Received: 3 January 2020 | Revised: 27 April 2020 | Accepted: 11 May 2020

DOI: 10.1002/ijc.33121

TUMOR MARKERS AND SIGNATURES

IJC International Journal of Cancer

Molecular characterisation of *TP53* mutated squamous cell carcinomas of the lung to identify putative targets for therapy

Vilde D. Haakensen[1,2] | Anand Khadse[1,3] | Vandana Sandhu[1,3,7] | Ann Rita Halvorsen[1,4,8] | Steinar K. Solberg[5] | Lars H. Jørgensen[5] | Odd Terje Brustugun[1,6] | Elin H. Kure[1,3] | Åslaug Helland[1,2]

[1]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

[2]Department of Oncology, Oslo University Hospital, Oslo, Norway

[3]University of South-Eastern Norway, Bø, Telemark, Norway

[4]Institute for Clinical Medicine, University of Oslo, Oslo, Norway

[5]Department of Cardiothoracic Surgery, Oslo University Hospital, Oslo, Norway

[6]Section of Oncology, Drammen Hospital, Vestre Viken Hospital Trust, Drammen, Norway

[7]University Health Network, Toronto, Ontario, Canada

[8]Department of Cancer Treatment, Section of Radiation Therapy, Oslo University Hospital, Oslo, Norway

Correspondence
Vilde D. Haakensen, Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Post box 4959 Nydalen, 0424 Oslo, Norway.
Email: vilde.haakensen@gmail.com

Funding information
Kreftforeningen, Grant/Award Number: 88503-2013; South-Eastern Norway Regional Health Authority, Grant/Award Number: 2010046

Abstract

Personalised cancer treatment depends on identification of therapeutically relevant biological subgroups of patients for assessing effect of treatment and to discover new therapeutic options. By analyses in heterogeneous patient populations, the effects may be lost in noise. Squamous cell carcinoma of the lung is a major killer worldwide. Despite recent advances, mortality is high and response to therapies varies greatly from patient to patient. Target search in biologically relevant subgroups may identify treatment options not so far discovered. A total of 198 patients undergoing surgery for squamous cell carcinomas of the lung were included in the study. The tumours were analysed for copy number alterations (n = 152) and gene expression from tumour (n = 188) and normal lung (n = 21), with both data levels present in 140 patients. We studied alterations in tumours harbouring mutations in *TP53* and in previously published gene expression subtypes. Genes with consistent alterations in both genomic levels were identified as putative biomarkers. Results were validated in TCGA. The most convincing biomarker in *TP53* mutated squamous cell carcinomas of the lung was *BIRC5* with amplification in 36% of mutated samples, 5% in wild-type samples and a 17%-fold change of expression between *TP53* mutated tumours and normal lung tissue. *BIRC5* was significantly altered in the classical and primitive subtypes. We suggest *BIRC5* as a putative predictive biomarker and putative druggable target in squamous cell lung carcinomas harbouring *TP53* mutation or classified as classical and primitive subtypes.

KEYWORDS

biomarker, squamous cell lung carcinoma, *BIRC5*, molecular profiling, personalised cancer treatment

1 | INTRODUCTION

Personalised cancer treatment means giving a treatment that is likely to be effective in a given patient based on individual characteristics of

Abbreviations: DAVID, the Database for Annotation, Visualisation and Integrated Discovery; FDR, false discovery rate; OUH, Oslo University Hospital; SAM, significance analysis of microarrays; TCGA, The Cancer Genome Atlas.

the tumour/patient. Studies testing inhibitors in unselected patient populations are likely to fail due to a proportion of the patients lacking a target for the treatment and hence an effect. This may hinder the identification of treatments that may be effective for smaller subgroups of patient. There is hence a need to enrich the study population for patients likely to benefit from the treatment by identifying useful predictive markers.

Squamous cell carcinoma (SCC) of the lung is a major killer worldwide. Despite recent advances, mortality is high and response to therapies varies greatly from patient to patient. For the recently studied combination of chemotherapy (carboplatin/paclitaxel) and immunotherapy (pembrolizumab) the response rate was found to be 58%.[1] For more traditional chemotherapy-regimens, the response rates are 20%-40%.[2,3] Although immunotherapy has revolutionised lung cancer treatment, most patients do not benefit from the treatment with response rates of 46% for the group with PD-L1 > 50% for monotherapy.[4] The greatly varying response rates reflect different underlying biologies. Tumour characteristics are known to be associated with treatment response in many cancers. For lung SCC, tumour expression of PD-L1 and tumour mutational burden have shown association with response to immunotherapy.[5,6]

TP53 is a well-known tumour suppressor gene and inactivation, primarily through genetic mutations or loss of heterozygosity, is frequent as an early step in carcinogenesis.[7] The gene product is essential in cellular response to stress by cell cycle regulation and induction of DNA repair and apoptosis.[8] As mutations in *TP53* are common in many cancer types, treatments linked to this genotype could potentially be effective in similar subgroups of various cancer diseases. Mutations in *TP53* have been linked to exposure of various carcinogens, including tobacco and the incidence of *TP53* mutations is particularly high in smoking-associated cancers, such as SCC of the lung.[7] Although *TP53* mutations have been widely studied in cancer, therapies targeting the gene alterations have not been introduced to the clinic and alternative approaches are needed to improve the predictive power and response to therapy in patients with *TP53* mutated lung cancers.

