Unsupervised Resolution of Histomorphologic Heterogeneity in Renal Cell Carcinoma Using a Brain Tumor–Educated Neural Network

Kevin Faust, HBS^{1,2}; Adil Roohi, BSc^{1,3}; Alberto J. Leon, PhD⁴; Emeline Leroux, MSc⁵; Anglin Dent, BSc⁵; Andrew J. Evans, MD, PhD^{5,6}; Trevor J. Pugh, PhD^{1,4,7}; Sangeetha N. Kalimuthu, MD^{5,6}; Ugljesa Djuric, PhD¹; and Phedias Diamandis, MD, PhD^{5,6,7}

PURPOSE Applications of deep learning to histopathology have proven capable of expert-level performance, but approaches have largely focused on supervised classification tasks requiring context-specific training and deployment. More generalizable workflows that can be easily shared across subspecialties could help accelerate and broaden adoption. Here, we hypothesized that histology-optimized feature representations, generated by a convolutional neural network (CNN) during supervised learning, are transferable and can resolve meaningful differences in large-scale, discovery-type unsupervised analyses.

METHODS We used a CNN, previously trained to recognize brain tumor histomorphologies, to extract 512 feature representations from > 550 digital whole-slide images (WSIs) of renal cell carcinomas (RCCs) from The Cancer Genome Atlas and other previously unencountered tumors. We use these extracted feature vectors to conduct unsupervised image-set clustering and analyze the clinical and biologic relevance of the intra- and interpatient subgroups generated.

RESULTS Within individual WSIs, feature-based clustering could reliably segment tumor regions and other relevant histopathologic subpatterns (eg, adenosquamous and poorly differentiated regions). Across the larger RCC cohorts, clustering extracted features generated subgroups enriched for clinically relevant subtypes (eg, papillary RCC) and outcomes (eg, survival). Importantly, individual feature activation mapping highlighted salient subtype-specific patterns and features of malignancies (eg, nuclear grade, sarcomatous change) contributing to subgroupings. Moreover, some proposed clusters were enriched for recurring, human-based RCC-subtype misclassifications.

CONCLUSION Our data support that CNNs, pretrained on large histologic datasets, can extend learned representations to novel scenarios and resolve clinically relevant intra- and interpatient tissue-pattern differences without explicit instruction or additional optimization. Repositioning of existing histology-educated networks could provide scalable approaches for image classification, quality assurance, and discovery of unappreciated patterns and subgroups of disease.

JCO Clin Cancer Inform 4:811-821. © 2020 by American Society of Clinical Oncology

INTRODUCTION

The ability to project previously learned knowledge to novel environments is an essential skill for advanced clinical decision-making and discovery. Detection of biologic differences among patients, especially those not previously encountered or anticipated, is also fundamental for preventing misclassifications and for characterizing and predicting the clinical behavior of rare and emerging diseases. Although recent breakthroughs in deep learning have allowed convolutional neural networks (CNNs) to conduct highly sophisticated, supervised, image-based classification tasks,¹⁻⁵ examples of unsupervised, clinically relevant inferences, extending outside their intended training context, are rare.

Pathologists frequently rely on recurring histomorphologic patterns to organize diseases into meaningful

subgroups. For example, because hallmark features of malignancy are shared across many cancer types, pathologists are often able to apply lessons learned from common pathologies to approximate clinical behaviors of rare atypical lesions or those outside their specific subspecialty. However, it is unclear if such unsupervised inferences are also operational in deep learning–based histomorphologic analysis. In other domains of computer visions,⁶⁻⁹ extracting learned feature representations from pretrained CNNs has helped facilitate unsupervised image-clustering tasks. If also effective in histopathology, it could help alleviate the need for laborious narrow and context-specific annotation and training steps and create sharable computational pathology tools across disciplines.

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on August 7, 2020 and published at ascopubs.org/journal/cci on September 18, 2020; DOI <https://doi.org/10.1200/CCI.20.00035>

CONTEXT

Key Objectives

The goal of this study was to address if previous, histology-optimized, deep convolutional neural networks can be applied to novel scenarios to help resolve clinically meaningful morphologic patterns of disease without the need for additional context-specific training.

Knowledge Generated

Image-based clustering highlighted that deep-learning feature representations can resolve pertinent intra- and intertumor patterns of tissue heterogeneity in an unsupervised manner. The morphology-based groupings proposed correlated with biologically distinct tumor and tissue types, varying degrees of tumor differentiation, and clinical aggressiveness.

