



# Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis

Yu Fu<sup>1</sup>, Alexander W. Jung<sup>1</sup>, Ramon Viñas Torne<sup>1,6</sup>, Santiago Gonzalez<sup>1,7</sup>, Harald Vöhringer<sup>1</sup>, Artem Shmatko<sup>1,2</sup>, Lucy R. Yates<sup>3</sup>, Mercedes Jimenez-Linan<sup>4</sup>, Luiza Moore<sup>1,3,4,8</sup> and Moritz Gerstung<sup>1,5,8</sup>

We use deep transfer learning to quantify histopathological patterns across 17,355 hematoxylin and eosin-stained histopathology slide images from 28 cancer types and correlate these with matched genomic, transcriptomic and survival data. This approach accurately classifies cancer types and provides spatially resolved tumor and normal tissue distinction. Automatically learned computational histopathological features correlate with a large range of recurrent genetic aberrations across cancer types. This includes whole-genome duplications, which display universal features across cancer types, individual chromosomal aneuploidies, focal amplifications and deletions, as well as driver gene mutations. There are widespread associations between bulk gene expression levels and histopathology, which reflect tumor composition and enable the localization of transcriptomically defined tumor-infiltrating lymphocytes. Computational histopathology augments prognosis based on histopathological subtyping and grading, and highlights prognostically relevant areas such as necrosis or lymphocytic aggregates. These findings show the remarkable potential of computer vision in characterizing the molecular basis of tumor histopathology.

The diagnosis of cancer is typically based on histopathological assessment of tissue sections, and supplemented by genetic and other molecular tests<sup>1–6</sup>. Modern computer vision algorithms have high diagnostic accuracy and the potential to enhance histopathology workflows<sup>7–9</sup>. Computational histopathology algorithms can process and cross-reference very large volumes of data, helping pathologists to navigate and assess slides more quickly and aid in quantifying aberrant cells and tissues<sup>10</sup>. Often based on convolutional neural networks (CNNs), these algorithms build an implicit quantification of histopathological image content, representing the patterns of the image as seen by the computer. These computational histopathological features are automatically learned for the original task of classifying the entire image and/or subregions of images into cancer or non-cancerous tissues. However, once learned, the feature representation may also be used to find similar images<sup>11</sup> and to quantify associations with traits beyond tissue types<sup>12,13</sup>. This approach, known as transfer learning, has been used to establish associations with genomic alterations<sup>14–19</sup>, transcriptomic changes<sup>20,21</sup> and survival<sup>22–24</sup>.

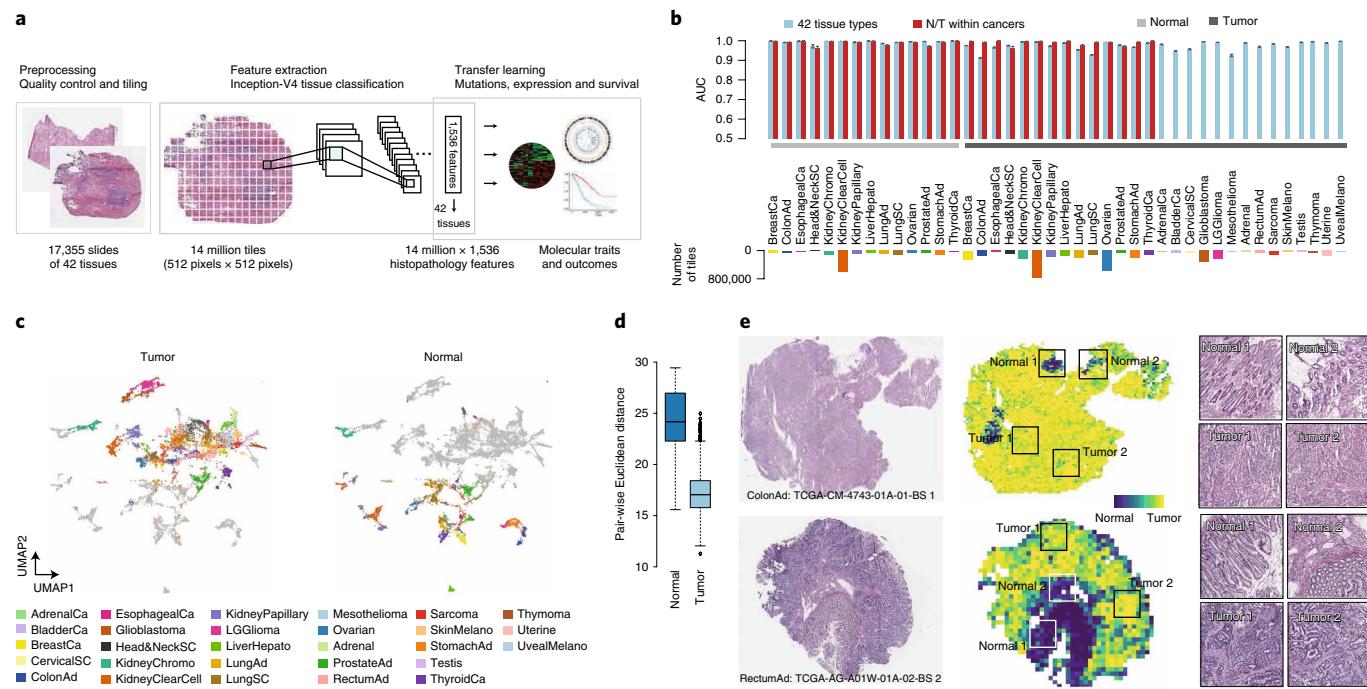
Here we performed a pan-cancer computational histopathology (PC-CHiP) analysis to assess the utility of computer vision and transfer learning across 28 cancer types, to study the associations of histopathology and genomic driver alterations, whole transcriptomes and survival. At its core, PC-CHiP is based on Inception-V4 (ref. <sup>25</sup>), an established CNN, which was used to extract a set of 1,536 image features for tissue classification and transfer learning (Fig. 1a). The algorithm was fine-tuned on 17,355 hematoxylin and eosin-stained fresh-frozen tissue image slides from The Cancer Genome Atlas (TCGA)<sup>26</sup>, containing specimens from 10,452 individuals from 28 tumor types and 14 normal tissues with matched

genomic, transcriptomic and clinical outcome data. Slides were tiled into more than 14 million 256 μm × 256 μm-sized tiles with a digital resolution of 512 pixels × 512 pixels. Tiles from 9,754 slides with a tumor purity of ≥85% were split into 80% training and 20% validation data for training based on 42 broadly defined tissue labels. The resulting quantitative histopathology representation accurately discriminated against different tissues, and displayed pervasive associations with underlying genomic alterations, widespread correlations with transcriptomic signatures and prognostic information across most cancer types.

## Results

**Pan-cancer tissue classification and spatial deconvolution.** For 14 cancers with normal and tumor images, the average tumor/normal tissue classification area under the receiver operating characteristic curve (AUC) was 0.99 (all values are given for the held-back validation set; range: 0.96 to 0.99; Fig. 1b). The average pan-tissue AUC of discriminating all 42 different tissues was 0.98 (range: 0.91 to 0.99), including the 14 cancer types without matched normal samples (Supplementary Table 1; see Discussion for potential limitations). To achieve this classification, PC-CHiP builds an image representation of each tile consisting of the output of the last 1,536 neurons of the network (Fig. 1a). Hereafter we refer to this output as ‘computational histopathological features’ and demonstrate that this representation enables us to derive quantitative associations with a range of molecular traits. As the network was trained to discriminate different tissues, a two-dimensional uniform manifold approximation and projection for dimension reduction (t-SNE) representation (see Methods) of the computational histopathological features showed clusters corresponding to each tissue class

<sup>1</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK. <sup>2</sup>Moscow State University, Moscow, Russia. <sup>3</sup>Cancer, Ageing and Somatic Mutation, Wellcome Sanger Institute, Hinxton, UK. <sup>4</sup>Department of Pathology, Addenbrooke's Hospital, Cambridge, UK. <sup>5</sup>Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. <sup>6</sup>Present address: Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. <sup>7</sup>Present address: Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain. <sup>8</sup>These authors jointly supervised this work: Luiza Moore, Moritz Gerstung. e-mail: [moritz.gerstung@ebi.ac.uk](mailto:moritz.gerstung@ebi.ac.uk)



**Fig. 1 | Pan-cancer computational histopathology quantifies tissue-specific morphology.** **a**, PC-CHiP transfer learning workflow for molecular and clinical outcome predictions based on fresh-frozen hematoxylin and eosin-stained tissue sections. **b**, Top: held-back classification accuracy (AUC) of 42 tissue types, including 14 normal tissues (marked in light gray) and 28 tumor tissues (marked in dark gray). The discrimination of 42 tissues is shown in blue, and organ-specific tumor/normal (T/N) classification is shown in red, with 95% bootstrap confidence intervals (CIs) for each. Bottom: number of tiles in the training dataset for each tissue type. **c**, UMAP dimensionality reduction representation<sup>46</sup> of the 1,536 histopathological features from  $n=200$  randomly selected tiles from each tissue type and JPEG quality. Tumors are highlighted and colored on the left, with normal tissues shown on the right. **d**, Pairwise Euclidean distance in the 1,536-feature space for normal tissues (left) and tumor tissues (right). Boxplots depict quartiles and median values, with whiskers extending to 1.5x the interquartile range, and outliers are denoted by asterisks. **e**, Examples of spatial tumor/normal tissue prediction. For each row, the original hematoxylin and eosin-stained slide is shown on the left, the predicted tumor probability of each tile is shown in the middle, and enlarged images of the four regions indicated by boxes in the middle panels are shown on the right. See Supplementary Table 2 for cancer type abbreviations.

with a certain resemblance of cancers from related organ sites (Fig. 1c and Extended Data Fig. 1). Generally, tumors tended to cluster together, indicating a convergent histological phenotype, usually characterized by a high cell density and loss of tissue architecture—as opposed to normal tissues, which were spread over the periphery in the reduced and original feature spaces (Fig. 1c,d).

PC-CHiP was trained using the consensus pathologist estimate of tumor content as soft labels for all tiles of a given slide. While this assigns the same training values for each tile on a given slide, remarkably, the network is capable of identifying which tiles on a given slide correspond to cancer and normal regions (Fig. 1e). This reflects the fact that, based on the comparison of millions of tiles, the algorithm recognizes that normal tiles from a tumor slide bear greater similarity with normal slides from the same organ site. Automatic deconvolution achieves a notable accuracy, with an average correlation between algorithm and pathologist-estimated tumor purity equal to 0.26 (range: 0.07 (for cervical cancer) to 0.60 (for uveal melanoma); Extended Data Fig. 2). The algorithm's ability to localize signals will be particularly useful when studying the nature of molecular and prognostic associations, as shown in the following sections.

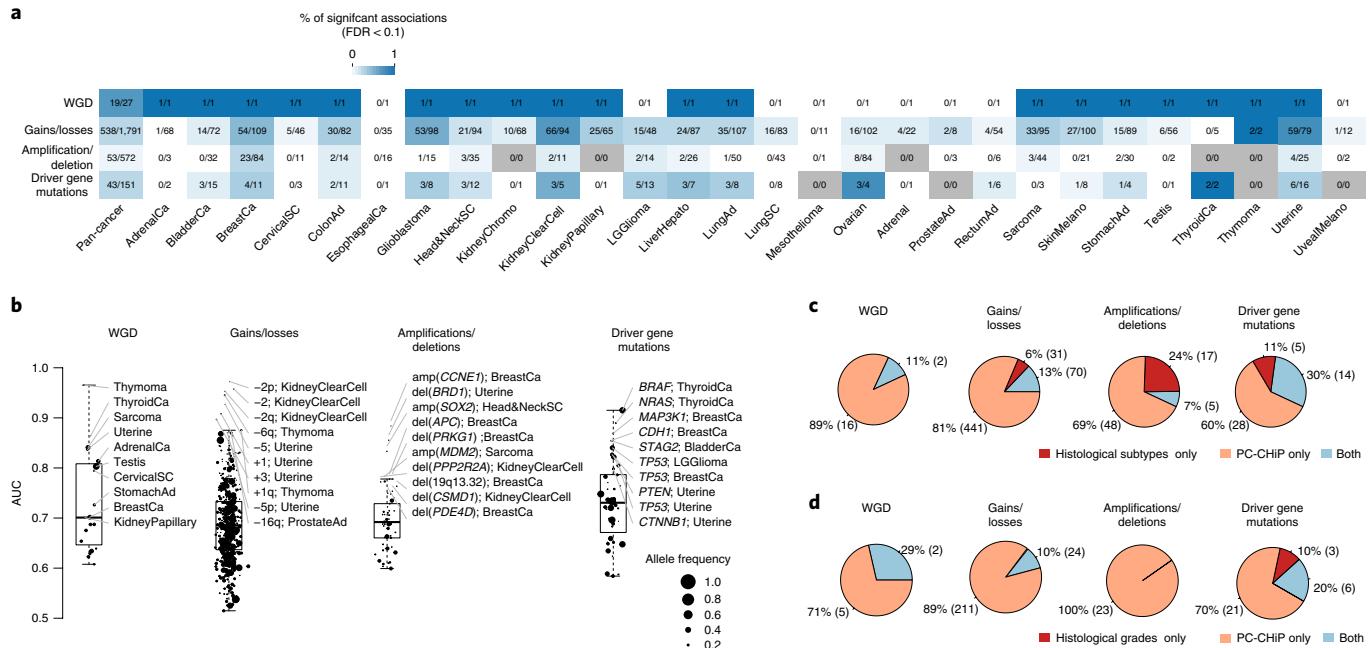
**Transfer learning uncovers associations between histopathology and genomic alterations.** Transfer learning describes the process of using PC-CHiP's 1,536-dimensional histopathological feature representation to discover novel associations with genomic, transcriptomic and prognostic traits. This amounts to using high-dimensional regression approaches, evaluated by fivefold cross-validation with

patient level splits for each fold and separate assessment of each cancer type (Methods).