Subgroups of lung cancer with biological differences relevant to cancer treatment can be identified in various ways. For targeted therapy, the presence of a targetable mutation is essential. Biological differences with therapeutic consequences other than targetable mutations require a different approach. For breast cancer, unsupervised clustering of the 535 most variable genes from a 44K gene expression microarray chip identified biological subgroups that have proven to be robust and clinically relevant.[9,10] Some of these subgroups have later been found to overlap with subgroups determined by genetic alterations suitable for targeted therapy. Similar subtypes with varying prognosis have been published for lung SCC.[11] Wilkerson and colleagues also used expression of genes with high reliability and variability to identify subtypes, but used a total of 2307 genes for unsupervised clustering. The four subtypes identified had specific molecular and clinical characteristics. Patients with tumours classified as the primitive subtype have a poor prognosis and the tumours display features associated with high proliferation. The classical subtype is more common and tumours of this subtype overexpress genes associated with xenobiotic metabolism. The secretory subtype is characterised by an immune profile and the basal subtype is dominated by cell adhesion. These subtypes have, however, not yet reached the clinic. In current clinical practice, PD-L1 and sometimes tumour mutational burden are the biomarkers used for SCC, both with limitations in predicting therapy response.[12]

In our study, we explore the biology of *TP53* mutated lung SCC to propose putative biomarkers or targets of therapy. We have studied early-stage lung carcinomas integrating information at the genetic and transcriptomic levels to identify alterations that could be predictive biomarkers in the presence of *TP53* mutations and in the published subtypes of SCC.[11] Based on these subgroups, we have identified putative biomarkers and plausible targets of therapy in these predefined biological subgroups.

What's new?

This study presents a target gene search combining copy number alteration and gene expression to identify putative genes for therapeutic and predictive approaches in *TP53* mutated lung squamous cell carcinoma (SCC) and published gene expression subtypes with high percentages of *TP53* mutations. Several potential biomarkers and therapeutic targets emerged from these pre-defined biological subgroups. The results suggest that *BIRC5* is one of the most appealing targets in *TP53* mutated cancers and in the classical and primitive subtypes and should be tested clinically in these subgroups. Testing in biologically defined subgroups may increase likelihood of discovering clinically relevant treatment effects.

2 | MATERIALS AND METHODS

Patients undergoing surgery for SCC from 2006 to 2015 were included in the study (n = 198). All patients signed informed consent. The study was approved by the Regional Ethical Committee (ref: S-06402b). Clinical data from medical journals including follow-up was available for all patients. Tumour tissue from the tumour centre was snap-frozen in liquid nitrogen and stored at −80°C until DNA extraction. A pathologist reviewed all samples to ensure a tumour percentage above 70%. Matched normal lung tissue was collected from 21 patients. Clinical data is shown in Table 1.

2.1 | Copy number alterations

DNA was extracted from the lung tumours using Maxwell® 16 DNA Purification Kits according to the standard Technical Manual, Literature # TM284 (http://www.promega.com) and using a Maxwell®16

TABLE 1 Patient characteristics

	Total	*TP53* mutated	*TP53* wild-type
Patients included (total)	198	76	35
Gender			
Male	133 (67%)	52 (68%)	23 (66%)
Female	65 (33%)	24 (32%)	12 (34%)
Age			
Mean (min-max)	67 (43-82)	67 (43-82)	68 (58-80)
Pack-years			
Mean (min-max)	40 (3-145)	38 (9-145)	37 (3-79)
pStage			
1	113 (57%)	39 (52%)	24 (69%)
2	59 (30%)	26 (34%)	7 (20%)
3	25 (13%)	10 (13%)	4 (11%)
4	1 (0%)	1 (1%)	0 (0%)
TP53			
wt	35 (18%)		
mut	76 (38%)		
NA	87 (44%)		
Survival status			
Lung cancer	57 (29%)	28 (37%)	8 (23%)
Cardiovascular	9 (5%)	4 (5%)	4 (11%)
Other	39 (20%)	16 (21%)	8 (23%)
Alive	92 (46%)	28 (37%)	15 (43%)
Follow-up time, months			
Mean (min-max)	56 (2-143)	62 (2-130)	68 (6-143)

instrument. Affymetrix Genome-Wide SNP Nsp/Sty 6.0 Arrays were used for copy number alterations performed by AROS Applied Biotechnology A/S (www.arosab.com; project no. A2714, samples no. 095-188). Raw signal intensities were extracted and quantile normalised using Affymetrix Power Tools (APT) and were converted to Log *R* ratio (LRR) and B Allele Frequency (BAF) using PennCNV-Affy package.[13] Log *R* values then corrected for GC bias. A locus is denoted as amplified if the copy number is 3 or higher.

2.2 | Gene expression

Gene expression analysis was performed as previously described.[14] For RNA extraction from 188 tumours and 21 matched normal samples, standard TRIZOL methods (Invitrogen, Carlsbad, California) were used as specified by the manufacturer's instructions. NanoDrop ND-1000 spectrometer (NanoDrop Technologies, Wilmington, Delaware) was used to determine RNA quantity and quality (yield, 260/280 ratio and 260/230 ratio). RNA integrity numbers (RIN) were measured by the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California) according to the manufacturer's protocol. Gene expression microarrays from Agilent technologies (SurePrint G3 human GE v3, 8 × 60 K) were used for gene expression. The raw data were processed with Agilent's Feature Extraction Software with default parameters (Agilent Feature Extraction version 10.7.3.1). Probes were collapsed by median, samples were quantile normalised, and the data were log2 transformed.

