ONLINE DATA SUPPLEMENT

COMPOSITE ANATOMIC-CLINICAL-MOLECULAR PROGNOSTIC MODEL IN NON-SMALL CELL LUNG CANCER

A multicenter study of 512 patients

Angel López-Encuentra, Fernando López-Ríos, Esther Conde, Ricardo García-Luján, Ana Suárez-Gauthier, Nuria Mañes, Guadalupe Renedo, José Luis Duque-Medina, Elena García-Lagarto, Ramón Rami-Porta, Guadalupe González-Pont, Julio Astudillo-Pombo, José Luis Maté-Sanz, Jorge Freixinet, Teresa Romero-Saavedra, Monserrat Sánchez-Céspedes, Agustin Gómez de la Cámara, on behalf of the Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery (GCCB-S)

INDEX

microarrays (TMA) and the population in which there was no material available to perform such study

b. Comparison of the clinical, histological and anatomic extent characteristics between the selected population (NSCLC in early pI-II stages with complete resection) where the immunohistochemical study with tissue microarrays (TMA) from 6 hospitals (population A; n=512) was performed and the population where the immunohistochemical study was not possible (population B; n=982) within the same working group (GCCB-S) from the rest of the GCCB-S population

5- Molecular markers – tissue microarrays

a. Selection of molecular markers
b. Performance of tissue microarrays
c. Assessment of samples
d. Analysis of interobserver agreement
e. Specific assessment methods by protein

1- BRONCHOGENIC CARCINOMA COOPERATIVE GROUP OF THE SPANISH SOCIETY OF PNEUMOLOGY AND THORACIC SURGERY (GCCB-S)

The GCCB-S began its work in 1993 and prospectively gathered 2,994 cases of non-small cell lung cancer (NSCLC) up till 1997, performing the following analyses during such period:

- Quality control with audit of cases and centres (1)
- Homogeneity in the classification of staging pN (1)
- Homogeneity of operability and resectability criteria in the surgical indication (2)
- Analysis of classification certainty in clinical and pathological staging (C factor – TNM-UICC classification) (3, 4)
- Homogeneity in the classification of surgery (complete, incomplete, etc) (5)
- Interhospital homogeneity in morbidity and mortality (6)
- Concordance study in clinical and pathological staging (7)

REFERENCES

1- Grupo Cooperativo de Carcinoma Broncogénico de SEPAR (GCCB-S). Control de calidad en un registro multiinstitucional de carcinoma broncogénico. *Arch Bronconeumol* 1996;32 (suppl 2):70.

2- López Encuentra A, and the Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery (GCCB-S). Criteria of functional and oncological operability in surgery for lung cancer. A multicenter study. *Lung Cancer* 1998;20:161-168

3- López Encuentra A, Gómez de la Cámara A, Varela de Ugarte A, Mañes N, Llobregat N y GCCB-S. El "fenómeno Will-Rogers". Migración de estadios en carcinoma broncogénico, tras aplicar criterios de certeza clasificatoria. *Arch Bronconeumol* 2002;38:166-171.

4- López Encuentra A, Gómez de la Cámara A, for the Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery (GCCB-S). Validation of a central review board of staging prior to surgery for non-small-cell lung cancer – impact on prognosis: a multicenter study. *Respiration* 2002;69:16-24.

5- Rami-Porta R, Mateu-Navarro M, Freixinet J de la Torre M, Torres-García AJ, Pun YW, Cantó Armengod A; on behalf of the Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery GCCB-S. Type of resection and prognosis in lung cancer. Experience of a multicenter study. *Eur J Cardiothorac Surg* 2005;28:622-628.

6- Freixinet JL, Julià-Serdà G, Rodríguez PM, Santana NB, Rodríguez de Castro F, Fiuza MD, López-Encuentra A, and Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery GCCB-S. Hospital volume: operative morbidity, mortality and survival in thoracotomy for lung cancer. A Spanish multicenter study of 2994 cases. *Eur J Cardio-thorac Surg* 2006;29:20-25.

7- López-Encuentra A, García-Luján R, Rivas JJ, Rodríguez-Rodríguez J, Torres-Lanza J, Varela-Simo G and GCCB-S. Comparison between clinical and pathological staging in 2,994 cases of lung cancer. *Ann Thorac Surg* 2005; 79: 974-9. **A.**

2- METHODS

Population

All patients included in the study had non-small cell lung cancer (NSCLC) in early stages and underwent thoracotomy with intent to cure in hospitals pertaining to the Bronchogenic Carcinoma Cooperative Group of the Spanish Society of Pneumology and Thoracic Surgery (GCCB-S). All patients treated surgically from October 1993 to September 1997 in hospitals participating in the GCCB-S were prospectively included. The annual cumulative number of cases was close to 50% of the surgical cases occurring in Spain. The participating GCCB-S centres had a wide variety of activities, including a representative range of number of beds, teaching or research activities (university and non-university hospitals), public and private ownership, and number of interventions per year. There were no differences between hospitals in terms of operative mortality or survival. The sample was complete, as verified by the inclusion in the registry of all patients undergoing surgery, including incomplete resections and exploratory thoracotomies.

The initial number of patients included in this study was 2,994. These patients are part of the worldwide database used by the International Staging Committee of the International Association for the Study of Lung Cancer (IASLC) for the creation of the 7th edition of the TNM classification for lung cancer, published in 2009. Given that the last case of this series underwent operation on the 30th of September 1997, we now have available a mortality experience or real survival rates of more than 10 years of follow-up.

In order to carry out this work, 512 patients with pathologic stage I and II non-small cell lung cancer (NSCLC) with complete resection (R0), who underwent thoracotomy in 6 hospitals of the GCCB-S between October 1993 and September 1997, have been prospectively and consecutively selected. No differences were detected in the characteristics of this population compared with those of the rest of the cases of the GCCB-S.

Patients in whom induction therapy had been administered and patients who had died in the postoperative period as a direct result of surgical treatment regardless of the time and place of death, were excluded.

Only 30 patients (5.8%) did not have adequate follow-up, but their characteristics did not differ from those of the remaining 482 patients.

TNM classification and clinical variables

The 7th edition of the tumour, node, and metastasis (TNM) classification for lung cancer published in 2010 (TNM2010) was used. In the proposal for the new TNM2010 classification, stages I and II are similar to the stages of the previous 1997 classification, except for the fact that nodules in the same lobe of the primary tumour are now situated within the T3 category; in the previous staging system that situation was classified as T4 and, therefore, outside stages I and II. Pathological staging is understood to be a classification obtained through findings observed upon thoracotomy with a microscopic examination of the excised specimens, together with the data obtained through the clinical staging.

In accordance with the initial design, the period of patient recruitment was short (1993-1997). The same criteria for the functional operability of patients and oncological operability of the tumour were used in all the GCCB-S hospitals.

The GCCB-S prospectively designed and used a single registry for all hospitals collecting NSCLC tumour extent data (TNM-stages classification). The presence-absence of each descriptor contained in the 1997-2010 TNM classification, whether qualitative (e.g., presence or absence of visceral pleural involvement) or quantitative (e.g., tumour size in cm), in the registry was identified in this section. The variable pT3di (direct invasion) was created for this work if any of the following were present: direct invasion of diaphragm, phrenic nerve, mediastinal pleura, pericardium, extrapericardial pulmonary artery or extrapericardial pulmonary vein involvement. The descriptor "other nodules in the same lobe of the primary tumour" has been considered in this study given that, in the TNM2010 staging, it is included in category T3.

