Award Number:
W81XWH-10-1-0634

TITLE: Blood-Based Biomarkers for Lung Cancer Early Detection and Evaluation of CT-Based Lesions

PRINCIPAL INVESTIGATOR:
Dr. Stephen Lam, M.D.

CONTRACTING ORGANIZATION: British Columbia Cancer Agency
Vancouver, BC Canada V5Z1L3

REPORT DATE:
December 2013

TYPE OF REPORT:Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
December 2013	Final	25 Sep 2010 - 24 Sep 2013

4. TITLE AND SUBTITLE
Blood-Based Biomarkers for Lung Cancer Early Detection and Evaluation of CT-Based Lesions

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-0634

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Stephen Lam, M.D., John Yee, M.D.,
Wan L. Lam, Ph.D., Don Wilson, M.D.
Calum MacAulay, Ph.D.,

email: slam2@bccancer.bc.ca

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

British Columbia Cancer Agency

Vancouver, BC, Canada V5Z 1L3

8.PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10.SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project has two major aims regarding blood based biomarkers: (1) develop and test biomarkers capable of detecting lung cancer up to 24 months prior to clinical diagnosis and (2) identify biomarkers that can discriminate benign from malignant lung nodules 5 to 30 mm in size identified by thoracic CT scans. The tasks at the BCCA are (1) integrate genomic profiles (mutation, miRNA, methylation and gene expression) and published data to identify over-expressed genes that may be potential protein targets and (2) select the best over-expressed miRNA for assessment in pre-validation studies to test for clinical applications.
The BCCA team has substantially contributed to successful accomplishment of the two overall aims of the consortium project headed by Dr. Sam Hanash in addition to validation of Pro-surfactant protein B using a unique, well-characterized (in terms of age, sex, detailed smoking history, family history, lung function), pathologically validated lung cancer screening dataset from 2,485 high risk current and former smokers with 138 lung cancers in the Pan-Canadian Early Detection of Lung Cancer Study that has 3.3 to 5.5 years of follow-up. In addition, we have discovered a unique set of miRNAs for future biomarker studies.

15. SUBJECT TERMS
Lung Cancer, Early Detection, MicroRNA, Gene expression, Genomics, Blood test, Biomarkers

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER**	**19a. NAME OF RESPONSIBLE PERSON:** USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
U	U	U	UU	38	

Table of Contents

INTRODUCTION:

The main objective of this multi-investigator, multi-site project is to evaluate, through blinded validation studies, candidate markers from genomic (mutation analysis, DNA methylation and microRNAs), proteomic (circulating proteins and auto-antibodies) and metabolomic (altered glycans, metabolites and volatile organic compounds) studies that show promise for yielding blood based tests for lung cancer. Two specific blood biomarker application goals are addressed: (1) Discrimination of benign and malignant lung nodules between 5 mm to 30 mm in size identified by thoracic CT scans and (2) Detection of non-small cell lung cancer in high risk individuals up to 24 months prior to clinical diagnosis.

Led by Dr. Wan Lam and Dr. Stephen Lam, the specific tasks for the project at the BCCA are: (a) generate microRNA (miRNA), methylation and gene expression profiles of tumor and matched non-malignant tissues to identify differentially expressed miRNAs and differentially methylated genes for further testing and validation in archival and prospectively collected blood samples; (b) Collect blood samples, clinical data and final diagnosis from 100 individuals with lung nodules ≤3 cm being evaluated with PET/CT. Within the 100 subjects evaluated for lung cancer with PET/CT imaging prior to consideration for surgical resection, paired lung tumor and adjacent non-tumor lung tissue will be collected from a subset of 40 patients.

Over the course of the entire research period, the BCCA has collected and performed multi-dimensional profiling (miRNA sequencing, DNA methylation and RNA seq) on sixty NSCLC cases. In addition to these samples, miRNA array profiles were obtained for a cohort of PET positive and negative tumors (30 in total) and diagnostic blood samples from over 100 subjects (either with cancer or benign lung disease) were collected and RNA extracted. Using these discovery and validation cohorts we've identified a number of genes and miRNAs recurrently altered in tumors and capable of discriminating between cancer and normal tissue or between metastatic and non metastatic lung cancer. Promising candidate biomarkers identified will serve as the basis for future validation studies as blood based predictive, diagnostic and prognostic biomarkers of lung cancer. In addition, we obtained plasma samples from 2,485 high risk current and former smokers with 138 lung cancers in the Pan-Canadian Early Detection of Lung Cancer Study that has 3.3 to 5.5 years of follow-up and validated one of the proteomic biomarkers identified by Dr. Hanash, namely, Pro-surfactant protein B (pro-SFTPB) as an important independent predictor of lung cancer in addition to existing lung cancer risk factors such as age, sex, smoking history, family history of lung cancer and pulmonary function. Pro-SFTPB levels were also found to discriminate benign and malignant lung nodules with the highest levels found in more aggressive interval lung cancers.

KEY WORDS: NSCLC, miRNAs, genomics, blood biomarker, early detection, metastasis, PET

OVERALL PROJECT SUMMARY:

Work accomplished in Year 1

Due to a 6 month delay in IRB approval (*Year 1 task 1*), many tasks from year 1 were delayed. Nevertheless, in year 1 we completed collection, micro-dissection and DNA and RNA extraction of the 30 locally invasive and 30 metastatic cases and 30 Stage 1 tumors with PET data for profiling (*Year 1 task 2*). At the time of the progress report submission, all profiling efforts were in progress (*Year 1 task 5*). Collection of the remaining sample sets; 40 paired lung cancer and blood specimens and 70 blood samples from patients with solitary pulmonary nodules were well underway (*Year 1 tasks 3 and 4*) and all other remaining tasks were in progress.

Work accomplished in Year 2

In Year 2, all outstanding year 1 tasks (genomic profiling and samples collection) were completed or very near completion and both year 2 tasks were in progress. Data integration of miRNA, DNA methylation, gene expression and mutation data identified promising candidate oncogenes and tumor suppressor genes, including EYA4, SIRPA, YEATS4, and ELF3 (*Year 2 task 1*). Manuscripts detailing the biological roles of these genes have been published or are in preparation and are appended. Analysis of miRNA sequencing data identified a panel of 18 miRNAs altered in >90% of tumor samples relative to matched non-malignant tissue. These miRNAs were validated in external cohorts and in pre validation samples to confirm their expression in blood and determine their concordance between tissue and blood samples (*Year 1 task 6 & Year 2 task 2*)

In addition, we performed a validation study on one of Dr. Hanash's blood biomarkers - Pro-surfactant protein B (pro-SFTPB) using blood samples from the Pan-Canadian Early Lung Cancer Detection Study. When adjusted for age, sex, body mass index, personal history of cancer, family history of lung cancer, forced expiratory volume in 1 second percent predicted, average number of cigarettes smoked per day, and smoking duration, plasma level of pro-SFTPB was found to be an important independent predictor of lung cancer (*J Clin Oncol* 31(36):4536-43).

Additional details about the specific tasks can be found in the Annual progress reports for Year 1 and 2.

Statement of Work for Year 3 (BC Cancer Agency component)

The tasks and timeline for the Vancouver site for Year Three are as follows:

1-Month 25-33. For markers that show promise in distinguishing between cancer vs. non cancer CT lesions, pre-validate whether the markers can distinguish between T1NOMO versus T1N+/M1 lung cancer.
2- Month 25-33.Test for discrimination between PET positive versus PET negative Stage IA lung cancers as well as PET positive benign lung nodules detected by screening spiral CT.
3- Month 25-36. Test validation specimens for clinical applications 1 and 2 to determine if the markers meet the statistical criteria compared to other biomarkers in the pre-validation study.

Work accomplished:

Overdue Work from Year two:

Due to a delay in IRB approval, most tasks were delayed by approximately 6 months. All tasks from year one and two have been completed with the exception of 1-7 and 2-2, which are validation tasks. Due to the continued identification of potential candidates these two tasks remain in progress as we attempt to identify the most robust candidates with potential clinical application. Year three tasks are all underway and the work accomplished on outstanding tasks as well as those for the past year are summarized below.

Year 1 Tasks:

Task 6 - Assay over-expressed miRNAs in blood from patients of the 100 tumor specimens and determine concordance with tumor tissue.
Status: Completed
Work accomplished: Analysis of miRNA sequencing data identified 13 miRNAs deregulated in over 85% of tumors relative to matched non-malignant tissue, 8 of which were over-expressed. miRNAs that also showed a high frequency of alteration (>80%) in the TCGA cohort (n=4) were assessed in blood samples. Concordance between miRNA expression in plasma and tumor tissue varied between miRNAs, but all miRNAs showed a strong correlation (r>0.55) between blood and tissue levels.

Task 7 - Assay miRNA in blood from 50 patients with benign CT detected lung nodules and compare the levels with patients with lung cancer.
Status: On going
Work accomplished: Due to small amount of RNA that can be extracted from blood samples (~200ng) and the nature of miRNA qPCR, which requires separate reverse transcriptase reactions using 10ng or RNA for each miRNA, we are limited in the number of miRNA that we can test in these blood samples. To ensure that we can test the candidates with the best promise to differentiate between benign and malignant lung nodules, testing in this cohort will be done as one of the final validation steps. The four miRNAs identified in task 6 along with a list of miRNA that differentiate metastatic from non-metastatic tumors and validate in blood samples (see Year 3 task 1) will be tested in these samples.

Task 8 - Select hypermethylated genes for further investigation in blood by Dr. Gazdar.
Status: Completed
Work accomplished: Our short list of differentially methylated genes was integrated with gene expression data and promising candidate genes with concordant methylation and expression changes were analyzed/validated by Dr. Gazdar at UTSW in his sub-contract.

Year 2 Tasks:

Task 1- Integration of miRNA, mutation status, methylation, gene expression and published data to deduce over-expressed genes to identify potential protein targets for investigation in blood by Dr. Hanash
Status: Completed
Work accomplished: Integration of the multiple genomic dimensions has identified a number of candidate oncogenes with DNA level alterations and corresponding changes in gene expression. These include *YEATS4, MARK2* and *ELF3* to name a few. These targets and others have been sent to Dr. Hanash for investigation of protein levels within blood to determine whether any candidates have potential application as biomarkers. Validation of one of Dr. Hanash's protein biomarkers (pro Surfactant B) was made possible through the use of blood samples from Dr. Stephen Lam's Pan-Canadian Early Cancer Detection Study and a manuscript detailing these findings has been published in the Journal of Clinical Oncology.

Task 2- Select the best over-expressed miRNA for assessment using aliquots of specimens assigned for pre-validation studies for clinical applications 1 and 2.
Status: On going
Work accomplished: Following identification of candidates and successful validation in external cohorts such as the TCGA, promising markers are assessed in pre-validation specimens to determine the most robust and clinically relevant candidates for further testing in precious blood specimens. miRNAs altered in >85% of tumors have been validated in these specimens and miRNA identified as differentially expressed between metastatic and non-metastatic (for more details see Year 3, task 1) tumors are currently being assessed.

Year 3 Tasks:

Task 1-For markers that show promise in distinguishing between cancer vs. non cancer CT lesions, pre-validate whether the markers distinguish between TINOMO versus TIN+/MI lung cancer.
Status: On going
Work Accomplished: Using miRNA seq profiles from the 30 locally invasive and 30 metastatic tumors we identified miRNAs that were 1) differentially expressed between tumors and non-tumor tissues and 2) differentially altered between the locally invasive and metastatic tumors. This analysis yielded 12 miRNA, of which 5 were specific to metastatic tumors and 7 to the locally invasive cohort. Target

analysis of these miRNAs revealed a number of common targets involved in the TGFB signaling pathway, which is known to play a role in invasion and metastasis (Figure 3). The ability of this panel of miRNAs to accurately discriminate between T1N0M0 and T1N+/M+ cases is currently being assessed in pre-validation sets and successful candidates will be further examined in blood samples (Year 1, task 7) to determine their potential as blood based biomarkers.

Task 2-Test for discrimination between PET positive versus PET negative Stage IA lung cancers as well as PET positive benign lung nodules detected by screening spiral CT.
Status: Complete
Work Accomplished: miRNA profiles were generated on an Agilent microarray for 30 formalin fixed paraffin embedded Stage I/II NSCLC cases (14 PET negative and 16 PET positive). To determine whether any of our candidate markers identified in earlier tasks (year 1, task 6 and year 3, task 1) are capable of discriminating between PET positive and PET negative cases, we assessed the levels of these 16 miRNA in this cohort. Of the 16 candidates, all but hsa-mir-4772 (associated with metastatic tumors) was measureable on the array. Four miRNA, let-7g, mir-29a, -34b and -141 were associated with PET status (Mann Whitney U, $p<0.05$); however none passed multiple testing correction (Table 3).

Task 3- Test validation specimens for clinical applications 1 and 2 if the markers meet the statistical criteria compared to other biomarkers in the pre-validation study.
Status: Awaiting candidates for analysis
Work Accomplished: As testing of candidates in the final validation cohorts has yet to be completed, statistical analysis on successful candidates has not yet been performed.

KEY RESEARCH ACCOMPLISHMENTS:

Year 1

- Nothing to report. In year 1 we focused on collecting samples, extracting genetic material and completing multi-omics profiling.

Year 2

- Identified 18 microRNAs (12 over-expressed and 6 under-expressed) that were altered in at least 90% of tumors and confirmed their ability to accurately discriminate between tumor and non-malignant tissue by Principal component analysis (See Table 1 and Figure 1).

Table 1. Alteration frequencies of the most frequently deregulated miRNAs in NSCLC (n=18)

		AC		SqCC		All Samples	
miRNA	**Alteration**	**Freq OE**	**Freq UE**	**Freq OE**	**Freq UE**	**Freq OE**	**Freq UE**
hsa-mir-210	OE	100%	0%	95%	0%	99%	0%
hsa-mir-96	OE	98%	0%	100%	0%	99%	0%
hsa-mir-130b	OE	95%	0%	100%	0%	97%	0%
hsa-mir-183	OE	94%	0%	100%	0%	95%	0%
hsa-mir-345	OE	94%	2%	95%	0%	94%	1%
hsa-mir-877	OE	94%	2%	95%	0%	94%	1%
hsa-mir-331	OE	94%	0%	91%	5%	93%	1%
hsa-mir-182	OE	94%	0%	86%	0%	92%	0%

hsa-mir-708	OE	89%	0%	100%	0%	92%	0%
hsa-mir-141	OE	92%	0%	86%	0%	91%	0%
hsa-mir-193b	OE	89%	3%	95%	0%	91%	2%
hsa-mir-301b	OE	89%	0%	95%	0%	91%	0%
hsa-mir-144	UE	0%	98%	0%	100%	0%	99%
hsa-mir-30a	UE	0%	97%	0%	100%	0%	98%
hsa-mir-451a	UE	0%	97%	0%	91%	0%	95%
hsa-mir-143	UE	0%	97%	0%	82%	0%	93%
hsa-mir-486	UE	0%	94%	0%	86%	0%	92%
hsa-mir-101	UE	0%	89%	0%	95%	0%	91%

Figure 1.

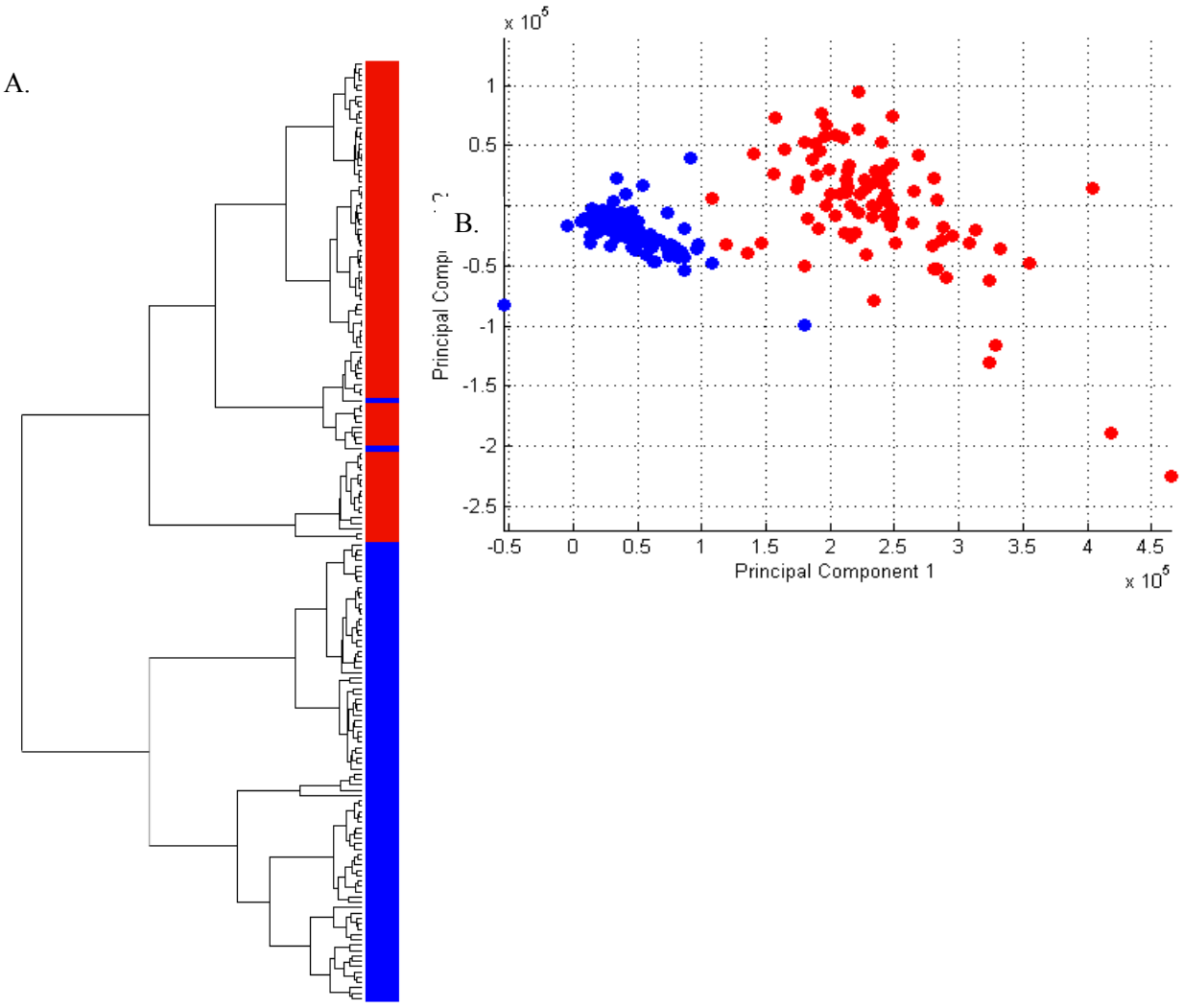


Figure 1. miRNA expression profiles accurately segregate tumor and non-malignant tissue
Clustering of 176 miRNA expression profiles revealed two distinct clusters associated with malignancy, one comprised of tumor samples (blue) and the other of all non-malignant tissues (red) and two tumor samples (A). Principal component analysis using expression of the 18 miRNA altered in >90% of all

cases accurately separated tumor and non-malignant tissue (B). Blue dots represent non-malignant tissue while red dots represent tumor samples.

- Validation of the 18 miRNAs in the TCGA cohort identified four miRNAs (miR-130b, -141, and -183 -210) frequently overexpressed in tumors. All four miRNA were detectable in blood and assessment of these candidates in pre-validation samples showed a strong correlation ($r>0.55$) between blood and tissue levels. Area under the curve analysis of these four miRNAs revealed all miRNAs were able to accurately discriminate between tumor and non-malignant tissue ($r>0.98$, $p<0.0001$) (Figure 2).

Figure 2.

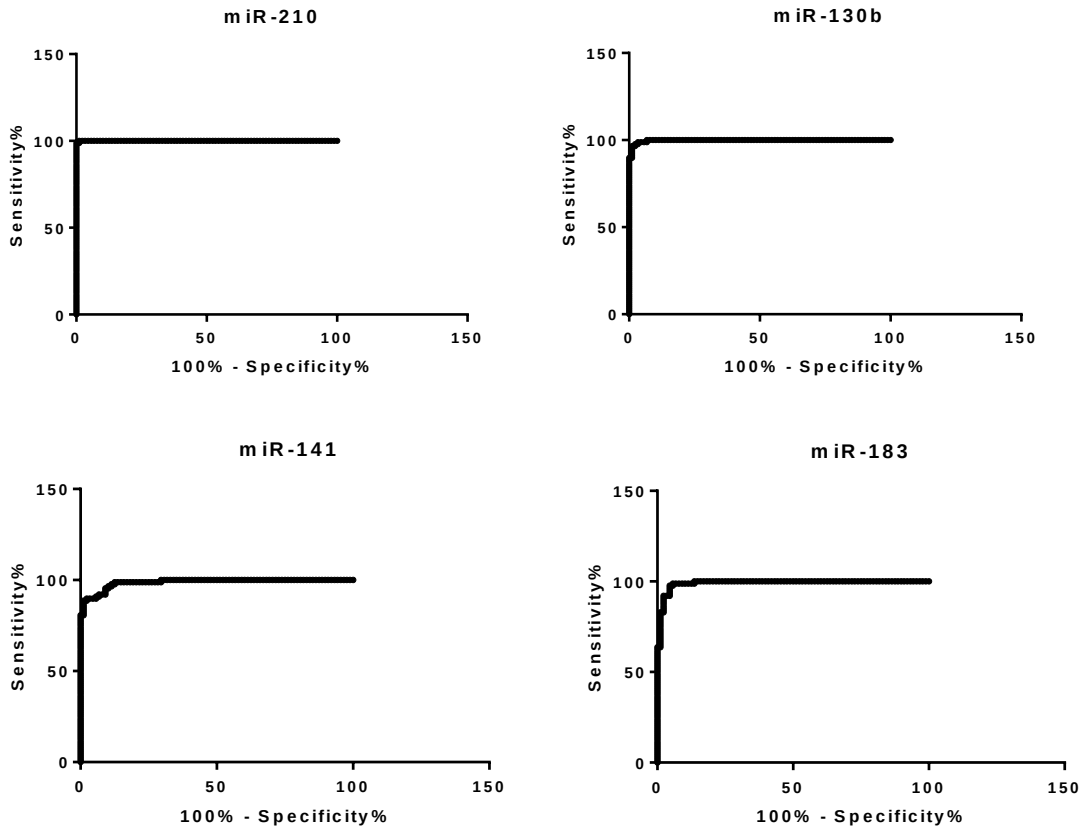


Figure 2. Receiver operator characteristics of validated miRNAs. Area under the curve analysis of the four miRNA frequently altered in tumors relative to non-malignant tissue and validated in TCGA cohort, revealed that all miRNAs are able to accurately discriminate between tumor and non-malignant tissue with high specificity and sensitivity.

