



Published in final edited form as:

*Sci Transl Med.* 2013 October 30; 5(209): 209ra153. doi:10.1126/scitranslmed.3006802.

## A Genomics-Based Classification of Human Lung Tumors

### The Clinical Lung Cancer Genome Project (CLCGP) and Network Genomic Medicine (NGM)

†

#### Abstract

We characterized genome alterations in 1255 clinically annotated lung tumors of all histological subgroups to identify genetically defined and clinically relevant subtypes. More than 55% of all cases had at least one oncogenic genome alteration potentially amenable to specific therapeutic intervention, including several personalized treatment approaches that are already in clinical evaluation. Marked differences in the pattern of genomic alterations existed between and within histological subtypes, thus challenging the original histomorphological diagnosis.

Immunohistochemical studies confirmed many of these reassigned subtypes. The reassignment

†Corresponding authors: roman.thomas@uni-koeln.de (R.T.); reinhard.buett-ner@uk-koeln.de (R.B.); juergen.wolf@uk-koeln.de (J.W.).

**Data and materials availability:** Segmented copy number data, gene expression data from expression arrays and RNAseq, and binary sequence alignment data of 300–base pair regions around all somatic mutations that were identified from whole-exome sequencing data are available at <http://www.uni-koeln.de/med-fak/clcgp/>. The data are also available through ArrayExpress at accession number E-MTAB-1999.

**Author contributions:** The CLCGP and NGM Initiatives contributed collectively to this study. The Biospecimens and Data Source Sites provided biospecimens or data for further analysis. Data were generated and analyzed by the biospecimen core resources and data analysis team led by D.S., T.Z., L.C.H., M.P., M.B., R.B., J.W., and R.K.T.

**Competing interests:** R.K.T. is a founder and shareholder of Blackfield AG. R.K.T. received consulting and lecture fees (Sanofi-Aventis, Merck, Roche, Lilly, Boehringer Ingelheim, AstraZeneca, Atlas-Biolabs, Daiichi-Sankyo, and Blackfield AG) as well as research support (Merck, EOS, and AstraZeneca). R.B. is a cofounder and owner of Targos Molecular Diagnostics and received honoraria for consulting and lecturing from AstraZeneca, Boehringer Ingelheim, Merck, Roche, Novartis, Lilly, Qiagen, and Pfizer. J.W. received consulting and lecture fees from Roche, Novartis, Boehringer Ingelheim, AstraZeneca, Bayer, Lilly, Merck, and Amgen and research support from Roche, Bayer, Novartis, and Boehringer Ingelheim. M.L.S. is a fellow of the International Association for the Study of Lung Cancer. M.S. received funds from Lilly, Roche, Boehringer Ingelheim, AstraZeneca, and Pfizer. P.N. is a founder, CEO, and shareholder of ATLAS Biolabs GmbH. W.P. received research funding from Enzon Xcovery, AstraZeneca, Symphogen, Clovis Oncology, and Bristol-Myers Squibb and had paid consulting relationships with Molecular MD, AstraZeneca, Bristol-Myers Squibb, Symphony Evolution, Clovis Oncology, Exelixis, and Clariant. Rights to EGFR T790M testing were licensed on behalf of W.P. and others by Memorial Sloan-Kettering Cancer Center to Molecular MD. M.P. is a founder and shareholder of Blackfield AG. J.H.C. is a member of the Advisory Board of Boehringer Ingelheim Pharma GmbH & Co KG. H.G. is a member of the Advisory Boards of Roche, Eli Lilly, and Pfizer. H.-U.S. has an advisory relationship with Roche, Abbott Molecular, and Pfizer and has received honoraria from Novartis, Roche, Abbott Molecular, and Pfizer. B.S. is a member of the Advisory Boards for Pfizer, Novartis, Roche, Boehringer Ingelheim, and AstraZeneca. J.W. is a member of the Advisory Boards for Novartis, Roche, AstraZeneca, Boehringer Ingelheim, Lilly, Bristol-Myers Squibb, and Bayer and has received speaking fees from Novartis, Roche, AstraZeneca, Boehringer Ingelheim, and Lilly and research support from Roche, Boehringer Ingelheim, Novartis, and Bayer. T.Z. has performed advisory work for Merck, Lilly, Amgen, Roche, and Boehringer Ingelheim. D.B. is a member of regional advisory boards for Roche, Lilly, and Boehringer Ingelheim. F.L. received consulting fees from Blackfield AG. S.M.-B. received honoraria and grants from Roche and Novartis. L.N. received speaking fees from Roche and Novartis and honoraria from Pfizer. M.D.W. had paid consulting relationships with GeneCentric Cancer Therapeutics Innovation Group.

<sup>16</sup>Technical University Dortmund, Department of Chemical Biology, 44227 Dortmund, Germany.

<sup>20</sup>Department for Internal Medicine II, Jena University Hospital, Friedrich-Schiller University, 07740 Jena, Germany.

<sup>21</sup>Center for Medical Genetics, Ghent University, 9000 Ghent, Belgium.

<sup>25</sup>Department of Pulmonary Diseases, VU University Medical Center Amsterdam, 1007 MB Amsterdam, the Netherlands.

<sup>39</sup>Department of Genome Sciences, University of Washington, Seattle, WA 98195–5065, USA.

<sup>43</sup>Department of Genetics, Stanford University, Stanford, CA, USA.

eliminated almost all cases of large cell carcinomas, some of which had therapeutically relevant alterations. Prospective testing of our genomics-based diagnostic algorithm in 5145 lung cancer patients enabled a genome-based diagnosis in 3863 (75%) patients, confirmed the feasibility of rational reassignments of large cell lung cancer, and led to improvement in overall survival in patients with *EGFR*-mutant or *ALK*-rearranged cancers. Thus, our findings provide support for broad implementation of genome-based diagnosis of lung cancer.

## INTRODUCTION

Lung cancer is traditionally classified into non-small cell lung cancer (NSCLC), small cell lung cancer (SCLC), and carcinoid (CA). NSCLC is further divided into adenocarcinoma (AD), squamous cell carcinoma (SQ), and large cell carcinoma (LC), which also includes tumors with neuroendocrine differentiation [large cell neuroendocrine carcinoma (LCNEC)] (1). These categories have been enriched with detailed histomorphological and immunohistochemical characteristics leading to the 2004 World Health Organization (WHO) classification. The most detailed subcategories exist for AD, mostly defined by growth patterns (2). This rather descriptive taxonomy has been complicated in the past decade because of the recognition of somatic genetic alterations occurring in some of these subtypes: *EGFR* mutations, *KRAS* mutations, and *EML4-ALK* fusions occur mainly in lung AD (3–7), whereas mutations in *DDR2*, *FGFR2*, and *NFE2L2* or amplifications of *FGFR1* and *SOX2* mainly affect SQ (8–11). The fact that some of these are associated with clinical response to molecularly targeted therapeutics (3–5, 12, 13) emphasizes the importance of adding genetic annotation to the current taxonomy. Systematic efforts to comprehensively characterize the cancer genome (14) constantly add genome alterations to the compendium of such potentially actionable alterations (9, 15–21). Similarly, immunohistochemical analyses have challenged some of the original histomorphological diagnoses, in particular in the case of LC (22, 23). We therefore sought to assess cancer genome alterations linked to histomorphological and immunohistochemical features of the disease as well as to patient outcome to identify genetically defined subtypes of lung tumors and optimize them for biologically informed patient stratification for personalized therapeutic approaches. We then tested the clinical relevance of these molecularly defined patient subgroups prospectively in a diagnostic outreach setting.

## RESULTS

### Cancer genome alterations in human lung tumors

Recent studies have provided analyses of genome alterations in lung cancer (20, 21, 24, 25). To establish a systematic relationship of such alterations across the different cancer subtypes, we collected a total of 1882 surgically resected, fresh-frozen human lung tumor specimens with clinical annotation, yielding 1255 specimens suitable for genetic analysis (Table 1, fig. S1A, and table S1). Visual inspection of somatic copy number alterations (SCNAs) in those tumors for which both single-nucleotide polymorphism (SNP) array and histology data were available (922 of 1032) revealed distinctive patterns in cases sorted according to their initial histological subtype (Fig. 1A): some SCNAs were present in all subtypes (for example, gains affecting 5p), whereas other SCNAs were subtype-specific (for

example, amplifications of 3q containing *SOX2* in SQs) (8). In addition, the overall pattern of SCNAs differed across histological entities; for example, SCLCs and some LCs exhibited a predominance of chromosome arm-level events, in contrast to focal events in AD and SQ. Stromal admixture could mask the detection and amplitude of SCNAs (26), but only partly accounted for the diversity of SCNA patterns between subtypes (Fig. 1A). LC, in contrast to other subtypes, did not exhibit a specific SCNA pattern. To identify significant copy number alterations across lung tumor subtypes, we applied a rank sum-based method that is insensitive to tumor purity and identified 8 regions of amplification and 12 regions of deletion (fig. S2 and tables S2 and S3) (11). In cases with only one deletion, *CDKN2A* (located at 9p21) was affected in 38% (fig. S3) (11, 27).

Similarly, most mutations showed histological subtype specificity (Fig. 1B). The most frequently mutated genes were *TP53* (53.6%), *KRAS* (16.1%), *STK11* (9.8%), *EGFR* (7.2%), *KEAP1* (6.6%), and *NFE2L2* (4.5%) (20, 21, 28). Seventeen genes were altered in at least two samples (figs. S4 and S5 and tables S1, S4, and S5). The *NFE2L2/KEAP1* axis was one of the most frequently mutated oncogenic pathways in lung cancer. Furthermore, gene set enrichment analysis revealed significant changes in the expression of *NFE2L2* and *KEAP1* target genes in cases harboring such mutations (false discovery rate,  $q = 0.0008$ ) (29). Either *NFE2L2* or *KEAP1* was mutated in 10.4% of AD (21) and 16.9% of SQ (20) in mutually exclusive fashion. *NFE2L2* mutations were mainly found in SQ (30–32) (Fig. 1B and fig. S4). In addition to known frequent driver mutations, we also found rare mutations that are possibly relevant for therapy. For example, we found an R248C mutation in the fibroblast growth factor (FGF)-binding domain of *FGFR3* (33, 34) in 3 of 365 SQ (0.8%), all of which were negative for *FGFR2* mutations (35). *FGFR3*-R248C was oncogenic in vitro and associated with sensitivity to *FGFR* inhibition (fig. S6). Overall, more than 55% of all malignant lung tumors harbored at least one genetic alteration with features of a possibly tractable target (fig. S7).

We observed clinically relevant associations related to histology and stage (1). The most frequent genome alterations had no significant impact on survival in AD, LC, and SQ (Fig. 1C). *TP53* mutations were associated with inferior survival in *EGFR*-mutant patients ( $P = 0.028$ ), which was partly driven by the higher stage of these patients. Because most of these patients were diagnosed before *EGFR* inhibitors were broadly available, this observation is likely to reflect a general aggressive behavior of *EGFR/TP53* double-mutant tumors (36). Similarly, patients with concurrent mutations of *TP53* and *RBI* loss had a particularly poor prognosis (Fig. 1C), independent of stage and histology (Cox regression analysis,  $P = 0.023$ ).

### Subtype-specific genome alterations

As a next step, we sought to determine which of the genomic alterations were significant in each subtype. Across all lung tumor cases, we identified amplified and deleted regions (fig. S2) that included known protooncogenes or tumor suppressor genes, which could be assigned to specific histological subtypes. Significantly amplified chromosomal regions in AD were 5p,7p(*EGFR*),8q(*MYC*),11q(*CCND1*),12q(*MDM2*),14q(*NKX2-1*), and 17q (*ERBB2*) (21, 37); in SCLC were 1p (*MYCL1*), 2p (*MYCN*), 5p, 8p (*FGFR1*), and 19q

(*CCNE1*) (24); and in SQ were 1p (*MYCL1*), 3q (*SOX2*), 7p (*EGFR*), 8p (*FGFR1*), and 11q (*CCND1*) (8, 11, 20). CA harbored no significant SCNA. As before, in LC, we observed amplifications typical of other histologies (for example, amplification of *NKX2-1*, which is specific for AD, or of *SOX2*, which is typical of SQ). Deletions in 9p (*CDKN2A*) were present in all subtypes except SCLC, which had deletions of 3p (*FHIT*) and 13q (*RBI*) as its hallmarks (Fig. 2A).

The specific co-occurrence of several of the alterations in individual lung cancer subtypes suggested that it might be possible to determine patterns of alterations that could be used to identify subtypes based on genomics alone. We integrated data on mutations and SCNAs that were significant above the assumed background mutation rate of 0.005 (binomial test) to identify significant signature alterations for each histological subtype (Fig. 2B). Some of these alterations mainly occurred in a certain subtype, such as alterations in *ALK* (3.4%), *BRAF* (2.7%), *EGFR* (15.3%), *ERBB2* (1.7%), *KRAS* (32.6%), and *STK11* (17.4%) in AD; *MYCN* amplifications (6.5%) in SCLC; and mutations in *DDR2* (1.1%), *FGFR3* (0.8%), and *NFE2L2* (10.6%) in SQ (Figs. 1B and 2B and fig. S4). Others were not only enriched in a given subtype (for example, *NKX2-1* amplification in AD, *MYCL1* amplification and *RBI* deletion in SCLC, or *SOX2* and *FGFR1* amplification in SQ) but also present in other histologies (Fig. 2B). By contrast, LC harbored alterations typical of all other subtypes (amplification of *ERBB2* and *NKX2-1* and mutations in *KRAS* and *STK11* as in AD; amplification of *MYCL1* and *RBI* as in SCLC; amplification of *CCND1*, *FGFR1*, and *SOX2* as in SQ) (Fig. 2B) but had no significant signature alterations.

