An RNA-based model for tertiary lymphoid structure (TLS) prediction and classification in pancreatic adenocarcinoma (PDAC)

BostonGene

Abstract 4909

Alexandra Livanova, Andrey Tyshevich, Andrey Kravets, Stanislav Kurpe, Nadezhda Lukashevich, Dmitry Ivchenkov, Daniil Dymov, Anna Belozerova, Kirill Kryukov, Aleksandr Sarachakov, Viktor Svekolkin, Vladimir Kushnarev

BostonGene Corporation, Waltham, Massachusetts, USA

Introduction

Tertiary lymphoid structures (TLS) are lymphoid aggregates composed mainly of B cells, T cells and dendritic cells (DC), which form in peripheral non-lymphoid tissues including malignant tumors[1]. Although TLS status is of prognostic significance in pancreatic adenocarcinoma (PDAC) and can potentially affect chemotherapy outcomes[1], there is currently a notable lack of RNA sequencing (RNA-seq) models that specialize in TLS identification and classification in PDAC. Here, we developed a model for predicting TLS status based on the RNA-seq data of patients (**Fig. 1**).

Methods

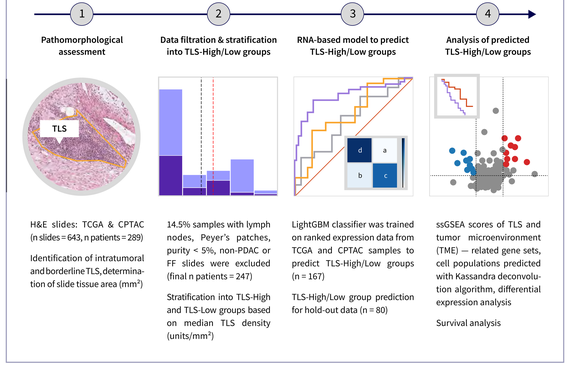
Hematoxylin and eosin (H&E) whole slide images of PDAC samples from TCGA (n = 118) and CPTAC (n = 129) were used for pathomorphological assessment of intratumoral and borderline TLS followed by TLS density measurements (units/mm^2) (**Fig. 2**). The samples were then stratified into TLS-High and TLS-Low groups based on median density values. The LightGBM gradient boosting classifier was then trained on ranked expression data with sequential feature selection to predict TLS-High and TLS-Low groups. We trained this RNA-based model with H&E staining annotations and ranked RNA expression data from TCGA or CPTAC samples (total n = 167). The remaining 80 samples were designated as hold-out dataset. The weighted classification metrics were computed to compare performance with TLS Functional Gene Expression Signatures (FGES). Next, we applied deconvolution by Kassandra algorithm[2] to identify cell subtypes abundant in each TLS group based on gene expression (RNA-seq) data. Calculation of ssGSEA scores for gene signatures corresponding to cell subtypes and TLS structures was performed, along with survival analysis. Finally, we analyzed differential expression and enrichment in predicted TLS-High and TLS-Low samples.

CI — Confidence Interval, CPTAC — Clinical Proteomic Tumor Analysis Consortium, DC — dendritic cells, DE — Differential Expression, DEGs — Differentially Expressed Genes, FF — fresh frozen, FGES — Functional Gene Expression Signature, GO — Gene Ontology, H&E — Hematoxylin and eosin, HR — Hazard Ratio, logFC — log$_2$FoldChange, MAD — Median Absolute Deviation, OS — Overall Survival, PDAC — Pancreatic Ductal Adenocarcinoma, ssGSEA — single sample Gene Set Enrichment Analysis, TCGA — The Cancer Genome Atlas, TLS — tertiary lymphoid structure, TME — tumor microenvironment.

