HHS Public Access

Author manuscript

Nat Commun. Author manuscript; available in PMC 2014 September 27.

Published in final edited form as:

Nat Commun. ; 5: 3518. doi:10.1038/ncomms4518.

Frequent mutations in chromatin-remodeling genes in pulmonary carcinoids

Lynnette Fernandez-Cuesta[#1], **Martin Peifer**[#1,2], **Xin Lu**[1], **Ruping Sun**[3], **Luka Ozretić**[4], **Danila Seidal**[1,5], **Thomas Zander**[1,6,7], **Frauke Leenders**[1,5], **Julie George**[1], **Christian Müller**[1], **Ilona Dahmen**[1], **Berit Pinther**[1], **Graziella Bosco**[1], **Kathryn Konrad**[8], **Janine Altmüller**[8,9,10], **Peter Nürnberg**[2,8,9], **Viktor Achter**[11], **Ulrich Lang**[11,12], **Peter M Schneider**[13], **Magdalena Bogus**[13], **Alex Soltermann**[14], **Odd Terje Brustugun**[15,16], **Åslaug Helland**[15,16], **Steinar Solberg**[17], **Marius Lund-Iversen**[18], **Sascha Ansén**[6], **Erich Stoelben**[19], **Gavin M. Wright**[20], **Prudence Russell**[21], **Zoe Wainer**[20], **Benjamin Solomon**[22], **John K Field**[23], **Russell Hyde**[23], **Michael PA. Davies**[23], **Lukas C Heukamp**[4,7], **Iver Petersen**[24], **Sven Perner**[25], **Christine Lovly**[26], **Federico Cappuzzo**[27], **William D Travis**[28], **Jürgen Wolf**[5,6,7], **Martin Vingron**[3], **Elisabeth Brambilla**[29], **Stefan A. Haas**[3], **Reinhard Buettner**[4,5,7], and **Roman K Thomas**[1,4,5]

[1]Department of Translational Genomics, Center of Integrated Oncology Cologne–Bonn, University of Cologne, 50924 Cologne, Germany [2]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany [3]Computational Molecular Biology Group, Max Planck Institute for Molecular Genetics, D-14195 Berlin, Germany [4]Department of Pathology, University Hospital Medical Center, University of Cologne, 50937 Cologne, Germany [5]Laboratory of Translational Cancer Genomics, Center of Integrated Oncology Cologne – Bonn, University of Cologne, 50924 Cologne, Germany [6]Department I of Internal Medicine, Center of Integrated Oncology Kö ln-Bonn, University of Cologne, 50924 Cologne, Germany [7]Network Genomic Medicine, University Hospital Cologne, Center of Integrated Oncology Cologne Bonn, 50924 Cologne, Germany [8]Cologne Center for Genomics (CCG), University of Cologne, Cologne, 50931

Correspondence: Roman K Thomas, Department of Translational Genomics, University of Cologne, Weyertal 115b, 50931 Cologne, Germany, +49-221-478-98771, roman.thomas@uni-Köln.de.

Author contributions

LFC and RKT conceived the project. LFC, MP and RKT analyzed, interpreted the data, and wrote the manuscript. LO, CM, ID, BP, KK, JA, and MB performed experiments. LFC, MP and XL performed computational analysis. MP, RS and SAH provided unpublished algorithms. LFC, MP, TZ, RB and RKT gave scientific input. AS, OTB, AH, SS, MLI, SA, ES, GMW, PR, ZW, BS, JKF, RH, MPAD, LCH, IP, SP, CL, FC, EB and RB contributed with samples. LO, WDT, EB, and RB performed pathology review. DS, FL, JG, GB, PN, VA, UL, PMS, SA, JW and MV helped with logistics. All the co-authors reviewed the manuscript.

Competing financial interests

RKT is a founder and shareholder of Blackfield AG. RKT received consulting and lecture fees (Sanofi- Aventis, Merck, Roche, Lilly, Boehringer Ingelheim, Astra-Zeneca, Atlas-Biolabs, Daiichi-Sankyo, MSD, Blackfield AG, Puma) as well as research support (Merck, EOS and AstraZeneca). RB is a cofounder and – owner of Targos Molecular Diagnostics and received honoraria for consulting and lecturing from AstraZeneca, Boehringer Ingelheim, Merck, Roche, Novartis, Lilly, and Pfizer. JW received consulting and lecture fees from Roche, Novartis, Boehringer Ingelheim, AstraZeneca, Bayer, Lilly, Merck, Amgen and research support from Roche, Bayer, Novartis, Boehringer Ingelheim. TZ received honoraria from Roche, Novartis, Boehringer Ingelheim, Lilly, Merck, Amgen and research support from Novartis. CML has served on an Advisory Board for Pfizer and has served as a speaker for Abbott and Qiagen. The remaining authors declare no competing financial interests.

Accession Codes

Sequence data have been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), which is hosted by the EBI, under accession number EGAS00001000650.

