



OPEN

Prognostic significance of AP-2 α / γ targets as cancer therapeutics

Damian Kołat , Żaneta Kałuzińska, Andrzej K. Bednarek & Elżbieta Płuciennik

Identifying genes with prognostic importance could improve cancer treatment. An increasing number of reports suggest the existence of successful strategies based on seemingly “untargetable” transcription factors. In addition to embryogenesis, AP-2 transcription factors are known to play crucial roles in cancer development. Members of this family can be used as prognostic factors in oncological patients, and AP-2 α / γ transcription factors were previously investigated in our pan-cancer comparative study using their target genes. The present study investigates tumors that were previously found similar with an emphasis on the possible role of AP-2 factors in specific cancer types. The RData workspace was loaded back to R environment and 3D trajectories were built via Monocle3. The genes that met the requirement of specificity were listed using `top_markers()`, separately for mutual and unique targets. Furthermore, the candidate genes had to meet the following requirements: correlation with AP-2 factor (through Correlation AnalyzerR) and validated prognostic importance (using GEPIA2 and subsequently KM-plotter or LOGpc). Eventually, the ROC analysis was applied to confirm their predictive value; co-dependence of expression was visualized via BoxPlotR. Some similar tumors were differentiated by AP-2 α / γ targets with prognostic value. Requirements were met by only fifteen genes (*EMX2*, *COL7A1*, *GRIA1*, *KRT1*, *KRT14*, *SLC12A5*, *SEZ6L*, *PTPRN*, *SCG5*, *DPP6*, *NTSR1*, *ARX*, *COL4A3*, *PPEF1* and *TMEM59L*); of these, the last four were excluded based on ROC curves. All the above genes were confronted with the literature, with an emphasis on the possible role played by AP-2 factors in specific cancers. Following ROC analysis, the genes were verified using immunohistochemistry data and progression-related signatures. Staining differences were observed, as well as co-dependence on the expression of e.g. *CTNNB1*, *ERBB2*, *KRAS*, *SMAD4*, *EGFR* or *MKI67*. In conclusion, prognostic value of targets suggested AP-2 α / γ as candidates for novel cancer treatment. It was also revealed that AP-2 targets are related to tumor progression and that some mutual target genes could be inversely regulated.

Activating enhancer-binding Protein 2 (AP-2) is a family of transcription factors (TFs) belonging to the basic Helix-Loop-Helix class (bHLH) of Superclass 1¹. In homeostatic conditions, its members regulate embryogenesis by managing proliferation, apoptosis or the cell cycle, thus ensuring the correct development of limbs, eyes or facial features^{2,3}. However, their altered functionality plays a crucial role in cancer and can be prognostic in oncological patients^{4–6}. The fact that each AP-2 member has a different profile within a given tumor, makes this complex field worth investigation. Moreover, an increasing number of reports indicate that there are successful strategies for seemingly “untargetable” transcription factors^{7,8}, which opens avenues for the future. Our previous investigations examined the properties of the two best-described AP-2 factors (AP-2 α and AP-2 γ) in various cancers⁹, they then compared the AP-2 α / γ targets between more than twenty tumor types from The Cancer Genome Atlas (TCGA)¹⁰. The findings shed light on the ability of AP-2 α and AP-2 γ to regulate the processes underlying the hallmarks of cancer, and the difference between tumor tissues regarding their target genes expression. They also indicated the potential for identifying cancer in cases where normal tissue samples had a distinct expression pattern compared to a corresponding tumor type. The current study provides a further examination of cancer types that were previously found to be similar. The study identifies mutual target genes that might be differently regulated by AP-2 α and AP-2 γ in specific cancers. It also explores unique target genes (i.e. for each AP-2 factor) that might have prognostic value in these tumors. In addition, it discusses the relevance of AP-2 α / γ as candidate transcription factors suitable for cancer treatment.

Department of Molecular Carcinogenesis, Medical University of Lodz, 90-752 Lodz, Poland. ✉email: damian.kolat@stud.umed.lodz.pl

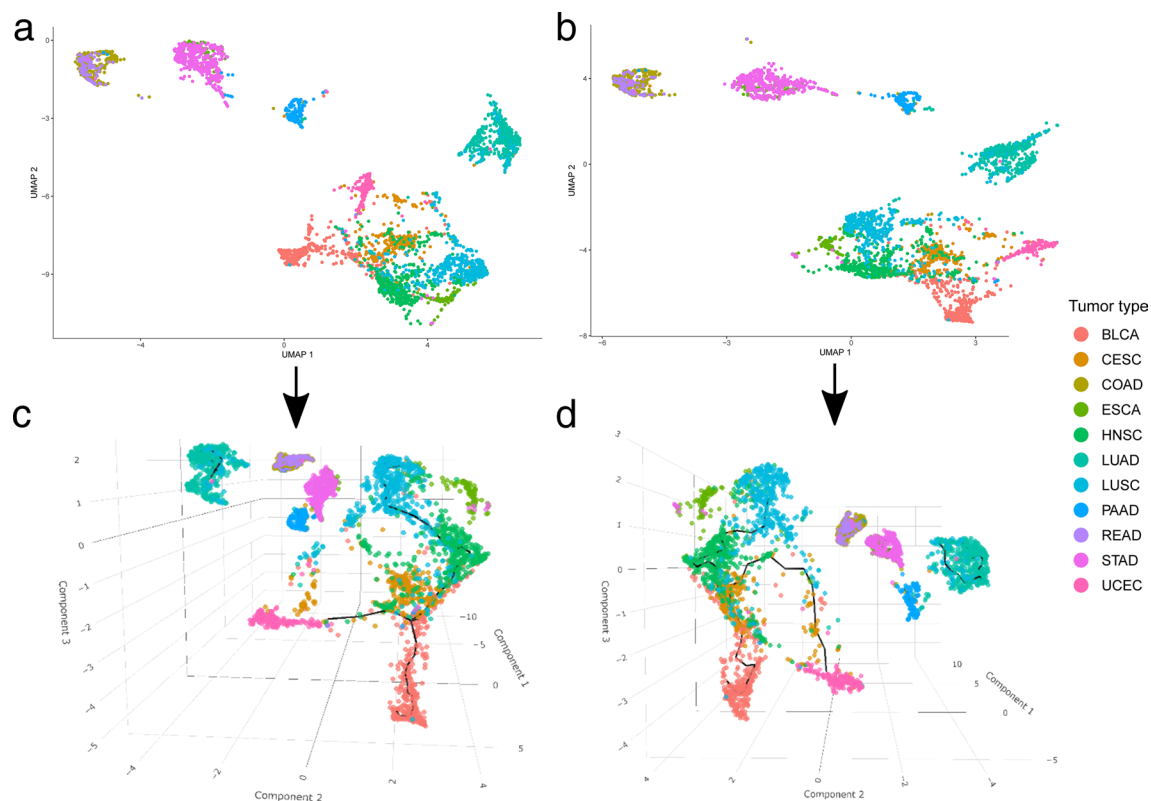


Figure 1. Spatial analysis showing differences between tumors. **(a)** AP-2 α target genes list. **(b)** AP-2 γ target genes list. **(c)** 3D trajectory of the first subfigure. **(d)** 3D trajectory of the second subfigure. Figure created using Monocle3 (<https://cole-trapnell-lab.github.io/monocle3/>).

Results

3D trajectories revealed the dissimilarities across tumor types.

As a preliminary remark, in our previous study¹⁰ the analysis included global profiling of twenty-one tumors via Monocle3 and was directed to functionally annotate whole gene sets. In the current research, it was decided to focus more specifically on individual AP-2 targets and their significance as cancer therapeutics. Only the tumors that formerly could not be clearly distinguished were included in this study—this has resulted in an analysis which uses RData from a previous study where visualization with Uniform Manifold Approximation and Projection (UMAP) was preliminarily performed. Due to the reduction in the number of cohorts to eleven, the graphs were re-learned to ensure the best dimensional distribution. The distribution was visualized with regard to the full list of gene targets for a given transcription factor; the trajectories are presented in Fig. 1A and B, while the example three-dimensional extensions are visualized in Fig. 1C and D. Complete and interactive 3D plots corresponding to Fig. 1C and D are deposited as Supplementary File S1 and Supplementary File S2, respectively.

Despite being very similar at first glance, subtle differences were observed. For example, some STAD samples were better separated from the rest of the cohort when visualized through AP-2 γ target genes. Likewise, UCEC was more distinct from other surrounding cohorts when AP-2 α target genes were applied. Nevertheless, the dissimilarities between tumor types can be seen more clearly in 3D trajectories. Additionally, the tumors were grouped based on the expression level of each AP-2 factor (a common median cut-off was applied for all cancers at once). The results indicate that some tumor types demonstrate higher AP-2 α/γ expression than others, and that some are more or less uniform in expression (Fig. 2).

For example in STAD or UCEC, a greater number of AP-2 α “high” samples can be seen compared to AP-2 γ “high”, while in BLCA or LUAD, more AP-2 γ “high” samples are found compared to AP-2 α . Generally, it seems that both AP-2 factors demonstrate higher expression in the heterogeneous mixed cluster (containing BLCA, CESC, ESCA, HNSC, LUSC, UCEC) compared to the other distinct clusters/cohorts.

Between tumors there are unique target genes of prognostic importance.

The study identified unique AP-2 α/γ targets between tumors or in a specific cancer. For tumors, this indicated AP-2 target genes which are the most specific for each tumor when compared to others. Only the genes that simultaneously satisfy the demands of specificity, correlation with an AP-2 factor and survival prediction (with external validation) are shown in detail (see penultimate subsection of Results for total number of genes included from each methodology branch). As indicated in Fig. 3A and B (respectively for AP-2 α and AP-2 γ), our findings indicate that while single gene can be expressed in a few tumors, most are associated more closely with a specific cancer type. For example, *CDX1* (an AP-2 α target) is highly expressed in colorectal carcinoma (COAD and READ cohorts) but it is also expressed to a certain extent in STAD. Likewise, the AP-2 γ target *ADAM23* is mainly expressed in LUSC

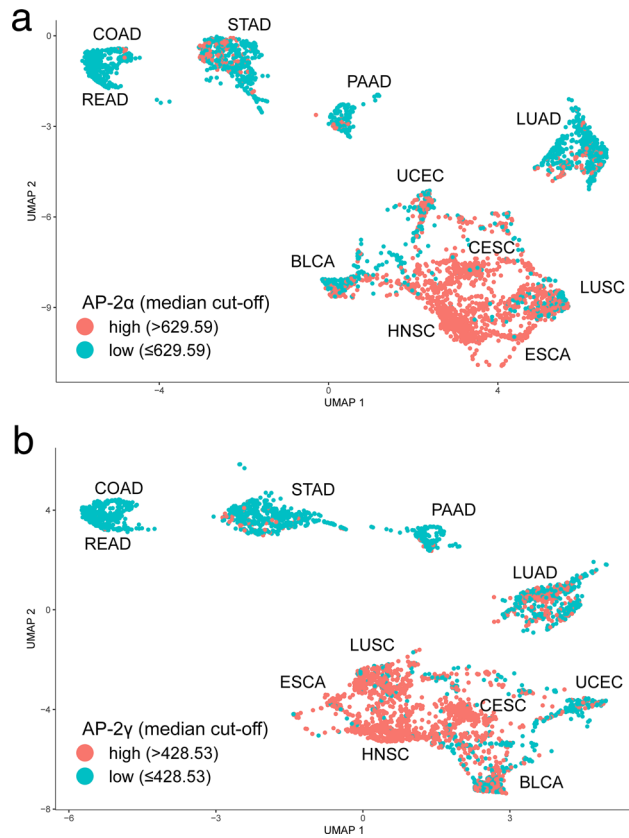


Figure 2. Spatial analysis showing differences between AP-2 factor level. (a) AP-2α. (b) AP-2γ. Figure created using Monocle3 (<https://cole-trapnell-lab.github.io/monocle3/>).

