

Deep Learning Features Encode Interpretable Morphologies within Histological Images

Ali Foroughi pour

The Jackson Laboratory for Genomic Medicine

Brian White

The Jackson Laboratory for Genomic Medicine

Jonghanne Park

The Jackson Laboratory for Genomic Medicine

Todd Sheridan

The Jackson Laboratory for Genomic Medicine

Jeffrey Chuang (✉ Jeff.Chuang@jax.org)

The Jackson Laboratory for Genomic Medicine <https://orcid.org/0000-0002-3298-2358>

Article

Keywords: Convolutional neural networks, hematoxylin and eosin, mones

Posted Date: September 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-865341/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 **DEEP LEARNING FEATURES ENCODE INTERPRETABLE**
2 **MORPHOLOGIES WITHIN HISTOLOGICAL IMAGES**

3 BY ALI FOROUGHI POUR¹ AND BRIAN WHITE¹ AND JONGHANNE PARK¹ AND TODD B.
4 SHERIDAN^{1,2} AND JEFFREY H. CHUANG^{1,3}

5 ¹*The Jackson Laboratory for Genomic Medicine, 10 Discovery Dr. Farmington, CT, 06032* ali.foroughipour@jax.org
6 jeff.chuang@jax.org

7 ²*Department of Pathology, Hartford hospital, 80 Seymour St, Hartford, CT, 06106*

8 ³*Department of Genetics and Genome Sciences, UCONN Health, Farmington, CT, 06032*

9 **ABSTRACT.** Convolutional neural networks (CNNs) are revolutionizing
10 digital pathology by enabling machine learning-based classification of a
11 variety of phenotypes from hematoxylin and eosin (H&E) whole slide images
12 (WSIs), but the interpretation of CNNs remains difficult. Most studies have
13 considered interpretability in a post hoc fashion, e.g. by presenting example
14 regions with strongly predicted class labels. However, such an approach does
15 not explain the biological features that contribute to correct predictions.
16 To address this problem, here we investigate the interpretability of H&E-
17 derived CNN features (the feature weights in the final layer of a transfer-
18 learning-based architecture), which we show can be construed as abstract
19 morphological genes (“mones”) with strong independent associations to
20 biological phenotypes. We observe that many mones are specific to individual
21 cancer types, while others are found in multiple cancers especially from
22 related tissue types. We also observe that mone-mone correlations are strong
23 and robustly preserved across related cancers. Importantly, linear mone-based
24 classifiers can very accurately separate 38 distinct classes (19 tumor types
25 and their adjacent normals, AUC=97.1% ± 2.8% for each class prediction),
26 and linear classifiers are also highly effective for universal tumor detection
27 (AUC=99.2% ± 0.12%). This linearity provides evidence that individual
28 mones or correlated mone clusters may be associated with interpretable
29 histopathological features or other patient characteristics. In particular, the
30 statistical similarity of mones to gene expression values allows integrative
31 mone analysis via expression-based bioinformatics approaches. We observe
32 strong correlations between individual mones and individual gene expression
33 values, notably mones associated with collagen gene expression in ovarian
34 cancer. Mone-expression comparisons also indicate that immunoglobulin
35 expression can be identified using mones in colon adenocarcinoma and that
36 immune activity can be identified across multiple cancer types, and we verify
37 these findings by expert histopathological review. Our work demonstrates that
38 mones provide a morphological H&E decomposition that can be effectively
39 associated with diverse phenotypes, analogous to the interpretability of
40 transcription via gene expression values.

41 **1. Introduction.** Deep learning has become an important methodology for analyzing
42 biomedical images, and in particular for analyzing hematoxylin and eosin (H&E) stained
43 whole slide images (WSIs). Deep neural networks have achieved classification accuracies
44 higher than classical machine learning models [1]. However, they are black-boxes that do not
45 directly reveal the morphological features they associate with labels, a significant concern
46 for mechanistic analysis and clinical decision making [2]. Identification of biologically
47 meaningful morphological features may be confounded by image artifacts [3], such as
48 blurring, noise, and lossy image compression [4]. Tissue damage, image quality, and dataset-
49 specific artifacts have also been suggested to affect feature representation and prediction

50 accuracy of neural networks (see [1], [5], and [6]). Given the impact of such artifacts on deep
51 learning-based predictors, it is of critical importance to be able to decompose CNNs into
52 features that can be biologically interpreted.

53 The majority of models for visualizing, analyzing, and interpreting CNNs reveal “where”
54 a network is “looking” to make its prediction, rather than revealing “what” information in
55 the region of interest is important. Some methods output pixel patterns that affect the value
56 of a neuron in a deep network (see [7]). However, such techniques tend to output different
57 predictive regions, can be difficult to validate, or have been suggested to be “fragile”, i.e.
58 extremely sensitive to small perturbations of the image [8]. Optimizing conventional deep
59 learning techniques to identify regions informative of class labels is a current theme in
60 digital pathology (see [9, 10] for examples). Other recent works have focused on visualizing
61 individual deep learning features as heatmaps [11].

62 Unlike natural image analysis [12], biomedical image analysis is complemented by
63 additional data modalities, such as multiplexed imaging, single cell and bulk sequencing,
64 and clinical information [13, 14]. These data may aid in interpreting the deep feature
65 representations of the H&E slide. However, models integrating these diverse modalities are
66 needed. The feasibility of doing so is supported by work establishing the connection between
67 modalities, for example by using CNNs to predict expression values of specific genes from
68 H&E images (see [15, 16, 17] for examples). Because of the architectural complexity of
69 CNNs, it has often been assumed that CNN-based decompositions of images into features
70 are not interpretable. However, there has been little empirical study of this question, e.g. by
71 testing whether CNN-derived features are correlated with simple biological features such as
72 gene expression values.

73 In this work, we investigate the interpretability of CNN-derived image features, which
74 we conceptualize as morphological genes (“mones”). We find that mones have strong
75 linear associations with phenotypic features, making them directly interpretable, which
76 we demonstrate in several analyses. We demonstrate that many mones can distinguish
77 cancer tumors from adjacent normal slides. These mones can be linearly combined for
78 reliable prediction of both pan-cancer and tissue-specific cancer phenotypes. Mone-mone
79 correlation analysis identifies clusters of highly correlated mones within cancer types,
80 and these correlations are strongly preserved among cancers from related tissues. Mone
81 values resemble gene expression data, allowing immediate use of many interpretable
82 bioinformatics tools and machine learning models to identify the underlying biology of
83 morphologies encoded by CNNs. For example, integrative mone-gene expression correlation
84 analysis reveals that collagen content and immune infiltration are linearly associated with
85 morphologies encoded by mones in several cancer types, and we confirm these relationships
86 by expert histopathological review. Our studies demonstrate the power of mones for
87 computationally deconstructing cancer images into interpretable biological features.

88 **2. Results.** We analyzed the InceptionV3 [18] features of tiles from whole slide images
89 of The Cancer Genome Atlas (TCGA) [1] for 19 cancers (see Supplementary File 1 for the
90 full list). We used features derived from this architecture because predictive models based
91 on Inception have shown high accuracy for identifying phenotypes in prior studies [1]. We
92 hereafter denote each of the 2048 outputs of the global average pooling layer of the Inception
93 V3 network (see Supplementary Figure 1) as mones (morphological genes). We use this
94 terminology because mones have analogies to genes with individual expression values. Tile
95 level mones can be combined to construct slide level mones (see methods). Unless otherwise
96 stated, in the studies below “mone” refers to a slide level characterization (see Supplementary
97 Figure 1).

