Triangular Analysis of Geographical Interplay of Lymphocytes (TriAnGIL): Predicting Immunotherapy Response in Lung Cancer

Sara Arabyarmohammadi[1,2], German Corredor[1], Yufei Zhou[2], Miguel López de Rodas[3], Kurt Schalper[3], and Anant Madabhushi[1,4(✉)]

[1] Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, USA
anantm@emory.edu
[2] Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, USA
[3] Department of Pathology, School of Medicine, Yale University, New Haven, USA
[4] Atlanta Veterans Administration Medical Center, Atlanta, USA

Abstract. Quantitative immunofluorescence (QIF) enables identifying immune cell subtypes across histopathology images. There is substantial evidence to show that spatial architecture of immune cell populations (e.g. CD4+, CD8+, CD20+) is associated with therapy response in cancers, yet there is a paucity of approaches to quantify spatial statistics of interplay across immune subtypes. Previously, analyzing spatial cell interplay have been limited to either building subgraphs on individual cell types before feature extraction or capturing the interaction between two cell types. However, looking at the spatial interplay between more than two cell types reveals complex interactions and co-dependencies that might have implications in predicting response to therapies like immunotherapy. In this work we present, Triangular Analysis of Geographical Interplay of Lymphocytes (TriAnGIL), a novel approach involving building of heterogeneous subgraphs to precisely capture the spatial interplay between multiple cell families. Primarily, TriAnGIL focuses on triadic closures, and uses metrics to quantify triads instead of two-by-two relations and therefore considers both inter- and intra-family relationships between cells. The TriaAnGIL's efficacy for microenvironment characterization from QIF images is demonstrated in problems of predicting (1) response to immunotherapy (N = 122) and (2) overall survival (N = 135) in patients with lung cancer in comparison with four hand-crafted approaches namely DenTIL, GG, CCG, SpaTIL, and deep learning with GNN. For both tasks, TriaAnGIL outperformed hand-crafted approaches, and GNN with AUC = .70, C-index = .64. In terms of interpretability, TriAnGIL easily beats GNN, by pulling biological insights from immune cells interplay and shedding light on the triadic interaction of CD4+-Tumor-stromal cells.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43987-2_77.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14225, pp. 797–807, 2023.
https://doi.org/10.1007/978-3-031-43987-2_77

1 Introduction

The tumor microenvironment (TME) is comprised of cancer, immune (e.g. B lymphocytes, and T lymphocytes), stromal, and other cells together with non-cellular tissue components [3,5,24,30]. It is well acknowledged that tumors evolve in close interaction with their microenvironment. Quantitatively characterizing TME has the potential to predict tumor aggressiveness and treatment response [3,23,24,30]. Different types of lymphocytes such as CD4+ (helper T cells), CD8+ (cytotoxic T cells), CD20+ (B cells), within the TME naturally interact with tumor and stromal cells. Studies [5,9] have shown that quantifying spatial interplay of these different cell families within the TME can provide more prognostic/predictive value compared to only measuring the density of a single biomarker such as tumor-infiltrating lymphocytes (TILs) [3,24]. Immunotherapy (IO) is the standard treatment for patients with advanced non-small cell lung cancer (NSCLC) [19] but only 27–45% of patients respond to this treatment [21]. Therefore, better algorithms and improved biomarkers are essential for identifying which cancer patients are most likely to respond to IO in advance of treatment. Quantitative features that relate to the complex spatial interplay between different types of B- and T-cells in the TME might unlock attributes that are associated with IO response. In this study, we introduce a novel approach called Triangular Analysis of Geographical Interplay of Lymphocytes (TriAnGIL), representing a unique and interpretable way to characterize the distribution, and higher-order interaction of various cell families (e.g., cancerous cells, stromal cells, lymphocyte subtypes) across digital histopathology slides. We demonstrate the efficacy of TriaAnGIL for characterizing TME in the context of predicting 1) response to IO with immune checkpoint inhibitors (ICI), 2) overall survival (OS), in patients with NSCLC, and 3) providing novel insights into the spatial interplay between different immune cell subtype. TriAnGIL source code is publicly available at http://github.com/sarayar/TriAnGIL.

