



# Radiological tumour classification across imaging modality and histology

Jia Wu<sup>1,2,3,11</sup>✉, Chao Li<sup>1,4,5,11</sup>, Michael Gensheimer<sup>1</sup>, Sukhmani Padda<sup>6</sup>, Fumi Kato<sup>7</sup>, Hiroki Shirato<sup>8</sup>, Yiran Wei<sup>5</sup>, Carola-Bibiane Schönlieb<sup>9</sup>, Stephen John Price<sup>5</sup>, David Jaffray<sup>2,10</sup>, John Heymach<sup>3</sup>, Joel W. Neal<sup>6</sup>, Billy W. Loo Jr<sup>1</sup>, Heather Wakelee<sup>6</sup>, Maximilian Diehn<sup>1</sup> and Ruijiang Li<sup>1</sup>✉

**Radiomics refers to the high-throughput extraction of quantitative features from radiological scans and is widely used to search for imaging biomarkers for the prediction of clinical outcomes. Current radiomic signatures suffer from limited reproducibility and generalizability, because most features are dependent on imaging modality and tumour histology, making them sensitive to variations in scan protocol. Here, we propose novel radiological features that are specially designed to ensure compatibility across diverse tissues and imaging contrast.** These features provide systematic characterization of tumour morphology and spatial heterogeneity. In an international multi-institution study of 1,682 patients, we discover and validate four unifying imaging subtypes across three malignancies and two major imaging modalities. These tumour subtypes demonstrate distinct molecular characteristics and prognoses after conventional therapies. In advanced lung cancer treated with immunotherapy, one subtype is associated with improved survival and increased tumour-infiltrating lymphocytes compared with the others. Deep learning enables automatic tumour segmentation and reproducible subtype identification, which can facilitate practical implementation. The unifying radiological tumour classification may inform prognosis and treatment response for precision medicine.

Radiological imaging is an integral component of cancer care and is used for screening, diagnosis and staging, as well as for evaluation of treatment response and surveillance for disease relapse. Beyond its routine clinical applications, imaging can also provide rich information about tumour phenotypes, which are fundamentally governed by the underlying biological processes of the malignancy<sup>1</sup>. This is achieved by the high-throughput extraction of quantitative image features from standard-of-care radiological scans. This approach, known as radiomics, has been extensively applied in various cancer types with the goal of predicting therapy response and outcomes<sup>2</sup>. In addition, specific radiomic features have been linked to genetic and molecular characteristics of the tumour and its microenvironment<sup>3–8</sup>.

Currently, radiomics analysis in any given study is focused on one imaging modality and one cancer type only. This is necessary, because most radiomics features (for example, texture) are highly sensitive to variations in image intensity, making them modality-dependent and histology-specific<sup>9,10</sup>. In clinical practice, each disease has a preferred imaging modality that maximizes tissue contrast, for instance, computed tomography (CT) in lung cancer and magnetic resonance imaging (MRI) in breast cancer. As a result, it has been challenging to apply radiomics signatures identified in a cohort with a given disease and imaging modality to other settings, which limits their reproducibility and generalizability<sup>11–13</sup>.

The Cancer Genome Atlas (TCGA) consortium recently performed an integrated molecular analysis of over 10,000 tumours in 33 cancer types<sup>14</sup>. With their greater statistical power, such pan-cancer studies may help identify commonly conserved patterns and unifying biological themes across cancers. Microsatellite instability, neurotrophin receptor tyrosine kinase gene fusions and tumour mutational burden are prime examples of tissue-agnostic biomarkers that are used to select patients for specific treatment, regardless of tumour histology<sup>15</sup>.

Given the availability of standardized genomic and transcriptomic data, current pan-cancer studies have mainly focused on molecular aspects of the tumour. In the context of radiological imaging, however, pan-cancer studies are notably lacking, primarily hampered by lack of standardization and diverse tissue contrast in different modalities. Compared with the conventional radiomic approach focusing on one cancer type and one imaging modality, a cross-histology and cross-modality strategy may lead to the discovery of unifying imaging phenotypes that are conserved across multiple cancer types and imaging modalities, which will have a broad impact on a larger population of patients.

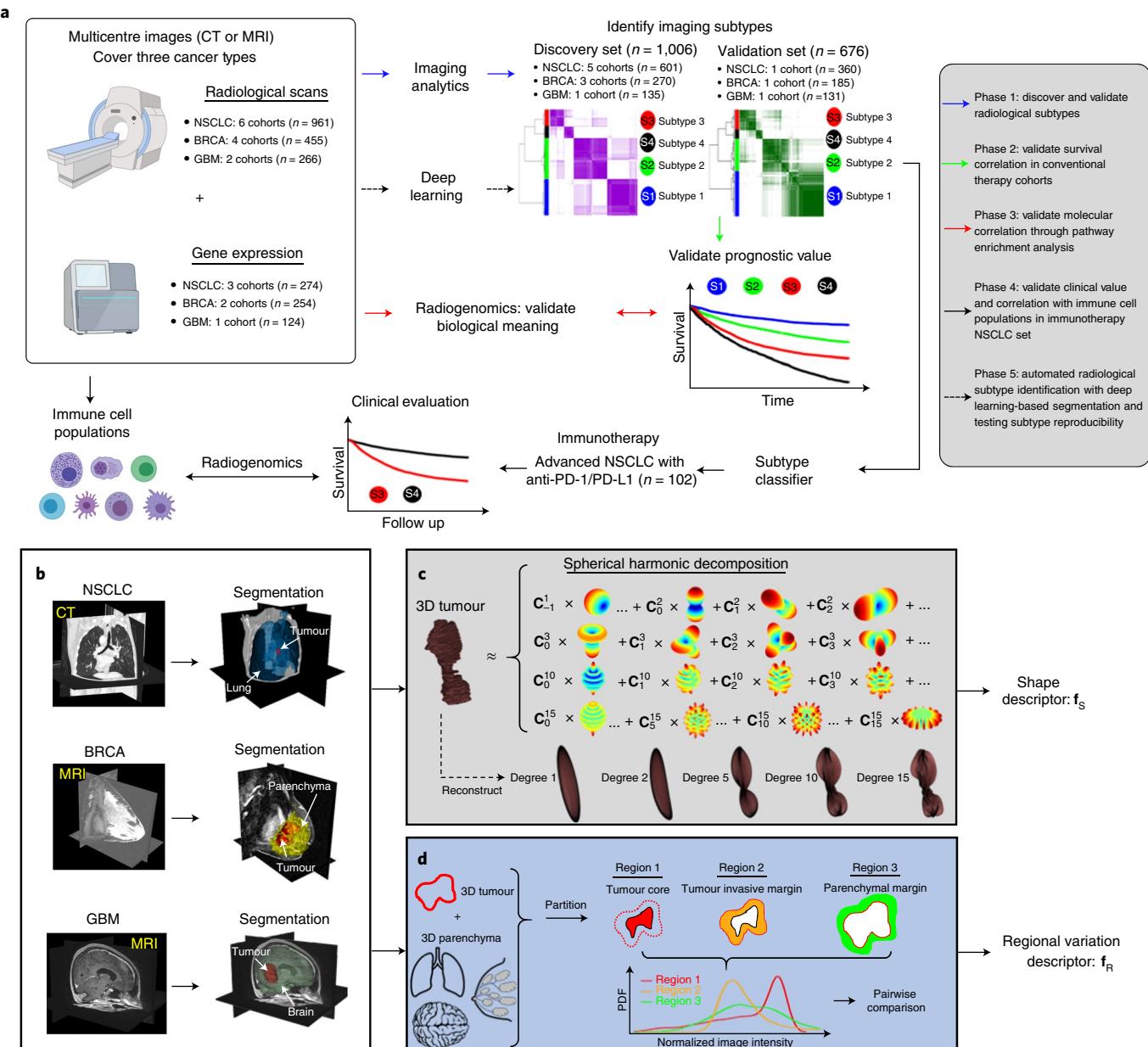
In this Article, we propose novel radiological features to systematically characterize tumour morphology and spatial heterogeneity. These features were meticulously designed to ensure that they are comparable across diverse tissues and imaging contrast. Based on multi-institutional cohorts that span three cancer types

<sup>1</sup>Department of Radiation Oncology, Stanford University School of Medicine, Palo Alto, CA, USA. <sup>2</sup>Department of Imaging Physics, MD Anderson Cancer Center, Houston, TX, USA. <sup>3</sup>Department of Thoracic and Head and Neck Medical Oncology, MD Anderson Cancer Center, Houston, TX, USA. <sup>4</sup>The Centre for Mathematical Imaging in Healthcare, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK.

<sup>5</sup>Cambridge Brain Tumor Imaging Laboratory, Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK.

<sup>6</sup>Department of Medicine, Division of Oncology, Stanford University School of Medicine, Stanford, CA, USA. <sup>7</sup>Department of Diagnostic and Interventional Radiology, Hokkaido University Hospital, Sapporo, Hokkaido, Japan. <sup>8</sup>Global Center for Biomedical Science and Engineering, Faculty of Medicine, Hokkaido University, Sapporo, Hokkaido, Japan. <sup>9</sup>Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. <sup>10</sup>Office of the Chief Technology and Digital Officer, MD Anderson Cancer Center, Houston, TX, USA. <sup>11</sup>These authors contributed equally: Jia Wu, Chao Li.

✉e-mail: [JWu1@mdanderson.org](mailto:JWu1@mdanderson.org); [rli2@stanford.edu](mailto:rli2@stanford.edu)



**Fig. 1 | Overview of the study design and quantitative imaging analysis.** **a**, Study design, comprising five phases. **b-d**, Illustration of the proposed image feature extraction pipeline. First, the primary tumour is manually delineated and the surrounding parenchymal tissues (lung, fibro-glandular and brain) are automatically segmented. Two broad categories of image feature are calculated (**b**): systematic shape descriptors through spherical harmonic decomposition (**c**) and spatial heterogeneity described by regional variations among tumour core, tumour invasive margin and parenchymal region (**d**). NSCLC, non-small cell lung cancer; BRCA, breast cancer; GBM, glioblastoma multiforme.

and two major modalities, we identify and validate four unifying imaging subtypes that are associated with distinct oncogenic processes and prognoses after conventional therapies. We further demonstrate their potential predictive value in patients treated with immunotherapy.

## Results

**Overview of study design.** We sought to define a radiological tumour classification scheme that is broadly applicable across cancer types and imaging modalities (Fig. 1a). To achieve this goal, we conducted an international multicentre study from 12 independent cohorts with a total of 1,682 patients with cancer recruited from the United States, Europe and Japan (Supplementary Table 1). Our analysis included three cancer types (lung, breast and brain malig-

nancies) and two imaging modalities (CT and MRI). All patients had a pathologically confirmed diagnosis of primary malignancy and had received standard therapies including surgery, radiation, chemotherapy and/or hormonal therapy. We collected their pre-treatment radiological scans and, if available, tumour molecular profiles and clinicopathological and outcome data (Supplementary Tables 2–4). To rigorously assess the reproducibility of our findings, we divided the 12 primary cohorts into discovery and independent validation sets as stratified by cancer types and imaging modalities. Furthermore, we assessed the clinical relevance of the imaging subtypes in a separate cohort of 102 patients with advanced lung cancer treated with anti-programmed cell death protein 1/programmed death ligand 1 (anti-PD-1/PD-L1) immune checkpoint blockade (Supplementary Table 5).

**Radiological features applicable across histology and imaging modality.** We proposed two broad categories of quantitative features to characterize radiological phenotypes: tumour morphology and spatial heterogeneity (Fig. 1b–d). These features are specifically designed to account for diversity in tissue origin and imaging contrast. For morphology, we transformed the 3D tumour shape into a structured sequence of coefficients through spherical harmonic decomposition (Extended Data Fig. 1a). These coefficients provide an unbiased, optimal representation of the original tumour shape in the spatial frequency domain and can be used to fully reconstruct any tumour shape via a unique one-to-one mapping (Extended Data Fig. 1b,c). For spatial heterogeneity, we analysed three distinct regions of interest: tumour core, intratumoral invasive margin and peritumoral parenchymal margin. Based on these non-overlapping regions, we defined 20 features to describe multiregional intensity variations through a pairwise comparison of normalized probability distribution functions. In total, 313 image features were extracted for each tumour.

We first confirmed that these image features were distributed similarly and thus comparable across imaging modality and cancer type (Extended Data Fig. 2). To test the reproducibility of extracted features against variation in tumour delineation, we randomly perturbed the original tumour contour and observed an overall high consistency with intraclass correlation coefficient of 0.83, indicating that our proposed features are robust to variation in tumour segmentation (Supplementary Fig. 1).

**Efficient encoding of radiological features using machine learning.** We split the patients in the 12 primary cohorts into a discovery set and a validation set stratified by cancer types, with the discovery set containing 601 patients with lung cancer, 269 with breast cancer and 136 with glioblastoma multiforme (GBM), and the validation set containing 360 patients with lung cancer, 185 with breast cancer and 131 with GBM. Given the large number of features, we trained an artificial neural network (ANN) called ‘autoencoder’ in the discovery cohort to learn efficient representations of the original features and mitigate the curse of dimensionality (Extended Data Fig. 3a). The trained autoencoder was tested in the validation set (Extended Data Fig. 3b). Instead of linear principal component analysis (PCA), we used autoencoder for feature representation due to its capability of modelling complex nonlinear patterns. After training, the autoencoder substantially reduced the dimension of features to 10 from 313, while at the same time maximally preserving information in the original data (Extended Data Fig. 3b,c). The feature dimension reduction has the dual advantages of effectively removing redundancy in the features and improving the stability and efficiency of subsequent clustering analysis.

**Discovery and validation of tumour subtypes across histology and imaging modality.** In the discovery set ( $n=1,006$ ), based on the consensus clustering of radiological tumour phenotypes encoded by 10 autoencoder features, we determined the optimal cluster number to be four (Fig. 2a and Supplementary Fig. 2a), which maximized consensus within clusters while minimizing ambiguity in cluster assignments. Next, we independently applied the same consensus clustering analysis and trained autoencoder in the validation set ( $n=676$ ) and also identified four clusters of patients (Fig. 2b and Supplementary Fig. 2b). To measure the reproducibility of clusters across the discovery and validation sets, we computed the in-group proportion (IGP) statistics, which showed high consistency for four clusters, with IGP values of 92.4% ( $P<1\times 10^{-10}$ ), 91.1% ( $P=0.002$ ), 83.2% ( $P=0.008$ ) and 78.0% ( $P=0.021$ ), respectively.