- Integration of multiple genomic levels (mutation, methylation and gene expression) identified candidate oncogenes (*YEATS4, ELF3*) and tumor suppressors (*SIRPA, EYA4*) that have been validated in our lab. Manuscripts detailing the biological roles of these genes have been published or are in preparation (see bibliography), and the published manuscripts for YEATS4 and EYA4 are appended.
- Performed a validation study on one of Dr. Hanash's blood biomarkers - Pro-surfactant protein B (pro-SFTPB) using 2,485 blood samples from the Pan-Canadian Early Lung Cancer Detection Study. Plasma level of pro-SFTPB was found to be an important independent predictor of lung cancer, and a

valuable addition to existing lung cancer risk prediction models (*J Clin Oncol* 31(36):4536-43, appended).

Year 3

- Identified 12 miRNAs that were differentially altered between the locally invasive and metastatic tumors (T1NOMO versus T1N+/MI); 5 specific to metastatic tumors and 7 to the locally invasive cohort and validated in an external cohort (Table 2). Target analysis revealed a number of common targets involved in the TGFB signaling pathway (Figure 3), which is known to play a role in invasion and metastasis.

Table 2. List of miRNAs differentially altered between TINOMO and TIN+M1

microRNA	Status
hsa-mir-127	OE in metastatic
hsa-mir-145	UE in metastatic
hsa-mir-34b	UE in metastatic
hsa-mir-508	UE in metastatic
hsa-mir-4772	UE in metastatic
hsa-mir-146a	OE locally invasive
hsa-mir-155	OE locally invasive
hsa-mir-664	OE locally invasive
hsa-let-7g	OE locally invasive
hsa-mir-125a	OE locally invasive
hsa-mir-29a	OE locally invasive
hsa-mir-362	OE locally invasive

Figure 3.

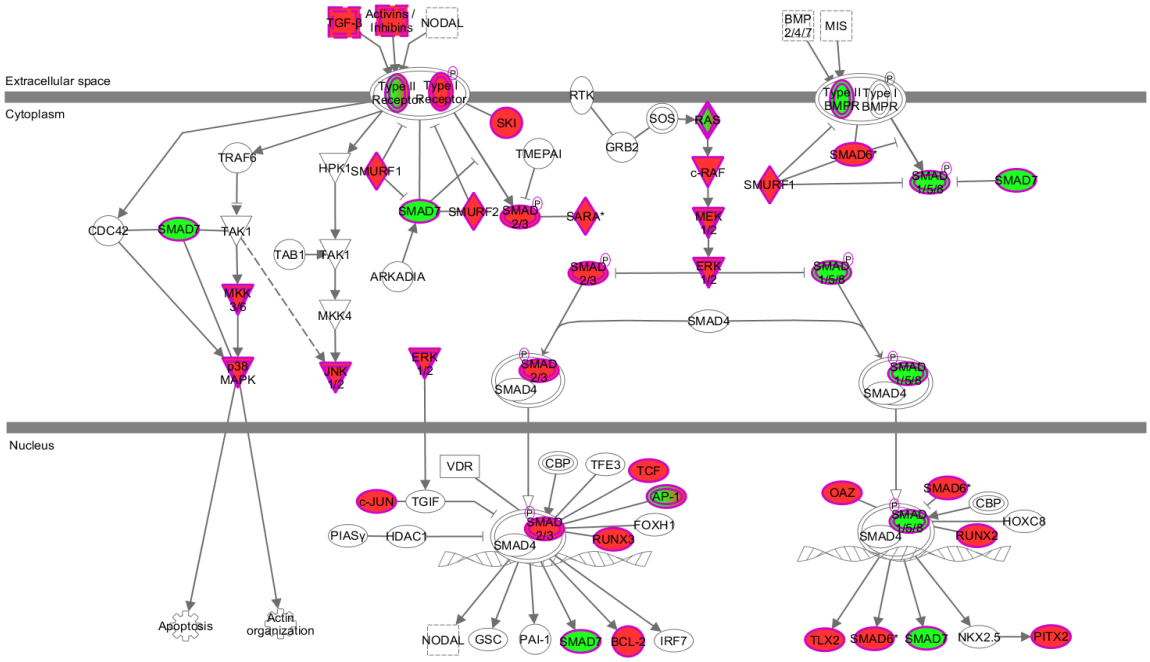
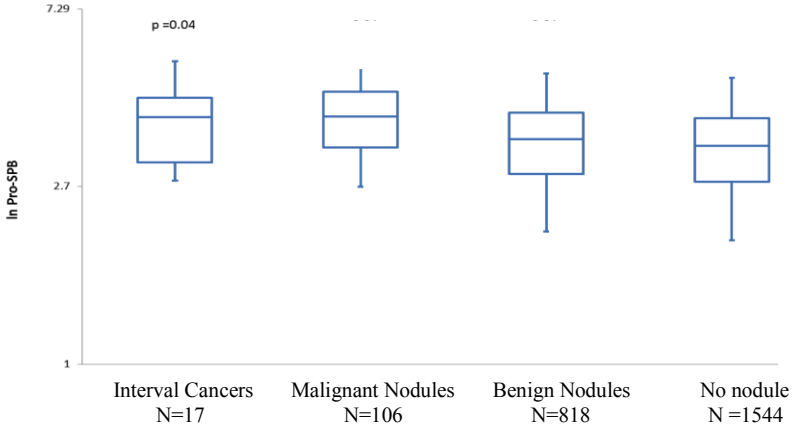


Figure 3. miRNAs associated with metastasis target the TGFB pathway. Pathway analysis of the 12 miRNAs differentially altered between metastatic and locally invasive adenocarcinomas revealed enrichment of mRNA targets in the TGFB pathway, a pathway known to be involved in epithelial to mesenchymal transition and metastasis.

- Identified 4/16 of our miRNAs; let-7g, mir-29a, -34b and -141 to be associated with PET status (Mann Whitney U, $p<0.05$). Although these miRNAs did not pass multiple testing correction

- Pro-SFTPB levels are significantly higher in interval cancers and malignant lung nodules versus benign lung nodules and those with no lung nodules.



CONCLUSION

Despite delayed IRB approval which caused a six month setback, we have completed the tasks in the statement of work. Our analysis has identified numerous potential candidates, some of the most promising of which are panels of miRNAs that distinguish tumor tissue from non-malignant tissue and aggressive tumors (metastatic) from less aggressive tumor (non-metastatic). These panels serve as the basis for future studies to determine their potential clinical relevance as blood based predictive, diagnostic or prognostic biomarkers for lung cancer. Through this grant, we have been able to collect unique cohorts of tumor and blood specimens for discovery and validation of biologically based biomarkers. We have identified promising miRNA candidates and will continue to assess the clinical potential of these and other candidate biomarkers over the next few years.

PUBLICATIONS, ABSTRACTS AND PRESENTATIONS:

A total of 22 peer reviewed journal publications and 12 published abstracts will be generated from grant # W81XWH-10-1-0634 . Specifically, 10 manuscripts and 12 abstracts have been published as a result of the work from this grant. An additional 5 manuscripts are currently in preparation or under review. These items are listed below. Moreover, the data published from this grant has enabled further analyses of tumor genomes in additional projects beyond the scope of W81XWH-10-1-0634 (7 publications listed below).

Peer-reviewed Scientific journals (n = 10):

1. Enfield KSS, Stewart GL, Pikor LA, Alvarez CE, Lam S, Lam WL, Chari R (2011). MicroRNA gene dosage alterations and drug response in lung cancer. Journal of Biomedicine and Biotechnology 2011: 474632, 1-15. [PMID: 21541180] *(Year 1)*

2. Gibb EA, Brown CJ, Lam WL (2011) The functional role of long non-coding RNA in human carcinomas. Molecular Cancer 10: 38, 1-17. [PMID: 21489289] *(Year 1)*

3. Gibb EA, Vucic, EA, Enfield KSS, Stewart GL, Lonergan KM, Kennett JY, Becker-Santos DD, MacAulay CE, Lam S, Brown CJ, Lam WL (2011) Human cancer long non-coding RNA transcriptomes. PLoS ONE 6: e25915, 1-10. [PMID: 21991387] *(Year 1)*

4. Hubaux R*, Becker-Santos DD*, Enfield KSS, Lam S, Lam WL, Martinez VD (2011) MicroRNAs as biomarkers for clinical features of lung cancer. Metabolomics 2:108, 1-11. (doi:10.4172/2153-0769.1000108). *(Year 1)*

5. Enfield KSS, Pikor LA, Martinez VD, Lam WL (2012) Mechanistic roles of non-coding RNAs in lung cancer biology and their clinical implications. Genetic Research International 2012: 737416, 1-16. *(Year 2)*

6. Yao Y, Zhihao W, Wu J, Wu H, Wang J, Lam S, Lam WL, Girard L, Minna J, Gazdar AF, Zhou Q (2012) Potential application of non-small cell lung cancer-associated autoantibodies to early cancer diagnosis. Biochemical and Biophysical Research Communications 423: 613–19. [PMID: 22713465] *(Year 2)*

7. Wilson IM*, Vucic EA*, Enfield KSS, Thu KL, Zhang Y-A, Chari R, Lockwood WW, Radulovich N, Starczynowski DT, Banáth JP, Zhang M, Pusic A, Fuller M, Lonergan KM, Rowbotham D, Yee J, English JC, Buys TPH, Selamat SA, Laird-Offinga I, Liu P, Anderson M, You M, Tsao-MS, Brown CJ, Bennewith KL, MacAulay CE, Karsan A, Gazdar AF, Lam S, Lam WL (2014) EYA4 is inactivated biallelically at a high frequency in sporadic lung cancer and is associated with familial lung cancer risk. Oncogene. In press. [PMID: 24096489] *(Year 3)*

8. Pikor LA, Lockwood WW, Thu KL, Vucic EA, Chari R, Gazdar AF, Lam S, Lam WL (2013) Integrative analysis identifies YEATS4, a novel oncogene in NSCLC that regulates the p53 pathway. Cancer Research 73:7301-12. [PMID: 24170126] *(Year 3)*

9. Pikor LA, Thu KL, Vucic EA, Lam WL (2013) The detection and implication of genome instability in cancer. Cancer Metastasis Reviews 32:341-52. [PMID: 23633034]. *(Year 3)*

10. Sin DD, Tammemagi CM, Lam S, Barnett MJ, Duan X, Tam A, Auman H, Feng Z, Goodman GE, Hanash S, Taguchi A (2013). Pro-surfactant protein B as a biomarker for lung cancer prediction. J Clin Oncol 31(36):4536-43. P[MID:24248694] *(Year 3)*

Peer reviewed scientific journal publications using published DOD data (n = 7):

11. Thu KL, Chari R, Lockwood WW, Lam S, Lam WL (2011) miR-101 DNA copy loss is a prominent subtype specific event in lung cancer. Journal of Thoracic Oncology 6:1594-98. [PMID: 21849855]

12. Pikor LA*, Enfield KSS*, Heryet C, Lam WL (2011) DNA extraction from paraffin embedded material for genetic and epigenetic analyses. Journal of Visualized Experiments 49: 2763 [PMID: 21490570]

13. Thu KL*, Pikor LA, Chari R, Wilson IM, MacAulay CE, Tsao-MS, Gazdar AF, Lam S, Lam WL, Lockwood WW (2011) Disruption of KEAP1/CUL3 E3 ubiquitin ligase complex components is a key mechanism for NFkB pathway activation in lung cancer. Journal of Thoracic Oncology 6:1521-1529. [PMID: 21795997]

14. Starczynowski DT, Lockwood WW, Delehouzee S, Chari R, Wegrzyn J, Fuller M, Tsao-MS, Lam S, Gazdar AF, Lam WL, Karsan A (2011) TRAF6 is an amplified oncogene bridging the RAS and nuclear factor-kB cascade in lung cancer. Journal of Clinical Investigation 121: 4095-105. [PMID: 21911935]

15. Lockwood WW, Wilson IM, Coe BP, Chari R, Pikor LA, Thu KL, Yee J, English J, Murray N, Tsao MS, Minna JD, Gazdar AF, MacAulay CE., Lam S, Lam WL (2012) Divergent genomic and epigenomic landscapes of lung cancer subtypes underscore the selection of different oncogenic pathways during tumor development. PLoS ONE 7(5): e37775, 1-18. [PMID: 22629454]

16. Thu KL, Vucic EA, Chari R, Zhang W, Lockwood WW, English JC, Fu R, Wang P, Feng Z, Macaulay CE, Gazdar AF, Lam S, Lam WL (2012) Lung adenocarcinoma of never smokers and smokers harbor differential regions of genetic alteration and exhibit different levels of genomic instability. PLoS ONE 7(3): e33003, 1-10. [PMID: 22412972]

17. Vucic, EA*, Thu KL*, Robison K, Rybaczyk L, Chari R, Alvarez CE, Lam WL (2012) Translating cancer omics to outcomes. Genome Research 22:188-95. [PMID: 22301133].

Abstracts (n=12)

1. Wilson I, Vucic EA, Enfield KSS, Chari R, Starczynowski D, Yee J, English JC, Selamat SA, Laird-Offringa IA, Tsao MS, Brown CJ, Bennewith KL, MacAulay CE, Karson A, Gazdar AF, Lam S, Lam WL (2010). Deleted and hypermethylation of a tumor suppressor gene EYA4 in 6q23-25 is associated with lung cancer risk and poor survival. AACR Cancer Prevention Conference, Philadelphia, PA (10-A-206).

2. Enfield KSS, Stewart GL, Pikor LA, Alvarez CE, Lam S, Chari R, Lam WL (2011) Deregulation of specific miRNAs correlates with systemic therapy response in NSCLC. World Conference on Lung Cancer, Amsterdam, NL.

3. Thu KL, Mosslemi M, Pikor LA, Vucic EA, Zhang W, Selamat SA, Laird-Offringa IA, Gazdar AF, Lam S, MacAulay CA, Lam WL (2012). Development of an integrative genomics algorithm to investigate differences in the molecular mechanisms contributing to lung adenocarcinoma of smokers and never smokers. American Thoracic Society Annual Meeting, Philadelphia, PA. Abstract A28

4. Vucic EA, Pikor LA, Thu KL, Ramnarine VR, Enfield KSS, MacAulay CE, Lam S, Lam WL (2012). Contribution of smoking to miRNA deregulation in lung tumors. Mechanisms and Models of Cancer Conference, Cold Spring Harbour, NY

5. Pikor LA, Lockwood WW, Vucic EA, Chari R, Lam S, Lam WL (2012) Integrative Analysis identifies GAS41 as a novel oncogene in NSCLC, localizing to the 12q15 amplicon. AACR-IASLC Joint conference on molecular origins of lung cancer, San Diego, CA

6. Pikor LA, Thu KL, Vucic EA, Lam S, Lam WL (2012). miRNAs display subtype specific patterns of aberrant expression in lung adenocarcinoma and squamous cell carcinoma. Mechanisms and Models of Cancer Conference, Cold Spring Harbour, NY, USA
7. Enfield KSS, Lam S, Lam WL (2012). MicroRNA deregulation associated with non-small cell lung cancer metastasis. Mechanisms and Models of Cancer Conference, Cold Spring Harbour, NY

8. Enfield KSS, Lam S, Lam WL (2012). The paralogous microRNA clusters, *miR-17-92* and *miR-106-25*, are specifically overexpressed in metastatic non-small cell lung carcinomas. AACR non-coding Conference, Miami, FL, USA

9. Katey SS Enfield, Varune R Ramnarine, David A Rowbotham, Stephen Lam, Wan L Lam (2013). MicroRNA deregulation associated metastasis and recurrence of non-small cell lung cancer. AACR-JCA Joint Conference: Breakthroughs in Basic and Translational Cancer Research, Maui, HI.

10. Vucic EA, Thu KL, Pikor LA, Enfield KSS, Ramnarine VR, Lam S, Lam WL (2013). Patterns of miRNA disruption specific to smoking and cancer in non-malignant and malignant lung tissues. AACR/JCA Joint Conference, Maui, HI. Abstract A063

11. Pikor LA, Thu KL, Vucic EA, Ramnarine VR, Lam S, Lam WL (2013). miRNA sequencing identifies miRNAs that differentiate histology and distinguish tumor from non-malignant tissue in NSCLC. AACR/JCA Joint Conference, Maui, HI. Abstract A062

12. Pikor LA, Lockwood WW, Thu KL, Vucic EA, Chari R, Gazdar AF, MacAulay CE, Lam S, Lam WL (2013). Amplification of YEATS4, a novel oncogene in NSCLC inhibits the p53 pathway and increases cisplatin resistance. World Conference on Lung Cancer. Sydney, Australia. Abstract # 1073

Manuscripts in preparation or under review (n=5):

1. Hubaux R, Thu KL, Vucic EA, Pikor LA, Kung SHY, Martinez VD, Mosslemi M, Becker-Santos DD, Gazdar AF, Lam S, Lam WL Microtubule affinity-regulating kinase 2 contributes to cisplatin sensitivity through modulation of the DNA damage response in non-small cell lung cancer. Oncotarget. Under Review.

2. Vucic EA, Thu KL, Pikor LA, Enfield KS, Yee J, English JC, MacAulay CE, Lam S, Jurisica I, Lam WL Smoking status impacts microRNA mediated prognosis and lung adenocarcinoma biology. Molecular Cancer. Under Review.

3. Taguchi A, Taylor AD, Rodriguez J, Çeliktaş M, Liu H, Ma X, Zhang Q, Wong CH, Chin A, Girard L, Behrens C, Lam WL, Lam S, Minna JD, Wistuba II, Gazdar AF, Hanash SM. A search for novel cancer/testis antigens in lung cancer identifies VCX/Y genes expanding the repertoire of potential immunotherapeutic targets. Cancer Research. Under Review.

4. Pikor LA, Thu KL, Vucic EA, MacAulay CE, Lam S, Lam WL miRNA sequencing identifies miRNAs that differentiate histology and distinguish tumor from non-malignant tissue in NSCLC. In preparation.

5. Enfield KSS, Hubaux RH, MacAulay CE, Lam S, Lam WL MicroRNA deregulation in node negative versus node positive non-small cell lung cancer. In preparation.

Presentations in 2013 (n=3)

Enfield KSS. MicroRNA deregulation in node negative versus node positive non-small cell lung cancer. 15th World Conference on Lung Cancer, International Association for the study of lung cancer, Sydney Australia, (2013). *

Lam WL. Molecular profiling/pathology of CT detected nodules. 15th World Conference on Lung Cancer, International Association for the study of lung cancer, Sydney Australia (2013).

Lam WL. Multi-Dimensional Omics Analysis of Tumor Genomes. Sao Paulo Advanced School of Comparative Oncology (ESPCA), Sao Paulo, Brazil (2013).

INVENTIONS, PATENTS AND LICENSES
Nothing to report

REPORTABLE OUTCOMES:
- Creation of the largest lung cancer miRNA sequencing dataset with patient matched tumor and non-malignant tissue

OTHER ACHIEVEMENTS:
Personnel supported by this award include:

Chad Malloff
Chunmei Zhang
Daiana Becker
Dorothy Hwang
Ewan Gibb
Heather Saprunoff
Kim Lonergan
Jennifer Kennett
Roland Hubaux
Sharon Gee
Victor Martinez

REFERENCES:
Guan P, Yin Z, Li X, Wu W, Zhou B (2012). Meta-analysis of human lung cancer microRNA expression profiling studies comparing cancer tissues with normal tissues. J Exp Clin Cancer Res Jun 6;31:54

Vosa U, Vooder T, Kolde R, Vilo J, Metspalu A, Annilo T (2013). Meta-analysis of microRNA expression in lung cancer. Int J Cancer 132(12):2884-93.

APPENDICES

Manuscripts detailing the integrative genomic analysis and biological characterization of two genes, YEATS4 and EYA4 are attached.

