Integrative genomic profiling of large-cell neuroendocrine carcinomas reveals distinct subtypes of high-grade neuroendocrine lung tumors

Julie George et al.#

Pulmonary large-cell neuroendocrine carcinomas (LCNECs) have similarities with other lung cancers, but their precise relationship has remained unclear. Here we perform a comprehensive genomic (n = 60) and transcriptomic (n = 69) analysis of 75 LCNECs and identify two molecular subgroups: “type I LCNECs” with bi-allelic TP53 and STK11/KEAP1 alterations (37%), and “type II LCNECs” enriched for bi-allelic inactivation of TP53 and RB1 (42%). Despite sharing genomic alterations with adenocarcinomas and squamous cell carcinomas, no transcriptional relationship was found; instead LCNECs form distinct transcriptional subgroups with closest similarity to SCLC. While type I LCNECs and SCLCs exhibit a neuroendocrine profile with ASCL1^high/DLL3^high/NOTCH1^low, type II LCNECs bear TP53 and RB1 alterations and differ from most SCLC tumors with reduced neuroendocrine markers, a pattern of ASCL1^low/DLL3^low/NOTCH1^high, and an upregulation of immune-related pathways. In conclusion, LCNECs comprise two molecularly defined subgroups, and distinguishing them from SCLC may allow stratified targeted treatment of high-grade neuroendocrine lung tumors.
Molecular characterization studies have provided invaluable insight into the relationship between the major lung tumor subtypes\textsuperscript{1–7}. These studies showed that morphologically defined lung adenocarcinomas, squamous cell carcinomas, and small cell carcinomas have distinct molecular phenotypes based upon their somatically altered genes\textsuperscript{8}. Furthermore, global transcriptional analyses have revealed intra-group consistency, as well as substantial differences in the patterns of expressed genes, which led to the discovery of novel intra-group subtypes\textsuperscript{9–11} and to the elimination of previous lung tumor categories (e.g., large-cell carcinoma)\textsuperscript{7}. Of the remaining lung cancer subtypes, only large-cell neuroendocrine carcinomas (LCNECs) have so far not been characterized in depth using both transcriptomic, as well as genomic approaches.

LCNECs account for 2–3\% of all resected lung cancers and belong to the category of neuroendocrine lung tumors, which also includes pulmonary carcinoids (PCa) and small cell lung cancer (SCLC)\textsuperscript{12,13}. Contrary to pulmonary carcinoids, LCNEC and SCLC are clinically aggressive tumors presenting in elderly heavy-smokers with 5-year survival rates below 15–25\% (LCNEC) and 5\% (SCLC), respectively\textsuperscript{12,13}. While therapy for both typical and atypical carcinoids and SCLC is primarily surgery and chemotherapy (in the case of SCLC), chemotherapy has limited efficacy in LCNEC patients and no standard treatment regimen exists for this tumor type\textsuperscript{14}. Thus, LCNECs share both commonalities (e.g., neuroendocrine differentiation) and discrepancies (e.g., limited response to chemotherapy) with SCLC; however, the underlying molecular basis of these shared and distinct features is only poorly understood. Further complicating the histological classification, LCNECs are sometimes found combined with adenocarcinoma or squamous cell carcinoma and some SCLCs are combined with a component of LCNEC\textsuperscript{12,13}. Thus, defining the molecular patterns of this tumor type presents the opportunity to not only reveal possible novel therapeutic targets, but also help clarifying the ontogeny and relationship of lung tumors in general.

Previous efforts in characterizing LCNECs through targeted sequencing of selected cancer-related genes\textsuperscript{15–17} and through gene expression profiling\textsuperscript{18} provided some first insights; however, global genomic studies combined with transcriptomic analyses have so far been lacking. Furthermore, given the lack of adequate therapeutic strategies in LCNECs, a precise delineation of the molecular boundaries between different neuroendocrine tumors is needed. We therefore aimed to comprehensively dissect both the mutational and the transcriptional patterns of this tumor type.

In this report, we show that LCNECs are composed of two mutually exclusive subgroups, which we categorize as “type I LCNECs” (with STKI\textsubscript{11}/KEAP1 alterations) and “type II LCNECs” (with RB1 alterations). Despite sharing genomic alterations with lung adenocarcinomas and squamous cell carcinomas, type I LCNECs exhibit a neuroendocrine profile with closest similarity to SCLC tumors. While type II LCNECs reveal genetic resemblance to SCLC, these tumors are markedly different from SCLC with reduced levels of neuroendocrine markers and high activity of the NOTCH pathway. Conclusively, LCNECs represent a distinct subgroup within the spectrum of high-grade neuroendocrine tumors of the lung, and our findings emphasize the importance of distinguishing LCNECs from other lung cancers subtypes.

**Results**

**Genomic alterations in LCNECs.** We collected 75 fresh-frozen tumor specimens from patients diagnosed with LCNEC under institutional review board approval (Supplementary Data 1). All tumors were thoroughly analyzed, and the histological features of pulmonary LCNECs were confirmed by expert pathologists (E.B., W.T., R.B.) according to the 2015 WHO classification\textsuperscript{13} (Supplementary Data 2). Most tumors were obtained from current or former heavy smokers, and enriched for stages I and II (68\%). Nineteen of 75 LCNECs included in this study showed additional histological components of lung adenocarcinoma (ADC) (n = 2), squamous cell carcinoma (SqCC) (n = 5) or SCLC (n = 12) (Supplementary Data 1–2). In subsequent analyses nucleic acids were extracted only from pure LCNEC regions (Methods section).

Early genomic profiling studies employing targeted sequencing of selected cancer-related genes aided in the identification of some prevalent mutations in LCNECs\textsuperscript{15–17}. In order to assess globally all genomic alterations in LCNECs and to compare them to those occurring in other lung tumors, we conducted whole-exome sequencing (WES) of 55 LCNEC tumor-normal pairs; we additionally performed whole-genome sequencing (WGS) in those cases where sufficient material was available (n = 11), thus amounting to sequencing data of 60 LCNECs in total (six tumors were both, genome- and exome-sequenced, Supplementary Fig. 1a). We furthermore performed Affymetrix 6.0 SNP array analyses of 60 and transcriptome sequencing of 69 tumors (Supplementary Data 1; Supplementary Fig. 1a). Despite initial review to include cases with a microscopic tumor content of >70\%, sequencing data analysis revealed a median tumor purity of 59.5\% and a median ploidy of 2.8 (Supplementary Data 1, Supplementary Fig. 1b, Methods section). On average, LCNECs exhibited an exonic mutation rate of 8.6 non-synonymous mutations per million base pairs and a C:G > A:T transversion rate of 38.7\% (Fig. 1a, Supplementary Data 1), indicative of tobacco exposure\textsuperscript{1–6}. We analyzed the signatures of mutational processes\textsuperscript{19,20} in LCNECs, which confirmed a prominent smoking-related signature (signature 419\textsuperscript{20}) that accounts for the majority of all somatic mutations, and which is in general comparable to most other lung tumors of heavy smokers (Supplementary Fig. 1c–f, Supplementary Data 3).

