

# SCIENTIFIC REPORTS



OPEN

## Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer

Received: 27 January 2015

Accepted: 14 May 2015

Published: 05 June 2015

Chintan Parmar<sup>1,3,4,\*</sup>, Ralph T. H. Leijenaar<sup>3,\*</sup>, Patrick Grossmann<sup>1,5</sup>, Emmanuel Rios Velazquez<sup>2</sup>, Johan Bussink<sup>6</sup>, Derek Rietveld<sup>7</sup>, Michelle M. Rietbergen<sup>8</sup>, Benjamin Haibe-Kains<sup>9,10</sup>, Philippe Lambin<sup>3</sup> & Hugo J.W.L. Aerts<sup>1,2,5</sup>

Radiomics provides a comprehensive quantification of tumor phenotypes by extracting and mining large number of quantitative image features. To reduce the redundancy and compare the prognostic characteristics of radiomic features across cancer types, we investigated cancer-specific radiomic feature clusters in four independent Lung and Head & Neck (H&N) cancer cohorts (in total 878 patients). Radiomic features were extracted from the pre-treatment computed tomography (CT) images. Consensus clustering resulted in eleven and thirteen stable radiomic feature clusters for Lung and H&N cancer, respectively. These clusters were validated in independent external validation cohorts using rand statistic (Lung RS = 0.92,  $p < 0.001$ , H&N RS = 0.92,  $p < 0.001$ ). Our analysis indicated both common as well as cancer-specific clustering and clinical associations of radiomic features. Strongest associations with clinical parameters: Prognosis Lung CI =  $0.60 \pm 0.01$ , Prognosis H&N CI =  $0.68 \pm 0.01$ ; Lung histology AUC =  $0.56 \pm 0.03$ , Lung stage AUC =  $0.61 \pm 0.01$ , H&N HPV AUC =  $0.58 \pm 0.03$ , H&N stage AUC =  $0.77 \pm 0.02$ . Full utilization of these cancer-specific characteristics of image features may further improve radiomic biomarkers, providing a non-invasive way of quantifying and monitoring tumor phenotypic characteristics in clinical practice.

Recent advances of medical and computational science have led to the emergence of ‘precision medicine’, which has revolutionized the cancer care and medical science in general. A major proportion of precision medicine research has centered on unveiling different molecular characteristics of the disease tissues by using genomic and proteomic technologies. In spite of their enormous potential, these techniques have found limited implementations in routine clinical practice<sup>1</sup>. A major challenge being the invasive nature, as biopsies, having high associated risk and cost, are often required.

Imaging on the other hand provides promising means of non-invasive tissue characterization and is furthermore routinely used for disease detection, diagnosis, and treatment purposes in clinical practice<sup>2–4</sup>. X-ray computed tomography (CT) is a frequently used imaging modality for oncology because it assesses tissue density in high resolution and exhibits strong contrasts among different tissue types. In

<sup>1</sup>Departments of Radiation Oncology. <sup>2</sup>Radiology, Dana-Farber Cancer Institute, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA. <sup>3</sup>Radiation Oncology (MAASTRO), Research Institute GROW, Maastricht University, Maastricht, the Netherlands. <sup>4</sup>Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. <sup>5</sup>Department of Biostatistics & Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA. <sup>6</sup>Department of Radiation Oncology, Radboud University Medical Center, Nijmegen, the Netherlands. <sup>7</sup>Department of Radiation Oncology, VU University Medical Center, Amsterdam, The Netherlands. <sup>8</sup>Department of Otolaryngology/Head and Neck Surgery, VU University Medical Center, Amsterdam, The Netherlands. <sup>9</sup>Ontario Cancer Institute, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. <sup>10</sup>Medical Biophysics Department, University of Toronto, Toronto, Ontario, Canada. \*These author contributed equally to this work. Correspondence and requests for materials should be addressed to H.J.W.L.A. (email: Hugo\_Aerts@dfci.harvard.edu) or C.P. (email: Chintan\_Parmar@dfci.harvard.edu)

routine clinical practice, tumor response to therapy is measured by the RECIST and/or WHO criteria, based on CT imaging. These descriptors measure the change in size of tumors, and often do not succeed in predicting overall survival<sup>5,6</sup>.

“Radiomics” is an emerging field of research that aims to utilize the full potential of medical imaging. Radiomics focuses on extracting a large number of quantitative features from medical images, providing a more detailed quantification of tumor phenotypic characteristics—effectively converting medical images into a high dimensional minable feature space<sup>7–9</sup>. Several studies have defined and quantified various image descriptors and stated their significance for treatment monitoring and outcome prediction in different cancer types<sup>10–14</sup>. Moreover, some studies have also reported an association between radiographic imaging phenotypes and tumor stage, metabolism<sup>15</sup>, hypoxia, angiogenesis<sup>16</sup> and the underlying gene and/or protein expression profiles<sup>17–19</sup>.

A main challenge in radiomics is to deal with feature redundancy in order to obtain a non-redundant set of imaging biomarkers. Consensus clustering<sup>20</sup> could address this issue by reducing the feature space into several non-redundant feature clusters. In this study we identified and validated radiomic feature clusters in cohorts of Lung cancer and Head & Neck (H&N) cancer patients. We also evaluated the clinical importance of these clusters by quantifying their association with important clinical parameters and patient survival. Moreover, we used the identified radiomic clusters to build cancer-specific multivariable radiomic signatures and tested their prognostic performance. Identification of cancer-specific radiomic clusters provides a crucial step towards stable and clinically relevant radiomic biomarkers, providing a non-invasive way of quantifying and monitoring tumor phenotypic characteristics in clinical practice.