Oncogene (2013), 1–10

www.nature.com/onc

ORIGINAL ARTICLE

EYA4 is inactivated biallelically at a high frequency in sporadic lung cancer and is associated with familial lung cancer risk

IM Wilson[1,12], EA Vucic[1,12], KSS Enfield[1], KL Thu[1], YA Zhang[2], R Chari[1,3], WW Lockwood[1,4], N Radulovich[5], DT Starczynowski[6], JP Banáth[1], M Zhang[1], A Pusic[1], M Fuller[1], KM Lonergan[1], D Rowbotham[1], J Yee[7], JC English[8], TPH Buys[1], SA Selamat[9], IA Laird-Offringa[9], P Liu[10], M Anderson[10], M You[10], MS Tsao[5], CJ Brown[11], KL Bennewith[1], CE MacAulay[1], A Karsan[1], AF Gazdar[2], S Lam[1] and WL Lam[1]

In an effort to identify novel biallelically inactivated tumor suppressor genes (TSGs) in sporadic invasive and preinvasive non-small-cell lung cancer (NSCLC) genomes, we applied a comprehensive integrated multiple 'omics' approach to investigate patient-matched, paired NSCLC tumor and non-malignant parenchymal tissues. By surveying lung tumor genomes for genes concomitantly inactivated within individual tumors by multiple mechanisms, and by the frequency of disruption in tumors across multiple cohorts, we have identified a putative lung cancer TSG, *Eyes Absent 4* (*EYA4*). *EYA4* is frequently and concomitantly deleted, hypermethylated and underexpressed in multiple independent lung tumor data sets, in both major NSCLC subtypes and in the earliest stages of lung cancer. We found that decreased *EYA4* expression is not only associated with poor survival in sporadic lung cancers but also that *EYA4* single-nucleotide polymorphisms are associated with increased familial cancer risk, consistent with *EYA4*s proximity to the previously reported lung cancer susceptibility locus on 6q. Functionally, we found that *EYA4* displays TSG-like properties with a role in modulating apoptosis and DNA repair. Cross-examination of *EYA4* expression across multiple tumor types suggests a cell-type-specific tumorigenic role for *EYA4*, consistent with a tumor suppressor function in cancers of epithelial origin. This work shows a clear role for *EYA4* as a putative TSG in NSCLC.

Oncogene advance online publication, 7 October 2013; doi:10.1038/onc.2013.396

Keywords: *EYA4*; two hit; hypermethylation; tumor suppressor; TSG; non-small-cell lung cancer

INTRODUCTION

Lung cancer is the leading cause of cancer mortality in the world, accounting for 1.5 million deaths each year.[1] Over 80% of lung cancers are non-small-cell lung cancer (NSCLC), of which adenocarcinomas (ACs) and squamous cell carcinomas (SqCCs) are the predominant subtypes.[2] Owing to late-stage diagnosis and paucity of effective treatments, the 5-year survival for lung cancer patients is $<15\%$. There remains an urgent, worldwide need for early detection markers and improved chemoprevention and therapeutic regimens for this disease.

Multiple genetic mechanisms contribute to the evolution of cancer genomes; therefore, integration of data from multiple 'omics' levels for an individual tumor represents a powerful approach for discovering genes selectively altered in tumors. Within individual tumor genomes, it is likely that genes selectively disrupted sustain biallelic or 'two-hit' disruptions, as commonly observed with many tumor suppressor genes (TSGs).[3] Therefore, we hypothesized that genes (i) sustaining frequent, high level and two-hit gene dosage and/or DNA methylation alterations and (ii) undergoing concomitant alterations at the mRNA level would be indicative of genes selectively inactivated in lung tumors and therefore relevant to lung tumor biology.

Applying this rationale to genome-wide copy number, DNA methylation and gene expression profiles from a large panel of NSCLC tumor clinical specimens with patient-matched, non-malignant parenchymal tissues, we discovered a novel putative lung cancer TSG *Eyes Absent 4* (*EYA4*). *EYA4*, a putative oncogene in tumors of neural origin, is an atypical, dual-functioning protein phosphatase that functions in mediating DNA repair, apoptosis and innate immunity in response to DNA damage, damaged cells and viruses. Our findings suggest a dual role for *EYA4* in carcinogenesis, likely dependent on cancer cell type of origin and strongly supportive of a tumor suppressor role in lung cancer. Collectively, our findings illustrate the utility of a multidimensional tumor systems approach to cancer 'omics' research applied to the discovery of novel lung cancer TSGs.

RESULTS

Few genes are inactivated by homozygous deletion in lung ACs

We sought to investigate whether homozygous deletion (HD) is a mechanism of recurrent gene inactivation in a panel of AC tumors and patient-matched, non-malignant lung parenchyma tissues ($n=77$ pairs), using Affymetrix SNP 6.0 arrays (Affymetrix, Santa

[1]Integrative Oncology Genetics Unit, British Columbia Cancer Research Centre, Vancouver, BC, Canada; [2]Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center, Dallas, TX, USA; [3]Department of Genetics, Harvard Medical School, Boston, MA, USA; [4]National Human Genome Research Institute, Cancer Genetics Branch, Bethesda, MD, USA; [5]Ontario Cancer Institute/Princess Margaret Hospital, Toronto, ON, Canada; [6]Division of Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital Research Foundation, Cincinnati, OH, USA; [7]Department of Surgery, Vancouver General Hospital, Vancouver, BC, Canada; [8]Department of Pathology, Vancouver General Hospital, Vancouver, BC, Canada; [9]Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA, USA; [10]Medical College of Wisconsin Cancer Center, Milwaukee, WI, USA and [11]Department of Medical Genetics, University of British Columbia, Life Sciences Centre, Vancouver, BC, Canada. Correspondence: Dr IM Wilson, Integrative Oncology Genetics Unit, British Columbia Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3.
E-mail: iwilson@bccrc.ca
[12]These authors contributed equally to this work.
Received 22 February 2013; revised 30 July 2013; accepted 6 August 2013

Clara, CA, USA). We calculate that a DNA copy number (CN) of 0.4 should represent an HD if tumor cell content is 80% in a given specimen. In our panel of 77 AC specimens, we identified only two genes, *CDKN2A* and *CDKN2B*, that were HD at CN <0.4 in more than two specimens, consistent with previous reports for these genes.[4] To compensate for cytological heterogeneity within tumors, we relaxed our HD detection threshold to CN <1.0, which yielded no further HD genes (Supplementary Table S1). Therefore, we reasoned that biallelic inactivation of TSGs must occur through a combination of other mechanisms such as DNA hypermethylation and single-copy loss.

EYA4 is frequently inactivated by deletion and hypermethylation in lung AC

To identify genes disrupted by deletion and promoter hypermethylation, we obtained genome-wide DNA methylation profiles (Illumina Infinium HumanMethylation27, Illumina, San Diego, CA, USA) for this same panel of 77 AC tumor pairs. We searched for frequent ($>15\%$) and concurrent CN loss and promoter hypermethylation events. We identified 114 genes that were frequently deleted and hypermethylated in the same tumor (Supplementary Table S2), which included both previously reported and novel putative lung TSGs. Integration with expression data revealed 37 genes that were significantly underexpressed, lost and hypermethylated in our cohort (indicated in Supplementary Table S2). Of these, we focused on the putative TSG *EYA4*, based on the high frequency of biallelic disruption (19.5%) and significant underexpression (32.5%) in our cohort (Figures 1a and b), frequency of inactivation by multiple mechanisms in other epithelial cancers and proximity to the lung cancer susceptibility locus at 6q23.[5–9] Overall, 67.5% of our AC panel sustained allelic inactivation of *EYA4* by either CN loss (26%) or by promoter hypermethylation (61%) (e.g., Figure 1a). We calculated the probability of observing a two-hit DNA level and gene expression event in a single tumor pair by multiplying the proportion of any probe we observed undergoing hypermethylation, CN loss and underexpression of alterations.[10] The average proportion of each of these events occurring in our cohort of 77 AC tumor pairs was: hypermethylation 0.0825, CN loss 0.1614 and underexpression 0.1176. Therefore, the probability of observing a two-hit inactivating DNA level alteration and underexpression event for a single gene in a tumor sample from our cohort was 0.0016. Moreover, the probability of randomly observing the frequency for which we detect *EYA4* inactivated by these mechanisms is extremely low (1.433×10^{-22}) (Supplementary Figure S1). These findings suggest that *EYA4* inactivation is strongly selected for in AC. We validated mechanistic control of *EYA4* expression by DNA methylation by observing re-expression of hypermethylated *EYA4* in AC cancer cells after treatment with a demethylating agent (5-azacytidine) (Supplementary Figure S2 and Supplementary Table S3).

EYA4 is inactivated by CN loss and hypermethylation in both major NSCLC subtypes

We also found that *EYA4* was significantly underexpressed ($P<0.0001$) in a panel of 45 SqCC tumors compared with 67 histologically normal bronchial epithelia samples, and also hypermethylated ($P<0.02$) in a panel of 8 SqCC tumors compared with 8 bronchial epithelia samples for which DNA methylation profiles were available (Figures 1c and d). We also applied our criteria to DNA methylation data downloaded from the recently published lung squamous The Cancer Genome Atlas (TCGA) study.[11] We limited our validation only to those TCGA SqCC

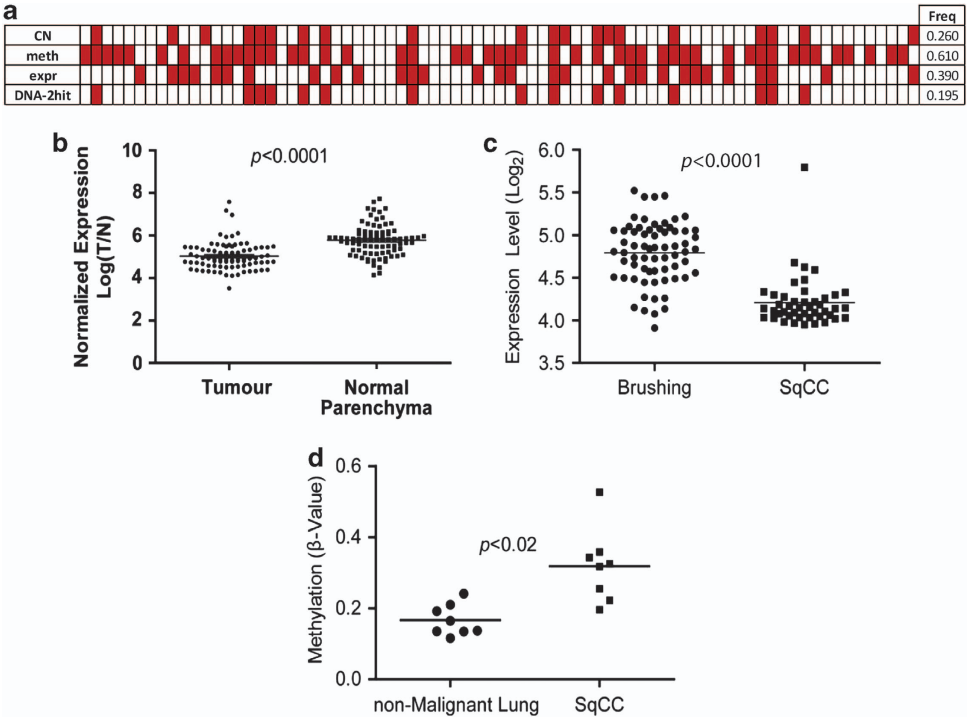


Figure 1. Identification of *EYA4* as a frequently inactivated TSG in lung cancer. (**a**) Summary of gene dosage, DNA methylation and gene expression data for 77 tumor-normal pairs. Each column represents one tumor sample, and red boxes indicate the presence of either a CN loss, hypermethylation or underexpression alteration in a tumor relative to its matched non-malignant parenchymal profile. (**b**) *EYA4* mRNA is significantly ($P<1\times10^{6}$, paired *t*-test) underexpressed in AC tumors ($n=83$) compared with patient-matched non-malignant lung specimens (see also Supplementary Figure S3). (**c**) *EYA4* is significantly underexpressed in SqCC tumors ($n=45$) compared with bronchial epithelia from small airways ($n=67$) ($P<0.0001$ Wilcoxon's signed rank test). (**d**) *EYA4* is significantly more methylated in promoters of SqCC tumors compared with matched non-malignant lung tissue ($n=8$ pairs) ($P<0.02$). Illumina GoldenGate probe β-values were averaged for each sample and plotted as a separate dot.

tumors with available matched normal specimens ($n = 27$ tumor pairs). We found that 15–37% of these tumors had a 20% or greater methylation increase at *EYA4* loci compared with their matched normal counterparts. We further examined an external set of NSCLC specimens ($n = 883$ tumors) from four additional data sets and found that 14% ($n = 123$ tumors) exhibited deletion of *EYA4*.[12–15] Reduced *EYA4* expression was confirmed in two additional independent NSCLC cohorts for which matched normal references were available (Supplementary Figure S3).[16,17] Taken together, these results strongly indicate that *EYA4* is disrupted in both major subtypes of NSCLC.

Mutation is not a major mechanism of *EYA4* disruption

We evaluated whether DNA sequence mutations occurred in *EYA4* by analyzing the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) and also by sequencing all 20 coding exons of *EYA4* in a panel of 38 AC cell lines (listed in Supplementary Table S3). COSMIC analysis revealed 20 confirmed somatic mutations (3%) in a cohort of 639 clinical NSCLC samples. In lung cancer cell lines, we identified only one likely pathogenic variant in exon 7 (*EYA4*: c.385G>C, p.Gly129Arg) of sample H23 that converts a Gly codon to Arg (Figure 2a). This mutation is predicted by PolyPhen2 and SIFT to be deleterious[18] based on the nature of the amino-acid change and the conservation at that residue. To examine the impact of this mutation in these cells, we assessed the locus for gene dosage, loss of heterozygosity, DNA methylation, and mRNA and protein expression. These analyses revealed that *EYA4* is diploid in H23, is not hypermethylated and resides within an area of acquired uniparental disomy. This is consistent with our observation of a homozygous *EYA4* mutation and moderate mRNA expression in these cells, but no detectable EYA4 protein, indicating perhaps that this mutation may impact protein stability (Figure 2b). Given the frequency of observed CN loss (43%) or hypermethylation (40%) events affecting *EYA4* relative to low frequency of *EYA4* mutations (3%), we posit that sequence level mutations are not a major mechanism contributing to *EYA4* disruption in lung cancer.

EYA4 inactivation is an early event in lung cancer

To assess whether somatic DNA level disruptions of *EYA4* occur early in tumorigenesis, we evaluated gene dosage levels in 20 carcinoma *in situ* (CIS) specimens. These rare samples were collected by fluorescent bronchoscopy (Figure 3a) and represent a stage of cancer development typically too early for detection using routine imaging procedures. Deletion of *EYA4* was observed in 35% of these preinvasive squamous CIS lesions, supporting the loss of *EYA4* as an early neoplastic event (Table 1). Transcriptome sequencing data generated in a previous study[19] revealed reduced *EYA4* expression in CIS and SqCC samples compared with histologically normal bronchial epithelial cells (Figure 3b). At the level of DNA methylation, we detected *EYA4* hypermethylation in 40% ($n = 10$) of cytologically normal bronchial epithelia from patients with NSCLC as well as in one high-risk patient with chronic obstructive pulmonary disease (Figure 3c). Inactivation of *EYA4* in the precursors of disease implicates *EYA4* as an early event in lung tumorigenesis.

EYA4 exhibits functional characteristics of a lung cancer TSG

Stable *EYA4* knockdown (*EYA4kd*) of lung lymphoblastoid (HCC-1954BL) cell lines and ectopic overexpression (EYA4+) of AC cells (H2122, H2405) were established for all *in vitro* and *in vivo* assays (Supplementary Figure S4A). AC lines were chosen based on the lack of detectable EYA4 protein expression (Supplementary Figure S2E), and lymphoblastoids as karyotypically normal models highly amenable to the assays performed. Notably, while more appropriate non-malignant lung models were available, including human bronchial epithelial cells, small airway epithelial cells and fetal lung fibroblasts (WI-38), we found these cells unsuitable for subsequent analyses because of unselectability based on our knockdown construct (human bronchial epithelial cells), or drastic changes to cell morphology and behavior (WI-38) upon *EYA4* knockdown. *EYA4* knockdown was also attempted in non-transformed small airway epithelial cell lines; however, loss of *EYA4* in these normal epithelial cells was not tolerated and resulted in cellular senescence, as described previously[20] (data not shown).

EYA4 loss negatively impacts DNA repair and genomic instability

EYA4 is an atypical phosphatase, containing a C-terminal tyrosine (Tyr) and a recently discovered N-terminal threonine (Thr)-phosphatase domain, that has distinct functions and catalytic activity conditions.[21,22] In response to double-stranded breaks, *EYA4* dephosphorylates the Tyr-142 residue of H2AX facilitating phosphorylation of Ser-139 of H2AX (forming γH2AX), leading to the recruitment of DNA repair complex components to sites of double-stranded break.[21,23–26] We assessed levels of γH2AX and Tyr-142 phosphorylation of H2AX in *EYA4kd* and control cells following DNA damage induced by irradiation. Cells lacking *EYA4* accumulate markedly more and longer lasting γH2AX than controls (Figure 4a). Consistently, we observed markedly higher levels of Tyr-142-phosphorylated H2AX in *EYA4kd* cells in response to irradiation compared with empty vector cells (Figures 4b–d). However, restoration of *EYA4* expression in EYA4+ cancer cell models had no significant effect on γH2AX levels following irradiation (Supplementary Figures S4C and D). These results suggest that while *EYA4kd* in karyotypically normal cells results in

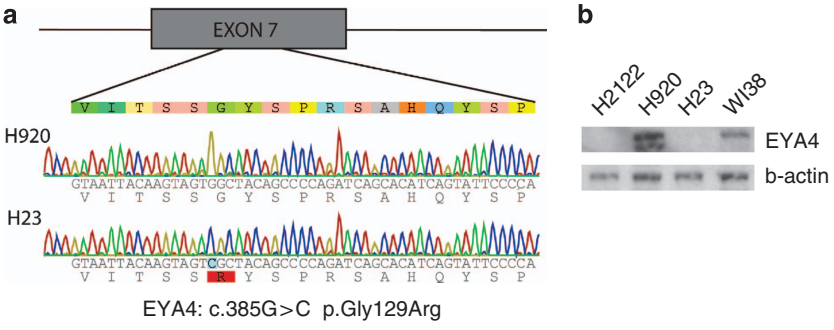


Figure 2. Mechanisms of two-hit *EYA4* inactivation. (**a**) Sample H23 shows a coding sequence mutation in *EYA4*. All exons and proximal intronic sequences were screened in 38 AC lines for mutations by Sanger sequencing. A likely pathogenic missense mutation in exon 7 was identified (*EYA4*: c.385G>C, p.Gly129Arg), which is shown along with wild-type H920. (**b**) This mutation is associated with a lack of protein production as shown by the immunoblot, despite mRNA production by H23. H2122 and H920 are negative and positive controls, respectively, whereas WI38 is a normal fibroblast reference line. This suggests that the mutation may result in premature protein degradation.

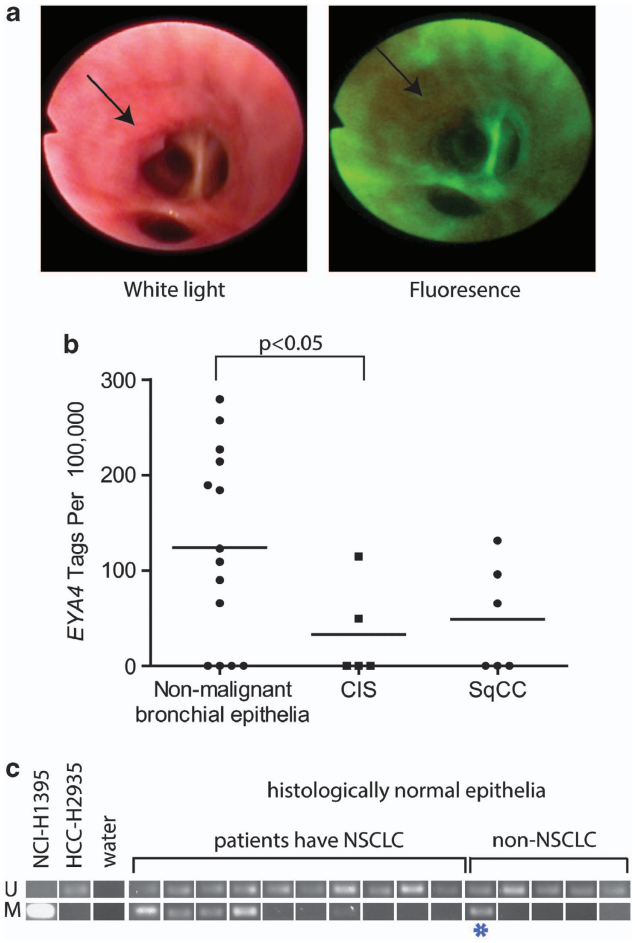


Figure 3. *EYA4* inactivation is an early event. (**a**) White light and autofluorescence bronchoscopy identifies CIS lesions. (**b**) *EYA4* expression was assessed by serial analysis of gene expression in normal bronchial epithelia ($n=14$), CIS ($n=5$) and invasive SqCC ($n=6$) samples. *EYA4* is underexpressed in the CIS group ($P<0.05$ compared with normal) and remains reduced in SqCC. (**c**) *EYA4* is hypermethylated in histologically normal bronchial epithelia of NSCLC patients ($n=10$) compared with patients without cancer ($n=5$) as detected by methylation-specific polymerase chain reaction, providing compelling evidence that inactivation of *EYA4* is a very early event. One high-risk chronic obstructive pulmonary disease (COPD) patient is indicated (*), and positive (H1395) and negative (HCC-2935) controls are shown.

increased double-stranded break during DNA damage onslaught (as measured by γH2AX accumulation), *EYA4* restoration in lung cancer cells is not sufficient to modulate γH2AX-mediated repair. This is likely due to high frequency of mutations in key DNA repair genes, such as *TP53* or *CDKN2A* (both mutated in H2122 and H2405).

