Cluster Analysis According to Immunohistochemistry is a Robust Tool for Non–Small Cell Lung Cancer and Reveals a Distinct, Immune Signature-defined Subgroup

William Sterlacci, MD, Michael Fiegl, MD,†‡ Darius Juskevicius, PhD,§ and Alexandar Tzankov, MD§*

Abstract: Clustering in medicine is the subgrouping of a cohort according to specific phenotypical or genotypical traits. For breast cancer and lymphomas, clustering by gene expression profiles has already resulted in important prognostic and predictive subgroups. For non–small cell lung cancer (NSCLC), however, little is known. We performed a cluster analysis on a cohort of 365 surgically resected, well-documented NSCLC patients, which was followed-up for a median of 62 months, incorporating 70 expressed proteins and several genes. Our data reveal that tumor grading by architecture is significant, that large cell carcinoma is likely not a separate entity, and that an immune signature cluster exists. For squamous cell carcinomas, a prognostically relevant cluster with poorer outcome was found, defined by a high CD4/CD8 ratio and lower presence of granzyme B+ tumor-infiltrating lymphocytes (TIL). This study shows that clustering analysis is a useful tool for verifying established characteristics and generating new insights for NSCLC. Importantly, for one "immune signature" cluster, the signature of the TIL (especially the amount of CD8+ TIL) was more crucial than the histologic or any other phenotypical aspect. This may be an important finding toward explaining why only a fraction of eligible patients respond to immunomodulating anticancer therapies.

Key Words: non–small cell lung cancer, cluster analysis, immunohistochemistry

(*Appl Immunohistochem Mol Morphol* 2019;00:000–000)

In medicine, clustering is known as the grouping of objects by calculating similarities; thus, objects in one cluster are more closely related than those found in other clusters. Especially in association with high-throughput protein and gene analyses, the utilization of clustering has become increasingly popular for characterizing different subgroups of cancer. Particularly in breast cancer and in lymphomas, cluster analysis has been reported as a powerful tool for discovering different classes.[1,2] Subclassification of heterogenous cohorts makes it possible to identify distinct prognostic or predictive subgroups that are, for example, potentially linked to specific treatment outcomes. In breast cancer, gene expression analyses based on clustering have made it possible to distribute morphologically similar tumors into groups showing different prognosis as well as different treatment responses.[3,4] In the future, such expression profiles will likely also be incorporated into the tumor classification systems.

The proteins that are present in tumor cells are defined by individual genetic programming. On the one hand, this programming underlies intrinsic tumor-associated factors, such as mutations (point mutations and copy number variations), especially of genes encoding transcription factors or genes involved in signaling cascades [eg, anaplastic lymphoma kinase (*ALK*), epidermal growth factor receptor (*EGFR*), hepatocyte growth factor receptor (*MET*), and *RAS*-family members]. In contrast, extrinsic factors are also influential, including interaction with immune cells, for example, programmed death 1—programmed death ligand 1 (PD1-PDL1) or tumor-infiltrating fibroblasts. Usually not the expression of a single protein, but rather of a whole group of proteins (ie, cluster) is involved. Using mass spectrometry, it would be possible to analyze the entire protein expression profile of a tumor sample; however, this method cannot differentiate the exact origin of these proteins [ie, tumor cells, tumor-infiltrating lymphocytes (TIL), macrophages, fibroblasts, endothelial cells, etc.].

Gene expression profiling data in non–small cell lung cancer (NSCLC) also point toward the possibility of subclassification and prognostic importance; however, this issue does not yet seem as relevant compared with breast cancer.[5] Nevertheless, profiling data of NSCLC contributed to pathogenesis understanding, showed the possibility of further subgrouping of the histologic types, and can predict disease-free and overall survival.[6,7]

Importantly, clustering analyses are also possible with immunohistochemical expression profiles and have been performed in several tumors including lymphomas and breast cancer.[1,8,9] For NSCLC, however, there are only a few reports on small cohorts with few analyzed parameters, or limited to a single stage.[10,11]

Received for publication September 24, 2018; accepted January 30, 2019.

From the *Institute of Pathology, Hospital Bayreuth, Preuschwitzerstrasse, Bayreuth, Germany; †Department of Internal Medicine, Division of Hematology and Oncology, Medical University Innsbruck, Anichstrasse, Innsbruck; ‡Department of Internal Medicine, Oncology and Geriatrics, Hospital Hochrum, Lärchenstrasse, Rum, Austria; and §Institute of Pathology, University Hospital Basel, Schoenbeinstrasse, Basel, Switzerland.

The authors declare no conflict of interest.

Reprints: William Sterlacci, MD, Institute of Pathology, Hospital Bayreuth, Preuschwitzerstrasse 101, Bayreuth 95445, Germany (e-mail: william.sterlacci@klinikum-bayreuth.de).

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 1. Antibodies Applied and Cut-off Scores

