A Deep Learning Approach for Rapid Mutational Screening in Melanoma

Randie H. Kim[1,2†], Sofia Nomikou[3†], Zarmeena Dawood[2], George Jour[1,2,4], Douglas Donnelly[2], Una Moran[2], Jeffrey S. Weber[2,5], Narges Razavian[6,7], Matija Snuderl[4], Richard Shapiro[2,8], Russell S. Berman[2,8], Nicolas Coudray[3,9], Iman Osman[1,2,5*], Aristotelis Tsirigos[2,3,4*]

[1]The Ronald O. Perelman Department of Dermatology

[2]Interdisciplinary Melanoma Cooperative Group

[3]Applied Bioinformatics Laboratories

[4]Department of Pathology

[5]Department of Medicine

[6]Deparmtent of Radiology

[7]Department of Population Health

[8]Department of Surgery

[9]Skirball Institute Department of Cell Biology

New York University School of Medicine, New York, New York.

†These authors contributed equally to this work.

*Corresponding Authors:

Aristotelis Tsirigos, PhD

Associate Professor of Pathology

Director, Applied Bioinformatics Laboratories

New York University School of Medicine

227 East 30th Street

New York, New York

Phone: 646-501-2693

Email: Aristotelis.Tsirigos@nyulangone.org

Iman Osman, MD

NYU School of Medicine

522 First Avenue

Phone: 212-263-9075; Fax: 212-263-9090

Email: Iman.Osman@nyulangone.org

Abstract

DNA-based molecular assays for determining mutational status in melanomas are time-consuming and costly. As an alternative, we applied a deep convolutional neural network (CNN) to histopathology images of tumors from 257 melanoma patients and developed a fully automated model that first selects for tumor-rich areas (Area under the curve AUC=0.98), and second, predicts for the presence of mutated *BRAF* or *NRAS*. Network performance was enhanced on *BRAF*-mutated melanomas ≤1.0 mm (AUC=0.83) and on non-ulcerated *NRAS*-mutated melanomas (AUC=0.92). Applying our models to histological images of primary melanomas from The Cancer Genome Atlas database also demonstrated improved performances on thinner *BRAF*-mutated melanomas and non-ulcerated *NRAS*-mutated melanomas. We propose that deep learning-based analysis of histological images has the potential to become integrated into clinical decision making for the rapid detection of mutations of interest in melanoma.

Introduction

Mutations in the *BRAF* oncogene are found in 50-60% of all melanomas[1], while *NRAS* mutations comprise an additional 15-20%. With the development of targeted therapies[2, 3], determining the mutational status of *BRAF* and *NRAS* has become an integral component for the management of Stage III/IV melanomas. DNA molecular assays such as Sanger sequencing, pyrosequencing, and next generation sequencing (NGS) are the current gold standard to determine mutational status[4]. However, these methods are costly and time-consuming. Immunohistochemistry, real-time polymerase chain reaction (PCR), and automated platforms[5, 6, 7] are rapid and less expensive alternatives, but are limited to screening for specific mutations, such as *BRAF*-V600E/K or *NRAS*-Q61R/L, and may potentially fail to identify rare mutational variants in patients that might have otherwise benefited from adjuvant targeted therapy.

Deep Convolutional Neural Network (CNN) methods to predict mutational status have been demonstrated in other solid tumors. CNNs utilize multiple layers of convolution operations, pooling layers, and fully connected layers to perform classification of images to classes of interest through identification of various image features often not directly detectable by the human eye. Deep CNNs, which utilize non-linear learning algorithms, have been successful in manipulating and processing large data sets, particularly for image analysis[8]. Using images from The Cancer Genome Atlas (TCGA), a collaborative cancer genomics database[9], our group has previously developed a machine learning algorithm that can predict for 6 different genes, including *EGFR* and *STK11,* in lung carcinoma[10]. In breast cancer, deep learning applied to tumor microarray images has been shown to predict for *ER* status with an 84% accuracy[11].

In this study, we adapt our previous deep learning algorithm to a different dataset comprised of histopathology images of primary melanomas resected from patients prospectively enrolled in a single-institution IRB-approved clinicopathological and biorepository in order to develop a model from tissue specimens that are more representative of what might be seen in routine clinical practice. While molecular testing is typically performed on the most recent metastatic sample, testing on the primary tumor can be performed if metastatic tissue is unavailable or carries a low tumor burden[12, 13]. We present our deep learning models for the screening of *BRAF* and *NRAS* mutations in primary melanomas, with the purpose of exploring its potential clinical utility.

Results

Patient characteristics

324 primary melanomas from 266 unique patients were included in this study and divided into training (n=182), validation (n=43), and independent (n=41) cohorts, without overlap between the patient subsets. Within each cohort, *BRAF*-mutant, *NRAS*-mutant, and *WT/WT* melanomas were represented (**Table 1**). The average ages of patients with *BRAF*-mutant melanoma were 59.4, 53.9, and 59-years old in training, validation, and independent cohorts, respectively. For patients with *NRAS*-mutant melanoma, the average ages were 61.2, 70.9, and 59.9-years old; and for *WT/WT* patients, the average ages were 65.1, 68.5, and 65.1-years old.

Dataset characteristics

365 formalin-fixed paraffin embedded (FFPE) hematoxylin and eosin (H&E)-stained slides from 324 primary melanomas were digitized and reviewed for quality control. After excluding images that were blurry, faded, or contained no tumor, 293 images from 257 melanomas were available for analysis. 103 *BRAF*-mutant, 94 *NRAS*-mutant, and 96 *WT/WT* melanomas images were included in the study. V600E comprised 70% of the *BRAF* mutations. *NRAS* Q61R/Q61K comprised 80% of the *NRAS* mutations.

Computational workflow for whole-slide histopathology image analysis

Our computational workflow with the CNN Inception v3 is shown in **Fig. 1.** and is common across all our classifiers (see Methods). The aim of our analytical approach was to: (1) to predict the presence of *BRAF* and *NRAS* mutations using manually annotated slides; and (2) to automate the annotation process by the CNN.

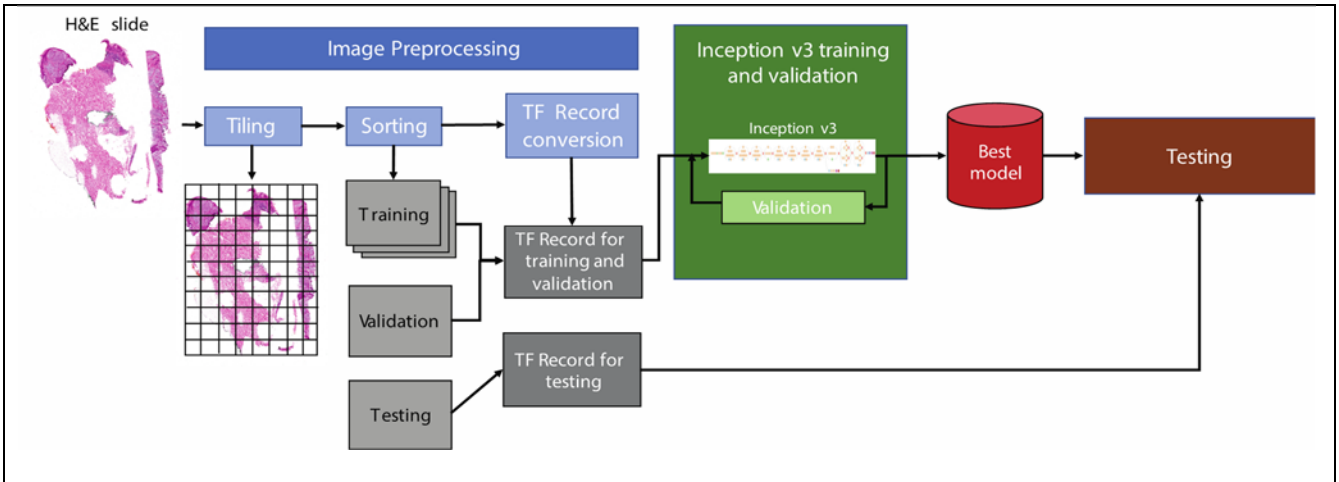


Figure 1. Computational Workflow To train the Inception v3 CNN, slides are tiled to non-overlapping 299x299 pixel tiles and assigned to training, validation and independent sets comprising of 70%, 15% and 15% of the total number of tiles, respectively. Tiles of slides belonging to the same patient are considered together in a data set. After conversion to TF Record format, training is performed. The best performing model on the validation data is evaluated on the independent set.