2.3 | *TP53* sequencing

Consecutive samples (n = 111) were tested for *TP53* mutations by the Sanger Sequencing method as previously described.[15] In brief, an Applied Biosystems 3730 DNA analyser was used according to the supplier's handbook, Applied Biosystems 3730/3730X/DNA Analysers Part 4331467 Rev.B (http://www.appliedbiosystems.com). Exons 2 to 11 were investigated. SeqScape v 2.5 was used to align and analyse the sequences (TP53 accession no: NM_000546 (http://www.ncbi.nlm.nih.gov/nuccore/NM_000546; TP53 ref_NC000017.9_NT010718.15). Two independent readers analysed each sequence.

2.4 | Bioinformatic analyses

Allele-specific copy number analysis of tumours (ASCAT) was applied to estimate copy number gains (amplifications) and losses (deletions).[16] Segmentation of the genome was obtained using the Allele-

Specific Piecewise Constant Fitting (ASPCF) algorithm with penalty parameter γ = 25 and the data were used in ASCAT to estimate tumour ploidy, aberrant cell fraction and the allele-specific copy number profile of the tumour samples. In the OUH cohort, flat profiles of 46 samples indicated insufficient percentage of tumour cells and these samples were excluded from further analysis, leaving available information about copy number alterations for 152 patients. Similarly, copy number aberration profiles were obtained for the TCGA cohort (n = 443).

2.4.1 | Identification of copy number gains and losses

For each sample, we obtained segmentation of the genome where each segment corresponds to a genomic region with a specific copy number. The copy number of the segment is the sum of the copy number of the major and minor allele in that region. Tumour ploidy was calculated as the median total copy number across the complete genome and was rounded to the nearest whole number. Segments with copy number greater than the ploidy were called as gains (amplifications) whereas the segments with copy number less than the ploidy were called as losses (deletions). The coordinates given in the Affymetrix array annotation file were used to map aberrant genomic regions to the genes.

2.4.2 | Frequency plots

Samples were assigned to two groups based on the presence of *TP53* mutations, as *TP53* mutant or wild-type. Frequency of gains (or losses) at every genomic position in a group was calculated as fraction of number of samples with gain (or loss) at a position to total number of samples in the group. Frequencies of gains are plotted on the *y*-axis in a positive scale while the frequencies of losses are plotted in a negative scale. Chromosome-wise genomic positions are represented on the *x*-axis. Regions with significant differences in gain or loss at given position in two groups were determined using two-proportions *z*-test implemented in *prop.test* function in R.

2.4.3 | Correlation analysis

The Spearman correlation coefficient was calculated using the R (v3.5.0) to estimate the correlation between copy number alteration and mRNA expression in 140 OUH samples and 443 samples in TCGA cohort. Expression data for 12 samples in OUH cohort were unavailable. Copy number value for the gene was calculated by subtracting the absolute ploidy from total copy number of the sample at the gene position. The quantile-normalised gene expression data in OUH cohort and the RSEM normalised gene values in TCGA cohort were used in correlation analysis. The *P* values and adjusted *P* values (FDR) were reported for the significant association at $P < .05$.

2.4.4 | Other analyses

Subgroups of lung SCC were identified by the presence of *TP53* mutations and by the previously published lung SCC subtypes.[11]

The candidate genes were tested for survival in patients with *TP53* mutated tumours receiving adjuvant platinum-based chemotherapy. Gene expression of the transcripts was divided in high and low by the median value.

To identify genes associated with subtype, significance analysis of microarrays (SAM) was performed with the four subtypes as stratifiers[17] and 100 permutations using the R-package *samr v3.0*. Two-sided SAM was used to identify genes differentially expressed between two groups of patients. A significance level of 0.05 was used and false-discovery rate (FDR) was used to correct for multiple testing. Only genes with both gene expression and copy number alteration data available were included.

Lists of putative genes were analysed using Reactome Analysis Tool[18,19] to identify pathways overrepresented in candidate genes.

3 | RESULTS

Gene expression data were available for 188 of 198 tumours and 21 matched normal samples. Information about copy number alterations was available for 152 patients, with both data levels available for 140 patients.

3.1 | *TP53* mutated tumours

Patient characteristics are listed in Table 1. *TP53* mutation status was available for 111 of 198 patients. Of these 111, 76 (68%) were mutated. Of the 198 samples, information about copy number alterations was available for 152 samples. *TP53* mutation status was available for 90 samples of which 72 (80%) were mutated. In addition, one silent mutation was counted as wild-type. Of the 18 wild-type samples, 10 (56%) had deletions in the region harbouring the *TP53* gene, 8 (44%) had no genomic alteration in the gene and there were no amplifications in this specific region. Of the samples with a mutation in *TP53*, four samples (6%) also had a genomic amplification in the same region. The TP53-mutations detected are listed in File S1.

