nature genetics

Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer

Martin Peifer[1,2,57], Lynnette Fernández-Cuesta[1,2,57], Martin L Sos[1–4], Julie George[1,2], Danila Seidel[1,2,5], Lawryn H Kasper[6], Dennis Plenker[1,2], Frauke Leenders[1,2,5], Ruping Sun[7], Thomas Zander[1–4], Roopika Menon[8], Mirjam Koker[1,2], Ilona Dahmen[1,2], Christian Müller[1,2], Vincenzo Di Cerbo[9], Hans-Ulrich Schildhaus[10], Janine Altmüller[11], Ingelore Baessmann[11], Christian Becker[11], Bram de Wilde[12], Jo Vandesompele[12], Diana Böhm[8], Sascha Ansén[3,4], Franziska Gabler[2], Ines Wilkening[2], Stefanie Heynck[2], Johannes M Heuckmann[1,2], Xin Lu[1,2], Scott L Carter[13], Kristian Cibulskis[13], Shantanu Banerji[13], Gad Getz[13], Kwon-Sik Park[14,15], Daniel Rauh[16], Christian Grütter[16], Matthias Fischer[17,18], Laura Pasqualucci[19,20], Gavin Wright[21], Zoe Wainer[21], Prudence Russell[22], Iver Petersen[23], Yuan Chen[23], Erich Stoelben[24], Corinna Ludwig[24], Philipp Schnabel[25], Hans Hoffmann[26], Thomas Muley[26], Michael Brockmann[27], Walburga Engel-Riedel[24], Lucia A Muscarella[28], Vito M Fazio[28], Harry Groen[29], Wim Timens[30], Hannie Sietsma[30], Erik Thunnissen[31], Egbert Smit[32], Daniëlle A M Heideman[31], Peter J F Snijders[31], Federico Cappuzzo[33], Claudia Ligorio[34], Stefania Damiani[34], John Field[35], Steinar Solberg[36], Odd Terje Brustugun[37,38], Marius Lund-Iversen[39], Jörg Sänger[40], Joachim H Clement[41], Alex Soltermann[42], Holger Moch[42], Walter Weder[43], Benjamin Solomon[44], Jean-Charles Soria[45], Pierre Validire[46], Benjamin Besse[45], Elisabeth Brambilla[47,48], Christian Brambilla[47,48], Sylvie Lantuejoul[47,48], Philippe Lorimier[47], Peter M Schneider[49], Michael Hallek[3–5], William Pao[50], Matthew Meyerson[13,51–53], Julien Sage[14,15], Jay Shendure[54], Robert Schneider[9,55], Reinhard Büttner[5,10], Jürgen Wolf[3–5], Peter Nürnberg[11,18,56], Sven Perner[8], Lukas C Heukamp[10], Paul K Brindle[6], Stefan Haas[7] & Roman K Thomas[1–5,10]

Small-cell lung cancer (SCLC) is an aggressive lung tumor subtype with poor prognosis[1–3]. We sequenced 29 SCLC exomes, 2 genomes and 15 transcriptomes and found an extremely high mutation rate of 7.4 ± 1 protein-changing mutations per million base pairs. Therefore, we conducted integrated analyses of the various data sets to identify pathogenetically relevant mutated genes. In all cases, we found evidence for inactivation of *TP53* and *RB1* and identified recurrent mutations in the *CREBBP*, *EP300* and *MLL* genes that encode histone modifiers. Furthermore, we observed mutations in *PTEN*, *SLIT2* and *EPHA7*, as well as focal amplifications of the *FGFR1* tyrosine kinase gene. Finally, we detected many of the alterations found in humans in SCLC tumors from *Tp53* and *Rb1* double knockout mice[4]. Our study implicates histone modification as a major feature of SCLC, reveals potentially therapeutically tractable genomic alterations and provides a generalizable framework for the identification of biologically relevant genes in the context of high mutational background.

Small-cell lung cancer (~15% of all lung cancer cases) typically occurs in heavy smokers and is characterized by aggressive growth, frequent metastases and early death[1,2,5]. Unfortunately, no single molecularly targeted drug has yet shown any clinical activity in SCLC[6]. Genomic analyses have revealed genetically altered therapeutic targets in lung adenocarcinoma[7–16] and in squamous cell lung carcinoma[17–19]. By contrast, little is known about the molecular events causing SCLC beyond the high prevalence of mutations in *TP53* and *RB1* (ref. 3). Systematic genomic analyses in SCLC are challenging because these tumors are rarely treated by surgery, resulting in a lack of suitable fresh-frozen tumor specimens.

We have established a global lung cancer genome research consortium[19], giving us access to approximately 6,600 surgically resected lung cancer specimens, out of which we retrieved 99 SCLC specimens. We conducted 6.0 SNP array analyses of 63 tumors, exome sequencing of 27 tumors and 2 cell lines, transcriptome sequencing of 15 tumors and genome sequencing of 2 tumors (**Supplementary Table 1**).

We applied a novel algorithm to identify significant broad (**Supplementary Fig. 1a**) and focal (**Fig. 1a** and **Supplementary Table 2**) somatic copy-number alterations (SCNAs) and observed almost universal deletions affecting 3p and 13q (containing *RB1*), frequent gains of 3q and 5p, and losses of 17p (containing *TP53*) (**Supplementary Fig. 1a**). Gains of 3q affected the region containing *SOX2*, which was recently shown to be focally amplified in squamous cell lung cancer[19,20]. However, 3q gains in SCLC were less focal

A full list of affiliations appears at the end of the paper.

Received 22 December 2011; accepted 9 August 2012; published online 2 September 2012; doi:10.1038/ng.2396

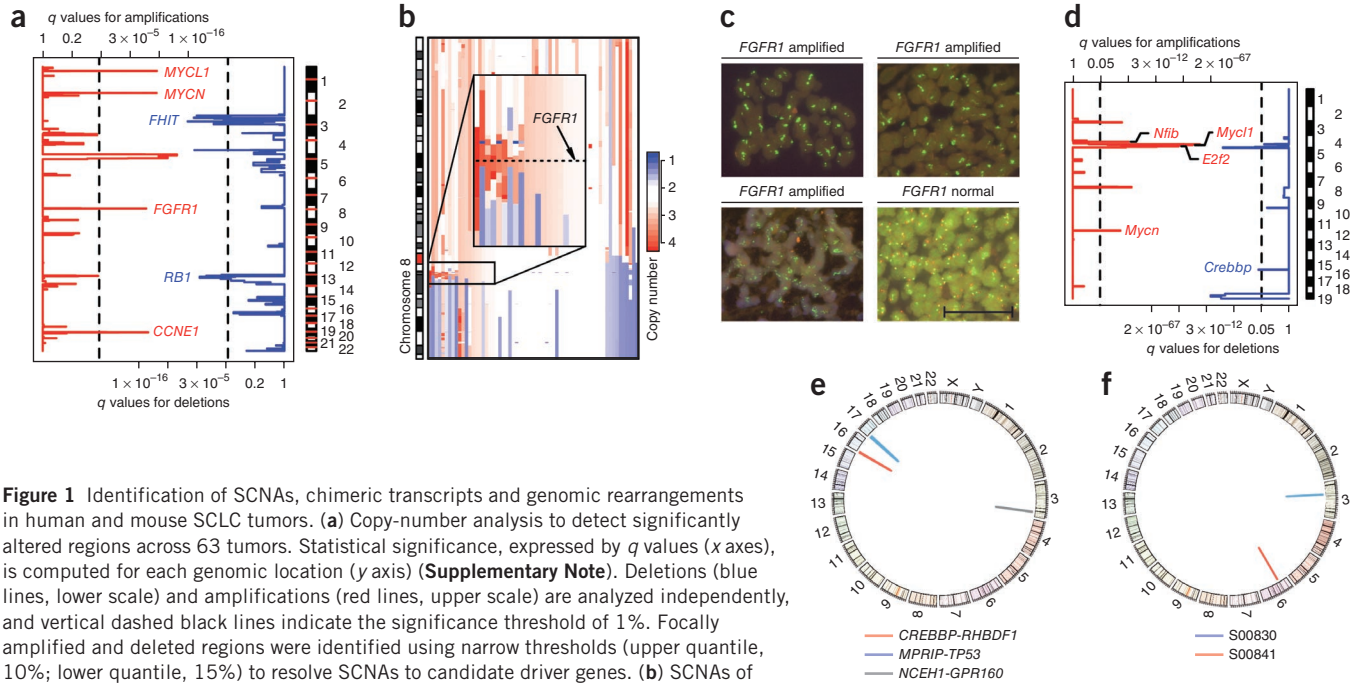

Figure 1 Identification of SCNAs, chimeric transcripts and genomic rearrangements in human and mouse SCLC tumors. (**a**) Copy-number analysis to detect significantly altered regions across 63 tumors. Statistical significance, expressed by *q* values (*x* axes), is computed for each genomic location (*y* axis) (**Supplementary Note**). Deletions (blue lines, lower scale) and amplifications (red lines, upper scale) are analyzed independently, and vertical dashed black lines indicate the significance threshold of 1%. Focally amplified and deleted regions were identified using narrow thresholds (upper quantile, 10%; lower quantile, 15%) to resolve SCNAs to candidate driver genes. (**b**) SCNAs of chromosome 8 containing *FGFR1* (8p12). Samples are sorted according to the amplitude of *FGFR1* amplification. (**c**) FISH analysis to screen for *FGFR1* amplifications in an independent set of 51 tumors. Quantification of green signals (*FGFR1*-specific probe) in comparison to red signals (centromere 8 probe) revealed three *FGFR1*-amplified samples. Scale bar, 100 µm. (**d**) Copy-number analysis based on array–comparative genomic hybridization (aCGH) data for 20 SCLC tumors derived from *Tp53* and *Rb1* double knockout mice. Data were analyzed as in **a**. Due to the small sample size, a significance threshold of 5% was used (vertical dashed lines). (**e**) Circos plot of all validated chimeric transcripts detected by transcriptome sequencing. (**f**) Circos plot of validated genomic rearrangements obtained from whole-genome sequencing. Both rearrangements affect only portions of the genome smaller than 500 kb. Whereas the structural variant in sample S00841 affects non-coding DNA, the rearrangement in S00830 leads to loss of exons 7–11 of the *FOXP1* gene.