Predicting BRAF and NRAS mutation on manually annotated whole-slide images

Because of tissue heterogeneity in skin specimens, tumor-rich areas were initially manually annotated as regions of interest (ROI). Normal skin and associated appendages, connective and subcutaneous tissue, necrosis, hemorrhage, and aggregates of dense inflammation were excluded from training. The network was trained on tiled images of manually annotated ROI, with 70% of images used for training, 15% used for validation, and 15% used for independent testing (**Supplemental Table 1**). Model performance achieved a per slide Area Under the Curve (AUC)=0.75 [95% CI: 0.60,0.90] for predicting $BRAF^{Mut}$ (**Supplemental Figure 1**) and AUC=0.77 [95% CI: 0.58,0.96] for predicting $NRAS^{Mut}$ (**Supplemental Figure 2**).

We next sought to elucidate some of the parameters that could influence network performance. To evaluate the role of tumor thickness, tumors from the independent cohort were sorted by Breslow depth. Model performance for predicting mutated *BRAF* improved for slides with a tumor thickness ≤1.0 mm, with an AUC=0.83 [95% CI: 0.45,1.0] (**Fig. 2A, *left***). Conversely, there were reductions in the AUC to 0.74 [95% CI: 0.58,0.89] for tumors >1.0-5.0 mm, and to 0.75 [95% CI: 0.60,0.90] for very thick tumors >5.0-10mm. One potential explanation for this difference is that *BRAF*-mutated melanomas are associated with a distinctive epidermal component, such as increased pagetoid scatter and intraepidermal nesting of melanocytes[14]. These histologic features may carry more weight in thinner tumors compared to deeper and more invasive melanomas. *NRAS* mutation prediction was not consistently dependent on tumor thickness (**Fig 2B, *left*; Supplemental Table 2)**.

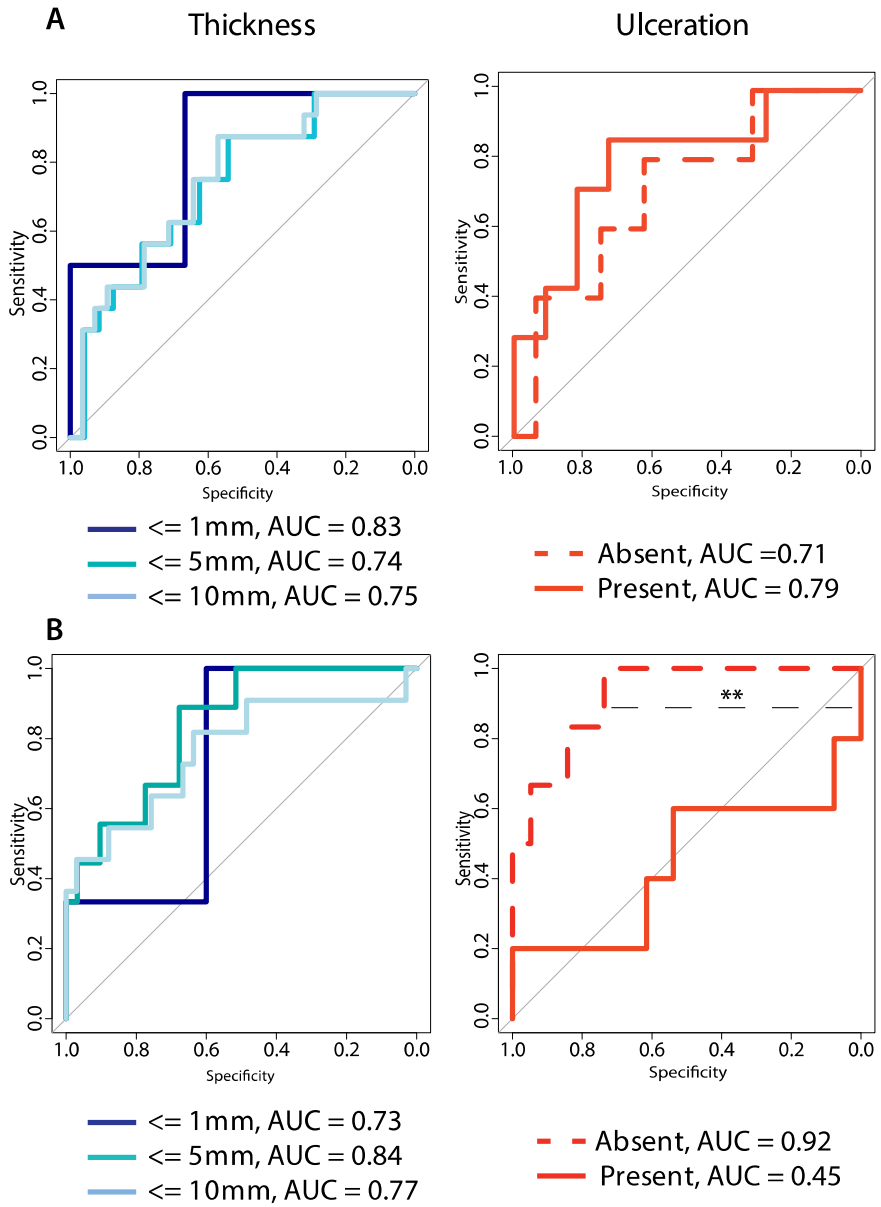


Figure 2. Exploration of parameters contributing to network performance A) BRAF mutation prediction. ROC plots for melanomas ≤1.0 mm, >1.0-5.0 mm and >5.0-10mm with the network trained on manually annotated slides demonstrate improved performance for thinner tumors (*left*). ROC plots for melanomas based on ulceration status show improved performance if ulceration was present (*right*). **B) NRAS mutation prediction.** ROC curves for melanomas based on Breslow depth show no significant differences (*left*). NRAS mutation prediction is improved for melanomas without ulceration (*right*).

We also examined whether ulceration status, as indicated by the original pathology report, can affect network performance. For the BRAF prediction model, an AUC=0.79 [95% CI:

0.56, 1.00] was achieved for melanomas with ulceration and an AUC= 0.71 [95% CI: 0.50, 0.92] for melanomas without ulceration (**Fig 2A, *right***). The opposite trend was observed with the NRAS model, where ulcerated melanomas led to a decreased AUC to 0.45 [95% CI 0.07-0.83] and non-ulcerated melanomas had an increased AUC to 0.92 [95% CI: 0.81,1.00] **(Fig 2B, *right*)**. Notably, these results were achieved on manually annotated ROIs that excluded areas of ulceration, indicating that the network is not learning from the presence of an ulceration on the slide itself. Rather, elements in the tumor microenvironment that influence the ulceration status are potentially playing an important role in determining *NRAS* mutation status.

In order to confirm that tumor thickness and ulceration alone are not predictors of mutational status, we built a multivariate logistic regression model where Breslow depth and ulceration status are the predictive variables for the presence of mutated *BRAF* or *NRAS*. The model was trained on the same training dataset used for our deep CNN. This model performed at random for predicting either mutated *BRAF* and *NRAS*, with AUC=0.53 [95% CI: 0.34,0.72] and AUC=0.52 [95% CI: 0.30,0.75], respectively (**Supplemental Figure 3**). This demonstrates that the necessary features for predicting mutation status are provided by the histopathological slide.