Relevance

The demonstrated transferability of deep-learning features in histomorphologic analysis provides support for the use of pretrained neural networks as generalizable, scalable, and sharable digital pathology tools for tissue annotation, tumor classification, quality assurance, and large-scale histomic profiling.

Here, we set out to address if a CNN educated in a particular pathology subspecialty could generalize learned concepts to another organ site and infer meaningful levels of intra- and intertumor heterogeneity. Indeed, we show that a neuropathology-educated CNN (CNN_{NP}), trained to differentiate brain-tumor types,¹⁰ could apply learned representations to find histopathologically relevant patterns in renal cell carcinoma (RCC) and other previously unencountered tumor and tissue types without the need for specific direction or additional training.

We previously used transfer learning and a diverse set of 838,644 human-annotated histopathologic image patches (516 μm^2) generated from 1,656 whole-slide images (WSIs) spanning 1,027 patients to fine-tune the VGG19 CNN¹¹ to recognize microscopic patterns of different brain tumors.¹⁰ By extracting 512-dimensional deep learning feature vectors (DLFVs) found in the CNN's final, global average pooling layer, we showed that many learned representations, defined by individual deep learning-engineered features (DLFs), correlated with salient, human-defined, cytoarchitectural constructs such as fibrosis, mucin, epithelium, and luminal structures.¹⁰ Given the shared diagnostic and prognostic value of many of these histologic patterns across cancer types, we reasoned that the learned features of this CNN_{NP} could be leveraged to make intelligent predictions in other neoplasms without the need for additional optimizations. To address this hypothesis, we use this pretrained CNN_{NP} to extract and cluster feature representations generated from a large collection of RCC and other previously unencountered tumor specimens and evaluate the biological relevance of the proposed groupings.

METHODS

Histopathology Image Source

We retrieved 914 digitally scanned, hematoxylin and eosin-stained WSIs from the RCC cohorts of The Cancer Genome Atlas (TCGA).^{12,13} We chose RCCs for the large-scale proof-of-concept analysis because they represent

neoplasms with robust clinicopathologic relationships, including survival differences associated with distinct tumor cytoarchitectures and nuclear features (Fig 1A).^{14,15} This TCGA cohort included the following subtypes: clear-cell RCC (TCGA-KIRC; n = 529 patients), papillary RCC (TCGA-KIRP; n = 299), and chromophobe (TCGA-KICH; n = 86). TCGA WSIs had variable compression qualities (specifically, 0.70 and 0.30) and apparent magnifications ($\times 20$ and $\times 40$). For compatibility with the pretrained CNN_{NP}, we included WSIs scanned at a compression quality of 0.70 and that could produce at least 20 lesional image patches. Cases with a compression quality of 0.30 were excluded because we found clustering biases for cases with a similar compression quality (Data Supplement). Images with an apparent magnification of $\times 40$ were down-sampled to $\times 20$. The final cohort, based on these criteria, included 396 KIRC, 154 KIRP, and 7 KICH tumors. Associated clinical information including histologic subtype and survival were also retrieved from TCGA. Representative slides (n = 5) from each RCC cohort were chosen to assess the ability of the pretrained CNN to conduct unsupervised segmentations or annotations. Given the small compatible TCGA-KICH cohort, this group was only used for segmentation studies and omitted from downstream interpatient analyses.

Subgrouping and survival analyses were conducted using confirmed lesional image patches (dimensions: 1,024 $\times$ 1,024 pixels [0.504 $\mu\text{m}/\text{pixel}$]; 516 μm^2) generated from available TCGA-KIRC/KIRP cases using a semiautomated approach.¹⁶ Briefly, a CNN was used to automate extraction of candidate tumor regions from each WSI, which were then manually reviewed by a trained human interpreter who removed erroneous or artifactual tiles. In total, this produced a collection of 219,718 image tiles spanning 396 KIRC and 154 KIRP tumors.¹⁷ For cases in which multiple slides were available, images from different WSIs were merged and analyzed together as a single case. We also included a separate collection of additional neoplasms not found in our initial training set from both local (specifically,