Analyses of chromosomal gene copy numbers revealed statistically significant amplifications of 1p34 (containing the MYCL1 gene, 12\%), 8p12 (containing FGFR1, 7\%), 8q24.21 (containing MYC, 5\%), 13q33 (containing IRS2, 3\%), and 14q13 (containing NXX2-1, also known as TTF-1, 10\%) (Q < 0.01, Supplementary Fig. 2a; Supplementary Data 4–5, Methods section). Statistically significant deletions affected CDKN2A (9p21, 8\%) and a putative fragile site at PTPRD (9p24, 7\%)\textsuperscript{21}. While amplifications of NXX2-1 and FGFR1 frequently occur in lung adenocarcinomas\textsuperscript{12,27,21} and squamous cell carcinomas\textsuperscript{27,21,22} respectively, MYCL1 amplifications are commonly found in SCLC\textsuperscript{4–6,23} Thus, LCNECs harbor significant copy-number alterations that occur in different lung cancer subtypes.

We next applied analytical filters to identify mutations with biological relevance in the context of a high-mutation rate and found eight significantly mutated genes (Q < 0.01, Methods section, Fig. 1a, Supplementary Data 6–7). TP53 was the most frequently mutated gene (92\%), followed by inactivating somatic events in RB1 (42\%); bi-allelic alterations in both genes, TP53 and RB1—a hallmark of SCLC\textsuperscript{4–6}—were found in 40\% of the cases (Supplementary Fig. 2b, Supplementary Data 6–9). Notably, LCNECs with admixtures of other histological components mostly had RB1 alterations (Fig. 1a). While genomic alterations in RB1 resulted in loss-of-nuclear Rb1 expression (P < 0.0001, Fisher’s exact test, Supplementary Fig. 3a), immunohistochemistry revealed that absence of Rb1 was not only confined to the LCNEC component, but also evident in the combined other histological subtype (6/7 cases, Supplementary Fig. 3b, Supplementary Data 2). This may implicate shared genetic features between LCNECs and the admixtures of other histological carcinoma types.

We furthermore identified—frequently deleterious—somatic alterations in functionally relevant domains of STKI\textsubscript{11} (30\%) and...
KEAP1 (22%)\(^1\)–\(^3\) (Fig. 1a, Supplementary Fig. 4a, Supplementary Data 6–9). Combined with loss-of-heterozygosity (LOH), bi-allelic alterations of STK11 and KEAP1 were found in 37% of the cases (Supplementary Fig. 2b, Supplementary Data 8). In those cases where WGS was performed, we were able to identify larger genomic rearrangements, which led to the inactivation of RB1, STK11, or KEAP1 (Fig. 1a, Supplementary Fig. 4a, Supplementary Data 9). Altogether, somatic alterations of RB1 and STK11/KEAP1 were detected in 82% of the cases (n = 49) and occurred in a mutually exclusive fashion (P < 0.0001, Fisher’s exact test, Fig. 1a). We furthermore observed a trend toward inferior outcome in patients with RB1-mutated tumors (P = 0.126, log-rank test, Supplementary Fig. 4b). The genomic profiling thus points to two distinct subgroups of LCNECs.

We additionally identified statistically significant mutations in the metalloproteinases ADAMTS2 (15%) and ADAMTS12 (20%),
**Fig. 2** Gene expression studies on lung cancer subtypes. 

(a) A schematic description of the unsupervised consensus clustering approach is provided on the left panel. The clustering results are displayed on the right panel as a heatmap, in which tumor samples are arranged in columns, grouped according to their expression clustering class, annotated for the histological subtype and for the somatic alteration status. Expression values of genes identified by CiaNC (Methods section) are represented as a heatmap; red and blue indicate high and low expression, respectively. Selected candidate genes are shown on the right.

(b) Significant enrichment of differentially expressed genes in signaling pathways is displayed for all clustering classes ($P < 0.0001$, Methods section). Expression values for key neuroendocrine differentiation markers are plotted for each clustering class as box-plots (median and interquartile range, whiskers: min–max values). Dashed black lines indicate the threshold for low expression (Methods section). $Q < 0.05$ (*), significance determined by SAM (Supplementary Dataset 12); $P < 0.001$ (***) Mann-Whitney U-test.

(d) The correlation of each sample to the centroid of its clustering class was calculated and displayed as box-plot (median and interquartile range, whiskers 5–95 percentile).
and in GAS7 (12%) and NTM (10%) (Q < 0.01, Methods section, Fig. 1a, Supplementary Fig. 4c, Supplementary Data 6–7), which so far have not been reported as significantly mutated in any other lung cancer subtype. The mutations affected functionally important protein domains, which may suggest a relevant role in the tumorigenesis of LCNECs (Supplementary Fig. 3c).

We also analyzed LCNECs for alterations in genes of known tumor-specific functions (e.g., CREBBP, EP300, NOTCH1, NOTCH3, NOTCH4, MGMT, BRCA1, BRCA2, RAD51) (Supplementary Fig. 2b, Supplementary Fig. 4d, Supplementary Data 6) and found oncogenic mutations of RAS family genes (KRAS-G12V, -G12C, NRAS-D57E, HRAS-G13R), NFE2L2 (2 cases with G31V and 1 case with E79Q) and BRAF (V600E, and G469V). Combined with focal amplifications, RAS genes were affected in 10% of the tumors (Fig. 1a; Supplementary Data 5–6). We also identified several private in-frame fusion events, e.g., involving the kinases NTRK1 and PTK6, which were, however, not recurrent (Supplementary Fig. 5, Supplementary Data 10). Thus, LCNECs harbor alterations of oncogenes which are commonly found in lung adenocarcinomas, but usually absent in neuroendocrine tumors like SCLC.

The distinct mutational patterns in LCNECs and the presence of other histological components may suggest a high level of intra-tumor heterogeneity. We analyzed the clonal distribution of somatic alterations and determined the cancer cell fraction (CCF) of each somatic mutation call (Methods section). Despite the fact that some LCNECs were found with admixtures of other histological subtypes (Fig. 1a, Supplementary Data 1–2), our studies on the LCNEC component of such composite tumors pointed to a low intra-tumor heterogeneity with a median of 7% sub-clonal mutations per sample (Supplementary Fig. 2b–c, Supplementary Data 1, Methods section). Furthermore, all relevant and significant mutations were found to be clonal within the tumor, thus suggesting these alterations as early events during tumorigenesis (Fig. 1b, Supplementary Data 6).

In summary, genome sequencing revealed distinct genomic profiles in LCNECs. While certain alterations (e.g., RB1, MYCL1) resemble patterns found in SCLC1–3,6,23, others are typical of lung adenocarcinoma or squamous cell carcinomas (e.g., STK11, KEAP1, NKX2-1, RAS, BRAF, and NFE2L2)1–3,7,21. Thus, LCNECs appear to divide into molecularly defined subsets of tumors with genomic similarities to other major lung cancer subtypes.

Transcriptional profiles of LCNECs and other lung cancers. Our sequencing efforts have revealed genomic alterations in LCNECs that were previously known as canonical alterations in either, lung adenocarcinomas, squamous cell carcinomas21, or SCLC1–4. In light of these distinct associations, it remained to be understood if these genomic correlates might reflect a relationship of LCNECs with these lung tumor subtypes on the level of gene expression. We therefore analyzed whether the transcriptional patterns in LCNECs are correlated with the expression profiles of other lung cancers.