We then split the patients based on imaging modalities, that is, CT versus MRI. Based on the discovery set of CT features in 961 lung cancer patients, we again identified four clusters (Fig. 2c and Supplementary Fig. 2c). Similarly, four clusters were identified in

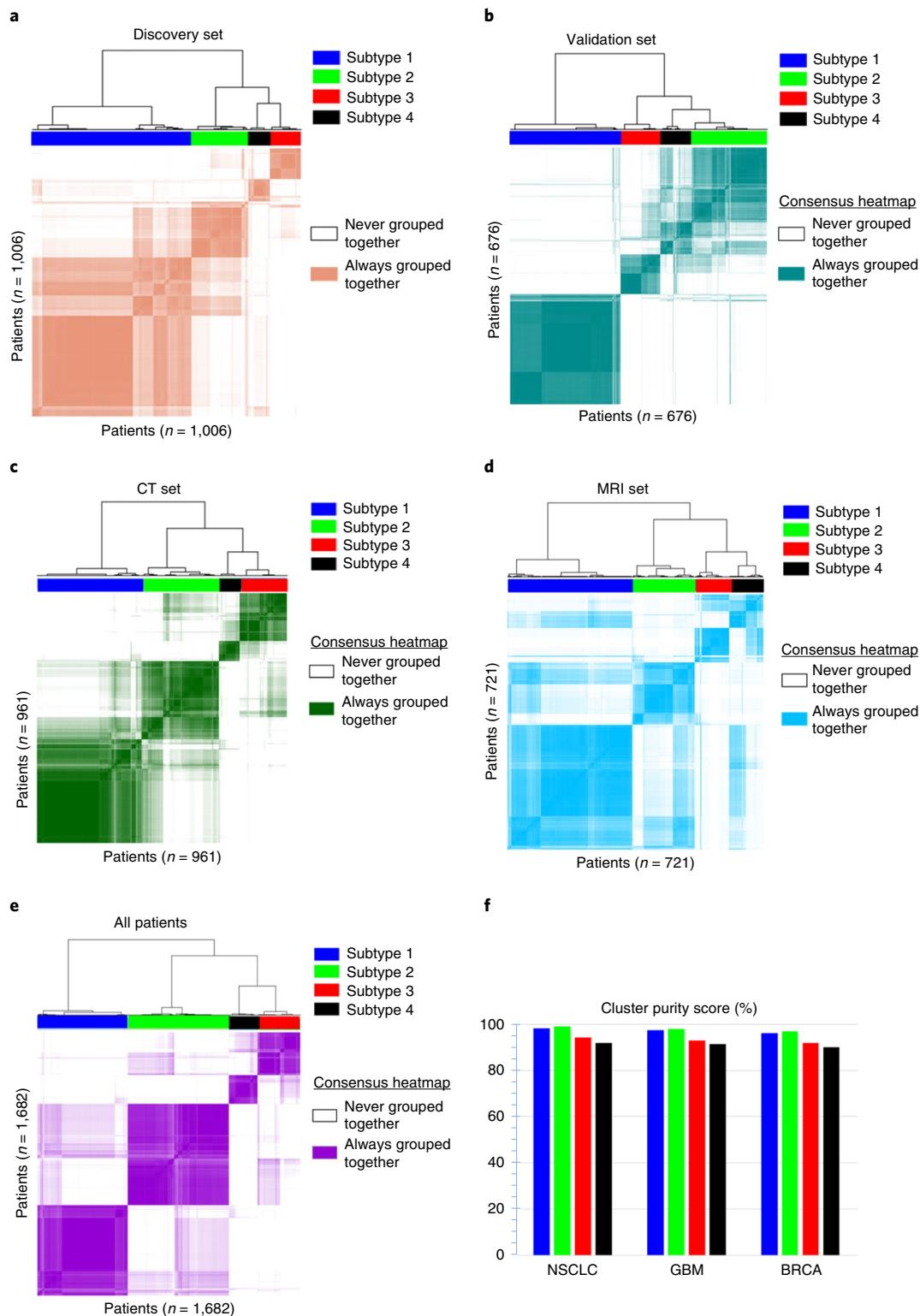
the validation set based on MRI features in 454 patients with breast cancer and 267 with GBM (Fig. 2d and Supplementary Fig. 2d). These clusters were found to be highly consistent across imaging modality, with IGP values of 81.0% ( $P=0.032$ ), 80.0% ( $P=0.041$ ), 91.8% ( $P=0.014$ ) and 91.2% ( $P=0.008$ ), respectively.

After confirming the consistency of patient clusters across imaging modalities and cancer types, we refined the subtypes by using all patients in the primary cohorts (Fig. 2e and Supplementary Fig. 2e). Overall, this resulted in 580 patients (34.5%) in cluster 1, 647 patients (38.5%) in cluster 2, 272 patients (16.2%) in cluster 3 and 183 patients (10.9%) in cluster 4. All four clusters were represented in each of the three cancer types (Extended Data Fig. 4a). The distribution of clusters was largely independent of cancer type (Cramér’s  $V=0.21$ ,  $P=0.073$ ). We further evaluated the accuracy of the clustering results (details are provided in the Methods). Overall, the cluster purity scores were 91–97%, confirming the validity of the clustering results (Fig. 2f). Within each cancer type and imaging modality, the clusters remained highly reproducible. These histology and modality-independent clusters are hereafter named imaging subtypes 1 to 4.

**Imaging characteristics distinguishing the tumour subtypes.** To better understand specifically which factors contribute to the distinction among the subtypes, we performed differential analysis by using the original 313 imaging features while controlling for multiple testing (Extended Data Fig. 5). We identified four main categories of image feature associated with the subtypes, which measure tumour volume, shape symmetry, shape regularity and regional variation (Fig. 3a,b). Based on this analysis, we summarize the radiological interpretations of four subtypes in Fig. 3c. Specifically, subtype 1 mainly consists of small tumours with large variations across the tumour–parenchyma interface (that is, a distinct, sharp margin). Subtype 2 consists of intermediate-sized tumours with a moderately well-defined margin. Subtype 3 consists of large tumours with an ill-defined, blurred margin. Subtype 4 is characterized as large tumours with a moderately well-defined margin. For tumour morphology, subtypes 1 to 3 had similarly smooth and regular shapes, while subtype 4 had the highest complexity with a rugged and irregular shape. Figure 3d shows a graphical representation of the four imaging subtypes in the feature space.

**Clinical validation for the prognostic significance of the imaging subtypes.** We tested the prognostic relevance of the imaging subtypes and observed significant differences in patient prognosis within each cancer type (Fig. 4a–c). In non-small cell lung cancer (NSCLC), subtype 1 was associated with the best survival, subtypes 2 and 3 with intermediate prognoses, and subtype 4 with the worst prognosis. Consistently, in breast cancer and GBM, patients in subtype 4 also had the worst prognosis. In multivariate analysis, subtype 4 remained an independent predictor for poor survival after adjusting for stage, tumour volume, clinicopathological factors and specific molecular features (Fig. 4d–f and Supplementary Tables 6 and 7).

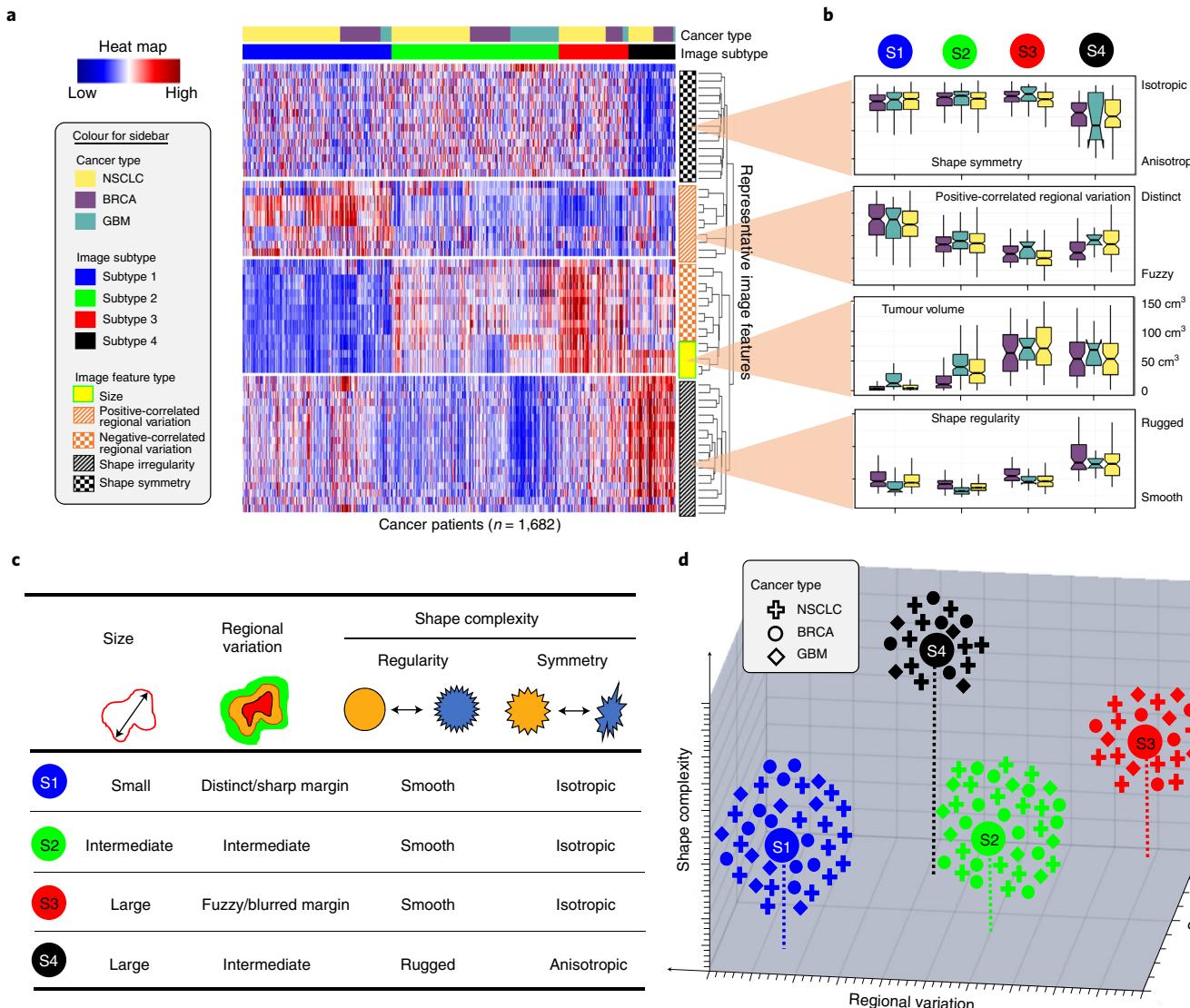
We further evaluated the prognostic significance of the imaging subtypes by focusing on patients in clinically relevant subgroups. Specifically, we confirmed that their prognostic value was largely independent of disease stage and primary treatment in NSCLC. The distribution of subtypes for clinical stage in NSCLC is shown in Extended Data Fig. 4b. Moreover, the imaging subtypes remained prognostic within early (stage I/II) as well as locally advanced (stage III) NSCLC (Extended Data Fig. 6a,b). One notable exception is subtype 4, which was associated with a worse prognosis compared with subtype 3 for patients treated with surgery (hazard ratio (HR)=1.82, 95% confidence interval (CI) 0.89–3.75,  $P=0.064$ ). On the other hand, subtype 4 appeared to have an improved prognosis that is similar to subtype 3 for patients treated with radiotherapy



**Fig. 2 | Identification of unifying tumour subtypes based on unsupervised consensus clustering of the extracted image features across three cancer types and across two modalities (CT and MRI).** **a-e**, The consensus matrix corresponding to the optimal cluster number ( $k=4$ ) for the discovery set (a), the validation set (b), the CT set (c), the MRI set (d) and the whole population (e). Patients are both rows and columns. The matrices are ordered by consensus-clustered groups, depicted as dendrograms above the heatmap. **f**, The cluster purity score of four tumour subtypes in three individual cancer types.

(HR = 0.94, 95% CI 0.61–1.46,  $P=0.8$ ; Extended Data Fig. 6c,d). Fewer patients in subtypes 3 and 4 received surgery compared with radiotherapy ( $n=67$  versus 160). We tested the interaction between

therapeutic regimen (surgery versus radiotherapy) and subtype 3 versus 4 and found a possible interaction effect ( $P=0.063$ ). Moreover, we performed subgroup analysis by stratifying tumours



**Fig. 3 | Radiological characteristics of the unifying tumour subtypes.** **a**, Heatmap of four subtypes with respect to the original imaging features. **b**, Boxplots of four representative groups of features including tumour volume, positive-correlated regional variation, shape symmetry and shape irregularity, stratified by imaging subtypes as well as cancer types. The whiskers correspond to minimum and maximum values, and boxes corresponding to lower quartile, median and upper quartile. **c**, Summary of the key imaging characteristics of the four subtypes. **d**, Schematic of the distribution of the imaging subtypes in a 3D space formed by tumour size, shape complexity and regional variation.