2 Previous Related Work and Novel Contributions

Many studies have only looked at the density of a single biomarker (e.g. TILs), to show that a high density of TILs is associated with improved patient survival and treatment response in NSCLC [3,24]. Other works have attempted to characterize the spatial arrangement of cells in TME using computational graph-based approaches. These approaches include methods that connect cells regardless of their type (1) using global graphs (GG) such as Voronoi that connect all nuclei [2,14], or (2) using Cell cluster graphs (CCG) [16] to create multiple nuclear subgraphs based on cell-to-cell proximity to predict tumor aggressiveness and patient outcome [16]. Others have explored (3) the spatial interplay between two different cell types [5].One example approach is Spatial architecture of TIL (SpaTIL) [9] which attempted to characterize the interplay between immune and cancer cells and has proven to be helpful in predicting the recurrence in early-stage NSCLC. All of these approaches point to overwhelming evidence that spatial architecture of cells in TME is critical in predicting cancer

outcome. However, these approaches have not been able to exploit higher-order interactions and dependencies between multiple cell types (> 2), relationships that might provide additional actionable insights. The contributions of this work include:

(1) TriAnGIL is a computational framework that characterizes the architecture and relationships of different cell types simultaneously. Instead of measuring only simple two-by-two relations between cells, it seeks to identify triadic spatial relations (hyperedges [18,20] instead of edges) between different cell types, thereby enabling the exploration of complex nuclear interactions of different cell types within TME. This in turn allows for development of machine classifiers to predict outcome and response in lung cancer patients treated with IO.
(2) TriAnGIL includes a set of quantitative metrics that capture the interplay within and between nuclei corresponding to different types/families. Previous works have focused primarily on intra-family relationships [2,5,9,14] while TriAnGIL measurements are able to consider inter- and intra-family relationships.
(3) Although deep learning (DL) models (e.g., graph neural networks(GNN)) have shown great capabilities in solving complex problems in the biomedical field, these tend to be black-box in nature. A key consideration in cancer immunology is the need for actionable insights into the spatial relationships between different types of immune cells. Not only does TriAnGIL provide predictions that are on par or superior compared to DL approaches, but also provides a way to glean insights into the spatial interplay of different immune cell types. These complex interactions enhance our understanding of the TME and will help pave the way for new therapeutic strategies that leverage these insights.

3 Description of TriAnGIL Methodology

3.1 Notation

Our approach consists of constructing heterogeneous graphs step by step and quantifying them by extracting features from them. The graphs are defined by G = (V, E), where V is the set of vertices (nodes) $V = \{v_1, ...v_N\}$ with τ_n vertex types, and E is the collection of pairs of vertices from V, $E = \{e_1, ...e_M\}$, which are called edges and ϕ_n is the mapping function that maps every vertex to one of n differential marker expressions in this dataset $\phi_n : V \rightarrow \tau_n$. G is represented by an adjacency matrix A that allows one to determine edges in constant time.

3.2 Node Linking and Computation of Graph-Interplay Features

The inputs of TriAnGIL are the coordinates of nuclear centroids and the corresponding cell types. In the TriAnGIL procedure, the centroid of each nucleus in a family is represented as a node of a graph. TriAnGIL is agnostic of the method used for identifying the coordinates and types. Once the different cell families are identified (Fig. 1-B), a list is generated for all possible sets comprising of

membership from three [12,18] different cell families. By focusing on every set, TriAnGIL allows for capturing higher-order and balanced spatial triadic relationships [4] between cell families, while keeping the computational complexity relatively low. Therefore, we initiate the process with the first set on the list (α, β, γ), build heterogeneous graphs, extract features, and then select another set until we have explored all possible triads of cell families on the list. The three main steps of TriAnGIL is as follows:

1) **Quantifying in absence of one family:** First, we build a proximity graph (G_1) on nodes of α, β, γ based on the Euclidean distance of every two nodes. Two nodes will be connected if their distance is shorter than a given "interaction rate", θ regardless of their family and cell type (Fig. 1-C, 1-C1). The interaction rate is a hyper-parameter that controls how close we expect the distance to be so that we consider some interaction between the nodes.

$$\{G_1 = (V, E) \mid \forall e \in E : |e| <= \theta\} \tag{1}$$

We then exclude the nodes of α from all the interactions by removing its edges from G1 and characterize the relationship of β and γ (Fig. 1-C2). Next, we extract a series of features including clustering coefficient, average degree from the resulting subgraph. We repeat this process by removing all the edges of β (Fig. 1-C3) and then γ (Fig. 1-C4). In this manner, a total of 126 features (supplemental Table 1) are extracted (42 features for absence of one family $\times 3$).