Performance for the mutation network can be visualized with a probability heat map, where the presence of the mutation of interest is shown in red and intensity of color corresponding to the probability of mutation. **Fig. 3A** demonstrates representative H&E sections of melanomas with their corresponding probability heat maps for *BRAF*-mutant

(*left*), and 2 correctly identified non-BRAF-mutant tumors: e.g. *NRAS*-mutant (*center*) and *WT/WT* melanoma (*right*), respectively. Similarly, **Fig. 3B** demonstrates probability heat maps for the NRAS prediction network, with correctly identified *NRAS*-mutant (*left*) and non-*NRAS*-mutant melanoma. Interestingly, in both *BRAF*-mutant (*center*) and *WT/WT* melanomas (*right*), there are regions identified by the network to harbor mutated *NRAS*, raising the possibility of intratumoral heterogeneity.

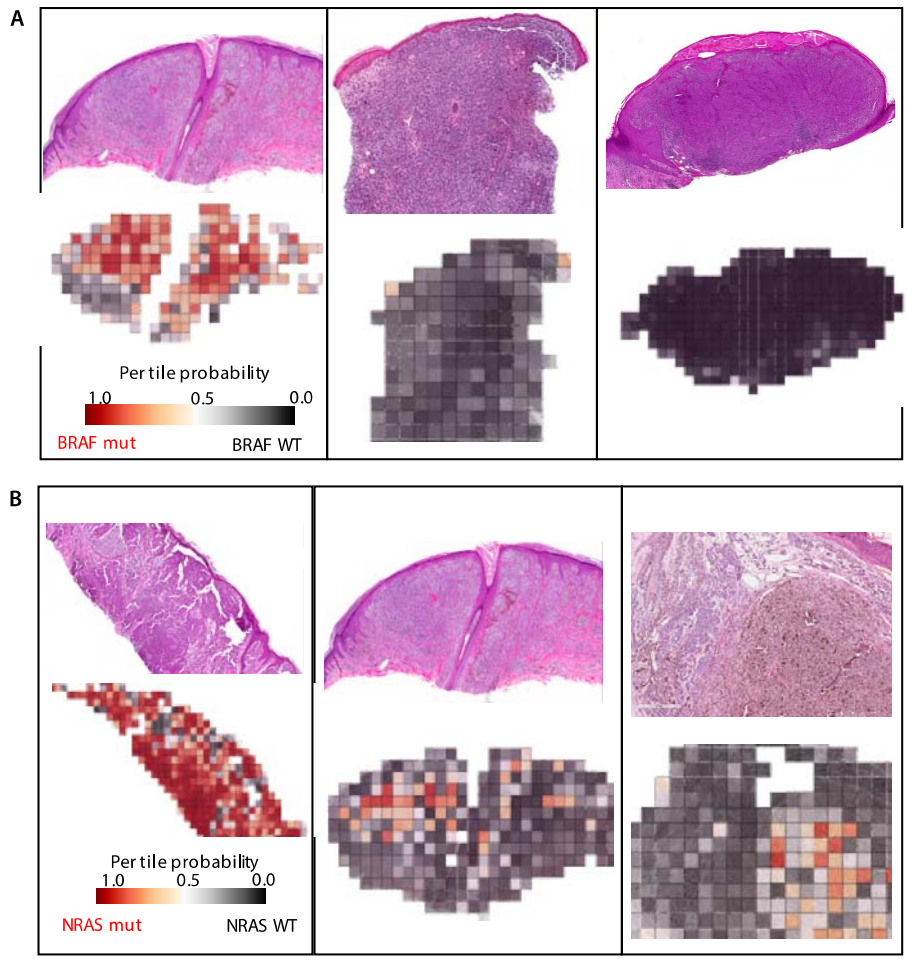


Figure 3. Probability heat maps for mutation prediction A) H&E slide section and corresponding heat map of a correctly classified $BRAF^{mut}$ melanoma (*left*), $NRAS^{mut}$ melanoma (*center*) and *WT/WT* melanoma (*right*) slides by the *BRAF* mutation prediction network. **B)** H&E slide section and corresponding heat map of a correctly classified $NRAS^{mut}$ melanoma (*left*), $BRAF^{mut}$ melanoma (*center*) and *WT/WT* melanoma (*right*) slides by the *NRAS* mutation prediction network.

Predicting mutated BRAF and NRAS using images from The Cancer Genome Atlas database

An image dataset of digitized FFPE H&E-stained slides of primary melanomas were retrieved from TCGA, a collaborative and publicly available research database comprised of tumor tissue and genomic data from multiple cancer types[9]. This dataset was used as an independent cohort, which comprised of 40 *BRAF*-mutant cases, 9 *NRAS*-mutant cases, and 22 *WT/WT* cases, summing up to 71 cases in total. After quality control, 68 images were approved for the independent validation.

Breslow depth information was available for 32 out of the 68 slides[15]. Melanomas from the TCGA database were skewed towards very thick tumors with a median of 7.5mm. There were no tumors less than 1.0mm in depth. In comparison, the median depths of tumors in our training, validation and test cohorts were 2.00, 1.45, and 1.90, respectively (**Supplemental Figure 4**). AUCs of mutation prediction were calculated for melanomas ≤1.0 mm, >1.0-5.0 mm, >5.0-10 mm, >10-15mm, and >15-20mm for NYU and TCGA cohorts (**Supplemental Table 2**). For *BRAF* mutation prediction on TCGA images, the network performed better for melanomas ≤ 5 mm. AUC values decreased with thicker tumors. This trend is similar to the effect tumor thickness had on our NYU test cohort (**Fig. 4A**). Tumor thickness did not affect network predictive ability for mutated *NRAS* on TCGA tumors, as was previously observed in our own cohort (**Supplemental Table 2**).

Regarding the effect of ulceration on network performance, TCGA melanomas without ulceration (i.e., T2a, T3a and T4a, n=10) were compared to melanomas with ulceration (i.e., T2b, T3b and T4b, n=56). The ulceration status of TCGA tumors did not have a significant impact on BRAF mutation prediction (**Supplemental Table 3**), consistent with our observations in the NYU cohort. Importantly, network performance for predicting mutated *NRAS* was significantly enhanced for non-ulcerated melanomas with an AUC=0.89 [95% CI: 0.67-1.0] compared to AUC=0.45 [95% CI: 0.24-0.67] for ulcerated melanomas, reproducing the difference that ulceration status has on NRAS mutation prediction in our own cohort (**Fig. 4B, Supplemental Table 3**).

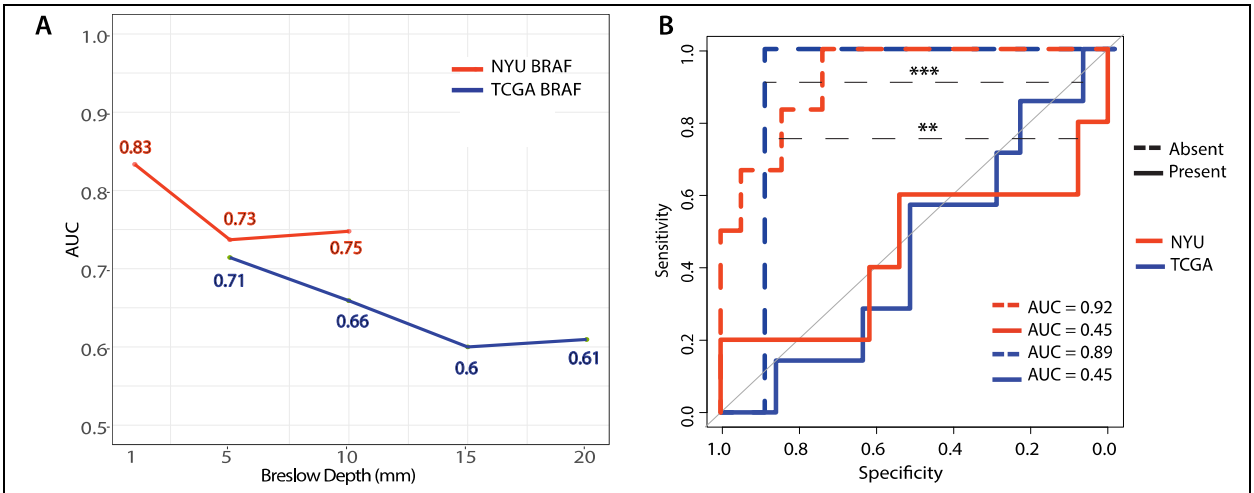


Figure 4. Validation of network performance on an independent cohort from TCGA. A) AUC variation for different Breslow depth values on the independent NYU test set and the TCGA cohort. *BRAF* mutation prediction is improved for thinner melanomas both on the NYU test set and the TCGA external validation cohort**. B)** *NRAS* mutation prediction is improved for melanomas without ulcerations both on the NYU independent cohort and on the TCGA external validation cohort.