We compared the expression data of LCNECs with lung adenocarcinomas23,25–27, squamous cell carcinomas3, SCLC6 and pulmonary carcinoids24 following extensive normalization of the transcriptome sequencing data (Fig. 2a, Methods section, Supplementary Data 11). Unsupervised consensus clustering yielded five consistent expression clusters, which correlated with the histological annotation of the tumors (P < 0.0001, Fig. 2a, Supplementary Fig. 6–7, Supplementary Data 12): pulmonary carcinoids, squamous cell carcinomas and adenocarcinomas formed distinct transcriptional classes (classes A, B, and C, respectively), with few LCNECs falling into these groups. However, the majority of SCLC and LCNECs clustered in two transcriptional subgroups (classes D and E) (Fig. 2a); a phenomenon that had previously been observed in other studies on high-grade neuroendocrine tumors6,18. While the majority of SCLC tumors formed consensus cluster E (75% of all SCLC cases analyzed), a fraction of SCLC tumors shared transcriptional similarities with LCNECs that predominantly formed cluster D. Thus, LCNECs appear to be more closely related to SCLCs than to adenocarcinomas or squamous cell carcinomas.

We next analyzed the transcriptome sequencing data for differentially expressed genes and their enrichment in biological pathways (Methods section). In line with previous observations22,29–31, this analysis showed that both adenocarcinomas and squamous cell carcinomas exhibited upregulation of pathways controlling cell differentiation, adhesion and immune responses, along with higher expression of ERBB2 and TP63 (Fig. 2b, Supplementary Fig. 8a, Supplementary Data 13–14, Q < 0.05, Methods section). Lung neuroendocrine tumors, on the contrary, showed significantly higher expression of neuroendocrine and endocrine markers, Hu antigens (ELAVL3 and ELAVL4) and the lineage transcription factor ASCL1, which is in agreement with previous studies on lung cancer subtypes11–13,18,29 (Q < 0.05, Methods section). Furthermore, particularly high expression of the neuronal and endocrine lineage transcription factors NEUROD1, NEUROD4, and NEUROG3,30,31 was found in SCLC and LCNECs of transcriptional class E (Fig. 2a, c, Supplementary Fig. 8b–e, Supplementary Data 13, Q < 0.05). While recent studies employing SCLC cell lines and mouse models indicated discordant expression patterns for ASCL1 and NEUROD131, our sequencing data of human high-grade neuroendocrine lung tumors revealed expression of both neuroendocrine lineage factors in class E (Supplementary Fig. 8f).

Within the spectrum of neuroendocrine lung tumors, pulmonary carcinoids formed a distinct subgroup with functional enrichment in pathways regulating cellular respiration and metabolism. LCNECs mostly shared similarities with SCLC, revealing upregulation of pathways and genes controlling cell cycle and mitosis (E2F transcription factors and checkpoint kinases), DNA damage response (RAD51, TOP2A, and BRCA1) and centrosomal functions (such as BUB1, PLK1, and ASPM), which, to some extent, were also found in squamous cell carcinomas (Fig. 2b; Supplementary Fig. 8g–i, Supplementary Data 13–14), and which is in agreement with previous studies18. Further supporting a molecular relationship of SCLC and LCNECs in a fraction of the cases, RB1-mutated LCNECs were enriched in classes D and E (P < 0.05, Fisher’s exact test). Although, LCNECs also harbored alterations commonly observed in adenocarcinomas and squamous cell carcinomas, even LCNECs with such alterations in KEAP1 or STK11 were primarily found in transcriptional subclasses shared with SCLC (Fig. 2a, Supplementary Fig. 7c, Supplementary Data 12). Therefore, this observation supports the view that despite the similarity in oncogenic mutations, LCNECs rather constitute their own biological class; and may not be considered as neuroendocrine versions of adenocarcinomas or squamous cell carcinomas.

We also quantified the consistency of the expression profiles for each sample with respect to its clustering group. Again, this analysis revealed a strong correlation for most LCNECs clustering with SCLC tumors (classes D and E); on the other hand, expression profiles of those few LCNEC samples clustering with lung adenocarcinomas, squamous cell carcinomas, and pulmonary carcinoids were less consistent (Fig. 2d). Furthermore, we performed separate transcriptional clustering of LCNECs with adenocarcinomas and squamous cell carcinomas only (excluding SCLC), which did not suggest any unrecognized similarities between these lung cancer subtypes (Supplementary Fig. 9). Thus,
despite sharing somatic alterations with other tumor subtypes, such as adenocarcinomas and squamous cell carcinomas, LCNECs were transcriptionally dissimilar with all non-neuroendocrine lung tumors and showed closest similarities to SCLC.

The transcriptional relationship of LCNEC and SCLC. In the previous section, we sought for a global approach to identify common and distinct transcriptional profiles of LCNECs in relationship with other lung tumors, which showed that LCNEC and SCLC appear to share most transcriptional patterns.
However, strongly divergent tumors (e.g., carcinoids, adenocarcinomas) may drive these clusters and mask important differences between LCNECs and SCLC. We therefore sought to directly compare LCNECs and SCLC on the transcriptional level (Fig. 3a). The resulting unsupervised clustering analysis revealed four consensus clusters of LCNEC and SCLC that we termed classes I–IV in order to distinguish them from the above-mentioned classes A–E (Fig. 3a, Supplementary Fig. 10–11, Supplementary Data 12). Class I exclusively included LCNECs with STK11 or KEAP1 alterations; yet, a few cases with these alterations fell into class II that predominantly consisted of LCNECs with RB1 loss (Fig. 3a). Some LCNECs, including tumors admixed with SCLC (“SCLC combined LCNECs”)—clustered with the majority of SCLC tumors in the classes III and IV; similarly, some SCLC tumors were part of class II that included LCNECs bearing RB1 alterations (Fig. 3a, Supplementary Fig. 11). Even though pathological review had been conducted to distinguish histological subtypes from one another, transcriptional clustering suggested high degrees of similarity for some LCNEC and SCLC cases; these tumors may therefore be considered as “SCLC-like” and “LCNEC-like” (Fig. 3a, Supplementary Fig. 11, Supplementary Data 11). Other major genome alterations (e.g., NXX2-1, MYCL1, RAS genes, NFE2L2, BRAF) did not segregate with the identified transcriptional subgroups (Supplementary Fig. 11). We further analyzed the consistency of the transcriptional subgroups by clustering LCNECs alone, which revealed high concordance with the transcriptional classes identified in Fig. 3a (62/66 cases, 94%, P < 0.001, Fisher’s exact test, Supplementary Fig. 13, Supplementary Data 12). Thus, despite the similarities between LCNECs and SCLCs, subtypes of LCNECs exist with profound differences to SCLC.

The transcriptional clustering heatmap pointed to a strong gene expression pattern shared by all LCNECs bearing STK11/KEAP1 alterations (Fig. 3a, Supplementary Fig. 12a, green box in upper left quadrant). We therefore conducted a supervised analysis of the gene expression data, in which LCNECs with STK11/KEAP1 alterations were compared to tumors bearing RB1 alterations. This analysis indicated specific expression profiles, which were similar to those observed in tumors constituting class I (Fig. 3b, Supplementary Fig. 12, Supplementary Data 13). We therefore assigned this genomic subset of tumors to one group, termed “type I LCNECs”.

Type I LCNECs exhibited high levels of calcitonin A (CALCA), a known marker of pulmonary neuroendocrine cells (Fig. 3a, Supplementary Fig. 12b, Supplementary Data 13). This subgroup furthermore displayed a pronounced upregulation of cellular metabolic pathways, which we also observed in pulmonary carcinoids (Fig. 2b), but which was less prominent in LCNECs and SCLC tumors with RB1 alterations (Fig. 3a, b, Supplementary Data 12–13). Other genes found in type I LCNECs included gastrointestinal transcription factors (e.g., HNF4A, HNF1A, and RFX6), which were previously reported to play a role in de-differentiated lung tumors (35,36) (Fig. 3b, Supplementary Fig. 12c, d, Supplementary Data 13).