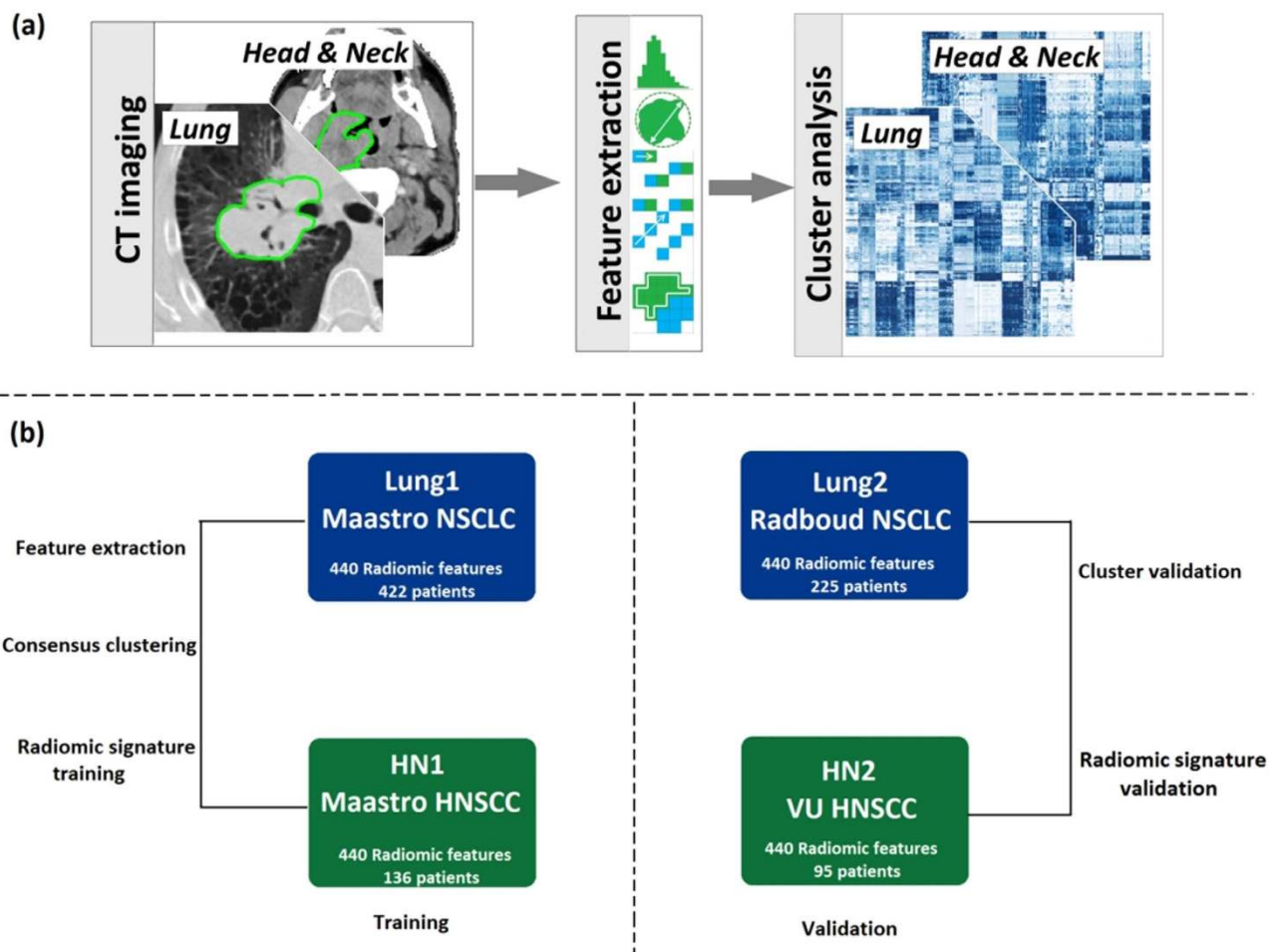
## Methods

**Radiomic features.** We defined 440 radiomic image features that quantify tumor characteristics. These features were divided in four groups: I) tumor intensity, II) shape, III) texture and IV) wavelet features. Tumor intensity based features, which are defined using first order statistics of the intensity histogram, quantified the density of the tumor region on CT image. Shape features described the 3D geometric properties of the tumor, whereas textural features quantified intra-tumor heterogeneity. Textural features were computed by analyzing the spatial distribution of voxel intensities in thirteen directions. These features are derived from gray level co-occurrence (GLCM)<sup>21</sup> and run length matrices (GLRLM)<sup>22</sup> and were computed by averaging their values over all thirteen directions. Wavelet features are the transformed domain representations of the intensity and textural features. These features were computed on different wavelet decompositions of the original image using a coiflet wavelet transformation. All image analysis was performed in Matlab R2012b (The Mathworks, Natick, MA) using an adapted version of CERR (Computational Environment for Radiotherapy Research)<sup>23</sup> and features were automatically extracted with in-house developed radiomics image analysis software. Mathematical definitions of all radiomic features as well as the extraction methods were previously described<sup>18</sup>.

**Datasets.** Briefly, we considered four image datasets (see overview in Figure 1) for this study, from different institutes in the Netherlands: (1) Lung1: 422 NSCLC patients treated at MAASTRO Clinic in Maastricht. (2) Lung2: 225 NSCLC patients treated at Radboud University Medical Center in Nijmegen. (3) HN1: 136 head and neck squamous cell carcinoma (HNSCC) patients treated at MAASTRO Clinic in Maastricht and (4) HN2: 95 HNSCC patients treated at the VU University Medical Center in Amsterdam. CT-scans, manual delineations and clinical data were available for all included patients. More details on the included datasets have been described earlier<sup>18</sup>.

**Data analysis.** *Comparison of the prognostic performance of radiomic features in Lung and H&N cancer.* In order to compare the prognostic utility of radiomic features across Lung and H&N cancer, for each feature, we computed and compared the concordance index (CI)<sup>24</sup>, which is the generalization of area under ROC curve. R package survcomp was used for the analysis<sup>25</sup>. P-values are corrected for multiple testing (FDR 5%).

*Consensus clustering.* We used consensus clustering to cluster the radiomic features extracted from the training cohorts Lung1 and HN1. Consensus clustering is a resampling based clustering methodology, which quantifies the consensus between several clustering iterations and provides means to estimate the number of clusters that best fit the data<sup>20</sup>. We estimated the range for the appropriate number of clusters from the delta area plots (supplementary figure S1). From this range, we chose the number of clusters which gave the highest median cluster consensus over all clusters. Cluster consensus was defined as the average consensus between all pairs of features belonging to the same cluster. Cluster consensus (range [0–1]) indicates the robustness (stability) of a cluster over resampling. We also computed the mean pairwise correlation (range [0–1]) between features of a cluster, which is a measure of the cluster compactness (similarity of features within the cluster). Qualitative categorization of cluster stability was defined as; consensus < 0.5, poor stability;  $0.5 \leq$  consensus < 0.75, moderate stability; and consensus  $\geq$  0.75, high stability. Cluster compactness was also assessed using the same qualitative categorization. We applied hierarchical clustering with agglomerative ward linkage, a Pearson correlation based dissimilarity measure  $(1 - r)$  and 10,000 resampling iterations. Consensus clustering was performed using the R package ConsensusClusterPlus<sup>26</sup>.



