Mutational analysis of pulmonary tumours with neuroendocrine features using targeted massive parallel sequencing: a comparison of a neglected tumour group

Claudia Vollbrecht1,8, Robert Werner2, Robert Fred Henry Walter2,3, Daniel Christian Christoph4, Lukas Carl Heukamp1,9, Martin Peifer5, Burkhard Hirsch6, Lina Burbat6, Thomas Mairinger7, Kurt Werner Schmid2, Jeremias Wohlschlaeger2 and Fabian Dominik Mairinger *,2,8

1Institute of Pathology, University Hospital Cologne, Cologne, Germany; 2Institute of Pathology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; 3Ruhrlandklinik, West German Lung Centre, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; 4Department of Oncology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany; 5Department of Translational Genomics, Cologne Center of Genomics, University of Cologne, Cologne, Germany; 6Institute of Pathology, Molecular Diagnostics, Charité Universitätsmedizin Berlin, Berlin, Germany and 7Institute of Pathology, Helios Klinikum Emil von Behring, Berlin, Germany

Background: Lung cancer is the leading cause of cancer-related deaths worldwide. The typical and atypical carcinoid (TC and AC), the large-cell neuroendocrine carcinoma (LCNEC) and the small-cell lung cancers (SCLC) are subgroups of pulmonary tumours that show neuroendocrine differentiations. With the rising impact of molecular pathology in routine diagnostics the interest for reliable biomarkers, which can help to differentiate these subgroups and may enable a more personalised treatment of patients, grows.

Methods: A collective of 70 formalin-fixed, paraffin-embedded (FFPE) pulmonary neuroendocrine tumours (17 TCs, 17 ACs, 19 LCNECs and 17 SCLCs) was used to identify biomarkers by high-throughput sequencing. Using the Illumina TruSeq Amplicon-Cancer Panel on the MiSeq instrument, the samples were screened for alterations in 221 mutation hot spots of 48 tumour-relevant genes.

Results: After filtering >26,000 detected variants by applying strict algorithms, a total of 130 mutations were found in 29 genes and 49 patients. Mutations in JAK3, NRAS, RB1 and VHL1 were exclusively found in SCLCs, whereas the FGFR2 mutation was detected in LCNEC only. KIT, PTEN, HNF1A and SMO were altered in ACs. The SMAD4 mutation corresponded to the TC subtype. We prove that the frequency of mutations increased with the malignancy of tumour type. Interestingly, four out of five ATM-mutated patients showed an additional alteration in TP53, which was by far the most frequently altered gene (28 out of 130, 22%). We found correlations between tumour type and IASLC grade for ATM-(P = 0.022; P = 0.008) and TP53-mutated patients (P<0.001). Both mutated genes were also associated with lymph node invasion and distant metastasis (P<0.005). Furthermore, PIK3CA-mutated patients with high-grade tumours showed a reduced overall survival (P = 0.040) and the mutation frequency of APC and ATM in high-grade neuroendocrine lung cancer patients was associated with progression-free survival (PFS) (P = 0.020).

Conclusions: The implementation of high-throughput sequencing for the analysis of the neuroendocrine lung tumours has revealed that, even if these tumours encompass several subtypes with varying clinical aggressiveness, they share a number of molecular features. An improved understanding of the biology of neuroendocrine tumours will offer the opportunity for novel approaches in clinical management, resulting in a better prognosis and prediction of therapeutic response.
Lung cancer is the leading cause of cancer-related deaths worldwide, with 15% of cases showing only a 5-year survival rate. Nearly one-fourth of all lung neoplasms originate from cells of the diffuse neuroendocrine system and are classified as neuroendocrine tumours (NET) (Rekhtman, 2010). This spectrum of neoplasms is composed of the high-grade small-cell lung cancer (SCLC), the large-cell neuroendocrine carcinoma (LCNEC), as well as the low-grade typical (TC) and atypical carcinoids (AC) (Takei et al, 2002).