Automated selection of primary melanomas on whole slide histopathology images

In order to improve the clinical application of our deep learning models, we attempted to automate the identification of melanoma by processing tiled images as “in” the ROI or “out” of the ROI. Model performance achieved a per slide AUC=0.98 [95% CI: 0.95,1.00]

and a per tile AUC=0.92 [95% CI: 0.922,0.924] (**Fig. 5A**). H&E-stained non-annotated whole slides of *BRAF*-mutant, *NRAS*-mutant and *WT/WT* melanomas are shown in **Fig. 5B-D.** along with their corresponding network-generated probability heat maps, where orange indicates tumor, gray indicates non-tumor, and the intensity of the color correlates with the probability gradient. Notably, there is excellent concordance between the pathologist and the network. Network performance was independent of melanoma mutational status (**Supplemental Figure 5**).

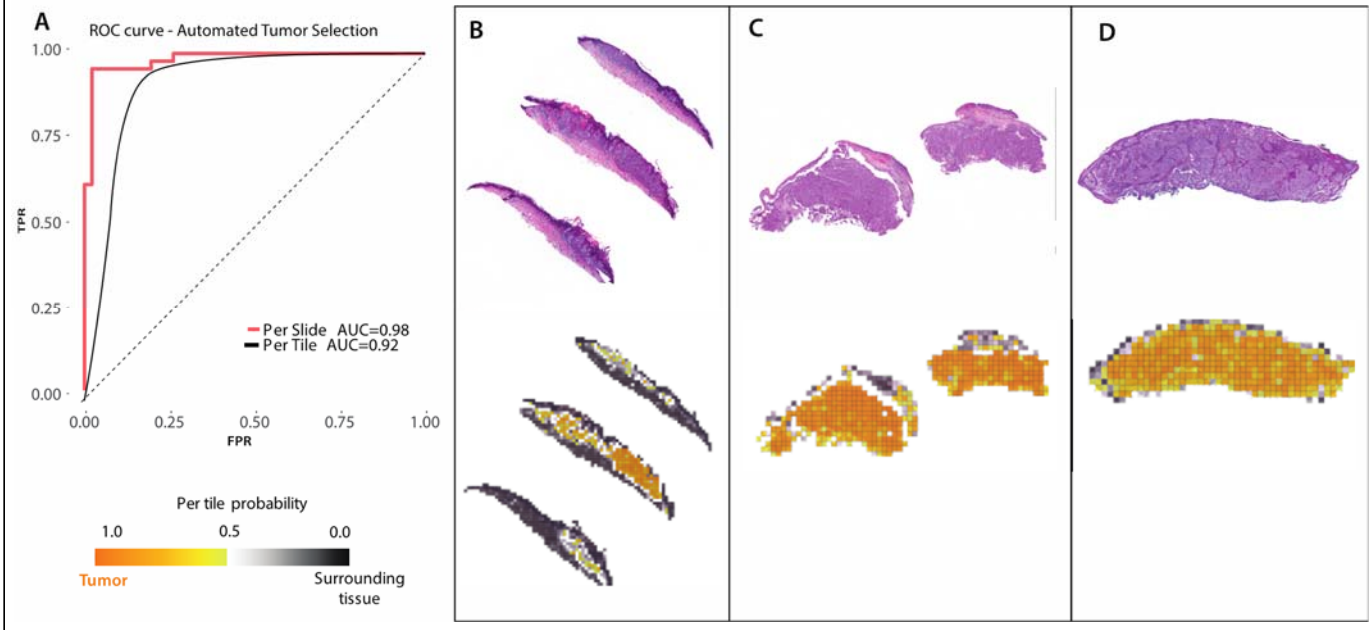


Figure 5. Automating tumor selection for a fully automated sequential workflow A) ROC and AUC for the automated tumor selection network. H&E slide section and corresponding heat map of tumor annotation on **B)** a *BRAF*-mutated slide **C)** a *NRAS*-mutated slide and **D)** a *WT/WT* slide.

We then examined whether the automated tumor selection network could be combined with the mutation prediction networks in a sequential manner. The computational workflow as outlined in **Fig. 1.** was repeated on all non-annotated images that were tiled

and passed through the automated tumor selection network. Tiles assigned with a probability of belonging to tumor area (probability >= 0.5) were filtered and split in training, validation and independent sets. The Inception v3 architecture was re-trained on tiles selected by the automated network for mutation prediction. Importantly, the 44 slides that comprised the independent set for the algorithm trained on manually annotated images were maintained as the independent set for the algorithm trained on network selected tumors. In this fully automated network, model performance achieved an AUC=0.75 [95% CI: 0.58, 0.89] for predicting mutated *BRAF* and an AUC=0.70 [95% CI: 0.47, 0.90] for predicting mutated *NRAS* (**Supplemental Figure 6**). These results are consistent with model performance trained on the manual annotations, indicating that the automated tumor selection network performs as well as the dermatopathologist and does not impact mutation prediction performance. **Fig. 6.** outlines the complete workflow of mutation prediction for melanoma H&E histopathology slides using sequential networks for automated tumor selection and mutation prediction.