98 2.1. *Individual mones differentiate phenotypes.* We first investigated to what extent
99 individual mones can differentiate phenotypes, focusing on TCGA tumor/normal slide
100 comparisons. We initially identified individual mones with significant differences in distribution
101 between breast cancer (BRCA) tumor and adjacent normal slides (>1800 mones were
102 statistically significant with $FDR < 5\%$). A clustermap of the top 100 such mones was
103 able to clearly separate these classes (see methods, Figure 1a, and Supplementary Figure
104 2, clustermap AUC=89%, rand score =96%, and adjusted rand score=85%). These results
105 were typical of mone behaviors in many tumor types – in any given tumor type, many mones
106 were able to separate frozen tumor from normal slides (see methods and Supplementary
107 File 1). We applied several statistical methods to test the robustness of such mones. In each
108 cancer, at least 40% of mones were statistically significant irrespective of the statistical test
109 used ($FDR = 5\%$, see Supplementary File 1), and 75.7% of mones significant by at least
110 one method were significant by all tests (see methods, see Figure 1b). A smaller subset
111 of these mones showed strong effect sizes, as identified by optimal Bayesian filter (OBF)
112 [19] statistics (see Methods). 22% of all mone-cancer pairs met this criterion based on
113 distributional differences between tumor and normal slides (see Supplementary Figure 3).

114 As an example, mone 983 is a tumor marker with a distributional difference between
115 frozen tumor and adjacent normal breast cancer (BRCA) slides (Figure 1c). It also behaves
116 similarly in several other tumor types (Supplementary Figure 4). It strongly correlates with
117 cell density in frozen BRCA slides (see methods, see Figures 1d-e, Pearson $r=0.69$, p -
118 value $<1e-200$, see Supplementary File 2) and significantly though with moderate magnitude
119 in FFPE BRCA tumor slides (Pearson $r=0.24$, p -value $=1.9e-13$). Mone 983 has higher
120 correlation with cell density in FFPE LUAD slides (Pearson $r=0.61$, p -value $=2.2e-34$) than
121 frozen LUAD slides (Pearson $r=0.28$, p -value $=4.1e-18$, see Supplementary File 2).

122 To further test whether mones exhibit consistent behavior in different cancer types, we
123 analyzed the four cancer families of [1]: pan-GYN, pan-KIDNEY, pan-LUNG, and pan-
124 GI (see Supplementary File 3 for cancers in each family). More than half of the mones
125 with distributional differences in each cancer type also have distributional differences in
126 all cancers of the family (see Supplementary File 3). Although cancer-specific mones are
127 uncommon, such mones still show clear distributional differences between tumor and normal
128 (see Supplementary Figure 5). Interestingly, most mones distinguishing tumor from normal
129 in both LUAD and LUSC also distinguish between the LUAD and LUSC cancers (see
130 Supplementary File 3), suggesting quantitative values are important. Individual mones can
131 also distinguish frozen from FFPE slides (see Supplementary File 4). We observed the ability
132 of some mones to distinguish between tumor and normal is impacted by differences between
133 frozen and FFPE modalities (421 ± 112 across all cancers, see methods, see Supplementary
134 File 4), though the majority of mones behave similarly in frozen and FFPE.

135 2.2. *Mone clusters provide robust encodings of cancer phenotypes.* We next investigated
136 to what extent mones are independent or encode behaviors together. We did this by
137 calculating pairs of mones that were significantly correlated, for each cancer type. We
138 analyzed this first by restricting to frozen slides (to avoid Simpson's paradox), and second
139 by combining frozen and FFPE slides (to avoid false correlations due to frozen-specific
140 artifacts). The results were generally robust between the two methods—a few cancers
141 had a non-trivial difference in the ratio of correlated mone pairs ascertained by the two
142 methods (ESCA, KICH, SARC, and PRAD $22.3\% \pm 2.22\%$), while the remaining cancers
143 had small differences ($7.8\% \pm 3.9\%$, see Supplementary File 6). Therefore we subsequently
144 analyzed frozen samples only unless otherwise specified. Overall, we found that correlated
145 mones are prevalent among tumor slides (see Figures 1f and 1g, and Supplementary File
146 6). For example, 83.9% and 88.7% of mone-mone pairs are correlated within LUAD and

147 within LUSC, respectively (see methods and Supplementary File 6). We observed similar
148 results for other cancers: $68.8\% \pm 13\%$ of all mone-pairs were statistically significant across
149 cancers (see Supplementary File 6). Remarkably, we observed that pairwise correlations are
150 preserved within cancer families – more than 45% of mone-pairs statistically significant in
151 one cancer are significant in all cancers of the family (pan-GYN, pan-KIDNEY, pan-LUNG,
152 and pan-GI, see Supplementary Figure 6). For example, the mone-mone correlations in lung
153 adenocarcinoma and lung squamous cell carcinoma are nearly identical (Figures 1f and 1g).
154 We also calculated mone-mone correlations in the normal slides associated with each cancer
155 type, hypothesizing that the difference in mone-mone correlations between tumor and normal
156 might be important to distinguishing tumor from normal images. However, these differential
157 mone correlations are weaker and less preserved across cancers (see Supplementary File 6
158 and Supplementary Figure 6).

159 Different cancers can be distinguished by different sets of mones. For example, while
160 mone 983 separates tumor and normal slides of BRCA, it does not differentiate COAD
161 tumor and normal slides (see Supplementary Figure 4). We identified 105 mones that are
162 highly correlated with mone 983 in BRCA (Null: $|r| \leq 0.5$, FDR= 0.1%, see methods),
163 among which 52, 14, and 29 also differentiate tumor from normal slides in COAD,
164 READ, and STAD, respectively (t-test FDR<0.1%). The first principal component of
165 these COAD-overlapping mones (explaining 41% of the variance across COAD samples)
166 strongly correlates with cell density in frozen slides (Cellpose cellularity estimates: Pearson
167 $r=0.22$, p-value= $2.2e-11$, HoverNet cellularity estimates: Pearson $r=0.43$, P-value= $1.9e-47$,
168 see Methods and Supplementary File 2). Thus the high cellularity in BRCA [20, 21] and
169 COAD [22, 23] involve incompletely overlapping mone sets.

170 *2.3. Linear models of mones can detect and distinguish tumors.* We investigated linear
171 models of mones for predicting phenotypes, as they allow direct interpretation of mone
172 values. We observed that linear models of mones can efficiently distinguish tumor from
173 adjacent normal slides, as well as the cancer type from which they are derived (19
174 cancers, 38 classes, see methods, see Figure 2a-c). We tested two linear models, multi-
175 class linear discriminant analysis [MLDA, One versus Rest (OVR)-AUC= $97.1\% \pm 4.6\%$,
176 see Supplementary File 7] and multinomial logistic regression with LASSO penalty (LR-
177 LASSO, OVR-AUC= $97.1\% \pm 4.2\%$, see Supplementary File 7). MLDA encodes mone
178 patterns indicative of class labels into a low dimensional space (i.e. the number of classes
179 - 1), yielding t-SNE visualizations with improved interpretability over naïve t-SNE (compare
180 Figures 2a,b and Supplementary Figures 7a,b). LR-LASSO, on the other hand, is a linear
181 model based on small mone sets, so its regression coefficients can be interpreted directly with
182 risks incurred by each mone. Although CNN methods typically use difficult-to-interpret fully
183 connected layers at the classification step, we found that efficiently designed linear models
184 can replace fully connected layers while still achieving high prediction AUCs. Combining
185 tumor probabilities of the LR-LASSO classifier we obtained a universal tumor detector with
186 extremely high AUC ($99.2\% \pm 0.12\%$, see methods), out-performing the fully deep learning
187 model of [24] (Reported AUC= 0.95 ± 0.02). Furthermore, LR-LASSO is effective at cross-
188 classification similar to CNNs with fully connected classification layers [1], i.e., LR-LASSO
189 trained to distinguish tumor/normal for one cancer type can distinguish tumor/normal for
190 other cancer types as well (Figure 2d and Supplementary Figure 7c). While the LR-LASSO
191 model has smaller average AUC (0.84) compared to the fully deep learning model of [1]
192 (0.88), logistic regression is more interpretable than a multi-layer perceptron. Our slide level
193 tumor detectors also produce meaningful tile level predictions. Independent review by our
194 pathology team supports most tumor regions having high tumor probability, and most non-
195 tumor regions have low tumor probability in these images, with the cases of misclassification

