

ORIGINAL ARTICLE

# DNA replication stress response involving *PLK1*, *CDC6*, *POLQ*, *RAD51* and *CLASPIN* upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients

C Allera-Moreau<sup>1,2,7</sup>, I Rouquette<sup>2,7</sup>, B Lepage<sup>3</sup>, N Oumouhou<sup>3</sup>, M Walschaerts<sup>4</sup>, E Leconte<sup>5</sup>, V Schilling<sup>1</sup>, K Gordien<sup>2</sup>, L Brouchet<sup>2</sup>, MB Delisle<sup>1,2</sup>, J Mazieres<sup>1,2</sup>, JS Hoffmann<sup>1</sup>, P Pasero<sup>6</sup> and C Cazaux<sup>1</sup>

Lung cancer is the leading cause of cancer deaths worldwide. Clinical staging classification is generally insufficient to provide a reliable prognosis, particularly for early stages. In addition, prognostic factors are therefore needed to better forecast life expectancy and optimize adjuvant therapeutic strategy. Recent evidence indicates that alterations of the DNA replication program contribute to neoplasia from its early stages and that cancer cells are frequently exposed to endogenous replication stress. We therefore hypothesized that genes involved in the replication stress response may represent an under-explored source of biomarkers. Expressions of 77 DNA replication-associated genes implicated in different aspects of chromosomal DNA replication, including licensing, firing of origins, elongation, replication fork maintenance and recovery, lesion bypass and post-replicative repair were determined in primary tumors and adjacent normal tissues from 93 patients suffering from early- or mid-stage non-small cell lung cancer (NSCLC). We then investigated a statistically significant interaction between gene expressions and survival of early-stage NSCLC patients. The expression of five genes, that is, *POLQ*, *PLK1*, *RAD51*, *CLASPIN* and *CDC6* was associated with overall, disease-free and relapse-free survival. The expression levels are independent of treatment and stage classification. Except *RAD51*, their prognostic role on survival persists after adjustment on age, sex, treatment, stage classification and conventional proliferation markers, with a hazard ratio of 36.3 for *POLQ* (95%CI 2.6–517.4,  $P = 0.008$ ), 23.5 for *PLK1* (95%CI 1.9–288.4,  $P = 0.01$ ), 20.7 for *CLASPIN* (95%CI 1.5–275.9,  $P = 0.02$ ) and 18.5 for *CDC6* (95%CI 1.3–267.4,  $P = 0.03$ ). We also show that a five-gene signature including *POLQ*, *PLK1*, *RAD51*, *CLASPIN* and *CDC6* separates patients into low- and high-risk groups, with a hazard ratio of 14.3 (95% CI 5.1–40.3,  $P < 0.001$ ). This ‘replication stress’ metamarker may be a reliable predictor of survival for NSCLC, and may also help understand the molecular mechanisms underlying tumor progression.

*Oncogenesis* (2012) 1, e30; doi:10.1038/oncsis.2012.29; published online 22 October 2012

**Subject Categories:** DNA damage and repair

**Keywords:** DNA repair; DNA replication; lung cancer; prognosis marker; replicative stress

## INTRODUCTION

Lung cancer, predominantly non-small cell lung cancer (NSCLC), is the first cause of cancer deaths worldwide, resulting in about one million deaths each year. Its incidence increases continuously, especially for nonsmoking women. Despite advances in prevention, screening, resection methodology and chemotherapy strategies, five-year survival of operated patient remains low, ranging from 59 (stage IA) to 14% (stage IIIA). According to clinicians and pathologists, the current clinical staging on the basis of ‘Tumor Node Metastasis’ system is insufficient to predict the outcome of patients, especially for early-stage diseases. It would thus be of great clinical interest to identify biomarkers that may help distinguish among operated patients those who can be considered cured by surgery from those who display a high risk of recurrence and need adjuvant treatment. Regarding the latter, identification of predictive

factors will either help select the appropriate drugs or prevent inappropriate use of chemotherapy. As a consequence of the absence of such predictive markers, the therapeutic strategies can vary from a country to another.

Studies profiling gene expression in lung cancer have been completed or are in progress to identify predictors of patient prognosis and response to therapy.<sup>1–4</sup> These unbiased microarray-based profilings of thousands of genes generally lead to multi-gene signatures including endpoint-selected cell proliferation genes, that is, genes that drive cell cycle progression or tumor differentiation at the latest stages of tumorigenesis. So far these signatures do not bring consistent prognostic molecular markers complementary to standard clinical-pathological markers such as Ki67, PCNA, tumor size, positive-involved nodes or distant metastasis, which also refer to cell proliferation status. Even if

<sup>1</sup>Cancer Research Center of Toulouse, Inserm U1037, CNRS ERL5294, University of Toulouse, University Paul Sabatier, Toulouse, France; <sup>2</sup>Rangueil-Larrey University Hospital, University of Toulouse, University Paul Sabatier, Toulouse, France; <sup>3</sup>Service of Epidemiology, INSERM U1027, University Paul Sabatier, Toulouse University Hospital, University of Toulouse, Toulouse, France; <sup>4</sup>Human Fertility Research Group, EA3694, University Paul Sabatier, University of Toulouse, Toulouse, France; <sup>5</sup>Toulouse School of Economics, GREMAC, University Capitole, University of Toulouse, Toulouse, France and <sup>6</sup>Institute of Human Genetics CNRS UPR1142, Montpellier, France. Correspondence: Professor C Cazaux, Cancer Research Center of Toulouse, Inserm U1037, CNRS ERL5294, University of Toulouse, University Paul Sabatier, 205 route de Narbonne (IPBS), Toulouse 31077, France.

E-mail: cazaux@ipbs.fr

<sup>7</sup>These authors contributed equally to this work.

Received 26 June 2012; revised 28 August 2012; accepted 4 September 2012

promising, the added value of gene profiling over routinely-used approaches thus remains of limited value and no reported gene expression signatures are still ready for clinical application.<sup>1</sup>

DNA replication represents an under-explored source of prognostic markers that could be used in combination with histo-pathological markers and/or multi-genetic signatures to predict prognosis. Faithful execution of the DNA replication program is critical to promote cell division and to limit cancer risk through the preservation of genome integrity.<sup>5</sup> Recent evidence indicates that oncogene-induced alterations of the DNA replication program triggers replicative stress and replication-associated DNA damage, favoring the accumulation of genetic alterations in cancer cells.<sup>6</sup> It is therefore likely that key replication proteins and associated repair/recombination factors limit cancer development by preventing genomic instability during DNA replication.

The cardinal role of DNA replication defects in tumorigenesis is illustrated by hereditary forms of colon, breast, ovary and skin cancers caused by mutations in DNA replication genes such as translesion DNA synthesis polymerases (for example, *POLH*), intra-S-phase replication checkpoints (for example, *ATR*, *BRCA1*) or post-replicative DNA repair and recombination genes (for example, *MLH1*, *BRCA2*, *BLM*).<sup>7–12</sup> In sporadic cancers, early mutations in replication-associated genes probably become diluted in the genetic chaos resulting from the progression of the disease, making the relationship between cancer and replication defects less obvious. Nevertheless, a large body of evidence indicates that alterations in genes involved in genome replication or surveillance favor sporadic cancers.<sup>13,14</sup> In addition, perturbation of the catalytic activity of the replicative DNA polymerase delta increases genomic instability and accelerates tumorigenesis in mice.<sup>15</sup> Mouse fibroblasts expressing altered levels of the origin licensing factors *Mcm4*, *Mcm7* and *Cdt1* show numerous chromosomal abnormalities and form tumors in nude mice.<sup>16–18</sup> Finally, non-invasive assays have been already developed to identify colorectal cancer by detection of *Mcm* proteins in colonocytes retrieved from the fecal surface,<sup>19</sup> which is currently tested in clinical trials.

In this study, we assessed specifically the expression of 77 replication-associated genes in primary tumors and adjacent normal tissues from a series of 93 operated NSCLC patients. We found that many of these genes were significantly deregulated in tumors. More importantly, we show that misregulation of a subset of these genes is determinant of overall survival (OS) after surgical treatment, independently of the therapeutic strategies or tumor stages.