**Figure 1.** (a) Radiomic analysis overview: For both Lung and H&N cancer datasets, we extracted radiomic features from pre-treatment CT images. Cluster analysis was performed on the feature data. (b) Datasets overview: Four independent radiomic cohorts of Lung and Head & Neck cancer were included in the analysis. Lung1 and HN1 were used for training; Lung2 and HN2 were used for validation.

**Radiomic cluster validation.** Radiomic feature clusters obtained on the training cohorts, Lung1 and HN1, were considered as the reference Lung and H&N clusters. For cluster validation, we clustered the Lung and H&N validation cohorts, Lung2 and HN2, using the same hierarchical clustering algorithm and the same number of clusters as for the corresponding training cohort. Rand Statistic (RS)<sup>27</sup> was used to assess the agreement between each reference clustering ( $P$ ) and the clustering obtained for its respective validation cohort ( $C$ ) and was defined as:

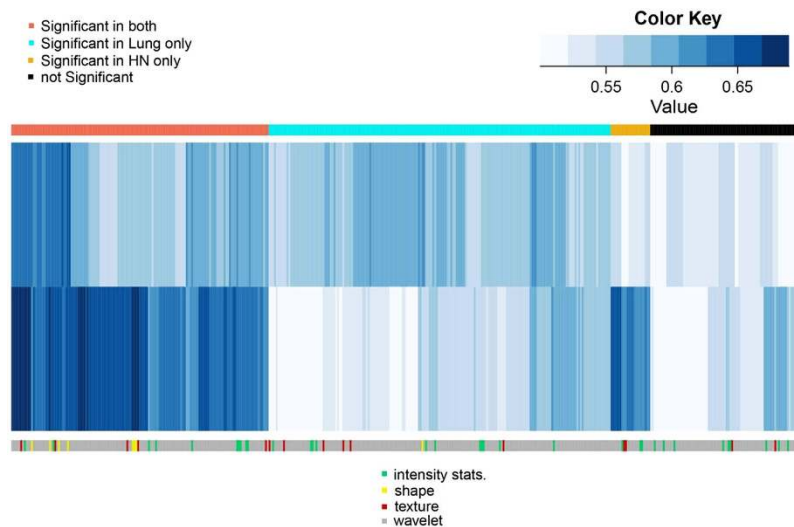
$$RS = \frac{|SS| + |DD|}{|SS| + |SD| + |DS| + |DD|} \quad (1)$$

where  $|SS|$  is the number of feature pairs that cluster together in both  $C$  and  $P$ ,  $|SD|$  is the number of feature pairs that cluster together in  $C$  but not in  $P$ ,  $|DS|$  is the number of feature pairs that cluster together in  $P$ , but not in  $C$  and  $|DD|$  is the number of feature pairs that do not cluster together in both  $C$  and  $P$ . Significance of RS was determined by a random permutation test using 1000 iterations.

**Similarity between Lung and H&N clusters.** Cluster overlap between the individual feature clusters of Lung and H&N radiomic cohorts were assessed using the Jaccard index<sup>27</sup>, which is defined as:

$$Jaccard(L, H) = \frac{|L \cap H|}{|L \cup H|} \quad (2)$$

where  $L$  and  $H$  are any feature clusters of Lung and H&N cohorts. Qualitative categorization of cluster overlap was defined as;  $Jaccard < 0.5$ , poor overlap;  $0.5 \leq Jaccard < 0.75$ , moderate overlap; and  $Jaccard \geq 0.75$  high overlap.



**Figure 2. Heatmap showing the prognostic performance of radiomic features in Lung2 and HN2 cohorts.** Prognostic performance was evaluated using the concordance index (CI). Note that a large number of features are prognostic in both cancer types. However, also a large number of features are cancer-type specific, e.g. prognostic only in Lung or only in H&N cancer.