196 tending to be prediction of non-tumor regions to be tumor (see Supplementary Figure 8 for
197 examples). These results indicate that the LR-LASSO slide level tumor markers are effective
198 at the tile level.

199 *2.4. Mones have interpretable correlations with gene expression.* We next investigated
200 whether mones are linearly associated with gene expression values, as this could provide
201 transcriptional interpretability for mones (see Methods for data stratification and pre-
202 processing). We observed many mones significantly correlated with individual genes. Across
203 five cancers analyzed (OV, COAD, KIRC, LUAD, and LUSC) between 83 (LUSC) and
204 1797 (KIRC) mones were associated with at least one gene, whereas between 332 (LUSC)
205 and 16474 (KIRC) genes were associated with at least one mone (see methods, see
206 Supplementary File 8). We then analyzed several cases of particular interest.

207 *2.4.1. Mones encode collagen content.* We first used unsupervised analysis to study
208 clusters of correlated mone-gene pairs. We identified a cluster of highly correlated mones
209 and collagen genes in OV (Figure 3a). These mone values can be efficiently combined for
210 association with phenotype using PCA (PC-1 explains 63% of the variance). Histopathological
211 review by our pathology team confirmed tiles with high PC-1 values as typically rich in
212 collagen (see Figure 3d and Supplementary Figure 9), and tiles with low PC-1 values
213 as having low collagen but increased cellularity. These mone-gene associations may be
214 clinically relevant, as high expression of collagen genes correlates with multi-drug resistance
215 [25] and poor prognosis in ovarian cancers [26]. Mone 1062—one of the mones in the
216 identified cluster—was additionally highly correlated with ECM2, THBS1 and THBS2, which
217 have been suggested to play a significant role in ovarian cancer drug resistance and metastasis
218 [26, 27, 28].

219 *2.4.2. Mones encode immune infiltration.* Supervised correlation analysis using fixed
220 gene sets can also be used to test if mones encode morphological features associated with
221 a biological phenotype. We tested whether mones encode immune infiltration in pan-GI
222 (COAD, READ, and STAD) and pan-LUNG (LUAD and LUSC) cancers, with immune
223 related gene sets taken from [29] and [30] (see Methods and Supplementary File 9).

224 We identified 19 mones significantly correlating (some positive, some negative) with
225 immune related genes in pan-GI cancers (Fig 3b and Supplementary File 10). These
226 mones have significant correlations with each other ($|r| = 0.56 \pm 0.13$), and 14 mones
227 significantly correlate with more than one immune gene. LAIR1, LCP2, MS4A4A, and
228 CCR1 correlate with 15, 9, 12, and 15 mones, respectively. All 19 mones are significantly
229 correlated with prior TCGA estimates of leukocyte fraction [31] ($FDR < 0.05, |r| = 0.29 \pm 0.07$)
230 and HoverNet estimates of immune cell quantity from the H&E images (see methods,
231 $FDR < 0.05, |r| = 0.45 \pm 0.12$) in COAD. Mone 179 had strong positive correlations with both
232 leukocyte fraction ($r = 0.2$) and HoverNet estimates ($r = 0.65$). Histopathological review
233 confirmed that mone 179 differentiates amongst COAD tiles according to their level of
234 immune infiltration (see Figure 3e and Supplementary Figure 10). PC-1 of the 19 mones
235 has significant correlation with leukocyte fraction ($r = 0.4, p\text{-value} = 2.4e - 35$) and
236 HoverNet estimates ($r = 0.17, p\text{-value} = 2.5e - 17$). Histopathological review validated that PC-
237 1 also differentiates COAD tiles according to differing levels of immune infiltration (see
238 Supplementary Figure 11). Thus PC-1 efficiently combines mones and provides a stronger
239 separation than individual mones.

240 We observed similar correlations, but smaller in magnitude, between mones and local
241 immune cytolytic activity genes in lung cancers (Fig 3c and Supplementary File 10).
242 We observed stronger correlations in LUAD than LUSC (LUAD $|r| = 0.20 \pm 0.05$, LUSC

243 $|r|=0.12 \pm 0.05$). We identified 31 mones significantly correlating (some positive, some
244 negative) with immune related genes in LUAD. 7 mones correlated with at least 3 genes
245 (see Figure 3c). Only one LUAD immune mone did not correlate with lymphocyte fraction
246 ($FDR=0.05$). PC-1 of the 31 mones correlates with lymphocyte fraction ($r=0.28$, p -value=
247 $2.2e-4$). Histopathological review of LUAD slides based on PC-1 suggests LUAD tiles with
248 high PC-1 show a strong tumor infiltrating lymphocyte presence and have inflammation,
249 while tiles with low PC-1 are typically not inflamed and show weaker immune infiltration
250 (Fig 3f, Supplementary Figure 12).

251 *2.4.3. Mones identify immunoglobulin gene expression in highly cellular colon*
252 *adenocarcinoma tumors.* Supervised correlation analysis can also clarify finer behaviors
253 within WSIs. For example, highly cellular COAD tumors typically show high expression
254 of immunoglobulin (IG) genes. We considered the 52 mones correlated with mone 983 and
255 which differentiate COAD tumor from normal (see section 2.2). The PC-1 dimension of these
256 mones significantly correlates with 87 genes in COAD ($FDR<0.05$, see Supplementary File
257 11 for the full gene list), as well with the average expression of these 87 genes ($r=0.33$,
258 p -value= $3.8e-12$). Immunoglobulin (IG) genes dominate this gene set, and we define their
259 average expression as a sample's IG score. B-cells express immunoglobulin, and as expected
260 we observed a statistically significant correlation between the log normalized B-cell estimates
261 of [31] and IG score (p -value= $2.4e-12$). Interestingly, however, B cells comprise only a
262 small fraction of cells in each sample ($0.86\% \pm 1.5\%$), and we did not observe a significant
263 correlation between the B-cell estimates and mone PC-1 (correlation coefficient= -0.009 , P -
264 value= 0.85). Thus mone analysis suggests that IG expression is not due solely to B-cells.
265 This supports recent studies suggesting colon cancer cells themselves express IG genes (see
266 [32]).

267 **3. Discussion.** Mones enable interpretation of deep neural networks because they linearly
268 correlate with phenotypes and gene expression profiles. We have empirically shown that
269 mones efficiently encode strong morphological features that can often be used to replace
270 multi-layer perceptrons with robust and interpretable linear classification models. Moreover,
271 we have demonstrated that integrative mone-gene correlation analysis can identify specific
272 transcriptional processes from images, and verified these through expert pathological review.

