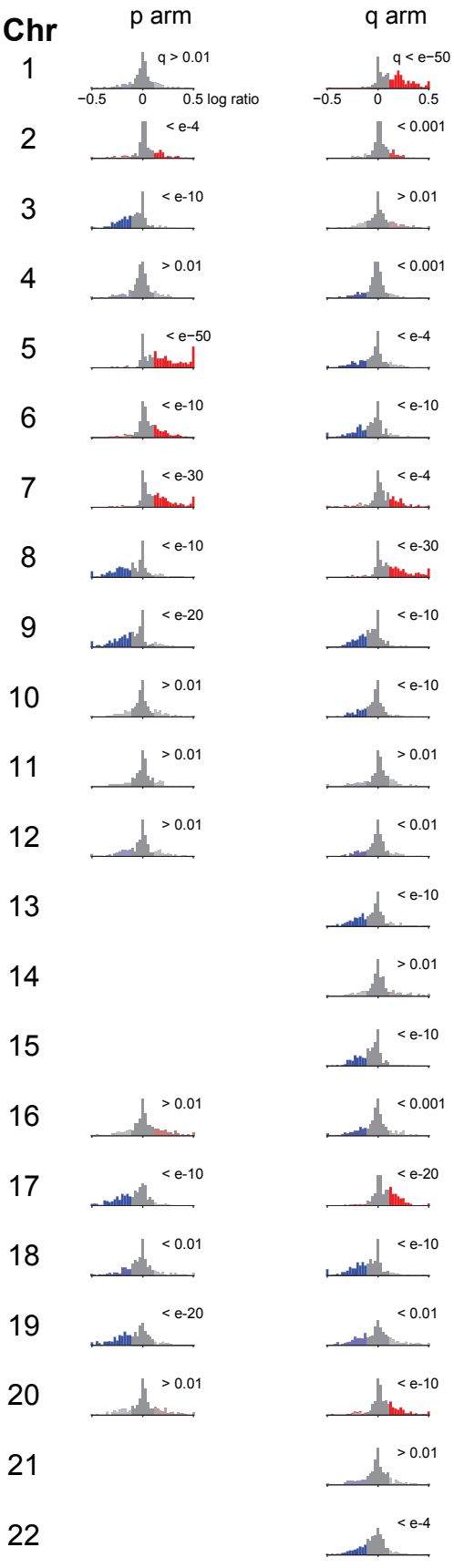
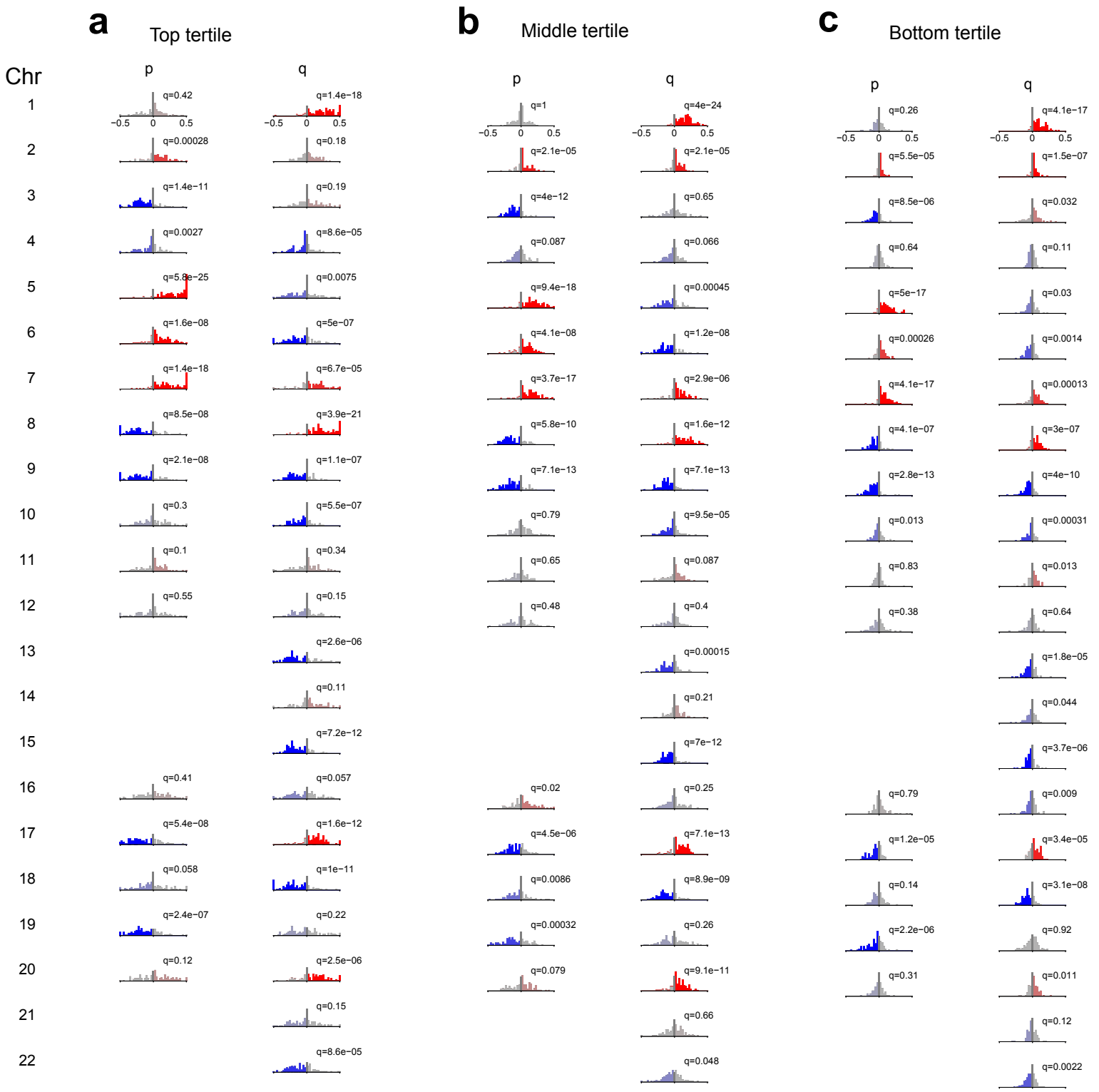
Characterizing the cancer genome in lung adenocarcinoma

Supplementary information

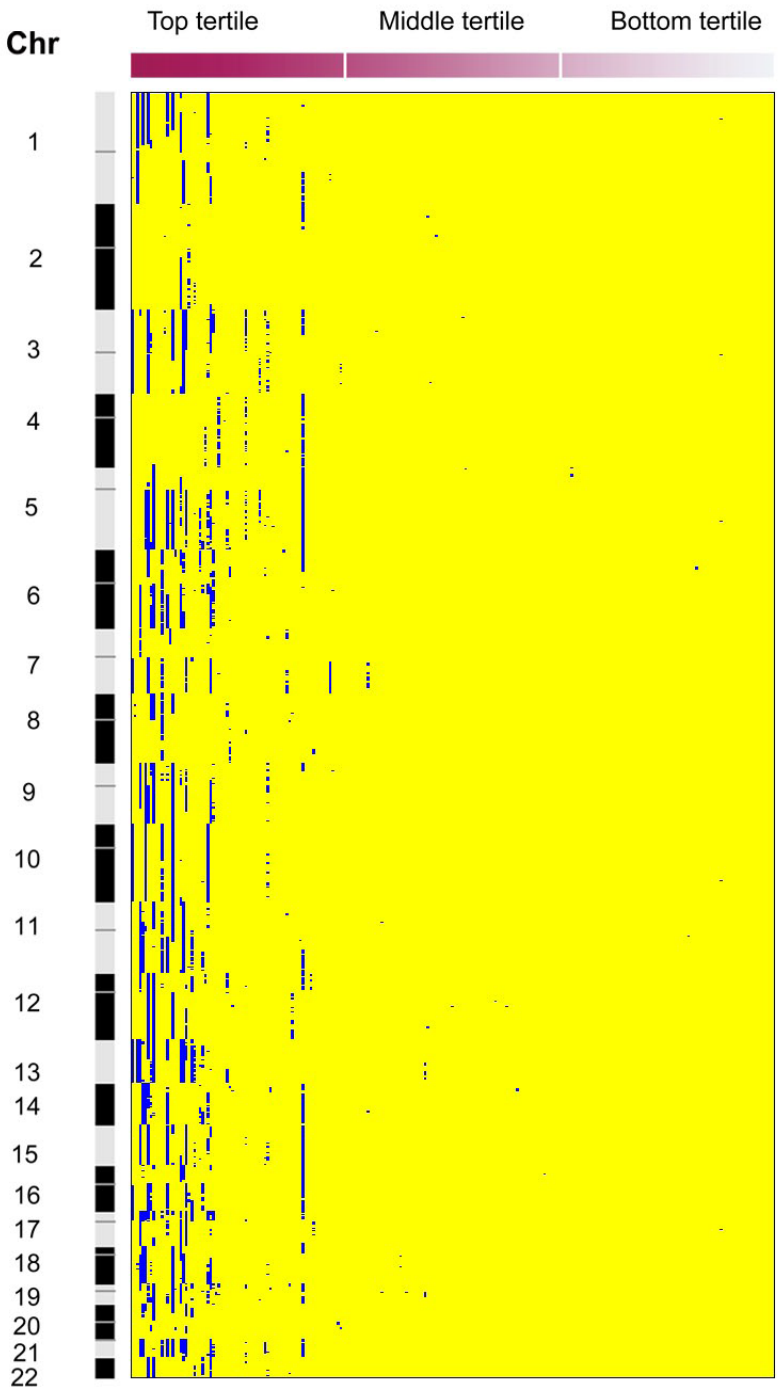
Table of contents



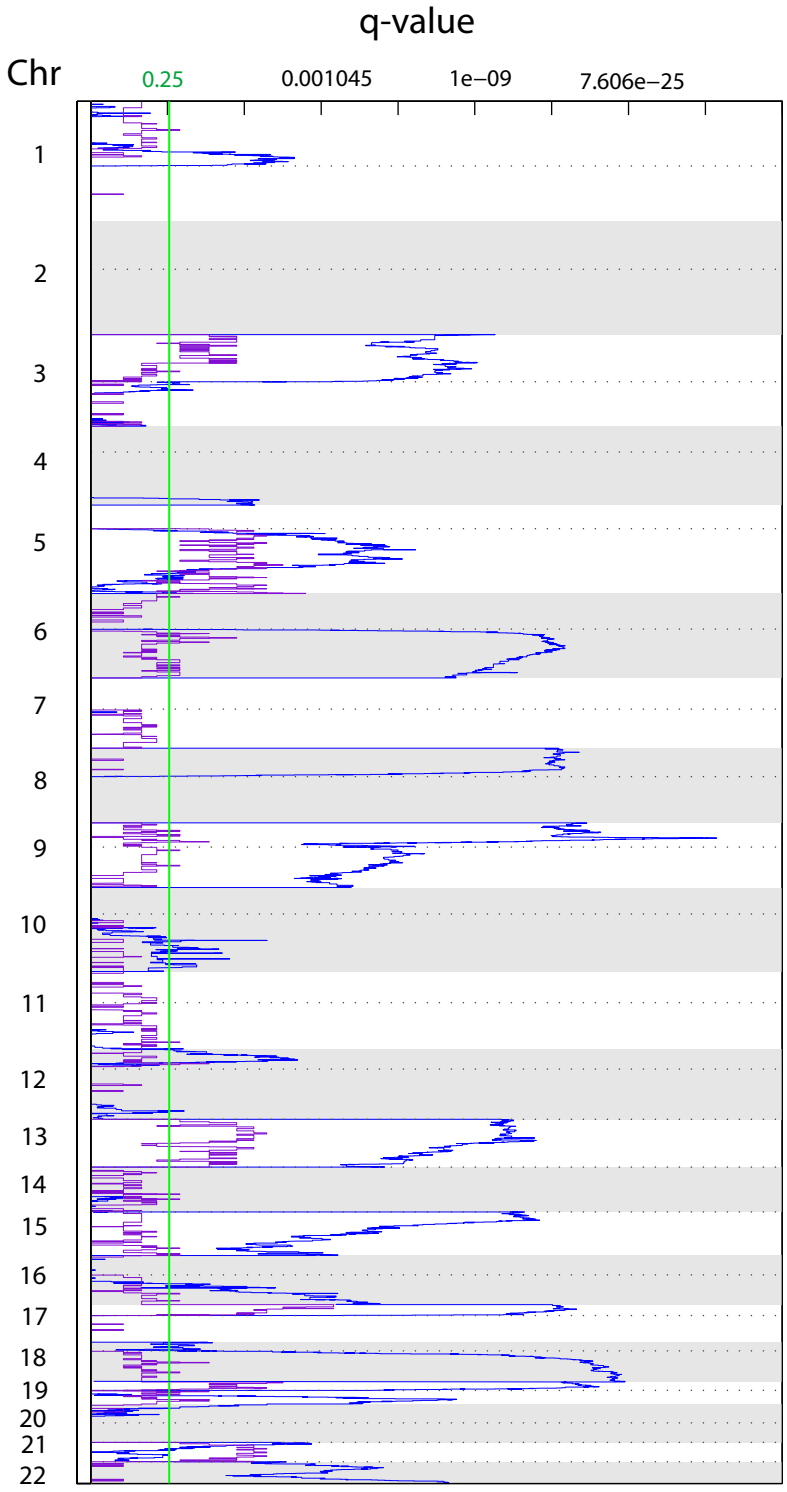
Supplementary Figure 1. Histograms of median copy number (log 2 ratio from -0.5 to 0.5, x-axis) are plotted for each arm; y axis values are trimmed at a value of 15. An imbalance toward amplification (red bars) or deletion (blue bars) is shown by the shift in histogram mass.