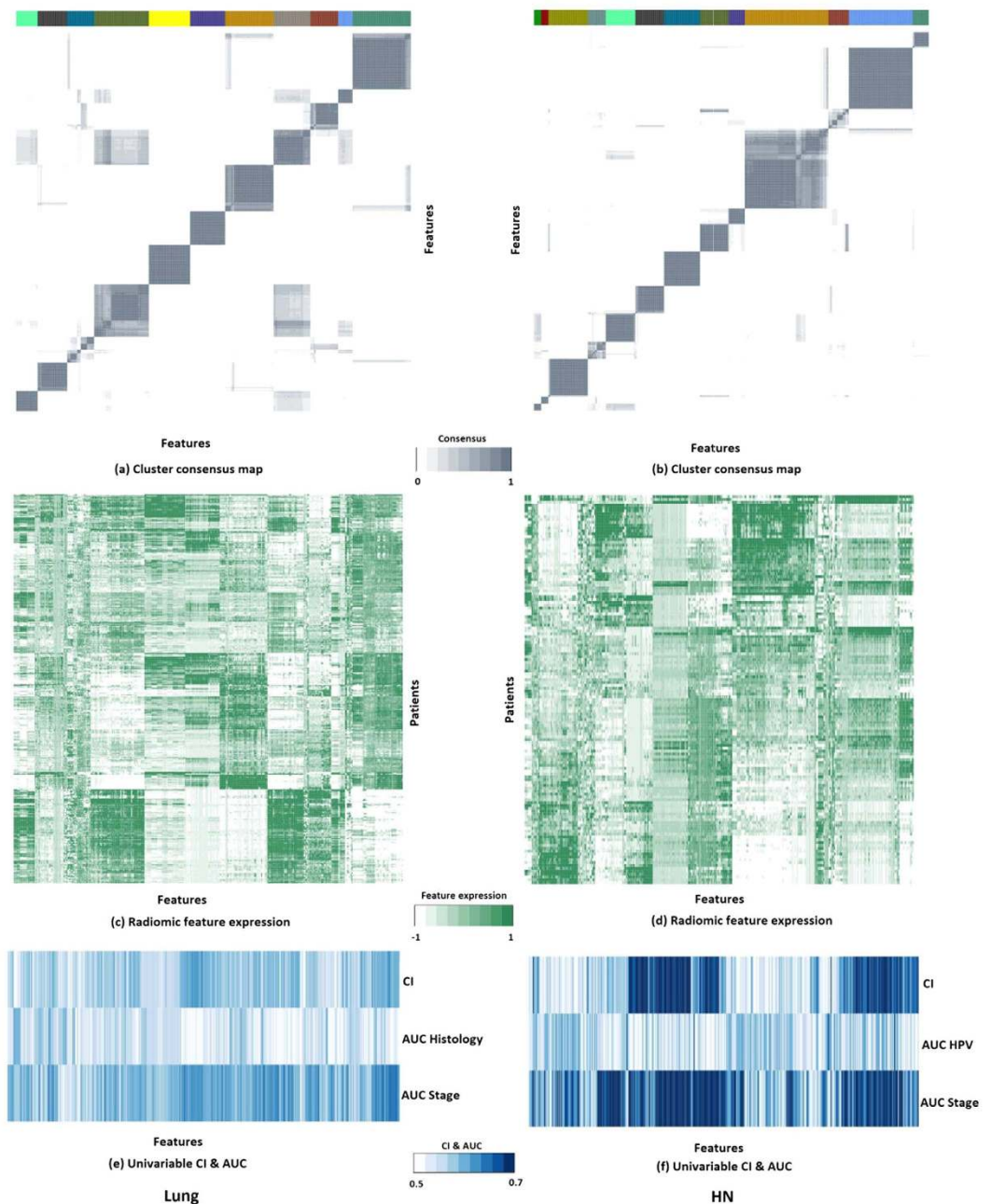
*Clinical relevance of radiomic clusters.* In order to quantify the association between radiomic feature clusters and patient survival, we used the concordance index (CI), whereas the association between a feature cluster and a categorical clinical parameter (i.e. Lung cancer histology, H&N HPV status or Lung/H&N tumor stage) was quantified using the area under the ROC curve (AUC). For clinical parameters having more than two categorical levels, a multi-class AUC was computed using a pairwise approach. Univariable CI and AUC were computed for each feature. A cluster's association with patient survival and clinical parameters was then quantified as the mean CI and mean AUC over all contained features. Significance was estimated using a random permutation test with 1000 iterations. R package `survcomp`<sup>25</sup> and `pROC`<sup>28</sup> was used for this analysis. Qualitative categorization of the prognostic or predictive performance was defined as poor (CI or AUC < 0.6), moderate ( $0.6 \leq$  CI or AUC < 0.75) and high (CI or AUC  $\geq$  0.75).

*Multivariable clinical relevance.* In order to select non-redundant imaging biomarkers, cancer-specific radiomic signatures were built using the medoids of the obtained clusters of Lung1 and HN1 cohorts. The medoid is a single representative feature, which has the highest average pairwise correlation within a cluster. To investigate the multivariable prognostic utility of these selected radiomic features, a multivariable Cox proportional hazards model was fitted on each training cohort (i.e. Lung1 and HN1) and their prognostic performance was tested on validation cohorts (i.e. Lung2 and HN2) using the CI. R package `survcomp`<sup>25</sup> was used for this analysis. For the prediction of categorical clinical parameters, we built multivariable classifiers on each training cohort (i.e. Lung1 or HN1) using logistic regression, with the medoids as independent variables. The predictive performance of a classifier was evaluated on the corresponding validation cohort (Lung2 or HN2) using AUC. For clinical parameters having more than two categorical levels, logistic regression was fitted using a pairwise approach and performance was evaluated using multiclass AUC<sup>29</sup>. R package `VGAM`<sup>30</sup> was used for this analysis.

## Results

In order to investigate radiomic features in Lung and H&N cancer cohorts, a total of 440 radiomic features were extracted from the segmented tumor regions of the pre-treatment CT images of Lung and Head and Neck cancer cohorts. In our analysis, we used datasets Lung1 (n = 422) and HN1 (n = 136) as training datasets, and Lung2 (n = 225) and HN2 (n = 95) as validation datasets (Fig. 1).

**Comparison of the prognostic performance of radiomic features in Lung and H&N cancer.** The prognostic utility of the radiomic features was assessed using the concordance index (CI). Figure 2 depicts a CI heatmap of radiomic features in the validation cohorts (Lung2 and HN2). We observed that 143 features had significant prognostic performance (CI > 0.5, p < 0.05 FDR corrected) in both the cancer types whereas 212 features (190 features in Lung and 22 features in H&N) showed significant prognosis only in one of the two cancer types. Eighty-five features turned out not to be significantly prognostic in either of the two cancer types.



**Figure 3.** Heatmaps for radiomic features of Lung and H&N training cohorts ordered with respect to the obtained Lung and H&N clusters. (a–b) Cluster consensus maps of Lung cancer (11 clusters) and H&N cancer (13 clusters) cohorts. (c–d) Radiomic feature expressions of Lung and H&N radiomic clusters. (e–f) Clinical relevance (CI & AUC) of radiomic clusters of Lung and H&N cancer.

**Identification and validation of radiomic feature clusters in Lung cancer.** In order to identify stable clusters of radiomic features, consensus clustering procedure was applied on Lung1 training cohort. We obtained eleven distinct clusters (size: 16 to 65 features per cluster). Heatmaps in Figure 3 depict the consensus maps (Fig. 3a) and normalized expression levels (Fig. 3c) of the obtained Lung radiomic feature clusters in Lung1 training cohort. These clusters were validated in the Lung2 validation cohort ( $RS=0.92$ , permutation test  $p\text{-value}<0.001$ ). We observed that four clusters (LCL-4, LCL-6, LCL-7, LCL-11) had a high cluster consensus (consensus  $\geq 0.75$ ) and within cluster correlation (correlation  $\geq 0.75$ ), whereas six clusters (LCL-1, LCL-2, LCL-3, LCL-5, LCL-8, LCL-10) showed high cluster consensus (consensus  $\geq 0.75$ ) but moderate within cluster correlation ( $0.5 \leq \text{correlation} < 0.75$ ). For cluster LCL-9, both the cluster consensus and within cluster correlation were poor (consensus = 0.41, correlation = 0.14). Details regarding the cluster size, associated feature categories, cluster consensus and within cluster correlation can be obtained from Table 1.