273 *3.1. Mones provide image-based interpretability.* Linear models based on mones have
274 several empirical and theoretical strengths for image analysis. Individual mones and small
275 mone clusters directly correlate with phenotypes (see section 2.1 and 2.3), enabling a simpler
276 interpretation of CNNs compared with methods that integrate all CNN features together.
277 Mones decompose arbitrary images into interpretable features, in contrast to approaches that
278 assume non-linearity and which therefore provide interpretability primarily through example
279 regions identified by problem-specific classifiers. A notable advantage of the mone approach
280 is that it does not require a trained classifier.

281 Some prior linear analyses of deep learning features have enabled partial interpretation
282 of CNNs trained on one cancer type (see [11] and [33]), but mones facilitate pan-
283 cancer interpretations. For example, our results demonstrate a pan-cancer mone can encode
284 conserved morphological features across multiple cancer types (see section 2.1), and a
285 conserved morphological feature can be encoded by distinct mones in different cancer types
286 (see section 2.2). Moreover, mones within strongly correlated clusters can be combined to
287 better identify shared encoded morphology (see section 2.2 and [34]).

288 Mone analysis improves low dimensional visualization of large image sets, such as via
289 t-SNE plots. Pre-trained CNNs are universal feature extractors that encode morphological

290 features predictive of a multitude of labels (see [1], [35], and [36]). Because not all high-
291 dimensional complex relations can be easily embedded in 2D, mone analysis can be used as
292 a first phase of targeted search to identify the mones relevant to classes of interest (see section
293 2.3), allowing more accurate visualization of class separations in mone space.

294 Correlation analysis of mones with gene expression values is a powerful approach for
295 interpreting mones. We identified clusters of highly correlated mone-gene sets, demonstrating
296 clear connection of mones to the underlying genetics. Some recent studies have used
297 exhaustive sets of deep learning features to predict expression profiles (see [15, 16, 17]),
298 but our work shows that small mone-gene clusters can be sufficient and provide simpler
299 interpretability. Both supervised and unsupervised analyses identify meaningful clusters (see
300 section 2.4). Supervised analysis using fixed gene sets is particularly interesting, as it enables
301 direct assessment of genes of interest via mones (see section 2.4.2).

302 *3.2. Mone analysis across architectures and data modalities.* Our analyses demonstrate
303 that mones provide an efficient and interpretable CNN embedding of image data, but a
304 caveat is that they have been restricted to Inception v3 mones. Architectures with fully
305 connected layers tend to increase non-linearity in feature representations. Therefore, models
306 that do not utilize multiple sequences of fully connected layers, such as Inception, are more
307 appropriate for linear mone analysis. For example, recent work suggests a small subset of
308 VGG19 features may also be interpreted via their direct association with phenotypes [11].
309 However, we believe Inception V3 mones are more appropriate for linear association studies
310 because they are the direct inputs to the classification layer. A few other studies have explored
311 correlations between deep autoencoder features and gene (see [33]) or protein expression
312 [37] for other architectures, but the relations between mones across architectures remains
313 a broad and open research topic. We have found that Inception V3 mone 983 in BRCA
314 can be reliably estimated via linear models using ResNet152V3 and DenseNet201 mones
315 ($R^2 > 0.95$). Furthermore, we have been able to convert Xception mones to Inception V3
316 mones using autoencoders with reasonable accuracy ($R^2 \approx 0.5$). These studies suggest that
317 integrative linear mone-gene correlation analysis can be made effective across a range of
318 deep learning architectures.

319 While this work focuses solely on H&E WSIs, we believe mone-based interpretation will
320 be valuable for extension to other spatial data types. Immunohistochemistry (IHC) images
321 are primed for rapid progress, as recent work has shown that several IHC markers can be
322 virtualized from H&E (see [38] and [39]). Generalizing mone analysis for other data types
323 such as spatial transcriptomics and multi-channel protein data is also an exciting and open
324 area, though new architectures will need to be explored to handle the high dimensionality of
325 such images. The interpretation of CNNs for these image types is a challenging but important
326 task, and we expect that integrative multi-modal multi-architecture mone analysis will be a
327 potent and informative approach.

328 **4. Methods.**

329 *4.1. Data acquisition and pre-processing.* The 20X H&E WSIs of TCGA are pre-
330 processed, tiled, and passed through an InceptionV3 model pre-trained on image-net data
331 in [1]. The cached 2048 global average pooling layer features of InceptionV3 (called mones
332 in this manuscript) were written to disk and analyzed. For each of these 2048 mones, the
333 median value across all tiles within a slide was computed to yield the slide-level mone value.

334 4.2. *Differential mone analysis.* Differential mone analysis identifies mones with
 335 statistically significant distributional differences across classes. Welch’s t-test, Kolmogorov
 336 Smirnov (KS) test, Wilcoxon Rank Sum (WRS) test, and optimal Bayesian Filter (OBF)
 337 (see [19] for details) were used for statistical analysis. t-test, WRS test, and KS test use the
 338 Benjamini-Hochberg procedure [40] for FDR correction. The scipy python package [41] was
 339 used to implement t-test, KS test, and WRS test. The statsmodels [42] implementation of the
 340 Benjamini-Hochberg procedure was used.

341 Minimal risk OBF (see [19] for details) identifies mones with posterior probabilities larger
 342 than $1 - \alpha$, where α is the FDR rate. FDR-OBF (see [43] for details) outputs the feature set
 343 that bounds the sample conditioned FDR by α . OBF can report the FDR of any arbitrary
 344 feature set. Unless otherwise stated FDR-OBF is used. OBF uses Jeffrey’s prior, assumes the
 345 prior probability of a mone having distributional differences is 50% (to model no preference
 346 on the identity of a mone, i.e., with or without distributional differences across classes), and
 347 sets the normalization constant of the prior to 0.1. Mones with distributional differences
 348 across classes are hereafter called markers, and mones without distributional differences
 349 are called non-markers. The posterior probabilities of OBF can be used to estimate the
 350 first two moments (mean and standard deviation) of the number of markers (see [43] for
 351 details). Assuming mone identities (marker or non-marker) are independent across cancers,
 352 the posterior probabilities can be multiplied to calculate the probability of a joint event.

353 OBF intrinsically computes the ratio between sample variance and weighted geometric
 354 mean of class-conditioned variances, hereafter denoted by $a(m)$ for mone m . Similar to many
 355 ANOVA-based analyses this ratio measures distributional differences and is closely related
 356 to Bhattacharyya distance (see [44]). It converges to 1 for non-marker mones and converges
 357 to larger values for markers. Larger $a(m)$ values denote larger distributional differences.
 358 Assuming balanced samples of sizes 200 and 100 we compute the $a(m)$ values resulting in a
 359 posterior of 0.95 as thresholds to distinguish moderate [$a(m)=1.088$] and strong [$a(m)=1.159$]
 360 mones separating tumor from normal slides (see Supplementary Figure 3).

361 Structured multi-class OBF (see [45] for details) considers the four possible relations
 362 (known as structures by OBF) between frozen normal, frozen tumor, and FFPE tumor
 363 slides: (A) a mone does not differentiate between slides (prior probability=0.5), (B) a mone
 364 has one distribution for frozen slides (both tumor and adjacent normal slides) and another
 365 distribution for FFPE slides (prior probability=0.5/3), (C) a mone has one distribution for
 366 tumor slides (both frozen and FFPE) and another distribution for frozen normal slides (prior
 367 probability=0.5/3), and (D) a mone has one distribution for FFPE tumor slides and frozen
 368 adjacent normal slides and another distribution for frozen tumor slides. Mones with structure
 369 B for which frozen tumor and FFPE tumors lie on both sides of frozen adjacent normal slides
 370 (based on mean values) are considered ineffective due to FFPE/frozen differences.

