

SCIENTIFIC REPORTS



OPEN

Computer-Aided Nodule Assessment and Risk Yield (CANARY) may facilitate non-invasive prediction of EGFR mutation status in lung adenocarcinomas

Ryan Clay¹, Benjamin R. Kipp², Sarah Jenkins³, Ron A. Karwoski⁴, Fabien Maldonado⁵, Srinivasan Rajagopalan⁴, Jesse S. Voss², Brian J. Bartholmai⁶, Marie Christine Aubry² & Tobias Peikert¹

Computer-Aided Nodule Assessment and Risk Yield (CANARY) is quantitative imaging analysis software that predicts the histopathological classification and post-treatment disease-free survival of patients with adenocarcinoma of the lung. CANARY characterizes nodules by the distribution of nine color-coded texture-based exemplars. We hypothesize that quantitative computed tomography (CT) analysis of the tumor and tumor-free surrounding lung facilitates non-invasive identification of clinically-relevant mutations in lung adenocarcinoma. Comprehensive analysis of targetable mutations (50-gene-panel) and CANARY analysis of the preoperative (≤ 3 months) high resolution CT (HRCT) was performed for 118 pulmonary nodules of the adenocarcinoma spectrum surgically resected between 2006–2010. Logistic regression with stepwise variable selection was used to determine predictors of mutations. We identified 140 mutations in 106 of 118 nodules. TP53 ($n = 48$), KRAS ($n = 47$) and EGFR ($n = 15$) were the most prevalent. The combination of Y (Yellow) and G (Green) exemplars, fibrosis within the surrounding lung and smoking status were the best discriminators for an EGFR mutation (AUC 0.77 and 0.87, respectively). None of the EGFR mutants expressing TP53 ($n = 5$) had a good prognosis based on CANARY features. No quantitative features were significantly associated with KRAS mutations. Our exploratory analysis indicates that quantitative CT analysis of a nodule and surrounding lung may noninvasively predict the presence of EGFR mutations in pulmonary nodules of the adenocarcinoma spectrum.

Lung cancer remains the leading cancer-related cause of death in the United States, and continues to have a poor prognosis with a 5-year survival rate of 18%¹. As the United States Preventative Services Task Force recommendations to screen high-risk individuals for lung cancer are implemented², we will see increased identification of early-stage lung cancer. The majority of these lesions are part of the lung adenocarcinoma spectrum³.

Adenocarcinomas frequently harbor driver mutations such as Kirsten rat sarcoma (KRAS) and epithelial growth factor receptor (EGFR) mutations. The clinical significance of these mutations is based on favorable therapeutic responses of EGFR-mutated tumors to EGFR tyrosine kinase inhibitor (TKI) therapy yielding improved survival with decreased toxicity compared to standard platinum-based cytotoxic chemotherapy^{4,5}. Conversely,

¹Pulmonary & Critical Care Medicine, Mayo Clinic, Rochester, MN, USA. ²Laboratory Medicine & Pathology, Mayo Clinic, Rochester, MN, USA. ³Biomedical Statistics & Informatics, Mayo Clinic, Rochester, MN, USA. ⁴Biomedical Imaging Resource, Mayo Clinic, Rochester, MN, USA. ⁵Allergy, Pulmonary & Critical Care Medicine, Vanderbilt University, Nashville, TN, USA. ⁶Thoracic Radiology, Mayo Clinic, Rochester, MN, USA. Correspondence and requests for materials should be addressed to T.P. (email: Peikert.Tobias@mayo.edu)

harboring a KRAS mutation predicts a poorer response to EGFR-targeted TKIs⁶. Invasive tissue sampling is required to investigate the presence of a targetable mutation.

While EGFR positivity is more prevalent among women and never smokers^{7–14}, KRAS is more prevalent in smokers and former smokers. KRAS and EGFR mutations tend to be mutually exclusive. It is also unclear how these driver mutations affect the natural history of non-small cell lung cancer (NSCLC)^{15–17}. Prior research suggests that lung adenocarcinomas harboring EGFR mutations are characterized by common radiologic features, such as an increased amount of ground glass opacity (GGO) however, published data have been inconsistent, and no definitive radiological pattern has emerged^{7–14,18}. There is limited data regarding the radiologic features of KRAS-positive tumors – though an association exists between spiculation of the nodule and KRAS positivity^{18,19}. Tumors interact with the tumor-free surrounding lung tissue in multiple ways. Field effects within the lung tissue may predispose to the development of the tumor. However, tumor growth pattern, stromal reactions and the effects cytokine-mediated changes on surrounding tissue²⁰ may alter the radiological characteristics of the lung tissue around the tumor. We hypothesize that these changes vary in the presence of different driver mutations, and that these differences can be detected by quantitative CT analysis.

Radiomics refers to reproducible quantitative CT analytic features that correlate with tumor biology and behavior²¹. A radiomics-based approach allows non-invasive comprehensive volumetric characterization of the tumor and surrounding lung tissue. Compared to tissue biopsy, it is more resilient to sampling error and tumor heterogeneity and may reflect molecular changes within the tumor including driver mutations²².