Frequency plots of *TP53* wild-type (n = 18 OUH, n = TCGA) and *TP53* mutant (n = 72 OUH, n = TCGA) tumours were generated separately (File S2, Figure S1). Significant differences in amplifications and deletions were observed in 9668 genes of 15 995 tested (60%) with a significance level of 0.05 (prop.test). The highest percentage of significant amplifications in *TP53* mutated vs wild-type samples was found

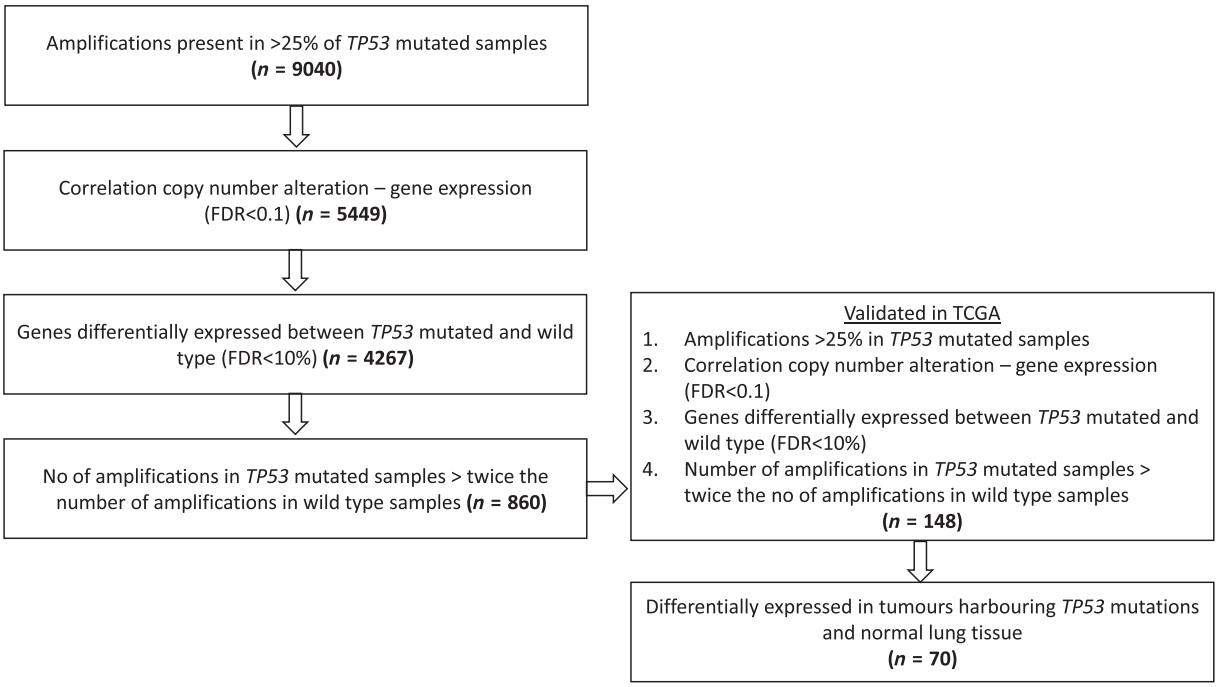

FIGURE 1 Target gene search for *TP53* mutated squamous cell carcinomas. Significant genes are validated in TCGA

on chromosomes 2 (10.7%), 12 (19.9%) and 17 (13.3%), respectively. *TP53* mutant tumours demonstrate a higher level of aberrations than wild-type tumours.

3.2 | Putative target genes

To identify putative target genes in *TP53* mutated tumours, we followed an approach specified in Figure 1. For *TP53* mutated samples, putative genes were those with amplification in more than 25%, correlation between copy number alterations and gene expression and, compared to wild-type tumours both overexpression of the transcript and more than the double frequency of amplifications. In the OUH cohort, 860 genes met the four criteria and 148 of these were validated in the TCGA cohort and are putative biomarkers. Of these 148 genes, 70 have differential expression between *TP53* mutated tumours and normal lung tissue in the OUH cohort, of which 67 are upregulated in tumours and constitute our candidate biomarkers (File S3 and Figure 1).

The 10 genes with highest fold change in gene expression between normal and tumour samples are listed in Table 2. The single gene with the highest fold change gene expression between tumour and normal samples is *BIRC5*. TKI1 has the highest gene expression in *TP53* mutated samples compared to all genes, but due to the relatively high expression in normal samples, the fold change is higher for *BIRC5*. All 67 genes are listed in Table 3 according to cytoband. In addition, three genes had significantly lower gene expression in normal tissue compared to tumour samples, notably *CNP*, *FBXL20* and *GABARAPL1*.

Reactome analysis of the 67 overexpressed genes identified 168 pathways overrepresented with an FDR < 0.05 (File S4). Among these, there was an overrepresentation of pathways involved in cell cycle, cell-cell communication, DNA replication, immune system and metabolism of RNA.

Looking for association to relapse (local or systemic) in the OUH cohort, log-rank tests were performed using gene expression from samples with *TP53* mutations (n = 105) and the list of 70 validated genes with differential expression in normal and tumour tissue. Genes with *P* values <.05 are listed in Table 4. None of the genes was significant after correction for multiple testing or in the TCGA cohort. The top gene, *CNP*, displayed a trend toward higher risk of relapse with lower expression of the gene; see Kaplan-Meier plots of the four top genes in File S2, Figure S2. The other top genes, *SNRPD1*, *EPCAM* and *CGREF1*, all showed a trend toward a higher risk of relapse with higher expression of the gene. Gene expression of *BIRC5* showed no association with relapse-free survival by log-rank test.

Analysing the expression of *BIRC5* with respect to relapse-free survival in all OUH samples with available gene expression and available survival data (n = 184), there is no significant difference in survival between patients with high and low expression of the gene (File S2, Figure S3). The same is seen using only patients with *TP53* mutated tumours (n = 74). In the TCGA cohort, we lacked information about relapse-free survival, but could confirm the lack of prognostic value of *BIRC5* expression with respect to overall survival.