than those in squamous cell lung cancer (**Supplementary Fig. 1b**). Focal amplifications affected *MYCL1* (5/63 cases) and *MYCN* (4/63 cases)[21,22] (**Fig. 1a**). A single case harbored a focal amplification of *MYC*. All *MYC* family member amplifications (16% of cases) were mutually exclusive, suggesting genetic epistasis[21–23]. Focal amplifications affected 8p12, including *FGFR1* (6% of cases with copy number of ≥3.5; **Fig. 1b**), and 19q12, containing *CCNE1* (ref. 24). FISH analyses in 51 independent specimens validated the occurrence of *FGFR1* amplifications in SCLC ($n = 3$, 6%; **Fig. 1c**). We and others have recently reported focal *FGFR1* amplifications in squamous cell lung cancer; FGF receptor inhibitors are currently being tested in such patients[17,19,25]. Thus, *FGFR1*-amplified SCLC might benefit from FGFR inhibition. The only significant focal deletion involved *FHIT*[26] (**Fig. 1a** and **Supplementary Table 2**).

Mice with conditional deletion of *Rb1* and *Tp53* develop SCLC[4,27–31] bearing amplifications of *Mycl1*, *Mycn* and *Nfib*, which were subsequently also found in human SCLC[28]. We analyzed SCNAs in 20 SCLC tumors (15 primary tumors and 5 metastases) from *Tp53* and *Rb1* conditional double knockout mice[4] to identify alterations shared by both human and mouse tumors. We found significant amplifications of *Mycl1*, *Mycn* and *Nfib* (**Fig. 1d**). In the 15 primary tumors (**Supplementary Fig. 2**), *Nfib* did not reach statistical significance, suggesting that *Nfib* amplifications occur later in tumor evolution. Although *NFIB* was not significantly amplified in the human tumors, three samples had copy-number gain at this locus (data not shown). Furthermore, we identified significant amplifications affecting *E2f2*, a mediator of RB1 function[32], and deletions of the histone acetyl transferase gene *Crebbp* in two mouse tumors (**Fig. 1d**).

By analyzing the transcriptome sequencing data of 15 human tumors, we next identified and validated 3 chimeric transcripts (**Fig. 1e** and **Supplementary Table 3**). Two contained a fusion partner that was also mutated, *MPRIP-TP53* and *CREBBP-RHBDF1* (**Fig. 1e**), both of which are predicted to cause loss of function of the genes involved (**Supplementary Fig. 3a**,**b**). Similarly, we also found a low genomic rearrangement frequency by reconstruction from paired-end whole-genome sequencing data of two specimens (**Fig. 1f**). This low frequency is in accordance with the spectrum of SCNAs in these samples that show almost exclusively chromosome arm–level events (**Supplementary Fig. 4a**).

To identify possible differences in the overall genomic architecture between surgically resected (early stage) samples ($n = 17$) and samples obtained by autopsy (late stage, $n = 10$), we compared the spectrum of broad SCNAs in these two sets. We computed absolute copy numbers from sequencing data to correct for admixture of non-tumor cells and for ploidy (**Supplementary Fig. 4b** and **Supplementary Note**) but found no significant difference between resected and autopsy cases (**Fig. 2a**). Furthermore, there was no difference in the total mutation frequency (**Fig. 2b**) and no segregation between resected and autopsy cases in an analysis of mutated 'driver' genes (**Fig. 2c**,**d**). We further identified 5 triploid and 2 near-tetraploid cases ($n = 29$) and found no statistically significant over-representation of samples with ploidy of >2 between resected and autopsy cases ($P = 0.15$). On average, we observed a ploidy of 2.3, which is in line with previously reported studies based on DNA cytometry[5]. Thus, resected early-stage tumors and late-stage tumors are genomically similar, underscoring the representative nature of our analysis.

Compared to other tumor types in global sequencing studies[33–41], SCLC exhibits an extremely high mutation rate of 7.4 protein-changing

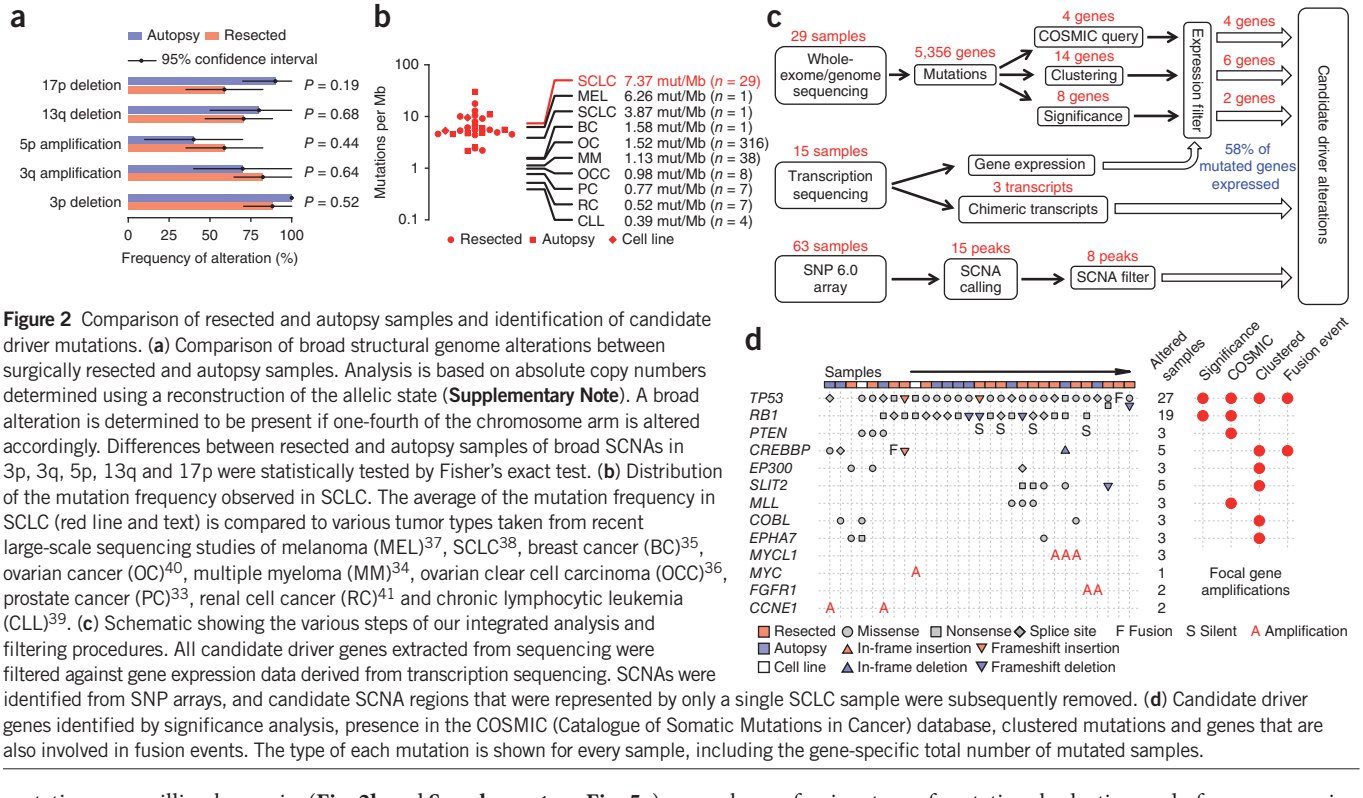

Figure 2 Comparison of resected and autopsy samples and identification of candidate driver mutations. (**a**) Comparison of broad structural genome alterations between surgically resected and autopsy samples. Analysis is based on absolute copy numbers determined using a reconstruction of the allelic state (**Supplementary Note**). A broad alteration is determined to be present if one-fourth of the chromosome arm is altered accordingly. Differences between resected and autopsy samples of broad SCNAs in 3p, 3q, 5p, 13q and 17p were statistically tested by Fisher's exact test. (**b**) Distribution of the mutation frequency observed in SCLC. The average of the mutation frequency in SCLC (red line and text) is compared to various tumor types taken from recent large-scale sequencing studies of melanoma (MEL)[37], SCLC[38], breast cancer (BC)[35], ovarian cancer (OC)[40], multiple myeloma (MM)[34], ovarian clear cell carcinoma (OCC)[36], prostate cancer (PC)[33], renal cell cancer (RC)[41] and chronic lymphocytic leukemia (CLL)[39]. (**c**) Schematic showing the various steps of our integrated analysis and filtering procedures. All candidate driver genes extracted from sequencing were filtered against gene expression data derived from transcription sequencing. SCNAs were identified from SNP arrays, and candidate SCNA regions that were represented by only a single SCLC sample were subsequently removed. (**d**) Candidate driver genes identified by significance analysis, presence in the COSMIC (Catalogue of Somatic Mutations in Cancer) database, clustered mutations and genes that are also involved in fusion events. The type of each mutation is shown for every sample, including the gene-specific total number of mutated samples.

