Lung Squamous Cell Carcinomas with Basaloid Histology Represent a Specific Molecular Entity

Christian Brambilla1, Julien Laffaire2, Sylvie Lantuejoul3, Denis Moro-Sibilot4, Hélène Mignotte5, François Arb1, Anne-Claire Toffart1, Fabien Petel2, Pierre Hainaut6, Sophie Rousseaux3, Saadi Khochbin5, Aurélien de Reynies7, and Elisabeth Brambilla3

Abstract

Purpose: The basaloid carcinoma (pure) and the (mixed) basaloid variant of lung squamous cell carcinoma (SCC) have a dismal prognosis but their underlying specific molecular characteristics remain obscure and no therapy has proven to be efficient.

Experimental Design: To assess their molecular specificity among other lung SCCs we analyzed DNA copy number aberrations and mRNA expression pangenomic profiles of 93 SCCs, including 42 basaloid samples (24 pure, 18 mixed).

Results: Supervised analyses reveal that pure basaloid tumors display a specific mRNA expression profile, encoding factors controlling the cell cycle, transcription, chromatin, and splicing, with prevalent expression in germline and stem cells, while genes related to squamous differentiation are underexpressed. From this signature, we derived a 2-genes (SOX4, IVL) immunohistochemistry-based predictor that discriminated basaloid tumors (pure and mixed) from non-basaloid tumors with 94% accuracy in an independent series. The pure basaloid tumors are also distinguished through unsupervised analyses. Using a centroid-based predictor, the corresponding molecular subtype was found in 8 independent public datasets (n = 58/533), and was shown to be associated with a very poor survival as compared with other SCCs (adjusted HR = 2.45; P = 0.000001).

Conclusion: This study enlightens the heterogeneity of SCCs that can be subclassified in mRNA expression subtypes. This study demonstrates for the first time that basaloid SCCs constitute a distinct histomolecular entity, which justifies its recognition and distinction from non-basaloid SCCs. In addition, their characteristic molecular profile highlights their intrinsic resistance to cytotoxic chemotherapy and could serve as a guide for targeted therapies. Clin Cancer Res; 20(22); 5777–86. ©2014 AACR.

Introduction

Recent advances in genetic profiling revealed characteristic molecular alterations in adenocarcinomas (1–5) and squamous cell carcinomas (6, 7). Small-cell lung carcinomas (8) have also been revisited at molecular levels: genomic mutations, DNA copy number aberrations, and mRNA expression, uncovering genetic alterations with potentially driver functions.

However, the genetic landscape of rare tumors entities remains largely obscure. Basaloid carcinoma account for about 5% of non–small cell lung carcinomas (NSCLC) and were identified as a histopathologic entity with a dismal prognosis (9, 10). Among NSCLC and lung squamous cell carcinomas, the basaloid carcinoma (pure type) and the squamous variant of basaloid carcinoma (mixed basaloid SCC; WHO 2004) show the poorest prognosis (11). In the absence of specific molecular characteristics, basaloid carcinomas, in their pure form, were presented as a subtype of large-cell carcinoma (WHO 2004), while they very likely present distinct biologic characteristics, whose definition...
Translational Relevance

The basaloid carcinoma (pure) and the (mixed) basaloid variant of lung squamous cell carcinoma (SCC) show a dismal prognosis compared with non-basaloid SCCs. Their molecular characteristics are largely unknown. No validated molecular marker has yet been identified for the diagnosis of this aggressive entity. Here, we show for the first time that pure basaloid carcinoma constitute a distinct histomolecular entity, with a distinct signature related to specific pathways that are highly coherent with the poorly differentiated status and aggressiveness of these tumors. These molecular characteristics highlight their intrinsic resistance to cytotoxic chemotherapy and could serve as a guide for targeted therapies. Finally, we propose an immunohistochemistry-based molecular predictor, based on two markers (SOX4, IVL), which reliably discriminates pure and mixed basaloid tumors from non-basaloid tumors.