Associations between recurrent mutations and histopathological patterns could be detected for all types of genomic alterations, including whole-genome duplications (WGDs), chromosome arm gains and losses, focal amplifications and deletions, as well as point mutations in driver genes in almost every cancer type (five-fold cross-validated  $AUC > 0.5$ ; false discovery rate (FDR)  $< 0.1$ ; Methods and Fig. 2a,b). Of note, conventional tumor histopathological subtypes and histopathological grade generally only accounted for a fraction of these associations (Fig. 2c,d), indicating that deep learning quantifies histopathological patterns not entirely reflected in conventional classification systems.

**WGDs alter nuclear morphology and histology.** WGDs occur in about 30% of solid tumors, leading to cells with a nearly tetraploid genome, probably as a result of a single failed mitosis<sup>27</sup>. WGD status could be predicted for 19 out of 27 informative cancer types (five-fold cross-validated  $AUC > 0.5$ ; FDR  $< 0.1$ ; Methods) with an average AUC of 0.73; four cancer types showed an AUC greater than 0.8 (Fig. 2a,b and Supplementary Table 2). Histopathological subtypes and grades provided an inferior predictive accuracy (average  $AUC = 0.59$ ; see also Fig. 2c,d), indicating that additional morphological characteristics underlie these associations.

Tiles with greater probability for WGD displayed an increased nuclear staining, probably due to the higher nuclear DNA content (Fig. 3a; high-resolution images available as Supplementary Data 1). Explicitly quantifying cell nucleus sizes and intensities



**Fig. 2 | Widespread associations between histopathology and genomic alterations.** **a**, Fraction of significant associations (Fisher's combined P value<sup>47</sup> from Wilcoxon's rank-sum test<sup>48,49</sup>, adjusted<sup>50</sup> to FDR < 0.1; the number of significant associations and the total number of tests used for multiple testing adjustments are indicated) between histopathological features and different types of genetic alterations across cancer types. The numbers of samples for each cancer type and alteration, as well as individual P values from each test can be found in Supplementary Table 2a–g. **b**, Boxplot with AUC values for all of the associations, grouped by alteration type, as indicated above. Each box corresponds to the distribution of significant associations for each alteration type. The sample size of each test can be found in Supplementary Table 2. All boxplots demarcate quartiles and median values, while whiskers extend to 1.5x the interquartile range. **c**, Pie charts showing the fraction (number) of genomic alterations by type that can be predicted by PC-ChIP only, by histological subtype only, and by both combined, in 20 cancer types with available histological subtype information. **d**, Pie charts showing the fraction (number) of genomic alterations by type that can be predicted by PC-ChIP only, by histological grades only, and by both combined, in ten cancer types with existing grading information.

using CellProfiler<sup>28</sup> showed that the average cell nucleus size, as well as its variation across the tumor<sup>29</sup>, was elevated in WGD samples, but these features provided a lower predictive accuracy in all but one cancer type compared with PC-ChIP (average AUC = 0.58; range: 0.37 to 0.78; Fig. 3b and Extended Data Fig. 3a,b). Nevertheless, the histopathological features indicative of WGD appeared largely independent of the specific tissue type, as, remarkably, WGD could be predicted for 18 cancer types (FDR < 0.1) by models entirely trained on all but the evaluated cancer type, albeit with slightly reduced accuracy (average AUC = 0.68; Fig. 3c and Extended Data Fig. 3c).

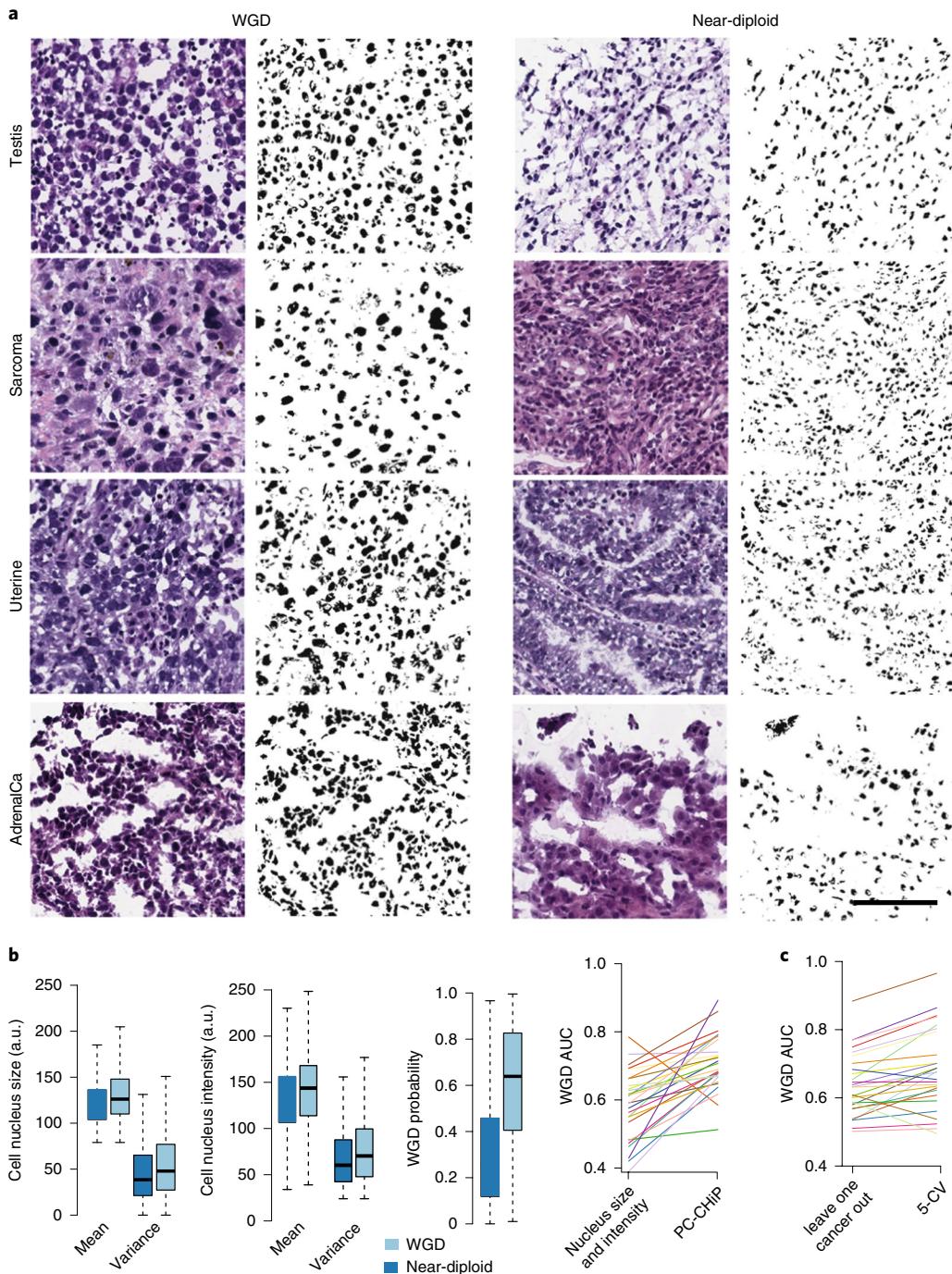
**Histopathological associations with copy number alterations.** The same approach revealed frequent associations between histopathological patterns and gain and losses of whole chromosomes or chromosome arms (Fig. 2a and Supplementary Table 2). Yet, no simple histopathological differences between individual aneuploidies were found, and around 10% of gains and losses were associated with a higher histopathological grade (Fig. 2d). The most frequently associated gain was +8q, as identified in 12 different cancer types. Conversely, loss of chromosome arm 17p, which harbors the *TP53* tumor suppressor gene, was identified in 12 cancers (Extended Data Fig. 4b).

Focal copy number alterations occur on the scale of several megabases and are thought to lead to oncogene amplification and tumor suppressor gene deletions. Based on a catalog of 140 variants (70 amplifications and deletions each)<sup>30</sup>, 53 out of 563 (9%) alteration–cancer pairs with more than ten recurrences showed significant histopathological associations, including ten amplifications and 43 deletions (FDR-adjusted P < 0.1; Fig. 2a,b and Supplementary

Table 2). The cancer type with the largest number of significant focal copy number alterations was breast invasive carcinoma (four amplifications and 19 deletions). Notable examples include deletion of *RB1* (13q14.2; AUC = 0.75; 95% confidence interval (95% CI) = 0.67 to 0.83) and deletion of *PTEN* (10q23.31; AUC = 0.69; 95% CI = 0.63 to 0.76), for which predictions based on histopathological subtypes were inferior (AUC = 0.67; 95% CI = 0.62 to 0.72 for *RB1* and AUC = 0.56; 95% CI = 0.46 to 0.66 for *PTEN*).

Recurrently detected focal deletions involve *CSMD1* (8p23.2; seven cancers) and *PPP2R2A* (8p21.2; seven cancers) (Supplementary Table 2). As these two deletions frequently co-occur (Cohen's  $\kappa$  = 0.88), it is possible that these histological associations reflect the same underlying alteration. Focal amplification of *EGFR* (located at 7p11.2) in glioblastoma (AUC = 0.74; 95% CI = 0.69 to 0.78) was characterized by a distinct small cell morphology of *EGFR*-amplified cancer cells (Fig. 4a). This histopathological association has been noted previously<sup>31</sup>, although *EGFR* amplifications do not exclusively define a molecular glioblastoma subtype<sup>32</sup>.

**Driver gene mutations.** Many oncogenic mutations are point mutations in cancer driver genes. Among all driver genes and cancers tested, 43 out of 151 (28%) gene–cancer pairs displayed significant histopathology associations involving 29 genes (FDR < 0.1; Fig. 2a and Supplementary Table 2). Interestingly, driver mutations in *TP53*—the most frequently mutated gene in cancers—could be detected in 12 out of 27 (44%) cancer types, including low-grade glioma (AUC = 0.84; 95% CI = 0.80 to 0.88), breast invasive carcinoma (AUC = 0.82; 95% CI = 0.78 to 0.87) and uterine cancer (AUC = 0.80; 95% CI = 0.73 to 0.87; Fig. 2b). Tumors with *TP53* mutations were

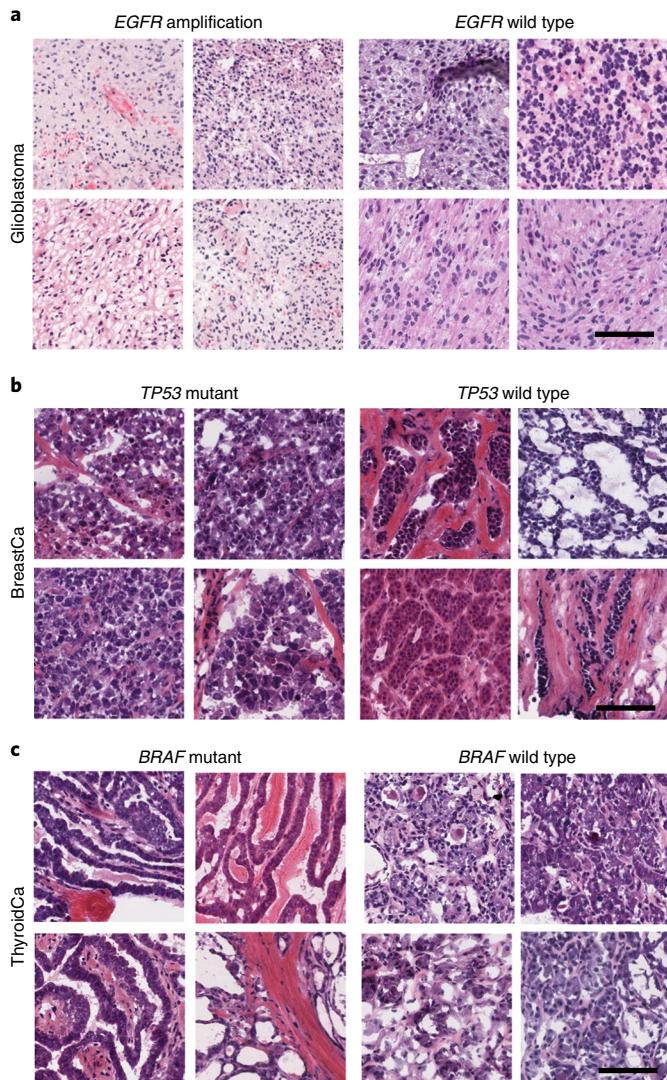


**Fig. 3 | WGDs are characterized by enlarged nuclei.** **a**, Example tiles with (left) and without (right) WGD. From top to bottom, each row corresponds to one cancer type, as indicated on the left. For each row, hematoxylin and eosin-stained tumor sections are shown (colored; left) next to a cell nucleus mark-up (black and white; right). Scale bar: 100  $\mu\text{m}$ . **b**, Left: boxplots showing comparisons of hard-coded morphological features, including nuclear size (left) and cell nucleus intensity (middle), and CNN-derived predictions of WGD probability (right), between WGD and near-diploid samples, for an average of  $n=463$  tiles from each cancer type. Far right: increases of predictive AUC from PC-CHiP compared with hard-coded features. Each line represents one cancer type, using the same colors as in Fig. 1. All boxplots demarcate quartiles and median values, while whiskers extend to 1.5 $\times$  the interquartile range. **c**, Comparison of AUC from leave-one-cancer-out and fivefold cross-validation (5-CV) using 100 tiles for each whole slide from  $n=6,541$  patients with cancer. Each line represents one of 28 cancer types, using the same colors as in Fig. 1.

generally less differentiated and showed higher-grade cell changes (Fig. 4b and Extended Data Fig. 4c).