The availability of large genomic data sets enabled us to conduct a systematic analysis of co-occurrence and exclusivity of genome alterations (Fig. 2C and fig. S8). In 5 of 21 lung tumors with *BRAF* mutations affecting residues other than V600E, either *NRAS* or *KRAS* was mutated (38); such cases might be particularly sensitive to mitogen-activated protein kinase kinase (MEK) inhibition (39). Furthermore, *ERBB2* mutations never co-occurred with mutations in *BRAF*, *HRAS*, *KRAS*, *NRAS*, or *STK11* (table S1). *FGFR2* mutations did not co-occur with *FGFR1* amplifications, in support of both alterations being oncogenic drivers (fig. S6). In our data set, *EGFR* amplifications correlated with *EGFR* mutations in AD ( $P = 0.0009$ ), but not in SQ (fig. S9). *EGFR* amplification predicted treatment response and outcome of patients receiving EGFR inhibitors in some studies (40, 41), although this could not be confirmed in a study focusing exclusively on ADs (42). Thus, whether patients with *EGFR*-amplified SQ may benefit from EGFR inhibition is currently unclear. Similarly, loss of *PTEN* may influence the dependency of tumors on mutant receptor tyrosine kinases (43, 44). *PTEN* was homozygously or hemizyously deleted in 5.5 and 11.1% of *EGFR*-mutant ADs, respectively (fig. S10). In summary, genome alterations that define specific lung tumor subgroups were determined in the major histological subgroups, except for LC, which has genomic features of all other subtypes.

### The heterogeneity of large cell lung cancer

Immunohistochemistry has become an indispensable method for lung cancer diagnosis. We therefore performed an independent immunohistochemistry-based pathology review of 583 cases confirming all subclasses of lung cancer except LC (fig. S11). In 42% of all LC cases,

pathology review led to reclassification to the other subtypes sharing similar immunohistochemical and genetic features.

To gain further insight into the heterogeneity of LC, we applied gene expression-based unsupervised hierarchical clustering to 261 lung tumors, including 31 initially diagnosed LCs. Whereas 87% of ADs, 92% of CA, 84% of SCLCs, and 76% of SQs formed distinct clusters, LCs were dispersed across all other clusters (Fig. 3A and tables S1 and S6). Except for one case, all LCs clustered with those tumors of the other subtypes that shared the same histology-defining signature alterations (Fig. 3, A and B, and table S7). Similar results were obtained when applying consensus clustering, where almost 98% of AD cases formed a distinct cluster, as well as 84% of SCLCs and 77% of SQs. Sample overlap between clusters obtained by the two methods was 95%, 93%, and 75% for the clusters, which mainly included tumors with neuro-endocrine differentiation (NEC), AD, or SQ cases, respectively (table S1). Applying recently described transcriptional classifiers of subtypes of AD (45) and SQ (46) revealed comparable results (figs. S12 and S13 and table S1). An integrative analysis of 209 cases including copy number and expression data using iCluster defined two subgroups that were mainly driven by amplifications on chromosome 3 (3q13.31-3q29) and chromosome 12 (12p13.33-12q15) and thus did not reveal distinct clinically relevant subtypes (47).

LCNEC exhibited substantial transcriptional similarity to SCLC when analyzed with unsupervised hierarchical clustering and classification (Fig. 3A, gray triangles, and figs. S12 and S13). Furthermore, similar amplified and deleted regions were observed in LCNEC and SCLC when compared with other histological subtypes (fig. S14).

Finally, LCNEC shared the significantly mutated genes *TP53*, *RBI*, and *EP300* with SCLC, as determined by whole-exome sequencing of 15 and transcriptome sequencing of 10 pathologically reviewed LCNECs (Fig. 3C and tables S8 and S9). We also found additional mutated genes in LCNEC that typically occurred in AD or SQ but did not reach significance in LCNEC (tables S8 and S9). Thus, LCNEC is most similar to SCLC, with individual cases bearing mutations of other subtypes. The genetic similarity between LCNEC and SCLC is also reflected by a similar overall survival (Fig. 3D). In summary, LC exhibits a general diagnostic plasticity when considering data on chromosomal copy number, gene mutations, gene expression, and immunohistochemistry. Combined immunohistochemical and genomic analysis is therefore ideal to classify this heterogeneous group as AD, SQ, or NEC.

### Automated genomics-based lung tumor classification

Given the strong correlation of specific genome alterations with certain histological subtypes, we devised a statistical model to test if subtypes could be predicted robustly based on such alterations alone. The diagnosis predicted by our model and the original diagnosis (Fig. 4A, left) or the diagnosis obtained from pathology review were highly similar for AD and SQ (Fig. 4A, right, fig. S15, and table S10). Only a few of the AD or SQ cases were reclassified according to the predominant genome alterations in these cases (fig. S15). Feature selection and automated reclassification using a similar model was highly stable when applied to validation data sets of 382 AD (21), SQ (20), and SCLC (25) cases (Fig. 4B, fig. S16, and table S11). Review of individual discrepant cases revealed subtype-

specific alterations, such as in one case that was predicted to be AD but classified by central pathology review as SQ, which harbored the oncogenic S310F *ERBB2* mutation (48); *ERBB2* mutations were typical of AD. Furthermore, one of three cases predicted to be SQ but classified as AD by pathology review exhibited amplification of *FGFR1* and *CCND1*, both predictive of SQ in our data set (fig. S16A). Nevertheless, most initial diagnoses of AD, SCLC, and SQ were confirmed by both our model and pathology review (Fig. 4, A and B, and fig. S11), but most of the LCs with at least one genomic alteration were reassigned to either AD, SQ, or aggressive neuroendocrine lung cancer/ SCLC in accordance with the pathological review (Fig. 4C, fig. S15, and table S10). Even in those cases where immunohistochemistry did not yield an unequivocal diagnosis, most remaining LC cases could be reassigned genetically to the other lung cancer subtypes (Fig. 4C).

### Clinical evaluation of genomics-based lung cancer diagnoses

To evaluate our combined genomic and immunohistochemical diagnostic approach, we enrolled 5145 lung cancer patients in a molecular screening outreach program run by Network Genomic Medicine (NGM) in the region of our cancer center between January 2010 and April 2013 (Table 1, fig. S1B, and table S12). The addition of immunohistochemistry to the diagnostic workup of tumors with LC features reduced the prevalence of this subgroup from 5.9% in the retrospective analysis (table S1) to 1.3% in the diagnostic NGM data set (Fig. 5A). The expression of TTF-1 and CK7 as immunohistochemical markers of AD, as well as p63 and CK5 as markers of SQ, allowed such cases to be assigned to either AD or SQ. Expression of neuroendocrine markers CD56, chromogranin A, and synaptophysin was tested to identify tumors with neuroendocrine differentiation.

We performed central genotyping for key alterations identified within our retrospective genomic study (*ALK*, *BRAF*, *DDR2*, *EGFR*, *ERBB2*, *FGFR1*, *KRAS*, and *PIK3CA*) (fig. S17). Genomic testing was feasible in 3863 (75%) paraffin-embedded tumor samples obtained by routine diagnostic procedures, yielding 1481 genomic alterations (table S12). Treatment recommendations were provided to the network partners to enable genetically tailored cancer therapy either with approved drugs or within clinical trials. Sixty-four of 84 advanced-stage (IIIB or IV) patients with an *EGFR* mutation (76%) received erlotinib or gefitinib, and 15 of 30 advanced-stage patients with *ALK* translocation (50%) received crizotinib. Furthermore, 34 patients with alterations in *BRAF* ( $n = 4$ ), *KRAS* ( $n = 10$ ), or *FGFR1* ( $n = 20$ ) were enrolled in clinical trials. The frequencies of genomic alterations in the diagnostic NGM data set were similar to those of the retrospective discovery data set (Fig. 5, A and B). *BRAF* mutations were predominantly activating (49, 50) (60%, possibly associated with sensitivity to BRAF or MEK inhibition). In 24% of the cases, *BRAF* mutations were predicted to be inactivating (50, 51), which might predict sensitivity to dasatinib (52) (fig. S18). In 34% of the remaining LC cases, which were not otherwise classifiable, we found signature alterations of AD or SQ (Fig. 5, A and B) that provided both a genetic diagnosis and a rationale for genetically tailored therapy. Thus, combined immunohistochemical and genetic diagnosis reduced the subgroup of LC to 1.1%.

We next compared the survival of patients in our historical and prospective data sets. Survival of stage IIIB/IV patients in the retrospective data set was better than that in the prospective data set ( $P < 0.02$ , fig. S19). The most likely reason for this difference is a higher proportion of patients undergoing surgery with curative intention in the retrospective data set (which consisted of surgically resected cases only), whereas very few patients underwent surgery in the prospective data set. In contrast, the improved survival of patients with earlier-stage disease in the prospective data set compared to the retrospective data set most likely reflects treatment improvements. Histology had no major impact on survival of advanced-stage patients in both data sets (Fig. 5C).

We next performed analyses within the prospective NGM cohort alone. When comparing the survival of patients in our prospective data set, whose tumors had been genotyped ( $n = 975$ ) to that of patients in the same cohort, in whom genetic diagnosis was not feasible (for example, because of insufficient tissue;  $n = 277$ ), genotyping alone had stage- and histology-independent impact on overall survival in multivariate analyses ( $P = 0.002$ ) (Fig. 5D, upper left). Although this observation most likely results from the favorable outcome in patients treated with kinase inhibitors, it demonstrates that genotyping is mandatory for patients to benefit from targeted therapeutic intervention. Accordingly, among the different genotypes determined within the prospective NGM cohort, *EGFR* mutations were associated with improved survival (Fig. 5D, upper right; hazard ratio, 0.617; 95% confidence interval, 0.442 to 0.859;  $P = 0.004$ ). We note, however, that by the time ALK fusion testing was introduced, ALK inhibitors were not yet broadly available, which may explain why ALK fusions were not generally associated with improved survival in our cohort. Patients with *EGFR*-mutant lung cancer treated with EGFR inhibitors survived longer than those not receiving EGFR inhibitors (median overall survival, 31.5 versus 9.6 months;  $P < 0.001$ ) (Fig. 5D, lower left). Similarly, the overall survival of patients with *ALK*-rearranged lung cancer treated with crizotinib was significantly better compared to *ALK*-positive patients not receiving crizotinib (median overall survival, 23 versus 11 months;  $P = 0.024$ ) (Fig. 5D, lower right). No difference in the number of therapeutic regimens and the number of platinum-based treatment regimens existed between patients with *EGFR*-mutant lung cancer treated with EGFR inhibitors and patients not treated with EGFR inhibitors (two-sided  $t$  test,  $P = 0.43$ ). Patients with lung cancer harboring *ALK* rearrangements treated with crizotinib did not differ in the number of platinum-based chemotherapy regimens, but differed in the total number of treatment regimens received: patients treated with crizotinib received an average of 2.17 regimens, compared to 1.3 regimens in patients who did not receive crizotinib (two-sided  $t$  test,  $P = 0.034$ ) (table S12). However, in a Cox regression analysis including treatment with kinase inhibitors and number of treatment regimens as variables, only treatment with kinase inhibitors had a significant impact on survival ( $P = 0.043$ ).

In summary, we validated the frequencies of signature alterations across lung cancer subtypes, demonstrated the feasibility of thorough and broad genome diagnostics in an academic-to-nonacademic outreach setting, confirmed the almost universal reassignment of LC to the other biologically relevant diagnoses, and showed that the introduction of a molecular diagnosis coupled to specific therapeutic intervention improves the overall

survival of patients with alterations in *EGFR* and *ALK* compared to standard chemotherapy (53, 54).

## DISCUSSION

Here, we defined a minimal set of genome alterations for genomic lung cancer diagnosis using comprehensive copy number analyses in combination with focused sequencing and RNA expression analysis. This approach yielded robust frequencies of genome alterations and afforded reassignment of LC to therapeutically relevant groups such as AD or SQ. Finally, introduction of genome diagnostics in an outreach setting not only confirmed these observations but also resulted in substantial improvement in overall survival of patients receiving genetically informed therapeutic intervention compared to standard chemotherapeutic treatment.

A major advantage of this combined histopathological and genetic analysis is the classification of the group of LCs, a subtype that is poorly defined, mainly because of a lack of specific morphologic features of AD, SCLC, or SQ. Applying immunohistochemistry (22, 23) (for example, TTF-1 for AD and p63 for SQ) helped assigning some of the LC cases to other categories. However, the addition of genome annotation not only confirmed several of the immunohistochemical assignments but also added information on possibly therapeutically relevant alterations and afforded classification in cases where definite pathological diagnosis was not possible. We have also observed a marked similarity between several LCNEC tumors and SCLC that shared the same pattern of SCNAs, the particularly poor survival, and the most significant gene mutations in this data set of limited size.

