

## RESEARCH ARTICLE

## Identifying transcriptomic correlates of histology using deep learning

Liviu Badea , Emil Stănescu

Artificial Intelligence and Bioinformatics Group, National Institute for Research and Development in Informatics, Bucharest, Romania

\* [badea.liviu@gmail.com](mailto:badea.liviu@gmail.com)

## Abstract

Linking phenotypes to specific gene expression profiles is an extremely important problem in biology, which has been approached mainly by correlation methods or, more fundamentally, by studying the effects of gene perturbations. However, genome-wide perturbations involve extensive experimental efforts, which may be prohibitive for certain organisms. On the other hand, the characterization of the various phenotypes frequently requires an expert's subjective interpretation, such as a histopathologist's description of tissue slide images in terms of complex visual features (e.g. 'acinar structures'). In this paper, we use Deep Learning to eliminate the inherent subjective nature of these visual histological features and link them to genomic data, thus establishing a more precisely quantifiable correlation between transcriptomes and phenotypes. Using a dataset of whole slide images with matching gene expression data from 39 normal tissue types, we first developed a Deep Learning *tissue classifier* with an accuracy of 94%. Then we searched for *genes whose expression correlates with features inferred by the classifier* and demonstrate that Deep Learning can automatically derive visual (phenotypical) features that are well correlated with the transcriptome and therefore *biologically interpretable*. As we are particularly concerned with *interpretability* and *explainability* of the inferred histological models, we also develop *visualizations* of the inferred features and compare them with gene expression patterns determined by immunohistochemistry. This can be viewed as a first step toward bridging the gap between the level of genes and the cellular organization of tissues.

 OPEN ACCESS

**Citation:** Badea L, Stănescu E (2020) Identifying transcriptomic correlates of histology using deep learning. PLoS ONE 15(11): e0242858. <https://doi.org/10.1371/journal.pone.0242858>

**Editor:** Yan Chai Hum, University Tunku Abdul Rahman, MALAYSIA

**Received:** August 10, 2020

**Accepted:** November 11, 2020

**Published:** November 25, 2020

**Copyright:** © 2020 Badea, Stănescu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data used in this study was downloaded from the GTEx portal and is already available at <https://gtexportal.org/home/>. See [S2 File](#) for list of samples.

**Funding:** This work was partially supported by the projects PN 1937-0601/2019 (LB, ES) and PN 1937-0301/CPN 301 300/2019 (LB) funded by the Romanian Ministry of Education and Research ([www.edu.ro](http://www.edu.ro)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

## Introduction

Histological images have been used for biological research and clinical practice already since the 19<sup>th</sup> century and are still employed in standard clinical practice for many diseases, such as various cancer types. On the other hand, the last few decades have witnessed an exponential increase in sophisticated genomic approaches, which allow the dissection of various biological phenomena at unprecedented molecular scales. But although some of these omic approaches have entered clinical practice, their potential has been somewhat tempered by the heterogeneity of individuals at the genomic and molecular scales. For example, omic-based predictors of cancer evolution and treatment response have been developed, but their clinical use is still

**Competing interests:** The authors have declared that no competing interests exist.

**Abbreviations:** CNN, Convolutional Neural Network; DL, Deep Learning; gBP, guided Backpropagation; GTEx, Genotype-Tissue Expression project; H&E, hematoxylin and eosin; IHC, immunohistochemistry; WSI, Whole Slide Image.

limited [1] and they have not yet replaced the century old practice of histopathology. So, despite the tremendous recent advances in genomics, histopathological images are still often the basis of the most accurate oncological diagnoses, information about the microscopic structure of tissues being lost in genomic data.

Thus, since histopathology and genomic approaches have their own largely non-overlapping strengths and weaknesses, a combination of the two is expected to lead to an improvement in the state of the art. Of course, superficial combinations could be easily envisioned, for example by constructing separate diagnosis modules based on histology and omics respectively, and then combining their predictions. Or, alternatively, we could regard both histopathological images and genomic data as features to be used jointly by a machine learning module [2].

However, for a more in-depth integration of histology and genomics, a better understanding of the relationship between genes, their expression and histological phenotypes is necessary. Linking phenotypes to specific gene expression profiles is an extremely important problem in biology, as the transcriptomes are assumed to play a causal role in the development of the observed phenotype. The definitive assessment of this causal role requires studying the effects of gene perturbations. However, such genome-wide perturbations involve extensive and complex experimental efforts, which may be prohibitive for certain model organisms.

In this paper we try to determine whether there is a link between the *expression of specific genes* and *specific visual features* apparent in histological images. Instead of observing the effects of gene perturbations, we use representation learning based on deep convolutional neural networks [3] to automatically infer a large set of visual features, which we correlate with gene expression profiles.

While gene expression values are easily quantifiable using current genomic technologies, determining and especially quantifying visual histological features has traditionally relied on subjective evaluations by experienced histopathologists. However, this represents a serious bottleneck, as the number of qualified experts is limited and often there is little consensus between different histologists analyzing the same sample [4]. Therefore, we employ a more objective visual feature extraction method, based on deep convolutional neural networks. Such more objective visual features are much more precisely quantifiable than any subjective features employed by human histopathologists. Therefore, although this correlational approach cannot fully replace perturbational studies of gene-phenotype causation, it is experimentally much easier and at the same time much more precise, due to the well-defined nature of the visual features employed.

In our study, we concentrate on gene expression rather than other multi-omic modalities, since the transcriptional state of a cell frequently seems to be one of the most informative modalities [5].

There are numerous technical issues and challenges involved in implementing the above-mentioned research.

The advent of *digital* histopathology [6] has enabled the large scale application of automated computer vision algorithms for analysing histopathological samples. Traditionally, the interpretation of histopathological images required experienced histopathologists as well as a fairly long analysis time. Replicating this on a computer was only made possible by recent breakthroughs in Deep Learning systems, which have achieved super-human performance in image recognition tasks on natural scenes, as in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [7, 8]. However, directly transferring these results to digital histopathology is hampered by the much larger dimensions of histological whole slide images (WSI), which must be segmented into smaller image patches (or “tiles”) to make them fit in the GPU memory of existing Deep Learning systems. Moreover, since there are far fewer annotated WSI

than natural images and since annotations are associated with the whole slide image rather than the individual image tiles, conventional supervised Deep Learning algorithms cannot always be directly applied to WSI tiles. This is particularly important whenever the structures of interest are rare and thus do not occur in all tiles. For example, tumor cells may not be present in all tiles of a cancer tissue sample. Therefore, we concentrate in this paper on normal tissue samples, which are more homogeneous at not too high magnifications and for which the above-mentioned problem is not as acute as in the case of cancer samples.

Only a few large-scale databases of WSI images are currently available, and even fewer with paired genomic data for the same samples. The Cancer Genome Atlas (TCGA) [9] has made available a huge database of genomic modifications in a large number of cancer types, together with over 10,000 WSI. The Camelyon16 and 17 challenges [10] aimed at evaluating new and existing algorithms for automated detection of metastases in whole-slide images of hematoxylin and eosin stained lymph node tissue sections from breast cancer patients. The associated dataset includes 1,000 slides from 200 patients, whereas the Genotype-Tissue Expression (GTEx) [11] contains over 20,000 WSI from various normal tissues.

Due to the paucity of large and well annotated WSI datasets for many tasks of interest (such as cancer prognosis), some research groups employed *transfer learning* by reusing the feature layers of Deep Learning architectures trained on ordinary image datasets, such as ImageNet. Although histopathological images look very different from natural scenes, they share basic features and structures such as edges, curved contours, etc., which may have been well captured by training the network on natural images. Nevertheless, it is to be expected that training networks on genuine WSI will outperform the ones trained on natural scenes [12].

The most active research area involving digital histopathology image analysis is computer assisted diagnosis, for improving and speeding-up human diagnosis. Since the errors made by Machine Learning systems typically differ from those made by human pathologists [13], classification accuracies may be significantly improved by *assisting* humans with such automated systems. Moreover, reproducibility and speed of diagnosis will be significantly enhanced, allowing more standardized and prompt treatment decisions. Other supervised tasks applied to histopathological images involve [14]: detection and segmentation of regions of interest, such as automatically determining the tumour regions in WSI [15], scoring of immunostaining [16], cancer staging [13], mitosis detection [17], gland segmentation [18], or detection and quantification of vascular invasion [19].

Machine Learning has also been used for discovering new clinical-pathological relationships by correlating histo-morphological features of cancers with their clinical evolution, by enabling analysis of huge amounts of data. Thus, relationships between morphological features and somatic mutations [12, 20, 21], as well as correlations with prognosis have been tentatively addressed. For example, [22] developed a prognosis predictor for lung cancer using a set of predefined image features. Still, this research field is only at the beginning, awaiting extensive validation and clinical application.

From a technical point of view, the number of *Deep Learning architectures*, systems and approaches used or usable in digital histopathology is bewildering. Such architectures include supervised systems such as Convolutional Neural Networks (CNN), or Recurrent Neural Networks (RNN), in particular Long Short Term Memory (LSTM) [23] and Gated Recurrent Units [24]. More sophisticated architectures, such as the U-net [25], multi-stream and multi-scale architectures [26] have also been developed to deal with the specificities of the domain. Among unsupervised architectures we could mention Deep Auto-Encoders (AEs) and Stacked Auto-Encoders (SAEs), Generative Adversarial Networks [3, 27], etc.

Currently, CNNs are the most used architectures in medical image analysis. Several different *CNN architectures*, such as AlexNet [7], VGG [28], ResNet [29], Inception v3 [30], etc.

have proved popular for medical applications, although there is currently no consensus on which performs best. Many open source Deep Learning *systems* have appeared since 2012. The most used ones are *TensorFlow* (Google [31]), *Keras* [32] (high level API working on top of TensorFlow, CNTK, or Theano) and *PyTorch* [33].

In this paper we look for gene expression-phenotype correlations for normal tissues from the Genotype-Tissue Expression (GTEx) project, the largest publicly available dataset containing paired histology images and gene expression data. The phenotypes consist in visual histological features automatically derived by various convolutional neural network architectures implemented in PyTorch and trained, validated and tested on 1,670 whole slide images at 10x magnification divided into 579,488 tiles. A supervised architecture was chosen, as the inferred visual features should be able to discriminate between the various tissues of interest instead of just capturing the components with the largest variability in the images.

## Materials and methods

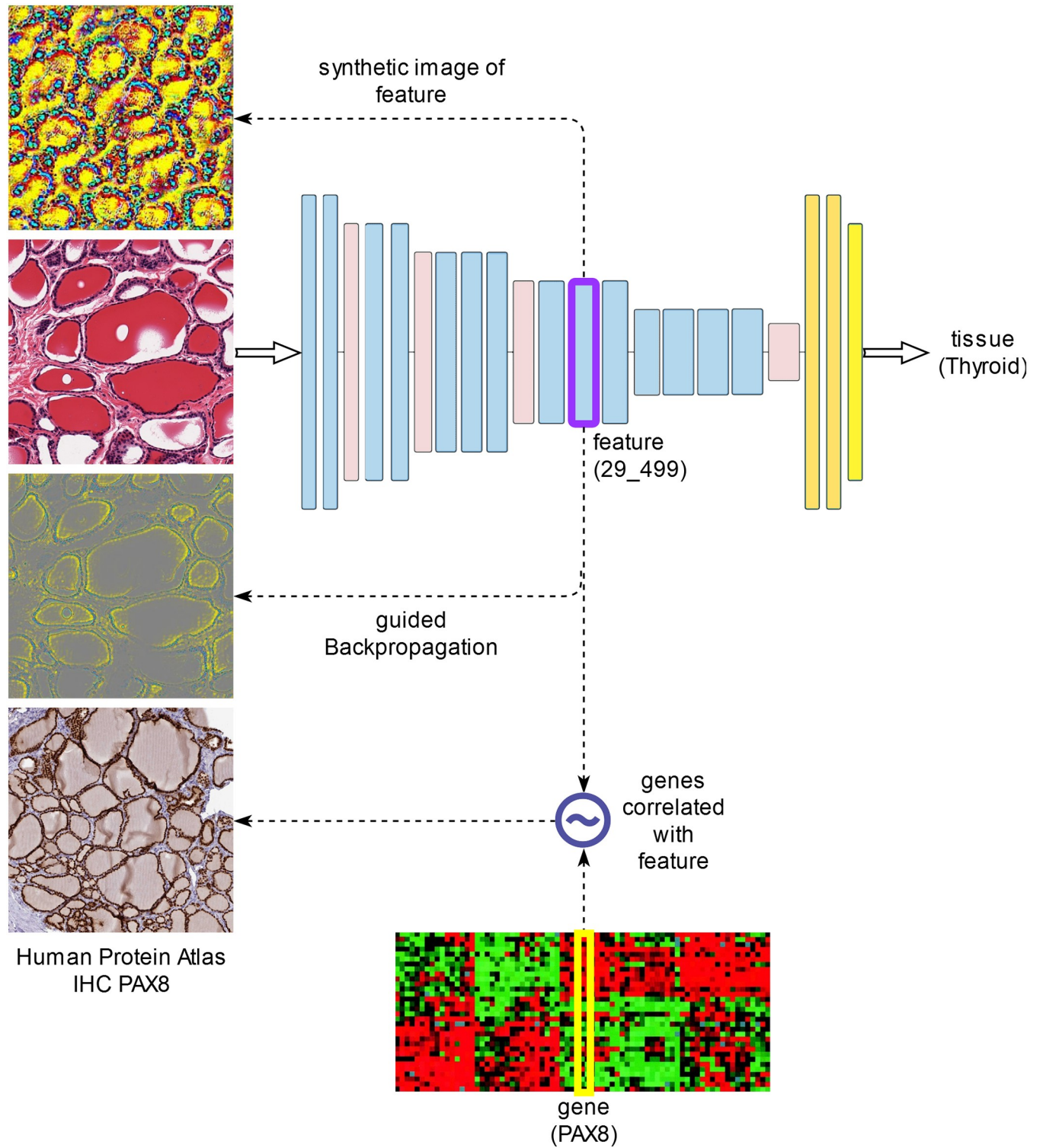
Our study of the correlations between visual histological features and gene expression involves several stages:

1. Automated *feature discovery* using representation learning based on a supervised Convolutional Neural Network (CNN) trained to classify histological images of various normal tissues; validation and testing of the tissue classifier on completely independent samples.
2. *Quantification* of the inferred histological features in the various layers of the network and their *correlation with paired gene expression data* in an independent dataset.
3. *Visualization of features* found correlated to specific gene expression profiles. Two different visualization methods are used: *guided backpropagation* on specific input images, as well as input-independent generation of *synthetic images* that maximize these features.