371 4.3. *Mone correlation analysis.* For each cancer type, we calculated correlations
 372 between mones over all samples of the given cancer type. This analysis was done for each
 373 cancer type. We use the Ledoit-Wolf shrinkage [46] implementation of the scikit-learn python
 374 package [47] for computing covariance matrices. We then compute the correlation matrix
 375 from the covariance matrix. We apply the Fisher transform to correlation coefficients and
 376 approximate the null with its asymptotic Gaussian distribution. Benjamini-Hochberg [40]
 377 procedure is used for FDR correction. We use seaborn package[48] with default values to
 378 generate clustergrams. Correlation matrices are averaged to compute the pooled correlation
 379 matrix. Statistically significant mone-mone correlations are referred to as “correlated mone
 380 pairs.”

381 Differential mone correlations denote the difference between the correlation coefficient of
 382 tumor and normal slides, i.e., for each cancer, differential matrix is computed by subtracting

383 the correlation coefficient matrix of normal slides form the correlation coefficient matrix of
384 tumor slides. Differential mone correlation analysis uses the asymptotic Gaussian distribution
385 of the difference between Fisher transformed correlation coefficients to compute the p-values.
386 Statistically significant differential mone correlations are referred to as “differentially correlated
387 mone pairs.”

388 4.4. *Linear classification models.* We implement MLDA and LR-LASSO using the scikit-
389 learn python package [47] with default values except we set $C = 100$ and use the “saga”
390 solver [49] in non-binary problems for LR-LASSO. We observed little sensitivity of AUCs
391 to the C values ranging from 1 to 1000, and hence use $C = 100$ throughout. We randomly
392 split data to train and test sets 10 times, which are then used to compute the mean and variance
393 of the AUCs. We use the Scipy package [41] to implement the Wilcoxon signed rank test.

394 4.5. *Cell segmentation and classification.* Cellpose [50] was used to segment and count
395 number of cells in BRCA, LUAD, and COAD tiles. Given the fixed magnification and tile size
396 the number of cells per tile captures tile level cell density. Median number of cells per tile was
397 used as slide level cell density index. HoverNet [51] was used to segment, count, and classify
398 nuclei within COAD tumor slides. HoverNet was executed using the pre-trained PanNuke
399 model ([52]), such that nuclei were classified into one of five types: neoplastic epithelial,
400 non-neoplastic epithelial, connective (including fibroblasts and endothelial), inflammatory
401 (including leukocytes, lymphocytes, and macrophages), and dead nuclei. Median number of
402 cells nuclei across tiles were used as cell density. Median number of predicted inflammatory
403 nuclei across tiles were used to characterize presence of immune cells.

404 4.5.1. *Integrative mone-gene analysis.* Gene expression data were downloaded from the
405 GDC portal [53]. We only used slide-gene expression pairs where both the slide and the
406 expression profile were from the same vial. Log normalized FPKMs were used. Genes with
407 zero counts in more than half the mone-gene pairs or expression standard deviation below
408 0.25 were removed. Given a set of mone-gene pairs, we stack the mone and gene vectors and
409 compute the covariance matrix using the Ledoit-Wolf shrinkage method [46] implemented in
410 the scikit-learn python package [47]. Correlation values are computed given the covariance
411 matrix similar to mone correlation analyses. Statistical significance tests are performed similar
412 to mone correlation analyses.

413 4.5.2. *Immune profiling and analysis.* Leukocyte fractions of TCGA samples were
414 obtained from [31]. All T-cell and B-cell categories were summed to obtain T-cell and B-
415 cell proportions, respectively. The fractions of T-cell and B-cells were summed to obtain
416 lymphocyte fractions. Log normalization of fractions were used throughout. Correlation
417 analysis of immune scores with mones and IG score was performed similar to mone
418 correlation analysis. B-cell percentages above 3% were removed for computing the B-cell
419 correlations as they were deemed outliers.

420 **Acknowledgements.** The authors would like to thank Jill Rubinstein for her feedback.
421 A. F. would like to thank The Jackson Laboratory for the JAX scholar award. J.H.C.
422 acknowledges support from grants R01CA230031 and P30CA034196.

423 **Conflict of interest.** The authors declare no conflict of interest.

REFERENCES

- 424 [1] Javad Noorbakhsh, Saman Farahmand, Ali Foroughi pour, Sandeep Namburi, Dennis Caruana, David Rimm,
425 Mohammad Soltanich-ha, Kourosh Zarringhalam, and Jeffrey H. Chuang.
426 Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological
427 images.
428 *nature communications*.
429 [1](#), [2](#), [3](#), [4](#), [7](#)
- 430 [2] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley.
431 Deep learning for healthcare: review, opportunities and challenges.
432 *Briefings in bioinformatics*, 19(6):1236–1246, 2018.
433 [1](#)
- 434 [3] Shidan Wang, Donghan M Yang, Ruichen Rong, Xiaowei Zhan, Junya Fujimoto, Hongyu Liu, John Minna,
435 Ignacio Ivan Wistuba, Yang Xie, and Guanghua Xiao.
436 Artificial intelligence in lung cancer pathology image analysis.
437 *Cancers*, 11(11):1673, 2019.
438 [1](#)
- 439 [4] Samuel Dodge and Lina Karam.
440 Understanding how image quality affects deep neural networks.
441 In *2016 eighth international conference on quality of multimedia experience (QoMEX)*, pages 1–6. IEEE,
442 2016.
443 [1](#)
- 444 [5] Tejal Nair, Ali Foroughi pour, and Jeffrey H. Chuang.
445 The effect of blurring on lung cancer subtype classification accuracy of convolutional neural networks.
446 In *IEEE conference on bioinformatics and biomedicine*, pages 2987–2989. IEEE, 2020.
447 [2](#)
- 448 [6] Frederick M Howard, James Dolezal, Sara Kochanny, Jefree Schulte, Heather Chen, Lara Heij, Dezheng
449 Huo, Rita Nanda, Olufunmilayo I Olopade, Jakob N Kather, et al.
450 The impact of digital histopathology batch effect on deep learning model accuracy and bias.
451 *bioRxiv*, 2020.
452 [2](#)
- 453 [7] Maximilian Alber, Sebastian Lapuschkin, Philipp Seegerer, Miriam Hägele, Kristof T Schütt, Grégoire
454 Montavon, Wojciech Samek, Klaus-Robert Müller, Sven Dähne, and Pieter-Jan Kindermans.
455 innvestigate neural networks!
456 *J. Mach. Learn. Res.*, 20(93):1–8, 2019.
457 [2](#)
- 458 [8] Amirata Ghorbani, Abubakar Abid, and James Zou.
459 Interpretation of neural networks is fragile.
460 In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3681–3688, 2019.
461 [2](#)
- 462 [9] Ming Y Lu, Melissa Zhao, Maha Shady, Jana Lipkova, Tiffany Y Chen, Drew FK Williamson, and Faisal
463 Mahmood.
464 Deep learning-based computational pathology predicts origins for cancers of unknown primary.
465 *arXiv preprint arXiv:2006.13932*, 2020.
466 [2](#)
- 467 [10] Miriam Hägele, Philipp Seegerer, Sebastian Lapuschkin, Michael Bockmayr, Wojciech Samek, Frederick
468 Klauschen, Klaus-Robert Müller, and Alexander Binder.
469 Resolving challenges in deep learning-based analyses of histopathological images using explanation
470 methods.
471 *Scientific reports*, 10(1):1–12, 2020.
472 [2](#)
- 473 [11] Kevin Faust, Adil Roohi, Alberto J Leon, Emeline Leroux, Anglin Dent, Andrew J Evans, Trevor J Pugh,
474 Sangeetha N Kalimuthu, Ugljesa Djuric, and Phedias Diamandis.
475 Unsupervised resolution of histomorphologic heterogeneity in renal cell carcinoma using a brain tumor-
476 educated neural network.
477 *JCO Clinical Cancer Informatics*, 4:811–821, 2020.
478 [2](#), [6](#), [7](#)
- 479 [12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei.
480 Imagenet: A large-scale hierarchical image database.