Surgical-pathological N0 (pN0) was determined by radical mediastinal lymph node dissection or systematic sampling of at least four lymph node areas (2 [only in right lung

cancer], 4, 7, and 10 on the same side as the tumour). Moreover, no nodes must be present in the aortopulmonary window nor in the anterior mediastinal areas (areas 5 and 6), if the NSCLC is left-sided (upper lobe or main left bronchus). In order to classify the presence or absence of mediastinal lymph node involvement, a randomised study demonstrated that systematic sampling had a similar value to that of radical mediastinal lymph node dissection. These minimum requirements (systematic sampling or radical mediastinal lymph node dissection) are essential to classify pN0 as there are no prognostic differences when compared to only random lymph node sampling.

Internal and external audits were made to survey the ratio between the number of patients undergoing surgical treatment and the patients included in the registry (standard over 95%), the presence and validity of the data recorded for each case (standard over 70%), including the consistency of tumour staging. The criterion for the validity of the survival data was established as the existence of a known follow-up for 85%, or more, of the patients registered in each hospital. In the hospitals that did not meet all these conditions, the cases corresponding to the period of problems were excluded. Finally, correct data transmission by a single central office from the paper record to the computer database was verified.

Definition of complete resection: free resection margins, no extracapsular nodal involvement, no involvement of the most distant removed nodes, and no positive pleural effusion. Incomplete resection: positive resection margins, extracapsular nodal involvement, unremoved positive nodes, involvement of the most distant removed nodes, positive pleural effusion, and pleural implants.

These procedures were designed to control the selection biases of surgical cases, registered cases out of the total number of surgical cases, sample size, type of hospital, prognostic migration due to the prolonged period of patient recruitment, classification with low or deficient degrees of certainty, contamination by data from incomplete series or erroneous data, and loss of long-term follow-up.

Histological and molecular study

For the histological study, all of the samples (pneumonectomy or lobectomy-bilobectomy specimens or segmental pulmonary resections) were fixed with 10% formol and embedded in paraffin. All surgical pieces were studied according to standard protocol.

Fifteen blocs of non-tumour lung and 3 amygdala samples without significant histological lesions were selected as controls for the performance of tissue microarrays (TMAs) with the aim of having internal controls of the technique in each TMA.

All histological preparations from each of the cases and controls were simultaneously and independently examined by three pathologists, who reassessed the anatomopathological diagnosis based on the latest Lung Tumour classification system of the World Health Organisation (WHO), which was developed in 2004. Any discrepancy between pathologists was resolved via consensus at the time of assessment.

The selection of proteins to be assessed for the molecular analysis in this study was performed based on a comprehensive review of the literature, as well as on the group's prior experience in molecular disturbances of lung carcinomas. The selection of molecular markers was based on their availability as well as their indication for use in paraffin-embedded material. To this end, immunohistochemistry has been used to study the levels of different proteins which play a key role in biological functions such as: a) the ability to generate their own mitogenic signals, b) resistance to exogenous influence of growth-inhibiting signals, c) evasion of apoptosis (programmed cell death), d) unlimited proliferation, e) ability to create their own vascular network (angiogenesis) and f) ability to invade other tissues.

A sample of about 500 patients was considered adequate for the purpose of this study. In choosing the sample size, the expected presence of a 40-50% event of death within a 5-year time inverval from time zero for the calculation of survival was taken into account, as well as the 35 variables left for multivariate analysis after univariate prognostic analysis. The initial available variables (over 200) were included in 4 different groups. The TNM-histology group contained all qualitative and quantitative descriptors that define each T-N category of the stages selected for this

work - 2010 pI-II stages - as well as tumour types (squamous cell carcinoma, adenocarcinoma) and the tumour differentiation grade (*Table E1. Supplementary appendix).*

Variables

The group of clinical variables included the presence of weight loss, the degree of performance status (ECOG), body mass index, comorbidity, etc. *(Table E1. Supplementary appendix).* The group of analytical and functional variables included haemoglobin, leukocyte, polymorphonuclear and lung function values, amongst others. *(Table E1. Supplementary appendix).* The group of molecular variables included 32 markers that explore five biochemical pathways *(Table E1. Supplementary appendix).*

Analysis procedures

Several steps were undertaken to build the predictive model: First, for univariate analysis for selection of significant prognostic variables, the Kaplan Meier method was used and a p value < 0.3 was chosen as threshold (1). Second, for each group a classification tree was built by recursive partitioning. We consider vital status at 5-year survival as dependent variable at each terminal node of the classification tree. This multivariate recursive partitioning was performed with the supervised learning classification algorithm C4.5 (2) constructed with the R interface to Weka (3). Minimal and maximal probabability terminal nodes of each group are displayed. Third, an Integrated Group was built with the obtained variables in every group of the second step. Fourth, five -year probability of survival via Kaplan Meier were calculated for terminal nodes of the Integrated Group (4). The model's ability to discriminate amongst patients with or without the event was assessed using the area under the receiver operative characteristics (ROC) curve (AUC) method, measured by the concordance index (C index) (5) and its overall predictive capacity with the coefficient of determination using Nalgerkerke's R^2 parameter, which expresses the amount of variability in results that can be explained by the selected predictive variables (6). The STATA programme was used for the remaining results (7).

These recursive partitioning method (supervised learning classification algorithm) was used for the multivariate analysis due to its ability to identify groups with similar prognosis, place the different prognostic factors in a hierarchy, and demonstrate the relationships between them. The recursive partitioning method may reutilize the same variable in different segments of the tree. In each segment, the most convenient cut-point that optimizes the partition criterion is chosen Continuous variables are dichotomized to create two groups with different probability of event. For ordinal and continuous predictors, the split is of the form X<c versus X>=c.

The classification tree was developed with the following conditions: In case a node is amplified, there should be two branches per node, only; the minimum number of individuals per branch had to be greater than 10; the confidence factors used for pruning were 0.05, 0.10, 0.15, 0.20, and 0.25 (smaller values provoke less tree depth, that is, fewer associated characteristics to the profile). The tree that combines complexity (fewer number of profiles) and discriminatory complexity (greater area under the curve) is selected. No statistical contrast was used to select the tree. For internal validation, stricter criteria were used because the process is automatic: in case a node is amplified, there should be two branches per node; the minimum number of individuals per branch had to be greater than 25; the confidence factor used for pruning was 0.15.

The internal validity of the model's estimations was calculated via bootstrapping. The same process that was performed on the development sample was performed on every bootstrap sample. Each model obtained in the bootstrap sample was assessed in the bootstrap sample and in the general sample. The expected optimism estimate was calculated from the average difference between estimations. The bootstrap procedure lead to estimates of the optimism-corrected performance, which was calculated as apparent performance minus optimism. Means and empirical standard errors were shown, 1000 bootstrap repetitions were used for calculation of both the mean and SE. The test performance was defined as the performance of the models from the bootstrap samples when applied to the original sample. The expected optimism was calculated as the difference between bootstrap performance and test performance. The optimism-corrected performance was defined as apparent performance-optimism.