The most striking difference was found in the expression levels of neuroendocrine genes: while type I LCNECs and the majority of SCLC tumors (class III + IV) harbored high levels of neuroendocrine genes (CHGA and SYP; Fig. 3c; Supplementary Fig. 12e; Supplementary Data 12), most LCNECs and some SCLC tumors with RB1 alterations in class II exhibited low levels of these genes (Fig. 3c, Supplementary Fig. 12e). By contrast, tumors in class II displayed elevated expression of genes associated with active Notch signaling (e.g., NOTCH1, NOTCH2, and HES1) and immune cell responses (e.g., PDCD1LG2, TLR4, and CTSB) (Fig. 3a, d, Supplementary Fig. 12f, Supplementary Data 12–13). Given the strong enrichment of LCNECs with STK11 or KEAP1 alterations in cluster I, and the prominent lack of expression of key neuroendocrine genes in most tumors of class II, we termed LCNECs within this transcriptional class as “type II LCNECs”.

We have recently demonstrated that SCLC tumors usually harbor inactive Notch signaling and that activation of Notch reduces expression of neuroendocrine genes (e.g., CHGA, SYP, and NCAM1) and Ascl1 (31). Consistent with this notion, we found that type II LCNECs and some SCLC within this transcriptional class exhibited signs of NOTCH upregulation and low expression of neuroendocrine markers, ASCL1 and DLL3, an inhibitor of the Notch signaling pathway (37) (Fig. 3d, and Supplementary Fig. 12f). Conversely, type I LCNECs and the majority of the SCLC samples (class III and IV) showed higher levels of neuroendocrine genes, as well as of ASCL1 and DLL3, and downregulation of NOTCH pathway genes (Fig. 3d, Supplementary Fig. 12f). Thus, despite the fact that type II LCNECs and some SCLC harbor bi-allelic loss of TP53 and RB1, their transcriptional signatures include low levels of neuroendocrine genes and a distinct profile of NOTCHhigh and ASCL1low/DLL3low, which differentiates these tumors from type I LCNECs and from the majority of SCLC cases. We did not identify any significant enrichment of somatic alterations in NOTCH pathway genes, which may explain these transcriptional differences (Supplementary Fig. 11). However, a recent study in a pre-clinical mouse model has established a central role of REST as a repressor of neuroendocrine markers in SCLC (38). Compatible with these findings, type II LCNECs displayed significantly higher levels of REST (clustering class II, Supplementary Data 12, Q < 0.05), which may explain the low neuroendocrine phenotype in type II LCNECs marked by ASCL1low/DLL3low/NOTCH3high. Given the important role of NOTCH signaling and ASCL1 in the decision of neuroendocrine fate and the development of neuroendocrine lung tumors (29,31,38), these findings provide further support for our distinction of type I and II LCNECs.

We next analyzed the relationship of the expression classes I–IV using hierarchical clustering, which revealed two major subgroups (Supplementary Fig. 11): one subgroup mainly consisting of LCNECs (type I and II LCNECs), and the other subgroup mainly containing SCLC tumors (classes III and IV). Thus, despite harboring distinct transcriptional subcategories, LCNEC and SCLC tumors largely followed their histological annotation and formed separate transcriptional subgroups. Differentially expressed genes included SOX1 and the neuroendocrine Hu genes (ELAVL3,
**Discussion**

Here we provide the first comprehensive molecular analysis of LCNECs, which allowed distinguishing between two genomic subgroups with specific transcriptional patterns, defined as “Type I LCNECs” and “Type II LCNECs” (Fig. 4).

Type I and II LCNECs harbor key genomic alterations and oncogenic mutations, which are commonly found in SCLC, lung adenocarcinoma, or squamous cell carcinoma (e.g., in RAS genes, BRAF, NFE2L2, as well as in STK11 and KEAP1 in the case of Type I LCNECs, and RB1 losses in the case of Type II LCNECs). One possible explanation for this observation might be a high level of intra-tumor heterogeneity, combined with occurrence of two tumor types in a single tumor. However, the key alterations that we found in LCNECs were mostly clonal, with limited genomic intra-tumor heterogeneity. Furthermore, thorough comparisons of gene expression profiles did not suggest similarities between LCNECs and lung adenocarcinomas or squamous cell carcinomas. Thus, the combinations of distinct sets of mutations with specific patterns of gene expression support the view that LCNECs are not a variant of the other types of lung cancer, but represent a distinct subgroup within the spectrum of neuroendocrine lung tumors.

In a more focused comparison with the most frequent neuroendocrine type of lung cancer, SCLC, Type I LCNECs with STK11 and KEAP1 alterations exhibited a high degree of similarity with these carcinomas, as well as high expression of neuroendocrine genes and a profile of ASCL1high/DLL3high/NOTCHlow. By contrast, Type II LCNECs with RB1 alterations revealed reduced expression of neuroendocrine genes and a pattern of ASCL1low/DLL3low/NOTCHhigh. Notch family members play a multifaceted role in the development of neuroendocrine tumors with cell-type specific tumor suppressor and oncogenic functions. We have shown in earlier studies that NOTCH serves as a tumor suppressor in SCLC, which mostly harbor high-level expression of the negative regulator of Notch, Dll3, has shown early signs of clinical activity in SCLC. A recent clinical trial with an antibody-drug conjugate targeting the non-canonical inhibitory NOTCH ligand, Dll3, has shown shared neuroendocrine pathways between SCLC and type I LCNECs, which may be similarly susceptible to this agent. On the other hand, Type II LCNECs with alterations in RB1 exhibited active Notch signaling. Clinical trials have assessed the efficacy of an antibody targeting Notch 2 and 3 in SCLC, but recently failed in demonstrating a clinical benefit. Therefore, future clinical trials involving therapeutics, targeting activating or inhibitory members of the Notch pathway will—in our view—require clear assignment of the respective molecular subtype.

Perhaps another noteworthy finding, Type II LCNECs exhibited a pattern of gene expression with upregulation of immune related pathways (Fig. 3b, Fig. 4), which has similarly been observed in various other tumor types and which may impact the response of patients to immunotherapy. Taken together, the precise distinction of high-grade neuroendocrine tumors representing as Type I LCNECs and as RB1-mutated SCLC or Type II LCNECs, may be pivotal to assess the efficacy of targeted therapeutics, including Notch pathway and immune checkpoint inhibitors. Our sequencing studies did not reveal any somatic events that may cause the transcriptional discrepancy observed in LCNEC and SCLC tumors with TP53 and RB1 alteration, which raises the
question if all neuroendocrine tumors share the same cell of origin. It remains to be understood whether distinct tumor-specific cell of origins or cellular processes allow for plasticity and trans-differentiation that consequently lead to distinct molecular phenotypes. Importantly, histologically trans-differentiation from lung adenocarcinoma to SCLC has been observed, both spontaneously or as resistance mechanisms to kinase inhibitors44,45; in some cases these were linked with a loss of RB146. Previous studies involving genetically engineered mouse models and human cell lines have emphasized the phenomenon of transcriptional heterogeneity in SCLC and pointed to discordant expression of key lineage factors (e.g. ASCL1, NEUROD1, REST)13,18. By contrast, human primary tumors revealed a more complex expression pattern with co-expression of these transcriptional regulators. As a limitation of bulk tumor sequencing, advances in single cell sequencing may further aid to resolve and study the level of transcriptional intra-tumor heterogeneity in high-grade neuroendocrine tumors. While our studies pointed to transcriptional correlates of genomically defined subsets in LCNECs (type I and type II LCNECs), additional analyses on a larger dataset are warranted to further interrogate subclonal or high-grade neuroendocrine tumor heterogeneity.