Although smoking is not correlating with carcinoids, most of LCNEC and all SCLC patients exhibit a smoking history (Travis, 2009; Swarts et al, 2012). Even though tumours are detected at an early stage, SCLCs and LCNECs show the poorest clinical outcome as compared with other lung tumours because of their high malignancy (Takei et al, 2002). The 5-year survival rate of patients with SCLC is <5% (Rekhtman, 2010). The same holds true for LCNECs that show a similar progression with 5-year survival rates varying from 15% up to 57%, depending on the source of information (Asamura et al, 2006; Rekhtman, 2010). In contrast, patients suffering from TC show excellent survival rates of up to 100%. Nevertheless, some of them present with lymph node metastasis (4–15%) (Travis et al, 1998; Beasley et al, 2000; Fink et al, 2001; Thomas et al, 2001; Cardillo et al, 2004; Pelosi et al, 2005; Asamura et al, 2006; Garcia-Yuste et al, 2007; Rea et al, 2007, Travis, 2009). Atypical carcinoids are more aggressive and show a higher frequency of nodal metastasis (35–64%) with decreased survival rates ranging from 61% to 88% (Beasley et al, 2000; Fink et al, 2001; Scott, 2003; Cardillo et al, 2004; Pelosi et al, 2005; Asamura et al, 2006; Garcia-Yuste et al, 2007; Rea et al, 2007).

During lung carcinogenesis, cells stepwise accumulate multiple genetic and epigenetic abnormalities that lead to immortalisation, uncontrolled proliferation and spread of these altered cells (Brambilla et al, 2003). This is often initiated by the presence of mutations, but can also be an effect of chromosomal aberrations (Esteller et al, 1999; Zochbauer-Muller et al, 2001). In particular, the most common mutated gene in lung cancer is the tumour suppressor TP53, encoding a protein that induces cell growth arrest and apoptosis and mediates cell cycle regulation. In contrast to pulmonary carcinoids, which are rarely TP53 altered, up to 50% of non-small-cell lung cancer (NSCLC) and >70% of SCLCs were associated with mutations in this gene (Travis et al, 2004). In addition, the crucial retinoblastoma (RB) pathway is often inactivated. Up to 100% of high-grade NETs like SCLCs or LCNECs are affected by the loss of the RB protein (Travis et al, 2004).

The advent of next-generation sequencing (NGS) systems allows a very sensitive analysis of multiple patients and targets simultaneously. Therefore, further mutational analysis of neuroendocrine lung tumours by a high-throughput sequencing approach helps to gain more insight into their genetic profile, enabling a further subdivision of these different groups. Here we subjected, for the first time, 70 neuroendocrine lung tumours to a targeted NGS approach based on sequencing-by-synthesis chemistry, feasible for formalin-fixed, paraffin embedded (FFPE) material.

**Materials and Methods**

Demographic data and study design. The study is based on a collective of 70 representative, FFPE neuroendocrine lung tumours (17 TCs, 17 ACs, 19 LCNECs and 17 SCLCs) used for massive parallel sequencing purposes. The initial diagnosis was re-evaluated by two experienced pathologists (IW and TM). Inclusion criteria were sufficient tumour material (tumour cell count 50–80%) and a minimum of contamination by lymphocytic and stromal cells. All analysed cases comprise surgical resection. The surgical resected SCLCs were resected as initial unknown tumoural lesion and were afterwards diagnosed as SCLC. All investigated SCLCs were classified as limited or very limited diseases (VLDL). Specimens were collected at the Institute of Pathology at the University Hospital Essen (Essen, Germany) from 2005 till 2012. The TNM staging was based on the WHO Classification Of Tumours guidelines (2004) (Travis et al, 2004). The mean age at date of diagnosis was 58.6 years (median age: 59.0 years; 95% CI: 50.8–66.9 months). Survival data were available for 34 patients with 12 reported deaths at the time of data collection. Survival data were collected for the high-grade group as well as overall survival (OS) for 3 carcinoid tumours showing an aggressive course with reported death after 4.4, 7.3 and 11.3 months, respectively. For the remaining carcinoids, no survival data were collected because of their excellent prognosis.

The study was conducted retrospectively for the identification of biomarkers. Patients who received chemotherapy before resection of tumour tissue specimens were excluded. The study was approved by the ethical committee of the University Hospital Essen (ID: 13-5382-BO). The investigations conform to the principles outlined in the declaration of Helsinki. Patient characteristics are summarised in Table 1.