The Computer-Aided Nodule Assessment and Risk Yield (CANARY) tool comprehensively analyses, voxel-by-voxel, the distributions of 9 texture-based exemplars within a nodule, as previously described²³. The exemplars are color coded as Violet (V), Indigo (I), Blue (B), Green (G), Yellow (Y), Orange (O), Red (R), Cyan (C), and Pink (P). Volumetric distributions of each exemplar are summarized in a glyph displaying the proportional makeup of the nodule. Histopathologic evaluation of adenocarcinoma for features such as lepidic growth (tumor growth along pre-existing alveolar structures) predicts improved disease free survival (DFS) after tumor resection with the best survival in adenocarcinoma *in-situ* (100% lepidic growth) compared with minimally-invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IA)^{24,25}. The distribution of the CANARY exemplars correlates well with consensus histopathology with B-C-G corresponding to lepidic growth (visually more ‘ground glass’ density) and V-I-R-O (generally more solid density) correlating with the invasive component of the tumor. Furthermore, natural clustering of these glyphs facilitates the risk stratification of lung adenocarcinomas into good (G), intermediate (I) and poor (P) survival groups independent of stage^{23,25,26}.

CANARY may add synergistic information regarding prognosis when paired with mutational analyses – and furthermore may be able to detect imaging signatures of common driver mutations, thus eliminating or reducing the need for further invasive testing to guide individualized therapy for lung cancer patients. Next-generation sequencing refers to multi-gene targeted massive parallel sequencing. Mayo Clinic Laboratories has clinically implemented a 50-gene panel that can be performed using as little as 10 ng of DNA using the Ampliseq Cancer Hotspot Panel v2 (Thermo Fischer Scientific) to amplify tumor DNA. This panel targets over 2800 possible somatic mutations within 50 cancer-associated genes facilitating individualized cancer management. Given that KRAS and EGFR are among the most clinically-relevant driver mutations, we selected these two mutations to look for CANARY signatures that could non-invasively identify these mutations.

We hypothesized that the V-I-R-O pattern will be seen more frequently in KRAS-positive tumors while B-C-G will be seen more frequently in EGFR-positive tumors. Additionally we performed quantitative textural analysis of the tumor-free surrounding lung parenchyma using Computer Aided Lung Informatics for Pathology Evaluation and Rating (CALIPER) to determine whether loco-regional lung parenchymal changes, particularly the presence of low attenuation areas and fibrosis, are predictive of EGFR or KRAS²⁷.

Material and Methods

Subject selection. In a previously-analyzed retrospective cohort of 264 clinical stage I cases of resected adenocarcinoma of the lung between January 2006 and 2009²⁵ we identified 129 adequate histopathological specimens in the Mayo Clinic tissue registry with a non-contrast preoperative high resolution computed tomography (HRCT) scan of the chest (≤ 3 months prior to resection). Clinical data including disease free survival (DFS) were collected from the Mayo Clinic electronic medical records. The study was approved by the Mayo Clinic Institutional Review Board for an informed consent waiver (protocol 14–000666). All subjects reviewed had previously consented to participate in retrospective research. All research was performed in accordance with relevant guidelines and regulations. Archival formalin-fixed, paraffin embedded (FFPE) tissue was available for all cases analyzed. Final analysis was performed on one nodule per subject.

Nodule analysis and CANARY development. The development of CANARY has been previously described²³. Briefly, an experienced thoracic radiologist (BJB) arbitrarily selected 774 regions of interest (ROI, 9×9 voxels) along the spectrum of histologically-proven lung adenocarcinomas from 37 randomly-selected tumors. The similarities between the ROIs were compared and clustered using pairwise similarity metric and affinity propagation clustering²⁸.

The location of all surgically-resected nodules was known a priori. Each selected nodule was extracted with a supervised approach using constrained region seed growing. Region growing was restricted to ground glass or reticular voxels connected to the seed voxel. After an initial mask was applied to the nodule, each nodule underwent editing, if required, by the user to ensure the nodule volume was captured in its entirety. Each voxel is analyzed and assigned the color code of the nearest exemplar (Fig. 1). Based on the distribution of the exemplar within a given nodule all nodules are assigned to one of the three ‘risk’ groups correlating with post-resection DFS^{23,25,26}. Most (89%) scans were volumetric noncontrast HRCT (less than 3 mm contiguous slices). The remaining scans were 3.75–5 mm contiguous slices. 95% had no edge-enhancing and 5% underwent a smoothing

