Prediction of TMB and an immune gene expression signature from melanoma diagnostic H&E images

Jason T Dean*

* Corresponding author

E-mail: jtdean@gmail.com

Affiliations: Independent

Abstract

Recently developed immune checkpoint inhibitors (ICI) have dramatically improved the treatment options and outcomes for cancer patients. Despite their remarkable success, only a subset of patients responds to ICI therapy, and identifying patients most likely to respond is an area of intense research interest. In this work I developed a machine learning classifier to predict non-synonymous tumor mutational burden (TMB) and a previously identified immune gene expression signature, two metrics that have been shown to correlate with ICI response, from melanoma digital hematoxylin and eosin (H&E) stain images from The Cancer Genome Atlas (TCGA). A deep convolutional neural network (CNN) was used to extract features from stain images via transfer learning and a logistic regression classifier was trained to predict TMB and a gene expression marker. Using this machine learning framework I achieved an average area under the curve (AUC) of 0.65 and 0.76 for classification of TMB and immune gene expression signature, respectively. This study contributes to the growing body of evidence that H&E images can be used to interrogate the mutational and transcriptional landscape of tumors.

Introduction

One way that cancer is able to evade the host immune system is through aberrant expression of immune checkpoint proteins [1,2]. Monoclonal antibodies (mAb) targeting programmed cell death protein 1 (PD1) and cytotoxic T-lymphocyte-associated antigen 4 (CTLA4), termed ICIs, can release T-cells from this inhibition and allow them to identify and kill cancer cells [3–5]. Ipilimumab and nivolumab are mAbs targeting CTLA4 and PD1, respectively, and have been approved for treatment of metastatic melanoma [6], one of the first cancers treated with ICIs. When effective, ICIs dramatically improve patient treatment and clinical outcomes [7]. Despite their great therapeutic success, only a fraction of patients respond to checkpoint inhibition. For example, ipilimumab and nivolumab were found to generate complete or total responses in only 19% and 43.7% of treated melanoma patients [8]. Therefore, identifying biomarkers that discriminate patients that are most likely to respond to ICIs is an area of intense active research.

Numerous studies have demonstrated that TMB is positively correlated with ICI response [9–11]. Since non-synonymous TMB is positively correlated with neoepitope production it is thought that tumors with high TMB are more susceptible to T-cell elimination [12]. Additionally, unique tumor multi-gene expression signatures that discriminate ICI responders from non-responders have been determined [12–14]. Although the genes contained in these signatures differ, they generally aim to capture markers of T-cell dysfunction and genes that are upregulated and implicated in cytolytic activity following CD8+ T-cell activation.

As described above, quantifying the mutational landscape and transcriptional profile of tumors has led to breakthroughs and identification of markers that may help identify patients best suited for ICI. H&E staining is the primary method of diagnosing many cancers [15] and digital H&E stain slides are a potentially rich source of data. Extracting quantitative information from digital H&E slides is challenging from a technical perspective because of the large size of the images (can be as large as 1e6 x 1e6 pixels), the heterogeneity of a tissue section, and batch to batch

differences in stain intensity, to name a few. These technical challenges are particularly suited to deep learning-based classifiers. Deep learning algorithms, when provided sufficient training data and computational resources, have demonstrated remarkable ability to learn complex non-linear relationships from high-dimensional feature spaces across diverse areas [16–18]. However, de novo construction of deep learning networks requires considerable domain expertise and end to end training of modern deep learning classifiers typically requires on the order of tens of millions of independent training observations and access to compute infrastructures with GPUs. Transfer learning, the process of using pre-trained machine learning models to solve different, but similar tasks than which they were originally intended has shown remarkable success in image classification [19]. Transfer learning, in the context of image classification, involves using a deep learning network that is pre-trained on a set of images to extract 'high-level' features from a new set of images. These new features can then be used to train more computationally manageable classifiers. Thus, transfer learning, when feasible, provides the advantages of avoiding long, computationally expensive training procedures and allows expertly constructed and trained deep-learning classifiers to be re-purposed.

Despite these challenges, significant progress has been made in recent years using digital H&E images for classification of non-small cell lung cancer [20] , predicting patient outcomes in colorectal cancer [21], predicting SPOP mutation state in prostate cancer [22], and predicting high and low TMB bladder cancer subgroups [23]. Digital pathology promises to transform pathology and is becoming increasingly more important in pathology workflows, and I refer you to the following detailed reviews for an in depth consideration of this exciting field [24,25]

In this work I evaluated the feasibility of using digital H&E images from TCGA from melanoma patients to predict non-synonymous TMB and a previously identified immune gene expression signature [13]. I extracted high level features from each image using a pre-trained CNN and transfer learning and evaluated multiple pooling strategies to condense image patch level features. Pooled features were used to train logistic regression classifiers and AUCs of 0.65 and 0.76 for classification of TMB and immune gene expression signature were achieved.