## RESULTS

Most of the replication-associated genes are deregulated in lung tumors

Gene expression profiles of 93 primary lung adenocarcinomas coupled with their normal counterparts at early or intermediate stages of the disease (Supplementary Table S1) were generated from a selection of 77 genes involved in different aspects of chromosome replication. These include genes involved in licensing and activation of replication origins, replication fork progression and restart, damage bypass by translesion DNA polymerases (TLS), post-replicative repair of DNA lesions and S-phase checkpoints. We then identified genes that were either up- or downregulated in tumors (T) compared with adjacent control tissues (N). Deregulated genes were stratified in three groups, on the basis of the number of tumors in which T/N expression ratio was either above or below 2. As indicated in Table 1, we evidenced genes that were downregulated ( $T/N < 1$ ), slightly upregulated ( $1 < T/N < 2$ ) or highly overexpressed ( $T/N > 2$ ). We did not find genes downregulated more than twofold, that is,  $T/N < 1/2$  (Table 1). Table 2 shows genes that were significantly

upregulated above a certain threshold (Th). Individual levels of expression in all the 93 coupled tumors are also shown for selected replication-associated genes (Supplementary Figure S1).

As indicated in Table 1, we confirmed that the DNA repair gene *ERCC1* and the tumor suppressor genes *APC* and *TP53* were downregulated in NSCLC tumors, which is consistent with the published data.<sup>20–22</sup> Conversely, we found that the *Ki67* gene, whose product is used as a proliferative marker by pathologists, was largely overexpressed in NSCLC tumors ( $T/N > 4$ ).

Genes involved in the licensing or the activation of replication origins were significantly overexpressed in tumors compared with controls (Table 1). Indeed, with the exception of *ORC4*, all the genes involved in initiation were overexpressed, including *SLD5*, *CDT1*, *CYCLIN A*, *CYCLIN E*, *CDC45*, *CDC6*, *PLK1*, and to a lesser extent *DBF4* and *MCM7*. Genes encoding the nonhistone proteins *HMG A1* and *HMG A2* were also overexpressed. These proteins interact with the origin recognition complex,<sup>23</sup> and have been implicated in replication origin function (JM Lemaitre, personal communication). Interestingly, we observed a concomitant repression of *CUL4*, an E3 ligase involved in the degradation of the replication licensing factor CDT1. Together, these data indicate that genes promoting origin firing are globally overexpressed in lung cancer cells.

Next, we investigated the expression of genes encoding different types of DNA polymerases, including replicative DNA polymerases (*POLA*, *POLD* and *POLE*), mitochondrial DNA polymerase (*POLG*) and specialized DNA polymerases involved in DNA lesions bypass or in DNA repair synthesis. Expression of the replicative DNA polymerases *POLD* and *POLE* and their processivity factor PCNA was slightly increased (Table 1), which is reminiscent of the overexpression (Table 1) of *MCM7*, *MCM8*, *SLD5* and *CDC45*, encoding components of the replicative helicase.<sup>24</sup> In contrast, the mitochondrial DNA polymerases *POLG* and all the so-called TLS were globally downregulated in NSCLC tumors (*POLZ*, *POLK*, *POLI*, *REV1*, *POLH*, *POLL* and *REV1*), as evidenced also for colorectal and breast cancers.<sup>25,26</sup> Similarly, the expression of factors involved in the ubiquitylation of PCNA and the recruitment of the Y-family TLS (*SHPRH*, *RAD6* and *RAD18*) was either downregulated or unchanged (Table 1). Exceptions to this rule include *POLM* and *POLB*, which are involved in DNA repair synthesis after base excision and nonhomologous end joining, respectively<sup>27,28</sup> and *POLQ*, which is the only DNA polymerase significantly overexpressed in NSCLC tumors. Together, these data indicate that unlike replicative DNA polymerases, specialized DNA polymerases are globally downregulated in NSCLC tumors, with the notable exception of *POLQ*.

We then investigated the expression of genes involved in intra S-phase checkpoints, fork restart and post-replicative DNA repair. Overall, the expression of sensors of the DNA damage response (DDR) was either downregulated (*ATM*, *RAD17*) or slightly increased (*53BP1*, *ATR*, *RAD9*). In contrast, mediators (*TIMELESS*, *CLASPIN*) and effectors (*CHK1*) of the DDR pathway were overexpressed. As these factors act primarily at stalled replication forks, we checked whether other factors involved in replication fork repair and restart were also overexpressed. Remarkably, we found that expression of fork-associated factors such as *SMARCA1*, *FANCM*, *SLX4*, *RUVBL1*, *BRCA2*, *BLM*, *RECQ4* and *RAD51* was globally upregulated to a variable extent. In contrast, the expression of factors involved in general DNA repair such as *BACH1/FANCI*, *RECQ1*, *SMC5*, *TIP60*, *XLF*, *LIG4*, *XRCC1*, *LIG3*, *XRCC4*, *DNAPKcs*, *LAMIN B*, *SIRT6*, *p300* and *SIRT1* was either reduced or unchanged (Table 1). These data suggest that unlike general DNA repair factors, genes involved in the surveillance or the repair of stalled replication forks are upregulated in NSCLC cells.

Exact binomial tests indicated that the expression of the 31 downregulated genes was reduced by less than twofold in tumors compared with the normal tissues (Tables 1 and 2), whereas expression of 17 genes was increased by more than twofold