To further explore the relationship between *EYA4* and DNA damage response in cancer cells, we examined the effect of a DNA damage-inducing agent, cisplatin, in response to *EYA4* overexpression. We observed that cancer cells made to overexpress *EYA4* were reproducibly more resistant to cisplatin than those without *EYA4*, indicating that in the absence of *EYA4*, cancer cells appear more sensitive to DNA damage onslaught (Figure 4c). Given the observed effect of *EYA4* on DNA repair and response to DNA damage, we hypothesized that tumors with reduced *EYA4* expression would exhibit a greater extent of genomic instability. To test this, we compared the proportion of the genome encompassed by segmental CN alterations in 83 lung AC tumors

Table 1. Deletion of *EYA4* locus in CIS specimens

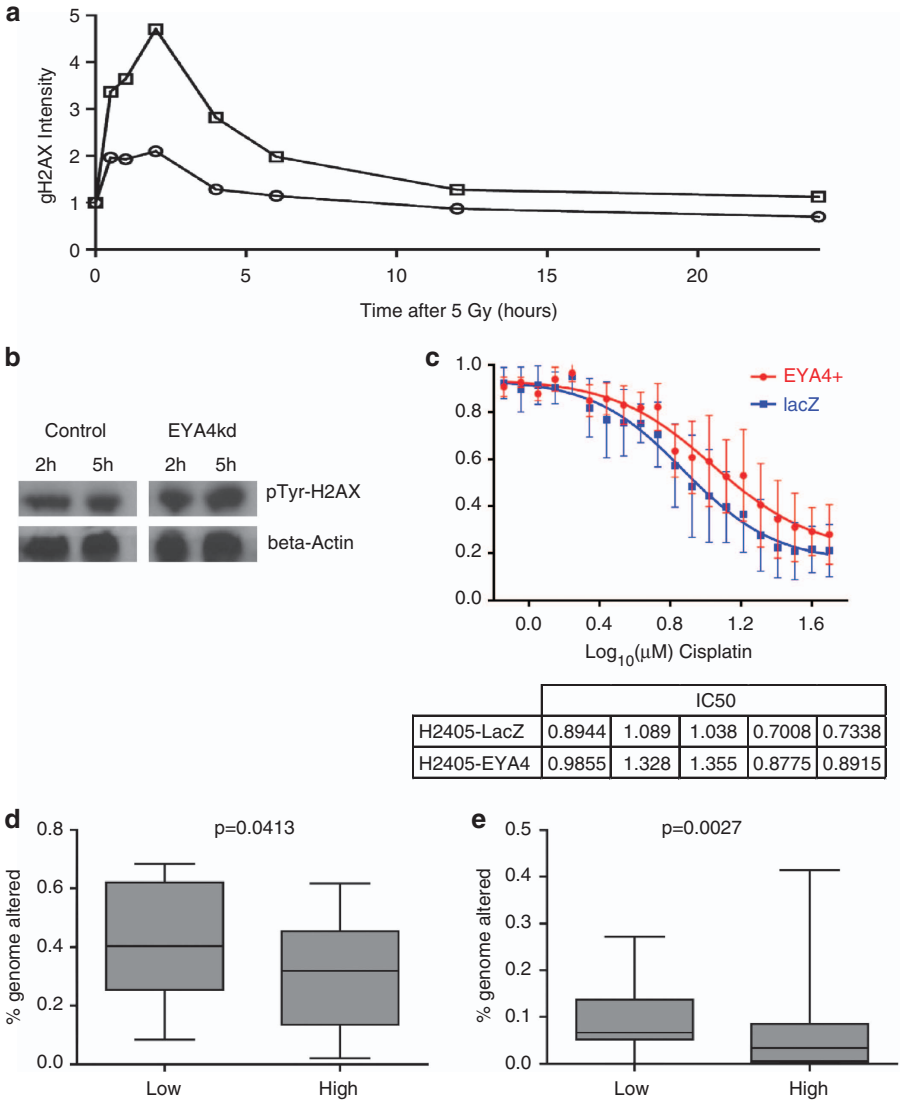
Clone name	*N0654A14*	*N0261J24*
Start (hg18 bp)	133 578 312	133 659 674
End (hg18 bp)	133 721 268	133 850 462
CIS 1	del	del
CIS 2	del	del
CIS 3	del	del
CIS 4	0	0
CIS 5	0	0
CIS 6	0	0
CIS 7	0	x
CIS 8	x	0
CIS 9	0	0
CIS 10	del	del
CIS 11	0	0
CIS 12	0	0
CIS 13	0	0
CIS 14	0	0
CIS 15	del	del
CIS 16	0	0
CIS 17	del	del
CIS 18	0	0
CIS 19	del	del
CIS 20	x	0

Abbreviations: CGH, comparative genomic hybridization; CIS, carcinoma *in situ*; del, clones that undergo copy number loss; EYA4, *Eyes Absent 4*; 0 is neutral; x, uninformative. DNA copy number of *EYA4* was assessed by array CGH in CIS ($n=20$). *EYA4* is frequently (35%) deleted in these very early lesions.

with high and low *EYA4* gene expression (3.2-fold expression difference between groups). This comparison revealed a significant trend toward increased genomic instability in tumors with low *EYA4* expression ($P=0.0413$, two-tailed *U*-test; Figure 4d), an association further supporting a role for *EYA4* in DNA damage repair. We found a similar trend in a second, independent lung AC cohort ($n=193$) downloaded from the Memorial Sloan Kettering Cancer Center[14] ($P=0.0027$) (Figure 4e).

EYA4 loss results in decreased induction of apoptosis. Previous literature implicates *EYA4* as a modulator of apoptotic response.[25,27] We tested this by fluorescence-activated cell sorting (FACS) of annexin V/propidium iodide (AV/PI)-stained *EYA4* modulated and control cells following serum starvation. EYA4+ cancer cells displayed no differences in the percentage of early or late apoptotic cells (Supplementary Figure S4B). As cancer cells have heavily disrupted genetic backgrounds that likely interfere with apoptotic pathways, we examined the effect of *EYA4* expression on apoptosis of karyotypically normal *EYA4kd* lines and controls. FACS results indicated that, contrary to our cancer cell lines, *EYA4kd* cells displayed a marked and reproducible decrease in the numbers of early (AV+/PI−) and late (AV+/PI+) apoptotic cells compared with pLKO ($26.6\pm4.6\%$ for control vs $14.5\pm1.7\%$ for *EYA4kd*) (Figures 5a–c), indicative of a proapoptotic role for *EYA4*.

Previous reports have suggested that a TSG role for *EYA4* may be mediated via its role as a transcriptional coactivator.[5,6,27,28] To identify genes correlated with *EYA4* expression (and potentially activated by *EYA4*), profiles for the 10 highest and 10 lowest *EYA4*-expressing normal bronchial specimens were compared using the significance analysis of microarrays algorithm.[29] Twenty-eight genes potentially coactivated by *EYA4* were identified (*q*-value percentage threshold of 5%) (Supplementary Table S4). One of the most highly correlated of these genes was *GADD45a*, which has a role in the apoptotic program, DNA damage and cell cycle arrest.[30] Indeed, we found that following serum starvation, *GADD45a* expression was attenuated in *EYA4kd* cells compared with control cells (Figure 5d).



	IC50				
H2405-LacZ	0.8944	1.089	1.038	0.7008	0.7338
H2405-EYA4	0.9855	1.328	1.355	0.8775	0.8915

Figure 4. *EYA4* promotes genomic stability and efficient DNA repair in karyotypically normal cells. (**a**) γH2AX levels in *EYA4kd* cells. DNA damage response was assessed in *EYA4kd* cells. Normalized fluorescence intensity for γH2AX in cells at different time points following irradiation with 5 Gy show cells lacking *EYA4* (square markers) accumulate markedly more γH2AX than wild-type cells (round markers). Cells lacking *EYA4* also demonstrate a marked reduction in the removal of γH2AX up to 6 h after irradiation. (**b**) Western blot of the levels of Tyr-142-phosphorylated H2AX in HCC-1954BL control and *EYA4kd* cell lysates at 2 and 5 h after irradiation. β-Actin was used as a loading control. (**c**) Cisplatin sensitivity assays for *EYA4* ectopically expressing H2405 lung AC cells and H2405 cells that do not express *EYA4* (LacZ controls) were performed in replicates of five. Lung adenocarinoma cells (no *EYA4*) are more sensitive to the DNA damage-inducing agent, cisplatin. *P*-value calculated by a paired one-tailed *t*-test. (**d**) Fraction of genome encompassed by segmental CN alterations (a measure for genomic instability) calculated for AC tumors ($n = 83$, BCCA) (see Materials and methods). Genomic instability was significantly higher in tumors with low vs high *EYA4*-expressing tumors ($P < 0.05$). (**e**) Low *EYA4* expression was also significantly associated with a greater extent of genomic instability ($P < 0.05$) in an additional, independent cohort of AC ($n = 193$, Memorial Sloan Kettering Cancer Center (MSKCC)).

These results are intriguing, although not explicitly demonstrative of any mechanistic link between the two. The observed relationship between *EYA4* and *GADD45a* expression warrants further exploration.

Restoration of EYA4 expression inhibits tumor growth in vitro and in vivo

Although restoration of *EYA4* expression in two AC cell lines had little effect on DNA damage and apoptosis, we sought to determine whether *EYA4* expression had any effect on anchorage-independent growth and tumor growth *in vivo*. Indeed, our EYA4+ AC cell lines had significantly impaired cell growth and colony formation ability as measured by soft agar colony formation (Figures 6a and b). When H2122 EYA4+ was implanted into non-obese diabetic/severe-combined immunodeficiency mice, *EYA4* overexpression significantly impeded tumor growth (Figure 6c and Supplementary Table S5).

EYA4 exhibits cell-type-specific expression

Intriguingly, in addition to its putative TSG functions in various cancers, *EYA4* has also been described as an oncogene in neural cancer.[31] We evaluated cancer cell-type-specific expression of *EYA4* in a panel of over 350 cancer cell lines encompassing a variety of cancer types (Supplementary Table S6). We detected high *EYA4* expression in cancer cells derived from sarcomas, autonomic ganglia and brain, relative to epithelial cancers such as the lung, gastrointestinal, pancreatic, head and neck and

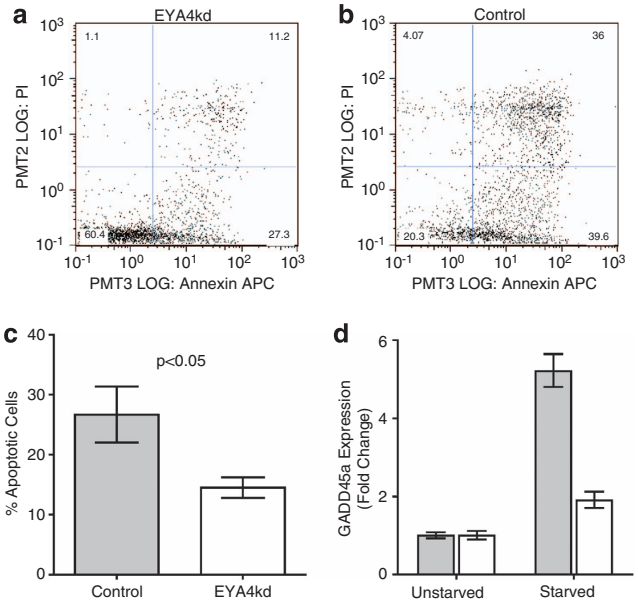


Figure 5. *EYA4* promotes apoptosis and genomic stability. (**a**) Apoptosis was assessed in *EYA4kd* cells. AV binding is depicted on the X axis and PI staining on the Y axis. FACS analysis following serum starvation indicates abrogation of the apoptotic program in *EYA4kd* HCC-1954BL cells compared with control cells. (**b**) Empty pLKO vector control cells show more apoptotic cells (upper right quadrant) vs *EYA4kd* cells. (**c**) Triplicate FACS comparisons show significantly more apoptotic cells in the control cells (gray) than in the *EYA4*kd cells (white) ($P<0.05$, *t*-test). (**d**) *GADD45a* expression levels in serum-starved *EYA4kd* (white) and control cells (gray) were assessed by quantitative real-time reverse transcription PCR (qRT–PCR). *GADD45a* expression change following 24 h of serum starvation increases substantially in control cells, and is attenuated in *EYA4kd* cells ($P<0.001$, *t*-test).

colorectal tumors (Figure 6d). These findings are consistent with a TSG role for *EYA4* in lung cancer, and interestingly support a dual role (TSG or oncogenic) for *EYA4* that is likely dependent on the cell type of origin. Of note, *EYA4* was not differentially expressed between small-cell lung cancer, which is thought to be derived from pulmonary neuroendocrine cells, and NSCLC (Figure 6e).

Clinical relevance of *EYA4*

We performed a Mantel–Cox survival analysis in two external lung AC data sets with survival data (GSE3141[16] and GSE12428[32]), and found that reduced *EYA4* expression was significantly and consistently associated with poor survival (Figures 7a and b), underscoring the clinical relevance of *EYA4* expression to AC patient prognosis. Although *EYA4* single-nucleotide polymorphisms (SNPs) have not been previously associated with familial lung cancer risk, given the frequency of *EYA4* inactivation in sporadic lung cancer, and the proximity of *EYA4* to the previously refined lung cancer susceptibility locus by You *et al.*[33] and Bailey-Wilson *et al.*[34] at 6q23–25, we sought to evaluate the significance of *EYA4* SNPs in familial lung cancer. A Cochrane–Armitage Trend test corrected for multiple comparisons ($P<0.05$) revealed that a cluster of SNPs in the *EYA4* gene (rs7743259, rs159420, rs35689029, rs1878551 and rs2677826) were indeed enriched in a panel of familial NSCLC cases (Figures 7c and d and Supplementary Table S7).[35]

DISCUSSION

Cancer genomes are frequently disrupted at multiple 'omic' levels—all of which may be differentially impacted by unique selective pressures occurring throughout the tumorigenic process. Within these tumor systems, it is likely that genes critical to abrogating tumor development undergo biallelic ('two-hit') disruption, as commonly observed with many TSGs. If two-hit events occur by differing mechanisms, the frequency of alteration for that gene may be low when assessed for only one mechanism, and thus likely overlooked. However, when multiple dimensions of disruption are considered simultaneously, alteration of the two-hit gene in question may be detected at a high frequency. Therefore, identification of these events in the complicated genomes of epithelial malignancies, such as lung cancer, requires the simultaneous interrogation of multiple 'omics' level data sets.

We applied such a multidimensional approach to a panel of patient-matched, paired tumor and non-malignant parenchymal samples from patients with NSCLC, and identified *EYA4*, a gene frequently and often simultaneously disrupted by CN loss and promoter hypermethylation in multiple clinical cohorts, representing over one thousand NSCLC tumors. Disruption of *EYA4* was also detected at the level of CN loss in SqCC precursor CIS specimens, and by promoter hypermethylation in cytologically normal small airway epithelia from patients with NSCLC. In addition to invasive lung AC, promoter hypermethylation of *EYA4* was also recently reported in atypical adenomatous hyperplasia and AC *in situ* (formerly known as bronchioloalveolar carcinoma), collectively pointing to the importance of *EYA4* in cancer initiation events of both major NSCLC subtypes.[5,36] Interestingly, early inactivation of *EYA4* by hypermethylation has been associated with Barrett's esophagus-related tumorigenesis, sporadic and colitic neoplasia in chronic ulcerative colitis and is currently being explored as an epigenetic biomarker for colorectal and pancreatic cancer screening.[6,9,37] Given our findings, exploration of *EYA4* as a maker for early NSCLC detection is warranted.

Consistent with TSG function, reconstitution of *EYA4* in NSCLC cell lines decreased soft agar colony formation and *in vivo* tumor growth. And, although we found lower *EYA4* expression associated with increased sensitivity to DNA-damaging cisplatin treatment in NSCLC cell lines, and increased proportions of genome altered in lung tumors, functional impact on both DNA repair and apoptosis was only significant when *EYA4* was abrogated in karyotypically normal cell models as opposed to restored in NSCLC models. Given the multiplicity of genomic aberrations in lung cancer cell lines models, particularly affecting DNA repair and apoptotic proteins (e.g., p53, ATM, PTEN), we were not surprised that modulation of one gene failed to cause significant phenotypic changes. However, in the context of an early tumor suppressor role for *EYA4*, it is possible that inactivation of *EYA4* could lead to increased lung cancer risk by promoting an impaired response to DNA-damaging agents such as cigarette smoke.

Although multiple lines of evidence in several cancers are supportive of a TSG role, *EYA4* is also an overexpressed putative oncogene in tumors of neural origin.[31] This discrepancy may be inherent to the unique and recently discovered dual phosphatase properties of *EYA4*, whereby a Tyr-phosphatase domain at the C terminal and a Thr-phosphatase domain at the N-terminal function independently and under distinct catalytic conditions.[21,22] Mutations of the Thr-phosphatase domain, but not the Tyr-phosphatase domain, have been shown to abolish the ability of *EYA4* to enhance the innate immune response to viruses and undigested double-stranded DNA from apoptotic cells,[22] and mutations to the Tyr-phosphatase domain, which normally promotes DNA repair in response to double-stranded break, are linked with neuronal developmental defects and deafness. Our findings in over 350 cancer cell lines from multiple tumor types demonstrate that *EYA4* is expressed at substantially lower levels in epithelial tumors compared with sarcomas or neurally derived tumors, supporting the seemingly contradictory findings and pointing towards an either tissue-specific or dual TSG or oncogene role for *EYA4* in cancer development. In the context of lung cancer,

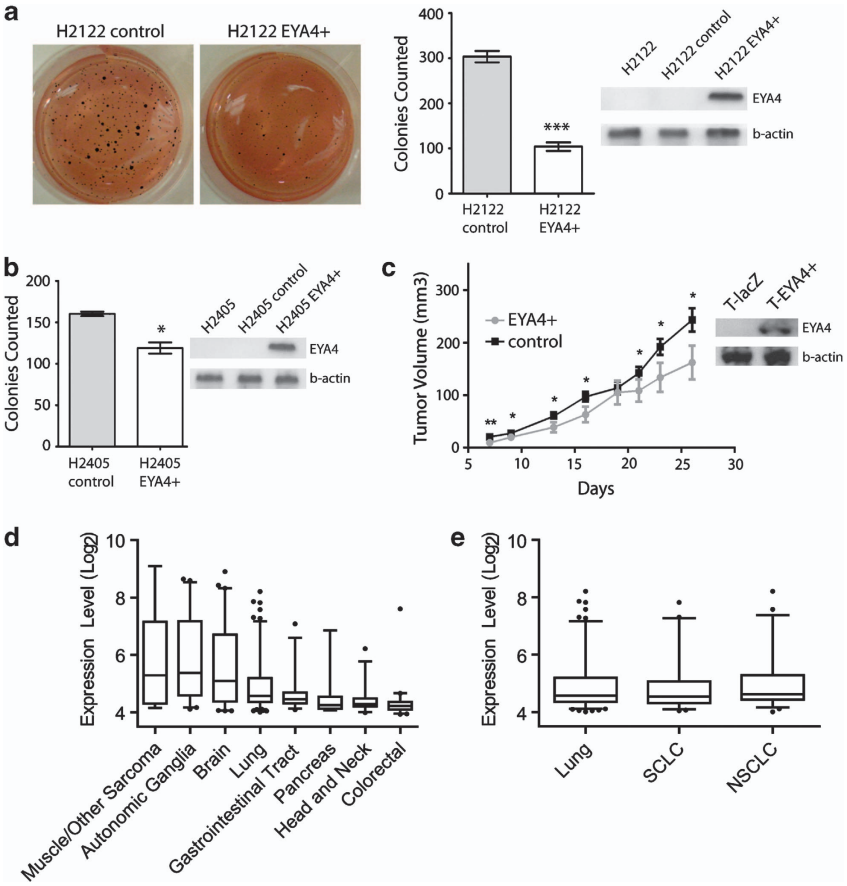
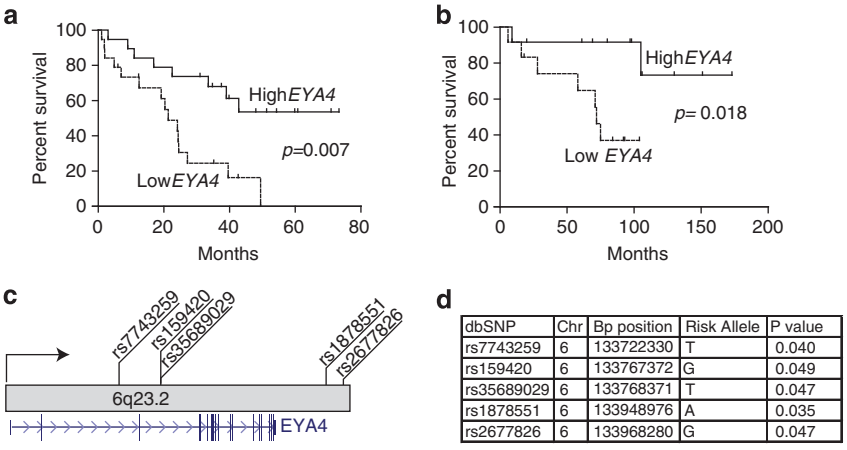


Figure 6. *EYA4* suppresses colony growth and is widely underexpressed in epithelial malignancies. (**a**) Reconstitution of *EYA4* suppresses colony formation. *EYA4* was overexpressed in the lung cancer cell H2122, which lacks *EYA4*, by stable integration of an *EYA4* vector (levels indicated by the corresponding immunoblot). Colony formation was significant ($P = 0.00037$, *t*-test). (**b**) Reconstitution of *EYA4* in the lung cancer cell lines, H2405 also significantly ($P = 0.0153$, *t*-test, significance indicated by asterisk) suppressed colony formation, consistent with a TSG role for *EYA4*. (**c**) *In vivo* tumor growth of EYA4+ and control H2122 cells (also see Supplementary Table S5). Overexpression of *EYA4* significantly ($P < 0.05$ indicated by an asterisk) impairs tumor growth *in vivo*. (**d**) *EYA4* expression in greater than 350 cancer cells from multiple tissues (also see Supplementary Table S6). Whiskers encompass middle 90% of data points. *EYA4* expression appears to be tissue specific and higher in sarcomas and in neurally derived cancers than in epithelial cancers. (**e**) Lung cancer histological subtype-specific expression was not observed.



dbSNP	Chr	Bp position	Risk Allele	P value
rs7743259	6	133722330	T	0.040
rs159420	6	133767372	G	0.049
rs35689029	6	133768371	T	0.047
rs1878551	6	133948976	A	0.035
rs2677826	6	133968280	G	0.047

Figure 7. *EYA4* is associated with lung cancer risk and poor survival. (**a**) Kaplan–Meier plot. Survival information from an external data set (GSE3141) for the highest and lowest tertiles (of *EYA4* expression) was compared using the Mantel–Cox log test. *EYA4* is significantly associated with poorer prognosis ($P = 0.007$). (**b**) The analysis was repeated using another external data set (GSE12428). Low *EYA4* expression is significantly associated with poor prognosis ($P = 0.018$). (**c**) *EYA4* SNPs are significantly associated with familial lung cancer risk. Genotype data for 6q-linked familial lung cancers ($n = 194$) and unrelated non-cancer controls ($n = 217$) were compared to determine whether *EYA4* allelotypes associate with risk. Five SNPs depicted were significantly associated ($P < 0.05$). Gene structure is shown in blue. (**d**) The allele associated with an increased risk is described as the risk allele. Demographic features, stage and histology are unavailable for the external data sets for panels **a** and **b**.

inherited or acquired *EYA4* disruption could potentially result in uncontrolled cellular division, accumulation of genetic damage from genotoxic agents such as cigarette smoke or impaired innate immunity from double-stranded DNA released from damaged respiratory cells.