Materials and methods

All code, data, and files required to reproduce this work can be found here:
https://github.com/JTDean123/immunoHE

Data availability

Skin cutaneous melanoma (TCGA-SKCM) RNA expression (HT-Seq FPKM) and image data (Diagnostic slide) were downloaded from the NCI Genomic Data Commons (GDC) [26]. TMB data was obtained from [27].

TMB and immune signature label assignment

Binary labels were assigned to each diagnostic image according to their TMB or immune signature values. It has recently been shown that melanoma patients with higher non-synonymous TMB levels (top 20%) respond better to ICI [9], therefore diagnostic images were classified as either low or high TMB using a cutoff of (number of non-synonymous mutations)/(Mb of exome) equal to 20, and 21% of the diagnostic images were found to have a TMB higher than this cutoff. An immune signature was calculated using genes that were previously identified [13]. For each diagnostic image the mean FPKM of CD247, CD2, CD3E, GZMH, NKG7, PRF1, and GZMK was determined. The mean was then log transformed and min-max scaled [28]. Diagnostic slides were classified as having either a low or high immune signature value using an arbitrarily defined cutoff of 0.60.

Image Processing

An image processing pipeline was constructed as shown in Fig 1. First, a diagnostic image was downloaded from the GDC. Raw, full-sized diagnostic images are not conducive to machine learning because of the large number of pixels per image, so a single diagnostic image at 20X magnification was first converted to 'patches' with pixel size 512x512 using OpenSlide [29]. Patches were only retained if the mean pixel intensity value across the three color channels was greater than 200 and if the standard deviation of pixel intensity across the three color channels was greater than 160. These filter cutoff values were calibrated by manually inspecting a subset of the diagnostic images and patches and finding values that filtered out patches that were homogenous in color and patches that contained mostly background and low amounts of tissue. Next, min(number of images, 1000) patches were retained for further analysis. Images were next normalized to standardize color intensities as previously described [30] (Fig 2) using a single chosen representative patch. This image processing pipeline yielded 372,881 normalized image patches from 423 diagnostic image slides.

Fig 1. Pipeline for Conversion of a Diagnostic H&E Slide to a Feature Vector. Diagnostic H&E slides were downloaded from GDC and tiled to create image patches with pixel size 512x512. Individual patches with low pixel variance and those with high amounts of background were removed and the retained patches were color normalized. High level features were next extracted from each individual patch using the pre-trained inception v3 model and patches from a single diagnostic image were combined by feature pooling.

Fig 2. Patch Color Normalization. Diagnostic slides were found to differ substantially in saturation, amount of background, and color intensity. To normalize patches from all diagnostic slides individual patches were normalized to the patch designated as 'Reference'.

Transfer learning

High level features were extracted from each normalized image patch using transfer learning and the Google developed inception v3 CNN trained on the ImageNet data set [31]. First, image patches were resized to a pixel size of 299x299 for consistency with the inception v3 framework. The last model layer (prior to the Softmax layer) from inception v3 was extracted for each patch, resulting in a 2048 features per image patch (Fig 1). Patch level feature vectors from individual

diagnostic slides were collated and a condensed to a single 2048 dimensional pooled vector by either max, average [32], or p-norm pooling [33].

Principal component analysis and logistic regression training

Principal component analysis (PCA) was used to reduce the dimensionality of the pooled feature vectors for each diagnostic image and the optimal number of principal components (PCs) was hyper-parameter tuned using 10-fold cross validation. First, diagnostic images were split 80:20 into a training and held out test set. Next, the training data features were scaled to a mean of zero and a variance of one, PCA was performed on the training data, and the training data was projected on to the PC coordinate space. The test data set features were then scaled by subtracting the mean of the training features and diving by the standard deviation. The scaled test data was then projected on to the PC coordinate system derived from the training data. The PCA reduced training data was used to train a L2 regularized logistic regression model to predict either TMB or immune signature using between 10 and 300PCs, in multiples of 10. The trained logistic regression model was used to predict the label of the held out PCA reduced test data, using the same number of PCs as was used for the training, and AUC was calculated.