A highly accurately predicted cancer driver gene was *BRAF* in thyroid tumors, with an AUC as high as 0.92 (CI = 0.87 to 0.96), which was seemingly associated with a papillary morphology (Fig. 4c). While it is known that the prevalence of *BRAF* mutations

is only about 25% in follicular thyroid carcinomas (as opposed to 75% in classical papillary and tall cell thyroid carcinomas<sup>15,33</sup>), these data indicate that the canonical histopathological classification may be further improved (AUC = 0.81; 95% CI = 0.69 to 0.93). A similar association was observed for *PTEN* in uterine cancers, with an AUC of 0.82 (95% CI = 0.76 to 0.89; Extended Data Fig. 4d), in part



**Fig. 4 | Histopathological characteristics of driver mutations.** **a–c,** Example tiles indicating the presence (left) and absence (right) of particular genomic alterations: *EGFR*-amplified glioblastoma (**a**); *TP53* mutant breast cancer (**b**); and *BRAF* mutant thyroid cancer (**c**). For each panel, four representative tiles from the 100 best-predicted tiles are shown. Scale bars: 100  $\mu\text{m}$ .

due to the enrichment of *PTEN* mutations in endometrial cancer<sup>34</sup>. When combined with histopathology subtypes, the AUC for *PTEN* mutations in uterine cancer increased to 0.92 (95% CI = 0.85 to 1). These widespread associations between genomics and histopathology illustrate how alterations either change cellular morphology or occur preferentially in a histopathologically distinct cellular context.

**Transcriptomic associations reflect tumor composition and proliferation.** Associations between gene expression and histopathology may not only reflect distinct tumor cell types with different morphological features, but also stromal and infiltrating immune cells. Overall, we found that 42% of all of the gene–cancer pairs tested showed an association between bulk transcriptome and histology across all cancer types (fivefold cross-validated  $\rho > 0.25$ ; FDR < 0.1; Fig. 5a; Supplementary Table 2). For 6% of gene–cancer pairs, a correlation of  $\rho > 0.5$  was found, and 0.2% displayed  $\rho > 0.75$ . Histopathological grade and subtype contributed to approximately 25% of these associations, but usually only accounted for a fraction of the signal (Fig. 5b,c). No obvious mechanistic insights were

provided by the highest-scoring genes, yet many associations correlated with the extent of normal tissue. Interpretable trends emerged at the level of gene sets, showing that genes were enriched in pathways related to the immune system (24 cancers;  $n = 106$  pathways), followed by cell cycle (11 cancers;  $n = 76$  pathways) and signal transduction (20 cancers;  $n = 59$  pathways; Fig. 5d), in broad agreement with recent reports<sup>21</sup>.

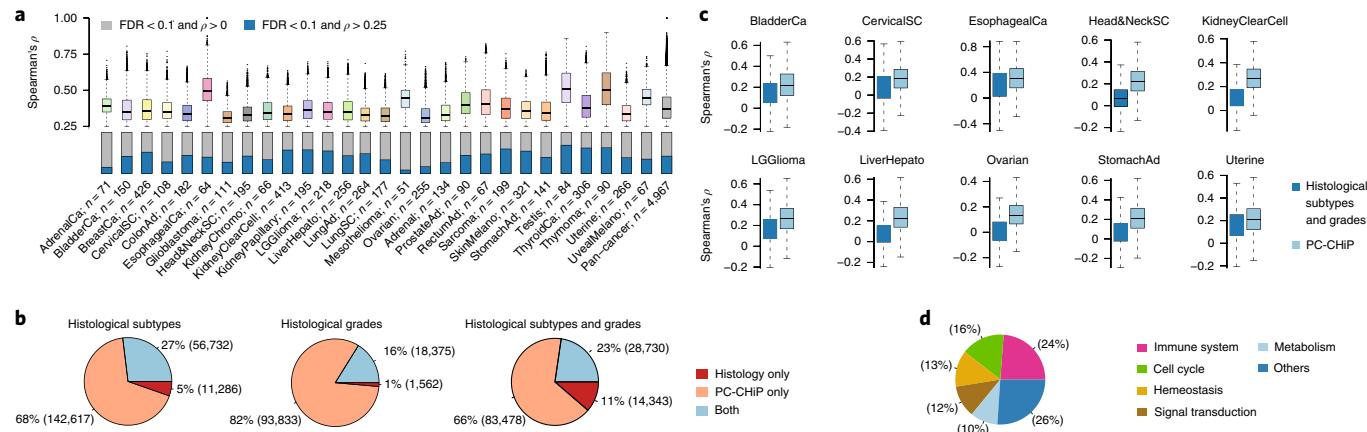
The emerging association between histopathology and transcriptional signals of cell proliferation was further supported by an analysis of transcriptomic proliferation scores<sup>35</sup>, which had histopathologically detectable signals in 25 out of 28 cancer types ( $\rho > 0$ ; FDR < 0.1; Supplementary Table 2 and Extended Data Fig. 5a). While the histopathological patterns associated with high proliferation differed between cancer types, a common feature was elevated grade in tumors with high transcriptomic proliferation (eight out of ten cancers with available data; FDR < 0.1; Extended Data Fig. 5b,c). No systematic trends were observed in relation to histopathological and molecular tumor content (Extended Data Fig. 5d).

**Localization of immune cells.** Tumor-infiltrating lymphocyte (TIL) scores<sup>35</sup> showed a significant  $\rho$  with histology for all 28 cancer types ( $\rho > 0$ ; FDR < 0.1), with  $\rho = 0.73$  (95% CI = 0.6 to 0.82) for thymoma and notable correlations for breast cancer ( $\rho = 0.59$ ; 95% CI = 0.52 to 0.64), bladder cancer ( $\rho = 0.63$ ; 95% CI = 0.51 to 0.72) and lung adenocarcinoma ( $\rho = 0.48$ ; 95% CI = 0.39 to 0.57; Supplementary Table 2). Tiles predictive of TILs indeed contained lymphocytes, which typically were relatively small cells with dark nuclei and scant cytoplasm, often occurring at high densities (Fig. 6a).

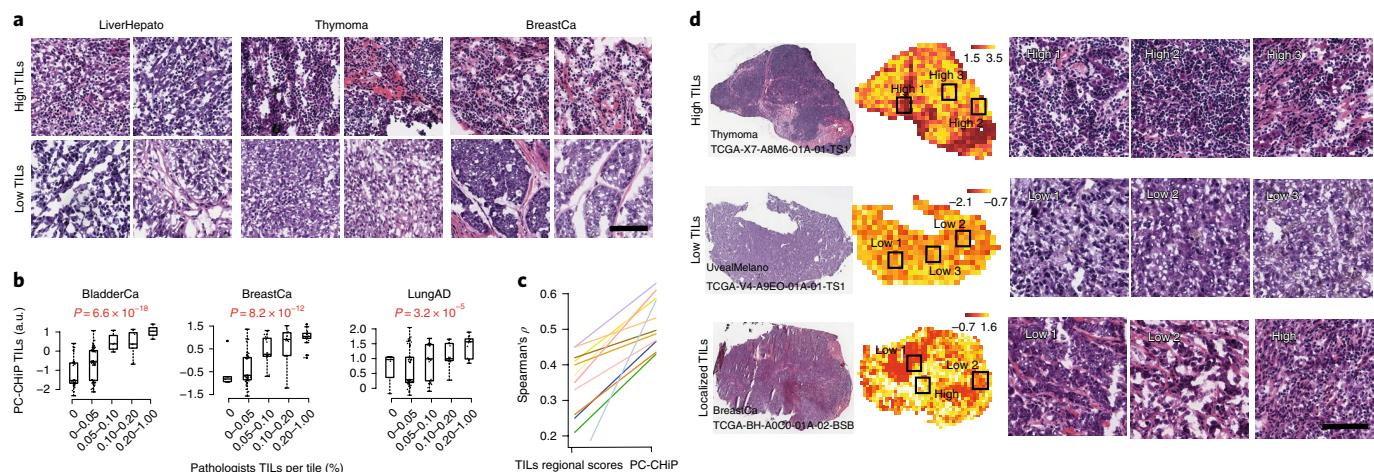
These associations were confirmed by a blinded evaluation of tile-level TIL counts and densities by two expert pathologists for three representative cancer types (Fig. 6b, Extended Data Fig. 6a,b and Methods). Of note, tile-level predictions were independent of the bulk molecular TIL score, confirming the algorithm's ability to localize the TIL signal to specific areas ( $P = 3 \times 10^{-3}$  for bladder;  $P = 6 \times 10^{-7}$  for breast;  $P = 3 \times 10^{-3}$  for lung adenocarcinoma; Wald test, including whole-slide TIL score). Although a transcriptomic training approach is likely to be inferior in its ability to map histopathological patterns of TILs compared with training on histopathologically labeled image areas<sup>12</sup>, it compares favorably in reconstructing overall molecular TIL signals (Fig. 6c).

Patterns of lymphocytic infiltration are recognized prognostic and therapeutic biomarkers<sup>12,36,37</sup>. On many slides, predicted TIL densities appeared relatively uniform, perhaps as a consequence of learning from bulk signals (Fig. 6d, top and middle panels); however, in other cases, the signals stemmed from confined regions containing defined lymphocytic aggregates (Fig. 6d, bottom panel). Given that the algorithm was never provided with explicitly labeled tiles, it is remarkable that these localized patterns of TILs were automatically learned to be a feature shared by a subset of tiles from different slides with a high overall transcriptomic TIL signal.

**Prognostic effects across cancer types.** There was a significant association of computational histopathological features with overall survival in 15 out of 18 cancer types with available data (family-wise error rate (FWER) < 0.05; mean concordance,  $C = 0.6$ ; range: 0.53 to 0.67; Methods, Fig. 7a and Supplementary Table 3). Compared with canonical histological subtypes and grades, which are routinely used to assess prognosis, the computational histopathological features showed a significant improvement in 10 out of 16 cancer types. This prognostic signal remained measurable in the majority of these cancer types, even when further including age, gender and tumor stage (Fig. 7b, Extended Data Fig. 7 and Supplementary Table 3). As illustrated by the survival curves, PC-ChIP may be used to refine existing stage-based prognosis in breast, head and neck, and stomach cancer, as well as (to a lesser extent) clear cell renal cell carcinoma.



**Fig. 5 | Widespread associations between histopathology and gene expression.** **a**, Boxplots of cross-validated Spearman's rank correlation ( $\rho$ ) between whole-tumor messenger RNA expression levels and histopathological features for significant associations (defined as  $\rho > 0.25$  and Fisher's combined  $P$  value<sup>47</sup> from a two-sided test<sup>48</sup>, adjusted<sup>50</sup> to FDR  $< 0.1$ ) across cancer types ( $n = 17,256$  genes tested for each cancer type; the number of samples for each cancer type is stated in the x-axis labels). The fraction of significant genes is shown at the bottom. Individual  $P$  values for each gene are available in Supplementary Table 2h. **b**, Pie charts showing the fraction (number) of genes (subtypes, grades, and subtypes and grades) that can be predicted by PC-ChIP only, by histology only, and by both combined, in cancer types with available histological information (20 for subtypes and ten for grades). **c**, Boxplots showing the distribution of predicted Spearman's  $\rho$  values for genes using PC-ChIP compared with those using histological subtypes and grades combined for each cancer type with available information ( $n = 207,258$  gene–cancer pairs from ten cancer types tested). All boxplots demarcate quartiles and median values, while whiskers extend to 1.5 $\times$  the interquartile range. **d**, Gene Ontology analysis for significant genes ( $n = 203,373$  transcript–cancer pairs tested; only pathways with FDR  $< 0.1$  from a paired t-test are shown).

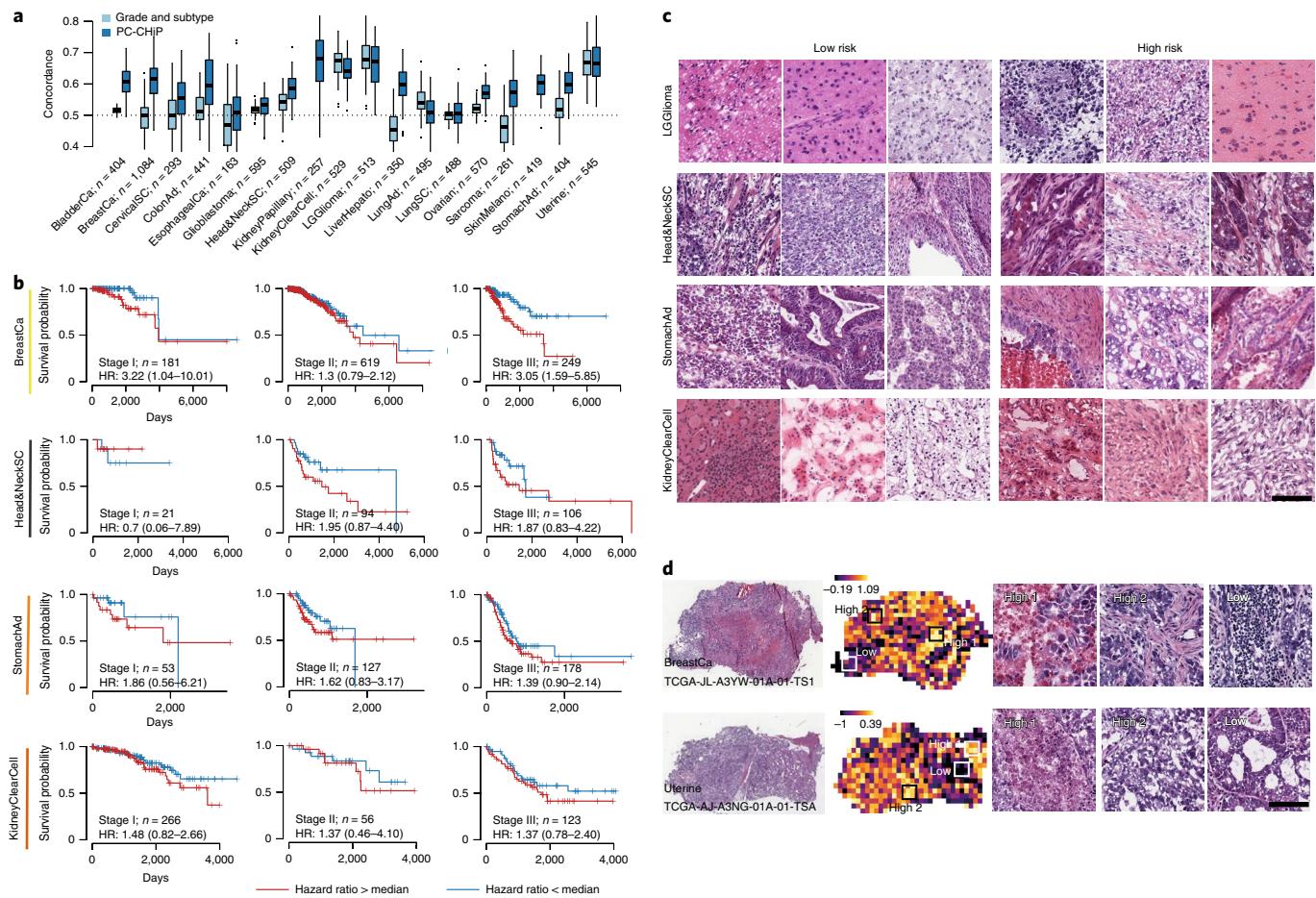


**Fig. 6 | Transcriptomic associations reveal immune infiltration and stromal cell types.** **a**, Example tiles of predicted TIL-rich (top) and TIL-poor regions (bottom) from hepatocellular carcinoma, thymoma and breast invasive carcinoma. For each cancer type, four representative examples selected from the 100 best-predicted tiles are shown. **b**, Systematic blinded assessment of TIL densities by two expert pathologists for three different cancer types ( $n = 150$  for each cancer). Each boxplot shows the predicted TIL scores from PC-ChIP for tiles with different TIL densities, as independently evaluated by pathologists. Boxplots depict quartiles and median values, while whiskers extend to 1.5 $\times$  the interquartile range.  $P$  values for the PC-ChIP score were determined from a binomial generalized linear model with TIL fraction as the response and the additional covariate of whole-slide molecular TIL score. **c**, Spearman's rank correlation ( $\rho$ ) between transcriptomic TIL scores and computational TIL regional scores derived from ref. <sup>12</sup> (left) and PC-ChIP predictions (right) for 13 cancer types. **d**, Example slides with high (top row), low (middle row) and localized TILs (bottom row). From left to right are the original hematoxylin and eosin-stained slides, predicted spatial TIL scores and example tiles with high and low TILs. Example tiles were manually selected from the best-predicted tiles. Scale bars in **a** and **d**: 100  $\mu$ m.

