ORIGINAL ARTICLE

Gene Expression Profiling of Large Cell Lung Cancer Links Transcriptional Phenotypes to the New Histological WHO 2015 Classification

Anna Karlsson, MSc,[a] Hans Brunnström, MD, PhD,[b,c] Patrick Micke, MD, PhD,[d] Srinivas Veerla, PhD,[a] Johanna Mattsson, PhD,[d] Linnea La Fleur, MSc,[d] Johan Botling, MD, PhD,[d] Mats Jönsson, PhD,[a] Christel Reuterswärd, MSc,[a] Maria Planck, MD, PhD,[a,e] Johan Staaf, PhD[a,*]

[a]*Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden*
[b]*Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden*
[c]*Department of Pathology, Regional Laboratories Region Skåne, Lund, Sweden*
[d]*Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden*
[e]*Department of Respiratory Medicine and Allergology, Skåne University Hospital, Lund, Sweden*

Received 1 March 2017; revised 26 April 2017; accepted 12 May 2017
Available online - 20 May 2017

ABSTRACT

Introduction: Large cell lung cancer (LCLC) and large cell neuroendocrine carcinoma (LCNEC) constitute a small proportion of NSCLC. The WHO 2015 classification guidelines changed the definition of the debated histological subtype LCLC to be based on immunomarkers for adenocarcinoma and squamous cancer. We sought to determine whether these new guidelines also translate into the transcriptional landscape of lung cancer, and LCLC specifically.

Methods: Gene expression profiling was performed by using Illumina V4 HT12 microarrays (Illumina, San Diego, CA) on samples from 159 cases (comprising all histological subtypes, including 10 classified as LCLC WHO 2015 and 14 classified as LCNEC according to the WHO 2015 guidelines), with complimentary mutational and immunohistochemical data. Derived transcriptional phenotypes were validated in 199 independent tumors, including six WHO 2015 LCLCs and five LCNECs.

Results: Unsupervised analysis of gene expression data identified a phenotype comprising 90% of WHO 2015 LCLC tumors, with characteristics of poorly differentiated proliferative cancer, a 90% tumor protein p53 gene (*TP53)* mutation rate, and lack of well-known NSCLC oncogene driver alterations. Validation in independent data confirmed aggregation of WHO 2015 LCLCs in the specific phenotype. For LCNEC tumors, the unsupervised gene expression analysis suggested two different transcriptional patterns corresponding to a proposed genetic division of LCNEC tumors into SCLC-like and NSCLC-like cancer on the basis of *TP53* and retinoblastoma 1 gene (*RB1*) alteration patterns.

Conclusions: Refined classification of LCLC has implications for diagnosis, prognostics, and therapy decisions. Our molecular analyses support the WHO 2015 classification of LCLC and LCNEC tumors, which herein follow different tumorigenic paths and can accordingly be stratified into different transcriptional subgroups, thus linking diagnostic immunohistochemical staining–driven classification with the transcriptional landscape of lung cancer.

Keywords: Lung cancer; Large cell lung carcinoma; LCNEC; Mutation; Gene expression; WHO classification

Introduction

NSCLC accounts for most lung cancers and is dominated by the histological subtypes adenocarcinoma, squamous cell carcinoma (SqCC), and large cell

*Corresponding author.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: Johan Staaf, PhD, Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Medicon Village, SE 22381 Lund, Sweden. E-mail: johan.staaf@med.lu.se

ISSN: 1556-0864
http://dx.doi.org/10.1016/j.jtho.2017.05.008

carcinoma with or without neuroendocrine features (LCNEC and LCLC, respectively) LCLC and LCNEC together account for 2% to 3% of all cases depending on cohort demographics and classification scheme.[1] In the WHO 2004 classification of lung cancer, LCLC was defined as an undifferentiated NSCLC lacking architectural and cytologic features of SCLC without glandular or squamous differentiation, whereas LCNEC was defined as an LCLC with neuroendocrine morphological features and at least one positive result of immunohistochemical (IHC) staining for a neuroendocrine marker.[2]

Controversy has existed as to whether LCLCs actually represent a truly distinct biological entity or are merely a group of very poorly differentiated tumors of other NSCLC histological subtypes (adenocarcinoma and/or SqCC).[3,4] The 2015 WHO classification scheme regrouped previously histologically defined LCLCs that express pneumocyte markers (thyroid transcription factor 1 [TTF1] and napsin A) as adenocarcinoma and those that test positive for a squamous marker (p40, CK5/6, or p63) as nonkeratinizing or basaloid SqCC, leaving surgically resected tumors lacking expression of these markers (referred to as *marker null*) as LCLCs.[5] Furthermore, LCNECs were separated from LCLCs because they instead share many similarities with neuroendocrine SCLC on the morphological, protein, mutational, DNA methylation, and transcriptional levels, albeit with some heterogeneity (see Rossi et al.,[6] Clinical Lung Cancer Genome Project and Network Genomic Medicine,[7] and Simbolo et al.[12] and the references therein).[6–12] Consistently, in the WHO 2015 update, LCNEC tumors have now been labeled *neuroendocrine* together with SCLC.[5]

Accurate distinction of the histological subtypes is of major clinical relevance. For SqCC, adenocarcinoma, and neuroendocrine tumors (LCNEC or SCLC), subtype-directed diagnostics and therapeutics are widely established. By the refined WHO 2015 classification, a minimization of the remaining marker-null LCLC group has been achieved. Recent studies have suggested poorer outcome for marker-null LCLC cases,[3,8] whereas others have not found this association.[13,14] Notably, all studies are still based on limited retrospective patient cohorts, which may explain the discrepancies. In advanced disease, the LCLC marker-null counterpart, NSCLC not otherwise specified, has been reported to have a worse patient outcome.[15] LCLC tumors remain fairly uncharacterized at the molecular level by high-dimensional genomic techniques, especially considering the otherwise strong molecular efforts made in the WHO 2015 classification scheme. Recent sequencing studies have demonstrated differences in oncogene mutation frequencies between WHO 2004 LCLC tumors expressing adenocarcinoma or SqCC markers and marker-null cases.[3,8,14,16] On the transcriptional level, no studies have thus far resolved the heterogeneity previously suggested for the LCLC group. Importantly, only with an improved molecular understanding can patients with marker-null LCLC benefit from the growing number of targeted treatments adopted in lung cancer.

In this study, we therefore aimed to investigate the transcriptional landscape of LCLC and LCNEC tumors in relation to other histological subgroups of lung cancer. On the basis of unsupervised analysis of global gene expression patterns in 159 surgically resected tumors we demonstrate a WHO 2015 LCLC transcriptional phenotype. Our results link the recent lung cancer classification scheme with the transcriptional landscape of the disease, now defining the poorly differentiated marker-null LCLC group as an entity with a specific gene expression phenotype (GEP).