We also found (exceptionally rare) SQ tumors bearing *EGFR* mutations or *ALK* rearrangements, which might be treatable with targeted therapies as well. Thus, genomic diagnosis should include all subtypes and all genome alterations to provide a therapeutic rationale for all possible patients. We propose to capture both genomic and immunohistochemical data to link these taxa to treatment benefit in trials (fig. S20).

In our outreach study, genotyping alone—as a prerequisite for personalized treatment—was associated with improved patient survival. These epidemiological results emphasize the need for broad availability of systematic and comprehensive genomic lung cancer diagnoses.

We note that the prospective part of our study was not a randomized clinical trial but an observational diagnostic intervention study. Unfortunately, obtaining overall survival data in a randomized fashion requires prohibiting patients from crossing to the other treatment arm—an irresponsible measure in this setting. Thus, registry data from observational studies like ours may be an approach to demonstrate differences in survival between two different therapeutic strategies.

Furthermore, the two cohorts analyzed in this study differ in important aspects: Whereas patient registration and genotyping occurred in a central study office in the prospective cohort, tumors of the retrospective cohort were obtained from multiple centers before targeted therapies became broadly available, thus giving us limited control over the quality of the clinical data. Additional differences existed in the time of treatment and distribution



of stages and rate of surgical treatment. Finally, clinical annotation was not complete, and data gaps exist for performance status, smoking status, and treatment. We note, however, that all major class-related genotypic and immunohistochemical findings were consistent across both cohorts, thus underscoring the validity of our approach.

In summary, we provide a blueprint for genomic diagnosis of lung tumors. Determination of immunohistochemical, genomic, and clinical features may thus be combined to yield classes of tumors that are biologically relevant, afford genomically tailored stratification of patients into clinical trials, and improve overall survival of patients with lung cancer.

## MATERIALS AND METHODS

### Study design

Detailed information on materials and methods, including study design and statistics, is given in the Supplementary Materials. In brief, we collected frozen tissue or genomic DNA from 1882 resected primary lung tumors (table S1) after obtaining informed consent. Mutations in 28 genes (*ABL1*, *AKT2*, *ALK*, *BRAF*, *CDK4*, *DDR2*, *EGFR*, *EPHA3*, *EPHA5*, *ERBB2*, *FGFR1*, *FGFR2*, *FGFR3*, *FLT3*, *HRAS*, *JAK2*, *KEAP1*, *KIT*, *KRAS*, *NFE2L2*, *NRAS*, *NTRK1*, *NTRK3*, *PDGFRA*, *PIK3CA*, *STK11*, *TP53*, and *RET*) were analyzed in 1127 tumor specimens using multiple technologies (table S13, A and B) (9, 28, 38). Rearrangements of *ALK* ( $n = 602$ ), *RET* ( $n = 362$ ), and *ROS1* ( $n = 211$ ) were detected by fluorescence in situ hybridization. Gene copy number of 1032 and whole-exome sequencing data of 15 LCNEC tumors were analyzed as described previously (11, 24). Gene expression analyses were performed in 261 samples. Five hundred eighty-three cases were independently reviewed by lung pathologists (E.B. and W.D.T.) (1, 2) (fig. S1). For the prospective diagnostic cohort, 5145 lung cancer patients were included for central genotyping for alterations in *ALK*, *BRAF*, *DDR2*, *EGFR*, *ERBB2*, *FGFR1*, *KRAS*, or *PIK3CA*, and written reports were provided to the treating oncologists, containing the detected mutation and a treatment recommendation (fig. S17 and table S12).

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank W. Vogel, C. Beschoner, C. Becker, B. Pinther, and P. Löhner for their technical assistance and M. Knowles for providing the FGFR3b pBabe plasmids. We also thank the regional computing center of the University of Cologne (Regional Computing Centre) for providing the CPU time on the German Research Foundation (DFG)-funded supercomputer “CHEOPS” as well as for support. The Victorian Cancer Biobank stored and processed the Australian biospecimens used in this study.

**Funding:** This work was supported by the EU-Framework Programme CURELUNG (HEALTH-F2-2010-258677 to R.K.T., R.B., and J.W.), by the Deutsche Forschungsgemeinschaft through TH1386/3-1 (to R.K.T. and M.L.S.) and through SFB832 (TP6 to R.K.T. and J.W.; TP5 to L.C.H.) and SB680 (to J.B.), by the German Ministry of Science and Education as part of the NGFNplus program (grant 01GS08100 to R.K.T. and J.W.; grant 01GS08101 to P.N.), by the Deutsche Krebshilfe as part of the Oncology Centers of Excellence funding program (to R.K.T., R.B., and J.W.) and SyBaCol (to J.B.), by the Max Planck Society, by the Behrens-Weise Foundation (M.I.F.A.NEUR8061 to R.K.T.), by Stand Up To Cancer—American Association for Cancer Research Innovative Research Grant (SU2C-AACR-IR60109 to W.P. and R.K.T.), by the Roy Castle Lung Cancer Foundation UK (to J.K.F.), and by an anonymous foundation (to R.K.T.). T.M. was supported by 82DZL00402 (DZL biobank

platform) from the German Centre for Lung Research. Funding of supercomputer CHEOPS was provided by the DFG and the Ministry of Research of the State of North Rhine–Westphalia.

## REFERENCES AND NOTES

1. Travis, W.; Brambilla, E.; Mueller-Hermelink, H.; Harris, C. World Health Organization Classification of Tumours. Pathology and Genetics: Tumours of the Lung, Pleura, Thymus and Heart. World Health Organization; Geneva: 2004.
2. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, Powell CA, Beer D, Riely G, Garg K, Austin JH, Rusch VW, Hirsch FR, Jett J, Yang PC, Gould M. American Thoracic Society, International Association for the Study of Lung Cancer/ American Thoracic Society/ European Respiratory Society: International multidisciplinary classification of lung adenocarcinoma: Executive summary. *Proc Am Thorac Soc.* 2011; 8:381–385. [PubMed: 21926387]
3. Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med.* 2004; 350:2129–2139. [PubMed: 15118073]
4. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. *EGFR* mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science.* 2004; 304:1497–1500. [PubMed: 15118125]
5. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol.* 2011; 12:175–180. [PubMed: 21277552]
6. Rodenhuis S, Slebos RJ, Boot AJ, Evers SG, Mooi WJ, Wagenaar SS, van Bodegom PC, Bos JL. Incidence and possible clinical significance of *K-ras* oncogene activation in adenocarcinoma of the human lung. *Cancer Res.* 1988; 48:5738–5741. [PubMed: 3048648]
7. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature.* 2007; 448:561–566. [PubMed: 17625570]
8. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, Kim SY, Wardwell L, Tamayo P, Gat-Viks I, Ramos AH, Woo MS, Weir BA, Getz G, Beroukhi R, O’Kelly M, Dutt A, Rozenblatt-Rosen O, Dziunycz P, Komisarof J, Chirieac LR, Lafargue CJ, Scheble V, Wilbertz T, Ma C, Rao S, Nakagawa H, Stairs DB, Lin L, Giordano TJ, Wagner P, Minna JD, Gazdar AF, Zhu CQ, Brose MS, Ceccanello I, Ribeiro U Jr, Marie SK, Dahl O, Shivdasani RA, Tsao MS, Rubin MA, Wong KK, Regev A, Hahn WC, Beer DG, Rustgi AK, Meyerson M. *SOX2* is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet.* 2009; 41:1238–1242. [PubMed: 19801978]
9. Hammerman PS, Sos ML, Ramos AH, Xu C, Dutt A, Zhou W, Brace LE, Woods BA, Lin W, Zhang J, Deng X, Lim SM, Heynck S, Peifer M, Simard JR, Lawrence MS, Onofrio RC, Salvesen HB, Seidel D, Zander T, Heuckmann JM, Soltermann A, Moch H, Koker M, Leenders F, Gabler F, Querings S, Ansén S, Brambilla E, Brambilla C, Lorimier P, Brustugun OT, Helland A, Petersen I, Clement JH, Groen H, Timens W, Sietsma H, Stoelben E, Wolf J, Beer DG, Tsao MS, Hanna M, Hatton C, Eck MJ, Janne PA, Johnson BE, Winckler W, Greulich H, Bass AJ, Cho J, Rauh D, Gray NS, Wong KK, Haura EB, Thomas RK, Meyerson M. Mutations in the *DDR2* kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov.* 2011; 1:78–89. [PubMed: 22328973]
10. Dutt A, Ramos AH, Hammerman PS, Mermel C, Cho J, Sharifnia T, Chande A, Tanaka KE, Stransky N, Greulich H, Gray NS, Meyerson M. Inhibitor-sensitive *FGFR1* amplification in human non-small cell lung cancer. *PLoS One.* 2011; 6:e20351. [PubMed: 21666749]
11. Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM, Ullrich RT, Menon R, Maier S, Soltermann A, Moch H, Wagener P, Fischer F, Heynck S, Koker M, Schöttle J, Leenders F, Gabler F, Dabow I, Querings S, Heukamp LC, Balke-Want H, Ansén S, Rauh D, Baessmann I, Altmüller J, Wainer Z, Conron M, Wright G, Russell P, Solomon B, Brambilla E, Brambilla C,

Lorimier P, Sollberg S, Brustugun OT, Engel-Riedel W, Ludwig C, Petersen I, Sanger J, Clement J, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman D, Cappuzzo F, Ligorio C, Damiani S, Hallek M, Beroukhim R, Pao W, Klebl B, Baumann M, Buettner R, Ernestus K, Stoelben E, Wolf J, Nurnberg P, Perner S, Thomas RK. Frequent and focal *FGFR1* amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med*. 2010; 2:62ra93.

12. Pao W, Chmielecki J. Rational, biologically based treatment of *EGFR*-mutant non-small-cell lung cancer. *Nat Rev Cancer*. 2010; 10:760–774. [PubMed: 20966921]
13. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou SH, Dezube BJ, Janne PA, Costa DB, Varella-Garcia M, Kim WH, Lynch TJ, Fidias P, Stubbs H, Engelman JA, Sequist LV, Tan W, Gandhi L, Mino-Kenudson M, Wei GC, Shreeve SM, Ratain MJ, Settleman J, Christensen JG, Haber DA, Wilner K, Salgia R, Shapiro GI, Clark JW, Iafrate AJ. Anaplastic lymphoma kinase inhibition in non–small-cell lung cancer. *N Engl J Med*. 2010; 363:1693–1703. [PubMed: 20979469]
14. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010; 11:685–696. [PubMed: 20847746]
15. Bergethon K, Shaw AT, Ou SH, Katayama R, Lovly CM, McDonald NT, Massion PP, Siwak-Tapp C, Gonzalez A, Fang R, Mark EJ, Batten JM, Chen H, Wilner KD, Kwak EL, Clark JW, Carbone DP, Ji H, Engelman JA, Mino-Kenudson M, Pao W, Iafrate AJ. *ROS1* rearrangements define a unique molecular class of lung cancers. *J Clin Oncol*. 2012; 30:863–870. [PubMed: 22215748]
16. Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, Kim YT, Kim JI, Kang JH, Seo JS. A transforming *KIF5B* and *RET* gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res*. 2012; 22:436–445. [PubMed: 22194472]
17. Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, Sakamoto H, Tsuta K, Furuta K, Shimada Y, Iwakawa R, Ogiwara H, Oike T, Enari M, Schetter AJ, Okayama H, Haugen A, Skaug V, Chiku S, Yamanaka I, Arai Y, Watanabe S, Sekine I, Ogawa S, Harris CC, Tsuda H, Yoshida T, Yokota J, Shibata T. *KIF5B-RET* fusions in lung adenocarcinoma. *Nat Med*. 2012; 18:375–377. [PubMed: 22327624]
18. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M, Curran JA, Balasubramanian S, Bloom T, Brennan KW, Donahue A, Downing SR, Frampton GM, Garcia L, Juhn F, Mitchell KC, White E, White J, Zwirko Z, Peretz T, Nechushtan H, Soussan-Gutman L, Kim J, Sasaki H, Kim HR, Park SI, Ercan D, Sheehan CE, Ross JS, Cronin MT, Janne PA, Stephens PJ. Identification of new *ALK* and *RET* gene fusions from colorectal and lung cancer biopsies. *Nat Med*. 2012; 18:382–384. [PubMed: 22327622]
19. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, Lim Choi Y, Satoh Y, Okumura S, Nakagawa K, Mano H, Ishikawa Y. *RET*, *ROS1* and *ALK* fusions in lung cancer. *Nat Med*. 2012; 18:378–381. [PubMed: 22327623]
20. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489:519–525. [PubMed: 22960745]
21. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, Sougnez C, Auclair D, Lawrence MS, Stojanov P, Cibulskis K, Choi K, de Waal L, Sharifnia T, Brooks A, Greulich H, Banerji S, Zander T, Seidel D, Leenders F, Ansen S, Ludwig C, Engel-Riedel W, Stoelben E, Wolf J, Goparju C, Thompson K, Winckler W, Kwiatkowski D, Johnson BE, Janne PA, Miller VA, Pao W, Travis WD, Pass HI, Gabriel SB, Lander ES, Thomas RK, Garraway LA, Getz G, Meyerson M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012; 150:1107–1120. [PubMed: 22980975]
22. Pardo J, Martinez-Peuella AM, Sola JJ, Panizo A, Gurpide A, Martinez-Peuella JM, Lozano MD. Large cell carcinoma of the lung: An endangered species? *Appl Immunohistochem Mol Morphol*. 2009; 17:383–392. [PubMed: 19444077]
23. Rekhtman N, Ang DC, Sima CS, Travis WD, Moreira AL. Immunohistochemical algorithm for differentiation of lung adenocarcinoma and squamous cell carcinoma based on large series of whole-tissue sections with validation in small specimens. *Mod Pathol*. 2011; 24:1348–1359. [PubMed: 21623384]