**Table 1.** Differential expression of DNA replication genes in coupled NSCLC tumors

DNA replication gene	DNA transactions	Coupled tumors	Slightly overexpressed or downregulated expression		Overexpression	Uncorrected P-value <sup>a</sup>
			0.5 < T/N < 1	1 < T/N < 2	T/N > 2	
		n	n (%)	n (%)	n (%)	
APC	Wnt signaling (control)	93	93 (100.0)	0 (0.0)	0 (0.0)	< 0.000001 <sup>b</sup>
KI67	Cell proliferation (control)	93	2 (2.2)	9 (9.7)	82 (88.2)	< 0.000001 <sup>b</sup>
REV3 (POLZ)	TLS	93	89 (95.7)	4 (4.3)	0 (0.0)	< 0.000001 <sup>b</sup>
POLK	TLS	93	82 (88.2)	11 (11.8)	0 (0.0)	< 0.000001 <sup>b</sup>
POLI	TLS	93	78 (83.9)	15 (16.1)	0 (0.0)	< 0.000001 <sup>b</sup>
POLG	TLS	93	74 (79.6)	18 (19.4)	1 (1.1)	< 0.000001 <sup>b</sup>
SHPRH	TLS	93	73 (78.5)	19 (20.4)	1 (1.1)	< 0.000001 <sup>b</sup>
REV1	TLS	93	78 (83.9)	13 (14.0)	2 (2.2)	< 0.000001 <sup>b</sup>
POLH	TLS	93	58 (62.4)	32 (34.4)	3 (3.2)	< 0.000001 <sup>b</sup>
POLL	TLS	93	59 (63.4)	31 (33.3)	3 (3.2)	< 0.000001 <sup>b</sup>
POLM	TLS/repair	93	29 (31.2)	51 (54.8)	13 (14.0)	< 0.000001 <sup>b</sup>
POLB	TLS/repair	93	27 (29.0)	45 (48.4)	21 (22.6)	< 0.000001 <sup>b</sup>
RAD6	TLS/PRR	93	42 (45.2)	48 (51.6)	3 (3.2)	< 0.000001 <sup>b</sup>
RAD18	TLS/PRR	93	36 (38.7)	52 (55.9)	5 (5.4)	< 0.000001 <sup>b</sup>
POLQ	TLS/DNA replication	93	3 (3.2)	15 (16.1)	75 (80.6)	< 0.000001 <sup>b</sup>
CUL4A	Initiation/licensing	93	52 (55.9)	38 (40.9)	3 (3.2)	< 0.000001 <sup>b</sup>
ORC4	Initiation/licensing	93	55 (59.1)	37 (39.8)	1 (1.1)	< 0.000001 <sup>b</sup>
MCM7	Initiation/licensing	93	21 (22.6)	43 (46.2)	29 (31.2)	0.0004 <sup>b</sup>
DBF4	Initiation/licensing	93	28 (30.1)	41 (44.1)	24 (25.8)	0.000003 <sup>b</sup>
CDC7	Initiation/licensing	93	24 (25.8)	35 (37.6)	34 (36.6)	0.01
GEMININ	Initiation/licensing	93	3 (3.2)	39 (41.9)	51 (54.8)	0.41
SLD5	Initiation/licensing	93	4 (4.3)	23 (24.7)	66 (71.0)	0.00006 <sup>b</sup>
CDC45	Initiation/licensing	93	0 (0.0)	9 (9.7)	84 (90.3)	< 0.000001 <sup>b</sup>
CDT1	Initiation/licensing	93	2 (2.2)	7 (7.5)	84 (90.3)	< 0.000001 <sup>b</sup>
CDC6	Initiation/licensing	93	1 (1.1%)	4 (4.3)	88 (94.6)	< 0.000001 <sup>b</sup>
HMGA1	Initiation/licensing	93	4 (4.3)	15 (16.1)	74 (79.6)	< 0.000001 <sup>b</sup>
HMGA2	Initiation/licensing	93	23 (24.7)	4 (4.3)	66 (71.0)	0.00006 <sup>b</sup>
PLK1	Initiation/licensing/cell cycle	93	3 (3.2)	4 (4.3)	86 (92.5)	< 0.000001 <sup>b</sup>
CYCLIN E	Initiation/licensing/cell cycle	93	4 (4.3)	19 (20.4)	70 (75.3)	0.000001 <sup>b</sup>
CYCLIN A	Initiation/licensing/cell cycle	93	3 (3.2)	11 (11.8)	79 (84.9)	< 0.000001 <sup>b</sup>
CDC25A	Initiation/licensing/cell cycle	93	8 (8.6)	12 (12.9)	73 (78.5)	< 0.000001 <sup>b</sup>
CDC25B	Cell cycle	93	83 (89.2)	8 (8.6)	2 (2.1)	< 0.000001 <sup>b</sup>
POLE	DNA replication	93	34 (36.6)	39 (41.9)	20 (21.5)	< 0.000001 <sup>b</sup>
POLD	DNA replication	93	35 (37.6)	35 (37.6)	23 (24.7)	0.000001 <sup>b</sup>
MCM8	DNA replication	93	21 (22.6)	39 (41.9)	33 (35.5)	0.007 <sup>b</sup>
POLA	DNA replication	93	50 (53.8)	35 (37.6)	8 (8.6)	< 0.000001 <sup>b</sup>
BACH1	DNA replication/repair	93	85 (91.4)	8 (8.6)	0 (0.0)	< 0.000001 <sup>b</sup>
RECQ1	DNA replication/repair	93	75 (80.6)	15 (16.1)	3 (3.2)	< 0.000001 <sup>b</sup>
SMC5	DNA replication/repair	93	73 (78.5)	19 (20.4)	1 (1.1)	< 0.000001 <sup>b</sup>
MCM9	DNA replication/repair	93	68 (73.1)	24 (25.8)	1 (1.1)	< 0.000001 <sup>b</sup>
CHTF18	DNA replication/repair	93	16 (17.2)	43 (46.2)	34 (36.6)	0.01
DSCC1	DNA replication/repair	93	23 (24.7)	28 (30.1)	42 (45.2)	0.41
PCNA	DNA replication/repair	93	13 (14.0)	57 (61.3)	23 (24.7)	0.000001 <sup>b</sup>
SMARCAL1	DNA replication/repair	93	44 (47.3)	48 (51.6)	1 (1.1)	< 0.000001 <sup>b</sup>
FANCM	DNA replication/repair	93	39 (41.9)	45 (48.4)	9 (9.7)	< 0.000001 <sup>b</sup>
SLX4	DNA replication/repair	93	27 (29.0)	54 (58.1)	12 (12.9)	< 0.000001 <sup>b</sup>
BRCA2	DNA replication/repair	93	18 (19.4)	44 (47.3)	31 (33.3)	0.002 <sup>b</sup>
BLM	DNA replication/repair	93	8 (8.6)	21 (22.6)	64 (68.8)	0.0004 <sup>b</sup>
RECQ4	DNA replication/repair	93	2 (2.2)	16 (17.2)	75 (80.6)	< 0.000001 <sup>b</sup>
ERCC1	DNA repair (control)	93	68 (73.1)	25 (26.9)	0 (0.0)	< 0.000001 <sup>b</sup>
ASF1	DNA repair/chromatin assembly	93	72 (77.4)	21 (22.6)	0 (0.0)	< 0.000001 <sup>b</sup>
P300	DNA repair/chromatin assembly	93	71 (76.3)	20 (21.5)	2 (2.2)	< 0.000001 <sup>b</sup>
RUVBL1	DNA repair/chromatin assembly	93	15 (16.1)	55 (59.1)	23 (24.7)	0.000001 <sup>b</sup>
TIP60	DSB repair	93	90 (96.8)	3 (3.2)	0 (0.0)	< 0.000001 <sup>b</sup>
XLF	DSB repair	93	75 (80.6)	18 (19.4)	0 (0.0)	< 0.000001 <sup>b</sup>
LIG4	DSB repair	93	60 (64.5)	30 (32.3)	3 (3.2)	< 0.000001 <sup>b</sup>
XRCC1	DSB repair	93	51 (54.8)	38 (40.9)	4 (4.3)	< 0.000001 <sup>b</sup>
LIG3	DSB repair	93	28 (30.1)	55 (59.1)	10 (10.8)	< 0.000001 <sup>b</sup>
XRCC4	DSB repair	17	6 (35.3)	9 (52.9)	2 (11.8)	0.002 <sup>b</sup>
LAMIN B	DSB repair	93	13 (14.0)	53 (57.0)	27 (29.0)	0.00006 <sup>b</sup>
DNAPKcs	DSB repair	93	16 (17.2)	46 (49.5)	31 (33.3)	0.002 <sup>b</sup>
RAD51	DSB repair	93	2 (2.2)	21 (22.6)	70 (75.3)	0.000001 <sup>b</sup>
MORF4	DNA repair/DDR	38	17 (44.7)	4 (10.5)	17 (44.7)	0.63
SIRT1	DNA repair/DDR	93	78 (83.9)	14 (15.1)	1 (1.1)	< 0.000001 <sup>b</sup>

**Table 1.** (Continued)

DNA replication gene	DNA transactions	Coupled tumors	Slightly overexpressed or downregulated expression		Overexpression	Uncorrected P-value <sup>a</sup>
			0.5 < T/N < 1	1 < T/N < 2		
			n	n (%)	n (%)	
<i>SIRT6</i>	DNA repair/DDR	93	54 (58.1)	39 (41.9)	0 (0.0)	<0.000001 <sup>b</sup>
<i>P53</i>	DDR (control)	93	74 (79.6)	15 (16.1)	4 (4.3)	<0.000001 <sup>b</sup>
<i>MRGX</i>	DDR	93	89 (95.7)	4 (4.3)	0 (0.0)	<0.000001 <sup>b</sup>
<i>ATM</i>	DDR	93	63 (67.7)	29 (31.2)	1 (1.1)	<0.000001 <sup>b</sup>
<i>RPA</i>	DDR	93	56 (60.2)	35 (37.6)	2 (2.2)	<0.000001 <sup>b</sup>
<i>RAD17</i>	DDR	93	47 (50.5)	44 (47.3)	2 (2.2)	<0.000001 <sup>b</sup>
<i>RAD9</i>	DDR	93	32 (34.4)	52 (55.9)	9 (9.7)	<0.000001 <sup>b</sup>
<i>53BP1</i>	DDR	93	39 (41.9)	41 (44.1)	13 (14.0)	<0.000001 <sup>b</sup>
<i>ATR</i>	DDR	93	23 (24.7)	55 (59.1)	15 (16.1)	<0.000001 <sup>b</sup>
<i>TOPBP1</i>	DDR	93	30 (32.3)	45 (48.4)	18 (19.4)	<0.000001 <sup>b</sup>
<i>MRG15</i>	DDR	93	26 (28.0)	49 (52.7)	18 (19.4)	<0.000001 <sup>b</sup>
<i>BRCA1</i>	DDR	93	22 (23.7)	33 (35.5)	38 (40.9)	0.10
<i>TIMELESS</i>	DDR	93	2 (2.2)	36 (38.7)	55 (59.1)	0.10
<i>FANCD2</i>	DDR	93	7 (7.5)	44 (47.3)	42 (45.2)	0.41
<i>CHK1</i>	DDR	93	2 (2.2)	14 (15.1)	77 (82.8)	<0.000001 <sup>b</sup>
<i>CLASPIN</i>	DDR	93	1 (1.1)	13 (14.0)	79 (84.9)	<0.000001 <sup>b</sup>

Abbreviations: DDR, DNA damage response; DSB, double-stranded DNA break; PRR, post-replicative repair; TLS, translesional DNA replication. <sup>a</sup>Bilateral binomial tests. <sup>b</sup>Significantly >50% of the population have T/N > 2 (red) or T/N < 2 (green and pink), according to the corrected overall critical P-value by the Benjamini and Yekutieli method (<0.0087).

(Table 2). The latter include *CYCLIN A*, *RECQ4*, *POLQ*, *CLASPIN* and *CHK1* (T/N > 3), *CDC45*, *CDC6*, *KI67*, *HMGA2* and *CDT1* (T/N > 4) and *PLK1* (T/N > 5). To determine whether the overexpression of these genes reflects the proliferation status of cancer tissues, we used a Pearson's test to compare their expression levels with that of the proliferation marker *KI67*. We found that it is indeed the case for the majority of them, but not for *HMGA2* ( $\rho = 0.1$ ), *SLD5*, *CYCLIN E*, *RECQ4* and *HMGA1* ( $\rho = 0.6$ ; data not shown).