TABLE 2 Top 10 putative targets genes amplified and overexpressed in *TP53* mutant, ranked according to the highest fold change in mutant compared to wild-type tumours by Significance Analysis of Microarrays (SAM)

		Average absolute expression values for each gene				Average expression values for each gene divided by average expression value of all genes in normal samples			
	SAM fold change	Normal	TP53 wild-type	*TP53* mut	*TP53* NA	Normal	TP53 wild-type	*TP53* mut	*TP53* NA
BIRC5	16.8	145	1857	3122	2637	0.17	2.17	3.65	3.09
TK1	11.4	768	6750	11 125	10 025	0.90	7.90	13.02	11.73
EPN3	5.8	55	233	401	386	0.06	0.27	0.47	0.45
TOP2A	5.2	44	176	281	267	0.05	0.21	0.33	0.31
CBX2	5.2	121	538	900	879	0.14	0.63	1.05	1.03
JUP	5.2	1010	3825	6833	6290	1.18	4.48	7.99	7.36
BUB1	4.8	55	212	295	275	0.06	0.25	0.35	0.32
PDK1	4.7	173	669	837	803	0.20	0.78	0.98	0.94
SPAG5	4.2	258	766	1141	1061	0.30	0.90	1.33	1.24
CENPA	4.1	40	104	186	169	0.05	0.12	0.22	0.20

Note: All have FDR < 0.01. Mean absolute gene expression is listed for normal lung tissue samples, tumours harbouring *TP53* mutations and *TP53* wild-type tumours and for tumours not analysed with respect to *TP53*. Average expression values for each gene are divided by the average expression value for all genes in all normal samples (854.7).

TABLE 3 Putative target genes amplified and overexpressed in *TP53* mutated lung squamous cell carcinomas, listed according to cytoband

2p11-13	2p14-16	2p21-25	2q11-14	2q31	12p12-13	17q11-12	17q21	17q25	18q11
BOLA3	MSH2	AGBL5	ANKRD23	ATP5G3	KIAA1467	ALDOC	BRCA1	ACTG1	SNRPD1
CCT7	MSH6	CENPA	ANKRD39	OLA1	LTBR	ERAL1	EPN3	BIRC5	
DGUOK	MTIF2	CGREF1	BUB1	PDK1	PLEKHA5	MRM1	ETV4	CBX2	
FBXO41	PSME4	EPCAM	CKAP2L		STRAP	POLDIP2	JUP	EIF4A3	
GMCL1		EPT1	CLASP1		STYK1	PSMD11	KLHL11	ICT1	
KDM3A		GEMIN6				SPAG5	NME1	MRPS7	
MTHFD2		KIF3C				TLCD1	NME2	TK1	
POLR1A		NOL10				TMEM97	PSMD3		
SMYD5		PREB					PSME3		
SNRNP27		SLC5A6					TOP2A		
SNRPG							UTP18		
SPR									

3.3 | Target gene relation to survival in patients receiving adjuvant chemotherapy

To look for primary chemoresistance, the 70 putative genes were tested for relation to relapse-free survival in patients receiving adjuvant platinum-based chemotherapy (n = 38 including nine relapses). One gene, *PDK1*, had a *P* value < .05, but none reached statistical significance after correction for multiple testing (FDR; File S2, Figure S4). There is a trend toward worse prognosis in platinum-treated patients if the tumour expresses high *BIRC5* with low expression of *PDK1*. Analysing the expression of *BIRC5* with respect to relapse-free survival in patients receiving adjuvant or neoadjuvant platinum therapy (n = 48), there is no significant difference in survival between patients with high and low expression of the gene (File S2, Figure S3). Removing the three neoadjuvant samples does not alter the result. Performing the same analysis in patients with *TP53* mutated tumours all receiving adjuvant chemotherapy (n = 19), there is a borderline significance toward better prognosis for patients with higher expression of the gene.

3.4 | Target gene search within specific gene expression subtypes

The presence of *TP53* mutations is one possible way of identifying subgroups for target gene search. Another plausible subgrouping of SCC lung carcinoma is based on gene expression and is published as intrinsic subtypes.[11] Using these subtypes, target gene search was performed according to the following criteria: (a) Genes amplified in more than 25% of the samples assigned to the subtype. (b) Correlation

TABLE 4 Log-rank test results for progression-free survival of the 70 putative target genes divided in high or low by the median gene expression

Gene	Median	*P* value	FDR
CNP	7.1	.01	0.25
SNRPD1	12.0	.01	0.25
EPCAM	11.7	.02	0.25
CGREF1	6.4	.02	0.25
SLC5A6	8.1	.03	0.25
MTIF2	9.1	.03	0.25
FBXO41	9.2	.03	0.25
PLEKHA5	6.7	.03	0.25
ATP5G3	13.2	.04	0.25
EPT1	7.4	.04	0.25
KLHL11	7.5	.04	0.25
BOLA3	11.6	.04	0.25
ANKRD23	7.3	.05	0.27

TABLE 5 The fraction of *TP53* mutations vary across gene expression subtypes, with the highest fractions in the classical and primitive subtypes, Fisher exact *P* value = .04