The Genetic Epidemiology of Lung Cancer Consortium (GELCC) has identified a susceptibility locus on 6q, which contains the over-expressed candidate oncogene, *RGS17*.[33–35,38–41] We discovered a cluster of SNPs within *EYA4*, immediately adjacent to this susceptibility locus, that are significantly associated with lung cancer risk in familial lung cancer cases. Patterns of *EYA4* disruption in lung cancer are similar to other important TSGs involving both sporadic and familial cancers, which also frequently sustain biallelic inactivation.[42] For example, *BRCA1/2* germline mutations and somatic promoter hypermethylation events occur in familial breast tumors, whereas biallelic inactivation by hypermethylation and deletion is common in sporadic breast cancer.[42–44] Taken together with evidence strongly indicating selective inactivation of *EYA4* in sporadic lung cancer, our findings raise the intriguing possibility of a novel familial lung cancer susceptibility gene.

In summary, the prevalence of biallelic inactivation of *EYA4* in NSCLC, its multiple tumor suppressor functions and association with survival in sporadic lung cancers and familial lung cancer risk suggest that *EYA4* is important to NSCLC development, and may be a promising marker for early lung cancer detection. Considering its diverse biological functions and likely multifaceted role in lung carcinogenesis, the development of appropriate *in vivo* models to assess the role and potential manipulation of *EYA4* pathways in lung cancer are needed, albeit highly challenging.

MATERIALS AND METHODS

Sample collection

Lung tumors and adjacent non-malignant tissue were obtained from freshly resected tumors and microdissected so that the tumor cell content exceeded 80%, and DNA and RNA were extracted using standard protocols. Bronchial epithelial specimens from airways ≤ 2 mm diameter, and biopsy of locally invasive SqCC and CIS specimens, were obtained during bronchoscopy as described previously.[19] This study was approved by the Review of Ethics Board of the University of British Columbia and the British Columbia Cancer Agency.

Molecular profiling

Genomic DNA for 83 lung ACs and matched non-malignant lung tissues from the same individuals was hybridized to Affymetrix SNP 6.0 arrays. Raw data were processed, segmented and CN alterations called using Partek Genomic Suite Software (Partek, Missouri, MO, USA), as defined previously.[12] SNP array data for the BCCA tumors is in compliance with the MIAME guidelines and has been deposited in the Gene Expression Omnibus. DNA methylation profiling for AC tumor pairs ($n = 77$) and cell lines ($n = 38$) was performed using the Illumina Infinium HumanMethylation27 chip, and SqCC tumor pairs ($n = 8$) on Illumina GoldenGate Cancer chip (Illumina) was performed using methods described previously, including bisulfite conversion using EZ DNA methylation-Gold kit (Zymo Research, Irvine, CA, USA).[45] Array data were analyzed and methylation levels determined using GenomeStudio software. $\beta =$ Methylated signal/(methylated signal + unmethylated signal + α); β with a detection *P*-value ≤ 0.05 were included. Hypermethylation was defined as $>20\%$ β-value difference in tumor relative to patient-matched non-malignant control. Methylation validation was performed by real-time methylation-specific polymerase chain reaction as described previously using the *EYA4* primers, 5′-TTGCGTAAGTGCGAGGTTGTC-3′ (forward), 5′-AACA ACGACAACTTCACGTAA-3′ (reverse), and 5′-FAM TCGTTTTCGGTTTTCGC GTAA BHQ1-3′ (probe), using non-methylated *MYOD1* as an internal reference standard. Standard methylation-specific polymerase chain reaction was performed using primers specific to methylated and unmethylated forms of *EYA4* as described previously.[28] Re-expression of methylated *EYA4* following demethylation by 10 μM 5-azacytidine every 2 days (Sigma Aldrich, St Louis, MO, USA) for 6 days was validated in NCI-H1395 and NCI-HCC2935 cells, otherwise cultured as per ATCC (Manassas, VA, USA) directions. Gene expression profiles were generated for AC tumor-normal pairs on the Illumina WG6 microarray (Illumina). *EYA4* DNA methylation status was also validated in an external data set downloaded from The Cancer Genome Atlas Data Portal (https://tcga-data.nci.nih.gov/tcga/) for all SqCC specimens that had methylation profiles for tumor and matched non-malignant specimens ($n = 27$ pairs).[46] Forty-five SqCC Affymetrix expression profiles (GSE3141) and 67 non-malignant bronchial epithelia samples (sigma.bccrc.ca) were retrieved from heavy current and former smokers. Publicly available Affymetrix CEL files were downloaded from NCBI Gene Expression Omnibus (GEO) (GSE3141, GSE10072 and GSE12428)[16,17,32,47] or from The Sanger Cell Line Project at the BROAD Institute (Cambridge, MA, USA). Where appropriate, CEL files were grouped, and RMA analysis[48] was performed using the 'affy' package from Bioconductor (Bioconductor, Seattle, WA, USA).[49] Gene expression profiles for lung AC cell lines ($n = 38$) were obtained by Human WG-6 gene expression chip. CN and gene expression data (*EYA4* SAGE tag: TAATTTGTGT) from CIS specimens were obtained from a previous study (GSE7898).[19] *EYA4* allelotypes data of 6q-linked familial lung cancers ($n = 194$) and unrelated non-cancer controls ($n = 217$) was obtained from a previous study.[50] Quantitative PCR validation was performed in triplicate using the primers: EYA4—Hs00187965_m1, 18s—Hs99999901_s1 and GADD45a—Hs00169255_m1. Protein lysates and western blots were prepared as described[51] (*EYA4* (Santa Cruz Biotechnology, Santa Cruz, CA, USA), sc-15106; pTyr-H2AX (Millipore, Billerica, MA, USA), 07-1590; β-actin (Abcam, Cambridge, UK), ab8226). DNA sequencing was performed for *EYA4* exons and proximal intronic sequences in 38 lung AC cell lines (listed in Supplementary Table S3) using previously described primers (Supplementary Table S8).

Statistical analyses

Correlation coefficients for *EYA4* DNA methylation and mRNA expression levels for 38 lung AC cell lines were obtained by Spearman's tests, for each of the eight *EYA4* DNA methylation probes against the average of the three *EYA4* expression probes (Supplementary Table S3). For genomic integrity assessment, tumors were segregated into tertiles based on *EYA4* expression and the proportion of genome altered was compared in tumors with low vs high *EYA4* expression using a *U*-test. This analyses was applied to the 83 AC tumor pairs as well as publicly available matched CN and gene expression data from the Memorial Sloan Kettering Cancer Center,[14] for which matched CN (Agilent 44K arrays, Agilent, Santa Clara, CA, USA) and gene expression data (Affymetrix U133A arrays, Affymetrix) for 193 lung AC tumors were available. Memorial Sloan Kettering Cancer Center CN data was segmented using the segmentation algorithm, FACADE.[52] For survival analyses, data were downloaded from NCBI GEO[47] (GSE3141 and GSE12428). Highest and lowest tertiles of samples, based on *EYA4* expression (probe 1561088_at), were analyzed using a log-rank (Mantel–Cox) test. Data from Liu *et al.*[50] were reanalyzed to determine whether *EYA4* allelotypes were associated with risk in familial lung cancers. The statistical significance of the association between SNP allele and disease status was assessed primarily with Cochran–Armitage trend test with one degree of freedom, implemented in PLINK software. Allelic odds ratios associated with each SNP and 95% confidence intervals were estimated.

In vitro assays

Lung AC (NCI-H2122, NCI-H2405) and lymphoblastoid (NCI-HCC1954BL) cells were obtained from ATCC. Stable knockdowns of *EYA4* were performed in triplicate for all cells using lentiviral vectors (clone i.d.: T TRCN0000051094), and a puromycin resistance selectable marker (Open Biosystems, Huntsville, AL, USA). Overexpression of *EYA4* was performed in AC cells by transfection using Invitrogen's Ultimate ORF with clone i.d.: IOH57275. *EYA4* was inserted by shuttle from entry vector pENTR221 to the lentiviral destination vector pLenti6.3/V5-DEST using LR recombination reaction. Lentiviral stock were produced using Invitrogen's ViraPower HiPerform Lentiviral expression system (Life Technologies, Carlsbad, CA, USA), after transfecting with lentiviral particles for 24 h and Blasticidin selection for 10–14 days. Soft agar colony formation was performed as described.[53] AV/PI staining for EYAkd and controls was performed (in triplicate), following 24 h serum starvation. Cells were washed in phosphate-buffered saline, resuspended in AV-binding buffer, PI and fluorescein isothiocyanate-conjugated anti-AV antibody (BD Bioscience, San Jose, CA, USA). Apoptotic cells were counted by flow cytometry in a FACS Canto II (BD Biosciences). γH2AX kinetics in response to radiation induced DNA damage was performed as described previously,[54] using 5 Gy radiation, and mouse monoclonal anti-phospho-Ser[13] H2AX primary antibody (Abcam; no. 18311, 1:4000 dilution; BD Bioscience Alexa Fluor 647-conjugated anti-γH2AX antibody, 4′,6-diamidino-2-phenylindole). To account for differences in radiation-induced changes in cell cycle

distribution that affect average γH2AX intensity measurements, γH2AX expression per cell was analyzed separately in G1-phase cells and results were expressed as a ratio of the signal intensity for irradiated vs unirradiated cells, and analyzed using the FACS Canto II flow cytometer. For cisplatin assays, 24 h after seeding, cells were treated for 72 h and then stained with Alamar Blue for absorbance analysis and IC50 calculation (GraphPad Prism 6 software, GraphPad, La Jolla, CA, USA).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Miwa Suzuki, Denise McDougal, Chad Malloff and Bradley Coe for assistance. Grant supports were from Canadian Institutes of Health Research (CIHR MOP86731, MOP77903, MOP94867), Canadian Cancer Society (CCS017076, CCS20485); Terry Fox Foundation (20395); NCI Early Detection Research Network (5U01 CA84971-10), Canary Foundation, NIH Genetic Epidemiology of Lung Cancer Consortium (U01CA76293); and scholarships from CIHR (IMW, EAV, KSSE and WWL) and Vanier Canada (RC, KLT and NR).

REFERENCES

1 Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380**: 2095–2128.
2 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *Cancer J Clin* 2011; **61**: 69–90.
3 Knudson Jr. AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971; **68**: 820–823.
4 Nagayama K, Kohno T, Sato M, Arai Y, Minna JD, Yokota J. Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution. *Genes Chromosomes Cancer* 2007; **46**: 1000–1010.
5 Selamat SA, Galler JS, Joshi AD, Fyfe MN, Campan M, Siegmund KD *et al.* DNA methylation changes in atypical adenomatous hyperplasia, adenocarcinoma *in situ*, and lung adenocarcinoma. *PLoS One* 2011; **6**: e21443.
6 Osborn NK, Zou H, Molina JR, Lesche R, Lewin J, Lofton-Day C *et al.* Aberrant methylation of the eyes absent 4 gene in ulcerative colitis-associated dysplasia. *Clin Gastroenterol Hepatol* 2006; **4**: 212–218.
7 Oster B, Thorsen K, Lamy P, Wojdacz TK, Hansen LL, Birkenkamp-Demtroder K *et al.* Identification and validation of highly frequent CpG island hypermethylation in colorectal adenomas and carcinomas. *Int J Cancer* 2011; **129**: 2855–2866.
8 Kim YH, Lee HC, Kim SY, Yeom YI, Ryu KJ, Min BH *et al.* Epigenomic analysis of aberrantly methylated genes in colorectal cancer identifies genes commonly affected by epigenetic alterations. *Ann Surg Oncol* 2011; **18**: 2338–2347.
9 Kisiel JB, Yab TC, Taylor WR, Chari ST, Petersen GM, Mahoney DW *et al.* Stool DNA testing for the detection of pancreatic cancer: assessment of methylation marker candidates. *Cancer* 2012; **118**: 2623–2631.
10 Thu KL, Radulovich N, Becker-Santos DD, Pikor LA, Pusic A, Lockwood WW *et al.* SOX15 is a candidate tumor suppressor in pancreatic cancer with a potential role in Wnt/beta-catenin signaling. *Oncogene* (e-pub ahead of print 14 January 2013; doi:10.1038/onc.2012.595).
11 Hammerman PS, Hayes DN, Wilkerson MD, Schultz N, Bose R, Chu A *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012; **489**: 519–525.
12 Thu KL, Vucic EA, Chari R, Zhang W, Lockwood WW, English JC *et al.* Lung adenocarcinoma of never smokers and smokers harbor differential regions of genetic alteration and exhibit different levels of genomic instability. *PLoS One* 2012; **7**: e33003.
13 Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* 2007; **450**: 893–898.
14 Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R *et al.* An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* 2009; **28**: 2773–2783.
15 Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM *et al.* Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Sci Transl Med* 2010; **2**: 62ra93.
16 Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D *et al.* Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006; **439**: 353–357.
17 Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* 2008; **3**: e1651.
18 Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat* 2011; **32**: 358–368.
19 Lonergan KM, Chari R, Coe BP, Wilson IM, Tsao MS, Ng RT *et al.* Transcriptome profiles of carcinoma-*in-situ* and invasive non-small cell lung cancer as revealed by SAGE. *PLoS One* 2010; **5**: e9162.
20 Chen Z, Trotman LC, Shaffer D, Lin HK, Dotan ZA, Niki M *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* 2005; **436**: 725–730.
21 Sadatomi D, Tanimura S, Ozaki K, Takeda K. Atypical protein phosphatases: emerging players in cellular signaling. *Int J Mol Sci* 2013; **14**: 4596–4612.
22 Okabe Y, Sano T, Nagata S. Regulation of the innate immune response by threonine-phosphatase of Eyes absent. *Nature* 2009; **460**: 520–524.
23 MacPhail SH, Banath JP, Yu TY, Chu EH, Lambur H, Olive PL. Expression of phosphorylated histone H2AX in cultured cell lines following exposure to X-rays. *Int J Radiat biology* 2003; **79**: 351–358.
24 Pignoni F, Hu B, Zavitz KH, Xiao J, Garrity PA, Zipursky SL. The eye-specification proteins So and Eya form a complex and regulate multiple steps in *Drosophila* eye development. *Cell* 1997; **91**: 881–891.
25 Clark SW, Fee BE, Cleveland JL. Misexpression of the eyes absent family triggers the apoptotic program. *J Biol Chem* 2002; **277**: 3560–3567.
26 Liu X, Sano T, Guan Y, Nagata S, Hoffmann JA, Fukuyama H. *Drosophila* EYA regulates the immune response against DNA through an evolutionarily conserved threonine phosphatase motif. *PLoS One* 2012; **7**: e42725.
27 De Carvalho DD, Sharma S, You JS, Su SF, Taberlay PC, Kelly TK *et al.* DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell* 2012; **21**: 655–667.
28 Zou H, Osborn NK, Harrington JJ, Klatt KK, Molina JR, Burgart LJ *et al.* Frequent methylation of eyes absent 4 gene in Barrett's esophagus and esophageal adenocarcinoma. *Cancer Epidemiol Biomarkers Prev* 2005; **14**: 830–834.
29 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.
30 Zhang X, Sun H, Danila DC, Johnson SR, Zhou Y, Swearingen B *et al.* Loss of expression of GADD45 gamma, a growth inhibitory gene, in human pituitary adenomas: implications for tumorigenesis. *J Clin Endocrinol Metab* 2002; **87**: 1262–1267.
31 Miller SJ, Lan ZD, Hardiman A, Wu J, Kordich JJ, Patmore DM *et al.* Inhibition of Eyes Absent Homolog 4 expression induces malignant peripheral nerve sheath tumor necrosis. *Oncogene* 2010; **29**: 368–379.
32 Boelens MC, van den Berg A, Fehrmann RS, Geerlings M, de Jong WK, te Meerman GJ *et al.* Current smoking-specific gene expression signature in normal bronchial epithelium is enhanced in squamous cell lung cancer. *J Pathol* 2009; **218**: 182–191.
33 You M, Wang D, Liu P, Vikis H, James M, Lu Y *et al.* Fine mapping of chromosome 6q23–25 region in familial lung cancer families reveals RGS17 as a likely candidate gene. *Clin Cancer Res* 2009; **15**: 2666–2674.
34 Bailey-Wilson JE, Amos CI, Pinney SM, Petersen GM, de Andrade M, Wiest JS *et al.* A major lung cancer susceptibility locus maps to chromosome 6q23–25. *Am J Hum Genet* 2004; **75**: 460–474.
35 Amos CI, Pinney SM, Li Y, Kupert E, Lee J, de Andrade MA *et al.* A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. *Cancer Res* 2010; **70**: 2359–2367.
36 Selamat SA, Chung BS, Girard L, Zhang W, Zhang Y, Campan M *et al.* Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res* 2012; **22**: 1197–1211.
37 Kisiel JB, Yab TC, Nazer Hussain FT, Taylor WR, Garrity-Park MM, Sandborn WJ *et al.* Stool DNA testing for the detection of colorectal neoplasia in patients with inflammatory bowel disease. *Aliment Pharmacol Therap* 2013; **37**: 546–554.
38 Sellers TA, Bailey-Wilson JE, Elston RC, Wilson AF, Elston GZ, Ooi WL *et al.* Evidence for Mendelian inheritance in the pathogenesis of lung cancer. *J Natl Cancer Instit* 1990; **82**: 1272–1279.
39 Sellers TA, Potter JD, Bailey-Wilson JE, Rich SS, Rothschild H, Elston RC. Lung cancer detection and prevention: evidence for an interaction between smoking and genetic predisposition. *Cancer Res* 1992; **52**(Suppl): 2694s–2697ss.
40 Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. *J Natl Cancer Inst* 1963; **30**: 289–312.
41 Tessema M, Willink R, Do K, Yu YY, Yu W, Machida EO *et al.* Promoter methylation of genes in and around the candidate lung cancer susceptibility locus 6q23–25. *Cancer Res* 2008; **68**: 1707–1714.
42 Bell DW, Erban J, Sgroi DC, Haber DA. Selective loss of heterozygosity in multiple breast cancers from a carrier of mutations in both BRCA1 and BRCA2. *Cancer Res* 2002; **62**: 2741–2743.
43 Bijron JG, van der Groep P, van Dorst EB, Seeber LM, Sie-Go DM, Verheijen RH *et al.* Promoter hypermethylation patterns in Fallopian tube epithelium of BRCA1 and BRCA2 germline mutation carriers. *Endocr Relat Cancer* 2011; **19**: 69–81.

44 Birgisdottir V, Stefansson OA, Bodvarsdottir SK, Hilmarsdottir H, Jonasson JG, Eyfjord JE. Epigenetic silencing and deletion of the BRCA1 gene in sporadic breast cancer. *Breast Cancer Res* 2006; **8**: R38.
45 Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R *et al.* Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* 2009; **1**: 177–200.
46 Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012; **489**: 519–525.
47 Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C *et al.* NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res* 2009; **37**: D885–D890.
48 Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**: 249–264.
49 Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80.
50 Liu P, Vikis HG, Wang D, Lu Y, Wang Y, Schwartz AG *et al.* Familial aggregation of common sequence variants on 15q24–25.1 in lung cancer. *J Natl Cancer Inst* 2008; **100**: 1326–1330.
51 Thu KL, Pikor LA, Chari R, Wilson IM, Macaulay CE, English JC *et al.* Genetic disruption of KEAP1/CUL3 E3 ubiquitin ligase complex components is a key mechanism of NF-kappaB pathway activation in lung cancer. *J Thorac Oncol* 2011 1521–1529.
52 Coe BP, Chari R, MacAulay C, Lam WL. FACADE: a fast and sensitive algorithm for the segmentation and calling of high resolution array CGH data. *Nucleic Acids Res* 2010; **38**: e157.
53 Lockwood WW, Wilson IM, Coe BP, Chari R, Pikor LA, Thu KL *et al.* Divergent genomic and epigenomic landscapes of lung cancer subtypes underscore the selection of different oncogenic pathways during tumor development. *PLoS One* 2012; **7**: e37775.
54 Olive PL, Banath JP. Phosphorylation of histone H2AX as a measure of radiosensitivity. *Int J Radiat Oncol Biol Phys* 2004; **58**: 331–335.