Materials and Methods

Patients and tumor samples

Ninety-three samples of SCCs of the lung, referred as the CIT (Cartes d’Identité des Tumeurs) cohort, obtained from 93 distinct patients, and collected from 1988 to 2006, were included in the present study. All tumor specimens were collected, stored, and used with the informed consent from the patients. All samples were independently reviewed according to the WHO criteria by two anatomopathologists: 100% concordance was obtained for the cases classified as basaloid carcinoma. For mixed basaloid carcinomas (basaloid component ≥ 50%), the two pathologists independently assessed the percentage of the basaloid component: their evaluation was concordant for 80% of the cases. A consensus was reached after a common review of the remaining cases.

To confirm the histopathologic diagnosis, all basaloid carcinoma cases were immunostained with antibodies against P40, P63, CK34betaE12 (recognizing the cytokeratins 1, 5, 10, 14), TTF-1, Chromogranin A, Synaptophysin, and CD56 (antibodies are described in Supplementary Table S1A). A tumor cell rate of more than 70% was used as criterion to select samples. This cohort was deliberately enriched in basaloid samples \( n = 42, 24 \) pure basaloid and 18 mixed), and the remaining samples \( n = 51 \) corresponded to non-basaloid SCC, including 36 well-differentiated SCC and 15 poorly differentiated SCC. Endoalveolar features were observed in 28 non-basaloid SCCs. Ninety-five percent of patients were males and the median age at diagnosis was 64 years (range: 44–82). All patients but 2 were current smokers \( n = 84 \) or former smokers \( n = 6 \). Forty non-basaloid SCCs (78%) and 10 basaloid SCCs (24%) were classified as T1N0M0. The main clinical and phenotypic characteristics of the patients are summarized in Supplementary Table S1B.

Immunohistochemistry-based predictor of basaloid histology

Four steps were followed to build and validate the predictor: (i) selection of discriminative markers using genome-wide expression data in a training set \( S1 \) (CIT series, \( n = 93 \) SCCs); (ii) measurement by immunohistochemistry (IHC) of the selected markers (IVL, SOX4) in a subpart \( S1’ \) of the training set \( S1 (n = 66 \) SCCs), including 26 basaloid SCCs (pure or mixed) and 40 non-basaloid SCCs (with no or less than 50% basaloid component) and in an independent validation series \( S2 (n = 15 \) basaloid SCCs + 10 non-basaloid SCCs + 10 lung adenocarcinomas); (iii) building of the predictor using immunohistochemical measures of the selected markers in the training set \( S1’ \); and (iv) validation of the predictor in the validation set \( S2 \). Of note, all cases included in the immunohistochemical series \( (S1’ + S2) \) were consensus cases (2 pathologists) and no borderline case was used, because our objective was to recognize...
basaloid carcinoma on the basis of the standard of morphologic classification. The tested antibodies are described in Supplementary Table S1A.

Detailed materials and methods are given in the Supplementary Data.

Results
Pure and mixed basaloid SCCs show the same dismal prognosis
The 93 samples of primary SCCs from the CIT cohort were classified according to the WHO criteria into 4 histologic classes, illustrated in (Fig. 1A–F): pure basaloid carcinoma (BAS_p, n = 24), mixed basaloid SCCs (BAS_m, n = 18), non-basaloid well-differentiated SCCs (SCC_wd, n = 36), and non-basaloid poorly differentiated SCCs (SCC pd, n = 15). All basaloid cases expressed P40, P63, and cytokeratins 1, 5, 10, and 14 but did not express TTF-1 and neuroendocrine markers (19).

Pure and mixed basaloid tumors showed the same poor prognosis in our dataset (Fig. 1G), indicating that the presence of a basaloid cellular contingent is of most interest for prognosis. Misclassified mixed basaloid samples could explain the lack of prognostic significance of the basaloid feature in previous studies (20).