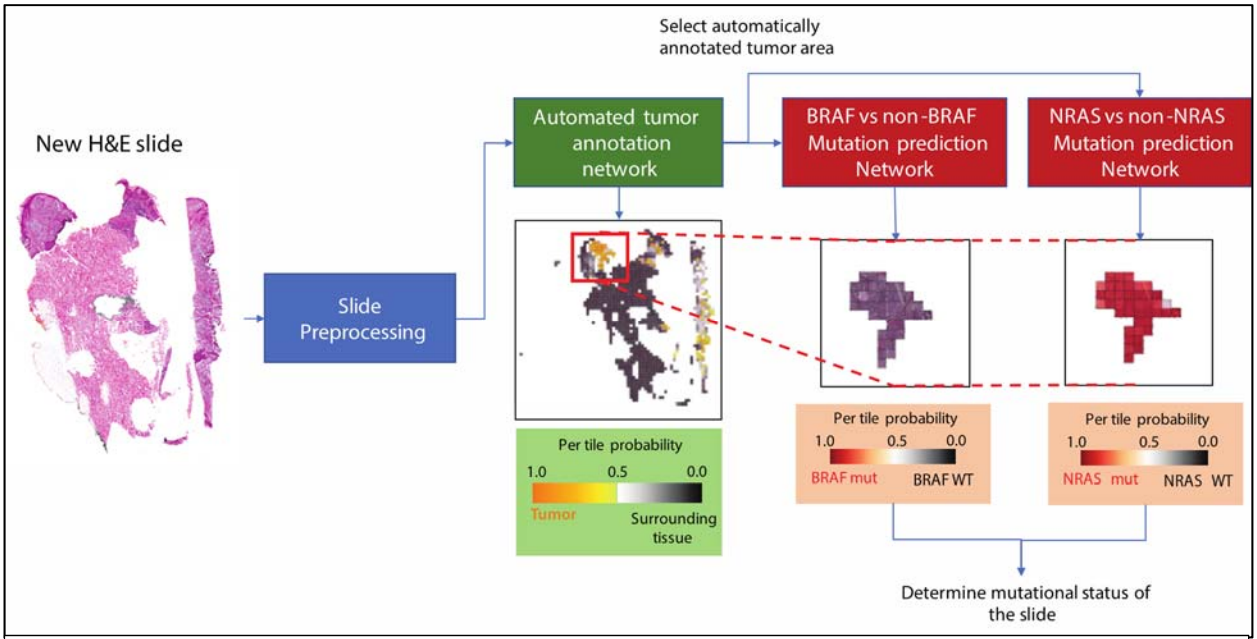


Figure 6. Fully Automated Sequential Workflow. Non-annotated whole slides are processed, tiled, and passed through the automated tumor annotation network which assigns a probability to each tile of belonging in the tumor. Tiles with probability > =0.5 are subsequently passed through the mutation prediction network for determining the mutational status of the slide of interest.

Association of network mutation localization with immunohistochemical analysis

To further corroborate network accuracy, we examined whether network-generated probability heat maps are true visual representations of mutation localization. An additional set of 39 *BRAF*^{V600E} cases underwent automated algorithmic prediction and immunohistochemical (IHC) analysis with the monoclonal VE1 antibody, a reliable screening tool for detecting the specific V600E mutation[16]. The tumor selection algorithm was applied with a threshold of 0.1 to remove tiles with very low tumor probability. The top 10 cases predicted to be *BRAF*^{Mut} were subjected to IHC analysis for *BRAF*^{V600E} staining. Regions of positive IHC staining were manually annotated by a single dermatopathologist blinded to the mutation status of the cases. In **Fig. 7A**, the annotated mask of positive IHC staining was overlaid on the network-generated probability heat map for our highest confidence prediction. The average probability of tiles falling inside vs. outside the selected mask was calculated (see Methods) and displayed as the corresponding box plot in **Fig. 7B.** Tiles containing *BRAF*^{Mut} were significantly more likely to fall within the IHC mask compared to outside the mask ($p<3e-08$), indicating that the network indeed localizes mutated *BRAF*. Similar results were obtained for the top 10 ranked predicted *BRAF*^{V600E} cases, for the majority of which there was statistically significant concordance between the heat map and IHC (**Fig. 7C**). Surprisingly, IHC failed to detect mutated *BRAF* in two of these ten high confidence cases (**Fig. 7D**).

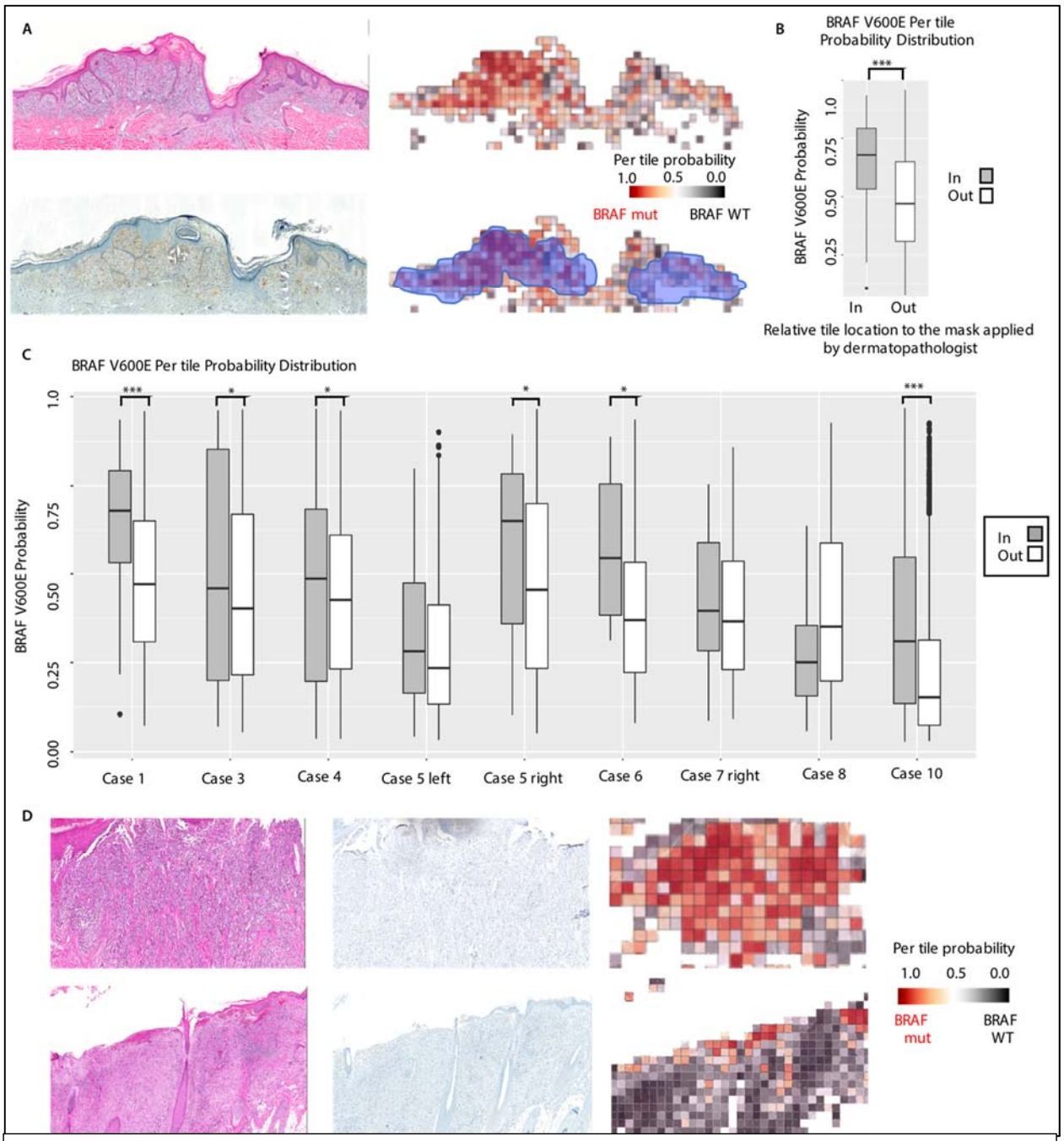


Figure 7. BRAF V600E-predicted tumor areas overlap with immunohistochemical V600E antibody staining. A) Annotated regions of positive IHC staining demonstrating overlap with the network-generated probability heat map. H&E-stained tissue section (*top left)*, IHC-stained tissue section (*bottom left*), probability heat map (*top right*), and overlay (*bottom right*) are shown. **B)** Boxplot of the probability distributions for tiles inside and outside the IHC mask. Tiles predicted to harbor *BRAF*^{Mut} were more likely to fall inside the IHC mask (*p<3e-09*). **C)** Boxplot distributions and IHC-heat map overlays of high-probability *BRAF*^{Mut} cases. P-values for statistically significant cases: Case 1: p=9e-08; Case 3: p=0.03; Case 5, right: p=0.012; Case 6: p=0.018; Case 10: p=6.6e-16. **D)** Case 2 with representative H&E-stained section, IHC-stained section, and probability heat map and Case 9 with representative H&E-stained section, IHC-stained section, and probability heat map.