Supplementary Information accompanies this paper on the Oncogene website (http://www.nature.com/onc)

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

Tumor and Stem Cell Biology

Cancer Research

YEATS4 Is a Novel Oncogene Amplified in Non–Small Cell Lung Cancer That Regulates the p53 Pathway

Larissa A. Pikor[1], William W. Lockwood[2], Kelsie L. Thu[1], Emily A. Vucic[1], Raj Chari[3], Adi F. Gazdar[4], Stephen Lam[1], and Wan L. Lam[1]

Abstract

Genetic analyses of lung cancer have helped found new treatments in this disease. We conducted an integrative analysis of gene expression and copy number in 261 non–small cell lung cancers (NSCLC) relative to matched normal tissues to define novel candidate oncogenes, identifying 12q13-15 and more specifically the *YEATS4* gene as amplified and overexpressed in ~20% of the NSCLC cases examined. Overexpression of YEATS4 abrogated senescence in human bronchial epithelial cells. Conversely, RNAi-mediated attenuation of *YEATS4* in human lung cancer cells reduced their proliferation and tumor growth, impairing colony formation and inducing cellular senescence. These effects were associated with increased levels of p21WAF1 and p53 and cleavage of PARP, implicating YEATS4 as a negative regulator of the p21-p53 pathway. We also found that YEATS4 expression affected cellular responses to cisplastin, with increased levels associated with resistance and decreased levels with sensitivity. Taken together, our findings reveal *YEATS4* as a candidate oncogene amplified in NSCLC, and a novel mechanism contributing to NSCLC pathogenesis. *Cancer Res; 73(24); 7301–12. ©2013 AACR.*

Introduction

Lung cancer is the leading cause of cancer death worldwide. The 5-year survival rate is a mere 15% and there exists a lack of therapies to effectively treat this deadly disease. However, within the last decade, characterization of lung cancer genomes has revealed a number of genes critical to tumorigenesis, resulting in significant changes to lung cancer treatment and a subsequent increase in progression free and overall survival for a subset of these patients. These successes have prompted a search for additional driver alterations, and have identified a number of recurrently mutated genes including *TP53, CDKN2A, PTEN, NRAS, BRAF, PIK3CA, DDR2, KEAP1*, and *NRF2* as well as gene fusions encompassing *RET* and *ROS* tyrosine kinases (1–5).

In addition to somatic mutations, copy number alterations such as recurrent amplifications and deletions occur in almost all lung cancers (6, 7). DNA amplification directly contributes to oncogene activation and the promotion of tumorigenesis, particularly for tumors driven by oncogene addiction. Oncogenes amplified at the DNA level therefore make ideal therapeutic targets as unlike loss of function tumor suppressor genes, they have the potential to be targeted directly. In non–small cell lung cancers (NSCLC), recurrent amplifications of several regions activate known oncogenes. These include: 1q21.2 (*ARNT*), 3q26.3-q27 (*PIK3CA* and *SOX2*), 5p15.33 (*TERT*), 7p11.2 (*EGFR*), 7q31.1(*MET*), 8p12 (*FGFR1*), 8q24.21 (*MYC*), 12q14.1 (*CDK4*), 14q13.3 (*NKX2-1*; refs. 7–13). Despite these discoveries, roughly 50% of lung cancers harbor no known targetable alterations, highlighting the need for a better understanding of the biology underlying lung tumorigenesis (2, 5).

To identify novel oncogenes in NSCLC, we performed a large-scale integrative analysis of DNA copy number and gene expression on 261 lung tumors, spanning both major NSCLC subtypes: adenocarcinoma (AC) and squamous cell carcinoma (SqCC). Our approach was based on the rationale that oncogenes selectively amplified and biologically relevant to NSCLC tumor biology would: (i) span regions of frequent high-level amplification, (ii) undergo frequent overexpression, and (iii) exert protumorigenic functions *in vitro* and *in vivo*. Our analysis identified a recurrent amplicon at 12q15, within which we identified the candidate oncogene *YEATS domain containing 4, glioma-amplified sequence 41* (*YEATS4/GAS41*). *In vivo* and *in vitro* functional assays were performed to characterize the biologic effects and investigate the oncogenic mechanism of YEATS4 in lung tumorigenesis. Based on the frequency of *YEATS4* amplification and overexpression in NSCLC tumors and cell lines, its role in viability, anchorage independent growth, senescence, and tumor formation, we propose that *YEATS4* is novel candidate oncogene in lung cancer.

Materials and Methods

NSCLC tumor samples and cell lines

A total of 261 formalin-fixed paraffin-embedded and fresh-frozen lung tumors (169 AC and 92 SqCC) were obtained under

Authors' Affiliations: [1]Integrative Oncology, BC Cancer Research Center, Vancouver, BC, Canada; [2]National Institutes of Health, Bethesda, Maryland; [3]Department of Genetics, Harvard Medical School, Boston, Massachusetts; and [4]Hamon Center of Therapeutics, University of Texas South Western, Dallas, Texas

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Larissa A. Pikor, BC Cancer Research Center, Vancouver, BC, Canada V5Z 1L3. Phone: 604-675-8000, ext. 7703; Fax: 604-675-8232; E-mail: lpikor@bccrc.ca

doi: 10.1158/0008-5472.CAN-13-1897

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

informed, written consent with approval from the University of British Columbia-BC Cancer Research and University of Toronto Ethics Board from patients undergoing surgical resection at the Vancouver General Hospital and the Princess Margaret Hospital, Toronto, Canada (14). Tissue sections were microdissected with the guidance of lung pathologists and matched nonmalignant lung tissue obtained for a subset of the primary tumors. DNA was extracted using standard phenol–chloroform procedures. RNA was extracted from tumor and matched nonmalignant normal tissue using RNeasy Mini Kits (Qiagen) or TRIzol reagent (Invitrogen). Quality and quantity of genomic material was assessed using a NanoDrop 1000 spectrophotometer and by gel electrophoresis and/or by Agilent 2100 Bioanalyzer. Demographic information for this cohort is summarized elsewhere (14). NSCLC cell lines H1993, H1355, H226, and A549 were obtained from American Type Culture Collection and HCC4011 from Dr. A. Gazdar and fingerprinted to confirm their identity (15). All lines were cultured in RPMI-1640 medium supplemented with 10% FBS and 0.1% penicillin–streptomycin (Invitrogen). Immortalized normal human bronchial epithelial cells (HBEC) with (HBEC-KT53) and without p53 knockdown (HBEC-KT), courtesy of Dr. J. Minna, were cultured in K-SFM media supplemented with 50 ng/μL bovine pituitary extract and 5 ng/μL EGF (Invitrogen). Demographic data for the panel of cell lines used in this study can be found at http://edrn.jpl.nasa.gov/ecas/data/dataset/urn:edrn:UTSW_MutationData.

Array comparative genomic hybridization and GISTIC analysis

Copy number profiles were generated for 261 NSCLC tumors using whole-genome tiling path array comparative genomic hybridization (aCGH), and were processed as previously described (16, 17). Probes were mapped to the March 2006 (Hg18) genomic coordinates and aCGH-Smooth was used to segment and smooth $\log_2$ ratio values (18). The corresponding segments and ratio values were analyzed using the GISTIC algorithm (19) and gene pattern software (http://www.broadinstitute.org/cancer/software/genepattern/) to identify regions of significant amplification across samples. Amplification threshold of 0.8, join segment size of 2, qv threshold 0.05, and removal of the X chromosome were the settings applied for analysis.

Gene expression profiling and data integration

Gene expression profiles were generated using custom Affymetrix microarrays for a subset (35 AC and 13SqCC) of the 261 tumors that had sufficient quantity and quality material for both tumor and matched nonmalignant tissue. Data were normalized using the Robust Multichip Average algorithm in R (20). Genes were classified as over- or under-expressed if the mRNA fold change in tumors relative to matched nonmalignant was greater or less than 2-fold. Mann–Whitney *U* tests with Benjamini Hochberg correction $P < 0.05$ were used to compare expression of 12q15 genes between tumor and nonmalignant tissue in 83 AC pairs (EDRN) and determine whether increased gene dosage resulted in increased gene expression. A Spearman's correlation conducted using MATLAB software was used to determine the strength of the correlation between copy number and expression, with a coefficient >0.55 considered significant.

In vitro and *in vivo* assays were performed as previously described (21–25). Detailed information can be found in the supplemental methods.

Results

Recurrently amplified regions in NSCLC

Copy number profiles for 169 AC and 92 SqCC were generated using aCGH. Significant regions of high-level amplification ($\log_2$ ratio >0.8) were identified using the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm, which calculates significance scores by considering both the amplitude and frequency of copy number alterations (19). GISTIC analysis of all 261 samples (NSCLC) identified 3 significant regions of focal amplification: 7p11.2 ($q = 0.00075$), 8p12 ($q = 0.036$), and 12q15 ($q = 0.036$). Subtype-specific analysis revealed 2 regions of amplification across the 169 AC tumors: 12q15 ($q = 4.5 \times 10^{-5}$) and 20q13.33 ($q = 0.017$) and 6 regions across the 92 SqCC tumors: 1p34.2 ($q = 0.044$), 3q27.1 ($q = 1.4 \times 10^{-10}$), 7p11.2 ($q = 0.029$), 8p11.23 ($q = 0.0042$), 8p12 ($q = 0.0042$), and 14q13.3 ($q = 0.03$; Fig. 1A–C). Amplification of these regions has been previously described in NSCLC, indicating our tumors display patterns of alteration characteristic of lung cancer (2, 7, 26, 27).

Although none of the regions identified were common between all 3 analyses, all of the regions identified in NSCLC were also significant in a subtype-specific manner. Further examination of these amplicons revealed that known oncogenes EGFR and BRF2, both of which are known to be preferentially amplified in SqCC (21, 28), were driving selection of the 7p11.2 and 8q12 amplicons, respectively. Intriguingly, the primary target of 12q15 amplification, which is believed to be MDM2—a ubiquitin ligase that targets TP53 for proteasomal degradation, and when overexpressed results in aberrant p53 inactivation, was excluded from both the focal and wide peak boundaries. The exclusion of MDM2 from this focal region suggested that a gene other than MDM2 may be driving selection of this amplicon. This combined with the fact that all other regions harbored known oncogenes 7p11.2 (*EGFR*), 8p11.23 (*FGFR1*), 8p12 (*BRF2*), 14q13.3 (*NKX2-1*), 20q13.3 (*EEF1A*), or are known to be subtype-specific regions of amplification (1p34.2 and 3q in SqCC; refs. 2, 7) prompted us to further explore the 12q15 amplicon.

Identification of YEATS4, the target of 12q15 amplification

The peak amplified region of 12q15 spanned a 432 kb interval (68,030,736–68,462,888) and contained 7 genes; *LYZ, YEATS4, FRS2, CCT2, LRRC10, BEST3*, and *RAB3IP*, none of which have been previously implicated in lung tumorigenesis (Fig. 1; Supplementary Table S1). Based on the notion that selectively amplified oncogenes would demonstrate elevated expression, we integrated copy number and gene expression data for adenocarcinoma tumors and matched nonmalignant tissue. Because of the limited size of our dataset with both copy number and expression data, identification of the 12q15 driver gene was performed in the largest dataset available (EDRN, $n =$

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

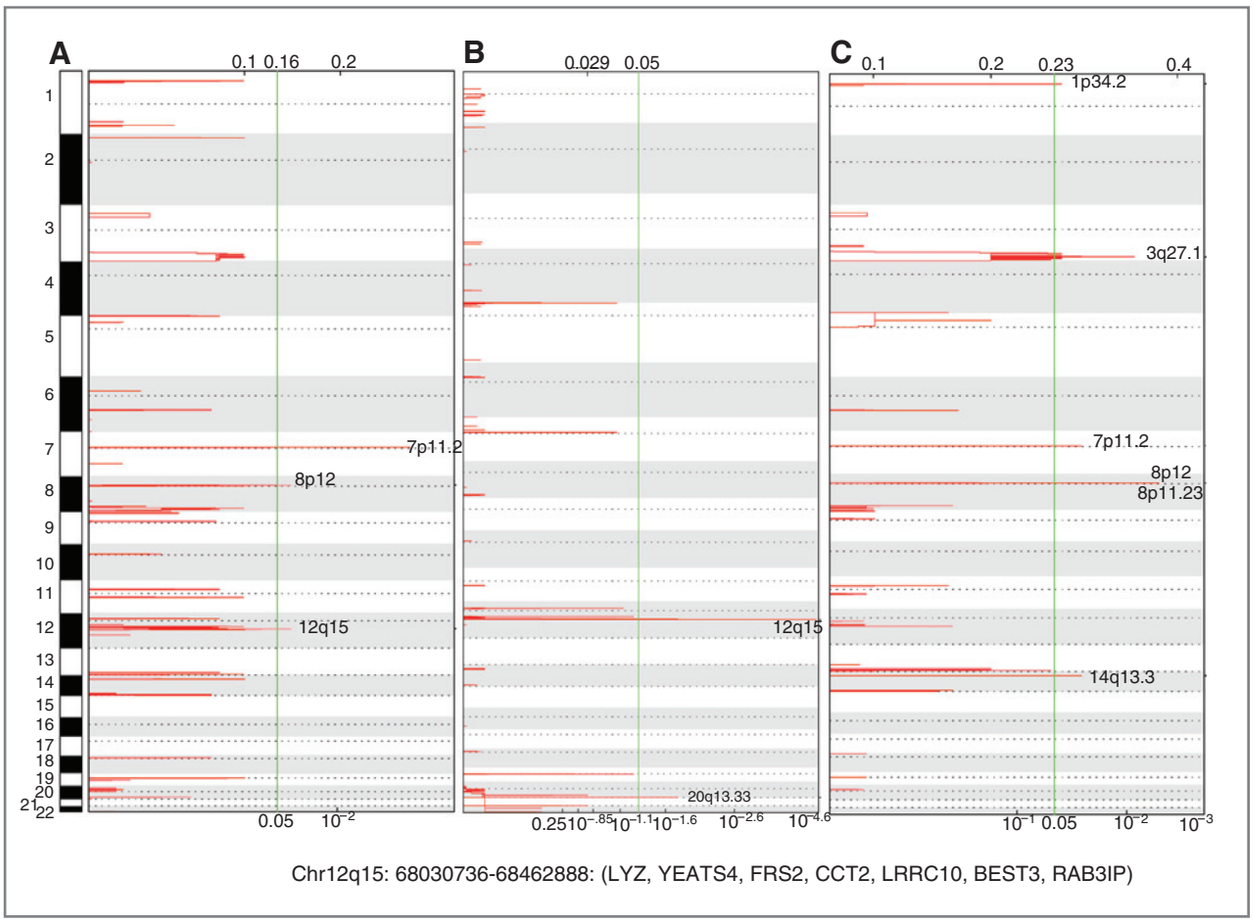


Figure 1. Recurrent amplifications in NSCLC. GISTIC plots for (A) 261 NSCLC, (B) 169 AC, and (C) 92 SqCC. Chromosomes are depicted as rows and chromosome numbers are indicated. Red peaks indicate frequently amplified regions and the green vertical line indicates the false discovery rate threshold ($q = 0.05$). Peaks extending beyond this line indicates a significant region. X-axis indicates the GISTIC score scale. Genomic coordinates and the genes located within the 12q15 amplicon are shown below.

83). Of the 7 genes within the amplicon, only *YEATS4* was both gained/amplified and concomitantly overexpressed in lung tumors relative to matched nonmalignant tissues (Fig. 2A–C). Although *YEATS4* has not been previously described in lung cancer, it is a well-established oncogene in cancers of neural origin (29, 30) and frequently amplified in liposarcomas (31).

YEATS4 is frequently amplified and overexpressed in NSCLC

YEATS4 was amplified in 18% (47/261) and overexpressed in 31% (15/48) of cases from our cohort. Although 12q15 was not significant in the GISTIC analysis of our 92 SqCC cases, to conclusively determine whether amplification of *YEATS4* was specific to AC, we compared copy number and expression data for both subtypes. Although no statistical difference in *YEATS4* copy number or expression was observed between subtypes (Supplementary Fig. S1B–S1D), on average AC tumors had a higher number of copies and greater fold change in expression compared with SqCC tumors. This suggests that although copy gain is a frequent event in both subtypes, it is likely a broader amplification event that occurs at a lower amplitude in SqCC relative to AC, which is why 12q15 failed to be identified by GISTIC in the SqCC tumors. Analysis of external datasets with both AC and SqCC data supported our findings, with gain/amplification and overexpression occurring at similar frequencies in both datasets (Table 1), indicating that amplification and overexpression of *YEATS4* is not subtype specific.

To gain further insight into the prevalence of *YEATS4* amplification, we investigated *YEATS4* copy number and expression in publicly available NSCLC tumor datasets. *YEATS4* was gained (2.3–5 copies) or amplified (>5 copies) at various frequencies across the 5 datasets, ranging from 5% to 22% and 0.4% to 5%, respectively (Table 1). A broader analysis of 508 human cancer cell lines revealed *YEATS4* copy gain/amp in 43/128 (33.6%) of lung cancer cell lines and in 117/508 (23%) of all cancer cell lines (Table 1). Expression analysis of the EDRN and TCGA datasets revealed *YEATS4* was overexpressed at comparable frequencies to our dataset; 18% (15/83) and 33% (14/42), respectively (Table 1). Taken together, these results show *YEATS4* is frequently gained and overexpressed in NSCLC, irrespective of subtype, as well as gained in many other human cancers.

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

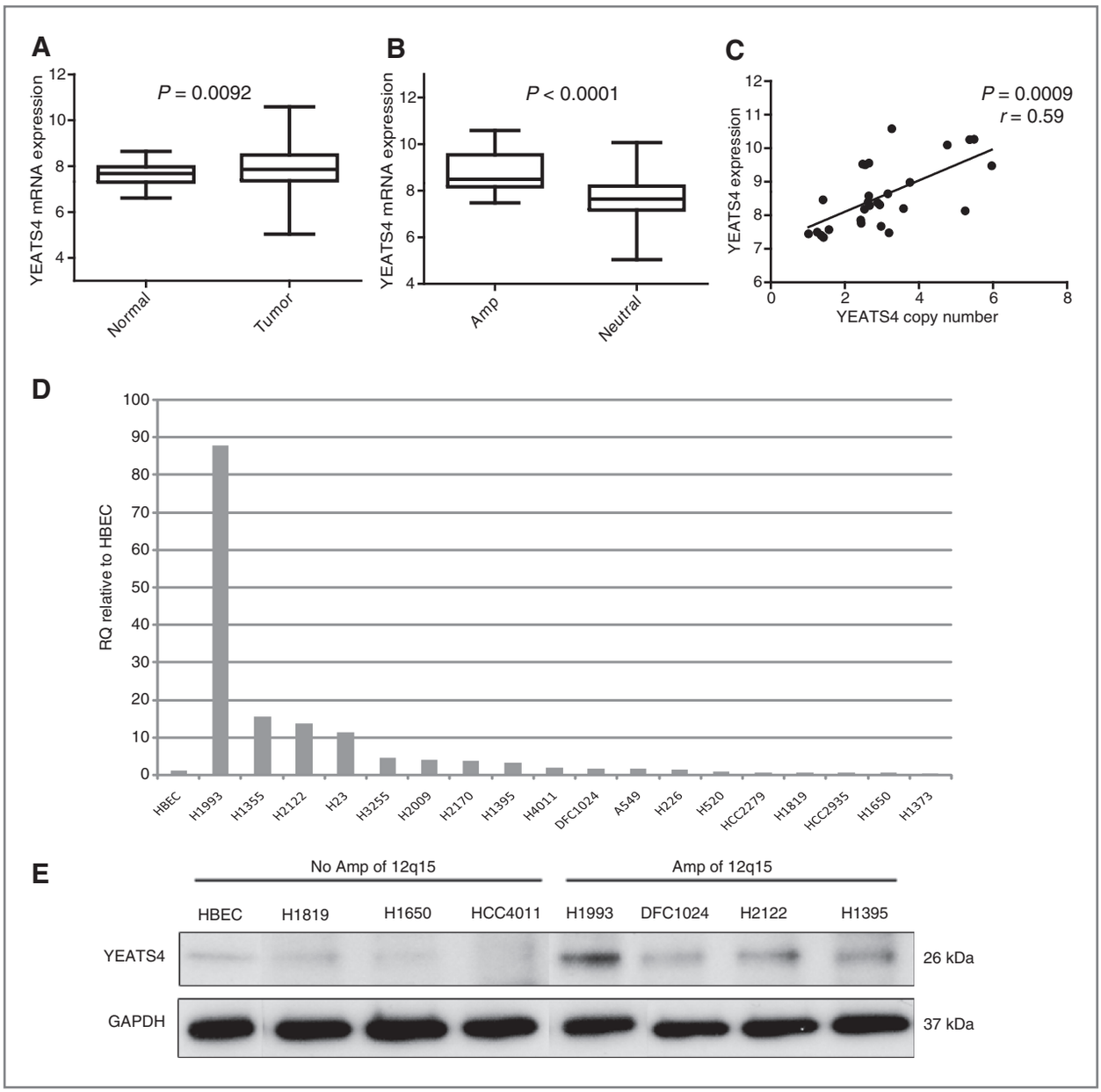


Figure 2. *YEATS4* is recurrently amplified and overexpressed in NSCLC and is the target of 12q15 amplification. A, comparison of mRNA expression in 83 AC tumors and matched nonmalignant tissue from the EDRN ($P = 0.0092$). B, *YEATS4* expression between tumors with gain/amplification and tumors with neutral copy number ($P < 0.0001$). C, Spearman correlation of copy number and expression for tumors with copy number alterations of *YEATS4* ($r = 0.59$, $P = 0.009$). Expression values for all plots are in $\log_2$ units. D, RT-qPCR of *YEATS4* expression in 18 NSCLC cell lines and nonmalignant HBEC cells. E, immunoblot of YEATS4 in NSCLC lines with and without amplification of 12q15 with GAPDH as a loading control.