Reassuringly, many of the prognostic histopathological associations automatically learned by computer vision reflect distinct cancer subtypes, as for low-grade gliomas (Fig. 7c). Other features, including necrosis<sup>38</sup> and high tumor grade, are associated with poor prognosis across tumor types. In contrast, a higher degree of differentiation<sup>39</sup> and the presence of TILs are usually associated with a favorable risk<sup>40</sup>. Favorable and unfavorable patterns can frequently be identified on the same slide. This highlights the ability of computer vision

to deconvolve the content of large tissue sections into molecularly and prognostically distinct areas (Fig. 7d), with necrosis and lymphocytic aggregates detected on the same specimen. Similarly, areas of low- and high-grade tumor differentiation identified on the same slide produced favorable and unfavorable risk predictions.

**Validation on external cohorts.** The PC-ChIP algorithm exhibited good generalization on two breast cancer validation cohorts,



**Fig. 7 | PC-CHiP provides complementary prognostic information.** **a**, Concordance index<sup>52</sup> of overall survival predictions using histopathological grades and subtypes (light blue) and PC-CHiP (dark blue) in 18 cancers. Shown is the distribution across 20 repetitions of fivefold cross-validation ( $n=100$  for each boxplot). All boxplots demarcate quartiles and median values, while whiskers extend to 1.5x the interquartile range. The number of patients for each cancer type is provided in the x axis labels. **b**, Kaplan-Meier plots with high (above median) and low (below median) PC-CHiP risk, shown for cancer stages I–III in four different cancer types. **c**, Example tiles with high and low estimated risk based on the histopathological features. Each row corresponds to a cancer type. The first three columns are tiles with low risk and the last three are tiles with high risk. **d**, Risk predictions across whole slide examples from breast invasive carcinoma and uterine endometrial carcinoma, with examples of high and low risk areas enlarged on the right. Scale bars in **c** and **d**: 100  $\mu$ m. HR, hazard ratio.

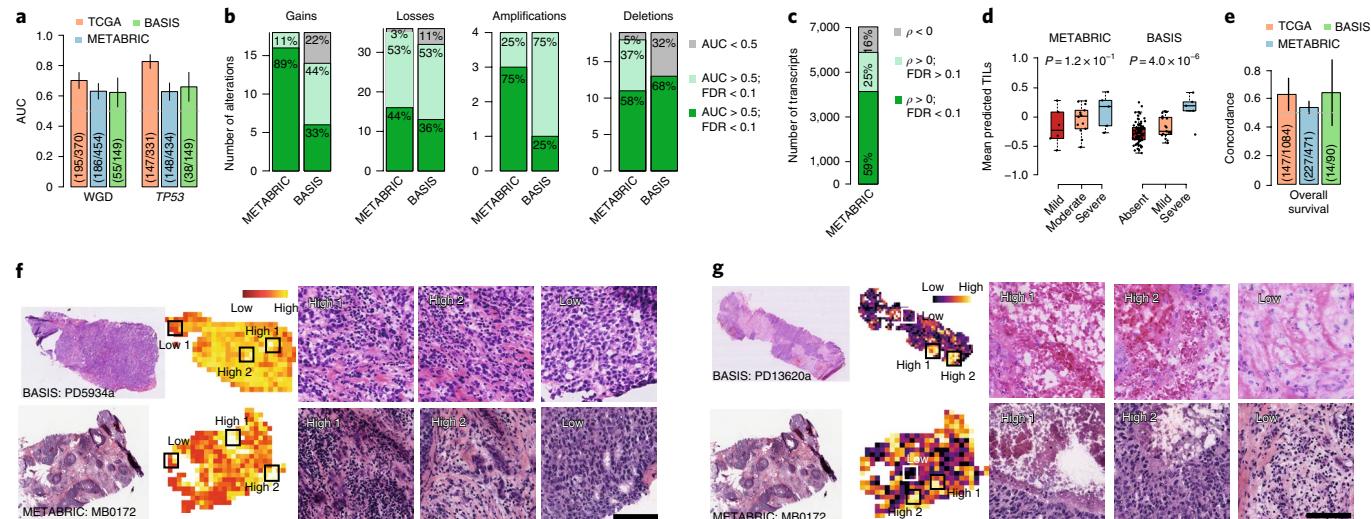
comprising fresh-frozen hematoxylin and eosin-stained slides from the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC;  $n=471$ )<sup>41</sup> and the Breast Cancer Somatic Genetics Study (BASIS) consortium ( $n=151$ )<sup>42</sup>. The vast majority of genomic associations could be replicated in both cohorts, including associations with TP53 mutations and WGD, with only a moderate drop in accuracy (Fig. 8a,b, Extended Data Fig. 8a and Supplementary Table 4). Similarly, at least 59% (FDR < 0.1) of transcriptomic predictions were recovered in the METABRIC cohort with comparable correlation levels (Fig. 8c and Extended Data Fig. 8b). A subset of both cohorts contained pathologist-evaluated TIL categories, which confirmed the predicted trends, and the algorithm's ability to localize TILs on a given slide was evident in both cohorts (Fig. 8d,f). Prognostic associations were replicated in both cohorts, although at reduced accuracy, particularly in METABRIC (Fig. 8e). As in the TCGA training data, the algorithm identified necrotic areas and TILs on slides from both cohorts as unfavorable and favorable prognostic markers (Fig. 8g).

Of note, there was a considerable tissue misclassification of tiles from the validation cohorts, possibly due to different file formats (Supplementary Table 4). Indeed, JPEG quality had a strong confounding effect on histopathological feature representation and

tissue classification in the TCGA cohort (Extended Data Fig. 9a,b). Yet, these biases were mostly confined to the initial classification task and were largely mitigated by PC-CHiP's tissue-aware transfer learning, as confirmed by the high validation rate of molecular associations. Biases of the histopathological feature representation could be further reduced using a file format-aware Inception-V4 architecture and additional data augmentation, which led to a slight drop in the tissue classification accuracy (average AUC = 0.95; Methods, Supplementary Table 4 and Extended Data Fig. 9c,d). Transfer learning using histopathological features from the modified architecture produced similar molecular and prognostic associations in TCGA and moderately improved the strength of associations in the METABRIC cohort, but not evidently in the BASIS cohort (Extended Data Fig. 10a–d).

## Discussion

The results presented here provide a molecular basis for the histopathological observation that tumors are a diverse cellular ecosystem, and offer new ways to histologically deconvolute and map its molecularly defined content. Our findings demonstrate that links between a tumor's morphology and its molecular composition can be found in every cancer type and for virtually every class



**Fig. 8 | External validation.** **a**, Predictive accuracy of WGD and *TP53* mutations for the cross-validated predictions in TCGA discovery and METABRIC and BASIS breast cancer validation cohorts. Each bar corresponds to the predicted AUC and the error bars correspond to 95% CIs. The numbers stated in each bar refer to mutated/total cases. **b**, Distribution of validated (dark green), indeterminate (light green) and invalid (gray) associations in the METABRIC ( $n=454$  tumor samples) and BASIS cohorts ( $n=149$  tumor samples) across different alteration types. P values for each alteration can be found in Supplementary Table 4b. **c**, Distribution of validated (dark green), indeterminate (light green) and invalid (gray) transcriptomic associations in the METABRIC cohort ( $n=14,756$  genes for  $n=456$  tumor samples), based on a two-sided rank correlation test. **d**, Slide average-predicted TIL scores (y axis) as a factor of pathology assessment. P values were estimated from one-sided analysis of variance. All boxplots demarcate quartiles and median values, while whiskers extend to 1.5× the interquartile range ( $n=36$  slides for METABRIC;  $n=129$  slides for BASIS with available TIL annotation). **e**, Prognostic accuracy for overall survival measured by the concordance index, C. The numbers given in each bar refer to deaths/total number of patients. All error bars denote 95% CIs. **f**, Example slides from the BASIS and METABRIC cohorts with high and low levels of TILs. From left to right are the original hematoxylin and eosin-stained slides, predicted spatial TIL scores and example tiles with high and low TILs. **g**, Example slides from the BASIS and METABRIC cohorts with high and low levels of prognostic risk. From left to right are the original hematoxylin and eosin-stained slides, predicted spatial risk and example tiles with high and low risk. Representative slides and tiles were selected from the 100 best-predicted ones. Scale bars in **f** and **g**: 100 μm.

of genomic and transcriptomic alteration. WGDs were characterized by (and likely caused) nuclear enlargement and increased nuclear intensities, reflecting abnormal chromatin content. Other alterations, such as *EGFR* amplifications or *BRAF* mutations, were associated with a distinct histology and it is currently unknown whether this is a consequence of the alteration or an indication that these mutations preferentially occur in particular cell types. Furthermore, a broad range of transcriptomic correlations were found reflecting stromal content, immune cell infiltration and cell proliferation.

In the majority of cancer types, computational histopathological features showed a good level of prognostic relevance, substantially improving prognostic accuracy compared with conventional grading and histopathological subtyping alone. While it is very remarkable that such predictive signals can be learned in a fully automated fashion, there was no measurable improvement over a full molecular and clinical workup. This might be a consequence of the far-ranging relationships between histopathological and molecular phenotypes described here, implying that histopathology is a reflection of the underlying molecular alterations rather than an independent trait. Yet, it probably also highlights the challenges of combining histopathological signals from different areas of the same tumor, which requires very large training datasets for each tumor entity.

One of the main current limitations of the study is that the training was performed on fresh-frozen tissue sections. These provide a better preservation of molecular content compared with formalin-fixed paraffin embedding, which is the diagnostic standard due to a better preservation of tissue morphology. Also, without further algorithmic amendments, there was a considerable dependence of the histopathological feature representation on image compression algorithms and their parameters, which could in part be mitigated by transfer learning. Given the sensitivity of

deep learning algorithms and the associated risk of overfitting, it appears mandatory to critically assess whether discoveries could be explained by technical artefacts, and also whether algorithms are applicable to new datasets with different image properties.

While the pervasiveness of associations between histopathology and molecular traits is remarkable, at present, they are too weak to replace genetic or transcriptomic tests. However, we expect at least some associations to become more accurate using improved algorithms and larger, ideally spatially annotated training cohorts. Yet histopathological annotation is possible only for patterns known beforehand, and molecular spatial annotation currently has low throughput, but may become possible with spatial transcriptomic<sup>43,44</sup> and sequencing technologies<sup>45</sup>. Direct training on the molecular trait of interest, or ideally multi-objective learning, may provide superior results compared with transfer learning, which risks missing patterns that are irrelevant for the initial classification task. Alternatively, convolutional rather than linear transfer learning has been shown to yield promising results for transcriptomic associations<sup>21</sup>. Also, less complex CNN architectures may suffice for classifying histopathology patterns<sup>9,18</sup>, because tumor sections have a defined scale, unlike everyday images.

Looking forward, our analyses reveal the potential for using computer vision alongside molecular profiling. While the eye of a trained human pathologist constitutes the gold standard for recognizing clinically relevant histopathological patterns and definitive diagnosis, computers have the capacity to augment these tasks by sifting through millions of images to retrieve similar patterns and establish associations with known and novel traits. Taking the overlays presented in this study as examples, it is possible to imagine computationally augmented and molecularly informed histopathology workflows enabling more precise and faster diagnosis and prognosis in the future.

## Methods

**Images.** We collected 17,355 hematoxylin and eosin-stained histopathology slides of 10,452 patients with 28 broadly defined cancer types from TCGA via the Genomic Data Commons Data Portal<sup>53</sup> (<https://portal.gdc.cancer.gov/>), including normal, tumor and metastatic tissue types. Sample inclusion criteria defined by TCGA required primary untreated samples, frozen and sufficiently sized resection samples and at least 60% tumor nuclei (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga/studied-cancers>). Scanned slides usually depict the top and bottom section of the tissue block used for molecular analysis. Only tissue types with at least 50 images with a magnification greater than 20 $\times$  were included in this study. We first cropped the whole slides into tiles of 512 pixels  $\times$  512 pixels, with an overlap of 50 pixels, at 20 $\times$  magnification. We then removed blurred and non-informative tiles by filtering on the weighted gradient magnitude (using Sobel operator, tiles with a weighted gradient magnitude smaller than 15 for more than half of the pixels were removed). Tiles from tumor samples with a tumor purity  $\geq 85\%$  were used in training/validation to avoid mislabeled tiles in the training process. To avoid bias caused by image preparation in individual laboratories, we randomly selected 80% of images from each center for training. In total, we used 6,502,594 tiles from 8,047 slides for training, 1,338,534 tiles from 1,682 slides for validation, and 6,547,172 tiles from 7,628 slides for testing.