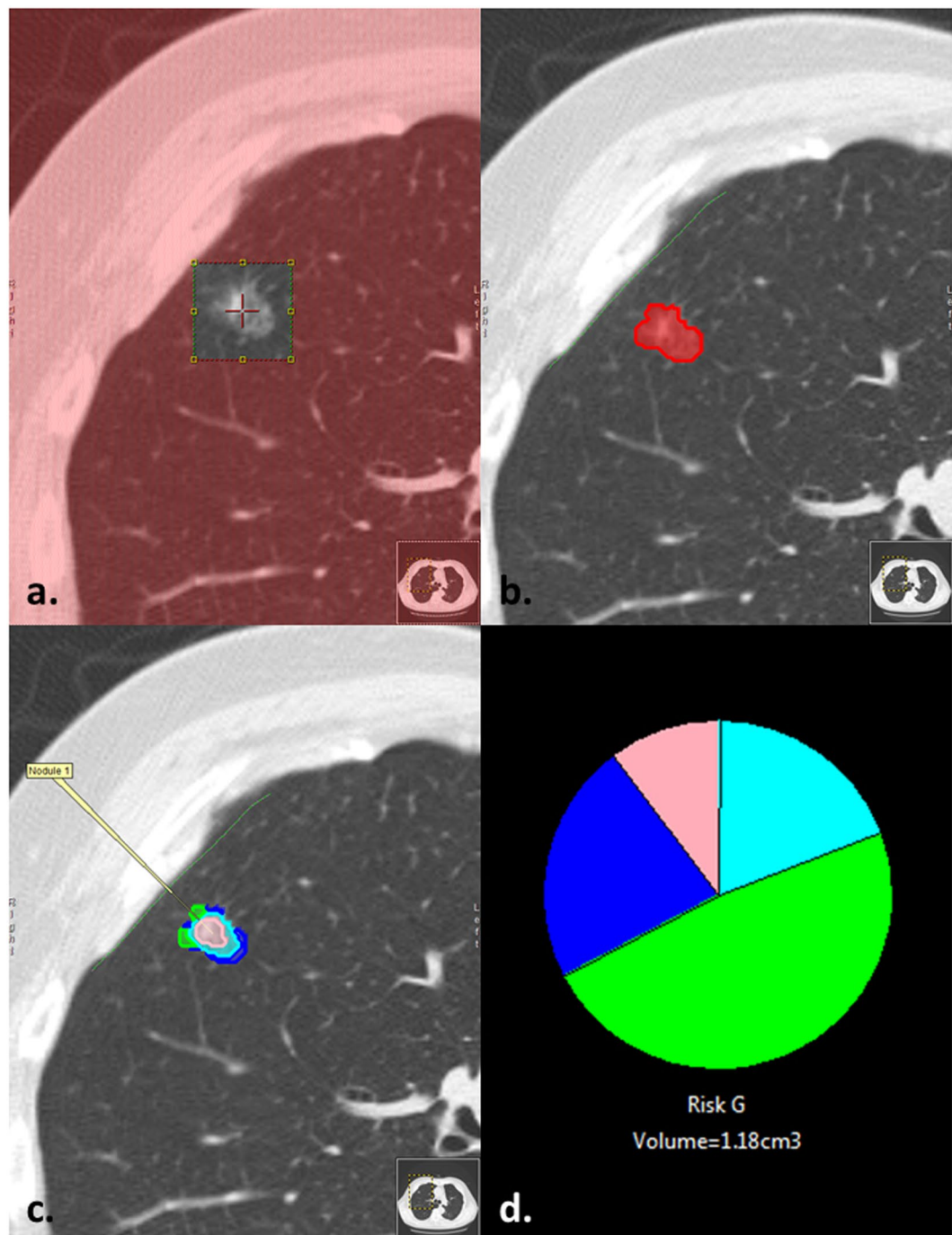


Figure 1. (a–d) Example of nodule characterization by CANARY in which the (a) user selects a seed in the center of the nodule guided by x, y and z axis, (b) a mask is generated encompassing the nodule's volume in which CANARY analysis (c) assigns each voxel the color code of the closest exemplar which is also represented by a glyph (d) displaying the relative proportion of each exemplar within a nodule.

algorithm (3×3 median filtering) to remove the kernel edge artifact to allow processing by CANARY. Data in submission (Nakajima, 2017) demonstrated excellent inter-user reproducibility of CANARY and data presented in abstract (Clay, 2017, World Congress on Thoracic Imaging) showed excellent repeatability of CANARY analysis across different acquisition techniques, slice thickness and reconstruction kernels.

Tumor-free surrounding lung analysis. CALIPER is quantitative CT analysis software that both segments the lung parenchyma classifies it into subtypes (normal (N), low attenuation (LA), ground glass (GG), reticular densities (R) and honeycomb change (HC). CALIPER development and validation is detailed previously – but in brief, radiologist-selected $15 \times 15 \times 15$ voxel volumes of interest (VOI) were allowed to cluster by affinity propagation and paired down to 5 basic clusters. These CALIPER classifications showed strong agreement with



Figure 2. Representation in red of the tumor-free surrounding lung for an adenocarcinoma in the right middle lobe. The area highlighted in red was analyzed by CALIPER for low attenuation and fibrosis shown in the axial, coronal and sagittal planes. Each nodule underwent analysis of the tumor-free surrounding lung characteristics in this manner.

radiologist classification, physiologic data and clinical phenotypes^{27,29}. CALIPER analysis of a 10 mm surrounding mask of tumor-free lung was performed. These results constituted an additional variable to consider in building a model to predict mutational status (Fig. 2).

Mutation analysis. DNA extraction was performed on archival FFPE tissue obtained at the time of initial surgery. We then performed targeted polymerase-chain reaction (PCR)-based sequencing with a 50 gene panel of common solid tumor driver mutations. DNA amplification was performed with the Ampliseq Hostspot Panel v2 (Life Technologies) to target common mutations in 50 known cancer-associated genes.

Statistical analysis. Mutation prevalence was compared with gender, prognostic categories, and smoking status using Fisher's exact tests or chi-square tests as appropriate. Age was compared with mutation status by Wilcoxon rank sum tests. Logistic regression with stepwise variable selection was used to determine the best exemplar predictors of mutations. Receiver operating characteristic (ROC) analyses were used to identify a cut-off for exemplars to achieve an 80% sensitivity to detect EGFR mutation. Post curative resection disease-free survival (DFS) was illustrated with Kaplan-Meier curves by CANARY prognosis and mutation status. Associations between DFS with prognosis, mutations, and exemplars were assessed with likelihood ratio tests from Cox proportional-hazards regression models. All p-values were two-tailed, and p-values less than 0.05 were considered statistically significant. Analyses were performed using SAS version 9.4 (copyright 2002–2012 by SAS Institute Inc., Cary, NC) and R (2014, R Foundation for Statistical Computing, Vienna, Austria).