Germany [9]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany [10]Institute of Human Genetics, University of Cologne, Cologne 50931, Germany [11]Computing Center, University of Cologne, 50931 Cologne, Germany [12]Department of Informatics, University of Cologne, 50931 Cologne, Germany [13]Institute of Legal Medicine, University of Cologne, 50823 Cologne, Germany [14]Institute for Surgical Pathology, University Hospital Zurich, 8091 Zurich, Switzerland [15]Institute of clinical medicine, Faculty of Medicine, University of Oslo, N-0424 Oslo, Norway [16]Department of oncology, Norwegian Radium Hospital, Oslo University Hospital, N-0310 Oslo, Norway [17]Department of Thoracic Surgery, Rikshospitalet, Oslo University Hospital, N-0027 Oslo, Norway [18]Department of pathology, Norwegian Radium Hospital, Oslo University Hospital, N-0310 Oslo, Norway [19]Thoracic Surgery, Lungenklinik Merheim, Kliniken der Stadt Köln gGmbH, 51109 Cologne, Germany [20]University of Melbourne Department of Surgery, St Vincent's Hospital, Melbourne, 3065 Victoria, Australia [21]Department of Pathology, St. Vincent's Hospital, Melbourne, 3065 Victoria, Australia [22]Department of Haematology and Medical Oncology, Peter MacCallum Cancer Centre, Melbourne, 3002 Victoria, Australia [23]Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, University of Liverpool Cancer Research Centre, Liverpool, L3 9TA, UK [24]Institute of Pathology, Jena University Hospital, Friedrich-Schiller-University, 07743 Jena, Germany [25]Department of Prostate Cancer Research, Institute of Pathology, University Hospital of Bonn, 53127 Bonn, Germany [26]Vanderbilt-Ingram Cancer Center, Nashville, TN37232, USA [27]Department of Medical Oncology, Istituto Toscano Tumouri, 57100 Livorno, Italy [28]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York 10065, USA [29]Department of Pathology, CHU Grenoble INSERM U823, Institute Albert Bonniot 38043 CS10217 Grenoble, France

[#] These authors contributed equally to this work.

Abstract

Pulmonary carcinoids are rare neuroendocrine tumors of the lung. The molecular alterations underlying the pathogenesis of these tumors have not been systematically studied so far. Here we perform gene copy number analysis (n=54), genome/exome (n=44) and transcriptome (n=69) sequencing of pulmonary carcinoids and observe frequent mutations in chromatin-remodeling genes. Covalent histone modifiers and subunits of the SWI/SNF complex are mutated in 40% and 22.2% of the cases respectively, with *MEN1*, *PSIP1* and *ARID1A* being recurrently affected. In contrast to small-cell lung cancer and large-cell neuroendocrine tumors, *TP53* and *RB1* mutations are rare events, suggesting that pulmonary carcinoids are not early progenitor lesions of the highly aggressive lung neuroendocrine tumors but arise through independent cellular mechanisms. These data also suggest that inactivation of chromatin remodeling genes is sufficient to drive transformation in pulmonary carcinoids.

Introduction

Pulmonary carcinoids are neuroendocrine tumors that account for about 2% of pulmonary neoplasms. Based on the WHO classification of 2004, carcinoids can be subdivided in

typical or atypical, the latter ones being very rare (about 0.2%)[1]. Most carcinoids can be cured by surgery; however, inoperable tumors are mostly insensitive to chemo- and radiation therapies1. Apart from few low-frequency alterations, such as mutations in *MEN1*[1], comprehensive genome analyses of this tumor type have so far been lacking.

Here we conduct integrated genome analyses[2] on data from chromosomal gene copy number of 54 tumors, genome and exome sequencing of 29 and 15 tumor-normal pairs respectively, as well as transcriptome sequencing of 69 tumors. Chromatin-remodeling is the most frequently mutated pathway in pulmonary carcinoids; the genes *MEN1*, *PSIP1* and *ARID1A* were recurrently affected by mutations. Specifically, covalent histone modifiers and subunits of the SWI/SNF complex are mutated in 40% and 22.2% of the cases respectively. By contrast, mutations of *TP53* and *RB1* are only found in 2 out of 45 cases, suggesting that these genes are not main *drivers* in pulmonary carcinoids.

Results

In total, we generated genome/exome sequencing data for 44 independent tumor-normal pairs, and for most of them, also RNAseq (n=39, 69 in total), and SNP 6.0 (n=29, 54 in total) data (**Supplementary Table S1)**. Although no significant focal copy number alterations were observed across the tumors analyzed, we detected a copy number pattern compatible with chromothripsis3 in a stage-III atypical carcinoid of a former smoker (**Fig. 1a; Supplementary Fig. S1**). The intensely clustered genomic structural alterations found in this sample were restricted to chromosomes 3, 12, and 13, and led to the expression of several chimeric transcripts (**Fig. 1b; Supplementary Table S2**). Some of these chimeric transcripts affected genes involved in chromatin remodeling processes, including out-of-frame fusion transcripts disrupting the genes, *ARID2*, *SETD1B*, and *STAG1*. Through the analyses of genome and exome sequencing data, we detected 529 non-synonymous mutations in 494 genes, which translates to a mean somatic mutation rate of 0.4 mutations per megabase (Mb) (**Fig. 1c; Supplementary Data 1**), which is much lower than the rate observed in other lung tumors (**Fig. 1c**)[2,4,5]. As expected, and in contrast to small-cell lung cancer (SCLC), no smoking-related mutation signature was observed in the mutation pattern of pulmonary carcinoids (**Fig. 1d**).

We identified *MEN1*, *ARID1A* and *EIF1AX* as significantly mutated genes[2] (q-value<0.2, see Methods section) (Fig. 2a; Supplementary Table S1 and S3; Supplementary Data 1). *MEN1* and *ARID1A* play important roles in chromatin remodeling processes. The tumor suppressor MEN1 physically interacts with MLL and MLL2 to induce gene transcription[6]. Specifically, MEN1 is a molecular adaptor that physically links MLL with the chromatin-associated protein PSIP1, an interaction that is required for MLL/MEN1-dependent functions7. MEN1 also acts as a transcriptional repressor through the interaction with SUV39H1[8]. We observed mutually exclusive frame-shift and truncating mutations in *MEN1* and *PSIP1* in 6 cases (13.3%), which were almost all accompanied by loss of heterozygosity (LOH) (**Supplementary Fig. S2**). We also detected mutations in histone methyltransferases (*SETD1B*, *SETDB1* and *NSD1*) and demethylases (*KDM4A*, *PHF8* and *JMJD1C*), as well as in the following members of the Polycomb complex[9] (Supplementary Table S1 and S2; Supplementary Data 1): *CBX6*, which belongs to the Polycomb repressive complex 1