Results

1

Figure 1. Data and analysis pipeline



2

Figure 2. Cohort overview

A. Clinical characteristics

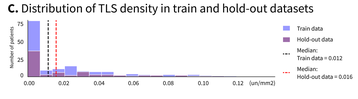
OS (month)		CPTAC	TCGA
median (interquartile range: Q1;Q3)	Train	13.1 (8.6, 21.4)	15.4 (8.9, 19.8)
	Hold-out	14.0 (9.3, 23.8)	12.9 (7.8, 21.9)
Stage I number (% in group)	Train	12 [15%]	5 [7%]
	Hold-out	4 [10%]	3 [9%]
Stage II number (% in group)	Train	40 [51%]	65 [88%]
	Hold-out	22 [56%]	32 [91%]
Stage III number (% in group)	Train	23 [30%]	1 [1%]
	Hold-out	10 [26%]	0 [0%]
Stage IV number (% in group)	Train	3 [4%]	3 [4%]
	Hold-out	3 [8%]	0 [0%]
Grade I number (% in group)	Train	7 [8%]	2 [3%]
	Hold-out	2 [5%]	2 [5%]
Grade II number (% in group)	Train	59 [67%]	40 [51%]
	Hold-out	33 [80%]	18 [46%]
Grade III number (% in group)	Train	22 [25%]	36 [46%]
	Hold-out	6 [15%]	19 [49%]

B. Demographic composition

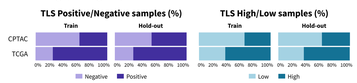
Age (years)		CPTAC	TCGA
median (interquartile range: Q1;Q3)	Train	66 (60, 72)	65 (56, 73)
	Hold-out	66 (59, 72)	65 (59, 74)

Gender		CPTAC F	CPTAC M	TCGA F	TCGA M
number (% in group)	Train	43 [49%]	45 [51%]	39 [49%]	40 [51%]
	Hold-out	20 [49%]	21 [51%]	12 [31%]	27 [69%]

C. Distribution of TLS density in train and hold-out datasets



D. TLS-High/Low groups occurrence in train and hold-out datasets



TLS status number (% in group)		CPTAC N	CPTAC P	TCGA N	TCGA P
	Train	54 [61%]	34 [39%]	19 [24%]	60 [76%]
	Hold-out	21 [51%]	20 [49%]	10 [26%]	29 [74%]

High ⇄ Low TLS number (% in group)		CPTAC L	CPTAC H	TCGA L	TCGA H
	Train	56 [64%]	32 [36%]	29 [37%]	50 [63%]
	Hold-out	25 [61%]	16 [39%]	14 [36%]	25 [64%]

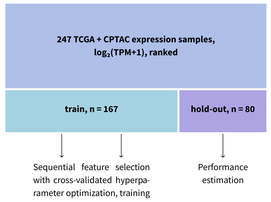
OS — Overall Survival (month); Gender F — Female, M — Male; TLS Status N — negative, P — Positive; High & Low TLS L — Low, H — High

Occurrence of intratumoral and borderline TLS was assessed by pathologist. (A) Clinical and **(B)** demographic characteristics of cohorts used. **(C)** Distribution of TLS density (un/mm^2) assessed by pathologist. **(D)** Occurrence of TLS-High/Low groups. TCGA cohort was enriched with TLS-positive and TLS-High samples compared to CPTAC.

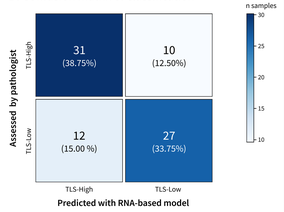
3

Figure 3. RNA-based model to predict TLS status

A. Data splitting workflow



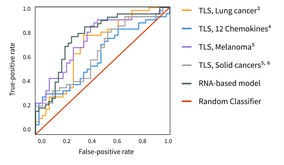
B. Confusion matrix of classification



C. Weighted classification metrics

	Accuracy	Precision	Recall	ROC-AUC	F1
TLS, Lung cancer, 2021[3]	0.66	0.67	0.66	0.70	0.65
TLS, 12 chemokines, 2011[4]	0.60	0.60	0.60	0.62	0.60
TLS, Melanoma, 2020[5]	0.61	0.63	0.61	0.75	0.60
TLS, Solid cancers, 2019[5,6]	0.53	0.53	0.53	0.65	0.52
RNA-based model	**0.73**	**0.73**	**0.72**	**0.77**	**0.72**