**PC-CHiP.** A pretrained Inception-V4 (ref. <sup>25</sup>)—a deep CNN—was used to classify tiles into 42 classes and to extract histopathological features from each tile. We applied sample-specific label smoothing—an adapted version of the label smoothing method first introduced in Inception-V3 (ref. <sup>54</sup>)—for model regularization, to avoid overfitting. In short, for a sample of tissue  $i$ , we set the ground-truth distribution  $q(k)$  to  $q(k) = p_i$  for  $k = i$  and  $q(k) = (1 - p_i)/(n - 1)$  for all  $k \neq i$ , where  $n = 42$  is the total number of classes, and  $p_i$  is the tumor purity of the sample. The model was trained in TensorFlow using Slim (<https://github.com/google-research/tf-slim>) with the default hyperparameters for 100,000 steps (~1 epoch). The scripts used for training are available from GitHub (<https://github.com/gerstung-lab/PC-CHiP>), and the retrained model checkpoints are available from the BioStudies database (<https://www.ebi.ac.uk/biostudies/>) under accession number S-BST292.

We retrieved the probability for all 42 classes for each tile, and the associated 1,536 histopathological features from the last hidden layer of the trained Inception-V4. As in practice the cancer site is usually known, we also computed tumor versus normal classification within cancer types for each tile by comparing only the probability of the sample being normal or tumorous for that cancer type.

To visualize the tiles represented by the 1,536 histopathological features, we applied UMAP<sup>46</sup> for dimension reduction for a subset of tiles (200 correctly classified tiles were randomly selected from high-tumor-purity images from each tissue type and image quality). As the distances between data points in the original dimension are not preserved in the low dimension generated by UMAP, we also calculated the mean pairwise Euclidean distance between tissue types in their original dimension.

**Algorithmic modifications.** To reduce the effect of confounding image quality, we modified the classification layers (both the last and auxiliary) for the probability to read  $P = \text{logit}^{-1}[X\beta + Z\alpha]$ , where  $X \in \mathbb{R}^{n \times 1536}$  denotes the matrix of PC-CHiP features and  $\beta \in \mathbb{R}^{1536 \times 42}$  is the weight matrix mapping features to the 42 labels, to include the additional term  $Z\alpha$ , where  $Z \in \{0, 1\}^{n \times f}$  is an indicator matrix of the confounding factors and  $\alpha \in \mathbb{R}^{f \times 42}$  is the corresponding weight matrix, to absorb the undesired effects of different image qualities and prevent them from being implicitly learned on the deeper layer of the model. The results shown in Extended Data Fig. 10 compare the original Inception-V4 architecture with default training parameters and the modified algorithm, with additional data augmentation to supersede any pre-existing JPEG patterns, and heavy color augmentation to overcome systematic differences in hematoxylin and eosin staining (random hue rotations by  $-90$  to  $90^\circ$ )<sup>55</sup>.

**Transfer learning.** Regularized generalized linear models, which are broadly analogous to Inception's original multinomial classification layer, were used to learn molecular associations for each image tile. These models used the set of 1,536 histopathological features and the tissue type encoded as additive indicator variables as predictors, and were fitted using the gllmnet R package<sup>56</sup>. Per-slide predictions were calculated by averaging the prediction of all tiles within that slide. To avoid normal contamination, only samples with a tumor purity of  $\geq 85\%$  were included during training. The model performance was reported by the mean predicted accuracy of fivefold cross-validation, split at the level of patients. A total of 100 tiles were randomly selected from each whole slide for genomic alterations, and 50 tiles were selected for gene expression. Within each fold, tenfold cross-validation was used to select the gllmnet regularization parameters (folds split at the patient level). For each of the five folds, a  $P$  value was calculated by evaluating model predictions on the held-back fifth using Wilcoxon's rank-sum test<sup>48,49</sup> for categorical predictions of genomic data (equivalent to using AUC as a readout), or Spearman's rank correlation test<sup>51</sup> for quantitative predictions of transcriptomic data. The resulting five  $P$  values from each test were combined into a single  $P$  value

statistic using Fisher's method to assess whether there was a measurable level of association across folds<sup>47</sup>. Combined  $P$  values were then adjusted to control the FDR across the entirety of cancer-alteration pairs tested using the method of Benjamini and Hochberg<sup>50</sup>. 95% CIs for the average AUC across folds were estimated using the cvAUC R package<sup>57</sup>. To compare the predictive accuracy of histopathological features with conventional histopathological evaluation methods, we built generalized linear models to predict genomic alterations using histological subtypes or histological grades. AUC values and adjusted  $P$  values were calculated as described above using the same fivefold split. Average rank correlations ( $\rho$ ) and the corresponding 95% CIs were calculated using a  $\tanh^{-1}$  Fisher transformation<sup>58</sup>. The predictive accuracy was evaluated within each cancer type to avoid reporting associations driven by different prevalence and levels of molecular traits across cancer types. Example tiles and slide overlays shown in this study were from held-back validation folds.

**Genomic alterations.** Point mutations (single-nucleotide variants and short deletions and insertions) were called using CaVEMan and Pindel algorithms plus a set of dedicated post-processing filters, as described previously<sup>59</sup>, for 8,769 patients from the TCGA cohort. Absolute copy number was called using the ASCAT algorithm<sup>60</sup>. WGD status was determined using the criteria described previously<sup>27</sup>. Chromosome and chromosome arm level gains and losses were retrieved from ref. <sup>61</sup>. Focal amplifications and deletions were based on the regions defined in ref. <sup>30</sup>. For each of the amplified regions, samples with an absolute copy number of at least ten were called amplified. For each of the deleted regions, only samples with  $\leq 1$  copy in the absence of WGD and samples with  $\leq 2$  copies in the presence of WGD were called deleted. We performed LASSO regularized multinomial regression to classify the gain, loss or non-alteration of 56 chromosomes or chromosome arms. We applied logistic regression with LASSO penalization for dichotomous genomic alterations. Per alteration, AUCs were then calculated in a one versus the rest fashion (for example, gain versus not altered and loss) for each cancer type using the statistical procedures described in the section ‘Transfer learning’.

**Gene expression.** Log-transformed upper-quantile normalized gene expression from RNA sequencing data was used as a readout. We performed linear regression with LASSO penalization on 17,256 genes that were expressed in at least 60% of the samples. The Spearman's rank correlation ( $\rho$ ) and predicted explained variance ( $R^2$ ) were calculated for each gene–cancer pair to evaluate the model performance. Associated  $P$  values for  $\rho > 0$  were estimated by Spearman's rank correlation test<sup>51</sup>. The  $P$  values were then corrected controlling the FWER using the method of Bonferroni. To identify functional classes of genes that can be predicted by histopathology features, we then performed gene set enrichment analysis<sup>62</sup> for a collection of REACTOME pathways<sup>63</sup>. A normalized enrichment score and  $P$  value were calculated for each pathway in each cancer type. The  $P$  values were corrected to control for the FDR. Finally, we performed regression on the gene expression-based proliferation score and TIL signature<sup>35</sup> using the same method used for single-gene expression.

**Prognostic associations.** Survival analysis was performed using penalized Cox's proportional hazard regression<sup>64</sup> using a mixture of  $L_1$  and  $L_2$  regularization, often referred to as the Cox elastic net<sup>65</sup>. To evaluate discriminative performance, we used Harrell's C index as a measure of the concordance between predicted and actual risk<sup>52</sup>. To obtain a scalable and sparse solution, we deployed proximal gradient descent for our parameter updates<sup>66</sup>. Due to the large-scale nature of the problem, an exhaustive hyperparameter search was infeasible. Therefore, hyperparameters, in particular  $L_1/L_2$  penalization strength, were automatically determined using Bayesian optimization<sup>67</sup>. Twenty repetitions of fivefold cross-validation were used to evaluate model performance. Each fold was further split into a training set (85%) and a validation set (15%).

A total of six models, each combining different combinations of variables, were evaluated for specific cancer types from TCGA. Cancers were included in the analysis if the sample size was at least 160 individuals and censorship was  $<90\%$ . The first model (histology) contained the histological subtype and the corresponding tumor grade information. Routine clinical information for each individual included age at cancer diagnosis, gender, cancer stage and the histopathology features form the second model (clinical). The third model (clinical + expression) was a combination of clinical and gene expression data. Model four (PC-CHiP) used the extracted histopathological features from the CNN. Model five (clinical + PC-CHiP) contained the histopathology features and the clinical data. Lastly, Model six (all) was a set of all covariates. If observations were missing, particularly for the gene expression data, mean imputation was applied. The gene expression data comprised the first 30 components of a principal component analysis. For the survival analysis with the histopathology features, each extracted tile was used as an individual observation. A global risk estimate was obtained using the average risk across the tiles from a patient.

Three different strategies were employed to assess the value of adding PC-CHiP to models based on conventional variables. First, it was tested whether the cross-validated linear predictor obtained using PC-CHiP alone added significant signal in a multivariate model. Second, it was assessed whether the pretrained predictor based on PC-CHiP improved the concordance  $C$  in a cross-validation

setting. Third, a likelihood boosting approach was used to train Cox models from scratch, combining clinical/gene expression data with the histopathology features. To compare predictive performance across models, we examined the distribution of concordance indices across folds, as well as the mean difference concordance within folds. Furthermore, we used a paired-Wilcoxon signed-rank test to compare C estimates across models. To account for multiple comparisons, we used the Holm-Bonferroni correction as the FWER procedure. Survival curves were estimated using the Kaplan-Meier estimator.

**External validation using the METABRIC and BASIS dataset.** Hematoxylin and eosin-stained slides from frozen tissue samples were downloaded and tiled into 512 pixel × 512 pixel tiles at 20× magnification in the same fashion as in TCGA. WGD status was calculated using the methods described previously<sup>68</sup>. The amended Inception-V4 architecture preprocessing scripts are available from GitHub (<https://github.com/gerstung-lab/PC-CHip>), and the retrained model checkpoints are available from the BioStudies database (<https://www.ebi.ac.uk/biostudies/>) under accession number S-BSST292. Per-slide predictions for METABRIC and BASIS were calculated using all tiles.

**Expert blinded assessment of TIL counts.** To assess whether predicted TIL levels reflected the true level of immune cells for individual tiles, we randomly selected 150 tiles with different levels of TILs from three cancer types: breast, bladder and lung cancer (each set of 50 tiles were from the highest 10% quantile, the 50–70% quantile and the lowest 10% quantile, respectively). The number of TILs for each tile was independently evaluated by two expert pathologists, who were blinded to the predicted scores. The total number of nuclei was automatically learned using the algorithm described<sup>69</sup>. The relationship between the predicted TIL scores from PC-CHip, the slide-level transcriptomic TIL scores and the true TIL scores was modeled using multiple linear regression.

**Statistics and reproducibility.** All statistical procedures were conducted as outlined above. Data from previously published studies were obtained from the sources stated below. Only tissue types from TCGA with at least 50 images with a magnification greater than 20× were included in this study. Histopathology slides with at least 85% tumor content (existing histopathological information) were used for training. Reproducibility of the image classification analysis was controlled using a randomly selected set of patients for evaluation. Fivefold cross-validation and subsequent hypothesis testing with multiple testing adjustment was performed to control for errors in genomic, transcriptomic and prognostic associations. The validity of discoveries was assessed on two validation cohorts (METABRIC and BASIS). No statistical method was used to predetermine the sample size. The experiments were not randomized. The investigators were not blinded to allocation during the experiments and outcome assessment, except for the histopathological assessment of TILs.

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

## Data availability

TCGA data (images, as well as genomic, transcriptomic and clinical data) are publicly available from <http://gdc.cancer.gov>. For METABRIC, images and genomic and transcriptomic data are available under controlled access at the European Genome-phenome Archive (<https://ega-archive.org/>) under study accession EGAS0000000098, and clinical data are available at <https://www.cbioportal.org/>. For BASIS, genomic data are freely available from <ftp://ftp.sanger.ac.uk/pub/cancer/Nik-ZainalEtAl-560BreastGenomes>, clinical data are published<sup>42</sup>, and histopathology images are available under controlled access at the European Genome-phenome Archive via accession EGAS00001001178. All other data supporting the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

## Code availability

The computational histopathology algorithm and analysis code are available at <https://github.com/gerstung-lab/PC-CHip>. The retrained checkpoints for Inception-V4 and amended Inception-V4 architecture are available from the BioStudies database (<https://www.ebi.ac.uk/biostudies/>) under accession number S-BSST292. Source data are provided with this paper.