**DNA extraction.** For DNA isolation, FFPE tissue slices of 4 μm were prepared using a microtome Cool Cut HM 355S (MICROM International, Walldorf, Germany). Three to five paraffin tissue slices per sample were deparaffinised with xylene before DNA extraction. Genomic DNA was isolated via Maxwell 16 Research (Promega, Madison, WI, USA) as recommended by the manufacturer. Following, DNA quantification was performed using...
Nanodrop ND-1000 (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit (Life Technologies, Carlsbad, CA, USA).

Targeted next-generation sequencing. Sample preparation was done using the TruSeq Amplicon–Cancer Panel (Illumina, San Diego, CA, USA). Therefore, 424 target-specific probes, covering 221 mutation hot spots of 48 tumour-relevant genes, were used to capture the regions of interest (Table 2). Furthermore, each sample was labelled with a barcode sequence consisting of eight individual nucleotides, in a final PCR step. A total of 212 amplicons per sample with a mean length of 180 base pairs were sequenced paired-end on the MiSeq instrument (Illumina) using the v1 chemistry. Sample preparation and sequencing procedure was performed according to the standard protocols provided by the manufacturer.

Sequencing data analysis. Fastq files generated by the MiSeq reporter software (Illumina) were used for further analysis with a customised bioinformatics pipeline based on the general cancer genome analysis algorithm published by Peifer et al (2012).

Subsequently, data filtering was done by excluding variants with <25 effective reads (total reads × (percent variant reads/100)) and by excluding variants with <5% allele frequency. Afterwards, synonymous alterations were removed and visual analysis of called variants was performed by means of the Integrative Genomic Viewer (IGV, Broad Institute, Cambridge, MA, USA). False positive variants, particularly in repetitive or highly homologous regions of the genome, variants in high background noise, as well as single-strand variants, were eliminated when they were clearly recognisable as artefacts. Finally, obtained variants were analysed for their functional impact on the protein by MutationAssessor (http://mutationassessor.org, release 2) (Reva et al, 2010). Only mutations with a functional impact as determined by at least two of the three applied algorithms as well as frameshift deletions/insertion and alterations resulting in a loss of frameshift were used for statistical analysis.

Statistical analysis. All statistical analyses were performed using the R environment for statistical computing and graphics (R Foundation, Vienna, Austria). For dichotomous factors and linear vectors the Wilcoxon rank sum test was applied. For variance analysis of variables with more than two categories, the Kruskal–Wallis test was performed. Double dependence of ranked parameters with more than two groups was verified by the Wald test, likelihood ratio test and Score (logrank) test (Supplementary Table S3). The level of statistical significance was defined as P ≤0.05.

RESULTS

Run parameters and sequencing quality. A total of 70 neuroendocrine lung cancers were sequenced with 4 MiSeq runs, containing 23 or 24 samples per run. Samples showing coverages with < 250 median read counts per amplicon (23 out of 70) were re-sequenced with an additional fifth run.

Bridge amplification of the target regions on the flow cell resulted in an average cluster density of 641 K/mm² with ~80% of the generated clusters passing the quality filter parameters. In total, the sequencing runs produced an average of 3.65 × 10⁶ reads and an average output of 1.1 gigabases. The runs showed read quality parameters with average Q30 scores (equals a 0.1% chance of a wrong base call at this position) of ~63% in a range of 8–85% (Supplementary Table S1). Approximately 60% of the targets showed coverages between 500 and 2000 reads (Supplementary Figure S1).

Variant filtering and mutation detection. The sequencing analysis of all five runs revealed a total of 26,945 variants. Removal of variants with <25% effective reads and variants with <5% allele frequency resulted in 643 variants. Finally, additional removal of variants clearly recognised as artefacts resulted in 254 variants that were further analysed according to their functional relevance, leading to a total of 130 mutations in 29 genes (Table 3) detected in 49 of 70 FFPE neuroendocrine lung tumour samples (Figure 1 and Supplementary Table S2). No variants were found in the remaining 19 genes.