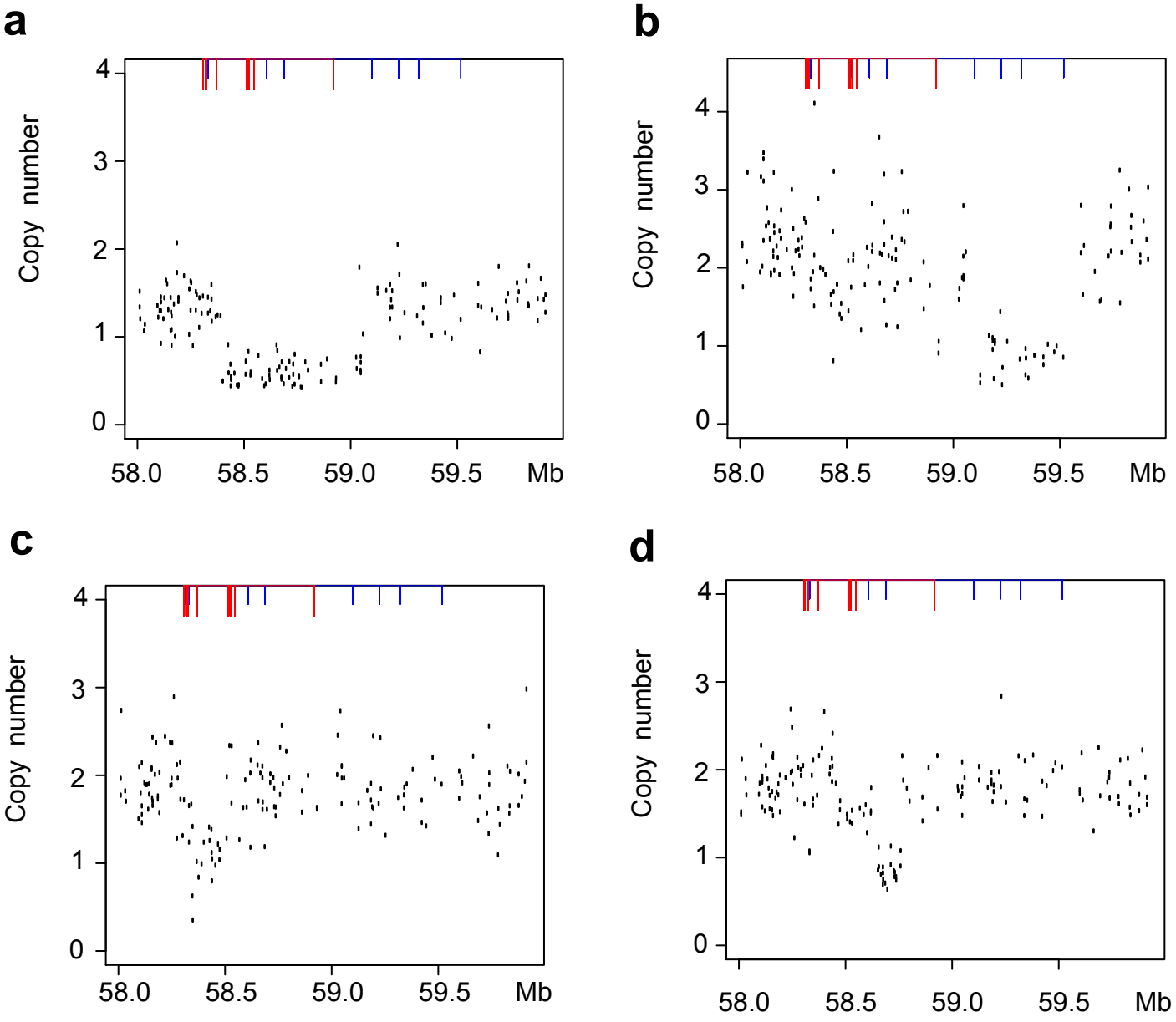
Supplementary Figure 2. Comparison of arm level events between sample tertiles. Histograms of median copy number (log 2 ratio from -0.5 to 0.5, x-axis) for each arm is plotted for the **a)** top tertile, **b)** middle tertile and **c)** bottom tertile. An imbalance toward amplification is shown by the histogram mass shifted toward the right (red bars), an imbalance toward deletion is shown by a shift to the left (blue bars). Amplification or deletion of an arm across the dataset was tested for significance by a 2-sided binomial test, after removing values between +/- 0.0125. p values were FDR corrected to give a false discovery rate q value, significance is set to a q value of 0.01.



Supplementary Figure 3. Inferred LOH for 237 tumor/normal sample pairs. Samples are sorted by degree of interchromosomal variation and divided into 3 groups (shown by shaded boxes along the top and designated top tertile, middle tertile and bottom tertile). Loss of heterozygosity is colored blue, rention of heterozygosity is colored yellow.



Supplementary Figure 4. Statistical analysis of LOH (purple) and copy loss (blue; copy number threshold =1.87) for 237 lung adenocarcinomas. False discovery rates (q values) for each alteration (x axes) are plotted at each genome position (y axis). Genomic positions corresponding to even-numbered chromosomes are shaded; dotted lines indicate the centromeres. The green line represents the q value cutoff (0.25) for significance.



Supplementary Figure 5: *PDE4D* gene deletions. **a), b), c)** and **d)** The copy number (x-axis) of four samples with *PDE4D* deletions are plotted from chromosome 5 position ~58 to 60 Mb (y-axis). Exons of all possible isoforms of the *PDE4D* gene are shown in blue at the top of the plots, the RefSeq entry for *PDE4D* is shown in red.

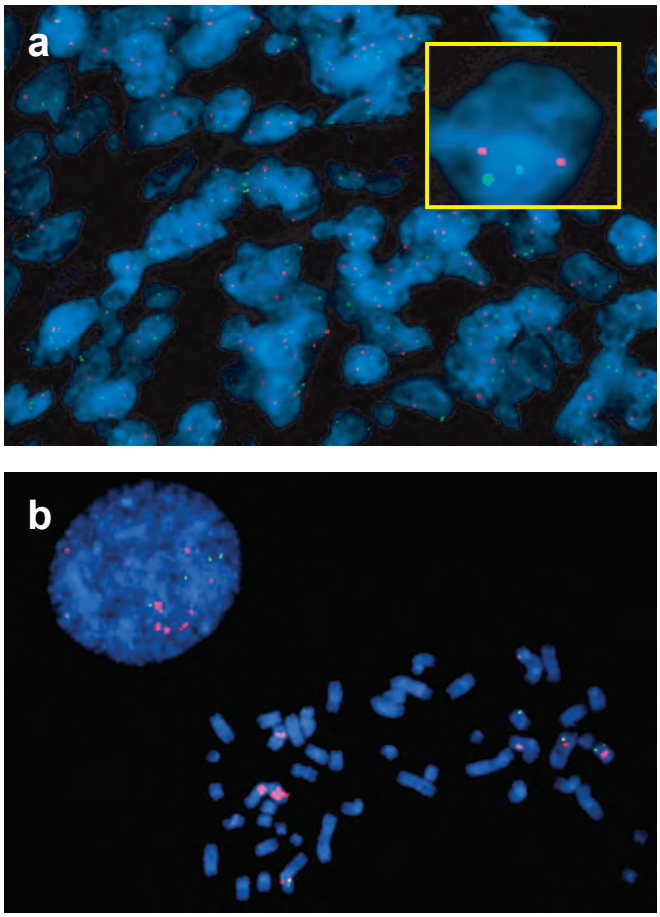
a P1810R

PTPRA / 686-710	SRQIRQFHFHGWPEV-GIPSDGKGMI
PTPRE / 584-608	VRVVRQFHFHGWPEI-GIPAEGKGMI
PTPRD / 1798-1822	SRTVRQFQFTDWPEQ-GVPKSGEGFI
PTPRD with mutation / 1798-1822	SRTVRQFQFTDW**R**EQ-GVPKSGEGFI
PTPRF / 1792-1816	SRTIRQFQFTDWPEQ-GVPKTGEGFI
PTPRC / 112-1136	SRTVYQYQYTNWSVE-QLPAEPKELI
PTPRU / 1333-1358	HLLVRHFQFLRWSAYRDTPDSKKAFL
PTPRG / 1308-1333	VLEVRHFQCPKWPNPDAPISSTFELI
PTPRH / 973-997	TLSVRQFHYQAWPDH-GVPSSPDTLL
PTPRB / 1857-1881	HRLIRHFHYTVWPDH-GVPETTQSLI

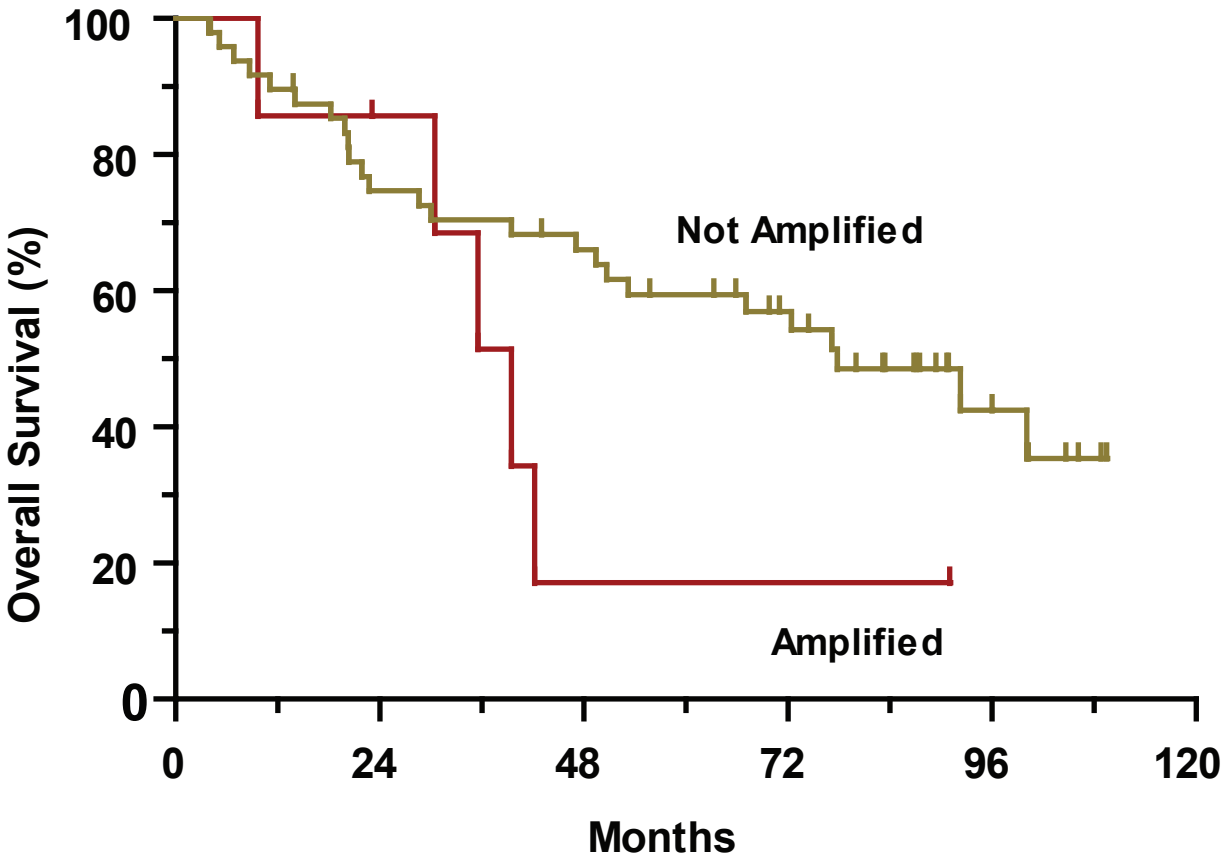
b R1537L

PTPRA / 407-445	SWPDFGVPFTPIGMLKFLKKVKACNPQYAGAIVVHCSAG
PTPRE / 300-338	SWPDFGVPFTPIGMLKFLKKVKTLNPVHAGPIVVHCSAG
PTPRD / 1519-1557	AWPDHGVPEHPTPFLAFLRRVKTCNPPDAGPMVVHCSAG
PTPRD with mutation / 1519-1557	AWPDHGVPEHPTPFLAFL**L**RVKTCNPPDAGPMVVHCSAG
PTPRF / 1513-1551	AWPDHGVPEYPTPILAFLRRVKACNPLDAGPMVVHCSAG
PTPRC / 816-854	SWPDHGVPEDPHLLLKLRRRVNAFSNFFSGPIVVHCSAG
PTPRU / 1050-1088	AWPEHGVPYHATGLLAFIRRVKASTPPDAGPIVIHCSAG
PTPRG / 1025-1063	QWPDMGVPEYALPVLTFVRRSSAARMPETGPVLVHCSAG
PTPRH / 700-736	VGGQRGSQDRSSCGEAVS--VLGLGPARSYPATITTIWD
PTPRB / 1568-1606	DGPLKPHTAYRISIRAFTQLFDEDLKEFTKPLYSDTFFS