In our database, a preference for digits was observed in the tumour size variable when the frequency distribution was observed. The procedure to categorize the continuous variable was established to achieve the objective that all created groups reflected the effect, that is, to collapse categories with no effect in order to obtain a classification determined by data. An incremental coding (nested coding) is then created and it can be useful when one wishes to compare each category against its immediate predecessors (8) and to select categories than can be excluded (8, 9). This statistical technique is useful for ordinal categorical variables or quantitative variables that we want to categorize, either for manageability or because of some problem such as terminal digit preference (10). The tumour size variable was categorized according to the Schoenfeld procedure (8, 9). The other quantitative variables did not present with this problem.

REFERENCES (Analysis procedures)

1- Kaplan EL, Meier P. Non parametric estimation from incomplete observations. *JAMA* 1958;53:457-81.

2- Quinlan R (1993). C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.

3- Hornik K, Buchta CH, Zeileis A. Open-Source Machine Learning: R Meets Weka. *Computational Statistics* 2009;24:225-232. doi:10.1007/s00180-008-0119-7

4- Ture M, Tokatli F, Kurt I. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications* 2009;36:2017–26.

5- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.

6- Nagelkerke, NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691-692.

7- StataCorp. 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.

8- Rothman KJ, Greenland S. Modern epidemiology. 1998 Lippincott Williams & Wilkins, 2nd edition.

9- Schoenfeld, DA. Analysis of categorical data: logistic model. In: Mike V, Stanley KE, eds. Statistics in medical research. New York, NY: Wiley, 1982; 443-454.

10- López-Encuentra A, Duque-Medina JL, Rami-Porta R, de la Cámara AG, Ferrando P; Bronchogenic Carcinoma Co-operative Group of the Spanish Society of Pneumology and Thoracic Surgery. Staging in lung cancer: is 3 cm a prognostic threshold in pathologic stage I non-small cell lung cancer? A multicenter study of 1,020 patients. *Chest* 2002;121:1515-20.

Other references

- Harrell FE, Jr. Regression modelling strategies. Springer- Verlag. New York. 2001.
- Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774-781.
- Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques, 2nd Edition. Morgan Kaufmann. San Francisco, CA. 2005.
- Breiman L, Friedman J, Olsen RA, Stone CJ. Classification and Regression Trees. Wadsworth. Monterey. CA. 1984.
- Cox DR. Regression models and life table. *J R Stat Soc* 1972;34:187-220.

- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2008 http://www.R-project.org.
- Therneau TM, Atkinson B. R port by Brian Ripley. rpart: Recursive Partitioning.R package version 3.1-42. 2008. http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm

3- DESCRIPTIVE DATA: CLINICAL, ANATOMIC EXTENT AND MOLECULAR BIOLOGY VARIABLES (n: 512)

Mean tumour size as measured in the surgical specimens from all 512 cases was 4.55 cm (SD:2.45), median tumour size was 4 cm with a 25 percentile of 3 cm, a 50 percentile of 4 cm and a 75 percentile of 5.5 cm. Other anatomic extent classification and histological type data, categorical or numeric clinical data, treatment-related data or data concerning molecular markers are presented in the following table.

Table E1 (Supp)
DESCRIPTIVE DATA. CATEGORICAL VARIABLES (n: 512)

T-Descriptor (pathologic)	**Frequency (%)**
Visceral pleural involvement	114 (22.3%)
Parietal pleural involvement	12 (2.3%)
Thoracic wall involvement	11 (2.1%)
Involvement at a bronchial level:	
-primary at <2 cm tracheal carina	7 (1.4%)
-primary at ≥2 cm tracheal carina	26 (5.1%)
-lobar	121 (23.6%)
-more distal	358 (69.9%)
Another nodule in the same lobe	13 (2.5%)
Atelectasis:	
-Less than entire lung	126 (24.6%)
-Complete lung	6 (1.2%)
Upper sulcus tumour	0 (0%)
Diaphragmatic involvement	1 (0.2%)
Vertebral involvement	0 (0%)
Phrenic nerve involvement	2 (0.4%)
Recurrent nerve involvement	0 (0%)
Mediastinal pleural involvement	1 (0.2%)
Pericardial involvement	3 (0.6%)
Atrial involvement	0 (0%)
Extrapericardial pulmonary artery involvement	7 (1.4%)
Extrapericardial pulmonary vein involvement	8 (1.6%)
Vena cava involvement	0 (0%)
Mediastinal tissue involvement	0 (0%)
Oesophageal involvement	0 (0%)
Aortic involvement	0 (0%)
Cytologically negative pleural effusion	4 (0.8%)
Cytologically negative pericardial effusion	3 (0.6%)

pT classification	**Frequency (%)**
pT1	107 (20.9%)
pT2	365 (71.3%)
pT3	40 (7.8%)

p: pathologic
(the presence of an additional tumour nodule in the same lobe of the primary tumour is not considered an element for pT classification)

pN classification	**Frequency (%)**
pN0	430 (84%)
pN1	82 (16%)

Table E1 (Supp) (Cont.)
DESCRIPTIVE DATA. CATEGORICAL VARIABLES (n: 512)

Histological type	**Frequency (%)**
Squamous cell carcinoma	324 (63.3%)
Adenocarcinoma	117 (22.9%)
Large cell carcinoma	62 (12.1%)
Others	9 (1.8%)
Specific histological components	**Frequency (%)**
Bronchiolo-alveolar component	42 (8.2%)
Presence of mucus	10 (2%)
Neuroendocrine component	19 (3.7%)
Clinical data (categorical)	**Frequency (%)**
Male sex	474 (92.6%)
Active smoker	274 (53.5%)
Previous tumour	93 (18.2%)
Weight loss >10%	58 (11.3%)
Ischemic heart disease	34 (6.6%)
COPD	234 (45.7%)
Peripheral vascular disease	54 (10.5%)
Systemic arterial hypertension	90 (17.6%)
Diabetes	58. (11.3%)
Casual finding	173 (33.8%)
Dyspnoea grade	
- no dyspnoea	351 (68.6%)
- grade 1: little dyspnoea	132 (25.8%)
- grade >1	29 (5.7%)
Right lung location	276 (53.9%)
ECOG (grade 0)	115 (22.5%)
(grade 1)	391 (76.4%)
(grade 2)	5 (1%)
(grade 3)	1 (0.1 %)
Treatment-related data (*)	**Frequency (%)**
Perioperative transfusion	120 (23.4%)
Postoperative morbidity	201 (39.3%)
Pneumonectomy	114 (22.3%)
Lobectomy or bilobectomy	336 (65.6%)
Sublobar resections or combination of all types of resections	62 (12.1%)

COPD: chronic obstructive pulmonary disease; ECOG: Eastern Co-operative Oncology Group

(*) as per the study's initial design, patient selection has only included patients with complete resections and excluded patients with incomplete resections, induction therapy and patients who died in the perioperative period at any time or place.