Antibodies	Source and Clone or ID	Dilution	Cut-off Score
ABCG5	Sigma HPA016514	1:200*	Any expression in tumor cells
ALDH1	Abcam EP1933Y	1:200	Any expression in tumor cells
ALK	Ventana ALK01	Ready to use	Any expression in tumor cells
Ber-Ep4	Ventana Ber-EP4	Ready to use	> 19% positive tumor cells
Cbl-b	Abcam 246C5a	1:50	> 3 TILs/mm^2
CD4	Cell Marque SP35	1:100	> 53 TILs/mm^2
CD8	DAKO C8/144B	1:400	> 35 TILs/mm^2
CD34	Ventana QBEnd/10	Ready to use	Any expression in tumor cells
CD24	ThermoFisher SN3b	1:20	> 9% positive tumor cells
CD44	Ventana SP37	Ready to use	Any expression in tumor cells
CD44v6	Medical Systems ABIN 123880	1:100	> 8% positive tumor cells
CD56 (NCAM)	Ventana MRQ-42	Ready to use	Any expression in tumor cells
CD95	Novocastra GM30	1:400	Any expression in tumor cells
CD105	Novocastra 4G11	1:80*	Any expression in tumor cells
CD166	Novocastra MOG/07	1:200*	> 9% positive tumor cells
Chromogranin A	Neomarkers MS-382-P	1:400	Any expression in tumor cells
CXCR4	Abcam ab2074	1:50	> 1.5% positive tumor cells
CXCR4 phosphorylated	Abcam ab74012	1:100	> 3% positive tumor cells
Cyclin D1	ThermoFisher DCS-6	1:50	> 15% positive tumor cells
Cyclin D2	Santa Cruz sc-181	1:8000	Any expression in tumor cells
Cyclin D3	Novocastra DCS-22	1:80	> 3.5% positive tumor cells
Cyclin E	Neomarkers 13A3	1:20	> 1% positive tumor cells
Cytokeratin 5/6	Ventana D5/16B4	Ready to use	Any expression in tumor cells
Cytokeratin 7	Progen Ks7.18	1:50†	Any expression in tumor cells
Cytokeratin 34βE12	Ventana 34bE12	Ready to use	Any expression in tumor cells
D2-40	Ventana D2-40	Ready to use	Any expression in tumor cells
E-cadherin	Ventana EP700Y	Ready to use	Any expression in tumor cells
EGFR	Ventana 3C6	Ready to use†	Any expression in tumor cells
ESA	Novocastra VU-1D9	1:800	Any expression in tumor cells
FoxP3	Abcam mAbcam 22510	1:50	Any TILs
Granzyme B	Novocastra 11F1	1:100	Any TILs
Ki67	DAKO MIB-1	1:100	> 3%
Mast cell tryptase	DAKO AA1	1:100	> 10 TI mast cells/mm^2
Mel-CAM	Novocastra N1238	1:25	Any expression in tumor cells
MET	Ventana SP44	Ready to use	> 49% positive tumor cells
Moc31	DAKO MOC-31	Ready to use	> 49% positive tumor cells
Mucin1	Ventana H23	Ready to use	Any expression in tumor cells
Mucin2	Ventana MRQ-18	Ready to use	Any expression in tumor cells
Mucin4	Biocare 8G-7	1100	Any expression in tumor cells
Mucin5AC	Ventana MRQ-19	Ready to use	Any expression in tumor cells
Mucin6	Ventana MRQ-20	Ready to use	Any expression in tumor cells
Nestin	AbD Serotec 10C2	1:200	Any expression in tumor cells
Neuron-specific enolase	Ventana MRQ-55	Ready to use	Any expression in tumor cells
OCT4	Ventana MRQ-10	Ready to use	Any expression in tumor cells
OPN	Novocastra OP3N	1:50	> 3.5% positive tumor cells
p16	CINtec E6H4	Ready to use	> 9% positive tumor cells
p21	ThermoFisher HZ52	1:400	> 15% positive tumor cells
p27	Neomarkers DCS-72.F6	1:250	> 45% nuclear positive tumor cells
p63	Ventana 4A4	Ready to use	Any expression in tumor cells
pAkt	Abcam ab8932	1:450	Any expression in tumor cells
PD1	Cell Marque NAT105	1:50	> 14 TILs/mm^2
PDL1	Cell signaling E1L3N	1:50	> 4% positive tumor cells
Pgp	Enzo Lifesciences P3II-26	1:4000*	> 0.5% positive tumor cells
PGP 9.5	Novocastra 10A1	1:20	Any expression in tumor cells
PI3K	BD Bioscience 4/PI3-Kinase	1:1000*	Any expression in tumor cells
pSTAT3	Cell Signaling D3A7	1:50*	Any expression in tumor cells
PTEN	Cell signaling 138G6	1:200	< 10% positive tumor cells
RHAMM	Novocastra 2D6	1:25*	> 4% positive tumor cells
SDF-1	Abcam ab135949	1:20	> 1.4% positive tumor cells
Snail	Abgent AP2054a	1:50*	Any expression in tumor cells
SOX2	Abcam EPR3131	1:100*	> 2% positive tumor cells
Synaptophysin	Novocastra 27G12	1:100	Any expression in tumor cells
TGF-β	Acris DM1047	1:50	> 3 TILs/mm^2, any expression in tumor cells
TIA1	Immunotech 2G9A10F5	1:100	> 7 TILs/mm^2
TTF1	Neomarkers 8G7G3/1	1:25	Any expression in tumor cells
VE-cadherin	Novocastra 3E1	1:100	Any expression in tumor cells
VEGF	DAKO VG1	1:40*	Any expression in tumor cells
VEGFR1	Neomarkers RB-1527-P0	1:20*	Any expression in tumor cells

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 1. (*continued*)

Antibodies	Source and Clone or ID	Dilution	Cut-off Score
VEGFR2	Neomarkers RB-10453-P1	1:10*	Any expression in tumor cells
β-catenin	Ventana 14	Ready to use	Nuclear expression in tumor cells

In all instances except for * and †, in which high pH buffers or protease have been applied, respectively, antigen retrieval was based on lower pH buffers and microwaving. ABCG5 indicates ATP binding cassette subfamily G member 5; *ALDH1*, aldehyde dehydrogenase 1; *ALK*, anaplastic lymphoma kinase; Cbl-b, casitas B-lineage lymphoma (protooncogene) B; CXCR4, chemokine receptor; *EGFR*, epidermal growth factor; ESA (Ber-EP4, MOC31), epithelial-specific antigen; FoxP3, forkhead-box protein 3; MET, hepatocyte growth factor receptor; OCT4, octamer binding transcription factor 4; OPN, osteopontine; PD1, programmed death (receptor) 1; PDL1, programmed death ligand 1; Pgp, P-glycoprotein; PGP 9.5, protein gene product 9.5; PI3K, phosphoinositide 3-kinase; pSTAT3, phosphorylated signal transducer and activator of transcription 3; PTEN, phosphatase and tensine homolog; RHAMM, receptor of hyaluronic acid mediated motility; *SOX2*, sex-determining region Y-box 2; TGF-β, transforming growth factor beta; TIA1, cytotoxic granule-associated RNA binding protein; TIL, tumor-infiltrating lymphocytes; TTF1, thyroid transcription factor 1; VEGF, vascular endothelial growth factor; VEGFR1, VEGF receptor 1.

Over the past years, we have analyzed over 70 proteins and protein combinations (Table 1), the gene status (by means of in situ hybridization) of *ALK*, cyclin D 1 (*CCND1*), *EGFR*, and *MET*, and the methylation status of *p16* of 405 clinically well-characterized NSCLC patients who underwent surgical resection with curative intention and who were followed-up over a median of 62 months (range, 0.1 to 223 mo) as a part of the retrospective Twenty Years Retrospective of Lung Cancer (TYROL) study.[12] Formalin-fixed and paraffin-embedded tumor tissue material of these patients was brought into a tissue microarray (TMA) platform,[13] and the number of analyzable cases decreased slightly from 405 to 365 over time due to TMA exhaustion. Besides markers associated with certain histologic subtypes, parameters crucial for neuroendocrine differentiation, cell cycle regulation, apoptosis, cell motility and adhesion, the immunologic microenvironment, stemness, and tumor microvasculature were analyzed. Certain predictive markers, for example, expression of ALK, PDL1, and MET, were also assessed.[13–26]

The aim of the present study was to apply clustering analysis—a potentially powerful tool, which has rarely been utilized for NSCLC so far—to all previously assessed parameters for further comprehensive characterization