To validate array findings and verify *YEATS4* is upregulated at the transcript level, we assessed *YEATS4* expression by quantitative reverse transcriptase PCR (RT-qPCR) in a panel of 59 lung ACs relative to matched nonmalignant tissue and in 18 NSCLC cell lines (2 SqCC and 16 AC) with reference to an immortalized normal human bronchial epithelial (HBEC) line. A total of 15 of 59 (25.4%) tumors and 8 of 18 (44.4%) cell lines showed a 2-fold or greater increase in *YEATS4* expression relative to their matched control (Fig. 2D; Supplementary Fig. S1A). Moreover, analysis of the 35 AC samples with expression data revealed a strong correlation between array findings and PCR results ($r = 0.75$, $P < 0.001$, Pearson correlation, data not shown), validating array findings and confirming frequent overexpression of *YEATS4*. Western blotting of cell lines with and without *YEATS4* amplification revealed increased YEATS4 expression in lines with amplification, demonstrating that amplification drives overexpression at both the mRNA and protein level (Fig. 2E).

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

Table 1. Frequency of YEATS4 gain and amplification across multiple datasets

Dataset	Histology	Samples with CN (*n*)	Gain (*n*) (>2.3 copies, <5)	Freq gain (%)	Amplification (*n*) (>5 copies)	Freq amp (%)	Samples with expression (*n*)	Overexpression	Freq OE (%)
BCCRC	AC and SqCC	16,992	28,19	16.5, 20.1	5, 2	3.0, 2.2	35, 13	11, 4	31, 30
EDRN	AC	83	18	21.7	4	4.8	83	15	18
GSE25016	AC and SqCC	77,155	10, 22	13, 14.2	0,1	0, 0.6	N/A	N/A	N/A
dbGAP	AC	354	18	5.1	4	1.1	N/A	N/A	N/A
MSKCC	AC	199	21	10.6	N/A	N/A	N/A	N/A	N/A
Broad	NSCLC	473	28	5.9	8	1.7	N/A	N/A	N/A
TCGA	AC and SqCC	277, 201	19, 22	6.9, 10.9	15, 6	5.4, 3.0	35, 17	11, 6	32, 35
Sanger lung lines	NSCLC and SCLC	128	42	32.8	1	0.78	N/A	N/A	N/A
All Sanger cell lines	All cancers	508	109	21.5	8	1.5	N/A	N/A	N/A

Association of YEATS4 and clinical features

Multivariate analysis of *YEATS4* copy number and expression revealed no significant associations between any clinical features (age, sex, stage, smoking status, race). Survival analysis of the Director's challenge expression datasets (32) using a Cox-regression analysis revealed a trend toward poorer survival in patients with *YEATS4* amplification, however this association failed to reach statistical significance in any of the datasets examined (data not shown).

YEATS4 displays oncogenic properties *in vitro* and *in vivo*

YEATS4 encodes a protein found in a number of multisubunit protein complexes involved in chromatin modification and transcriptional regulation and has also been shown to be involved in the regulation of TP53. To assess its oncogenic potential, *YEATS4* was stably transfected into 2 immortalized HBEC lines; HBEC-KT and HBEC-KT53 (KT-YEATS and KT53-YEATS). Empty vector transfected cells were used as controls (KT-EV and KT53-EV). YEATS4 gene and protein expression was confirmed by qPCR and Western blot (Fig. 3A and B). Relative to controls, ectopic expression of YEATS4 had no effect on viability and failed to induce anchorage-independent growth in HBECs (data not shown), indicating that in immortalized normal cells YEATS4 overexpression alone is incapable of inducing colony formation. However, a dramatic inhibition of senescence in overexpressing cells relative to controls was observed in both lines (Student *t* test, $P < 0.05$; Fig. 3C and D), suggesting elevated *YEATS4* expression is capable of inducing a phenotype associated with malignant transformation.

Complimentary knockdown experiments using lentiviral shRNAs were performed in lung cancer cell lines with (H1993, H1355, H226) and without (A549, HCC4011) *YEATS4* amplification and various *TP53* backgrounds (Supplementary Table S2). Empty vector transfected cells were used as controls (PLKO) and knockdown was confirmed by qPCR and Western blotting (Fig. 4A and B). Knockdown significantly decreased cell viability in H1993 and H1355 ($P = 0.0127$ and 0.0172, respectively), both of which harbor *YEATS4* amplification and mutant *p53* (Fig. 4C), but had no effect on A549, HCC4011, or H266 lines ($P = 0.428$, 0.45, and 0.49, respectively) which do not harbor *YEATS4* amplification (A549 &H4011), or have *YEATS4* amplification with wild-type (wt) *p53* (H226; Fig. 4C). Similarly, knockdown resulted in a significant decrease in anchorage-independent colony formation in H1993 ($P = 7.26 \times 10^{-6}$) and H1355 ($P = 6.06 \times 10^{-10}$) cells, but not in A549 ($P = 0.97$), H4011 ($P = 0.21$), or H226 ($P = 0.74$) cells, indicating wt *p53* may abrogate the effect of *YEATS4* knockdown on viability and colony formation in lines with amplification (Fig. 4D). A significant increase in senescence was observed in all 3 lines with amplification; H1993 ($P = 5.71 \times 10^{-6}$), H1355 ($P = 0.0012$), and H226 ($P = 1.21 \times 10^{-13}$) as well as moderate increase in A549 ($P = 2.99 \times 10^{-7}$). No difference in senescence was observed in HCC4011 ($P = 0.06$; ref. Fig. 4E). The finding that A549 cells showed a modest increase in senescence is not surprising given the role of YEATS4 in the p53 pathway (discussed below) and the wt p53 background of

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

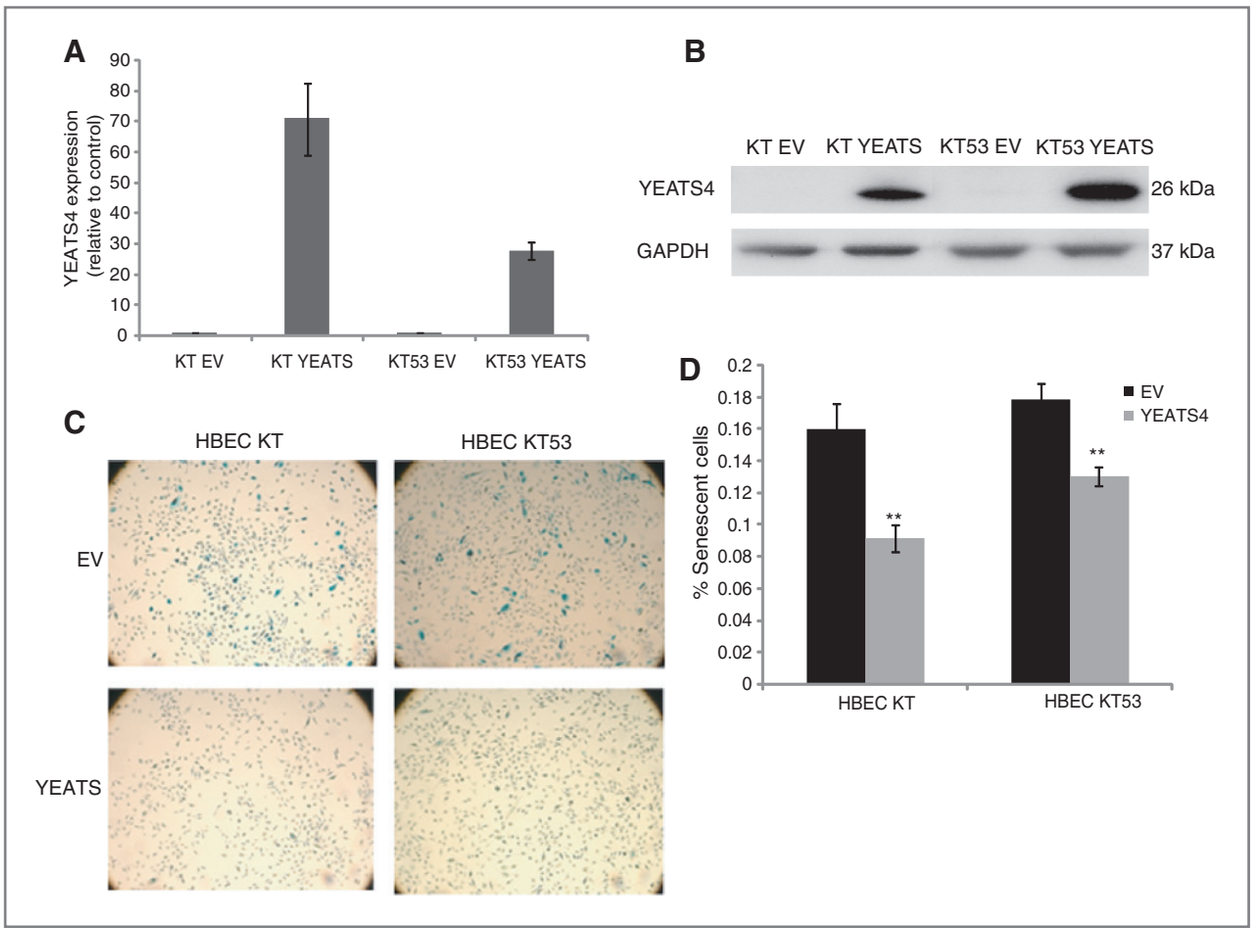


Figure 3. Overexpression of YEATS4 induces a malignant phenotype. Ectopic expression of YEATS4 increases (A) mRNA expression (mean ± SEM of triplicate replicates) and (B) protein levels relative to EV controls. GAPDH was used as a loading control. C, β-Gal staining for cellular senescence in EV and YEATS4 expressing HBECs. Cells stained blue indicate senescence. Original magnification, 10×. D, quantification of cellular senescence in YEATS4 and control cells. The mean of the proportion of senescent cells (senescent cells/total cells) for YEATS4 and EV lines is shown for triplicate experiments ± SEM. $^{**}P < 0.01$, Student *t* test.

this line, which enables pathway activation and cellular senescence.

To explore the oncogenic potential of *YEATS4 in vivo*, tumor formation in NOD/SCID mice was examined by subcutaneous flank injections of H1993 and H1355 control and shY4 cells. Tumor formation was significantly reduced in shY4 cells of both cell lines at all time points (Fig. 4F and G). Our results demonstrate that knockdown of *YEATS4* in cell lines with amplification effectively inhibits tumorigenesis, with a significant inhibition in viability, tumor and anchorage-independent growth, and increased cellular senescence, strongly supporting *YEATS4* as an oncogene in NSCLC.

YEATS4 suppresses p53 and p21

Inactivation of the p53 pathway is one of the most frequent alterations in lung cancer, with somatic mutations occurring in approximately 50% of all cases (28, 33). *p53* is a key tumor suppressor that regulates cell cycle, DNA repair, apoptosis, and senescence and inhibits aberrant proliferation and the propagation of damaged cells. A study by Park and colleagues showed that under normal, unstressed conditions, *YEATS4* binds to and inhibits the promoters of *p14* and *p21*, subsequently repressing the p53 tumor suppressor pathway (34). To assess whether this interaction occurs in NSCLC, we assessed these proteins in cell lines with *YEATS4* manipulation. Upon *YEATS4* knockdown, p21 and p53 protein levels were increased, with the greatest increases in expression of p21 and p53 observed in cell lines harboring *YEATS4* amplification or wt *p53*, respectively (Fig. 5A). No change in p14 levels was observed upon knockdown. Overexpressing lines showed a modest reduction of p21 and p14 as well as a reduction of p53 levels in HBEC-KT (Fig. 5B). MDM2 levels remained unchanged following knockdown or overexpression of *YEATS4*, indicating that the observed changes in p21, p14, and p53 were a direct result of *YEATS4* manipulation.

YEATS4 alters the sensitivity of cell lines to cisplatin and nutlin

To determine whether the downstream effects of *YEATS4* manipulation alters cellular sensitivity to chemotherapy, cell

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

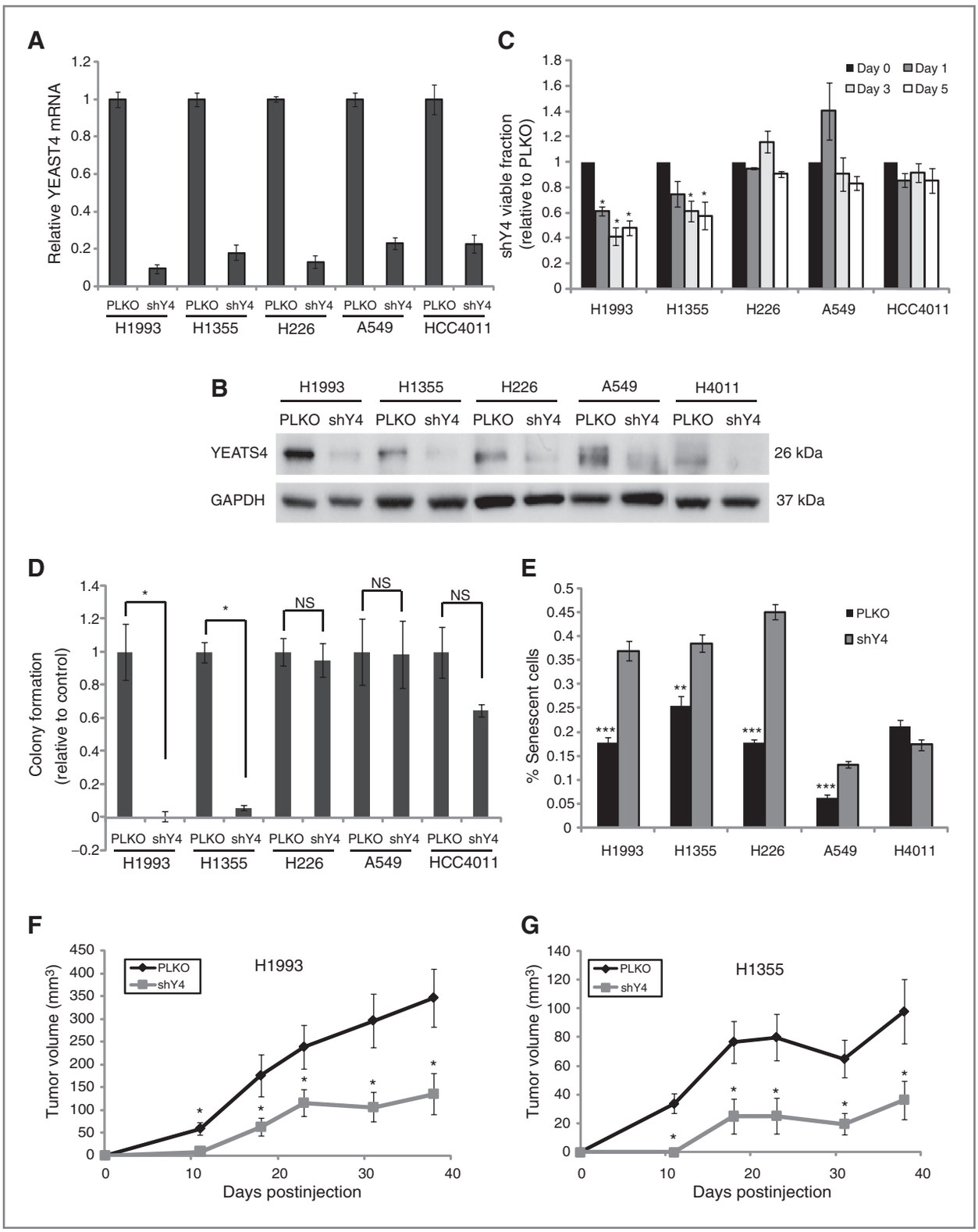


Figure 4. YEATS4 knockdown impairs growth and induces senescence. shRNA targeting YEATS4 significantly reduces (A) mRNA expression and (B) protein levels in all cell lines relative to controls (PLKO). GAPDH was used as a loading control. C, viability of cell lines with knockdown (shY4) relative to controls as measured by MTT. D, colony formation ability of shY4 cell lines relative to controls. E, quantification of cellular senescence based on β-Gal staining. Values reported as mean ± SEM of triplicate experiments. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, Student t test of shY4 cells relative to PLKO. F, G, effect of YEATS4 knockdown on tumor growth in mice injected with H1993 or H1355 PLKO and shY4 cells. Error bars indicate SEM of each group of 10 mice, $^{*}P < 0.05$.

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

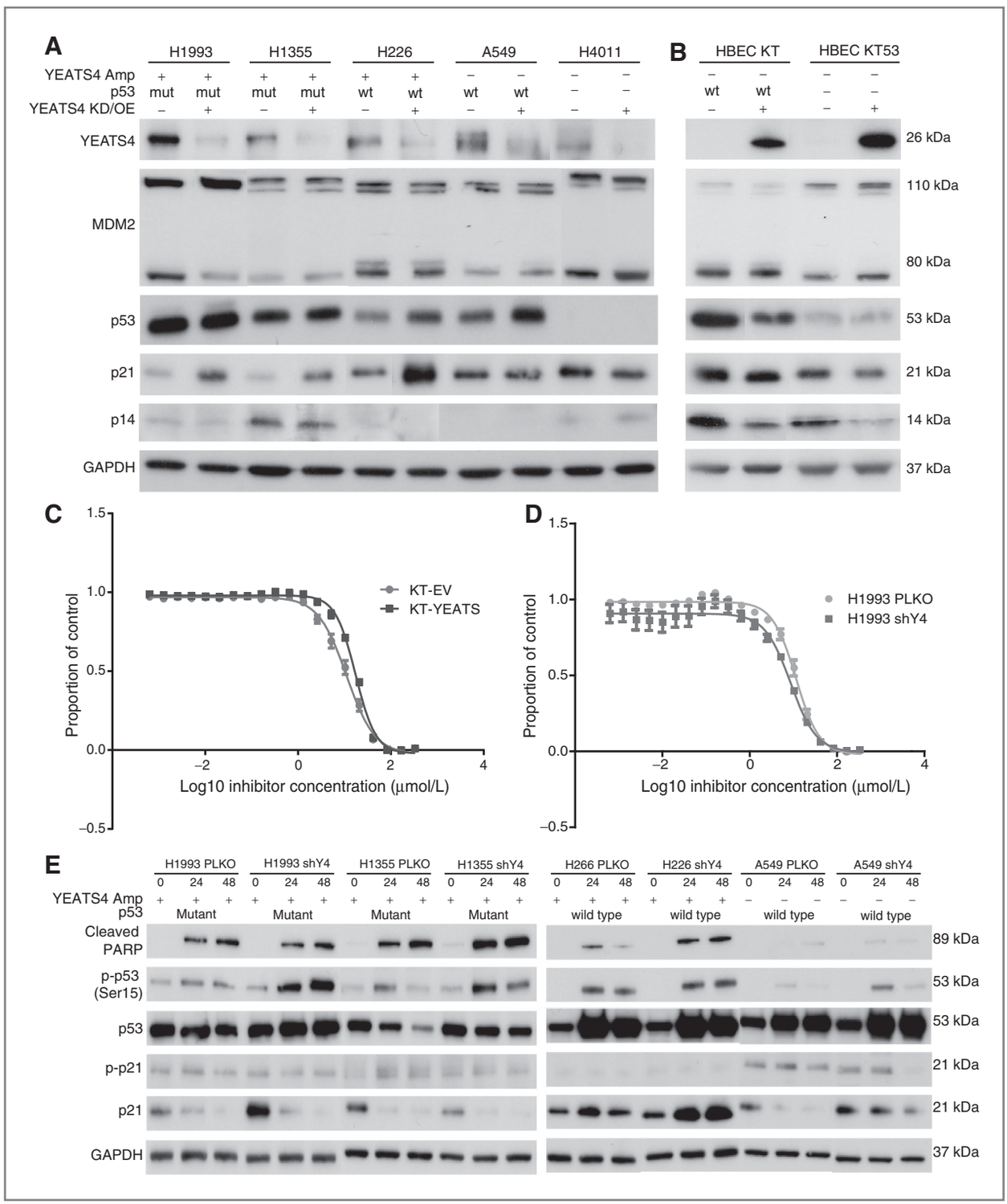


Figure 5. YEATS4 alters p21 and p53 protein levels. A, knockdown of YEATS4 increases expression of p21 in cell lines with YEATS4 amplification and increases p53 in all lines that express p53. B, overexpression of YEATS4 reduces p14 and p21 levels in both HBEC lines, and p53 only in the HBEC KT line. Dose–response curves of HBEC KT (C) and H1993 (D) cells treated with 2-fold dilutions of cisplatin for 72 hours. Viability is shown as a proportion of treated cells against untreated controls (mean ± SEM of triplicate experiments). E, immunoblot of PLKO and shY4 cell lines treated with 40 μm of cisplatin for 0, 24, or 48 hours. Cisplatin treatment induces apoptosis as measured by the increase in cleaved PARP, p53, and phosphorylated p53 (Ser15). GAPDH was used as a loading control for all blots.

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

Table 2. Cisplatin and nutlin IC_{50}s

		Cisplatin			Nutlin			
Cell line	**YEATS4**	**IC_{50}**	**SEM**	***t* Test**	**IC_{50}**	**SEM**	***t* Test**	**Trend**
H1993	PLKO	11.45	1.11					
H1993	shY4	8.649	0.488	0.004				Sensitive
H1355	PLKO	9.111	0.491					
H1355	shY4	15.91	0.905	1.6E-06				Resistant
H226	PLKO	5.788	0.276		3.266	0.130		
H226	shY4	9.965	0.716	0.0003	4.325	0.376	0.033	Resistant in both
A549	PLKO	10.88	0.452		7.58	0.316		
A549	shY4	11.4	0.705	0.204	6.908	0.123	0.084	Not significant in both
H4011	PLKO	8.952	0.326					
H4011	shY4	10.32	0.566	0.055				Not significant
HBEC KT	EV	11.09	1.472		16.55	1.405		
HBEC KT	YEATS	17.32	0.696	0.007	26.01	2.696	0.023	Resistant in both
HBEC KT53	EV	15.41	1.612		19.63	1.367		
HBEC KT53	YEATS	20.38	0.718	0.029	24.85	1.385	0.034	Resistant in both

lines were treated with serial dilutions of cisplatin, a commonly prescribed first line chemotherapy for lung cancer patients that crosslinks DNA triggering apoptosis, or nutlin, a *cis*-imidazoline analog that inhibits the interaction of p53 and MDM2, stabilizing p53. Based on the observed effects on p53 and p21 protein levels following manipulation of *YEATS4* expression and the notion that cells with *YEATS4* amplification may be dependent on *YEATS4* for growth and survival, we hypothesized that HBEC-KT/KT53-Y cells would be more resistant to treatment, whereas shY4 cells harboring *YEATS4* amplification would be more sensitive.