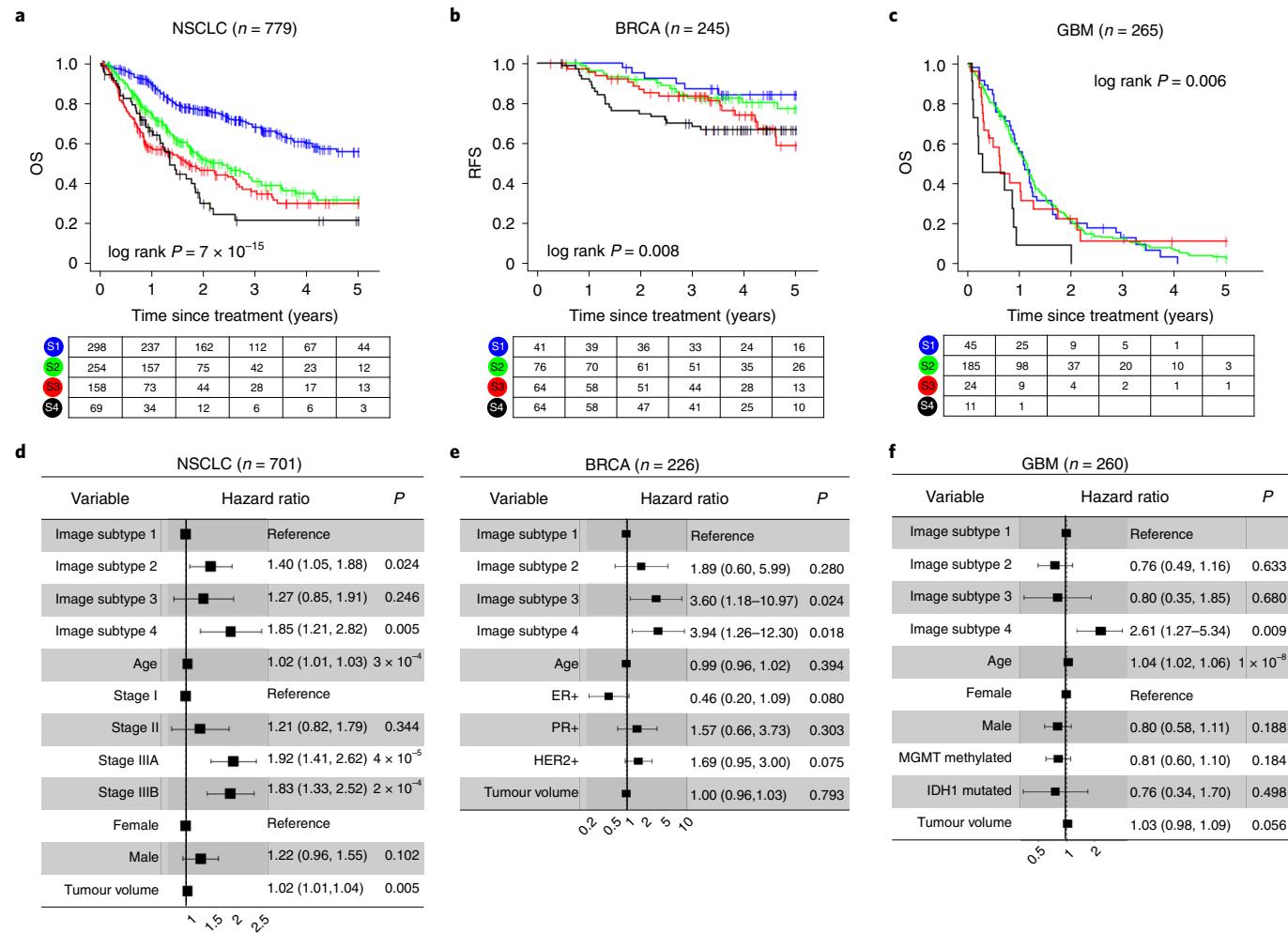
according to driver gene alterations in NSCLC, such as *EGFR* and *ALK* (Extended Data Fig. 7a–c).

Similarly, we tested the imaging subtypes in different histological subtypes of breast cancer and confirmed prognostic significance, especially in human epidermal growth factor receptor 2 (HER2)-positive patients (Extended Data Fig. 7d–f). For GBM, we tested the survival stratification by imaging subtypes in different patient subgroups defined by *O<sup>6</sup>-methylguanine-methyltransferase (MGMT)* methylation or *IDH1* mutation status (Extended Data Fig. 7g–i).

To further demonstrate the advantage of using these cross-modality and cross-histology image features, we compared the performance for prognostic prediction using the conventional radiomic approach. Specifically, we trained a radiomic model in a supervised manner to predict the survival of patients with lung cancer (Extended Data Fig. 8a,b). We observed that the radiomics risk score was highly correlated with tumour size (Pearson correlation coefficient = 0.92, Extended Data Fig. 8c), and the radiomics model

had a C-index of 0.60 (95% CI: 0.56–0.64) in the validation cohort. By comparison, our proposed imaging subtypes were independent of tumour volume and achieved a C-index of 0.67 (95% CI: 0.63–0.72) in the validation cohort, which is significantly better than the radiomics model ( $P < 0.001$ , Extended Data Fig. 8d).

**Biological validation for the molecular basis of the imaging subtypes.** We performed gene set enrichment analysis (GSEA) to identify molecular pathways associated with the imaging subtypes. In NSCLC, for subtype 1 with the most favourable prognosis, the majority of cancer hallmark pathways, including proliferation, angiogenesis and hypoxia, were significantly downregulated compared with other subtypes (Extended Data Fig. 9a and Supplementary Table 9). By contrast, many cancer hallmarks, including glycolysis and metastasis-related pathways, were upregulated in subtype 4, suggesting a more aggressive phenotype consistent with the poor survival in these patients (Extended Data Fig. 9b and Supplementary Table 10). Of note, we also observed an increased expression of



**Fig. 4 | Evaluation of prognostic value of the four imaging subtypes in three individual cancer types.** **a–c**, Kaplan–Meier curves for overall survival (OS) in NSCLC (**a**), recurrence-free survival (RFS) in BRCA (**b**) and overall survival in GBM (**c**). **d–f**, Forest plots showing the hazard ratio and *P* values obtained from a multivariate Cox regression analysis including the proposed imaging subtypes and established clinicopathologic factors in NSCLC (**d**), BRCA (**e**) and GBM (**f**).

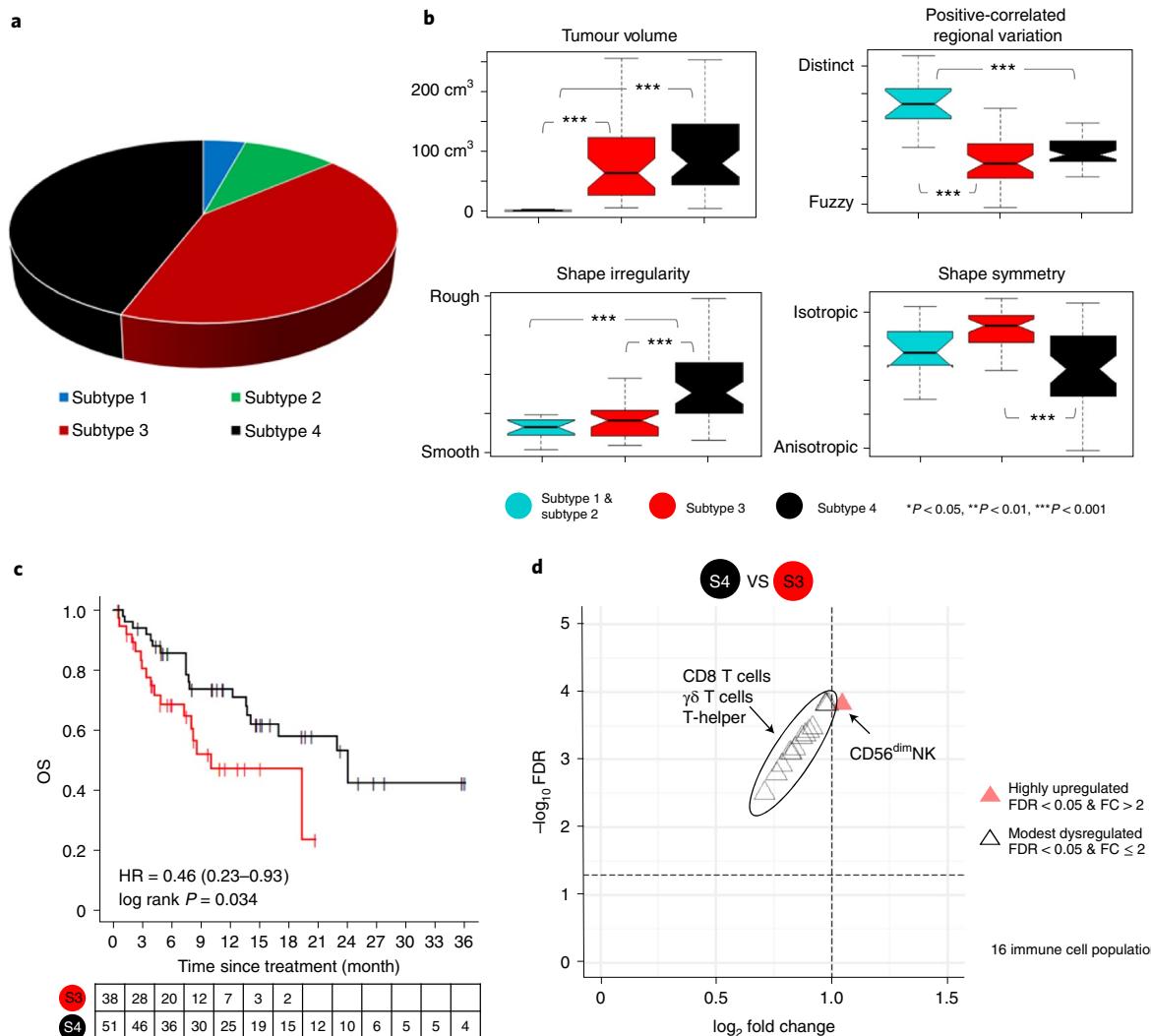
genes related to radiation response, such as cell cycle, apoptosis and DNA repair, consistent with their increased survival when treated with radiotherapy.

Similar enrichment analyses were also performed separately for breast cancer and GBM. However, we did not observe any pathways that reached the predefined statistical significance level at a false discovery rate (FDR) of  $<0.05$ , probably due to the limited statistical power given the small number of samples (Extended Data Fig. 9c–f). We further investigated the distribution of established molecular subtypes among the different imaging subtypes (Extended Data Fig. 4c,d), and these were independent of each other, with Cramér's  $V=0.18$  ( $P=0.539$ ) for breast cancer and 0.16 ( $P=0.490$ ) for GBM.

**Imaging subtypes stratify survival in patients treated with immunotherapy.** The imaging subtypes were discovered and validated in 12 primary cohorts of patients treated with conventional therapies. We sought to further evaluate their clinical relevance in the immunotherapy setting in a completely independent cohort. To this end, we collected data from an institutional cohort of 102 patients with advanced NSCLC who were treated with immune checkpoint blockade. Given the extracted CT features of primary tumours, we trained a multiclass classification model using XGBoost (Supplementary Methods) to predict the imaging subtype of each new patient.

The majority of these patients with advanced NSCLC were classified as either subtype 3 (37%) or subtype 4 (50%), as shown in Fig. 5a, and both subtypes had large tumours. This is consistent with the fact that advanced-stage tumours tend to be larger than early-stage tumours. We then confirmed that the distributions of characteristic imaging features followed the same patterns in both the primary and immunotherapy cohorts (Fig. 5b). In particular, tumours in subtype 3 had a regular shape and ill-defined, blurred margin, whereas those in subtype 4 had irregular shape and a moderately well-defined margin.

There was no statistically significant difference in survival between patients in subtypes 1&2 and subtype 4 (Extended Data Fig. 10). Strikingly, we observed that patients in subtype 4 had significantly better survival compared with subtype 3 (HR = 0.46, 95% CI: 0.23–0.93,  $P=0.034$ ), as shown in Fig. 5c. This is in stark contrast with the results in our primary cohorts treated with conventional therapies, where patients in subtype 4 had a significantly worse prognosis. We further compared the immune cell composition in the tumour microenvironment between subtypes 3 and 4. Tumours in subtype 4 had a higher infiltration of several adaptive immune cell populations, including activated CD56<sup>dim</sup> natural killer (NK) cells, cytotoxic CD8 T cells, CD4 T helper cells and  $\gamma\delta$  T cells (Fig. 5d and Supplementary Table 11).



**Fig. 5 | Clinical evaluation of the imaging subtypes in advanced lung cancer treated with immunotherapy.** **a**, Distribution of inferred imaging subtypes. **b**, Representative features stratified by imaging subtypes, including tumour volume, positive-correlated regional variation, shape symmetry and shape irregularity. **c**, Kaplan-Meier curves of overall survival stratified by imaging subtype 3 versus 4. **d**, Comparison of tumour-infiltrating immune cell populations between subtypes 4 and 3.

**Deep learning enables automated tumour segmentation and reproducible identification of imaging subtypes.** Our previous analysis of the imaging subtypes requires the tumour contour, which involves manual delineation and is subject to inter- and intra-rater variability. To overcome this issue and facilitate practical implementation, we trained a deep learning model based on the two-dimensional (2D) U-Net (Fig. 6a) to perform automated tumour segmentation. The U-Net model performed well at segmenting tumours in NSCLC, with average Sørensen-Dice coefficients of 0.90, 0.88 and 0.84 for training, validation and testing, respectively (Fig. 6b). This model performed especially well for large tumours ( $>10\text{ cm}^3$ ), with Sørensen-Dice coefficient  $>0.9$ , and for tumours not attached to mediastinum, with Sørensen-Dice coefficient  $>0.85$  (Fig. 6c,d). On the other hand, the U-Net model did not perform well in breast cancer, with Sørensen-Dice coefficients of 0.73, 0.72 and 0.69 for training, validation and testing, respectively (Supplementary Fig. 3). For GBM, tumour segmentation was performed with a pre-trained deep learning model.

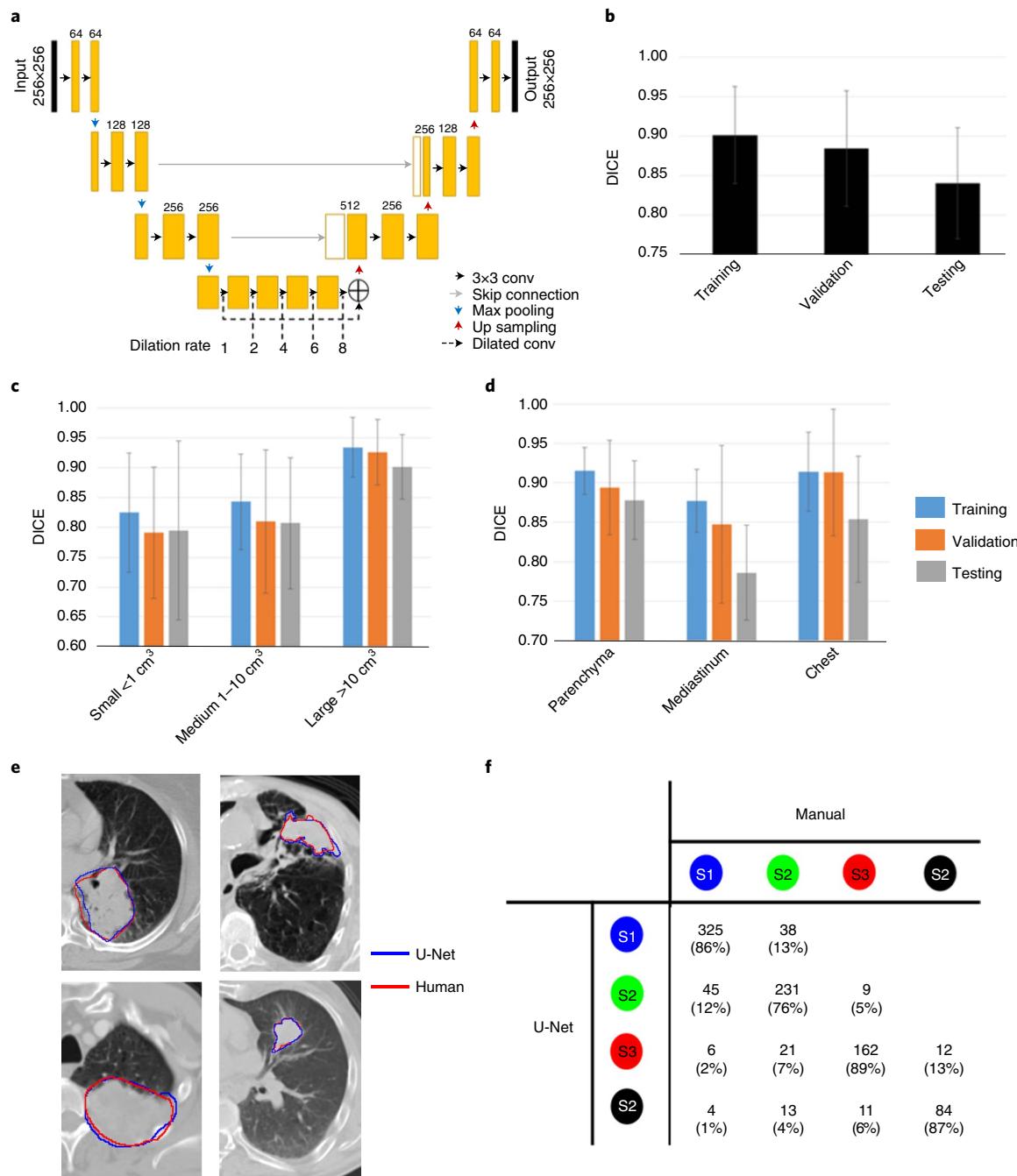
Finally, we applied the automatically generated tumour segmentations of lung patients and repeated the entire pipeline for imaging

subtype discovery (Fig. 6e). The clustering results were highly reproducible, with the cluster assignment for the majority (83.5%) of patients remaining unchanged based on manual versus deep learning generated segmentations (Fig. 6f).