Figure 1. Pathology review. A, basaloid carcinoma: tumor cells are small, monomorphic, cuboidal, or fusiform, with a scant cytoplasm and peripheral palisading. B, mixed basaloid carcinoma (SYN. Basaloid variant of squamous cell carcinoma): basaloid component on the right field, squamous cell component on the left field. C, well-differentiated SCC: squamous differentiation and keratinization are frequent and obvious. D, poorly differentiated SCC: most of the tumor shows absence of definite differentiation with few intercellular bridges and keratinization. E and F, PEA SCC: the SCC shows predominant intra-alveolar topography entrapping numerous strands of type 1 alveolar residual pneumocytes (E). G, Kaplan–Meier curves of overall survival (OS) in the CIT cohort, stratified by subhistology. P values refer to log-rank tests comparing OS between subhistologies.
Pure basaloid SCCs show a specific transcriptome profile related to the deregulation of specific pathways

Expression profiles of 42 basaloid (24 BAS_p, 18 BAS_m) and 51 non-basaloid SCCs (36 SCC_wd, 15 SCC_pd) were obtained. A high proportion of genes were found differentially expressed between BAS_p and non-basaloid SCCs (H1 proportion = 47%; Supplementary Table S2A), indicating a specific mRNA profile for BAS_p. Significantly upregulated genes in BAS_p (limma q value < 0.05) included genes related to TP53 mutation signature (TMSB15A, PROM1), transcription factors (SOX4, SOX9, SOX11, MYB, E2F3, E2F5), embryonic development (FGF3, FGF19), methylation regulation (TET1, DNMT1, DNMT3A), cell cycle (MKI67, RUB1, DTL), splicing (SFRS1, 2, 3), and survival (BCL2), whereas the most downregulated genes were related to epithelial cell and keratinocyte differentiation (KRT6, IVL, SPRR genes, and SFN). In addition, the recently defined ectopic male germ cell/placenta–specific gene expression signature, indicative of very aggressive lung tumors (21), also highly correlated with pure basaloid SCCs in our cohort ($\chi^2; P < 0.0005$, Supplementary Data). An extensive pathway analysis also confirmed these findings (Supplementary Table S3). The comparison between pure basaloid and non-basaloid SCCs either well-differentiated or poorly differentiated (Fig. 2), confirmed the upregulation of pathways related to cell cycle, transcription factors, mRNA splicing, and chromatin modifications and the downregulation of the squamous differentiation pathway. It also revealed the upregulation of signatures associated with testis and embryonic stem (ES) cells and poorly differentiated tumor markers (NANOG, OCT4, SOX2, and c-MYC targets) and the downregulation of signatures related to the Polycomb gene silencing system (SUZ12, EED, H3K27ME3, and PRC2 gene sets; Supplementary Fig. S1). Similarly, mixed basaloid were compared with non-basaloid SCCs (H1 proportion = 20%) and as in pure basaloid tumors, pathways related to cell cycle, spliceosome, and ES cells were upregulated (Supplementary Fig. S1).

A 2-genes immunohistochemistry-based predictor discriminates pure and mixed basaloid tumors from non-basaloid tumors

To identify molecular markers of basaloid tumors, we selected overexpressed transcripts showing high AUC (>75%) and specificity (>95%), while optimizing sensitivity (cutoff yielding a Fisher test $P < 1e-6$; Supplementary Table S2A). This selection yielded 5 genes (SOX4, CBX5, PATZ1, RBM6, SFRS3), all showing a higher sensitivity in pure basaloid than in mixed basaloid tumors. The transcription factor SOX4 showed 100% specificity and 50% sensitivity.

### Table

<table>
<thead>
<tr>
<th>Signatures</th>
<th>Up</th>
<th>Down</th>
<th>Wilcoxon test</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metastasis</td>
<td></td>
<td></td>
<td></td>
<td>8.5E-9</td>
</tr>
<tr>
<td>Chromatin</td>
<td></td>
<td></td>
<td></td>
<td>1.0E-6</td>
</tr>
<tr>
<td>Splicing</td>
<td></td>
<td></td>
<td></td>
<td>5.1E-6</td>
</tr>
<tr>
<td>Testis</td>
<td></td>
<td></td>
<td></td>
<td>4.2E-5</td>
</tr>
<tr>
<td>Cell cycle</td>
<td></td>
<td></td>
<td></td>
<td>5.6E-4</td>
</tr>
<tr>
<td>Transcription factors</td>
<td></td>
<td></td>
<td></td>
<td>2.4E-6</td>
</tr>
<tr>
<td>Methylation</td>
<td></td>
<td></td>
<td></td>
<td>3.6E-4</td>
</tr>
<tr>
<td>NANOG-OCT4-SOX2 TARGETS</td>
<td></td>
<td></td>
<td></td>
<td>2.2E-3</td>
</tr>
<tr>
<td>Senescence</td>
<td></td>
<td></td>
<td>ns</td>
<td>2.6E-2</td>
</tr>
<tr>
<td>Keratinocyte</td>
<td></td>
<td></td>
<td>ns</td>
<td>3.5E-2</td>
</tr>
<tr>
<td>TP53 targets</td>
<td></td>
<td></td>
<td>ns</td>
<td>4.9E-2</td>
</tr>
<tr>
<td>Lung</td>
<td></td>
<td></td>
<td>ns</td>
<td>3.9E-4</td>
</tr>
<tr>
<td>Drug metabolism</td>
<td></td>
<td></td>
<td>ns</td>
<td>2.3E-5</td>
</tr>
</tbody>
</table>