Results

DNA was successfully extracted and analyzed in 118 of the 129 cases. Patient demographics are summarized in Table 1. 106 of the 118 nodules had at least one mutation detected and a total of 140 mutations were identified. 47 tumors harbored the KRAS mutation while 15 tumors harbored the EGFR mutation. These two mutations were mutually exclusive. Of the 15 EGFR mutants, 6 were the L858R point mutation in exon 21 and the rest were exon 19 deletions. Additional identified mutations included TP53 (n = 48), STK11 (n = 11), BRAF (n = 4), ATM (n = 3), PTEN (n = 3), PIK3 (n = 2), SMAD4 (n = 1), MET (n = 1), APC (n = 1), GRAS (n = 1), CDK2NA (n = 1), RB (n = 1), and PTPN11 (n = 1). EGFR mutations were more common among never smokers, while KRAS was more frequently mutated in current and former smokers ($p < 0.0001$, $p = 0.02$, respectively). There

Demographics	n = 118
Age at diagnosis: median years (range)	68 (35–91)
Gender n (%)	
Women	65 (55%)
Men	53 (45%)
Smoking n (%)	
Current	27 (23%)
Former	77 (65%)
Never	14 (12%)
Pathologic TNM stage n (%)	
I	94 (80%)
II	12 (10%)
III	11 (9%)
IV	1 (1%)

Table 1. Patient demographics and pathological stage.

	EGFR positive (n = 15)	Wild type (n = 103)	p
Gender			0.17
Male	4 (26.7%)	49 (47.6%)	
Female	11 (73.3%)	54 (52.4%)	
Age at diagnosis			0.11
Years: median (range)	73 (42–89)	68 (35–91)	
Smoking status			<0.0001
Never	8 (53.3%)	6 (5.8%)	
Former	7 (46.7%)	70 (68.0%)	
Current	0 (0.0%)	27 (26.2%)	
Nodule volume (cc)	2.9 (1.8–6.0)	2.1 (1.1–3.5)	0.192
CANARY prognosis			0.16
G	2 (13.3%)	9 (8.7%)	
I	12 (80.0%)	65 (63.1%)	
P	1 (6.7%)	29 (28.1%)	

Table 2. Characteristics of EGFR mutants. Nodule volume displayed as median + interquartile range. P values calculated by chi square or Fischer exact for <5 counts and Wilcoxon Rank Sum for continuous variables.

was no significant gender or age difference between patients with EGFR or KRAS mutations versus the wild type. (Tables 2 and 3) There was no significant difference in median nodule volume between EGFR versus wild type tumors ($p = 0.192$).

CANARY analysis was performed on each nodule generating a representative glyph that shows the proportional distribution of the 9 CANARY exemplars (Fig. 3). This data represents a subset of previously reported data²⁵. While CANARY prognostic categories Good (G), intermediate (I) and Poor (P) predicted disease free survival ($p = 0.002$) independent of stage, there were no statistical DFS differences based on EGFR, KRAS or any detected mutation by our 50 gene panel using Kaplan Meier analysis ($p = 0.26, 0.48, 0.78$, respectively, Fig. 4). While CANARY prognosis categories did not differ significantly between tumors with EGFR and KRAS mutations ($p = 0.16, p = 0.06$, respectively), we found that an increase of the V-I-R-O or a decrease of the Y-P component (which is negatively correlated with V-I-R-O correlation = -0.78) within a tumor was associated with a lower likelihood of EGFR positivity ($p = 0.01$ for V-I-R-O (area under the curve (AUC) = 0.70), $p = 0.02$ for Y-P (AUC = 0.68)). Each 10% decrease of V-I-R-O component per nodule was associated with a 23% increase in the odds of containing an EGFR mutation (OR = 1.23, 95% CI 1.04–1.46). In contrast the B-C-G exemplars did not significantly affect the odds of harboring an EGFR mutation ($p = 0.16$). Using receiver operating characteristic (ROC) analysis, we identified that a cut-off for V-I-R-O of $\leq 71\%$ tumor volume identifies EGFR mutations with a sensitivity of 80% and a specificity of 52% (AUC = 0.66). Similarly, since V-I-R-O and Y-P are strongly negatively correlated, we identified a cut-off for Y-P of $\geq 23.5\%$ tumor volume with a sensitivity of 80% and specificity of 53% (AUC = 0.67).

EGFR-positive tumors also had significantly less fibrosis (summed GG + R + HC) and low attenuation areas in the tumor-free surrounding lung, ($p = 0.007$ and 0.001 , respectively).

Using logistic regression and stepwise variable selection to choose among the 9 individual exemplars, none were found to be significant in predicting KRAS positivity. Using the same methods to analyze the relationship between the exemplars and EGFR status we found both Y ($p = 0.002$) and G ($p = 0.008$) to be significant and that the odds of harboring an EGFR mutation increase as the percentage of Y and G in a nodule increase. Using

	KRAS positive (<i>n</i> = 47)	Wild type (<i>n</i> = 71)	<i>p</i>
Gender			0.43
Male	19 (40.4%)	34 (47.9%)	
Female	28 (59.6%)	37 (52.1%)	
Age at diagnosis			0.48
Years: median (range)	68 (46–82)	69 (35–91)	
Smoking status			0.02
Never	1 (2.1%)	13 (18.3%)	
Former	36 (76.6%)	41 (57.7%)	
Current	10 (21.3%)	17 (23.9%)	
Nodule volume (cc)	2.5 (0.2–22.4)	2.0 (0.2–14.5)	0.19
CANARY prognosis			0.06
G	7 (14.9%)	4 (5.6%)	
I	25 (53.2%)	52 (73.2%)	
P	15 (31.9%)	15 (21.1%)	

Table 3. Characteristics of KRAS mutants.