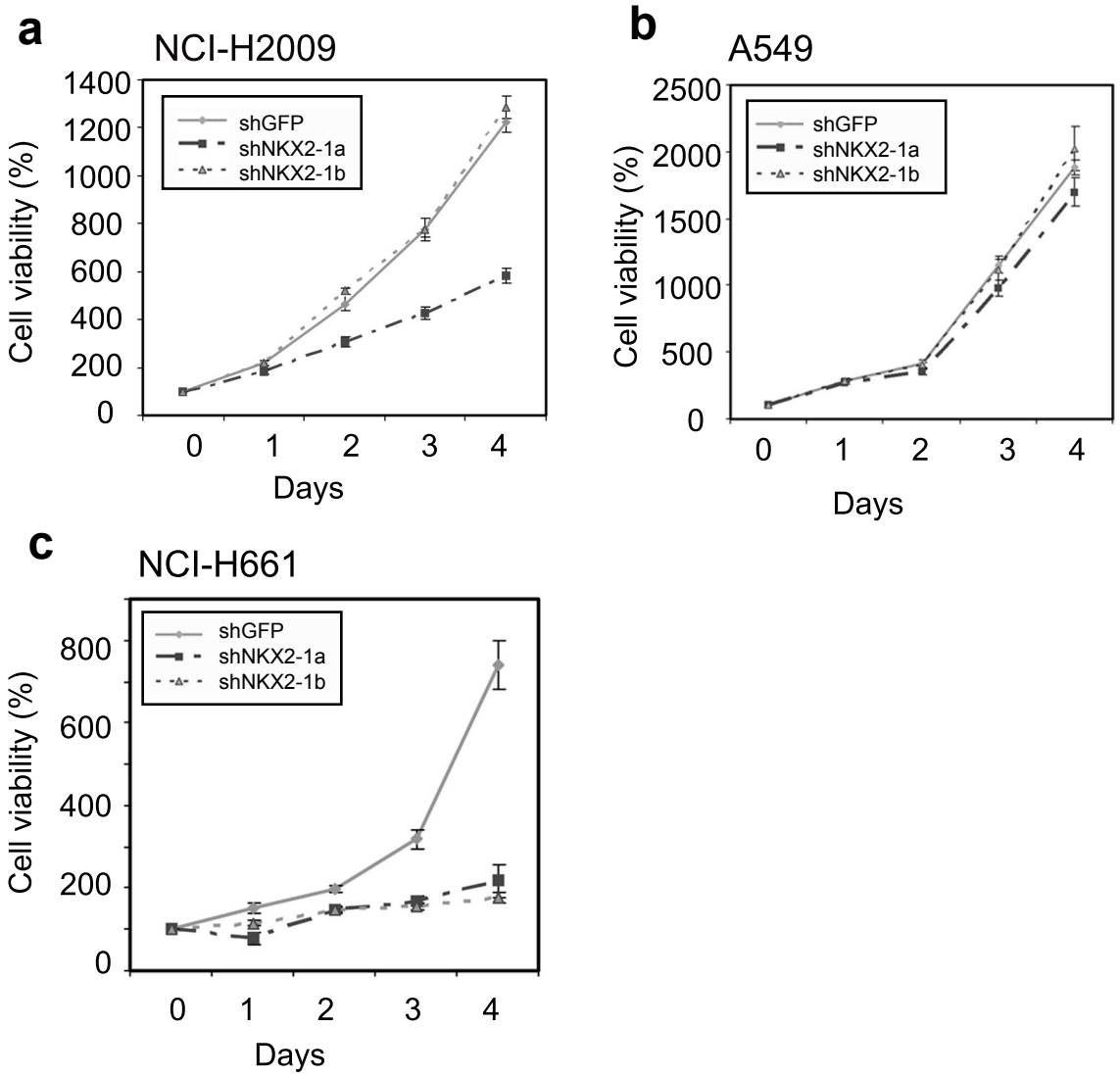
Supplementary Figure 6. Alignment of the sequence of two mutations, **a)** P1810R and **b)** R1537L within the tyrosine phosphatase domain with other protein tyrosine phosphatase receptor genes. Mutated residues shown in red, Ensembl transcript ENST00000356435 was used for annotating the mutations.



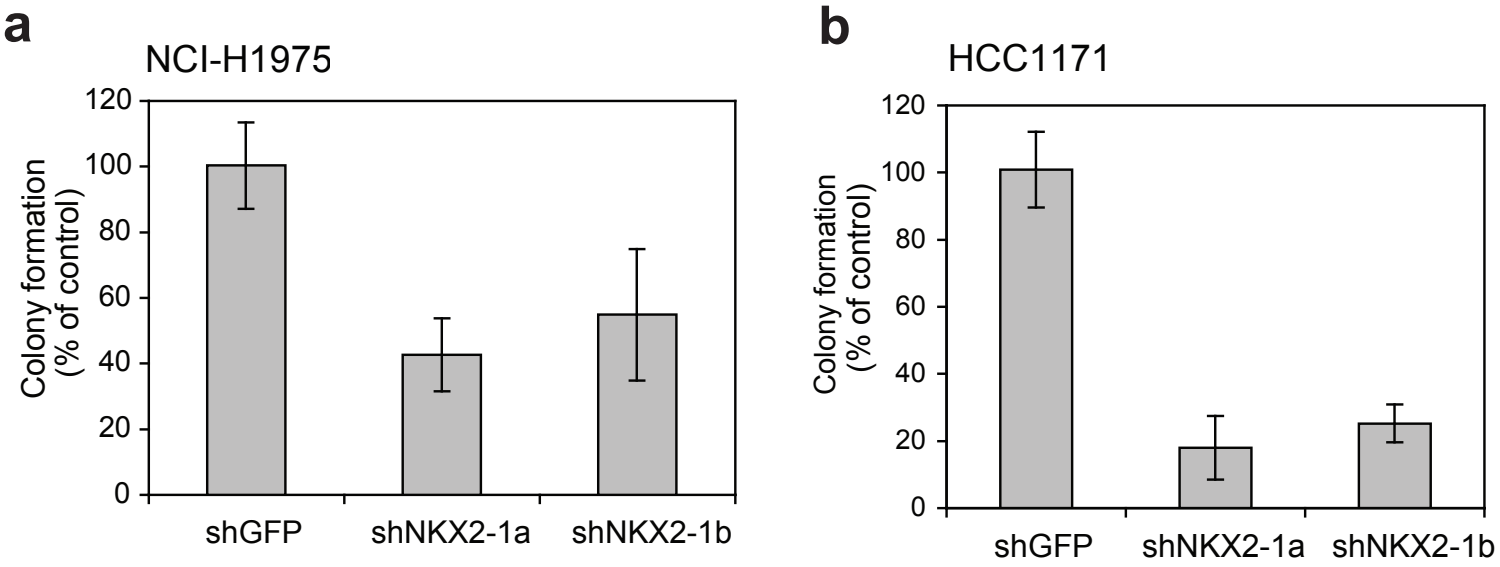
Supplementary Figure 7. Validation of *MBIP/NKX2-1* amplification by fluorescence *in situ* hybridization (FISH). FISH for *NKX2-1* (red) and a chromosome 14 reference probe (green) **a)** on lung adenocarcinoma tissue without amplification and **b)** of the NCI-H1819 cell line showing both a metaphase and interphase nucleus are shown. Nuclei are stained with DAPI (blue) and the yellow boxed areas show a single nucleus.



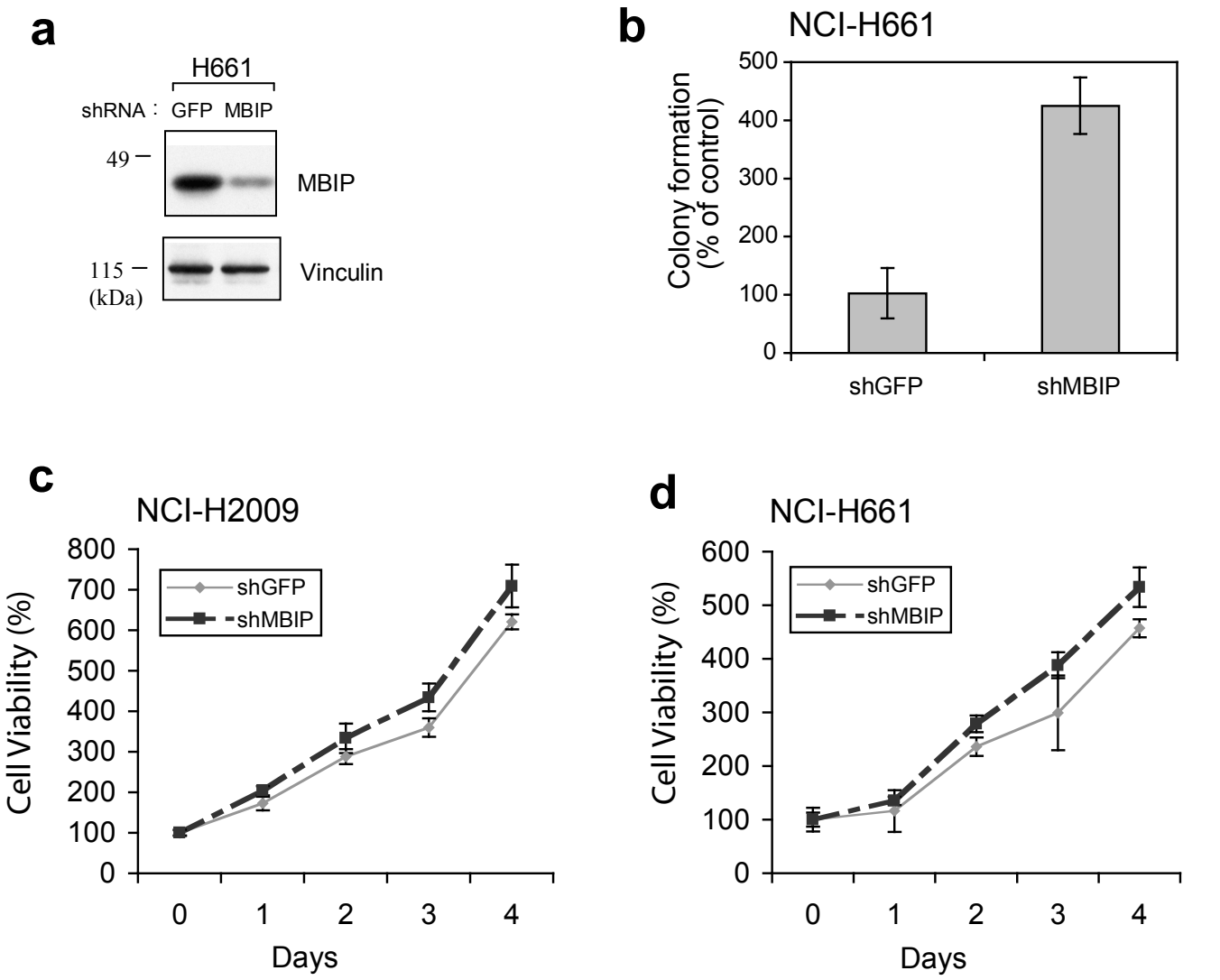
Supplementary Figure 8: Overall survival for patients with lung adenocarcinoma with and without *NKX2-1* amplification (p=0.15). The median survival for patients with *NKX2-1* amplification was 39.5 months and the median survival for patients with no *NKX2-1* amplification was 77.8 months.



Supplementary Figure 9. Functional consequence of *NKX2-1* down-regulation. WST measurements of cell viability in the **a)** NCI-H2009, **b)** A549 and **c)** NCI-H661 cell line.



Supplementary Figure 10. NKX2-1 knockdown leads to reduced anchorage-independent growth of NCI-H1975 and HCC1171 cells. **a)** Soft agar colony formation by NCI-H1975 cells expressing the indicated shRNA is shown relative to the shGFP control as a mean percentage (+/- standard deviation in triplicate samples; $p = 1.1 \times 10^{-5}$ when comparing shGFP to shNKX2-1a and p=0.00135 when comparing to shNKX2-1b) **b)** Soft agar colony formation by HCC1171 cells expressing the indicated shRNA is shown relative to the shGFP control as a mean percentage (+/- standard deviation in triplicate samples, n=1; $p = 7.1 \times 10^{-4}$ when comparing shGFP to shNKX2-1a and p = 0.0021 when comparing to shNKX2-1b).