The following sections describe these analysis steps in more detail (see also Fig 1).

### Developing a classifier of histological images

First, we developed a normal tissue classifier based on histological whole slide images stained with hematoxylin and eosin (H&E). Such a tissue classifier is also useful in its own right, since it duplicates the tasks performed by skilled histologists, whose expertise took years or even decades to perfect. Such a task could not even be reliably performed by computer before the recent advent of Deep Learning. But although Deep Learning has been able to surpass humans in classifying objects from natural images (as in the ImageNet competition [8]), transferring these results to digital histopathology is far from trivial [14]. This is mainly because histopathology images are far larger (up to tens of billions of pixels) and labelled data much sparser than for natural images used in other visual recognition tasks, such as ImageNet. Whole slide images (WSI) thus need to be divided in tiles of sizes small enough (typically 224x224, or 512x512 pixels) to allow Deep Learning mini-batches to fit in the memory of existing GPU boards. Moreover, since there are far fewer annotated WSI than natural images and since annotations are associated with the whole slide image rather than the individual image tiles, conventional supervised Deep Learning algorithms cannot always be directly applied to WSI tiles. This is particularly important whenever the structures of interest are rare and thus do not occur in all tiles. (For example, tumor cells may not be present in all tiles of a cancer tissue sample.) Thus, the main strength of Deep Learning in image recognition, namely the huge amounts of labelled data, is not available in the case of histopathological WSI. More sophisticated algorithms based on multiple instance learning [34] or semi-supervised learning are



**Fig 1. The main analysis steps.**

<https://doi.org/10.1371/journal.pone.0242858.g001>

needed, but haven't been thoroughly investigated in this domain. In multiple instance learning, labels are associated to *bags* of instances (i.e. to the whole slide image viewed as a bag of tiles), rather than individual instances, making the learning problem much more difficult. In contrast, semi-supervised learning employs both labelled and unlabelled data, the latter for obtaining better estimates of the true data distribution.

Therefore, we concentrate in this paper on *normal tissue* samples, which are more homogeneous at not too high magnifications and for which the above-mentioned problem is not as acute as in the case of cancer samples. In the following, we used data from the *Genotype-Tissue Expression* (GTEx) project, the largest publicly available dataset containing paired histology images and gene expression data.

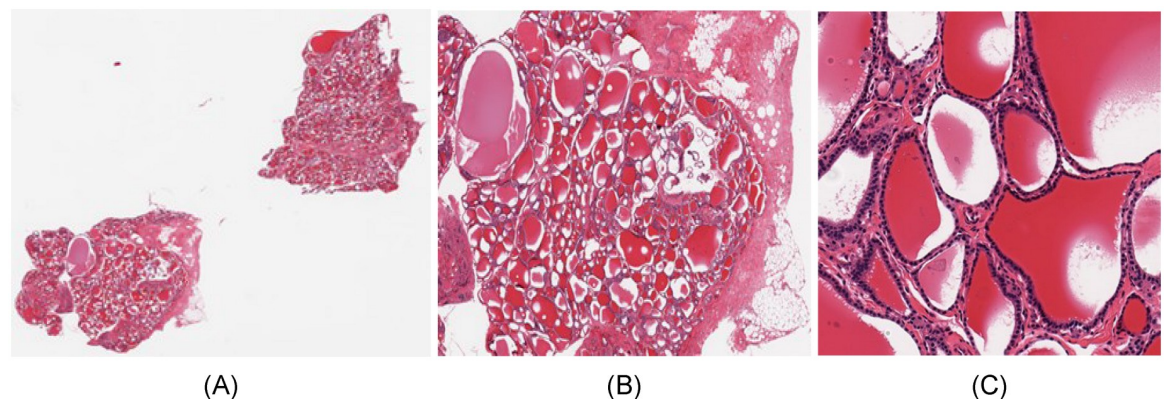
**The GTEx dataset.** GTEx is a publicly available resource for data on tissue-specific gene expression and regulation [35]. Samples were collected from over 50 normal tissues of nearly 1,000 individuals, for which data on whole genome or whole exome sequencing, gene expression (RNA-Seq), as well as histological images are available.

In our application, we searched for subjects for which both histological images and gene expression data (RNA-Seq) are available. We selected from these 1,670 histological images from 39 normal tissues, with 1,778 associated gene expression samples (for 108 of the histological images there were duplicate gene expression samples). The 1,670 histological images were divided into three data sets:

- *training set* (1,006 images, representing approximately 60% of the images),
- *validation set* (330 images, approximately 20% of images),
- *test set* (334 images, approximately 20% of images).

Fig 2A and 2B show such a histological whole slide image (WSI) and a detail of a thyroid gland sample.

Another very important aspect specific to WSI is the optimal level of *magnification* to be fed as input to the image classification algorithms. As tissues are composed of discrete cells, critical information regarding cell shape is best captured at high resolutions, whereas more complex structures, involving many cells, are better visible at lower resolutions. The maximum image acquisition resolution in GTEx SVS files corresponds to a magnification of 20x (0.4942 mpp—microns per pixel). In this paper, we chose an intermediate magnification level of 10x (0.9884 mpp), which represents a reasonable compromise between capturing enough cellular



**Fig 2. Whole slide image of thyroid sample GTEX-11NSD-0126.** (A) Whole slide. (B) Detail. (C) Tile of size 512x512.

<https://doi.org/10.1371/journal.pone.0242858.g002>

details, and allowing a field of view of the image tiles large enough to include complex cellular structures. The whole slide images were divided into tiles of size 512x512 pixels (506 $\mu$ m x 506 $\mu$ m), which were subsequently rescaled to 224x224, the standard input dimension for the majority of convolutional neural network architectures. Image tiles with less than 50% tissue, as well as those with a shape factor other than 1 (with a non-square shape, originating from WSI edges) were removed so as not to distort the training process, and the remaining tiles were labelled with the identifier of the tissue of origin of the whole image. A 512x512 image tile of the WSI from Fig 2A can be seen in Fig 2C.

The image tiles dataset contains 579,488 such tiles (occupying 72GB of disk space) and was further split into distinct training, validation and test datasets as follows:

- 349,340 tiles in the training set (taking up 44GB of disk space),
- 114,123 tiles in the validation set (14GB),
- 116,025 tiles in the test set (15GB).

S1 Table shows the numbers of images and respectively image tiles in the three datasets (training, validation and test) for each of the 39 tissue types (see also S2 File for the list of GTEx samples). By comparison, the ImageNet dataset (from the ILSVRC-2012 competition) had 1.2 million images and 1,000 classes (compared to 39 classes in the GTEx dataset).

Since to the best of our knowledge, we could not find, for validation purposes, other comparable datasets with paired histological images and gene expression data, we did not perform a color normalization optimized to the GTEx dataset, as it might not extrapolate well to other datasets with different color biases. Therefore, we used the standard color normalization employed in the ImageNet dataset. Such a normalization may be slightly suboptimal, but this is expected to have little impact on the classification results, due to the complexity of the CNNs employed.

**Training and validation of deep learning tissue classifiers.** We trained and validated several different convolutional neural network architectures on the image tile dataset described above (Table 1).

The VGG16\_1FC model is a modification of the VGG16 model, in which the final three fully connected layers have been replaced by a single fully connected layer, preceded by a drop-out layer. This not only significantly reduces the very large number of parameters of the VGG16 architecture (from 134,420,327 to 15,693,159), but also attempts to obtain more intuitive representations on the final convolutional layers.

Since the last convolutional layer of VGG16 is of the form 7x7 x 512 channels and since location is completely irrelevant in tissue classification, we also extended the convolutional part of VGG16 by an average pooling layer (over the 7x7 spatial dimensions), followed as in VGG16\_1FC by a single fully connected layer. We refer to this new architecture as VGG16\_avg1FC.

Architectures with the suffix “bn” use “batch normalization” [36].

Note that in the case of ResNet34, the internal layers of the residual blocks do not make up complete representations, since their outputs are *added* to the skip connections. Therefore, the ResNet features considered here involve only layers that are not “skipped over”.

On the other hand, the Inception\_v3 architecture contains many *redundant* representations, since it combines by concatenation individual multi-scale filter outputs (the individual filter outputs are repeated in the concatenated layer). To avoid this representational redundancy, we only consider as features the *elementary* layers with non-negative outputs (ReLU, max- or average pooling) that are not on the auxiliary branch.

**Table 1. Convolutional neural network architectures used in this paper.** For ResNet34, the features considered involve only layers that are not “skipped over”, while for Inception\_v3 only elementary layers with non-negative outputs (ReLU, max/avg-pooling) that are not on the auxiliary branch.

Network	Number of features $f_z$	Number of trainable parameters	Reference
AlexNet	2,816	57,163,623	Krizhevsky, 2012 [7]
VGG11	6,976	128,926,119	Simonyan, 2014 [28]
VGG13	7,360	129,110,631	Simonyan, 2014 [28]
VGG16	9,920	134,420,327	Simonyan, 2014 [28] (0)Conv2d(3,64) (1)ReLU (2)Conv2d(64,64) (3)ReLU (4)MaxPool2d (5)Conv2d(64,128) (6)ReLU (7)Conv2d(128,128) (8)ReLU (9)MaxPool2d (10)Conv2d(128,256) (11)ReLU (12)Conv2d (256,256) (13)ReLU (14)Conv2d(256,256) (15)ReLU (16)MaxPool2d (17)Conv2d (256,512) (18)ReLU (19)Conv2d (512,512) (20)ReLU (21)Conv2d(512,512) (22)ReLU (23)MaxPool2d (24)Conv2d(512,512) (25)ReLU (26)Conv2d(512,512) (27)ReLU (28)Conv2d(512,512) (29)ReLU (30)MaxPool2d (31)Linear (25088,4096) (32)ReLU (33)Dropout(0.5) (34)Linear(4096,4096) (35)ReLU (36)Dropout(0.5) (37)Linear(4096,39)
VGG19	12,480	139,730,023	Simonyan, 2014 [28]
VGG11_bn	9,728	128,931,623	Simonyan, 2014 [28]; Ioffe, 2015 [36]
VGG13_bn	10,304	129,116,519	Simonyan, 2014 [28]; Ioffe, 2015 [36]
VGG16_bn	14,144	134,428,775	Simonyan, 2014 [28]; Ioffe, 2015 [36]
VGG19_bn	17,984	139,741,031	Simonyan, 2014 [28]; Ioffe, 2015 [36]
ResNet34	28,992	21,304,679	He, 2016 [29]
Inception_v3	27,712	24,453,166	Szegedy, 2016 [30]
VGG16_1FC	9,920	15,693,159	This paper: same convolutional part as VGG16, but with a single fully connected layer: Conv(VGG16); Dropout(0.5); <b>Linear(25088,39)</b>
VGG16_avg1FC	10,432	14,734,695	This paper: the convolutional part as VGG16, followed by an average pooling layer and a single fully connected layer: Conv(VGG16); <b>AvgPool2d(7,7,512;1,1,512)</b> ; Dropout(0.5); Linear(512,39)

<https://doi.org/10.1371/journal.pone.0242858.t001>

We report, in the following, the accuracies of the classifiers on both *validation* and *test* datasets, after training for 90 epochs on the *training* dataset, with the initial learning rate of 0.01 and reducing it by a factor of 10 every 30 epochs. The mini-batch sizes were chosen according to the available GPU memory (11GB): 256 for AlexNet and 40 for the other models, respectively.

Note that although the validation dataset has not been used for model *construction*, it has been employed for model *selection* (corresponding to the training epoch with the best model accuracy on the validation data). Therefore, we use a completely independent *test* dataset for model evaluation.

The models were implemented in Pytorch 1.1.0 (<https://pytorch.org/>) [33], based on the torchvision models library (<https://github.com/pytorch/vision/tree/master/torchvision/models>).

As the histological images contain multiple tiles, we also constructed a *tissue classifier for whole slide images (WSI)* by applying the tile classifier on all tiles of the given WSI and aggregating the corresponding predictions using the majority vote.

## Correlating histological features with gene expression data

From a biological point of view, the visual characteristics of a histological image, its phenotype, is determined by the gene expression profiles of the cells making up the tissue. But the way in which transcriptomes determine phenotypes was practically impossible to determine automatically until the advent of machine learning methods for learning representations based on



Deep Learning, mainly because most histological phenotypes were hard to define precisely and especially difficult to quantify automatically.