(PRC1); *EZH1*, which is part of the Polycomb repressive complex 2 (PRC2); and *YY1*, a member of the PHO repressive complex 1 that recruits PRC1 and PRC2. *CBX6* and *EZH1* mutations were also accompanied by LOH (**Supplementary Fig. S2**). In addition, we also detected mutations in the histone modifiers *BRWD3* and *HDAC5* in one sample each. In total, 40% of the cases carried mutually exclusive mutations in genes that are involved in covalent histone modifications (*q*-value=$8x10^{-7}$, see Methods section) (**Fig. 2a; Supplementary Table S4**). In order to evaluate the impact of these mutations on histone methylation, we compared the levels of the H3K9me3 and H3K27me3 on 7 mutated and 6 wild-type samples, and observed a trend towards lower methylation in the mutated cases (**Table 1; Fig. 2b**).

Truncating and frame-shift mutations in *ARID1A* were detected in 3 cases (6.7%). ARID1A is one of the two mutually exclusive ARID1 subunits, believed to provide specificity to the ATP-dependent SWI/SNF chromatin-remodeling complex[10,11]. Truncating mutations of this gene have been reported at high frequency in several primary human cancers[12]. In total, members of this complex were mutated in mutually exclusive fashion in 22.2% of the specimens (*q*-value=$8x10^{-8}$, see Methods section) (**Fig. 2a; Supplementary Table S4**). Among them were the core subunits *SMARCA1*, *SMARCA2*, and *SMARCA4*, which carry the ATPase activity of the complex, as well as the subunits *ARID2*, *SMARCC2*, *SMARCB1*, and, *BCL11A* (Fig. 2a; Supplementary Table S1 and S2; Supplementary Data 1)[13,14]. Another recurrently affected pathway was sister-chromatid cohesion during cell cycle progression with the following genes mutated (Fig. 2a; Supplementary Table S1 and S2; Supplementary Data 1; Supplementary Fig. S3): the cohesin subunit *STAG1*[15], the cohesin loader *NIPBL*[16]; the ribonuclease and microRNA processor *DICER*, necessary for centromere establishment[17]; and ERCC6L, involved in sister chromatid separation[18]. In addition, although only few chimeric transcripts were detected in the 69 transcriptomes analyzed (**Supplementary Table S5),** we found one sample harboring an inactivating chimeric transcript leading to the loss of the mediator complex gene *MED24* (**Supplementary Fig. S4**) that interacts both physically and functionally with cohesin and NIPBL to regulate gene expression[19]. In summary, we detected mutations in chromatin remodeling genes in 23 (51.1%) of the samples analyzed. The specific role of histone modifiers in the development of pulmonary carcinoids was confirmed by the lack of significance of these pathways in SCLC[2] (**Supplementary Table S4**). This was further supported by a gene expression analysis including 50 lung adenocarcinomas (unpublished data), 42 SCLC[2,20], and the 69 pulmonary carcinoids included in this study (**Supplementary Data 2**). Consensus k-means clustering revealed that although both SCLC and pulmonary carcinoids are lung neuroendocrine tumors, both tumor types as well as adenocarcinomas formed statistically significant separate clusters (**Fig. 3a**). In support of this notion, we recently reported that the early alterations in SCLC universally affect *TP53* and *RB1*[2], whereas in this study these genes were only mutated in two samples (Fig 2a; Supplementary Table S1; Supplementary Data 1). Moreover, when examining up- and down-regulated pathways in SCLC versus pulmonary carcinoids by Gene Set Enrichment Analysis (GSEA)[21], we found that in line with the pattern of mutations, the RB1 pathway was statistically significantly altered in SCLC (q-value=$5x10^{-4}$, see Methods section) but not in pulmonary carcinoids (**Fig. 3b; Supplementary Table S6**).

Another statistically significant mutated gene was the eukaryotic translation initiation factor 1A (*EIF1AX*) mutated in 4 cases (8.9%). Additionally, *SEC31A*, *WDR26*, and the E3-ubiquitin ligase *HERC2* were mutated in two samples each. Further supporting a role of E3 ubiquitin ligases in the development of pulmonary carcinoids we found mutations or rearrangements affecting these genes in 17.8% of the samples analyzed (Fig. 2a; Supplementary Table S1 and S7; Supplementary Data 1). All together, we have identified candidate driver genes in 73.3% of the cases. Of note, we did not observe any genetic segregation between typical or atypical carcinoids, neither between the expression clusters generated from the two subtypes, nor between these clusters and the mutated pathways (**Supplementary Fig. S5**). However, it is worth mentioning that only 9 atypical cases were included in this study. The spectrum of mutations found in the discovery cohort, was further validated by transcriptome sequencing of an independent set of pulmonary carcinoid specimens (**Supplementary Table S1 and S8**). Due to the fact that many nonsense and frame-shift mutations may result in nonsense-mediated decay[22,23], the mutations detected by transcriptome sequencing were only missense. Due to this bias, accurate mutation frequencies could not be inferred from these data.

Discussion

This study defines recurrently mutated sets of genes in pulmonary carcinoids. The fact that almost all of the reported genes were mutated in a mutually exclusive manner and affected a small set of cellular pathways, defines these as the key pathways in this tumor type. Given the frequent mutations affecting the few signaling pathways described above and the almost universal absence of other cancer mutations, our findings support a model where pulmonary carcinoids are not early progenitor lesions of other neuroendocrine tumors, such as small-cell lung cancer or large-cell neuroendocrine carcinoma, but arise through independent cellular mechanisms. More broadly, our data suggest that mutations in chromatin remodeling genes, which in recent studies were found frequently mutated across multiple malignant tumours[24], are sufficient to drive early steps in tumorigenesis in a precisely defined spectrum of required cellular pathways.