## Discussion

In this international multi-cohort study of 1,682 patients, we discovered and independently validated novel tumour subtypes that are broadly applicable across major imaging modalities and three cancer types. These subtypes demonstrate distinct radiological and molecular features, as well as survival outcomes after conventional therapies. Moreover, their prognostic value was independent of established clinical risk factors, including tumour volume. Importantly, we showed that specific imaging subtypes are associated with differential outcomes after immunotherapies. Finally, we demonstrated that deep learning can be used to automate tumour segmentation, which will help standardize subtype identification and facilitate its implementation in clinical practice.

Radiologic-based classification has several important advantages over histopathologic assessment or molecular tumour profiling.



**Fig. 6 | Deep learning to automate 3D tumour segmentation.** **a**, Detailed architecture of the proposed U-Net with dilation at bottleneck layers. conv, convolutional layer. **b**, Sørensen-Dicecoefficients of the trained U-Net model in NSCLC in training, validation and testing sets. **c,d**, Sørensen-Dice (DICE) coefficients of U-Net stratified by tumour size (**c**) and anatomical location of the tumour (**d**). **e**, Representative CT slices and tumour contours of four different testing patients. **f**, Confusion matrix of clustering results based on manual tumour contouring and automated segmentation for NSCLC patients.

First, imaging evaluation is performed for the entire 3D tumour *in situ*, which overcomes the sampling bias of a small biopsy due to intratumoral spatial heterogeneity. Second, imaging provides a non-invasive means for evaluation, which would allow longitudinal monitoring and follow-up of disease. Because radiomic features provide a representation of tumour phenotypes from a radiological perspective, the proposed radiologic-based tumour classification could complement current clinical and molecular classification. We envision that integration of information from different domains may further enhance the prediction of treatment response and prognosis for precision oncology<sup>16,17</sup>.

It is important to emphasize that the purpose of our work is not to develop a predictive model for a specific clinical endpoint (such as prognosis) in a particular cancer as is done in the traditional radiomics approach with supervised machine learning. Rather, our study is aimed at the discovery of unifying radiological phenotypes across different imaging modalities and cancer types. By identifying these converging radiographic hallmarks, we further revealed biological insights and established their connection with prognosis and response to different cancer treatments including surgery, radiation and chemotherapy, as well as immunotherapy.

Previous works on radiomics have been limited to one cancer type and one imaging modality in any given study. Consequently, it is difficult to generalize the findings, which are often based on relatively small datasets<sup>13</sup>. Here, we aimed to discover common radiological patterns with biological and clinical relevance across histology and modality. To achieve this goal, we meticulously designed image features that are robust to tissue contrast. This study provides a conceptual framework that will allow aggregation of datasets with disparate modalities and cancer types, similar to integration of molecular data in pan-cancer studies by TCGA.

Despite extensive investigations in radiomics, progress towards clinical translation has been slower than anticipated<sup>12,13</sup>. One major reason is a lack of reproducibility and rigorous validation, given the exploratory nature and inclusion of small numbers of patients in many studies<sup>12,18</sup>. Textural features, which usually represent the overwhelming majority of features in a typical radiomics study, have been shown to be highly sensitive to imaging protocol and technical factors<sup>9</sup>. This has precluded meaningful comparison of existing radiomics signatures across studies with different modalities and various cancer types<sup>11</sup>.

To overcome these issues, we focused on two broad categories of radiological feature (tumour morphology and spatial heterogeneity) and made special efforts to ensure their compatibility across imaging modalities. Geometric features are computed based on a binary image of the tumour contour, and are insensitive to tissue contrast and can be readily comparable across different imaging modalities and cancer types<sup>19</sup>. Specifically, we applied the spherical harmonic decomposition method to approximate the 3D tumour shape with a series of orthonormal basis functions defined on the surface of the sphere. Similar to Fourier analysis of time series, this allows us to interrogate, in an unbiased manner, the geometric patterns with an ordered sequence of coefficients in the frequency domain. Previous morphological features only focus on the low-frequency component (Extended Data Fig. 1c, degree 1) and disregard rich information captured at higher spatial frequencies (degrees 2 to 15).

Although texture features are widely used as a measure of tumour heterogeneity, current approaches typically do not distinguish different regions within the tumour; most are focused on the primary tumour, ignoring the surrounding parenchyma. There is emerging evidence supporting the clinical relevance of these different regions<sup>20,21</sup>. Here, we measure tumour heterogeneity by performing spatially explicit analysis of intratumoral and peritumoral regions, as well as the invasive margin. By making pairwise comparison of normalized density functions and using each region as an internal control, we generate features that are robust to variability in tissue contrast. It is worth noting that our tumour subregions are defined from a ‘geometric’ perspective, which is simple to implement and is generally applicable across cancer types and imaging modalities. On the other hand, physiological tumour subregions or habitats such as hypoxia and hypermetabolic activity can provide more meaningful biological information, and they do not necessarily follow this simple geometric paradigm<sup>22,23</sup>. One caveat is that these physiological tumour subregions are probably cancer type-specific and imaging modality-dependent, and reliable identification of these subregions requires sophisticated algorithms, such as habitat imaging<sup>24,25</sup>.

Here, we have shown that the imaging subtypes are associated with distinct survival outcomes in three cancer types after conventional therapies. Furthermore, the prognostic value of our imaging subtype is independent of tumour volume and other clinical factors. A key limitation of previous radiomic studies is that some signatures may be correlated with tumour burden, which would diminish their clinical relevance<sup>26</sup>. Similarly, we also observed that the radiomics model is highly correlated with tumour size (Extended Data Fig. 8c). This underscores the importance that any useful biomarker must provide additional value beyond established risk factors.

The new imaging subtypes may have therapeutic implications. In particular, we found that subtype 4 was associated with improved survival in NSCLC patients treated with immune checkpoint blockade. Consistently, we found a higher infiltration of cytotoxic NK and T lymphocytes in the tumour microenvironment of subtype 4. These results suggest that subtype 4 tumours are associated with a pre-existing antitumour immunity and thus may preferentially respond to immunotherapy<sup>27</sup>. Interestingly, a recent study developed a CT-based radiomics signature specifically for estimating tumour-infiltrating CD8 T cells, which was correlated to survival after immunotherapy<sup>4</sup>. These findings are based on retrospective analysis and will require prospective validation in future randomized trials.

For clinical implementation, different imaging modalities used in different malignancies may introduce some practical challenges. For instance, in lung cancer imaging, normal breathing can induce some degree of tumour motion depending on its anatomic location, which may cause blurring artefacts in the CT image. Methods for dynamic lung modelling and tumour tracking may be useful<sup>28</sup>. Modern multi-slice CT scanners allow fast imaging with a breath-hold protocol and can largely mitigate this issue. On the other hand, MRI is subject to its own source of variability due to the use of various sequence protocols and parameters. Further development in quantitative imaging with standardized acquisition should improve image quality and the reproducibility of radiomic biomarkers<sup>29</sup>. In breast cancer, MRI may be acquired after tumour biopsy, which can introduce imaging artefacts and affect the calculation of certain features related to spatial variation.

One limitation of the radiogenomic analysis is that gene expression profiling was performed for the bulk tumour from surgical specimens. Single-cell gene expression analysis can provide much refined details and offer a deeper insight about tumour heterogeneity<sup>30</sup>. In future work, the radiogenomic association findings should be confirmed at the protein level using immunohistochemistry or immunofluorescence assays. Furthermore, it will be important to establish a ‘mechanistic’ link between the imaging subtypes and their biological underpinnings, which will require gene knock-in/knockout experiments using *in vivo* animal studies.

Deep learning, such as convolutional neural networks, has emerged as a powerful technique for medical image analysis and achieved promising performance in various clinical applications<sup>31–33</sup>. Future development of advanced deep learning techniques including physics-reinforced or physics-aware algorithms<sup>34–36</sup> can lead to further improvement for more reliable automated tumour segmentation<sup>37</sup>, which will allow consistent identification of the imaging subtypes defined here.

Finally, we note that, while finding unifying imaging phenotypes across cancers is useful, as demonstrated here, there certainly exist modality- and histology-specific features that are also important in determining disease outcomes. For example, the oedematous and diffusive growth pattern as visualized by specific MRI sequences is unique to GBM<sup>38</sup>. In this study, we mainly investigated anatomical imaging with CT and MRI, the two most widely used modalities in clinical oncology practice. Our result does not contradict the ongoing or future investigation of imaging characteristics that are specific to modality or cancer type<sup>39</sup>. More advanced, specialized imaging technologies may allow direct visualization of the functional and molecular characteristics, which could provide complementary information about the disease.

In conclusion, we have proposed a radiological tumour classification system that is applicable across imaging modality and histology. These imaging subtypes are associated with distinct oncogenic and microenvironmental features as well as survival patterns. Future studies are needed to validate the potential of this system to identify patients who are likely to benefit from immunotherapy.

## Methods

**Study design and patient cohorts.** For subtype discovery and validation, we included a total of 12 cohorts: 6 NSCLC cohorts ( $n=961$ ) with CT scans and 4 breast cancer cohorts ( $n=455$ ) and 2 GBM cohorts ( $n=266$ ) with MRI scans. Among these, data for three cohorts (one for each cancer type) were retrospectively collected from each participating centre and the remaining nine cohorts are publicly available. Follow-up and outcome data were available for 1,289 patients. We evaluated the prognostic significance of the imaging subtypes in different clinical settings. Furthermore, using the tumour gene expression data available for 652 patients, we investigated molecular correlates of the imaging subtypes. Finally, we collected data from an independent cohort of 102 patients with advanced lung cancer treated with anti-PD-1/PD-L1 immune checkpoint blockade and assessed the clinical relevance of imaging subtypes in the immunotherapy setting. Detailed clinical characteristics for the discovery cohorts and the immunotherapy validation cohort are summarized in Supplementary Tables 2 to 5 (Supplementary Methods).

**Scan acquisition and image processing.** We collected pre-treatment scans for the most commonly used imaging modality for each cancer, that is, CT for NSCLC and MRI for breast and GBM. Given the heterogeneous scan protocols, we applied a series of image-processing algorithms to harmonize the image data and facilitate robust feature extraction. The primary tumour was manually delineated in all slices by experienced physicians, and the surrounding lung, brain and breast parenchyma were automatically segmented. Details about the scan protocols, image processing and segmentation are presented in the Supplementary Methods.

**Radiological features.** We defined two broad categories of quantitative features: tumour morphology and spatial heterogeneity. For morphology, we transformed the 3D tumour shape into a structured sequence of coefficients through spherical harmonic decomposition (Extended Data Fig. 1a). Special efforts were made to ensure that the shape decompositions were invariant to scale (size), translation and rotation. Based on these coefficients, we proposed additional second-order features to summarize the shape complexity, including shape irregularity, which measures the boundary smoothness as well as shape symmetry that measures directionally dependent tumour expansion. We extracted five features to characterize tumour burden, resulting in 293 shape descriptors ( $f_s$ ). For spatial heterogeneity, we analysed three distinct regions of interest given the tumour contour and its surrounding parenchyma, designated as tumour core, intratumoral invasive margin and parenchymal (peritumoral) margin. Based on these non-overlapping regions, we defined 20 features ( $f_R$ ) to describe multiregional intensity variations through pairwise comparison of normalized probability distribution functions. The details of the image features and their interpretations are explained in the Supplementary Methods and Supplementary Table 12. In total, 313 quantitative image features were extracted for each tumour.

To evaluate the sensitivity of feature extraction with respect to variation in tumour segmentation, we generated a new set of tumour segmentation through random perturbation of the original tumour contour by applying elastic deformation in 100 randomly selected patients via the Matlab function *affine3d*. We then repeated the pipeline for image feature extraction and computed the intraclass correlation coefficient between the two sets of extracted image features.

We trained and validated an autoencoder, a type of artificial neural network used for dimensionality reduction and representation learning, to efficiently encode the original 313 image features in a low-dimensional space (Extended Data Fig. 3). We chose autoencoder rather than linear PCA to account for the complex nonlinear relationship among high-dimensional features. The dimensionality reduction will avoid undue influence of redundant features and also reduce noise. Details are described in the Supplementary Methods.

**Discovery and validation of imaging subtypes.** We identified the intrinsic imaging subtypes by applying unsupervised consensus clustering to study patients from 12 multicentre cohorts (details are provided in the Supplementary Methods). The patients were divided into discovery and validation sets based on cancer types and imaging modalities to assess the reproducibility of the clustering analysis and prevent information leakage. As there were multiple cohorts, the largest cohort within each cancer type was preselected as the validation set and the remaining ones were merged into the discovery set. The consensus clustering analysis was carried out separately in the discovery and validation cohorts, with identical configurations. In detail, the partition around the medoids clustering algorithm<sup>40</sup> with the Pearson's correlation distance metric was used as the basis for clustering. We performed 10,000 bootstraps with 80% item resampling of the autoencoder features. The optimal number of clusters was determined by varying from 2 to 10 and finding the one that produced the most stable consensus matrices and the most unambiguous cluster assignments across permuted runs. The IGP statistic<sup>41</sup> was applied to measure the similarity of imaging subtypes identified between the discovery and validation sets. IGP ranges from 0 to 100%, where a higher value indicates more similarity between two datasets.