mutations per million basepairs (**Fig. 2b** and **Supplementary Fig. 5a**). This high mutation rate is likely linked to tobacco carcinogens, reflected by an elevated rate of C:G>A:T transversions compared to the neutral mutation rate observed in evolution (**Supplementary Fig. 5b**)[38,42–44]. To identify pathogenetically relevant driver genes in the context of frequent background mutations, we applied several filters, including analyses of a signature of mutational selection and of gene expression (**Fig. 2c** and **Supplementary Note**). In particular, significantly mutated genes showing an expression level lower than 1 FPKM (fraction per kilobase of exon per million fragments mapped) in more than half of the 15 transcriptomes were removed. Using these adjustments, only two genes had q values of ≤0.1: *TP53* and *RB1* (**Fig. 2d**)[22,29,30,45,46].

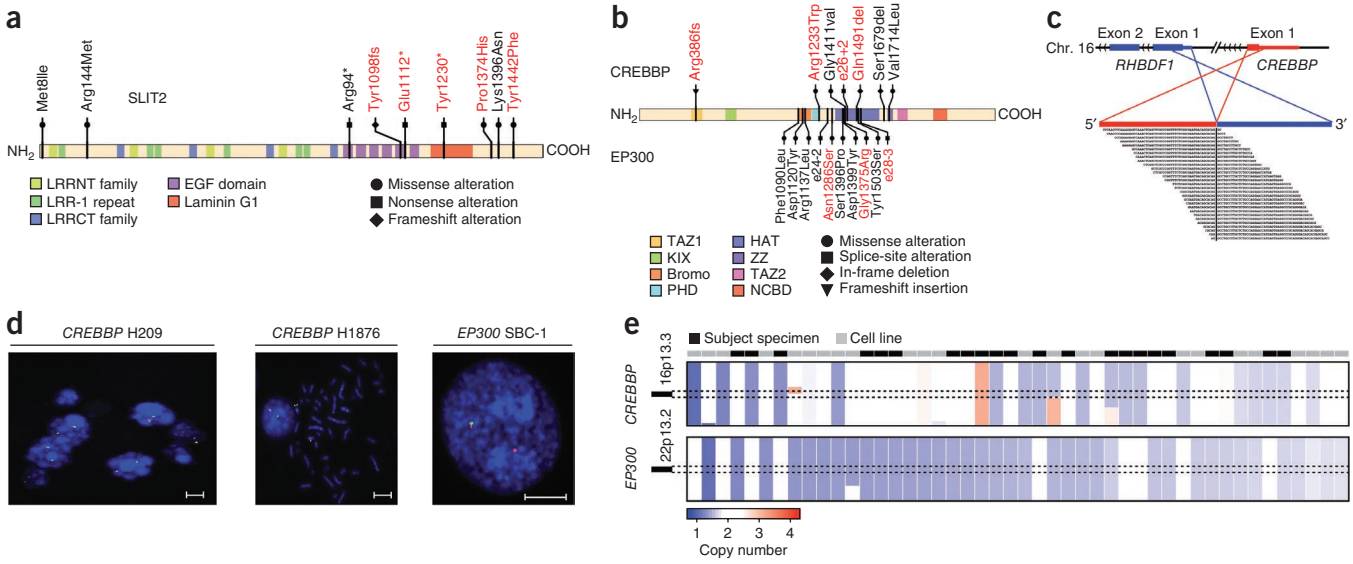

Figure 3 Recurrent mutations affecting *SLIT2*, *CREBBP* and *EP300*. (**a**) The spectrum of alterations affecting SLIT2. Alterations detected by exome sequencing are shown in red, and the results of the extended screen using 454 sequencing are shown in black. (**b**) Alterations in CREBBP and EP300. Alterations identified by whole-exome sequencing are shown in red, and the results from extended sequencing around the region encoding the HAT domain are shown in black. (**c**) The structure of the chimeric transcript affecting *RHBDF1* and *CREBBP* is shown. The genomic scale was adapted to accommodate exons from both genes (axis break, dashes). Chimeric reads are shown below. The chimeric breakpoint is indicated by the vertical black line. (**d**) Cell lines showing abnormal signals in the break-apart FISH assay of *CREBBP* and *EP300*. For *CREBBP*, both cell lines showed loss of the telomeric signal (red). For *EP300*, one cell line also showed loss of the telomeric signal (green). Break-apart FISH results for *CREBBP* in H209 are shown as a control[38]. Scale bars, 5 μm. (**e**) Copy-number status for *CREBBP* and *EP300* in all samples that showed a deletion in one of the two genes (copy numbers of ≤1.6 are considered to indicate deletion). Copy numbers are sorted with respect to the minimal copy number between *CREBBP* and *EP300*.

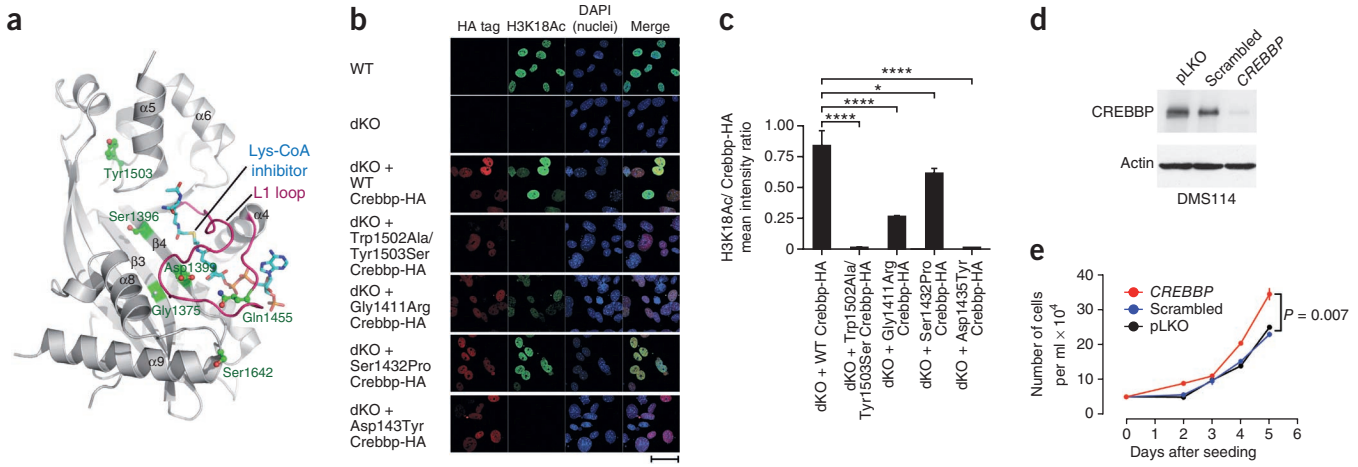

Figure 4 Functional analysis of CREBBP and EP300. (**a**) CREBBP and EP300 alterations mapped to the crystal structure of the EP300 HAT domain[56]. All alterations are located at the molecular interface involved in Lys-CoA inhibitor binding. In particular, Asp1399 and Gln1455 (equivalent to CREBBP Asp1435 and Gln1491) are located on the substrate-binding loop L1 (magenta). (**b**) Immunofluorescence was applied to measure the amount of acetylated lysine 18 on histone H3 (H3K18Ac) in wild-type (WT) MEFs, *Crebbp* and *Ep300* double knockout MEFs (dKO) and double knockout MEFs transduced with retroviruses expressing wild-type or SCLC-derived mutants of mouse *Crebbp*. Human mutations were made to alter the equivalent mouse amino acid, but human numbering is shown. Crebbp-HA, red (CY3); H3K18Ac, green (Alexa Fluor 488); nuclei, blue (DAPI). The functionally defective Trp1502Ala/Tyr1503Ser alteration[81] was included as a control. Scale bar, 50 μm. (**c**) Quantification of H3K18Ac mean signal intensity per nucleus relative to the HA-tagged Crebbp mean signal intensity. Error bars, 1 s.d. of the mean. *P* values shown are from the Bonferroni *post-hoc* test of one-way ANOVA. $*P < 0.05$; $****P < 0.0001$. (**d**) Whole-cell lysates of DMS114 cells stably infected with lentiviruses expressing short hairpin RNAs (shRNAs) targeting *CREBBP* were analyzed for CREBBP protein levels by immunoblotting. (**e**) DMS114 cells stably infected with lentiviral shRNAs targeting *CREBBP* or the indicated control cells were seeded in 6-well plates and counted as triplicates at the indicated time points (*x* axis). Absolute numbers are given on the *y* axis; error bars show 1 s.d. of the mean.