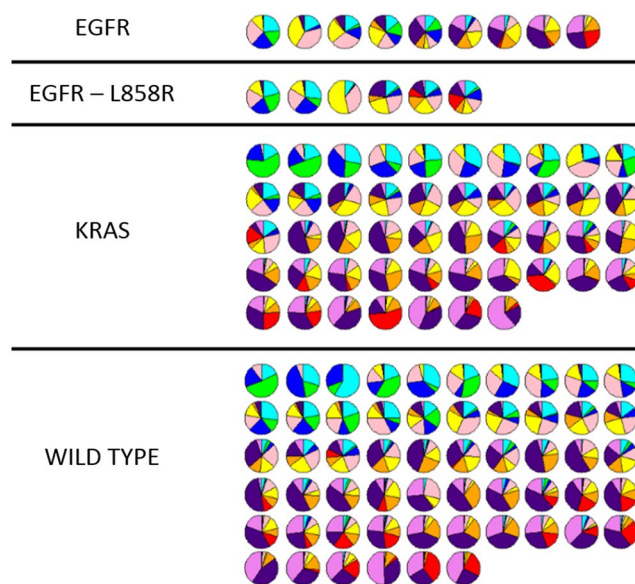


Figure 3. CANARY glyphs representing each unique nodule demonstrate proportionate representation of each CANARY exemplar. Glyphs are arranged by mutation status from left to right in order of parametric signatures correlating with progressively more invasive histopathology. Wild type denotes WT status for both EGFR and KRAS.

recursive partitioning as an additional method to determine the best discriminator of EGFR status, the best predictor was found to be the Y exemplar. In univariate performance, the Y exemplar had an AUC of 0.69. Using a combined Y and G model increased the AUC to 0.77. Adding smoking status to the Y + G model increased the AUC to 0.85. Y retained significance in this model ($p = 0.003$) whereas G did not ($p = 0.06$). Y and smoking status remained significant in this model (OR = 1.09, $p = 0.05$; and OR = 20.0, $p = 0.0001$, respectively). Neither fibrosis nor the G exemplar were significant in this model ($p = 0.09$, 0.22, respectively). Fibrosis and low attenuation were not combined in a model due to their high correlation of 0.95 – which would add to model instability. The degree of fibrosis found in the tumor-free surrounding lung was also predictive of EGFR status with an AUC 0.72. A combined model of smoking status, Y + G and CALIPER-determined fibrosis predicted having an EGFR mutation with an AUC of 0.87. Both fibrosis and smoking status retained significance in this model ($p = 0.04$, $p < 0.0001$, respectively).

TP53 mutations were detected in 40.7% of our study. None of the TP53 mutant nodules were classified as a good (G) by CANARY ($p = 0.002$) and all 5 of the 15 EGFR-mutant tumors with concurrent TP53 mutations were classified as intermediate (I) by CANARY.

We performed additional exploratory analysis comparing the CANARY exemplars of tumors with the L858R EGFR point mutation ($n = 6$) to the other EGFR mutants ($n = 9$) and wild type cases ($n = 103$). While the median V-I-R-O component was lower in the L858R group compared with other EGFR mutants (17.6% versus 49.1%),

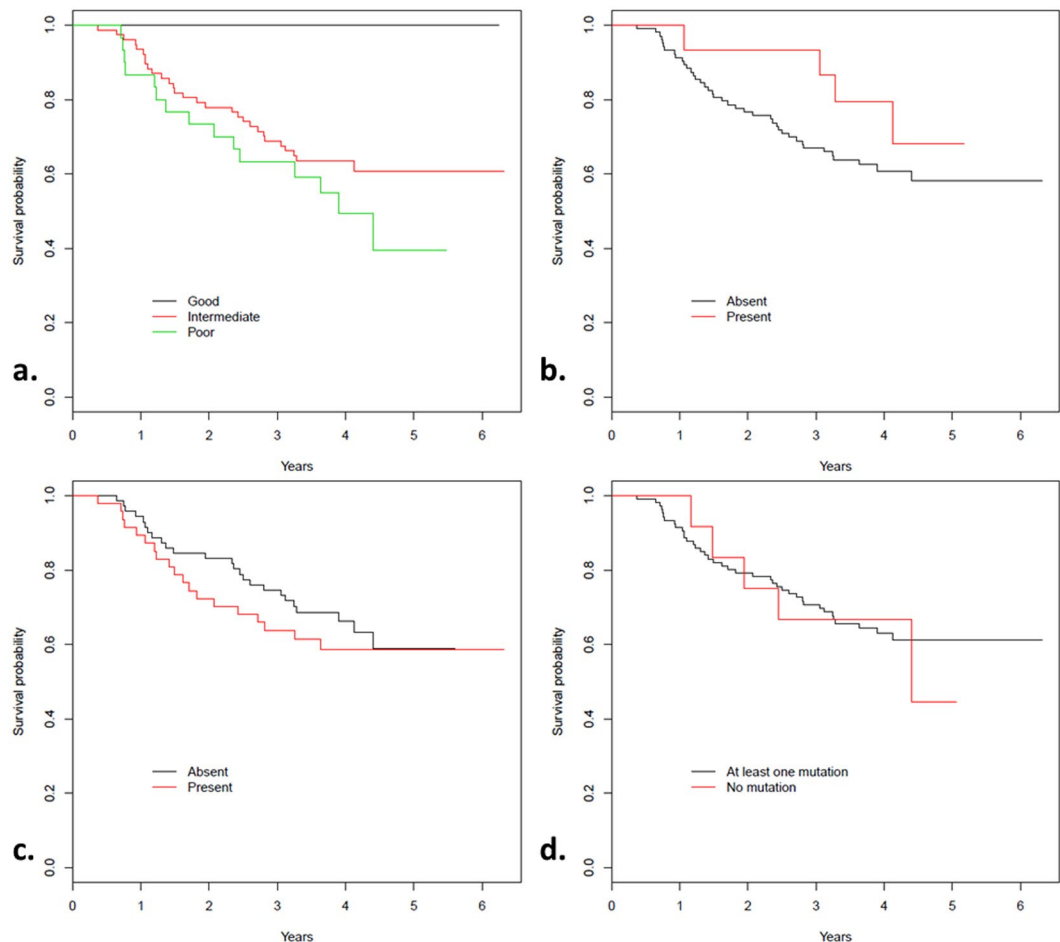


Figure 4. (a–d) Kaplan Meier survival curves depict likelihood of survival by (a) CANARY prognosis ($p = 0.002$), (b) presence of EGFR mutation ($p = 0.26$), (c) presence of KRAS mutation ($p = 0.48$) or (d) presence of any driver mutation ($p = 0.78$).

this difference did not reach statistical significance ($p = 0.38$). However there was significantly less V-I-R-O component among the L858R mutants than the wild type cases (17.6% versus 73.0%, $p = 0.02$). Additionally there was no difference in DFS between these groups ($p = 0.71$).