Software and code

All analysis was performed with python version 3.6.8 using the following libraries – numpy, pandas, and sklearn. Plotting was performed with python using matplotlib and inception v3 was used for feature extraction using keras version 2.2.4

All of the code used for this work has been made publicly available:
https://github.com/JTDean123/immunoHE

Results and discussion

Diagnostic slide feature extraction

There are numerous technical challenges in converting a H&E diagnostic image into a feature vector that is suitable for downstream classifier training. In particular, diagnostic images are large (1e6 x 1e6 pixels), the tissue section does not completely cover the slide, there is substantial image to image heterogeneity in stain intensity and background, and a significant fraction of the image is likely to be uninformative from a diagnostic perspective. A pipeline was developed to attempt to overcome these issues and convert a H&E diagnostic slide image into a feature vector (Fig 1). TCGA-SKCM diagnostic images were first downloaded from the GDC [26]. Images were next tiled into patches with pixel size 512x512, thus converting a single large diagnostic image into a collection of smaller images. Patches that were mostly background or contained low amounts of tissue were discarded (Materials and methods). Each image patch was next normalized [30] using a randomly chosen single patch from a single image (Fig 2). This process resulted in 372,881 normalized image patches from 423 diagnostic image slides.

Following tiling and normalization, high level features were extracted from each patch using the inception v3 CNN trained on the ImageNet data set [31]. Numerous attempts were made to fully train the inception v3 CNN model. Training the CNN from scratch using either randomly initialized or previously trained model weights resulted in rapid overfitting despite extensive image augmentation. Additionally, I evaluated training by fine-tuning by adding a fully connected layer and training only this final layer, however this model resulted in poor generalization to the test data set. One of the main challenges in generating features from a H&E diagnostic slide is that not all patches are informative and not all regions in an informative patch are informative. In an attempt capture both informative patches and informative regions a feature pooling strategy, inspired by the work of Xu et al [33], was constructed. The output from the last layer of inception v3, a vector of size 2048, was extracted from each normalized image patch. Extracted feature vectors from each patch of a single diagnostic image were then concatenated to build, separately for each diagnostic image, a (number of patches) x (2048) feature matrix. Three different pooling strategies (max, average, and p-norm) were evaluated to condense the diagnostic feature matrix to a single pooled feature vector of size 2048 (Fig 1). Pooling provided two advantages: 1) not every slide had the same number of patches, and pooling allowed an arbitrary number of patches to be converted to a standardized feature vector and 2) it increased generalization of model performance to test data.

TMB and immune signature determination

Gene expression values (FPKM) corresponding to each diagnostic image were obtained from the GDC and converted to an immune signature score between zero and one as described in Materials and Methods (Fig 3A and 3B). This immune signature, described by Davoli et al. [13], is comprised of the expression values of seven genes (CD247, CD2, CD3E, GZMH, NKG7, PRF1, and GZMK) and captures the expression of markers of cytolytic activity and of parts of the T-cell receptor. In this work, Davoli et al. discovered that patients with elevated tumor aneuploidy had a worse prognosis with ICI treatment. By stratifying tumor samples according to aneuploidy Davoli et al. further showed that aneuploidy is negatively correlated with the immune signature score. I reasoned that developing techniques to predict the immune signature score from additional data sources may complement and build upon these earlier works. Here I stratified tumor samples into low and high immune signature score values using an arbitrary cutoff of 0.6. It has been shown that non-synonymous TMB is positively correlated with ICI response [9–11]. TMB corresponding to each tumor diagnostic image was obtained from previous work [27] and diagnostic images were labeled as either low or high TMB using a cutoff of 20 (Fig 3C). I did not find a significant correlation between TMB and immune signature (Fig 3D).

Fig 3. Immune Signature and TMB for Melanoma (TCGA-SKCM) Diagnostic H&E Slides. (A) The immune signature score is comprised of the FPKM expression values of seven genes. (B) Distribution of immune signatures across diagnostic slides. Diagnostic slides were classified as low or high immune signature using a cutoff of 0.6. (C) Non-synonymous TMB (non-silent TMB) corresponding to the diagnostic slides. Diagnostic slides were classified as having low or high TMB using a cutoff of 20. (D) TMB is not correlated with Immune Signature.