Supplementary Figure 11: MBIP RNAi leads to increased colony formation and has no effect on the viability of NCI-H2009 and NCI-H661 cells. **a)** Anti-MBIP and anti-vinculin immunoblots of lysates from NCI-H661 cells expressing the MBIP shRNA. **b)** Soft agar colony formation by NCI-H661 MBIP knockdown cells ($p = 6.2 \times 10^{-7}$ when comparing shGFP to shMBIP). **c)** Effect of MBIP knockdown on NCI-H2009 cell viability as assayed by WST assay. **d)** Effect of MBIP knockdown on NCI-H661 cell viability as in c).

Supplementary Results

Large-scale lesions

The most common genomic alteration is copy-number gain of chromosome 5p, which is found in 60% of total samples and over 80% of the top tertile (**Supplementary Table 5**). Other frequent copy-number gains are seen within chromosome arms 1q, 7p/q, 8q, and 17q (**Supplementary Table 5**). The most common copy-number losses occur on chromosome arms 3p, 6q, 8p, 9p/q, 13q, 15q, 17p, 18q, and 19p, with each seen in at least one-third of the total samples. Despite their high frequency, only a handful of these large-scale events have been clearly related to functional effects on specific genes. Based on our current understanding of cancer progression, loss of a chromosome arm is likely to act by uncovering an inactivated tumor suppressor gene on the other homolog or possibly by causing haploinsufficiency. Yet, tumor suppressor gene mutations in lung adenocarcinoma have been well-established in only three of the 16 deleted chromosome arms (*CDKN2A* on 9p, *TP53* on 17p and *STK11* on 19p) [15-17]. Two other chromosome arms are known to harbour tumor suppressor genes that are mutated in small cell lung carcinoma (*PTEN* on 10q and *RB1* on 13q) [42-44], but not yet established in lung adenocarcinoma. Moreover, it remains to be established clearly whether those tumors with arm-level losses consistently show inactivation of the corresponding genes. Such demonstration will require assessing deletions, point mutations and epigenomic modification status, because, for example, it has been reported that *CDKN2A* is often inactivated by methylation [45, 46].

We wanted to test whether the multiple chromosomal copy number aberrations in lung adenocarcinoma are tightly associated with one another and thereby define specific subclasses of this disease. However, association with the limited set of demographic and clinical data for these anonymized specimens failed to find any statistically significant correlation between lesions and clinical parameters that passed correction for multiple hypothesis testing. The most significant correlations without correction are listed in **Supplementary Table 7.**

Visual inspection reveals substantial differences among samples in the amplitude of copy-number variation seen: some samples show dramatic changes across the genome, while others

show a narrower, attenuated range (**Figure 1a**, left vs. right side). Various lines of evidence indicate that these levels represent differing levels of non-tumor cell admixture and/or variations in ploidy. When the samples are partitioned into three tertiles based on the overall copy-number amplitude, we find that each tertile shows the same genome-wide pattern of sites of amplification and loss. The lower tertiles show a pattern of allelic balance consistent with stromal admixture and/or single-copy variations on a background of higher ploidy. One can estimate (roughly, given the assumptions involved) the level of this signal attenuation relative to unmixed diploid tumors as ~50% in the top tertile, ~65% in the middle tertile and ~78% in the bottom tertile.

LOH analysis

The significant stromal admixture in this tumor DNA collection makes it difficult to assess loss of heterozygosity (LOH) comprehensively across the genome. We have previously shown that LOH this measurement can not be determined accurately in the setting of 30% or more non-tumor DNA [47]. Our current results show LOH can only be called in only a subset of the top tertile of samples, consistent with our estimated signal attenuation (**Supplementary Figure 3**). Applying the GISTIC method to allelic loss, we identified several LOH regions including chromosome arms 5q, 13q, 17p, 19p, and 21q, each of which is also identified as a region of large-scale copy number loss **(Supplementary Table 6, Supplementary Figure 4)**. Within each region, there are individual cases of copy-neutral LOH as well as deletion-associated LOH. For example, 4 out of 17 tumor DNA specimens with chromosome 17p LOH and 4 out of 14 specimens with chromosome 19p LOH retain a neutral copy number for these regions, with no evident deletion.

Comparison with large-scale events found in previous studies

The overall pattern of large-scale genomic events seen in this study is consistent with the literature on lung cancer [17, 23-27, 32-38]. For example, although all 10 of the large-scale copy number gains and 13 of the16 deleted chromosome arms were identified previously in lung adenocarcinoma, no single previous study identified more than 5 of the gained arms or 11 of the deleted arms [13, 48-52] **(Supplementary Table 2).** The most commonly reported NSCLC-specific

large-scale gains, on chromosome arms 1q, 3q, 5p and 8q, and large-scale losses, on chromosome arms 3p, 8p, 9p, 13q, and 17p are seen in our analysis[12]. Chromosome 3q gain, which is more prevalent in squamous cell carcinoma than in adenocarcinoma [53], and several other alterations reported as arm-level in these lower-resolution studies **(Supplementary Table 2)** are seen here as focal events, including amplification within chromosome arms 11q and 19q and deletions within chromosome arms 1p and 4q **(Figure 1b)**.

Comparison with focal alterations found in previous studies

Comparison of our results with three recent studies [8, 13, 14] underscores the importance of a large sample set for cancer genome analysis. Twenty-three of the 31 recurrent focal events seen here are among the ~200 events reported in three previous, modestly powered studies, but these regions were not reliably and reproducibly identified. In fact, only 4 events were seen in primary adenocarcinomas and/or unspecified primary NSCLCs in multiple studies [8, 14], amplification of *EGFR*, *CCNE1*, *MDM2*, and 8p11; each of these 4 events is found by the current study **(Supplementary Table 3)**. In addition, small sample sizes in these studies made it unfeasible to estimate the frequency of the focal alteration, which we could now more precisely localize. For example, the minimum 5p13.33 amplified region seen in the three previous studies was ~4.5 Mb, but the current study narrows this to ~870 kb (**Supplementary Table 10**). Of the six proto-oncogenes with known mutations in lung adenocarcinoma, three (*EGFR*, *ERBB2*, and *KRAS*) were found to lie within focal amplifications in our study; three (*BRAF*, *PIK3CA* and *PTPN11*) did not lie within such amplifications. In contrast, previous analyses each found one to two of these mutated proto-oncogenes to be amplified, and only one proto-oncogene amplified in both previous studies.

Focal amplifications

The focal amplification events are seen at a frequency of about 1-7% across the entire sample collection, although the frequency and copy number are likely to be an underestimate due to stromal admixture. The frequency in the top tertile is nearly twice as high, ranging from 1-12%.

Each of the amplification events is seen in at least two samples and all but 8 are seen in at least five samples. In the 13 most significant amplifications ($q < 0.01$), the regions can be localized to relatively small genomic segments containing 15 or fewer genes.

The results highlight the limits of our current knowledge about the genetic basis of lung adenocarcinoma. Although a known proto-oncogene can be associated with 14 of the 24 regions of recurrent amplification (**Supplementary Table 10**), only three of these genes (*EGFR*, *KRAS* and *ERBB2*) have been previously reported to be mutated in lung adenocarcinoma. Conversely, among proto-oncogenes previously reported to be mutated in lung adenocarcinoma, only three (*BRAF, PIK3CA* and *PTPN11)* do not lie within the significant regions of focal amplification. We note that it is essential to systematically analyse all genes within the amplified regions and the presence of a known proto-oncogene does not prove that it is the functional target of the amplification. For example, the amplification of 8p11 contains the *FGFR1* proto-oncogene, but *FGFR1* is not required for survival of lung cancer cells while the nearby *WHSC1L1* gene is required [14].

VEGFA, which encodes the vascular endothelial growth factor is amplified in a region containing only one other gene and the region encoding the VEGF receptor (the *KDR* locus) also shows amplification, although the signal lies just below our threshold for genome-wide significance. These observations suggest a possible molecular basis for increased angiogenesis and response to angiogenic inhibitors in this disease, including the efficacy of the anti-VEGF antibody, bevacizumab, in treatment of non-small cell lung cancer [23, 24].

Chromosome 7q21.2 shows amplification in 3 samples and delineates a region containing 11 genes. The center of the amplified region is *CDK6*, which encodes a cyclin-dependent kinase. This amplification is intriguing because it supports the role of cell cycle alterations as a result of genomic changes in the pathogenesis of lung adenocarcinoma. One of the most common amplifications affects the closest homolog to *CDK6*, the *CDK4* gene; another affects the gene for cyclin D1 (*CCND1*), the major cyclin activating subunit for Cdk4 and Cdk6, while the leading homozygous deletion targets the genes, *CDKN2A* and *CDKN2B*, whose products inhibit Cdk4 and Cdk6.

Exon re-sequencing of NKX2-1 and MBIP

Many genes in recurrent amplifications are associated with somatic mutations, often genes involved in signal transduction pathways (e.g. *EGFR, KRAS, ERBB2*), while other presumptive amplified proto-oncogenes, often transcription factors, are not (e.g. *MYC, MDM2*). To determine the occurrence of mutations within *NKX2-1*, a homeodomain-containing transcription factor [26] and *MBIP*, a putative JNK/SAP kinase pathway inhibitor [54], we subjected these genes to exon-based sequencing in 384 lung adenocarcinoma DNA samples, including 232 that were characterized by SNP array analysis. We found no somatic mutations in either gene, suggesting that any oncogenic function might be exerted by the wild-type gene and therefore could best be elucidated by functional assays.

Supplementary methods

Primary lung specimens. A total of 575 DNA specimens were obtained from primary lung tumors (all of them with the original diagnosis of lung adenocarcinoma, 528 of which were confirmed to be lung adenocarcinomas), 439 matched normals and 53 additional normal specimens. These DNAs were labelled and hybridized to SNP arrays (see below) without prior whole genome amplification. Each of the selected tumor samples were determined to have greater than 70% tumor percentage by pathology review.

Of the 575 selected tumors, 384 anonymous lung tumor and matched normal DNAs for the Tumor Sequencing Project (TSP) were collected from five sites: Memorial-Sloan Kettering Cancer Center (102 tumors and paired normals), University of Michigan (101 tumors and paired normals), MD Anderson Cancer Center (29 tumors and paired normals), Washington University (84 tumors and paired normals) and Dana-Farber Cancer Institute/The Broad Institute (68 tumors and paired normals). Additional anonymous lung adenocarcinoma samples or DNAs were collected from the Brigham and Women's Hospital tissue bank (19 tumors and 18 paired normal samples), Hidefumi Sasaki at the Nagoya City University Medical School (112 tumors and 37 paired normal samples) and from the University Health Network in Toronto (60 tumor samples). In addition to the matched normal samples, 53 unmatched normal tissue or blood samples were used for SNP array normalization purposes (sources include Josef Llovet, Scott Pomeroy, Sam Singer, the Genomics Collaborative, Inc, Massachusetts General Hospital and Rameen Beroukhim). All tumor samples were surgically dissected and frozen at -80^{o}C until use.

Single nucleotide polymorphism array experiments. For each sample, SNPs were genotyped with the StyI chip of the 500K Human Mapping Array set (Affymetrix Inc). Array experiments were performed according to manufacturer's directions. In brief, for each sample 250 ng of genomic DNA was digested with the StyI restriction enzyme (New England Biolabs). The digested DNA was then ligated to an adaptor with T4 ligase (New England Biolabs) and PCR amplified using an Applied Biosystems 9700 Thermal Cycler I and Titanium Taq (Clontech) to

achieve a size range of 200-1100 bp. Amplified DNA was then pooled, concentrated and put through a clean up set. The product was then fragmented using DNaseI (Affymetrix Inc) and subsequently labeled, denatured and hybridized to arrays. Hybridized arrays were scanned using the GeneChip Scanner 3000 7G (Affymetrix Inc.). Batches of 96 samples were processed as a single plate using a Biomek FX robot with dual 96 and span-8 heads (Beckman Coulter) and a GeneChip Fluidics Station FS450 (Affymetrix Inc). Samples and plates were tracked using ABgene 2D barcode rack and single tube readers (ABGene). Tumor and paired normal sample (where applicable) were always placed in adjacent wells on the same plate to minimize experimental differences. Raw data (.cel and .txt files) are available through the website http://www.broad.mit.edu/tsp.

Primary SNP array data analysis. SNP arrays were processed as a plate of 96 samples using the GenePattern software package [31] with modules based on dChipSNP algorithms [9, 10]. GenePattern modules are available at http://www.broad.mit.edu/cancer/software/genepattern/. Intensity (.CEL) files were normalized and modelled using the PM-MM difference modelling method [9] with the SNPfileCreator module. Array normalization, similar to quantile normalization was performed [32]; 6000 matching quantiles from the probe density distributions of two arrays were used to fit a running median normalization curve for normalization of each array to a common baseline array [10].