2) **Triangulation-based connections:** A Delaunay triangulation is constructed by the nodes of α, β, γ (Fig. 1-D, and 1-D1). Delaunay triangulation is a planar graph formed by connecting the vertices in a way that ensures no point lies within the circumcircle of any triangle formed by the vertices [7]. We then extract 10 features relating to edge length and vertex count. Next, we prune long edges (D_1) where the Euclidean distance between connected nodes is more than the "interaction rate" (Fig. 1-D2). Next, a series of features were extracted from the remaining subgraph (e.g. number of edges between the nodes of α and β, β and γ, α and γ; complete list of features in supplemental Table 1).

$$\{D_1 = (V, E) \mid \forall e \in E : |e| <= \theta\} \tag{2}$$

3) **Triangular interactions:** As illustrated in Algorithm 1, from the unpruned Delaunay triangulation that includes the nodes of α, β, γ, we select those triangles (closed triads [8,25]) that link nodes from three distinct families (Fig. 1-E, 1-E1). In other words, we remove triangles with more than one vertex from a single family. Next, we call GetTriangleFeatures() function to quantify triangular relationships by extracting features from the resulting subgraphs (e.g. perimeter and area of triangles; complete list of features in supplemental Table 1).

4 Experimental Results and Discussion

4.1 Dataset

The cohort employed in this study was composed of pre-treatment tumor biopsy specimens from patients with NSCLC from five centers (two centers for training

Algorithm 1: Finding Triangles

INPUT: A jagged array Del : Delaunay graph with three vertices of every triangle in each row, A hashmap ϕ : maps nodes to their type
OUTPUT: Triangle features $TriFeatSet$
let $TriIndex \leftarrow \emptyset$ be the list for triangle indices
for $i = 1$ *to* $i = length(Del)$ **do**
  let $Marker \leftarrow \emptyset$ be a auxiliary list to keep the viewed markers
  for $j = 1$ *to* 3 **do**
    if $\phi(Del(i,j)) \notin Marker$ **then**
      $Marker \leftarrow Marker \bigcup \phi(Del(i,j))$
    end
  end
  if $length(Marker) == 3$ **then**
    $TriIndex \leftarrow TriIndex \bigcup i$
  end
end
$TriFeatSet \leftarrow$`GetTriangleFeatures`$(Del(TriIndex, :))$
return $TriFeatSet$

(S_t) and three centers for independent validation (S_v)). The entire analysis was carried out using 122 patients in Experiment 1 (73 in S_t, and 49 in S_v) and 135 patients in Experiment 2 (81 in S_t, and 54 in S_v). Specimens were analyzed with a multiplexed quantitative immunofluorescence (QIF) panel using the method described in [22]. From each whole slide image, 7 representative tiles were obtained and used to train the software InForm to define background, tumor and stromal compartments. Then, individual cells were segmented based on nuclear DAPI staining and the segmentation performance was controlled by direct visualization of samples by a trained observer. Next, the software was trained to identify cell subtypes based on marker expression (CD8, CD4, CD20, CK for tumor epithelial cells and absence of these markers for stromal cells).

4.2 Comparative Approaches

The efficacy of TriAnGIL was compared against five different approaches.

TIL density (DenTIL): For every patient, multiple density measures including the number of different cells types and their ratios are calculated [3,24] (supplemental Table 2).

GG: A Delaunay triangulation, a Voronoi diagram, and a Minimum Spanning Tree were constructed [2,14] on all nuclei regardless of their type. Architectural features (e.g., perimeter, triangle area, edge length) were then calculated on these global graphs for each patient.

CCG: For every patient, subgraphs are built on nuclei regardless of their type and only based on their Euclidean distance. Local graph metrics (e.g. clustering coefficient) [16] are then calculated from these subgraphs.