![Figure 2](image-url)

Figure 2. Deregulated pathways in pure basaloid tumors. Illustration of 13 signatures/pathways found among the most deregulated ones between pure basaloid tumors (black) and non-basaloid tumors (poorly differentiated SCC (light gray); well-differentiated SCC (dark gray)) in our dataset. Samples ($n = 75$) are independently ordered for each signature, on the mean expression of the corresponding signature. References of the signatures and pathways are indicated in Supplementary Table S3B. The Wilcoxon test $P$ values for upregulated (left) and downregulated (right) mean expression in basaloid versus non-basaloid SCCs are shown for each pathway.
sensitivity to discriminate basaloid tumors in our cohort; it was thus selected for further analyses.

To identify negative markers of basaloid tumors, we selected underexpressed transcripts showing high AUC (>75%) and sensitivity (>90%), while optimizing specificity (cutoff yielding a Fisher test \(P < 1e^{-5}\); Supplementary Table S2A): this selection yielded 5 genes (IVL, KRT16, TOM1, C10orf224, KCNK6). The squamous differentiation marker IVL (involutin), showing a high fold change (>2) between non-basaloid and basaloid SCCs, was selected for further analyses.

We measured by IHC the protein expression of SOX4 and IVL in a training set of 66 tumors common to the CIT transcriptome series (26 basaloid + 40 non-basaloid SCCs). Using this training set, we built a predictor based on the quick scores (QS) related to these two genes, with the following formula: if \(QS(SOX4) \geq 110\) then Basaloid; if \(QS(SOX4) < 50\) then Non-basaloid; if \(QS(SOX4) \in [50;110]\) and QS (SOX4) - QS(IVL) \(\geq -55\) then Basaloid else Non-basaloid (Fig. 3 and Supplementary Table S2B). In the training set, the predictor correctly classified all basaloid tumors (pure and mixed) and correctly classified 90% of the SCC samples. The predictor was then applied to an independent validation series of 35 tumors (15 basaloid + 10 non-basaloid SCCs + 10 adenocarcinomas), where it correctly classified all samples except 2 mixed basaloid tumors (classified as non-basaloid). The accuracy of this predictor was found to be 94% in the validation series (sensitivity: 87%; specificity: 100%; positive predictive value: 100%), in line with the performances observed in the training series (accuracy: 94%; sensitivity: 100%; specificity: 90%; positive predictive value: 87%).

**SCC tumors mostly share a similar genomic aberration profile.**

DNA from 27 basaloid (14 Bas_p, 13 Bas_m) and 37 non-basaloid SCCs (25 SCC_wd, 12 SCC_pd) were hybridized on Illumina HumanCNV370 SNP arrays. The most frequent copy number aberrations (CNA) on the whole dataset were identified with GISTIC2.0. This analysis identified previously reported CNA (7) such as gains of 3q, 5p, and 8q and losses of 1p, 3p, 4p, 4q, 5q, 6p, and 9p. Similarly, it pointed out known target genes, including gains of SOX2, MYC, CCND1, MDM1, and FGFR1 and losses of CDKN2A, PCDH10, RB1, PTEN (Supplementary Fig. S2). Very few CNAs were found in significantly different proportions among histologic classes (Supplementary Table S4). At the single gene level, gains of MYB, JUN, FGFR1, PK3C3, DSC/DSC genes were found more frequent in pure basaloid tumors. However, none of these differences reached significance after correction for multiple testing. From these data we were not able to identify CNA being specific of basaloid SCCs, either pure or mixed.