FIG 1. Analyzing histologic patterns of renal cell carcinoma (RCC) using a brain tumor-educated CNN. (A) Study schematic. Kaplan-Meier survival curves summarizing the known clinicopathologic correlates across different histologic subtypes and nuclear grades of TCGA-RCC cases. Only cases with whole-slide images (WSIs) scanned with a compression quality 0.70 were included. Image patches from cases were analyzed by a CNN, previously trained on brain tumors, to generate 512-dimensional deep learning feature vectors (DLFVs). Clustering approaches were used to explore overall data structures. (B) Representative H&E-stained WSI of a TCGA-KIRC. (C, D) Complete-linkage hierarchical clustering and *t*-SNE plot of the DLFVs generated from non-overlapping, 516- μm^2 image patches extracted from WSI in (B) for $k = 8$ subgroups. (E) Unsupervised annotation of WSI in (B) for $k = 8$ subgroups. Tiles with mostly blank space have been omitted for clarity. (F) Representative image patches (516 μm^2) from (E) highlighting the principal tissue patterns associated with the DLFV-driven tissue subgroups. CNN, convolutional neural network; H&E, hematoxylin and eosin; IFTA, interstitial fibrosis and tubular atrophy; KICH, chromophobe renal cell carcinoma; KIRC, clear-cell renal cell carcinoma; KIRP, papillary renal cell carcinoma; Max, maximum; ReLU, rectified linear unit; *t*-SNE, *t*-distributed Stochastic Neighbor Embedding; TCGA, The Cancer Genome Atlas.

pancreatic cancer and retinoblastoma) and other TCGA cohorts (namely, uterine carcinosarcoma and endometrial carcinoma) to highlight generalizability of this unsupervised approach beyond RCCs. Images and medical data used in this study were retrieved from previously published, anonymized, publicly available resources; thus, additional institutional research ethics board approval was not applicable.

CNN Selection

We use a pathology-optimized version¹⁰ of the pretrained VGG19 CNN¹¹ to extract feature representations. In a previous study, we used transfer learning to fine-tune the existing ImageNet-based node weightings toward histologic patterns found in a set of 838,644 pathologist-annotated

FIG 2. Unsupervised and automated resolution of intratissue heterogeneity across various tumor types. (A, B) Representative hematoxylin and eosin–stained whole-slide images (WSIs) of (A) metastatic pancreatic adenosquamous carcinoma to the adrenal gland and (B) uterine endometrial carcinoma. (C, D) Unsupervised annotations of WSIs shown in (Figure 2A and 2B for k = 8 and k = 10 subgroups, respectively) (see Data Supplement). (continued on following page)

image patches generated from 1,656 WSIs spanning 1,027 tumor specimens that included 74 different lesional and nonlesional tissue types.^{10,16} Although this image resource did include cases of metastatic clear-cell RCC, other important parameters we uncovered in this study were never part of the initial supervised training steps. This includes absence of renal parenchyma, other subtypes of renal neoplasms (ie, papillary and chromophobe RCC, oncocytoma) or survival information. Similarly, we further decoupled previously learned class labels from our analysis by focusing completely on feature patterns found within the global average pooling layer of the CNN prior to softmax reduction and classification (see the next section on DLFV generation). As such, we believe the groupings generated in this study best represent unsupervised associations. The CNN used did not undergo additional training or optimization for this study.¹⁸

DLFV Generation and Unsupervised Clustering

To generate a quantitative histomorphologic signature (or DLFV) for each case, we averaged the DLF values extracted from the final global average pooling layer as each case's associated tiles were passed through the network, as previously described.¹⁰ We previously showed that individual DLFs within this vector often approximate human-like histologic features such as fibrosis, epithelium, and mucoid patterns. Importantly, these activations are maintained across different unrelated cancer types (eg, primary and metastatic tumors) allowing DLFVs to serve as transferable, objective, and quantitative histopathologic signatures for characterizing microscopic patterns. For interpatient analyses, we randomly select 20 images to generate a global averaged DLFV. To organize reoccurring patterns within our case cohorts, we use DLFVs to conduct both complete linkage and Ward hierarchical clustering (using the L2 metric), as previously described.¹⁰ To understand clinicopathologic correlations of image clusters, we used a silhouette approach to define the optimal number of subgroups. For many instances, we also used a stepwise k-means clustering approach to reveal additional subtle or complex subgroupings. Generation and clustering of DLFVs for all experiments was repeated multiple times and yielded similar results, trends, and conclusions throughout the study. Python code used to generate DLFVs is available at BitBucket.¹⁸

Our presented approach is notably distinct from traditional, supervised deep learning workflows in pathology, which would require generating individual training image sets and separate CNN optimizations for each of the presented segmentation, subclassification, and prognostication tasks.