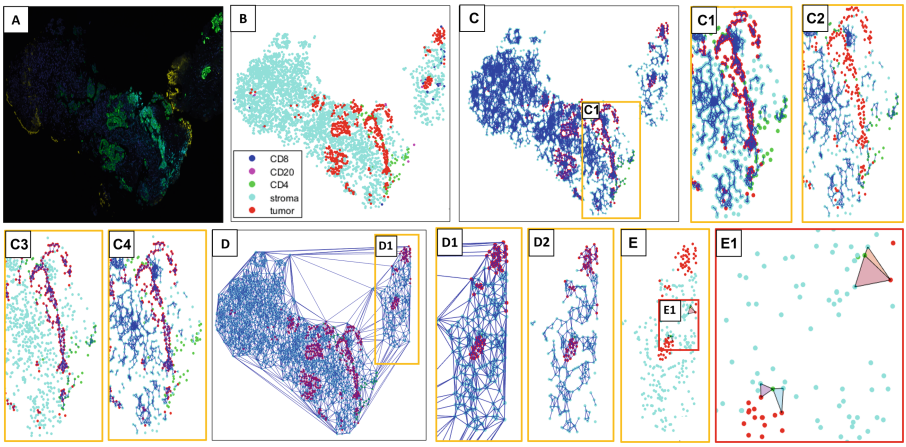

Fig. 1. Illustration of workflow for TriAnGIL. (A) Representative QIF image. (B) Nuclei subtypes based on differential marker expression. (C) Proximity graph on all CD4+-tumor-stroma nuclei. (C1) A zoomed-in region showing the edges based on Euclidean distances. (C2) The first family (tumor cells) is removed to enable characterization of other two nuclear types. (C3) Stroma cells are removed. (C4) CD4+ cells are removed. (D) A Delaunay triangulation on all CD4+, tumor, and stroma nuclei. (D1) Shows a zoomed-in region. (D2) Shows the same region in the pruned Delaunay subgraph. (E) From all Delaunay triangles, only those that connect CD4+-tumor-stroma are identified. (E1) A zoomed-in region showing the CD4+-tumor-stroma triangles.

SpaTIL: For each patient, first, subgraphs are built on individual cell types based on a distance parameter. The convex hulls are then constructed on these subgraphs. After selecting every two cell types, features are extracted from their convex hulls (e.g. the number of clusters of each cell type, area intersected between clusters [9]; complete list of combinations in supplemental Table 3).

GNN: A recent study [31] demonstrated that Transformer-based [29] GNNs are able to learn the arrangement of tiles across pathology images for survival analysis. Here, for each tile in the slide, a Delaunay graph was constructed regardless of cell subtypes, and tile-level feature representations (e.g.side length minimum, maximum, mean, and standard deviation, triangle area minimum, maximum, mean, and standard deviation) were aggregated by a Transformer according to their spatial arrangement [31]. Our approach utilized the Weisfeiler-Lehman (WL) test [15] for embedding graphs into Euclidean feature space. Well-known approaches, such as GraphSage [10], are considered as continuous approximations to the WL test. Therefore, our GNN is a valid baseline for heterogeneous graphs.

4.3 Experiment 1: Immunotherapy Response Prediction in Lung Cancer

Design: TriAnGIL was also trained to differentiate between patients who responded to IO and those who did not. For our study, the responders to IO were identified as those patients with complete response, partial response, and

stable disease, and non-responders were patients with progressive disease. A linear discriminant analysis (LDA) classifier was trained on S_t to predict which patients would respond to IO. For creating the model, the minimum redundancy maximum relevance (mRMR) method [1] was used to select the top features. The same procedure using mRMR and LDA was performed for the comparative hand-crafted approaches. The ability to identify responders post-IO was assessed by the area under the receiver operating characteristic curve (AUC) in S_v.

Results: The two top predictive TriAnGIL features were found to be the number of edges between stroma and CD4+ cells, and the number of edges between stroma and tumor cells with more interactions between stromal cells and both CD4+ and tumor cells being associated with response to IO. This finding is concordant with other studies [13,17,22,27] that stromal TILs were significantly associated with improved OS. Therefore, TriAnGIL approach is not only predictive of treatment response but more critically it enables biological interpretations that a DL model might not be able to provide. In S_v, this LDA classifier was able to distinguish responders from non-responders to IO with AUC_{Tri}=0.70 that was higher than all other hand-crafted and DL approaches with $AUC_{Den} = 0.45$, $AUC_{GG} = 0.52$, $AUC_{Spa} = 0.53$, $AUC_{CCG} = 0.65$, $AUC_{GNN} = 0.65$.