Methods

Tumor specimens

The study as well as written informed consent documents had been approved by the Institutional Review Board of the University of Cologne. Additional biospecimens for this study were obtained from the Victorian Cancer Biobank, Melbourne, Australia; the Vanderbilt-Ingram Cancer Center, Nashville, USA; and Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Center, Liverpool, UK. The Institutional Review Board (IRB) of each participating institution approved collection and use of all patient specimens in this study.

Nucleic acid extraction and sample sequencing

All samples in this study were reviewed by expert pathologists. Total RNA and DNA were obtained from fresh-frozen tumor and matched fresh-frozen normal tissue or blood. Tissue

was frozen within 30 min after surgery and was stored at –80 °C. Blood was collected in tubes containing the anticoagulant EDTA and was stored at –80 °C. Total DNA and RNA were extracted from fresh-frozen lung tumor tissue containing more than 70% tumor cells. 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 treated accordingly. DNA from sections and blood was extracted using the Puregene Extraction kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 1× TE buffer (Qiagen), diluted to a working concentration of 150 ng—l and stored at –80 °C. For whole exome sequencing we fragmented 1 μg of DNA with sonification technology (Bioruptor, diagenode, Liège, Belgium). The fragments were endrepaired and adaptor-ligated, including incorporation of sample index barcodes. After size selection, we subjected the library to an enrichment process with the SeqCap EZ Human Exome Library version 2.0 kit (Roche NimbleGen, Madison, WI, USA). The final libraries were sequenced with a paired-end 2×100 bp protocol. On average, 7 Gb of sequence were produced per normal, resulting in 30x coverage of more than 80% of target sequences (44Mb). For better sensitivity, tumors were sequenced with 12Gb and 30x coverage of more than 90%. We filtered primary data according to signal purity with the Illumina Realtime Analysis software. Whole genome sequencing was also performed using a read length of 2x 100bp for all samples. On average, 110 Gb of sequence were produced per sample, aiming a mean coverage of 30x for both tumor and matched-normal. RNAseq was performed on cDNA libraries prepared from PolyA+ RNA extracted from tumor cells using the Illumina TruSeq protocol for mRNA. The final libraries were sequenced with a paired-end 2×100 bp protocol aiming at 8.5 Gb per sample, resulting on a 30x mean coverage of the annotated transcriptome. All the sequencing was carry on an Illumina HiSeq™ 2000 sequencing instrument (Illumina, San Diego, CA, USA).

Sequence data processing and mutation detection

Raw sequencing data are aligned to the most recent build of the human genome (NCBI build 37/hg19) using BWA (version: 0.5.9rc1)[25] and possible PCR-duplicates are subsequently removed form the alignments. Somatic mutations were detected using our in-house developed sequencing analysis pipeline. In brief, the mutation calling algorithm incorporates parameters such as local copy number profiles, estimates of tumor purity and ploidy, local sequencing depth, as well as the global sequencing error into a statistical model with which the presence of a mutated allele in the tumor is determined. Next, the absence of this variant in the matched normal is assessed by demanding that the corresponding allelic fraction is compatible with the estimated background sequencing error in the normal. In addition, we demand that the allelic fractions between tumor and normal differ significantly. To finally remove artificial mutation calls, we apply a filter that is based on the forward-reverse bias of the sequencing reads. Further details of this approach are given in Peifer *et* al.[2]

Genomic rearrangement reconstruction from paired-end data

To reconstruct rearrangements from paired-end data, we refined our initial method[2] by adding breakpoint-spanning reads. Here, locations of encompassing read pairs are screened for further reads where only one pair aligns to the region and the other pair either does not align at all or is clipped by the aligner. These reads are then realigned using BLAT to a

1000bp region around the region defined by the encompassing reads. Rearrangements confirmed by at least one spanning read are finally reported. To filter for somatic rearrangements, we subtracted those regions where rearrangements are present in the matched-normal and in all other sequenced normals within the project.

Analysis of significantly mutated genes and pathways

The analysis of significantly mutated genes is done in a way that both gene expression and the accumulation of synonymous mutations are considered to obtain robust assessments of frequently mutated, yet biologically relevant genes. To this end, the overall background mutation rate is determined first, from which the expected number of mutations for each gene is computed under the assumption of a purely random mutational process. This gene specific expected number of mutations defines the underlying null model of our statistical test. To account for misspecifications, e.g., due to a local variation of mutation rates, we also incorporated the synonymous to non-synonymous ratio into a combined statistical model to determine significantly mutated genes. Since mutation rates in non-expressed genes are often high than the genome-wide background rate[2,26], genes that are having a median FPKM value less than one in our transcriptome sequencing data are removed prior testing. To account for multiple hypothesis testing, we are using the Benjamini-Hochberg approach[27]. Mutation data of the total of 44 samples, for which either WES or WGS was performed, were used for this analysis.

In case of the pathway analysis, gene lists of the methylation- and the SWI/SNF complex were obtained from recent publications[9,13,14,28]. To assess whether mutations in these pathways are significantly enriched, all genes of the pathway are grouped together as if they represent a ”single gene” and subsequently tested if the total number of mutation exceed mutational background of the entire pathway. To this end, the same method as described above was used. Mutation data of the total of 44 samples, for which either WES or WGS was performed, were used for this analysis.

Analysis of chromosomal gene copy number data

Hybridization of the Affymetrix SNP 6.0 arrays was carried out according to the manufacturers' instructions and analyzed as follows: raw signal intensities were processed by applying a log-linear model to determine allele-specific probe affinities and probe-specific background intensities. To calibrate the model, a Gauss-Newton approach was used and the resulting raw copy number profiles are segmented by applying the circular binary segmentation method[29].