	TP53	
	Wild-type, n (%)	Mutated, n (%)
Subtype		
Basal	10 (38)	16 (62)
Classical	11 (23)	37 (77)
Primitive	0 (0)	5 (100)
Secretory	13 (50)	13 (50)
	34 (32)	71 (68)

between copy number alteration and gene expression with (FDR < 0.05). (c) Significant differential expression of genes between subtypes (SAM q < 5%) and contrast in the specific subtype > 0 (signifying a higher gene expression in this subtype). For validation in TCGA, the following criteria were applied: (a) Genes amplified in more than 25% of the samples assigned to the subtype. (b) Correlation between copy number alteration and gene expression with (FDR < 0.05). Genes for which gene expression was missing in either dataset were removed. The candidate genes for each subtype validated in the TCGA are listed in File S5. For the classical and primitive subtypes, *BIRC5* is listed as a top candidate gene. Testing the occurrence of *TP53* mutations in the gene expression subtypes, we found an uneven distribution of mutations with the highest frequencies in the classical and primitive subtypes (Fisher exact *P* value = .04; Table 5).

The candidate genes for each subtype were entered into Reactome for pathway analyses. Pathways with FDR < 0.05 are listed in File S4. While most subtypes have a large representation of pathways involved in the cell cycle, for the Secretory subtype, all significant pathways are involved in the immune system. In addition to cell cycle-related pathways, the Basal, Classical and Primitive subtypes have some pathways involved in mitochondrial elongation and metabolism of proteins.

We wanted to explore if the expression of *BIRC5* would vary between previously published subtypes.[11] The gene expression of the transcript was not significantly different between subtypes by Student's *t* test (File S2, Figure S5).

4 | DISCUSSION

As an effort to meet the challenges of heterogeneity of squamous cell lung cancer and the lack of useful biomarkers, we have performed a focused target gene search in surgically removed samples harbouring *TP53* mutations, using a TCGA cohort for validation. We have identified several putative target genes, focusing particularly on *BIRC5* which is amplified and overexpressed in *TP53* mutated squamous cell lung carcinomas as well as in the two gene expression subtypes with highest percentage of *TP53* mutations; Classical and Primitive.

The gene *BIRC5* (Baculoviral IAP Repeat Containing 5) is a member of the inhibitor of apoptosis (IAP) gene family and encodes a protein called survivin. The transcript is highly expressed during foetal development and in malignant tumours, but has a low expression in adult normal tissues. The gene is included in numerous prognostic and predictive cancer signatures such as the Oncotype DX used for node-negative breast cancer,[20] the proliferation score for breast cancer[21] and found prognostic in stage III NSCLC.[22] In NSCLC cells, *BIRC5* has been identified as a target of miR-195 and as a regulator of apoptosis, senescence, migration and invasion.[23,24] Analyses of gene expression pathways has identified *BIRC5* as a hub gene and a promising therapeutic target in lung adenocarcinomas.[25]

We do not find any indication that *BIRC5* expression is prognostic in itself, but it remains worth exploring as a predictive marker and as a target of therapy.

Downregulation of *BIRC5* by various micro-RNAs have increased chemo-sensitivity in cancer cell lines.[26,27] Similarly, one study found that inhibition of NF-κB reduced expression of *BIRC5* in lung cancer stem cells and hence induced apoptosis.[28]

BIRC5-targeting therapy includes various inhibitors and vaccines. Despite promising preclinical studies, none of the drugs have found their way into the clinic so far. Phase I-II studies have been performed, but have lacked sufficient evidence of effect, and some have even proven to be too toxic.[29,30] Vaccines targeting survivin (BIRC5 protein) have been tested in patients with glioma (peptide based vaccine)[31] and in patients with prostate cancer (dendritic cell based vaccine),[32] both with good tolerance and effect.

One possible approach in the future studies of *BIRC5*-targeting therapies could be to improve patient selection to include patients with tumours harbouring *TP53* mutations or displaying the classical or primitive gene expression subtypes only, or to stratify the survival analyses on these features.

Targeting BIRC5 to increase chemosensitivity has been tried with success on cell lines from various cancer types chemotherapy.[33,34] We found a trend toward a prognostic value in the subpopulations of patients who had received adjuvant or neoadjuvant platinum-based therapy and had a *TP53* mutated tumour ($P = .08$) that could indicate a primary resistance to platinum-based chemotherapy. To our surprise, higher expression of the gene conferred better prognosis in this population. The sample size was small (n = 19) and the results needs to be confirmed. As we lack information about adjuvant chemotherapy in the TCGA cohort, we could not validate this result. If high gene expression is actually associated with favourable prognosis after adjuvant chemotherapy, BIRC5-inhibition to increase chemo-sensitivity does not seem like a good idea. One study of acute myeloid leukaemia cells found that the combination of an inhibitor of BIRC5 (YM155) and chemotherapeutics were either synergistic or antagonistic depending upon the type of chemotherapy used and the type of cancer cell studied.[35]

Other genes present themselves as candidate genes combining overexpression and amplification in *TP53* mutated lung squamous cell carcinomas (Table 2). After *BIRC5*, Thymidine Kinase 1 (*TK1*) has the highest fold change between wild-type and mutated samples. The gene is also a candidate for the primitive subtype (File S5). This gene resides in the same cytoband as *BIRC* (17q25) and is involved in DNA repair. It is a known biomarker of cancer which can be measured in the serum in early stages of a malignant disease[36] and it has a prognostic value.[37] There is, however, no treatment strategy available targeting this molecule. Other genes among those with highest overexpression also reside on 17q (*TOP2A* and *JUP*). These genomic regions are on the same chromosome as *TP53* (17p13.1), but on a different chromosome arm. Since there is not only amplification, but also overexpression, we expect the genes to have an individual role, not only as passengers due to chromosomal localisation close to a driver.