Received: 3 April 2020; Accepted: 26 May 2020;

Published online: 27 July 2020

## References

- Lindeman, N. I. et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. *J. Thorac. Oncol.* **8**, 823–859 (2013).
- Woodman, S. E., Lazar, A. J., Aldape, K. D. & Davies, M. A. New strategies in melanoma: molecular testing in advanced disease. *Clin. Cancer Res.* **18**, 1195–1200 (2012).
- Russnes, H. G., Lingjærde, O. C., Børresen-Dale, A.-L. & Caldas, C. Breast cancer molecular stratification: from intrinsic subtypes to integrative clusters. *Am. J. Pathol.* **187**, 2152–2162 (2017).
- Dienstmann, R. et al. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nat. Rev. Cancer* **17**, 79–92 (2017).
- Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
- Bailey, P. et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
- Esteve, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
- 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).
- Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
- Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).
- Hegde, N. et al. Similar image search for histopathology: SMILY. *NPJ Digit. Med.* **2**, 56 (2019).
- Saltz, J. et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**, 181–193.e7 (2018).
- Shia, J. et al. Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology–molecular associations: clinical and biological implications. *Mod. Pathol.* **30**, 599–609 (2017).
- Schaumberg, A. J., Rubin, M. A. & Fuchs, T. J. H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer. Preprint at bioRxiv <https://doi.org/10.1101/064279> (2018).
- Tsou, P. & Wu, C.-J. Mapping driver mutations to histopathological subtypes in papillary thyroid carcinoma: applying a deep convolutional neural network. *J. Clin. Med. Res.* **8**, 1675 (2019).
- Chang, P. et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am. J. Neuroradiol.* **39**, 1201–1207 (2018).
- Kather, J. N. et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **25**, 1054–1056 (2019).
- Kather, J. N., Heij, L. R., Grabsch, H. I. & Kooreman, L. F. S. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* <https://doi.org/10.1038/s43018-020-0087-6> (2020).
- Kather, J. N., Schulte, J., Grabsch, H. I., Loeffler, C. & Mutti, H. S. Deep learning detects virus presence in cancer histology. Preprint at bioRxiv <https://doi.org/10.1101/690206> (2019).
- Yu, K.-H. et al. Association of omics features with histopathology patterns in lung adenocarcinoma. *Cell Syst.* **5**, 620–627.e3 (2017).
- Schmauch, B., Romagnoni, A., Pronier, E. & Saillard, C. Transcriptomic learning for digital pathology. Preprint at bioRxiv <https://doi.org/10.1101/760173> (2019).
- Mobadersany, P. et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl Acad. Sci. USA* **115**, E2970–E2979 (2018).
- Cheng, J. et al. Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. *Cancer Res.* **77**, e91–e100 (2017).
- Yu, K.-H. et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **7**, 12474 (2016).
- Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI Press* **4**, 4278–4284 (2017).
- Cooper, L. A. et al. PanCancer insights from The Cancer Genome Atlas: the pathologist's perspective. *J. Pathol.* **244**, 512–524 (2018).
- Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
- Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
- Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* **171**, 950–965.e28 (2017).
- Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
- Burger, P. C. et al. Small cell architecture—a histological equivalent of EGFR amplification in glioblastoma multiforme? *J. Neuropathol. Exp. Neurol.* **60**, 1099–1104 (2001).

32. Verhaak, R. G. W. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
33. Kebabew, E. et al. The prevalence and prognostic value of BRAF mutation in thyroid cancer. *Ann. Surg.* **246**, 466–470 (2007).
34. O’Hara, A. J. & Bell, D. W. The genomics and genetics of endometrial cancer. *Adv. Genomics Genet.* **2012**, 33–47 (2012).
35. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
36. Nawaz, S., Heindl, A., Koelble, K. & Yuan, Y. Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Mod. Pathol.* **28**, 766–777 (2015).
37. Cabrita, R. et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* **577**, 561–565 (2020).
38. Pollheimer, M. J. et al. Tumor necrosis is a new promising prognostic factor in colorectal cancer. *Hum. Pathol.* **41**, 1749–1757 (2010).
39. Jögi, A., Vaapil, M., Johansson, M. & Pähkinen, S. Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors. *Ups. J. Med. Sci.* **117**, 217–224 (2012).
40. Gooden, M. J. M., de Bock, G. H., Leffers, N., Daemen, T. & Nijman, H. W. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *Br. J. Cancer* **105**, 93–103 (2011).
41. Curtis, C. et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
42. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
43. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
44. Bayraktar, O. A., Bartels, T., Polioudakis, D. & Holmqvist, S. Single-cell *in situ* transcriptomic map of astrocyte cortical layer diversity. Preprint at *bioRxiv* <https://doi.org/10.1101/432104> (2018).
45. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
46. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
47. Elston, R. C. On Fisher’s method of combining *P*-values. *Biom. J.* **33**, 339–345 (1991).
48. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1**, 80–83 (1945).
49. Mann, H. B. & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
50. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Stat. Methodol.* **57**, 289–300 (1995).
51. Best, D. J. & Roberts, D. E. Algorithm AS 89: the upper tail probabilities of Spearman’s rho. *J. R. Stat. Soc. C Appl. Stat.* **24**, 377–379 (1975).
52. Harrell, F. E. Jr, Califff, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *J. Am. Med. Assoc.* **247**, 2543–2546 (1982).
53. Grossman, R. L. et al. Toward a shared vision for cancer genomic data. *N. Engl. J. Med.* **375**, 1109–1112 (2016).
54. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2818–2826 (IEEE, 2016).
55. Tellez, D. et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Preprint at <https://arxiv.org/abs/1902.06543> (2019).
56. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
57. LeDell, E., Petersen, M. & van der Laan, M. Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates. *Electron. J. Stat.* **9**, 1583–1607 (2015).
58. Fieller, E. C., Hartley, H. O. & Pearson, E. S. Tests for rank correlation coefficients. I. *Biometrika* **44**, 470–481 (1957).
59. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041.e21 (2017).
60. Van Loo, P. et al. Allele-specific copy number analysis of tumors. *Proc. Natl Acad. Sci. USA* **107**, 16910–16915 (2010).
61. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* **33**, 676–689.e3 (2018).
62. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
63. Croft, D. et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
64. Cox, D. R. in *Breakthroughs in Statistics* (eds Kotz, S. & Johnson, N. L.) 527–541 (Springer, 1992).
65. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for Cox’s proportional hazards model via coordinate descent. *J. Stat. Softw.* **39**, 1–13 (2011).
66. Singer, Y. & Duchi, J. C. Efficient learning using forward-backward splitting. *Adv. Neural Inf. Proc. Syst.* **22**, 495–503 (2009).
67. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Proc. Syst.* **25**, 2951–2959 (2012).
68. Dentro, S. C. et al. Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. Preprint at *bioRxiv* <https://doi.org/10.1101/312041> (2018).
69. Caicedo, J. C. et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* **16**, 1247–1253 (2019).

## Acknowledgements

A.W.J. and M.G. are supported by grant NNF17OC0027594 from the Novo Nordisk Foundation. L.M. is a recipient of a Cancer Research UK Clinical PhD Fellowship (C20/A20917). L.R.Y. is funded by a Wellcome Trust Clinical Research Career Development Fellowship (214584/Z/18/Z). The results shown here are in part based on data generated by the TCGA Research Network (<https://www.cancer.gov/tcga>). We thank C. Caldas, S.-F. Chin, Y. Yuan and the METABRIC consortium, as well as M. Stratton, M. Van de Vijver and the BASIS consortium for assistance and sharing data. We also thank all members of the Gerstung laboratory, I. Martincorena and A. Lawson for critical comments on the manuscript.

## Author contributions

Y.F. retrieved and quality controlled all images, developed and trained the deep learning algorithms, performed statistical tests for genomic and molecular association and created all of the figures. A.W.J. performed the survival analysis, reviewed the statistical procedures and applied multiple testing adjustments. R.V.T. and M.G. extended the Inception-V4 algorithm. S.G. provided copy number and annotated mutation data. H.V. extracted mutational signature data. A.S. performed nuclei segmentation. L.R.Y. curated validation data. L.M. oversaw the histopathology review, including blinded assessment of TILs, with help from M.J.-L. M.G. conceived of and supervised the study. Y.F., A.W.J. and M.G. wrote the manuscript with input from L.M. and all other authors, who also approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s43018-020-0085-8>.

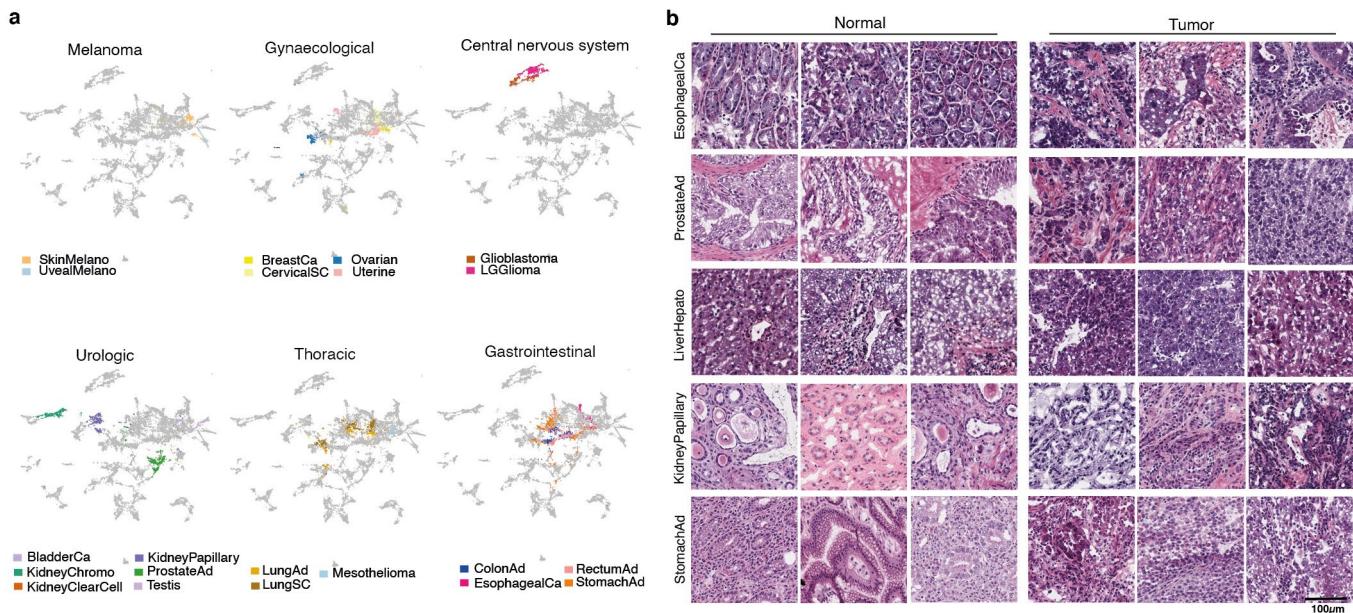
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s43018-020-0085-8>.

**Correspondence and requests for materials** should be addressed to M.G.

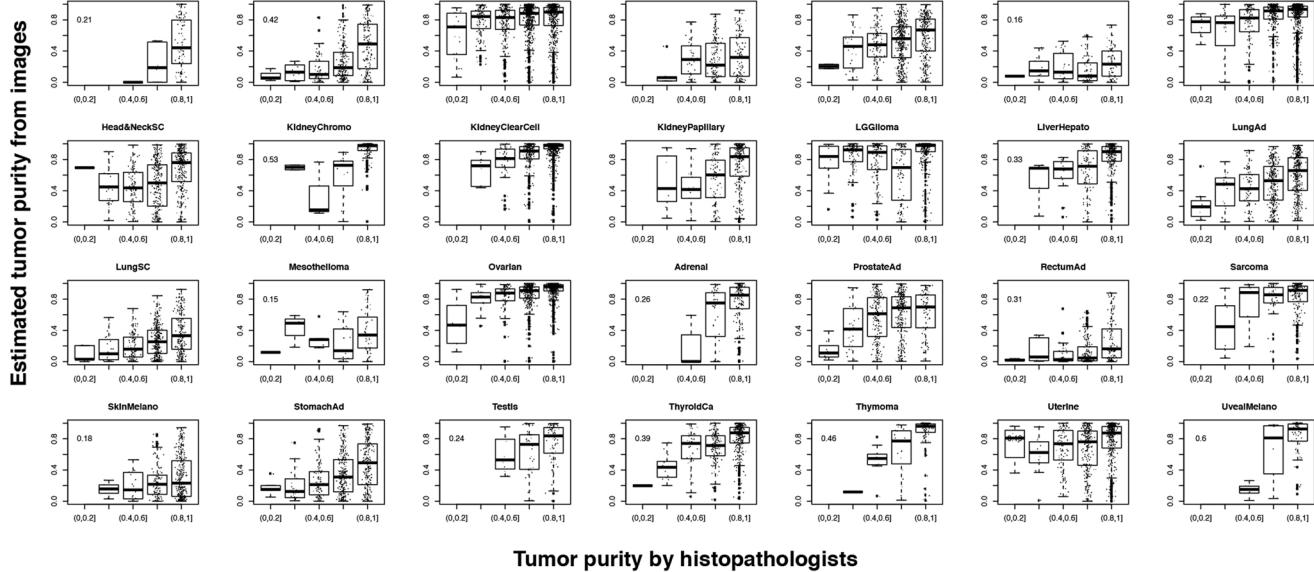
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

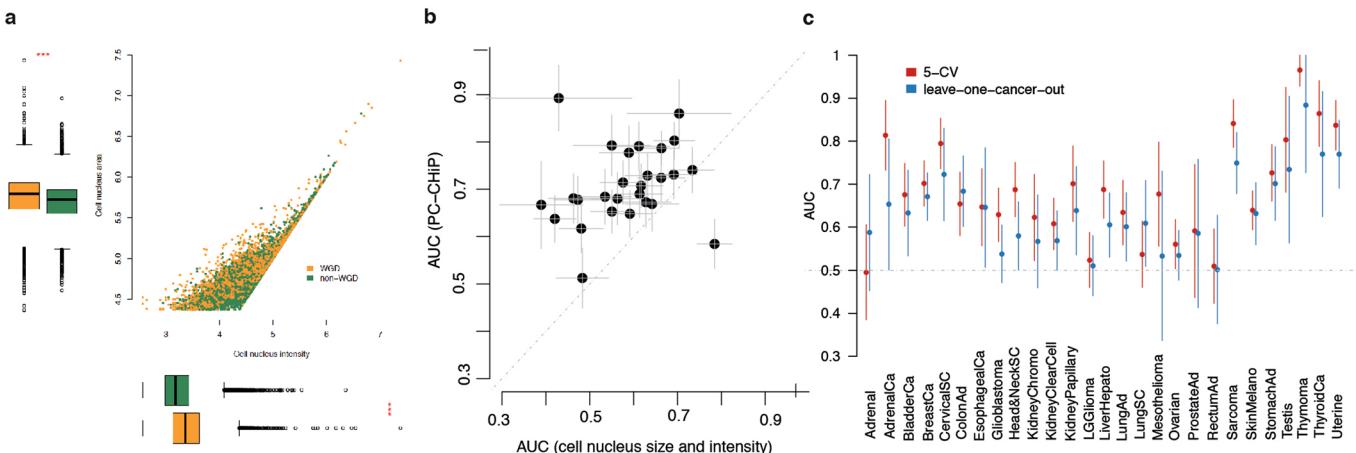
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020