With confirmation of the highly consistent imaging subtypes across discovery and validation cohorts, we refined the final imaging subtypes pulling all patients in the 12 cohorts. To obtain a robust subtype definition, the reproducibility of clustering analysis was evaluated in a repeated five-fold cross-validation scheme,

where patients were randomly separated into training and testing sets at a ratio of 4:1. The training set was used to build the clustering model, which in turn was used to predict patient labels in the hold-out testing set. Cluster purity<sup>42</sup> was used to evaluate the clustering robustness. Cluster purity ranges from 0 to 100%, where a higher value indicates greater robustness of the clustering results. Finally, we identified specific radiological features that were significantly associated with the newly identified imaging subtypes. Specifically, the single-sample GSEA (ssGSEA) algorithm<sup>43</sup> was applied to compute the enrichment scores, and the limma algorithm<sup>44</sup> was implemented to model the differential expression in imaging features.

**Clinical validation for prognostic significance.** For 1,289 patients with detailed follow-up information, we evaluated the prognostic relevance of the imaging subtypes. Within each of the three cancer types, we adjusted for established clinicopathological risk factors in the multivariate analysis. For NSCLC ( $n=701$ ), we adjusted for age, gender, stage and tumour volume. For breast cancer ( $n=226$ ), we included age, estrogen receptor (ER), progesterone receptor (PR) HER2 status and tumour volume. For GBM ( $n=260$ ), we considered age, gender, *MGMT* methylation, *IDH1* mutation status and tumour volume. We also evaluated the prognostic significance of the imaging subtypes in clinically relevant subgroups, such as primary treatment (surgery versus radiotherapy) in NSCLC and established molecular subtypes in breast cancer and GBM.

**Biological validation to identify molecular correlates.** We performed GSEA to identify molecular pathways associated with each of the imaging subtypes. This was done in a subset of patients for whom both imaging and gene expression data were available (NSCLC ( $n=274$ ), breast cancer ( $n=254$ ) and GBM ( $n=124$ )). Specifically, the ssGSEA algorithm<sup>43</sup> was applied to compute the enrichment scores of 50 cancer hallmark pathways curated from MSigDB. Moreover, ssGSEA is a powerful way to mitigate batch effects and overcome discrepancies between different technologies (RNA sequencing versus microarray). The limma algorithm<sup>44</sup> was implemented to model the differential expression pathways.

**Clinical evaluation in lung cancer immunotherapy.** We analysed an independent cohort of 102 patients with advanced NSCLC treated with anti-PD-1 or anti-PD-L1 immune checkpoint blockade, and assessed clinical outcomes in relation to the previously identified imaging subtypes. For imaging feature extraction, we focused on the lesion with the largest size. Because this cohort was not included in the imaging subtype discovery and validation sets, we built an ensemble learning classifier (XGBoost, Supplementary Methods) to predict the imaging subtype label of each new patient given the radiological features bypassing the procedure of dimension reduction and clustering. Because the immune cell composition in the tumour microenvironment may influence the response to immunotherapy, we assessed the enrichment of tumour-infiltrating immune cells for each individual imaging subtype. In particular, we took the gene expression for 274 primary NSCLC tumour samples and then estimated the abundance of 16 immune cell populations by using a previously curated list of genes (Supplementary Table 13).

**Evaluation of subtype reproducibility with deep learning of automated tumour segmentation.** We trained a deep convolutional neural network (U-Net<sup>45</sup>) to fully automate the tumour segmentation process (details are provided in the Supplementary Methods). The main goal was to reduce the inter-observer variations and improve consistency in tumour segmentation by standardizing this process. Furthermore, automated segmentation can facilitate clinical implementation of the tumour subtyping approach by saving time and annotation efforts. The Sørensen–Dice coefficient was used to measure the quality of automated segmentation, with manual contours serving as the ground truth. The automated contours were used as input to repeat the entire computational analysis and assess reproducibility of the previously identified imaging subtypes.

**Statistical analysis.** Kaplan–Meier analysis and log-rank tests were used to evaluate the statistical significance of patient stratification by the imaging subtypes. The Cox proportional hazard regression model was used to adjust for relevant clinicopathologic variables in multivariable analysis. For differential expression analysis, the Benjamini–Hochberg method was used to adjust for multiple statistical testing and control the FDR. The Cramér's V statistic was used to assess the correlation between imaging clusters and cancer types. All statistical tests were two-sided, with a *P* value of  $<0.05$  or FDR of  $<0.05$  considered statistically significant. All statistical analyses were performed in R.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

## Data availability

The data are available within the Article or the Supplementary Information. The imaging data for 9 out of a total of 13 cohorts used in this study are publicly available through the TCIA website (<https://www.cancerimagingarchive.net/>), as described in the Supplementary Information. The imaging data for the breast cancer cohort from Hokkaido University, Japan are publicly available at

<https://drive.google.com/drive/folders/1AsI-bvUWwdmwMd7SHXzJtUsKqmImAGz?usp=sharing>. The imaging data for the Stanford Lung Cancer, Lung Cancer Immunotherapy and Cambridge GBM cohorts are not publicly available because they contain sensitive information that may compromise patient privacy as well as the ethical restrictions or regulation policy of local institutions. These data will be made available to individuals who contact the corresponding authors with a reasonable request, for example, for non-commercial, research purposes. The gene expression data and mutational data of TCGA samples are publicly available in the Genomic Data Commons (<https://gdc.cancer.gov/>). The gene expression data for the other cohorts are available from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), accession nos. GSE22226, GSE103584 and GSE58661.

## Code availability

For the spherical harmonic decomposition, we used the SPHARM-MAT software (<http://www.iu.edu/~spfmat/>). For autoencoder, XGboost and consensus clustering, we used R software (version 3.5.3, R Foundation for Statistical Computing, Vienna, Austria), the package autoencoder (version 1.1), XGboost (version 1.1.1.1) and ConsensusClusterPlus (version 1.52.0). The U-Net architecture is available at <https://github.com/lyakaap/Kaggle-Carvana-3rd-place-solution>. Custom codes<sup>16</sup> are available at <https://github.com/WuLabMDA/PanCancer>.

Received: 28 January 2021; Accepted: 24 June 2021;

Published online: 09 August 2021

## References

1. Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446 (2012).
2. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278**, 563–577 (2015).
3. Itakura, H. et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Sci. Transl. Med.* **7**, 303ra138 (2015).
4. Sun, R. et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* **19**, 1180–1191 (2018).
5. Jiang, Y. et al. Noninvasive imaging evaluation of tumor immune microenvironment to predict outcomes in gastric cancer. *Ann. Oncol.* **31**, 760–768 (2020).
6. Vaidya, P. et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multi-cohort study for outcome prediction. *Lancet Digit. Health* **2**, e116–e128 (2020).
7. Fan, M., Xia, P., Clarke, R., Wang, Y. & Li, L. Radiogenomic signatures reveal multiscale intratumour heterogeneity associated with biological functions and survival in breast cancer. *Nat. Commun.* **11**, 4861 (2020).
8. Wu, J. et al. Magnetic resonance imaging and molecular features associated with tumor-infiltrating lymphocytes in breast cancer. *Breast Cancer Res.* **20**, 101 (2018).
9. Berenguier, R. et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology* **288**, 407–415 (2018).
10. Mackin, D. et al. Measuring computed tomography scanner variability of radiomics features. *Invest. Radiol.* **50**, 757–765 (2015).
11. Traverso, A., Wee, L., Dekker, A. & Gillies, R. Repeatability and reproducibility of radiomic features: a systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* **102**, 1143–1158 (2018).
12. Limkin, E. et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann. Oncol.* **28**, 1191–1206 (2017).
13. Lambin, P. et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
14. Hoadley, K. A. et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304 (2018).
15. Pestana, R. C., Sen, S., Hobbs, B. P. & Hong, D. S. Histology-agnostic drug development—considering issues beyond the tissue. *Nat. Rev. Clin. Oncol.* **17**, 555–568 (2020).
16. O'Connor, J. P. B. et al. Imaging biomarker roadmap for cancer studies. *Nat. Rev. Clin. Oncol.* **14**, 169–186 (2017).
17. Wu, J., Mayer, A. T. & Li, R. *Seminars in Cancer Biology* (Elsevier, 2020).
18. Chalkidou, A., O'Doherty, M. J. & Marsden, P. K. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS ONE* **10**, e0124165 (2015).
19. Zhang, Y. J. *Geometric Modeling and Mesh Generation from Scanned Images* (CRC Press, 2018).
20. Wu, J. et al. Intratumoral spatial heterogeneity by perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Radiology* **288**, 26–35 (2018).
21. Braman, N. M. et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res.* **19**, 57 (2017).
22. Wu, J. et al. Robust intra-tumor partitioning to identify high-risk subregions in lung cancer: a pilot study. *Int. J. Radiat. Oncol. Biol. Phys.* **95**, 1504–1512 (2016).
23. Yankeelov, T. E. et al. Clinically relevant modeling of tumor growth and treatment response. *Sci. Transl. Med.* **5**, 187ps19 (2013).
24. Wu, J. et al. Tumor subregion evolution-based imaging features to assess early response and predict prognosis in oropharyngeal cancer. *J. Nucl. Med.* **61**, 327–336 (2020).
25. Syed, A. K., Whisenant, J. G., Barnes, S. L., Sorace, A. G. & Yankeelov, T. E. Multiparametric analysis of longitudinal quantitative MRI data to identify distinct tumor habitats in preclinical models of breast cancer. *Cancers* **12**, 1682 (2020).
26. Welch, M. L. et al. Vulnerabilities of radiomic signature development: the need for safeguards. *Radiother. Oncol.* **130**, 2–9 (2019).
27. Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593 (2018).
28. Zhang, Y. J., Jing, Y. M., Liang, X. H., Xu, G. L. & Dong, L. in *Computational Modelling of Objects Represented in Images: Fundamentals, Methods and Applications III* (eds Di Giambardino, P. et al.) 215–220 (2012).
29. Shukla-Dave, A. et al. Quantitative imaging biomarkers alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J. Magn. Reson. Imaging* **49**, e101–e121 (2019).
30. Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* **20**, 1349–1360 (2018).
31. Lou, B. et al. An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. *Lancet Digit. Health* **1**, e136–e147 (2019).
32. Coudray, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
33. Jiang, Y. et al. Radiographic assessment of tumor stroma and treatment outcomes using deep learning: a retrospective multicohort study. *Lancet Digit. Health* **3**, e371–e382 (2021).
34. Li, A., Chen, R., Farimani, A. B. & Zhang, Y. J. Reaction diffusion system prediction based on convolutional neural network. *Sci. Rep.* **10**, 3894 (2020).
35. Li, A., Farimani, A. B. & Zhang, Y. J. Deep learning of material transport in complex neurite networks. *Sci. Rep.* **11**, 11280 (2021).
36. Tajdari, M. et al. Image-based modelling for adolescent idiopathic scoliosis: mechanistic machine learning analysis and prediction. *Comput. Methods Appl. Mech. Eng.* **374**, 113590 (2021).
37. Kickingereder, P. et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol.* **20**, 728–740 (2019).
38. Barajas, R. F. et al. Regional variation in histopathologic features of tumor specimens from treatment-naïve glioblastoma correlates with anatomic and physiologic MR Imaging. *Neuro Oncol.* **14**, 942–954 (2012).
39. Nasha, Z. R. et al. Early response evaluation using primary tumor and nodal imaging features to predict progression-free survival of locally advanced non-small cell lung cancer. *Theranostics* **10**, 11707–11718 (2020).
40. Reynolds, A. P., Richards, G., de la Iglesia, B. & Rayward-Smith, V. J. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. *J. Math. Model. Algorithms* **5**, 475–504 (2006).
41. Kapp, A. V. & Tibshirani, R. Are clusters found in one dataset present in another dataset? *Biostatistics* **8**, 9–31 (2007).
42. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830 (2018).
43. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
44. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
45. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (eds Navab, N. et al.) 234–241 (Springer, 2015).
46. WuLabMDA/PanCancer: first release (Zenodo); <https://doi.org/10.5281/zendodo.4906510>

## Acknowledgements

This research was partially supported by the National Institutes of Health (NIH) grants R01 CA233578, R01 CA222512 and R01 CA193730 (R.L.). J.W. acknowledges the NIH K99/R00 CA218667 and University of Texas MD Anderson Cancer Center Lung Moon Shot Program. S.J.P. is funded by a National Institute for Health Research (NIHR), Career Development Fellowship (CDF-18-11-ST2-003) and NIHR Brain Injury MedTech Co-operative based at Cambridge University Hospitals NHS Foundation Trust and University of Cambridge. The views expressed are those of the authors and

not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. C.-B.S. acknowledges support from EPSRC Centre grant no. EP/N014588/1. C.L. acknowledges Cancer Research UK grant no. CRUK/A19732 and EPSRC Centre grant no. EP/N014588/1. We thank TCGA and TCIA for sharing the imaging and genomics data for a subset of patients used in this study.

### Author contributions

J.W. and R.L. contributed to the conception and design of the study. J.W. developed the imaging features and analytics tools. J.W., C.L., M.G., S.P., F.K., H.S., Y.W., C.-B.S., S.J.P., J.W.N., B.W.L., H.W. and M.D. contributed to the acquisition, annotation and curation of the data. J.W., C.L., D.J., B.W.L., M.D. and R.L. contributed to the analysis and interpretation of the data. J.W., C.L. and R.L. wrote the initial manuscript. J.W. and R.L. supervised the work. All authors contributed to the revision of the manuscript.

### Competing interests

J.H. reports fees for advisory committees from AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Catalyst, EMD Serono, Foundation Medicine, Hengrui Therapeutics, Genentech, GSK, Guardant Health, Eli Lilly, Merck, Novartis, Pfizer, Roche, Sanofi, Seattle Genetics, Spectrum and Takeda, research support from AstraZeneca, GlaxoSmithKline, Spectrum, and royalties and licensing fees from

Spectrum. M.D. reports research funding from AstraZeneca, Varian Medical Systems and Illumina, ownership interest in CiberMed and Foresight Diagnostics, patent filings related to cancer biomarkers, and paid consultancy from Roche, Genentech, AstraZeneca, Novartis, Boehringer Ingelheim, Gritstone Oncology, RefleXion and BioNTech.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42256-021-00377-0>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42256-021-00377-0>.