24. Peifer M, Fernández-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, Müller C, Di Cerbo V, Schildhaus HU, Altmüller J, Baessmann I, Becker C, de Wilde B, Vandesompele J, Böhm D, Ansén S, Gabler F, Wilkening I, Heynck S, Heuckmann JM, Lu X, Carter SL, Cibulskis K, Banerji S, Getz G, Park KS, Rauh D, Grütter C, Fischer M, Pasqualucci L, Wright G, Wainer Z, Russell P, Petersen I, Chen Y, Stoelben E, Ludwig C, Schnabel P, Hoffmann H, Muley T, Brockmann M, Engel-Riedel W, Muscarella LA, Fazio VM, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman DA, Snijders PJ, Cappuzzo F, Ligorio C, Damiani S, Field J, Solberg S, Brustugun OT, Lund-Iversen M, Sängler J, Clement JH, Soltermann A, Moch H, Weder W, Solomon B, Soria JC, Validire P, Besse B, Brambilla E, Brambilla C, Lantuejoul S, Lorimier P, Schneider PM, Hallek M, Pao W, Meyerson M, Sage J, Shendure J, Schneider R, Büttner R, Wolf J, Nürnberg P, Perner S, Heukamp LC, Brindle PK, Haas S, Thomas RK. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet.* 2012; 44:1104–1110. [PubMed: 22941188]
25. Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, Bergbower EA, Guan Y, Shin J, Guillory J, Rivers CS, Foo CK, Bhatt D, Stinson J, Gnad F, Haverty PM, Gentleman R, Chaudhuri S, Janakiraman V, Jaiswal BS, Parikh C, Yuan W, Zhang Z, Koeppen H, Wu TD, Stern HM, Yauch RL, Huffman KE, Paskulin DD, Illei PB, Varella-Garcia M, Gazdar AF, de Sauvage FJ, Bourgon R, Minna JD, Brock MV, Seshagiri S. Comprehensive genomic analysis identifies *SOX2* as a frequently amplified gene in small-cell lung cancer. *Nat Genet.* 2012; 44:1111–1116. [PubMed: 22941189]
26. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhi R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol.* 2012; 30:413–421. [PubMed: 22544022]
27. Iwakawa R, Kohno T, Anami Y, Noguchi M, Suzuki K, Matsuno Y, Mishima K, Nishikawa R, Tashiro F, Yokota J. Association of p16 homozygous deletions with clinicopathologic characteristics and EGFR/KRAS/p53 mutations in lung adenocarcinoma. *Clin Cancer Res.* 2008; 14:3746–3753. [PubMed: 18559592]
28. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, Fulton L, Fulton RS, Zhang Q, Wendl MC, Lawrence MS, Larson DE, Chen K, Dooling DJ, Sabo A, Hawes AC, Shen H, Jiangiani SN, Lewis LR, Hall O, Zhu Y, Mathew T, Ren Y, Yao J, Scherer SE, Clerc K, Metcalf GA, Ng B, Milosavljevic A, Gonzalez-Garay ML, Osborne JR, Meyer R, Shi X, Tang Y, Koboldt DC, Lin L, Abbott R, Miner TL, Pohl C, Fewell G, Haipek C, Schmidt H, Dunford-Shore BH, Kraja A, Crosby SD, Sawyer CS, Vickery T, Sander S, Robinson J, Winckler W, Baldwin J, Chiriac LR, Dutt A, Fennell T, Hanna M, Johnson BE, Onofrio RC, Thomas RK, Tonon G, Weir BA, Zhao X, Ziaugra L, Zody MC, Giordano T, Orringer MB, Roth JA, Spitz MR, Wistuba II, Ozenberger B, Good PJ, Chang AC, Beer DG, Watson MA, Ladanyi M, Broderick S, Yoshizawa A, Travis WD, Pao W, Province MA, Weinstock GM, Varmus HE, Gabriel SB, Lander ES, Gibbs RA, Meyerson M, Wilson RK. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature.* 2008; 455:1069–1075. [PubMed: 18948947]
29. Malhotra D, Portales-Casamar E, Singh A, Srivastava S, Arenillas D, Happel C, Shyr C, Wakabayashi N, Kensler TW, Wasserman WW, Biswal S. Global mapping of binding sites for Nrf2 identifies novel targets in cell survival response through ChIP-Seq profiling and network analysis. *Nucleic Acids Res.* 2010; 38:5718–5734. [PubMed: 20460467]
30. Kim YR, Oh JE, Kim MS, Kang MR, Park SW, Han JY, Eom HS, Yoo NJ, Lee SH. Oncogenic *NRF2* mutations in squamous cell carcinomas of oesophagus and skin. *J Pathol.* 2010; 220:446–451. [PubMed: 19967722]
31. Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, Tsuta K, Asamura H, Yamamoto M, Hirohashi S. Cancer related mutations in *NRF2* impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. *Proc Natl Acad Sci USA.* 2008; 105:13568–13573. [PubMed: 18757741]
32. Ohta T, Iijima K, Miyamoto M, Nakahara I, Tanaka H, Ohtsuji M, Suzuki T, Kobayashi A, Yokota J, Sakiyama T, Shibata T, Yamamoto M, Hirohashi S. Loss of Keap1 function activates Nrf2 and

- provides advantages for lung cancer cell growth. *Cancer Res.* 2008; 68:1303–1309. [PubMed: 18316592]
33. Cappellen D, De Oliveira C, Ricol D, de Medina S, Bourdin J, Sastre-Garau X, Chopin D, Thiery JP, Radvanyi F. Frequent activating mutations of *FGFR3* in human bladder and cervix carcinomas. *Nat Genet.* 1999; 23:18–20. [PubMed: 10471491]
  34. Karoui M, Hofmann-Radvanyi H, Zimmermann U, Couvelard A, Degott C, Faridoni-Laurens L, Ahomadegbe JC, Gazzeri S, Brambilla E, Clerici T, Charbonnier P, Tresallet C, Mitry E, Penna C, Rougier P, Boileau C, Thiery JP, Nordlinger B, Franc B, Radvanyi F. No evidence of somatic *FGFR3* mutation in various types of carcinoma. *Oncogene.* 2001; 20:5059–5061. [PubMed: 11526491]
  35. Dutt A, Salvesen HB, Chen TH, Ramos AH, Onofrio RC, Hatton C, Nicoletti R, Winckler W, Grewal R, Hanna M, Wyhs N, Ziaugra L, Richter DJ, Trovik J, Engelsens IB, Stefansson IM, Fennell T, Cibulskis K, Zody MC, Akslen LA, Gabriel S, Wong KK, Sellers WR, Meyerson M, Greulich H. Drug-sensitive *FGFR2* mutations in endometrial carcinoma. *Proc Natl Acad Sci USA.* 2008; 105:8713–8717. [PubMed: 18552176]
  36. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, Lee J, Jung YJ, Kim JO, Shin JY, Yu SB, Kim J, Lee ER, Kang CH, Park IK, Rhee H, Lee SH, Kim JI, Kang JH, Kim YT. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* 2012; 22:2109–2119. [PubMed: 22975805]
  37. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhi R, Lin WM, Province MA, Kraja A, Johnson LA, Shah K, Sato M, Thomas RK, Barletta JA, Borecki IB, Broderick S, Chang AC, Chiang DY, Chirieac LR, Cho J, Fujii Y, Gazdar AF, Giordano T, Greulich H, Hanna M, Johnson BE, Kris MG, Lash A, Lin L, Lindeman N, Mardis ER, McPherson JD, Minna JD, Morgan MB, Nadel M, Orringer MB, Osborne JR, Ozenberger B, Ramos AH, Robinson J, Roth JA, Rusch V, Sasaki H, Shepherd F, Sougnez C, Spitz MR, Tsao MS, Twomey D, Verhaak RG, Weinstock GM, Wheeler DA, Winckler W, Yoshizawa A, Yu S, Zakowski MF, Zhang Q, Beer DG, Wistuba II, Watson MA, Garraway LA, Ladanyi M, Travis WD, Pao W, Rubin MA, Gabriel SB, Gibbs RA, Varmus HE, Wilson RK, Lander ES, Meyerson M. Characterizing the cancer genome in lung adenocarcinoma. *Nature.* 2007; 450:893–898. [PubMed: 17982442]
  38. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, Wang M, Feng W, Zander T, MacConaill L, Lee JC, Nicoletti R, Hatton C, Goyette M, Girard L, Majumdar K, Ziaugra L, Wong KK, Gabriel S, Beroukhi R, Peyton M, Barretina J, Dutt A, Emery C, Greulich H, Shah K, Sasaki H, Gazdar A, Minna J, Armstrong SA, Mellinghoff IK, Hodi FS, Dranoff G, Mischel PS, Cloughesy TF, Nelson SF, Liau LM, Mertz K, Rubin MA, Moch H, Loda M, Catalona W, Fletcher J, Signoretti S, Kaye F, Anderson KC, Demetri GD, Dummer R, Wagner S, Herlyn M, Sellers WR, Meyerson M, Garraway LA. High-throughput oncogene mutation profiling in human cancer. *Nat Genet.* 2007; 39:347–351. [PubMed: 17293865]
  39. Su F, Viros A, Milagre C, Trunzer K, Bollag G, Spleiss O, Reis-Filho JS, Kong X, Koya RC, Flaherty KT, Chapman PB, Kim MJ, Hayward R, Martin M, Yang H, Wang Q, Hilton H, Hang JS, Noe J, Lambros M, Geyer F, Dhomen N, Niculescu-Duvaz I, Zamboni A, Niculescu-Duvaz D, Preece N, Robert L, Otte NJ, Mok S, Kee D, Ma Y, Zhang C, Habets G, Burton EA, Wong B, Nguyen H, Kockx M, Andries L, Lestini B, Nolop KB, Lee RJ, Joe AK, Troy JL, Gonzalez R, Hutson TE, Puzanov I, Chmielowski B, Springer CJ, McArthur GA, Sosman JA, Lo RS, Ribas A, Marais R. *RAS* mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. *N Engl J Med.* 2012; 366:207–215. [PubMed: 22256804]
  40. Cappuzzo F, Hirsch FR, Rossi E, Bartolini S, Ceresoli GL, Bemis L, Haney J, Witta S, Danenberg K, Domenichini I, Ludovini V, Magrini E, Gregorc V, Doglioni C, Sidoni A, Tonato M, Franklin WA, Crino L, Bunn PA Jr, Varella-Garcia M. Epidermal growth factor receptor gene and protein and gefitinib sensitivity in non-small-cell lung cancer. *J Natl Cancer Inst.* 2005; 97:643–655. [PubMed: 15870435]
  41. Tsao MS, Sakurada A, Cutz JC, Zhu CQ, Kamel-Reid S, Squire J, Lorimer I, Zhang T, Liu N, Daneshmand M, Marrano P, da Cunha Santos G, Lagarde A, Richardson F, Seymour L, Whitehead M, Ding K, Pater J, Shepherd FA. Erlotinib in lung cancer—Molecular and clinical predictors of outcome. *N Engl J Med.* 2005; 353:133–144. [PubMed: 16014883]
  42. Fukuoka M, Wu YL, Thongprasert S, Sunpaweravong P, Leong SS, Sriuranpong V, Chao TY, Nakagawa K, Chu DT, Saijo N, Duffield EL, Rukazenzov Y, Speake G, Jiang H, Armour AA, To

KF, Yang JC, Mok TS. Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/ paclitaxel in clinically selected patients with advanced non–small-cell lung cancer in Asia (IPASS). *J Clin Oncol*. 2011; 29:2866–2874. [PubMed: 21670455]