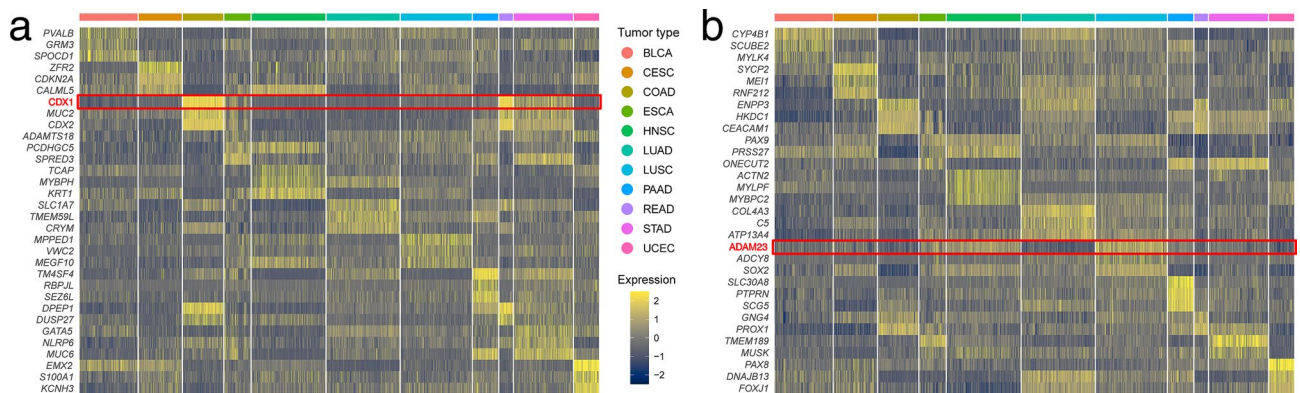


Figure 3. Top three the most specific genes per tumor. (a) AP-2α unique targets. (b) AP-2γ unique targets. In-text examples are marked with red. Figure created using Seurat 4.0.4 (<https://CRAN.R-project.org/package=Seurat>).

but also in ESCA or HNSC. As expected, some tumors demonstrate more specific expression of AP-2 target genes than others.

All sixty-six genes were subjected to downstream investigation. Three met all the requirements: the AP-2α-regulated *EMX2* for UCEC, and the AP-2γ targets *PTPRN* and *SCG5* for PAAD. Not only were they negatively correlated with the corresponding AP-2 factor, but they also significantly affected Disease-Free Survival (DFS), as confirmed externally. Regarding specificity, *PTPRN* was found to be the most specific (0.73), followed by *EMX2* (0.72) and *SCG5* (0.65); this can be seen in Fig. 4 (the last graph in each subfigure).

Unique targets can predict patient survival for a specific cancer type. Unique AP-2α or AP-2γ targets were also independently analyzed within a specific cancer by classifying AP-2a/γ expression as “high” or “low”, with regard to a median cut-off point. The three genes with the highest specificity per “high” or “low”

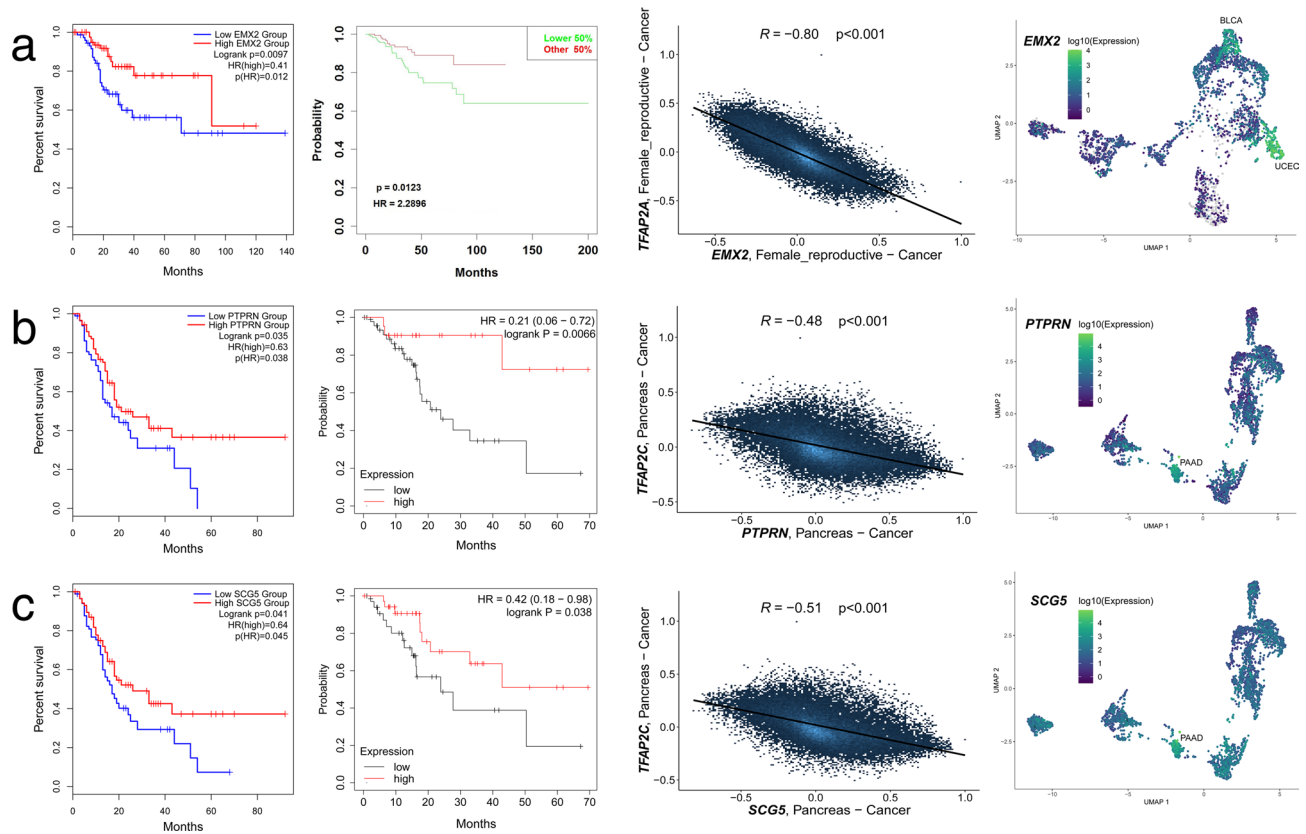


Figure 4. The best three candidate genes from unique targets. **(a)** *EMX2* (AP-2 α target). **(b)** *PTPRN* (AP-2 γ target). **(c)** *SCG5* (AP-2 γ target). Each subfigure contains results from survival analysis (left), which was further validated (middle-left). The candidate gene was correlated with AP-2 factor (middle-right) and the specificity of gene expression in tumors was assessed (right).

group were identified for each AP-2 factor, giving six genes per tumor. As some AP-2 targets did not meet the minimal specificity requirement (Table 1), fewer genes were included in the downstream analysis.

Eventually, it was found that seven targets of AP-2 α (*KRT1*, *COL7A1*, *TMEM59L*, *KRT14*, *SLC12A5*, *GRIA1*, *SEZ6L*) were of prognostic importance for patients having BLCA, LUAD, LUSC or PAAD. Regarding genes regulated by AP-2 γ , three targets (*PPEF1*, *ARX*, *COL4A3*) had prognostic value for individuals from LUAD, PAAD or UCEC cohort. Most of the AP-2 α unique targets concerned PAAD; interestingly, they were all negatively correlated with transcription factor-encoding gene. In contrast, the remaining part (*KRT1* for BLCA, *COL7A1* for LUAD and *TMEM59L* for LUSC) was positively correlated with AP-2 α . When it comes to AP-2 γ , it correlated positively with *PPEF1* and *ARX* while negatively with *COL4A3*. Figure 5 presents a summary for targets of both TFs.

Differently-regulated mutual target genes were found in ESCA and LUAD. To identify genes whose expression could be regulated in opposite directions by AP-2 α and AP-2 γ within a single tumor type, each cancer was examined individually. Samples of specific tumor were compared regarding high/low phenotypes for the two AP-2 factors; this allowed changes in expression of mutual target genes to be established between phenotypes. Only two genes satisfied all criteria: *DPP6* and *NTSR1*. AP-2 α strongly downregulates the former (Fig. 6A), while AP-2 γ appears to downregulate the latter (Fig. 6B). High *DPP6* expression is associated with shorter survival in ESCA patients and *NTSR1* with shorter survival in LUAD.

ROC curves confirmed that most of the selected genes are good predictors. The relevant targets from previous sections were first demonstrated in Table 2 to infer the possible AP-2 role in specific cancer, and then were subjected to Receiver Operating Characteristic (ROC) analysis to confirm predictive abilities. Binary classification depended on methodology branch e.g. for genes identified within a specific cancer the “high” and “low” groups of AP-2 factor(s) were used. The targets identified by the tumor vs tumor comparisons (*EMX2*, *PTPRN*, *SCG5*) required a representative cohort for ROC analysis; these three genes concern only UCEC and PAAD tumors. For *EMX2* expression, UCEC was compared with CESC, an other carcinoma of the female reproductive system. For *PTPRN* and *SCG5* expression, PAAD was compared to ESCA, since it was the closest cohort in terms of size (no tumor with a similar site of origin to PAAD was included in this study). Out of fifteen genes that were suitable for ROC analysis, *COL4A3*, *PPEF1* and *TMEM59L* did not predict a binary outcome sufficiently: their Area Under the Curve (AUC) was respectively 0.57, 0.56 and 0.57. The usefulness of *ARX* is also

Transcription factor	Tumor	“High” group	“Low” group
AP-2α	BLCA	<u>KRT1, KRT14, NOTUM</u>	<u>TCAP, SLC16A12, EMX2</u>
	CESC	<u>KRT14, GLDC, SOX15</u>	<u>DPEP1, TLL10, NOTUM</u>
	COAD	<u>MUC6, SLC26A9, TM4SF4</u>	<u>TCAP, TMEM132C, ISM2</u>
	ESCA	<u>KRT1, KRT14, CALML5</u>	<u>NLRP6, DPEP1, CDX1</u>
	HNSC	<u>KRT1, PCDHGC5, PLA2G4D</u>	<u>MUC6, NOTUM, ALDH1A1</u>
	LUAD	<u>KRT14, COL7A1, LAMA1</u>	<u>NOTUM, KCNE4, SLC16A12</u>
	LUSC	<u>PCDHGC5, VASH2, KRT14</u>	<u>NOTUM, DLL3, TMEM59L</u>
	PAAD	<u>MUC2, KRT14, CGB5</u>	<u>SLC12A5, GRIAL, SEZ6L</u>
	READ	<u>GRIK2, TM4SF4, SHC2</u>	<u>EDA, GPR55, SCN4A</u>
	STAD	<u>TGM1, S100A2, KRT14</u>	<u>SMTNL2, SLC16A12, PYGM</u>
	UCEC	<u>KRT14, ADAMTS18, KCNH3</u>	<u>DLL3, SYN1, ADAM33</u>
AP-2γ	BLCA	<u>MYLK4, ALOXE3, CEACAM6</u>	<u>FAM78B, SIGLEC15, PAX8</u>
	CESC	<u>PDE6A, PRIMA1, AKR1B15</u>	<u>HKDC1, HABP2, PAX8</u>
	COAD	<u>FAM83A, PTPRN, PRIMA1</u>	<u>KCNK17, ENPP3, STAG3</u>
	ESCA	<u>AKR1B15, CDH26, ALDH3B2</u>	<u>HABP2, NPC1L1, FOXA3</u>
	HNSC	<u>GSTM5, GSDMA, PLA2G4E</u>	<u>INSM1, SLC7A2, KCNJ10</u>
	LUAD	<u>KRT15, PRSS27, SCUBE3</u>	<u>NPC1L1, INSM1, PPEF1</u>
	LUSC	<u>PRSS27, AKR1B15, SYCP2</u>	<u>INSM1, SPIB, ENPP3</u>
	PAAD	<u>FAM83A, IRX3, PTGES</u>	<u>COL11A2, ARX, VGF</u>
	READ	<u>CYP24A1, MYO3A, FPR1</u>	<u>F5, RPL39L, HKDC1</u>
	STAD	<u>PRSS27, KRT15, FAM83A</u>	<u>SPIB, CCR7, KCNJ10</u>
	UCEC	<u>VIPR1, RUNDC3B, COL14A1</u>	<u>ONECUT2, CEACAM6, COL4A3</u>

Table 1. The top three most specific genes per “high” or “low” group in tumors for both AP-2α and AP-2γ. Underlined and in bold are genes that met >0.6 requirement in terms of specificity estimated by top_markers().