Notably, many of the significantly mutated genes were actually not expressed (**Supplementary Table 4**), and none of these mutations were called in the transcriptomes. By contrast, all known tumor suppressors showed expression in the upper part of the overall distribution (**Supplementary Fig. 6**), supporting the use of our strategy for the elimination of 'passenger' mutations. Additional filters included an analysis of regional clustering of mutations in a given gene (defining a mutational hotspot) and integration with orthogonal data sets and databases (**Fig. 2c**)[47]. As in the analysis of significantly mutated genes, we discarded genes that were enriched for silent mutations. Together, these filters yielded a list of likely driver genes in SCLC: *TP53*, *RB1*, *PTEN*, *CREBBP*, *EP300*, *SLIT2*, *MLL*, *COBL* and *EPHA7* (**Fig. 2d**).

SLIT2 showed a pronounced clustering of mutations (5/29 cases). The observed mutation spectrum (two nonsense, one frameshift deletion and two missense; **Fig. 3a**) together with frequent genomic losses (**Supplementary Fig. 7a**) suggests that *SLIT2* may be a novel tumor suppressor gene in SCLC. We sequenced *SLIT2* in 26 additional tumors and 34 cell lines and found an overall mutation frequency of 10% ($n = 89$). Slit proteins are secreted ligands for Robo receptors, which are involved in axon guidance and cellular migration[48,49]. Supporting the notion of a tumor suppressive function of SLIT2-ROBO1 in the lung, *Robo1*-knockout mice do not develop normal lungs; surviving mice exhibit bronchial hyperplasia[50]. Accordingly, a tumor suppressive role for SLIT2 has recently been implied in lung cancer cell lines[51]. Furthermore, ROBO1 was recently found to be a specific serum biomarker of SCLC[52]. *EPHA7* was recently described as a tumor suppressor gene that is frequently lost in lymphomas[53]. Given the role of EPHA7 in embryonic development and neural tube closure[54], mutations in this gene may contribute to the invasive phenotype of SCLC.

Mutations in *CREBBP* and *EP300* were significantly clustered around the sequence encoding the histone acetyltransferase (HAT) domain (**Fig. 3b**). Of these mutations, those affecting the homologous Asp1399 (EP300) and Asp1435 (CREBBP) residues both affect acetylase activity *in vitro*[55–57]. Furthermore, the p.Gly1411Glu alteration in CREBBP has previously been identified in lung cancer[58] and follicular lymphoma[59], and p.Gly1411Val as well as p.Asp1435Gly were found in relapsed acute lymphoblastic leukemia[60], suggesting a mutational hotspot. By contrast, the p.Arg386fs alteration and the *CREBBP-RHBDF1* gene fusion truncate the protein at the N terminus (**Fig. 3c** and **Supplementary Fig. 3a**). Together with the observation of *Crebbp* deletions in mouse SCLC (**Fig. 1d**) and the recently described *CREBBP-BTBD12* gene fusion in the NCI-H209 SCLC cell line[38], inactivation of *CREBBP* and *EP300* likely has a major role in SCLC. Focused sequencing of the HAT domain–encoding exons of *CREBBP* and *EP300* in a validation set of 26 additional SCLC tumor specimens and 45 cell lines, as well as break-apart FISH performed in 34 SCLC cell lines, confirmed an overall mutation frequency of 18% (point mutations, insertions and/or deletions (indels) and gene rearrangements) (**Fig. 3b–d**). *CREBBP* and *EP300* mutations have recently been described in relapsed acute lymphoblastic leukemia and B-cell lymphoma[57,61] but have not been observed at such high frequency in solid tumors to date. Furthermore, all mutations and most of the deletions in *CREBBP* and *EP300* occurred in a mutually exclusive fashion in the total set of 101 samples analyzed, suggesting epistasis (**Fig. 3e**). The observed alterations are predominantly heterozygous, supporting haploinsufficiency[57,62]. Thus, even hemizygous deletions occurring in at least 10% of non-mutant samples (**Fig. 3e** and **Supplementary Fig. 7b**) may be considered inactivating.

Further supporting the relevance of *CREBBP* and *EP300* mutations in SCLC, all but one (the mutation encoding the p.Asn1286Ser alteration in *EP300*) of the missense mutations were classified as being damaging by computational analyses[63]. Furthermore, all HAT domain alterations were located at the interface of substrate binding[56] (**Fig. 4a**), thus supporting the notion that they may affect catalytic activity. We assessed the functional impact on histone acetylation of

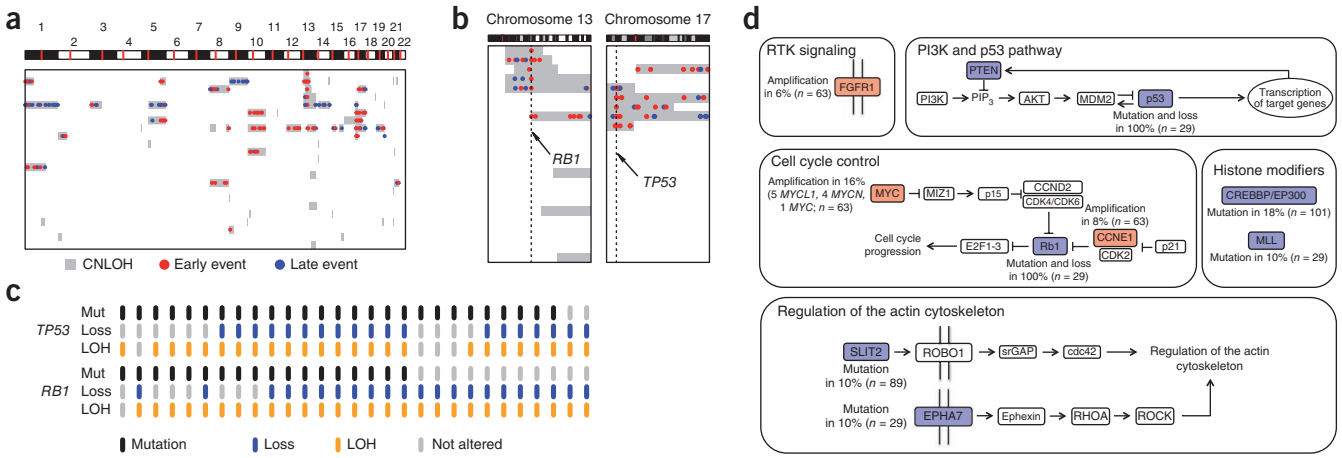

Figure 5 Mutation spectra of *TP53* and *RB1* and genetic pathways altered in SCLC. (**a**) Analysis of CNLOH events in SCLC. The allelic state of each exome-sequenced sample was reconstructed by applying a detailed mathematical model (**Supplementary Note**). Genomic portions that showed loss of heterozygosity (LOH) and an absolute copy number equal to the overall ploidy of the sample are classified as CNLOH events. Only samples showing at least one CNLOH event are shown. An analysis of the allelic fraction of somatic mutations in CNLOH regions yields information about the timing of these mutational events. (**b**) *TP53* and *RB1* mutations in CNLOH regions. Symbols are defined as in **a**. (**c**) Distribution of mutations (including rearrangements), hemizygous deletions and LOH affecting *TP53* and *RB1* across all exome-sequenced samples. (**d**) SCLC driver genes and their mutation frequency mapped to signaling pathways. We classified the occurring mutations into five major groups: receptor tyrosine kinase (RTK) signaling, PI3K and p53 pathway, cell cycle control, histone modification and regulation of actin cytoskeleton.

the p.Gly1411Arg, p.Asp1435Tyr and p.Ser1432Pro CREBBP alterations (homologous to p.Gly1375Arg, p.Asp1399Tyr and p.Ser1396Pro alterations in EP300) in reconstitution experiments in *Crebbp*$^{fl/fl}$; *Ep300*$^{fl/fl}$ (Crebbp and Ep300 Cre-deleted double knockout) mouse embryonic fibroblasts (MEFs)[64–66]. All three mutations significantly reduced acetylation of histone 3 lysine 18 (H3K18) (**Fig. 4b**,**c**). Specifically, p.Asp1435Tyr induced complete, p.Gly1411Arg pronounced and p.Ser1432Tyr moderate loss of H3K18 acetylation. Furthermore, knockdown of *CREBBP* in the DMS114 SCLC cell line that lacks CREBBP HAT domain alterations resulted in a moderate but significant increase in cell proliferation (**Fig. 4d**,**e**). Tumors with mutations and hemizygous deletions in *CREBBP* and *EP300* did not exhibit a significantly different pattern of gene expression compared to wild-type tumors after correcting for multiple hypothesis testing (data not shown), suggesting that global changes in gene expression are not the predominant mechanism by which loss of HAT activity contributes to SCLC pathogenesis. Taken together, these results support a role for loss of CREBBP and EP300 function in the biology of SCLC.