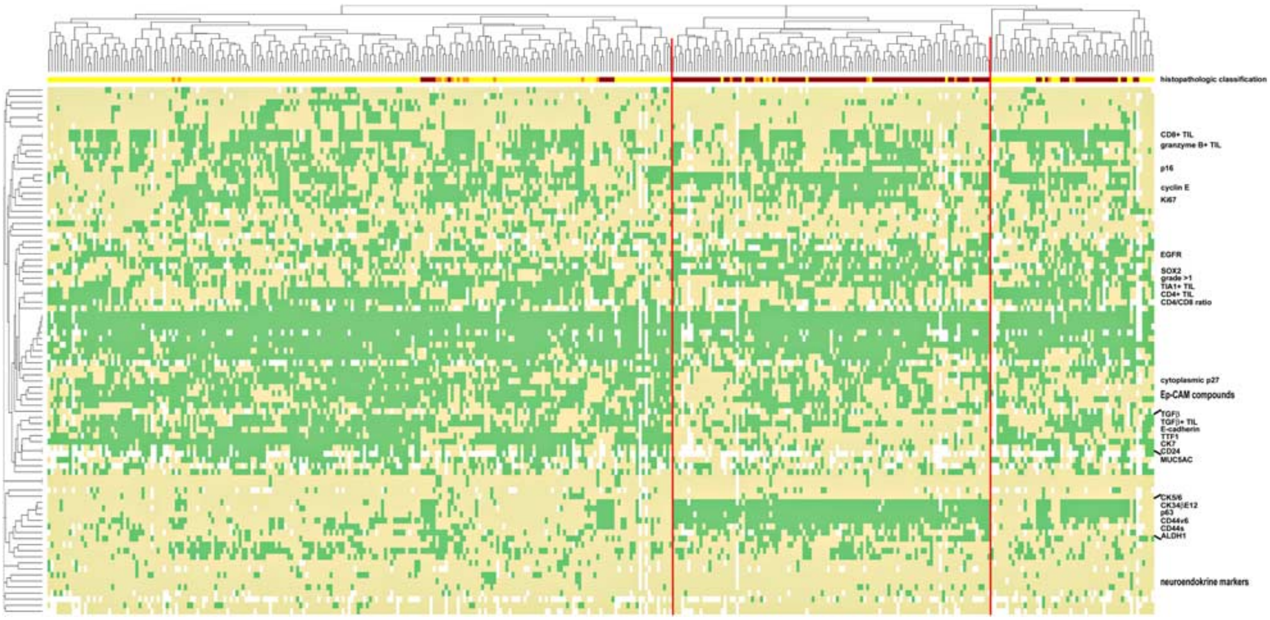

FIGURE 1. Heatmap of unsupervised hierarchical clustering analysis of the studied 365 non–small cell lung cancers. The upper dendrogram shows the patients split into 3 clusters, which are separated by red vertical lines. The left-handed dendrogram reflects marker clustering. For markers, yellow indicates lacking (below cut-off) expression/negativity or CD4/CD8 ratio <1, while green indicates expression/positivity or CD4/CD8 ratio >1. Empty spots depict markers that were not analyzable in respective cases. Only significant ($P < 0.000055$) discriminating markers or marker groups as well as some other informative markers (grade >1, Ki67, neuroendocrine markers) are annotated. The 3-colored line between the upper dendrogram and the heatmap shows the corresponding histopathologic classification of cases: yellow—ACA, red—SCC, and orange—LCC. Note that ACA mostly cluster to cluster 1 (ACA-like) and SCC to cluster 2 (SCC-like), while LCC are distributed throughout the unsupervised clusters 1, 2, and 3. Interestingly, a third unsupervised cluster 3 (right-handed) mostly defined by TIL signatures and containing ACA, SCC, and LCC appears. ACA indicates adenocarcinoma; *ALDH1*, aldehyde dehydrogenase 1; Ep-CAM, epithelial cell adhesion molecule; LCC, large cell carcinoma; SCC, squamous cell carcinoma; *SOX2*, sex-determining region Y-box 2; TIL, tumor-infiltrating lymphocyte.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

and subgrouping of our cohort to reveal additional biological and prognostic stratifications. This represents the largest NSCLC collective to date analyzed by protein expression clustering analysis, and incorporates the widest panel of markers, providing a first thorough look for this technique in surgically resected NSCLC.

METHODS

Cases

The archival samples derived from NSCLC patients with radical surgical resection with curative intent between 1992 and 2004 and diagnosed at the Institute of Pathology, Medical University of Innsbruck, were studied.[13] The cohort reported here consists of 365 cases. Carcinoids were excluded from this analysis. Cases were selected only on the basis of tissue preservation. Hematoxylin and eosin (H&E)-stained slides from all available specimens were reclassified by 2 pathologists (W.S. and A.T.) without knowledge of patient data, according to the current (2015) WHO classification of tumors of the lung.[27] Tumor differentiation was graded as well, as moderate or poor. The clinical information (including parameters such as tumor stage, disease recurrence, and overall survival) was documented within the TYROL survey, a project aiming to analyze various features of a large number of lung cancer patients.[12] Approval for data acquisition and analysis was obtained from the Ethics Committee of the Medical University of Innsbruck.

TMA Construction

The tumor material consisted of paraffin-embedded tissue after fixation in 10% neutral buffered formalin. The TMA was constructed, as previously described.[13] The first sections were stained by H&E, Periodic Acid-Schiff (PAS), and alcian-blue-PAS to confirm validity; the following were used for immunohistochemistry.

Phenotypic and Genotypic Marker Analyses

Starting in the year 2009[13] and ending in the year 2018,[26] over 70 single-protein expression analyses have been performed by means of immunohistochemistry, and 4 NSCLC-relevant genes, that is, *ALK*, *CCND1*, *EGFR*, and *MET*, have been studied by means of interphase fluorescence in situ hybridization (FISH) in that cohort,[13–26] and clinical follow-up has been updated until 2016. Importantly, for all markers rational, diagnostically useful, prognostically relevant, or biologically meaningful cut-off scores have been applied; Table 1 incorporates all antibodies and FISH-probes applied, pretreatment conditions, cut-off scores used, and number of positive cases.

Statistical Analysis

Hierarchical cluster analysis was used to organize markers and cases, respectively, according to their similarities. Immunohistochemical marker expression and FISH data were converted to binary format representing positivity or negativity based on the predefined cut-offs (Table 1). Distance matrix was calculated using the Jaccard's distance measure, and hierarchical clustering with complete linkage was performed applying R statistical software (version 3.3.0, https://www.r-project.org/).[8,9,28] Analysis was performed on the entire data set row-wise (cases) and column-wise (markers) and visualized as a binary heatmap with adjunct dendrograms. To evaluate the stability of selected clusters, that is, reproducibility of clustering, NSCLC cases were randomly split into 2 groups and reclustered. Allocation of randomly selected cases into the initially obtained clusters was evaluated by the Cohen κ, $\kappa > 0.61$ implying substantial, > 0.41, moderate, and > 0.21 fair reproducibility. Hierarchical clustering was also performed, as described above in a supervised manner on case subsets according to their histologic subtype, that is, squamous cell carcinoma (SCC), adenocarcinoma (ACA), and large cell carcinomas (LCC).

Distribution analysis of clinicopathologic, immunohistochemical, and FISH parameters with regard to clusters was performed using the χ^2 test, corrected for multiple testing with the Bonferroni method, considering $P \leq 0.000055$ as statistically significant in the unsupervised entire data set, and ≤ 0.000018 in the supervised cluster analysis of separate histologic subtypes.