Materials and Methods

Patient Material Discovery Cohort

A total of 159 patients with early-stage lung cancer surgically treated at the Skåne University Hospital in Lund, Sweden, were collected. A total of 116 cases have been described in previous studies[8,10,17] (Table 1, Supplementary Table 1, and Supplementary Methods). Classification of LCLC and LCNEC was originally performed according to the WHO 2004 scheme[2] and later updated to the WHO 2015 scheme.[5]

Ethics Statement

The study was approved by the Regional Ethical Review Board in Lund, Sweden (registration nos. 2004/762, 2008/702, and 2014/748).

Reclassification of LCLC by IHC Staining

Lung cancer cases with neuroendocrine morphological features that were classified as LCLC according to the WHO 2004 guidelines were evaluated for IHC staining of the neuroendocrine markers chromogranin A, synaptophysin, and CD56 (see Karlsson et al.[8] and Supplementary Methods). WHO 2004–classified LCLC cases without neuroendocrine features were analyzed for IHC staining of CK5/P40 (squamous cell markers) and TTF1/napsin A (adenocarcinoma markers) as described in the Supplementary Methods and in Karlsson et al.,[8] Micke et al.,[18] and Brunnström et al.[19] to classify them according to the WHO 2015 guidelines. No LCLC case was classified as the WHO 2015 uncertain phenotype thanks to successful stains of all cases. In addition, LCNEC cases were also analyzed by retinoblastoma 1 (RB1) immunohistochemistry (see Supplementary Methods).

Table 1. Patient Characteristics and Clinicopathological Data

Variable	Total Cohort	LCLC
No. of patients	159	47
Histological subtype		
Adenocarcinoma	83	–
Squamous cell carcinoma	26	–
SCLC	3	–
LCLC[a]	33	33
LCNEC	14	14
LCLC immunomarker profile (excluding LCNEC)		
Adenocarcinoma-like	–	19 (58%)
Squamous cell carcinoma-like	–	4 (12%)
Marker null	–	10 (30%)
Tumor stage		
I	120 (76%)	24 (51%)
II	27 (17%)	16 (34%)
III	8 (5%)	6 (13%)
IV	2 (1%)	1 (2%)
Smoking history		
Never-smokers	19	0
Smokers	114	23
Not available	26	24
Sex		
Female	85	24
Male	73	23
Median age (range), y	67 (34-84)	63 (34-77)
Patients evaluable for		
Gene expression	159	47
Mutations		

Note: LCLC refers to the WHO 2004 classification.
[a]Basaloid (n = 6) and lymphoepithelioma-like (n = 1) cases are included in the LCLC sample numbers.
LCLC, lung cell lung cancer; LCNEC, large cell neuroendocrine carcinoma.

Global Gene Expression Analysis Discovery Cohort

RNA and DNA from fresh frozen tissue were extracted with the Qiagen Allprep extraction kit (Qiagen, Hilden, Germany). Gene expression data for 43 of the cases were generated by using the Illumina HT12 V4 microarrays (Illumina, San Diego, CA) at the Swegene Center for Integrative Biology at Lund University. Gene expression data were pooled with previously reported data for 116 cases analyzed by the same expression platform[10] as described (see Supplementary Methods) and is available as Gene Expression Omnibus series GSE94601. Consistency in pooling of the two cohorts was confirmed by principal component analysis[20] and analysis of 16 overlapping technical replicate samples between the two cohorts (see Supplementary Fig. 1). Consensus clustering[21] was performed as previously described[22] on mean-centered data (centering across samples) by using probe sets with different log2 SD cutoffs (see Supplementary Table 1). Differentially expressed probe sets between subgroups were identified by significance analysis of microarrays with a false discovery rate threshold of 1%. For independent validation of identified transcriptional subgroups, gene expression centroids were created as mean averages for each gene across all samples in the respective subgroup, as described.[23]

In addition, tumors were also scored according to six expression metagenes in lung cancer representing different biological processes and reported GEPs.[10,24,25] Functional classification was performed as described in Supplementary Methods.

Gene Expression Analysis Validation Cohort

Validation of gene expression subgroups was performed in data provided by Djureinovic et al.[26] (see Supplementary Methods). Histological classification of the samples was updated according to the WHO 2015 guidelines as previously described.[18]

Mutational Analysis

All cases were analyzed by the Illumina TruSight Tumor 26-gene next-generation sequencing (NGS) panel (Illumina), as described.[8] In addition, LCNEC cases were screened for retinoblastoma 1 gene (*RB1*) mutations by using a custom-designed bidirectional NGS panel (Illumina).

Results

IHC Reclassification of WHO 2004 LCLC

Thirty-three lung cancer cases classified as LCLC according to the WHO 2004 guidelines were included in the discovery cohort, of which 70% were reclassified as variants of adenocarcinoma or SqCC on the basis of the WHO 2015 guidelines. Specifically, 19 cases (58%) were reclassified as adenocarcinoma on the basis of positive expression of TTF1/napsin A, four (12%) were reclassified as SqCC on the basis of positive expression of CK5/P40, and 10 (30%) did not express any of these IHC markers (hereon referred to as *marker-null cases*) (see Table 1).

Unsupervised Gene Expression Analysis Stratifies LCLC in Accordance with Immunomarker Expression

To investigate whether the WHO 2015 guidelines translated into a better transcriptional subgrouping of LCLC, we performed unsupervised consensus clustering of a discovery cohort comprising 159 lung cancers of all histological subtypes, including 33 lung cancer cases that were classified as LCLCs and 14 classified as LCNECs according to the WHO 2004 guidelines. We first performed iterative consensus clustering without respect to sample annotations by using variable number of genes (Illumina probes displaying large variation in expression

across tumors [range 300–12581]) and evaluated cluster solutions (k = 5–11) to investigate sample cluster stability versus gene selection. Stable sample clusters formed across a wide range of different gene sets for different cluster solutions, indicating that gene selection has less influence on sample grouping in this cohort (Supplementary Fig. 2*A*). A similar stability was also observed when clustering only WHO 2004 LCLC and LCNEC cases separately (Supplementary Fig. 2*B*).

Next, we performed an in-depth comparison of unsupervised transcriptional subgroups with sample molecular and clinicopathological data. Acknowledging that the histological subtypes of lung cancer strongly influence the transcriptional landscape[7] and that subgroups within the histological subtypes likely exist, we chose a 10-group consensus cluster solution (k = 10 with 2730 Illumina probes, corresponding to a log2ratio standard deviation cutoff of 0.5) to be able to also study characteristics for minor subgroups. Consistent with previous studies,[7,27] we observed a clear separation of adenocarcinoma, SqCC, and SCLC cases into subclusters driven by specific transcriptional programs (Fig. 1). In agreement with recent studies,[7] LCNEC tumors clustered strongly (79% of cases) with SCLC tumors, forming a neuroendocrine subcluster (see Fig. 1). For LCLC, 84% of the WHO 2004 cases reclassified as adenocarcinoma-like clustered in an adenocarcinoma-dominated subcluster, whereas 50% of the LCLC SqCC-like cases clustered in an SqCC-dominated subcluster (see Fig. 1). Notably, 90% of marker-null cases (nine of 10) aggregated in a separate transcriptional cluster (see Fig. 1 [*cluster 8*]), hereon referred to as the *marker-null–enriched subtype*.