Evaluation of feature pooling strategies

Feature extraction with the inception v3 model results in a feature matrix with dimensions (number of patches) x (2048) for each diagnostic image. Max, average, and p-norm pooling strategies were evaluated to convert a feature matrix into a feature vector that represents a single diagnostic slide (Fig 1). Eighty percent of the diagnostic slides (338) were used for training and the remaining 20% were used for testing (85). PCA was used for dimensionality reduction since the dimension of the feature space (2048) was significantly larger than the number of diagnostic slides used for training. It was not obvious how many principal components should be retained, so a range between ten and 300 were evaluated (see Materials and Methods for details). Interestingly, I found that retaining 30 principal components yielded the best average classifier performance, determined from AUC, for both TMB and immune signature classification (Fig 4A and 4B), however the optimal pooling strategy differed. In particular, I found that max and average pooling were optimal for TMB and immune signature classification, respectively (Fig 4A and 4B).

Fig 4. Determination of Optimal Number of Principal Components and Pooling Strategy. A patch feature matrix from a diagnostic slide was converted to a feature vector by pooling. Diagnostic slides were split 80/20 into training and test sets and PCA was performed to reduce the dimensionality of the feature space. For each split, between 10 and 300 principal components were retained and a logistic regression classifier was fit to each reduced dimension diagnostic slide feature matrix and AUC was calculated by evaluating predictions on the held-out test data set. The splitting of the slides was performed 10 times and each point represents the mean for (A) TMB and (B) immune signature prediction.

Logistic regression model training

Logistic regression classifiers were next trained to predict either the immune signature or TMB using the optimal number of principal components as described above. The diagnostic images were randomly split 10 times into training and test sets and a logistic regression classifier was trained on the pooled feature vectors to predict either TMB or the immune signature score (low or high) (Fig 5A). As shown in Fig 5A, I found that max and average pooling were the optimal pooing strategies for TMB and immune signature predictions, using AUC to evaluate model performance. Using the optimal number of principal components and pooling strategy for each classification objective an average ROC curve was constructed. The average AUC for TMB classification was 0.65 and for immune signature was 0.76 (Fig 5B, C). An advantage of pooling and PCA during training is a possible increase in model generalization, however a downside is that information is potentially lost. Indeed, we found that for the 30PCs retained for TMB and immune signature classification explained 56% and 91% of the variance in the input data, respectively S1 Fig.

Fig 5. Logistic Regression Model Performance using a 30 Principal Component Reduced Feature Space. (A) Logistic regression model performance, as determined from AUC, for max, average, and p-norm feature pooling using 30 principal components. Each data point represents AUC from one random split of the diagnostic slides into training and testing data sets. (B) and

(C) Average AUC from 10 random train test splits of the diagnostic images using optimal pooling and principal component analysis decomposition strategies. Grey lines represent the AUC obtained from each split of the and solid (blue – TMB (B) , red – immune signature (C)) the mean AUC across the ten random splits.

S1 Fig. Cumulative Proportion of Variance Explained by Number of Principal Components. After max pooling TMB and average pooling immune signature feature data 30PCs were found to explain 56% and 91% of the variance.

Conclusions

In this work I show that diagnostic H&E images are a data source that can be used to investigate the mutation and transcriptional profile of tumors. In particular, I show that TMB and an immune gene expression signature can be predicted from a diagnostic H&E image using transfer learning. Since most solid tumor cancers are diagnosed with using H&E slides, these slides represent a rich source of data that is readily available for most cancer patients. Here I show that TMB and an immune signature score can be predicted with an AUC of 0.65 and 0.76, respectively.

Despite these results this study has limitations that are important to consider. First, the performance has not been evaluated on an independent, held-out validation data set, the gold-standard for machine learning model performance. Second, the slides evaluated in this study have been FFPE preserved, and it is unknown how model performance will be impacted by other preservation techniques like freezing. Third, the TCGA slides were imaged on an Aperio Imagescope, and it is unknown how the strategies described in this work will perform with data generated on other imaging instruments. Finally, and most important, it is critical to note that the classification strategy devised in this work merely classifies images into high/low TMB and gene expression signature groups, thus it is unknown if the classifiers developed here will provide any clinical benefit.

In this study I only evaluated melanoma diagnostic H&E slides, however future work evaluating additional cancer types will allow for further development of the approach described here. Although the performance described here is likely not usable in a clinical setting, it is possible that the addition of more training data will drive increases in performance. Furthermore, additional multi-gene expression panels have been found that correlate with ICI response [12–14], and I hope that the feature extraction and machine learning pipeline I describe here will allow for their evaluation. In conclusion, I hope that the methodology described here for extracting genomic information from H&E images will contribute to this rapidly growing field and hopefully someday improve patient outcomes.