Expression of BIRC5 was not associated with survival in our cohort. After correction for multiple testing, none of our putative target genes were significant (Table 4), and their prognostic role is therefore questionable. The therapeutic value is, however, independent of the prognostic value.

Several of the putative genes identified in our study (Table 2) have previously been tested as targets of therapy. Anticancer therapies targeting TOP2A have been developed (Delgado), but have not been widely used due to increased risk of developing leukaemia (Chen). Although there are no drugs targeting CBX2 directly, there are inhibitors of histone deacetylase that indirectly modulate the targets of CBX2.[38] Increased effect of several anticancer drugs is seen after inhibition of BUB1.[39] For further exploration of these drugs, identification of a relevant subtype for testing is crucial and our study may help in that regard.

Driver genes have previously been identified by using Pearson correlation between copy number alteration and mRNA expression value across different TCGA datasets.[40] They used a study population of different cancer types, identifying genes that were amplified in two or more datasets, narrowing the list down to cancer-associated genes with druggable properties. Searching for targets across datasets with several cancer types is reasonable, since similar genetic alterations may occur in different cancer diseases leading to a target-driven treatment rather than diagnosis-driven treatment. We have chosen a different approach, subgrouping the patients in order to identify less heterogeneous groups that may allow for identification of targets that may be missed due to noise in bigger populations.

Pathway analyses of candidate genes for each gene expression subtype identified an overrepresentation of pathways involved with the immune system for the samples assigned to the Secretory pathway. This could be linked to response of immunotherapy. Unfortunately, there is not enough information to analyse clinical effect of check-point inhibition in our data, but this should be elucidated in future studies.

5 | CONCLUSION

Personalised therapy requires identification of biologically relevant subgroups for discovery of new therapies. The gene *TP53* is frequently mutated in various cancer diseases, including squamous cell lung carcinomas (68% in our cohort). As an important player in DNA repair and DNA maintenance, deficiency in *TP53* affects the biology of the cancer cells profoundly. Indeed, we do observe a high level of genomic aberrations in the *TP53* mutated cancers. Since therapies targeting *TP53* have not found their way into the clinic, alternative strategies for developing new therapies in *TP53* mutated tumours are needed. Our study presents a target gene search combining copy number alteration and gene expression to identify putative genes for therapeutic and predictive approaches. *BIRC5* is confirmed as an appealing target of therapy in *TP53* mutated squamous cell carcinomas and in the gene expression subtypes *classical* and *primitive*. Future clinical studies should consider stratifying on these parameters or preselecting patients to increase the possibility of identifying patients that will benefit from *BIRC5*-directed treatment strategies. Other putative target genes are identified and our study may help decide which groups of patients should be selected for further testing of treatments targeting these genes.

ACKNOWLEDGEMENTS

This project was supported by the Norwegian Cancer Society (88503-2013 to Å. H.) and South-Eastern Norway Regional Health Authority (2010046 to Å. H.). We like to thank all participating patients. We would also like to thank Ingjerd Solvoll for collecting tissue and data, Åsa Øjlert for the estimation of the previously published squamous lung cancer subtypes and Arne Pladsen and Ole Christian Lingjærde for discussion of copy number analyses and statistics.

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest.

DATA ACCESSIBILITY

Gene expression data are entered into Array Express for public release, accession code: E-MTAB-8615. Other data will be made available from the corresponding author upon reasonable request.

ETHICS STATEMENT

All patients signed informed consent. The study was approved by the Regional Ethical Committee (ref: S-06402b).

ORCID

Vilde D. Haakensen https://orcid.org/0000-0003-0864-3628
Ann Rita Halvorsen https://orcid.org/0000-0003-1223-7948