The internal representations constructed by a convolutional neural network can be viewed as visual features detected by the network in a given input image. These visual features, making up the phenotype, can be correlated to paired gene expression data to determine the most significant gene-phenotype relationships. However, to compute these correlations, we need a precise *quantification* of the visual features.

The *quantification of the inferred visual features* for a given input image is non-trivial, as it needs to be *location invariant*—the identity of a given tissue should not depend on spatial location. Although the activations  $Y_{lzy}(X)$  of neurons  $(x,y,z)$  in layer  $l$  for input image  $X$  obviously depend on the location  $(x,y)$ , we can construct location invariant aggregates of these values of the form:

$$f_l(X) = \sum_{x,y} Y_{lzy}(X)^p.$$

We have experimented with various values of  $p$ , but in the following we show the simplest version,  $p = 1$ . This is equivalent to computing the spatial averages of neuron values in a given layer  $l$  and channel  $z$ . Thus, we employ *spatially invariant features* of the form  $f_{lz}$ , which are quantified by forward propagation of histological images  $X$  using the formula above.

Since histological whole slide images  $W$  were divided into multiple tiles  $X$ , we also sum over these tiles to obtain the feature value for  $W$ :

$$f_{lz}(W) = \sum_{X \in W} f_{lz}(X).$$

After quantification of features  $f_{lz}(W_i)$  over all samples  $i$ , we compute pairwise gene-feature Pearson correlations  $r(g, f_{lz})$ , where  $g_i$  are the gene expression values of gene  $g$  in the samples  $i$  (for each sample  $i$ , we have simultaneous histological whole slide image  $W_i$  and gene expression data  $g_i$  for virtually all human genes  $g$ ).

To avoid any potential data leakage, we compute gene-feature correlations on the *test dataset*, which is completely independent from the datasets used for model construction and respectively selection (*training* and respectively *validation* datasets).

Note that since we are in a ‘small sample’ setting (the number of variables greatly exceeds the number of samples), we performed a univariate analysis rather than a multivariate one, since multivariate regression is typically not recommended for small samples. For example, in the case of the VGG16 architecture, we have 9,920 features  $f_{lz}$ , 56,202 genes (totaling 557,523,840 gene-feature pairs) and just 334 samples in the test dataset.

The selection of significantly correlated gene-feature pairs depends on the correlation- and gene expression thresholds used. Genes with low expression in all tissues under consideration should be excluded, but determining an appropriate threshold is non-trivial, since normal gene expression ranges over several orders of magnitude. (For example, certain transcription factors function at low concentrations).

We compared the different CNN architectures from [Table 1](#) in terms of the numbers of significant gene-feature pairs (as well as unique genes, features and respectively tissues involved in these relations) using fixed correlation- ( $R_T = 0.8$ ) and  $\log_2$  expression thresholds ( $E_T = 10$ ). More precisely, the expression threshold  $E_T$  constrains the gene expression value  $E$  as follows:  $\log_2(1+E) \leq E_T$ .

We also assessed the significance of gene-feature correlations using permutation tests (with  $N = 1,000$  permutations).

Since one may expect the final layers of the CNN to be better correlated with the training classes (i.e. the tissue types), we studied the *layer* distribution of features with significant gene correlations.

We also studied the reproducibility of gene-feature correlations w.r.t. the dataset used. More precisely, we determined the Pearson correlation coefficient between the (Fisher transformed) gene-feature correlations computed w.r.t. the validation and respectively test datasets.

## Visualization of histological features

The issues of *interpretability* and *explainability* of neural network models is essential in many applications. Deep learning based histological image classifiers can be trained to achieve impressive performance in days, while a human histopathologist needs years, even decades, to achieve similar performance. However, despite their high classification accuracies, neural networks remain opaque about the precise way in which they manage to achieve these classifications. Of course, this is done based on the visual characteristics automatically inferred by the network, but the details of this process cannot be easily explained to a human, who might want to check not just the end result, but also the reasoning behind it. This is especially important in clinical applications, where *explainable* reasoning is critical for the adoption of the technology by clinicians, who need explanations to integrate the system's findings in their global assessment of the patient.

In our context, it would be very useful to obtain a better understanding of the visual features found to be correlated to gene expression, since these automatically inferred visual features make up the phenotype of interest. As opposed to features detected by human experts which are much more subjective and much harder to quantify, these automatically derived features have a precise mathematical definition that allows a precise quantification. Their visualization would also enable a comparison with expert domain knowledge.

In contrast to fully connected network models, convolutional networks tend to be easier to interpret and visualize. We have found two types of visualizations to be particularly useful in our application.

1. The first looks for elements (pixels) of the original image that affect a given feature most. These can be determined using backpropagation of the feature of interest w.r.t. a given input image. In particular, we employ *guided backpropagation* [37], which tends to produce better visualizations by zeroing negative gradients during the standard backpropagation process through ReLU units. Such visualizations, which we denote as  $gBP(f, X) = \nabla_x^+ f$  depend on the given input image  $X$  and were constructed for images from the *test dataset*, which is completely independent from the datasets used for model construction and respectively selection (training and respectively validation datasets).
2. The second visualization is independent of a specific input and amounts to determining a *synthetic input image* that maximizes the given feature  $\arg \max_x f(X)$  [38].

Note that unfortunately, guided backpropagation visualization cannot be applied exhaustively because it would involve an unmanageably large number of feature-histological image ( $f, X$ ) pairs. To deal with this problem, we developed an algorithm for selecting a small number of representative histological image tiles to be visualized with guided backpropagation (see [S1 File](#)). The problem is also addressed by using the second visualization method mentioned above, which generates a single synthetic image per feature.

Guided backpropagation visualizations of select features are compared with synthetic images and immunohistochemistry stains for the genes found correlated with the selected feature.

Since the network features were optimized to aid in tissue discrimination, while many genes are also tissue-specific, it may be possible that certain gene-feature correlations are indirect via the tissue variable. To assess the prevalence of such indirect gene( $g$ )-tissue( $t$ )-feature( $f$ ) correlations, we performed conditional independence tests using partial correlations of the form  $r(g, f | t)$  with a significance threshold  $p = 0.01$ .

## Results

First, we developed a classifier of histological images that is able to discriminate between 39 tissues of interest. A convolutional neural network enables learning visual representations that can be used as visual features in the subsequent stages of our analysis.

The resulting visual features are then quantified on an independent test dataset and correlated with paired gene expression data from the same subjects.

Finally, the features that are highly correlated to genes are visualized for better interpretability.

The following sections describe the results of these analysis steps in more detail.

### Deep learning tissue classifiers achieve high accuracies

We trained and validated several different convolutional neural network architectures (from Table 1) on the GTEx image tile dataset. Table 2 shows the accuracies of the classifiers on both *validation* and *test* datasets, after training for 90 epochs on the *training* dataset. Note that the test dataset is completely independent, while the validation dataset has been used for model selection.

As expected, AlexNet turned out to be the worst classifier among the ones tested. The various VGG architectures produced comparable results, regardless of their size. This may be due to the smaller number of classes (39) compared to ImageNet (1,000). On the other hand, batch normalization leads to improved classification accuracies, but also to representations that show much poorer correlation with gene expression data (see next section).

The modified architectures VGG16\_1FC and VGG16\_avg1FC perform slightly better than the original VGG16.

**Table 2. Classification accuracies for image tiles.**

Network	accuracy (tiles)	
	<i>validation</i>	<i>test</i>
AlexNet	77.02%	74.76%
VGG11	80.18%	77.49%
VGG13	80.77%	78.19%
VGG16	80.35%	78.00%
VGG19	79.61%	76.93%
VGG11_bn	83.24%	80.74%
VGG13_bn	83.94%	81.56%
VGG16_bn	84.24%	81.46%
VGG19_bn	84.44%	81.54%
ResNet34	82.51%	80.09%
Inception_v3	81.13%	78.99%
VGG16_1FC	82.66%	79.82%
VGG16_avg1FC	83.79%	80.92%

<https://doi.org/10.1371/journal.pone.0242858.t002>

**Table 3. Classification accuracies for whole slides.**

Network	accuracy (slides)	
	validation	test
AlexNet	91.82%	90.42%
VGG11	93.33%	92.22%
VGG13	93.94%	92.81%
VGG16	92.73%	93.71%
VGG19	93.33%	93.11%
VGG11_bn	93.94%	92.81%
VGG13_bn	93.03%	92.51%
VGG16_bn	93.64%	92.22%
VGG19_bn	93.94%	92.81%
ResNet34	93.33%	92.51%
Inception_v3	92.12%	92.22%
VGG16_1FC	93.33%	93.11%
VGG16_avg1FC	93.64%	94.01%

<https://doi.org/10.1371/journal.pone.0242858.t003>

Overall, accuracies of 80–82% in identifying the correct tissue based on a single image tile are significant, but still not comparable to human performance. This is because single image tiles allow only a very limited field of view of the tissue.

However, as the images contain multiple tiles, we constructed a *tissue classifier for whole slide images (WSI)* by applying the tile classifier on all tiles of the given WSI and aggregating the corresponding predictions using the majority vote. Table 3 show the accuracies of the whole slide image classifiers obtained using the convolutional neural network models from Table 1.

Note that all architectures, except AlexNet, lead to similar WSI classification accuracies (in the range 92–94%), with the best around 93–94%. (See S4 Table for the associated confusion matrix.) Most of the errors concern confusions between quite similar tissues, such as ‘Skin—Sun Exposed’ versus ‘Skin—Not Sun Exposed’. Merging the following groups of histologically similar tissues (Artery—Aorta, Artery—Coronary, Artery—Tibial), (Colon—Sigmoid, Colon—Transverse), (Esophagus—Gastroesophageal Junction, Esophagus—Muscularis), (Heart—Atrial Appendage, Heart—Left Ventricle), (Skin—Not Sun Exposed, Skin—Sun Exposed) leads to a WSI classification accuracy for 33 tissues of 97.9% in the case of VGG16.

### Histological features inferred by VGG architectures correlate well with gene expression data

We calculated the correlations between the visual features of convolutional networks trained on histopathological images and paired gene expression data. We found numerous genes whose expression is correlated with many visual histological features.

Table 4 shows the numbers of significant gene-feature pairs (as well as unique genes, features and respectively tissues involved in these relations) for the various CNN architectures using fixed correlation- ( $R_T = 0.8$ ) and  $\log_2$  expression thresholds  $E_T = 10$  (where  $\log_2(1+E) \leq E_T$ ). Histograms of features, genes and their correlations are shown in S1 Fig.

Note that although batch normalization slightly improved *tile* classification accuracies (but not *WSI* classification accuracies), it also produced visual representations (features) with dramatically poorer correlation with gene expression data. For example, for VGG16\_bn, we could identify only 51 significant gene-feature pairs, involving just 20 unique genes and 11 unique

**Table 4. Numbers of significantly correlated gene-feature pairs for various network architectures.** Fixed correlation- and  $\log_2$  gene expression thresholds are used:  $R_T = 0.8$ ,  $E_T = 10$ . Numbers of unique genes, features and tissues involved are also shown (test dataset).

Network	gene-feature pairs	unique genes	unique features	unique tissues
AlexNet	424	55	112	11
VGG11	1,175	70	204	12
VGG13	2,149	70	307	13
VGG16	2,176	74	346	13
VGG19	1,227	71	321	14
VGG11_bn	34	19	7	3
VGG13_bn	59	22	17	6
VGG16_bn	51	20	11	4
ResNet34	3	2	3	2
Inception_v3	23	12	7	2
VGG16_1FC	2,714	69	363	13
VGG16_avg1FC	3,114	83	448	15

<https://doi.org/10.1371/journal.pone.0242858.t004>

features, compared to 2,176 gene-feature pairs implicating 74 unique genes and 346 unique features for VGG16 without batch normalization.

Unless stated otherwise, we concentrate in the following on visual representations obtained using the VGG16 architecture. (The results for AlexNet and the other VGG architectures without batch normalization are similar).

VGG16 was selected not just based on it achieving one of the highest WSI classification accuracies on the test dataset (93.71%, Table 3), but also since its last convolutional layer is not directly connected to the output classes (as in the case of VGG16\_1FC and VGG16\_avg1FC). This is assumed to enable a higher degree of independence of the final convolutional layers from the output classes, rendering them more data oriented.

Table 5 shows the numbers of significantly correlated gene-feature pairs for various correlation- and  $\log_2$  gene expression thresholds (for the VGG16 network).