References

1. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018. doi:10.1126/science.aar4060

2. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nature Reviews Cancer. 2012. doi:10.1038/nrc3239
3. Wolchok JD, Kluger H, Callahan MK, Postow MA, Rizvi NA, Lesokhin AM, et al. Nivolumab plus Ipilimumab in Advanced Melanoma. N Engl J Med. 2013; doi:10.1056/NEJMoa1302369
4. Honjo T, Ishida M, Iwai Y, Tanaka Y, Okazaki T, Minato N. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. Proc Natl Acad Sci. 2002; doi:10.1073/pnas.192461099
5. Leach DR, Krummel MF, Allison JP. Enhancement of antitumor immunity by CTLA-4 blockade. Science (80-). 1996; doi:10.1126/science.271.5256.1734
6. Jenkins RW, Barbie DA, Flaherty KT. Mechanisms of resistance to immune checkpoint inhibitors. Br J Cancer. 2018; doi:10.1038/bjc.2017.434
7. Shields BD, Mahmoud F, Taylor EM, Byrum SD, Sengupta D, Koss B, et al. Indicators of responsiveness to immune checkpoint inhibitors. Sci Rep. 2017; doi:10.1038/s41598-017-01000-2
8. Bastholt L, McArthur GA, Hill A, Grossmann K, Hodi FS, Ferrucci PF, et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med. 2015; doi:10.1056/nejmoa1504030
9. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nature Genetics. 2019. doi:10.1038/s41588-018-0312-8
10. Goodman AM, Frampton GM, Bazhenova L, Stephens PJ, Kurzrock R, Miller V, et al. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. Mol Cancer Ther. 2017; doi:10.1158/1535-7163.mct-17-0386
11. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science (80-). 2015; doi:10.1126/science.aaa1348
12. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015; doi:10.1016/j.cell.2014.12.033
13. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science (80-). 2017; doi:10.1126/science.aaf8399
14. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018; doi:10.1038/s41591-018-0136-1
15. Fischer AH, Jacobson KA, Rose J, Zeller R. Hematoxylin and eosin staining of tissueand cell sections. Cold Spring Harb Protoc. 2008; doi:10.1101/pdb.prot4986
16. Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, et al. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat Genet. 2019; doi:10.1038/s41588-019-0420-0
17. Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nature Medicine. 2019. doi:10.1038/s41591-018-0279-0
18. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol. 2015;

doi:10.1038/nbt.3300
19. Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: An astounding baseline for recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2014. doi:10.1109/CVPRW.2014.131
20. 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. Nat Med. 2018; doi:10.1038/s41591-018-0177-5
21. Bychkov D, Linder N, Turkki R, Nordling S, Kovanen PE, Verrill C, et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep. 2018; doi:10.1038/s41598-018-21758-3
22. Schaumberg AJ, Rubin MA, Fuchs TJ. H&E-stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer. bioRxiv. 2017; doi:10.1101/064279
23. Xu H, Park S, Lee SH, Hwang TH. Using transfer learning on whole slide images to predict tumor mutational burden in bladder cancer patients. bioRxiv. 2019; doi:10.1101/554527
24. Niazi MKK, Parwani A V., Gurcan MN. Digital pathology and artificial intelligence. The Lancet Oncology. 2019. doi:10.1016/S1470-2045(19)30154-8
25. Zarella MD, Bowman D, Aeffner F, Farahani N, Xthona A, Absar SF, et al. A practical guide to whole slide imaging a white paper from the digital pathology association. Archives of Pathology and Laboratory Medicine. 2019. doi:10.5858/arpa.2018-0343-RA
26. Grossman RL, Heath AP, Ferretti V, Varmus H, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016; doi:10.1056/NEJMp1002530
27. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The Immune Landscape of Cancer. Immunity. 2018; doi:10.1016/j.immuni.2018.03.023
28. Cao XH, Stojkovic I, Obradovic Z. A robust data scaling algorithm to improve classification accuracies in biomedical data. BMC Bioinformatics. 2016; doi:10.1186/s12859-016-1236-x
29. Satyanarayanan M, Goode A, Gilbert B, Jukic D, Harkes J. OpenSlide: A vendor-neutral software foundation for digital pathology. J Pathol Inform. 2013; doi:10.4103/2153-3539.119005
30. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009. 2009. doi:10.1109/ISBI.2009.5193250
31. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2015. doi:10.1109/CVPR.2015.7298594
32. Boureau Y-L, Ponce J, LeCun Y. A Theoretical Analysis of Feature Pooling in Visual Recognition. The 27th International Conference on Machine Learning. 2010. doi:10.1.1.170.864
33. Xu Y, Jia Z, Wang LB, Ai Y, Zhang F, Lai M, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinformatics. 2017; doi:10.1186/s12859-017-1685-x