Table E1 (Supp) (Cont.)
DESCRIPTIVE DATA. CONTINUOUS VARIABLES

	Mean	SD	Median	Percentile 25	Percentile 50	Percentile 75
Age	65.5	8.3	67	60	67	71
Body mass index	25.5	3.9	25.4	22.9	25.4	27.8
Smoking:packs/year	52	36.4	50	29.3	50	75
FEV1%	82.5	20	81.8	68.3	81.8	95.8
Haemoglobin (gr)	13.9	1.78	14.1	12.8	14.1	15.1
Leukocytes (mm^3)	8.776	2.842	8.330	6.865	8.330	10.000
Albumin (gr)	3.97	0.7	4.07	3.5	4	4.5

FEV1: Forced expiratory volume in 1 second.

Molecular markers (immunohistochemical)	Total number	Negative	Positive (%)	Not assessable
CELL CYCLE				
-Cyclin A	334	67	259 (77.5%)	8
-Cyclin B_1	143	112	27 (18.9%)	4
-Cyclin D_1	334	142	184 (55.1%)	8
-Cyclin E	334	178	148 (44.3%)	8
-CDK 2	334	86	239 (71.6%)	9
-CDK 6	334	122	192 (57.5%)	20
-P16	334	199	119 (35.6%)	16
-P21	143	87	52 (36.4%)	4
-P27	512	228	27.5 (55.7%	9
-RB	512	94	413 (80.7%)	5
-Ki67	512	150	362 (70.7%)	0
-CDC6	512	269	229 (44.7%)	14
APOPTOSIS				
-P53	512	214	295 (57.6%)	3
-MDM2	143	124	11 (7.7%)	5
-BCL2	334	225	105 (31.4%)	4
-Caspase 3	512	260	234 (45.7%)	12
-Survivin	334	185	142 (42.5%)	7
-NFKβ (p65)	320	154	144 (45%)	22
-FAS (CD95)	334	281	40 (12%)	13
ADHESION MOLECULES				
-β-catenin	333	225	98 (29.84%)	10
-E-cadherin	333	182	142 (42.5%)	10
SIGNAL TRANSDUCTORS				
-phospho-AKT	512	192	307 (60%)	13
-phospho-mTOR	512	229	267 (52.1%)	16
-EGFR	512	224	275 (53.7%)	13
-HER2-Herceptest	512	470	29 (5.7%)	13
-phospho-ACC	512	276	224 (43.8%)	12
-phospho-S6	512	288	201 (39.3%)	23
-LKB1	334	52	249 (74.6%)	33
OTHERS				
-COX2	512	375	118 (23%)	19
-TTF1	334	255	67 (20%)	12
-P63	512	215	284 (55.5%)	13
-γCP4	143	105	35 (24.5%)	3

4- ANALYSIS OF POTENTIAL SELECTION BIASES

4-a.- Comparison of clinical, histological and anatomic extent characteristics between the selected population (NSCLC in initial pI-II stages with complete resection) in which it was possible to perform a immunohistochemical analysis using tissue microarrays (TMA) and in the population in which there was no material available to perform such study

To assess the potential population selection bias in a tissue matrices study all patients with NSCLC in early stages (pI-II) who underwent surgical treatment in a centre of the GCCB-S between October 1993 and September 1997.

The total population was divided into 2 groups: one group for which there was sufficient material to perform the immunohischemistry study in tissue matrices and another group for which there was not. Of the total 180 cases, in 34 (18%) cases it was not possible to retrieve sufficient material. In 32 cases the sample had been exhausted in the conventional histological study and in the 2 remaining cases the material was very necrotic.

The comparative analysis between both groups detected significant differences in the smaller tumour size in the group in which it was not possible to perform immunohistochemistry analysis (mean 3.3 cm versus 4.5 cm; $p = 0.04$) and in the extension of resection with a higher probability of lobectomy in the group in which it was not possible to conduct this type of studies (79% versus 57%; $p = 0.02$).

No statistically significant differences were found between the 2 groups regarding the following variables: sex, age, presence of active smoking, chronic obstructive pulmonary disease, peripheral valvulopathy, arterial hypertension, diabetes, perioperative transfusion, FEV1 percentage over theoretical value (FEV1%), pathologic stage, histological type or survival.

REFERENCE: García Luján R, López-Encuentra A, Conde Gallego E, López-Ríos F, Martín de Nicolas JL, Sánchez Céspedes M, Echave Sustaeta JM, Villena Garrido V, García Quero C. Análisis de marcadores moleculares en carcinoma broncogénico no microcítico resecado: Estudio de sesgos de selección y expresión según la histología. *Rev Patol Respir* 2008;11:161-167.

4-b.- Comparison of the clinical, histological and anatomic extent characteristics between the selected population from 6 hospitals of the GCCB-S (NSCLC in early pI-II stages with complete resection) where the immunohistochemical study with tissue microarrays (TMA) (population A; n=512) was performed and the population of the GCCB-S with the same characteristics where the immunohistochemical study was not possible (population B; n=982).

Table E2
DESCRIPTIVE DATA

T-Descriptor (pathologic)	**Population A Frequency (%)**	**Population B Frequency (%)**
Visceral pleural involvement	114 (22.3%)	264 (27.1%)
Parietal pleural involvement	12 (2.3%)	19 (1.9%)
Thoracic wall involvement	11 (2.1%)	18 (1.8%)
Involvement at a bronchial level:		
-primary at <2 cm tracheal carina	7 (1.4%)	3 (0.3%)
-primary at ≥2cm tracheal carina	26 (5.1%)	83 (8.5%)
-lobar	121 (23.6%)	229 (23.3%)
-more distal	358 (69.9%)	667 (67.9%)
Another nodule in the same lobe	13 (2.5%)	20 (2%)
Atelectasis:		
-less than the entire lung	126 (24.6%)	280 (28.5%)
-complete pulmonary	6 (1.2%)	10 (1%)
Upper sulcus tumour	0 (0%)	2 (0.2%)
Diaphragmatic involvement	1 (0.2%)	3 (0.3%)
Vertebral involvement	0 (0%)	0 (0%)
Phrenic nerve involvement	2 (0.4%)	2 (0.2%)
Recurrent nerve involvement	0 (0%)	0 (0%)
Mediastinal pleural involvement	1 (0.2%)	6 (0.6%)
Pericardial involvement	3 (0.6%)	7 (0.7%)
Atrial involvement	0 (0%)	0 (0%)
Extrapericardial pulmonary artery involvement	7 (1.4%)	8 (0.8%)
Extrapericardial pulmonary vein involvement	8 (1.6%)	10 (1%)
Vena cava involvement	0 (0%)	0 (0%)
Mediastinal tissue involvement	0 (0%)	0 (0%)
Oesophageal involvement	0 (0%)	0 (0%)
Aortic involvement	0 (0%)	0 (0%)
Cytologically negative pleural effusion	4 (0.8%)	17 (1,7%)
Cytologically negative pericardial effusion	3 (0.6%)	2 (0.2%)

pT classification	**Population A Frequency (%)**	**Population B Frequency (%)**
pT1	107 (20.9%)	181 (18.4%)
pT2	365 (71.3%)	743 (75.7%)
pT3	40 (7.8%)	58 (5.9%)

(The presence of an additional tumour nodule in the same lobe of the primary tumour is not considered an element for classification)