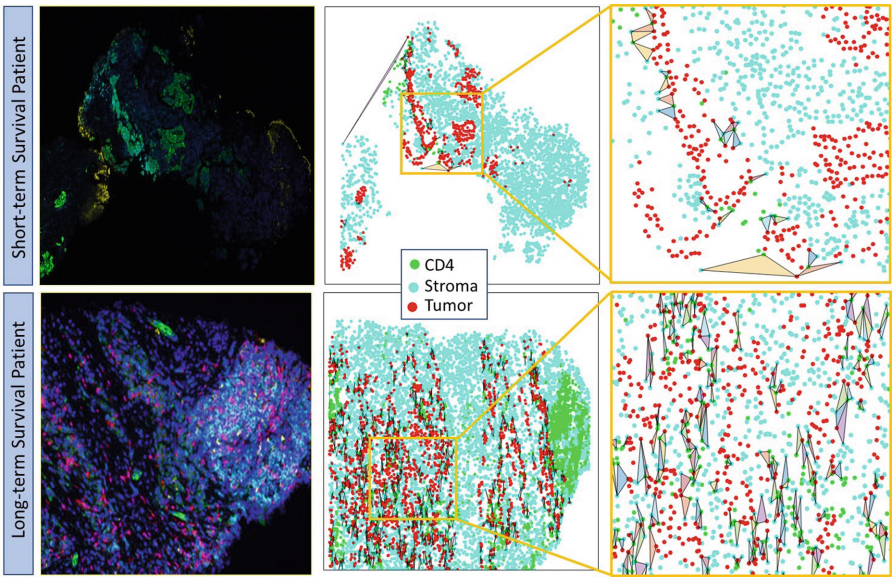

Fig. 2. Visual illustration of the qualitative difference in feature representations between NSCLC patients with short-term (top) and long-term survival (lower row). The leftmost column shows a part of a QIF image. The second column shows the triangular relationships formed by CD4, tumor, and stromal cell families. The third column shows a zoomed-in region. The number, size, and interplay of triangles formed on the three cell types are clearly different between the two cases. (Color figure online)

4.4 Experiment 2: Predicting Survival in Lung Cancer Patients Treated with Immunotherapy

Design: S_t was used to construct a least absolute shrinkage and selection operator (LASSO) [28] regularized Cox proportional hazards model [6] using the TriAnGIL features, to obtain risk score for each patient. LASSO features are listed in supplemental Table 4. The median risk score in S_t was used as a threshold in both S_t and S_v to dichotomize patients into low-risk/high-risk categories. Kaplan-Meier (KM) survival curves [26] were plotted and the model performance was summarized by hazard ratio (HR), with corresponding (95% confidence intervals (CI)) using the log-rank test, and Harrell's concordance index (C-index) on S_v. The C-index evaluates the correlation between risk predictions and survival times, aiming to maximize the discrimination between high-risk and low-risk patients [11]. OS is the time between the initiation of IO to the death of the patient. The patients were censored if the date of death was unknown.

Result: Figure 2 presents some TriAnGIL features in a field of view for a patient with long-term survival and another with short-term survival. More triangular relationships, shorter triangle edges, and smaller triangles with smaller perimeters are found in the long-term survival case when analyzing the triadic interactions within tumor-stroma-CD4, thereby suggesting higher relative presence and closer interaction of these cell families. Figure 3 illustrates the KM plots for the six approaches. We also calculated the concordance index (C-index) for the two prognostic approaches in S_v. The C-index for TriAnGIL and GNN methods

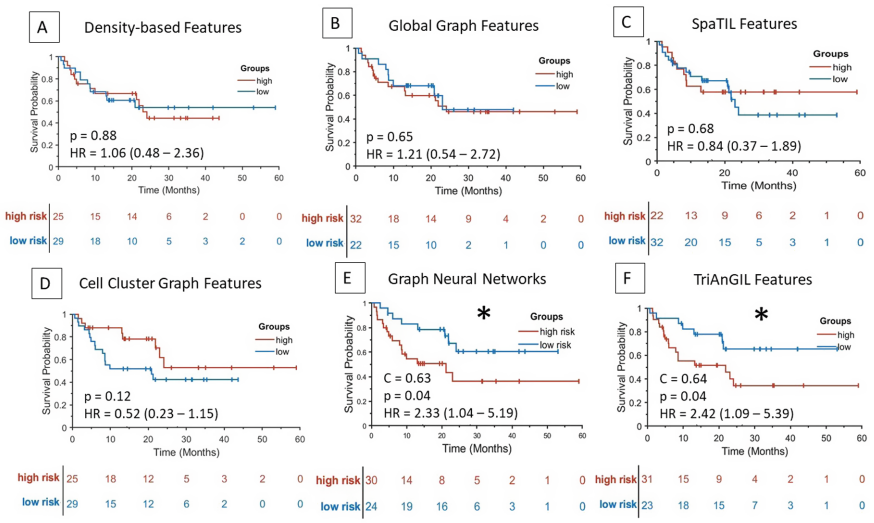

Fig. 3. The KM curves of the high-risk (red) and low-risk groups (blue) in S_v (N=54) using (A) DenTIL, (B) GG,(C) SpaTIL, (D) CCG,(E) GNN,(F) TriAnGIL features. Asterisks * indicate the models with $p < 0.05$ that statistically significantly prognosticate OS post-IO. (Color figure online)

