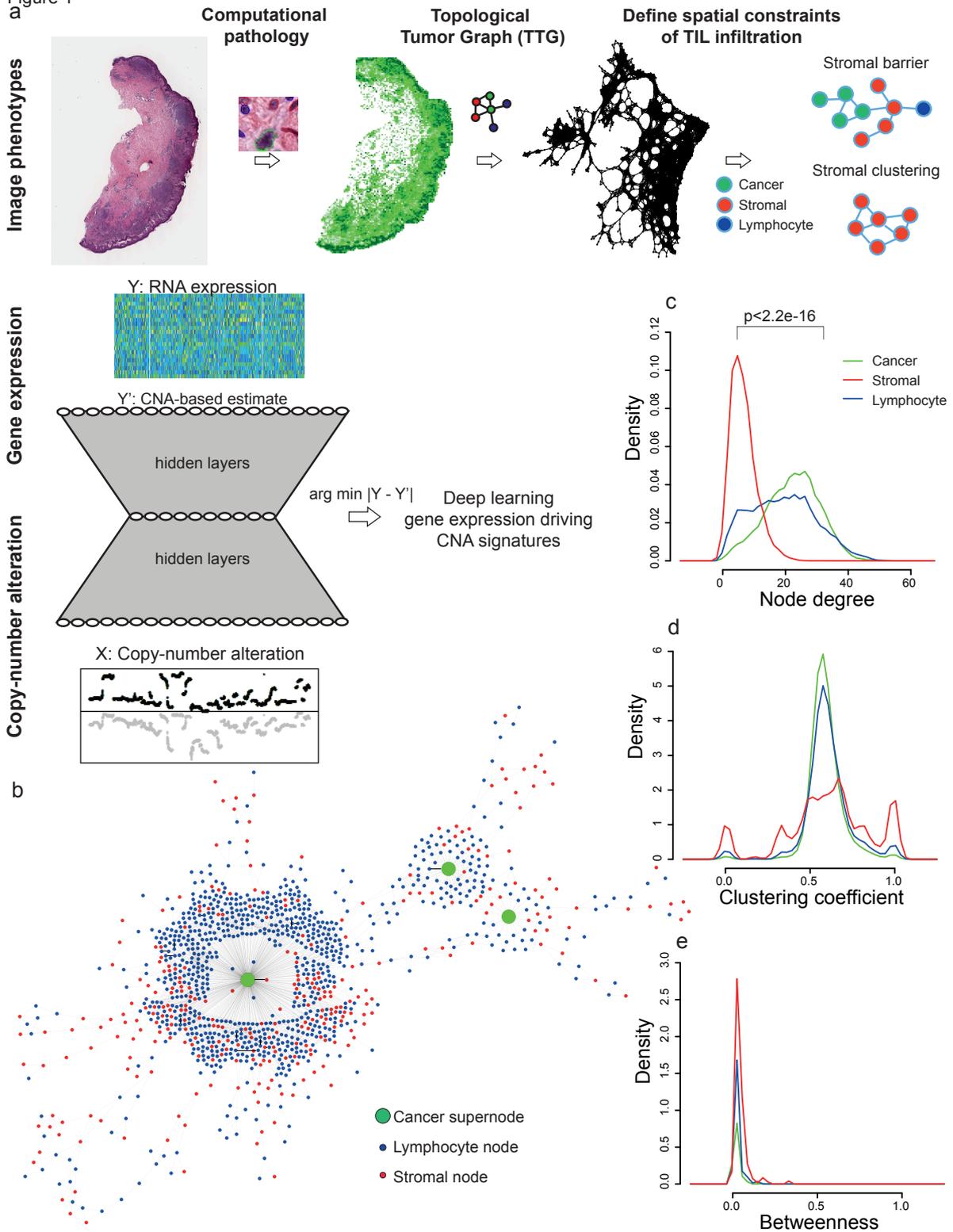
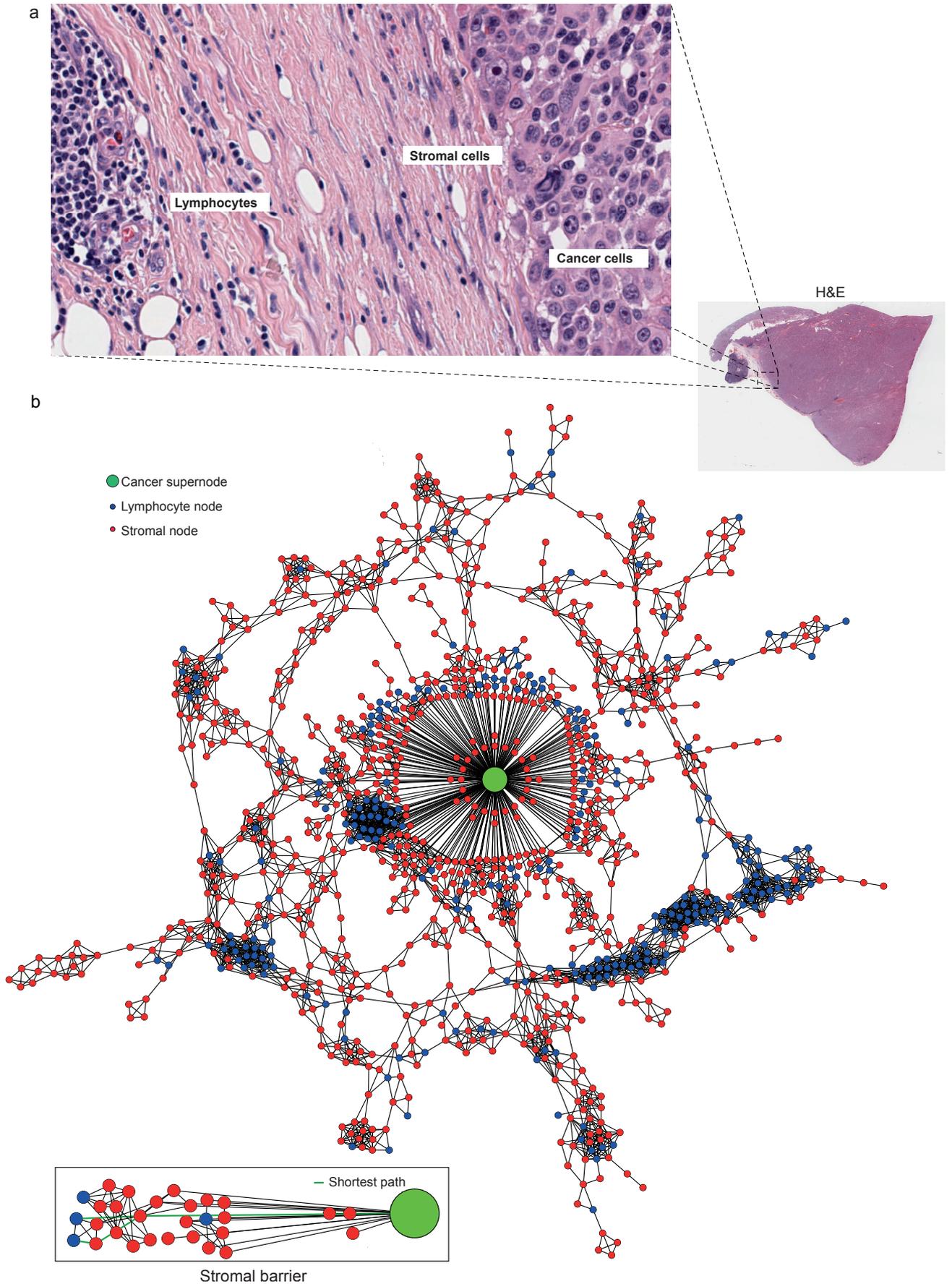


Figure 2



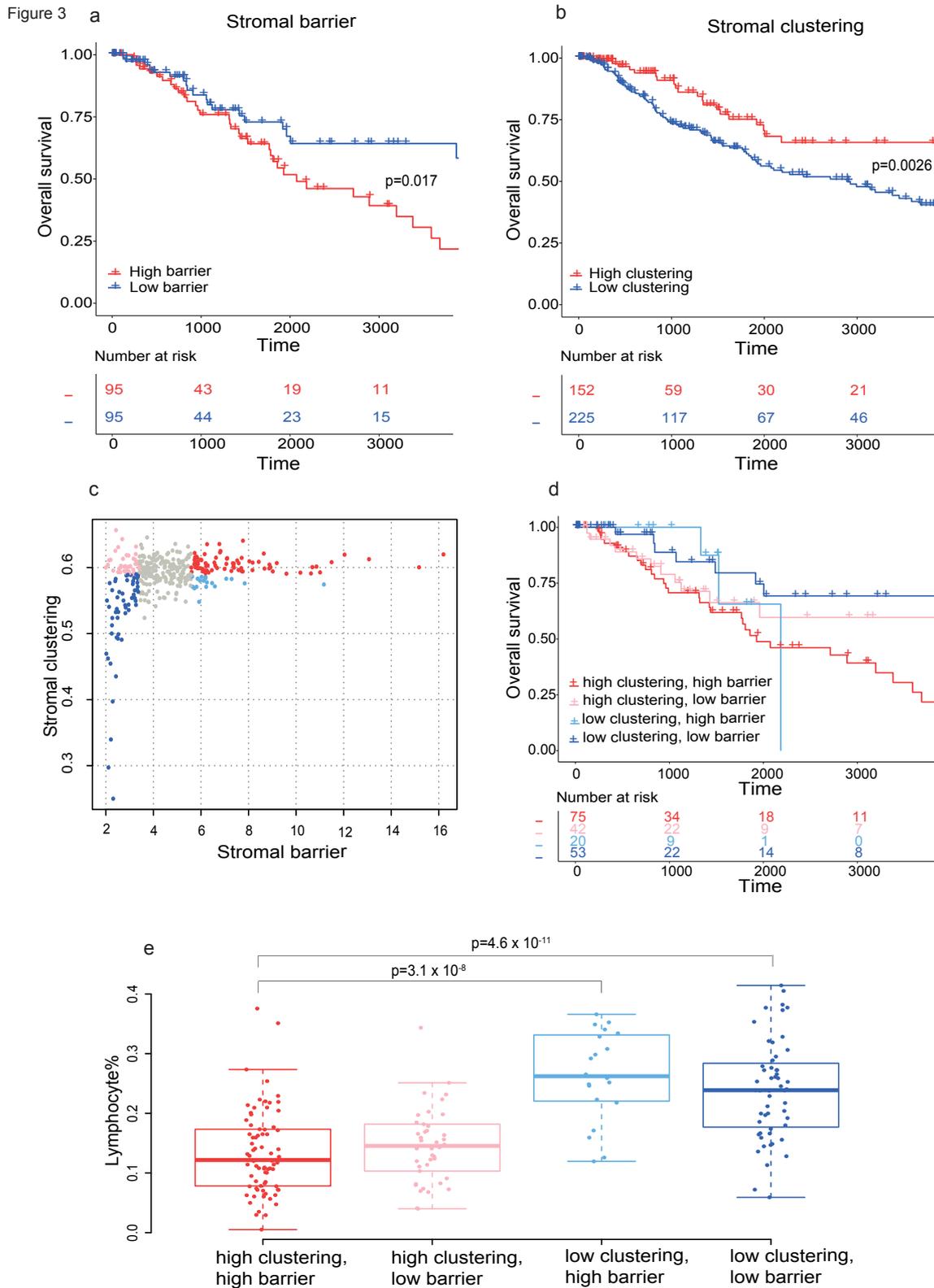
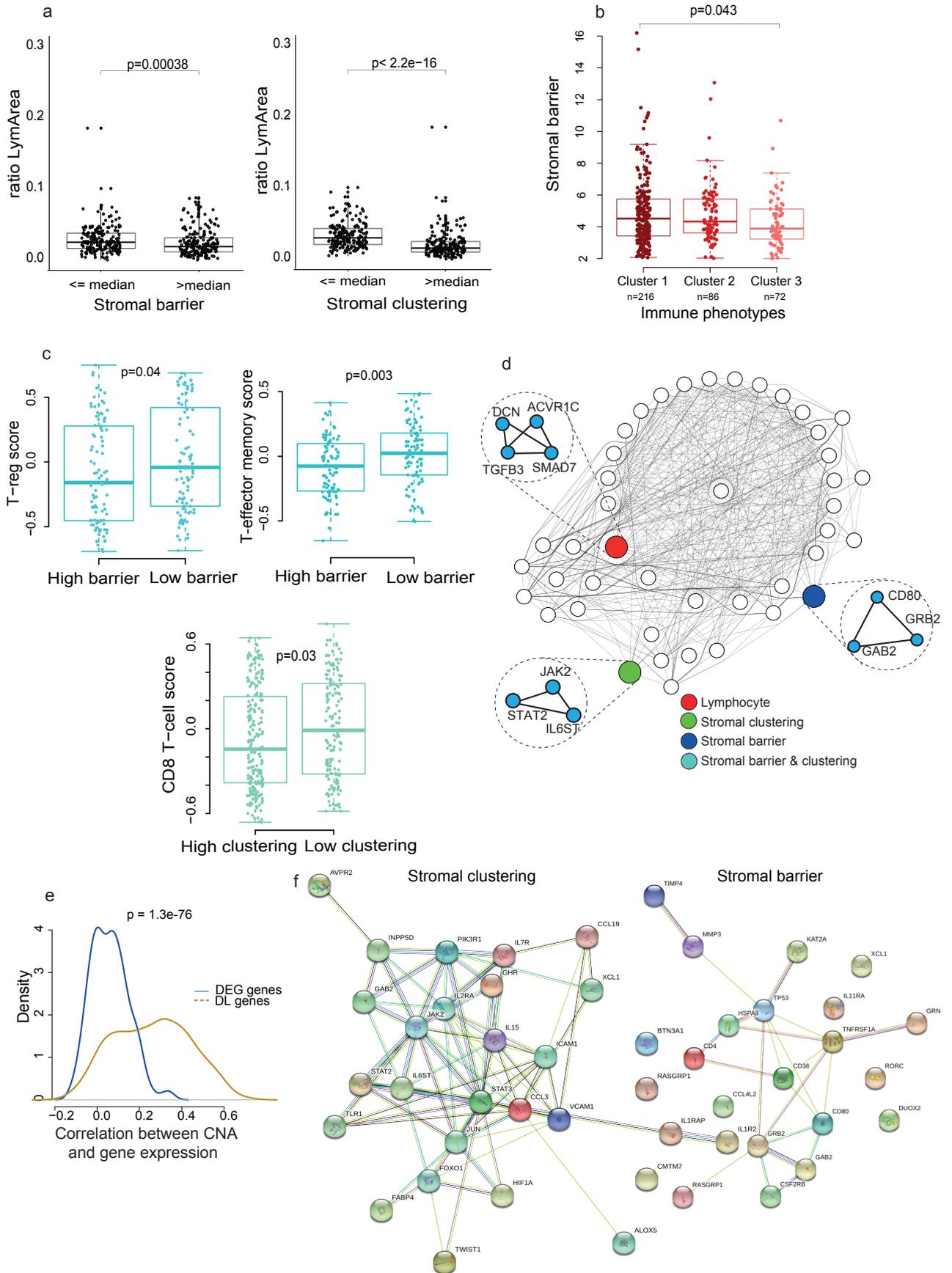


Figure 4



Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Topological Tumor Graphs: a graph-based spatial model to infer stromal recruitment for immunosuppression in melanoma histology

Henrik Failmezger, Sathya Muralidhar, Antonio Rullan, et al.

Cancer Res Published OnlineFirst December 24, 2019.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-19-2268
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2019/12/24/0008-5472.CAN-19-2268.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/early/2019/12/24/0008-5472.CAN-19-2268 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.