Discussion

Our study indicates that CANARY, especially the absence of V-I-R-O and the presence of Y-G exemplars within HRCT-imaged adenocarcinoma of the lung may noninvasively predict the presence of an EGFR mutation. This prediction is strengthened by analysis of the tumor-free surrounding lung. Radiomics features may become valuable adjuncts to patient care especially since these features (CANARY exemplars) have been proven to be more predictive of post-resection DFS when compared with EGFR or KRAS mutation status alone^{25,26}. Currently the American College of Pathologists, International Association for the Study of Lung Cancer, the Association for Molecular Pathology and other major organizations recommend the routine testing of targetable molecular abnormalities for all lung adenocarcinomas³⁰. This approach requires an invasive tissue biopsy, exposing patients to iatrogenic complications. In addition, molecular testing of invasive tissue samples or resected tumor specimens typically only includes a minute portion of the tumor and is susceptible to sampling error and tumor heterogeneity. Furthermore, small samples may be insufficient to perform these ancillary studies, potentially resulting in the need to re-expose patients to invasive procedures to obtain adequate material^{31,32}.

Consistent with prior literature, we found that KRAS correlated with increased tobacco exposure while EGFR correlated with decreased tobacco exposure³³. This may explain why we saw increased low attenuation lung (emphysema) surrounding the non-EGFR-mutated tumors. Invasion of fibroblasts into the tumor-free surrounding lung driven by carcinogenic cytokines is thought to facilitate tumor growth and invasion²⁰. Our finding of increased fibrosis in the tumor-free surrounding lung of wild-type adenocarcinoma, particularly in a tobacco-exposed cohort, may represent the radiologic correlate of this phenomenon. Our study had a low incidence of EGFR mutations – though this may be due to our North American patient population and not specifically enriching our cohort for EGFR mutations. The majority of our subjects were current or former smokers, also

lowering the likelihood of harboring an EGFR mutation and increasing the likelihood of KRAS mutations. We did not have any ALK mutations, likely due to our small n and its relative infrequency in NSCLC^{10,34}.

Non-invasive comprehensive analysis of the tumor volume using cross-sectional imaging carries a decreased risk of morbidity compared to biopsy and can account for tumor heterogeneity. Other investigators previously demonstrated a number of clinical-radiological characteristics and more recently described radiomic features of the tumor can predict the presence of molecular abnormalities specifically EGFR mutations^{7-9,11-14,17,22,35}. These studies, however, are quite heterogeneous and make use of different imaging modalities such as positron-emission tomography – whereas the CANARY exemplars are robust in their reproducibility and have been previously validated among large datasets^{25,26}.

EGFR has multiple possible mutations, however the Exon 21 L858R point mutation and Exon 19 deletion account for the majority of EGFR mutations in NSCLC³⁶. The L858R EGFR mutation has distinct clinical behavior compared with the 19 Exon deletion, with a lower likelihood of responding to EGFR-targeted therapy and worse DFS³⁷. Lee and colleagues recently reported that L858R EGFR mutant cases may have unique radiologic and histopathologic features, specifically increased percentage of GGO within the nodule, and a predominantly lepidic pattern of growth, compared with cases carrying other EGFR mutations⁸. Other studies yielded mixed results regarding the radiologic characteristics of tumors harboring the L858R mutation^{7,12}. Although we observed a trend towards less V-I-R-O among the small group (n = 6) of L858R EGFR mutant cases, the difference did not reach statistical significance and additional cases are needed to explore this association. EGFR has diverse mutations and as a whole is not clearly tied to outcome in adenocarcinoma of the lung³⁸. We did not find a relationship between BCG and EGFR status as originally hypothesized – but rather the individual exemplars Y and G. BCG is tied to indolent histopathology and good prognosis in adenocarcinoma of the lung^{23,25}, so the lack of a relationship between EGFR and BCG makes sense. The histopathologic correlate of Y and G is not entirely clear—a mix of an exemplar associated with good prognosis (G) and associated with intermediate prognosis (Y) – however EGFR does not clearly impact prognosis either³⁸. Perhaps this mix of exemplars suggests increased tumor heterogeneity in EGFR-mutated tumors – but this needs further determination.

There was a notable absence of TP53 mutation in any lesions classified as good (G) by CANARY (DFS = 100%). This finding correlates with the known association of TP53 with poor outcomes in lung adenocarcinoma and unfavorable response to EGFR-TKIs^{39,40}, and may have influenced the CANARY signature in the 5 EGFR mutant cases harboring concurrent TP53 mutations.

Limitations of our study include its retrospective nature, single center design and the relatively small number of cases. The small number of EGFR mutations and other less common molecular changes may make it difficult to detect differences between mutation subtypes. We are currently planning a larger multicenter study to mitigate these limitations. This will also allow us to evaluate the radiological features of tumors with less common molecular abnormalities such ALK translocations, which have been demonstrated to have less GGO¹⁴. Though the definitive treatment in early stage lung cancer is resection, applying quantitative CT analysis tools to late stage cancer may facilitate therapy choice without the need for a biopsy, and its application to early stage cancer could open the door to explore additional adjuvant therapy.