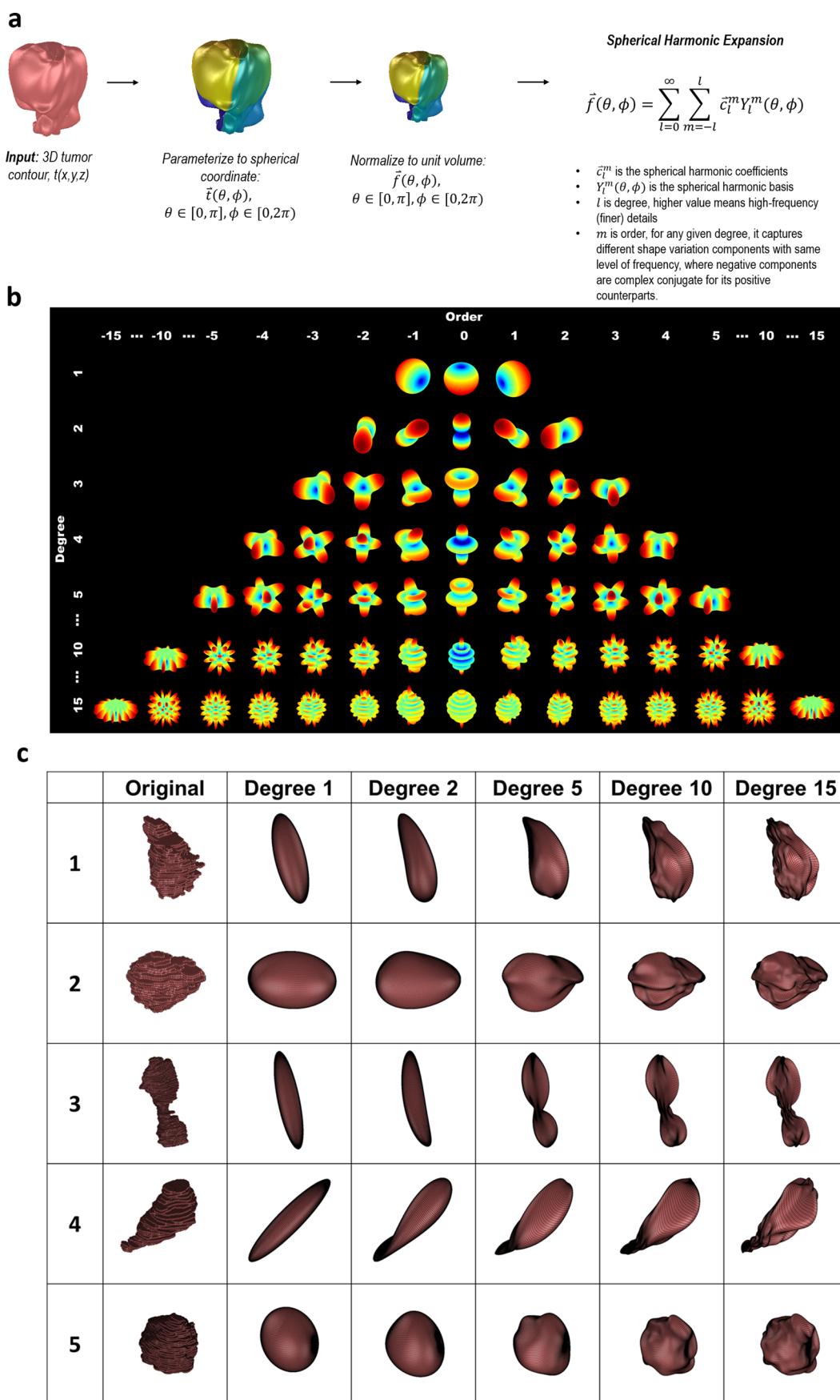
**Correspondence and requests for materials** should be addressed to J.W. or R.L.

**Peer review information** *Nature Machine Intelligence* thanks Anum Kazerouni, Yue Wang and Yongjie Zhang for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

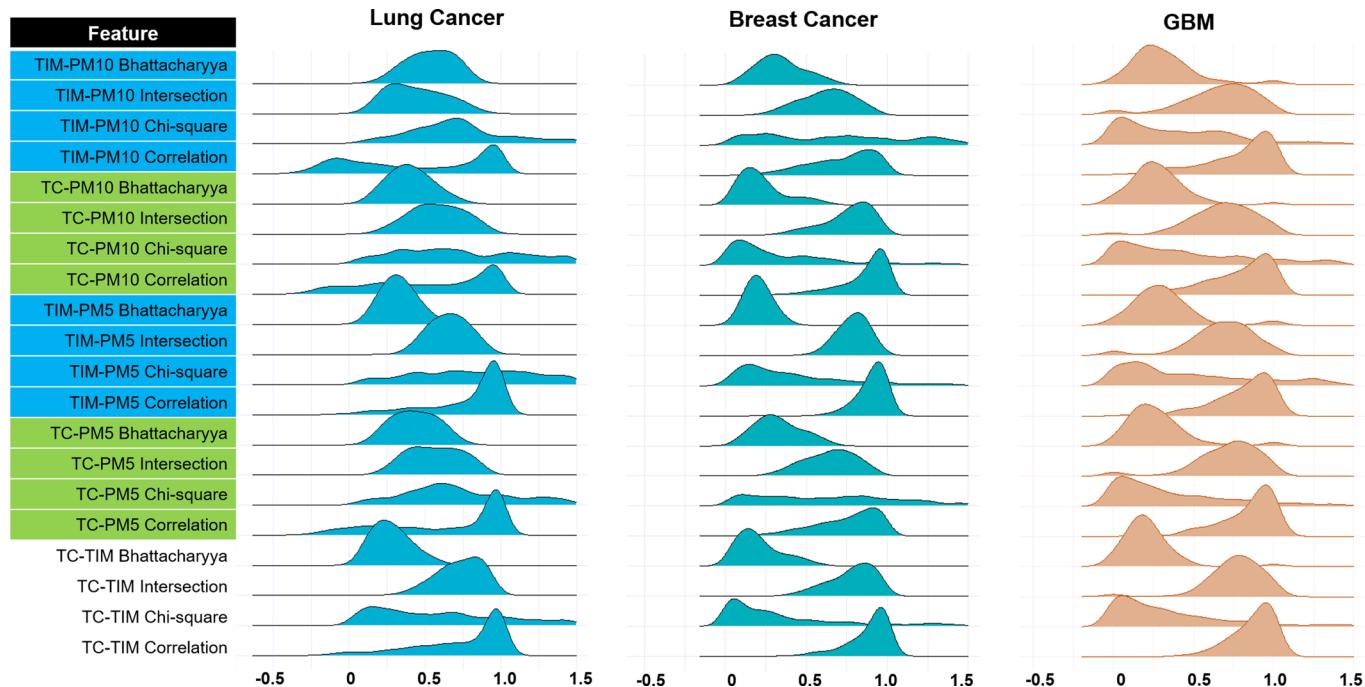
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

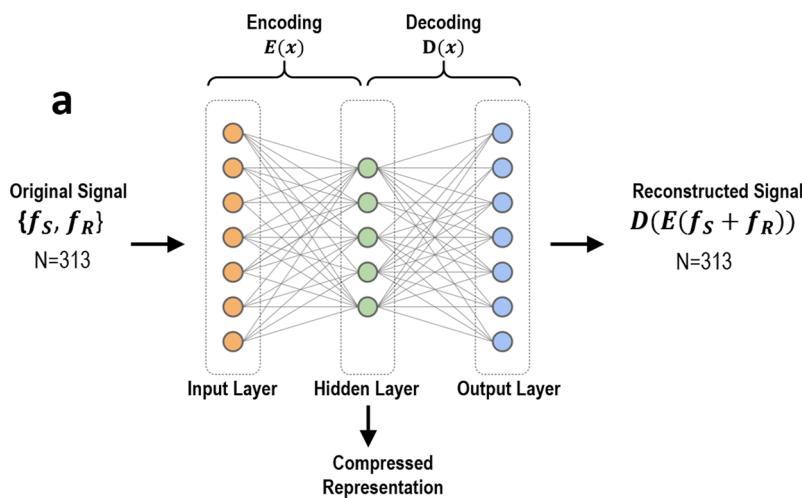


Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Morphological characterization of tumours by spherical harmonic decomposition.** Morphological characterization of tumours by spherical harmonic decomposition. **a)** Overall design of morphological analysis; **b)** Illustration of 3D spherical harmonic basis functions at different degrees and orders; **c)** Illustration of 3D tumours reconstructed by coefficients obtained from spherical harmonic decomposition. Each row represents a selected 3D tumour, which is reconstructed using decomposition results at 5 different degree levels. Here, lower degree captures more global patterns and higher degree corresponds to more detailed morphological patterns.



**Extended Data Fig. 2 | The ridgeline plots present the distribution of 20 regional variation features in three different cancer types.** The ridgeline plots present the distribution of 20 regional variation features in three different cancer types. Here, we investigate 2 tumour regions, tumour core (TC) and tumour invasive margin (TIM), plus 2 peritumour regions, parenchymal margin at 5 mm or 10 mm (PM5 or PM10). In total, 5 pair-wise regions are considered, namely, TC-TIM, TC-PM5, TC-PM10, TIM-PM5, TIM-PM10. Variation for each pair-wise region was quantified with four measures (chi-square, Bhattacharyya distance, correlation, intersection), yielding  $5 \times 4 = 20$  regional variation features. TC-PM5 and TC-PM10 related features are colored in green, while TIM-PM5 and TIM-PM10 related features are colored in blue.



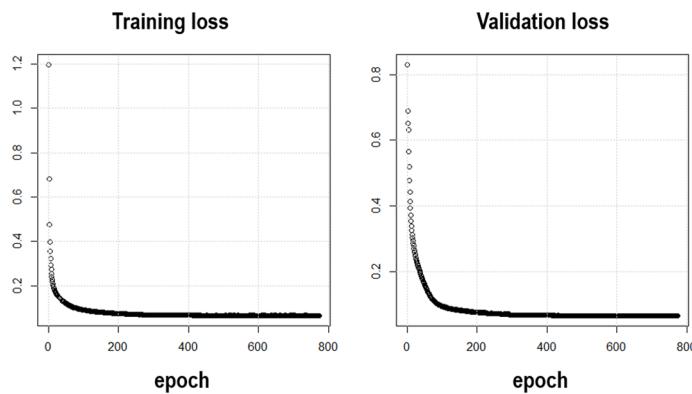
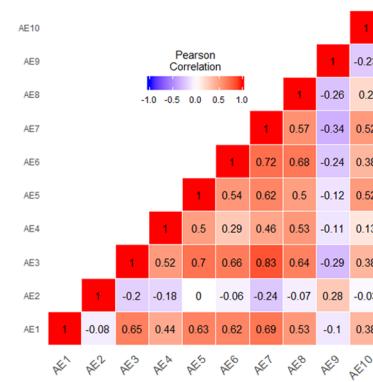
### Hyper-parameter tuning with cross validation

- Input/output layer: {100, **150**, 200}
- Hidden layer: {2, 5, **10**, 15}
- Active function: {tanh, **relu**}
- Learning rate: {0.005, **0.001**, 0.0005}
- Optimize function: {rmsprop, **adam**}
- Loss function: {binary cross entropy, **mean square error**}

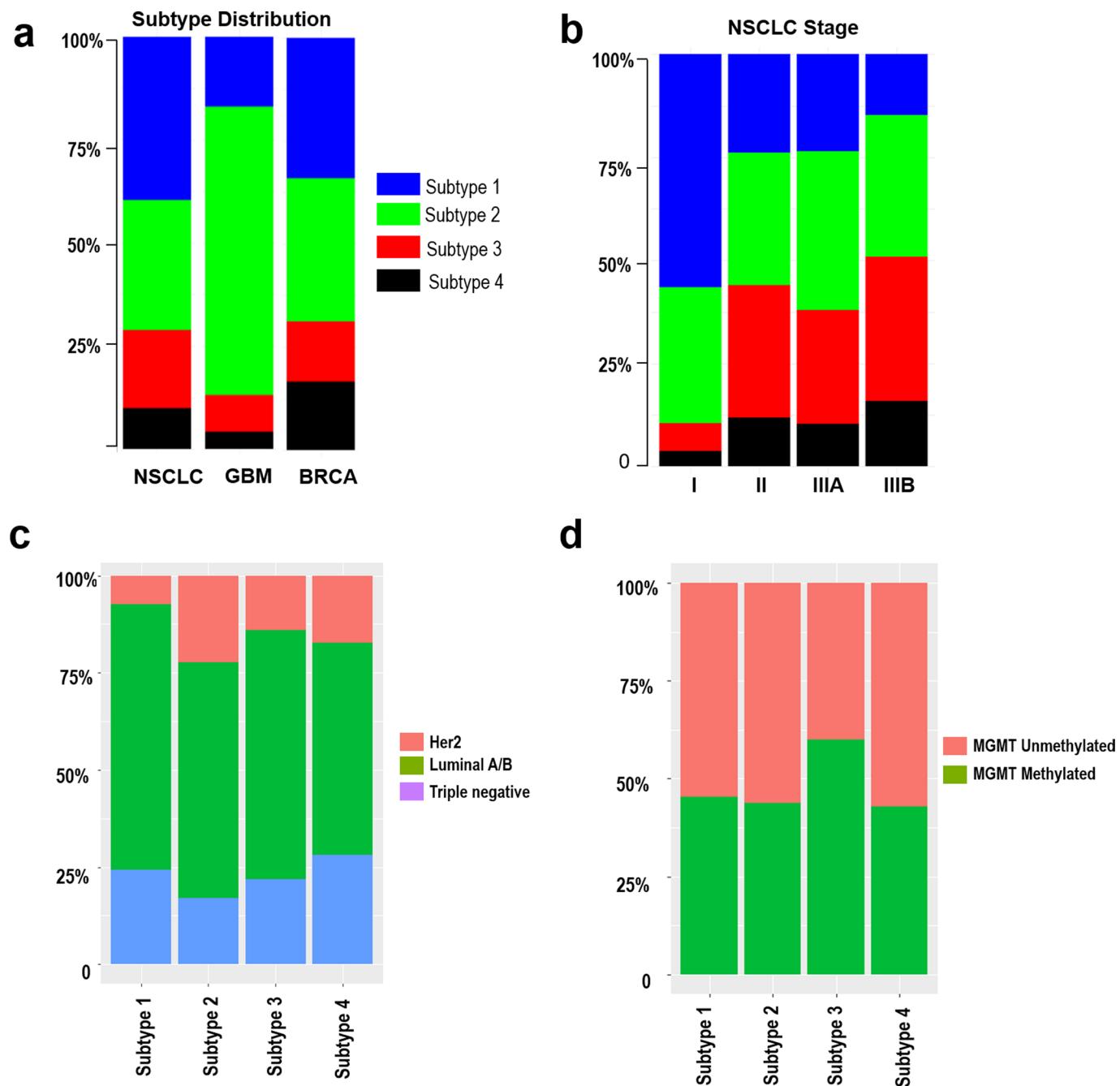
*Red indicates the configuration of final optimized model*