D. ROC-AUC curves

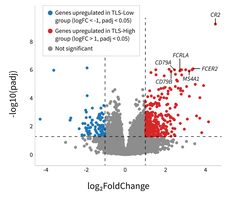


RNA-based model was trained with expression data to predict TLS-High/Low status in PDAC patients. (A) TCGA and CPTAC samples were used to train RNA-based classifier. **(B)** Confusion matrix for TLS-High/Low classification in hold-out dataset. **(C)** Weighted metrics confirm that the model outperformed previously published TLS FGES. **(D)** ROC-AUC curves of the model and TLS FGES in the hold-out dataset.

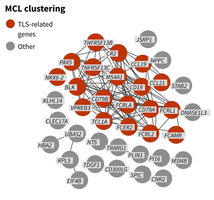
4

Figure 4. Revealing transcriptomic features associated with lymphocytes and DCs in predicted TLS-High/Low groups

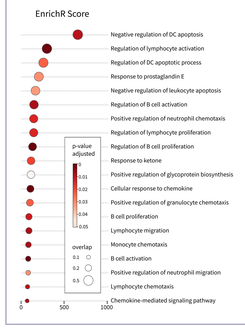
A. DE for TLS-Low vs TLS-High group



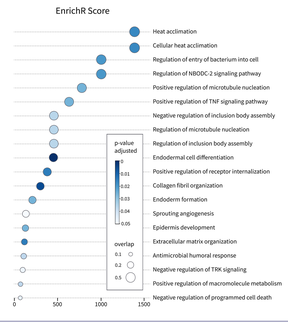
B. TLS-High top genes, STRING net



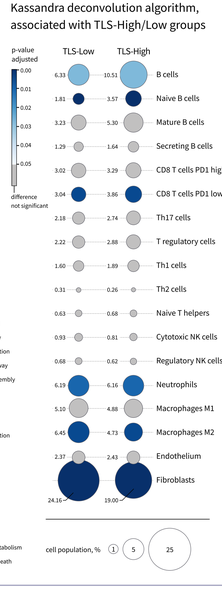
C. Top20 GO pathways enriched in TLS-High group



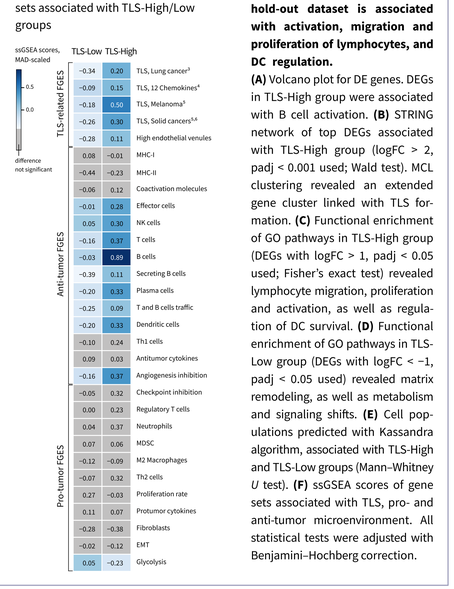
D. Top20 GO pathways enriched in TLS-Low group



E. Cell populations predicted with Kassandra deconvolution algorithm, associated with TLS-High/Low groups



F. MAD-scaled ssGSEA scores of gene sets associated with TLS-High/Low groups



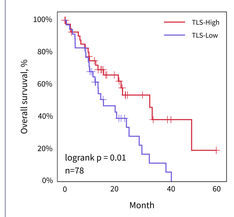
Predicted TLS-High group in hold-out dataset is associated with activation, migration and proliferation of lymphocytes, and DC regulation.