Table E2 (Cont)
DESCRIPTIVE DATA

pN classification	**Population A Frequency (%)**	**Population B Frequency (%)**
pN0	430 (84%)	812 (82.7%)
pN1	82 (16%)	170 (17.3%)
Histological type	**Population A Frequency (%)**	**Population B Frequency (%)**
Squamous cell carcinoma	324 (63.3%)	590 (60.1%)
Adenocarcinoma	117 (22.9%)	259 (26.4%)
Others	71 (13.9%)	133 (13.5%)
Clinical data (categorical)	**Population A Frequency (%)**	**Population B Frequency (%)**
Male sex	474 (92.6%)	904 (92.1%)
Active smoker	274 (53.5%)	551 (56.1%)
Previous tumour	93 (18.2%)	143 (14.6%)
Weight loss >10%	58 (11.3%)	60 (6.1%)
Ischemic heart disease	34 (6.6%)	81 (8.2%)
COPD	234 (45.7%)	473 (48.2%)
Peripheral vascular disease	54 (10.5%)	100 (10.2%)
Systemic arterial hypertension	90 (17.6%)	197 (20.1%)
Diabetes	58. (11.3%)	84 (8,6%)
Casual finding	173 (33.8%)	316 (32.2%)
Dyspnoea grade - no dyspnoea	351 (68.6%)	581 (59.2%)
- grade 1: great effort	132 (25.8%)	332 (33.8%)
- grade >1	29 (5.7%)	69 (7%)
Right lung location	276 (53.9%)	512 (52.1%)
ECOG (grade 0)	120 (23.4%)	239 (24.3%)
(grade 1)	381 (74.4%)	739 (75.3%)
(grade >1)	11 (2.2%)	4 (0.4%)
Treatment-related data (*)	**Population A Frequency (%)**	**Population B Frequency (%)**
Perioperative transfusion	120 (23.4%)	158 (16.1%)
Postoperative morbidity	201 (39.3%)	290 (29.5%)
Pneumonectomy	114 (22.3%)	241 (24.5%)
Lobectomy or bilobectomy	336 (65.6%)	642 (65.4%)
Sublobar resections or combination of all types of resection	62 (12.1%)	99 (10%)

Non-categorical clinical data	**Population A**	**Population A**	**Population A**	**Population B**	**Population B**	**Population B**
	Mean	**SD**	**Median**	**Mean**	**SD**	**Median**
Age	65.5	8.3	67	64.9	8.9	66
Body mass index	25.5	3.9	25.4	25.5	3.73	25.3
Smoking:packs/year	52	36.4	50	50	33.9	47
FEV1%	82.5	20	81.8	81.7	19.7	80.6
Haemoglobin (gr)	13.9	1.78	14.1	14	1.68	14.1
Leukocytes (mm^3)	8,776	2,842	8,330	9,360	6,428	8,300
Albumin (gr)	3.97	0.7	4.07	4.05	0.62	4.1

5- MOLECULAR MARKERS – TISSUE MICROARRAYS

SELECTION OF MOLECULAR MARKERS

The selection of the molecules to be studied was performed based on a comprehensive review of the literature and on the prior experience of the authors regarding the molecular alterations of lung cancer. The selection of molecular markers was based on their availability and on their indication for use in paraffin-embedded material. To this end, immunohistochemistry has been used to study the levels of different proteins that play a key role in biological functions such as: a) the ability to generate their own mitogenic signals, b) resistance to the exogenous influence of growth inhibiting signals, c) evasion of apoptosis (programmed cell death), d) unlimited proliferation, e) ability to create their own vascular network (angiogenesis) and f) ability to invade other tissues.

a) *Ability to generate their own mitogenic signals*
Membrane receptors or molecules that intervene in the intracellular transmission of different biological signals are included within this category. Mutated genes in this category include *PTEN, EGFR, KRAS, ERBB2* and *LKB1*. In this project the levels of the EGFR, HER2 (protein coded by the ERBB2 gene) and LKB1 proteins have been tested. There are no commercial antibodies with reproducible results available for PTEN and mutations in KRAS do no alter the protein's expression. However, there are indirect markers of the presence of disturbances in these genes. Hence, the inactivation of PTEN leads to the activation of AKT, and, therefore, to increased protein levels of phosphorylated phospho-AKT, phospho-mTOR (mammalian target of rapamycin) and phospho-S6 (ribosomal protein), markers that have been assessed in this project. Similarly, the inactivation of LKB1 leads to decreased levels of phosphorylated and inactive phospho-ACC (acetyl-CoA carboxylase). Other studies also indicate an association between mutations/genic amplification in *EGFR, ERBB2* and *KRAS* and increased levels of phospho-AKT.

b) Resistance to the exogenous influence of growth-inhibiting signals
Cell cycle markers are included within this category. In lung cancer, p16 inactivating mutations (in non-small cell lung cancer [NSCLC]) and retinoblastoma (RB) (in small-cell lung cancer) have been detected. Mutations in RB are usually associated with high p16 levels. p16 inactivation is perfectly reflected in a total absence of protein. The levels of different markers of the cell cycle are not good indicators of mutations in RB and p16, but do give us an idea of the changes that take place in the regulation of the cell cycle. There are reports that suggest that the presence of Cdc6 overexpression could also be associated with loss of p16 protein expression.

c) *Evasion of apoptosis* (programmed cell death)
Until now, the only marker that has been determined to be mutated in this pathway in lung cancer is *TP53*. P53 protein levels give an approximate idea of the state of the gene. In this case the overexpression of protein *P53* usually indicates the presence of inactivating mutations in the gene. Another indicator of the presence of mutations in *TP53* is the absence of the P21 protein (which controls the cell cycle). P21 levels are controlled at a transcriptional level by *P53*. Overexpression of MDM2 is another marker that indicates the presence of mutations in *TP53*.

d) Unlimited proliferation
The overexpression of telomerase is included within this function. This protein has not been included in the study because it is usually overexpressed in over 90% of tumours.

e) *Ability to create their own vascular network (angiogenesis)*
No markers have been included in this category given that their analysis requires the assessment of the entire section; the analysis cannot be performed using only core biopsy material (TMA cylinders). Until now no mutations of the genes involved in this function have been described.

f) Ability to invade other tissues

Mutated genes included in this category have not yet been identified. However, there are several studies indicating that disturbances in the expression of the genes involved in the degradation of the extracellular matrix or the inhibition of cell adhesions with other cells or with the extracellular environment are factors that favour metastasis. β-catenin and E-cadherin markers have been analysed within this category in this work.

Based on this background, it was decided to include 32 molecular markers grouped by biochemical pathways in this study.