**Consensus unsupervised clustering identifies a poor prognosis molecular subtype corresponding to pure basaloid SCCs.**

To assess whether basaloid tumors could correspond to a molecular subtype obtained without a priori knowledge, we performed unsupervised analyses of the mRNA expression profiles. We identified consensus partitions in \(k = 2–8\) clusters of the 93 SCC expression profiles from the CIT cohort. The consensus partition in \(k = 4\) clusters was selected (Fig. 4A) based on the gap statistic method (22; Supplementary Fig. S3A). The underlying coclassification matrix showed a great level of agreement between the raw data and the predicted classification.

---

**Figure 3. IHC-based predictor of the basaloid entity.** A, to discriminate basaloid from non-basaloid tumors, the following predictor is used, based on the QS values of SOX4 and IVL: if QS (SOX4) \(\geq 110\) then Basaloid; if QS (SOX4) < 50 then Non-basaloid; if QS(SOX4) in [50;110] and QS (SOX4) - QS(IVL) \(\geq -55\) then Basaloid else Non-basaloid. B, pie charts showing the results obtained (red, predicted basaloid; gray, predicted non-basaloid) when applying this predictor to the tumors from the training series (top part) and the validation series (bottom part), in basaloid (left) and non-basaloid (right) tumors.

C, illustration of the immunohistochemical measures of SOX4 (top part) and IVL (bottom part) in a basaloid (left part) and a non-basaloid SCC (right part) tumor.
partitions in $k = 4$ clusters obtained using the different experimental settings (Fig. 4B). This partition was significantly associated with histology (Fisher, $P = 1e$–$8$). The 4 clusters were named basaloid-like ($BL; n = 21$), peripheral endoalveolar (PEA; $n = 29$), Classical_1 ($n = 21$), and Classical_2 ($n = 22$).

The BL cluster contained almost only basaloid tumors (90%), and mostly contained pure basaloid tumors (72%), contrary to other clusters (<15%). It also showed enrichment in tumors with stage 2 or higher (62%; Fig. 4A). Its expression pattern was the most singular compared with other clusters (Supplementary Fig. S3C). Tumors from the PEA cluster mostly showed an alveolar contingent in their microenvironment (83%), most were stage 1 (66%) without basaloid features (79%). The PEA cluster was found enriched in non-basaloid poorly differentiated SCCs (Fisher, $P = 3e$–$4$) tumors with a high (>75%) diploid tumor cell rate (Fisher, $P = 1e$–$5$), and TP53 wild-type tumors (Fisher, $P = 0.02$). Non-basaloid well-differentiated SCCs were mostly found in the Classical_1 and Classical_2 clusters (67% and 45%) and absent in the BL cluster (0%). Tumors with exon 4 TP53 mutations were found enriched in the Classical_1 group (Fisher, $P = 0.005$). Mixed basaloid tumors were not associated with any particular cluster.

To assign independent SCC datasets to these four molecular subtypes, we built a 139-genes nearest-centroid predictor (Supplementary Table S5), and classified all SCC...
samples \( (n = 533) \) of 8 public expression profiling datasets (Supplementary Table S6A). Samples from each of the four subtypes were found in all datasets (Supplementary Table S6A and S6B). Overall, 58 samples (11\%) were classified in the BL subtype, 152 (28\%) in the PEA subtype, 215 (41\%) in the Classical_1 subtype, and 108 (20\%) in the Classical_2 subtype. A significantly poorer prognosis was observed in the BL subtype compared with other molecular subtypes in the validation datasets (log-rank, \( P < 6e^{-8} \); Fig. 4C) and in the discovery cohort (Supplementary Fig. S4). Among the 3 other subtypes, Classical_2 and PEA subtypes presented similar outcome and Classical_1 showed a significant better prognosis. These results were conserved in stage 1 tumors (Fig. 4D).