Another histone-modifying enzyme mutated in SCLC was the methyltransferase gene *MLL*, which was recurrently mutated to alter Ile960 (p.Ile960Met)[47]. *MLL* rearrangements occur in acute leukemia[67,68]. Similarly, recurrent genetic alterations in histone modifying genes seem to be a newly identified hallmark of SCLC.

Confirming previous reports[69], we found mutations in *PTEN* (3/29 cases), all of which are likely (p.Gly165Glu) or proven (p.His61Arg and p.Arg130Gly) to affect phosphatase activity[70], thereby activating the phosphatidylinositol 3-kinase (PI3K) pathway. We did not observe any mutations in *PIK3CA*[71].

We developed a mathematical model that gives insight into the allelic state of each tumor and yields estimates of tumor heterogeneity (**Supplementary Note**). On average, we observed a rather low heterogeneity of approximately 6.5% (**Supplementary Table 5**). Using the reconstructed allelic states of each tumor, we found that copy-neutral loss of heterozygosity (CNLOH) events (complete loss of one allele at a given locus combined with a match of the absolute copy number at that locus with the overall ploidy of the sample) were enriched at the *TP53* and *RB1* loci (**Fig. 5a**,**b**). Furthermore, all *TP53* and *RB1* mutations in CNLOH regions were early events (**Fig. 5b**), as their allelic fractions were compatible with the tumor purity. By integrating the different data sets, we found that at least one allele of *TP53* and *RB1* was affected by any genomic event (mutation (including rearrangement) or hemizygous deletion (LOH)) in all cases (**Fig. 5c**). Thus, similar to genetically manipulated mouse models of SCLC, inactivation of *TP53* and *RB1* are early and necessary events in the development of SCLC in humans as well[4,27–31]. Finally, we identified one case, in which the affected individual had undergone surgery for lung adenocarcinoma 3 years before diagnosis with SCLC. Whereas both tumors contained the same TP53 alteration (p.Val73fs), the RB1 alteration (p.Arg251*) was restricted to the SCLC tumor (**Supplementary Fig. 8**), compatible with trans-differentiation of adenocarcinoma cells to SCLC cells, mediated in part through loss of *RB1*. Acquired resistance of EGFR-mutant lung adenocarcinomas to EGFR inhibition has been linked with trans-differentiation to SCLC[72,73]. It is tempting to speculate that loss of *RB1* may be mechanistically involved in such cases of acquired resistance as well.

Despite methodological challenges (limited sample set and high mutation frequency), integrative genome analyses of human and mouse SCLCs afforded a molecular map of this tumor type, condensed in five categories (**Fig. 5d**). The tumor-suppressive functions of p53 rely on its acetylation by CREBBP or EP300 (refs. 74–79). However, given the universal loss of p53 function in SCLC, the tumor suppressive functions of CREBBP that we observed are likely independent of p53. One of the best-studied functions of SLIT2 is its involvement in actin polymerization mediated by Cdc42 (ref. 80). We speculate that this property might enhance invasive capabilities and thus contribute to the aggressiveness of SCLC. The reported functions of EPHA7 (refs. 53,54) may also contribute to this phenotype. Beyond universal losses of *TP53* and *RB1* and amplifications of *MYCL1*, *MYCN* and *MYC*, we present *PTEN* mutations and *FGFR1* amplifications as potentially therapeutically tractable genome alterations. Finally, we define genomic alterations that affect the histone-modifying enzymes *CREBBP*, *EP300* and *MLL* as the second most frequently mutated class of genes in SCLC. In summary, our study represents a

considerable extension of the current molecular concept of SCLC and, more broadly, provides an example of how integrative computational genome analyses can provide functionally tractable information in the context of a highly mutated cancer genome.

URLs. ATCC, http://www.atcc.org/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Binary sequence alignment data of 300-bp regions around all identified somatic mutations, segmented human SNP array data and segmented mouse aCGH data can be downloaded from http://www.uni-koeln.de/sclc/SCLC_Data.tgz.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are indebted to the individuals donating their tumor specimens as part of the Clinical Lung Cancer Genome Project initiative. Additional biospecimens for this study were obtained from the Victorian Cancer Biobank (Melbourne, Australia). The Institutional Review Board (IRB) of each participating institution approved collection and use of all specimens in this study. We also thank our colleagues at The Cancer Genome Atlas Research Network (TCGA) and A.L. Kung for invaluable discussion and many helpful comments. This work was supported by the German Ministry of Science and Education (BMBF) as part of the NGFNplus program (grant 01GS08100 to R.K.T. and 01GS08101 to J.W. and P.N.), by the Max Planck Society (M.I.F.A.NEUR8061 to R.K.T.), by the Deutsche Forschungsgemeinschaft (DFG) through SFB832 (TP6 to R.K.T. and TP5 and Z1 to L.C.H. and R.B.) and TH1386/3-1 (to R.K.T. and M.L.S.), by the European Union's Framework Programme CURELUNG (HEALTH-F2-2010-258677 to R.K.T., J.F., E.B., C. Brambilla, S.L., B.B. and J.W.), by Stand Up To Cancer–American Association for Cancer Research Innovative Research Grant (SU2C-AACR-IR60109 to R.K.T. and W.P.), by the Behrens-Weise Foundation (to R.K.T.) and by an anonymous foundation to R.K.T. M.L.S. is a fellow of the International Association for the Study of Lung Cancer (IASLC). P.K.B. and L.H.K. thank the St. Jude Cell and Tissue Imaging facility and acknowledge support from the US National Institutes of Health (NIH) Cancer Center (grant P30 CA021765) and the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital. F.C. was supported by Associazione Italiana Ricerca sul Cancro (AIRC, grant IG 9425).