Although the CNN model we use was initially trained using labeled data, this supervised component focused on differentiating between brain tumor and tissue types—a task largely distinct from those presented in this work. Instead, we use this diversely trained CNN as a histopathology feature extractor and generate all groupings using unsupervised learning techniques (eg, dimensionality reduction and clustering) rather than relying on further optimization of this CNN using context-specific supervised image labels.

Molecular Classification of DLFV-Defined TCGA-KIRC Subgroup Tumors Exhibiting Clinically Indolent Behavior

To molecularly classify clinically indolent RCC cases grouped together by our unsupervised approach, we retrieved RNA-sequencing gene-expression files from the KIRC, KIRP, and KICH TCGA projects available from the BROAD Firehose data repository. The R package model Tsnr was used to select the 1,000 most informative genes (broadest interquartile range), perform t-SNE (t-distributed Stochastic Neighbor Embedding¹⁹) analysis, and plot the resulting 2-dimensional configuration. A nearest-neighbor approach was used to assign an objective molecular class to these outlier cases.

DLF Selection and Activation Mapping

We used DLFs enriched in specific subgroups to understand morphologic features that were potentially responsible for the observed clustering. First, we identified the most significantly enriched features (*t* test) between subgroups of interest (eg, DLFs enriched in TCGA-KIRC *v* TCGA-KIRP, clinically indolent *v* aggressive). To visually assess what morphology this DLF is potentially detecting, we retrieved images at the extremes of DLF values within our entire cohort and asked two or more blinded pathologists to describe morphologic differences between the two extremes. We further validated this qualitative exercise by generating an isolated feature activation map (FAM) of candidate DLFs of interest to ensure that the subtle coordinates of the DLF matched the location of the histologic feature proposed by the study pathologists. Python code used to generate the FAMs can be found at BitBucket.¹⁸

RESULTS

To first assess if the CNN_{NP} could independently differentiate RCCs from other surrounding tissue elements, we selected representative WSIs from the TCGA-KIRC, TCGA-KICP, and TCGA-KICH RCC TCGA cohorts (Fig 1B). We automated tiling of these WSIs into 516 μm^2 image patches and subgrouped them using complete-linkage hierarchical clustering of their corresponding DLFV values generated by

FIG 2. (Continued). (E, F) Proximity-based comparisons of image patches on *t*-distributed Stochastic Neighbor Embedding plots mirror human interpretations of defined subgroups. (G, H) Representative tumor image patches (516 μm^2) from Figure 2C and 2D highlighting different tumor patterns enriched in the deep learning feature vector-driven tissue subgrouping.

FIG 3. Unsupervised subgrouping of renal cell carcinomas (RCCs) using deep learning feature vectors (DLFVs) correlating with TCGA-defined histotypes. (A, B) Ward hierarchical clustering and *t*-SNE visualization of the DLFVs generated for each of the 396 TCGA-KIRC and 154 TCGA-KIRP cases. (C, D) An illustrative, individual DLF over-represented in the TCGA-KIRC-enriched cluster. Comparison of cytologic patterns found (continued on following page)

the CNN_{NP} (Fig 1C-1E; Data Supplement). This unsupervised approach initially divided tissue into two major groups composed of cellular (eg, RCC, renal parenchyma) and paucicellular (eg, adipose, hemorrhage, necrosis) elements ($k = 2$ subgroups, silhouette method; Data Supplement). Comparing finer DLFV-driven subgroupings ($k = 3-9$) to expert-level domain knowledge demonstrated resolution of other relevant histomorphologies not defined a priori. This included differentiating nonneoplastic renal parenchymal tissue from RCC (teal v yellow clusters) and defining distinct areas of renal parenchyma showing interstitial fibrosis and tubular atrophy (green cluster) and inflammatory infiltrates (magenta cluster; Fig 1E and 1F; Data Supplement). Grouping of cancer regions from nonneoplastic tissue elements mirrored expert annotations across multiple randomly selected cases ($n = 5$; normalized mutual information [NMI] score = 1.0; Data Supplement). This unsupervised workflow also generalized well when further extended and validated on other previously unencountered cancer types, including retinoblastoma, pancreatic, and uterine cancers and their associated nonneoplastic tissue (Fig 2; Data Supplement). Together, this hierarchical approach to subgrouping various sophisticated histomorphologies provides a dynamic, generalizable, and objective framework for exploring complex intratumoral patterns of heterogeneity.