In summary, we provide the first comprehensive characterization of neuroendocrine lung tumors, which integrates the molecular phenotypes of less frequent lung tumor subtypes. Despite the fact that LCNEC and SCLC tumors share some common clinical and histological characteristics, our study emphasizes pronounced differences in the pattern of genomic alterations and in their transcriptome profiles. The precise distinction of type I and type II LCNECs from SCLC is consequently pivotal to evaluate the response of patients to treatment options and to further understand morphological trans-differentiation processes in lung cancer patients.

Methods

Human specimens. The institutional review board (IRB) of the University of Cologne approved this study. Patient samples were obtained under IRB-approved protocols following written informed consent from all human participants. We collected and analyzed fresh-frozen samples of 75 LCNEC patients, which were provided by multiple collaborating institutions; 42 tumors were previously subject of other studies conducted by Rousséau et al.47 (n = 25) and Seidel et al.7 (n = 37) (Supplementary Data 1). Clinical data were available for most patients, who were predominantly male (approximate ratio of 4:1) and current or former heavy smokers. All tumors were re-evaluated by independent expert pathologists (E.B., W.T., and R.B.), and the diagnosis of LCNEC and the assessment of combined histological components were confirmed by H&E staining and immunohistochemistry, including markers for chromogranin A, synaptophysin, CD56 and Ki67. All tumors were positive for at least one neuroendocrine differentiation marker (Supplementary Data 1–2). Specimens containing >70% of tumor cells were processed for DNA and RNA extractions. DNA was extracted from matching normal material that was provided in the form of blood or adjacent non-tumorigenic lung tissue, which through pathological evaluation was confirmed to be free of tumor contaminants.

Nucleic acid extraction. Total DNA and RNA were obtained from fresh-frozen tumor tissue and matched fresh-frozen normal tissue or blood. Depending on the size of the tissue, 15–30 sections, each 20 μm thick, were cut using a cryostat (Leica) at −20 °C. The matched normal sample obtained from frozen tissue was processed the same way. Nineteen LCNEC cases were identified with mixed histological components of SCLC, lung adenocarcinomas and squamous carcinomas (Supplementary Data 1); in these cases nucleic acids were extracted from pure LCNEC regions by only dissecting the LCNEC component. DNA was extracted with the Gentra Puregene DNA extraction kit (Qiagen) and diluted to a working concentration of 100 ng/μl. The DNA was analyzed by agarose gel electrophoresis and confirmed to be of high-molecular weight (>10kb). The RNA of tumor and normal material was confirmed to originate from the same patient by short tandem repeat (STR) analysis which was conducted at the Institute for Legal Medicine at the University of Cologne (Cologne, Germany), or by subsequent Affymetrix 6.0 SNP array and sequencing analyses.

DNA was isolated from fresh-frozen tissue by first freezing and homogenizing tissue sections with the Tissue Lyzer (Qiagen). The RNA was then extracted with the Qiagen RNAeasy Mini Kit. The RNA quality was analyzed at the Bioanalyon 2100 DNA Chip 7500 (Agilent Technologies) and cases with a RNA integrity number (RIN) of over seven were considered for RNA-seq experiments.

Next-generation sequencing (NGS). WES was performed by first fragmenting 1 μg of DNA (Bio rotor, digenade, Liège, Belgium). The DNA fragments were then end-repaired and adaptor-ligated with sample index barcodes. Following size selection, the SeqCap EZ Human Exome Library version 2.0 kit (Roche NimbleGen, Madison, WI, USA) was used to enrich for the whole exome. The DNA libraries were then sequenced in a paired-end 2 x 100 bp protocol aiming for an average coverage of 90% and 120% for the normal and tumor DNA, respectively. The primary data were filtered for signal purity with the Illumina Realtime Analysis software.

WGS was performed with a read length of 2 x 100 bp. The samples were processed to provide 110 Gb of sequence, thus amounting to a mean coverage of 30x–99x for both tumor and matched normal.

For RNA-seq, cDNA libraries were prepared from PolyA + RNA following the Illumina TruSeq protocol for mRNA (Illumina, San Diego, CA, USA). The libraries were sequenced with a paired-end 2 x 100 bp protocol resulting in 8.5 Gb per sample, and thus in a 30x mean coverage of the annotated transcriptome.

Whole genome, whole exome and transcriptome sequencing reactions were performed on an Illumina HiSeq 2000 sequencing instrument (Illumina, San Diego, CA, USA).

Copy-number analysis by Affymetrix SNP 6.0 arrays. Human DNA from fresh-frozen tumors was analyzed with Affymetrix Genome-Wide Human SNP 6.0 arrays to determine copy-number alterations. Raw copy number data were compared to dividing tumor-derived signals by the means obtained from a subset of normal samples which were hybridized to the array in the same batch. Circular binary segmentation was applied to obtain segmented raw copy numbers46. Significant copy-number alterations were assessed with CGARS47 at a threshold of Q < 0.01 (Supplementary Data 4).

Data processing and analyses of DNA sequencing data. The sequencing reads were aligned to the human reference genome NCBI build 37 (NCBI37/hg19) with BWA (version 0.6.1-r104)48. Possible PCR-duplicates were masked and not included for subsequent studies. We applied our in-house analysis pipeline49,50 to analyze the data for somatic mutations, copy number alterations and genomic rearrangements. In brief, the mutation calling algorithm considers local sequencing depth, forward-reverse bias, and global sequencing error, to thus determine the presence of a mutated allele. We determined the somatic status of these mutations by assessing the absence of these variants in the sequencing data of the matched normal.

We determined genomic rearrangements from WGS data of 11 human LCNECs following the procedure as previously described45. In brief, the sequencing data were analyzed for discordant read-pairs, which were not located within the expected mapping distance (>600 base pairs) or which revealed an incorrect orientation. Discordant read-pairs were analyzed for breakpoint-spanning reads, in which one read-pair shows partial alignments to two distinct genomic loci. Rearranged genomic loci were then reported at instances where at least one breakpoint-spanning read was identified. The genomic loci that were rearranged from each tumor sample were further filtered against the sequencing data of a matched normal and additionally against a library of normal genomes to thus minimize the detection of false-positive rearrangements.

Significantly mutated genes were analyzed as previously described46. In brief, we first determined the overall mutation rate of each gene by computing its expected number of mutations assuming that all mutations are uniformly distributed across the genome. We also considered the ratio of synonymous to non-synonymous mutations into a combined statistical model to determine significantly mutated genes. Since mutation rates in non-expressed genes are often higher than the genome-wide background rate, we furthermore filtered for the expression of genes by referring to the transcriptome sequencing data of LCNECs. Only genes with a median FPKM (Fragments Per Kilobase Million) value of >1 in at least 35 out of 60 samples were considered (Methods section: RNA sequencing data processing and analyses). The significance of recurrently mutated genes was determined at a Q-value of 0.01 (Supplementary Data 7). Following previously described methods, we furthermore analyzed the data for significant enrichment of damaging mutations (including splice site, non-sense, and frameshift mutations)50 and for significant clustering of mutations in genomic hotspots following a re-sampling based approach51. Significance was determined at a Q-value of 0.01, if the gene was affected in >10% of the samples (Supplementary Data 7). The damaging impact of mutations was further assessed by Polyphen2.