43. Mellinghoff IK, Wang MY, Vivanco I, Haas-Kogan DA, Zhu S, Dia EQ, Lu KV, Yoshimoto K, Huang JH, Chute DJ, Riggs BL, Horvath S, Liau LM, Cavenee WK, Rao PN, Beroukhi R, Peck TC, Lee JC, Sellers WR, Stokoe D, Prados M, Cloughesy TF, Sawyers CL, Mischel PS. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N Engl J Med*. 2005; 353:2012–2024. [PubMed: 16282176]
44. Sos ML, Koker M, Weir BA, Heynck S, Rabinovsky R, Zander T, Seeger JM, Weiss J, Fischer F, Frommolt P, Michel K, Peifer M, Mermel C, Girard L, Peyton M, Gazdar AF, Minna JD, Garraway LA, Kashkar H, Pao W, Meyerson M, Thomas RK. *PTEN* loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR. *Cancer Res*. 2009; 69:3256–3261. [PubMed: 19351834]
45. Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, Hayward MC, Miller CR, Socinski MA, Parsons AM, Thorne LB, Haitcock BE, Veeramachaneni NK, Funkhouser WK, Randell SH, Bernard PS, Perou CM, Hayes DN. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS One*. 2012; 7:e36530. [PubMed: 22590557]
46. Wilkerson MD, Yin X, Hoadley KA, Liu Y, Hayward MC, Cabanski CR, Muldrew K, Miller CR, Randell SH, Socinski MA, Parsons AM, Funkhouser WK, Lee CB, Roberts PJ, Thorne L, Bernard PS, Perou CM, Hayes DN. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin Cancer Res*. 2010; 16:4864–4875. [PubMed: 20643781]
47. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*. 2009; 25:2906–2912. [PubMed: 19759197]
48. Greulich H, Kaplan B, Mertins P, Chen TH, Tanaka KE, Yun CH, Zhang X, Lee SH, Cho J, Ambrogio L, Liao R, Imielinski M, Banerji S, Berger AH, Lawrence MS, Zhang J, Pho NH, Walker SR, Winckler W, Getz G, Frank D, Hahn WC, Eck MJ, Mami DR, Jaffe JD, Carr SA, Wong KK, Meyerson M. Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of *ERBB2*. *Proc Natl Acad Sci USA*. 2012; 109:14476–14481. [PubMed: 22908275]
49. Dahlman KB, Xia J, Hutchinson K, Ng C, Hucks D, Jia P, Atefi M, Su Z, Branch S, Lyle PL, Hicks DJ, Bozov V, Glaspy JA, Rosen N, Solit DB, Netterville JL, Vnencak-Jones CL, Sosman JA, Ribas A, Zhao Z, Pao W. *BRAF*<sup>L597</sup> mutations in melanoma are associated with sensitivity to MEK inhibitors. *Cancer Discov*. 2012; 2:791–797. [PubMed: 22798288]
50. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R. Cancer Genome Project. Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell*. 2004; 116:855–867. [PubMed: 15035987]
51. Heidorn SJ, Milagre C, Whittaker S, Nourry A, Niculescu-Duvas I, Dhomen N, Hussain J, Reis-Filho JS, Springer CJ, Pritchard C, Marais R. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell*. 2010; 140:209–221. [PubMed: 20141835]
52. Sen B, Peng S, Tang X, Erickson HS, Galindo H, Mazumdar T, Stewart DJ, Wistuba I, Johnson FM. Kinase-impaired *BRAF* mutations in lung cancer confer sensitivity to dasatinib. *Sci Transl Med*. 2012; 4:136ra70.
53. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, Palmero R, Garcia-Gomez R, Pallares C, Sanchez JM, Porta R, Cobo M, Garrido P, Longo F, Moran T, Insa A, De Marinis F, Corre R, Bover I, Illiano A, Dansin E, de Castro J, Milella M, Reguart N, Altavilla G, Jimenez U, Provencio M, Moreno MA, Terrasa J, Muñoz-Langa J, Valdivia J, Isla D, Domine M, Molinier O, Mazieres J, Baize N, Garcia-Campelo R, Robinet G, Rodriguez-Abreu D, Lopez-Vivanco G, Gebbia V, Ferrera-Delgado L, Bombaron P, Bernabe R, Bearz A, Artal A, Cortesi E, Rolfo C, Sanchez-Ronco M, Drozdowskyj A, Queralt C, de Aguirre I, Ramirez JL, Sanchez JJ, Molina MA, Taron M, Paz-Ares L. Spanish Lung Cancer Group in collaboration with Groupe Français de

Pneumo-Cancérologie and Associazione Italiana Oncologia Toracica, Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): A multicentre, open-label, randomised phase 3 trial. *Lancet Oncol.* 2012; 13:239–246. [PubMed: 22285168]

54. Shaw AT, Kim DW, Nakagawa K, Seto T, Crinó L, Ahn MJ, De Pas T, Besse B, Solomon BJ, Blackhall F, Wu YL, Thomas M, O'Byrne KJ, Moro-Sibilot D, Camidge DR, Mok T, Hirsh V, Riely GJ, Iyer S, Tassell V, Polli A, Wilner KD, Jänne PA. Crizotinib versus chemotherapy in advanced *ALK*-positive lung cancer. *N Engl J Med.* 2013; 368:2385–2394. [PubMed: 23724913]
55. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA, Stratton MR. Patterns of somatic mutation in human cancer genomes. *Nature.* 2007; 446:153–158. [PubMed: 17344846]
56. Ney JT, Froehner S, Roesler A, Buettner R, Merkelbach-Bruse S. High-resolution melting analysis as a sensitive prescreening diagnostic tool to detect *KRAS*, *BRAF*, *PIK3CA*, and *AKT1* mutations in formalin-fixed, paraffin-embedded tissues. *Arch Pathol Lab Med.* 2012; 136:983–992. [PubMed: 22938585]
57. Querings S, Altmüller J, Ansén S, Zander T, Seidel D, Gabler F, Peifer M, Markert E, Stemshorn K, Timmermann B, Saal B, Klose S, Ernestus K, Scheffler M, Engel-Riedel W, Stoelben E, Brambilla E, Wolf J, Nürnberg P, Thomas RK. Benchmarking of mutation diagnostics in clinical lung cancer specimens. *PLoS One.* 2011; 6:e19601. [PubMed: 21573178]
58. Göke F, Perner S. Translational research and diagnostics in lung cancer. *Pathologe.* 2012; 33(Suppl 2):269–272. [PubMed: 22968733]
59. Schildhaus HU, Heukamp LC, Merkelbach-Bruse S, Riesner K, Schmitz K, Binot E, Paggen E, Albus K, Schulte W, Ko YD, Schlesinger A, Ansén S, Engel-Riedel W, Brockmann M, Serke M, Gerigk U, Huss S, Göke F, Perner S, Hekmat K, Frank KF, Reiser M, Schnell R, Bos M, Mattonet C, Sos M, Stoelben E, Wolf J, Zander T, Buettner R. Definition of a fluorescence in-situ hybridization score identifies high- and low-level *FGFR1* amplification types in squamous cell lung cancer. *Mod Pathol.* 2012; 25:1473–1480. [PubMed: 22684217]
60. Perner S, Wagner PL, Demichelis F, Mehra R, Lafargue CJ, Moss BJ, Arbogast S, Soltermann A, Weder W, Giordano TJ, Beer DG, Rickman DS, Chinnaiyan AM, Moch H, Rubin MA. *EML4-ALK* fusion lung cancer: A rare acquired event. *Neoplasia.* 2008; 10:298–302. [PubMed: 18320074]
61. Perner S, Demichelis F, Beroukhi R, Schmidt FH, Mosquera JM, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG, Kuefer R, Vessella R, Sun XW, Meyerson M, Lee C, Sellers WR, Chinnaiyan AM, Rubin MA. *TMPRSS2:ERG* fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res.* 2006; 66:8337–8341. [PubMed: 16951139]
62. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102:15545–15550. [PubMed: 16199517]

## Appendix

**Data analysis team:** Danila Seidel,<sup>1,2\*</sup> Thomas Zander,<sup>1,3,52\*</sup> Lukas C. Heukamp,<sup>8,52\*</sup> Martin Peifer,<sup>1\*</sup> Marc Bos,<sup>3,52\*</sup> Lynnette Fernández-Cuesta,<sup>1</sup> Frauke Leenders,<sup>1,2</sup> Xin Lu,<sup>1</sup> Sascha Ansén,<sup>3</sup> Masyar Gardizi,<sup>3,52</sup> Chau Nguyen,<sup>5,11</sup> Johannes Berg,<sup>5</sup> Prudence Russell,<sup>46</sup> Zoe Wainer,<sup>45</sup> Hans-Ulrich Schildhaus,<sup>8,52,76</sup> Toni-Maree Rogers,<sup>13</sup> Benjamin Solomon,<sup>44</sup> William Pao,<sup>38</sup> Scott L. Carter,<sup>42</sup> Gad Getz,<sup>42</sup> D. Neil Hayes,<sup>51</sup> Matthew D. Wilkerson,<sup>51,77</sup> Erik Thunnissen,<sup>24</sup> William D. Travis,<sup>26</sup> Sven Perner,<sup>14</sup> Gavin Wright,<sup>45,78</sup> Elisabeth Brambilla,<sup>36,37</sup> Reinhard Büttner,<sup>2,8,52</sup> Jürgen Wolf,<sup>2,3,52</sup> and Roman K. Thomas<sup>1,2,8,52</sup>

**Biospecimens core resources:** Franziska Gabler,<sup>2,4</sup> Ines Wilkening,<sup>4</sup> Christian Müller,<sup>1</sup> Ilona Dahmen,<sup>1</sup> Roopika Menon,<sup>14</sup> Katharina König,<sup>8,52</sup> Kerstin Albus,<sup>8,52</sup> Sabine Merkelbach-Bruse,<sup>8,52</sup> Jana Fassunke,<sup>8,52</sup> Katja Schmitz,<sup>8,52</sup> Helen Kuenstlinger,<sup>8,52</sup> Michaela A. Kleine,<sup>8,52</sup> Elke Binot,<sup>8,52</sup> Silvia Querings,<sup>2,4</sup> Janine Altmüller,<sup>6,10</sup> Ingelore Bäßmann,<sup>6</sup> Peter Nürnberg,<sup>6,9,10</sup> Peter M. Schneider,<sup>12</sup> and Magdalena Bogus<sup>12</sup>

**Pathology committee:** Reinhard Büttner,<sup>2,8</sup> Sven Perner,<sup>14</sup> Prudence Russell,<sup>46</sup> Erik Thunnissen,<sup>24</sup> William D. Travis,<sup>26</sup> and Elisabeth Brambilla<sup>36,37</sup>

**Biospecimens and data source sites:** Alex Soltermann,<sup>33</sup> Holger Moch,<sup>33</sup> Odd Terje Brustugun,<sup>31,32</sup> Steinar Solberg,<sup>30</sup> Marius Lund-Iversen,<sup>73</sup> Åslaug Helland,<sup>31,32</sup> Thomas

<sup>1</sup>Department of Translational Genomics, Center of Integrated Oncology Köln-Bonn, University of Cologne, 50931 Cologne, Germany.

<sup>2</sup>Laboratory of Translational Cancer Genomics, Center of Integrated Oncology Köln-Bonn, University of Cologne, 50937 Cologne, Germany.

\*These authors contributed equally to this work.

<sup>3</sup>Department I of Internal Medicine, Center of Integrated Oncology Köln-Bonn, University of Cologne, 50937 Cologne, Germany.

<sup>52</sup>Network Genomic Medicine, University Hospital Cologne, Center of Integrated Oncology Köln Bonn, 50937 Cologne, Germany.

<sup>8</sup>Institute of Pathology, Center of Integrated Oncology Köln-Bonn, University of Cologne, 50937 Cologne, Germany.

<sup>5</sup>Institute for Theoretical Physics, University of Cologne, 50937 Cologne, Germany.

<sup>11</sup>Bonn-Cologne Graduate School of Physics and Astronomy, 53115 Bonn, Germany.

<sup>46</sup>Department of Pathology, St. Vincent's Hospital, Melbourne, 3065 Victoria, Australia.

<sup>45</sup>The University of Melbourne Department of Surgery, St. Vincent's Hospital, Melbourne, 3065 Victoria, Australia.

<sup>76</sup>Institute of Pathology, University Hospital Goettingen, Goettingen, Germany.

<sup>13</sup>Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, 3002 Victoria, Australia.

<sup>44</sup>Department of Haematology and Medical Oncology, Peter MacCallum Cancer Centre, Melbourne, 3002 Victoria, Australia.

<sup>38</sup>Vanderbilt-Ingram Cancer Center, Nashville, TN, USA.

<sup>42</sup>The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.

<sup>51</sup>Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

<sup>77</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

<sup>24</sup>Department of Pathology, VU University Medical Center Amsterdam, 1007 MB Amsterdam, the Netherlands.

<sup>26</sup>Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.

<sup>14</sup>Department of Prostate Cancer Research at the Institute of Pathology, University Hospital of Bonn, 53127 Bonn, Germany.

<sup>78</sup>Division of Cancer Surgery, Peter MacCallum Cancer Centre, East Melbourne, 3002 Victoria, Australia.

<sup>36</sup>Department of Cancer Medicine, Institute Gustave Roussy, 94800 Villejuif, France.

<sup>37</sup>Pole de Cancérologie et Médecine Aigue Communautaire, CHU Grenoble, CS 10217, 38043 Grenoble, France and INSERM U823, Grenoble, France.

<sup>4</sup>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, 50931 Cologne, Germany.