TABLE 2. Unsupervised Cluster Analysis

	Cluster 1 (n = 206)		Cluster 2 (n = 105)		Cluster 3 (n = 54)	
	Increased	**Decreased**	**Increased**	**Decreased**	**Increased**	**Decreased**
Parameters	CK7	CD24−/CD44s+	CK5/6	—	CD8+ TIL	CD4/CD8
	TTF1	—	CK34βE12	—	TGF-β+ TIL	Ber-EP4
	TGF-β	—	p63	—	—	MOC31
	EGFR	—	CD44s	—	—	ESA
	E-cadherin	—	CD44v6	—	—	—
	Ber-EP4	—	CCNE	—	—	—
	MOC31	—	p21	—	—	—
	ESA	—	ALDH1	—	—	—
	MUC5AC	—	SOX2	—	—	—
	p16	—	—	—	—	—
	p27c	—	—	—	—	—
	CD24	—	—	—	—	—

ALDH1 indicates aldehyde dehydrogenase 1; *c*, cytoplasmic; *CCNE*, cyclin E; *CD4/CD8*, CD4/CD8 ratio; *ChrA*, chromogranin A; *CK7*, cytokeratin 7; *EGFR*, epidermal growth factor; *ESA(Ber-EP4, MOC31)*, epithelial-specific antigen; MUC5AC, mucin 5AC; *SOX2*, sex-determining region Y-box 2; *TGF-β*, transforming growth factor beta; *TIL*, tumor-infiltrating lymphocytes; *TTF1*, thyroid transcription factor 1.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Kaplan-Meier curves were calculated for survival estimates, and exploratory (ie, not corrected for multiple testing) log-rank statistics was used to determine prognostic differences between clusters. When not corrected for multiple testing, $P<0.05$ were considered as significant. Two-sided tests were used throughout. Statistical calculations with regard to Cohen κ, distribution, and survival were performed using SPSS 22.0 software (SPSS, Chicago, IL).

RESULTS AND DISCUSSION

Unsupervised cluster analysis revealed 3 NSCLC clusters (Fig. 1 and Table 2): one consisting of 206 cases showing higher expression of CK7, thyroid transcription factor 1 (TTF1), transforming growth factor beta (TGF-β), EGFR, E-cadherin, and epithelial cell adhesion molecule (Ep-CAM) compounds (Ber-EP4, MOC31 and ESA), MUC5AC, p16, cytoplasmic p27, and CD24 and low presence of cases with the stem cell–like phenotype (CD24−/CD44s+), and thus

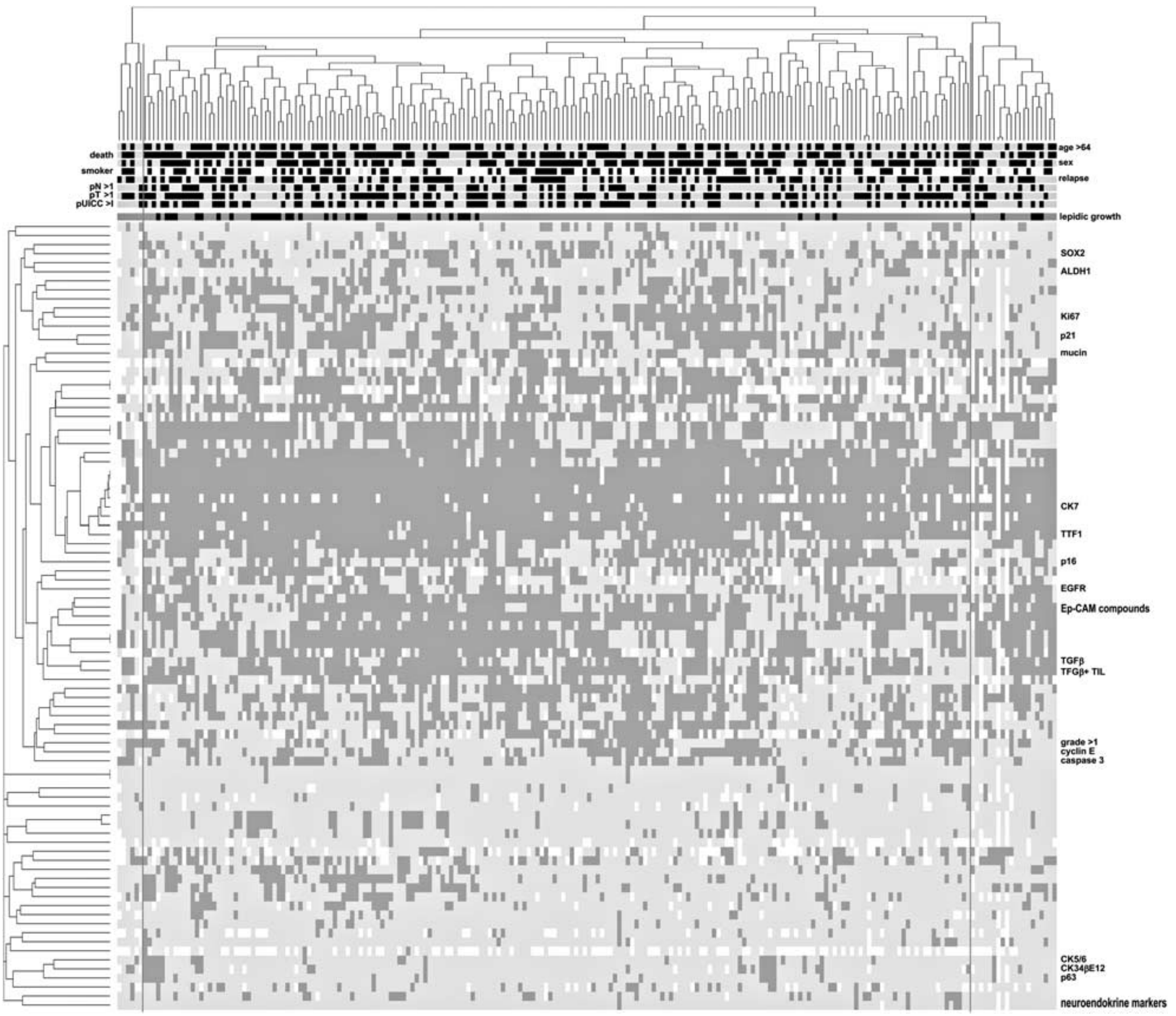

FIGURE 2. Heatmap of hierarchical clustering analysis of the studied 218 ACA. Patients split into 3 clusters, which are separated by vertical lines. For markers, light grey indicates lacking (below cut-off) expression/negativity, whereas dark grey indicates expression/positivity. Only significant ($P<0.000018$) discriminating markers or marker groups as well as some other informative markers (grade >1, CK5/6, CK34βE12, p63, ACA markers and neuroendocrine markers, stainable mucins) are annotated. The grayscale colored heatmap between the upper dendrogram and the marker heatmap codes for relevant clinicopathologic data, black indicating patients above 65 years old, male individuals, smokers, patients suffering from relapses, higher pathological nodal, pathological tumor, and pathological Union Internationale Contre le Cancer stages or cases displaying lepidic growth patterns, respectively. Empty spots depict markers that were not analyzable in respective cases or in cases with missing clinicopathologic data. ACA indicates adenocarcinoma; *ALDH1*, aldehyde dehydrogenase 1; *EGFR*, epidermal growth factor; Ep-CAM, epithelial cell adhesion molecule; *SOX2*, sex-determining region Y-box 2; TIL, tumor-infiltrating lymphocyte; TTF1, thyroid transcription factor 1.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