AUTHOR CONTRIBUTIONS

M.P., L.F.-C., M.L.S., J.G., D.S., L.H.K., F.L.,T.Z., R.M., J.V., P.S., J. Sage, R. Schneider, R.B., S.P., L.C.H., P.K.B. and R.K.T. conceived and designed the experiments. L.F.-C., M.L.S., J.G., D.S., L.H.K., D.P., R.M., M.K., I.D., C.M., V.D.C., H.-U.S., J.A., I.B., C. Becker, B.d.W., D.B., F.G., I.W., S. Heynck, J.M.H. and P.M.S. performed experiments. M.P., L.F.-C., M.L.S., J.G., D.S., L.H.K., D.P., F.L., R. Sun, T.Z., R.M., V.D.C., B.d.W., J.V., X.L., W.P., M.M., J. Sage, R. Schneider, S.P., L.C.H., P.K.B., S. Haas and R.K.T. analyzed the data. M.P., R. Sun, S.A., S.L.C., K.C., S.B., G.G., K.-S.P., D.R., C.G., M.F., L.P., G.W., Z.W., P.R., I.P., Y.C., E. Stoelben, C. Ludwig, P.S., H.H., T.M., M.B., W.E.-R., L.A.M., V.M.F., H.G., W.T., H.S., E.T., E. Smit, D.A.M.H., P.J.F.S., F.C., C. Ligorio, S.D., J.F., S.S., O.T.B., M.L.-I., J. Sänger, J.H.C., A.S., H.M., W.W., B.S., J.-C.S., P.V., B.B., E.B., C. Brambilla, S.L., P.L., M.H., J. Sage, J. Shendure, R. Schneider, R.B., S.P., L.C.H., J.W., P.N., L.C.H., P.K.B. and S. Haas contributed reagents, materials or analysis tools. M.P., L.F.-C. and R.K.T. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Published online at http://www.nature.com/doifinder/10.1038/ng.2396.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Gustafsson, B.I., Kidd, M., Chan, A., Malfertheiner, M.V. & Modlin, I.M. Bronchopulmonary neuroendocrine tumors. *Cancer* **113**, 5–21 (2008).
2. Travis, W.D. Lung tumours with neuroendocrine differentiation. *Eur. J. Cancer* **45** (suppl. 1), 251–266 (2009).
3. van Meerbeeck, J.P., Fennell, D.A. & De Ruysscher, D.K. Small-cell lung cancer. *Lancet* **378**, 1741–1755 (2011).
4. Park, K.S. *et al.* A crucial requirement for Hedgehog signaling in small cell lung cancer. *Nat. Med.* **17**, 1504–1508 (2011).
5. World Health Organization. *Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart* (eds. Travis, W.D., Brambilla, E., Müller-Hermilink, H.K. & Harris, C.C.) 344 (IARC Press, Lyon, France, 2004).
6. Tiseo, M. & Ardizzoni, A. Current status of second-line treatment and novel therapies for small cell lung cancer. *J. Thorac. Oncol.* **2**, 764–772 (2007).
7. Kwak, E.L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
8. Lynch, T.J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
9. Paez, J.G. *et al. EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
10. Pao, W. & Chmielecki, J. Rational, biologically based treatment of *EGFR*-mutant non-small-cell lung cancer. *Nat. Rev. Cancer* **10**, 760–774 (2010).
11. Pao, W. *et al.* EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA* **101**, 13306–13311 (2004).
12. Soda, M. *et al.* Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
13. Bergethon, K. *et al. ROS1* rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **30**, 863–870 (2012).
14. Kohno, T. *et al. KIF5B-RET* fusions in lung adenocarcinoma. *Nat. Med.* **18**, 375–377 (2012).
15. Lipson, D. *et al.* Identification of new *ALK* and *RET* gene fusions from colorectal and lung cancer biopsies. *Nat. Med.* **18**, 382–384 (2012).
16. Takeuchi, K. *et al. RET*, *ROS1* and *ALK* fusions in lung cancer. *Nat. Med.* **18**, 378–381 (2012).
17. Dutt, A. *et al.* Inhibitor-sensitive *FGFR1* amplification in human non-small cell lung cancer. *PLoS One* **6**, e20351 (2011).
18. Hammerman, P.S. *et al.* Mutations in the *DDR2* kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov.* **1**, 78–89 (2011).
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. Bass, A.J. *et al. SOX2* is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat. Genet.* **41**, 1238–1242 (2009).
21. Kim, Y.H. *et al.* Combined microarray analysis of small cell lung cancer reveals altered apoptotic balance and distinct expression signatures of *MYC* family gene amplification. *Oncogene* **25**, 130–138 (2006).
22. Wistuba, I.I., Gazdar, A.F. & Minna, J.D. Molecular genetics of small cell lung carcinoma. *Semin. Oncol.* **28**, 3–13 (2001).
23. Gazzeri, S. *et al.* Activation of *myc* gene family in human lung carcinomas and during heterotransplantation into nude mice. *Cancer Res.* **51**, 2566–2571 (1991).
24. Zhao, X. *et al.* Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res.* **65**, 5561–5570 (2005).
25. Voortman, J. *et al.* Array comparative genomic hybridization–based characterization of genetic alterations in pulmonary neuroendocrine tumors. *Proc. Natl. Acad. Sci. USA* **107**, 13040–13045 (2010).
26. Hassan, M.I., Naiyer, A. & Ahmad, F. Fragile histidine triad protein: structure, function, and its association with tumorogenesis. *J. Cancer Res. Clin. Oncol.* **136**, 333–350 (2010).
27. Calbo, J. *et al.* A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* **19**, 244–256 (2011).
28. Dooley, A.L. *et al.* Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev.* **25**, 1470–1475 (2011).
29. Meuwissen, R. *et al.* Induction of small cell lung cancer by somatic inactivation of both *Trp53* and *Rb1* in a conditional mouse model. *Cancer Cell* **4**, 181–189 (2003).
30. Schaffer, B.E. *et al.* Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer Res.* **70**, 3877–3883 (2010).
31. 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).
32. Iaquinta, P.J. & Lees, J.A. Life and death decisions by the E2F transcription factors. *Curr. Opin. Cell Biol.* **19**, 649–657 (2007).
33. Berger, M.F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
34. Chapman, M.A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
35. Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
36. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma. *Science* **330**, 228–231 (2010).
37. Pleasance, E.D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
38. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
39. Puente, X.S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).

40. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
41. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
42. Karro, J.E., Peifer, M., Hardison, R.C., Kollmann, M. & von Grunberg, H.H. Exponential decay of GC content detected by strand-symmetric substitution rates influences the evolution of isochore structure. *Mol. Biol. Evol.* **25**, 362–374 (2008).
43. Hecht, S.S. Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol.* **21**, 160–171 (2008).
44. Rodin, S.N. & Rodin, A.S. Origins and selection of p53 mutations in lung carcinogenesis. *Semin. Cancer Biol.* **15**, 103–112 (2005).
45. Horowitz, J.M. *et al.* Frequent inactivation of the retinoblastoma anti-oncogene is restricted to a subset of human tumor cells. *Proc. Natl. Acad. Sci. USA* **87**, 2775–2779 (1990).
46. Mori, N. *et al.* Variable mutations of the RB gene in small-cell lung carcinoma. *Oncogene* **5**, 1713–1717 (1990).
47. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
48. Wong, K., Park, H.T., Wu, J.Y. & Rao, Y. Slit proteins: molecular guidance cues for cells ranging from neurons to leukocytes. *Curr. Opin. Genet. Dev.* **12**, 583–591 (2002).
49. Zhou, W.J. *et al.* Slit-Robo signaling induces malignant transformation through Hakai-mediated E-cadherin degradation during colorectal epithelial cell carcinogenesis. *Cell Res.* **21**, 609–626 (2011).
50. Xian, J. *et al.* Inadequate lung development and bronchial hyperplasia in mice with a targeted deletion in the *Dutt1/Robo1* gene. *Proc. Natl. Acad. Sci. USA* **98**, 15062–15066 (2001).
51. Tseng, R.C. *et al.* SLIT2 attenuation during lung cancer progression deregulates β-catenin and E-cadherin and associates with poor prognosis. *Cancer Res.* **70**, 543–551 (2010).
52. Taguchi, A. *et al.* Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell* **20**, 289–299 (2011).
53. Oricchio, E. *et al.* The Eph-receptor A7 is a soluble tumor suppressor for follicular lymphoma. *Cell* **147**, 554–564 (2011).
54. Holmberg, J., Clarke, D.L. & Frisen, J. Regulation of repulsion versus adhesion by different splice forms of an Eph receptor. *Nature* **408**, 203–206 (2000).
55. Muraoka, M. *et al.* p300 gene alterations in colorectal and gastric carcinomas. *Oncogene* **12**, 1565–1569 (1996).
56. Liu, X. *et al.* The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature* **451**, 846–850 (2008).
57. Pasqualucci, L. *et al.* Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature* **471**, 189–195 (2011).
58. Kishimoto, M. *et al.* Mutations and deletions of the *CBP* gene in human lung cancer. *Clin. Cancer Res.* **11**, 512–519 (2005).
59. Morin, R.D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).
60. Inthal, A. *et al.* CREBBP HAT domain mutations prevail in relapse cases of high hyperdiploid childhood acute lymphoblastic leukemia. *Leukemia* **26**, 1797–1803 (2012).
61. Mullighan, C.G. *et al.* *CREBBP* mutations in relapsed acute lymphoblastic leukaemia. *Nature* **471**, 235–239 (2011).
62. Tillinghast, G.W. *et al.* Analysis of genetic stability at the *EP300* and *CREBBP* loci in a panel of cancer cell lines. *Genes Chromosom. Cancer* **37**, 121–131 (2003).
63. Ng, P.C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
64. Kang-Decker, N. *et al.* Loss of CBP causes T cell lymphomagenesis in synergy with $p27^{Kip1}$ insufficiency. *Cancer Cell* **5**, 177–189 (2004).
65. Kasper, L.H. *et al.* Conditional knockout mice reveal distinct functions for the global transcriptional coactivators CBP and p300 in T-cell development. *Mol. Cell Biol.* **26**, 789–809 (2006).
66. Kasper, L.H. *et al.* CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation. *EMBO J.* **29**, 3660–3672 (2010).
67. Thirman, M.J. *et al.* Rearrangement of the *MLL* gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *N. Engl. J. Med.* **329**, 909–914 (1993).
68. Yang, X.J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **32**, 959–976 (2004).
69. Yokomizo, A. *et al.* *PTEN/MMAC1* mutations identified in small cell, but not in non-small cell lung cancers. *Oncogene* **17**, 475–479 (1998).
70. Han, S.Y. *et al.* Functional evaluation of *PTEN* missense mutations using *in vitro* phosphoinositide phosphatase assay. *Cancer Res.* **60**, 3147–3151 (2000).
71. Yamamoto, H. *et al.* *PIK3CA* mutations and copy number gains in human lung cancers. *Cancer Res.* **68**, 6913–6921 (2008).
72. Zakowski, M.F., Ladanyi, M. & Kris, M.G. *EGFR* mutations in small-cell lung cancers in patients who have never smoked. *N. Engl. J. Med.* **355**, 213–215 (2006).
73. Sequist, L.V. *et al.* Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).
74. Gu, W., Shi, X.L. & Roeder, R.G. Synergistic activation of transcription by CBP and p53. *Nature* **387**, 819–823 (1997).
75. Brooks, C.L. & Gu, W. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr. Opin. Cell Biol.* **15**, 164–171 (2003).
76. Sakaguchi, K. *et al.* DNA damage activates p53 through a phosphorylation-acetylation cascade. *Genes Dev.* **12**, 2831–2841 (1998).
77. Kruse, J.P. & Gu, W. Modes of p53 regulation. *Cell* **137**, 609–622 (2009).
78. Grossman, S.R. *et al.* Polyubiquitination of p53 by a ubiquitin ligase activity of p300. *Science* **300**, 342–344 (2003).
79. Lill, N.L., Grossman, S.R., Ginsberg, D., DeCaprio, J. & Livingston, D.M. Binding and modulation of p53 by p300/CBP coactivators. *Nature* **387**, 823–827 (1997).
80. Wong, K. *et al.* Signal transduction in neuronal migration: roles of GTPase activating proteins and the small GTPase Cdc42 in the Slit-Robo pathway. *Cell* **107**, 209–221 (2001).
81. Bordoli, L. *et al.* Functional analysis of the p300 acetyltransferase domain: the PHD finger of p300 but not of CBP is dispensable for enzymatic activity. *Nucleic Acids Res.* **29**, 4462–4471 (2001).