Cluster no	Cluster size (Intensity stats + Shape + Texture + Wavelet features)	Within cluster correlation	Cluster consensus	CI	Histology AUC	Stage AUC	Cluster medoid
LCL-1	65 (3+2+13+47)	0.62 ± 0.21	0.92 ± 0.12	<b>*0.58 ± 0.02</b>	0.54 ± 0.02	<b>*0.60 ± 0.03</b>	Wavelet_HLL_glcm_invdifnorm
LCL-2	16 (3+0+5+8)	0.70 ± 0.23	0.97 ± 0.03	<b>*0.54 ± 0.02</b>	0.51 ± 0.02	<b>*0.59 ± 0.03</b>	GLCM_clusTend
LCL-3	31 (1+2+1+27)	0.50 ± 0.26	0.85 ± 0.17	<b>*0.55 ± 0.03</b>	0.54 ± 0.03	<b>*0.59 ± 0.03</b>	Wavelet_LHH_rgl_lowGrayLevelRunEmphasis
LCL-4	41 (2+1+8+30)	0.78 ± 0.12	0.91 ± 0.09	<b>*0.59 ± 0.01</b>	0.52 ± 0.01	<b>*0.61 ± 0.02</b>	GLCM_diffEntro
LCL-5	54 (2+0+4+48)	0.70 ± 0.15	0.91 ± 0.12	<b>*0.58 ± 0.01</b>	0.55 ± 0.02	<b>*0.60 ± 0.02</b>	Wavelet_LHL_stats_uniformity
LCL-6	39 (2+3+2+32)	0.78 ± 0.16	1 ± 0	<b>*0.60 ± 0.01</b>	0.52 ± 0.01	<b>*0.61 ± 0.01</b>	RLGL_runLengthNonuniformity
LCL-7	46 (2+0+0+44)	0.86 ± 0.07	1 ± 0	<b>*0.56 ± 0.01</b>	0.55 ± 0.01	<b>*0.58 ± 0.01</b>	Wavelet_HLL_glcm_sumAvg
LCL-8	61 (0+0+0+61)	0.71 ± 0.15	0.79 ± 0.17	<b>*0.57 ± 0.01</b>	0.55 ± 0.02	<b>*0.58 ± 0.02</b>	Wavelet_HLH_glcm_diffEntro
LCL-9	30 (0+0+0+30)	0.14 ± 0.30	0.41 ± 0.36	<b>*0.54 ± 0.02</b>	<b>*0.55 ± 0.03</b>	<b>*0.54 ± 0.03</b>	Wavelet_HHH_stats_min
LCL-10	33 (0+0+0+33)	0.66 ± 0.19	0.95 ± 0.08	<b>*0.57 ± 0.02</b>	<b>*0.56 ± 0.03</b>	<b>*0.58 ± 0.03</b>	Wavelet_LHL_rgl_longRunHighGrayLevelEmpha
LCL-11	24 (0+0+0+24)	0.75 ± 0.17	0.96 ± 0.04	<b>*0.55 ± 0.02</b>	0.55 ± 0.02	<b>*0.59 ± 0.02</b>	Wavelet_HLH_stats_std

**Table 1.** Cluster table describing cluster size, size of each feature category, within cluster correlation, cluster consensus, average CI and AUC and selected representative features in Lung cancer. Bold values with an asterisk indicate the significant clinical relevance of a cluster.

**Identification and validation of radiomic feature clusters in HN cancer.** Consensus clustering in HN1 training cohort resulted in thirteen distinct radiomic feature clusters (size: 8 to 93 features per cluster), which were validated in independent HN2 validation cohort (RS = 0.92, permutation test p-value < 0.001). Heatmaps in Figure 3b and 3d show the consensus maps and normalized expression levels of the obtained radiomic feature clusters in HN1 training cohort. Six clusters (HNCL-1, HNCL-2, HNCL-6, HNCL-7, HNCL-12, HNCL-13) had high cluster consensus (consensus ≥ 0.75) and within cluster correlation (correlation ≥ 0.75), whereas five other clusters (HNCL-4, HNCL-5, HNCL-8, HNCL-9, HNCL-11) showed high cluster consensus (consensus ≥ 0.75) but moderate within cluster correlation (0.5 ≤ correlation < 0.75). Cluster HNCL-10 had moderate cluster consensus (consensus = 0.65) and poor cluster correlation (correlation = 0.04), whereas clusters HNCL-3 showed poor cluster consensus (consensus = 0.41) and correlation (correlation = 0.12) (see Table 2).

**Similarity between Lung and H&N clusters.** In order to assess the overlap between individual Lung and H&N radiomic clusters, we compared the Lung and H&N clusters, pairwise, using the Jaccard index. We observed that Lung cluster LCL-6 and H&N cluster HNCL-7 had high overlap (Jaccard = 0.98). Cluster pairs LCL-7 & HNCL-2 (Jaccard = 0.65), LCL-8 & HNCL-4 (Jaccard = 0.66) and LCL-5 & HNCL-11 (Jaccard = 0.66) and LCL-3 & HNCL-8 (Jaccard = 0.66) showed moderate overlap, whereas the remaining pairs had poor overlap (Jaccard < 0.5) (see Fig. 4).