Discussion

In melanoma, deep learning has previously been applied to classify pigmented lesions as benign vs. malignant using clinical[17] or dermoscopic[18] images with impressive accuracy. Nevertheless, histopathological examination remains the gold standard for the diagnosis of melanoma. In patients with localized disease, surgical excision is curative. For advanced melanomas, the development of targeted therapies, such as BRAF and MEK inhibitors, and immunotherapies, such as anti-CTLA4 and anti-PD1 antibodies, have substantially increased median overall survival[3]. Selecting the optimal treatment in these patients depends, in part, on determining the mutational status of the *BRAF* oncogene. While the ideal treatment regimen for *NRAS*-mutated melanomas is still unclear, combination therapy with MEK inhibition is under investigation[3]. Mutational testing is therefore routinely performed on Stage III and IV melanomas. Here, we use a deep learning approach on whole slide histopathology images to predict for *BRAF* and *NRAS* driver mutations in primary melanomas.

Specific morphologic signatures associated with mutated *BRAF* have been described independently with dermoscopy[19], reflectance confocal microscopy[20], and histology[14, 21]. Histologic features include greater pagetoid scatter, intraepidermal nesting, epidermal thickening, better circumscription, larger rounder and more pigmented melanocytes, and less solar elastosis. However, attempts to develop binary decision trees to predict for the *BRAF* mutation using histology alone achieved a predictive accuracy of only 60.3%[21]. In our study, we corroborate that *BRAF* mutations lead to specific morphologic changes that can be detected through deep learning and demonstrate that network performance for

predicting mutated *BRAF* is improved with thinner tumors with an accuracy of 83%. As several of the morphologic features described to be specific for mutated *BRAF* predominantly affect the epidermis, it may be the case that it is easier for the network to detect these features in thinner tumors. In studies correlating *BRAF* mutations and tumor thickness, some have found *BRAF* mutations to be associated with thinner tumors[22] while others have reported either an inverse[23, 24] or no relationship[25, 26] with Breslow depth.

Evaluating the effect of ulceration status on our institutional cohort suggests an association between ulceration and mutated *BRAF* melanomas, as there was a modest improvement in AUC for predicting mutated *BRAF* in ulcerated melanomas. In a logistic regression model using clinicopathological features, only ulceration and histologic subtype were found to be significant predictors for mutated *BRAF*[26]. Ulceration may be due to downregulation of genes involved in cell adhesion pathways through copy number losses on chromosomes 6q and 10q. Losses at 10q23-26 have been connected with *BRAF* mutations[27], providing support for an association with an ulcerative state.

So far, *NRAS*-mutated melanomas have non-specific histologic findings, such as greater mitotic index[28], fewer tumor-infiltrating lymphocytes[29], and nodular histologic subtypes[30]. Not surprisingly, attempts to predict for NRAS mutation using pre-defined morphologic features performed at random[14, 21] and studies examining *NRAS*-mutated melanomas and parameters such as tumor thickness and ulceration have yielded conflicting results[25, 31, 32]. Using deep learning, we demonstrate that *NRAS* mutations can be predicted from histopathology images, indicating that these specific morphologic features have not yet

been fully described. It is possible these features are detectable on the nuclear or chromosomal level, as *NRAS* mutations more frequently exhibit chromosomal loss of the 11q23.3-11q25 region, whereas *BRAF* mutations are associated with loss at 10q23-26 and gains at chromosome 7 and 1q23-q25[27]. This provides a structural basis that could explain, in part, how our deep learning methods are able to classify these mutations. Intriguingly, in non-ulcerated melanomas, network performance to predict mutated *NRAS* reached AUC=0.92 in our institutional cohort. Ulceration may represent a distinct biological subtype as it is the second most significant prognostic factor in melanoma survival. In addition to genetic alterations, it has been proposed that the tumor-infiltrating lymphocytes (TILs) are a critical factor in ulcerated melanomas[33]. The importance of TILs for prognosis and response to treatment is an area of active investigation in immunooncology; and deep learning has been used to create spatial maps of TILs and correlating TIL patterns with survival[34]. Thus, there may be contributions from the tumor microenvironment that influence network performance in a more substantial manner for *NRAS*-mutated melanomas.

Cross-validating our network on all images of primary melanomas from TCGA resulted in a reduced performance compared to our institutional cohort. TCGA primary melanoma specimens are enriched for thicker tumors, with a median of 2.7mm and a mean of 4.9mm[15]. We verified that this cohort contained thicker melanomas, with a median depth of 7.5mm at the time of initial diagnosis. Our network model was trained on melanomas with a more equitable distribution among all tumor stages (**Supplemental Figure 3**). Nevertheless, we were able to corroborate some of the observed trends with subgroup

330 analysis. Although there were no TCGA melanomas thinner than 1.0 mm, network
331 performance for predicting mutated *BRAF* was greatest in melanomas ≤ 5.0 mm, with a
332 continual reduction of performance as tumor thickness increased. Interestingly,
333 extrapolating BRAF network performance on the TCGA dataset in **Fig 4A** leads to a
334 predicted AUC of approximately 0.80 on melanomas ≤1.0 mm, similar to the AUC of 0.83
335 obtained on our institutional cohort. In addition, we again found that the absence of an
336 ulceration is an important factor for predicting mutated *NRAS*, with an AUC=0.89 on
337 TCGA images.
338
339 With respect to existing rapid screening tests, it is unclear to what extent
340 immunohistochemistry is being used in clinical practice. Although antibodies to detect
341 $BRAF^{V600E}$ and $NRAS^{Q61R}$ specific mutations have reported high sensitivities and
342 specificities[5], known limitations of interpretation include: variations in staining, equivocal
343 or ambiguous staining in tumors with high melanin content (>10%), samples with <10%
344 of tumor content[35], and false negatives in inappropriately fixed tissue. Furthermore,
345 despite shorter turnaround times[6], sample preparation and slide cutting still incur
346 additional time and cost. For these reasons, IHC requires optimized and standardized
347 testing protocols[36] and interpretation of results by an experienced pathologist.
348
349 In our study, we utilize IHC analysis with the monoclonal VE1 antibody to further
350 substantiate the accuracy of our model by assessing the overlay between positive IHC
351 staining of $BRAF^{V600E}$ on tissue sections and network-generated probability heat maps.
352 In 10 high probability $BRAF^{Mut}$ cases, 6 cases demonstrated excellent concordance

between positive IHC staining and the heat map, 2 cases showed no statistically significant overlap, and 2 cases were misidentified by IHC as negative. One of these false negative cases was noted by the pathologist to contain high amounts of background pigment, highlighting certain advantages deep learning may have over current screening methods.

Because whole slide image analysis will be a crucial feature for clinical adaptability, we fully automated our mutation prediction by first applying a tumor selection model on non-annotated images, achieving an AUC=0.98. The high discriminatory power of our model is demonstrated by the ability of the network to identify melanomas independent of mutation status (**Supplemental Figure 5**). Importantly, the performance of the fully automated model was comparable to the manual annotation model across all our analyses.

With the recent FDA approval of the first whole slide imaging system for primary diagnosis in pathology[37], the digitization of slides seems poised to be integrated into routine clinical practice. In the context of our mutation prediction model, primary melanomas could be rapidly screened on initial H&E slides. While we did not utilize metastatic melanoma samples in this study over concerns of suboptimal training on a smaller dataset, a number of studies demonstrate mutational testing on the primary tumor is an acceptable alternative[38, 39, 40]. Our BRAF model could potentially be used in conjunction with IHC screening, where concordant cases do not require confirmatory sequencing. Alternatively, cases that were negative for the *BRAFV600E* mutation by IHC can be

376 analyzed by deep learning in order to identify false negatives or non-V600E *BRAF*
377 mutants in patients who would benefit from targeted therapy. As others have advocated
378 using multiple detection methods for challenging samples[41] or to minimize technique-
379 related discordancy[42], additional rapid and cost-effective mutational screening techniques
380 would be highly valuable. Regarding our NRAS model, additional training with an
381 increased sample size at 40x image magnification can improve overall network
382 performance. Intriguingly, our network performs particularly well on non-ulcerated, *NRAS*-
383 mutated melanomas, suggesting there may be contributions in the tumor
384 microenvironment that warrant further study.