- 481 In 2009 *IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
482 2
- 483 [13] Bisakha Ray, Mikael Henaff, Sisi Ma, Efstratios Efstathiadis, Eric R Peskin, Marco Picone, Tito Poli,
484 Constantin F Aliferis, and Alexander Statnikov.
485 Information content and analysis methods for multi-modal high-throughput biomedical data.
486 *Scientific reports*, 4(1):1–10, 2014.
487 2
- 488 [14] Jun Kong, Lee AD Cooper, Fusheng Wang, David A Gutman, Jingjing Gao, Candace Chisolm, Ashish
489 Sharma, Tony Pan, Erwin G Van Meir, Tahsin M Kurc, et al.
490 Integrative, multimodal analysis of glioblastoma using tcga molecular data, pathology images, and clinical
491 outcomes.
492 *IEEE Transactions on Biomedical Engineering*, 58(12):3469–3474, 2011.
493 2
- 494 [15] Benoît Schmauch, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro,
495 Aurélie Kamoun, Meriem Sefta, Sylvain Toldo, Mikhail Zaslavskiy, et al.
496 A deep learning model to predict rna-seq expression of tumours from whole slide images.
497 *Nature communications*, 11(1):1–15, 2020.
498 2, 7
- 499 [16] Alona Levy-Jurgenson, Xavier Tekpli, Vessela N Kristensen, and Zohar Yakhini.
500 Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in
501 breast and lung cancer.
502 *Scientific reports*, 10(1):1–11, 2020.
503 2, 7
- 504 [17] Liviu Badea and Emil Stănescu.
505 Identifying transcriptomic correlates of histology using deep learning.
506 *PLOS ONE*, 15(11):1–30, 11 2020.
507 2, 7
- 508 [18] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna.
509 Rethinking the inception architecture for computer vision.
510 In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
511 2
- 512 [19] Ali Foroughi pour and Lori A. Dalton.
513 Theory of optimal bayesian feature filtering.
514 *Bayesian Analysis*, 15(4):1169–1197, 2020.
515 3, 8
- 516 [20] Robert A Ambros and Ronald Trost.
517 Cellularity in breast carcinoma.
518 *American journal of clinical pathology*, 93(1):98–100, 1990.
519 4
- 520 [21] Kanji Tanaka, Daigo Yamamoto, Masanori Yamada, and Homa Okugawa.
521 Influence of cellularity in human breast carcinoma.
522 *The Breast*, 13(4):334–340, 2004.
523 4
- 524 [22] A Serrablo, P Paliogiannis, F Pulighe, S Saudi-Moro Moro, V Borrego-Estella, F Attene, F Scognamillo,
525 and C Hörndler.
526 Impact of novel histopathological factors on the outcomes of liver surgery for colorectal cancer metastases.
527 *European Journal of Surgical Oncology (EJSO)*, 42(9):1268–1277, 2016.
528 4
- 529 [23] Frederikke Petrine Fliedner, Trine Bjørnbo Engel, Henrik H El-Ali, Anders Elias Hansen, and Andreas
530 Kjaer.
531 Diffusion weighted magnetic resonance imaging (dw-mri) as a non-invasive, tissue cellularity marker to
532 monitor cancer treatment response.
533 *BMC cancer*, 20(1):1–9, 2020.
534 4
- 535 [24] Jeonghyuk Park, Yul Ri Chung, Seo Taek Kong, Yeong Won Kim, Hyunho Park, Kyungdoc Kim, Dong-II
536 Kim, and Kyu-Hwan Jung.
537 Aggregation of cohorts for histopathological diagnosis with deep morphological analysis.
538 *Scientific reports*, 11(1):1–11, 2021.
539 4

- 540 [25] Timon PH Buys, Raj Chari, Eric HL Lee, May Zhang, Calum MacAulay, Stephen Lam, Wan L Lam, and
541 Victor Ling.
542 Genetic changes in the evolution of multidrug resistance for cultured human ovarian cancer cells.
543 *Genes, Chromosomes and Cancer*, 46(12):1069–1079, 2007.
544 5
- 545 [26] Wei Zhang, Yi Liu, Na Sun, Dan Wang, Jerome Boyd-Kirkup, Xiaoyang Dou, and Jing-Dong Jackie Han.
546 Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor
547 prognosis in ovarian cancer.
548 *Cell reports*, 4(3):542–553, 2013.
549 5
- 550 [27] Su Li, Hua Li, Ying Xu, and Xiaomei Lv.
551 Identification of candidate biomarkers for epithelial ovarian cancer metastasis using microarray data.
552 *Oncology letters*, 14(4):3967–3974, 2017.
553 5
- 554 [28] Karolina Sterzyńska, Andrzej Klejewski, Karolina Wojtowicz, Monika Świerczewska, Marta Nowacka,
555 Dominika Kaźmierczak, Małgorzata Andrzejewska, Damian Rusek, Maciej Brzązert, Jacek Brzązert,
556 et al.
557 Mutual expression of *aldh1a1*, *lox*, and collagens in ovarian cancer cell lines as combined cscs-and ecm-
558 related models of drug resistance development.
559 *International journal of molecular sciences*, 20(1):54, 2019.
560 5
- 561 [29] Michael S Rooney, Sachet A Shukla, Catherine J Wu, Gad Getz, and Nir Hacohen.
562 Molecular and genetic properties of tumors associated with local immune cytolytic activity.
563 *Cell*, 160(1-2):48–61, 2015.
564 5
- 565 [30] Luigi Cari, Francesca De Rosa, Maria Grazia Petrillo, Graziella Migliorati, Giuseppe Nocentini, and Carlo
566 Riccardi.
567 Identification of 15 t cell restricted genes evaluates t cell infiltration of human healthy tissues and cancers
568 and shows prognostic and predictive potential.
569 *International journal of molecular sciences*, 20(20):5242, 2019.
570 5
- 571 [31] Vésteinn Thorsson, David L Gibbs, Scott D Brown, Denise Wolf, Dante S Bortone, Tai-Hsien Ou Yang,
572 Eduard Porta-Pardo, Galen F Gao, Christopher L Plaisier, James A Eddy, et al.
573 The immune landscape of cancer.
574 *Immunity*, 48(4):812–830, 2018.
575 5, 6, 9
- 576 [32] Zi-Han Geng, Chun-Xiang Ye, Yan Huang, Hong-Peng Jiang, Ying-Jiang Ye, Shan Wang, Yuan Zhou, Zhan-
577 Long Shen, and Xiao-Yan Qiu.
578 Human colorectal cancer cells frequently express *igg* and display unique *ig* repertoire.
579 *World journal of gastrointestinal oncology*, 11(3):195, 2019.
580 6
- 581 [33] Jordan T Ash, Gregory Darnell, Daniel Munro, and Barbara E Engelhardt.
582 Joint analysis of expression levels and histological images identifies genes associated with tissue
583 morphology.
584 *Nature communications*, 12(1):1–12, 2021.
585 6, 7
- 586 [34] Ellery Wulczyn, David F Steiner, Melissa Moran, Markus Plass, Robert Reihls, Fraser Tan, Isabelle Flament-
587 Auvigne, Trissia Brown, Peter Regitnig, Po-Hsuan Cameron Chen, et al.
588 Interpretable survival prediction for colorectal cancer using deep learning.
589 *NPJ digital medicine*, 4(1):1–13, 2021.
590 6
- 591 [35] Yu Fu, Alexander W Jung, Ramon Viñas Torne, Santiago Gonzalez, Harald Vöhringer, Artem Shmatko,
592 Lucy R Yates, Mercedes Jimenez-Linan, Luiza Moore, and Moritz Gerstung.
593 Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis.
594 *Nature Cancer*, 1(8):800–810, 2020.
595 7
- 596 [36] Jakob Nikolas Kather, Lara R Heij, Heike I Grabsch, Chiara Loeffler, Amelie Echle, Hannah Sophie Muti,
597 Jeremias Krause, Jan M Niehues, Kai AJ Sommer, Peter Bankhead, et al.
598 Pan-cancer image-based detection of clinically actionable genetic alterations.