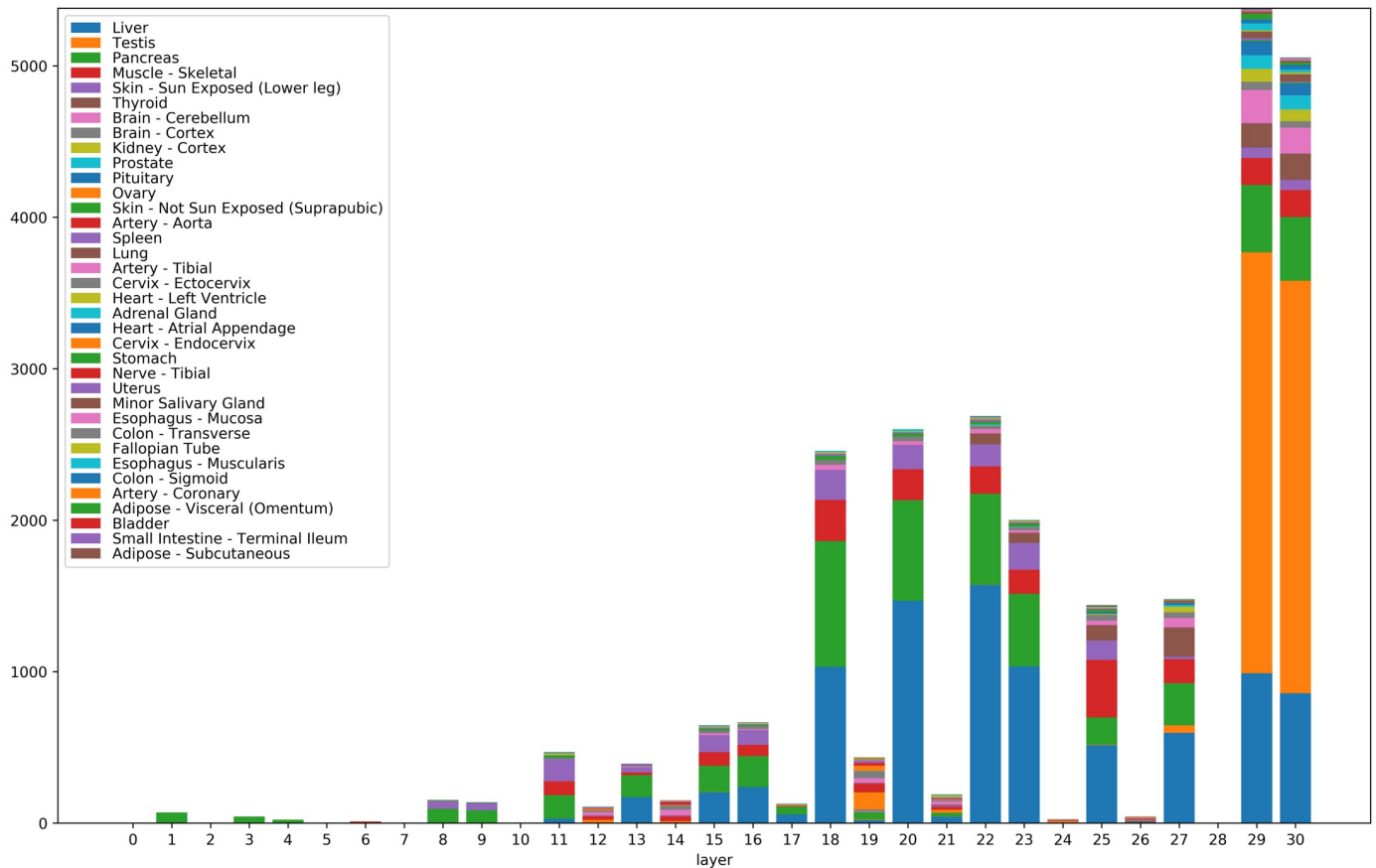
Using a correlation threshold  $R_T = 0.7$  and a  $\log_2$  expression threshold  $E_T = 7$ , we obtained 27,671 significant correlated gene-feature pairs involving 995 unique genes and 1,947 features (for the VGG16 architecture and the test dataset). We call these gene-feature pairs significant, because the p-values computed using permutation tests (with  $N = 1,000$  permutations) were all  $p < 10^{-3}$ .

Fig 3 shows the numbers of significantly correlated genes for the 31 layers of VGG16 (numbered 0 to 30). All features (channels) of a given layer were aggregated, since VGG16 has a too

**Table 5. Numbers of significantly correlated gene-feature pairs for various correlation- and  $\log_2$  gene expression thresholds.** Numbers of unique genes, features and tissues involved are also shown (VGG16 network, test dataset).

Correlation threshold	$\log_2$ expression threshold	gene-feature pairs	unique genes	unique features	unique tissues
0.8	10	2,176	74	346	13
0.75	10	4,308	115	647	22
0.7	10	7,984	213	1,146	28
0.75	8	9,624	312	926	26
0.8	7	8,055	365	576	18
0.75	7	15,062	535	1,046	31
0.7	7	27,671	995	1,947	36

<https://doi.org/10.1371/journal.pone.0242858.t005>



**Fig 3. Numbers of significantly correlated genes for the 31 layers of VGG16.** Significantly correlated genes were aggregated for all features (channels) belonging to a given layer (i.e. for all features of the form [layer]\_[channel]). The correlated genes were broken down according to their tissues of maximal expression. Tissues are color-coded. Colors of the figure bars scanned bottom-up correspond to colors in the legend read top to bottom.

<https://doi.org/10.1371/journal.pone.0242858.g003>

large number of features (9,920 in total). (See also [S2A Fig](#) for numbers of correlated genes for *individual* features.) The correlated genes were broken down according to their tissues of maximal expression. Note that the highest numbers of correlated genes correspond to layers with non-negative output (ReLU or MaxPool2d –cf. layer numbers in the VGG16 entry in [Table 1](#)).

We also remark that the features with significant gene correlations do not necessarily belong to the final layers of the convolutional neural network, even though one may have expected the final layers to be better correlated with the training classes (i.e. the tissue types). Also note that certain tissues with a simpler morphology, such as liver (blue in [Fig 3](#)) show correlated genes with features across many layers of the network, from lower levels to the highest levels. Other morphologically more complex tissues, such as testis (orange in [Fig 3](#)) express genes correlated almost exclusively with the final layers 29 and 30 (which are able to detect such complex histological patterns). See also examples of histological images for these tissues in [Fig 6](#) (column 1, rows 1 and 2).

For an increased *specificity* and in order to reduce the number of gene-feature pairs subject to human evaluation, we initially performed an analysis with higher thresholds  $R_T = 0.8$  and  $E_T = 10$  (resulting in 2,176 gene-feature pairs, involving 74 genes and 346 features).

A second analysis with lower thresholds ( $R_T = 0.75$ ,  $E_T = 7$ ) was performed with the aim of improving the *sensitivity* of the initial analysis (15,062 gene-feature pairs, involving 535 genes and 1,046 features—S2 Table).

Since gene-feature *correlations* do not necessarily indicate *causation* (i.e. the genes involved are not necessarily causal factors determining the observed visual feature), we tried to select a subset of genes, based on known annotations that are more likely to play a causal role in shaping the histological morphology of the tissues. In the following, we report on the subset of genes with Gene Ontology ‘*developmental process*’ or ‘*transcription regulator activity*’ annotations (either direct, or inherited) [39, 40] (Table 6 and S3 Table). While some developmental genes are turned off after completion of the developmental program, many continue to be expressed in adult tissue, to coordinate and maintain its proper structure and function.

Many of the genes significantly correlated with histological features have crucial roles in the development and maintenance of the corresponding tissues. The transcription factors ZIC1, ZIC2, ZIC4, NEUROD1 and NEUROD2 are known to be involved in brain and more specifically cerebellar development [41, 42], NKX2-5 and BMP10 are implicated in heart development [43, 44], POU1F1 regulates expression of several genes involved in pituitary development and hormone expression [45], NKX2-1, PAX8 and FOXE1 are key thyroid transcription factors, with a fundamental role in the proper formation of the thyroid gland and in maintaining its functional differentiated state in the adult organism [46], etc. (Table 6, S3 Table). While a correlational approach such as the present one cannot replace perturbational tests of causality involved in shaping tissue morphology, it can provide appropriate candidates for such tests.

We also studied the reproducibility of gene-feature correlations w.r.t. the dataset used. More precisely, we determined the Pearson correlation coefficient between the (Fisher transformed) gene-feature correlations computed w.r.t. the validation and respectively test datasets (S3 Fig):

$$\begin{aligned} r(z(r(\log_2(1+g), f))|_{val}, z(r(\log_2(1+g), f))|_{test}) &= 0.9233 \\ r(r(\log_2(1+g), f)|_{val}, r(\log_2(1+g), f)|_{test}) &= 0.9205 \end{aligned}$$

where  $z(r)$  is the Fisher transform of correlation  $r$ . Gene-feature correlations thus show a good reproducibility across datasets.

## Visualization of histological features

To investigate the biological relevance of the discovered gene-feature correlations, we analyzed in detail different methods of visualizing the features inferred by convolutional networks. We applied two different visualization methods. The first based on ‘*guided backpropagation*’ determines the regions of a histopathological image that affect the feature of interest most. However, this visualization method cannot be applied exhaustively because it would involve an unmanageably large number of feature-histological image pairs. To deal with this problem, we developed an algorithm for selecting a small number of representative histological image tiles to be visualized with guided backpropagation (S1 File). We also considered a second feature visualization method that is independent of any input image. The method involves generating a synthetic input image that optimizes the network response to the visual feature of interest. Such synthetic images look similar to real histological images that strongly activate the visual feature of interest.

Figs 4–6 show such visualizations of select histological features. Note that *guided backpropagation* (column 2) emphasizes important structural features of the original histological images (column 1). For example, performing guided backpropagation (gBP) of feature 29\_499 on a

Table 6. Developmental and transcription regulation genes correlated with visual features.

Gene	Gene name	Tissue with highest expression	Highest median tissue expression ( $\log_2$ )	Best correlated feature	r
FABP4	Fatty acid-binding protein, adipocyte	Adipose—Visceral (Omentum)	12.6	1_14	0.767
CYP17A1	Steroid 17-alpha-hydroxylase/17,20 lyase	Adrenal Gland	12.4	20_467	0.805
HSD3B2	3 beta-hydroxysteroid dehydrogenase/Delta 5—>4-isomerase type 2	Adrenal Gland	11.2	29_123	0.791
STAR	Steroidogenic acute regulatory protein, mitochondrial	Adrenal Gland	12.5	29_132	0.770
TPM4	Tropomyosin alpha-4 chain	Artery—Aorta	9.1	19_468	0.767
S100A6	Protein S100-A6	Artery—Aorta	11.6	19_162	0.760
YAP1	Transcriptional coactivator YAP1	Artery—Aorta	7.6	19_162	0.758
MARVELD1	MARVEL domain-containing protein 1	Artery—Tibial	7.8	19_162	0.764
GNG12	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	Artery—Tibial	7.1	12_209	0.755
ZIC4	Zinc finger protein ZIC 4	Brain—Cerebellum	7.0	29_447	0.922
NEUROD2	Neurogenic differentiation factor 2	Brain—Cerebellum	7.2	29_463	0.915
NEUROD1	Neurogenic differentiation factor 1	Brain—Cerebellum	7.8	29_479	0.873
CRTAM	Cytotoxic and regulatory T-cell molecule	Brain—Cerebellum	7.2	29_479	0.868
ZIC2	Zinc finger protein ZIC 2	Brain—Cerebellum	7.9	29_463	0.854
SLC12A5	Solute carrier family 12 member 5	Brain—Cerebellum	7.5	27_140	0.850
ZIC1	Zinc finger protein ZIC 1	Brain—Cerebellum	8.1	29_479	0.844
ELAVL3	ELAV-like protein 3	Brain—Cerebellum	8.0	29_463	0.834
PVALB	Parvalbumin alpha	Brain—Cerebellum	9.1	29_463	0.781
HPCAL4	Hippocalcin-like protein 4	Brain—Cerebellum	7.5	29_463	0.781
CPLX2	Complexin-2	Brain—Cerebellum	8.7	27_140	0.762
SPOCK2	Testican-2	Brain—Cerebellum	8.6	27_148	0.761
SLC1A2	Excitatory amino acid transporter 2	Brain—Cortex	8.7	18_8	0.844
GRIN1	Glutamate receptor ionotropic, NMDA 1	Brain—Cortex	7.9	18_8	0.812
HPCA	Neuron-specific calcium-binding protein hippocalcin	Brain—Cortex	7.5	29_485	0.811
PACSIN1	Protein kinase C and casein kinase substrate in neurons protein 1	Brain—Cortex	8.5	29_463	0.810
SLC17A7	Vesicular glutamate transporter 1	Brain—Cortex	9.3	18_8	0.799
CAMK2A	Calcium/calmodulin-dependent protein kinase type II subunit alpha	Brain—Cortex	9.2	23_417	0.786
DDN	Dendrin	Brain—Cortex	7.9	25_478	0.775
CEND1	Cell cycle exit and neuronal differentiation protein 1	Brain—Cortex	8.6	29_485	0.755
KIF5A	Kinesin heavy chain isoform 5A	Brain—Cortex	10.0	29_485	0.754
NBL1	Neuroblastoma suppressor of tumorigenicity 1	Cervix—Ectocervix	9.6	19_162	0.758
CRTAP	Cartilage-associated protein	Cervix—Ectocervix	8.1	12_223	0.756
HMGB1	High mobility group protein B1	Cervix—Ectocervix	7.4	19_468	0.753
PIAS3	E3 SUMO-protein ligase PIAS3	Cervix—Endocervix	7.0	19_468	0.764
SPIN1	Spindlin-1	Cervix—Endocervix	7.2	19_333	0.752
PLXNB2	Plexin-B2	Cervix—Endocervix	8.0	12_164	0.750
NKX2-5	Homeobox protein Nkx-2.5	Heart—Atrial Appendage	7.0	30_214	0.903
BMP10	Bone morphogenetic protein 10	Heart—Atrial Appendage	8.3	29_313	0.855
NPPA	Natriuretic peptides A	Heart—Atrial Appendage	14.9	29_481	0.777
NMRK2	Nicotinamide riboside kinase 2	Heart—Atrial Appendage	9.8	30_72	0.751
MYBPC3	Myosin-binding protein C, cardiac-type	Heart—Left Ventricle	10.8	23_270	0.783
TNNI3	Troponin I, cardiac muscle	Heart—Left Ventricle	12.0	25_28	0.780
TNNT2	Troponin T, cardiac muscle	Heart—Left Ventricle	11.6	25_31	0.759