The clonal status of mutations was assessed by computing for every mutation the "cancer cell fraction" (CCF), which defines within a tumor the fraction of cancer cells harboring that particular mutation45. The CCF was computed following our approach. In brief, the CCF estimates tumor purity, ploidy, and absolute copy numbers, and computes for each mutation in a given sample the expected allele frequency under the assumption of clonality. The CCF is the quotient of the observed allelic fraction and the expected allelic fraction of a mutation. The distribution of CCFs for every mutation in a sample allows to further identify distinct clonal and subclonal populations. The analysis described in Supplementary Fig. 2e considers mutations, which were assigned to clonal and subclonal fractions with a
probability >90%. In consideration of the sequencing coverage and the overall distribution of CCFs of every mutation in a sample, we furthermore determined the significant enrichment of mutations in a subclone at a P-value of 0.01 (Fig. 1b).

Mutational signatures analyses. Mutational signatures were analyzed in lung cancer subtypes applying previously described methods and referring to the reads were further normalized to the overall upper quantile (Fragments Per Kilobase Million) and the expression values served as a estimation method. ClaNC was used to identify genes whose expression patterns were all similarly processed to analyze for gene expression patterns.

Differential expression analysis. The expression data on lung cancer subtypes (n = 341) was analyzed and the 0.75 quantile of all log(mean(RSEM)) values was used to identify highly expressed genes, while the 0.9 quantile of log(variance(RSEM)) was used as a threshold to identify clustering gene sets that have highly variable expression patterns, which yielded a set of 1854 genes (Supplementary Fig. 6a). The samples were clustered with ConsensusClusterPlus following partition around medoids (PAM), and the ConsensusClusterPlus output along with gene expression heatmaps, principal components analysis, and silhouette plots was analyzed. Manual review of ConsensusClusterPlus output suggested a possible clustering solution based on the number of clusters (k = 6). However, two of the six groups included mainly lung adenocarcinoma samples and the gene expression heatmaps and PCA plots showed that these groups were quite similar. Thus, we chose to collapse these groups, thereby producing a five-class solution. The consensus clusters highly correlated with the histological subtypes as determined by Fisher’s exact test (P < 0.001, 10,000 permutations): class A (n = 66; enriched for pulmonary carcinoids), class B (n = 65, enriched for lung squamous cell carcinomas), class C (n = 108, enriched for lung adenocarcinomas; data generated by different institutes), class D (n = 38, enriched for LCNEC and SCLC cases), and class E (n = 64, enriched for SCLC and LCNEC cases) (Supplementary Fig. 6b, Supplementary Data 12). ClANC led to the identification of 875 classifier genes, which are displayed in the expression heatmaps (Fig. 2, Supplementary Fig. 6–7, Supplementary Data 13).

We then conducted consensus clustering of LCNECs, SCLC, lung adenocarcinomas, and squamous cell carcinomas. The unsupervised clustering approach was repeated for a subset of lung subtypes using the respective expression heatmaps and the respective electro- pherograms were analyzed by visual inspection using 4 Peaks or Genioese.

Analysis of RNA sequencing data. In order to detect chimeric transcripts, RNA-seq data were processed using TRUP v.4–5. In brief, paired-end RNA-seq reads were aligned to the human reference genome (GRCh38/hg19). We used TRUP to identify potential chimeric transcripts. Gene expression levels were determined with Cufflinks v2.2.2 referring only to paired-end reads that uniquely mapped within the expected mapping distance. The expression was quantified as FPKM (Fragments Per Kilobase Million) and the expression values served as a filter for identifying significantly mutated genes (Methods section: Data processing and analyses of DNA sequencing data).

Gene expression profiling and clustering studies. We analyzed transcriptomic sequencing data from a total of n = 341 lung cancer samples. N = 221 samples referred to the data generated at the University of Cologne, Department of Translational Genomics, which included 41 lung adenocarcinomas, 61 pulmonary carcinoids, 139 LCNECs, and 66 LUSCs. The remaining 120 samples were randomly selected from both the TCGA lung squamous cell carcinoma (n = 60) and TCGA lung adenocarcinoma (n = 60) cohorts v.4–5. We refer to the Tumor Data of the Genomics Commons Legacy Archive. Sequencing data of lung adenocarcinomas from two different platforms aided in controlling for potential batch effects in subsequent studies.

We performed consensus clustering of LCNECs and SCLC through unsupervised clustering of the expression data of LCNEC and SCLCs alone (n = 119). Exploratory analyses of the gene expression data suggested the use of the 0.9 quantile of both the log(mean(RSEM)) and log(variance(RSEM)) values as thresholds for highly expressed and highly variably expressed genes. This produced a set of 1416 clustering genes. The Consensus Clustering approach included partitioning around medoids and yielded four gene expression subtypes: class I (n = 19, only LCNECs), class II (n = 49, LCNEC and some SCLC tumors), class III (n = 10, LCNECs and some SCLCs), and class IV (n = 41, mainly SCLC and some LCNECs) (Fig. 3, Supplementary Fig. 10–11, Supplementary Data 12). Hierarchical clustering of these cases revealed two main subgroups: one mainly formed by class I and II (enriched for LCNECs) and one mainly formed by class III and IV (enriched for SCLC) (Supplementary Fig. 11). 300 classifier genes were identified by ClANC and are displayed in the expression heatmaps (Fig. 3, Supplementary Fig. 11, Supplementary Data 13).

We also performed consensus clustering of LCNECs with lung adenocarcinomas or lung squamous cell carcinomas. A gene set of (a) 1335 and (b) 1388 highly variable (0.85 quantile) and expressed genes (0.925 quantile) was identified in subsets of lung cancer tumors, including (a) LCNECs and lung adenocarcinomas (n = 167) and (b) LCNECs and lung squamous cell carcinomas (n = 126). The consensus clustering approach through PAM (partitioning around medoids) suggested in both cases two transcriptional subclasses: for approach (a) class A (n = 70, mainly LCNECs) and class B (n = 97, mainly lung adenocarcinomas); and for approach (b) class A (n = 58, mainly LCNECs) and class B (n = 68, mainly lung squamous cell carcinomas). ClANC identified 100 classifier genes in each approach, which were used for the expression heatmaps (Supplementary Fig. 9).

We furthermore performed consensus clustering of LCNECs alone. The transcriptional data on LCNECs was analyzed and hierarchical clustering referred to 475 very highly expressed (0.875 quantile) and very highly variable (0.975 quantile) genes. The consensus clustering approach yielded a k = 4 clustering solution: class 1 (n = 11), class 2 (n = 21), class 3 (n = 24), and class 4 (n = 10). ClANC was then applied to the clustering solution, which further identified 540 classifier genes (Supplementary Fig. 13, Supplementary Data 13).

Differential expression analysis. The SAMR R package was used to identify genes that were differentially expressed in the expression subtypes using 1000 permutations and a Q-value threshold of 0.05 (Supplementary Data 13). We then used the DAVID annotation database to identify pathways that were enriched for differentially expressed genes at P < 0.0001 (Supplementary Data 14).