To investigate the robustness of the marker-null–enriched subtype in the selected consensus clustering solution, we applied histological annotations to the previously performed iterative consensus clustering (varying number of genes and cluster solutions). Reassuringly, the marker-null subtype was present in all analyses when six consensus clusters or more were used (Supplementary Fig. 3*A*). Similarly, performing consensus clustering in only the LCLC (WHO 2004) and LCNEC tumors also identified the marker-null cases as a separate distinct cluster (Supplementary Fig. 3*B*). Together, these results indicate that we identified, within a general lung cancer population, stable transcriptional subgroups describing LCLC and LCNEC in accordance with the WHO 2015 classification.

Molecular and Clinicopathological Characteristics of LCNEC Tumors

The LCNEC/SCLC-dominated tumor cluster (cluster 3 in Fig. 1) was on the transcriptional level characterized by high expression of neuroendocrine genes, proliferation-related genes, and gene clusters 11 and 8 metagene/Metacore (potentially representing an SRY-box 2 transcription factor–driven gene cluster based on Metacore analysis) and by lower expression of the napsin A/surfactant metagene and gene clusters 2 and 9 metagene/Metacore (potentially representing a MYC transcription factor driven gene cluster based on Metacore analysis) (Figs. 1 and 2 and Supplementary Fig. 4). In tumor cluster 3, tumor protein p53 gene (*TP53*) and *RB1* mutations were found in 91% and 82% of LCNEC tumors, respectively, with *RB1* mutations always concurrently with *TP53* mutations, and 91% of cases showed absent RB1 protein expression.

Only three LCNEC tumors were not present in the neuroendocrine cluster despite IHC expression of one or more tested neuroendocrine markers. Interestingly, two of three outlier cases showed distinct napsin A immunostaining, with the third case showing some focal positive cells (all three cases were TTF1 positive but without any clear histological adenocarcinoma component). All three outlier cases showed *TP53* mutations but positive RB1 protein expression, although one case (located in cluster 7) had a concurrent c.841C>A *RB1* mutation (COSM5658729) together with a serine/threonine kinase 11 gene (*STK11*) frameshift mutation (c.164_165insG). Notably, cluster 7 was strongly enriched for *STK11*-mutated adenocarcinomas (see Fig. 1). Consistently, the average expression of a 16-gene STK11 gene loss signature,[28] indicating serine/threonine kinase 11 inactivation, was higher in both LCNEC cases located in cluster 7 (see Fig. 2*B*). Finally, the three non–cluster 3 cases showed lower expression of the proliferation metagene and generally lower expression of the neuroendocrine metagene compared to the 11 LCNEC cases in tumor cluster 3 (see Fig. 2*C* and *D*).

Molecular and Clinicopathological Characteristics of LCLCs Reclassified as Adenocarcinoma and SqCC

As evident from Figure 1, adenocarcinoma-like or SqCC-like cases classified as LCLC according to the WHO 2004 guidelines did not form separate clusters. This finding indicates that despite their undifferentiated morphological features, these tumors retain transcriptional similarities with different reported adenocarcinoma or SqCC gene expression subtypes.[24,25]

Of the four LCLCs reclassified as SqCC (see Fig. 1), two grouped in a cluster dominated by secretory GEP–classified SqCCs[24] (cluster 4 in Fig. 1) whereas the remaining two cases fell in the neuroendocrine cluster (n = 1 [cluster 3]) and in the marker-null LCLC cluster (n = 1 [cluster 8]). Although our transcriptional analysis suggests that gene expression profiling might add to current SqCC IHC classification, the numbers are too low to allow any definite