(Continued)



Table 6. (Continued)

Gene	Gene name	Tissue with highest expression	Highest median tissue expression ( $\log_2$ )	Best correlated feature	r
CSRP3	Cysteine and glycine-rich protein 3	Heart—Left Ventricle	9.4	30_72	0.755
AQP2	Aquaporin-2	Kidney—Cortex	7.6	29_280	0.890
UMOD	Uromodulin	Kidney—Cortex	7.4	27_274	0.874
PLG	Plasminogen	Liver	8.7	29_234	0.959
APOA5	Apolipoprotein A-V	Liver	8.8	29_192	0.954
SERPINC1	Antithrombin-III	Liver	10.1	29_192	0.952
HRG	Histidine-rich glycoprotein	Liver	9.2	29_234	0.952
F2	Prothrombin	Liver	9.4	29_192	0.949
AHSG	Alpha-2-HS-glycoprotein	Liver	10.7	29_192	0.946
APCS	Serum amyloid P-component	Liver	10.7	29_192	0.942
BAAT	Bile acid-CoA:amino acid N-acyltransferase	Liver	7.7	29_192	0.938
ANGPTL3	Angiopoietin-related protein 3	Liver	7.3	29_234	0.934
APOA2	Apolipoprotein A-II	Liver	12.2	29_192	0.931
CYP4A11	Cytochrome P450 4A11	Liver	8.3	29_234	0.924
CPB2	Carboxypeptidase B2	Liver	8.5	29_192	0.922
PROC	Vitamin K-dependent protein C	Liver	7.9	29_234	0.919
APOH	Beta-2-glycoprotein 1	Liver	11.7	29_234	0.919
FGB	Fibrinogen beta chain	Liver	12.6	29_192	0.896
FGL1	Fibrinogen-like protein 1	Liver	10.2	18_231	0.879
G6PC	Glucose-6-phosphatase	Liver	7.1	29_234	0.872
FGA	Fibrinogen alpha chain	Liver	12.2	30_192	0.858
ASGR2	Asialoglycoprotein receptor 2	Liver	8.8	30_192	0.854
CRP	C-reactive protein	Liver	12.6	29_192	0.852
VTN	Vitronectin	Liver	11.2	18_279	0.851
FGG	Fibrinogen gamma chain	Liver	12.2	30_192	0.829
IGFBP1	Insulin-like growth factor-binding protein 1	Liver	7.2	29_75	0.802
APOB	Apolipoprotein B-100	Liver	8.5	30_438	0.784
CREB3L3	Cyclic AMP-responsive element-binding protein 3-like protein 3	Liver	8.2	20_481	0.777
CPS1	Carbamoyl-phosphate synthase [ammonia], mitochondrial	Liver	8.6	29_192	0.775
ARG1	Arginase-1	Liver	9.0	20_194	0.768
SFTPB	Pulmonary surfactant-associated protein B	Lung	12.2	29_383	0.789
SCGB1A1	Uteroglobin	Lung	9.2	29_155	0.776
STATH	Statherin	Minor Salivary Gland	7.7	25_200	0.767
MYF6	Myogenic factor 6	Muscle—Skeletal	7.2	25_409	0.872
NEB	Nebulin	Muscle—Skeletal	9.9	25_409	0.851
XIRP2	Xin actin-binding repeat-containing protein 2	Muscle—Skeletal	8.0	18_102	0.828
RYR1	Ryanodine receptor 1	Muscle—Skeletal	8.6	29_362	0.827
TMOD4	Tropomodulin-4	Muscle—Skeletal	8.6	30_16	0.827
KLHL40	Kelch-like protein 40	Muscle—Skeletal	7.8	25_409	0.824
SMTNL1	Smoothelin-like protein 1	Muscle—Skeletal	7.2	30_362	0.820
TTN	Titin	Muscle—Skeletal	8.7	18_136	0.810
MYPN	Myopalladin	Muscle—Skeletal	7.3	30_72	0.805
LMOD3	Leiomodin-3	Muscle—Skeletal	7.4	18_136	0.802
MYLPF	Myosin regulatory light chain 2, skeletal muscle isoform	Muscle—Skeletal	10.9	30_161	0.798
LMOD2	Leiomodin-2	Muscle—Skeletal	8.9	30_72	0.798

(Continued)

Table 6. (Continued)

Gene	Gene name	Tissue with highest expression	Highest median tissue expression ( $\log_2$ )	Best correlated feature	r
NRAP	Nebulin-related-anchoring protein	Muscle—Skeletal	10.0	18_136	0.794
MYL2	Myosin regulatory light chain 2, ventricular/cardiac muscle isoform	Muscle—Skeletal	13.7	18_136	0.781
MYLK2	Myosin light chain kinase 2, skeletal/cardiac muscle	Muscle—Skeletal	7.7	30_161	0.777
TNNT1	Troponin T, slow skeletal muscle	Muscle—Skeletal	12.5	18_272	0.776
TNNI1	Troponin I, slow skeletal muscle	Muscle—Skeletal	10.0	30_161	0.773
MB	Myoglobin	Muscle—Skeletal	13.1	30_72	0.772
MYH7	Myosin-7	Muscle—Skeletal	12.4	18_136	0.770
KLHL41	Kelch-like protein 41	Muscle—Skeletal	11.6	20_171	0.766
ANP32B	Acidic leucine-rich nuclear phosphoprotein 32 family member B	Nerve—Tibial	8.2	19_468	0.775
CNTF	Ciliary neurotrophic factor	Nerve—Tibial	8.3	29_36	0.769
MXRA8	Matrix remodeling-associated protein 8	Nerve—Tibial	8.5	19_468	0.751
RBPJL	Recombining binding protein suppressor of hairless-like protein	Pancreas	9.5	30_467	0.935
INS	Insulin	Pancreas	10.8	18_158	0.913
CELA2A	Chymotrypsin-like elastase family member 2A	Pancreas	12.6	23_110	0.895
CEL	Bile salt-activated lipase	Pancreas	14.0	23_324	0.866
REG3G	Regenerating islet-derived protein 3-gamma	Pancreas	9.6	30_290	0.815
FSHB	Follitropin subunit beta	Pituitary	7.1	29_279	0.930
POU1F1	Pituitary-specific positive transcription factor 1	Pituitary	7.3	29_279	0.913
GHRHR	Growth hormone-releasing hormone receptor	Pituitary	8.9	29_279	0.900
TSHB	Thyrotropin subunit beta	Pituitary	9.2	29_436	0.869
LHB	Lutropin subunit beta	Pituitary	12.0	30_429	0.844
MYO15A	Unconventional myosin-XV	Pituitary	7.2	29_436	0.803
PRL	Prolactin	Pituitary	15.5	29_436	0.799
TGFBR3L	Transforming growth factor-beta receptor type 3-like protein	Pituitary	8.0	29_436	0.787
GH1	Somatotropin	Pituitary	16.8	30_429	0.763
COL22A1	Collagen alpha-1(XXII) chain	Pituitary	7.2	30_436	0.753
KLK3	Prostate-specific antigen	Prostate	12.8	29_458	0.900
KLK4	Kallikrein-4	Prostate	9.4	29_430	0.821
HOXB13	Homeobox protein Hox-B13	Prostate	7.3	30_458	0.801
KRT77	Keratin, type II cytoskeletal 1b	Skin—Not Sun Exposed	8.6	18_78	0.884
FLG2	Filaggrin-2	Skin—Sun Exposed	9.6	18_78	0.883
LCE2C	Late cornified envelope protein 2C	Skin—Sun Exposed	8.2	18_78	0.878
LCE1A	Late cornified envelope protein 1A	Skin—Sun Exposed	8.9	18_78	0.878
CDSN	Corneodesmosin	Skin—Sun Exposed	8.3	18_78	0.875
LCE2B	Late cornified envelope protein 2B	Skin—Sun Exposed	9.4	18_78	0.874
LCE1B	Late cornified envelope protein 1B	Skin—Sun Exposed	8.0	18_78	0.874
LCE1C	Late cornified envelope protein 1C	Skin—Sun Exposed	9.4	18_78	0.873
LCE6A	Late cornified envelope protein 6A	Skin—Sun Exposed	7.6	18_78	0.870
LCE1F	Late cornified envelope protein 1F	Skin—Sun Exposed	7.6	18_78	0.868
LCE5A	Late cornified envelope protein 5A	Skin—Sun Exposed	7.3	18_78	0.864
LCE2D	Late cornified envelope protein 2D	Skin—Sun Exposed	7.5	18_78	0.863
LCE2A	Late cornified envelope protein 2A	Skin—Sun Exposed	7.2	18_78	0.856
DSC1	Desmocollin-1	Skin—Sun Exposed	8.2	18_78	0.849

(Continued)

Table 6. (Continued)

Gene	Gene name	Tissue with highest expression	Highest median tissue expression ( $\log_2$ )	Best correlated feature	r
KRT2	Keratin, type II cytoskeletal 2 epidermal	Skin—Sun Exposed	12.8	18_78	0.840
FLG	Filaggrin	Skin—Sun Exposed	9.0	18_78	0.835
SERPINB7	Serpin B7	Skin—Sun Exposed	7.4	18_78	0.830
KRT10	Keratin, type I cytoskeletal 10	Skin—Sun Exposed	14.6	18_78	0.826
CASP14	Caspase-14	Skin—Sun Exposed	9.4	18_78	0.820
PSAPL1	Proactivator polypeptide-like 1	Skin—Sun Exposed	7.4	22_265	0.812
CST6	Cystatin-M	Skin—Sun Exposed	9.6	20_329	0.773
ASPRV1	Retroviral-like aspartic protease 1	Skin—Sun Exposed	9.6	18_78	0.772
ALOX12B	Arachidonate 12-lipoxygenase, 12R-type	Skin—Sun Exposed	7.4	25_227	0.757
NR5A1	Steroidogenic factor 1	Spleen; Adrenal Gland	8.0	27_423	0.780
CD19	B-lymphocyte antigen CD19	Spleen	7.9	27_389	0.759
KCNE2	Potassium voltage-gated channel subfamily E member 2	Stomach	8.3	29_287	0.782
DDX4	Probable ATP-dependent RNA helicase DDX4	Testis	7.5	29_39	0.951
DMRTB1	Doublesex- and mab-3-related transcription factor B1	Testis	7.1	29_39	0.949
SHCBP1L	Testicular spindle-associated protein SHCBP1L	Testis	7.7	29_39	0.945
CALR3	Calreticulin-3	Testis	7.1	29_246	0.945
ZBP2	Zona pellucida-binding protein 2	Testis	7.2	29_39	0.941
SPEM1	Spermatid maturation protein 1	Testis	7.5	29_246	0.939
ACSBG2	Long-chain-fatty-acid—CoA ligase ACSBG2	Testis	8.0	29_39	0.938
SPATA19	Spermatogenesis-associated protein 19, mitochondrial	Testis	8.0	29_168	0.937
RNF151	RING finger protein 151	Testis	8.5	29_39	0.937
FSCN3	Fascin-3	Testis	7.2	29_246	0.933
TCP11	T-complex protein 11 homolog	Testis	8.7	29_258	0.930
ODF1	Outer dense fiber protein 1	Testis	9.6	29_39	0.929
IQCF1	IQ domain-containing protein F1	Testis	7.8	29_246	0.926
PRM3	Protamine-3	Testis	7.0	29_246	0.926
TDRG1	Testis development-related protein 1	Testis	7.0	29_39	0.923
SYCP3	Synaptonemal complex protein 3	Testis	7.2	29_39	0.917
CABS1	Calcium-binding and spermatid-specific protein 1	Testis	7.9	29_73	0.915
DKKL1	Dickkopf-like protein 1	Testis	9.3	29_39	0.914
RPL10L	60S ribosomal protein L10-like	Testis	7.5	29_246	0.908
SYCE3	Synaptonemal complex central element protein 3	Testis	7.9	29_39	0.906
CAPZA3	F-actin-capping protein subunit alpha-3	Testis	8.7	29_258	0.906
GTSF1	Gametocyte-specific factor 1	Testis	7.0	29_246	0.905
CCDC42	Coiled-coil domain-containing protein 42	Testis	7.0	29_73	0.903
TXNDC2	Thioredoxin domain-containing protein 2	Testis	7.1	29_246	0.898
TPPP2	Tubulin polymerization-promoting protein family member 2	Testis	8.2	29_73	0.888
AKAP3	A-kinase anchor protein 3	Testis	7.3	29_39	0.880
TNP1	Spermatid nuclear transition protein 1	Testis	13.1	29_39	0.874
TSSK6	Testis-specific serine/threonine-protein kinase 6	Testis	8.2	29_246	0.869
MAEL	Protein maelstrom homolog	Testis	7.3	29_39	0.864
TCP10L	T-complex protein 10A homolog 1	Testis	8.1	29_39	0.864
CCIN	Calicin	Testis	7.3	29_73	0.862
PRAME	Melanoma antigen preferentially expressed in tumors	Testis	7.3	30_39	0.860
PRM1	Sperm protamine P1	Testis	14.3	29_73	0.850

(Continued)

Table 6. (Continued)

Gene	Gene name	Tissue with highest expression	Highest median tissue expression ( $\log_2$ )	Best correlated feature	r
PRM2	Protamine-2	Testis	14.3	29_73	0.849
GGN	Gametogenetin	Testis	7.5	29_39	0.845
ROPN1L	Ropporin-1-like protein	Testis	9.0	29_39	0.844
CABYR	Calcium-binding tyrosine phosphorylation-regulated protein	Testis	8.0	29_39	0.840
INSL3	Insulin-like 3	Testis	8.9	29_39	0.802
ACRBP	Acrosin-binding protein	Testis	9.3	30_246	0.798
PHOSPHO1	Phosphoethanolamine/phosphocholine phosphatase	Testis	7.3	29_246	0.782
SPATA24	Spermatogenesis-associated protein 24	Testis	7.5	29_73	0.778
SPINK2	Serine protease inhibitor Kazal-type 2	Testis	9.2	29_73	0.773
PCSK4	Proprotein convertase subtilisin/kexin type 4	Testis	8.2	30_39	0.773
PRSS21	Testisin	Testis	7.0	30_168	0.773
NKX2-1	Homeobox protein Nkx-2.1	Thyroid	8.5	29_202	0.870
TG	Thyroglobulin	Thyroid	12.3	30_93	0.869
PAX8	Paired box protein Pax-8	Thyroid	10.4	29_499	0.802
FOXE1	Forkhead box protein E1	Thyroid	7.5	27_376	0.778
TRIP6	Thyroid receptor-interacting protein 6	Uterus	7.6	12_209	0.762

Genes with Gene Ontology ‘developmental process’ or ‘transcription regulator activity’ annotations are grouped w.r.t. tissues and ordered by correlation—for each gene we only show the best correlated feature, in the form [layer]\_[channel] (for the architecture VGG16 and the test dataset; correlation threshold = 0.75,  $\log_2$  gene expression threshold = 7). Genes with ‘transcription regulator activity’ are shown in red.