Analysis of RNAseq data

For the analysis of RNAseq data, we have developed a pipeline that affords accurate and efficient mapping and downstream analysis of transcribed genes in cancer samples (Lynnette Fernandez-Cuesta and Ruping Sun, personal communication). In brief, paired-end RNAseq reads were mapped onto hg19 using a sensitive gapped aligner, GSNAP[30]. Possible breakpoints were called by identifying individual reads showing split-mapping to distinct locations as well as clusters of discordant read pairs. Breakpoint assembly was performed to

leverage information across reads anchored around potential breakpoints. Assembled contigs were aligned back to the reference genome to confirm *bona fide* fusion points.

Dideoxy sequencing

All non-synonymous mutations found in the genome/exome data were checked in RNAseq data when available. Genes recurrently mutated involved in pathways statistically significantly mutated, or interesting because of their presence in other lung neuroendocrine tumors, were selected for validation. 158 mutations were considered for validation: 115 validated and 43 did not (validation rate 73%). Sequencing primer pairs were designed to enclose the putative mutation (**Supplementary Data 1**), or to encompass the candidate rearrangement (**Supplementary Table S7**) or chimeric transcript (**Supplementary Table S2 and S5**). Sequencing was carried out using dideoxy-nucleotide chain termination (Sanger) sequencing, and electropherograms were analyzed by visual inspection using 4 Peaks.

Gene expression data analyses

Unsupervised consensus clustering was applied to RNAseq data of 69 pulmonary carcinoids, 50 AD, and 42 SCLC[2,20] samples. The 3000 genes with highest variation across all samples were filtered out before performing consensus clustering. We used the clustering module from GenePattern[31] and the consensus CDF[32,33]. Significance was obtained by using SigClust[34]. Fisher's exact test[35] was used to check for associations between clusters and histological subtypes. GSEA[21] were performed on 69 pulmonary carcinoids and 42 SCLC[2,20] samples; and the gene sets *oncogenic signatures* were used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are indebted to the patients donating their tumor specimens as part of the Clinical Lung Cancer Genome Project initiative. We thank Philipp Lorimier, Elisabeth Kirst, Emilia Müller, and Juana Cuesta Valdes for their technical assistance. We furthermore thank the regional computing center of the University of Cologne (RRZK) for providing the CPU time on the DFG-funded supercomputer 'CHEOPS' as well as the support.

This work was supported by the Deutsche Krebshilfe as part of the small-cell lung cancer genome-sequencing consortium (grant ID: 109679 to RKT, MP, RB, PN, MV and SAH). Additional funds were provided by the EU-Framework program CURELUNG (HEALTH-F2-2010-258677 to RKT, JW, JKF and EB); by the German federal state North Rhine Westphalia (NRW) and the European Union (European Regional Development Fund: *Investing In Your Future*) within PerMed NRW (grant 005-1111-0025 to RKT, JW, RB); by the Deutsche Forschungsgemeinschaft through TH1386/3-1 (to RKT) and through SFB832 (TP6 to RKT and JW; TP5 to LCH); by the German Ministry of Science and Education (BMBF) as part of the NGFNplus program (grant 01GS08101 to RKT, JW, PN); by the Deutsche Krebshilfe as part of the *Oncology Centers of Excellence* funding program (RKT, RB, JW); by *Stand Up To Cancer*–American Association for Cancer Research Innovative Research Grant (SU2C-AACR-IR60109 to RKT); by an NIH K12 training grant (K12 CA9060625) and by an *Uniting Against Lung Cancer* grant, and a Damon Runyon Clinical Investigator Award (to CML); and by AIRC and Istituto Toscano Tumori project F13/16 (to FC).

References

1. Swarts, D. R. a; Ramaekers, FCS.; Speel, E-JM. Molecular and cellular biology of neuroendocrine lung tumors: evidence for separate biological entities. Biochim. Biophys. Acta. 2012; 1826:255–271. [PubMed: 22579738]
2. Peifer M, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat. Genet. 2012; 44:1104–1110. [PubMed: 22941188]
3. Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell. 2011; 144:27–40. [PubMed: 21215367]
4. Imielinski M, et al. Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. Cell. 2012; 150:1107–1120. [PubMed: 22980975]
5. Hammerman PS, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012
6. Marx SJ. Molecular genetics of multiple endocrine neoplasia types 1 and 2. Nat. Rev. Cancer. 2005; 5:367–376. [PubMed: 15864278]
7. Yokoyama A, Clearly M. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell. 2008; 8:2469.
8. Yang Y-J, et al. Menin mediates epigenetic regulation via histone H3 lysine 9 methylation. Cell Death & Disease. 2013; 4:e583. [PubMed: 23579270]
9. Lanzuolo C, Orlando V. Memories from the polycomb group proteins. Annu. Rev. Genet. 2012; 46:561–89. [PubMed: 22994356]
10. Roberts CWM, Orkin SH. The SWI/SNF complex — chromatin and cancer. Nat. Rev. Cancer. 2004; 4:133–142. [PubMed: 14964309]
11. Wu JI, Lessard J, Crabtree GR. Understanding the words of chromatin regulation. Cell. 2009; 136:200–206. [PubMed: 19167321]
12. Wilson BG, Roberts CWM. Epigenetics and genetics SWI/SNF nucleosome remodellers and cancer. Nat. Rev. Cancer. 2011; 11:481–492. [PubMed: 21654818]
13. Tang J, Yoo AS, Crabtree GR. Reprogramming human fibroblasts to neurons by recapitulating an essential microRNA-chromatin switch. Current opinion in genetics & development. 2013; 23:591–8. [PubMed: 24035011]
14. Kadoch C, et al. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat. Genet. 2013; 45:1–11. [PubMed: 23268125]
15. Peters J, Tedeschi A, Schmitz J. The cohesin complex and its roles in chromosome biology. Genes Dev. 2008; 22:3089–3114. [PubMed: 19056890]
16. Ciosk R, et al. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell. 2000; 5:243–254. [PubMed: 10882066]
17. Fukagawa T, et al. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat. Cell Biol. 2004; 6:784–791. [PubMed: 15247924]
18. Baumann C, Korner R, Hofmann K, Nigg EA. PICH, a centromere-associated SNF2 family ATPase, is regulated by Plk1 and required for the spindle checkpoint. Cell. 2007; 128:101–114. [PubMed: 17218258]
19. Kagey MH, et al. Mediator and Cohesin Connect Gene Expression and Chromatin Architecture. Nature. 2010; 467:430–435. [PubMed: 20720539]
20. Rudin CM, et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat. Genet. 2012; 44:1111–1116. [PubMed: 22941189]
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL. Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide expression profiles. PNAS. 2005
22. Nicholson P, et al. Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cellular and molecular life sciences: CMLS. 2010; 67:677–700. [PubMed: 19859661]
23. Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Muhlemann O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. RNA (New York, N.Y.). 2011; 17:2108–18.

24. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nature Reviews Cancer. 2013; 13:497–510.
25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009; 25:1754–60.
26. Lawrence MS, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013:10–14.
27. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57
28. Black JC, Van Rechem C, Whetstine JR. Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol. Cell. 2012; 48:491–507. [PubMed: 23200123]
29. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics (Oxford, England). 2004; 5:557–72.
30. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics (Oxford, England). 2010; 26:873–81.
31. Kuehn H, Liberzon A, Reich M, Mesirov JP. Using GenePattern for gene expression analysis. Curr. Protoc. Bioinformatics. 2008; 22:7.12.1–7.12.39.
32. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics (Oxford, England). 2010; 26:1572–3.
33. Monti S, et al. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning Journal. 2003; 52(1-2):91–118.
34. Liu Y, et al. Statistical significance of clustering for high-dimension, low-sample size data. J Am. Stat. Assoc. 2008; 103:1281–1293.
35. Fisher RA. Statistical Methods for Research Workers. Oliver and Boyd. 1954
36. Karro JE, et al. Exponential decay of GC content detected by strand-symmetric substitution rates influences the evolution of isochore structure. Mol. Biol. Evol. 2008; 25:362–374. [PubMed: 18042807]

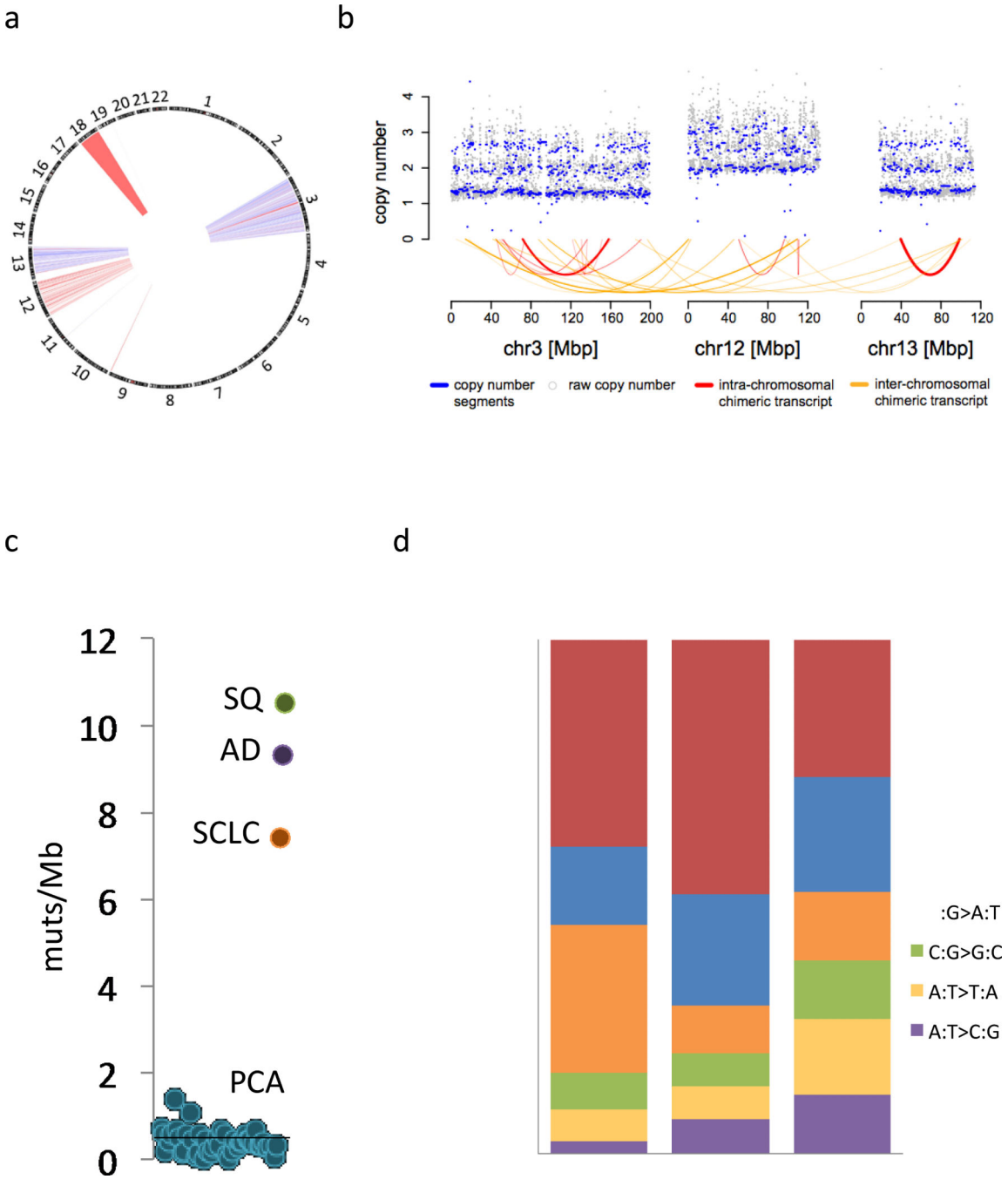

Figure 1.