Array quality control analysis. Further analysis was performed on arrays that met certain quality control criteria. As a first step, non-adenocarcinoma samples (n = 47) from the TSP set of 384 tumors were removed from further analysis (leaving 528 adenocarcinomas). Technical failure criteria (removing 33 tumors) included a requirement for correct tumor/normal matching, genotyping call rates (% of SNPs that a genotype call can be inferred for) greater than 85% and a score measuring copy number variation between neighboring SNPs of less than 0.5. The measure of local SNP copy number variation is calculated by the formula: Variation score = Mean [($\log(RC_i)-\log(RC_{i+1}))^2+(\log(RC_i)-\log(RC_{i-1}))^2$], where RC_i is the raw copy number at SNP i and the mean is taken over all SNPs. Criteria also included a requirement that after taking the log2 ratio and performing segmentation by GLAD [11], the number of times the smoothed copy number

crossed +/-0.1 on the log scale in the genome of tumor samples was < 100 (removing 73 tumors). The same test was used to exclude normal samples, with the number of times the smoothed copy number crossed +/- 0.1 decreased to < 45 (removing 50 normals). A histogram quality control step, as part of the GISTIC procedure, then removed tumors (n=51) with high degrees of non tumor DNA contamination by looking for samples with only 1 peak of copy number across its whole genome. This histogram quality control step also removed normals (n=20) with tumor DNA contamination by looking for samples with greater than 1 peak of copy number across its whole genome.

GISTIC analysis. GISTIC analysis (described in detail in Beroukhim et al., in press) was performed on arrays that met certain quality control criteria. Raw intensity value files from the GenePattern SNPfileCreator module were used as input into the GISTIC algorithm. In brief, batch correction, data normalization, copy-number determination using either the paired normal sample or the average of the five closest normal samples and copy number segmentation. Dataset specific copy number polymorphisms were identified by running GISTIC on the set of normal samples alone; the regions identified from this analysis were then also removed from the subsequent analysis of tumors. GISTIC then assigns G^{AMP} and G^{DEL} scores to each locus, respectively representing the frequency of amplifications (deletions) seen at that locus, multiplied by the average increase (decrease) in the log2 ratio in the amplified (deleted) samples. The score (G) is based on the average amplitude (a) of the lesion type (amplification or deletion) and its frequency (f) in the dataset according to the formula: $G_i^{(\text{lesion type})} = f_i^{(\text{lesion type})}\, a_i^{(\text{lesion type})}$. The significance of each score is determined by comparison to similar scores obtained after permuting the data within each sample. The resulting q-value is an upper bound for the expected fraction of false positives among all regions with a particular q-value or less. GISTIC also implements a peel-off step, which identifies additional secondary peaks within a region.

GISTIC analysis was performed essentially the same as is described in Beroukhim, Getz et al, with the following exceptions. Copy number determination was performed for each tumor using its matched normal sample when available and of good quality (n=242). For all others, the average of

the 5 closest normal samples was used (n=129). Copy number segmentation was performed using the GLAD algorithm with parameter d=10. GLAD segments less than 8 SNPs in length were also removed.

Regions identified by GISTIC were also compared to known copy-number polymorphisms [33] and were manually reviewed for the presence of the alteration in the paired normal sample. Focal deletion regions with events that occurred in tumor samples that did not have paired normals were considered presumed polymorphisms and also removed from the list. Secondary peaks, known and presumed germline copy number polymorphisms are listed in **Supplementary Tables 12 and 13**.

GISTIC analysis of large-scale regions. Significant broad regions of amplification and deletion were identified by applying GISTIC with the default thresholds of 2.14 / 1.87 (log2 ratio of +/- 0.1). Regions identified by GISTIC that were greater than 50% of a chromosome arm were considered large-scale. Region frequencies were calculated by determining the number of samples that had a median log2 ratio greater/less than the threshold (+/- 0.1), for those SNPs within the region.

GISTIC analysis of focal regions. Significant focal regions of amplification and deletion were identified by applying GISTIC with a threshold of 3.6 / 1.2 (log2 ratio of 0.848/-0.737).

Data visualization. Normalized raw copy number from GISTIC analysis was used as input for visualization in the GenePattern SNPviewer (http://www.broad.mit.edu/cancer/software/genepattern/) [31]. Mapping information for SNP, Refgene and cytoband locations are based on Affymetrix annotations and hg17 build of the human genome sequence from the University of California, Santa Cruz (http://genome.ucsc.edu).

Chromosome arm analysis. After segmentation by GLAD, the median of each chromosome arm for each sample was calculated. Amplification or deletion of an arm across the dataset was tested

for significance by a 2-sided binomial test, after removing log2 copy number ratios between +/- 0.1. *p* values were FDR corrected to give a false discovery rate *q* value, significance is set to a *q* value of 0.01. The standard deviation of the median copy number of significant arms was then used to sort samples into 3 groups. Higher standard deviation implies higher interchromosomal variation, which correlates with less stromal contamination. Frequencies were then calculated for the total set and for only the top 1/3 least stromally contaminated samples to give a better idea of true frequencies in the context of attenuated signal due to stromal contamination.

Comparison between tertiles. A similar chromosome arm analysis was performed independently on the 3 sample groups, separated according to the standard deviations of their median arm log2 copy number ratios. Amplification or deletion of an arm across the dataset was tested for significance by a 2-sided binomial test, after removing values between +/- 0.0125. *p* values were FDR corrected to give a false discovery rate *q* value, significance is set to a *q* value of 0.01.

Estimation of stromal contamination. To attempt to estimate stromal contamination, we calculated the allele-specific copy-numbers by taking all informative SNPs in each of the 237 tumors which have a paired normal (removing 5 bad pairs) by dividing the allele-specific signal from the tumor by that of the normal. Then for each SNP we found M, the minimum between the copy numbers of the A and B alleles. In regions in which one allele has zero copies (e.g one copy loss in diploid cells) M represents the stromal contamination level (since the stroma has one copy of each allele). We calculated the median value of M across each of the chromosome arms and then estimated the stromal contamination by taking their minimum.

LOH analysis. Inferred LOH calls using an HMM algorithm for 242 tumor/normal sample pairs were generated using dChipSNP[34]. Default parameters were used, except the genotyping error rate was set to 0.2. Five bad quality sample pairs were removed prior to visualization and GISTIC analysis. GISTIC analysis of LOH calls and copy loss for 237 samples were performed as described (Beroukhim, Getz et al, in press).

Correlation analysis. Associations were tested between each large-scale alteration identified by GISTIC and certain clinical parameters. A Fisher's exact test was used to determine association of large-scale copy number lesions with the binary clinical parameters (gender and smoking status). A chi^2 test was used to determine whether each large-scale copy number alteration was independent of each non-binary clinical parameter (age range, differentiation, tumor stage or patient's reported ancestry). p-values were FDR corrected to give a false discovery rate *q* value, significance is set to a *q* value of 0.05.

Correlation of clinical features and *NKX2-1* amplification. The analysis included 123 consecutive patients with lung adenocarcinoma treated at Brigham and Women's Hospital between January 1997 and December 1999. Fifty-two of these cases had a FISH amplification status that was not assessable (6 cases showed no tumor on the tissue cores and 46 cases had insufficient hybridization). Of the remaining 71 cases, 10 cases had *NKX2-1* amplification, 1 had a *NKX2-1* deletion, and 60 cases showed no *NKX2-1* alteration. All cases for which the *NKX2-1* amplification status was not assessable and the one case that showed a *NKX2-1* deletion were excluded, bringing the final number of cases included in the analysis to 70.

All cases were histologically confirmed as lung adenocarcinomas. For cases that showed a pure solid growth pattern, a mucicarmine and immunohistochemical stains were performed to confirm that the tumor was an adenocarcinoma. Well-differentiated tumors were defined as tumors with a purely bronchioloalveolar growth pattern or mixed tumors with an acinar component with cytologic atypia equivalent to that seen with bronchioloalveolar carcinoma. Poorly-differentiated tumors were defined as tumors that showed any amount of solid growth. All other tumors were classified as moderately-differentiated. Patient demographics, smoking status, tumor location, type of surgical resection, tumor stage (according to the 6th Edition of the American Joint Committee on Cancer system for lung carcinoma), and nodal status were recorded.

Overall survival of patients with *NKX2-1* amplification. We excluded from the survival analysis three cases with *NKX2-1* amplification and 11 cases that had no *NKX2-1* alterations.

Exclusion criteria included: cancer was a recurrence, patients received neoadjuvant treatment, or died within the first 30 days after surgery, or patients had another cancer diagnosed in the five years prior to the diagnosis of lung adenocarcinoma. Survival was plotted by Kaplan-Meier method using the date of resection and date of death or last follow-up.

Sequencing. *NKX2-1*, *MBIP*, and *AUTS2* were sequenced in all 384 TSP lung adenocarcinomas. Primers were designed in an automated fashion using Primer 3 [35] and characterized by amplification in genomic DNA from three Coriell cell lines. Primers that show an agarose gel band for at least 2 of the 3 DNAs were then used for production PCR. Passing primers were arrayed into 384 well PCR plates along with samples and PCR master mix. A total of 5 ng of whole-genome amplified sample DNA was PCR amplified over 35 cycles in Thermo-Hybaid units, followed by a SAP/Exo cleanup step. *NKX2-1* PCR reactions for sequencing contained an addition of 5% DMSO. The resulting purified template is then diluted and transferred to new plates for the sequencing reaction. After cycling (also performed on Thermo-Hybaids), the plates are cleaned up with an ethanol precipitation, re-hydrated, and detected on ABI 3730xl's (Applied Biosystems). Output from the detectors is transferred back to the directed sequencing platform's informatics pipeline. SNPs and/or mutations are then identified using three mutation-detecting algorithms in parallel: PolyPhred [36] and PolyDHAN (Richter et al., manuscript in preparation), which are bundled into the in-house software package SNP Compare, and the commercially available Mutation Surveyor (SoftGenetics, LLC.). Candidates were filtered to remove silent variants, intronic variants (with the exception of potential splice site mutations), and validated SNPs registered in dbSNP or confirmed as SNPs in our previous experiments.

Mutation validation by genotyping. hME genotyping for validation of sequencing candidates was performed in 96-well plates with up to 7-plex reactions. PCR was performed with a final concentrations of 0.83mM dNTP’s, 1.56X of 10X Buffer, 3.38mM MgCl, 0.03 units/uL HotStar Taq (Qiagen), 0.10uM PCR primers. Thermocycling was performed at 92°C for 15 min, followed by 45 cycles of 92°C for 20 sec, 56°C for 30 sec and 72°C for 1 min, with an additional extension at 72°C for 3 min. Shrimp Alkaline Phosphatase (SAP) clean-up was performed using a master

mix made up of 0.5X buffer and SAP. Reactions were performed at 34°C for 20 min, 85°C for 5 min and then held at 4°C. Following the SAP clean-up, hME reaction was performed using Thermosequenase and final concentrations of 0.06 mM Sequenom Termination Mix (specific to the pool being used), and 0.64uM extension primer Reactions were cycled at 94°C for 2 min, followed by 55 cycles of 94°C for 5 sec, 52°C for 5 sec and 72°C for 5 sec. Samples then were put through a resin clean-up step, then the purified primer extension reaction was loaded onto a matrix pad (3-hydroxypicoloinic acid) of a SpectroCHIP (Sequenom, San Diego, CA) and detected by a Bruker Biflex III MALDI-TOF mass spectrometer (SpectroREADER, Sequenom).

***PTPRD* mutation discovery and validation.** The *PTPRD* gene was sequenced in 188 lung adenocarcinoma samples. Sequence traces (Reads) were aligned to human reference sequence using cross-match. PolyPhred [36] and PolyScan were used to predict SNPs and indels. Identified SNPs were validated using Illumina Goldengate assay. ENST00000356435 is the transcript used for annotating the mutations. Both synonymous and non-synonymous candidates were identified, but only non-synonymous mutations were validated.

Tissue microarray Fluorescent in-situ hybridization (TMA–FISH). A Biotin-14-dCTP labeled BAC clone RP11-1083E2 (conjugated to produce a red signal) was used for the NKX2-1 probe and a Digoxin-dUTP labeled BAC clone RP11-72J8 (conjugated to produce a green signal) was used for the reference probe. Tissue hybridization, washing, and color detection were performed as described previously [7, 37]. NKX2-1 amplification by FISH was assessed using a total of 935 samples (represented by 2818 tissue microarray cores).

The BAC clones were obtained from the BACPAC Resource Center, Children's Hospital Oakland Research Institute (CHORI, Oakland, CA). Prior to tissue analysis, the integrity and purity of all probes were verified by hybridization to metaphase spreads of normal peripheral lymphocytes. The samples were analyzed under a 60x oil immersion objective using an Olympus BX-51 fluorescence microscope equipped with appropriate filters, a CCD (charge-coupled device) camera and the CytoVision FISH imaging and capturing software (Applied Imaging). Semi quantitative

evaluation of the tests was independently performed by two evaluators (SP and LAJ); at least 100 nuclei for each case were analyzed when possible. Cases with significant differences between the two independent evaluations were refereed by a third person (MAR). The statistical analysis was performed using SPSS 13.0 for Windows (SPSS Inc.) with a significance level of 0.05.