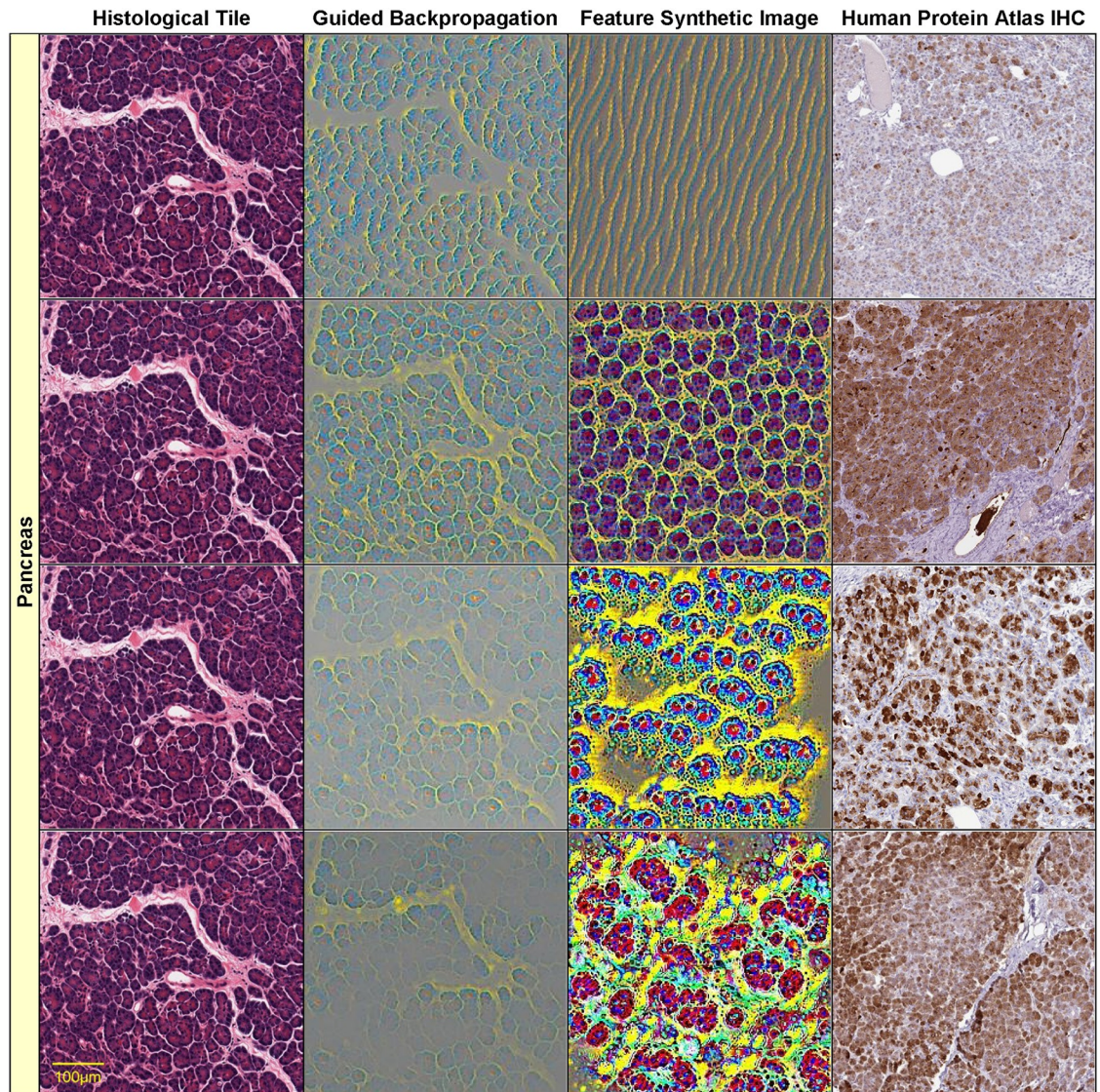
<https://doi.org/10.1371/journal.pone.0242858.t006>

thyroid sample produces patterns that clearly overlap known histological structures of thyroid tissues, namely follicular cells in blue and follicle boundaries in yellow (row 3, column 2 of Fig 5). More precisely, the blue “dots” in the gBP image precisely overlap the follicular cells, while the yellow patterns correspond to boundaries of follicles (please compare with the original image from column 1).

On the other hand, column 3 displays a synthetically generated image that maximizes the corresponding feature (29\_499). The image was generated by optimization of the feature using incremental changes to an initially random input image. Still, the synthetic image shows a remarkable resemblance to the histological structure of thyroid tissue (except for color differences due to color normalization).

We also show in column 4 the expression of PAX8 in an immunohistochemistry (IHC) stain of thyroid tissue from the Human Protein Atlas [47]. Note that PAX8 was identified as a correlate of the above-mentioned feature (29\_499) with correlation  $r = 0.802$  (Table 6) and is specifically expressed in the follicular cells (corresponding to the blue “dots” in the gBP image from column 3). PAX8 encodes a member of the paired box family of transcription factors and is known to be involved in thyroid follicular cell development and expression of thyroid-specific genes. Although PAX8 is expressed during embryonic development and is subsequently turned off in most adult tissues, its expression is maintained in the thyroid [46].

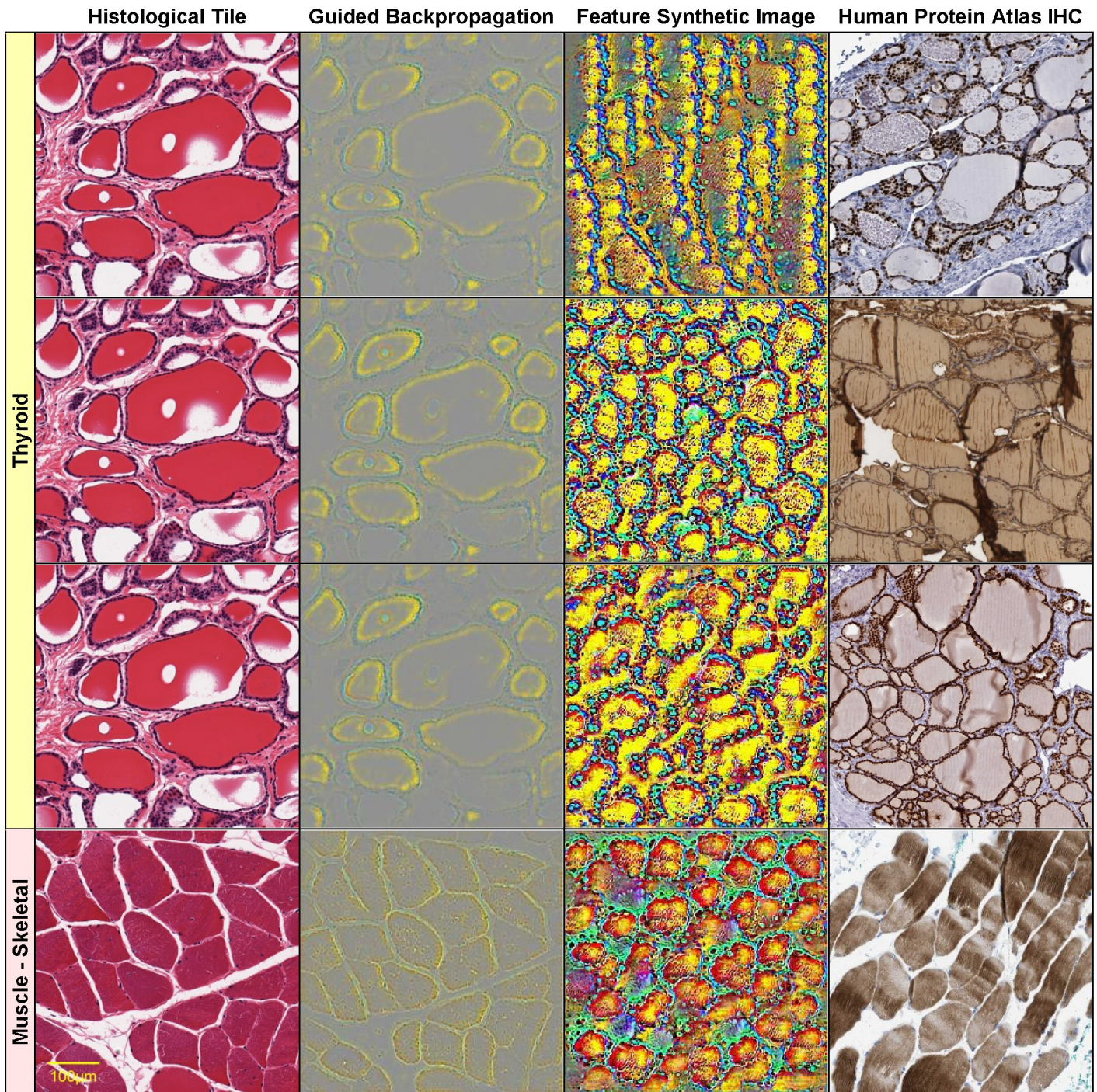
Note that the features with significant gene correlations do not necessarily belong to the final layers of the convolutional neural network (as one may expect the final layers to be better correlated with the training classes, i.e. the tissue types). Fig 4 illustrates this by showing four features at various layers (8, 18, 23 and 30) of the VGG16 architecture (which has in total 31 layers, numbered 0 to 30) for a pancreas slide. Lower layers tend to capture simpler visual features, such as blue/yellow edges in the synthetic image of feature 8\_55 (row 1, column 3 of



**Fig 4. Visualizations of select histological features.** The following features (of the form [layer]\_[channel]) found correlated with specific genes are visualized on each row: row 1: 8\_55—CTRC ( $r = 0.85$ ), row 2: 18\_13—CUZD1 ( $r = 0.866$ ), row 3: 23\_324—CELA3B ( $r = 0.868$ ), row 4: 30\_467—AMY2A ( $r = 0.928$ ). Original image (column 1), guided backpropagation of the feature on the original image (column 2), synthetic image of the feature (column 3), immunohistochemistry image for the corresponding gene from the Human Protein Atlas (column 4). All visualizations are for *pancreas* sample tile GTEX-11NSD-0526\_32\_5.