- 599 *Nature Cancer*, 1(8):789–799, 2020.
600 7
- 601 [37] Luke Ternes, Mark Dane, Marilyne Labrie, Gordon Mills, Joe Gray, Laura Heiser, and Young Hwan Chang.
602 Me-vae: Multi-encoder variational autoencoder for controlling multiple transformational features in single
603 cell image analysis.
604 *bioRxiv*, 2021.
605 7
- 606 [38] Christopher R Jackson, Aravindhan Sriharan, and Louis J Vaickus.
607 A machine learning algorithm for simulating immunohistochemistry: development of sox10 virtual ihc and
608 evaluation on primarily melanocytic neoplasms.
609 *Modern Pathology*, pages 1–11, 2020.
610 7
- 611 [39] Zhaoyang Xu, Carlos Fernández Moro, Béla Bozóky, and Qianni Zhang.
612 Gan-based virtual re-staining: a promising solution for whole slide image analysis.
613 *arXiv preprint arXiv:1901.04059*, 2019.
614 7
- 615 [40] Yoav Benjamini and Yosef Hochberg.
616 Controlling the false discovery rate: a practical and powerful approach to multiple testing.
617 *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.
618 8
- 619 [41] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni
620 Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stefan J. van der Walt, Matthew Brett,
621 Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern,
622 Eric Larson, CJ Carey, Ihan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef
623 Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald,
624 Antonio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors.
625 Scipy 1.0: Fundamental algorithms for scientific computing in python.
626 *Nature Methods*, 2020.
627 8, 9
- 628 [42] Skipper Seabold and Josef Perktold.
629 statsmodels: Econometric and statistical modeling with python.
630 In *9th Python in Science Conference*, 2010.
631 8
- 632 [43] A. Foroughi Pour and L. A. Dalton.
633 Bayesian error analysis for feature selection in biomarker discovery.
634 *IEEE Access*, 7:127544–127563, 2019.
635 8
- 636 [44] Ali Foroughi pour and Lori A Dalton.
637 Optimal bayesian filtering for biomarker discovery: Performance and robustness.
638 *IEEE/ACM transactions on computational biology and bioinformatics*, 17(1):250–263, 2018.
639 8
- 640 [45] Ali Foroughi pour and Lori A Dalton.
641 Biomarker discovery via optimal bayesian feature filtering for structured multiclass data.
642 In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and*
643 *Health Informatics*, pages 331–340, 2018.
644 8
- 645 [46] Olivier Ledoit and Michael Wolf.
646 A well-conditioned estimator for large-dimensional covariance matrices.
647 *Journal of multivariate analysis*, 88(2):365–411, 2004.
648 8, 9
- 649 [47] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer,
650 R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and
651 E. Duchesnay.
652 Scikit-learn: Machine learning in Python.
653 *Journal of Machine Learning Research*, 12:2825–2830, 2011.
654 8, 9
- 655 [48] Michael Waskom and the seaborn development team.
656 mwaskom/seaborn, September 2020.
657 8

- 658 [49] Aaron Defazio, Francis Bach, and Simon Lacoste-Julien.
659 Saga: A fast incremental gradient method with support for non-strongly convex composite objectives.
660 *arXiv preprint arXiv:1407.0202*, 2014.
661 [9](#)
- 662 [50] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu.
663 Cellpose: a generalist algorithm for cellular segmentation.
664 *Nature Methods*, 18(1):100–106, 2021.
665 [9, 15](#)
- 666 [51] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and
667 Nasir Rajpoot.
668 Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images.
669 *Medical Image Analysis*, 58:101563, 2019.
670 [9](#)
- 671 [52] Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benes, Simon Graham, Mostafa Jahanifar, Syed Ali
672 Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot.
673 Pannuke dataset extension, insights and baselines.
674 *arXiv preprint arXiv:2003.10778*, 2020.
675 [9](#)
- 676 [53] Robert L Grossman, Allison P Heath, Vincent Ferretti, Harold E Varmus, Douglas R Lowy, Warren A Kibbe,
677 and Louis M Staudt.
678 Toward a shared vision for cancer genomic data.
679 *New England Journal of Medicine*, 375(12):1109–1112, 2016.
680 [9](#)