Immunohistochemistry. FFPE tissue sections of 3-μm thickness were stained for hematoxylin and eosin (H&E) and immunohistochemistry (IHC) was conducted for CD56 (NCAM1), Synaptophysin (SYP), Chromogranin A (CHGA, clone DAK- A3), TTF-1 (NKX2-1, clone 8G7G3/1), and Rb1 (Rb1, clone 1F8 (ab81701; Abcam, LCNEC and some LCNECs tumors), class III (n = 10, LCNECs and some SCLCs), and class IV (n = 41, mainly SCLC and some LCNECs) (Supplementary Fig. 11). 300 classifier genes were identified by ClANC and are displayed in the expression heatmaps (Fig. 3, Supplementary Fig. 11, Supplementary Data 13).

We also performed consensus clustering of LCNECs with lung adenocarcinomas or lung squamous cell carcinomas. A gene set of (a) 1335 and (b) 1388 highly variable (0.85 quantile) and expressed genes (0.925 quantile) was identified in subsets of lung cancer tumors, including (a) LCNECs and lung adenocarcinomas (n = 167) and (b) LCNECs and lung squamous cell carcinomas (n = 126). The consensus clustering approach through PAM (partitioning around medoids) suggested in both cases two transcriptional subclasses: for approach (a) class A (n = 70, mainly LCNECs) and class B (n = 97, mainly lung adenocarcinomas); and for approach (b) class A (n = 58, mainly LCNECs) and class B (n = 68, mainly lung squamous cell carcinomas). ClANC identified 100 classifier genes in each approach, which were used for the expression heatmaps (Supplementary Fig. 9).

We furthermore performed consensus clustering of LCNECs alone. The transcriptional data on LCNECs was analyzed and hierarchical clustering referred to 475 very highly expressed (0.875 quantile) and very highly variable (0.975 quantile) genes. The consensus clustering approach yielded a k = 4 clustering solution: class 1 (n = 11), class 2 (n = 21), class 3 (n = 24), and class 4 (n = 10). ClANC was then applied to the clustering solution, which further identified 540 classifier genes (Supplementary Fig. 13, Supplementary Data 13).

Differential expression analysis. The SAMR R package was used to identify genes that were differentially expressed in the expression subtypes using 1000 permutations and a Q-value threshold of 0.05 (Supplementary Data 13). We then used the DAVID annotation database to identify pathways that were enriched for differentially expressed genes at P < 0.0001 (Supplementary Data 14).

Immunohistochemistry. FFPE tissue sections of 3-μm thickness were stained for hematoxylin and eosin (H&E) and immunohistochemistry (IHC) was conducted for CD56 (NCAM1), Synaptophysin (SYP), Chromogranin A (CHGA, clone DAK-A3), TTF-1 (NKX2-1, clone 8G7G3/1), and Rb1 (Rb1, clone 1F8 (ab81701; Abcam, NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-03099-x | www.nature.com/naturecommunications

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
4. Peifer, M. et al. Integrative genome analyses identify key somatic driver mutations in lung adenocarcinoma. *Oncogene* **36**, 1384–1393 (2017).

5. Rudin, C. M. et al. Comprehensive genomic analysis identifies a mutational landscape in lung adenocarcinoma. *Cell* **155**, 2272–2283 (2013).

6. George, J. et al. Comprehensive genomic profiling of lung adenocarcinoma identifies novel actionable targets. *Clin. Cancer Res.* **20**, 1371–1380 (2014).

7. Augustyn, A. et al. ASCL1 is a lineage oncogene providing therapeutic targets for high-grade pulmonary neuroendocrine tumor-initiating cells in vivo. *Clin. Cancer Res.* **23**, 5250–5260 (2017).

8. Westerman, B. A. et al. Basic helix-loop-helix transcription factor profiling identifies novel therapeutic targets in human lung adenocarcinoma. *Semin. Oncol*. **40**, 2208–2217 (2013).

9. Hong, W. C. et al. Comprehensive genomic profiling of lung adenocarcinoma identifies novel therapeutic targets in high-grade pulmonary neuroendocrine tumors. *Clin. Cancer Res.* **22**, 3681–3692 (2016).

10. Wilkerson, M. D. et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin. Cancer Res.* **16**, 4864–4875 (2010).

11. Chen, F. et al. Multiplatform-based molecular subtypes of non-small-cell lung cancer. *Oncogene* **36**, 1384–1393 (2017).

12. Travis, W. D. Advances in neuroendocrine lung tumors. *Ann. Oncol.* **21**, vi65–71 (2010).

13. Travis, W. D. et al. The 2015 World Health Organization Classification of lung tumors. *J. Thorac. Oncol.* **10**, 1243–1260 (2015).

14. Pasaro, M. et al. Pulmonary large-cell neuroendocrine carcinoma: from epidemiology to therapy. *J. Thorac. Oncol.* **10**, 1133–1141 (2015).

15. Karlsson, A., Brunnhölz, M., Lindquist, K. E. & Jirström, K. Mutational and gene fusion analyses of primary large cell and large cell neuroendocrine lung cancer Patient material. *Oncotarget* **6**, 22028–22037 (2015).

16. Kekhtman, N. 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.* **22**, 3681–3692 (2016).

17. Miyoshi, T. et al. Genomic profiling of large-cell neuroendocrine carcinoma of the lung. *Clin. Cancer Res.* **23**, 757–765 (2017).

18. Jones, M. E. 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 Oncol.* **16**, 757–761 (2015).

19. Weiss, J. et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci. Transl. Med.* **2**, 62ra93 (2010).

20. Wustuba, I. L., Gazdar, A. F. & Minna, J. D. Molecular genetics of small cell lung carcinoma. *Semin. Oncol.* **28**, 3–13 (2001).

21. Fernandez-Cuesta, L. et al. Frequent mutations in chromatin-remodeling genes in pulmonary carcinoids. *Nat. Commun.* **5**, 3518 (2014).

22. Imielinski, M. et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).

23. Sutherland, K. D. et al. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cell Rep.* **1**, 1272–1282 (2016).

24. Park, K. et al. Characterization of the cell of origin for small cell lung cancer. *Proc. Natl Acad. Sci. USA* **111**, 14788–14793 (2014).

25. Western, B. A. et al. Basic helix-loop-helix transcription factor profiling of lung tumors shows aberrant expression of the proneural gene atonal homolog 1 (ATOH1, HATH1, MATH1) in neuroendocrine tumors. *Int. J. Biol. Markers* **22**, 218–222 (2007).

26. Borromeo, M. D. et al. ASCL1 and NEUROD1 reveal heterogeneity in pulmonary neuroendocrine tumors and regulate distinct genetic programs. *Cell Rep.* **16**, 1259–1272 (2016).

27. Sutherland, K. D. et al. Cell of origin of small cell lung cancer: inactivation of Trp53 and Rb1 in distinct cell types of adult mouse lung. *Cancer Cell* **19**, 754–764 (2011).

28. Park, K. et al. Characterization of the cell of origin for small cell lung cancer. *Cell Cycle* **10**, 2806–2815 (2011).

29. Song, H. et al. Functional characterization of pulmonary neuroendocrine cells in lung development, injury, and tumorigenesis. *Proc. Natl Acad. Sci. USA* **109**, 17531–17536 (2012).

30. Sugano, M., Nagasaka, T. & Sasaki, E. HNF4a as a marker for invasive pulmonary neuroendocrine tumor. *J. Clin. Oncol.* **29**, 17536 (2011).