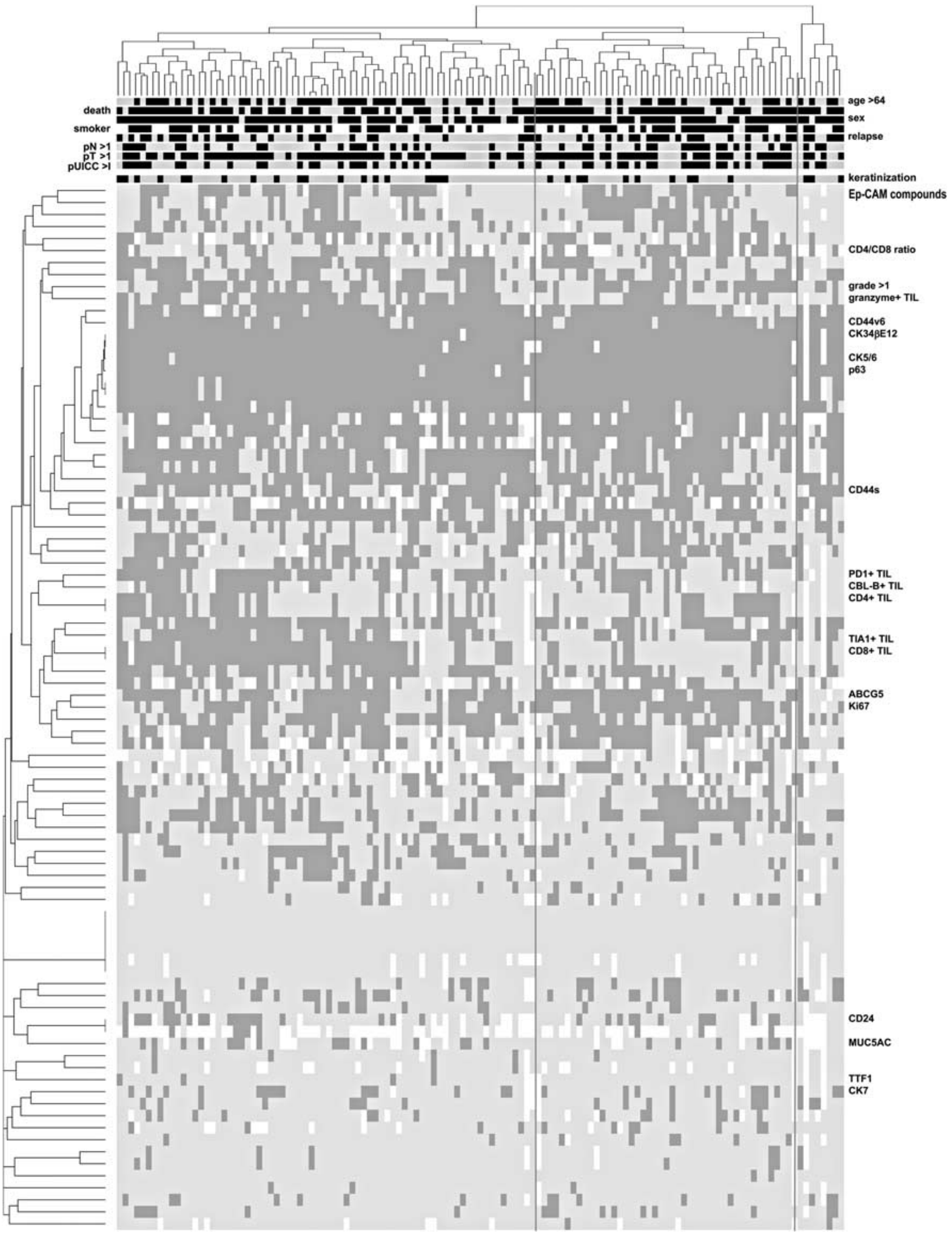

FIGURE 3. Heatmap of hierarchical clustering analysis of the studied 125 SCC. Patients split into 3 clusters, which are separated by vertical lines. For markers, light grey indicates lacking (below cut-off) expression/negativity or CD4/CD8 ratio <1, while dark grey indicates expression/positivity or CD4/CD8 ratio >1. Only significant ($P<0.000018$) discriminating markers or marker groups as well as some other informative markers (grade >1, CK7, TTF1, Ki67, SCC-markers, subpopulations of TIL) are annotated. The grayscale colored heatmap between the upper dendrogram and the marker heatmap codes for relevant clinicopathologic data, black indicating patients above 65 years old, male individuals, smokers, patients suffering from relapses, higher pathological nodal, pathological tumor, and pathological Union Internationale Contre le Cancer stages, or cases displaying keratinization, respectively. Empty spots depict markers that were not analyzable in respective cases or cases with missing clinicopathologic data. CK7 indicates cytokeratin 7; Ep-CAM, epithelial cell adhesion molecule; SCC indicates squamous cell carcinoma; TIL, tumor-infiltrating lymphocyte; TTF1, thyroid transcription factor 1.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

displaying an ACA-like phenotype (unsupervised cluster 1); another consisting of 105 cases with higher expression of CK5/6, CK34βE12, p63, CD44s and CD44v6, cyclin E (CCNE) and p21, as well as aldehyde dehydrogenase 1 (ALDH1) and sex-determining region Y-box 2 (SOX2), and thus displaying an SCC-like phenotype (unsupervised cluster 2); and a third cluster of 54 cases with higher intratumoral contents of CD8-positive T cells, lower CD4/CD8 ratios, higher amounts of TGF-β-positive lymphocytes, and lower expression of Ep-CAM compounds (unsupervised "immune signature" cluster 3). *P*-values for the linkage of all the above markers to the respective clusters were <0.000055. As to be expected, skewing toward male sex was observable in the unsupervised cluster 2 (SCC-like), while neuroendocrine markers were more commonly present in the unsupervised clusters 1 and 3.

Reanalysis of the randomly dichotomized entire NSCLC collective to test reproducibility of clustering yielded a κ-value of 0.733 for the ACA-like cluster 1 and 0.692 for the SSC-like cluster 2, indicating substantial reproducibility, whereas the smaller cluster 3 was only fairly reproducible with a κ-value of 0.371.