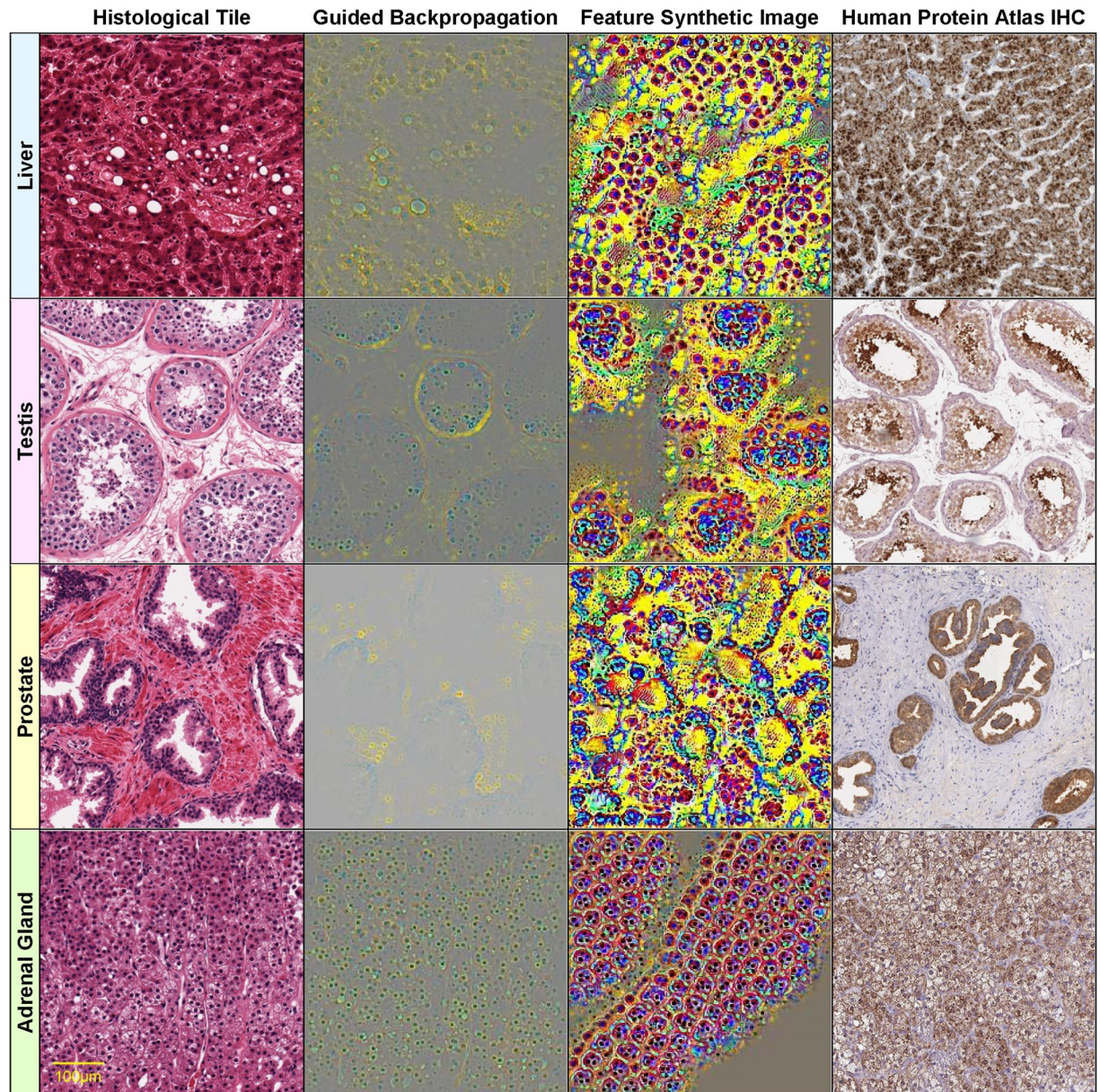
<https://doi.org/10.1371/journal.pone.0242858.g004>

**Fig 4).** In the corresponding guided backpropagation image (column 2), these edges are sufficient to emphasize the yellow boundaries between the blue pancreatic acinar cells. Higher layer features detect more complex histological features, such as round “cell-like” structures that tile the entire image (feature 18\_13, row 2, column 3 of Fig 4), or more complex blue cell-like structures (including red nuclei) separated by yellow interstitia (feature 23\_324, row 3, column 3 of Fig 4). The highest level feature, 30\_467 (row 4, column 3 of Fig 4) seems to detect even more complex patterns involving mostly yellow connective tissue and surrounding red/blue “cells”.



**Fig 5. Visualizations of select histological features.** The following features (of the form [layer]\_[channel]) found correlated with specific genes are visualized on each row: row 1: 22\_405 NKX2-1 ( $r = 0.726$ ) Thyroid GTEX-11NSD-0126\_31\_16, row 2: 27\_343 TG ( $r = 0.801$ ) Thyroid GTEX-11NSD-0126\_31\_16, row 3: 29\_499 PAX8 ( $r = 0.802$ ) Thyroid GTEX-11NSD-0126\_31\_16, row 4: 25\_260 NEB ( $r = 0.831$ ) Muscle—Skeletal GTEX-145ME-2026\_39\_19. Original image (column 1), guided backpropagation of the feature on the original image (column 2), synthetic image of the feature (column 3), immunohistochemistry image for the corresponding gene from the Human Protein Atlas (column 4).

<https://doi.org/10.1371/journal.pone.0242858.g005>



**Fig 6. Visualizations of select histological features.** The following features (of the form [layer]\_[channel]) found correlated with specific genes are visualized on each row: row 1: 29\_234 AGXT ( $r = 0.939$ ) Liver GTEX-Q2AG-1126\_9\_7, row 2: 29\_118 CALR3 ( $r = 0.829$ ) Testis GTEX-11NSD-1026\_5\_27, row 3: 30\_244 KLK3 ( $r = 0.828$ ) Prostate GTEX-V955-1826\_8\_13, row 4: 20\_137 CYP11B1 ( $r = 0.850$ ) Adrenal Gland GTEX-QLQW-0226\_25\_9. Original image (column 1), guided backpropagation of the feature on the original image (column 2), synthetic image of the feature (column 3), immunohistochemistry image for the corresponding gene from the Human Protein Atlas (column 4).

<https://doi.org/10.1371/journal.pone.0242858.g006>

Other examples of visualizations of features from different layers are shown in rows 1–3 of Fig 5 for a thyroid slide. The lower level feature 22\_405 detects boundaries between blue follicular cells and yellow follicle lumens, while the higher level features 27\_343 and 29\_499 are activated by more complex, round shapes of blue follicular cells surrounding yellow-red follicle interiors. All of these features are significantly correlated with thyroglobulin TG expression  $r(TG,22_405) = 0.805$ ,  $r(TG,27_343) = 0.801$ ,  $r(TG,29_499) = 0.781$ , but also with other key

thyroidal genes, such as TSHR, NKX2-1, PAX8 and FOXE1:  $r(\text{TSHR}, 22\_405) = 0.828$ ,  $r(\text{TSHR}, 27\_343) = 0.819$ ,  $r(\text{TSHR}, 29\_499) = 0.824$ ,  $r(\text{FOXE1}, 22\_405) = 0.744$ ,  $r(\text{NKX2-1}, 22\_405) = 0.726$ ,  $r(\text{NKX2-1}, 29\_499) = 0.711$ ,  $r(\text{PAX8}, 29\_499) = 0.802$ . Note that NKX2-1, FOXE1 and PAX8 are three of the four key thyroid transcription factors, with a fundamental role in the proper formation of the thyroid gland and in maintaining its functional differentiated state in the adult organism [46]. The fourth thyroid transcription factor, HHEX, is missing from our analysis due to its expression slightly below the  $\log_2$  expression threshold used (the median HHEX expression is 6.84, just below the threshold 7). For illustration purposes, we show, in column 4 of Fig 5, IHC stains for distinct thyroid genes (NKX2-1 for feature 22\_405, TG for 27\_343 and respectively PAX8 for 29\_499). Note that all of these IHC images are remarkably similar to the corresponding guided backpropagation (gBP) images from column 2 and the synthetic images from column 3.

It is remarkable that spatial expression patterns of genes, as assessed by IHC, are frequently very similar to the gBP images of their correlated features. Still, not all significantly correlated gene-feature pairs display such a good similarity between the gBP image of the feature and the spatial expression pattern of the gene. This is due to the complete lack of spatial specificity of the RNA-seq gene expression data, as well as to the rather coarse-grained classes used for training the visual classifier—just tissue labels, without spatial annotations of tissue substructure. This is the case of tissue-specific genes, which may display *indirect gene-tissue-feature correlations* with distinct spatial specificities of the gene-tissue and respectively feature-tissue correlations. (The gene may be specifically expressed in a certain tissue substructure, while the visual feature may correspond to a distinct substructure of the same tissue. As long as the two different tissue substructures have similar distributions in the tissue slides, we may have indirect gene-tissue-feature correlations without perfect spatial overlap of gene expression and the visual feature).

For example, the high-level feature 30\_467 seems to detect primarily yellow connective tissue in pancreas samples, rather than acinar cells, which express the AMY2A gene (row 4, column 2 of Fig 4). The partial correlation  $r(g, f | t) = 0.107$  is much lower than  $r(g, f) = 0.928$ , with a p-value of the conditional independence test  $p = 0.0501 > 0.01$ . Therefore, the AMY2A expression and the feature 30\_467 are independent conditionally on the tissue (i.e. their correlation is indirect via the tissue variable).

To assess the prevalence of such indirect gene( $g$ )-tissue( $t$ )-feature( $f$ ) correlations, we performed conditional independence tests using partial correlations of the form  $r(g, f | t)$  with a significance threshold  $p = 0.01$ . Table 7 shows the resulting numbers of indirect dependencies for two different significance thresholds of the conditional independence test ( $\alpha = 0.01$  and 0.05). The gene-tissue-feature ( $g-t-f$ ) case corresponds to the indirect gene-feature correlations mediated by the tissue variable, discussed above. Indirect feature-gene-tissue ( $f-g-t$ ) dependencies correspond to genes that mediate the feature-tissue correlation (for which the classifier is

**Table 7. Numbers of significant gene-feature pairs with indirect dependencies gene-tissue-feature ( $g-t-f$ ), feature-gene-tissue ( $f-g-t$ ), gene-feature-tissue ( $g-f-t$ ). Conditional independence tests with  $\alpha = 0.01$  and respectively  $\alpha = 0.05$ .**

Thresholds	$g-f$ pairs	$g-t-f$		$f-g-t$		$g-f-t$	
		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$
corr 0.7 $\log_2$ expr 7	27,671	12,442	10,133	5,068	1,486	931	647
corr 0.8 $\log_2$ expr 10	2,176	1,495	1,254	37	12	1	1

<https://doi.org/10.1371/journal.pone.0242858.t007>



responsible). There are significantly fewer such indirect dependencies, and even fewer ones of the form  $g-f-t$ .

The visualizations of the prostate sample from row 3 of Fig 6 illustrate an interesting intermediary case in which although conditioning on the tissue leads to a big drop in correlation (from  $r(g,f) = 0.828$  to  $r(g,f|t) = 0.167$ ), the conditional independence test still rejects the null hypothesis ( $p = 0.00215 < 0.01$ , i.e. there is still conditional *dependence*, although a weak one). The gene-feature correlation is therefore not entirely explainable by the indirect gene-tissue-feature influence. This can also be seen visually in the gBP image of feature 30\_244, which detects blue glandular cells and surrounding yellow stroma (row 3, column 2 of Fig 6), similar to the spatial expression pattern of the KLK3 gene, which is specifically expressed in the glandular cells (see IHC image from column 4). The imperfect nature of the dependence is due to the feature detecting mostly basally situated glandular cells (blue) rather than all glandular cells.

## Discussion

Due to the limited numbers of high-quality datasets comprising both histopathological images and genomic data for the same subjects, there are very few research publications trying to combine visual features with genomics. Also, the relatively small numbers of samples compared to genes and image features make such integration problems difficult ('small sample problem').

The sparse canonical correlation analysis (CCA) used in [48] is an elegant and very general technique of determining correlations between two *sets* of variables (more precisely, finding linear combinations of variables from each of the two sets that are maximally correlated to one another). Since the sets of genes and respectively image features are very large, CCA is ideal for determining *general trends* (CCA components) in the complicated correlation structure of genes and features. On the other hand, in this paper we are interested in determining *individual* visual features correlated with genes, rather than CCA *composites* of such features, which may be hard to visualize and interpret. Such individual visual features might also be essential for discriminating between visually similar tissues (especially in the case of the number of variables exceeding the number of samples—sparsity constraints mitigate this problem, but do not eliminate it altogether). Moreover, multivariate regression analysis is typically not recommended for small samples. Also, while we use spatially invariant feature encodings obtained by aggregating feature values over both *spatial* and *tile* dimensions, [48] average their feature encodings only over the tile ("window") dimension and not over the spatial dimensions.

Other work searched for relationships between histological images and somatic cancer mutations. For example, [12] trained a CNN to predict the mutational status of just 10 genes (the most frequently mutated genes in lung cancer, rather than a genome-wide study). Similar mutation predictors were developed for hepatocellular carcinoma [20] and prostate cancer [21]. But the focus of these studies was obtaining mutation predictors, rather than correlating the mutational status with visual histological features.

There is a much larger body of research on purely visual algorithms for analyzing histological slides (without considering correlations with transcriptomics). However, a direct comparison of tissue classification accuracies is probably not very meaningful, since different sets of tissues were used in most papers.

The GTEx dataset was also used to obtain tissue classifiers in [49], for 5, 10, 20 and respectively 30 tissues. The highest *tile classification accuracies* reported in [49] for 30 tissues are 61.8% for VGG pretrained on Imagenet and respectively 77.1% for the network retrained from scratch on histological images, compared to our 81.5% accuracy for classifying 39 tissues (30 tissue *WSI* classification accuracies are not reported in [49]). The dataset in [49] contains

53,000 tiles obtained from 787 WSI images from GTEx, for 30 tissues. However, as already mentioned above, some of the additional tissues in our extended set of 39 tissues are hard to distinguish (e.g. ‘Artery—Aorta’ and ‘Artery—Coronary’ versus ‘Artery—Tibial’).

While many researches have already covered normal and diseased tissue classification in digital histopathology, only a small minority have begun to address the problem of interpretability of the resulting models. From this perspective, there are two main contributions of this work, one dealing with the *visual interpretability* of the models, the second involving their *biological interpretability*.

Firstly, we develop *visualization methods* for the features that are inferred automatically by deep learning architectures. These histological features have certain domain-specific characteristics, namely they are *location independent*, as well as precisely *quantifiable*. Quantifiability involves being able to estimate the numbers of occurrences of a specific feature in a given histological image, such as the numbers of specific cellular structures in a slide.

The second main contribution consists in assessing the *biological interpretability* of the deep learning models by correlating their inferred features with matching gene expression data. Such features correlated with gene expression have more than a visual interpretation—they correspond to biological processes at the level of the genome.

It is remarkable that the synthetic images of certain features resemble the corresponding tissue structures extremely well. Genes predominantly expressed in tissues with simpler morphologies tend to be correlated with simpler, lower level visual features, while genes specific to more complex tissue morphologies correlate with higher level features.