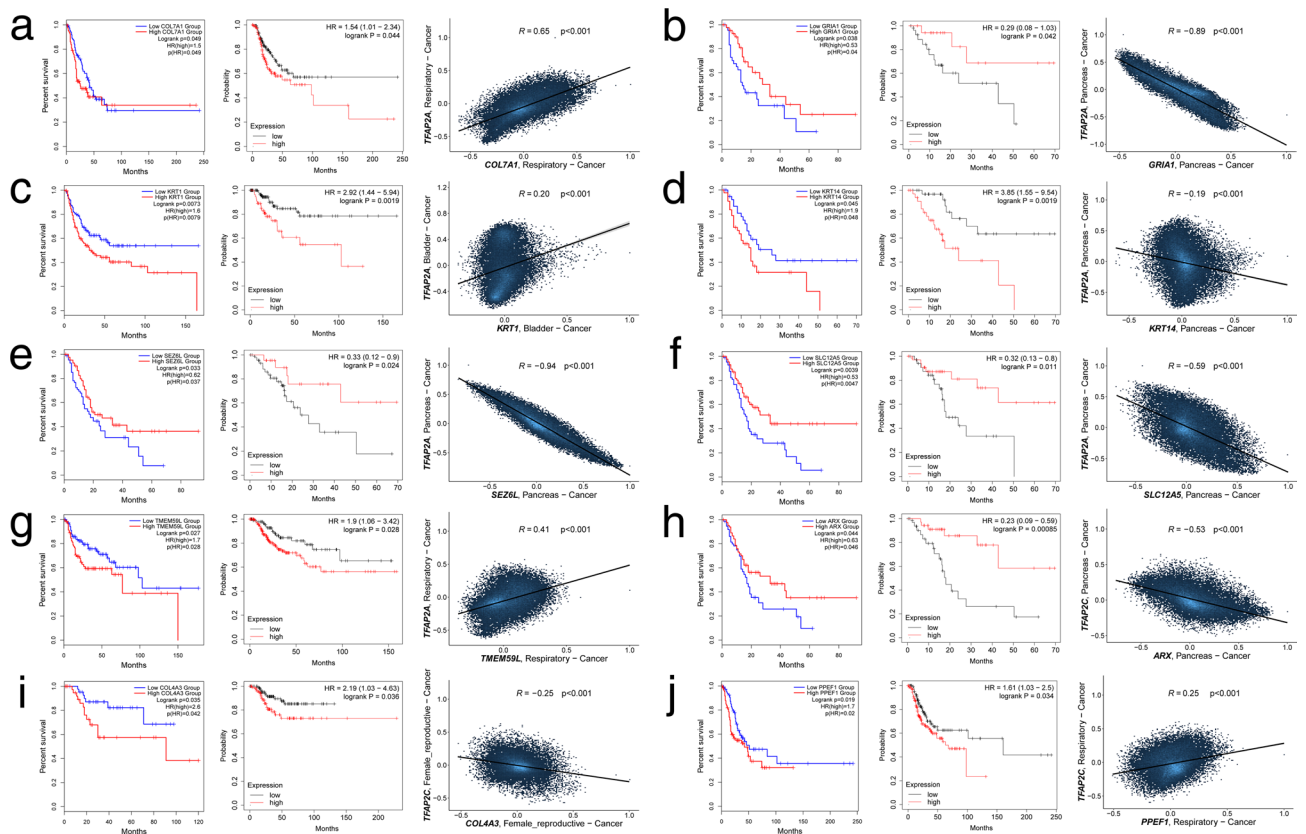


Figure 5. Relevant genes identified within a specific cancer. (a) COL7A1 (AP-2α target). (b) GRIAL (AP-2α target). (c) KRT1 (AP-2α target). (d) KRT14 (AP-2α target). (e) SEZ6L (AP-2α target). (f) SLC12A5 (AP-2α target). (g) TMEM59L (AP-2α target). (h) ARX (AP-2γ target). (i) COL4A3 (AP-2γ target). (j) PPEF1 (AP-2γ target).

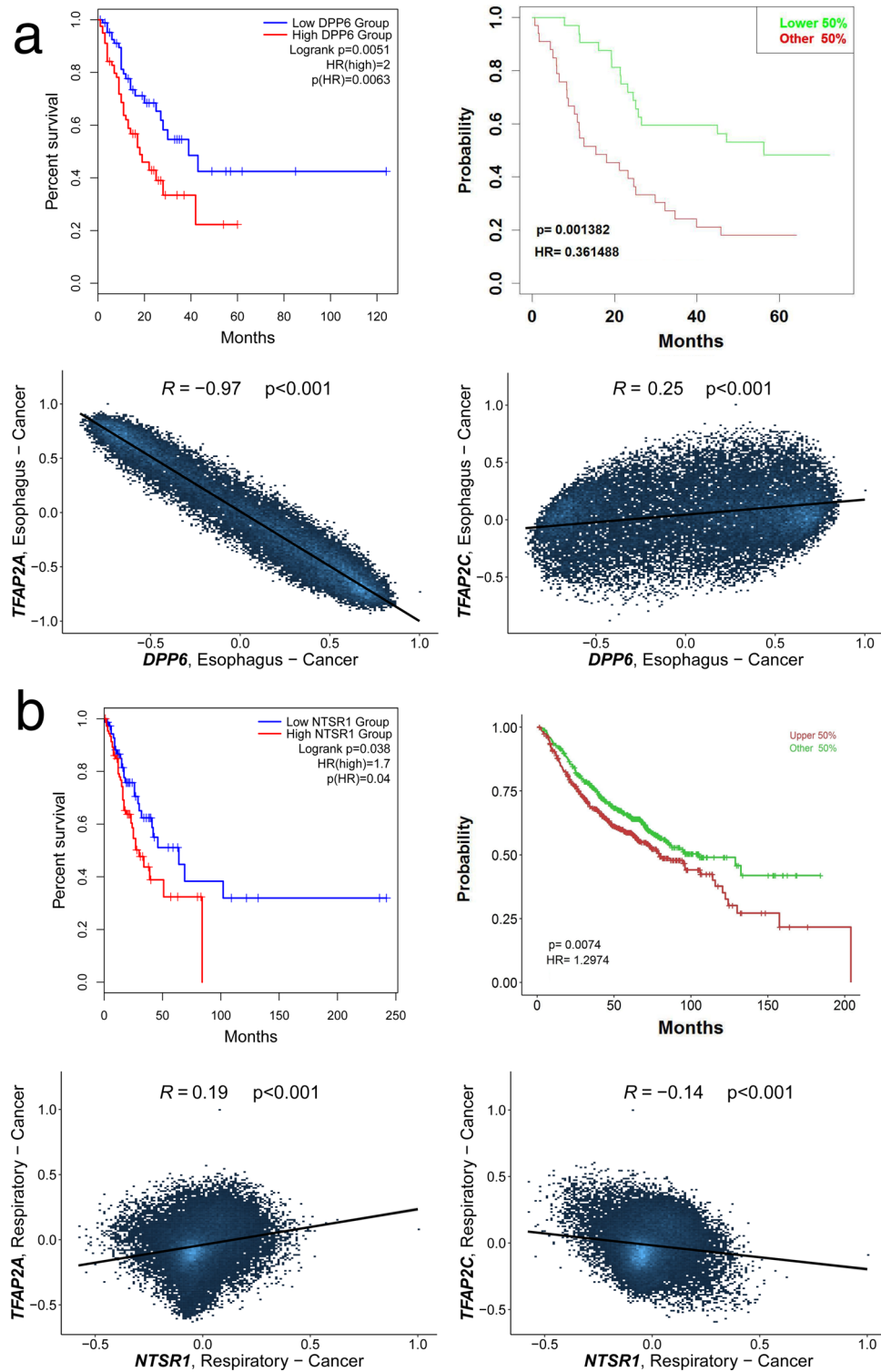


Figure 6. Differently regulated mutual targets. (a) *DPP6*. (b) *NTSR1*.

questionable since the AUC was 0.61. Nevertheless, the remaining AP-2 targets presented AUC > 0.65 including some even above 0.9. ROC curves are collected in Fig. 7.

Immunohistochemistry showed staining differences and the genes were found to be related to tumor progression. To complement the results of the survival analysis, all genes that met the AUC requirement (herein denoted as post-ROC genes) were subjected to further analysis of their immunohistochemical (IHC) data (Fig. 8). Out of eleven genes, there was insufficient or no data for *DPP6* or *SEZ6L* and *NTSR1* in

Transcription factor	Tumor	Gene of interest	Prognosis when gene is highly expressed	Correlation with AP-2 factor	Possible role of AP-2 factor in cancer
AP-2α	BLCA	<i>KRT1</i>	Unfavorable	Positive	Pro-tumorigenic
	ESCA	<i>DPP6</i>	Unfavorable	Negative	Anti-cancer
	LUAD	<i>COL7A1</i>	Unfavorable	Positive	Pro-tumorigenic
		<i>NTSR1</i>	Unfavorable	Positive	Pro-tumorigenic
	LUSC	<i>TMEM59L</i>	Unfavorable	Positive	Pro-tumorigenic
	PAAD	<i>KRT14</i>	Unfavorable	Negative	Anti-cancer
		<i>GRIA1</i>	Favorable	Negative	Pro-tumorigenic
		<i>SEZ6L</i>	Favorable	Negative	Pro-tumorigenic
<i>SLC12A5</i>		Favorable	Negative	Pro-tumorigenic	
UCEC	<i>EMX2</i>	Favorable	Negative	Pro-tumorigenic	
AP-2γ	ESCA	<i>DPP6</i>	Unfavorable	Positive	Pro-tumorigenic
	LUAD	<i>NTSR1</i>	Unfavorable	Negative	Anti-cancer
		<i>PPEF1</i>	Unfavorable	Positive	Pro-tumorigenic
	PAAD	<i>ARX</i>	Favorable	Negative	Pro-tumorigenic
		<i>PTPRN</i>	Favorable	Negative	Pro-tumorigenic
	UCEC	<i>SCG5</i>	Favorable	Negative	Pro-tumorigenic
		<i>COL4A3</i>	Unfavorable	Negative	Anti-cancer

Table 2. Supposed role of AP-2α and AP-2γ in cancer based on genes that met the study requirements.