Thus, the unsupervised cluster analysis of this large collective of NSCLC cases demonstrates the very robust and substantially reproducible signature of the biological diversity determined by their histogenesis, that is, being either ACA or SCC, which was expected and not surprising. There were only 7 ACA (all but one belonging to the supervised cluster ACA2, demonstrating expression of SOX2, markers of stemness and increased proliferation and apoptosis) found among the SCC-like cluster, all of which were CK7 positive and showed PAS and alcian-blue-PAS–positive inclusions in their cytoplasm and nucleus. CK5/6 was expressed by 5/7, and p63 by 4/7, and 2 were positive for TTF1. Interestingly, both TTF1-positive ACAs were also positive for CK5/6 and p63. For all 7 ACA cases that clustered in the SCC-like cluster, classical histomorphology combined with mucin positivity, and CK7 and—if applicable—TTF1 positivity were the determining factors for the diagnosis of ACA. On the opposite, 11 SCC (all belonging to the supervised cluster SCC2) were encountered in the ACA-like cluster. All these cases were positive for CK5/6 and p63, and negative for TTF1. These 11 SCC were mostly positive for SOX2 and showed a profile typical of stem cells (CD44+/CD24−). Obviously, stemness-like signatures of either ACA or SCC

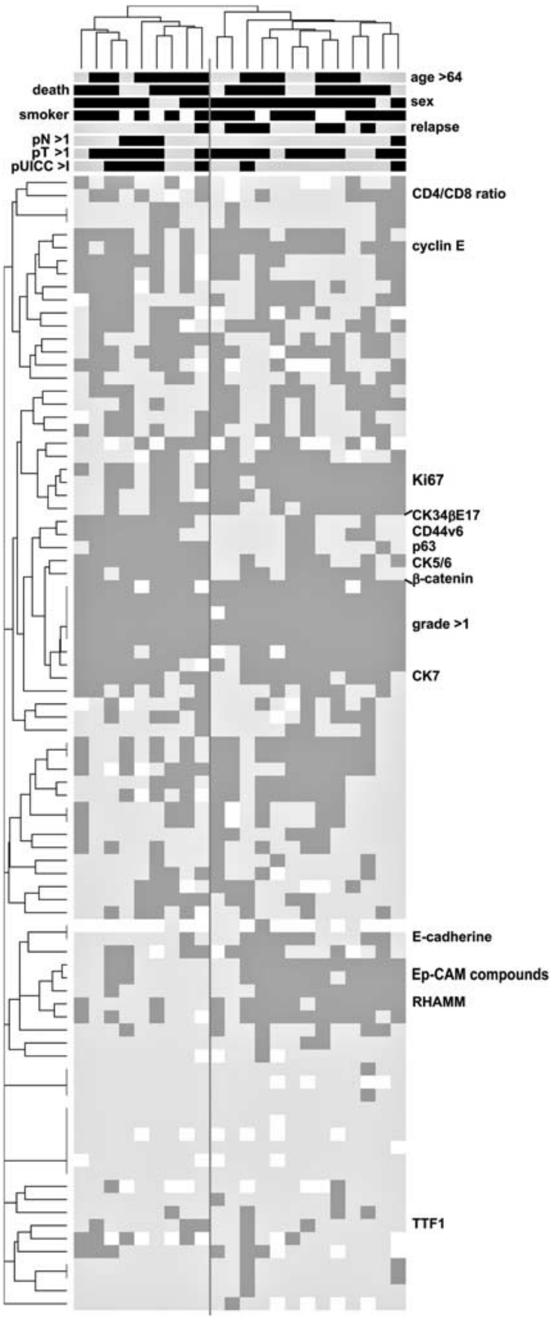

FIGURE 4. Heatmap of hierarchical clustering analysis of the studied 22 LCC. Patients split into 3 clusters, which are separated by vertical lines. For markers, light grey indicates lacking (below cutoff) expression/negativity or CD4/CD8 ratio <1, whereas dark grey indicates expression/positivity or CD4/CD8 ratio >1. Only significant ($P < 0.008$) discriminating markers or marker groups as well as some other informative markers (grade >1, CD4/CD8 ratio, CK7, TTF1) are annotated. The grayscale colored heatmap between the upper dendrogram and the marker heatmap codes for relevant clinicopathologic data, black indicating patients above 65 years old, male individuals, smokers, patients suffering from relapses, or higher pathological nodal, pathological tumor, or pathological Union Internationale Contre le Cancer stages, respectively. Empty spots depict markers that were not analyzable in respective cases or cases with missing clinicopathologic data. Note that LCC clearly cluster into 2 groups, which are mainly defined by expression of either squamous cell/basal markers (β-catenin, CD44v6, CK5/6, CK34βE12, and/or p63,) or more simple epithelia-like markers (E-cadherine and/or epithelial cell adhesion molecule compounds: Ber-EP4, ESA, MOC31) and higher proliferative activity. CK7 indicates cytokeratin 7; LCC, large cell carcinoma; RHAMM, receptor for hyaluronan-mediated motility; TTF1, thyroid transcription factor 1.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 3. Supervised Cluster Analysis According to Histologic Tumor Type

	Adenocarcinoma (n = 218)			**Squamous Cell Carcinoma (n = 125)**			**Large Cell Carcinoma (n = 22)**	
	Cluster1 (n = 6)	**Cluster 2 (n = 192)**	**Cluster 3 (n = 20)**	**Cluster 1 (n = 8)**	**Cluster 2 (n = 45)**	**Cluster 3 (n = 72)**	**Cluster 1 (n = 9)**	**Cluster 2 (n = 13)**
Parameters	**Grade**	**m/f**	**mucin**	CD24 – /CD44s+	**CD4/CD8**	**GrB+ TIL**	**CK5/6**	**Ki67**
	Csp3	**Ki67**	**lepidic**	Ber-EP4	**Ber-EP4**	***CD8+TIL***	**CK34βE12**	**CCNE**
	TGF-β	**ALDH1**	**TTF1**	MOC31	**MOC31**	***TIA+ TIL***	**p63**	**Ber-EP4**
	m/f	**SOX2**	**EGFR**	ESA	**ESA**	***CBL-B+ TIL***	**β-catenin**	**MOC31**
	mucin	—	**Ber-EP4**	MUC5AC	**CD24 – /CD44s+**	***PDL1+ TIM***	**CD44v6**	**ESA**
	lepidic	—	**MOC31**	—	OS	CD4/CD8	—	**RHAMM**
	p16	—	**ESA**	—	GrB+ TIL	Ber-EP4	—	**E-cadherin**
	p21	—	grade	—	—	MOC31	—	CK5/6
	EGFR	—	Ki67	—	—	ESA	—	CK34βE12
	Ber-EP4	—	CCNE	—	—	CD24 – (CD44s+)	—	p63
	MOC31	—		—	—	—	—	—
	ESA	—		—	—	—	—	—

Bold signifies increase, normal print signifies decrease, factors being of significance only when not corrected for multiple testing are in italics.

ALDH1 indicates aldehyde dehydrogenase 1; *CBL-B*, Casitas B-lineage lymphoma B; *CCNE*, cyclin E; *CD4/CD8*, CD4/CD8 ratio; *CK*, cytokeratin; *Csp3*, caspase 3; *EGFR*, epidermal growth factor; *ESA (Ber-EP4, MOC31)*, epithelial-specific antigen; *GrB*, granzyme B; m/f, male/female ratio; MUC5AC, mucin 5AC; *OS*, overall survival; *PDL1*, programmed death ligand 1; *RHAMM*, receptor for hyaluronan-mediated motility; *SOX2*, sex-determining region Y-box 2; *TGF*-β, transforming growth factor beta; *TIL*, tumor-infiltrating lymphocytes; *TIM*, tumor-infiltrating mononuclear cells; *TTF1*, thyroid transcription factor 1.

cause difficulties in assigning tumors to a specific entity in clustering analyses.