were 0.64, and 0.63 respectively. Therefore, overall TriAnGIL worked marginally better than GNN, with much higher biological interpretability.

5 Concluding Remarks

We presented a new approach, Triangular Analysis of Geographical Interplay of Lymphocytes (TriAnGIL), to quantitatively chartacterize the spatial arrangement and relative geographical interplay of multiple cell families across pathological images. Compared to previous spatial graph-based methods, TriAnGIL quantifies the spatial interplay between multiple cell families, providing a more comprehensive portrait of the tumor microenvironment. TriAnGIL was predictive of response after IO (N = 122) and also demonstrated a strong correlation with OS in NSCLC patients treated with IO (N = 135). TriAnGIL outperformed other graph- and DL-based approaches, with the added benefit of provoding interpretability with regard to the spatial interplay between cell families. For instance, TriAnGIL yielded the insight that more interactions between stromal cells and both CD4+ and tumor cells appears to be associated with better response to IO. Although five cell families were studies in this work, TriAnGIL is flexible and could include other cell types (e.g., macrophages). Future work will entail larger validation studies and also evaluation on other use cases.

Acknowledgements. Research reported in this publication was supported by the National Cancer Institute under award numbers R01CA268287A1, U01 CA269181, R01 CA26820701A1, R01CA249992- 01A1, R01CA202752- 01A1, R01CA208236- 01A1, R01CA216579- 01A1, R01CA220581-01A1, R01CA257612- 01A1, 1U01CA239055- 01, 1U01CA248226- 01, 1U54CA254566- 01, National Heart, Lung and Blood Institute 1R01HL15127701A1, R01HL15807101A1, National Institute of Biomedical Imaging and Bioengineering 1R43EB028736- 01, VA Merit Review Award IBX004121A from the United States Department of Veterans Affairs Biomedical Laboratory Research and Development Service the Office of the Assistant Secretary of Defense for Health Affairs, through the Breast Cancer Research Program (W81XWH- 19- 1-0668), the Prostate Cancer Research Program (W81XWH- 20-1- 0851), the Lung Cancer Research Program (W81XWH-18-1-0440, W81XWH-20-1-0595), the Peer Reviewed Cancer Research Program (W81XWH- 18-1-0404, W81XWH- 21-1-0345, W81XWH- 21-1-0160), the Kidney Precision Medicine Project (KPMP) Glue Grant and sponsored research agreements from Bristol Myers-Squibb, Boehringer-Ingelheim, Eli-Lilly and Astrazeneca. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the U.S. Department of Veterans Affairs, the Department of Defense, or the United States Government.

References

1. Auffarth, B., López, M., Cerquides, J.: Comparison of redundancy and relevance measures for feature selection in tissue classification of CT images. In: Perner, P. (ed.) ICDM 2010. LNCS (LNAI), vol. 6171, pp. 248–262. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-14400-4_20