385

386 There is great promise for advanced computational approaches to be integrated into
387 clinical care. Beyond predicting mutations, our study lays the groundwork for more
388 sophisticated deep learning models based on histopathology images, such as predicting
389 for treatment responders vs. non-responders or even survival outcomes, as has been
390 previously demonstrated in lung cancers[43] and gliomas[44]. We present a fully automated
391 deep CNN model that accurately differentiates melanomas from benign tissue and uses
392 morphologic features to predict the presence of *BRAF* or *NRAS* driver mutations. Not only
393 has this approach provided additional insight into how these mutations may affect tumor
394 structural characteristics and its surrounding environment, our models have the potential
395 to complement existing mutation screening assays, with the advantage of significantly
396 reducing costs, and importantly, expediting the decision-making process for treatment.

397

398

Materials and Methods

Dataset of whole-slide images

All patients were enrolled in an IRB-approved clinicopathological database and biorepository in the Interdisciplinary Melanoma Cooperative Group (IMCG) at NYU Langone Health. The IMCG collects prospective clinical, pathological, and follow-up data from melanoma patients who present for diagnosis and/or treatment[45]. 365 H&E-stained FFPE whole-slides from 324 primary melanomas diagnosed between 1994 to 2013 were retrieved and digitized at 20x magnification. A single board-certified dermatopathologist (RHK) reviewed all digitized slides for image quality and excluded images that were blurry, faded, or did not contain any tumor. 293 images from 257 melanomas were subsequently annotated by RHK for tumor-rich regions of interest (ROIs) using Aperio ImageScope software. Driver mutations were previously determined by Sanger sequencing.

Software availability

We utilized the adapted Tensorflow pipeline (https://github.com/ncoudray/DeepPATH.git) to perform our analysis using the Inception v3 CNN architecture.

Image pre-processing

Whole-slide images were partitioned at 20x magnification into non-overlapping 299x299 pixel “tiles”. This process generated 794,588 total tiles in our dataset, after removing tiles with more than 50% background (white area of slides). All tiles take the label of the slide they belong to and are sorted in training, validation and independent sets comprising of

70%, 15% and 15% of the total number of tiles correspondingly. Tiles of images coming from the same patient are all included in the same set. Tiles in the train and validation sets were then converted to TF record format, which is necessary for training of Inception v3, in groups of 1024 tiles in each TF record file for the training set and 128 tiles for the validation set.

Deep learning with Convolutional Neural Network

The Inception v3 architecture is a Convolutional Neural Network (CNN) that utilizes modules comprised of various convolutions with different kernel sizes and a max pooling layer. The network was trained on 70% of the tiles from each data set, with 15% of the tiles used for validation and 15% used for independent testing. The network was trained for maximum 500,000 iterations on batches of 30 images with a step of 5,000 iterations. The activation function used in the output layer was softmax. The network's performance was monitored based on the precision on the validation set. The best performing model was chosen when the difference between the precision of the current model and the minimum precision of the previous 5 models was less than 0.01, indicating a plateau in precision. The performance of the best model was then evaluated on the independent set (44 slides) and the AUC was calculated. The network outputs a probability value for every tile for each class of interest. The tile is assigned to the class with the highest probability. A heat map for each slide in the test set is generated. The color intensity is analogous to the probability value of the tile to belong in each class.

Network performance on the data from The Cancer Genome Atlas

71 FFPE slides of primary melanomas from the TCGA were downloaded and tiled into non-overlapping tiles of 299x299 pixels. All tiles were sorted for testing and TFRecord files were generated. The slides were passed through the mutation prediction networks and the average probabilities per slide were used for the AUC calculation.

Automated tumor selection

Whole images were tiled in non-overlapping tiles of 299x299 pixels. The tiles were sorted based on their position compared to the manual tumor selection applied by the dermatopathologist as 'in' and 'out' tumor, and were divided in train, validation and independent sets the same way as for the mutation prediction networks. The same 44 slides were kept as the independent set. The Inception v3 model was trained on these two classes of tiles and the performance of the best performing model on the validation set was measured on the independent set.

Annotated and automated mutation prediction

For the annotated model, only tiles belonging inside the annotated tumor area were taken into consideration. For the automated model, only tiles belonging inside the tumor area as determined by the tumor annotation network were selected for training. The tiles were sorted in two categories depending on the mutation classification task and based on the true label of the slide they belong to. They were also divided in train, validation and independent set as before. Inception v3 was trained on the tiles and performance was

467 monitored based on the precision on the validation set. The best performing model was
468 obtained and evaluated on the independent set.
469
470 *Sequential network*
471 To apply the sequential model, non-annotated slides of interest were tiled in non-
472 overlapping 299x299 pixel tiles. The tiles were first passed through the automated tumor
473 selection network which will output a probability for each tile belonging in the tumor area.
474 Tiles with probability of belonging in the tumor higher or equal than 0.5 were obtained and
475 passed through the BRAF and NRAS mutation prediction networks to assess the
476 mutational profile of the slide.
477
478 *Statistical analysis*
479 After training and choosing the best performing model on the validation set, model
480 performance was evaluated using the independent dataset, which is comprised of a held-
481 out population of tiles coming from 44 slides. The probabilities for each slide were
482 aggregated by the average of probabilities of the corresponding tiles or by the percentage
483 of tiles positively classified. Receiver Operative Characteristic (ROC) curves and the
484 corresponding Area Under the Curve (AUC) were generated as a measure of accuracy.
485 Heat maps allowed visualization of probability differences and regions of interest.
486
487 *Multivariate model*
488 The multivariate logistic regression model was built using the *glm* function in R from the
489 package ROCR.

Receiver Operating Characteristic Curves

ROC curves were generated using the pROC package in R and the p-values were calculated using the roc.test() function.

Immunohistochemical analysis of mutated BRAF V600E

Immunohistochemistry (IHC) was performed on 10% neutral buffered FFPE, 4-µm human archival melanoma sample sections collected on plus slides (Fisher Scientific, Cat# 22-042-924) and stored at room temperature. Unconjugated, mouse anti-human Serine-Threonine-Protein Kinase B-raf (BRAF) V600E, clone VE1 (Abcam Cat# ab228461, Lot# GR32335840-6) raised against a synthetic peptide within human BRAF (amino acids 550-650) containing the glutamic acid substitution, was used for IHC[38, 46]. BRAF antibody was optimized on known positive and negative colon samples and subsequently validated on a mixed set 20 known positive/negative samples. Chromogenic immunohistochemistry was performed on a Ventana Medical Systems Discovery Ultra using Ventana's reagents and detection kits unless otherwise noted. In brief, slides were deparaffinized online and antigen retrieved for 24 minutes at 95°C using Cell Conditioner 1 (Tris-Borate-EDTA pH8.5). BRAF was diluted 1:50 in Ventana antibody diluent (Ventana Medical Systems, Cat# 251-018) and incubated for 16 minutes at 36°C. Endogenous peroxidase activity was post-primary blocked with 3% hydrogen peroxide for 4 minutes. Primary antibody was detected using Optiview linker followed by multimer-HRP incubated for 8 minutes each, respectively. The complex was visualized with 3,3 diaminobenzidene for 8 minutes and enhanced with copper sulfate for 4 minutes. Slides were counterstained online with

hematoxylin for 8 minutes and blued for 4 minutes. Slides were washed in distilled water, dehydrated and mounted with permanent media. Positive and negative (diluent only) controls were run in parallel with study sections. Blinded analysis of staining was performed by a single dermatopathologist (GJ).

Calculation of BRAF V600E-predicted tumor areas overlap with immunohistochemical V600E antibody staining.