Interestingly, with all caveats of more limited case numbers and fair reproducibility, there appears cluster 3. It encompasses (i) a subgroup of TTF1+ and CK7+ cases, (ii) another subgroup of p63+ and CK5/6+ tumors, with or without expression of TTF1 and/or CK7, that primarily clustered together because of their very high amounts of intratumoral CD8+ and TGF-β+ (and TIA1+)-positive lymphocytes, and (iii) a third, small subgroup of CK7+ and mostly TTF1+ cases with expression of pAKT, chemokine receptor type 4 (CXCR4), and presence of higher amounts of TGF-β+ (but not CD8+) lymphocytes that particularly displayed higher relapse frequencies (because of further subgrouping, no statistical testing has been performed for these instances). For this cluster 3, obviously, the signature of the TIL prevailed over the signature of histogenesis. This may be especially important in the light of current immunomodulating therapies for NSCLC.[29] As only a portion of patients respond to such treatment strategies, a successful therapy may be linked to the presence of a certain immunologic tumor phenotype/signature. The important role of intratumoral contents of CD8+ T cells to define a distinct unsupervised NSCLC cluster, fits with recent data showing that the prognostic impact of PDL1 expression in NSCLC cases depends on the presence or absence of CD8+TIL and—together with the very encouraging in vivo data on the efficacy of immune checkpoint modulation in subsets of NSCLS patients—points toward a real biological existence of such an NSCLC cluster with decelerated but potentially "awakenable" immunity.[24,30]

Importantly, in this unsupervised setting, neither a neuroendocrine cluster nor an LCC cluster appeared, fitting with the current concepts of NSCLC classification, which challenged and to a great part abolished respective entities, and which is also in line with our previous work as well as other reports (see also discussion on the supervised LCC cluster analysis).[13,19,27]

Supervised cluster analysis could further identify relevant clusters within the predetermined NSCLC histotypes, that is, ACA, SCC, and LCC (Figs. 2–4 and Table 3).

For ACA (n = 218), 3 groups were apparent: 1 very large group, and 2 groups with far less cases (Fig. 2). One cluster (ACA1, n = 6) was the only 1 with an inversed male/female ratio with twice as many women than men, and encompassed cases without mucin and without lepidic growth, with higher grade and lower expression of p16 and p21, higher active caspase 3 presence, lower to lacking EGFR and Ep-CAM compounds, and higher tumor expression of TGF-β ($P < 0.000018$ for all listed markers). The second smaller cluster (ACA3, n = 20) contained a higher proportion of cases with detectable mucin and of cases with lepidic predominance, and, thus, was of lower grade. Furthermore, this cluster displayed high TTF1 positivity, a low Ki67 proliferation rate, and low expression of CCNE, but was linked to the expression of EGFR and Ep-CAM compounds ($P < 0.000018$ for all listed markers). The third, very large ACA cluster (ACA2, n = 192) showed the highest male predominance, had the highest Ki67 proliferation rate, and higher ALDH1 and SOX2 positivity ($P < 0.000018$ for all listed markers). Notably, these results show that a grading system for ACA according to architecture, particularly lepidic growth, as previously proposed and reproduced on our cohort as well, is also reflected by clustering analysis.[19,27]

Among the SCC (n = 125) also, 3 main clusters were identified (Fig. 3). One small group (SCC1, n = 8) was defined by complete lack of the stem cell–like phenotype (CD24−/CD44s+), lacking expression of Ep-CAM compounds, and MUC5AC ($P < 0.000018$ for all listed

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

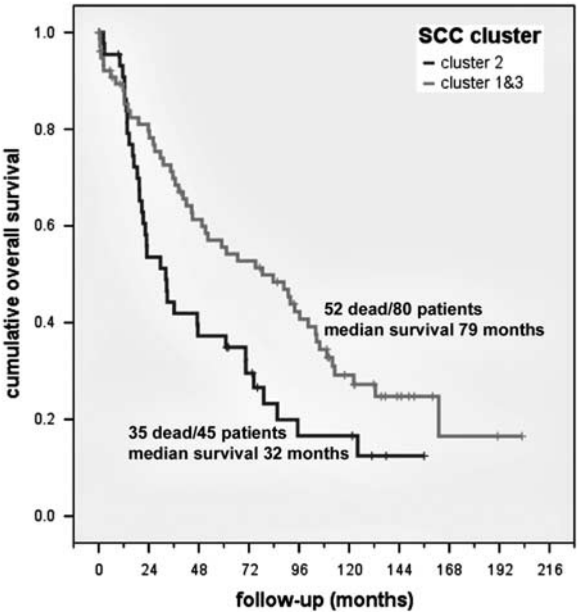

FIGURE 5. Survival analysis of SCC clusters. Note the poorer outcome of SCC cluster 2 (linked to high CD4/CD8 ratio, lower presence of granzyme B+ tumor-infiltrating lymphocytes, higher expression of epithelial cell adhesion molecule compounds, and higher expression of stemness markers (ABCG5+, CD24−, D44s+) compared with cluster 1 and 3 (for simplicity and because of having almost the same slope, shown together); $P=0.041$ (for comparison between all 3 clusters). ABCG5 indicates ATP binding cassette subfamily G member 5; SCC, squamous cell carcinoma.

markers). Another cluster (SCC2, n = 45) was linked to high CD4/CD8 ratio, lower presence of granzyme B+ TIL, higher expression of Ep-CAM compounds, and the presence of the proportionally most cases with a stem cell–like phenotype (CD24−/CD44s+) ($P<0.000018$ for all listed markers), and was the only cluster with a specific—in that case worse—prognosis (Fig. 5). The third cluster (SCC3, n = 72) showed the opposite with a low CD4/CD8 ratio and a higher presence of granzyme B+ TIL ($P<0.000018$ for all listed markers), and—yet statistically not significant when corrected for multiple testing—higher amounts of CD8+ and/or TIA+, CBL-B+ and PDL1+ TIL, as well as lower expression of the Ep-CAM compounds and less cases with stem cell–like phenotype ($P<0.000018$ for the latter two).

For the LCC (n = 22), 2 groups were distinguished by our cluster analysis (Fig. 4): one with frequent expression of basal markers (CK5/6, CK34βE12, p63), β-catenin, and CD44v6 (LCC1, n = 9), that is, SCC-like, and another with diminished basal markers but higher Ki67 proliferation rate and expression of CCNE, as well as increased expression of Ep-CAM compounds, receptor for hyaluronan-mediated motility (RHAMM), and E-cadherin (LCC2, n = 13), that is, more ACA-like. As only 22 LCC entered the analysis, the marker distribution was assessed in an exploratory setting, and *P*-values were corrected for multiple testing of 3 histotypes with 2 clusters and not for 70 markers, that is, $P<0.008$. Importantly, when linked to the unsupervised clusters, 6 of the 9 LCC1 appeared in the unsupervised cluster 2 (SCC-like) and only 1 and 2 in the unsupervised cluster 1 (ACA-like) and 3 (immune signature), respectively, whereas 11 of the 13 LCC2 emerged in the unsupervised cluster 1 (ACA-like) and 2 in the unsupervised cluster 2 (SCC-like, $P=0.002$). These results suggest that it is generally possible to further, more specifically phenotypically assign LCC (diagnosed based on histologic analysis) to favor either SCC or ACA, which is currently carried out in the routine diagnosis of NSCLC, in which TTF1 and p63 or p40 expression is regarded as important as morphology to classify cases.