Table E3

CELL CYCLE	cyclin A
	cyclin B1
	cyclin D1
	cyclin E
	CDK2
	CDK6
	P16
	P21
	P27
	RB
	Ki-67
	CDC6
APOPTOSIS	P53
	MDM2
	BCL2
	caspase-3 (active)
	survivin
	NFKβ (p65)
	FAS (CD95)
ADHESION MOLECULES	β-catenin
	E-cadherin
SIGNAL RECEPTORS/TRANSDUCTION	phospho-AKT
	phospho-TOR
	EGFR
	HER2/Herceptest
	phopho-S6
	LKB1
	COX2
OTHERS	TTF-1
	phospho-ACC (Ser79)
	P63
	γCP4

PERFORMANCE OF TISSUE MICROARRAYS (TMA)

1.- Selection of material
1.1.- Selection of histological preparations

For each surgical specimen, all histological preparations contained in the Pathology archives of the six participating hospitals were assessed. Those which contained an ideal amount of tumour and were representative of the entire lesion were selected, excluding areas with necrosis, inflammation or extensive keratinisation.

Two tumour foci, each measuring 1 mm in diameter, were selected from each preparation. It was attempted that these sites be representative areas of each tumour, including different degrees of differentiation or different histological areas in those tumour cases with high histological heterogeneity.

1.2. - Selection of donor blocks

Using the most representative histological preparations of each case, the respective blocs were selected from each archive. From each of the cases and controls the block that best represented the thickness-tumour representativity ratio was selected. To try and homogenise the technique to the best possible all blocks with a thickness of less than 2mm were rearranged. In doing so, mean thickness of the donor blocks used for the construction of the TMA was 2mm.

1.3.- Selection of the area to be biopsied in the donor block

Using the best pair of histological preparation-paraffin block, the chosen area in the preparation was transferred to the corresponding block. The lung is a heterogeneous organ that has characteristic structures that serve as reference points when analysing the surface of a histological block. These structures are the bronchi, vessels, lymph nodes, and pleural surface. The selected sites were marked on each block using permanent ink, as was done in the preparations. The same procedure was followed for the controls.

1.4.- TMA design

This is one of the most critical parts in the performance of this technique. The design included the following points:

- Total number of cases: 512.
- Total number of controls: 59 non-tumour lung, 1 breast carcinoma and 3 amygdalas.
- Cylinder diameter: 1mm
- Separation between cylinders: 1.5mm

Templates were created using a software programme (Excel), identifying all cylinders with coordinates according to row and column, with letters and numbers. Each case had an identification code, which would be used, after obtaining TMA results, to fuse data with the other groups of variables (TNM anatomical extension, clinical and therapeutic variables).

2.- Performance of the receptor bloc

To perform the tissue microarrays (TMA) an arrayer from Beecham Instruments (Hackensack, USA) equipped with micrometric digital measurement was used.

Plain paraffin blocks were constructed to be later used as receptor blocks or master blocks. Two 1mm in diameter cores were taken from each selected block in the previously marked area. The cores were then inserted into the receptor blocks, at a distance separation of 1.5mm between

each other. Two or 3 blocks were selected from the cases with the greatest histological heterogeneity per case. Hence, from 260 cases and 36 controls of non-tumour lung tissue 1 block was selected (2 cylinders / per case); from 12 cases, 2 blocks were selected (4 cylinders / per case); and from 3 cases, 3 blocks were selected (6 cylinders / per case).

From the remaining number of cases (n=261) with 23 controls, three 1mm in diameter cores were taken from one single selected block in the previously marked areas. The tissue cores were then inserted into the receptor blocks, at a distance separation of 1.5mm between each other.

The process consists of the following steps:
STEP 1: A "biopsy" of the area marked in the donor block is performed using a hollow needle.
STEP 2: A needle similar to the one used in Step 1 but with a finer calibre is used to obtain a paraffin mould of the receptor block. Specific coordinates measured with micrometric precision are used.
STEP 3: The cylinder obtained from the donor block is inserted into the gap we have left in the receptor block. The cylinder is pushed using a piston located inside the needle.
STEP 4: Once finished with the receptor block, it is placed inside a 37º C heater for approximately 15 minutes in order to enable a better adherence of the cylinders to the paraffin in the receptor block's holes. Next, using a microscope slide the introduced cylinders are pressed down with the aim of pushing them down to the same height. This avoids loosing the first cuts.

A total of three receptor blocks were used. Two of them contained 176 cylinders, distributed in 11 rows and 16 columns. After the analysis of these two TMAs, a third receptor block with 46 cylinders, distributed in 6 rows and 8 columns, was built. The third block included 8 cases (16 cylinders) which had been previously included in the two prior receptor blocks, but which had tuned out to be non-assessable after the preliminary study (due to tissue loss).

There was a minimal loss of tissue after puncturing the donor blocks, which made it possible to perform new tests, either using via the conventional method or by means of tissue microarrays. Subsequent to this first study, 9 more have been performed, obtaining a total of 12 receptor blocks.

3.- Cutting of TMAs

Seventy-five initial sections (3 micron each) were taken from the receptor blocks following the microtomy traditional histological technique and mounted on xylenated slides. The use of ancillary dissection systems was ruled out from the start (self-adhesive tape, etc).

4.- Storage of the sections

To protect tissue sections from oxidation or other damage and in order to maintain antigenicity following the excision of the section, microscope slides were inserted in liquid paraffin and stored. For their immediate subsequent use, paraffin was removed and the slides left overnight in the heater at 56º C. Next, the slides were inserted in xylol and left to stand inside a heater at 56º C, for 15-30 min.

5.- Immunohistochemical method protocol

Immunohistochemical analysis was carried out on formol-fixed material and embedded in paraffin on the 3-micron TMA sections placed on the xylated microscope slides. The 32 antibodies used, their clones, suppliers, dilutions and unmasking protocol are shown in the following table.

Table E4

Antibody	Clone	Supplier	Dilution	UNMASKING
cyclin A	6 E6	Novocastra	1:50	Citrate buffer 10mM pH6.5
cyclin B1	7A9	Novocastra	1:25	Citrate buffer 10mM pH6.5
cyclin D1	Sp4	Neomarkers	1:10	Citrate buffer 10mM pH6.5 + PK
cyclin E	13A3	Novocastra	1:10	Citrate buffer 10mM pH6.5
CDK2 Ab-2	8D4	NeoMarkers	1:200	No heat
CDK6	Poli-rabbit	BD PharMingen	1:350	Citrate buffer 10mM pH6.5
P16 (F-12)	F-12	Santa Cruz	1:25	Citrate buffer 10mM pH6.5
P21 (WAF1)	EA10	Oncogen	1:25	Citrate buffer 10mM pH6.5
P27	57	BD Transduction Lab	1:1000	Citrate buffer 10mM pH6.5
RB	G3-245	BD PharMingen	1:100	Citrate buffer 10mM pH6.5
Ki-67	MIB-1	DAKO	1:50	Citrate buffer 10mM pH6.5
CDC6	180.2	Santa Cruz	1:25	EDTA 1mM pH 8
P53	DO-7	Novocastra	1:50	Citrate buffer 10mM pH6.5
MDM2	IF2	Oncogen	1:25	Citrate buffer 10mM pH6.5
BCL2	124	DAKO	1:25	Citrate buffer 10mM pH6.5
caspase-3 act	C92-605	BD PharMingen	1:25	Citrate buffer 10mM pH6.5
survivin	Poli-rabbit	RD System	1:1000	Citrate buffer 10mM pH6.5
NFKβ	F-6	Santa Cruz	1:350	Citrate buffer 10mM pH6.5
FAS (CD95)	GM30	Novocastra	1:25	Citrate buffer 10mM pH6.5
β-catenin	14	BD Transduction Lab	1:500	Citrate buffer 10mM pH6.5
E-cadherin	4A2C7	Zymed	1:50	Citrate buffer 10mM pH6.5
phospho-AKT	Poli-rabbit	Cell Signaling	1:25	Citrate buffer 10mM pH6.5
phospho-mTOR	Poli-rabbit	Cell Signaling	1:10	Citrate buffer 10mM pH6.5
EGFR	EGFR.113	Novocastra	1:10	Citrate buffer 10mM pH6.5, 16h, 4ºC
HER2-Herceptest	Policlonal	DAKO	prediluted	Citrate buffer 10mM pH6
phospho-ACC	Poli-rabbit	Cell Signaling	1:25	AR: Trilogy
phospho-S6	Poli-rabbit	Cell Signaling	1:50	Citrate buffer 10mM pH6.5
LKB1	LEY37		1:10	Citrate buffer 10mM pH6.5 + PK
COX2	SP21	NeoMarkers	1:25	Citrate buffer 10mM pH6.5
TTF-1	8G7G3/1	DAKO	1:25	Citrate buffer 10mM pH6.5
P63	4A4	DAKO	1:50	Citrate buffer 10mM pH6.5
γCP4		Zymed	1:1000	