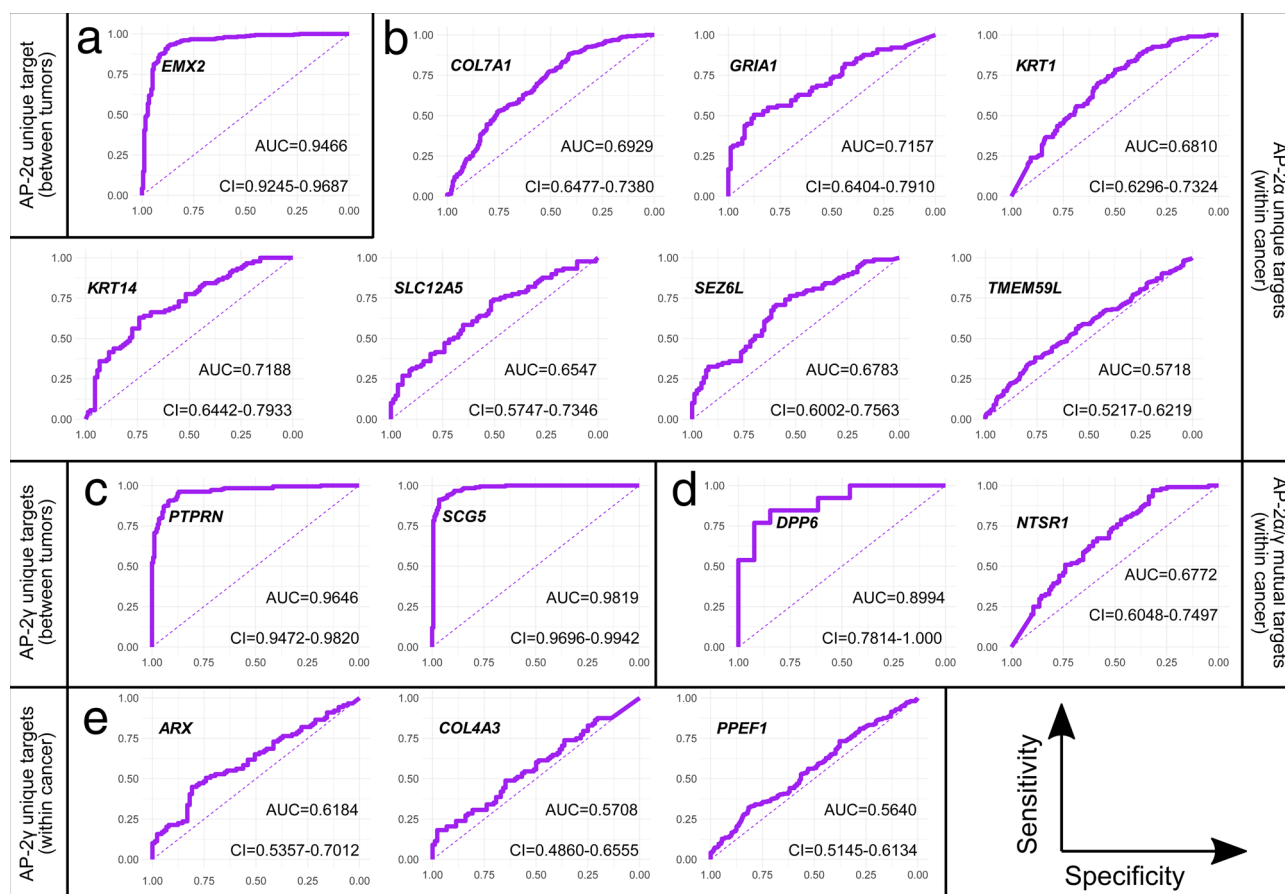


Figure 7. Receiver operating characteristic curves of genes that met the study requirements. (a) Unique for AP-2α (between-tumor comparison). (b) Unique for AP-2α (comparison within a specific cancer type). (c) Unique for AP-2γ (between-tumor comparison). (d) Mutual for AP-2α and AP-2γ. (e) Unique for AP-2γ (comparison within a specific cancer type).

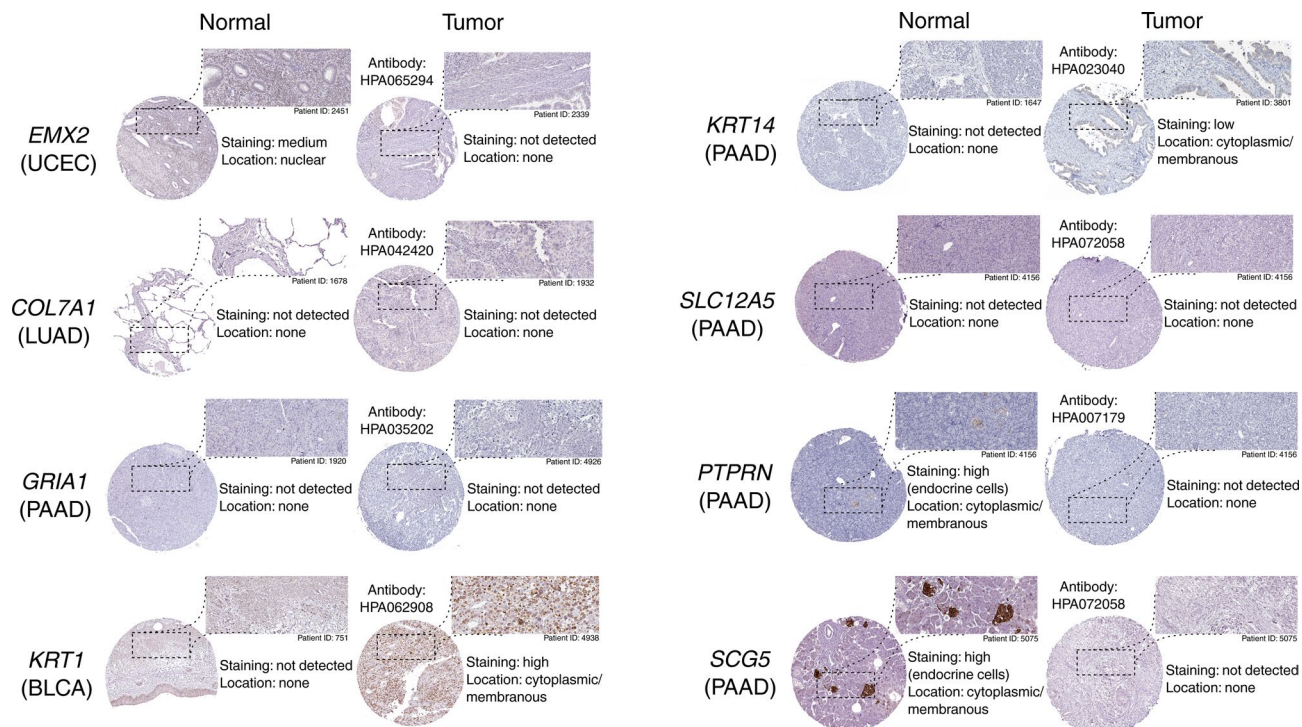


Figure 8. Representative immunostaining data for post-ROC genes from both normal and tumor tissues. For each gene, the cohort abbreviation for which it was identified throughout the study is given in brackets.

the tumor tissues for which they were identified throughout the study; thus, no comparison was possible. For the remainder, three genes (*SLC12A5*, *COL7A1*, *GRIA1*) showed no differences between tumor and normal tissue, while five of them (*EMX2*, *KRT1*, *KRT14*, *PTPRN*, *SCG5*) presented various staining.

The study also examined whether the expression of the post-ROC genes could co-depend on expression of other genes known to play a role in tumor progression (a group of five or six genes was selected for each tumor based on literature data). It was found that patients with high or low expression of each post-ROC gene presented varied expression of at least two progression-related genes (Fig. 9).

Discussion

Apart from their role in human development, AP-2 transcription factors are also known to influence carcinogenesis; as such, they have prognostic value for cancer patients^{11,12}. Once regarded as “undruggable” DNA-binding proteins, transcription factors are now the object of studies examining selective modulators of their activity^{8,13,14}. Among these factors, AP-2 is of particular significance since one of the members from this family, AP-2 δ (encoded by the *TFAP2D* gene), is one of less than three hundred candidate transcription factors suitable for studies focused on examining selective modulators of transcription factor activity¹⁵.

This suggests that other AP-2 representatives also deserve attention with regard to TF-based targeted therapy. Our previous research on AP-2 α and AP-2 γ in twenty-one tumors from TCGA identified evident differences in the expression of their target genes between tissues¹⁰. The present study examined tumors whose independence was not so obvious. The identification of AP-2 α/γ mutual or unique targets provided an indication of their relevance for a specific cancer and indirectly, whether AP-2 α or AP-2 γ might be worth consideration in forthcoming TF-based therapy.

To identify the most relevant AP-2 α/γ targets, mutual and unique genes were considered separately. Only fifteen genes met the initial requirements of this study; some were excluded based on ROC curves. Prior to ROC analysis, these fifteen genes were confronted with the literature data, revealing their potential in cancer therapy. Below, they are discussed in sequence.

One of methodology branches was to investigate unique AP-2 α/γ target genes between tumors. Only *EMX2* (for AP-2 α) and *PTPRN* or *SCG5* (both for AP-2 γ) were identified; however, they demonstrated very good discriminatory properties. *EMX2* encodes a homeodomain-containing transcription factor essential for growth and differentiation¹⁶. It is also a fundamental protein necessary for the development of the reproductive tract^{17,18}; it is therefore not surprising that it was found to be specific for uterine corpus endometrial carcinoma in the current study: *EMX2* expression was noted mainly within UCEC and to a small extent within BLCA. High expression was found to be favorable for survival in UCEC patients, which is in agreement with previous data indicating that it acts as a tumor suppressor in lung, kidney, colorectal, gastric cancers or in sarcoma^{16,19–22}. Moving forward, *PTPRN* encodes transmembrane receptor-type protein tyrosine phosphatase, mainly expressed in neuroendocrine tissues such as pancreas²³. Its expression was found to be PAAD-specific in this study. High *PTPRN* expression was found to be favorable for survival in pancreatic adenocarcinoma patients. This contradicts data