2. Basavanhally, A.N., et al.: Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. IEEE Trans. Biomed. Eng. **57**(3), 642–653 (2010)
3. Brambilla, E., et al.: Prognostic effect of tumor lymphocytic infiltration in resectable non-small-cell lung cancer. J. Clin. Oncol.: Official J. Am. Soc. Clin. Oncol. **34**, 1223–30 (2016)
4. Cartwright, D., Harary, F.: Structural balance: a generalization of Heider's theory. Psychol. Rev. **63**(5), 277 (1956)
5. Corredor, G., et al.: Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non-small cell lung cancer. Clin. Cancer Res. **25**(5), 1526–1534 (2019)
6. Cox, D.R.: Regression models and life-tables. J. Royal Stat. Soc.: Ser. B (Methodol.) **34**(2), 187–202 (1972)
7. Delaunay, B.: Sur la sphère vide. Izvestiya Akademii Nauk SSSR. Otdelenie Matematicheskikh i Estestvennykh Nauk **7**(4), 793–800 (1934)
8. Dimitrova, T., Petrovski, K., Kocarev, L.: Graphlets in multiplex networks. Sci. Rep. **10**(1), 1928 (2020)
9. Ding, R., et al.: Image analysis reveals molecularly distinct patterns of tils in NSCLC associated with treatment outcome. NPJ Precis. Oncol. **6**(1), 1–15 (2022)
10. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems 30 (2017)
11. Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. JAMA **247**(18), 2543–2546 (1982)
12. Kitts, J.A., Huang, J.: Triads. Encyclopedia Soc. Netw. **2**, 873–874 (2010)
13. Krishnamurti, U., Wetherilt, C.S., Yang, J., Peng, L., Li, X.: Tumor-infiltrating lymphocytes are significantly associated with better overall survival and disease-free survival in triple-negative but not estrogen receptor-positive breast cancers. Human Pathol. **64**, 7–12 (2017)
14. Lee, G., Veltri, R.W., Zhu, G., Ali, S., Epstein, J.I., Madabhushi, A.: Nuclear shape and architecture in benign fields predict biochemical recurrence in prostate cancer patients following radical prostatectomy: Preliminary findings. Eur. Urol. Focus **3**, 457–466 (2017)
15. Leman, A., Weisfeiler, B.: A reduction of a graph to a canonical form and an algebra arising during this reduction. Nauchno-Technicheskaya Informatsiya **2**(9), 12–16 (1968)
16. Lewis, J.S., Ali, S., Luo, J., Thorstad, W.L., Madabhushi, A.: A quantitative histomorphometric classifier (quhbic) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. Am. J. Surg. Pathol. **38**(24145650), 128–137 (2014)
17. Luen, S., et al.: Prognostic implications of residual disease tumor-infiltrating lymphocytes and residual cancer burden in triple-negative breast cancer patients after neoadjuvant chemotherapy. Ann. Oncol. **30**(2), 236–242 (2019)
18. Ma, Y., Tang, J.: Deep Learning on Graphs. Cambridge University Press, Cambridge (2021)
19. Malhotra, J., Jabbour, S.K., Aisner, J.: Current state of immunotherapy for non-small cell lung cancer. Transl. Lung Cancer Res. **6**(2), 196 (2017)
20. Newman, M.: Networks. Oxford University Press, Oxford (2018)
21. Reck, M., et al.: Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N. Engl. J. Med. **375**, 1823–1833 (2016)

22. de Rodas, M.L., et al.: Role of tumor infiltrating lymphocytes and spatial immune heterogeneity in sensitivity to PD-1 axis blockers in non-small cell lung cancer. J. ImmunoTherapy Cancer **10**(6), e004440 (2022)
23. Sato, J., et al.: CD20+ tumor-infiltrating immune cells and CD204+ M2 macrophages are associated with prognosis in thymic carcinoma. Cancer Sci. **111**(6), 1921–1932 (2020)
24. Schalper, K.A., et al.: Objective measurement and clinical significance of TILs in non-small cell lung cancer **107**(3). https://doi.org/10.1093/jnci/dju435
25. Sherwin, R.G.: Introduction to the graph theory and structural balance approaches to international relations. University of Southern California Los Angeles, Tech. Rep. (1971)
26. Simon, R.M., Subramanian, J., Li, M.C., Menezes, S.: Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data. Briefings Bioinform. **12**(3), 203–214 (2011)
27. Tavares, M.C., et al.: A high CD8 to FOXP3 ratio in the tumor stroma and expression of PTEN in tumor cells are associated with improved survival in non-metastatic triple-negative breast carcinoma. BMC Cancer **21**(1), 1–12 (2021)
28. Tibshirani, R.: The lasso method for variable selection in the cox model. Stat. Med. **16**(4), 385–395 (1997)
29. Vaswani, A., et al.: Attention is all you need (2017). https://doi.org/10.48550/arxiv.1706.03762
30. Whiteside, T.: The tumor microenvironment and its role in promoting tumor growth. Oncogene **27**(45), 5904–5912 (2008)
31. Zhou, Y., et al.: Transformer as a spatially aware multi-instance learning framework to predict the risk of death for early-stage non-small cell lung cancer. In: Digital and Computational Pathology. No. 12471–33, SPIE (TBD 2023), accepted for publication