The robust detection of these subtle histologic patterns suggested that learned representations stored within the CNN_{NP} may also be able to resolve interpatient clinicopathologic correlates in RCC, including histologic subtypes and International Society of Urologic Pathologists nuclear grades.²⁰ Therefore, next we assessed if the CNN_{NP}-derived DLFVs, could also subgroup the major histologic subgroups of RCC recognized by human experts. We extracted tumor-containing image patches from the TCGA-KIRC ($n = 396$) and TCGA-KIRP ($n = 154$) cohorts, digitally phenotyped and subgrouped them using their averaged DLFV (Fig 3A and 3B; Data Supplement). Indeed, we noted nonrandom grouping of RCCs into clusters enriched for either TCGA-KIRP or TCGA-KIRC using various clustering parameters (Rand Index = 0.66; $P = 4.5e^{-75}$ [χ^2 test, $k = 2$, Ward clustering] and $P = 9.8e^{-20}$ - $1.5e^{-37}$ [χ^2 test, $k = 2-7$, complete-linkage clustering]). Similarly, clustering of TCGA-KIRP cases alone revealed subgroups enriched for either type 1 or type 2 papillary RCCs ($P = 2.5e^{-3}$ [χ^2 test, $k = 3$]; Data Supplement). Importantly, we noted DLFs enriched within the clusters composed primarily of TCGA-KIRC and TCGA-KIRP correlated with clear-cell and papillary

morphologies, respectively, both on review by our pathologists and by using individual salient FAMs (Fig 3C-3F). Interestingly, cases of TCGA-KIRC with low values for clear-cell correlate DLF219 (or high values for the papillary correlate DLF309) appeared to represent potentially misclassified chromophobe RCCs or oncocytomas—more indolent renal neoplasms (Fig 3C and 3E). Similarly, TCGA-KIRP cases with high DLF219 values resembled the newly defined clear-cell papillary RCC subtype that shows indolent clinical behavior (Fig 3D).²¹

To ensure this RCC subgrouping could not be solely attributed to cytopathological features of metastatic clear-cell RCC learned during the original training period, we reclustered cases using only a small fraction ($n = 10$ of 512) of KIRP-enriched DLFs. Even this limited DLF signature, not directly activated by clear-cell RCC patterns, could achieve excellent separation between subtypes (Data Supplement). In fact, even individual DLFs such as epithelial-correlated feature DLF165 (enriched in non-RCC metastatic tumors in the original model) and paranuclear clearing-correlated feature DLF113 (enriched in oligodendrocytes of white matter) could predictably differentiate these 2 common RCC subtypes (Data Supplement). We further validated this unsupervised subgrouping approach in uterine cancers to highlight generalizability (NMI = 1.0; Data Supplement). Together, these results show how previously learned representations can be intelligently extended to find pertinent and analogous cytologic patterns in other tissue types without the need for explicit direction or a particular task in mind.

In addition to these major cytoarchitectural RCC subtypes, hallmark cellular changes of cancer progression (eg, nuclear grade) are also well-recognized predictors of clinical aggressiveness among clear-cell RCCs. To explore if the CNN_{NP} learned and could apply such features of malignancy to further stratify RCCs, we subgrouped the TCGA-KIRC cohort ($n = 396$) on the basis of their DFLV signatures. Complete-linkage hierarchical clustering produced 2 to 3 major subgroups with distinct survival intervals ($P = 3.0e^{-05}$, log-rank test; Fig 4A and 4B; Data Supplement). To understand which learned representations contributed to this clustering, we examined histologic correlates of DLFs enriched in the computer-derived subgroups showing indolent and aggressive clinical behaviors ($k = 3$; yellow v green clusters in Figure 4C-4E). Specific features found in the poor prognostic subgroup histologically correlated with nuclear (Fuhrman/International Society of Urologic

FIG 3. (Continued). on H&E-stained image patches with high (top panels) and low (low panels) DLF219 values highlight a relationship with a clear-cell morphology in (C) TCGA-KIRC and (D) TCGA-KIRP cases. FAMs localized DLF219 to tumor cells within the image patches. (E, F) Similarly, H&E and FAM images of a DLF over-represented in the TCGA-KIRP-enriched cluster. This DLF detected a papillary-like pattern in both TCGA-KIRC and TCGA-KIRP cohorts. FAM, feature activation map; H&E, hematoxylin and eosin; KIRC, chromophobe renal cell carcinoma; KIRC, clear-cell renal cell carcinoma; KIRP, papillary renal cell carcinoma; TCGA, The Cancer Genome Atlas.