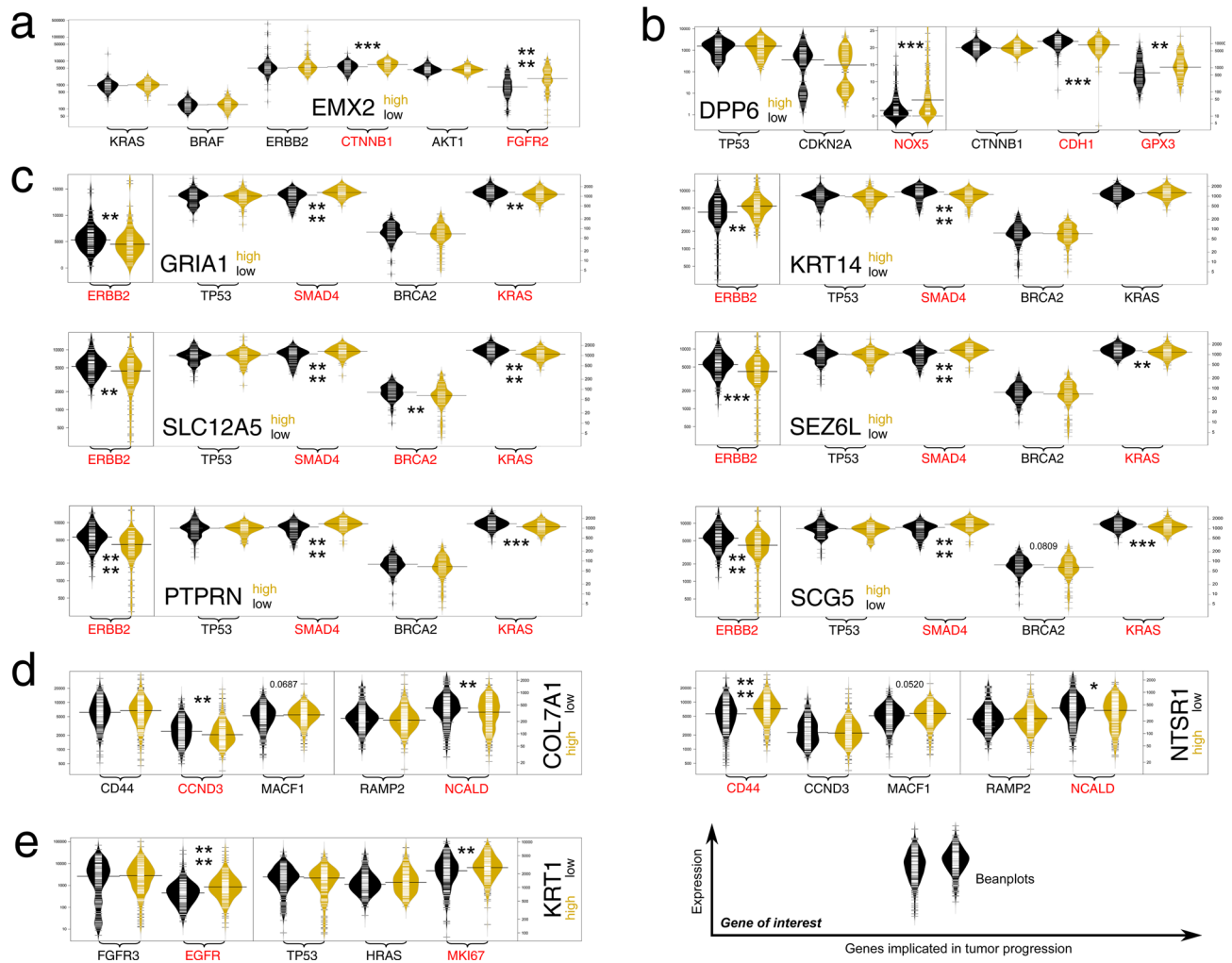


Figure 9. Co-dependence of post-ROC genes and tumor progression-related genes. Genes with an impact on patient survival in: (a) UCEC, (b) ESCA, (c) PAAD, (d) LUAD, (e) BLCA. In most cases, two separate scales (separated by a continuous line) are used on the Y axis due to large differences in expression level.

from other tumors; for example, its presence favors metastasis and migration promotion in LUAD²⁴, and it is associated with worse prognosis of patients with hepatocellular carcinoma²⁵ or progression of gastric cancer²⁶. However, little or no research has been performed on the role of *PTPRN* in PAAD, and it may be the case that the context is tissue-dependent, similar to *GLDC* gene function across tumors²⁷. Lastly, *SCG5* gene encodes secretogranin V, an essential chaperone involved in signaling that influences proliferation, among other things^{28,29}. *SCG5* was found to be implicated in polyposis syndrome, which is associated with colorectal cancer (CRC)³⁰. Our present findings indicate that *SCG5* expression was specific for the PAAD cohort. This is in line with data regarding ability of secretogranin V to regulate prohormone convertase 2 (PC2), a neuroendocrine-specific proteinase³¹. In addition, the higher *SCG5* expression was found to improve survival among PAAD patients, which is consistent with the literature³².

The other group of unique AP-2 α/γ targets identified throughout this study were derived from comparisons within a specific cancer. From genetic information of *COL7A1*, the alpha chain of basement-membrane protein, type VII collagen, is formed. Its high expression was found to be unfavorable for survival among lung adenocarcinoma patients. Despite the lack of literature data on the role of *COL7A1* in LUAD, similar observations have been made in patients with squamous-cell skin cancer³³, laryngeal cancer³⁴ and gastric cancer³⁵. In contrast, high expression of *GRIA1*, an ionotropic receptor involved in glutamate signaling, was found to be beneficial in PAAD; indeed, *GRIA1* has previously been included in a gene-based risk score system constructed for patients with that tumor^{36,37}. However, its function may depend on tissue type, since *GRIA1* was shown to promote tumor progression in glioma³⁸. Moving forward, *KRT1* encodes a member of the intermediate filament superfamily, i.e. the clade comprising the cytokeratins: common markers of differentiation, migration and proliferation in epithelial cells³⁹. *KRT1* is used to molecularly distinguish muscle-invasive urothelial carcinoma into luminal or basal subtypes⁴⁰; as such, the identification of this target as an unfavorable prognostic marker in BLCA is useful. *KRT1* was previously found to be involved in bladder tumorigenesis⁴¹, and to be associated with advanced tumor stage and worse prognosis in melanoma patients⁴². The other cytokeratin, *KRT14*, was also found to worsen the outcome in patients with PAAD. This is in line with data from ovarian and lung cancers^{43,44}; however, a study on

pancreatic carcinoma suggests that *KRT14* is not predictive of outcome (hazard ratio was similar as in our study but statistical significance was not met)⁴⁵. Another gene indicated in the PAAD cohort was *SEZ6L*, encoding a transmembrane protein implicated in signal transduction, protein–protein interactions and complement regulation (via inhibition of C3 convertases and promotion of C3b degradation)⁴⁶. High expression was favorable for PAAD patients, which is in line with other studies on pancreatic carcinoma⁴⁷ or lung cancer⁴⁸. Similarly, a potassium chloride cotransporter encoded by *SLC12A5*, also improved DFS in PAAD; however, no previous research exists on this tumor (various members of the solute carrier family have been discussed in pancreas neoplasm but *SLC12A5* is not among them⁴⁹). In fact, previous studies indicate that this gene promotes tumor invasion and metastasis in BLCA⁵⁰ or proliferation and G₁/S cell cycle transition in colon cancer⁵¹. The last target unique for AP-2 α was the transmembrane protein-encoding gene *TMEM59L* (also known as *C19orf4*). However, its role in LUSC and other cancers remains unclear⁵². The protein is believed to regulate apoptosis⁵³ and the GeneCards website indicates that *TMEM59L* modulates the glycosylation of amyloid precursor protein. The remaining genes from unique targets were found to be dependent on AP-2 γ . *ARX* encodes a transcription factor that is crucial in regulating the endocrine pancreas development; it is mainly expressed in the central nervous system, skeletal muscles and aforementioned lineage of pancreatic cells^{54,55}. This gene correlates with an aggressive course and frequent relapses in pancreatic neuroendocrine tumors^{56,57}. Our present survival analysis suggests that high *ARX* expression is favorable for PAAD patients, which might suggest a subtype-specific behavior that is worth investigation since pancreatic tumors arising from endocrine or epithelial portion have different properties^{58–60}. The prognostic outcome of *COL4A3* could also vary depending on cancer type. Our findings indicate that high expression was associated with shorter survival of UCEC patients. This corresponds to research on non-small cell lung (NSCLC) and breast cancers^{61,62} but contradicts data from HNSC⁶³. Finally, while the molecular function of phosphatase encoded by *PPEF1* is not known in detail, it has been connected to apoptosis regulation and the response to calcium (Ca²⁺ is also second messenger controlling cell death)⁶⁴. It exerts a tumorigenic role in breast cancer⁶⁵ and was found to be overexpressed in pancreatic carcinoma⁶⁶ which confirms both its prognostic value in LUAD (poorer survival when highly expressed) and literature data regarding lung cancer⁶⁴.

The presence of any mutual targets that are differently regulated by AP-2 α and AP-2 γ within a specific cancer type was also investigated. Only two genes were found to be inversely regulated by both transcription factors and met all the other requirements: *DPP6* (in ESCA cohort) and *NTSRI* (in LUAD). The first gene stores information about membrane glycoprotein of dipeptidyl peptidase IV family which regulates apoptosis, differentiation or proliferation^{67,68}. Our present findings indicate that high *DPP6* expression worsens survival of ESCA patients which is in line with previous data regarding colon cancer progression⁶⁷. However, as no DFS data was present in the validation cohort, the survival analysis of DFS with regard to *DPP6* level was validated using overall survival outcome. Interestingly, this gene was found to be overexpressed in long-term survivors of study on esophageal cancer compared to those with shorter survival⁶⁹, and *DPP6* hypomethylation or hypermethylation has been noted depending on cancer type⁶⁷. It is undoubtedly an important gene in esophagus neoplasms since, in addition to opposite regulation by AP-2 α / γ , its expression is regulated by ARID3A, ZNF354C: two out of five key transcription factors crucial for carcinogenesis and development of esophageal squamous cell carcinoma⁷⁰. Moreover, the gene encoding AP-2 α (*TFAP2A*) was also found to significantly correlate with longer survival rate⁷⁰. Considering how strong *TFAP2A* and *DPP6* correlated in this study ($R = -0.97$) and that AP-2 γ is suspected to contribute in esophageal cancer progression⁹, they clearly deserve further investigation in ESCA. Last but not least, *NTSRI* gene encodes the seven-transmembrane G-protein coupled receptor, through which neurotensin acts on proliferation, DNA synthesis or migration⁷¹. This receptor has been the subject of more study than other neurotensin receptors⁷²; it has been found to play a tumorigenic role in PAAD, HNSC, NSCLC or CRC^{73,74}. In LUAD, it has been found to correlate with poor prognosis and to participate in cancer progression^{71,75}; this corresponds to the survival analysis performed in this study. As such, *NTSRI* has often been proposed as a potential therapeutic or diagnostic target^{71,74,76}.

All genes with satisfactory AUC were subjected to analysis of IHC data. Their influence on tumor progression was also evaluated; each cancer was considered separately, with progression-related genes being selected for BLCA⁷⁷, ESCA⁷⁸, LUAD⁷⁹, PAAD^{80,81} and UCEC⁸² based on literature. The prognostic value of *EMX2* suggests it as a favorable marker for DFS and the staining confirms that it is present at a higher level in normal endometrial tissue than in tumor, as noted previously⁸³. Similarly, high *PTPRN* or *SCG5* expression was found to be favorable for PAAD patients; more intense staining was observed in normal pancreatic tissue, suggesting their expression is lowered during tumorigenesis. In addition, *SCG5* expression has been found to be decreased in primary pancreatic cancer, and even lower in metastatic carcinoma³². In contrast, *KRT1* or *KRT14* staining was more intense in BLCA or PAAD than in corresponding normal specimens, confirming their observed unfavorable impact on patient outcome. *KRT1* was found to be elevated in urospheres (which contain cancer initiating cells) compared to the parent non-tumorigenic UROtsa cell line, which was exposed to arsenite in order to acquire transformed cells⁸⁴. Likewise, basal keratins (including *KRT14*) are expressed in a subset of pancreatic ductal adenocarcinoma but are undetectable in normal pancreas⁴⁵.