<sup>6</sup>Cologne Center for Genomics, University of Cologne, 50931 Cologne, Germany.

<sup>10</sup>Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases, University of Cologne, Cologne, Germany.

<sup>9</sup>Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany.

<sup>12</sup>Institute of Legal Medicine, University of Cologne, 50823 Cologne, Germany.

<sup>33</sup>Institute for Surgical Pathology, University Hospital Zurich, 8091 Zurich, Switzerland.

<sup>31</sup>Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, N-0424 Oslo, Norway.



Muley,<sup>18,48</sup> Hans Hoffmann,<sup>18</sup> Philipp A. Schnabel,<sup>17,48</sup> Yuan Chen,<sup>15</sup> Harry Groen,<sup>22</sup> Wim Timens,<sup>23</sup> Hannie Sietsma,<sup>23</sup> Joachim H. Clement,<sup>71</sup> Walter Weder,<sup>33</sup> Jörg Sanger,<sup>19</sup> Erich Stoelben,<sup>7</sup> Corinna Ludwig,<sup>7</sup> Walburga Engel-Riedel,<sup>7</sup> Egbert Smit,<sup>24</sup> Danielle A. M. Heideman,<sup>24</sup> Peter J. F. Snijders,<sup>24</sup> Lucia Nogova,<sup>3,52</sup> Martin L. Sos,<sup>3,52,53</sup> Christian Mattonet,<sup>3,52</sup> Karin Topelt,<sup>3,52</sup> Matthias Scheffler,<sup>3,52</sup> Eray Goekkurt,<sup>52,55</sup> Rainer Kappes,<sup>52,56,62</sup> Stefan Kruger,<sup>52,56</sup> Kato Kambartel,<sup>52,57</sup> Dirk Behringer,<sup>52,58</sup> Wolfgang Schulte,<sup>52,59</sup> Wolfgang Galetke,<sup>52,60</sup> Winfried Randerath,<sup>52,61</sup> Matthias Heldwein,<sup>52,63</sup> Andreas Schlesinger,<sup>52,64</sup> Monika Serke,<sup>52,65</sup> Khosro Hekmat,<sup>52,63</sup> Konrad F. Frank,<sup>52,66</sup> Roland Schnell,<sup>52,67</sup> Marcel Reiser,<sup>52,68</sup> Ali-Nuri Hunerliturkoglu,<sup>52,69</sup> Stephan Schmitz,<sup>52,70</sup> Lisa Meffert,<sup>52,40</sup> Yon-Dschun Ko,<sup>52,40</sup> Markus Litt-Lampe,<sup>52,72</sup> Ulrich Gerigk,<sup>52,41</sup> Rainer Fricke,<sup>54</sup> Benjamin Besse,<sup>36</sup> Christian Brambilla,<sup>37</sup> Sylvie Lantuejoul,<sup>36,75</sup> Philippe Lorimier,<sup>36</sup> Denis Moro-Sibilot,<sup>37</sup> Federico Cappuzzo,<sup>27</sup> Claudia Ligorio,<sup>28</sup> Stefania Damiani,<sup>28</sup> John K. Field,<sup>29</sup> Russell Hyde,<sup>29</sup> Pierre Validire,<sup>35</sup> Philippe Girard,<sup>35</sup> Lucia A. Muscarella,<sup>47</sup> Vito M. Fazio,<sup>47,74</sup> Michael Hallek,<sup>2,3</sup> Jean-Charles Soria,<sup>34</sup> Scott L. Carter,

<sup>32</sup>Department of Oncology, Norwegian Radium Hospital, Oslo University Hospital, N-0310 Oslo, Norway.

<sup>30</sup>Department of Thoracic Surgery, Rikshospitalet, Oslo University Hospital, N-0027 Oslo, Norway.

<sup>73</sup>Department of Pathology, Norwegian Radium Hospital, Oslo University Hospital, N-0310 Oslo, Norway.

<sup>18</sup>Department of Thoracic Surgery, Thoraxklinik am Universitatsklinikum Heidelberg, 69126 Heidelberg, Germany.

<sup>48</sup>Translational Lung Research Center Heidelberg, Member of the German Center for Lung Research, Heidelberg, Germany.

<sup>17</sup>Institute of Pathology, University of Heidelberg, 69120 Heidelberg, Germany.

<sup>15</sup>Institute of Pathology, Jena University Hospital, Friedrich-Schiller-University, 07743 Jena, Germany.

<sup>22</sup>Department of Pulmonary Diseases, University of Groningen, University Medical Centre Groningen, 9713 GZ Groningen, the Netherlands.

<sup>23</sup>Department Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, 9713 GZ Groningen, the Netherlands.

<sup>71</sup>Department Hematology and Oncology, Jena University Hospital, Friedrich-Schiller University, 07747 Jena, Germany.

<sup>19</sup>Institute for Pathology Bad Berka, 99438 Bad Berka, Germany.

<sup>7</sup>Thoracic Surgery, Lungenklinik Merheim, Kliniken der Stadt Koln gGmbH, 51109 Cologne, Germany.

<sup>53</sup>Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA 94158, USA.

<sup>55</sup>Hematologisch-Onkologische Praxis Eppendorf, 20249 Hamburg, Germany.

<sup>56</sup>Klinik fur Pneumologie, Florence Nightingale Krankenhaus, 40489 Dusseldorf, Germany.

<sup>62</sup>Schwerpunktpraxis fur Lungen- und Bronchialheilkunde, Allergologie, Friedrichstr. 33-35, 40217 Dusseldorf, Germany.

<sup>57</sup>Lungenklinik, Krankenhaus Bethanien Moers, 47441 Moers, Germany.

<sup>58</sup>Klinik fur Hematologie, Onkologie und Palliativmedizin, Augusta Kliniken Bochum, 44791 Bochum, Lungenkrebszentrum Herne-Bochum, Germany.

<sup>59</sup>Abteilung fur Pneumologie, Malteser Krankenhaus, 53123 Bonn, Germany.

<sup>60</sup>Klinik fur Innere Medizin, Krankenhaus der Augustinerinnen, 50678 Koln, Germany.

<sup>61</sup>Klinik fur Pneumologie, Krankenhaus Bethanien Solingen, 42669 Solingen, Germany.

<sup>63</sup>Department of Cardiothoracic Surgery, University Hospital Cologne, Center of Integrated Oncology Koln Bonn, 50924 Cologne, Germany.

<sup>64</sup>Abteilung fur Pneumologie, Evangelisches Krankenhaus Kalk, 51103 Koln, Germany.

<sup>65</sup>Abteilung Pneumologie III/Thorakale Onkologie, Lungenklinik Hemer, 58675 Hemer, Germany.

<sup>66</sup>Department III of Internal Medicine, University Hospital Cologne, Center of Integrated Oncology Koln Bonn, 50924 Cologne, Germany.

<sup>67</sup>PIOH Frechen, Kolner Strae 9, 50226 Frechen, Germany.

<sup>68</sup>PIOH Koln, Richard Wagner Strae 13-17, 50674 Koln, Germany.

<sup>69</sup>Lukaskrankenhaus Neuss, Innere Medizin II, 41464 Neuss, Germany.

<sup>70</sup>Gemeinschaftspraxis fur Hematologie und Onkologie, Sachsenring 69, 50677 Koln, Germany.

<sup>40</sup>Klinik fur Internistische Onkologie Evangelische Kliniken Bonn, 53113 Bonn, Germany.

<sup>72</sup>Innere Medizin II Gastroenterologie/ Pneumologie/Onkologie, Marienhospital Bruhl, 50321 Bruhl, Germany.

<sup>41</sup>Klinik fur Thoraxchirurgie, Malteser Krankenhaus, 53123 Bonn, Germany.

<sup>54</sup>Epidemiologisches Krebsregister NRW gGmbH, 48149 Munster, Germany.

<sup>75</sup>Universite Joseph Fourier-INSERM U 823 Institut Albert Bonniot, 38700 La Tronche, France.

<sup>27</sup>Department of Medical Oncology, Ospedale Civile, 57100 Livorno, Italy.

<sup>28</sup>Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, 40138 Bologna, Italy.

<sup>29</sup>Roy Castle Lung Cancer Research Programme, The University of Liverpool Cancer Research Centre, University of Liverpool Cancer Research Centre, Department of Molecular and Clinical Cancer Medicine, Institute of Translational Medicine, The University of Liverpool, Liverpool L3 9TA, UK.

<sup>35</sup>France Service d'Anatomie-Pathologie, Institut Mutualiste Montsouris, 75014 Paris, France.

<sup>42</sup> Gad Getz,<sup>42</sup> D. Neil Hayes,<sup>51</sup> Matthew D. Wilkerson,<sup>51,77</sup> Viktor Achter,<sup>50</sup> and Ulrich Lang<sup>49,50</sup>

**Writing committee:** Danila Seidel,<sup>1,2\*</sup> Thomas Zander,<sup>1,3,52\*</sup> Lukas C. Heukamp,<sup>8,52\*</sup> Martin Peifer,<sup>1\*</sup> Marc Bos,<sup>3,52\*</sup> William Pao,<sup>38</sup> William D. Travis,<sup>26</sup> Elisabeth Brambilla,<sup>36,37</sup> Reinhard Büttner,<sup>2,8,52</sup> Jürgen Wolf,<sup>2,3,52</sup> and Roman K Thomas<sup>1,2,8,52</sup>

**Study leaders:** Reinhard Büttner,<sup>2,8,52</sup> Jürgen Wolf,<sup>2,3,52</sup> and Roman K. Thomas<sup>1,2,8,52</sup>

---

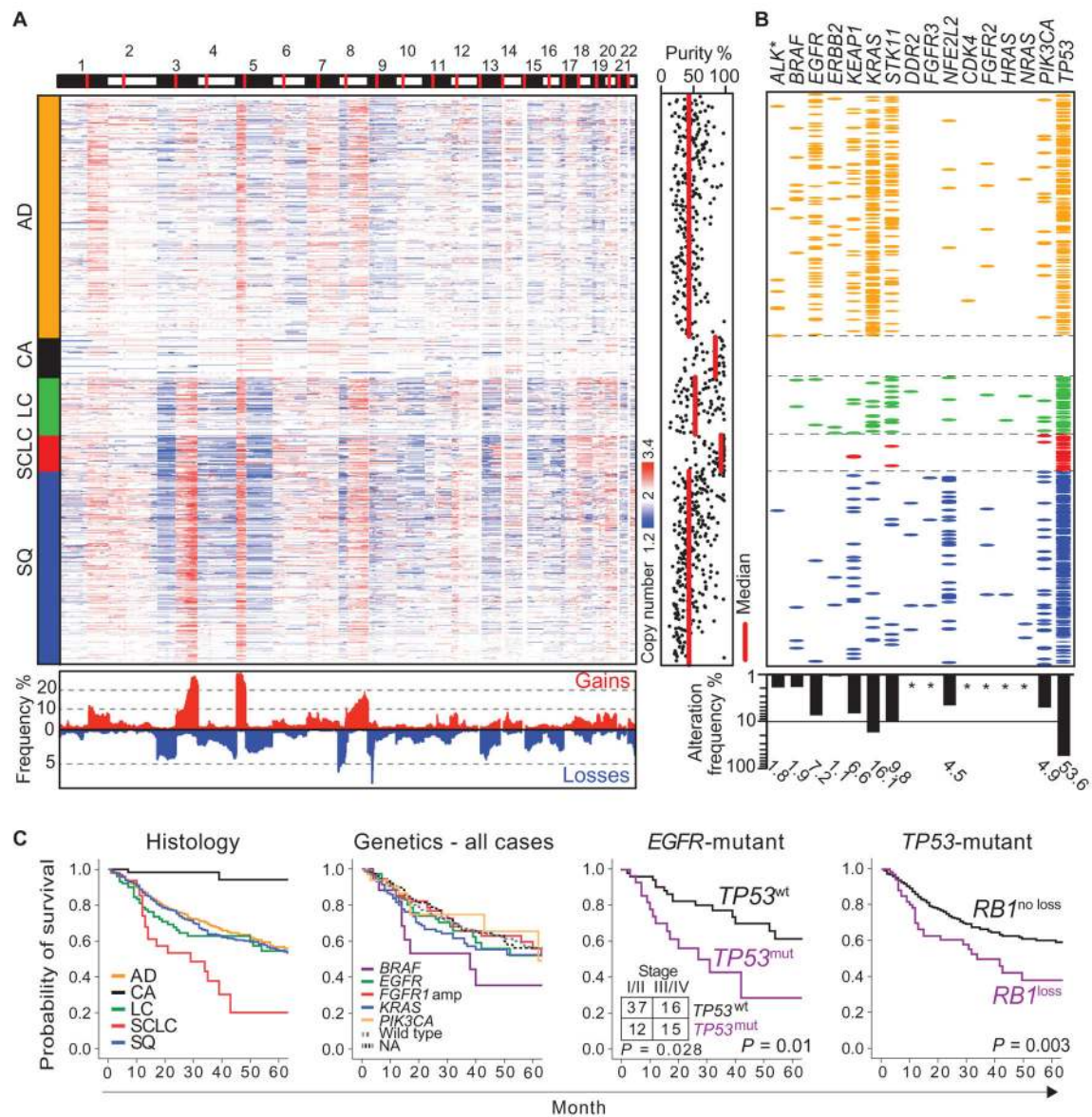
<sup>47</sup>Laboratory of Oncology, Istituto di Ricovero e Cura a Carattere Scientifico Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy.