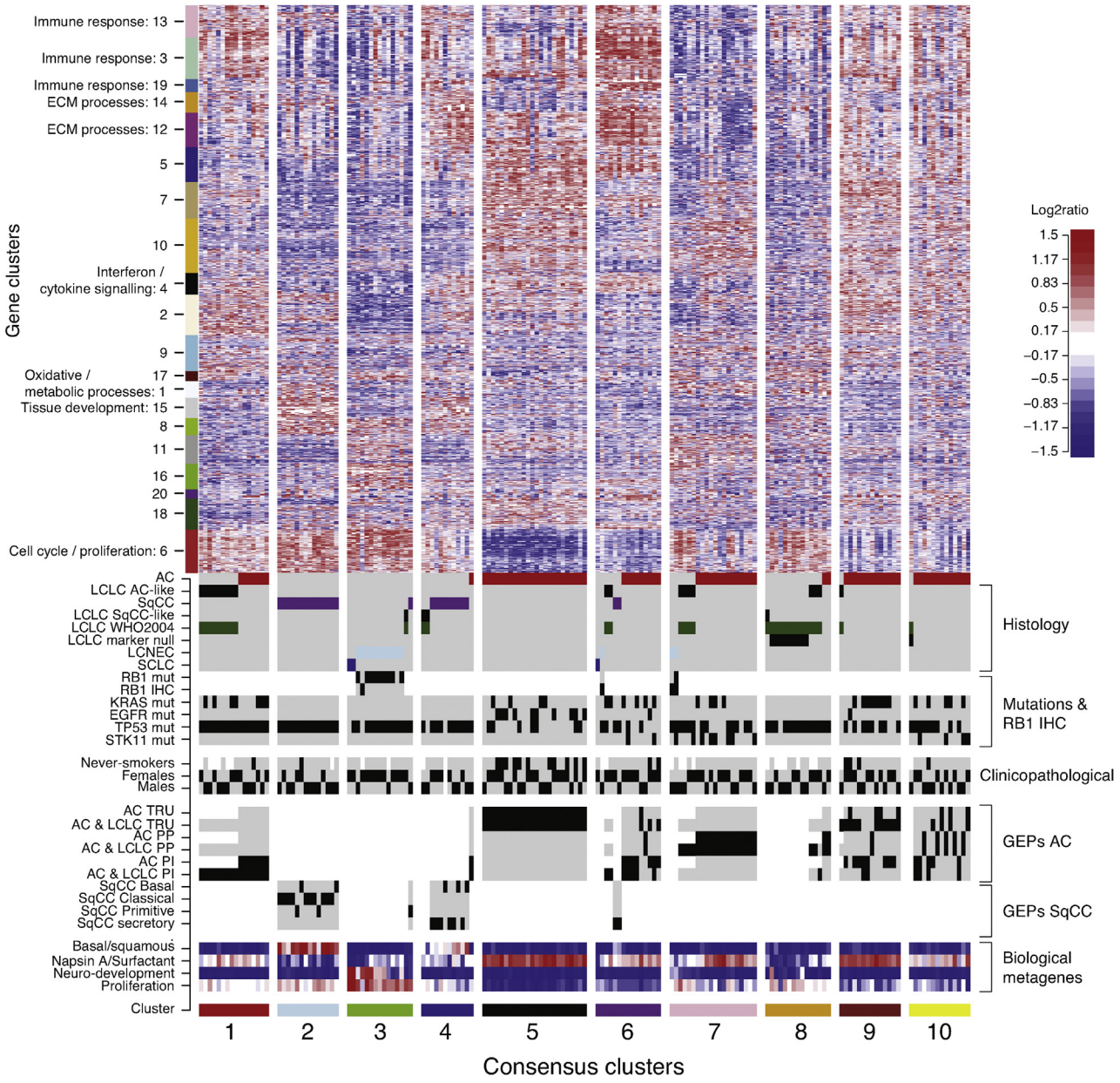

Figure 1. Unsupervised gene expression analysis stratifies large cell lung cancer (LCLC) and large cell neuroendocrine carcinoma (LCNEC) into molecular subgroups. Gene expression heatmap of 2730 Illumina probes across 159 lung cancers stratified by 10 specified consensus clusters. The 2730 probes correspond to a log2ratio standard deviation cutoff of more than 0.5 Annotations for histological subtypes, clinicopathological variables, selected mutations, retinoblastoma 1 immunohistochemistry (RB1 IHC), classification according to reported gene expression phenotypes (GEPs) for adenocarcinoma (AC) and squamous cell carcinoma (SqCC), and expression of selected biological metagenes are provided. For annotations, black corresponds to a positive/presence call, gray to a negative call, and white to not applicable or not available. Gene cluster functional annotations are provided for some specific clusters in the heatmap. ECM, extracellular matrix; mut, mutation; *RB1* mut, retinoblastoma 1 gene mutation; *TP53*, tumor protein p53 gene; *STK11*, serine/threonine kinase 11 gene; TRU, terminal respiratory unit; PP, proximal proliferative; and PI, proximal inflammatory.

conclusions to be drawn about the heterogeneity of SqCC- or SqCC-reclassified LCLC in this cohort.

For adenocarcinoma-like WHO 2004 LCLCs (n = 19), these clustered primarily in adenocarcinoma-dominated clusters (84% of cases [clusters 1, 6, 7, and 9 in Fig. 1]). Importantly, these clusters showed distinct clinicopathological, mutational, and GEP subtype characteristics shared by both the original adenocarcinomas and the reclassified WHO 2004 LCLCs. Three adenocarcinoma-like LCLCs fell in the LCLC marker-null–dominated gene expression cluster (cluster 8 in Fig. 1). However, in-depth analysis of

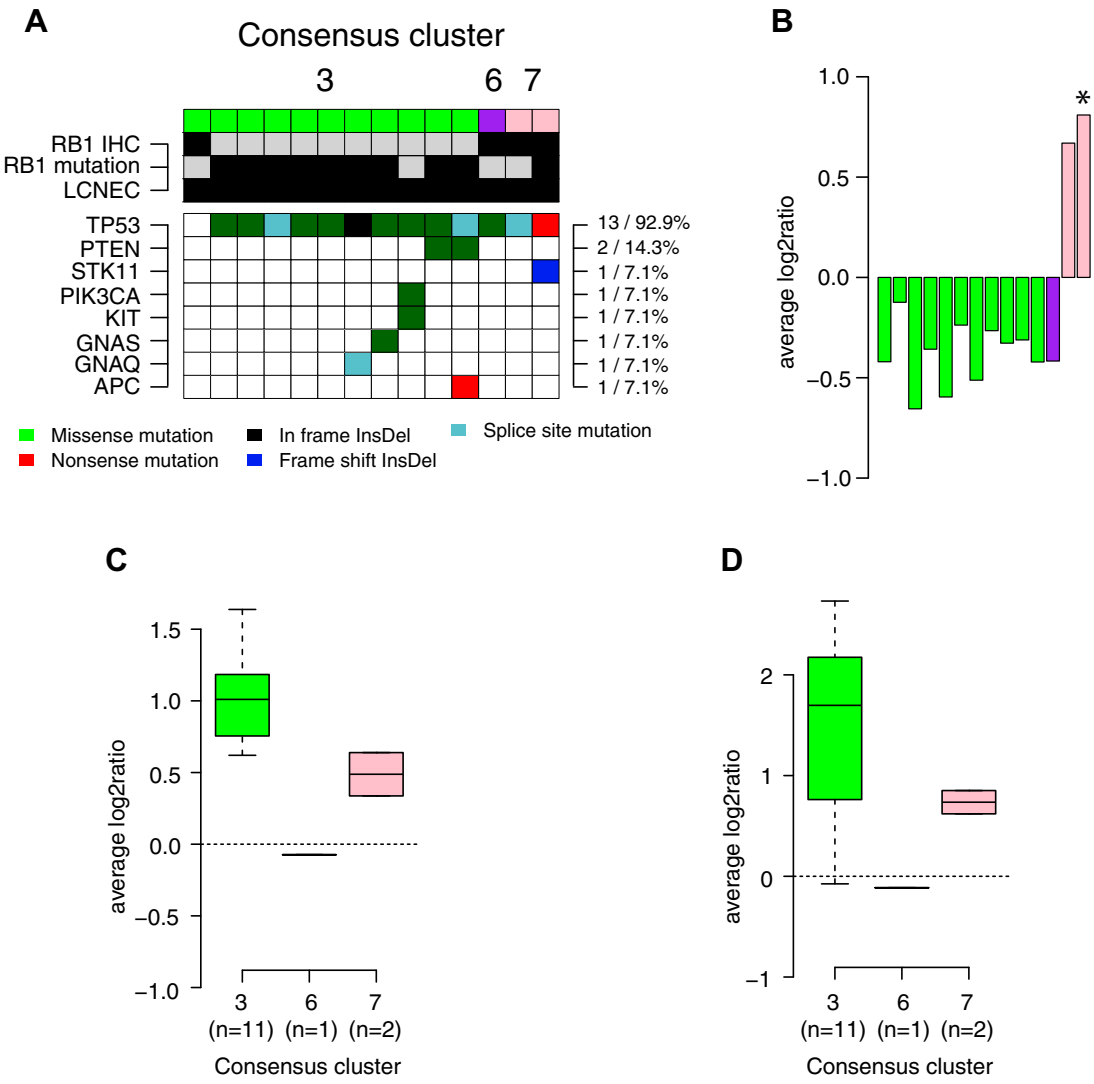

Figure 2. Mutations and gene expression characteristics of large cell neuroendocrine carcinoma (LCNEC) tumors with respect to consensus clusters. (*A*) Mutational map of detected mutations in 26 tumor suppressor and oncogenes for 14 LCNEC tumors stratified by consensus cluster assignment. Only genes with one or more mutations are shown. (*B*) Average log2ratio expression of a 16-gene serine/threonine kinase 11 gene (*STK11*) loss gene signature[28] for LCNEC tumors. Tumors are colored by their consensus cluster as in (*A*). Asterisk indicates a case with *STK11* mutation (*C*) Expression of the proliferation metagene for LCNEC tumors stratified by consensus cluster. (*D*) Expression of the neuroendocrine metagene for LCNEC tumors stratified by consensus cluster. Statistical testing in *B*–*D* was not performed on account of the small group sizes. In *A*–*D*, consensus cluster 3 cases are labeled green, cluster 6 cases purple, and cluster 7 cases pink. RB1 IHC, retinoblastoma 1 immunohistochemistry; *RB1* mutation, retinoblastoma 1 gene mutation; *TP53*, tumor protein p53 gene; *PTEN*, phosphatase and tensin homolog gene; *PIK3CA*, phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha gene; *KIT*, KIT proto-oncogene receptor tyrosine kinase gene; *GNAS*, GNAS complex locus gene; GNAQ, G Protein Subunit Alpha, Q; *APC*, adenomatous polyposis coli gene; APC, WNT signaling pathway regulator gene; InsDel, insertion/deletion.

mutational and gene expression patterns (expression of biological metagenes and gene cluster metagenes) in these few samples compared with adenocarcinoma-like LCLCs in other expression clusters failed to demonstrate any specific or clear differences.