The mutation status of some genes was associated with a special subtype of the pulmonary NETs (Table 3 and Figure 2A). Mutations in JAK3 (exon 16: p.W690*), NRAS (exon 2: p.G13D, p.S17N), RBT1 (exon 18: p.P595L; exon 20: H673Y) and VHL1 (exon 3: Q164*) were exclusively found in SCLCs, whereas the detected FGFR2 mutation p.A264T in exon 7 was only associated with LCNEC. The KIT (exon 10: p.K546delinsG; exon 13 p.M638fs), PTK (exon 8: p.P393S), HNF1A (exon 4: p.G288fs) and SMO (exon 5; p.L367R) alterations were only found in ACs and the SMAD4 (exon 5: p.Q169*; exon 5) mutation corresponded to the TC subtype.

Furthermore, the total mutation frequency increased with the malignancy of the tumour type. Of the 130 mutations, 11 (9%) were related to TCs (mean: 0.6 per sample), 21 (16%) were found in ACs (mean: 1.2 mutations per sample) and 38 (29%) were related to LCNECs (mean: 1.9 per sample). Patients associated to SCLC showed the highest mutation frequency of 46% with 60 detected mutations (mean: 3.5 per sample). This effect was especially observed for ATM, ERBB2, ERBB4, PIK3CA, RET and TP53 (Figure 2B). Interestingly, four of five ATM mutated patients

| Table 2. Genes included in the TruSeq Amplicon-Cancer Panel (Illumina) |
|---------------------------------|
| ABL1 | AKT1 | ALK | APC | ATM | BRAF | CDH1 | CDKN2A |
| CSFTR | CTNNB1 | EGFR | ERBB2 | ERBB4 | FBXW7 | FGFR1 | FGFR2 |
| FGFR3 | FLT3 | GNA11 | GNAQ | GNAS | HNF1A | HRAS | IDH1 |
| JAK2 | JAK3 | KDR | KIT | KRAS | MET | MLH1 | MPL |
| NOTCH1 | NPM1 | NRAS | PDGFRB | PIK3CA | PTEN | PTPN11 | RB1 |
| RET | SMAD4 | SMARCB1 | SMO | SRC | STK11 | TP53 | VHL |

The panel consists of 2 × 212 probes covering 221 mutation hot spots in 48 genes known to be important in human carcinogenesis.
also showed an alteration in the tumour suppressor TP53, and this was by far the most frequently altered gene (31 out of 130; 24%) followed by ERBB2 (11 out of 130, 9%), PIK3CA with 8 (6%) and EGFR present with 7 mutations (5%).

We found significant correlations between tumour type and IASLC grade for ATM (P = 0.022; P = 0.008) and TP53 mutated patients (P > 0.001; P < 0.001). In addition, TP53 showed significant more multiple mutations (P < 0.001). Both genes were also associated with lymph node invasion and distant metastasis when they were mutated (both P ≤ 0.005).

Moreover, the frequency of variants in ATM and APC in high-grade neuroendocrine lung cancer patients was associated with PFS (P = 0.020; Figure 3A). Finally, PIK3CA mutated patients with high-grade tumours showed a reduced OS compared with their wild-type counterparts (P = 0.040; Figure 3B).

**DISCUSSION**

In contrast to NETs of other organ systems (e.g., gastrointestinal), pulmonary NETs form two prognostically different groups: the low-grade carcinoid tumours (TC and AC) and the high-grade carcinomas (LCNEC and SCLC) (Rindi et al, 2014). The detailed molecular characteristic of these subgroups still remains unknown.

Here, for the first time, we describe a targeted high-throughput parallel sequencing approach used to analyse 70 neuroendocrine lung tumours for alterations in 221 hot spots of 48 genes suspected to play an important role in the tumourigenesis. Entity-specific mutations identified by this method can potentially serve as markers for classification of neuroendocrine lung tumours in diagnostics or therapy decision.