<sup>74</sup>University Campus Bio-Medico of Roma, Center for Integrated Research, Laboratory for Molecular Medicine and Biotechnology, 00128 Rome, Italy.

<sup>34</sup>Phase I Unit—Department of Medicine, Institute Gustave Roussy, 94800 Villejuif, France.

<sup>50</sup>Computing Center, University of Cologne, 50931 Cologne, Germany.

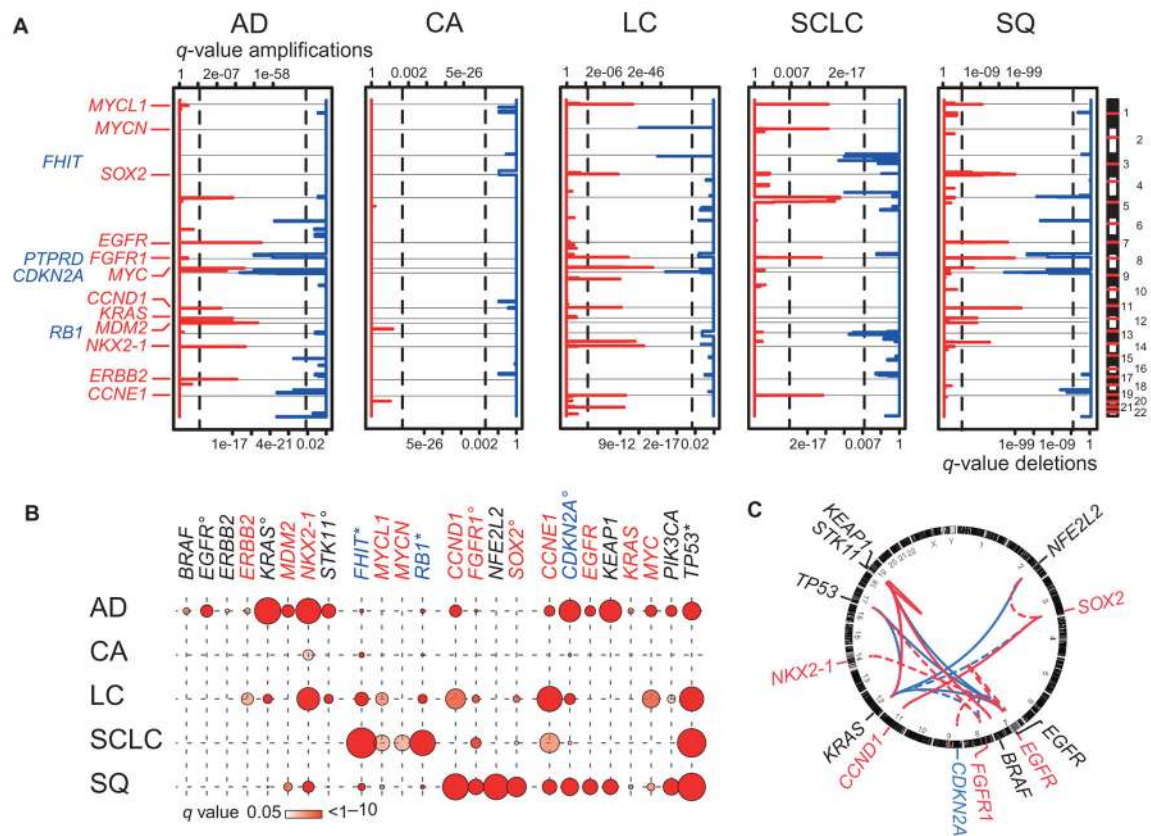
<sup>49</sup>Department of Informatics, University of Cologne, 50931 Cologne, Germany.



**Fig. 1. A global view of the lung cancer genome**

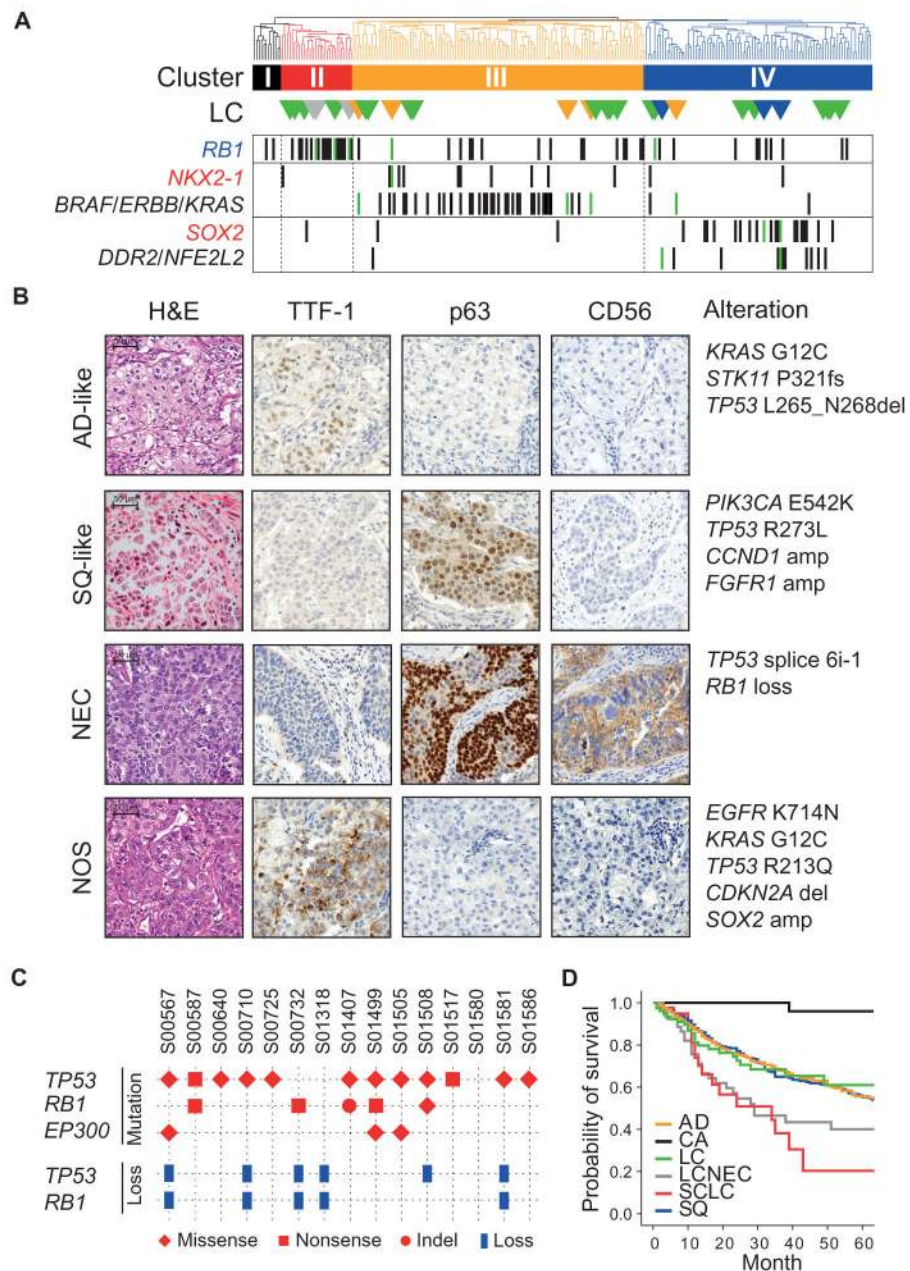
(A) Copy number profiles of lung cancer specimens of the major histological subtypes ( $n = 992$ ) (red, increases; blue, decreases) are plotted along the genome (horizontal axis: chromosomes as indicated, centromeres in red). Vertical colored bars on the left indicate lung cancer subtypes. Bottom: The frequencies (y axis) of copy number gains (red; cutoff, 2.7) and losses (blue; cutoff, 1.3) across all samples, calculated for adjoining 1-Mb fragments using segmented copy number data, are represented along the genome. Purity of tumor samples determined through SNP array-derived copy number data (26) is shown on the right, with the median purity calculated for each histological subgroup indicated in red. (B) Mutations and *ALK* rearrangements (*ALK*\*) are depicted per sample per gene as colored ellipses. Sample order was conserved from (A), and colors were chosen consistent with lung cancer subtypes. Total mutation frequencies per gene expressed as a percentage of all cases are shown as a bar graph at the bottom. Frequencies below 1% are marked with an asterisk. (C) Kaplan-Meier curves for overall survival are shown for the overall population per histological subtype (LC includes LCNEC), per genotype, for *EGFR*-mutant cases according to their *TP53* mutation status, and for *TP53*-mutant cases according to their *RB1* alteration status (from left to right) ( $P$  values for survival were calculated using the log-rank test). Numbers of cases with wild-type (wt) and mutant (mut) *TP53* in early stages (I

and II) and late stages (III and IV) are given for *EGFR*-mutant cases (inset; *P* value was calculated using the Pearson  $\chi^2$  test). Color code for histology: orange, AD; black, CA; green, LC (including LCNEC); red, SCLC; blue, SQ.



**Fig. 2. Genomic alterations in histological subgroups of lung cancer**

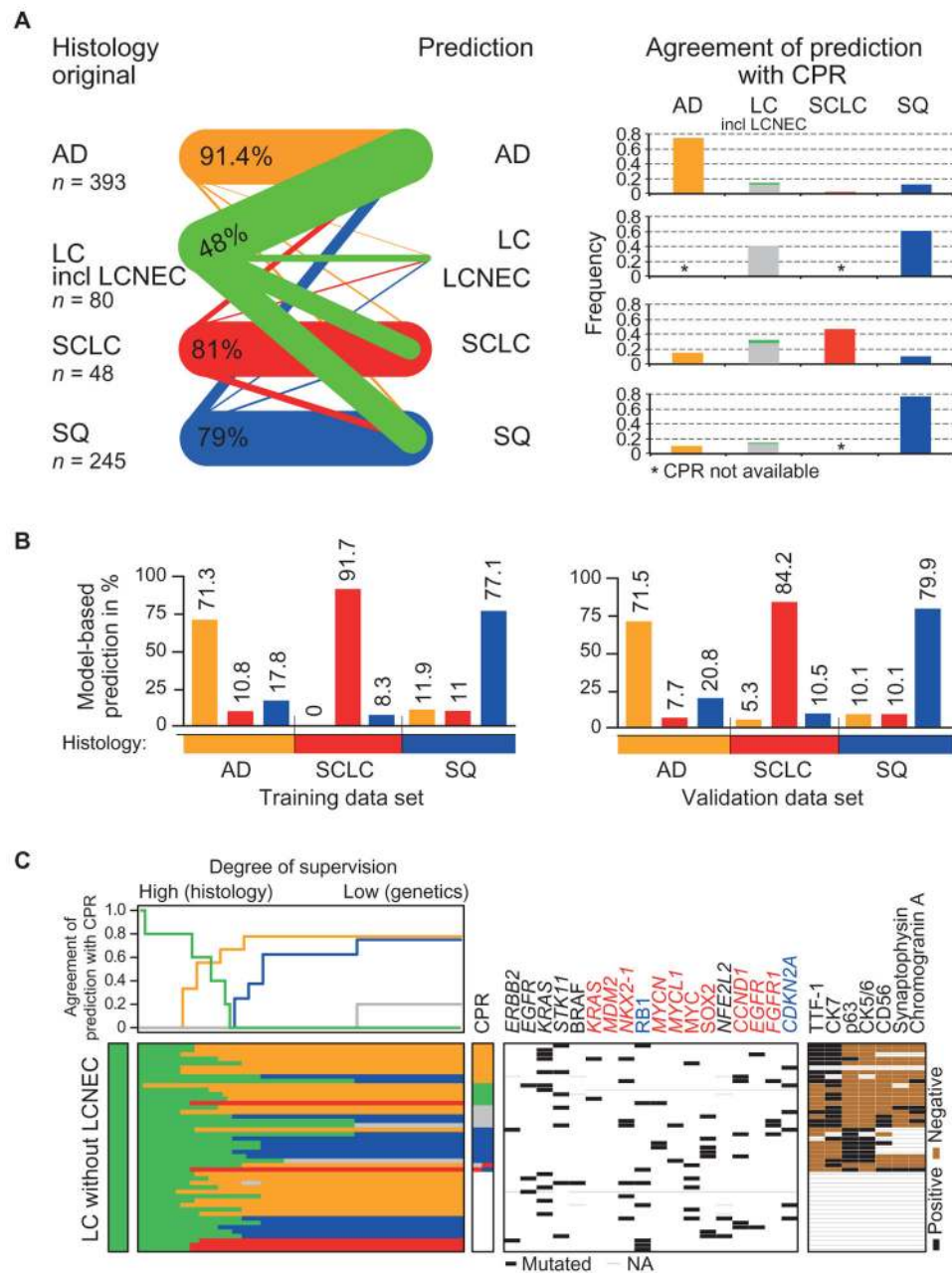
(A) Significantly amplified (red) and deleted (blue) regions calculated using a rank sum-based algorithm (24) are plotted along the genome (y axis) for the five major lung cancer subtypes [AD ( $n = 421$ ), CA ( $n = 69$ ), LC ( $n = 101$ ), SCLC ( $n = 63$ ), and SQ ( $n = 338$ )]. Statistical significance, expressed by  $q$  values (x axes: amplification, upper scale; deletion, lower scale), was computed for each genomic location. Known or potential oncogenes (red) or tumor suppressor genes (blue) are given at respective locations. Vertical lines indicate level of significance of  $q = 0.01$ . (B) Frequencies of significant genomic alterations are given per gene per histological subtype. Colors of gene names are encoded as follows: red, amplified; blue, deleted; and black, mutated. Frequencies of alterations correspond to circle size [frequencies of deletions of *FHIT* and *RB1* and mutations in *TP53* were adapted by dividing values by three (asterisks); frequencies of mutations in *EGFR*, *KRAS*, and *STK11*, of deletions in *CDKN2A*, and of amplifications in *FGFR1* and *SOX2* were adapted by dividing values by two (circles)]. Significant mutations were determined using a binomial test with a background mutation rate of 0.5%.  $P$  values were adjusted for multiple hypothesis testing using the Benjamini and Hochberg method across each histological subtype.  $q$  values of significant results ( $q < 0.05$ ) are indicated by the color code of the symbols (color key provided below the chart). (C) Associations of copy number alterations and mutations calculated using Fisher's exact test followed by Benjamini and Hochberg adjustment are represented with a Circos plot. Involved genes are named at corresponding genomic locations (copy number gains in red, copy number deletions in blue, and mutations in black) outside the ring representing the genome. Internal lines show significant co-occurring (red) and mutually exclusive (blue) events ( $q < 0.05$ ) between two copy number alterations or two frequently mutated genes (solid lines) or between a copy number alteration and a mutation (dashed lines) found in lung cancer.