Molecular and Clinicopathological Characteristics of Marker-Null LCLC

Nine out of 10 marker-null–classified LCLCs aggregated in a specific gene expression cluster, accounting for 60% of all tumors in this cluster (cluster 8 in Fig. 1). On the transcriptional level, this cluster was characterized mainly by high expression of proliferation-related genes (see Fig. 1 and Supplementary Fig. 4 for extensive comparison of metagenes and gene clusters). On the genomic level, besides *TP53* mutations in 90% of cases, only one *BRAF* mutation (p.Q456K), one phosphatase and tensin homolog gene (*PTEN*) mutation (p.F257L), and one APC, adenomatous polyposis coli gene (*APC*) mutation (p.E1080Ter) were found (Fig. 3*A*). A total of 445 genes were identified

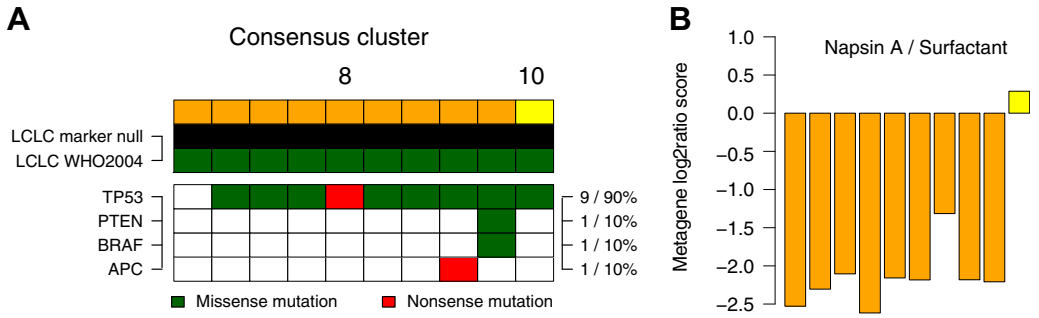

Figure 3. Genomic characteristics of large cell lung cancer (LCLC) marker-null cases. (*A*) Mutational map of detected mutations among 26 tumor suppressor and oncogenes in the 10 LCLC marker-null cases stratified by the 10 defined consensus clusters. Only genes with one or more mutations are shown. (*B*) Expression of the napsin A/surfactant biological metagene for the 10 marker-null cases stratified by consensus clusters. Colors of the individual samples correspond to cluster colors in (*A*). LCLC, large cell lung cancer; *TP53*, tumor protein p53 gene; *PTEN*, phosphatase and tensin homolog gene; *APC*, APC, adenomatous polyposis coli gene; WNT signaling pathway regulator gene.

by significance analysis of microarray to be differentially expressed between tumors in cluster 8 versus all other tumors (false discovery rate <1%). Consistent with our observations from biological metagenes and gene clusters, 79% of genes were down-regulated in cluster 8 tumors, whereas the 95 up-regulated genes (21%) were enriched for biological gene ontology processes such as nucleosome assembly, mitotic cell cycle, DNA replication, cell division, and cellular response to stress and oxidation-reduction processes (Panther overrepresentation test[29] [Bonferroni $p < 0.05$]).

The marker-null LCLC case that did not cluster in the marker-null expression cluster (instead located in cluster 10 in Fig. 1) was further analyzed for transcriptional characteristics by using biological metagene and gene cluster metagene expression patterns. One specific feature was observed, namely, higher expression of the napsin A/surfactant biological metagene than in the remaining cases in cluster 8 (Fig. 3*B*). This finding suggests that this case has some adenocarcinoma-like characteristics consistent with its clustering.

Validation of Transcriptional LCLC Subgroups

To validate our findings we analyzed a reported 199-sample NSCLC cohort profiled by RNA sequencing,[26] including 19 WHO 2004 LCLC cases and five LCNECs. WHO 2015 reclassification identified a similar proportion of LCLC marker-null cases in this cohort as in our discovery set (32% [n = 6]). The remaining cases (68%) were reclassified as adenocarcinoma (37%), SqCC (5%), adenosquamous (16%), or sarcomatoid (10%). To investigate the reproducibility of our discovery cohort findings, we created gene expression centroids for each consensus cluster in the discovery cohort (see Fig. 1) and classified the 199 tumors by a nearest centroid approach (Fig. 4 and Supplementary Table 1). Three of five LCNEC and four of six LCLC marker-null cases were classified into the LCNEC and LCLC marker-null clusters, respectively. The two LCNEC cases not in the neuroendocrine cluster did not express high mRNA levels of neuroendocrine marker genes. For the two outlier LCLC marker-null cases, one case was found in predicted cluster 10 (i.e., similar to the outlier in the discovery set), whereas the second was found in predicted cluster 3 (the neuroendocrine cluster) despite not showing increased neuroendocrine metagene expression (see Fig. 4).

To compare our supervised classification with unsupervised analysis of the validation cohort we performed an independent broad consensus clustering (similar to the discovery cohort). For higher standard deviations (using ~1000 to 2000 genes in the clustering), we did observe aggregation of marker-null LCLC tumors in specific clusters that was comparable to our classification results (≥67% of marker-null tumors in a single cluster). In these instances, marker-null cases comprised approximately one-third of cases in these clusters (see Supplementary Fig. 5). This finding is again similar to our results of classification of this cohort (LCLCs comprise 36% of tumors in predicted cluster 8). In contrast, in the unsupervised analysis of the validation cohort we did not find a similar enrichment of LCNEC tumors (<50% aggregation in a cluster by consensus clustering versus 60% for supervised classification) (see Supplementary Fig. 5).