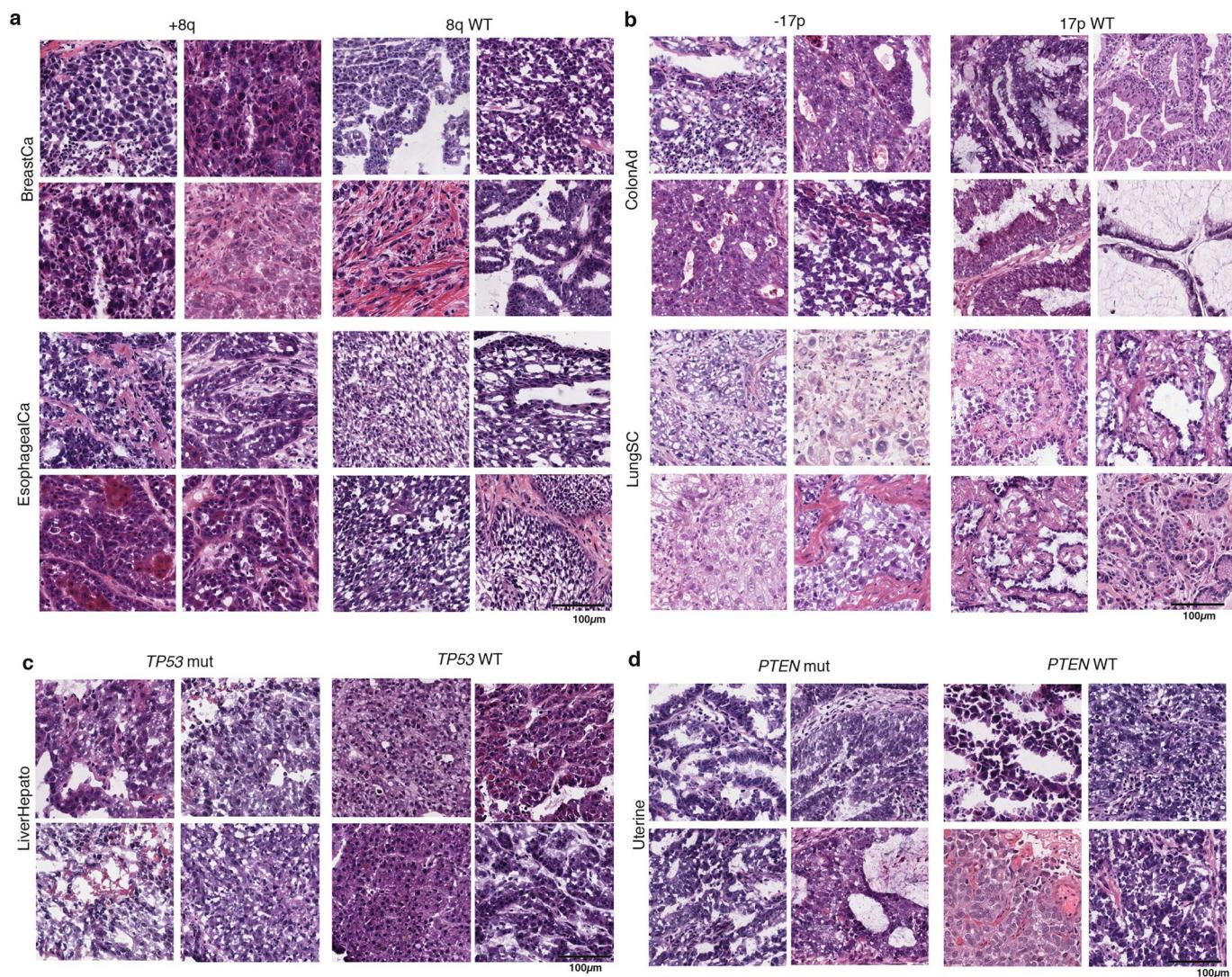
**Extended Data Fig. 1 | Computational histopathological features discriminate between different tissue types.** **a**, UMAP dimensionality reduction representation of the 1,536 histopathological features from randomly selected tiles colored by groups of cancer types (n=200 tiles per tissue type and JPEG quality). **b**, Example tiles from H&E-stained tissue sections of normal and tumor samples from different cancer types (arranged by row, manually selected from best predicted tiles). All tiles are manually selected from best predicted tiles.



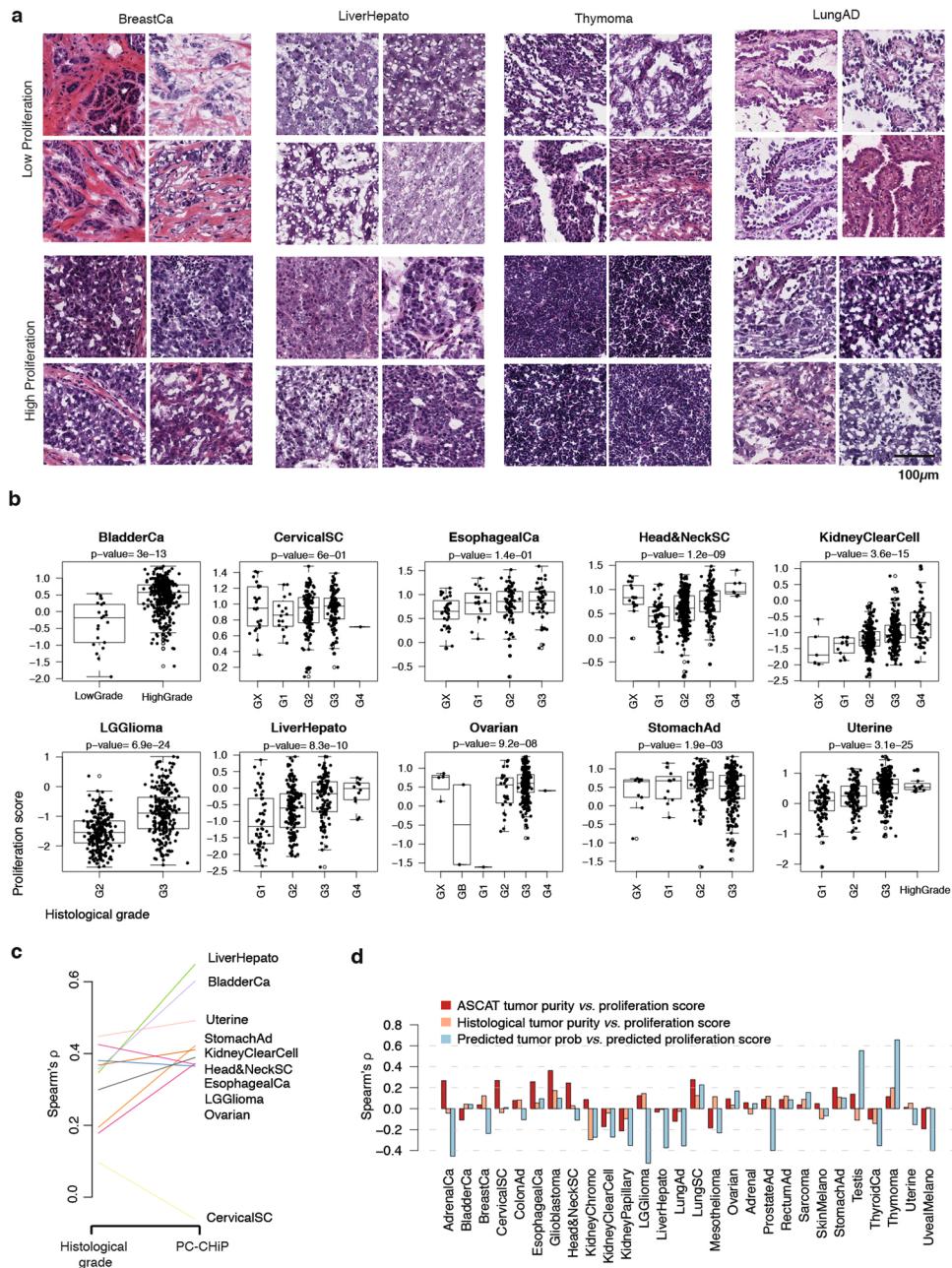
**Extended Data Fig. 2 | The distribution of predicted tumor purity by histopathological features for samples with different histopathologists evaluated tumor purity.** Each boxplot corresponds to one cancer type, each box corresponds to the predicted tumor purity from histopathological features for samples with the histopathologist evaluated tumor purity indicated on x-axis (total number of slides n=14,862). Boxplots depict the quartiles and median, whiskers extend to 1.5× the inter quartile range.



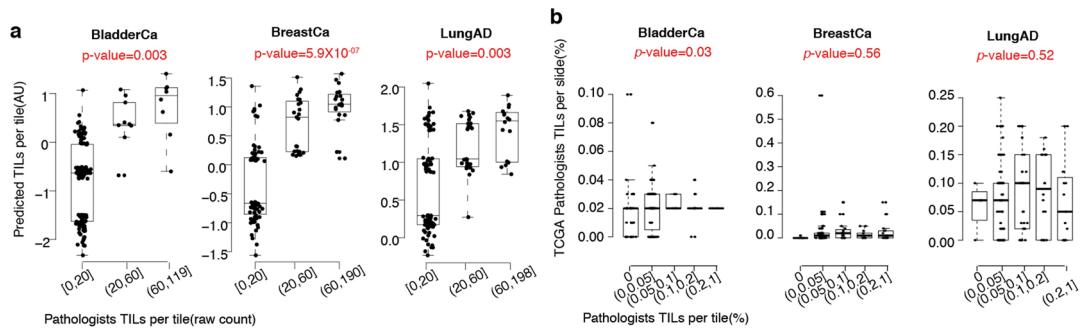
**Extended Data Fig. 3 | Pan-cancer morphological features of whole genome duplications.** **a**, Distribution of cell nucleus size and intensity of samples with and without WGD. Each dot in the scatter plot corresponds to one of 12,000 tiles that were randomly selected across cancer types. The cell nucleus size and intensity were calculated using Cell Profiler with a pipeline provided by the software provider. Boxplots depict the quartiles and median, whiskers extend to 1.5 $\times$  the inter quartile range. **b**, AUC from PC-CHiP (y-axis) compared to hard coded features (x-axis) for a set of n=500 randomly selected tiles for each cancer type. Each dot represents a cancer type. Error bars correspond to 95% confidence intervals. **c**, Histopathological prediction of WGD using 5-fold cross validation (red) and models trained leaving out one cancer type (blue). Error bars correspond to 95% confidence intervals estimated by bootstrap resampling.



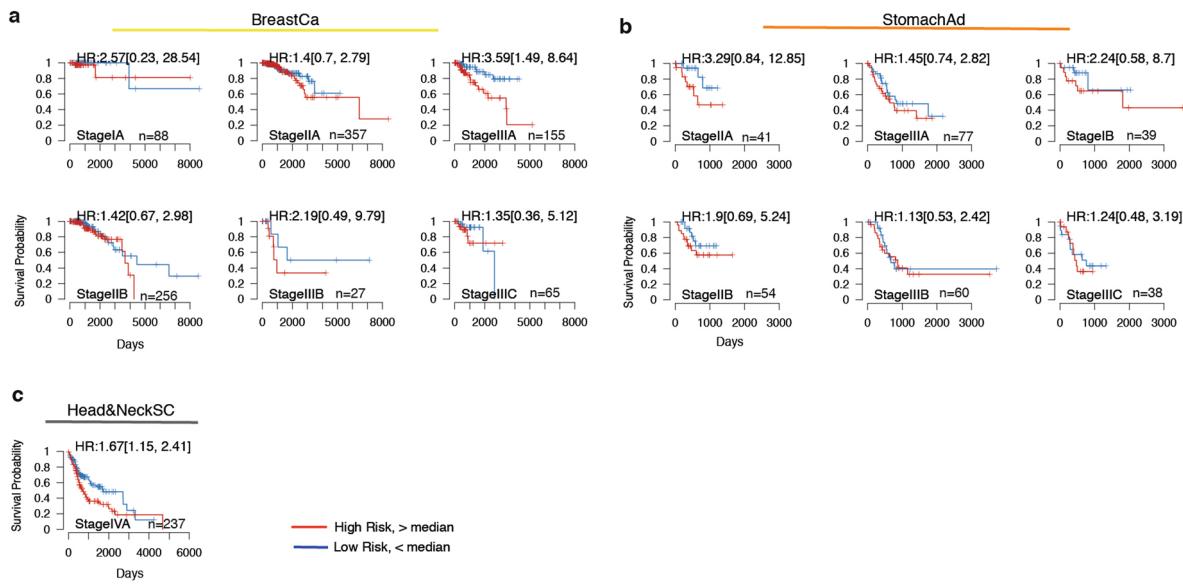
**Extended Data Fig. 4 | Example tiles for associations between computational histopathological and genomic alterations.** **a**, Four example tiles for chromosome 8q gain (left column) and wild type (right column) breast invasive carcinoma (top row) and esophageal carcinoma (bottom row). **b**, Four example tiles for chromosome 17p loss (left column) and wild type (right column) for colon adenocarcinoma (top row) and lung squamous cell carcinoma (bottom row). **c**, Four example tiles for *TP53* mutated (left column) and wild type (right column) liver cancer (hepatocellular carcinomas). **d**, Four example tiles for *PTEN* mutation (left column) and wild type (right column) for uterine cancer. Representative tiles are selected from 100 best predicted tiles.