[1]Department of Translational Genomics, University of Cologne, Cologne, Germany. [2]Max Planck Institute for Neurological Research with Klaus-Joachim-Zülch Laboratories of the Max Planck Society and the Medical Faculty of the University of Cologne, Cologne, Germany. [3]Department I of Internal Medicine, University of Cologne, Cologne, Germany. [4]Center of Integrated Oncology Köln–Bonn, University of Cologne, Cologne, Germany. [5]Laboratory of Translational Cancer Genomics, Center of Integrated Oncology Kölm–Bonn, University of Cologne, Cologne, Germany. [6]Department of Biochemistry, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [7]Max Planck Institute for Molecular Genetics, Berlin, Germany. [8]Institute of Prostate Cancer Research at the Institute of Pathology, University Hospital of Bonn, Bonn, Germany. [9]Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. [10]Department of Pathology, University of Cologne, Cologne, Germany. [11]Cologne Center for Genomics, University of Cologne, Cologne, Germany. [12]Center for Medical Genetics, Ghent University, Ghent, Belgium. [13]The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [14]Department of Pediatrics, Stanford University, Stanford, California, USA. [15]Department of Genetics, Stanford University, Stanford, California, USA. [16]Department of Chemical Biology, Technical University Dortmund, Dortmund, Germany. [17]Department of Pediatric Oncology and Hematology, University Children's Hospital of Cologne, Cologne, Germany. [18]Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany. [19]Institute for Cancer Genetics, Columbia University, New York, New York, USA. [20]Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York, USA. [21]Department of Surgery, St. Vincent's Hospital, The University of Melbourne, Melbourne, Victoria, Australia. [22]Department of Pathology, St. Vincent's Hospital, Melbourne, Victoria, Australia. [23]Institute of Pathology, Jena University Hospital, Friedrich-Schiller-University, Jena, Germany. [24]Department of Thoracic Surgery, Lungenklinik Merheim, Kliniken der Stadt Köln, Cologne, Germany. [25]Institute of Pathology, Heidelberg, Germany. [26]Department of Thoracic Surgery, Thoraxklinik-Heidelberg, Heidelberg, Germany. [27]Department of Pathology, Hospital Merheim, Kliniken der Stadt Köln, Cologne, Germany. [28]Laboratory of Oncology, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy. [29]Department of Pulmonary Diseases, University Medical Centre Groningen, Groningen, The Netherlands. [30]Department of Pathology, University Medical Centre Groningen, Groningen, The Netherlands. [31]Department of Pathology, VU University Medical Center Amsterdam, Amsterdam, The Netherlands. [32]Department of Pulmonary Diseases, VU University Medical Center Amsterdam, Amsterdam, The Netherlands. [33]Department of Medical Oncology, Ospedale Civile, Livorno, Italy. [34]Department of Haematology and Oncologic Science, University Hospital Bologna, Bologna, Italy. [35]Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Centre, University of Liverpool, Liverpool, UK. [36]Department of Thoracic Surgery, Rikshospitalet, Oslo University Hospital, Oslo, Norway. [37]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway. [38]Department of Oncology, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway. [39]Department of Pathology, Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway. [40]Institute for Pathology Bad Berka, Bad Berka, Germany. [41]Department of Internal Medicine II, Jena University Hospital, Friedrich-Schiller University, Jena, Germany. [42]Institute for Surgical Pathology, University Hospital Zurich, Zurich, Switzerland. [43]Division of Thoracic Surgery, University Hospital Zurich, Zurich, Switzerland. [44]Department of Haematology and Medical Oncology, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. [45]Phase I Unit–Department of Medicine, Institute Gustave Roussy, Villejuif, France. [46]France Service d'Anatomie-Pathologie, Institut Mutualiste Montsouris, Paris, France. [47]Department of Pathology, Université Joseph Fourier, Grenoble, France. [48]Institut Albert Bonniot, Institute National de la Santé et de la Recherche Médicale (INSERM) U823, Université Joseph Fourier, Grenoble, France. [49]Institute of Legal Medicine, University of Cologne, Cologne, Germany. [50]Vanderbilt-Ingram Cancer Center, Nashville, Tennessee, USA. [51]Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. [52]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [53]Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [54]Department of Genome Sciences, University of Washington, Seattle, Washington, USA. [55]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche (UMR) 7104, INSERM U964, Université de Strasbourg, Illkirch, France. [56]Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany. [57]These authors contributed equally to this work. Correspondence should be addressed to R.K.T. (roman.thomas@uni-koeln.de).

ONLINE METHODS

Sample preparation, DNA and RNA extraction and Illumina sequencing. 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. For autopsy cases, tumors were derived within a few hours after death. 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 60% 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/µl and stored at −80 °C.

We used the SPRIworks system (Beckman Coulter) for automated library construction. For whole-exome sequencing, exome enrichment was performed using the SureSelect Human All Exon 38Mbp Kit (Agilent), following the manufacturer's protocol. Exon-enriched libraries were subsequently paired-end sequenced using mostly a read length of 2 × 95 bp on the Illumina Genome Analyzer IIx (**Supplementary Table 5**). Whole-genome sequencing was performed on the Illumina HiSeq 2000 using a read length of 2 × 100 bp for all samples.

Sections for RNA extraction were disrupted and homogenized for 2 min at 20 Hz with the Tissue Lyser (Qiagen), and RNA was extracted using the Qiagen RNeasy Mini kit. RNA quality was assessed in a Bioanalyzer, and only samples showing an RNA integrity number (RIN) of >8 were retained for transcriptome sequencing. RNA sequencing (RNA-seq) was performed on cDNA libraries prepared from PolyA$^+$ RNA extracted from tumor cells. We aimed for a library with an insert size of 250 bp, allowing us to sequence 95-bp paired-end reads without overlap. All RNA-seq libraries were sequenced on the Illumina Genome Analyzer IIx.

Processing of whole-exome and whole-genome sequencing data. Raw sequencing data were aligned to the human genome (NCBI Build 36/hg18) using MAQ[82] (version: 0.7.1) for whole-exome data and the Burrows-Wheeler Aligner (BWA; version: 0.5.9rc1)[83] for whole-genome sequencing data. To prevent miscalls that might be caused by duplicated sequencing errors, possible PCR duplicates were removed form the alignments. The quality of the sequencing data was assessed by evaluating criteria such as on-target coverage (exome), average coverage and insert size. These quality metrics are summarized in **Supplementary Table 5**.

Mutation detection. We implemented a new variant caller to identify somatic mutations from the aligned sequencing (M.P. *et al.*, unpublished data). To this end, tumor-specific characteristics, including local SCNAs, tumor purity, and total aneuploidy were incorporated into a mathematical model that controls variant calling. Details of our approach are presented in the **Supplementary Note**.

Reconstruction of rearrangements from whole-genome paired-end data. To reconstruct rearrangements from paired-end data, we first screened for read pairs that were either separated by at least 600 bp or showed incorrect orientation. For the regions encompassing this type of read pair, we next examined whether the region had aberrant reorganization in the matched normal sample. The remaining genomic locations were then annotated for repetitive elements. We discarded those locations where both pairs mapped to the same repeat type, as the sequences showed a very high degree of homology, which might lead to an elevated rate for the alignment of artifacts. The remaining candidates were finally filtered by comparing the coverage of the read pairs to the total read coverage in the region that encompassed the reads. To validate these candidates by genomic PCR, two candidate-specific primer sets encompassing the fusion points were designed: one for each of the two possible orientations of the rearrangements. All validated genomic rearrangements are shown in **Supplementary Fig. 9** and **Supplementary Table 6**.

Validation of somatic mutations and frequently mutated genes. Because of the high mutation rate, we only systematically validated by dideoxy sequencing those mutations that were detected in the candidate driver genes *TP53*, *RB1*, *PTEN*, *CREBBP*, *EP300*, *SLIT2*, *MLL*, *COBL* and *EPHA7*. For *CREBBP*, *EP300* and *SLIT2*, we extended our sequencing efforts to an independent validation cohort. To this end, we sequenced the regions around the gene sequence encoding the HAT domain (exons 18–30) for *CREBBP* and *EP300* by dideoxy sequencing. For *SLIT2*, the full-length gene was screened for mutations using 454 sequencing. Further details and results from our validation strategy are given in the **Supplementary Note**.