Discussion

Previous extensive gene expression analyses have not discriminated LCLC as a distinct expression phenotype in NSCLC. Instead, LCLC tumors have often been dispersed in various adenocarcinoma, SqCC, or LCNEC subgroups (for example, see Clinical Lung Cancer Genome Project et al.,[7] Djureinovic et al.,[26] and Botling et al.[30]). In the current study, we explored the transcriptional and genomic spectrum of LCLC and LCNEC in the context of

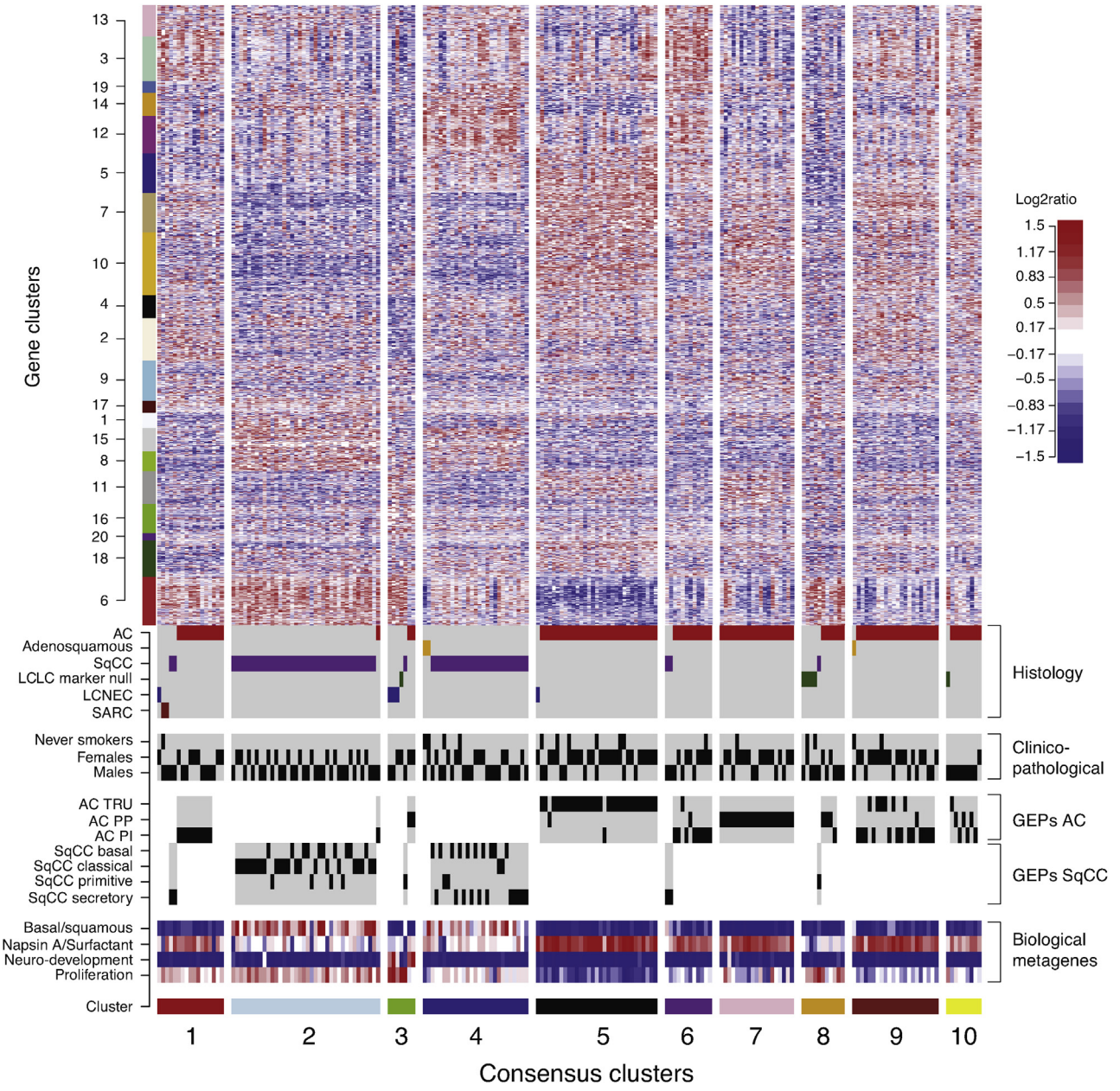

Figure 4. Validation of the large cell neuroendocrine carcinoma (LCNEC) and large cell lung cancer (LCLC) marker-null gene expression phenotypes (GEPs). Gene expression heatmap of 2196 genes across 199 lung cancers from Djureinovic et al.[26] classified by a nearest centroid predictor into 10 consensus clusters. Six cases were set as unclassified and are excluded from the heatmap (n = 193 cases in the heatmap). Annotations for histological subtypes, clinicopathological variables, classification according to reported GEPs, for adenocarcinoma (AC) and squamous cell carcinoma (SqCC), and expression of selected biological metagenes are provided. For annotations, black corresponds to a positive/presence call, gray to a negative call, and white to not applicable or not available. SARC, sarcomatoid, TRU, terminal respiratory unit; PP, proximal proliferative; PI, proximal inflammatory.

recent molecular findings, refined pathological classification schemes, and other histological subgroups of lung cancer. Through gene expression profiling of 159 primary lung tumors of all histological subtypes and subsequent independent validation in 199 additional cases, we found that the WHO 2015 guidelines translated into a better transcriptional subgrouping of LCLC. Our results endorse the recently refined definition of LCLC on a transcriptional level, demonstrating the presence of a GEP highly enriched for poorly differentiated highly proliferative tumors that do not express diagnostic immunomarkers for adenocarcinoma or SqCC.

An important conclusion from this study is that the updated WHO 2015 classification translates to the

transcriptional landscape, as the transcriptional phenotypes mimic the WHO classification on a general level. On a detailed level, transcriptional phenotypes likely provide additional stratification within histological subtypes that may be associated with prognosis and therapeutic options. Within both adenocarcinoma and SqCC different GEPs have been proposed (e.g., by Wilkerson et al.,[24] Wilkerson et al.,[25] Bhattacharjee et al.,[27] Takeuchi et al.,[31] Garber et al.,[32] and Raponi et al.[33]), although the consensus between phenotypes is not absolute in independent multicohort analysis.[34] Our data suggest that additional subgroup stratification appears possible in at least lung adenocarcinoma compared with the most used gene expression-based classification.[25,35]

The new WHO 2015 classification system has substantially reduced the proportion of LCLC (marker-null) NSCLC cases. The reclassification frequencies observed in the discovery and validation cohorts (70% and 68%, respectively) are consistent with previous studies (59%–90%).[3,14,16,36–40] Our transcriptional analysis of reclassified WHO 2004 LCLC cases demonstrates that these cases have heterogeneous profiles matching different molecular subsets/GEPs of adenocarcinoma and SqCC, providing additional biological information compared to current diagnostic immunomarkers. Generally, reclassified WHO 2004 LCLCs followed the more aggressive transcriptional phenotypes, such as proximal proliferative or proximal inflammatory in adenocarcinoma.[35] Specifically, no LCLC according to the WHO 2004 guidelines that has been reclassified as adenocarcinoma fell in the cluster characterized by *EGFR* mutations, never-smokers, and low-proliferative terminal respiratory unit–like tumors[35] (cluster 5). These observations appear consistent with the generally undifferentiated morphological state of WHO 2004 LCLC tumors and the well-known association of LCLC with smoking.