A replication gene signature is associated with poor prognosis

The ultimate aim of this study was to identify DNA replication genes whose expression in tumors can be used to predict the OS of NSCLC patients. A log-rank test for equality of survival functions identified a group of 10 genes associated with a higher overall morbidity (Figure 1 and Supplementary Table S2). This group includes *POLQ* ( $P = 0.0008$ ), *PLK1* ( $P = 0.0062$ ), *RAD51* ( $P = 0.007$ ), *CYCLIN A* ( $P = 0.0128$ ), *CDC25A* ( $P = 0.0196$ ), *CLASPIN* ( $P = 0.0233$ ), *CDC6* ( $P = 0.0404$ ), *POLL* ( $P = 0.0464$ ) and *RPA* ( $P = 0.0458$ ). It also includes the DNA repair gene *ERCC1* ( $P = 0.0256$ ), which is consistent with published data.<sup>22</sup> Disease-free survival was associated with the expression of seven of these genes, that is, *POLQ*, *PLK1*, *RAD51*, *CYCLIN A*, *CDC25A*, *CLASPIN* and *CDC6* (Supplementary Table S2). Misregulation of five of these genes correlated with relapse-free survival, with the exception of *CYCLIN A* and *CDC25A* (Supplementary Table S2). Altogether, these data indicate that the expression of five genes, that is, *POLQ*, *PLK1*, *RAD51*, *CLASPIN* and *CDC6* was associated with overall, disease-free and relapse-free survival (Table 3).

Next, we asked whether the expression of these five genes is co-regulated in NSCLC cells. Using a Pearson's test, we show that it is indeed the case (Pearson's coefficient > 0.7; Supplementary Figure S2A). This was confirmed with an independent hierarchical ascending classification, following Ward's criteria of the most deregulated genes (T/N > 2). These data indicate that *POLQ*, *CDC6* and *PLK1*, and to a lesser extent *CLASPIN* and *RAD51*, behave as a single 'metamarker' (Supplementary Figure S2B).

We used  $\chi^2$ -test and Fisher's exact tests to examine the dependence between gene expression and treatment (chemotherapy, chemotherapy plus radiotherapy or no adjuvant treatment) or node stages (N). This bivaried analysis revealed that the expression of *POLQ*, *PLK1*, *RAD51*, *CLASPIN* or *CDC6* did not correlate ( $P > 0.05$ , Supplementary Table S3) with the therapeutic strategy or with the number of metastasis-containing nodes. In contrast, the expression levels of *CDC25B* ( $P = 0.0307$ ), *GEMININ* ( $P = 0.0263$ ) and *APC* ( $P = 0.0404$ ) correlated with anti-cancer treatment and the expression levels of *CDC25B* ( $P = 0.0344$ ), *SLX4* ( $P = 0.0466$ ), *GEMININ* ( $P = 0.0192$ ) and *MCM7* ( $P = 0.0364$ ) correlated with the clinical classification (Supplementary Table S4).

Finally, we used a Cox multivariate regression model to examine the effect of age, sex, treatment, tumor stage and expression of *KI67* and *PCNA* on the association between OS and expression of these five DNA replication genes. This analysis revealed that a strong association was still observed after adjustment to covariates for four of these genes (*POLQ*, *PLK1*, *CLASPIN* and *CDC6*), except *RAD51* (Table 4), presumably because *RAD51* expression correlates with that of *PCNA* and *KI67* (Pearson's coefficients  $\rho = 0.54$  and  $\rho = 0.81$ , respectively). Remarkably, patients with tumors strongly overexpressing either *POLQ*, *PLK1*, *CLASPIN* and *CDC6* show a much poorer prognosis than patients with normal expression levels, with a hazard ratio of 36.3 for *POLQ* (95%CI 2.6–517.4,  $P = 0.008$ ), 23.5 for *PLK1* (95%CI 1.9–288.4,  $P = 0.01$ ), 20.7 for *CLASPIN* (95%CI 1.5–275.9,  $P = 0.02$ ) and 18.5 for *CDC6* (95%CI 1.3–267.4,  $P = 0.03$ ). A five-gene signature of the OS was also determined using a single multivariate Cox regression model, including the expression levels of the five genes. For each patient, the predictive value of the Cox model was used as a risk score. Patients with a high-risk score exhibit the poorer survival prognosis, with a hazard ratio of 14.3 (95% CI 5.08–40.3,  $P < 0.001$ ) compared with the low-risk group (Figure 2).

Validation of the prognosis significance

The prognostic significance of our markers was tested in a subset of the reference Director's Challenge Consortium training-testing,



**Table 2.** Exact binomial tests after setting four different T/N thresholds

Replication genes	DNA transaction	n	Under expression (T/N < 1/Th) <sup>a</sup>	Neither over- nor under-expression <sup>a</sup>	Over expression (T/N > Th) <sup>a</sup>	Uncorrected P-value <sup>b</sup>	T/N threshold (Th)
<i>PLK1</i>	Initiation/licensing	93	0	25	68	4.69E-06	5
<i>CDC45</i>	Initiation/licensing	93	0	23	70	5.51E-07	4
<i>CDC6</i>	Initiation/licensing	93	0	24	69	1.65E-06	4
<i>CDT1</i>	Initiation/licensing	93	0	27	66	0.000032	4
<i>HMGA2</i>	DNA replication/repair	93	10	26	57	0.000013	4
<i>Ki67</i>	Cell proliferation	93	0	24	69	1.65E-06	4
<i>CYCLIN A</i>	Initiation/licensing	93	0	21	72	5.19E-08	3
<i>CLASPIN</i>	DDR	93	0	28	65	0.000079	3
<i>RECO4</i>	DNA replication/repair	93	0	26	67	0.000013	3
<i>POLQ</i>	TLS	93	0	27	66	0.000032	3
<i>CHK1</i>	DDR	93	0	29	64	0.000183	3
<i>CYCLIN E</i>	Initiation/licensing	93	0	23	70	5.51E-07	2
<i>SLD5</i>	Initiation/licensing	93	0	27	66	0.000032	2
<i>CDC25A</i>	DDR	93	1	19	73	3.86E-09	2
<i>BLM</i>	DNA replication/repair	93	0	29	64	0.000183	2
<i>RAD51</i>	DSB repair	93	0	23	70	5.51E-07	2
<i>HMGA1</i>	DNA replication/repair	93	0	19	74	3.86E-09	2

Abbreviations: DDR, DNA damage response; DSB, double-stranded DNA break; TLS, translesional DNA replication. <sup>a</sup>T/N > Th means a more than Th-fold overexpression. <sup>b</sup>Significant over-expression compared with the corrected overall P-value by Benjamini and Yekutieli method ( $P < 0.00021$  for Th = 5;  $P < 0.00074$  for Th = 4;  $P < 0.0013$  for Th = 3; and  $P < 0.0019$  for Th = 2).

multisite, blinded validation study, which provides the largest available set of microarray data with extensive pathological and clinical annotation.<sup>29</sup> To validate our data, we only considered Director's Challenge Consortium patients with completely resected stage IA to IIIA NSCLC who did not receive neo-adjuvant therapy (total  $n = 400$ ). We found that *POLQ* ( $P = 0.002$ ), *PLK1* ( $P = 0.003$ ) and *CDC6* ( $P = 0.0001$ ) expression successfully classified patients into low- and high-risk groups (Supplementary Table S5, Supplementary Figure S3).