Cell lines and cell culture conditions. NCI-H2009 [38], NCI-H661 [39], NCI-H1975 [38] and HCC1171[8] have been previously described. A549 cells were purchased from American Type Culture Collection. NSCLC cells were maintained in RPMI growth media consisting of RPMI 1640 plus 2 mM L-glutamine (Mediatech) supplemented with 10% fetal bovine serum (Gemini Bio-Products), 1 mM sodium pyruvate, and penicillin/streptomycin (Mediatech).

RNAi knockdown. shRNA vectors targeted against *NKX2-1, MBIP* and *GFP* were provided by TRC (The RNAi Consortium). The sequences targeted by the NKX2-1 shRNAs are as follows: shNKX2-1a (TRCN0000020449), 5'-CGCTTGTAAATACCAGGATTT-3', and shNKX2-1b (TRCN0000020453) 5'- TCCGTTCTCAGTGTCTGACAT-3'. The sequence targeted by the MBIP shRNA and GFP shRNA are 5'-CCACCGGAAGGAAGATTTATT-3' (TRCN0000003069) and 5'-GCAAGCTGACCCTGAAGTTCAT-3', respectively. Lentiviruses were made by transfection of 293T packaging cells with a three plasmid system [40, 41]. Target cells were incubated with lentiviruses for 4.5 hours in the presence of 8 μg/ml polybrene. After the incubation, the lentiviruses were removed and cells were fed fresh medium. Two days after infection, puromycin (0.75 μg/ml for NCI-H1975, 1.0 μg/ml for NCI-H661, 1.5 μg/ml for NCI-H2009, 1.0 μg/ml for NCI-H661, and 2.0 μg/ml for A549 and HCC1171) was added. Cells were grown in the presence of puromycin for three days or until all of the non-infected cells died. Twenty-five micrograms of total cell lysates prepared from the puro-selected cell lines was analyzed by Western blotting using anti-NKX2-1 polyclonal antibody (Santa Cruz Biotechnology), anti-MBIP polyclonal antibody (Proteintech Group, Inc.) and anti-vinculin monoclonal antibody (Sigma).

Soft Agar Anchorage-Independent Growth Assay. NCI-H2009 (1×10^4), NCI-H661 (2.5×10^4), A549 (3.3×10^3), NCI-H1975 (5×10^4) or HCC1171 (1×10^4) cells expressing shRNAs

targeting *NKX2-1*, *MBIP* or *GFP* were suspended in a top layer of RPMI growth media and 0.4% Noble agar (Invitrogen) and plated on a bottom layer of growth media and 0.5 % Noble agar in 35 mm wells. Soft agar colonies were counted 3 to 4 weeks after plating. The data are derived from two independent experiments unless otherwise noted and are graphed as the percentage of colonies formed relative to the shGFP control cells (set to 100%) +/- one standard deviation of the triplicate samples. p values between shGFP and shNKX2-1 or shMBIP samples were calculated using a t-test.

Cell proliferation assays. NCI-H2009 (500 cells/well), A549 (400 cells/well), and NCI-H661 (600 cells/well) cells expressing shRNAs targeting *NKX2-1*, *MBIP* or *GFP* were seeded in 6 wells in a 96 well plate. Cell viability was determined at 24 hour time points for a total of 4 days using the WST-1 based colorimetric assay (Roche Applied Science). The percentage of cell viability is plotted for each cell line +/- one standard deviation of the reading from six wells, relative to Day 0 readings. Experiments were performed two or more times and a representative experiment is shown.

Supplementary Table 1: Clinical Summary

Age	# samples
<40	4
40-49	29
50-59	91
60-69	176
70-79	141
80-89	39
NA	48

Gender	# samples
F	256
M	225
NA	47

Smoking status (pack-years)	# samples
0	82
1-10	25
11-20	28
21-30	29
31-40	35
41-50	38
51-60	19
61-70	5
71-80	15
81-90	8
91-100	3
>100	19
NA	222

Stage	# samples
1	33
1A	74
1B	72
2	17
2A	13
2B	35
3A	44
3B	21
4	12
NA	207

T stage	# samples
1	176
2	226
3	34
4	30
NA	62

N stage	# samples
0	305
1	85
>1	20
NA	118

Supplementary Table 2: Comparison of large-scale regions with literature*

	Current study	Garnis et al. [13]	Luk et al. [50]	Balsara et al. [48]	Bjorkqvist et al. [49]	Petersen et al. [51]	Testa et al. [52]	Present in at least 1 study
Amplified arms	1q, 5p, 6p, 7p, 7q, 8q, 16p, 17q, 20p, 20q	1q, 5p, 7p, 8q, 20q, 2p	1q, 5p, 8q, 20q, 3q, 11q, 15q, 19q	1q, 5p, 7p, 8q, 20p, 2p, 3q	1q, 5p, 6p, 7p, 8q	1q, 5p, 8q, 16p, 17q, 11q, 19q	1q, 7p, 7q, 11q	1q, 5p, 6p, 7p, 7q, 8q, 16p, 17q, 20p, 20q
# of amplified arms	10	5 / 6	4 / 8	5 / 7	5 / 5	5 / 7	3 / 4	10 / 15
Deleted arms	3p, 5q, 6q, 8p, 9p, 9q, 10q, 12p, 13q, 15q, 17p, 18q, 19p, 19q, 21q, 22q	3p, 8p, 9p, 13q, 18q, 4q, 10p	5q, 6q, 9p, 9q, 13q, 18q			3p, 5q, 6q, 8p, 9p, 9q, 13q, 18q, 19p, 19q, 21q 1p, 4q, 10p, 3q	3p, 6q, 8p, 9p, 9q, 13q, 17p, 18q, 19p, 21q, 22q	3p, 5q, 6q, 8p, 9p, 9q, 13q, 17p, 18q, 19p, 19q, 21q, 22q
# of deleted arms	16	5 / 7	7 / 7	0 / 0	0 / 0	11 / 15	11 / 11	13 / 17

* The numerator of each indicates number of unique regions that overlap the current study, the denominator for each indicates the total number of unique lesions found for that study; only NSCLC studies are used, but are not restricted lung adenocarcinoma. Underlined arms are not present in the current study.

Supplementary table 3: Comparison of focal regions with literature

	Current study	Tonon et al. [14] primary NSCLC alterations *	Zhao et al.[8] primary lung adenocarcinoma alterations *	In both previous studies and current study ^	In both previous studies but not current study ^	In only one previous and current study ^	In only one previous but not current study ^
Amplifications	24	12 / 73	5 / 13	4 / 82	0 / 82	9 / 82	62 / 82
Amplifications containing a mutated gene in lung adenocarcinoma	3 / 24	2 / 73	1 / 13	1 / 82	0 / 82	2 / 82	0 / 82
Deletions	7	2 / 20	0	0 / 20	0 / 20	2 / 20	18 / 20
Deletions containing a mutated gene in lung adenocarcinoma	1 / 7	1 / 20	0	0 / 20	0 / 20	1 /20	0 / 20

* The numerator indicates number of unique regions that overlap the current study, the denominator for each indicates the total number of unique lesions found;

^ The numerator indicates number of unique regions for the column, the denominator for each indicates the total number of unique lesions found.

Supplementary Table 4: Arm level events with q value <0.01

Amplifications

Chromosome arm	Frequency *	q value
5p	60.4%	6.6E-56
1q	51.7%	2.1E-52
7p	44.6%	4.2E-39
8q	42.9%	3.9E-34
17q	34.6%	2.4E-26
6p	26.4%	1.1E-13
20q	30.3%	1.0E-12
7q	26.9%	2.5E-09
2p	16.5%	6.7E-07
2q	13.4%	6.9E-04

Deletions

Chromosome arm	Frequency ^	q value
9p	42.9%	2.2E-24
19p	43.3%	4.6E-22
17p	40.5%	4.3E-14
18q	39.2%	4.3E-14
9q	35.1%	1.1E-13
15q	30.5%	1.5E-13
8p	41.1%	1.9E-13
6q	31.0%	1.9E-13
3p	34.0%	1.9E-13
13q	33.3%	1.9E-13
10q	21.0%	4.5E-11
22q	27.7%	2.5E-09
5q	25.1%	4.6E-05
4q	16.9%	0.00064
16q	21.9%	0.00072
18p	21.2%	0.004
12q	15.4%	0.008
19q	28.4%	0.0099

* Frequency for amplifications is the percent of samples per arm with copy # median ≥2.14;

^ Frequency for deletions is the percent of samples per arm with copy # median ≤1.87.

Supplementary Table 5: Large-scale events from GISTIC^

Amplifications

Arm	Region*	Region seen previously **	# samples (total, top 1/3) #	% samples (total, top 1/3) #
1q	120.88-245.528	NSCLC, SCLC	209, 86	56.3, 69.4
5p	1-52.43	NSCLC, SCLC	232, 103	62.5, 83.1
6p	1-57.74	NSCLC	108, 50	29.1, 40.3
7p + 7q	1-158.63	NSCLC	139, 67	37.5, 54.0
8q	41.06-146.27	NSCLC, SCLC	162, 90	43.7, 72.6
16p	1-34.07	NSCLC	93, 41	25.1, 33.1
17q	22.37-78.77	NSCLC	142, 66	38.3, 53.2
20p + 20q	1-62.44	NSCLC	111, 64	29.9, 51.6

Deletions

Arm	Region*	Region seen previously **	# samples (total, top 1/3) #	% samples (total, top 1/3) #
3p	1-96.55	NSCLC, SCLC	138, 67	37.2, 54.0
5q	51.72-166.38	NSCLC, SCLC	103, 45	27.8 ,36.3
6q	63.85-170.98	NSCLC	127, 59	34.2, 47.6
8p	1-43.36	NSCLC, SCLC	168, 74	45.3, 59.7
9p + 9q	1-138.43	NSCLC	140, 58	37.7, 46.8
10q	80.63-135.41	SCLC	90, 45	24.3, 36.3
12p	1.12-24.05		88, 33	23.7, 26.6
13q	1-114.14	NSCLC, SCLC	133, 69	35.8, 55.6
15q	1-100.34	SCLC †	126, 72	34.0, 58.1
17p	1-22.03	NSCLC, SCLC	159, 74	42.9, 59.7
18q	12.88-76.12	NSCLC	165, 71	44.5, 57.3
19p	1-32.84	NSCLC	163, 67	43.9, 54.0
19q	37.48-63.81	NSCLC	122, 53	32.9, 42.7
21q	1-33.39	NSCLC	74, 44	19.9, 35.5
22q	1-49.55	NSCLC	108, 55	29.1, 44.4

^ Only regions > 50% of a chromosome arm (by physical distance) are shown (thresholds used are 2.14 and 1.87); * based on hg17 human genome assembly, positions in Mb; ** Based on a non-comprehensive, but representative group of publications [12, 13, 48-52, 55, 56], Loss of heterozygosity (LOH) studies are indicated with a †; # Top third numbers and percentages refer to the top 1/3 least stromally contaminated samples, as assayed by standard deviation measurements.

Supplementary Table 6: Top regions of loss of heterozygosity

Cytoband	q-value	Region *	# samples with copy neutral LOH [%]	# samples with LOH associated with deletion [%]
17p13.2	0.00051019	1-18.82	4	13
5q35.3	0.0023507	175.74-180.62	4	11
5q31.1	0.00684	46.21-144.95	4	10
19p13.3	0.00684	1-21.53	4	10
13q14.11	0.013344	28.86-62.40	2	11
21q22.11	0.013344	1-46.90	4	9

[%] All LOH events occur within the top tertile of least stromally contaminated samples, as assayed by standard deviation measurements;

* based on hg17 human genome assembly, positions in Mb.

Supplementary Table 7: Top clinical associations

fisher's test for association

Clinical feature	Lesion	p value	FDR q value	n
Female	10q deletion	0.040838	0.939274	330
Neversmoker	10q deletion	0.005545	0.127535	259
Neversmoker	16p amplification	0.023095	0.265593	259
Neversmoker	15q deletion	0.023969	0.183762	259
Neversmoker	7p-q amplification	0.026513	0.15245	259

chi^2 test for independence

Clinical feature*	Lesion	p value	FDR q value	n
Differentiation	3p deletion	0.013536	0.311328	94
Differentiation	9p-q deletion	0.018598	0.213877	94
Differentiation	19p deletion	0.022181	0.170054	94
Differentiation	19q deletion	0.022286	0.128145	94
Differentiation	8q amplification	0.030301	0.139385	94
Differentiation	21q deletion	0.048607	0.186327	94
Stage	8p deletion	0.037705	0.867215	201
Age range	22q deletion	0.036286	0.834578	330
Reported race	5q deletion	0.00974	0.22402	126

* Differentiation – well, moderate, poor or undifferentiated; Stage – 1, 2, 3, or 4; Age range – <40, 40-49, 50-59, 60-69, 70-79, 80-89; Reported race – White, African American or Japanese.