The LSAB (Labelled streptavidin-biotin) visualisation technique was used (DAKO, Glostrup, Denmark) for the following antibodies: cyclin B1, cyclin D1, P21, P 27, Ki-67,CDC6, P53, MDM2, BCL2, caspase-3, survivin, phospho-AKT, and TTF-1. The DAKO EnVision method was used for the remaining antibodies; cyclin A, cyclin E, CDK2, CDK6, P16, RB, NFKβ, FAS, β-catenin, E-cadherin, phospho-mTOR, EGFR, HER2, phospho-ACC, phospho-S6, LKB1, COX-2, P63 and γCP4.

Reference: Nadji M and Morales AR. Immunoperoxidase techniques, a practical approach to tumour diagnosis. Amer Soc Clin Pathol Press. Chicago. Illinois.1986.

ASSESSMENT OF SAMPLES

Two, three, four or six disks measuring 1mm in diameter from each tumour and two cores from each control were analysed. With the aim of ensuring the reproducibility of results simple and reproducible assessment criteria were used. Only three categories were considered in immunohistochemical results: "positive" (1), "negative" (0) and "not assessable (NA)" (6), based on the criteria established in the literature for the correlation between protein expression and the genetic status in the three remaining markers (LKB1, phospho-ACC y phospho-S6).

All immunohistochemical preparations were independently assessed by two pathologists who semiquantitatively assessed the intensity of protein expression and the percentage of positive cells in each cylinder. The assessment was conducted without prior knowledge of any clinical data or the distribution of the cylinders of the same case within the TMA. Cylinders with non-tumour pulmonary parenchyma were also assessed, evaluating the protein expression in alveoli, bronchial and bronchiolar epithelium, as well as inflammatory and structural cells (fibroblasts and endothelial cells).

ANALYSIS OF INTEROBSERVER AGREEMENT

There are, therefore, two types of assessment of results susceptible to the concordance analysis:

- Intratumour agreement; most of them with two results per case and per observer and, in some cases, with 3, 4 or 6 results.
- Interobserver agreement between the two pathologists.

In an interobserver analysis study with the different cores of the same case in 4 markers, a Kappa index for interobserver agreement between 0.9066 (95% CI, 0.80-1) and 1 (95% CI, 0.89-1) was detected (Table E5).

Table E5

Protein	Kappa Index	Standard error	95% Confidence interval
Cyclin A			
- Core 1 (n=157)	1	0.0798	0.84 – 1
- Core 2 (n=152)	1	0.0811	0.84 – 1
Ki-67			
- Core 1 (n=353)	0.9066	0.0532	0.80 – 1
- Core 2 (n=346)	0.9320	0.0538	0.83 – 1
- Core 3 (n=248)	0.9172	0.0635	0.79 – 1
E-Cadherin			
- Core 1 (n=151)	0.9867	0.0814	0.83 – 1
- Core 2 (n=148)	0.9865	0.0822	0.83 – 1
phospho-mTOR			
- Core 1 (n=342)	1	0.0541	0.89 – 1
- Core 2 (n=338)	1	0.0544	0.89 – 1
- Core 3(n=245)	1	0.0639	0.87 – 1

SPECIFIC ASSESSMENT METHODS BY PROTEIN

The following table describes the criteria followed for each marker, based on the percentage of cells expressed by the molecular markers (MM) and the visual intensity of staining (subjective criterion), as well as the intracellular location of the protein and the bibliographic references used to select the cut-off values for negative (0) or positive (1) consideration. NV: Not valid (Table E6).

Table E6

	Percentage of cells (%)	**Intensity**	**Categories**	**Location**
Cyclin A	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
Cyclin B1	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Cytoplasm
Cyclin D1	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
Cyclin E	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
CDK2	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
CDK6	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
P16	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
P21	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
P27	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
RB	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
Ki-67	0-20% (-) >20% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
P53	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
MDM2	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
BCL2	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Cytoplasm

Table E6 (Cont)

	Percentage of cells (%)	Intensity	Categories	Location
Caspase-3	0 = (-) 1 = < 25 % 2 = 25-50 % 3 = 50-75 % 4 = > 75 %	0 = (-) 1 = + 2 = ++ 3 = +++ 4 = ++++	6 = NV 0 = (-) ---- [0] 1 = (+) --- [1-4] 2 = (+) --- [5-8]	Cytoplasm
Survivin	0-10% (-) >10% (+)	0 = - / + 1 = ++ / +++	Cytoplasm (C) and nucleus (N) 6 = NV 0 = (-) [0] 1 = (+) [1-2]	Cytoplasm and nuclear
NFKB	0 = (-) 1 = focal, <1/3 2 = multifocal, <2/3 3 = diffuse, >2	0 = (-) 1 = + 2 = ++	Cytoplasm (C) and nucleus (N) 6 = NV 0 = (-) ---- [0] 1 = (-) --- [2] 2 = (+) --- [3-5]	Cytoplasm and nuclear
FAS	0 = (-) 1 = 1- 25 % 2 = 26-50 % 3 = > 50 %	0 = (-) 1 = + 2 = ++ 3 = +++	6 = NV 0 = (-) [0-3] 1 = (+) [4-6]	Membranous or membranous and cytoplasmatic; not only cytoplasmatic
		Nucleus: 0 = (-) 1 = + 2 = ++ 3 = +++	Nucleus (N):% x intensity 6 = NV 0 = (-) [- / <5] 1 = low - [5-100] 2 = medium[101-200] 3 = high [201-300]	
B-catenin	Specified	Membrane/ cytoplasm: 0 = (-) 1 = trazes 2 = + 3 = ++ 4 = +++	Membrane (Mb)/ cytoplasm (C): % x intensity 6 = NV 0 = (-) [- / <5] 1 = (-) traces [5-100] 2 = (-) low [101-200] 3 = (+) medium [201-300] 4 = (+) high [301-400]	Cytoplasm and nuclear

Table E6 (Cont)