Pathologists–like) grading (eg, DLF499) and sarcomatoid change (DLF442; Fig 4D, 4F, and 4G). Conversely, DLFs enriched within the TCGA-KIRC subgroup showing indolent clinical behavior correlated with cystic spaces and eosinophilic change (Fig 4E, 4H, and 4I).

Although cystic change is a known favorable prognostic biomarker of RCC, the eosinophilic morphology detected by DLF254 again was suspicious for misclassified renal neoplasms within the TCGA-KIRC cohort. To test this hypothesis, we more closely explored the histomorphologies and RNA signatures of TCGA-KIRC cases found in the CNN_{NP} cluster that displayed an indolent clinical course (Fig 4J; Data Supplement). Indeed, of the nine cases, only two had a clear-cell morphology; the remaining cases had papillary (n = 1), chromophobe (n = 4), or oncocytic (n = 2) features (Data Supplement). Molecularly, comparison of RNA signatures of these cases to patterns of the entire TCGA-RCC cohorts confirmed only two of nine cases had RNA profiles supportive of classic clear-cell RCCs, with the remaining having chromophobe RCC (n = 5) or outlier (n = 2) RNA signatures (Fig 4J).

Last, to explore if at least some of these individual neuropathology-optimized DLFs could independently serve as biomarkers in clear-cell RCC, we examined their prognostic power across the complete TCGA-KIRC data set. Indeed, cases registering high (top 20%; n = 79) versus low (bottom 20%; n = 78) DLF values showed significantly distinct survival outcomes and often correlated with Fuhrman grading (Fig 4K–4M; Data Supplement). Importantly, these trends were maintained even at more conservative cutoffs (quartile, median DLF values), highlighting the potential value of such quantitative histologic correlates over traditional discrete grading systems. Last, consistent with recent reports,²² the defined nuclear grade correlate (DLF499) was transferable and also associated with survival in the TCGA-KIRC cohort (Data Supplement).

DISCUSSION

Deep learning has the potential to automate and objectify many manual and qualitative tasks across all aspects of human life, including medical image analysis.²³ However, much of the progress in computational pathology has

focused on highly defined and supervised classification tasks that require manual, human-designed training data and discrete goals. Here, using RCC and various other tumors as a proof of concept, we show that meaningful patterns of intra- and intertumor heterogeneity can be automatically proposed by a CNN from principles learned in other environments. Moreover, this approach highlighted how unappreciated and recurring anomalies in large patient cohorts (eg, misclassified oncocytomas and chromophobe RCCs in TCGA-KIRC) could be autonomously detected in expert-annotated data sets. Finally, we highlight that these decisions are indeed interpretable and appear to be driven by discrete pathologist-like histologic constructs (eg, sarcomatoid and nuclear features) that can be predictably applied to independent settings.

Image-clustering workflows can provide automation for complex decision-making tasks in digital pathology. First, unsupervised approaches to image segmentation provide complementary benefits to traditional supervised workflows, including the ability to stratify tissue patterns into meaningful subregions that may not have been anticipated or included during training. Indeed, we show resolution of unexpected subregions of renal parenchyma with interstitial fibrosis and tubular atrophy and squamous and poorly differentiated regions in pancreatic and uterine carcinomas. Similarly, these approaches can effectively detect diagnostically challenging atypical or untrained classes for closer human review, molecular workup, and error reduction.

The ability of unsupervised approaches to automate sorting of massive collections of histologic patterns also has significant implications for research and discovery. Because most medical imaging data sets produced are only partially labeled, efficient organization of cases and subregions, based on unbiased similarities, offers scalable ways to improve and accelerate cohort developments for subsequent supervised tasks.²⁴ Similarly, tools that can quantitatively survey hundreds to thousands of histologic patterns across large clinical trial cohorts offer automated solutions to defining histologic biomarkers that predict differential therapy responses. Moreover, similar to other image-based tasks,^{6,25–28} image-clustering approaches can