31. Sasaki, E. H. & Sugano, M. HNF4a expression in pulmonary carcinoids. *Oncogene* **22**, 218–222 (2003).

32. Snyder, E. L. et al. Article Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol. Cell* **50**, 185–199 (2013).

33. Saunders, L. R. et al. A DLL3-targeted antibody-drug conjugate eradicates high-grade pulmonary neuroendocrine tumor-initiating cells in vivo. *Sci. Transl. Med.* **7**, 302ra136 (2015).

34. Lim, J. S. et al. Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. *Nature* **545**, 360–364 (2017).
This work was supported by the German Cancer Aid (Deutsche Krebshilfe) as part of the small cell lung cancer genome sequencing consortium (grant ID: 109679 to R.K.T., M.P., R.B., P.N., M.V., and S.A.H.), by the German Ministry of Science and Education (BMBF) as part of the eMed program (grant no. 01ZX130A to R.K.T., R.B., U.L., M.P., and J. W., and grant no. 01ZX1406 to M.P.), by the EU-Framework program CURELUNG (HEALTH-F2-2010-258677 to R.K.T., J.W., E.B., and L.R.), by the Deutsche Krebshilfe as part of the Oncology Centers of Excellence funding program (R.K.T.), by the National Institute of Health (NIH U01CA181009 to D.N.H.), by the German Cancer Consortium (DKTK) Joint Funding program, by Associazione Italiana per la Ricerca sul Cancro (AIRC, IG 16847) to L.R., and by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy National Nuclear Security Administration under Contract No. DE-AC52-06NA25396 (L.B.A.). J. G. received funding as part of the LASLC Young Investigator award. L.B.A. is supported through a J. Robert Oppenheimer Fellowship at Los Alamos National Laboratory. We are indebted to the patients donating their tumor specimens as part of the Clinical Lung Cancer Genome Project initiative. We thank the regional computing center of the University of Cologne (ZRKZ) for providing the CPU time on the DFG-funded supercomputer, "CHEOPS", as well as the support. We would like to acknowledge that Australian specimens were provided with assistance of the Victorian Cancer Biobank. We furthermore thank Johannes Berg, Chau Nguyen, Philipp Lorimier, Elisabeth Kirtz, and César Tejerina Álvarez for their technical assistance.

**Author contributions**

R.K.T., L.F.-C., J.G., and E.B. conceived and designed the experiment. J.G., V.W., S.P., S.A.H., M.F., D.N.H., W.D.T., L.F.C., E.B., and R.K.T. supervised the work and gave scientific input. J.W., M.O., Y.C., and L.F.-C. performed computational and statistical analyses. R.B., W.D.T., and E.B. performed pathological review. D.M.S., C.G.B., S.L., A.S., W.V., M.C., O.T.B., M.L.I., A.H., S.S., A.G.W., B.S., L.R., U.P., J.H.C., J.S., R.B., N.L., and E.B. contributed with samples. D.S., G.B., F.L., M.L., N.L., V.A., U.L., P.N., P.M., S.D.M.W., J.W., M.V., and T.Z. helped with logistics. J.G., R.K.T., and L.F.-C. wrote the manuscript, which was reviewed by all the co-authors.

**Additional information**

**Supplementary Information** accompanies this paper at https://doi.org/10.1038/s41467-018-03099-x.

**Competing interests:** L.F.-C. and R.K.T. are inventors on a patent application related to findings described in this manuscript. R.K.T. is a founder of NEO New Oncology GmbH, now part of Siemens Healthcare. R.K.T. received consulting and lecture fees (Merck, Roche, Lilly, Boehringer Ingelheim, AstraZeneca, Daiichi-Sankyo, MSD, NEO New Oncology, Puma, Clovis). R.B. is a co-founder and owner of Targos Molecular Diagnostics and received honoraria for consulting and lecturing from AstraZeneca, Boehringer Ingelheim, Merck, Roche, Novartis, Lilly, and Pfizer. J.W. received consulting and lecture fees from Roche, Novartis, Boehringer Ingelheim, AstraZeneca, Bayer, Lilly, Merck, Amgen and research support from Roche, Bayer, Novartis, Boehringer Ingelheim. T.Z. received honoraria from Roche, Novartis, Boehringer Ingelheim, Lilly, Merck, Amgen and research support from Novartis. B.S. received consulting fees from AstraZeneca, Roche-Genentech, Pfizer, Novartis, Merck, and Bristol Myers Squibb. The remaining authors declare no competing financial interest.

**Reprints and permission information** is available online at http://npg.nature.com/reprintsandpermissions/

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

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

© The Author(s) 2018

**NATURE COMMUNICATIONS** | DOI: 10.1038/s41467-018-03099-x

**ARTICLE**
Author/s:
George, J; Walter, V; Peifer, M; Alexandrov, LB; Seidel, D; Leenders, F; Maas, L; Mueller, C; Dahmen, I; Delhomme, TM; Ardin, M; Leblay, N; Byrnes, G; Sun, R; De Reynies, A; McLeer-Florin, A; Bosco, G; Malchers, F; Menon, R; Altmuller, J; Becker, C; Nurnberg, P; Achter, V; Lang, U; Schneider, PM; Bogus, M; Soloway, MG; Wilkerson, MD; Cun, Y; McKay, JD; Moro-Sibilot, D; Brambilla, CG; Lantuejoul, S; Lemaitre, N; Soltermann, A; Weder, W; Tischler, V; Brustugun, OT; Lund-Iversen, M; Helland, A; Solberg, S; Ansen, S; Wright, G; Solomon, B; Roz, L; Pastorino, U; Petersen, I; Clement, JH; Saenger, J; Wolf, J; Vingron, M; Zander, T; Perner, S; Travis, WD; Haas, SA; Olivier, M; Foll, M; Buettner, R; Hayes, DN; Brambilla, E; Fernandez-Cuesta, L; Thomas, RK

Title:
Integrative genomic profiling of large-cell neuroendocrine carcinomas reveals distinct subtypes of high-grade neuroendocrine lung tumors

Date:
2018-03-13

Citation:
George, J; Walter, V; Peifer, M; Alexandrov, LB; Seidel, D; Leenders, F; Maas, L; Mueller, C; Dahmen, I; Delhomme, TM; Ardin, M; Leblay, N; Byrnes, G; Sun, R; De Reynies, A; McLeer-Florin, A; Bosco, G; Malchers, F; Menon, R; Altmuller, J; Becker, C; Nurnberg, P; Achter, V; Lang, U; Schneider, PM; Bogus, M; Soloway, MG; Wilkerson, MD; Cun, Y; McKay, JD; Moro-Sibilot, D; Brambilla, CG; Lantuejoul, S; Lemaitre, N; Soltermann, A; Weder, W; Tischler, V; Brustugun, OT; Lund-Iversen, M; Helland, A; Solberg, S; Ansen, S; Wright, G; Solomon, B; Roz, L; Pastorino, U; Petersen, I; Clement, JH; Saenger, J; Wolf, J; Vingron, M; Zander, T; Perner, S; Travis, WD; Haas, SA; Olivier, M; Foll, M; Buettner, R; Hayes, DN; Brambilla, E; Fernandez-Cuesta, L; Thomas, RK, Integrative genomic profiling of large-cell neuroendocrine carcinomas reveals distinct subtypes of high-grade neuroendocrine lung tumors, NATURE COMMUNICATIONS, 2018, 9