In conclusion, we could show that clustering analysis is generally feasible for NSCLC and is a useful tool for verifying established characteristics as well as generating new standpoints. Specifically, our data show that tumor grading of ACA by architecture is significant, that LCC is presumably not a separate entity, that an immunologic cluster—depending on patients' immunity and the complex interaction between the tumor and the lymphocytes—likely exists, and that one prognostically relevant SCC cluster with high CD4/CD8 ratio and lower presence of granzyme B+ TIL and poorer outcomes exists.

REFERENCES

1. Abd El-Rehim DM, Ball G, Pinder SE, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int J Cancer*. 2005;116:340–350.
2. Alizadeh AA, Ross DT, Perou CM, et al. Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol*. 2001;195:41–52.
3. Güler EN. Gene expression profiling in breast cancer and its effect on therapy selection in early-stage breast cancer. *Eur J Breast Healt*. 2017;13:168–174.
4. Soria D, Garibaldi JM, Ambrogi F, et al. A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients. *Comput Biol Med*. 2010;40:318–330.
5. Tomida S, Koshikawa K, Yatabe Y, et al. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. *Oncogene*. 2004;23:5360–5370.
6. Hamamoto J, Soejima K, Yoda S, et al. Identification of microRNAs differentially expressed between lung squamous cell carcinoma and lung adenocarcinoma. *Mol Med Rep*. 2013;8:456–462.
7. Markou A, Sourvinou I, Vorkas PA, et al. Clinical evaluation of microRNA expression profiling in non small cell lung cancer. *Lung Cancer*. 2013;81:388–396.
8. Menter T, Dickenmann M, Juskevicius D, et al. Comprehensive phenotypic characterization of PTLD reveals potential reliance on EBV or NF-κB signalling instead of B-cell receptor signalling. *Hematol Oncol*. 2017;35:187–197.
9. Menter T, Ernst M, Drachneris J, et al. Phenotype profiling of primary testicular diffuse large B-cell lymphomas. *Hematol Oncol*. 2014;32:72–81.
10. Au NH, Cheang M, Huntsman DG, et al. Evaluation of immunohistochemical markers in non-small cell lung cancer by unsupervised hierarchical clustering analysis: a tissue microarray study of 284 cases and 18 markers. *J Pathol*. 2004;204:101–109.
11. Grossi F, Spizzo R, Bordo D, et al. Prognostic stratification of stage IIIA pN2 non-small cell lung cancer by hierarchical clustering analysis of tissue microarray immunostaining data: an Alpe Adria Thoracic Oncology Multidisciplinary Group study (ATOM 014). *J Thorac Oncol*. 2010;5:1354–1360.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

12. Kocher F, Hilbe W, Seeber A, et al. Longitudinal analysis of 2293 NSCLC patients: a comprehensive study from the TYROL registry. *Lung Cancer*. 2015;87:193–200.
13. Sterlacci W, Fiegl M, Hilbe W, et al. Clinical relevance of neuroendocrine differentiation in non-small cell lung cancer assessed by immunohistochemistry: a retrospective study on 405 surgically resected cases. *Virchows Arch*. 2009;455:125–132.
14. Sterlacci W, Fiegl M, Hilbe W, et al. Deregulation of p27 and cyclin D1/D3 control over mitosis is associated with unfavorable prognosis in non-small cell lung cancer, as determined in 405 operated patients. *J Thorac Oncol*. 2010;5:1325–1336.
15. Sterlacci W, Tzankov A, Veits L, et al. The prognostic impact of sex on surgically resected non-small cell lung cancer depends on clinicopathologic characteristics. *Am J Clin Pathol*. 2011;135: 611–618.
16. Sterlacci W, Tzankov A, Veits L, et al. A comprehensive analysis of p16 expression, gene status, and promoter hypermethylation in surgically resected non-small cell lung carcinomas. *J Thorac Oncol*. 2011;6:1649–1657.
17. Sterlacci W, Fiegl M, Tzankov A. Prognostic and predictive value of cell cycle deregulation in non-small-cell lung cancer. *Pathobiology*. 2012;79:175–194.
18. Sterlacci W, Wolf D, Savic S, et al. High transforming growth factor β expression represents an important prognostic parameter for surgically resected non-small cell lung cancer. *Hum Pathol*. 2012;43: 339–349.
19. Sterlacci W, Savic S, Schmid T, et al. Tissue-sparing application of the newly proposed IASLC/ATS/ERS classification of adenocarcinoma of the lung shows practical diagnostic and prognostic impact. *Am J Clin Pathol*. 2012;137:946–956.
20. Sterlacci W, Savic S, Fiegl M, et al. Putative stem cell markers in non-small-cell lung cancer: a clinicopathologic characterization. *J Thorac Oncol*. 2014;9:41–49.
21. Augustin F, Fiegl M, Schmid T, et al. Receptor for hyaluronic acid-mediated motility (RHAMM, CD168) expression is prognostically important in both nodal negative and nodal positive large cell lung cancer. *J Clin Pathol*. 2015;68:368–373.
22. Pomme G, Augustin F, Fiegl M, et al. Detailed assessment of microvasculature markers in non-small cell lung cancer reveals potentially clinically relevant characteristics. *Virchows Arch*. 2015;467: 55–66.
23. Sterlacci W, Saker S, Huber B, et al. Expression of the CXCR4 ligand SDF-1/CXCL12 is prognostically important for adenocarcinoma and large cell carcinoma of the lung. *Virchows Arch*. 2016; 468:463–471.
24. Sterlacci W, Fiegl M, Droeser RA, et al. Expression of PD-L1 identifies a subgroup of more aggressive non-small cell carcinomas of the lung. *Pathobiology*. 2016;83:267–275.
25. Sterlacci W, Fiegl M, Gugger M, et al. MET overexpression and gene amplification: prevalence, clinico-pathological characteristics and prognostic significance in a large cohort of patients with surgically resected NSCLC. *Virchows Arch*. 2017;471:49–55.
26. Sterlacci W, Fiegl M, Veits L, et al. Diagnostic and prognostic impact of mucin 1-6 expression in non-small cell lung cancer. *Indian J Pathol Microbiol*. 2018;61:187–191.
27. Travis W, Brambilla E, Burke A, et al. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart*. Lyon: IARCPress; 2015.
28. Sneath PH. Some thoughts on bacterial classification. *J Gen Microbiol*. 1957;17:184–200.
29. Pabani A, Butts CA. Current landscape of immunotherapy for the treatment of metastatic non-small-cell lung cancer. *Curr Oncol*. 2018; 25:S94–S102.
30. Durgeau A, Virk Y, Corgnac S, et al. Recent advances in targeting CD8 T-cell immunity for more effective cancer immunotherapy. *Front Immunol*. 2018;9:14.

Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2019 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.