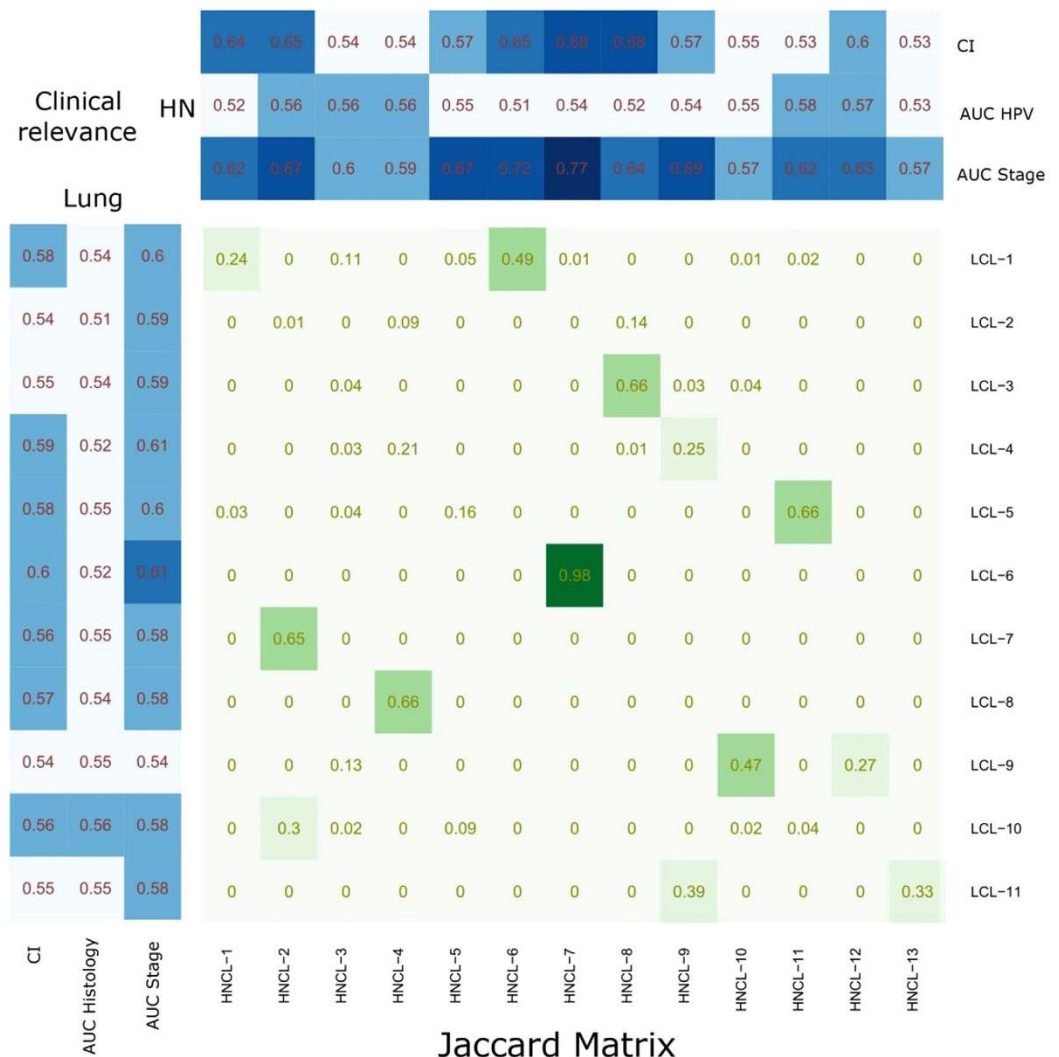
**Clinical relevance of radiomic clusters.** Mean CI and mean AUC values for the obtained Lung and H&N clusters are depicted in Table 1 & Table 2. All eleven Lung clusters had a significant prognostic association with patient survival. However, only four Lung clusters (LCL-1, LCL-4, LCL-5 and LCL-6) had a mean CI higher or equal to 0.58. Two Lung clusters (LCL-9 and LCL-10) were significantly

Cluster no	Cluster size (Intensity stats + Shape + Texture + Wavelet features)	Within cluster correlation	Cluster consensus	CI	HPV status AUC	Stage AUC	Cluster medoid
HNCL-1	18 (0+0+8+10)	0.80 ± 0.25	1 ± 0.21	<b>*0.64 ± 0.03</b>	0.52 ± 0.03	<b>*0.62 ± 0.02</b>	Wavelet_LLL_glcm_sumAvg
HNCL-2	71 (2+0+1+68)	0.80 ± 0.14	1 ± 0.01	<b>*0.65 ± 0.03</b>	0.56 ± 0.04	<b>*0.67 ± 0.02</b>	Wavelet_LHH_glcm_sumVar
HNCL-3	23 (4+0+1+18)	0.12 ± 0.30	0.41 ± 0.27	0.54 ± 0.04	0.56 ± 0.04	<b>*0.60 ± 0.04</b>	Wavelet_LLL_stats_mean
HNCL-4	93 (4+0+7+82)	0.70 ± 0.17	0.83 ± 0.19	0.54 ± 0.03	0.56 ± 0.03	0.59 ± 0.02	Wavelet_LHL_glcm_diffEntro
HNCL-5	18 (2+0+5+11)	0.61 ± 0.19	0.96 ± 0.04	0.57 ± 0.04	0.55 ± 0.03	<b>*0.68 ± 0.01</b>	Wavelet_LLL_rgl_longRunEmphasis
HNCL-6	32 (0+0+4+28)	0.76 ± 0.13	1 ± 0.06	<b>*0.65 ± 0.05</b>	0.51 ± 0.05	<b>*0.72 ± 0.03</b>	Wavelet_LHL_glcm_invDiffnorm
HNCL-7	40 (2+4+2+32)	0.98 ± 0.07	1 ± 0	<b>*0.68 ± 0.01</b>	0.54 ± 0.03	<b>*0.77 ± 0.02</b>	Stats_totalenergy
HNCL-8	32 (1+3+3+25)	0.65 ± 0.21	0.97 ± 0.13	<b>*0.68 ± 0.06</b>	0.52 ± 0.03	<b>*0.64 ± 0.02</b>	Wavelet_LLH_rgl_lowGrayLevelRunEmphasis
HNCL-9	33 (0+0+2+31)	0.71 ± 0.18	0.95 ± 0.09	0.57 ± 0.02	0.54 ± 0.03	<b>*0.69 ± 0.02</b>	Wavelet_LHL_stats_std
HNCL-10	20 (0+1+0+19)	0.04 ± 0.26	0.65 ± 0.22	0.55 ± 0.03	0.55 ± 0.05	0.57 ± 0.03	Wavelet_HHH_glcm_correl1
HNCL-11	44 (0+0+0+44)	0.68 ± 0.16	0.99 ± 0.15	0.53 ± 0.02	0.58 ± 0.03	<b>*0.62 ± 0.02</b>	Wavelet_HHL_stats_uniformity
HNCL-12	8 (0+0+0+8)	0.76 ± 0.19	1 ± 0.03	<b>*0.60 ± 0.05</b>	0.57 ± 0.04	<b>*0.63 ± 0.02</b>	Wavelet_HHH_stats_min
HNCL-13	8 (0+0+0+8)	0.84 ± 0.15	1 ± 0	0.53 ± 0.02	0.53 ± 0.02	0.57 ± 0.01	Wavelet_HLH_stats_std