	Percentage of cells (%)	Intensity	Categories	Location
E-cadherin	0-75% (-) >75% (+)	Negative (-) Weak (- / +) Moderate (+) Strong (++)	6 = NV 0 = (-) <75, (-) ,(-/+), (+) 1 = (+) > o = 75%, (++)	Membranous
phospho-AKT	NV	0 = (-) 1 = + 2 = ++	6 = NV 0 = (-) 1 = (+) 2 = (++)	Cytoplasm
phospho - mTOR	NV	0 = (-) 1 = + 2 = ++	6 = NV 0 = (-) 1 = (+) 2 = (++)	Cytoplasm
EGFR	0-10% (-) >10% (+)	NV	6 = NV 0 = (-) 1 = (+)	Membranous
HER2-Herceptest	0 = (-) / (+) <10 % 1= (+) weak / incomplete in >10% 2= (+)weak-moderate / complete in >10% 3= (+) strong / complete in >10%		6 = NV 0 = (-) 1 = (-) 2 = (+) 3 = (+)	Membranous
COX-2	0 = (-) 1 = 1- 25 % 2 = 26-50 % 3 = 51-75 % 4 = 76-100 %	0 = (-) 1 = + 2 = ++ 3 = +++	NV 0 = (-) [0-1] 1 = (+) [>2]	Cytoplasm
TTF-1	0-5% (-) >5% (+)	NV	6 = NV 0 = (-) 1 = (+)	Nuclear
P63	NV	0 = (-) 1 = + / ++ 2 = +++	6 = NV 0 = (-) 1 = + / ++ 2 = +++	Nuclear
LKB1	0-5% (-) >5% (+)	0 = (-) 1 = +/ ++ /+++ NV?	6 = NV 0 = (-) 1 = (+)	Cytoplasm

Table E6 (Cont)

	Percentage of cells (%)	Intensity	Categories	Location
phospho-ACC	NV	0 = (-) 1 = + 2 = ++ 3 = +++	(-)= 0 and 1 / (+)=2 and 3 or (-)= 0 / (+)=1, 2 and 3	Cytoplasm
phospho-S6	1=<50% 2=50-75% 3=76-85% 4=>86%	0 = (-) 1 = + 2 = ++ 3 = +++	6 = NV 0 = (-) <5 1 = (+) > = 5 (Sum of % + intensity)	Cytoplasm
CDC6	NV	0 = (-) 1 = + 2 = ++ 3 = +++	6 = NV 0 = (-) 0 and 1 1 = (+) 2 and 3	Cytoplasm and nuclear
γCP4	0 = (-) 1 = 1- 25 % 2 = 26-50 % 3 = 51-75 % 4 = 76-100 %	0 = (-) 1 = + 2 = ++ 3 = +++ 4 = ++++	6 = NV 0 = (-) 0 and 1 1 = (+) 2 -4	Cytoplasm

All tumour markers are assigned three single results: 6, 0, 1. All groups were performed with the criterion described below. With regard to the values assigned to the tumour markers for analysis purposes and in terms of immunohistochemical data interpretation, value assignation was as follows:

- **Value 6**: NV: Not valid (for all markers)

- **Value 0:** interpreted as negative (absence of marker expression), has been assigned as follows:

  a) For Cyclin A, B1, D1, E , CDK2, CDK6, P16, P21, P27, RB, Ki 67, P53, MDM2, BCL2, Caspase 3, Survivin, FAS, E-cadherin, phospho-AKT, phospho-mTOR, EGFR, COX2, TTF1, P63, LKB1, phospho-S6, CDC6, γCP4, it corresponds to category 0 of each marker's IHC.
  b) For NFKβ, it corresponds to IHC categories 0 and 1.
  c) For β-catenin, it corresponds to IHC categories 0, 1 and 2.
  d) For HER2, it corresponds to IHC categories 0 and 1.

- **Value 1:** interpreted as positive (marker expression) has been assigned as follows:

  a) For Cyclin A, B1, D1, E, CDK2, CDK6, P16, P21, P27, RB, Ki 67, P53, MDM2, BCL2, Survivin, FAS, E-cadherin, EGFR, COX2, TTF1, LKB1, phospho-S6, CDC6, γCP4, it corresponds to IHC category 1.
  b) For Caspase 3, it corresponds to IHC categories 1 and 2.
  c) For NFKβ, it corresponds to IHC category 2.
  d) For β-catenin, it corresponds to IHC categories 3 and 4.

e) For phospho-AKT, it corresponds to IHC categories 1 and 2.
f) For phospho-mTOR, it corresponds to IHC categories 1 and 2.
g) For HER2, it corresponds to IHC categories 2 and 3.
h) For P63, it corresponds to IHC categories 1 and 2.

An exception is made for the ACC marker which has been granted two interpretations:

phospho-ACC (1): Value 0: IHC categories 0 and 1.
Value 1: IHC categories 2 and 3.

phospho-ACC (2): Value 0: IHC category 0.
alue 1: IHC categories 1,2 and 3.

In cases where tissue cores from the same tumour presented variable results regarding protein expression, each case was examined individually, selecting as a rule the greatest value and, when in doubt, the most positive result.

Global reference:

Leversha MA, Fielding P, Watson S, Gosney JR, Field JK. Expression of p53, pRB, and p16 in lung tumours: a validation study on tissue microarrays. *J Pathol* 2003;200:610-619

Specific references:

Cyclin A	*Am J Pathol* 1998;153:963-972
Cyclin B1	*Cancer Res* 2000;60:4000-4004
Cyclin D1	*Lung Cancer* 2001;31:3-8
Cyclin E	*J Pathol* 2003;199:208-220
CDK2	*Am J Pathol* 1998;153:963-972
CDK6	*Am J Pathol* 1998;153:963-972
P16	*J Pathol* 2003;200:610-619
P21	*Jpn J Cancer Res* 2001;92:285-292
P27	*Clin Can Res* 2000;6:4073-4081
RB	*J Pathol* 2003;200:610-619
Ki-67	*Cancer Res* 2000;60:4000-4004
P53	*J Pathol* 2003;200:610-619
MDM2	*Jpn J Cancer Res* 2001;92:285-292
BCL2	*Hum Pathol* 2002;33:105-110
Caspase-3	*Int J Cancer* 2001;93:192-198
Survivin	*Oncology Reports* 2002;9:835-838
	Br J Cancer 2004;91:537-540
NFKB	*Carcinogenesis* 2003;24:665-671
FAS	*Int J Cancer* 1999;84:239-243
B-catenin	*Cancer* 2002;94:752-758
	J Clin Oncol 2002;20:2417-2428
E-cadherin	*Clin C Res* 2003;9:3606-3614
phospho-AKT	*Cancer Res* 2003;63:2742-2746
phospho -mTOR	*Cancer Res* 2003;63:2742-2746
EGFR	*Modern Pathology* 2004;17:430-439
COX-2	*Lung Cancer* 2003;42:163-170
TTF-1	*Hum Pathol* 2003;34:597-604
P63	*Cancer Res* 2003;63:7113-7121
CDC6	*Am J Pathol* 2004;165:1351-1365
γCP4	*Cancer Res* 2004;64:4171-4179