**b**

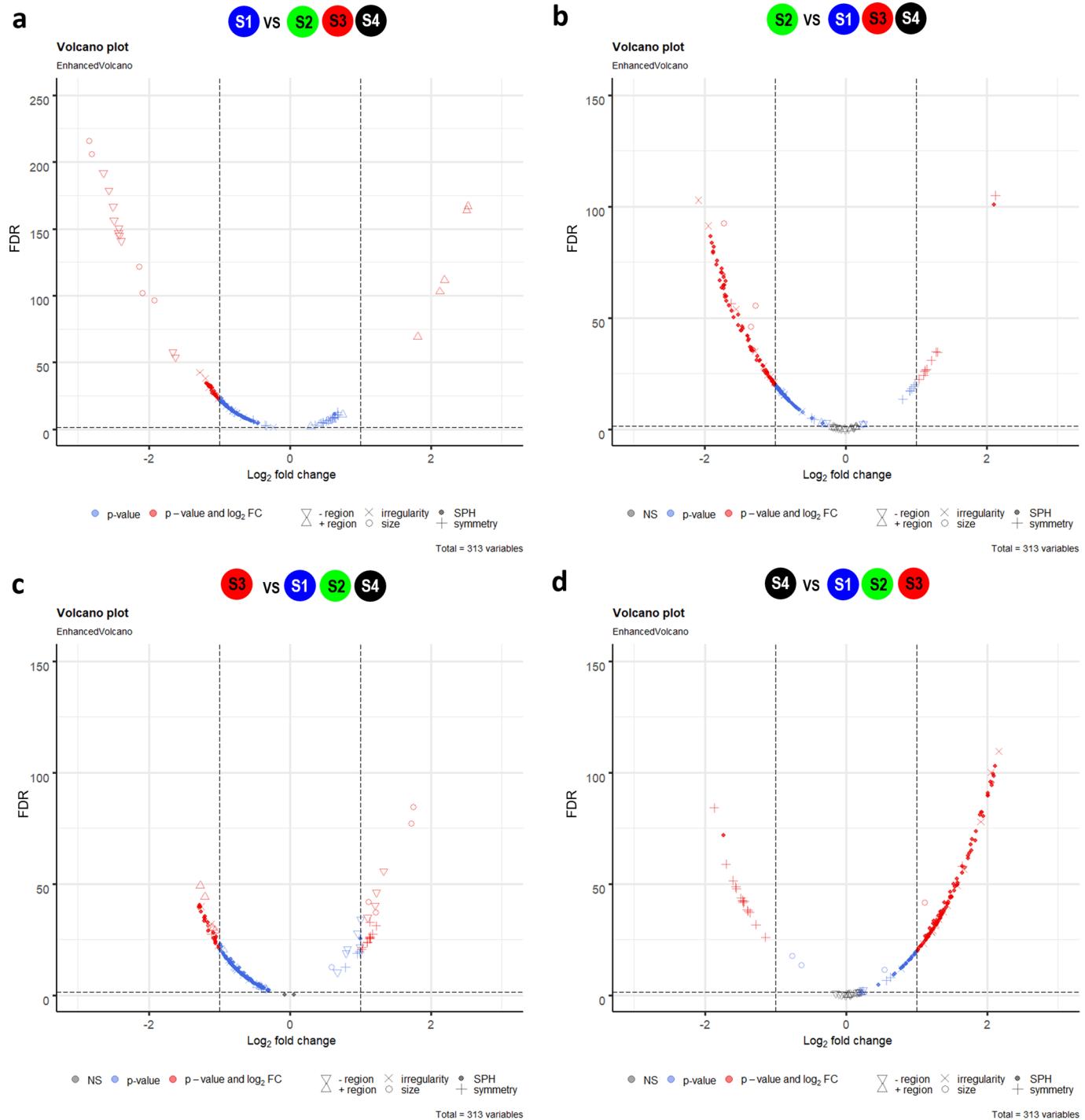
Loss Curve

**c**Correlation of Compressed Representation  $E(f_S + f_R)$ 

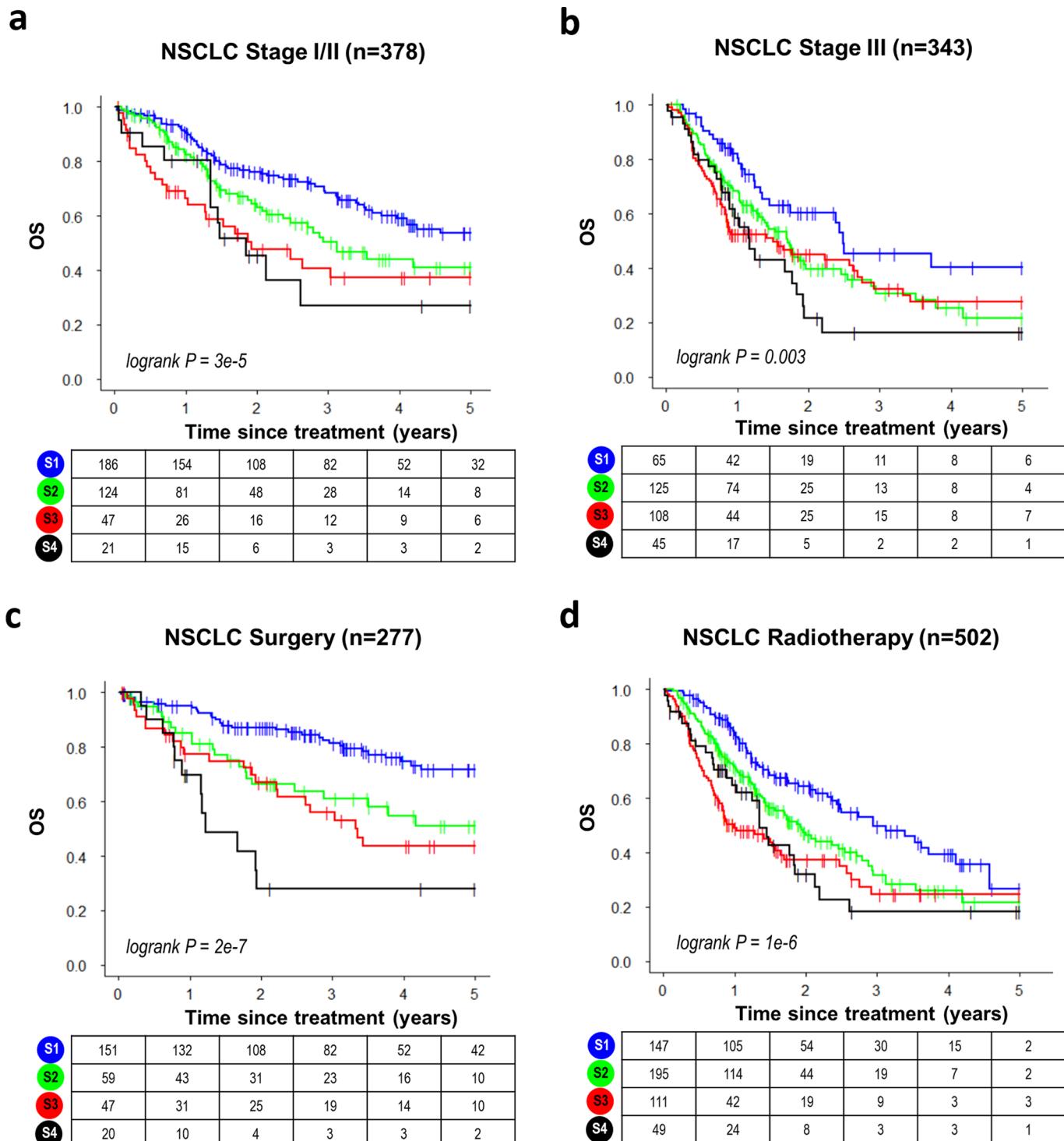
**Extended Data Fig. 3 | Details of imaging feature dimension reduction via an autoencoder model.** Details of imaging feature dimension reduction via an autoencoder model. **a**) The structure of autoencoder used to learn a low-dimensional mapping of the original feature signals with detailed tuning hyperparameters; **b**) The optimal autoencoder loss curves in training and validation; **c**) Heatmap of pairwise correlations between 10 autoencoded features.



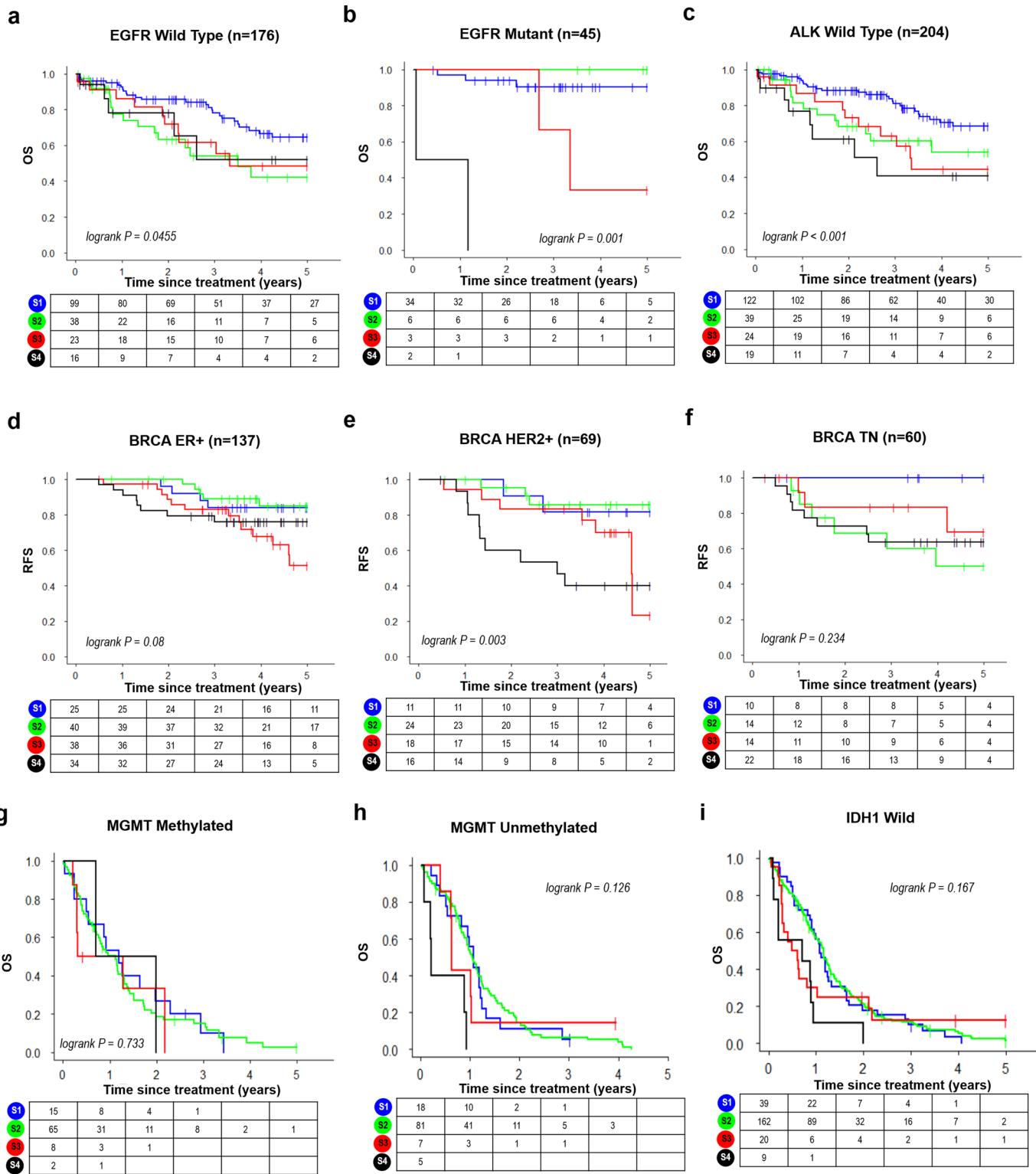
**Extended Data Fig. 4 | Distribution of imaging clusters (subtypes) in different clinical groups.** Distribution of imaging clusters (subtypes) in different clinical groups. **a**) The distribution of all patients in four clusters (subtypes) across three cancer types; **b**) The distribution of lung cancer patients in four clusters (subtypes) across different clinical stage; The molecular subtype distribution in four imaging subtypes for **c**) breast cancer with luminal A/B, Her2+, and triple negative; **d**) GBM with different MGMT methylation status.



**Extended Data Fig. 5 | Volcano plot of enrichment scores through single-sample Gene Set Enrichment Analysis (ssGSEA) of 313 proposed imaging features in all three cancer types.** Volcano plot of enrichment scores through single-sample Gene Set Enrichment Analysis (ssGSEA) of 313 proposed imaging features in all three cancer types. **a**) imaging subtype 1 versus rest, **b**) subtype 2 versus rest, **c**) subtype 3 versus rest, and **d**) subtype 4 versus rest. The data for all enrichment scores are plotted as log<sub>2</sub> fold change versus the  $-\log_{10}$  of the adjusted p-value. Thresholds are shown as dashed lines. Pathways deemed as significantly different (false discovery rate or FDR  $< 0.05$ ) are highlighted with different color schemes.