In this study, marker-null LCLC cases formed a specific, reproducible, transcriptional cluster linking the immunomarker classification with a transcriptional phenotype featuring characteristics of poorly differentiated proliferative cancer (see Figs. 1 and 4), which typically is linked to poorer patient outcome. In the validation cohort, the enrichment of marker-null tumors in the proposed phenotype was lower than in the discovery cohort (67% versus 90%, respectively). This might be due to low sample numbers in the former cohort causing strong proportional shifts by individual samples (16% and 20% shifts per sample for marker-null and LCNEC comparisons, respectively). The clearer identification of the marker-null phenotype in the discovery cohort may thus be due to a higher enrichment of marker-null tumors than in the validation cohort (6% versus 3%, respectively). Considering the overall low sample numbers, we acknowledge that additional validation is warranted.

Three main differentiation lineages (bronchioid/glandular, epidermoid/squamoid/squamous/keratinizing, and neuroendocrine) are generally recognized in lung cancer (for example, see the references in Pelosi et al.[14]). Recent mutational analyses of marker-null LCLC cases have suggested that genetic profiling could further reduce the marker-null group beyond IHC scoring by detection of typical adenocarcinoma or SqCC mutations, often favoring an adenocarcinoma lineage.[3,8,14,16] Although recent NGS-based studies[14,16] have reported a notably higher frequency of (especially) adenocarcinoma-linked mutations in marker-null cases than in our cohort, it may be noted that 83% of non–marker-null cases in our marker-null–enriched gene expression cluster (cluster 8) were adenocarcinoma or LCLCs according to the WHO 2004 guidelines that have been reclassified as adenocarcinoma. This observation may lend some support to a hypothesis that marker-null LCLCs represent a variant of undifferentiated TTF1-negative adenocarcinoma.[14,41] Clearly, deeper molecular/genetic characterization of marker-null LCLC may further reduce the WHO 2015 marker-null group, benefitting both clinical patient management and basic understanding of differentiation lineages in lung cancer.

On the basis of massive parallel sequencing studies, it is becoming evident that a subset of LCNEC tumors share mutational patterns with SCLC, whereas others carry mutations typically altered in non-neuroendocrine tumors.[7,9,11,12] Rekhtman et al.[9] recently hypothesized a genetic division of LCNECs into SCLC-like and NSCLC-like subgroups. The SCLC-like group was defined by concomitant *TP53* and *RB1* alterations, reflecting their ubiquitous inactivation in SCLC.[42] In contrast, the NSCLC-like subset was characterized by the lack of concomitant *TP53* and *RB1* alterations and occurrence of other NSCLC-type mutations (e.g., *KRAS* and *STK11* mutations). Interestingly, our data indicate that this stratification may potentially be mimicked also on the transcriptional level, providing a speculative link between the mutational and transcriptional landscape of LCNEC. An earlier gene expression study has suggested the presence of a good and poor outcome group in neuroendocrine tumors independent of LCNEC and SCLC status.[43] However, how these subgroups relate to the division of LCNEC by Rekhtman et al. is not clear,[9] as genetic data from the former study are not available. Acknowledging the limited number of LCNEC cases in the current study, consensus cluster 3 and non–cluster 3 cases still share intriguingly many features of the proposed SCLC-like and NSCLC-like LCNEC groups, respectively. These features include (1) frequent co-occurrence of *TP53* mutations and *RB1* mutations/protein loss in cluster 3, (2) higher

expression of proliferation-related genes and SCLC marker genes in cluster 3, (3) more pronounced expression of a potential SRY-box 2 (a suggested lineage survival oncogene in SCLC) transcription factor–driven gene cluster in cluster 3 (see Rudin et al.[44]), (4) indication of frequent *STK11* inactivation in non–cluster 3 cases, (5) distinct napsin A staining in two of three non–cluster 3 cases with few additional focal positive cells in the third case (all cases TTF1-positive), and (6) a trend of worse patient outcome in cluster 3 (with 55% of patients dead within 5 years versus 33% of non–cluster 3 patients). Notably, LCNEC has hitherto been consistently reported with negative napsin A expression,[4,45] and how this relates to the proposed NSCLC-like subgroup and whether the speculative association with transcriptional patterns indicated in the current study stands in larger validation studies remain to be determined.

In summary, the current study provides a novel link between the recent WHO 2015 diagnostic classification scheme and the transcriptional landscape of lung cancer. Our study confirms that WHO 2004–classified LCLCs may be refined by genomic patterns into clinically relevant subgroups that may have implications for diagnosis, predictive testing, and therapy decisions. Specifically, we demonstrate a gene expression profile that further defines a phenotype associated with poor patient outcome and comprises undifferentiated LCLC cases not expressing adenocarcinoma or SqCC markers. A continued search for molecular targets for therapeutic inhibition in WHO 2015 LCLC cases is highly warranted.

Acknowledgments

Financial support for this study was provided by the Swedish Cancer Society, the Mrs. Berta Kamprad Foundation, the Gunnar Nilsson Cancer Foundation, the Crafoord Foundation, BioCARE a Strategic Research Program at Lund University, the Gustav V's Jubilee Foundation, and governmental funding (ALF). Dr. Staaf, Mr. Karlsson, and Dr. Planck conceived of the study. Mr. Karlsson performed the experimental microarray and next-generation sequencing analyses with the assistance of Mr. Reuterswärd. Dr. Mattsson performed the validation experiments. Drs. Micke, Botling, Mr. La Fleur, and Mattsson provided the validation cases. Dr. Brunnström performed the pathological evaluations, and Drs. Brunnström, Jönsson, and Planck evaluated the immunohistochemical staining results. Mr. Karlsson and Dr. Staaf performed statistical analyses with support of Dr. Veerla. Dr. Staaf drafted and wrote the manuscript together with Mr. Karlsson. All authors approved the final manuscript.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the *Journal of Thoracic Oncology* at www.jto.org and at http://dx.doi.org/10.1016/j.jtho.2017.05.008.