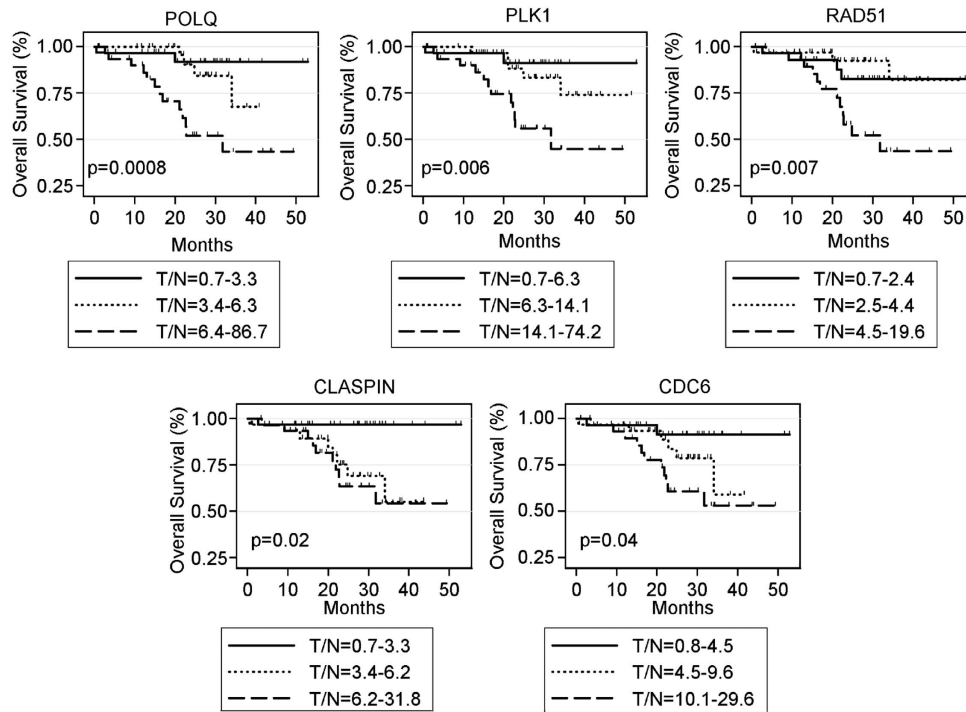
## DISCUSSION

DNA replication is at the heart of the cancer process. It is perturbed by the initial events driving aberrant cancer cell proliferation, namely the deregulated expression of oncogenes.<sup>6,30–33</sup> Oncogene-induced replication stress promotes the constitutive activation of the DNA damage response in pretumoral cells, leading to the bypass of anticancer barriers such as apoptosis or senescence.<sup>34</sup> Replication problems also increase genetic instability in the course of tumor progression through reduction of polymerase fidelity, replication fork collapse or alteration of chromatin assembly on newly replicated DNA. Replication stress also promotes cancer development by inducing breaks at common fragile sites, which are specific regions of the genome that show increased fragility when DNA replication is perturbed. Common fragile site instability also increases when proteins involved in replication fork maintenance are mutated,<sup>35</sup> or when the balance between replicative and TLS is altered.<sup>36</sup>

Although chronic replication stress promotes cancer development at its earliest stages, it is also a burden for tumor cells as it impedes cell proliferation. Recent evidence indicates that cancer cells often rely on an efficient replication stress response for viability, especially when they are deficient for p53,<sup>37</sup> which can be exploited for selective treatment.<sup>38</sup> Moreover, it has been recently reported that increased levels of the Chk1 kinase protects mouse cells against replication stress and favors transformation.<sup>39</sup> Together, these data suggest that genes involved in DNA

replication and in the replication stress response could be used as prognosis factors for aggressive cancers such as NSCLC. We have, therefore, monitored the expression of a subset of genes involved in different aspects of DNA replication, including initiation, elongation, replication fork maintenance and recovery, lesion bypass and post-replicative repair. We show that most of these genes are significantly deregulated in NSCLC cells relative to adjacent normal tissue. This observation supports the view that cancer cells adapt to replication stress by modifying key aspects of their replication program and by increasing their ability to restart arrested replication forks.

We also show that overexpression of *POLQ*, *PLK1*, *RAD51*, *CLASPIN* and *CDC6* genes, either considered individually or as a signature, is associated with a bad prognosis in NSCLC. We propose that the upregulation of these genes helps cancer cells tolerate spontaneous replication stress. The mechanisms by which overexpression of these five genes protects cells from replication stress is currently unclear, but recent studies suggest that they could modulate key aspects of the replication stress response, including fork recovery and licensing of dormant replication origins (Figure 3). Dormant origins are backup origins that are licensed in G<sub>1</sub>, but are only used when cells are exposed to replication stress. Recent studies indicate that the viability of cancer cell lines depends on an excess of MCM proteins in the presence of the replication inhibitor hydroxyurea.<sup>40–42</sup> As the recruitment of MCMs to replication origins depends on Cdc6,<sup>6</sup> an attractive possibility could be that overexpression of the *CDC6* gene increases the licensing of dormant origins in aggressive NSLCC cells. Claspin and Rad51 are two other key players of the replication stress response, which promote the maintenance and recovery of arrested forks, respectively.<sup>43–45</sup> Plk1 is a key regulator of the G2/M checkpoint, which prevents cells from entering mitosis with an incompletely replicated genome.<sup>46</sup> Besides its checkpoint function, recent evidence indicates that Plk1 contributes to the licensing of dormant origins through the phosphorylation of the Orc2 protein,<sup>47</sup> and promotes the recruitment of the Rad51 recombinase on DNA breaks.<sup>48</sup> Finally, overexpression of the



**Figure 1.** Relationship between OS and expression levels of *POLQ*, *PLK1*, *RAD51*, *CLASPIN* or *CDC6*. Probabilities for OS were estimated using Kaplan–Meier method. OS corresponds to the interval between the date of lung surgery and the date of death, related or not to the NSCLC. Expression levels were classified into three categories according to the T/N distribution, and the log-rank test was used to compare survival curves. Differences were considered significant at 5% level.

**Table 3.** Survival of patients according to the level of DNA replication gene expression

Replication gene	DNA transaction	T/N (terciles)	n	Overall survival			Disease-free survival			Relapse-free survival		
				Events	%	P-value (Log Rank)	Events	%	P-value (Log Rank)	Events	%	P-value (Log Rank)
<i>POLQ</i>	TLS	(0.70, 3.34)	31	2	6.45	0.0008	8	25.81	0.003	7	22.58	0.002
		(3.36, 6.31)	31	4	12.9		17	54.84		16	51.61	
		(6.42, 86.70)	31	13	41.94		20	64.52		14	45.16	
<i>PLK1</i>	Initiation Licensing Cell cycle	(0.68, 6.28)	31	2	6.45	0.006	6	19.35	0.0001	5	16.13	0.001
		(6.31, 14.12)	31	5	16.13		18	58.06		16	51.61	
		(14.13, 74.19)	31	12	38.71		21	67.74		16	51.61	
<i>CLASPIN</i>	DDR	(0.74, 3.27)	31	1	3.23	0.02	7	22.58	0.0005	6	19.35	0.002
		(3.39, 6.15)	31	8	25.81		17	54.84		14	45.16	
		(6.25, 31.81)	31	10	32.26		21	67.74		17	54.84	
<i>RAD51</i>	DSB repair	(0.70, 2.40)	31	4	12.9	0.007	9	29.03	0.0001	7	22.58	0.0009
		(2.48, 4.40)	31	3	9.68		13	41.94		12	38.71	
		(4.46, 19.65)	31	12	38.71		23	74.19		18	58.06	
<i>CDC6</i>	Initiation Licensing	(0.75, 4.47)	31	2	6.45	0.04	10	32.26	0.01	9	29.03	0.03
		(4.55, 9.62)	31	6	19.35		14	45.16		11	35.48	
		(10.08, 29.65)	31	11	35.48		21	67.74		17	54.84	

Abbreviations: DDR, DNA damage response; DSB, double-stranded DNA break; TLS, translesional DNA replication.

*POLQ* gene could allow the error-prone bypass of DNA lesions by facilitating extension from mismatches.<sup>49,50</sup> Interestingly, PolQ and its helicase domain could also promote the activation of dormant origins through its interaction with components of the

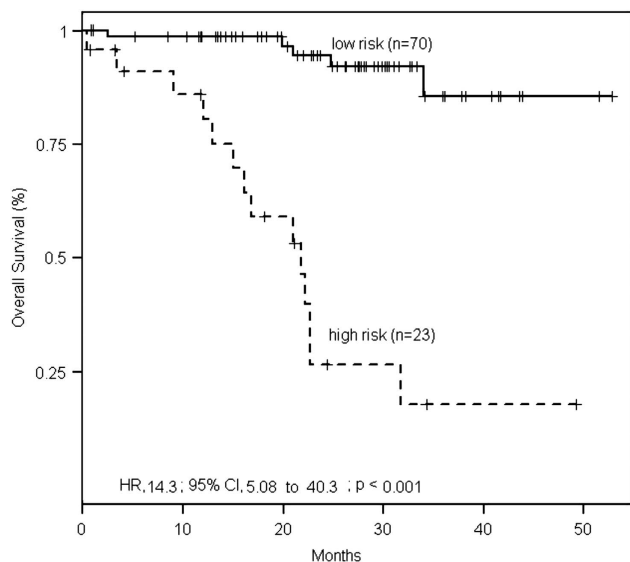
pre-replication complex (A Vidal-Fernandez and C Cazaux, unpublished observations).