Our data confirm previous findings (Lohmann et al, 1993; Przygodzki et al, 1996; Couce et al, 1999; Leotella et al, 2003; Swarts et al, 2012), showing that mutations in the TP53 gene increase with malignancy of the tumour subtype. The TP53 alterations detected in this study ranged from none in carcinoid tumours to 11 in the 17 analysed SCLC samples. Interestingly, the same trend was observed for ATM mutations. Furthermore, four out of five ATM mutated patients showed co-alterations in TP53 (Figure 1). ATM is a serine/threonine kinase that coordinates an integrated cellular response to DNA damage by double-stranded DNA breaks and acts upstream of TP53 which leads to the activation of an apoptotic pathway (Pettitt et al, 2001). Furthermore, it activates several TP53-independent cellular pathways and it is conceivable that mutations in both genes could increase the survival advantage of a malignant tumour cell clone. Subsequently, additional mutations in APC, which is a downstream target of TP53, could possibly enhance the survival benefit for certain tumour subclones. Actually, two out of seven APC mutations detected in our study were observed coincidentally with an ATM or with ATM and TP53 mutations, pointing out a second or third hit mechanism during tumourigenesis.

Interestingly, the group of R Thomas (Fernandez-Cuesta et al, 2014) identified mutations in genes related to chromatin remodelling (including MEN1, ARID1A and EIF1AX) as frequent events in pulmonary carcinoids. In addition, mutations in the histone methyltransferases (including SETD1B, SETDB1 and NSD1) as well as demethylases (KDM4A, PHF8, JMJD1C and others) were determined (Fernandez-Cuesta et al, 2014). Likewise, his working group proofed the appearance of CREBBP and EP300 histone acetyltransferases as well as the histone-modifying enzyme MLL in high-grade lung carcinomas in another study (Peifer et al, 2012), indicating that histone modifications play a crucial role in neuroendocrine lung cancer tumourigenesis and progression.

Activating mutations in the EGFR gene have been widely studied in NSCLC (Engle and Kolesar, 2014; Karachaliou and Rosell, 2014; Revannasiddaiah et al, 2014). Unfortunately, although new therapeutic strategies are urgently needed for metastatic pulmonary carcinoid tumours, this subgroup of lung neoplasia is totally neglected, so far. Instead of common activating EGFR mutations, we found alterations located in the TOPO domain in all four tumour subtypes. In particular, the p.A859V in exon 21 may activate the EGFR receptor similarly to the well-known therapeutically relevant activating p.L858R mutation (Peng et al, 2014). Of note, in one patient we observed that EGFR mutations were located in different regions of the gene. Importantly, these complex events are also under suspicion to predict a response to tyrosine kinase inhibitor (TKI) therapy (Peng et al, 2014).

Similar to EGFR, both other investigated members of the ERBB superfamily (ERBB2 and ERBB4) also showed mutations in the extracellular TOPO domain (11 samples; 2 TCs and ACs, 3 LCNECs and 4 SCLCs). In addition, the growth factor receptor PDGFRα is impaired in four samples (one TC, one LCNEC and two SCLCs). Furthermore, KDR (VEGFR2) is mutated in one AC and two LCNECs. Of note, KDR p.Q427H variant in exon 11, located in the protein kinase domain, occurred in approximately one-third of all carcinoids and the half of all investigated carcinomas. Although all three applied algorithms predicted a benign or tolerated alteration of
Figure 1. Genetic profile of four analysed neuroendocrine lung tumour subtypes carrying gene mutations determined by targeted massive parallel sequencing are shown. Columns represent mutations per patient; rows summarise mutations occurring in a particular gene. Dark blue boxes indicate more than one mutation per patient in the corresponding gene.