Analysis of significantly mutated genes and detection of mutational hotspots within a gene. The analysis of significantly mutated genes was an extension of a previously described method[84] to correct for gene expression and the accumulation of synonymous mutations. Conceptual and mathematical details are outlined in the **Supplementary Note**.

Mutational hotspots within a gene were detected by resampling positions of observed mutations. *P* values were computed by comparing the observed variance of the mutations with the distribution of the variance derived from resampled mutations. We restricted this analysis to genes with at least three somatic mutations and those that did not show enrichment of silent mutations; frameshift indels were not considered. The results of this analysis for all genes having a *P* value of ≤0.05 are shown in **Supplementary Table 7**.

Analysis of RNA-seq data. For analysis of RNA-seq data, we have developed a pipeline that affords accurate and efficient mapping and downstream analysis of transcribed genes in cancer samples (R. Sun *et al.*, unpublished data). Details of this method are presented in the **Supplementary Note**.

FISH analyses. A dual-color break-apart FISH assay was developed to assess *CREBBP* and *EP300* (chromosomes 16 and 22, respectively) rearrangements on the chromosomal level, as has been described previously[85]. For the *CREBBP* break-apart assay, we used the BAC clone RP11-962J17 for centromeric labeling with digoxigenin (green) and RP11-363A1 for telomeric labeling with biotin (red). Similarly, for the *EP300* break-apart assay, we used BAC clone RP11-928B9 for telomeric labeling with digoxigenin (green) and RP11-844C16 for centromeric labeling with biotin (red). Further information about the break-apart FISH assay is given in the **Supplementary Note**. *FGFR1* FISH analysis was carried out as described elsewhere[86].

Analysis of SNP 6.0 data. Genomic DNA was hybridized to Affymetrix SNP 6.0 arrays according to the manufacturer's instructions. Raw signal intensities were processed using the same approach as in a previous publication[19], with modifications in the normalization of SNP probes. Allele-specific copy numbers were estimated using an adaption of the PLASQ algorithm[87] to the design of the SNP 6.0 arrays. Parameters of the log linear model that account for allele-specific probe affinities and probe-specific background were calibrated by a Gauss-Newton approach. The resulting raw copy numbers were segmented using circular binary segmentation[88]. Significantly amplified and deleted regions were assessed by a novel rank sum–based method (**Supplementary Note**).

aCGH analysis of mouse tumors. We extracted tumor and normal DNA form explanted p53- and Rb1-deficient SCLC mouse tumors using the Gentra DNA extraction kit (Qiagen, Gentra Puregene). In total, 20 tumors were analyzed. Among the 20 tumors, we analyzed 15 primary tumors and 5 tumors at metastatic sites. Of the five metastases, two tumors were harvested from an independent mouse, whereas the remaining three were explanted from the same mouse from which a primary tumor had been extracted. Arrays were hybridized and analyzed as described previously[4].

Analysis of histone acetylation by indirect immunofluorescence. MEFs with conditional alleles of *Crebbp* and *Ep300* were transduced with retrovirus expressing HA-tagged Crebbp protein (either with or without alterations in the HAT domain). After retroviral transduction, the endogenous *loxP*-flanked (floxed) alleles of *Crebbp* and *Ep300* were recombined using Cre-expressing adenovirus to produce MEFs lacking endogenous *Crebbp* and *Ep300* (double knockout MEFs). Four days after deletion of endogenous *Crebbp* and *Ep300*, cells were seeded on slides. The following morning, cells were fixed (in 3%

paraformaldehyde for 10 min at room temperature), permeabilized (0.1% Triton X-100 in PBS) and blocked in 3% nonfat milk in PBS for 30 min. Cells were incubated with primary antibody against the HA tag (1:500 dilution; mouse monoclonal HA-11, Covance) or H3K18Ac (1:1,000 dilution; rabbit polyclonal ab1191, Abcam) for 3 h at room temperature, and cells were then washed and incubated with secondary antibody (1:500 dilution; donkey antibody to mouse conjugated to CY3, Jackson or 1:500 dilution; goat antibody to rabbit conjugated to Alexa Fluor 488, Invitrogen) and DAPI for 1 h. Confocal images were acquired, and individual nuclei were masked on the basis of the DAPI signal. The mean signal intensity for the nuclei was assessed using Slidebook software. The background mean signal intensities for HA (CY3) and H3K18Ac (Alexa 488) were determined for nuclei from double knockout MEFs not transduced with retrovirus. Data were expressed as the ratio of the mean H3K18Ac signal intensity for each nucleus to the mean HA signal intensity for the same nucleus. Background signal was subtracted from each mean signal intensity value before the ratio was calculated. Only nuclei with a Crebbp-HA mean signal intensity that was at least twofold above background were analyzed. Graphs and statistics were produced using Prism GraphPad software. The immunofluorescence protocol (with some modifications) was described previously[29,66].

***CREBBP* knockdown and growth analysis.** *Cell lines:* The SCLC cell line DMS114 was cultivated in RPMI medium (Sigma Aldrich) supplemented with 10% FCS (PAA Laboratories) and 1% penicillin-streptomycin (Invitrogen). The cell line was confirmed to be free of cross-contaminations based on a short tandem repeat (STR) analysis conducted at the Leibniz-Institute DSMZ. The DMS114 cell line lacks *CREBBP* mutations and deletions, as determined by cDNA sequencing and copy-number analysis, respectively.

shRNA-mediated knockdown of CREBBP in DMS114: A *CREBBP*-specific shRNA (AAATGCCAGTGACAAGCGAAACCAACAAA, OriGENE) and a scrambled shRNA control sequence (AACAAGATGAAGAGCACCAA, Sigma-Aldrich) were inserted into a pLKO.1-puro–based vector (Sigma) and cotransfected with pMD.2 and pCMVd.8.9 helper plasmids into HEK 293 TL cells using the TransIT-LT1 reagent (Mirius). Similarly, pLKO.1-puro vectors without any shRNA inserts were applied and served as an additional control. After 48 h, replication-incompetent lentiviruses were collected from the supernatant to infect DMS114 cells in the presence of 10 μg/ml polybrene (Millipore). To select for transduced cell clones, medium was replenished with growth medium containing 3 μg/ml puromycin (Sigma) 24 h after infection.

Protein blot analysis: Equal protein amounts of cellular lysates were separated on 4–12% Tris-glycine gels (Invitrogen) and subjected to protein blot analysis to detect endogenous amounts of CREBBP (A-22, SC-369, Santa Cruz Biotechnology) and actin (691001, MP Biomedicals).

Cell growth analysis: Virally transduced DMS114 cells were seeded into 6-well dishes (50,000 cells per well) and maintained for 5 d in selective growth medium. Cell growth was assessed by counting the cellular particles (Z1 Particle Counter, Beckman Coulter) in triplicate every 1–2 d.

Cell culture. Cell lines were obtained in part from the American Type Culture Collection (ATCC) or were received as a kind gift from X.X. Ninomiya (University Hospital Okayama) and were cultured as described previously[89], using either RPMI or HITES cell culture medium supplemented with 10–20% FBS. Whole-genome DNA was extracted from cell lines using the Puregene kit (Qiagen) as described previously[89].

Dideoxy sequencing. For validation sequencing, primer pairs were designed to enclose the putative mutation. For resequencing, we designed primer pairs that covered the desired amplicons. Sequencing was carried out as described previously[90], and electropherograms were analyzed by visual inspection using Mutation Surveyor 2.03 software (SoftGenetics).

Additional data are given in **Supplementary Figures 10–12**, **Supplementary Tables 8–13** and the **Supplementary Note**.

82. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
83. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
84. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
85. Scheble, V.J. *et al. ERG* rearrangement is specific to prostate cancer and does not occur in any other common tumor. *Mod. Pathol.* **23**, 1061–1067 (2010).
86. Schildhaus, H.-U. *et al.* Definition of a fluorescence *in situ* hybridization score identifies high- and low-level *FGFR1* amplification types in squamous cell lung cancer. *Mod. Pathol.* published online, doi:10.1038/modpathol.2012.102 (8 June 2012).
87. Laframboise, T., Harrington, D. & Weir, B.A. PLASQ: a generalized linear model–based procedure to determine allelic dosage in cancer cells from SNP array data. *Biostatistics* **8**, 323–336 (2007).
88. Olshen, A.B., Venkatraman, E.S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
89. Sos, M.L. *et al.* Predicting drug susceptibility of non-small cell lung cancers based on genetic lesions. *J. Clin. Invest.* **119**, 1727–1740 (2009).
90. Querings, S. *et al.* Benchmarking of mutation diagnostics in clinical lung cancer specimens. *PLoS One* **6**, e19601 (2011).

doi:10.1038/ng.2396