<table>
<thead>
<tr>
<th>Pathway</th>
<th>CIT</th>
<th>WI</th>
<th>LEE</th>
<th>RA</th>
<th>RO</th>
<th>CIT</th>
<th>LEE</th>
<th>RA</th>
<th>RO</th>
<th>CIT</th>
<th>LEE</th>
<th>RA</th>
<th>RO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basaloid-like</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classical_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classical_2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PeriEndoAlveolar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Heatmap of most deregulated pathways in the 4 molecular subtypes, in the CIT cohort and 4 validation datasets. Heatmap representing the mean gene expression level of 31 signatures and pathways (rows) found differentially deregulated (see Materials and Methods) across molecular subtypes in our dataset and 4 validation datasets (columns) by pathway analysis (Supplementary Table S7). Sample groups are ordered according to their molecular subtypes and dataset membership (top annotation): CIT, discovery dataset; WI, Wilkerson et al. dataset; LEE, Lee et al. dataset; RA, Raponi et al. dataset; RO, Roepman et al. dataset. NTL_B and NTL_L refer to non-tumoral bronchi and lung samples from the CIT cohort. References of signatures and pathways are indicated in Supplementary Table S7B. Color scale: blue/white/red gradient corresponding to minimum/intermediate/maximum value for each row including non-tumoral samples.
Pathways found consistently deregulated in the BL cluster across datasets (Fig. 5 and Supplementary Table S7) were highly coherent with those identified in pure basaloid tumors compared with non-basaloid SCCs (Fig. 2 and Supplementary Table S3). As expected, normal bronchial basal cell signatures supposed to be the stem cell of bronchi (23) were found highly and significantly overexpressed in the BL subtype, whose expression profile was found clearly distinct from that of normal lung (alveoli) thus confirming the proposed derivation of basaloid carcinoma from a basal stem cell progenitor at the time of first description (9).

Finally, we applied a SCC classification system, recently published by Wilkerson and colleagues (6), to all SCC profiles from the discovery and validation cohorts (n = 626). Overall, the association between the CIT and Wilkerson classification systems was found to be very high (all series; Fisher test, P ≈ 0). Wilkerson subtypes were also found very associated with histology (CIT series; Fisher test, P = 2e−9). To match subtypes of both systems, we then used pairwise Cohen Kappa statistics. A fair and good agreement was respectively observed between BL and primitive subtypes (κ = 0.56) and PEA and secretory subtypes (κ = 0.72). Other subtypes, related to non-basaloid well-differentiated SCCs, showed a poor agreement (κ ≤ 0.40). Despite the fair agreement between the BL (n = 78/626; 12%) and the primitive (n = 104/626, 17%) subtypes, the BL subtype was far more associated with overall survival on the validation cohort (dataset adjusted HR = 2.45, 95%CI = (1.72–3.5); Wald P < 1e−6; C-index = 0.67, 95%CI = (0.52–0.73)) than the primitive subtype (dataset adjusted HR = 1.38; 95%CI = (1.00–1.89); Wald P = 0.04; C-index = 0.58, 95%CI = (0.45–0.70)). Concerning subtypes related to non-basaloid well-differentiated SCCs, contrary to that of Wilkerson, our classification reveals a subtype [Classical_1, n = 240/626 (38%)] showing a higher overall survival rate than other subtypes in the validation cohort [dataset adjusted HR = 0.61, 95%CI = (0.47–0.79), Wald P = 2e−4; C-index = 0.40, 95%CI = (0.30–0.51); Supplementary Fig. S5], even after excluding the BL subtype [dataset adjusted HR = 0.69; 95% CI = (0.53–0.95); Wald P = 7e−3].

In conclusion, pure basaloid tumors correspond to a specific molecular subtype, either in our classification system or in that of Wilkerson. This finding further demonstrates that histologically pure basaloid tumors constitute a specific molecular entity.