Altogether, these data support the view that the overexpression of *CDC6*, *PLK1*, *RAD51*, *CLASPIN* and *POLQ* genes promotes cancer

**Table 4.** Multivaried Cox regression analysis of the relationship between survival and DNA replication gene expression by taking into account age, sex, treatment (in three categories: none/chemotherapy/chemotherapy and radiotherapy), T classification (two categories: T0 or T1/T2 or T3), KI67 and PCNA clinical markers

T/N (terciles)	Overall survival			Disease-free survival			Relapse-free survival		
	HR	(95% CI)	P-value	HR	(95% CI)	P-value	HR	(95% CI)	P-value
<i>POLQ</i>									
0.70–3.34 (ref)	1.00			1.00			1.00		
3.36–6.31	4.59	(0.56–37.69)	0.16	1.07	(0.35–3.29)	0.90	0.93	(0.27–3.15)	0.90
6.42–86.70	36.31	(2.55–517.38)	0.008	1.57	(0.40–6.18)	0.51	0.94	(0.20–4.41)	0.94
<i>PLK1</i>									
0.68–6.28 (ref)	1.00			1.00			1.00		
6.31–14.12	4.43	(0.50–39.56)	0.18	3.10	(0.78–12.32)	0.11	3.26	(0.73–14.53)	0.12
14.13–74.19	23.49	(1.91–288.38)	0.01	11.14	(2.35–52.75)	0.002	11.11	(1.97–62.70)	0.006
<i>RAD51</i>									
0.70–2.40 (ref)	1.00			1.00			1.00		
2.48–4.40	0.72	(0.11–4.74)	0.73	0.98	(0.31–3.13)	0.98	1.08	(0.30–3.98)	0.90
4.46–19.65	2.94	(0.51–16.84)	0.23	3.44	(1.00–11.83)	0.05	4.00	(0.96–16.68)	0.06
<i>CLASPIN</i>									
0.74–3.27 (ref)	1.00			1.00			1.00		
3.39–6.15	18.44	(1.52–223.05)	0.02	4.68	(1.51–14.55)	0.008	4.23	(1.23–14.56)	0.02
6.25–31.81	18.50	(1.28–267.42)	0.03	4.93	(1.50–16.17)	0.008	4.73	(1.30–17.25)	0.02
<i>CDC6</i>									
0.75–4.47 (ref)	1.00			1.00			1.00		
4.55–9.62	5.77	(0.97–34.29)	0.05	1.25	(0.44–3.56)	0.67	0.98	(0.30–3.23)	0.97
10.08–29.65	20.65	(1.54–275.91)	0.02	2.25	(0.46–10.92)	0.31	2.34	(0.38–14.33)	0.36

Abbreviations: CI, confidence interval; HR, hazard ratio.

**Figure 2.** OS of low- and high-risk groups on the basis of five-gene signature. Probabilities for OS were estimated using the Kaplan-Meier method. The hazard ratio (HR) has been obtained using a Cox regression model with the group as covariate.

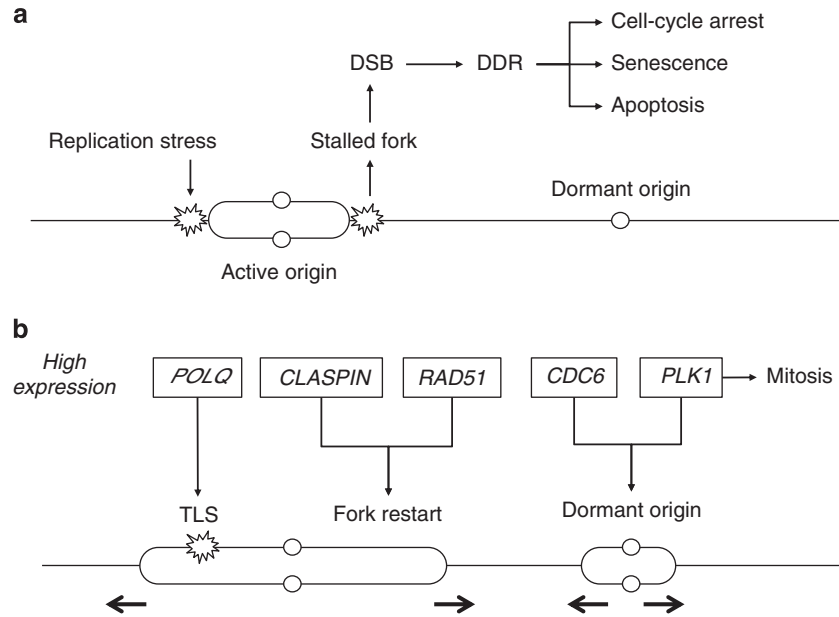
progression by increasing resistance to endogenous replication stress in a concerted manner. Besides this five-gene prognostic signature, this view is also consistent with the overexpression of many genes involved in the initiation of DNA replication and fork repair, and the downregulation of TLS polymerase genes, with the notable exception of *POLQ*.

Because of the high risk of recurrence, adjuvant chemotherapy is always recommended for stage II and III patients and for a subset of stage I patients, depending on the legislation of the country. Many among these patients will receive needless chemotherapy, either because of primary chemoresistance or because of their good intrinsic prognostics. Although recent multigene signatures look promising,<sup>2–4</sup> there is still no reliable predictive genetic marker that could help clinicians choose the adequate strategy.<sup>1</sup> In patient treated with platinum-based adjuvant chemotherapy, significant prognostic values have been associated to the expression of the nucleotide excision repair gene *ERCC1*,<sup>22</sup> Combined with the mismatch repair gene *MSH2*, *ERCC1* expression also helps identify patients who would benefit from chemotherapy.<sup>51</sup> Here, we show that genes involved in the response to DNA replication stress may help refine lung cancer prognosis after surgery by efficiently discriminating between low-risk and high-risk patients, independently of the therapeutic strategies or tumor stages.

## MATERIALS AND METHODS

### Patients, tumor samples

Patients were treated and followed-up at the Thoracic Oncology Department of the Toulouse University Hospital (France). Coupled tumoral and normal lung tissue samples were collected from chemo-naïve patients surgically treated from 2006–2010 for stage I to III primary lung adenocarcinoma then analyzed in the Pathological Department by applying the latest WHO classification. The characteristics of the patients and tumors are described in Supplementary Table S1. Immediately after surgery, tumoral and normal lung tissue samples were snap-frozen in liquid nitrogen, and stored at  $-80^{\circ}\text{C}$ . Normal lung tissues were taken from the surgical specimens, at  $>3$  cm of distance from the tumor. Diagnosis was assessed by a lung cancer pathologist by applying the last WHO classification, and clinicopathological stage was assigned according



**Figure 3.** Model of the consequences of the overexpression of *POLQ*, *PLK1*, *CLASPIN*, *RAD51* and *CDC6* genes on the resistance to replication stress. **(a)** Deregulated oncogene expression induces replication stress in cancer cells. Replication fork stalling leads to the formation of DNA double-strand breaks and to the constitutive activation of the DNA damage response (DDR), driving cells into senescence, apoptosis or irreversible checkpoint arrest. **(b)** Cells overexpressing the five-gene signature are protected from replication stress in different ways. Upregulation of *POLQ* would favor replication past DNA lesions. Overexpression of *CLASPIN* and *RAD51* would favor the maintenance and the recovery of stalled replication forks. Finally, upregulation of *CDC6* and *PLK1* could increase the number of licensed backup origins. *PLK1* overexpression could also promote cell division by helping cells bypass the G2/M checkpoint.

to the Tumor-Node-Metastasis classification. Frozen samples were analyzed on hematoxylin- and eosin-colored slides, and tumor-rich areas were marked by the pathologist to ensure that the maximum amount of tumor material was collected for the genetic testing. All specimens contained at least 80% tumor cells, and necrosis was removed. Exclusion criteria included non-adenocarcinoma, stages IIIb and IV tumors. Written informed consent was obtained from all patients before testing. All informed consents were collected and stored in the Pathology Department. This study was approved by the Ethics of Human Research Committee at the Pathology Toulouse Hospital. The experiments conformed to the principles set out in the WMA Declaration of Helsinki.