Figure 2. (A) Genes within entity-specific mutation frequency are shown. The JAK3, NRAS, RB1 and VHL variants were just detected in SCLCs, whereas ATM, TP53, ALK and BRAF mutations were found in both high-grade tumour entities. The FGFR2 mutation occurred exclusively in LCNEC. The KIT, PTEN, HNF1A and SMO variants were only found in ACs, and the SMAD4 variant was only present in TC. The GNAS mutations were limited to carcinoids. (B) Specific genes showing a higher mutation frequency with increasing tumour malignancy. The bars indicate the distribution of the mutation between the different tumour entities. The TP53 and ATM variants were just detected in high-grade tumours with ATM variants mostly in SCLCs; PIK3CA and ERBB4 showed additional variants in AC; RET mutations were present in all four entities, but mainly in SCLCs, whereas ERBB2 showed an increasing mutation frequency with an increase of the tumour malignancy.
the protein, the KDR p.Q427H mutation is reported to be associated with an increased risk of coronary heart disease and has been detected in patients with NSCLC (Wang et al., 2007; Glubb et al., 2011). Glubb et al. (2011) demonstrated that this alteration can modestly increase VEGFR-2 signaling and may be also associated with increased lung tumour vascularisation. The high proportion of neuroendocrine lung tumours showing this alteration underlines the hypothesis that it might be a potential marker for future risk assessment in a diagnostic setting.

Interestingly, downstream of these receptor tyrosine kinases, well-known activating mutations were observed in both NRAS and KRAS. Pathway activation downstream of EGFR, for example, by KRAS mutations, is known to be a resistance mechanism for EGFR-TKI therapeutic approaches (Engelman et al., 2006). Activation downstream of BRAF, primarily by NRAS mutations, is reported to act as a potential resistance mechanism against vemurafenib therapy (Sos et al., 2014). Even continued activation of PI3K signalling by a PIK3CA oncogenic mutant was shown to be sufficient to abrogate gefitinib-induced apoptosis (Engelman et al., 2006). Additional studies confirmed mutations in the PI3K-ACT domain and the PI3K/PI4K domain of the protein, the RET proto-oncogene, a member of the cadherin superfamily, encodes a receptor tyrosine kinase. Mutations in this gene are associated with the disorders multiple endocrine neoplasms of the lung: a prognostic spectrum. (Krampitz and Norton, 2014). In line with previous findings in other endocrine tumour entities (Rotondi et al., 2009; Schulte et al., 2010; Colombo et al., 2015; Yeganeh et al., 2015), we determined RET variants in ~9% (6 out of 70) of pulmonary NET patients. By the way, the activating p.M918T RET mutation was previously assessed in one metastatic SCLC specimen (Dabir et al., 2014). In contrast, Futami et al. (2003) could not proof mutations but LOH of RET in the majority of SCLCs. Nevertheless, the analysed collective was very small (8 informative samples) (Futami et al., 2003). In carcinoid tumours, high gene expression of RET was associated with low long time survival but was not proven on protein level and also mutations in the RET oncogene could not be identified in any samples of the study (Swarts et al., 2013). In our collective we found one RET mutation in each TC and AC.

Summing up, the implementation of targeted high-throughput sequencing for the analysis of the neuroendocrine lung tumours has revealed that these tumours, even if they vary in their clinical aggressiveness, share a number of molecular features. The impact of these alterations for tumourigenesis and progression of the tumour has to be addressed in further studies, may be smoothing the way for enhanced therapeutic option in the near future.

**AUTHOR CONTRIBUTIONS**

All authors have read and approved the manuscript.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**REFERENCES**

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4): 248–249.

Asamura H, Kameya T, Matuno Y, Noguchi M, Tada H, Ishikawa Y, Yokose T, Jiang SX, Inoue T, Nakagawa K, Tajima K, Nagai K (2006) Neuroendocrine neoplasms of the lung: a prognostic spectrum. *J Clin Oncol* 24(1): 70–76.

BosLeod MB, Thouinis FB, Brambilla E, Hasleton P, Steele R, Hammar SP, Colby TV, Sheppard M, Shimosato Y, Koss MN, Falk R, Travis WD (2000) Pulmonary atypical carcinoid: predictors of survival in 106 cases. *Hum Pathol* 31(10): 1255–1265.

Brambilla C, Fievet F, Jeanmart M, de Fraipont P, Lantuejoul S, Frappat V, Ferretti G, Brichon PY, Merodio Sibilot D (2003) Early detection of lung cancer: role of biomarkers. *Eur Respir J* Suppl 39: 36s–44s.

Capodanno A, Boldrini L, Ali G, Pellicioni S, Mussi A, Fontanini G (2012) Phosphatidylinositol-3-kinase alpha catalytic subunit gene somatic