681 **Figures.**

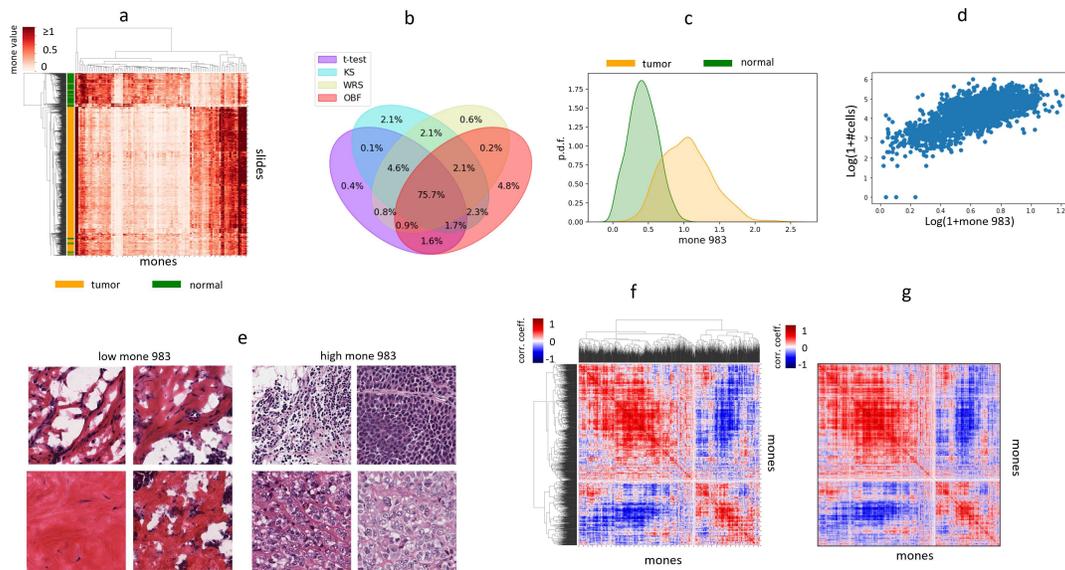


Figure 1: Individual mones and mone pairs encode and distinguish phenotypes. (a) Clustermap of BRCA slides using the top100 mones differentiating the slides. 100 mones are sufficient to separate frozen normal (green) from frozen tumor (orange) slides. (b) Venn diagram of statistically significant mones differentiating tumor from adjacent normal frozen slides, comparing different statistical tests. Venn diagrams were calculated for each cancer type, and the observed plot shows the average across all cancer types. On average the statistical tests agree on 75% of mones differentiating between tumor and normal slides. (c) Probability density function of mone 983 among frozen tumor (orange) and adjacent normal (green) BRCA slides. (d) Log-normalized scatter plot of slide level mone 983 and cellpose (see [50]) estimates of cellularity across BRCA frozen slides. (e) Example tiles from slides with extreme mone 983 values (high and low). (f) Cluster map of the mone-mone correlation matrix of LUAD tumor slides, demonstrating that many mones pairs are highly correlated. (g) Mone-mone correlation matrix of LUSC slides, with mones ordered identically to the Figure 1f cluster map.

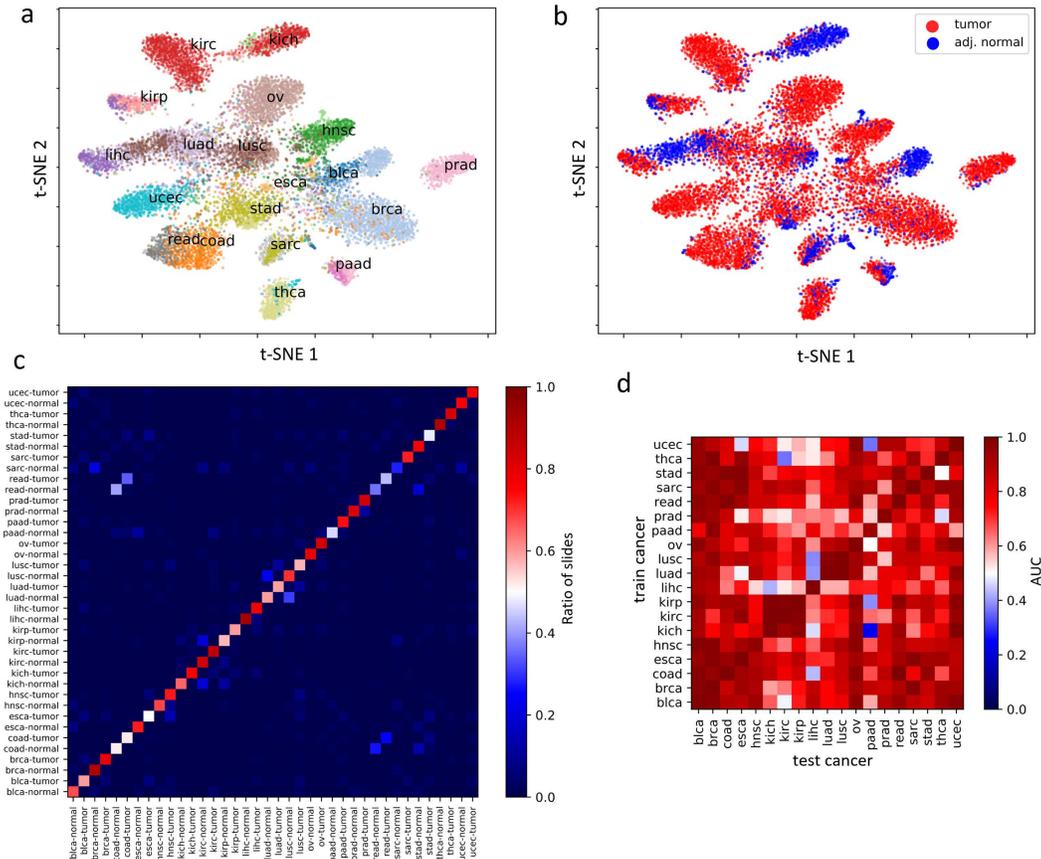


Figure 2: The joint distribution of mones reliably separates tumor and normal slides and the underlying cancer. 2D t-SNE plots of the mones-based MLDA feature space distinguishing 38 classes (19 cancers, tumor/normal status) based on (a) cancer type and (b) tumor/normal status. (c) Normalized confusion matrix of the 38-class mones-based logistic regression classifier. The color depicts the ratio of slides with a given true class predicted as any of the possible classes. The large diagonal values suggest the classifier has high accuracy. (d) The cross-classification AUCs of mones-based logistic regression tumor/normal classifier trained on each cancer and applied to all cancers.

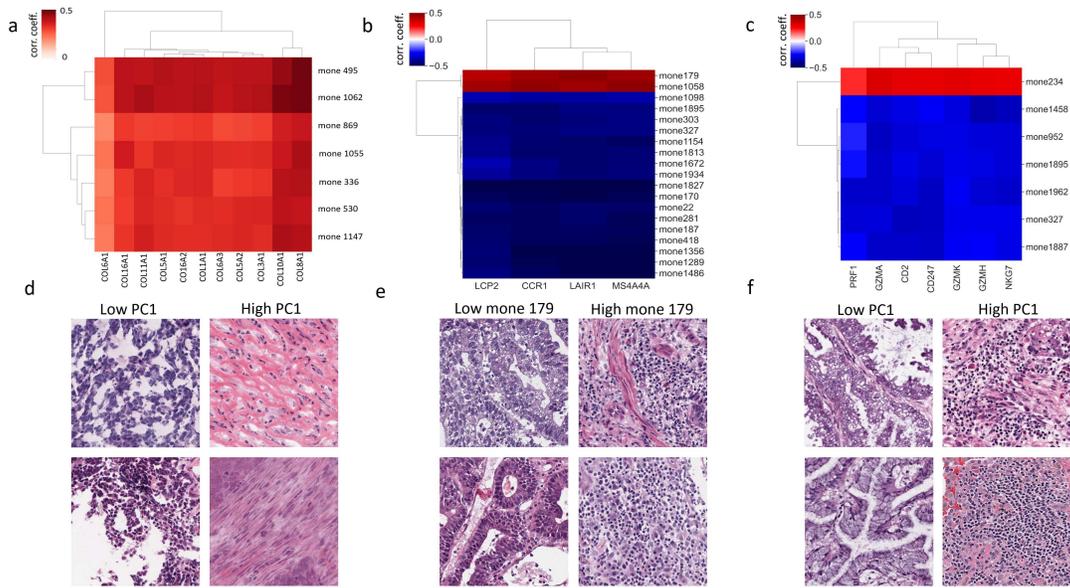


Figure 3: Mone-gene correlation analysis identifies highly correlated mone-gene clusters. Correlation matrix of (a) a cluster of highly correlated mones and collagen genes in OV , and a cluster of highly correlated mones and immune-related genes in (b) COAD and (c) LUAD. See Supplementary Figure 9 for adjusted p-values. Example tiles from slides with (d) high and low PC-1 in OV, (e) high and low mone 179 in COAD, and (f) high and low PC-1 in LUAD. Histopathology review identifies that mone-predicted (d) OV tiles with high PC-1 are rich in collagen, and (e) COAD with high mone 179 and (f) luad tiles with high PC-1 have a strong lymphocyte presence. See Supplementary Figures 10-13 for additional examples at both high and low mone values.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Sfiles.zip](#)
- [fmp1sf.pdf](#)