Genomic characterization of pulmonary carcinoids. (a) CIRCOS plot of the chromothripsis case. The outer ring shows chromosomes arranged end to end. Somatic copy number alterations (gains in red and losses in blue) detected by 6.0 SNP arrays are depicted in the inside ring. (**b**) Copy numbers and chimeric transcripts of affected chromosomes. Segmented copy number states (blue points) are shown together with raw copy number data averaged over 50 adjacent probes (grey points). To show the different levels of strength for the identified chimeric transcripts all curves are scaled according to the sequencing coverage

at the fusion-point. (**c**) Mutation frequency detected by genome and exome sequencing in pulmonary carcinoids (PCA). Each blue dot represents the number of mutations per megabase in one pulmonary carcinoid sample. Average frequencies are also shown for adenocarcinomas (AD), squamous (SQ), and small-cell lung cancer (SCLC) base on previous studies[2,4,5] (**d**) Comparison of context independent transversion and transition rates (an overall strand symmetry is assumed) between rates derived from molecular evolution (evol)[36], from a previous SCLC sequencing study[2], and from the pulmonary carcinoids (PCA) genome and exome sequencing. All rates are scaled as such that their overall sum is one.

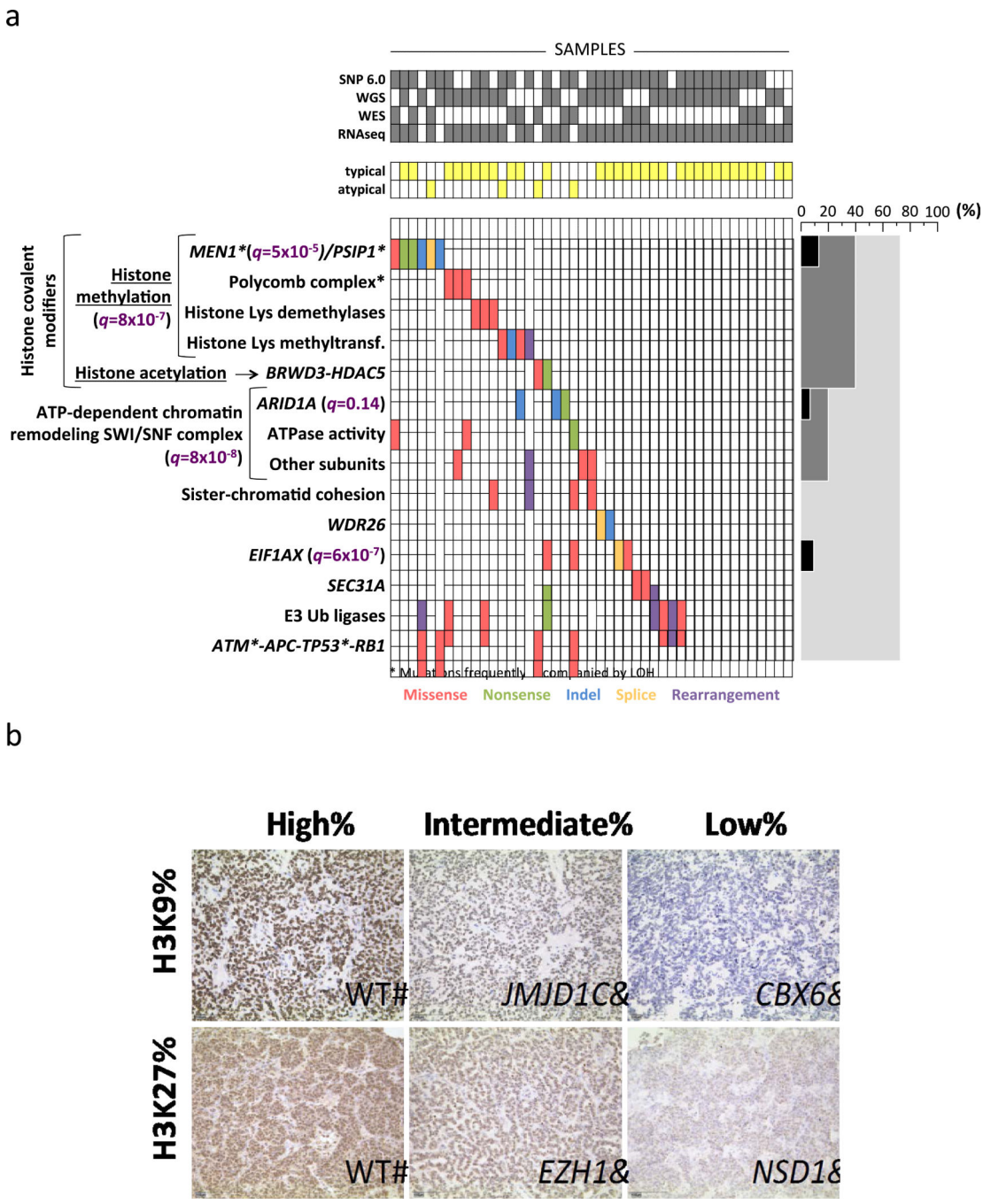

Figure 2.

Significant affected genes and pathways in pulmonary carcinoids. (a) Significantly mutated genes and pathways identified by genome (n=29), exome (n=15) and transcriptome (n=69) sequencing. The percentage of pulmonary carcinoids with a specific gene or pathway mutated is noted at the right side. The *q*-values of the significantly mutated genes and pathways are shown in brackets (see Methods section). Samples are displayed as columns and arranged to emphasize mutually exclusive mutations. (**b**) Methylation levels of H3K9me3 and H3K27me3 in pulmonary carcinoids. Representative pictures of different

degrees of methylation (high, intermediate, and low) for some of the samples summarized in Table 1. The mutated gene is shown in italics at the bottom right part of the correspondent picture. Wild-type samples are denoted by WT.