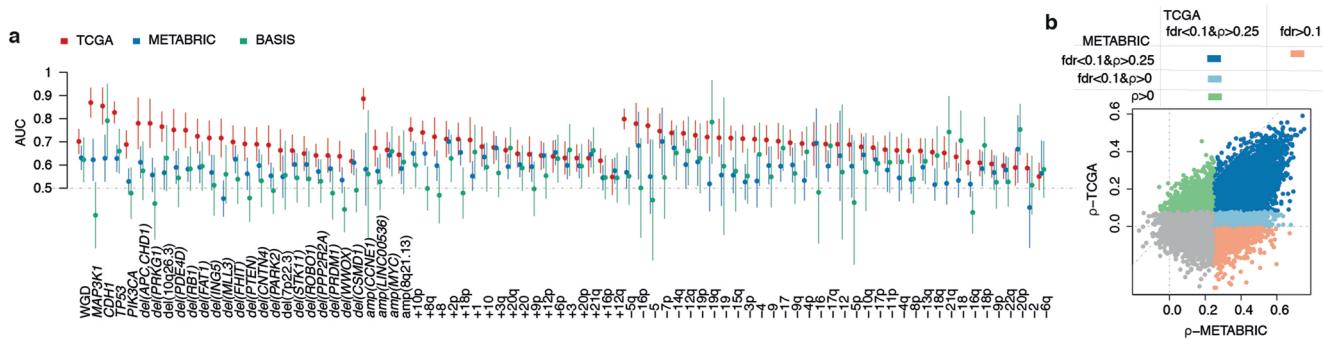
**Extended Data Fig. 5 | Histopathological associations with transcriptomic cell proliferation scores.** **a**, Example tiles for low proliferation (top row) and high (bottom row) for breast invasive carcinoma, liver hepatocellular carcinoma, thymoma and lung adenocarcinoma. Four example tiles manually selected from best predicted tiles are shown for each tumour type. **b**, Boxplots show the different transcriptomic proliferation score for tumors with different histological grades for 10 cancer types with available data ( $n=11,080$ ). G1-G4 corresponds to different grades with G1 being the lowest and G4 the highest, GX stands for “Grade cannot be assessed”, GB stands for “Borderline grade”.  $p$ -values were calculated by ANOVA. Boxplots depict the quartiles and median, whiskers extend to  $1.5 \times$  the inter quartile range. **c**, Figure shows the increases of predictive accuracy of proliferation score from PC-CHiP compared to conventional histological grades. Each line represents one cancer type with the same colors as in Fig. 1e-g. **d**, Barplots showing the correlation of transcriptomic proliferation score and the tumor purity estimated by ASCAT (at patient level), histopathology (at patient level) and predicted tumor probability from PC-CHiP (at tile level) in each cancer ( $n=10,762$  tumor samples for ASCAT,  $n=11,080$  tumor samples for histopathology and  $n=6,188$  tumor samples for PC-CHiP).



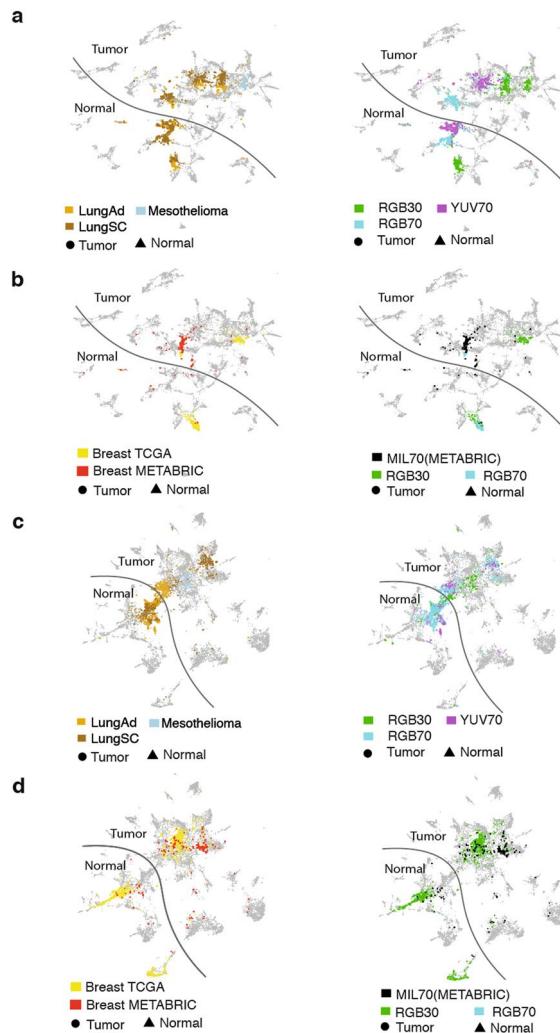
**Extended Data Fig. 6 | Accuracy of TIL scores predicted by PC-ChIP.** **a**, Systematic blinded assessment of TIL raw counts by two expert pathologists for three different cancer types ( $n=150$  for each cancer). Each box plot shows the predicted TIL scores from PC-ChIP for tiles with different TIL raw counts, as independently evaluated by pathologists. **b**, Publicly available slide-level TIL data displays lower concordance compared to with systematic blinded assessment of TIL ( $n=372$  tiles). Each box plot shows the slide level TILs evaluation from TCGA for tiles with different TIL raw counts. Boxplots depict the quartiles and median, whiskers extend to  $1.5\times$  the inter quartile range.



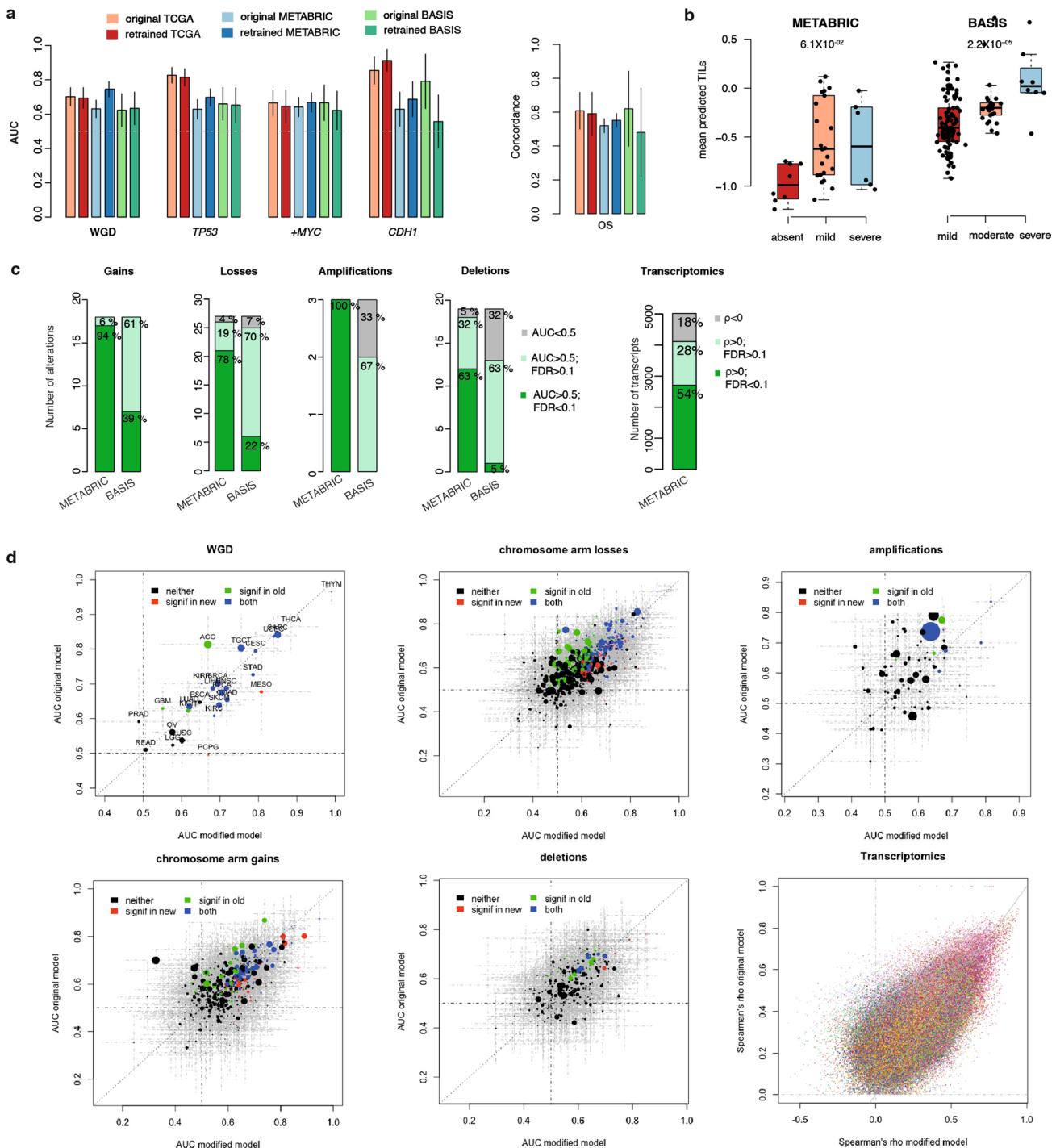
**Extended Data Fig. 7 | Patient risk stratification using histopathological features.** Kaplan-Meier curves for high and low risk groups in different tumor types and stages. **a**, breast invasive carcinoma. **b**, stomach adenocarcinoma. **c**, head and neck squamous cell carcinoma. Only tumor stages with at least 20 patients are shown. Hazard ratios (HR) and the corresponding 95% confidence interval were computed using a Cox proportional hazards model.



**Extended Data Fig. 8 | Overall performance of PC-ChIP in validation datasets.** **a**, The validation accuracy in METABRIC (blue) and BASIS (green) datasets compared to TCGA dataset (red) for each significant association discovered in TCGA indicated at the bottom (total number of genomic alterations tested n=82). Each point corresponds to the predicted AUC for the genomic alteration indicated at the bottom. Error bars correspond to 95% confidence intervals. *p*-value estimated from Wilcoxon's rank sum test and adjusted using FDR. **b**, The distribution of correlation between predicted and true transcript level in METABRIC (x-axis) compared to those in TCGA (y-axis). Each dot represents a gene (n=14,756 genes); blue dots are the genes that can be validated in METABRIC (*Spearman's rank correlation*  $\rho > 0$ , *p*-value estimated using two sided t-test, adjusted FDR<0.1).



**Extended Data Fig. 9 | Histopathological feature representations before and after retraining of Inception-V4.** UMAP representation of the histopathological features from the original Inception model ( $n=200$  tiles randomly selected for each tissue type/JPEG quality) (**a, b**) and the modified, retrained architecture (**c, d**). **a**, lung adenocarcinoma, squamous cell carcinoma and normal lung tissue highlighted. **b**, breast tumor and normal from TCGA and breast tumor from METABRIC highlighted. **c**, as in **a**, but for the modified architecture. **d**, as **c** based on the modified architecture. In each figure, the plot on the right side is colored by tissue type and the plot on the left side is colored by jpeg quality.



**Extended Data Fig. 10 | Molecular associations before and after retraining of Inception-V4.** **a**, AUC for selected genetic alterations and survival for the original and modified Inception architecture. Error bars denote 95% confidence intervals. Sample sizes are n=149 tumor samples for BASIS; for METABRIC, n=454 tumor samples were used for WGD status and copy number alterations; n=434 tumor samples were used for driver gene mutations. Additional details can be found in Supplementary Table 4. **b**, Whole-slide average histopathology predictions for TILs from the modified network (x-axis) relative to expert pathologist categories (y-axis). Boxplots depict the quartiles and median, whiskers extend to 1.5× the inter quartile range. Shown are n=36 tumor samples for METABRIC and n=129 tumor samples for BASIS with available TIL annotation **c**, Distribution of validated (deep green), indeterminate (light green) and invalid (gray) associations in METABRIC and BASIS across different alteration types. Distribution of validated (deep green), indeterminate (light green) and invalid (gray) transcriptomic associations in METABRIC. Sample sizes for genomic associations as in **a**; n=456 tumor samples were used for transcriptomics. Details can be found in Supplementary Table 4. **d**, Scatterplots of genomic and transcriptomic association strengths based on the original (x-axis) and modified (y-axis) Inception model for the TCGA cohort. Predictions from the original model are five-fold cross-validated, while those of the modified architecture are evaluated on a single 70% training / 30% testing split. Sample sizes and the number of alterations tested can be found in Supplementary Table 5.

## 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 [Authors & Referees](#) 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

As this study was a re-analysis of available data, no additional software was used for data collection.

Data analysis

OpenCV 4.1.1  
OpenSlide 3.4.1  
Python 3.4  
Numpy 1.17.3  
tensorflow 1.12  
TF-slim 1.0  
R 3.3.1  
glmnet 3.0.2  
pROC 1.15.3  
cvAUC 1.1.0

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/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

TCGA data (images, genomic, transcriptomic and clinical data) is freely available from <http://gdc.cancer.gov>. For METABRIC, images, genomic and transcriptomic

data is available at the European Genome-Phenome Archive EGA <https://ega-archive.org/> under accession number EGAS00000000098 through controlled access; clinical data is available from <https://www.cbiportal.org/>. For BASIS, genomic data is freely available from <ftp://ftp.sanger.ac.uk/pub/cancer/Nik-ZainalEtAl-560BreastGenomes>; clinical data is published; histopathology images are available as controlled access data at the EGA via study accession EGAS00001001178.

## 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](http://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 was an exploratory analysis using available data from TCGA. No formal sample size calculations were made. Therefore the analysis does not provide evidence for absence of effects.
Data exclusions	No exclusion criteria were pre-established. Only tissue types with at least 50 images with a magnification greater than 20X are included in this study. Tiles with poor quality were removed (Methods, Images). Slides were excluded for the initial classification and molecular association testing if the tumour purity was lower than 85%. All slides were included for survival analyses.
Replication	The replication of the findings was assessed using 5 fold cross validation (classification, genomic, transcriptomic associations and survival) Findings for breast cancer were validated using data from the METABRIC and BASIS studies. 80/82 and 66/82 genomic associations replicated in METABRIC and BASIS, respectively ( $AUC > 0.5$ ), as well as 85% of transcriptomic associations replicated in METABRIC ( $\rho > 0$ ). These results are shown in Figure 8, Extended Data Figure 8 and Supplementary Table 4.
Randomization	No experimental groups were defined.
Blinding	A blinded evaluation of Tumor Infiltrating Lymphocytes was performed by 2 expert pathologists independently on a set of randomly selected tiles for low, intermediate and high TIL predictions for 3 cancer types (bladder, breast and lung cancer).

## 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
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data

### 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