FIG 4. Clustering of TCGA-KIRC cohort using deep learning feature vectors (DLFVs) defined prognostically relevant subgroups and deep learning–patterned histomorphologies. (A) Complete-linkage hierarchical clustering and division of the TCGA-KIRC DLFV signatures into 3 subgroups. (B) Kaplan-Meier curves show divergent survival patterns of the DLFV-defined subgroups. (C) Volcano plot of the DLFs between the clinically aggressive and clinically indolent subgroups. (D, E) Dot plots highlighting relative value of DLFs enriched in either (D) clinically aggressive and (E) clinically indolent subgroups of Figure 4A and 4B. (F, G) Comparison of hematoxylin and eosin–stained images and feature activation maps with high (top panels) and low (low panels) values of selected DLFs enriched in the clinically aggressive cluster. DLF499 and DLF442 were deemed to correlate with high nuclear grade and sarcomatoid morphologies, respectively. (H, I) DLFs (DLF263/254) enriched in the clinically indolent cluster were deemed to correlate with cystic and oncocytoma-like morphologies, respectively. (J) *t*-SNE plot comparing the RNA patterns of the entire TCGA renal cell carcinoma cohort. The 9 TCGA-KIRC cases belonging to the DLFV-defined clinical indolent cluster are labeled for reference. (K–N) Kaplan-Meier curves of upper and lower pentiles (20%; n = 39) of individual DLFs across all cases. Log-rank test *P* values are displayed for each comparison. Additional cutoffs are shown in the Data Supplement. Ctrl, control; DLF, deep learning–engineered feature; RNA-seq, RNA sequencing; *t*-SNE, *t*-distributed Stochastic Neighbor Embedding; TCGA, The Cancer Genome Atlas.

be leveraged to conduct similarity searches of rare tumors in large, archival, digital databases for unique cohort development and downstream molecular profiling.

Overall, we show that, similar to other areas of computer vision,⁶⁻⁹ feature representations extracted from pretrained CNNs facilitate powerful, unsupervised image-clustering outputs for pathology. Given their diverse applications presented in this study, development of CNN feature extractors, pretrained on large diverse pathology data sets, creates generalizable and transferable histomorphologic knowledge that could offset the need for repeated task-specific image-set generation and CNN training. This DLFV-driven workflow is theoretically generalizable to other

existing CNNs and offers powerful and democratized digital histology tools for rapid dissemination to a wide range of unsupervised and supervised applications within the pathology informatics community.

Data Availability

Histopathology images and clinical information used in this study are available in a public repository from the TCGA Data Portal.²⁹ The pretrained CNN used in this study is publicly available at <https://bitbucket.org/diamandislabi/faust-feature-vectors-2019>. The Data Supplement can be found in our BitBucket repository¹⁷ or by contacting the corresponding authors.

AFFILIATIONS

¹Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada

²Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

³Harvard Extension School, Cambridge, MA

⁴Ontario Institute for Cancer Research, Toronto, Ontario, Canada

⁵Laboratory Medicine Program, University Health Network, Toronto, Ontario, Toronto, Canada

⁶Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

⁷Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

CORRESPONDING AUTHOR

Phedias Diamandis, MD, PhD, University Health Network, Canada, 200 Elizabeth St, Rm 11E427, Department of Pathology, Toronto General Hospital, Toronto, ON M5G 2C4, Canada; e-mail: p.diamandis@mail.utoronto.ca.

EQUAL CONTRIBUTION

K.F. and A.R. contributed equally to this work.

SUPPORT

Support for the Diamandis Laboratory and trainees is provided by the Princess Margaret Cancer Foundation, an American Society of Clinical Oncology Career Development Award, and The Brain Tumour Charity Expanding Theories Research Grant (No. GN-000560 [P.D.]). Support for the Princess Margaret Cancer Centre-Ontario Institute for Cancer Research Translational Genomics Laboratory is provided by the Ontario Institute for Cancer Research and Princess Margaret Cancer Foundation.

AUTHOR CONTRIBUTIONS

Conception and design: Kevin Faust, Ugljesa Djuric, Phedias Diamandis

Collection and assembly of data: Kevin Faust, Adil Roohi, Anglin Dent, Phedias Diamandis

Data analysis and interpretation: Kevin Faust, Adil Roohi, Alberto J. Leon, Emeline Leroux, Andrew J. Evans, Trevor J. Pugh, Sangeetha N. Kalimuthu, Phedias Diamandis

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](http://OpenPayments.org)).

Andrew J. Evans

Honoraria: AstraZeneca

Trevor J. Pugh

Honoraria: Merck

Consulting or Advisory Role: Chrysalis Biomedical Advisors, Axiom Healthcare Strategies

Research Funding: Roche

Patents, Royalties, Other Intellectual Property: Hybrid-capture sequencing for determining immune cell clonality

No other potential conflicts of interest were reported.