REFERENCES

1. Paz-Ares L, Luft A, Vicente D, et al. Pembrolizumab plus chemotherapy for squamous non-small-cell lung cancer. *N Engl J Med*. 2018;379:2040-2051.
2. Goffin J, Lacchetti C, Ellis PM, Ung YC, Evans WK. First-line systemic chemotherapy in the treatment of advanced non-small cell lung cancer: a systematic review. *J Thorac Oncol*. 2010;5:260-274.
3. Schiller JH, Harrington D, Belani CP, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med*. 2002;346:92-98.
4. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med*. 2016;375:1823-1833.
5. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Updated analysis of KEYNOTE-024: Pembrolizumab versus platinum-based chemotherapy for advanced non-small-cell lung cancer with PD-L1 tumor proportion score of 50% or greater. *J Clin Oncol*. 2019;37:537-546.
6. Alborelli I, Leonards K, Rothschild SI, et al. Tumor mutational burden assessed by targeted NGS predicts clinical benefit from immune checkpoint inhibitors in non-small cell lung cancer. *J Pathol*. 2020;250:19-29.
7. Olivier M, Hollstein M, Hainaut P. Origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol*. 2010;2:1-17.
8. Levine AJ. p53, the cellular gatekeeper for growth and division. *Cell*. 1997;88:323-331.
9. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*. 2001;98:10869-10874.
10. Perou CM, Parker JS, Prat A, Ellis MJ, Bernard PS. Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol*. 2010;11:718-719. author reply 720-1.
11. Wilkerson MD, Yin X, Hoadley KA, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to Normal cell types. *Clin Cancer Res*. 2010;16:4864-4876.
12. Chen Y, Liu Q, Chen Z, et al. PD-L1 expression and tumor mutational burden status for prediction of response to chemotherapy and targeted therapy in non-small cell lung cancer. *J Exp Clin Cancer Res*. 2019;38:1-14.
13. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17:1665-1674.
14. Ojlert AK, Halvorsen AR, Nebdal D, et al. The immune microenvironment in non-small cell lung cancer is predictive of prognosis after surgery. *Mol Oncol*. 2019;13:1166-1179.
15. Halvorsen AR, Silwal-Pandit L, Meza-Zepeda LA, et al. TP53 mutation spectrum in smokers and never smoking lung cancer patients. *Front Genet*. 2016;7:1-10.
16. van Loo P, Nordgard SH, Lingjaerde OC, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA*. 2010;107:16910-16915.
17. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA*. 2001;98:5116-5121.
18. Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics*. 2017;18:142.
19. Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46:D649-D655.
20. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817-2826.
21. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160-7.
22. Cho HJ, Kim HR, Park YS, Kim YH, Kim DK, Park S. Prognostic value of survivin expression in stage III non-small cell lung cancer patients treated with platinum-based therapy. *Surg Oncol*. 2015;24:329-334.
23. Yu X, Zhang Y, Wu B, Kurie JM, Pertsemlidis A. The miR-195 axis regulates chemoresistance through TUBB and lung cancer progression through BIRC5. *Mol Ther Oncolyt*. 2019;14:288-298.
24. Altieri DC. Survivin, cancer networks and pathway-directed drug discovery. *Nat Rev Cancer*. 2008;8:551-560.
25. Ni K, Sun G. The identification of key biomarkers in patients with lung adenocarcinoma based on bioinformatics. *Math Biosci Eng*. 2019;16:7671-7687.
26. Xu X, Chen X, Xu M, et al. miR-375-3p suppresses tumorigenesis and partially reverses chemoresistance by targeting YAP1 and SP1 in colorectal cancer cells. *Aging (Albany NY)*. 2019;11:7357-7385.
27. Abdi J, Rastgoo N, Chen Y, Chen GA, Chang H. Ectopic expression of BIRC5-targeting miR- 101-3p overcomes bone marrow stroma- mediated drug resistance in multiple myeloma cells. *BMC Cancer*. 2019;19:975.
28. Zakaria N, Yusoff NM, Zakaria Z, Widera D, Yahana BH. Inhibition of NF-κB signaling reduces the stemness characteristics of lung cancer stem cells. *Front Oncol*. 2018;8:1-12.
29. Li F, Aljahdali I, Ling X. Cancer therapeutics using survivin BIRC5 as a target: what can we do after over two decades of study ? *J Exp Clin Cancer Res*. 2019;38:368.
30. Wheatley SP, Altieri DC. Survivin at a glance. *J Cell Sci*. 2019;132:jcs223826.
31. Fenstermaker RA, Ciesielski MJ, Qiu J, et al. Clinical study of a survivin long peptide vaccine (SurVaxM) in patients with recurrent malignant glioma. *Cancer Immunol Immunother*. 2016;65:1339-1352.
32. Xi H, Wang G, Fu B, Liu W, Li Y. Survivin and PSMA loaded dendritic cell vaccine for the treatment of prostate cancer. *Biol Pharm Bull*. 2015;38:827-835.
33. Yu M, Xu B, Yang H, et al. MicroRNA-218 regulates the chemo-sensitivity of cervical cancer cells through targeting survivin. *Cancer Manag Res*. 2019;11:6511-6519.
34. Ferrario A, Luna M, Rucker N, et al. Targeting survivin enhances chemosensitivity in retinoblastoma cells and orthotopic tumors. *PLoS One*. 2016;11:1-16.
35. Huang J, Lyu H, Wang J, Liu B. Influence of survivin-targeted therapy on chemosensitivity in the treatment of acute myeloid leukemia. *Cancer Lett*. 2015;366:160-172.
36. Alegre MM, Robison RA, O'Neill KL. Thymidine kinase 1: a universal marker for cancer. *Cancer Clin Oncol*. 2013;2:159-167.
37. Chen F, Tang L, Xia T, et al. Serum thymidine kinase 1 levels predict cancer-free survival following neoadjuvant, surgical and adjuvant treatment of patients with locally advanced breast cancer. *Mol Clin Oncol*. 2013;1:894-902.
38. di Costanzo A, del Gaudio N, Conte L, et al. The HDAC inhibitor SAHA regulates CBX2 stability via a SUMO-triggered ubiquitin-mediated pathway in leukemia. *Oncogene*. 2018;37:2559-2572.
39. Siemeister G, Mengel A, Fernandez-Montalv AE, et al. Inhibition of BUB1 kinase by bay 1816032 sensitizes tumor cells toward taxanes, ATR, and PARP inhibitors in vitro and in vivo. *Clin Cancer Res*. 2019;25:1404-1414.

40. Chen Y, McGee J, Chen X, et al. Identification of druggable cancer driver genes amplified across TCGA datasets. *PLoS One*. 2014;9: e98293.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Haakensen VD, Khadse A, Sandhu V, et al. Molecular characterisation of *TP53* mutated squamous cell carcinomas of the lung to identify putative targets for therapy. *Int. J. Cancer*. 2020;147:2957–2966. https://doi.org/10.1002/ijc.33121