The observations regarding the co-dependence of post-ROC genes and tumor progression-related genes are generally consistent with both prognostic outcome and IHC data; however, *EMX2* presented intriguing tendencies. Namely, while survival analysis and immunohistochemistry are both consistent with literature data suggesting *EMX2* as endometrial tumor suppressor, the *EMX2*^{high} group of patients were found to have higher expression of *CTNNB1* and *FGFR2*, two known oncogenes implicated in endometrial oncogenesis⁸². This clearly requires future investigation. Nevertheless, the three genes identified in PAAD for which IHC data was available (*PTPRN*, *SCG5*, *KRT14*) indicated that an important switch exists between *ERBB2* and *SMAD4* expression i.e. the oncogene and tumor suppressor, respectively^{80,81}. If the expression of *PTPRN* and *SCG5* (both favorable for prognosis) was high, *SMAD4* is elevated but *ERBB2* is lowered, similar to the *KRAS* oncogene. In contrast, when *KRT14* expression (being unfavorable) is high, the opposite tendency is seen for *ERBB2* and *SMAD4*. Similar

Cohort	Description
BLCA	Bladder urothelial carcinoma
CESC	Cervical and endocervical cancers
COAD	Colon adenocarcinoma
ESCA	Esophageal carcinoma
HNSC	Head and neck squamous cell carcinoma
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
PAAD	Pancreatic adenocarcinoma
READ	Rectum adenocarcinoma
STAD	Stomach adenocarcinoma
UCEC	Uterine corpus endometrial carcinoma

Table 3. Cohorts selected from previous study.

observations were made for the genes *SEZ6L*, *GRIA1* and *SLC12A5* in PAAD, whose impact on survival suggested anti-cancer properties but insufficient data or no differences were observed in IHC specimens. The same changes in *ERBB2*, *SMAD4* and *KRAS* level are observed between the “high” and “low” groups of these genes and conforms that of *PTPRN* and *SCG5*. The remaining genes in which IHC data was insufficient or staining was not detected, concern *DPP6* (for ESCA) and *COL7A1* or *NTSR1* (both for LUAD), all being unfavorable for patient outcome. High *DPP6* expression was found in the group that also demonstrated increased *NOX5* and *GPX3*; while both genes belong to the same biological pathway i.e. regulate reactive oxygen species levels, the former gene is upregulated during esophageal carcinogenesis while the latter is downregulated⁸⁵. Nevertheless, the metabolism of oxygen by-products is dysregulated and this is known to affect tumor progression^{86,87}. Undoubtedly, patients with high *DPP6* expression demonstrate altered adhesion, as *CDH1* expression is significantly decreased, which is associated with poor survival⁸⁸. In LUAD, “high” expressing groups of both *COL7A1* and *NTSR1* demonstrated reduced *NCALD*, the gene whose low expression worsens patient outcome⁸⁹. Furthermore, *CCND3*, whose high expression improves survival⁷⁹, was reduced in the *COL7A1*^{high} group while *CD44*, a promoter of KRAS-dependent lung tumorigenesis⁹⁰, was elevated in the *NTSR1*^{high} group. Lastly, *KRT1*^{high} BLCA patients had significantly higher expression of *EGFR* and *MKI67*, suggesting that proliferation is potentiated in this group; this complements the survival analysis (high *KRT1* is unfavorable in BLCA) and IHC data (*KRT1* staining is higher in cancer than normal specimens).

Finally, there is a need to determine whether AP-2 α / γ upregulate or downregulate all the above genes, as estimated by correlation analysis, and thus the role of AP-2 factors in particular cancer types. It appears that while both anti-cancer and pro-tumorigenic roles are possible, the latter is predominant (Table 2), suggesting AP-2 α / γ are potential candidates for cancer treatment. Moreover, the presence of “anti-cancer” next to “pro-tumorigenic” within a single tumor only complicates the final conclusion regarding AP-2 role in that cancer (Table 2; e.g. *KRT14* vs *GRIA1*, *SEZ6L*, *SLC12A5* for AP-2 α within PAAD or *NTSR1* vs *PPEF1* for AP-2 γ within LUAD). The only unequivocal “anti-cancer” cases are *DPP6* in ESCA (for AP-2 α) or *COL4A3* in UCEC (for AP-2 γ) but this cannot be concluded only on the basis of the single TF–target example. However, three genes per AP-2 factor were of favorable prognostic value for PAAD and all were negatively regulated by the TF, implying both AP-2 α and AP-2 γ could play tumorigenic role in this tumor. Our previous findings suggest that AP-2 α may have an oncogenic role in pancreatic cancer⁹; as such, further studies are needed of these two AP-2 factors in this tumor type. In addition, they may have different roles in other neoplasms and further research could confirm their value as novel candidate TFs suitable for targeting in cancer treatment, as aforementioned in terms of AP-2 δ ¹⁵.

Conclusions

Our findings indicate that genetic targets of AP-2 α and AP-2 γ differ between seemingly similar tumors. These differences can be of prognostic importance, being implicated in tumor progression, and they may be of value in targeted therapy. The study also paved the way for these two AP-2 transcription factors to be considered as candidates for developing TF-based cancer treatment.

Methods

Data collection, identification of AP-2 target genes, building trajectories. The expression and clinical data of patients (level 3 RNA-seqV2, RSEM normalized) from TCGA-dedicated GDAC Firehose Repository (gdac.broadinstitute.org), together with a list of AP-2 α / γ targets (combined from GTRD v19.10, TRANSFAC v2019.2 and TRRUST v2—see Supplementary File S3) were loaded back to R environment in the form of RData workspace (available at github.com/koldam/AP2-prognostic-significance). Thus, the entire workflow (with objects) of the Monocle3 R toolkit (cole-trapnell-lab.github.io/monocle3), performed in the former research¹⁰, was automatically exported. The additional (i.e. not previously performed) part within Monocle3 included e.g. the generation of subsets using `choose_cells()`. At first, the subset of the `cell_data_set` was created in order to include only selected tumors (Table 3). Further comparisons depended on methodology branch (Fig. 10). For example, if the research focused on identifying unique AP-2 α / γ target genes within a specific

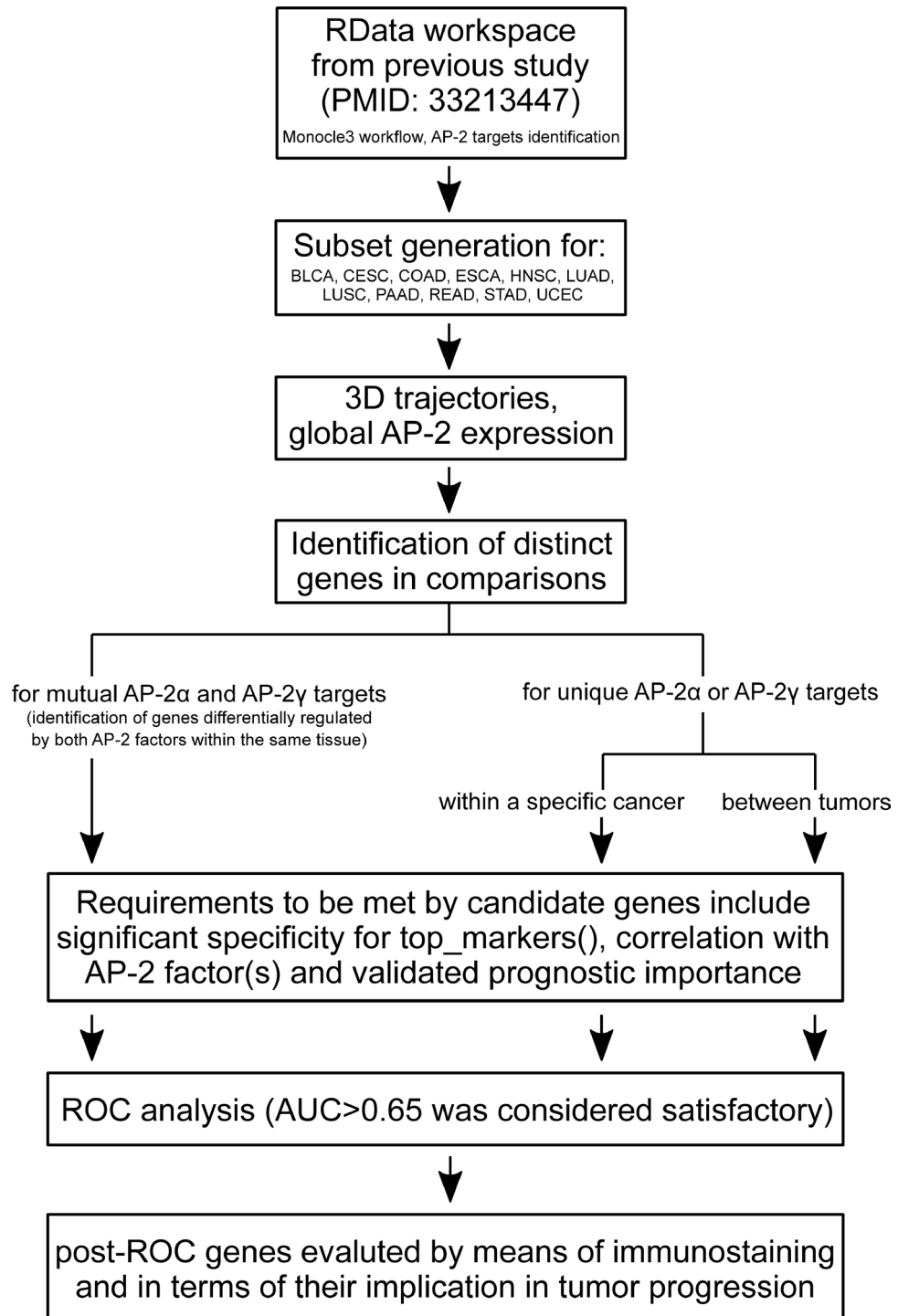


Figure 10. Methodology overview.

tumor, the remaining cohorts were temporarily excluded from the subset. The study was carried out in accordance with relevant guidelines/regulations.

The `plot_cells_3d()` was used to visualize the differences between tumors; this relied on the `learn_graph()` function (with `use_partition` parameter set as “TRUE”) that was preceded by the `reduce_dimension()` function (with `umap.metric` set as “cosine”) and `cluster_cells()`, both with the reduction method set on UMAP. Finally, the most specific genes (specificity > 0.6) in comparisons were listed using `top_markers()` with `fraction_expressing` ≥ 0.5 and `marker_test_p_value` < 0.05. For tumor vs tumor comparisons, gene expression was plotted on heatmaps generated with the use of the `DoHeatmap()` function (`scale_fill_viridis` option “E” was used for aesthetics) after `CreateSeuratObject()` was used within the Comprehensive R Archive Network (CRAN) Seurat R-package.

Analysis of correlation, prognostic importance (with validation) and classification model. Genes that met the requirement of specificity > 0.6 were individually correlated with gene encoding AP-2 α (*TFAP2A*), AP-2 γ (*TFAP2C*) or both (depending on whether the gene was unique or mutual AP-2 target). Correlation AnalyzeR (gccri.bishop-lab.uthscsa.edu/correlation-analyzer) was used to correlate genes using Pearson's correlation coefficient in desired tissue and sample type ("Gene vs gene" mode was used). Since this tool uses ARCHS4 repository as RNA-seq data source, this could advantageously show relationships independent of TCGA. Survival analysis performed in GEPIA2 (gepia2.cancer-pku.cn) was validated using separate web tools i.e. either pan-cancer RNA-seq KM plotter (kmplot.com) or LOGpC (bioinfo.henu.edu.cn/DatabaseList.jsp), depending on the data availability for specific cancer (e.g. for prognostic endpoints, disease-free or recurrence-free survival (DFS; RFS) was primarily used, with a few examples of disease-specific survival (DSS) and single overall survival (OS)). ROC curves were constructed to evaluate the genes not excluded in previous steps of analysis. Estimation of AUC and 95% Confidence Interval (CI) was done using pROC package with curve visualization made with ggroc and ggplot2 in R environment.

Evaluation of both immunostaining data and influence on tumor progression. Representative IHC data were obtained from publicly-available Human Protein Atlas (proteatlas.org); the same antibody for both normal and tumor specimens was selected. The "Tissue" or "Pathology" atlas was used for normal or tumor tissue data, respectively. Genes with prognostic significance were also analyzed on beanplots generated via the BoxPlotR (shiny.chemgrid.org/boxplotr), a web-tool which uses beanplot R-package. The median expression was used as a cut-off value for the post-ROC genes to identify differences in the expression of other genes representing a progression-related signature in a specific tumor. In most cases, two separate scales were used on the Y axis due to large differences in expression level.

Data availability

The datasets supporting the conclusions of this article are available in the GDAC Firehose repository (<https://gdac.broadinstitute.org/>) and GitHub (<https://github.com/koldam/AP2-prognostic-significance>).