However, not all CNN architectures infer features that tend to be well correlated with gene expression profiles, for example VGG networks with batch normalization, Inception\_v3 or ResNet (which need batch normalization to deal with their extreme depth). It seems that batch normalization produces slight improvements in tile classification accuracies (though not for *whole* slides) at the expense of biological interpretability (i.e. significantly worse correlations of inferred features with the transcriptome—see [Table 4](#)).

## Conclusions

Current artificial intelligence systems are still not sufficiently developed to fully take over the tasks of the histopathologist, who is ultimately responsible for the final clinical decision. But the AI system could prove invaluable in assisting the clinical decision process. To do so, it needs to provide the histologist with as much meaningful knowledge as possible. Unfortunately, there is at present no established way to easily explain why a specific decision was made by a network when dealing with a given histopathology image. This is generally unacceptable in the medical community, as clinicians typically need to understand and justify the reasons for a specific decision. A reliable diagnosis must be transparent and fully comprehensible. This is also especially important for obtaining regulatory approval for use in clinical practice [50]. Of particular concern in the medical field is the uncertainty of the decisions taken by a deep network, which can be radically affected even by the change of very few pixels in an image, in case of a so-called adversarial attack [51].

Visualizations of the features automatically inferred by a deep neural network are a first step toward making the decisions of the network more *interpretable* and *explainable*. Moreover, correlating the visual histological features with specific gene expression profiles increases the confidence in the *biological interpretability* of these features, since the genes and their expression are responsible for the cell structures that make up these visual features.

This paper deals with identifying transcriptomic correlates of histology using Deep Learning, at first just in normal tissues, as a small step towards bridging the wide explanatory gap

between genes, their expression and the complex cellular structures of tissues that make up histological phenotypes. Of course, the relationships discovered in this study are only correlational. For elucidating causality, more complicated, perturbational experiments are needed, but the methodology used here is still applicable.

## Supporting information

**S1 Fig. Histograms of features, genes and their correlation.** (A) Histogram of feature values. (B) Histogram of  $\log_2$  transformed gene expression values  $\log_2(1+g)$ . (C) Histogram of gene-feature correlations (for genes with highest median tissue  $\log_2$  expression over 10). (D) Histogram of gene-feature correlations separately for non-negative features (e.g. outputs of ReLU units) and potentially negative features.

(PDF)

**S2 Fig. Numbers of correlated genes for individual features and respectively correlated features per gene.** (A) Numbers of correlated genes for individual features. Features are sorted in decreasing order of the corresponding numbers of correlated genes. (B) Numbers of correlated features per gene. Genes are sorted in decreasing order of the corresponding numbers of correlated features. Various correlation thresholds are applied (from 0.7 to 0.95).

(PDF)

**S3 Fig. Reproducibility of gene-feature correlations between datasets.** Scatter plots of gene-feature correlations computed on the validation- and respectively test dataset. (A) scatter plot of *correlations* ( $R = 0.9205$ ). (B) scatter plot of *Fisher transformed* correlations ( $R = 0.9233$ ).

(PDF)

**S1 Table. The numbers of slides and respectively tiles for each data set and tissue type.**

(PDF)

**S2 Table. Genes correlated with visual histopathological features.** Genes are grouped w.r.t. tissues and ordered by correlation—for each gene we only show the best correlated feature, in the form [layer]\_[channel] (for the architecture VGG16 and the test dataset; correlation threshold = 0.75,  $\log_2$  gene expression threshold = 7).

(PDF)

**S3 Table. Developmental and transcription regulation genes correlated with visual features.** Genes with Gene Ontology ‘*developmental process*’ or ‘*transcription regulator activity*’ annotations are grouped w.r.t. tissues and ordered by correlation—for each gene we only show the best correlated feature, in the form [layer]\_[channel] (for the architecture VGG16 and the test dataset; correlation threshold = 0.75,  $\log_2$  gene expression threshold = 7). Genes with ‘*transcription regulator activity*’ are shown in red.

(PDF)

**S4 Table. Confusion matrix for the whole slide tissue classifier.** VGG16 architecture, test dataset.

(PDF)

**S1 File. Supporting information.**

(PDF)

**S2 File. List of GTEx samples.** List of GTEx samples with their associated dataset (training, validation, test).

(TXT)

## Acknowledgments

We are deeply grateful to the GTEx consortium for making the gene expression data and histological images publicly available. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal (<https://gtexportal.org/home/>). LB is forever grateful to Eugenia Badea.

## Author Contributions

**Conceptualization:** Liviu Badea.

**Data curation:** Liviu Badea, Emil Stănescu.

**Formal analysis:** Liviu Badea.

**Investigation:** Liviu Badea, Emil Stănescu.

**Methodology:** Liviu Badea.

**Project administration:** Liviu Badea.

**Software:** Liviu Badea.

**Supervision:** Liviu Badea.

**Validation:** Liviu Badea, Emil Stănescu.

**Visualization:** Liviu Badea, Emil Stănescu.

**Writing – original draft:** Liviu Badea.

**Writing – review & editing:** Liviu Badea, Emil Stănescu.

## References

1. Xin L, Liu YH, Martin TA, Jiang WG. The era of multigene panels comes? The clinical utility of Oncotype DX and MammaPrint. *World journal of oncology*. 2017 Apr; 8(2):34. <https://doi.org/10.14740/wjon1019w> PMID: 29147432
2. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Vega JE, et al. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*. 2018 Mar 27; 115(13):E2970–9. <https://doi.org/10.1073/pnas.1717139115> PMID: 29531073
3. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In *Advances in neural information processing systems 2014* (pp. 2672–2680).
4. Baak JP, Lindeman J, Overdip SH, Langley FA. Disagreement of histopathological diagnoses of different pathologists in ovarian tumors—with some theoretical considerations. *European Journal of Obstetrics & Gynecology and Reproductive Biology*. 1982 Feb 1; 13(1):51–5. [https://doi.org/10.1016/0028-2243\(82\)90037-5](https://doi.org/10.1016/0028-2243(82)90037-5)
5. Yuan Y, Van Allen EM, Omberg L, Wagle N, Amin-Mansour A, Sokolov A, et al. Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nature biotechnology*. 2014 Jul; 32(7):644–52. <https://doi.org/10.1038/nbt.2940> PMID: 24952901
6. Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to digital image analysis in whole-slide imaging: a white paper from the digital pathology association. *Journal of pathology informatics*. 2019;10.
7. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems 2012* (pp. 1097–1105).
8. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*. 2015 Dec 1; 115(3):211–52.

9. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, et al. Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. *Nature genetics*. 2013 Sep 26; 45(10):1113. <https://doi.org/10.1038/ng.2764> PMID: 24071849
10. Litjens G, Bandi P, Ehteshami Bejnordi B, Geessink O, Balkenhol M, Bult P, et al. 1399 H&E-stained sentinel lymph node sections of breast cancer patients: the CAMELYON dataset. *GigaScience*. 2018 Jun; 7(6):giy065.
11. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nature genetics*. 2013 Jun; 45(6):580–5. <https://doi.org/10.1038/ng.2653> PMID: 23715323
12. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nature medicine*. 2018 Oct; 24(10):1559–67. <https://doi.org/10.1038/s41591-018-0177-5> PMID: 30224757
13. Wang C, Yang H, Bartz C, Meinel C. Image captioning with deep bidirectional LSTMs. In *Proceedings of the 24th ACM international conference on Multimedia 2016* Oct 1 (pp. 988–997).
14. Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*. 2018 Jan 1; 16:34–42. <https://doi.org/10.1016/j.csbj.2018.01.001> PMID: 30275936
15. Spanhol FA, Oliveira LS, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. In *2016 international joint conference on neural networks (IJCNN) 2016* Jul 24 (pp. 2560–2567). IEEE.
16. Sheikhzadeh F, Guillaud M, Ward RK. Automatic labeling of molecular biomarkers of whole slide immunohistochemistry images using fully convolutional networks. *arXiv preprint arXiv:1612.09420*. 2016 Dec 30.
17. Shah M, Wang D, Rubadue C, Suster D, Beck A. Deep learning assessment of tumor proliferation in breast cancer histological images. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2017* Nov 13 (pp. 600–603). IEEE.
18. Chen H, Qi X, Yu L, Heng PA. DCAN: deep contour-aware networks for accurate gland segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition 2016* (pp. 2487–2496).
19. Caie PD, Turnbull AK, Farrington SM, Oniscu A, Harrison DJ. Quantification of tumour budding, lymphatic vessel density and invasion through image analysis in colorectal cancer. *Journal of translational medicine*. 2014 Dec 1; 12(1):156. <https://doi.org/10.1186/1479-5876-12-156> PMID: 24885583
20. Chen M, Zhang B, Topatana W, Cao J, Zhu H, Juengpanich S, et al. Classification and mutation prediction based on histopathology H&E images in liver cancer using deep learning. *npj Precision Oncology*. 2020 Jun 8; 4(1):1–7.
21. Schaumberg AJ, Rubin MA, Fuchs TJ. H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer. *BioRxiv*. 2018 Jan 1:064279.
22. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature communications*. 2016 Aug 16; 7(1):1–0. <https://doi.org/10.1038/ncomms12474> PMID: 27527408
23. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*. 1997 Nov 15; 9(8):1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
24. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. 2014 Jun 3.
25. 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 2015* Oct 5 (pp. 234–241). Springer, Cham.
26. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*. 2017 Feb 1; 36:61–78. <https://doi.org/10.1016/j.media.2016.10.004> PMID: 27865153
27. Gadermayr, M., Gupta, L., Klinkhammer, B. M., Boor, P. and Merhof, D., 2018. Unsupervisedly Training GANs for Segmenting Digital Pathology with Automatically Generated Annotations. *arXiv preprint arXiv:1805.10059*.
28. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014 Sep 4.
29. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2016* (pp. 770–778).

30. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition 2016 (pp. 2818–2826).
31. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467. 2016 Mar 14.
32. Chollet F. et al. Keras: Deep learning library for theano and tensorflow. <https://keras.io/>.
33. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. In Advances in neural information processing systems 2019 (pp. 8026–8037).
34. Xu Y, Li Y, Shen Z, Wu Z, Gao T, Fan Y, et al. Parallel multiple instance learning for extremely large histopathology image analysis. BMC bioinformatics. 2017 Dec; 18(1):1–5.
35. GTEx. The Genotype-Tissue Expression (GTEx) project. (<https://gtexportal.org/home/>)
36. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167. 2015 Feb 11.
37. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806. 2014 Dec 21.
38. Olah C, Mordvintsev A, Schubert L. Feature visualization. Distill. 2017 Nov 7; 2(11):e7.
39. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nature genetics. 2000 May; 25(1):25–9. <https://doi.org/10.1038/75556> PMID: 10802651
40. Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. Nucleic acids research. 2019 Jan 8; 47(D1):D330–8. <https://doi.org/10.1093/nar/gky1055> PMID: 30395331
41. Aruga J, Millen KJ. ZIC1 Function in Normal Cerebellar Development and Human Developmental Pathology. Advances in experimental medicine and biology. 2018; 1046:249–68. [https://doi.org/10.1007/978-981-10-7311-3\\_13](https://doi.org/10.1007/978-981-10-7311-3_13) PMID: 29442326
42. Pieper A, Rudolph S, Wieser GL, Götze T, Mießner H, Yonemasu T, et al. NeuroD2 controls inhibitory circuit formation in the molecular layer of the cerebellum. Scientific reports. 2019 Feb 5; 9(1):1–7.
43. Akazawa H, Komuro I. Cardiac transcription factor Csx/Nkx2-5: Its role in cardiac development and diseases. Pharmacology & therapeutics. 2005 Aug 1; 107(2):252–68. <https://doi.org/10.1016/j.pharmthera.2005.03.005> PMID: 15925411
44. Chen H, Shi S, Acosta L, Li W, Lu J, Bao S, et al. BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. Development. 2004 May 1; 131(9):2219–31. <https://doi.org/10.1242/dev.01094> PMID: 15073151
45. Kelberman D, Rizzoti K, Lovell-Badge R, Robinson IC, Dattani MT. Genetic regulation of pituitary gland development in human and mouse. Endocrine reviews. 2009 Dec 1; 30(7):790–829. <https://doi.org/10.1210/er.2009-0008> PMID: 19837867
46. Fernandez LP, Lopez-Marquez A, Santisteban P. Thyroid transcription factors in development, differentiation and disease. Nature Reviews Endocrinology. 2015 Jan; 11(1):29–42. <https://doi.org/10.1038/nrendo.2014.186> PMID: 25350068
47. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015 Jan 23; 347 (6220). <https://doi.org/10.1126/science.1260419> PMID: 25613900
48. Ash JT, Darnell G, Munro D, Engelhardt BE. Joint analysis of gene expression levels and histological images identifies genes associated with tissue morphology. bioRxiv. 2018 Jan 1:458711.
49. Bizzego A, Bussola N, Chierici M, Maggio V, Francescato M, Cima L, et al. Evaluating reproducibility of AI algorithms in digital pathology with DAPPER. PLoS computational biology. 2019 Mar 27; 15(3): e1006269. <https://doi.org/10.1371/journal.pcbi.1006269> PMID: 30917113
50. Tizhoosh HR, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. Journal of pathology informatics. 2018;9.
51. Athalye A, Engstrom L, Ilyas A, Kwok K. Synthesizing robust adversarial examples. In International Conference on Machine Learning 2018 Jul 3 (pp. 284–293).