References

1. Howlader N, Noone AM, Krapcho M, et al, eds. SEER cancer statistics review, 1975-2013. http://seer.cancer.gov/csr/1975_2013/. Accessed 18 Jan 2017.
2. Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, eds. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart*. Lyon, France: IARC Press; 2004.
3. Rekhtman N, Tafe LJ, Chaft JE, et al. Distinct profile of driver mutations and clinical features in immunomarker-defined subsets of pulmonary large-cell carcinoma. *Mod Pathol*. 2014;26:511–522.
4. Rossi G, Mengoli MC, Cavazza A, et al. Large cell carcinoma of the lung: clinically oriented classification integrating immunohistochemistry and molecular biology. *Virchows Arch*. 2014;464:61–68.
5. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG, eds. *WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart*. Lyon, France: IARC Press; 2015.
6. Rossi G, Cavazza A, Marchioni A, et al. Role of chemotherapy and the receptor tyrosine kinases KIT, PDGFRalpha, PDGFRbeta, and Met in large-cell neuroendocrine carcinoma of the lung. *J Clin Oncol*. 2005;23:8774–8785.
7. Clinical Lung Cancer Genome Project (CLCGP), Network Genomic Medicine (NGM). A genomics-based classification of human lung tumors. *Sci Transl Med*. 2013;5:209ra153.
8. Karlsson A, Brunnstrom H, Lindquist KE, et al. Mutational and gene fusion analyses of primary large cell and large cell neuroendocrine lung cancer. *Oncotarget*. 2015;6:22028–22037.
9. Rekhtman N, Pietanza MC, Hellmann MD, et al. Next-generation sequencing of pulmonary large cell neuroendocrine carcinoma reveals small cell carcinoma-like and non-small cell carcinoma-like subsets. *Clin Cancer Res*. 2016;22:3618–3629.
10. Karlsson A, Jonsson M, Lauss M, et al. Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome. *Clin Cancer Res*. 2014;20:6127–6140.
11. Miyoshi T, Umemura S, Matsumura Y, et al. Genomic profiling of large-cell neuroendocrine carcinoma of the lung. *Clin Cancer Res*. 2016;23:757–765.
12. Simbolo M, Mafficini A, Sikora KO, et al. Lung neuroendocrine tumours: deep sequencing of the four WHO histotypes reveals chromatin remodelling genes as major players and a prognostic role for TERT, RB1, MEN1 and KMT2D. *J Pathol*. 2017;241:488–500.
13. Hwang DH, Szeto DP, Perry AS, Bruce JL, Sholl LM. Pulmonary large cell carcinoma lacking squamous differentiation is clinicopathologically indistinguishable from solid-subtype adenocarcinoma. *Arch Pathol Lab Med*. 2014;138:626–635.

14. Pelosi G, Fabbri A, Papotti M, et al. Dissecting pulmonary large-cell carcinoma by targeted next generation sequencing of several cancer genes pushes genotypic-phenotypic correlations to emerge. *J Thorac Oncol*. 2015;10:1560–1569.
15. Righi L, Vavala T, Rapa I, et al. Impact of non-small-cell lung cancer-not otherwise specified immunophenotyping on treatment outcome. *J Thorac Oncol*. 2014;9:1540–1546.
16. Driver BR, Portier BP, Mody DR, et al. Next-generation sequencing of a cohort of pulmonary large cell carcinomas reclassified by World Health Organization 2015 criteria. *Arch Pathol Lab Med*. 2016;140:312–317.
17. Salomonsson A, Jonsson M, Isaksson S, et al. Histological specificity of alterations and expression of KIT and KITLG in non-small cell lung carcinoma. *Genes Chromosomes Cancer*. 2013;52:1088–1096.
18. Micke P, Mattsson JS, Djureinovic D, et al. The impact of the fourth edition of the WHO classification of lung tumours on histological classification of resected pulmonary NSCCs. *J Thorac Oncol*. 2016;11:862–872.
19. Brunnström H, Johansson L, Jirström K, Jönsson M, Jönsson P, Planck M. Immunohistochemistry in the differential diagnostics of primary lung cancer: an investigation within the Southern Swedish Lung Cancer Study. *Am J Clin Pathol*. 2013;140:37–46.
20. Lauss M, Visne I, Kriegner A, Ringnér M, Jönsson G, Höglund M. Monitoring of technical variation in quantitative high-throughput datasets. *Cancer Inform*. 2013;12:193–201.
21. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*. 2010;26:1572–1573.
22. Karlsson A, Ringnér M, Lauss M, et al. Genomic and transcriptional alterations in lung adenocarcinoma in relation to smoking history. *Clin Cancer Res*. 2014;20:4912–4924.
23. Planck M, Isaksson S, Veerla S, Staaf J. Identification of transcriptional subgroups in EGFR-mutated and EGFR/KRAS-wild type lung adenocarcinoma reveals gene signatures associated with patient outcome. *Clin Cancer Res*. 2013;19:5116–5126.
24. Wilkerson MD, Yin X, Hoadley KA, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin Cancer Res*. 2010;16:4864–4875.
25. Wilkerson MD, Yin X, Walter V, et al. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS One*. 2012;7:e36530.
26. Djureinovic D, Hallstrom BM, Horie M, et al. Profiling cancer testis antigens in non-small-cell lung cancer. *JCI Insight*. 2016;1:e86837.
27. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A*. 2001;98:13790–13795.
28. Kaufman JM, Amann JM, Park K, et al. LKB1 loss induces characteristic patterns of gene expression in human tumors associated with NRF2 activation and attenuation of PI3K-AKT. *J Thorac Oncol*. 2014;9:794–804.
29. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res*. 2016;44:D336–D342.
30. Botling J, Edlund K, Lohr M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis and tissue microarray validation. *Clin Cancer Res*. 2012;19:194–204.
31. Takeuchi T, Tomida S, Yatabe Y, et al. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. *J Clin Oncol*. 2006;24:1679–1688.
32. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. *Proc Natl Acad Sci U S A*. 2001;98:13784–13789.
33. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res*. 2006;66:7466–7472.
34. Ringner M, Staaf J. Consensus of gene expression phenotypes and prognostic risk predictors in primary lung adenocarcinoma. *Oncotarget*. 2016;7:52957–52973.
35. The Cancer Genome Atlas Network, Collisson EA, Campbell JD, et al. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543–550.
36. 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.
37. Barbareschi M, Cantaloni C, Del Vescovo V, et al. Heterogeneity of large cell carcinoma of the lung: an immunophenotypic and miRNA-based analysis. *Am J Clin Pathol*. 2011;136:773–782.
38. Monica V, Ceppi P, Righi L, et al. Desmocollin-3: a new marker of squamous differentiation in undifferentiated large-cell carcinoma of the lung. *Mod Pathol*. 2009;22:709–717.
39. Monica V, Scagliotti GV, Ceppi P, et al. Differential thymidylate synthase expression in different variants of large-cell carcinoma of the lung. *Clin Cancer Res*. 2009;15:7547–7552.
40. Pardo J, Martinez-Penuela AM, Sola JJ, et al. Large cell carcinoma of the lung: an endangered species? *Appl Immunohistochem Mol Morphol*. 2009;17:383–392.
41. Pelosi G, Fraggetta F, Pasini F, et al. Immunoreactivity for thyroid transcription factor-1 in stage I non-small cell carcinomas of the lung. *Am J Surg Pathol*. 2001;25:363–372.
42. George J, Lim JS, Jang SJ, et al. Comprehensive genomic profiles of small cell lung cancer. *Nature*. 2015;524:47–53.
43. Jones MH, Virtanen C, Honjoh D, et al. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet*. 2004;363:775–781.
44. Rudin CM, Durinck S, Stawiski EW, et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat Genet*. 2012;44:1111–1116.
45. Masai K, Tsuta K, Kawago M, et al. Expression of squamous cell carcinoma markers and adenocarcinoma markers in primary pulmonary neuroendocrine carcinomas. *Appl Immunohistochem Mol Morphol*. 2013;21:292–297.