#### RNA extraction and quantification

Thick frozen sections of tissues were obtained by using a Leica CM3050S cryostat (Nanterre, France). Frozen tissue sections of thickness 10–300  $\mu\text{m}$  were crisped for 90 s at room temperature by using a 5-mm diameter stainless steel beads and a tissue lyser (Qiagen, Courtaboeuf, France); then total RNA was extracted with the RNeasy extraction kit according to the manufacturer (Qiagen). The quality of total RNA (DO260/DO280 > 1.7) was assessed with the Agilent 2100 bio-analyzer using the RNA Nano Lab chip, 6000 Nano Assay kit (Agilent Technologies, Santa Clara, CA, USA). Its quantity was estimated with Nanodrop (Thermo Scientific, Brebières, France). The four most stable control housekeeping genes (*GUSB*, *IPO8*, *HMBS*, *UBS*) were selected by the GeNorm and BestKeeper softwares among 16 tested on the TaqMan Low Density Human Endogenous Control Array (Applied Biosystems, Carlsbad, CA, USA) after amplifying in triplicate 800 ng cDNA from four tumor and their four counterpart normal biopsies using the TaqMan Universal PCR Master Mix, the TaqMan Low Density Array technology (Applied Biosystems) and the 7900HT fast real-time PCR system.

To quantify RNAs from tumor and normal tissues, cDNAs were first preamplified in the presence of the 3R probes (TaqMan gene Expression Assays, Applied Biosystems) in TaqMan Preamp Master Mix (Early Access, Applied Biosystems). These products were then amplified using the Dynamic Array technology (Fluidigm, BioMark, Pueblo, CO, USA). Preamplified products were incubated in DNA-Binding Sample Loading reagent (BioMark), Master Mix (Applied Biosystems) and probes, then injected in nanotubes-containing Integrated Fluidic Circuit loader and amplified with the BioMark amplifier. Fluidigm data were analyzed with the

GenEx software after normalizing tumor (T) and normal (N) tissues the levels of transcripts to the mean level of the four selected stable genes. Relative levels of expression in the tumor sample compared with the adjacent normal tissue were expressed by either  $\Delta(\Delta\text{Ct})$  or T/N ratios. For a given transcript,  $\Delta(\Delta\text{Ct}) = (\text{Ct}_\text{T} - \text{Ct}_\text{HKN}) - (\text{Ct}_\text{T} - \text{Ct}_\text{HKT})$  where  $\text{Ct}_\text{T}$  and  $\text{Ct}_\text{N}$  are the Cycle Thresholds (Ct) in tumor and adjacent normal tissue, respectively, and  $\text{Ct}_\text{HKN}$  and  $\text{Ct}_\text{HKT}$  are the mean Ct's from the expression of the four housekeeping (HK)/stable genes, that is, *GUSB*, *IPO8*, *HMBS*, *UBS* in tumor and counterpart, respectively. T/N was calculated from the  $\Delta(\Delta\text{Ct}) = \log_2(T/N)$  transformation.  $T/N > 1$  ( $< 1$ ) indicates a higher (lower) expression in the tumor sample compared with the adjacent normal tissue.

#### Statistical analysis

Statistical analyses were performed using the free statistical software R (version 2.9.2), including the 'survival' packages (R development Team, <http://cran.r-project.org/>) and Stata SE 11.2 software (Stata Corporation, College Station, TX, USA). When comparing the expressions in cancer tissues, the major parameters were the individual T/N ratios. The probability to observe > 50% of patients who displayed a misregulation ratio higher ( $T/N > 2$ ) or lower ( $T/N < 2$ ) than 2 was assessed by a bilateral binomial test. Other thresholds were tested by unilateral binomial tests ( $T/N > 5$  or  $T/N < 1/5$ ,  $T/N > 4$  or  $T/N < 1/4$ ,  $T/N > 3$  or  $T/N < 1/3$ , and  $T/N > 2$  or  $T/N < 1/2$ ). Correlations between genes were assessed with a Pearson's correlation coefficient. A clustering algorithm was also applied: hierarchical ascending classification. This clustering was carried out for genes using Ward's method of linkage and correlation distance as a metric. Expression levels were classified in three categories, according to the terciles of the T/N distribution. Expression levels were compared by  $\chi^2$ - or Fisher's exact test in relation to treatment (surgery only, surgery-chemotherapy-radiotherapy, or surgery-chemotherapy) or tumor stage. Survival probabilities were estimated using Kaplan-Meier method by investigating the OS (time between surgery and death due to any cause, that is, both related and unrelated to the NSCLC), disease-free survival (time between surgery and either death due to any cause, loco-regional recurrence or detection of distant metastases) and relapse/recurrence-free survival (time between surgery and loco-regional/distant recurrence or distant metastases). The log-rank test was used to compare survival curves. We used a multivariate Cox's proportional hazards regression model adjusted on sex,



age, treatment, tumor grade and expression level of *Ki67* and *PCNA* genes. To build the five-gene signature, we used a multivariate Cox regression model, including the five prognostic genes split in tertiles. We used the linear predictive value of the Cox model as a risk score, which is the expression level for each gene weighted by the associated coefficients of the Cox model. The cutoff of the risk score was determined to obtain the most significant difference between the two groups using the log-rank test (maxstat package in R). For all statistical tests, differences were considered significant at the 5% level. The Benjamini and Yekutieli correction for multiple testing was applied with binomial tests.

### Validation of prognostic signature

Our signature was tested in a reference and independent microarray data set, that is, the US National Cancer Institute Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma, which includes University of Michigan Cancer Center, Moffitt Cancer Center, Memorial Sloan-Kettering Cancer Center and the Dana-Farber Cancer Institute. To reflect our patient population as much as possible, validation was restricted to NSCLC stage IA to IIIA patients who did not receive any neo-adjuvant treatment. Following sorting, the DDC validation data set included 400 patients. CEL files data extraction and normalization were performed with R Bioconductor ([www.bioconductor.org](http://www.bioconductor.org)), AMP (<http://compbio.dfci.harvard.edu/ampand>) and Cluster 3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>). Each data set was independently normalized using RMA procedure. Cluster 3.0 was used to filter data (removing all genes that have missing values in >80% of the patients), log<sub>2</sub> transformation and normalization. Link between probe sets and gene names were identified with Clone Gene ID Converter (<http://idconverter.bioinfo.cnio.es>) and GeneAnnot (<http://genecards.weizmann.ac.il/geneannot/index.shtml>). The probability to observe >50% of patients who displayed a gene under/over-regulated was assessed by a bilateral binomial test. The threshold was the median expression value of the given control genes (generated from the Universal Human Reference RNA from Stratagene, La Jolla, CA, USA). Kaplan-Meier OS analysis and log-rank test were performed with R (survival package), in order to determine the prognostic significance of our five genes in the validation set. For a given gene, the 'low risk group' represents the half population whose expression value was under the median expression value. The 'high risk group' represents the other half population. For statistical tests, differences were considered significant at the 5% level. The Benjamini and Yekutieli correction for multiple testing was applied with binomial tests.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

This work was supported by the National Institute of Cancer (INCa GSO, grant ACI 3R to CC) and Ligue Nationale Contre le Cancer ('équipe labellisée 2010' to JSH). We acknowledge C Boursier (Ranguelil-Larrey hospital) for technical assistance and Yea-Lih Lin for critical reading of the manuscript.

### REFERENCES

- Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* 2010; **102**: 464-474.
- Zhu C, Ding K, Strumpf D, Weir B, Meyerson M, Pennell N *et al*. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 2010; **28**: 4417-4424.
- Kratz JR, He J, Van Den Eeden SK, Zhu ZH, Gao W, Pham PT *et al*. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. *Lancet* 2012; **379**: 823-832.
- Chen DT, Hsu YL, Fulp WJ, Coppola D, Haura EB, Yeatman TJ *et al*. Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. *J Natl Cancer Inst* 2011; **103**: 1859-1870.
- Burhans WC, Carr AM, Wahl GM. DNA replication and cancer. In: DePamphilis M (ed.). *DNA replication and human disease*. Cold Spring Harbor: New York, 2006.
- Bartkova J, Rezaei N, Lontos M, Karakaidos P, Kleitas D, Issaeva N *et al*. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature* 2006; **444**: 633-637.