REFERENCES

1. Esteva A, Kuprel B, Novoa RA, et al: Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115-118, 2017 [Erratum: *Nature* 546:686, 2017]
2. Yu K-H, Zhang C, Berry GJ, et al: Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 7:12474, 2016
3. Campanella G, Hanna MG, Geneslaw L, et al: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 25:1301-1309, 2019
4. Hollon TC, Pandian B, Adapa AR, et al: Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. *Nat Med* 26:52-58, 2020

5. Coudray N, Ocampo PS, Sakellaropoulos T, et al: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 24:1559-1567, 2018
6. Guérin J, Gibaru O, Thiery S, et al: CNN features are also great at unsupervised classification. 2017. <http://arxiv.org/abs/1707.01700>
7. Razavian AS, Azizpour H, Sullivan J, et al: CNN features off-the-shelf: An astounding baseline for recognition. <http://arxiv.org/abs/1403.6382>
8. Yosinski J, Clune J, Bengio Y, et al: How transferable are features in deep neural networks? 2014. <https://arxiv.org/abs/1411.1792>
9. Donahue J, Jia Y, Vinyals O, et al: DeCAF: A deep convolutional activation feature for generic visual recognition. 31st Int Conf Mach Learn ICML 2014 2:988-996, 2013
10. Faust K, Bala S, van Ommeren R, et al: Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning. *Nat Mach Intell* 1:316-321, 2019
11. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. 2014. <http://arxiv.org/abs/1409.1556>
12. Linehan WM, Spellman PT, Ricketts CJ, et al: Comprehensive molecular characterization of papillary renal-cell carcinoma. *N Engl J Med* 374:135-145, 2016
13. Creighton CJ, Morgan M, Gunaratne PH, et al: Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499:43-49, 2013
14. Samarasinghe H, Gianduzzo T, Delahunt B: The ISUP system of staging, grading and classification of renal cell neoplasia. *J Kidney Cancer VHL* 1:26-39, 2014
15. Amin MB, Amin MB, Tamboli P, et al: Prognostic impact of histologic subtyping of adult renal epithelial neoplasms: An experience of 405 cases. *Am J Surg Pathol* 26:281-291, 2002
16. Faust K, Xie Q, Han D, et al: Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. *BMC Bioinformatics* 19:173, 2018
17. BitBucket: <https://bitbucket.org/diamandislabbii/faust-unsupervised-2020>
18. BitBucket: <https://bitbucket.org/diamandislabbii/faust-feature-vectors-2019>
19. GitHub: <https://github.com/oicr-gsi/modelTsne>
20. Klatte T, Rossi SH, Stewart GD: Prognostic factors and prognostic models for renal cell carcinoma: A literature review. *World J Urol* 36:1943-1952, 2018
21. Gill S, Kauffman EC, Kandel S, et al: Incidence of clear cell papillary renal cell carcinoma in low-grade renal cell carcinoma cases: A 12-year retrospective clinicopathologic study from a single cancer center. *Int J Surg Pathol* 24:207-212, 2016
22. Cornejo KM, Dong F, Zhou AG, et al: Papillary renal cell carcinoma: Correlation of tumor grade and histologic characteristics with clinical outcome. *Hum Pathol* 46:1411-1417, 2015
23. Sarwar S, Dent A, Faust K, et al: Physician perspectives on integration of artificial intelligence into diagnostic pathology. *npj Digit Med* 2:28, 2019
24. Wang X, Lu L, Shin HC, et al: Unsupervised joint mining of deep features & image labels for large-scale radiology image categorization & scene recognition. *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*. Institute of Electrical and Electronics Engineers Inc., 2017, pp 998-1007
25. Kim G, Sigal L, Xing EP: Joint summarization of large-scale collections of web images and videos for storyline reconstruction, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE Computer Society, 2014, pp 4225-4232
26. Yong JL, Grauman K: Shape discovery from unlabeled image collections. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Institute of Electrical and Electronics Engineers, 2010, pp. 2254
27. Gong Y, Pawlowski M, Yang F, et al: Web scale photo hash clustering on a single machine. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2015, pp 19-27
28. Flickner M, Sawhney H, Niblack W, et al: Query by Image and Video Content: The QBIC System. *IEEE Comput* 28:23-32, 1995
29. TCGA Data Portal: <https://tcga-data.nci.nih.gov/tcga/>