Conclusions

In conclusion, the distribution of density-based CANARY exemplars and quantitative CT analysis of the immediate tumor-free surrounding lung may predict the presence of EGFR mutations in lung adenocarcinomas. As volumetric CT-based CANARY analysis is non-invasive and accounts for tumor heterogeneity, CANARY may prove to be a useful radiologic biomarker beyond its validated role for non-invasive histological assessment and stratification of lung adenocarcinomas. The specified prediction cutoffs found in our trial and whether radiomics features can predict tumor response to molecularly-targeted therapy deserves future study.

CANARY software availability. CANARY software is currently licensed to Imbio LLC (Minneapolis, MN). The software is available through Imbio by request. In addition the Mayo Clinic has been sharing this software with interested research collaborators by request and we hope to expand this process. Please address requests to the corresponding author.

References

1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, *CA Cancer J Clin* **66**, 7–30, <https://doi.org/10.3322/caac.21332> (2016).
2. National Lung Screening Trial Research, T. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395–409, <https://doi.org/10.1056/NEJMoa1102873> (2011).
3. Moyer, V. A. & Force, U. S. P. S. T. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **160**, 330–338, <https://doi.org/10.7326/M13-2771> (2014).
4. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129–2139, <https://doi.org/10.1056/NEJMoa040938> (2004).
5. Zhou, C. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* **12**, 735–742, [https://doi.org/10.1016/S1470-2045\(11\)70184-X](https://doi.org/10.1016/S1470-2045(11)70184-X) (2011).
6. Massarelli, E. *et al.* KRAS mutation is an important predictor of resistance to therapy with epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. *Clin Cancer Res* **13**, 2890–2896, <https://doi.org/10.1158/1078-0432.CCR-06-3043> (2007).
7. Hsu, K. H. *et al.* Epidermal growth factor receptor mutation status in stage I lung adenocarcinoma with different image patterns. *J Thorac Oncol* **6**, 1066–1072, <https://doi.org/10.1097/JTO.0b013e31821667b0> (2011).
8. Lee, H. J. *et al.* Epidermal growth factor receptor mutation in lung adenocarcinomas: relationship with CT characteristics and histologic subtypes. *Radiology* **268**, 254–264, <https://doi.org/10.1148/radiol.13112553> (2013).
9. Liu, Y. *et al.* CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. *Radiology* **280**, 271–280, <https://doi.org/10.1148/radiol.2016151455> (2016).