- Bertwistle D, Ashworth A. Functions of the BRCA1 and BRCA2 genes. *Curr Opin Genet Dev* 1998; **8**: 14-20.
- Marra G, Boland CR. Hereditary non-polyposis colorectal cancer (HNPCC): the syndrome, the genes and historical perspectives. *J Natl Cancer Inst* 1995; **87**: 1114-1125.
- Sancar A. Mechanisms of DNA excision repair. *Science* 1994; **266**: 1954-1956.
- Masutani C, Kusumoto R, Yamada A, Dohmae N, Yokoi M, Yuasa M *et al*. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase  $\epsilon$ . *Nature* 1999; **399**: 700-704.
- O'Driscoll M, Jeggo PA. Clinical impact of ATR checkpoint signalling failure in humans. *Cell Cycle* 2003; **2**: 194-195.
- Kamath-Loeb AS, Loeb LA. DNA replication and cancer. In: DePamphilis M (ed). *DNA helicases and human diseases*. Cold Spring Harbor: New York, 2006.
- Kunkel TA. Considering the cancer consequences of altered DNA polymerase function. *Cancer Cell* 2003; **3**: 105-110.
- Mitchell JR, Hoeijmakers JH, Niedernhofer LJ. Divide and conquer: nucleotide excision repair battles cancer and ageing. *Curr Opin Cell Biol* 2003; **15**: 232-240.
- Venkatesan RN, Treuting PM, Fuller ED, Goldsby RE, Norwood TH, Gooley TA *et al*. Mutation at the polymerase active site of mouse DNA polymerase  $\delta$  increases genomic instability and accelerates tumorigenesis. *Mol Cell Biol* 2007; **27**: 7669-7682.
- Arentson E, Faloon P, Seo J, Moon E, Studts JM, Fremont DH *et al*. Oncogenic potential of the DNA replication licensing protein Cdt1. *Oncogene* 2002; **21**: 1150-1158.
- Honeycutt KA, Chen Z, Koster MI, Miers M, Nuchtern J, Hicks J *et al*. Deregulated minichromosomal maintenance protein MCM7 contributes to oncogene driven tumorigenesis. *Oncogene* 2006; **25**: 4027-4032.
- Shima N, Alcaraz A, Liachko I, Buske TR, Andrews CA, Munroe RJ *et al*. A viable allele of Mcm4 causes chromosome instability and mammary adenocarcinomas in mice. *Nat Genet* 2007; **39**: 93-98.
- Gonzalez MA, Tachibana KE, Laskey RA, Coleman N. Control of DNA replication and its potential clinical exploitation. *Nat Rev Cancer* 2005; **5**: 135-141.
- Fong KM, Zimmerman PV, Smith PJ. Tumor progression and loss of heterozygosity at 5q and 18q in non-small cell lung cancer. *Cancer Res* 1995; **55**: 220-223.
- Esposito V, Baldi A, De Luca A, Micheli P, Mazzeo G, Baldi F *et al*. Prognostic value of p53 in non-small cell lung cancer: relationship with proliferating cell nuclear antigen and cigarette smoking. *Hum Pathol* 1997; **28**: 233-237.
- Zheng Z, Chen TC, Li X, Haura E, Sharma A, Bepler G. DNA synthesis and repair genes RRM1 and ERCC1 in lung cancer. *N Engl J Med* 2007; **356**: 800-808.
- Thomae AW, Pich D, Brocher J, Spindler MP, Berens C, Hock R *et al*. Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. *Proc Natl Acad Sci USA* 2008; **105**: 1692-1697.
- Maiorano D, Cuvier O, Danis E, Mechali M. MCM8 is an MCM2-7-related protein that functions as a DNA helicase during replication elongation and not initiation. *Cell* 2005; **120**: 315-328.
- Pillaire MJ, Selves J, Gordien K, Gouraud PA, Gentil C, Danjoux M *et al*. A 'DNA replication' signature of progression and negative outcome in colorectal cancer. *Oncogene* 2010; **29**: 816-887.
- Lemee F, Bergoglio V, Fernandez-Vidal A, Machado-Silva A, Pillaire MJ, Bieth A *et al*. DNA polymerase theta is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proc Natl Acad Sci USA* 2010; **107**: 13390-13395.
- Sobol RW, Horton JK, Kuhn R, Gu H, Singhal RK, Prasad R *et al*. Requirement of mammalian DNA polymerase-beta in base-excision repair. *Nature* 1996; **379**: 183-186.
- Capp JP, Boudsocq F, Besnard AG, Lopez B, Cazaux C, Hoffmann JS *et al*. Involvement of DNA polymerase mu in the repair of a specific subset of DNA double strand breaks in mammalian cells. *Nucleic Acids Res* 2007; **35**: 3551-3560.
- Shedden K, Taylor JMG, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL *et al*. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008; **14**: 822-827.
- Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F, Zieger K *et al*. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* 2005; **434**: 864-870.
- Gorgoulis VG, Vassiliou LV, Karakaidos P, Zacharatos P, Kotsinas A, Liloglou T *et al*. Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* 2005; **434**: 907-913.
- Karakaidos P, Taravira S, Vassiliou LV, Zacharatos P, Kastrinakis NG, Kougiou D *et al*. Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: synergistic effect with mutant p53 on tumor growth and chromosomal instability—evidence of E2F-1 transcriptional control over hCdt1. *Am J Pathol* 2004; **165**: 1351-1365.
- Tsimaratou K, Kleitas D, Kastrinakis NG, Tsantoulis PK, Evangelou K, Sideridou M *et al*. Evaluation of claspin as a proliferation marker in human cancer and normal tissues. *J Pathol* 2007; **211**: 331-339.

- 34 Halazonetis TD, Gorgoulis VG, Bartek J. An oncogene-induced DNA damage model for cancer development. *Science* 2008; **319**: 1352–1355.
- 35 Glover TW, Arlt MF, Casper AM, Durkin SG. Mechanisms of common fragile site instability. *Hum Mol Genet* 2005; **15**: 197–205.
- 36 Rey L, Sidorova J, Puget N, Boudsocq F, Biard D, Monnat RJ *et al*. Human DNA polymerase eta is required for common fragile site stability during unperturbed DNA replication. *Mol Cell Biol* 2009; **29**: 3344–3354.
- 37 Toledo LI, Murga M, Fernandez-Capetillo O. Targeting ATR and Chk1 kinases for cancer treatment: a new model for new (and old) drugs. *Mol Oncol* 2011; **5**: 368–373.
- 38 Bunting S, Callén E, Wong N, Chen H, Polato F, Gunn A *et al*. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* 2010; **141**: 243–254.
- 39 López-Contreras AJ, Gutierrez-Martinez P, Specks J, Rodrigo-Perez S, Fernandez-Capetillo O. An extra allele of Chk1 limits oncogene-induced replicative stress and promotes transformation. *J Exp Med* 2012; **209**: 455–461.
- 40 Di Paola D, Rampakakis E, Chan MK, Arvanitis DN, Zannis-Hadjopoulos M. Increased origin activity in transformed versus normal cells: identification of novel protein players involved in DNA replication and cellular transformation. *Nucleic Acids Res* 2010; **38**: 2314–2331.
- 41 Ge XQ, Jackson DA, Blow JJ. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev* 2007; **21**: 3331–3341.
- 42 Ibarra A, Schwob E, Méndez J. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci USA* 2008; **105**: 8956–8961.
- 43 Izawa N, Wu W, Sato K, Nishikawa H, Kato A, Boku N *et al*. HERC2 interacts with Claspin and regulates DNA origin firing and replication fork progression. *Cancer Res* 2011; **71**: 5621–5625.
- 44 Kumagai A, Dunphy WG. Claspin, a novel protein required for the activation of Chk1 during a DNA replication checkpoint response in *Xenopus* egg extracts. *Mol Cell* 2000; **6**: 839–849.
- 45 Costanzo V. Brca2, Rad51 and Mre11: performing balancing acts on replication forks. *DNA Repair (Amst)* 2011; **10**: 1060–1065.
- 46 Barr FA, Silljé HH, Nigg EA. Polo-like kinases and the orchestration of cell division. *Nat Rev Mol Cell Biol* 2004; **5**: 429–440.
- 47 Song B, Liu XS, Davis K, Liu X. Plk1 phosphorylation of Orc2 promotes DNA replication under conditions of stress. *Mol Cell Biol* 2011; **31**: 4844–4856.
- 48 Yata K, Lloyd J, Maslen S, Bleuyard JY, Skehel M, Smerdon SJ *et al*. Plk1 and CK2 act in concert to regulate Rad51 during DNA double strand break repair. *Mol Cell* 2012; **45**: 371–383.
- 49 Seki M, Masutani C, Yang LW, Schuffert A, Iwai S, Bahar I *et al*. High-efficiency bypass of DNA damage by human DNA polymerase Q. *EMBO J* 2004; **23**: 4484–4494.
- 50 Seki M, Wood RD. DNA polymerase theta (POLQ) can extend from mismatches and from bases opposite a (6-4) photoproduct. *DNA Repair (Amst)* 2008; **7**: 119–127.
- 51 Kamal NS, Soria JC, Mendiboure J, Planchart D, Olausson KA, Rousseau V *et al*. MutS homologue 2 and the long-term benefit of adjuvant chemotherapy in lung cancer. *Clin Cancer Res* 2010; **16**: 1206–1215.



*Oncogenesis* is an open-access journal published by Nature Publishing Group. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on the *Oncogenesis* website (<http://www.nature.com/oncsis>).