(A) Volcano plot for DE genes. DEGs in TLS-High group were associated with B cell activation. **(B)** STRING network of top DEGs associated with TLS-High group (logFC > 2, padj < 0.001 used; Wald test). MCL clustering revealed an extended gene cluster linked with TLS formation. **(C)** Functional enrichment of GO pathways in TLS-High group (DEGs with logFC > 1, padj < 0.05 used; Fisher's exact test) revealed lymphocyte migration, proliferation and activation, as well as regulation of DC survival. **(D)** Functional enrichment of GO pathways in TLS-Low group (DEGs with logFC < −1, padj < 0.05 used) revealed matrix remodeling, as well as metabolism and signaling shifts. **(E)** Cell populations predicted with Kassandra algorithm, associated with TLS-High and TLS-Low groups (Mann–Whitney *U* test). **(F)** ssGSEA scores of gene sets associated with TLS, pro- and anti-tumor microenvironment. All statistical tests were adjusted with Benjamini–Hochberg correction.

5

Figure 5. TLS-High/Low groups and overall survival

A. Kaplan–Meier curves



B. Cox regression for survival prediction models comparison

	C-index (base = 0.64)	log(HR)	log(HR) lower CI	log(HR) upper CI	p-value
TLS, Lung cancer[3]	0.68	0.35	-0.31	1.00	0.30
TLS, 12 chemokines[4]	0.68	0.10	-0.58	0.78	0.77
TLS, Melanoma[5]	0.69	0.35	-0.50	1.20	0.42
TLS, Solid cancers[5, 6]	**0.70**	0.62	-0.06	1.29	0.07
RNA-based model	0.68	**0.76**	**0.04**	**1.48**	**0.04**

Predicted TLS-Low status is associated with worse overall survival in hold-out dataset. (A) Kaplan–Meier plot. **(B)** Metrics of Cox-regression model adjusted for clinical and demographical parameters (grade, stage, sex, age) and TLS status. Predictions of TLS status performed with the RNA-based model show stronger associations with OS compared to TLS FGES. "Base" C-index refers to the model not adjusted for TLS status. Best performance in terms of clinical benefit is designated in bold.

Our RNA-based classifier predicted TLS-High/Low groups with a F1 weighted score of 0.72 and ROC-AUC score of 0.77, outperforming TLS FGES (**Fig.3**). Kassandra deconvolution revealed B-cell abundance, but fibroblast and macrophage depletion in the TLS-High group (**Fig.4**).

SsGSEA scores of previously described TLS gene signatures, along with those of different B-cell subtypes and DCs, showed strong association with the TLS-High group. Genes associated with B-cell proliferation, differentiation, and signaling were also upregulated in this group (**Fig.4**).

Patients in the predicted TLS-Low group had worse overall survival compared to the TLS-High group (Log(HR) = 0.76; 95% CI [0.04; 1.48]; p = 0.04). Moreover, RNA-based model predictions demonstrated the best association with overall survival compared to those obtained from FGES of previously known TLS-related gene sets (**Fig.5**).

Conclusions

We present an RNA-based model capable of stratifying pancreatic ductal adenocarcinoma samples into TLS-High/Low groups with results consistent with pathological annotations. Predicted TLS-low group is associated with worse survival. Thus, the model may be used as an objective tool to predict outcomes for PDAC patients based on TLS status.

References:

1. Zou X. et al. Characterization of intratumoral tertiary lymphoid structures in pancreatic ductal adenocarcinoma: cellular properties and prognostic significance. J Immunother Cancer 2023; 11: e006698.
2. Zaitsev A. et al. Precise reconstruction of the TME using bulk RNA-seq and a machine learning algorithm trained on artificial transcriptomes. Cancer Cell. 2022; 40(8): 879-894. e16.
3. Feng H. et al. Prognostic Significance of Gene Signature of Tertiary Lymphoid Structures in Patients With Lung Adenocarcinoma. Front Oncol. 2021; 11: 693234.
4. Coppola D. et al. Unique ectopic lymph node-like structures present in human primary colorectal carcinoma are identified by immune gene array profiling. Am J Pathol. 2011; 179(1): 37-45.
5. Cabrita, R. et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature. 2020; 577: 561-565.
6. Sautès-Fridman C. et al. Tertiary lymphoid structures in the era of cancer immunotherapy. Nat Rev Cancer. 2019; 19: 307-325.

Corresponding author:
Vladimir Kushnarev
vladimir.kushnarev@bostongene.com

Poster electronic version