Received: 24 November 2021; Accepted: 24 March 2022

Published online: 31 March 2022

References

- Wingender, E., Schoeps, T., Haubrock, M., Krull, M. & Donitz, J. TFClass: expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* **46**, D343–D347. <https://doi.org/10.1093/nar/gkx987> (2018).
- Hilger-Eversheim, K., Moser, M., Schorle, H. & Buettner, R. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis and cell-cycle control. *Gene* **260**, 1–12. [https://doi.org/10.1016/s0378-1119\(00\)00454-6](https://doi.org/10.1016/s0378-1119(00)00454-6) (2000).
- Eckert, D., Buhl, S., Weber, S., Jager, R. & Schorle, H. The AP-2 family of transcription factors. *Genome Biol.* **6**, 246. <https://doi.org/10.1186/gb-2005-6-13-246> (2005).
- Wu, H. R. & Zhang, J. AP-2 α expression in papillary thyroid carcinoma predicts tumor progression and poor prognosis. *Cancer Manag. Res.* **10**, 2615–2625. <https://doi.org/10.2147/CMAR.S167874> (2018).
- Perkins, S. M. *et al.* TFAP2C expression in breast cancer: correlation with overall survival beyond 10 years of initial diagnosis. *Breast Cancer Res. Treat.* **152**, 519–531. <https://doi.org/10.1007/s10549-015-3492-2> (2015).
- Wang, X. *et al.* TFAP2C promotes stemness and chemotherapeutic resistance in colorectal cancer via inactivating hippo signaling pathway. *J. Exp. Clin. Cancer Res.* **37**, 27. <https://doi.org/10.1186/s13046-018-0683-9> (2018).
- Bushweller, J. H. Targeting transcription factors in cancer - from undruggable to reality. *Nat. Rev. Cancer* **19**, 611–624. <https://doi.org/10.1038/s41568-019-0196-7> (2019).
- Henley, M. J. & Koehler, A. N. Advances in targeting "undruggable" transcription factors with small molecules. *Nat. Rev. Drug Discov.* **20**, 669–688. <https://doi.org/10.1038/s41573-021-00199-0> (2021).
- Kolat, D., Kaluzinska, Z., Bednarek, A. K. & Pluciennik, E. The biological characteristics of transcription factors AP-2 α and AP-2 γ and their importance in various types of cancers. *Biosci Rep.* <https://doi.org/10.1042/BSR20181928> (2019).
- Kolat, D., Kaluzinska, Z., Orzechowska, M., Bednarek, A. K. & Pluciennik, E. Functional genomics of AP-2 α and AP-2 γ in cancers: in silico study. *BMC Med. Genomics* **13**, 174. <https://doi.org/10.1186/s12920-020-00823-9> (2020).
- Pellikainen, J. M. & Kosma, V. M. Activator protein-2 in carcinogenesis with a special reference to breast cancer—a mini review. *Int. J. Cancer* **120**, 2061–2067. <https://doi.org/10.1002/ijc.22648> (2007).
- Beggs, A. D. *et al.* Methylation changes in the TFAP2E promoter region are associated with BRAF mutation and poorer overall & disease free survival in colorectal cancer. *Oncoscience* **2**, 508–516. <https://doi.org/10.18632/oncoscience.149> (2015).
- Shiroma, Y., Takahashi, R. U., Yamamoto, Y. & Tahara, H. Targeting DNA binding proteins for cancer therapy. *Cancer Sci.* **111**, 1058–1064. <https://doi.org/10.1111/cas.14355> (2020).
- Islam, Z. *et al.* Transcription factors: The fulcrum between cell development and carcinogenesis. *Front. Oncol.* **11**, 681377. <https://doi.org/10.3389/fonc.2021.681377> (2021).
- Lambert, M., Jambon, S., Depauw, S. & David-Cordonnier, M. H. Targeting transcription factors for cancer treatment. *Molecules* <https://doi.org/10.3390/molecules23061479> (2018).
- Aykut, B. *et al.* EMX2 gene expression predicts liver metastasis and survival in colorectal cancer. *BMC Cancer* **17**, 555. <https://doi.org/10.1186/s12885-017-3556-2> (2017).
- Daftary, G. S. & Taylor, H. S. EMX2 gene expression in the female reproductive tract and aberrant expression in the endometrium of patients with endometriosis. *J. Clin. Endocrinol. Metab.* **89**, 2390–2396. <https://doi.org/10.1210/jc.2003-031389> (2004).
- Lagana, A. S., Sturlese, E., Retto, G., Sofo, V. & Triolo, O. Interplay between misplaced mullerian-derived stem cells and peritoneal immune dysregulation in the pathogenesis of endometriosis. *Obstet. Gynecol. Int.* **2013**, 527041. <https://doi.org/10.1155/2013/527041> (2013).
- Okamoto, J. *et al.* EMX2 is epigenetically silenced and suppresses growth in human lung cancer. *Oncogene* **29**, 5969–5975. <https://doi.org/10.1038/ncr.2010.330> (2010).
- Jiang, H., Chen, H., Wan, P., Song, S. & Chen, N. Downregulation of enhancer RNA EMX2OS is associated with poor prognosis in kidney renal clear cell carcinoma. *Aging (Albany NY)* **12**, 25865–25877. <https://doi.org/10.18632/aging.202151> (2020).
- Li, J. *et al.* Adenoviral delivery of the EMX2 gene suppresses growth in human gastric cancer. *PLoS ONE* **7**, e45970. <https://doi.org/10.1371/journal.pone.0045970> (2012).

22. Jimenez-Garcia, M. P., Lucena-Cacace, A., Otero-Albiol, D. & Carnero, A. Regulation of sarcomagenesis by the empty spiracles homeobox genes EMX1 and EMX2. *Cell Death Dis.* **12**, 515. <https://doi.org/10.1038/s41419-021-03801-w> (2021).
23. Gomi, H., Kubota-Murata, C., Yasui, T., Tsukise, A. & Torii, S. Immunohistochemical analysis of IA-2 family of protein tyrosine phosphatases in rat gastrointestinal endocrine cells. *J. Histochem. Cytochem.* **61**, 156–168. <https://doi.org/10.1369/0022155412466872> (2013).
24. Song, X. *et al.* Overexpression of PTPRN promotes metastasis of lung adenocarcinoma and suppresses NK cell cytotoxicity. *Front. Cell Dev. Biol.* **9**, 622018. <https://doi.org/10.3389/fcell.2021.622018> (2021).
25. Zhangyuan, G. *et al.* Prognostic value of phosphotyrosine phosphatases in hepatocellular carcinoma. *Cell Physiol. Biochem.* **46**, 2335–2346. <https://doi.org/10.1159/000489625> (2018).
26. Stebbing, J. *et al.* The regulatory roles of phosphatases in cancer. *Oncogene* **33**, 939–953. <https://doi.org/10.1038/onc.2013.80> (2014).
27. Kaluzinska, Z., Kolat, D., Bednarek, A. K. & Pluciennik, E. PLEK2, RRM2, GCSH: A novel WWOX-dependent biomarker triad of glioblastoma at the crossroads of cytoskeleton reorganization and metabolism alterations. *Cancers (Basel)*. <https://doi.org/10.3390/cancers13122955> (2021).
28. Yang, H. *et al.* Meta-analysis of the rs4779584 polymorphism and colorectal cancer risk. *PLoS ONE* **9**, e89736. <https://doi.org/10.1371/journal.pone.0089736> (2014).
29. Segerstolpe, A. *et al.* Single-cell transcriptome profiling of human pancreatic islets in health and Type 2 diabetes. *Cell Metab.* **24**, 593–607. <https://doi.org/10.1016/j.cmet.2016.08.020> (2016).
30. Yusuf, I. *et al.* Genetic risk factors for colorectal cancer in multiethnic Indonesians. *Sci. Rep.* **11**, 9988. <https://doi.org/10.1038/s41598-021-88805-4> (2021).
31. Portela-Gomes, G. M., Grimelius, L. & Stridsberg, M. Prohormone convertases 1/3, 2, furin and protein 7B2 (Secretogranin V) in endocrine cells of the human pancreas. *Regul. Pept.* **146**, 117–124. <https://doi.org/10.1016/j.regpep.2007.09.019> (2008).
32. Xu, J. S., Liao, K. L., Wang, X., He, J. & Wang, X. Z. Combining bioinformatics techniques to explore the molecular mechanisms involved in pancreatic cancer metastasis and prognosis. *J. Cell Mol. Med.* **24**, 14128–14138. <https://doi.org/10.1111/jcmm.16023> (2020).
33. Martins, V. L. *et al.* Increased invasive behaviour in cutaneous squamous cell carcinoma with loss of basement-membrane type VII collagen. *J. Cell Sci.* **122**, 1788–1799. <https://doi.org/10.1242/jcs.042895> (2009).
34. Tao, Y. *et al.* Identification of novel enriched recurrent chimeric COL7A1-UCN2 in human laryngeal cancer samples using deep sequencing. *BMC Cancer* **18**, 248. <https://doi.org/10.1186/s12885-018-4161-8> (2018).
35. Oh, S. E. *et al.* Prognostic value of highly expressed Type VII collagen (COL7A1) in patients with gastric cancer. *Pathol. Oncol. Res.* **27**, 1609860. <https://doi.org/10.3389/pore.2021.1609860> (2021).
36. Hu, H. *et al.* Hypoxia-inducible factors enhance glutamate signaling in cancer cells. *Oncotarget* **5**, 8853–8868. <https://doi.org/10.18632/oncotarget.2593> (2014).
37. Dou, D., Yang, S. & Zhang, J. Prognostic prediction of a 12-methylation gene-based risk score system on pancreatic adenocarcinoma. *Oncol. Lett.* **20**, 85–98. <https://doi.org/10.3892/ol.2020.11575> (2020).
38. de Groot, J. F., Piao, Y., Lu, L., Fuller, G. N. & Yung, W. K. Knockdown of GluR1 expression by RNA interference inhibits glioma proliferation. *J. Neurooncol.* **88**, 121–133. <https://doi.org/10.1007/s11060-008-9552-2> (2008).
39. Palko, E., Poliska, S., Sziklai, I. & Penyige, A. Analysis of KRT1, KRT10, KRT19, TP53 and MMP9 expression in pediatric and adult cholesteatoma. *PLoS ONE* **13**, e0200840. <https://doi.org/10.1371/journal.pone.0200840> (2018).
40. Iyyanki, T. *et al.* Subtype-associated epigenomic landscape and 3D genome structure in bladder cancer. *Genome Biol.* **22**, 105. <https://doi.org/10.1186/s13059-021-02325-y> (2021).
41. Wu, J. *et al.* Low expression of Keratin17 is related to poor prognosis in bladder cancer. *Oncotargets Ther.* **14**, 577–587. <https://doi.org/10.2147/OTT.S287891> (2021).
42. Han, W., Hu, C., Fan, Z. J. & Shen, G. L. Transcript levels of keratin 1/5/6/14/15/16/17 as potential prognostic indicators in melanoma patients. *Sci. Rep.* **11**, 1023. <https://doi.org/10.1038/s41598-020-80336-8> (2021).
43. Bilandzic, M. *et al.* Keratin-14 (KRT14) positive leader cells mediate mesothelial clearance and invasion by ovarian cancer cells. *Cancers (Basel)*. <https://doi.org/10.3390/cancers11091228> (2019).
44. Dong, Y. *et al.* Desmoglein 3 and Keratin 14 for distinguishing between lung adenocarcinoma and lung squamous cell carcinoma. *Oncotargets Ther.* **13**, 11111–11124. <https://doi.org/10.2147/OTT.S270398> (2020).
45. Martinelli, P. *et al.* GATA6 regulates EMT and tumour dissemination, and is a marker of response to adjuvant chemotherapy in pancreatic cancer. *Gut* **66**, 1665–1676. <https://doi.org/10.1136/gutjnl-2015-311256> (2017).
46. Pìgoni, M. *et al.* Seizure protein 6 and its homolog seizure 6-like protein are physiological substrates of BACE1 in neurons. *Mol. Neurodegener.* **11**, 67. <https://doi.org/10.1186/s13024-016-0134-z> (2016).
47. Liu, Y., Zhu, D., Xing, H., Hou, Y. & Sun, Y. A 6gene risk score system constructed for predicting the clinical prognosis of pancreatic adenocarcinoma patients. *Oncol. Rep.* **41**, 1521–1530. <https://doi.org/10.3892/or.2019.6979> (2019).
48. Nishioka, M. *et al.* Identification of a 428-kb homozygously deleted region disrupting the SEZ6L gene at 22q12.1 in a lung cancer cell line. *Oncogene* **19**, 6251–6260. <https://doi.org/10.1038/sj.onc.1204031> (2000).
49. Xie, J., Zhu, X. Y., Liu, L. M. & Meng, Z. Q. Solute carrier transporters: potential targets for digestive system neoplasms. *Cancer Manag. Res.* **10**, 153–166. <https://doi.org/10.2147/CMAR.S152951> (2018).
50. Liu, J. Y. *et al.* Solute carrier family 12 member 5 promotes tumor invasion/metastasis of bladder urothelial carcinoma by enhancing NF-kappaB/MMP-7 signaling pathway. *Cell Death Dis.* **8**, e2691. <https://doi.org/10.1038/cddis.2017.118> (2017).
51. Xu, L. *et al.* Increased expression of Solute carrier family 12 member 5 via gene amplification contributes to tumour progression and metastasis and associates with poor survival in colorectal cancer. *Gut* **65**, 635–646. <https://doi.org/10.1136/gutjnl-2014-308257> (2016).
52. Schmit, K. & Michiels, C. TMEM proteins in cancer: A review. *Front. Pharmacol.* **9**, 1345. <https://doi.org/10.3389/fphar.2018.01345> (2018).
53. Mannherz, O., Mertens, D., Hahn, M. & Lichter, P. Functional screening for proapoptotic genes by reverse transfection cell array technology. *Genomics* **87**, 665–672. <https://doi.org/10.1016/j.ygeno.2005.12.009> (2006).
54. Courtney, M., Rabe, T., Collombat, P. & Mansouri, A. Pax4 and Arx represent crucial regulators of the development of the endocrine pancreas. *New J. Sci.* **1–6**, 2014. <https://doi.org/10.1155/2014/981569> (2014).
55. Balic, A., Dorado, J., Alonso-Gomez, M. & Heeschen, C. Stem cells as the root of pancreatic ductal adenocarcinoma. *Exp. Cell Res.* **318**, 691–704. <https://doi.org/10.1016/j.yexcr.2011.11.007> (2012).
56. Hackeng, W. M. *et al.* Non-functional pancreatic neuroendocrine tumours: ATRX/DAXX and alternative lengthening of telomeres (ALT) are prognostically independent from ARX/PDX1 expression and tumour size. *Gut* <https://doi.org/10.1136/gutjnl-2020-322595> (2021).
57. Hackeng, W. M. *et al.* Assessment of ARX expression, a novel biomarker for metastatic risk in pancreatic neuroendocrine tumors, in endoscopic ultrasound fine-needle aspiration. *Diagn. Cytopathol.* **48**, 308–315. <https://doi.org/10.1002/dc.24368> (2020).
58. Starzynska, T. *et al.* Differences between well-differentiated neuroendocrine tumors and ductal adenocarcinomas of the pancreas assessed by multi-omics profiling. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms21124470> (2020).
59. Pulvirenti, A. *et al.* Limited role of Chromogranin A as clinical biomarker for pancreatic neuroendocrine tumors. *HPB (Oxford)* **21**, 612–618. <https://doi.org/10.1016/j.hpb.2018.09.016> (2019).