Supplementary table 8: Focal regions of deletion

Cytoband*	q value	Peak region *	Min. inferred copy #	# of samples (percentage) below threshold[%]	# of most variable samples (percentage) below threshold[%]	# of genes *^	Region reported in lung cancer?	Minimal previously defined regional boundary*	Known tumor suppressor gene in region*[#]	Genomic evidence in cancer [$]	Genomic evidence in lung adenocarcinoma[$]	Single gene deletion previously seen in cancer	New candidate gene(s)
9p21.3	3.35E-13	21.80-22.19	0.7	11 (3.0)	8 (6.5)	3	Yes [8, 13, 14, 57, 58]	21.79-22.09 [8]	*CDKN2A/ CDKN2B*	mut.[29, 30]	mut. [15, 59, 60]		
9p23	0.001149	9.41-10.40	0.4	5 (1.4)	5 (4.0)	1	Yes [8, 13]	10.03-10.07 [8]				*PTPRD* [8, 18, 61]	
5q11.2	0.005202	58.40-59.06	0.6	3 (0.8)	2 (1.6)	1	No						*PDE4D*
7q11.22	0.025552	69.50-69.62	0.7	3 (0.8)	3 (2.4)	1	No						*AUTS2*
10q23.31	0.065006	89..67-89.95	0.5	2 (0.5)	2 (1.6)	1	Yes [8, 13, 42]	89.68-89.73 [8]	*PTEN*	mut. [29, 30]			
13q14.2	0.16313	47.89-48.02	0.6	2 (0.5)	2 (1.6)	2	Yes [13, 14, 62, 63]	45.82-57.20 [14]	*RB1*	mut. [29, 30]			
18q23	0.19267	73.96-76.10	1.1	3 (0.8)	3 (2.4)	9	Yes [13]	17.3-76.12 [13]					

[%] Most variable numbers and percentages refer to the top 1/3 least stromally contaminated samples, as assayed by standard deviation measurements;

* based on hg17 human genome assembly, positions in Mb; ^ RefSeq genes only;

[#] Known tumor suppressor genes defined as found in either COSMIC [29], CGP Census [30] or other evidence;

[$] Abbreviations are del. = deletion and mut. = mutation.

Supplementary Table 9: *PTPRD* somatic mutations

Genomic position (bp)	Reference genotype	Normal genotype	Tumor genotype	Validation	Copy number	Amino acid change*	SIFT [64, 65]	Domain
8623394	AA	AA	AG		1.72	D92G	Deleterious	InterPro domain: IPR013106; Pfam domain: Ig_V-set
8514867	CC	CC	AC	Yes	2.05	T229K	Tolerated	
8508382	AA	AA	AG	Yes	1.86	T338A	Tolerated	InterPro domain: IPR003961; Pfam domain: FN_III
8508187	GG	GG	GT	Yes	1.56	G403W	Deleterious	InterPro domain: IPR003961; Pfam domain: FN_III
8507943	TT	TT	AT	Yes	1.59	V484E	Tolerated	InterPro domain: IPR003961; Smart domain: FN_III
8494333	CC	CC	AC	Yes	1.69	R585S	Tolerated	
8475273	TT	TT	AT	Yes	2.13	L1037Q	Deleterious	
8450516	CC	CC	CG	Yes	2.16	P1258R	Deleterious	
8365990	GG	GG	GT	Yes	2.11	R1537L	Deleterious	InterPro domain: IPR000242; Smart domain: Tyr_PP
8321690	CC	CC	CG	Yes	1.92	P1810R	Deleterious	InterPro domain: IPR000242; pfscan: Tyr_PP

* ENST00000356435 is the transcript used for annotating the mutations.

Supplementary table 10: Focal regions of amplification

Cytoband*	q value	Peak region	Max. inferred copy #	# of samples (percentage) above threshold[%]	# of top tertile (percentage) above threshold[%]	# of genes *^	Region reported in lung cancer?	Minimal previously defined regional boundary*	Known proto-oncogene in region*#	Genomic evidence in cancer[$]	Genomic evidence in lung adenocarcinoma[$]	siRNA evidence in lung adenocarcinoma	New candidate gene(s)
14q13.3	2.26E-29	35.61-36.09	13.7	23 (6.2)	15 (12.1)	2	Yes[8, 13, 14]	34.64-36.22[14]				*NKX2-1* (this study)	*NKX2-1, MBIP*
12q15	1.78E-15	67.48-68.02	9.7	15 (4.0)	7 (5.6)	3	Yes [8, 14, 66]	67.42-67.95 [14]	*MDM2*	amp.[67]			
8q24.21	9.06E-13	129.18-129.34	10.3	14 (3.8)	10 (8.1)	0	Yes [8, 13, 14, 68]	128.77-129.00 [8]	*MYC* [$$]	amp., transl.[30]			
7p11.2	9.97E-11	54.65-55.52	8.7	11 (3.0)	8 (6.5)	3	Yes [8, 13, 14, 69, 70]	54.44-55.66 [8]	*EGFR*	amp., mut.[29, 30]	mut. [71-73]	*EGFR*[74]	
8q21.13	1.13E-07	80.66-82.55	10.4	9 (2.4)	6 (4.8)	8	Yes [8, 14]	48.86-97.23 [14]					
12q14.1	1.29E-07	56.23-56.54	10.4	11 (3.0)	5 (4.0)	15	Yes [8, 75]	56.26-56.75 [8]	*CDK4*	mut. [29, 30]			
12p12.1	2.83E-07	24.99-25.78	10.4	8 (2.2)	5 (4.0)	6	Yes [14, 76]	24.61-26.95 [14]	*KRAS*	mut. [29, 30]	mut. [77]		
19q12	1.60E-06	34.79-35.42	6.7	7 (1.9)	5 (4.0)	5	Yes [8, 14]	34.79-35.55 [8]	*CCNE1*	amp.[78, 79]			
17q12	2.34E-05	34.80-35.18	16.1	5 (1.3)	1 (0.8)	12	Yes [8, 70]	34.80-35.99 [8]	*ERBB2*	amp., mut.[29, 30]	mut. [80]	*ERBB2*[81]	
11q13.3	5.17E-05	68.52-69.36	6.5	7 (1.9)	2 (1.6)	9	Yes [8, 13, 14, 82, 83]	68.58-69.34 [8]	*CCND1*	transl. [30]			
5p15.33	0.000279	0.75-1.62	4.2	8 (2.2)	5 (4.0)	10	Yes [13, 14]	1-4.50 [13]	*TERT*	mut.[29]			
22q11.21	0.001461	19.06-20.13	6.6	6 (1.6)	3 (2.4)	15	Yes [8, 14]	19.45-20.31 [8]					
5p15.31	0.007472	8.88-10.51	5.6	5 (1.3)	4 (3.2)	7	Yes [8]	8.88-14.31 [8]					
1q21.2	0.028766	143.48-149.41	4.6	5 (1.3)	4 (3.2)	86	No		*ARNT*	transl., mut. [29, 30]			
20q13.32	0.0445	55.52-56.30	4.4	5 (1.3)	4 (3.2)	6	No						
5p14.3	0.064673	19.72-23.09	3.8	5 (1.3)	3 (2.4)	2	No						
6p21.1	0.078061	43.76-44.12	7.7	4 (1.1)	3 (2.4)	2	No						*VEGFA*

6p21.33	0.10468	30.24-30.53	6.2	3 (0.8)	3 (2.4)	5	No				
2p15	0.12296	61.87-63.04	13.1	2 (0.5)	2 (1.6)	5	Yes [14]	49.10-64.73 [14]			
7q21.2	0.12296	91.38-92.69	7.7	3 (0.8)	3 (2.4)	11	Yes [8]	90.81-92.22 [8]	*CDK6*	mut., transl., amp. [29, 30]	
3q26.2	0.12892	171.56-172.26	5.8	3 (0.8)	2 (1.6)	5	No		*SKIL*	mut. [29]	
19q13.12	0.135	40.27-40.43	7.7	3 (0.8)	1 (0.8)	5	No				
18q11.2	0.135	21.54-21.90	6.3	3 (0.8)	2 (1.6)	1	No		*SS18*	transl. [30]	
8p11.23	0.14247	38.16-40.88	8.1	4 (1.1)	4 (3.2)	16	Yes [8, 14]	38.24-38.45 [14]	*FGFR1*	mut, transl. [29, 30]	*WHSC1L1* [14]

* based on hg17 human genome assembly, positions in Mb;

[#] Known proto-oncogenes defined as found in either COSMIC [29], CGP Census [30] or other evidence; if there is more than one known proto-oncogene in the region, only one is listed (priority for listing is, in order: known lung adenocarcinoma mutation, known lung cancer mutation, other known mutation (by COSMIC frequency), listing in CGP Census);

[%] Most variable numbers and percentages refer to the top 1/3 least stromally contaminated samples, as assayed by standard deviation measurements;

^ RefSeq genes only;

[$] Abbreviations are amp. = amplification, transl. = translocation and mut. = mutation;

[$$] *MYC* is near, but not within the peak region.

Supplementary Table 11: Clinical features of patients with *NKX2-1*-amplified^ lung adenocarcinomas

	***NKX2-1* Not amplified # of patients (%)* N=60**	***NKX2-1* Amplified # of patients (%)* N=10**	**p value**
Sex - no. (%)			0.99
Male	24 (40.0)	4 (40.0)	
Female	36 (60.0)	6 (60.0)	
Age - yr			0.63
Mean	63.7	65.7	
Range	(36-83)	(38-82)	
Smoking Status†			0.90
Nonsmoker	9 (14.3)	1 (17.7)	
Smoker	42 (85.7)	6 (82.4)	
Pack-years (mean)	47.9	42.8	0.74
Tumor Size - cm			0.65
Mean	2.5	2.3	
Range	(0.6-6.0)	(1.0-6.5)	
Resection Type - no. (%)			0.73
Wedge Resection	14 (23.3)	2 (20.0)	
Lobectomy	43 (71.7)	7 (70.0)	
Pneumonectomy	3 (5.0)	1 (10.0)	
Tumor Differentiation - no. (%)			0.87
Well	5 (8.3)	0 (0)	
Moderate	26 (43.3)	6 (60.0)	
Poor	29 (48.3)	4 (40.0)	
pT Category - no. (%)‡			0.19
T1	20 (35.7)	2 (22.2)	
T2	29 (51.8)	4 (44.4)	
T3	4 (7.1)	0 (0)	
T4	3 (5.4)	3 (33.3)	
pN Category - no. (%)§			0.60
N0	36 (66.7)	5 (55.6)	
N1 and N2	18 (33.3)	4 (44.4)	

^ NKX2-1 amplification status determined by FISH; * Due to rounding not all percentages total 100; † Smoking status was unknown for 12 patients (3 with amplified and 9 with non-amplified tumors); ‡ Five patients were excluded from the staging analysis (4 patients had neoadjuvant treatment and 1 tumor was a recurrence); § Lymph node status was unknown for seven patients (1 with amplified and 6 with non-amplified tumors).

Supplementary table 12: Additional focal regions of amplification

Cytoband*	q value	Peak region *	Max. inferred copy #	# of samples (percentage) above threshold[%]	% of most variable samples (percentage) above threshold[%]	# of genes*^	Region reported in lung cancer?	Minimal previously defined regional boundary*	Notes	Known oncogene in region*[#]	Genomic evidence in cancer [$]	Genomic evidence in lung adenocarcinoma[$]	siRNA evidence in lung adenocarcinoma	New candidate gene(s)
8q21.11	0.000654	71.98-77.59	10.1	6 (1.6)	5 (4.0)	17	Yes [8, 14]	48.86-97.23 [14]	Secondary peak in GISTIC region					
8q24.12	0.002656	121.49-123.97	10.4	6 (1.6)	5 (4.0)	5	Yes [14]	123.37-126.14 [14]	Secondary peak in GISTIC region					
17q21.31	0.040588	41.56-41.71	4.4	5 (1.3)	3 (2.4)	1	No		known polymorphism					

* based on hg17 human genome assembly, positions in Mb;

[#] Known oncogenes defined as found in either COSMIC [29], CGP Census [30] or other evidence; if there is more than one known oncogene in the region, only one is listed (priority for listing is, in order: known lung adenocarcinoma mutation, known lung cancer mutation, other known mutation (by COSMIC frequency), listing in CGP Census);

[%] Most variable numbers and percentages refer to the top 1/3 least stromally contaminated samples, as assayed by standard deviation measurements;

^ RefSeq genes only;

[$] Abbreviations are amp. = amplification, transl. = translocation and mut. = mutation.

Supplementary table 13: Additional focal regions of deletion

Cytoband*	q value	Peak region *	Min. inferred copy #	# of samples (percentage) below threshold[%]	% of most variable samples (percentage) above threshold[%]	# of genes *^	Region reported in lung cancer?	Minimal previously defined regional boundary*	Notes	Known tumor suppressor gene in region*[#]	Evidence for candidate gene(s)[$]	Genomic evidence in lung adenocarcinoma[$]	Single gene deletion previously seen in cancer	New candidate gene(s)
8p23.2	1.93E-06	2.54-3.95	0.8	8 (2.2)	8 (6.5)	1	Yes [8, 13, 14]	0.18-2.57 [8]	Potential polymorphism				*CSMD1*[84]	
4q34.3	0.025552	180.14-191.41	0.6	2 (0.5)	2 (1.6)	31	Yes [8, 13]	182.66-183.20 [8]	Potential polymorphism					
16q23.3	0.028983	81.38-81.65	0.8	3 (0.8)	2 (1.6)	1	No		Potential polymorphism				*CDH13* [85]	
12p13.31	0.065006	6.90-10.65	0.9	3 (0.8)	2 (1.6)	59	No		Potential polymorphism					
5q23.1	0.071159	115.02-118.78	0.9	3 (0.8)	2 (1.6)	7	Yes [8]	114.60-115.05 [8]	Potential polymorphism					
22q11.23	0.030226	23.98-24.24	0.8	3 (0.8)	2 (1.6)	1	No		Known polymorphism				*LRP5L*	
15q11.2	0.012513	1-20.08	0.9	4 (1.1)	3 (2.4)	3	No		Known polymorphism					

[%] Most variable numbers and percentages refer to the top 1/3 least stromally contaminated samples, as assayed by standard deviation measurements;

* based on hg17 human genome assembly, positions in Mb;

^ RefSeq genes only;

[$] Abbreviations are del. = deletion and mut. = mutation.