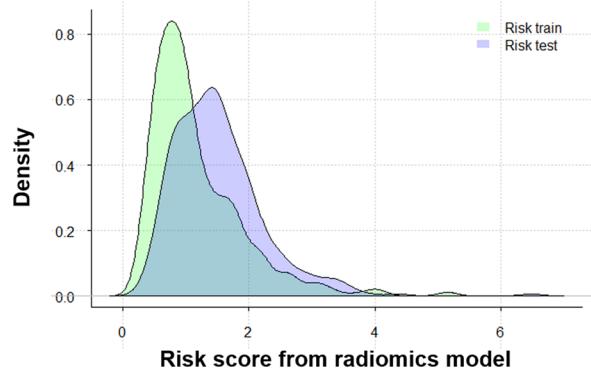
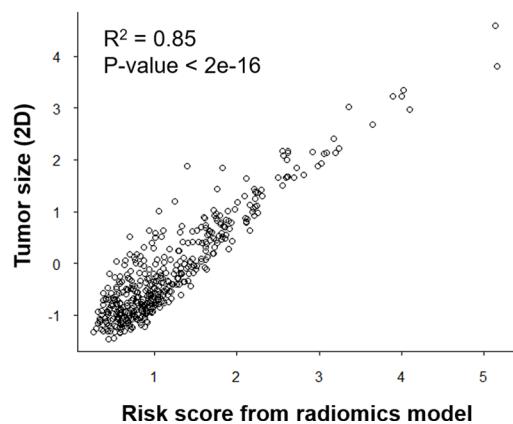
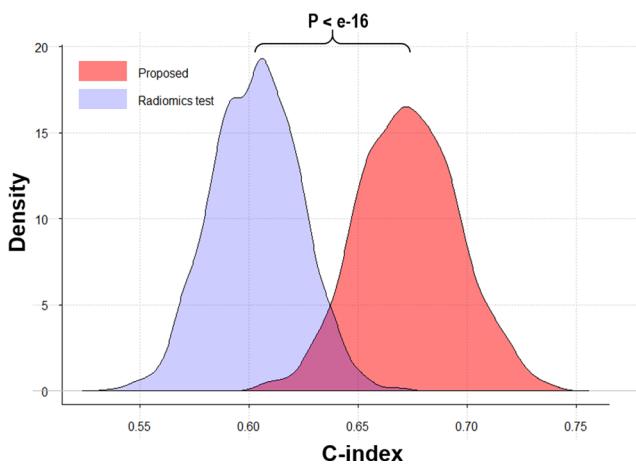
**Extended Data Fig. 6 | Evaluation of prognostic value of the four imaging subtypes in lung cancer subgroups.** Evaluation of prognostic value of the four imaging subtypes in lung cancer subgroups. Kaplan-Meier curves for **a**) stage I + II; **b**) Stage III; **c**) Patients treated with surgery; **d**) Patients treated with radiation.



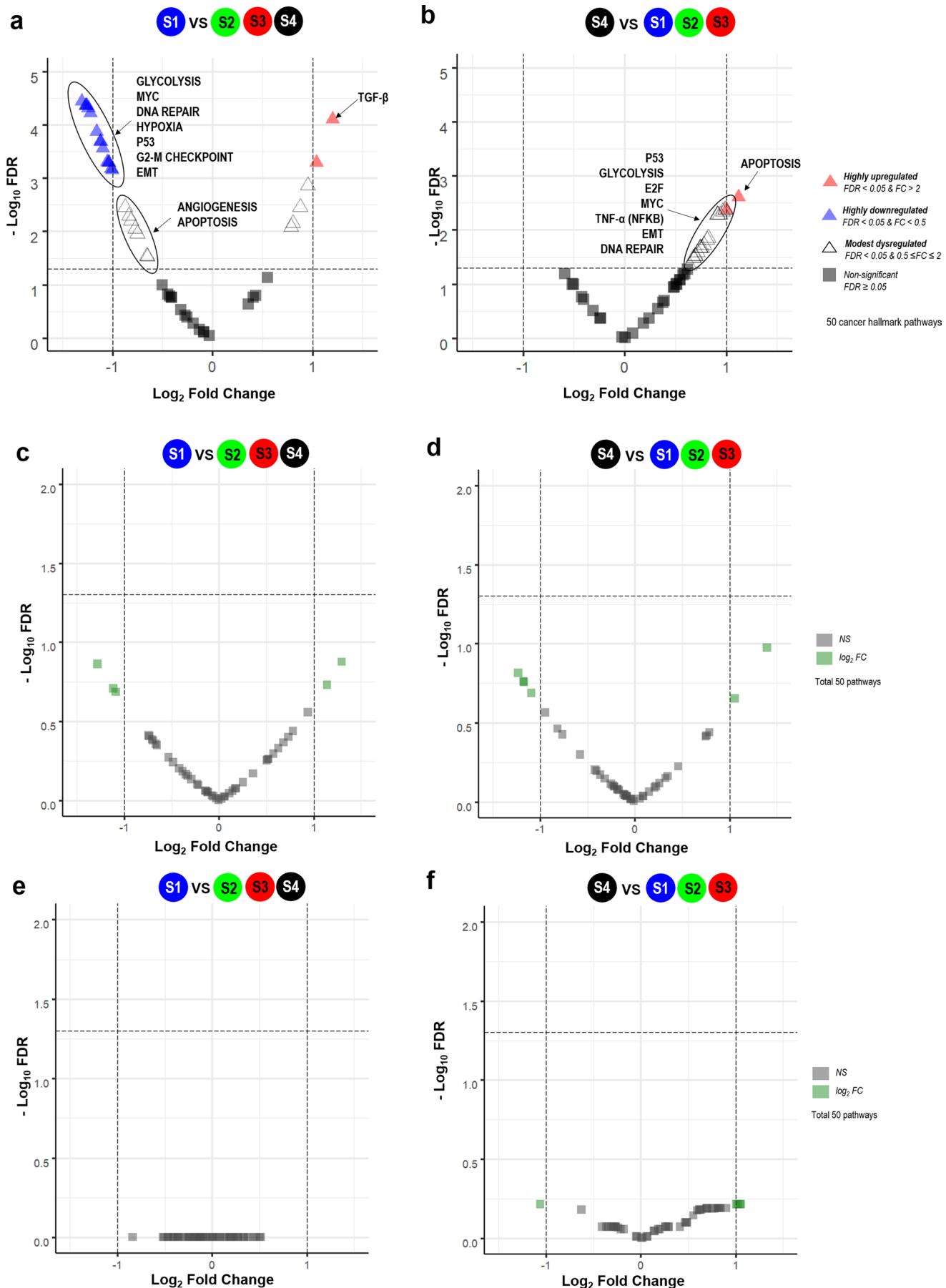
**Extended Data Fig. 7 | Evaluation of prognostic value of the four imaging subtypes in subgroups within three cancer types.** Evaluation of prognostic value of the four imaging subtypes in subgroups within three cancer types. Kaplan-Meier curves for lung cancer subgroups: **a**) EGFR Wild Type; **b**) EGFR Mutant; **c**) ALK Wild Type; for breast cancer subgroups: **d**) ER+ group; **e**) HER2+ group; **f**) Triple Negative (TN) group; for GBM cancer subgroups: **g**) MGMT Methylated group; **h**) MGMT Unmethylated group; **i**) IDH1 Wild group.

**a**

Features	Coef	Exp(coef)	P-value
Shape: MajorAxisLength	0.287	1.332	<b>0.008</b>
Shape: SurfaceVolumeRatio	-0.100	0.905	0.397
Firstorder Mean	0.252	1.286	<b>0.004</b>

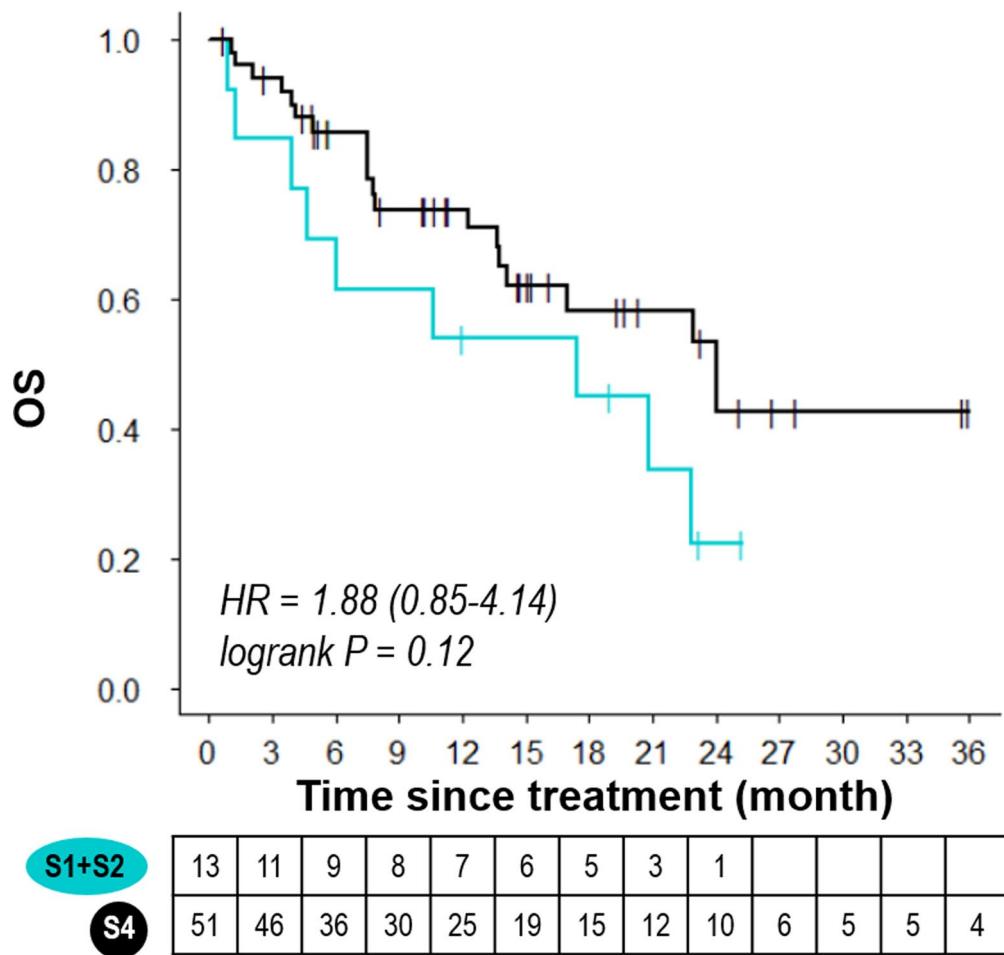
**b****c****d**

**Extended Data Fig. 8 | Comparison between the proposed imaging subtypes and conventional radiomics analysis for survival prediction in lung cancer cohorts.** Comparison between the proposed imaging subtypes and conventional radiomics analysis for survival prediction in lung cancer cohorts. **a**) Details of the final radiomic model; **b**) Distribution of the radiomic risk score in training and validation cohorts; **c**) Scatterplot shows the correlation between radiomic risk score and tumour size measured in 2D; **d**) Distribution and comparison of c-index for the radiomic signature and the proposed imaging subtypes in the validation cohort.



Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | Oncogenic processes associated with the imaging subtypes in three cancer types.** Oncogenic processes associated with the imaging subtypes in three cancer types. Limma-modeled enrichment analysis by single-sample Gene Set Enrichment Analysis (ssGSEA) of 50 cancer hallmark pathways is applied. Volcano plot of enrichment scores in lung cancer: **a**) subtype 1 versus rest, and **b**) subtype 4 versus rest; in breast cancer: **c**) subtype 1 versus rest, and **d**) subtype 4 versus rest; in GBM: **e**) subtype 1 versus rest, and **f**) subtype 4 versus rest. The enrichment scores of 50 cancer hallmark pathways are plotted as log<sub>2</sub> fold change versus the  $-\log_{10}$  of the adjusted p-value. Thresholds are shown as dashed lines. Pathways deemed as significantly different (false discovery rate [FDR] < 0.05) are highlighted with different color schemes.



**Extended Data Fig. 10 | Evaluation of imaging subtypes in the advanced lung cancer treated with immunotherapy.** Evaluation of imaging subtypes in the advanced lung cancer treated with immunotherapy. Kaplan-Meier curves of overall survival stratified by imaging subtype 1 and 2 versus 4.

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

## Software and code

Policy information about [availability of computer code](#)

Data collection

All imaging data used in this study are stored in standard DICOM formats and are collected from The Cancer Imaging Archive or the picture archiving and communication system (PACS) at each institution. Genomics data are publicly available and are collected from the Genomic Data Commons or Gene Expression Omnibus.

Data analysis

For the spherical harmonic decomposition, we used the SPHARM-MAT software (<http://www.iu.edu/~spharm/>). For autoencoder, XGboost, consensus clustering, we used R software (version 3.5.3, R Foundation for Statistical Computing, Vienna, Austria); package autoencoder (version 1.1), XGBoost (version 1.1.1.1), ConsensusClusterPlus (version 1.52.0), respectively. The U-Net architecture is available at <https://github.com/lyakaap/Kaggle-Carvana-3rd-place-solution>. Radiomic features were computed using the open-source software PyRadiomics (version 3.0). Custom codes developed for the study are publicly available at GitHub <https://github.com/WuLabMDA/PanCancer>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The imaging data for 9 out of the total 13 cohorts used in this study are publicly available through TCIA website (<https://www.cancerimagingarchive.net/>). In addition, the imaging data for the breast cancer cohort from Hokkaido University is publicly available. The imaging data for the remaining 3 cohorts are not publicly

available because they contain sensitive information that may comprise patient privacy. These data will be made available to individuals who contact the corresponding authors with a reasonable request, e.g., for non-commercial, research purposes. The gene expression data and mutational data of TCGA samples are publicly available in the Genomic Data Commons (<https://gdc.cancer.gov/>). The gene expression data for the other cohorts are available from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>; [accession number: GSE22226, GSE103584, GSE58661]).

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This is an exploratory analysis of publicly available data in TCGA/TCIA and institutional data for patients of similar characteristics. A formal power calculation was not performed in this study. Instead, sample size was determined empirically by collecting the largest number of samples possible given the inclusion/exclusion criteria described in the Methods.
Data exclusions	The data exclusion criteria are described in detail for each of the 13 cohorts in the Supplementary Information. In general, patients with poor-quality image or no analyzable tumor were excluded.
Replication	We assessed reproducibility of the clustering results through repeated 5-fold cross validation. Cluster purity was used to evaluate the clustering accuracy. Overall, the cluster purity scores were 91-97% (Fig. 3D).
Randomization	This is not relevant for our study as there was no experimental group.
Blinding	The clinical and molecular data were blinded to the investigator during imaging analysis.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

### Population characteristics

This information is presented in detail for each of the 13 patient cohorts in Supplementary Tables 2-5.

### Recruitment

Information about patient recruitment for the TCGA study is available in the Genomic Data Commons (<https://gdc.cancer.gov/>). Inclusion criteria for the remaining institutional cohorts are presented in the Supplementary Information and generally match the characteristics of patients in the publicly available datasets. The validation cohort included 102 metastatic NSCLC patients who were consecutively treated with anti-PD-1 or anti-PD-L1 immunotherapy at a single institution during a 4-year period.

### Ethics oversight

The IRB at each of the participating centers (Stanford University School of Medicine, USA, Hokkaido University Hospital, Japan, and University of Cambridge, UK) approved the study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

[onlineservice@springernature.com](mailto:onlineservice@springernature.com)