Discussion

The worse prognosis of basaloid SCCs as compared with non-basaloid SCCs suggests the existence of underlying biologic differences. Incidentally, we show here that mixed basaloid SCCs, representing 45% of the basaloid tumors in our series, share the same poor prognosis as pure basaloid SCCs, meaning that the presence of basaloid contingents in SCCs is informative for prognosis. To identify the molecular characteristics of basaloid SCCs, we analyzed mRNA expression and DNA CNA profiles of a large cohort of this rare entity (n = 42) with tumor controls (51 non-basaloid SCCs). Contrary to pure basaloid SCCs, mixed basaloid SCCs are expected to be heterogeneous at the molecular level. Thus, we mainly concentrated on pure basaloid tumors to identify molecular specificities of this histologic group. Supervised analyses revealed that pure basaloid SCCs display a specific mRNA expression profile as compared with non-basaloid SCCs. Enrichment in TP53 mutation signature, upregulation of transcription, epigenetic, cell cycle, splicing, and survival factors, as well as that of male germ cells, embryonic development, stemness/poor differentiation genes characterized pure basaloid SCCs, which also downregulated keratinocyte differentiation genes. Mixed basaloid SCCs also showed upregulation of proliferation and embryonic development–related genes as compared with non-basaloid SCCs, but to a lesser extent than pure basaloid SCCs, likely due to their heterogeneous cellular content. These molecular observations are in line with the poorly differentiated status and aggressiveness of the basaloid SCCs.

DNA CNAs were found non-informative to distinguish between SCC histologic subgroups. All SCCs, either basaloid or not, showed very similar CNA profiles, pointing at regions previously reported (7), such as gain or amplicon of the 3q region around SOX2, found in almost all SCC samples. However, CNA are potentially highly informative concerning targeted therapies (24). In particular, FGFR1 (8p12) and MYB (6q22-q23), found here in peak regions of gain, are targetable: MYB could be targeted by several drugs under development (25, 26), and FGFR1 amplified NSCLCs have been shown to be sensitive to FGFR1 inhibitors (27–28).

By unsupervised analyses of SCC mRNA expression profiles, we identified four molecular subtypes: BL, PEA, Classical_1, and Classical_2. The BL cluster was shown highly associated with (pure) basaloid histology. Similarly, among Wilkerson molecular subtypes (primitive, secretory, basal, classical), the primitive cluster was found associated with basaloid histology, a key characteristic that was not reported by Wilkerson and colleagues (6). Altogether, analysis of the molecular subtypes obtained either from our series or the literature unambiguously shows that pure basaloid SCCs do correspond to a specific molecular subtype.

Our classification system, applied to a large validation cohort via a centroid-based predictor, outperforms that of Wilkerson in predicting overall survival. The CIT BL subtype, while agreeing with Wilkerson primitive subtype, is far more associated with a poor prognosis. A substantial proportion of the samples classified as primitive were predicted as non-BL (44%). Expression-based analyses support that these samples could be more differentiated than BL primitive samples (data not shown). Moreover, contrary to Wilkerson, among well-differentiated SCCs, we identify a subtype (Classical_1) showing a significantly better prognosis. These clinical observations suggest that our subtypes are more homogeneous at the molecular level than that of Wilkerson. Indeed, assuming that tumor molecular characteristics may greatly determine prognosis, then defining...
more homogeneous molecular subgroups may yield greater prog nostic differences between subgroups. In particular, it supports that the BL subtype is more specifically related to basaloid SCCs than the primitive subtype of Wilkerson, even if not significantly different in the discovery cohort (NB, unavailable data in validation series).

Interestingly, most of the poorly differentiated non-basaloid SCCs of our series (11/15) were found in the PEA subtype, shown to be very similar to Wilkerson secretory subtype. Contrary to Wilkerson, suggesting secretory properties for this tumor subtype, our study reveals that these tumors show a PEA microenvironment, which much more likely explains the "secreting" signatures of this group. Moreover comparison of PEA profiles with that of bronchi and lung controls unambiguously shows that the PEA subtype—contrary to all other subtypes—has much more in common with lung profiles (which include alveoli) than with bronchi profiles. From our data, it seems excluded that PEA could be driven by non-tumor cell–related artifacts, both because a tumor cell rate higher than 70% criterion was strictly applied within our series, and that the percentage of tumor cells was not found different across subtypes. Interestingly, this subtype was characterized by a smaller fraction of tumors with genome instability and a smaller TP53 mutation rate. Accordingly, pathway analysis revealed upregulation of TP53 targets signatures and down-regulation of cell cycle in this subtype (Supplementary Data). Upregulation of immune signatures was also found to be very significantly associated with the PEA subtype.