Supplementary Notes

Dataset

Raw data and other related files are available at http://www.broad.mit.edu/tsp.

Raw data from the Tumor sequencing project (TSP) sample set only is available from http://caintegrator-info.nci.nih.gov/csp.

References:

1. Weir, B., Zhao, X. & Meyerson, M. Somatic alterations in the human cancer genome. Cancer Cell 6, 433-8 (2004).
2. Sawyers, C. Targeted cancer therapy. Nature 432, 294-7 (2004).
3. Pinkel, D. et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20, 207-11 (1998).
4. Pollack, J. R. et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 23, 41-6 (1999).
5. Bignell, G. R. et al. High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Res 14, 287-95 (2004).
6. Zhao, X. et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res 64, 3060-71 (2004).
7. Garraway, L. A. et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 436, 117-22 (2005).
8. Zhao, X. et al. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 65, 5561-70 (2005).
9. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 98, 31-6 (2001).
10. Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2, RESEARCH0032 (2001).
11. Hupe, P., Stransky, N., Thiery, J. P., Radvanyi, F. & Barillot, E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 20, 3413-22 (2004).
12. Balsara, B. R. & Testa, J. R. Chromosomal imbalances in human lung cancer. Oncogene 21, 6877-83 (2002).

13. Garnis, C. et al. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. Int J Cancer 118, 1556-64 (2006).
14. Tonon, G. et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A 102, 9625-30 (2005).
15. Hayashi, N., Sugimoto, Y., Tsuchiya, E., Ogawa, M. & Nakamura, Y. Somatic mutations of the MTS (multiple tumor suppressor) 1/CDK4l (cyclin-dependent kinase-4 inhibitor) gene in human primary non-small cell lung carcinomas. Biochem Biophys Res Commun 202, 1426-30 (1994).
16. Sanchez-Cespedes, M. et al. Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. Cancer Res 62, 3659-62 (2002).
17. Takahashi, T. et al. p53: a frequent target for genetic abnormalities in lung cancer. Science 246, 491-4 (1989).
18. Sato, M. et al. Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. Genes Chromosomes Cancer 44, 405-14 (2005).
19. Cox, C. et al. A survey of homozygous deletions in human cancer genomes. Proc Natl Acad Sci U S A 102, 4542-7 (2005).
20. Barnes, A. P. et al. Phosphodiesterase 4D forms a cAMP diffusion barrier at the apical membrane of the airway epithelium. J Biol Chem 280, 7997-8003 (2005).
21. Zhu, C. Q. et al. Amplification of telomerase (hTERT) gene is a poor prognostic marker in non-small-cell lung cancer. Br J Cancer 94, 1452-9 (2006).
22. Zhang, A. et al. Frequent amplification of the telomerase reverse transcriptase gene in human tumors. Cancer Res 60, 6230-5 (2000).
23. Johnson, D. H. et al. Randomized phase II trial comparing bevacizumab plus carboplatin and paclitaxel with carboplatin and paclitaxel alone in previously untreated locally advanced or metastatic non-small-cell lung cancer. J Clin Oncol 22, 2184-91 (2004).
24. Sandler, A. et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 355, 2542-50 (2006).
25. Bingle, C. D. Thyroid transcription factor-1. Int J Biochem Cell Biol 29, 1471-3 (1997).
26. Ikeda, K. et al. Gene structure and expression of human thyroid transcription factor-1 in respiratory epithelial cells. J Biol Chem 270, 8108-14 (1995).
27. Yuan, B. et al. Inhibition of distal lung morphogenesis in Nkx2.1(-/-) embryos. Dev Dyn 217, 180-90 (2000).
28. Mullighan, C. G. et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature (2007).
29. Bamford, S. et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer (2004).
30. Futreal, P. A. et al. A census of human cancer genes. Nat Rev Cancer 4, 177-83 (2004).
31. Reich, M. et al. GenePattern 2.0. Nat Genet 38, 500-1 (2006).
32. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185-93 (2003).
33. Iafrate, A. J. et al. Detection of large-scale variation in the human genome. Nat Genet 36, 949-51 (2004).

34. Lin, M. et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 20, 1233-40 (2004).
35. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132, 365-86 (2000).
36. Nickerson, D. A., Tobe, V. O. & Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 25, 2745-51 (1997).
37. Rubin, M. A. et al. Overexpression, amplification, and androgen regulation of TPD52 in prostate cancer. Cancer Res 64, 3814-22 (2004).
38. Phelps, R. M. et al. NCI-Navy Medical Oncology Branch cell line data base. J Cell Biochem Suppl 24, 32-91 (1996).
39. Banks-Schlegel, S. P., Gazdar, A. F. & Harris, C. C. Intermediate filament and cross-linked envelope expression in human lung tumor cell lines. Cancer Res 45, 1187-97 (1985).
40. Naldini, L. et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272, 263-7 (1996).
41. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat Biotechnol 15, 871-5 (1997).
42. Kohno, T., Takahashi, M., Manda, R. & Yokota, J. Inactivation of the PTEN/MMAC1/TEP1 gene in human lung cancers. Genes Chromosomes Cancer 22, 152-6 (1998).
43. Yokomizo, A. et al. PTEN/MMAC1 mutations identified in small cell, but not in non-small cell lung cancers. Oncogene 17, 475-9 (1998).
44. Mori, N. et al. Variable mutations of the RB gene in small-cell lung carcinoma. Oncogene 5, 1713-7 (1990).
45. Herman, J. G. et al. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res 55, 4525-30 (1995).
46. Merlo, A. et al. 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nat Med 1, 686-92 (1995).
47. Lindblad-Toh, K. et al. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. Nat Biotechnol 18, 1001-5 (2000).
48. Balsara, B. R. et al. Comparative genomic hybridization analysis detects frequent, often high-level, overrepresentation of DNA sequences at 3q, 5p, 7p, and 8q in human non-small cell lung carcinomas. Cancer Res 57, 2116-20 (1997).
49. Bjorkqvist, A. M. et al. Comparison of DNA copy number changes in malignant mesothelioma, adenocarcinoma and large-cell anaplastic carcinoma of the lung. Br J Cancer 77, 260-9 (1998).
50. Luk, C., Tsao, M. S., Bayani, J., Shepherd, F. & Squire, J. A. Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization. Cancer Genet Cytogenet 125, 87-99 (2001).
51. Petersen, I. et al. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. Cancer Res 57, 2331-5 (1997).

52. Testa, J. R. et al. Cytogenetic analysis of 63 non-small cell lung carcinomas: recurrent chromosome alterations amid frequent and widespread genomic upheaval. Genes Chromosomes Cancer 11, 178-94 (1994).
53. Bjorkqvist, A. M. et al. DNA gains in 3q occur frequently in squamous cell carcinoma of the lung, but not in adenocarcinoma. Genes Chromosomes Cancer 22, 79-82 (1998).
54. Fukuyama, K. et al. MAPK upstream kinase (MUK)-binding inhibitory protein, a negative regulator of MUK/dual leucine zipper-bearing kinase/leucine zipper protein kinase. J Biol Chem 275, 21247-54 (2000).
55. Stanton, S. E., Shin, S. W., Johnson, B. E. & Meyerson, M. Recurrent allelic deletions of chromosome arms 15q and 16q in human small cell lung carcinomas. Genes Chromosomes Cancer 27, 323-31 (2000).
56. Petersen, I. et al. Small-cell lung cancer is characterized by a high incidence of deletions on chromosomes 3p, 4q, 5q, 10q, 13q and 17p. Br J Cancer 75, 79-86 (1997).
57. Kamb, A. et al. A cell cycle regulator potentially involved in genesis of many tumor types. Science 264, 436-40 (1994).
58. Nobori, T. et al. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. Nature 368, 753-6 (1994).
59. Cairns, P. et al. Rates of p16 (MTS1) mutations in primary tumors with 9p loss. Science 265, 415-7 (1994).
60. Spruck, C. H., 3rd et al. p16 gene in uncultured tumours. Nature 370, 183-4 (1994).
61. Stallings, R. L. et al. High-resolution analysis of chromosomal breakpoints and genomic instability identifies PTPRD as a candidate tumor suppressor gene in neuroblastoma. Cancer Res 66, 3673-80 (2006).
62. Tamura, K. et al. Deletion of three distinct regions on chromosome 13q in human non-small-cell lung cancer. Int J Cancer 74, 45-9 (1997).
63. Testa, J. R. & Graziano, S. L. Molecular implications of recurrent cytogenetic alterations in human small cell lung cancer. Cancer Detect Prev 17, 267-77 (1993).
64. Ng, P. C. & Henikoff, S. Predicting deleterious amino acid substitutions. Genome Res 11, 863-74 (2001).
65. Ng, P. C. & Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. Genome Res 12, 436-46 (2002).
66. Marchetti, A. et al. mdm2 gene amplification and overexpression in non-small cell lung carcinomas with accumulation of the p53 protein in the absence of p53 gene mutations. Diagn Mol Pathol 4, 93-7 (1995).
67. Fakharzadeh, S. S., Trusko, S. P. & George, D. L. Tumorigenic potential associated with enhanced expression of a gene that is amplified in a mouse tumor cell line. Embo J 10, 1565-9 (1991).
68. Little, C. D., Nau, M. M., Carney, D. N., Gazdar, A. F. & Minna, J. D. Amplification and expression of the c-myc oncogene in human lung cancer cell lines. Nature 306, 194-6 (1983).
69. Reissmann, P. T., Koga, H., Figlin, R. A., Holmes, E. C. & Slamon, D. J. Amplification and overexpression of the cyclin D1 and epidermal growth factor receptor genes in non-small-cell lung cancer. Lung Cancer Study Group. J Cancer Res Clin Oncol 125, 61-70 (1999).

70. Shiraishi, M., Noguchi, M., Shimosato, Y. & Sekiya, T. Amplification of protooncogenes in surgical specimens of human lung carcinomas. Cancer Res 49, 6474-9 (1989).
71. 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, 2129-39 (2004).
72. Paez, J. G. et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304, 1497-500 (2004).
73. 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 U S A 101, 13306-11 (2004).
74. Sordella, R., Bell, D. W., Haber, D. A. & Settleman, J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. Science 305, 1163-7 (2004).
75. Wikman, H. et al. CDK4 is a probable target gene in a novel amplicon at 12q13.3-q14.1 in lung cancer. Genes Chromosomes Cancer 42, 193-9 (2005).
76. Pulciani, S., Santos, E., Long, L. K., Sorrentino, V. & Barbacid, M. ras gene Amplification and malignant transformation. Mol Cell Biol 5, 2836-41 (1985).
77. Rodenhuis, S. et al. Mutational activation of the K-ras oncogene. A possible pathogenetic factor in adenocarcinoma of the lung. N Engl J Med 317, 929-35 (1987).
78. Keyomarsi, K. & Pardee, A. B. Redundant cyclin overexpression and gene amplification in breast cancer cells. Proc Natl Acad Sci U S A 90, 1112-6 (1993).
79. Snijders, A. M. et al. Genome-wide-array-based comparative genomic hybridization reveals genetic homogeneity and frequent copy number increases encompassing CCNE1 in fallopian tube carcinoma. Oncogene 22, 4281-6 (2003).
80. Stephens, P. et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. Nature 431, 525-6 (2004).
81. Kao, J. & Pollack, J. R. RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. Genes Chromosomes Cancer 45, 761-9 (2006).
82. Berenson, J. R. et al. Frequent amplification of the bcl-1 locus in poorly differentiated squamous cell carcinoma of the lung. The Lung Cancer Study Group. Oncogene 5, 1343-8 (1990).
83. Betticher, D. C. et al. Prognostic significance of CCND1 (cyclin D1) overexpression in primary resected non-small-cell lung cancer. Br J Cancer 73, 294-300 (1996).
84. Toomes, C. et al. The presence of multiple regions of homozygous deletion at the CSMD1 locus in oral squamous cell carcinoma question the role of CSMD1 in head and neck carcinogenesis. Genes Chromosomes Cancer 37, 132-40 (2003).
85. Sato, M., Mori, Y., Sakurada, A., Fujimura, S. & Horii, A. The H-cadherin (CDH13) gene is inactivated in human lung cancer. Hum Genet 103, 96-101 (1998).