60. Stanculeanu, D. L. *et al.* Adenocarcinoma versus pancreatic neuroendocrine tumor—Case report. *Rom. J. Morphol. Embryol.* **58**, 1091–1097 (2017).
61. Jiang, C. P. *et al.* High COL4A3 expression correlates with poor prognosis after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. *Tumour Biol.* **34**, 415–420. <https://doi.org/10.1007/s13277-012-0565-2> (2013).
62. Yadav, V. K. *et al.* Computational analysis for identification of the extracellular matrix molecules involved in endometrial cancer progression. *PLoS ONE* **15**, e0231594. <https://doi.org/10.1371/journal.pone.0231594> (2020).
63. Yang, X., Wu, Q., Wu, F. & Zhong, Y. Differential expression of COL4A3 and collagen in upward and downward progressing types of nasopharyngeal carcinoma. *Oncol. Lett.* **21**, 223. <https://doi.org/10.3892/ol.2021.12484> (2021).
64. Park, S. Y. *et al.* Protein serine/threonine phosphatase PPEF-1 suppresses genotoxic stress response via dephosphorylation of PDCD5. *Sci. Rep.* **7**, 39222. <https://doi.org/10.1038/srep39222> (2017).
65. Ye, T. *et al.* The clinical significance of PPEF1 as a promising biomarker and its potential mechanism in breast cancer. *Onco Targets Ther.* **13**, 199–214. <https://doi.org/10.2147/OTT.S229432> (2020).
66. Hang, J. *et al.* The role of phosphoprotein phosphatases catalytic subunit genes in pancreatic cancer. *Biosci. Rep.* <https://doi.org/10.1042/BSR20203282> (2021).
67. Sheikh, M. A. *et al.* Epigenetic regulation of Dpp6 expression by Dnmt3b and its novel role in the inhibition of RA induced neuronal differentiation of P19 cells. *PLoS ONE* **8**, e55826. <https://doi.org/10.1371/journal.pone.0055826> (2013).
68. Neumiller, J. J., Wood, L. & Campbell, R. K. Dipeptidyl peptidase-4 inhibitors for the treatment of type 2 diabetes mellitus. *Pharmacotherapy* **30**, 463–484. <https://doi.org/10.1592/phco.30.5.463> (2010).
69. Ashida, A. *et al.* Expression profiling of esophageal squamous cell carcinoma patients treated with definitive chemoradiotherapy: clinical implications. *Int. J. Oncol.* **28**, 1345–1352 (2006).
70. Zhang, Y. *et al.* Identification of the key transcription factors in esophageal squamous cell carcinoma. *J. Thorac. Dis.* **10**, 148–161. <https://doi.org/10.21037/jtd.2017.12.27> (2018).
71. Nikolaou, S. *et al.* The role of Neurotensin and its receptors in non-gastrointestinal cancers: a review. *Cell Commun. Signal* **18**, 68. <https://doi.org/10.1186/s12964-020-00569-y> (2020).
72. Wu, Z., Martinez-Fong, D., Tredaniel, J. & Forgez, P. Neurotensin and its high affinity receptor 1 as a potential pharmacological target in cancer therapy. *Front Endocrinol (Lausanne)* **3**, 184. <https://doi.org/10.3389/fendo.2012.00184> (2012).
73. Dupouy, S. *et al.* The potential use of the neurotensin high affinity receptor 1 as a biomarker for cancer progression and as a component of personalized medicine in selective cancers. *Biochimie* **93**, 1369–1378. <https://doi.org/10.1016/j.biochi.2011.04.024> (2011).
74. Kim, J. T., Weiss, H. L. & Evers, B. M. Diverse expression patterns and tumorigenic role of neurotensin signaling components in colorectal cancer cells. *Int. J. Oncol.* **50**, 2200–2206. <https://doi.org/10.3892/ijo.2017.3990> (2017).
75. Alifano, M. *et al.* Neurotensin receptor 1 determines the outcome of non-small cell lung cancer. *Clin. Cancer Res.* **16**, 4401–4410. <https://doi.org/10.1158/1078-0432.CCR-10-0659> (2010).
76. Akter, H., Yoon, J. H., Yoo, Y. S. & Kang, M. J. Validation of neurotensin receptor 1 as a therapeutic target for gastric cancer. *Mol. Cells* **41**, 591–602. <https://doi.org/10.14348/molcells.2018.0025> (2018).
77. Li, Y. *et al.* Frontiers in bladder cancer genomic research. *Front. Oncol.* **11**, 670729. <https://doi.org/10.3389/fonc.2021.670729> (2021).
78. Testa, U., Castelli, G. & Pelosi, E. Esophageal cancer: Genomic and molecular characterization, stem cell compartment and clonal evolution. *Medicines (Basel)*. <https://doi.org/10.3390/medicines4030067> (2017).
79. Song, Z., Zhang, Y., Chen, Z. & Zhang, B. Identification of key genes in lung adenocarcinoma based on a competing endogenous RNA network. *Oncol. Lett.* **21**, 60. <https://doi.org/10.3892/ol.2020.12322> (2021).
80. Cowan, R. W. & Maitra, A. Genetic progression of pancreatic cancer. *Cancer J.* **20**, 80–84. <https://doi.org/10.1097/PPO.0000000000000011> (2014).
81. Hruban, R. H., Goggins, M., Parsons, J. & Kern, S. E. Progression model for pancreatic cancer. *Clin. Cancer Res.* **6**, 2969–2972 (2000).
82. Okuda, T. *et al.* Genetics of endometrial cancers. *Obstet. Gynecol. Int.* **2010**, 984013. <https://doi.org/10.1155/2010/984013> (2010).
83. Qiu, H. *et al.* EMX2 is downregulated in endometrial cancer and correlated with tumor progression. *Int. J. Gynecol. Pathol.* **32**, 193–198. <https://doi.org/10.1097/PGP.0b013e31825d8049> (2013).
84. Hoggarth, Z. E. *et al.* Enrichment of genes associated with squamous differentiation in cancer initiating cells isolated from urothelial cells transformed by the environmental toxicant arsenite. *Toxicol. Appl. Pharmacol.* **374**, 41–52. <https://doi.org/10.1016/j.taap.2019.04.021> (2019).
85. Kalatskaya, I. Overview of major molecular alterations during progression from Barrett's esophagus to esophageal adenocarcinoma. *Ann. N. Y. Acad. Sci.* **1381**, 74–91. <https://doi.org/10.1111/nyas.13134> (2016).
86. Klaunig, J. E., Kamendulis, L. M. & Hocevar, B. A. Oxidative stress and oxidative damage in carcinogenesis. *Toxicol. Pathol.* **38**, 96–109. <https://doi.org/10.1177/0192623309356453> (2010).
87. Kryston, T. B., Georgiev, A. B., Pissis, P. & Georgakilas, A. G. Role of oxidative stress and DNA damage in human carcinogenesis. *Mutat. Res.* **711**, 193–201. <https://doi.org/10.1016/j.mrfmmm.2010.12.016> (2011).
88. Ishiguro, H. *et al.* Decreased expression of CDH1 or CTNBN1 affects poor prognosis of patients with esophageal cancer. *World J. Surg. Oncol.* **14**, 240. <https://doi.org/10.1186/s12957-016-0956-8> (2016).
89. Chen, Y. *et al.* Clinicopathological and prognostic significance of NCALD protein expression in lung adenocarcinoma. **2019**, 752–760 (2019).
90. Zhao, P. *et al.* CD44 promotes Kras-dependent lung adenocarcinoma. *Oncogene* **32**, 5186–5190. <https://doi.org/10.1038/onc.2012.542> (2013).

Author contributions

D.K. and A.K.B. conceptualized the article. D.K. and A.K.B. established the methodology. D.K. was responsible for software. A.K.B. and E.P. supervised the article. D.K. and Ž.K. visualized the results. D.K. wrote the original draft. D.K., Ž.K., A.K.B. and E.P. reviewed and edited the article. All authors read and approved the final manuscript.

Funding

This research was funded by Medical University of Lodz, grant number: 503/0-078-02/503-01-001-19-00. The funding body had no role in study design, collection, analysis, and interpretation of data and in writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09494-1>.

Correspondence and requests for materials should be addressed to D.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022