Our study also completely renews the molecular subtyping of non-basaloid well-differentiated SCCs. In the discovery cohort, the Classical_1 subtype is specifically mutated in the exon 4 of TP53 (NB, unavailable information in validation series). Pathway analysis across 5 series (Supplementary Data) revealed that TP53 target genes are more expressed in Classical_1 than in Classical_2 and BL subtypes, while having a similar TP53 mutation rate. These observations suggest that mutations of the fourth exon of TP53 could be less deleterious than mutations of other exons observed within our series. Biologic consequences of a more functional p53 protein are also supported by pathway analysis as we found a relative higher expression of proapoptotic genes and lower expression of cell-cycle genes in Classical_1 tumors as compared with BL and Classical_2 subgroups. Finally, a clinical support of these biologic hypotheses is given by the better prognosis observed in the Classical_1 subtype both in our cohort and in the validation series.

Mixed basaloid SCCs were not found associated with any particular molecular subtype, as expected due to their cellular heterogeneity. Accordingly, mRNA expression profiles yielded very specific but moderately sensitive markers for the identification of basaloid SCCs as a whole, due to a lack of sensitivity concerning mixed basaloid SCCs. Using IHC-based markers we could overcome this limitation: we derived a 2-genes predictor based on a positive (SOX4) and a negative (IVL) marker of the basaloid SCCs, which showed a great accuracy (94%) in an independent validation set (n = 35). This simple predictor, used in addition to histologic examination, should considerably help the identification of basaloid tumors in clinical routine.

Of note, none of the present data on basaloid carcinoma (pure and mixed) are applicable to large-cell carcinoma (NOS) of the WHO 2004 classification nor to the revised one [to be published in 2015 (WHO classification: tumors of the lung, pleura, thymus, mediastinum and heart. Ed.: William D. Travis, Elisabeth Brambilla, Andrew Nicholson, Alexander Marx, Allen Burke)], because a thorough revision of the concept of large-cell carcinoma based on immunohistochemical differentiation markers and on the last genomic-based proposed classification of lung cancer restricted this class to a few cases with no clear differentiation phenotype, which is not the case for basaloid carcinoma.

In conclusion, our results establish that pure basaloid SCCs correspond to a specific molecular entity, fully justifying its histologic recognition and distinction from non-basaloid SCCs. Related deregulated pathways enlighten its intrinsic resistance to cytotoxic chemotherapy and should serve as a guide to targeted therapies.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors' Contributions
Conception and design: C.G. Brambilla, A. de Reynies, E. Brambilla
Development of methodology: C.G. Brambilla, A. de Reynies, E. Brambilla
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C.G. Brambilla, S. Lantuejoul, D. Moro-Sibilot, H. Nagy-Mignotte, F. Arbib, P. Hainaut, E. Brambilla
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C.G. Brambilla, J. Laffaire, F. Petel, P. Hainaut, S. Rousseaux, S. Khochbin, A. de Reynies, E. Brambilla
Writing, review, and/or revision of the manuscript: C.G. Brambilla, J. Laffaire, D. Moro-Sibilot, A.-C. Toffart, S. Rousseaux, A. de Reynies, E. Brambilla
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C.G. Brambilla, H. Nagy-Mignotte, F. Petel, E. Brambilla
Study supervision: C.G. Brambilla, A. de Reynies, E. Brambilla
Other (pathology review of cases): E. Brambilla

Grant Support
This work is part of the program CIT (http://cit.ligue-cancer.net) funded and developed by the Ligue Nationale Contre le Cancer. This study was supported by grants from Curelung European Project (2010–2013), PNEs cancers du poumon from Institut National Du Cancer (2010), Conseil Scientifique Agrédom, projet libre from Association pour la Recherche sur le Cancer (2010, Saadi Khochbin), and PHRC Biomarck (2003).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 4, 2014; revised June 25, 2014; accepted July 27, 2014; published OnlineFirst September 4, 2014.

References
Lung Squamous Cell Carcinomas with Basaloid Histology Represent a Specific Molecular Entity

Christian Brambilla, Julien Laffaire, Sylvie Lantuejoul, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-14-0459

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/09/05/1078-0432.CCR-14-0459.DC1

Cited articles
This article cites 28 articles, 10 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/20/22/5777.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/20/22/5777.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.