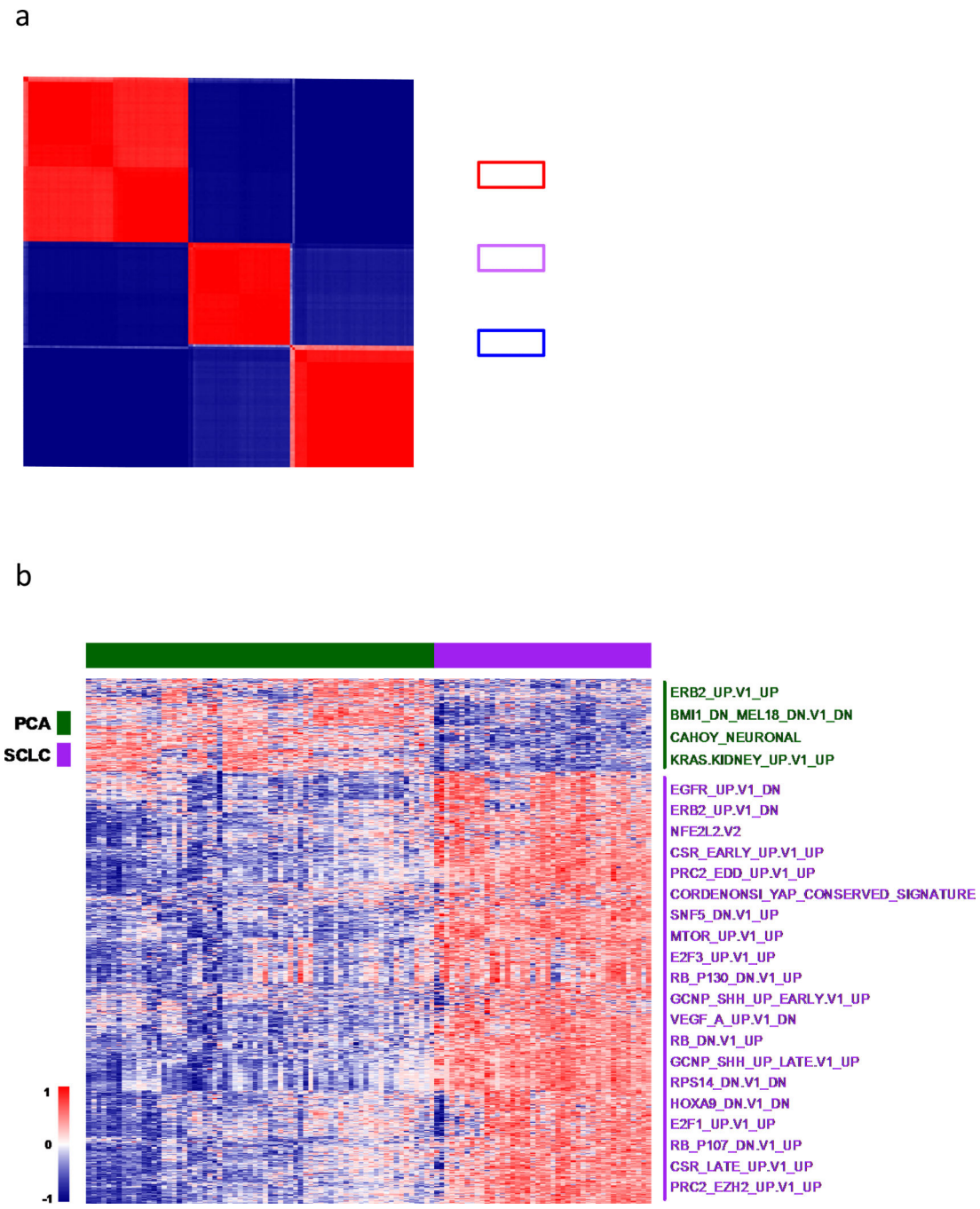

Figure 3.

Expression data analysis of pulmonary carcinois based on RNAseq data. (a) Consensus Kmeans clustering[32,33] using RNAseq expression data of 50 adenocarcinomas (AD, in blue), 42 small-cell lung cancer (SCLC, in red), and 69 pulmonary carcinoids (PCA, in purple) identified 3 groups using the clustering module from GenePattern[31] and consensus CDF[32,33] (left panel). The significance of the clustering was evaluated by using SigClust34 with a $p<0.0001$. Fisher's exact test35 was used to check associations between the clusters and the histological subtypes (right panel). (**b**) Gene Set Enrichment Analysis (GSEA)[21] for

SCLC versus PCA using RNAseq expression data. Low gene expression is indicated in blue and high expression, in red. On the right side are named the altered pathways in PCA (green) and SCLC (purple).

Table 1

Overview of samples annotated for mutations in genes involved in histone methylation, and correspondent levels of H3K9me3 and H3K27me3 detected by immunohistochemistry.

SAMPLE	MUTATION	H3K9me3	H3K27me3
S02333	*JMJD1C*_H954N	Intermediate	Low
S01502	*KDM4A*_I168T	Intermediate	N/A
S02323	*MEN1*_e3+1 and LOH	Low	Low
S02339	*NSD1*_A1047G	Intermediate	Low
S02327	*CBX6*_P302S and LOH	Low	Low
S01746	*EZH1*_R728G and LOH	Low	Intermediate
S02325	YY1_E253K	Low	Intermediate
S01501	Wild type	N/A	High
S01731	Wild type	Low	Low
S01742	Wild type	High	High
S02334	Wild type	Intermediate	High
S02337	Wild type	High	High
S02338	Wild type	High	Intermediate