**Fig. 3. Genetic features typical of other lung cancer subtypes in LC**

(A) Unsupervised hierarchical clustering using 294 highly variable ( $SD/mean > 2.1$ ) expressed genes identified four gene expression subgroups containing mainly CA (I), SCLC (II), AD (III), and SQ (IV). LC samples are indicated as triangles at corresponding positions below the cluster dendrogram. They are colored orange if they have AD-specific alterations, blue if they have SQ-specific alterations, gray if the case was initially diagnosed as an LCNEC, and green if they have no known alteration. Genetic alterations (label: red, amplified; blue, deleted; black, mutated; *ERBB* includes mutation in *EGFR* or *ERBB2*) are given for selected genes per sample as vertical lines (LC cases in green; others in black). (B) Typical immunohistochemistry is shown for LC specimens with immunohistochemical and genetic characteristics of AD (AD-like), SQ (SQ-like), and NEC, as well as LC lacking features of other lung cancer subtypes (NOS, not otherwise specified). The corresponding genetic alterations are indicated on the right. H&E, hematoxylin and eosin. (C) Distribution of mutations (in red, symbols according to type of

mutation: diamond for missense, square for nonsense, and circle for indel) and copy number loss (in blue) of *TP53*, *RBI*, and *EP300* across all whole exome-sequenced LCNECs. (D) Overall survival corresponding to each histological lung cancer subtype, with LC separated into LCs with neuroendocrine (gray) and without neuroendocrine (green) features.

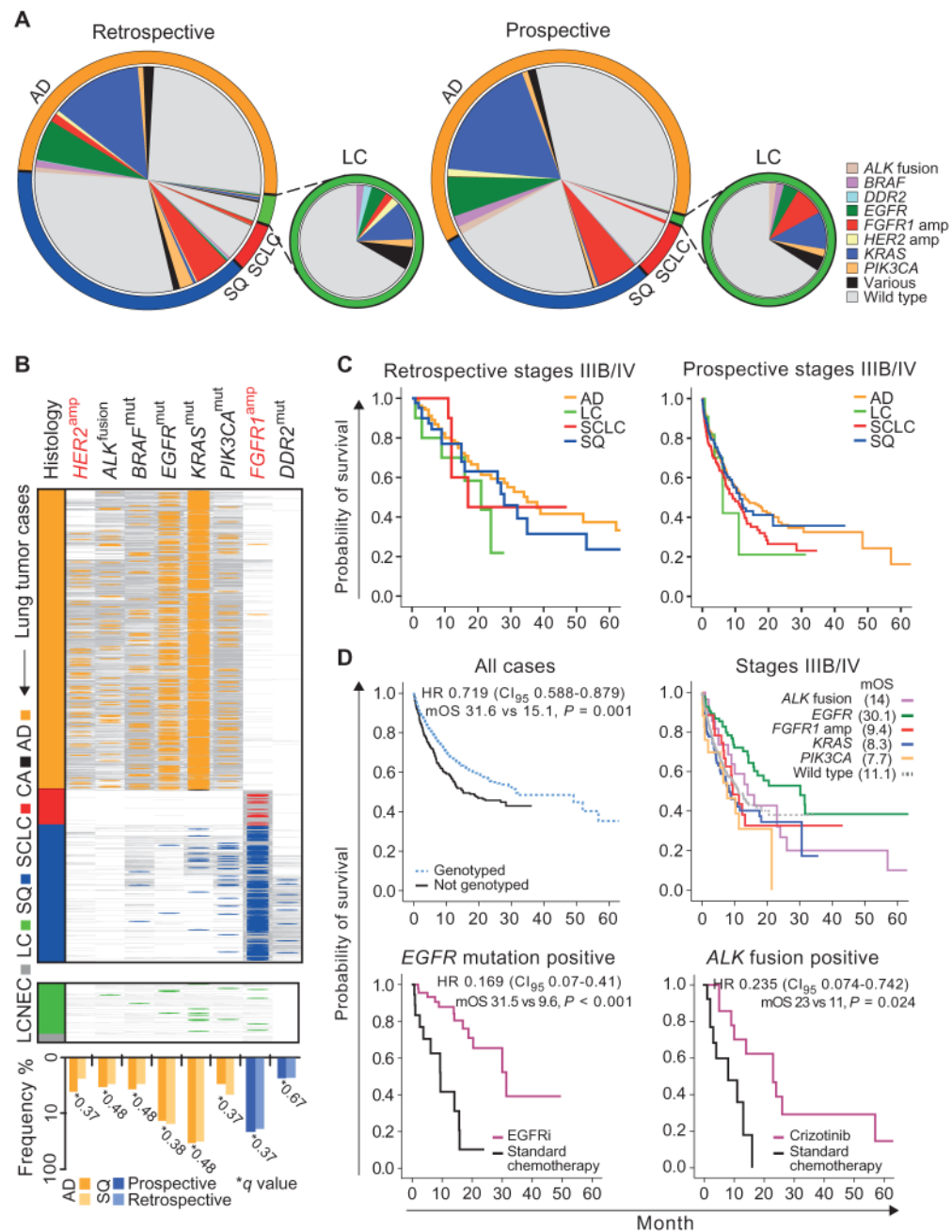


**Fig. 4. Genomics-based classification of lung cancer**

(A) Semisupervised reclassification of lung tumor samples. The relative proportion of cases per histological subtype (left; the LC group includes LCNEC cases) that were reclassified on the basis of 18 genetic alterations (table S11) to a certain subgroup (labels in the middle) is illustrated as lines. The weight of the lines is proportional to the fraction of cases classified to the respective subgroup. All cases that were predicted to be LC were histological LCNEC. Bars in the right graph give the concordance of each predicted class with the central pathological review (CPR). Subtypes for which no CPR was available are denoted with asterisks. (B) Supervised in silico classification of lung cancer specimens based on genetic features for 637 tumor samples with at least one genetic alteration and validation of the classifier in independent data sets of all three subgroups (20, 21, 25). Original histological subtypes defined groups for supervised learning. Bars indicate classification frequencies relative to the original histology. Classification results for the CLCGP data set are shown on the left, and results of the three validation data



sets are on the right. (C) Semi-supervised genetics-based reclassification of LC specimens without neuroendocrine features. For each sample (rows), prediction to a certain subtype (color per row in accordance to the color code used for histological subtypes, see below) is given (lower graph). Degree of supervision ( $x$  axis, upper graph) decreases continuously from left to right, the farthest right representing a genetics-based prediction. Agreement of the prediction with the CPR is plotted for each stage of supervision (upper part). Detailed information is given in Supplementary Materials and Methods. Genome alterations (black lines) and immunohistochemistry results (black, positive; brown, negative; thin gray line, not available) are indicated for each sample (middle and right panels). Genes are sorted according to their predictive value for histological subtypes. For the cases in the lower part of the figure, immunohistochemistry was not performed. Color code for predicted classes and CPR: orange, AD; black, CA; green, LC; gray, LCNEC; red, SCLC; blue, SQ; combination of colors, mixed subtype; white, no CPR.



**Fig. 5. Clinically relevant genome alterations in lung cancer subtypes**

(A) Genetic alterations per histological subtype in retrospective and prospective sample sets. Each chart represents the overall population with proportions of histological subtypes color-coded in the outer ring. Frequencies of alterations (wild type: no alteration in *ALK*, *EGFR*, *FGFR1*, *KRAS*, or *PIK3CA*) are given per gene relative to all cases within each histological subtype. Distribution of alterations for the LC population is shown separately. (B) Genotyping results of 3590 patients enrolled in the prospective screening effort are sorted according to the histological subtype [AD (2250), CA (3), SCLC (265), SQ (1018), LC (47), and LCNEC (7)]. Colored lines indicate alterations, and gray lines indicate wild type. Frequencies of alterations for AD (orange) and SQ (blue) are plotted below the respective genes, comparing the mutation frequency in the prospective data set (dark colors) to the retrospective data set (light colors). No significant difference between the data sets was seen (*q* values are

given on the graph; *P* values were adjusted for multiple hypothesis testing using the Benjamini and Hochberg method). Color code: orange, AD; black, CA; green, LC; gray, LCNEC; red, SCLC; blue, SQ. (C) Kaplan-Meier curves for overall survival are shown for stage IIIB/IV patients per histological subtype for the retrospective (left) and prospective (right) sample sets. No significant difference was seen between subtypes within each data set. (D) Prospective sample set: Kaplan-Meier curves for overall survival are shown for all patients who were genetically tested versus those without available genetic information (top left). Overall survival is shown for stage IIIB/IV patients with alterations in given genes versus patients with wild type in the given genes (top right). Overall survival was statistically significantly longer in *EGFR*-mutant cases compared to all other (log-rank test,  $P < 0.05$ ) except *ALK*-rearranged cases ( $P = 0.065$ ). Overall survival is shown for patients with *EGFR* mutation treated with an EGFR inhibitor or standard chemotherapy (bottom left) and patients with *ALK* translocations treated with crizotinib or standard chemotherapy (bottom right). Gain of overall survival in the patient group treated with kinase inhibitors versus standard chemotherapy is given by the median overall survival (mOS). *P* values were corrected using the Bonferroni adjustment. HR, hazard ratio.

**Table 1**  
**Clinical characteristics of lung cancer patients in the retrospective [Clinical Lung Cancer Genome Project (CLCGP)] and prospective (NGM) data sets**

UICC, Union Internationale Contre le Cancer.

		Retrospective CLCGP	Prospective NGM
Total <i>n</i>		1255	3863
Age at diagnosis	<i>n</i>	1225	3019
	Median (range)	65 (18–94)	67 (28–96)
Sex, <i>n</i> (%)	Female	419 (33.5)	1397 (37.2)
	Male	831 (66.5)	2354 (62.8)
	Unknown	5	112
Histology (WHO 2004), <i>n</i> (%)	Adenocarcinoma	537 (44.5)	2250 (62.7)
	Carcinoid	71 (6)	3 (0.08)
	Large cell carcinoma *	129 (10.7)	54 (1.5)
	Small cell carcinoma	65 (5.4)	265 (7.4)
	Squamous cell carcinoma	403 (33.4)	1018 (28.3)
	Other/unknown	50	273
Stage (UICC), <i>n</i> (%)	IA	305 (25.4)	105 (10.4)
	IB	316 (26.3)	72 (7.1)
	IIA	33 (2.8)	56 (5.5)
	IIB	190 (15.8)	49 (4.8)
	IIIA	195 (16.2)	131 (13.0)
	IIIB	89 (7.4)	115 (11.4)
	IV	74 (6.1)	479 (47.5)
	Unknown	53	2856
Survival, median no. of months	I	152	Not reached
	II	66	49 <sup>†</sup>
	III	34	28
	IV	18	9.5
	Unknown	58	11.6
Smoking history, <i>n</i> (%)	Current/former	896 (86)	102 (89)
	Never	147 (14)	13 (11)
	Unknown	212	3748

\* Including LCNEC.

<sup>†</sup> Median follow-up time for stage II patients was 16.3 months. Because of short median follow-up time for stage II patients, accuracy of median overall survival in this stage was low (fig. S19).