**Table 2.** Cluster table describing cluster size, size of each feature category, within cluster correlation, cluster consensus, average CI and AUC and selected representative features in H&N cancer. Bold values with an asterisk indicate the significant clinical relevance of a cluster.

associated with tumor histology. All Lung clusters had significant association with tumor stage (see Table 1). For H&N cancer, six clusters (HNCL-1, HNCL-2, HNCL-6, HNCL-7, HNCL-8 and HNCL-12) were significantly associated with patient survival. We did not observe any association between the H&N clusters and HPV status. However, except for the three H&N clusters (HNCL-4, HNCL-10, HNCL-13), all the other H&N clusters were significantly associated with tumor stage (see Table 2). Univariable CI and AUC values of radiomic features in Lung1 and HN1 cohorts are represented by heatmaps in Fig. 3e,f, respectively.

**Multivariable clinical relevance.** We built two radiomic signatures, one each for Lung and H&N cohort, using the medoids of the obtained Lung and H&N radiomic feature clusters, respectively. To evaluate the multivariate prognostic performance, we trained a multivariable Cox proportional hazards model on both training cohorts (Lung1 and HN1). Prognostic performance of each model was externally tested on validation cohorts (Lung2 and HN2). We observed that the prognostic performance of Lung based multivariable model (CI = 0.61) was higher than the H&N model (CI = 0.56) in Lung2 validation cohort, whereas for HN2 validation cohort, H&N based multivariable model (CI = 0.63) performed better than the Lung model (CI = 0.57). We also used the cluster medoids for the prediction of clinical parameters. We trained classifiers on training cohort (i.e. Lung1 or HN1) using logistic regression and



**Figure 4.** Heatmap depicting cluster overlap and clinical relevance (CI & AUC). Center matrix in green & white color represents the overlap (Jaccard index) between the clusters of Lung (rows) and H&N (columns) radiomic cohorts. Top and left side panels in blue & white color depicts the average CI & AUC of the corresponding Lung and H&N radiomic clusters.

evaluated the predictive performance in the corresponding validation cohort (Lung2 or HN2). The Lung signature had a moderate performance for the prediction of tumor histology (AUC = 0.64) and tumor stage (AUC = 0.64). The H&N signature was highly predictive for tumor stage (AUC = 0.80) and had a moderate predictive performance for HPV status (AUC = 0.60).

## Discussion

Medical imaging plays an important role in medical care and science due to its ability to assess tissue characteristics and organ anatomy non-invasively. It is therefore widely used in disease diagnosis, progression assessment and treatment monitoring in clinical oncology. Radiomics, a high throughput approach, can quantify the differences between oncologic tissues and hence provide prognostic or predictive imaging biomarkers<sup>7,8</sup>.

In this study we investigated clustering as a means to deal with the high dimensional feature space generated with radiomics, as well as to investigate common and cancer-type specific radiomic patterns. We applied consensus clustering on 440 radiomic features extracted from Lung cancer and Head & Neck cancer patient cohorts. Furthermore these clusters were externally validated on independent validation cohorts. For both cancer types, many clusters showed high cluster consensus and high within cluster correlation, which indicates the high robustness (stability) and compactness of these clusters. These results indicate that consensus clustering could provide robust radiomic feature clusters and hence reduce the feature redundancy. The majority of the obtained Lung and H&N radiomic clusters were significantly associated with patient survival and tumor stage. Two Lung clusters also showed significant association



with tumor histology. Our multivariable analysis showed that cancer-specific multivariable radiomic signatures displayed moderate or high prognostic (predictive) performance.

Comparing the individual Lung and H&N feature clusters, we observed that five cluster pairs had substantial overlap ( $Jaccard \geq 0.6$ ) between the Lung and H&N cancer, whereas the overlap for other cluster pairs was poor. These results demonstrate both common as well as cancer-specific clustering characteristics of radiomic features.

It can be observed from our analysis that radiomic features also have cancer-specific prognostic ability. We compared the univariable CI values of radiomic features across the two cancer types and observed that several radiomic features have significant prognostic utility in only one of the two cancer types. Furthermore the multivariable radiomic signatures performed better in validation cohorts of the same cancer type in our multivariable analysis.

Recently, Aerts *et al.*<sup>18</sup>, built a radiomic signature of four features, which were selected based on their stability across test-retest image scans and multiple tumor delineations, as well as their prognostic performance in a training dataset. However, the feature selection was only applied to one cancer type (i.e. lung cancer patients). The primary objective of our study was to separately investigate and compare radiomic feature subgroups in Lung and H&N cancer. Our analysis reveals a cancer-specific grouping and prognostic trends of radiomic features, which could be exploited to potentially improve the performance of prognostic models. In another radiomic study of Lung cancer cohorts, Balagurunathan *et al.*<sup>31</sup>, used a three step procedure to select the most reproducible, informative and non-redundant features. In this method, it was required to empirically choose three threshold values corresponding to the concordance correlation coefficient, dynamic range and pairwise correlation. On the contrary, our method of feature selection only depends on the number of clusters. In general, the estimation of number of clusters is done by a visual inspection of consensus matrices and the delta area plot<sup>20</sup>. However, we used a more deterministic method for the estimation of number of clusters. Overall, the parameters of consensus clustering procedure were carefully chosen, keeping in mind the larger applicability of the obtained clustering results in radiomics. However, these parameters are still just design choices and there may be other relevant parameter configurations, which could be tried and the resultant clustering outputs could be analyzed and compared. These interesting research issues do not fall within the scope of this study, which is primarily focused on the cancer specific comparison of radiomic features.