10. Rizzo, S. *et al.* CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. *Eur Radiol* **26**, 32–42, <https://doi.org/10.1007/s00330-015-3814-0> (2016).
11. Sugano, M. *et al.* Correlation between computed tomography findings and epidermal growth factor receptor and KRAS gene mutations in patients with pulmonary adenocarcinoma. *Oncol Rep* **26**, 1205–1211, <https://doi.org/10.3892/or.2011.1412> (2011).
12. Yang, Y. *et al.* EGFR L858R mutation is associated with lung adenocarcinoma patients with dominant ground-glass opacity. *Lung Cancer* **87**, 272–277, <https://doi.org/10.1016/j.lungcan.2014.12.016> (2015).
13. Yano, M. *et al.* Epidermal growth factor receptor gene mutation and computed tomographic findings in peripheral pulmonary adenocarcinoma. *J Thorac Oncol* **1**, 413–416 (2006).
14. Zhou, J. Y. *et al.* Comparative analysis of clinicoradiologic characteristics of lung adenocarcinomas with ALK rearrangements or EGFR mutations. *Eur Radiol* **25**, 1257–1266, <https://doi.org/10.1007/s00330-014-3516-z> (2015).
15. Marks, J. L. *et al.* Prognostic and therapeutic implications of EGFR and KRAS mutations in resected lung adenocarcinoma. *J Thorac Oncol* **3**, 111–116, <https://doi.org/10.1097/JTO.0b013e318160c607> (2008).
16. Chen, Y. M. *et al.* Advanced non-Small cell lung cancer patients at the extremes of age in the era of epidermal growth factor receptor tyrosine kinase inhibitors. *Lung Cancer* **98**, 99–105, <https://doi.org/10.1016/j.lungcan.2016.05.020> (2016).
17. Weiss, G. J. *et al.* Noninvasive image texture analysis differentiates K-ras mutation from pan-wildtype NSCLC and is prognostic. *PLoS One* **9**, e100244, <https://doi.org/10.1371/journal.pone.0100244> (2014).
18. Glynn, C., Zakowski, M. F. & Ginsberg, M. S. Are there imaging characteristics associated with epidermal growth factor receptor and KRAS mutations in patients with adenocarcinoma of the lung with bronchioloalveolar features? *J Thorac Oncol* **5**, 344–348, <https://doi.org/10.1097/JTO.0b013e3181ce9a7a> (2010).
19. Wang, H. *et al.* Association Between Computed Tomographic Features and Kirsten Rat Sarcoma Viral Oncogene Mutations in Patients With Stage I Lung Adenocarcinoma and Their Prognostic Value. *Clin Lung Cancer* **17**, 271–278, <https://doi.org/10.1016/j.clc.2015.11.002> (2016).
20. Bremnes, R. M. *et al.* The role of tumor stroma in cancer progression and prognosis: emphasis on carcinoma-associated fibroblasts and non-small cell lung cancer. *J Thorac Oncol* **6**, 209–217, <https://doi.org/10.1097/JTO.0b013e3181f8a1bd> (2011).
21. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**, 4006, <https://doi.org/10.1038/ncomms5006> (2014).
22. Aerts, H. J. *et al.* Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. *Sci Rep* **6**, 33860, <https://doi.org/10.1038/srep33860> (2016).
23. Maldonado, F. *et al.* Noninvasive characterization of the histopathologic features of pulmonary nodules of the lung adenocarcinoma spectrum using computer-aided nodule assessment and risk yield (CANARY)—a pilot study. *J Thorac Oncol* **8**, 452–460, <https://doi.org/10.1097/JTO.0b013e3182843721> (2013).
24. Boland, J. M. *et al.* Adenocarcinoma *in situ*, minimally invasive adenocarcinoma, and invasive pulmonary adenocarcinoma—analysis of interobserver agreement, survival, radiographic characteristics, and gross pathology in 296 nodules. *Hum Pathol* **51**, 41–50, <https://doi.org/10.1016/j.humpath.2015.12.010> (2016).
25. Raghunath, S. *et al.* Noninvasive risk stratification of lung adenocarcinoma using quantitative computed tomography. *J Thorac Oncol* **9**, 1698–1703, <https://doi.org/10.1097/JTO.0000000000000319> (2014).
26. Maldonado, F. *et al.* Noninvasive Computed Tomography-based Risk Stratification of Lung Adenocarcinomas in the National Lung Screening Trial. *Am J Respir Crit Care Med* **192**, 737–744, <https://doi.org/10.1164/rccm.201503-0443OC> (2015).
27. Raghunath, S. *et al.* Quantitative stratification of diffuse parenchymal lung diseases. *PLoS One* **9**, e93229, <https://doi.org/10.1371/journal.pone.0093229> (2014).
28. Frey, B. J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972–976, <https://doi.org/10.1126/science.1136800> (2007).
29. Bartholmai, B. J. *et al.* Quantitative computed tomography imaging of interstitial lung diseases. *J Thorac Imaging* **28**, 298–307, <https://doi.org/10.1097/RTI.0b013e3182a21969> (2013).
30. Lindeman, N. I. *et al.* Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J Thorac Oncol* **8**, 823–859, <https://doi.org/10.1097/JTO.0b013e318290868f> (2013).
31. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892, <https://doi.org/10.1056/NEJMoa1113205> (2012).
32. Taniguchi, K., Okami, J., Kodama, K., Higashiyama, M. & Kato, K. Intratumor heterogeneity of epidermal growth factor receptor mutations in lung cancer and its correlation to the response to gefitinib. *Cancer Sci* **99**, 929–935, <https://doi.org/10.1111/j.1349-7006.2008.00782.x> (2008).
33. Tam, I. Y. *et al.* Distinct epidermal growth factor receptor and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. *Clin Cancer Res* **12**, 1647–1653, <https://doi.org/10.1158/1078-0432.CCR-05-1981> (2006).
34. Ess, S. M. *et al.* Epidermal growth factor receptor and anaplastic lymphoma kinase testing and mutation prevalence in patients with advanced non-small cell lung cancer in Switzerland: A comprehensive evaluation of real world practices. *Eur J Cancer Care (Engl)*, <https://doi.org/10.1111/ecc.12721> (2017).
35. Yip, S. S. *et al.* Associations between somatic mutations and metabolic imaging phenotypes in non-small cell lung cancer. *J Nucl Med*, <https://doi.org/10.2967/jnumed.116.181826> (2016).
36. Kosaka, T. *et al.* Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res* **64**, 8919–8923, <https://doi.org/10.1158/0008-5472.CAN-04-2818> (2004).
37. Liu, Y., Ren, Z., Wang, J. & Zhang, S. Epidermal growth factor receptor-tyrosine kinase inhibitor therapy is especially beneficial to patients with exon 19 deletion compared with exon 21 L858R mutation in non-small-cell lung cancer: Systematic review and meta analysis. *Thorac Cancer* **7**, 406–414, <https://doi.org/10.1111/1759-7714.12343> (2016).
38. Kim, J. S. *et al.* Prognostic impact of EGFR mutation in non-small-cell lung cancer patients with family history of lung cancer. *PLoS One* **12**, e0177015, <https://doi.org/10.1371/journal.pone.0177015> (2017).
39. Canale, M. *et al.* Impact of TP53 Mutations on Outcome in EGFR-Mutated Patients Treated with First-Line Tyrosine Kinase Inhibitors. *Clin Cancer Res*, <https://doi.org/10.1158/1078-0432.CCR-16-0966> (2016).
40. Vietri, M. T., Riegler, G., Pellino, G., Molinari, A. M. & Cioffi, M. Arg72Pro polymorphism of p53 may predict poor response to medical treatment in ulcerative colitis. *Ann Ital Chir* **85**, 513–517 (2014).

Acknowledgements

Study made possible by a Mayo Clinic internal research grant.

Author Contributions

R.C. and T.P. prepared the manuscript, S.J. prepared data analysis, R.C., T.P. and S.R. prepared the figures, B.R.K., M.C.A. and J.S.V. analyzed biological specimens, R.C., R.A.K., F.M., B.J.B. and T.P. executed image analysis. All authors contributed to editing and agreed on manuscript content.

Additional Information

Competing Interests: CANARY software, currently licensed to Imbio LLC (annual royalties <\$5000). This COI applies to Mayo Clinic and Brian Bartholmai, Fabien Maldonado, Ron Karwoski, Tobias Peikert and Srinivasan Rajagopalan. Ryan Clay, Benjamin Kipp, Jesse Voss and Marie Christine Aubry have no relevant conflicts of interest to disclose.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017