As expected, HBEC-KT-YEATS and HBEC-KT53-YEATS lines were significantly more resistant to both cisplatin and nutlin than their control counterparts (Fig. 5C; Table 2). Differences in sensitivity were less consistent in the lung cancer cell lines, likely because of the fact these cell lines harbor numerous genomic alterations that could influence drug sensitivities. Although H1993 shY4 cells were significantly more sensitive to cisplatin (IC_{50} PLKO:11.45 vs. shY4:8.65; Fig. 5D) supporting our hypothesis, knockdown in both H1355 and H226, showed the opposite trend resulting in greater resistance relative to controls (Table 2). As anticipated, A549 and HCC4011 shY4 cells showed no difference in sensitivity (Table 2). As specimens with mutant p53 are resistant to nutlin, only A549 and H226 were treated. Similar to the cisplatin results, A549 shY4 cells showed no significant difference in sensitivity to nutlin (PLKO: 7.58 vs. shY4:6.91; $P = 0.84$), whereas H226 shY4 cells were unexpectedly significantly more resistant (PLKO:3.27, shY4:4.33, $P = 0.033$; Table 2). Analysis of lung cancer cell line IC_{50} data from the Sanger drug sensitivity project failed to reveal a significant association between YEATS4 amplification and response to cisplatin or nutlin. However, based on the fact that transformed bronchial epithelial cells, which harbor minimal genetic alterations, were significantly more resistant to cisplatin and nutlin following overexpression of YEATS4, and H1993 shY4 cells (which harbor the greatest amplification of *YEATS4*) were more sensitive to cisplatin compared with controls, we feel this data supports the notion that YEATS4 alters the *in vitro* sensitivity of lung cells to cisplatin and nutlin.

Sensitivity to cisplatin is not mediated solely through the p53-p21 pathway

To gain further insight into the potential mechanisms of altered sensitivity to cisplatin, we treated cell lines with 40μmol/L cisplatin for 48 hours, and collected protein lysates at 0, 24, and 48 hours posttreatment. As expected, cisplatin treatment of HBECs resulted in an increase in p53, p53 Ser15 phosphorylation (a marker of stabilization), p21, and induced apoptosis as measured by cleaved PARP. However, no differences between HBEC-EV and HBEC-YEATS cells were observed for any of the proteins examined (Supplementary Fig. S2). In shY4 cells with amplification, treatment with cisplatin led to a greater induction of p53 and phospho-p53 (Ser15), and in H226 also led to a significant increase in p21 levels relative to control cells (Fig. 5E). As no significant differences in protein levels between HBEC-EV and HBEC-YEATS were observed, despite a significant increase in resistance following overexpression, our results suggest that although the p53-p21 signaling pathway may be involved, resistance is likely mediated through the interaction of *YEATS4* with other signaling pathways.

YEATS4 knockdown phenotypes are independent of p21 signaling in mutant p53 cells

To explore the effect of increased p21 expression on the observed phenotypes following knockdown, siRNA knockdown of *CDKN1A* was performed on shY4 and PLKO cells for cell lines with *YEATS4* amplification (H1993, H1355, and H226). Knockdown of *CDKN1A* showed no effect on viability or colony formation in any of the lines (data not shown), but significantly

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

altered senescence levels in the presence of wt p53 (Supplementary Fig. S2A). *CDKN1A* siRNA reduced senescence in both H226 shY4 and PLKO cells relative to nontargeting control siRNA-treated cells, such that the percent of senescent H226 shY4-p21 cells was similar to H226 PLKO-NTC (Supplementary Fig. S2B). The findings from these experiments suggest that in a wild-type p53 background, the increase in senescence following *YEATS4* knockdown occurs in a p53-dependent manner and is the direct result of increased p21-expression. As *CDKN1A* knockdown failed to rescue viability, colony formation, and senescence in cell lines with mutant p53, these findings further support the notion that the phenotypes observed following knockdown of YEATS4 are not solely because of changes in p53-p21 signaling. Based on these findings, and the prominent role of *Rb* in senescence, we investigated whether the increased senescence following YEATS knockdown could be because of altered *Rb* signaling. We observed modest reductions in Rb Ser807/811 phosphorylation following *YEATS* knockdown, which in mutant p53 cell lines seems to be due in part to reduced levels of p27 (Supplementary Fig. S2C).

Identification of additional cellular networks regulated by YEATS4

In an attempt to gain a better understanding of other pathways *YEATS4* is involved in, we performed expression profiling on shY4 and PLKO cells for the 3 cell lines with *YEATS4* amplification. To identify significantly enriched pathways/networks and gene sets affected by *YEATS4* knockdown, Ingenuity Pathway Analysis and Gene Set Enrichment Analysis (GSEA) were performed. A total of 32 genes (27 overexpressed and 5 underexpressed) were differentially expressed between knockdown and control cells across all 3 cell lines. Because of the small number of input genes, none of the significantly enriched canonical pathways passed multiple testing correction. However, network analysis, which assesses regulatory relationships existing between genes and proteins, identified 2 networks associated with protumorigenic functions: (1) cancer and (2) cell death, survival, cell cycle, and cell morphology. These networks were centered around known targets or binding partners of *YEATS4* including *p53*, *CDKN1A*, and *MYC* (Supplementary Fig. S3), further supporting our *in vitro* findings. Preranked GSEA revealed significant enrichment of a number of transcription factor gene sets including *MYCN*, which has been shown to be a binding partner of *YEATS4* and all 6 *serum response factor (SRF)* gene sets. *SRF* is a ubiquitously expressed transcription factor implicated in cell proliferation, differentiation, metastasis, and clinically associated with castration-resistant prostate cancer (35, 36). Interestingly, *PDLIM7*, which contains a serum response element and is transcribed upon induction of *SRF*, was shown to inhibit p53 and p21 through the inhibition of MDM2 self-ubiquitination. Although neither *MYCN*, *SRF*, or *PDLIM7* were differentially disrupted at the mRNA level following knockdown, our downstream analysis suggests the target genes of these 2 transcription factors could be involved in *YEATS4*-mediated tumorigenesis and warrant investigation in future studies to elucidate additional mechanisms through which YEATS4 promotes tumorigenesis.

Discussion

Although single-dimensional genomic analyses have been instrumental in cancer gene discovery, this type of analysis often overlooks genes disrupted at low frequencies, and is unlikely to distinguish causal from passenger events. The integration of multiple parallel genomic dimensions enables the identification of genes with concurrent DNA and expression alterations, which are likely selected for because of their roles in driving cancer phenotypes (37). Toward this end, we integrated copy number and gene expression data in an attempt to identify novel oncogenes important in lung tumorigenesis. Although our analysis revealed gains/amplifications in a number of regions previously reported in NSCLC, the amplicon at 12q15 was the only one without a candidate driver gene located within the amplicon boundaries and was therefore the only regions we pursued further. Integration of expression and copy number data for the 7 genes located within 12q15 identified *YEATS4* as the candidate target gene of this amplicon.

First identified and isolated in the glioblastoma multiforme cell line TX3868, *YEATS4* is a highly conserved nuclear protein essential for cell viability that is frequently amplified in gliomas, astrocytomas, and liposarcomas (29, 31, 38). A member of a protein family characterized by the presence of an N-terminal YEATS domain, *YEATS4* shares high homology with transcription factor family members *AF-9* and *ENL* (39). Like other family members, *YEATS4* is involved in chromatin modification and transcriptional regulation through its incorporation into multisubunit complexes; specifically the human TIP60/TRRAP and SRCAP complexes (40, 41), which mediate the incorporation of an H2A variant histone protein into nucleosomes, altering chromatin structure and controlling transcriptional regulation.

In addition to its role in transcriptional regulation, yeast 2 hybrid screens have revealed a number of *YEATS4* binding partners. These include *MYC*, *MYCN*, *TACC1, TACC2*, *NuMa*, *AF10*, *PFDN1*, and *KIAA1009* (42–46). Analysis of expression data before and after *YEATS4* knockdown showed no effect on expression of any binding partners, suggesting that *YEATS4* does not control the expression of its binding partners at the mRNA level. To date, the majority of work surrounding *YEATS4* has focused primarily on the identification of *YEATS4* binding partners with only a few studies having explored the phenotypic effects of *YEATS4* amplification, none of which have been performed in lung (34, 46).

Our study is the first to show gain/amplification and overexpression of *YEATS4* in NSCLC and the first to implicate amplification of *YEATS4* in lung cancer tumorigenesis. We observed frequent gain/amplification of *YEATS4* in multiple independent tumor cohorts in addition to our own, as well as a strong correlation between gain and overexpression in both tumors and cell lines (Fig. 2). Analysis of the catalogue of somatic mutations in cancer (COSMIC) revealed *YEATS4* is rarely mutated in lung (0.23%) or any cancer type (0.17%), suggesting that DNA amplification is the predominant

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

mechanism of activation. In addition to the genomic evidence supporting selection of *YEATS4* in NSCLC, we demonstrate the oncogenic potential of *YEATS4* both *in vitro* and *in vivo* (Figs. 3 and 4). Ectopic expression resulted in a significant reduction in senescence, suggesting overexpression of *YEATS4* is sufficient to induce phenotypic changes characteristic of malignant transformation (Fig. 3), whereas knockdown in cell lines with amplification and mutant p53 showed reduced viability and colony formation along with an increase in senescence, consistent with oncogenic function. Although wt p53 abrogates the effects on viability and colony formation on *YEATS4* knockdown lines with amplification, a significant increase in senescence was still observed. In addition to these phenotypic effects, we also demonstrated that *YEATS4* inhibits p21 thereby repressing p53 activity, consistent with the findings of Park and Roeder who demonstrated this interaction in unstressed conditions (34). siRNA-mediated knockdown of *CDKN1A* failed to rescue viability, colony formation, and senescence in mutant p53 backgrounds, suggesting the phenotypic effects of *YEATS4* amplification occur through a mechanism other than p21.

MDM2, an E3 ubiquitin ligase, is the major negative regulator of p53, mediating its ubiquitination and subsequent degradation (47, 48). Overexpression results in inactivation of p53 and is a common mechanism of *p53* inactivation in cancer. *MDM2* is frequently amplified and overexpressed in human cancers including lung cancer, and is largely considered to be the driver gene of the 12q15 amplicon (7, 49). We were therefore intrigued to discover that despite being frequently gained in our dataset, MDM2 did not fall within the boundaries of the 12q15 amplicon identified in our cohort. This led us to suppose that an alternative oncogene was being selected for in this region. When looking at high-resolution copy number profiles, although the majority of cases showed identical copy number for both *YEATS4* and *MDM2*, a small number of cases (3/83) had more copies of *YEATS4* than *MDM2*, suggesting *YEATS4* is selected as the target of amplification in these samples and that amplification of *YEATS4* is not merely a passenger event of *MDM2* amplification. Of note, 4 of 83 cases had higher level gain/amplification of *MDM2* relative to *YEATS4*. For cancers with amplification of 12q15 spanning both *YEATS4* and *MDM2*, these genes may work synergistically to suppress p53, however further experimentation is required to investigate this hypothesis. Along with the many tumor promoting effects of *YEATS4*, of immediate clinical interest is our discovery of a YEATS4 dependent mechanism of reduced cisplatin and nutlin sensitivity, which seems to occur at least in part through inhibition of p21 and subsequent suppression of the p53 pathway.

In summary, we have shown *YEATS4* is frequently amplified and overexpressed in NSCLC and together with its multiple oncogenic functions suggests it is a novel oncogene in lung cancer. Given the effect of *YEATS4* on the p53 pathway, we propose that *MDM2* is not the sole driver gene targeted by amplification of 12q15 and that suppression of the p53 pathway can be achieved through amplification of *YEATS4*. Additional investigation into the signaling pathways altered as a result of YEATS4 amplification will provide further insight into the mechanism underlying *YEATS4*-mediated tumorigenesis.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: L.A. Pikor, W.W. Lockwood, R. Chari, W.L. Lam
Development of methodology: L.A. Pikor, R. Chari, A.F. Gazdar
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L.A. Pikor, A.F. Gazdar, S. Lam
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L.A. Pikor, W.W. Lockwood, K.L. Thu, E.A. Vucic, R. Chari, W.L. Lam
Writing, review, and/or revision of the manuscript: L.A. Pikor, K.L. Thu, E.A. Vucic, R. Chari, A.F. Gazdar, S. Lam, W.L. Lam
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): L.A. Pikor, S. Lam
Study supervision: L.A. Pikor, W.L. Lam

Acknowledgments

The authors thank M. Zhang and M. Suzuki for their assistance.

Grant Support

Funding for this work is provided by the Canadian Institutes for Health Research (MOP 86731, 94867), Canadian Cancer Society (CCS20485), NCI Early Detection Research Network (U01CA0864021), and the Canary Foundation. L.A. Pikor and K.L. Thu are supported by Vanier Canada Graduate Scholarship, W.W. Lockwood by CIHR Jean-Francois Saint Denis Fellowship in Cancer Research, E. A. Vucic by CIHR Frederick Banting & Charles Best Canada Graduate Scholarship, and R. Chari by Banting Postdoctoral Fellowship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 9, 2013; revised September 19, 2013; accepted October 7, 2013; published OnlineFirst October 29, 2013.

References

1. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 2008;455:1069–75.
2. Hammerman PS, Hayes DN, Wilkerson MD, Schultz N, Bose R, Chu A, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489:519–25.
3. Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, et al. KIF5B-RET fusions in lung adenocarcinoma. Nat Med 2012;18:375–7.
4. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, et al. RET, ROS1 and ALK fusions in lung cancer. Nat Med 2012;18:378–81.
5. Pao W, Hutchinson KE. Chipping away at the lung cancer genome. Nat Med 2012;18:349–51.
6. Lockwood WW, Chari R, Coe BP, Girard L, Macaulay C, Lam S, et al. DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. Oncogene 2008;27:4615–24.
7. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature 2007;450:893–8.
8. Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, et al. SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat Genet 2009;41:1238–42.
9. Hirsch FR, Varella-Garcia M, Franklin WA, Veve R, Chen L, Helfrich B, et al. Evaluation of HER-2/neu gene amplification and protein expression in non-small cell lung carcinomas. Br J Cancer 2002; 86:1449–56.

Published OnlineFirst October 29, 2013; DOI: 10.1158/0008-5472.CAN-13-1897

10. Okudela K, Suzuki M, Kageyama S, Bunai T, Nagura K, Igarashi H, et al. PIK3CA mutation and amplification in human lung cancer. Pathol Int 2007;57:664–71.
11. Onozato R, Kosaka T, Kuwano H, Sekido Y, Yatabe Y, Mitsudomi T. Activation of MET by gene amplification or by splice mutations deleting the juxtamembrane domain in primary resected lung cancers. J Thorac Oncol 2009;4:5–11.
12. Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM, et al. Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. Sci Transl Med 2010;2:62ra93.
13. Yamamoto H, Shigematsu H, Nomura M, Lockwood WW, Sato M, Okumura N, et al. PIK3CA mutations and copy number gains in human lung cancers. Cancer Res 2008;68:6913–21.
14. Lockwood WW, Wilson IM, Coe BP, Chari R, Pikor LA, Thu KL, et al. Divergent genomic and epigenomic landscapes of lung cancer subtypes underscore the selection of different oncogenic pathways during tumor development. PLoS One 2012;7:e37775.
15. Gazdar AF, Girard L, Lockwood WW, Lam WL, Minna JD. Lung cancer cell lines as tools for biomedical discovery and research. J Natl Cancer Inst 2010;102:1310–21.
16. Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, et al. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 2004;36:299–303.
17. Khojasteh M, Lam WL, Ward RK, MacAulay C. A stepwise framework for the normalization of array CGH data. BMC Bioinform 2005;6:274.
18. Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B. Breakpoint identification and smoothing of array comparative genomic hybridization data. Bioinformatics 2004;20:3636–7.
19. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A 2007;104:20007–12.
20. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249–64.
21. Lockwood WW, Chari R, Coe BP, Thu KL, Garnis C, Malloff CA, et al. Integrative genomic analyses identify BRF2 as a novel lineage-specific oncogene in lung squamous cell carcinoma. PLoS Med 2010;7:e1000315.
22. Lockwood WW, Thu KL, Lin L, Pikor LA, Chari R, Lam WL, et al. Integrative genomics identified RFC3 as an amplified candidate oncogene in esophageal adenocarcinoma. Clin Cancer Res 2012;18:1936–46.
23. Thu KL, Pikor LA, Chari R, Wilson IM, Macaulay CE, English JC, et al. Genetic disruption of KEAP1/CUL3 E3 ubiquitin ligase complex components is a key mechanism of NF-κB pathway activation in lung cancer. J Thorac Oncol 2011;6:1521–9.
24. Coe BP, Chari R, MacAulay C, Lam WL. FACADE: a fast and sensitive algorithm for the segmentation and calling of high resolution array CGH data. Nucleic Acids Res 2010;38:e157.
25. Chari R, Lonergan KM, Pikor LA, Coe BP, Zhu CQ, Chan TH, et al. A sequence-based approach to identify reference genes for gene expression analysis. BMC Med Genom 2010;3:32.
26. Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhim R, Garraway L, et al. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 2005;65:5561–70.
27. Job B, Bernheim A, Beau-Faller M, Camilleri-Broet S, Girard P, Hofman P, et al. Genomic aberrations in lung adenocarcinoma in never smokers. PLoS One 2010;5:e15145.
28. Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med 2008;359:1367–80.
29. Fischer U, Heckel D, Michel A, Janka M, Hulsebos T, Meese E. Cloning of a novel transcription factor-like gene amplified in human glioma including astrocytoma grade I. Hum Mol Genet 1997;6:1817–22.
30. Munnia A, Schutz N, Romeike BF, Maldener E, Glass B, Maas R, et al. Expression, cellular distribution and protein binding of the glioma amplified sequence (GAS41), a highly conserved putative transcription factor. Oncogene 2001;20:4853–63.
31. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet 2010;42:715–21.
32. Shedden K, Taylor JM, 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;14:822–7.
33. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat 2007;28:622–9.
34. Park JH, Roeder RG. GAS41 is required for repression of the p53 tumor suppressor pathway during normal cellular proliferation. Mol Cell Biol 2006;26:4006–16.
35. Medjkane S, Perez-Sanchez C, Gaggioli C, Sahai E, Treisman R. Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. Nat Cell Biol 2009;11:257–68.
36. Prencipe M, Madden SF, O'Neill A, O'Hurley G, Culhane A, O'Connor D, et al. Identification of transcription factors associated with castration-resistance: is the serum responsive factor a potential therapeutic target? Prostate 2013;73:743–53.
37. Chari R, Thu KL, Wilson IM, Lockwood WW, Lonergan KM, Coe BP, et al. Integrating the multiple dimensions of genomic and epigenomic landscapes of cancer. Cancer Metastasis Rev 2010;29:73–93.
38. Zimmermann K, Ahrens K, Matthes S, Buerstedde JM, Stratling WH, Phi-van L. Targeted disruption of the GAS41 gene encoding a putative transcription factor indicates that GAS41 is essential for cell viability. J Biol Chem 2002;277:18626–31.
39. Schulze JM, Wang AY, Kobor MS. YEATS domain proteins: a diverse family with many links to chromatin modification and transcription. Biochem Cell Biol 2009;87:65–75.
40. Cai Y, Jin J, Florens L, Swanson SK, Kusch T, Li B, et al. The mammalian YL1 protein is a shared subunit of the TRRAP/TIP60 histone acetyltransferase and SRCAP complexes. J Biol Chem 2005;280:13665–70.
41. Doyon Y, Selleck W, Lane WS, Tan S, Cote J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol 2004;24:1884–96.
42. Piccinni E, Chelstowska A, Hanus J, Widlak P, Loreti S, Tata AM, et al. Direct interaction of Gas41 and Myc encoded by amplified genes in nervous system tumours. Acta Biochim Pol 2011;58:529–34.
43. Lauffart B, Howell SJ, Tasch JE, Cowell JK, Still IH. Interaction of the transforming acidic coiled-coil 1 (TACC1) protein with ch-TOG and GAS41/NuBI1 suggests multiple TACC1-containing protein complexes in human cells. Biochem J 2002;363:195–200.
44. Harborth J, Weber K, Osborn M. GAS41, a highly conserved protein in eukaryotic nuclei, binds to NuMA. J Biol Chem 2000;275:31979–85.
45. Debernardi S, Bassini A, Jones LK, Chaplin T, Linder B, de Bruijn DR, et al. The MLL fusion partner AF10 binds GAS41, a protein that interacts with the human SWI/SNF complex. Blood 2002;99:275–81.
46. Schmitt J, Fischer U, Heisel S, Strickfaden H, Backes C, Ruggieri A, et al. GAS41 amplification results in overexpression of a new spindle pole protein. Genes Chromosomes Cancer 2012;51:868–80.
47. Marine JC, Lozano G. Mdm2-mediated ubiquitylation: p53 and beyond. Cell Death Differ 2010;17:93–102.
48. Manfredi JJ. The Mdm2-p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor. Genes Dev 2010;24:1580–9.
49. Momand J, Jung D, Wilczynski S, Niland J. The MDM2 gene amplification database. Nucleic Acids Res 1998;26:3453–9.