We expect that the obtained feature clusters of our study combined with feature stability information could further enhance the feature selection process, providing more reliable and prognostic radiomic signatures. However, due to the unavailability of test-retest and multiple delineation H&N cancer cohorts, we could not compute the stability of H&N based radiomic features and hence could not include a cancer-specific stability analysis in our study. Machine learning based advanced feature selection methods could also enhance the radiomic biomarker selection process. Although identifying the optimal feature selection method is very important for radiomics research, it does not fit well within the clustering framework of this study and hence should be addressed in future studies, investigating machine learning methods for radiomic data.

We focused only on CT derived radiomic features in this study. However, imaging is routinely performed in clinical practice using different modalities (e.g. CT, PET, MRI, etc.). An integrated radiomic analysis of different cancer types using multiple modalities could therefore provide even more information to facilitate medical decision support with imaging biomarkers.

This study identified different Lung and H&N radiomic feature subgroups and quantified their clinical significance. Our investigation revealed that clustering and prognostic characteristics of radiomic features are cancer-specific. Cancer-specific prognostic characteristics of radiomic features should be considered for building prognostic models, which could improve the prognosis in cancer care. In general, high throughput medical image data mining research like “Radiomics” can influence the cancer research greatly, as it provides a promising non-invasive way of quantifying and monitoring tumor phenotypic characteristics across different cancer types in clinical oncology.

## References

1. Chung, C. H., Levy, S., Chaurand, P. & Carbone, D. P. Genomics and proteomics: emerging technologies in clinical cancer research. *Crit. Rev. Oncol. Hemat.* **61**, 1–25 (2007).
2. Kurland, B. F. *et al.* Promise and pitfalls of quantitative imaging in oncology clinical trials. *Magn. Reson. Imaging* **30**, 1301–1312 (2012).
3. Buckler, A. J., Bresolin, L., Dunnick, N. R. & Sullivan, D. C. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. *Radiology* **259**, 875–884 (2011).
4. Buckler, A. J., Bresolin, L., Dunnick, N. R. & Sullivan, D. C. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology* **258**, 906–914 (2011).
5. Zhao, B. *et al.* A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. *Clin. Cancer Res.* **16**, 4647–4653 (2010).
6. Birchard, K. R., Hoang, J. K., Herndon, J. E. & Patz, E. F. Early changes in tumor size in patients treated for advanced stage nonsmall cell lung cancer do not correlate with survival. *Cancer* **115**, 581–586 (2009).
7. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446 (2012).
8. Kumar, V. *et al.* Radiomics: the process and the challenges. *Magn. Reson. Imaging* **30**, 1234–1248 (2012).
9. Lambin, P. *et al.* Predicting outcomes in radiation oncology—multifactorial decision support systems. *Nat. Rev. Clin. Oncol.* **10**, 27–40 (2012).

10. El Naqa, I. *et al.* Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recogn.* **42**, 1162–1171 (2009).
11. Tixier, F. *et al.* Intra-tumor heterogeneity on baseline 18F-FDG PET images characterized by textural features predicts response to concomitant radio-chemotherapy in esophageal cancer. *J. Nucl. Med.* **52**, 369–378 (2011).
12. Vaidya, M. *et al.* Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother. Oncol.* **102**, 239–245 (2012).
13. Ganeshan, B., Panayiotou, E., Burnand, K., Dizdarevic, S. & Miles, K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur. Radiol.* **22**, 796–802 (2012).
14. Nicolajilwan, M. *et al.* Addition of MR imaging features and genetic biomarkers strengthens glioblastoma survival prediction in TCGA patients. *J. Neuroimaging*. doi: 10.1016/j.neurad.2014.02.006 (2014)
15. Ganeshan, B., Abaleke, S., Young, R. C., Chatwin, C. R. & Miles, K. A. Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage. *Cancer Imaging* **10**, 137 (2010).
16. Ganeshan, B. *et al.* Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology* **266**, 326–336 (2013).
17. Segal, E. *et al.* Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat. Biotechnol.* **25**, 675–680 (2007).
18. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* doi:10.1038/ncomms5006 (2014).
19. Zinn, P. O. *et al.* Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme. *PLOS ONE* **6**, e25451 (2011).
20. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **52**, 91–118 (2003).
21. Haralick, R. M., Shanmugam, K. & Dinstein, I. H. Textural features for image classification. *IEEE Trans. Syst., Man, Cybern.* **6**, 610–621 (1973).
22. Galloway, M. M. Texture analysis using gray level run lengths. *Comput. Vision Graph.* **4**, 172–179 (1975).
23. Deasy, J. O., Blanco, A. I. & Clark, V. H. CERR: a computational environment for radiotherapy research. *Med. Phys.* **30**, 979–985 (2003).
24. Harrell, F. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis, Ch. **10**, 247–249 (Springer, 2001).
25. Schröder, M. S., Culhane, A. C., Quackenbush, J. & Haibe-Kains, B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* **27**, 3206–3208 (2011).
26. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
27. Theodoridis, S. & Koutroumbas, K. *Pattern Recognition* 4th edn, Ch. **16**, 866–872 (Elsevier Science, 2009).
28. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
29. Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **45**, 171–186 (2001).
30. Yee, T. W. The VGAM package for categorical data analysis. *J. Stat. Softw.* **32**, 1–34 (2010).
31. Balagurunathan, Y. *et al.* Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Transl. Oncol.* **7**, 72–87 (2014).

## Acknowledgements

Authors also acknowledge financial support from the National Institute of Health (NIH-USA U01 CA 143062-01, and NIH-USA U01 CA 190234-01), EU 7th framework program (EURECA, ARTFORCE), Kankeronderzoekfonds Limburg from the Health Foundation Limburg and the Dutch Cancer Society (KWF UM 2009-4454, KWF MAC 2013-6425).

## Author Contributions

H.J.W.L.A., C.P., R.T.H.L. and P.G. conceived of the project, analysed the data, and wrote the paper. E.R.V., J.B., D.R., M.M.R., B.H.K., and P.L. provided expert guidance, data, or analysis tools and reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Parmar, C. *et al.* Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci. Rep.* **5**, 11044; doi: 10.1038/srep11044 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>