By looking at the relative positioning between the IHC and H&E slides, a direction of shift was chosen and the slides were shifted by a small shift of 1 or maximum 2 tiles towards the observed direction to better align the slides for overlap. Next, the probability distributions for the tiles falling into the mask applied by a dermatopathologist to select for the V600E antibody stained area and the probabilities of the ones outside of the mask were generated. The p value was calculated using a one-sided Wilcoxon rank sum test. The p values were also adjusted for the number of potential conformations for each slide (shift by 1, shift by 2 and no-shift) by multiplication with a factor of 3 (multiple testing correction).

Supplementary Materials

References:

1. Ascierto PA*, et al.* The role of BRAF V600 mutation in melanoma. *J Transl Med* **10**, 85 (2012).

2. Sun J, Zager JS, Eroglu Z. Encorafenib/binimetinib for the treatment of BRAF-mutant advanced, unresectable, or metastatic melanoma: design, development, and potential place in therapy. *Onco Targets Ther* **11**, 9081-9089 (2018).

3. Luke JJ, Flaherty KT, Ribas A, Long GV. Targeted agents and immunotherapies: optimizing outcomes in melanoma. *Nat Rev Clin Oncol* **14**, 463-482 (2017).

4. Cheng L, Lopez-Beltran A, Massari F, MacLennan GT, Montironi R. Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine. *Mod Pathol* **31**, 24-38 (2018).

5. Barel F, Guibourg B, Lambros L, Le Flahec G, Marcorelles P, Uguen A. Evaluation of a Rapid, Fully Automated Platform for Detection of BRAF and NRAS Mutations in Melanoma. *Acta Derm Venereol* **98**, 44-49 (2018).

6. Bisschop C*, et al.* Rapid BRAF mutation tests in patients with advanced melanoma: comparison of immunohistochemistry, Droplet Digital PCR, and the Idylla Mutation Platform. *Melanoma Res* **28**, 96-104 (2018).

7. Colomba E*, et al.* Detection of BRAF p.V600E mutations in melanomas: comparison of four methods argues for sequential use of immunohistochemistry and pyrosequencing. *J Mol Diagn* **15**, 94-100 (2013).

8. Khosravi P, Kazemi E, Imielinski M, Elemento O, Hajirasouliha I. Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images. *EBioMedicine* **27**, 317-328 (2018).

9. Wang Z, Jensen MA, Zenklusen JC. A Practical Guide to The Cancer Genome Atlas (TCGA). *Methods Mol Biol* **1418**, 111-141 (2016).

10. Coudray N*, et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* **24**, 1559-1567 (2018).

11. Couture HD*, et al.* Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ Breast Cancer* **4**, 30 (2018).

12. Menzies AM*, et al.* Intrapatient homogeneity of BRAFV600E expression in melanoma. *Am J Surg Pathol* **38**, 377-382 (2014).

13. Cormican D, Kennedy C, Murphy S, Werner R, Power DG, Heffron C. High concordance of BRAF mutational status in matched primary and metastatic melanoma. *J Cutan Pathol* **46**, 117-122 (2019).

14. Viros A*, et al.* Improving melanoma classification by integrating genetic and morphologic features. *PLoS Med* **5**, e120 (2008).

15. Cancer Genome Atlas N. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681-1696 (2015).

16. Capper D*, et al.* Assessment of BRAF V600E mutation status by immunohistochemistry with a mutation-specific monoclonal antibody. *Acta Neuropathol* **122**, 11-19 (2011).

17. Esteva A*, et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115-118 (2017).

18. Haenssle HA*, et al.* Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* **29**, 1836-1842 (2018).

19. Armengot-Carbo M, Nagore E, Garcia-Casado Z, Botella-Estrada R. The association between dermoscopic features and BRAF mutational status in cutaneous melanoma: significance of the blue-white veil. *J Am Acad Dermatol*, (2018).

20. Colombino M*, et al.* Dermoscopy and confocal microscopy for metachronous multiple melanomas: morphological, clinical, and molecular correlations. *Eur J Dermatol* **28**, 149-156 (2018).

21. Broekaert SM*, et al.* Genetic and morphologic features for melanoma classification. *Pigment Cell Melanoma Res* **23**, 763-770 (2010).

22. Liu W*, et al.* Distinct clinical and pathological features are associated with the BRAF(T1799A(V600E)) mutation in primary melanoma. *J Invest Dermatol* **127**, 900-905 (2007).

23. Ponti G*, et al.* BRAF, NRAS and C-KIT Advanced Melanoma: Clinico-pathological Features, Targeted-Therapy Strategies and Survival. *Anticancer Res* **37**, 7043-7048 (2017).

24. Mitchell B, Leone DA, Feller JK, Yang S, Mahalingam M. BRAF and epithelial-mesenchymal transition in primary cutaneous melanoma: a role for Snail and E-cadherin? *Hum Pathol* **52**, 19-27 (2016).

25. Heppt MV*, et al.* Prognostic significance of BRAF and NRAS mutations in melanoma: a German study from routine care. *BMC Cancer* **17**, 536 (2017).

26. Spathis A*, et al.* BRAF Mutation Status in Primary, Recurrent, and Metastatic Malignant Melanoma and Its Relation to Histopathological Parameters. *Dermatol Pract Concept* **9**, 54-62 (2019).

27. Lazar V*, et al.* Marked genetic differences between BRAF and NRAS mutated primary melanomas as revealed by array comparative genomic hybridization. *Melanoma Res* **22**, 202-214 (2012).

28. Devitt B*, et al.* Clinical outcome and pathological features associated with NRAS mutation in cutaneous melanoma. *Pigment Cell Melanoma Res* **24**, 666-672 (2011).

29. Thomas NE, *et al.* Association Between NRAS and BRAF Mutational Status and Melanoma-Specific Survival Among Patients With Higher-Risk Primary Melanoma. *JAMA Oncol* **1**, 359-368 (2015).

30. Lee JH, Choi JW, Kim YS. Frequencies of BRAF and NRAS mutations are different in histological types and sites of origin of cutaneous melanoma: a meta-analysis. *Br J Dermatol* **164**, 776-784 (2011).

31. Si L, *et al.* Prevalence of BRAF V600E mutation in Chinese melanoma patients: large scale analysis of BRAF and NRAS mutations in a 432-case cohort. *Eur J Cancer* **48**, 94-100 (2012).

32. Ellerhorst JA, *et al.* Clinical correlates of NRAS and BRAF mutations in primary human melanoma. *Clin Cancer Res* **17**, 229-235 (2011).

33. de Moll EH, *et al.* Immune biomarkers are more accurate in prediction of survival in ulcerated than in non-ulcerated primary melanomas. *Cancer Immunol Immunother* **64**, 1193-1203 (2015).

34. Saltz J, *et al.* Spatial Organization and Molecular Correlation of Tumor-Infiltrating Lymphocytes Using Deep Learning on Pathology Images. *Cell Rep* **23**, 181-193 e187 (2018).

35. Fisher KE, Cohen C, Siddiqui MT, Palma JF, Lipford EH, 3rd, Longshore JW. Accurate detection of BRAF p.V600E mutations in challenging melanoma specimens requires stringent immunohistochemistry scoring criteria or sensitive molecular assays. *Hum Pathol* **45**, 2281-2293 (2014).

36. Long E, *et al.* Why and how immunohistochemistry should now be used to screen for the BRAFV600E status in metastatic melanoma? The experience of a single institution (LCEP, Nice, France). *J Eur Acad Dermatol Venereol* **29**, 2436-2443 (2015).

37. Evans AJ, *et al.* US Food and Drug Administration Approval of Whole Slide Imaging for Primary Diagnosis: A Key Milestone Is Reached and New Questions Are Raised. *Arch Pathol Lab Med*, (2018).

38. Nielsen LB, Dabrosin N, Sloth K, Bonnelykke-Behrndtz ML, Steiniche T, Lade-Keller J. Concordance in BRAF V600E status over time in malignant melanoma and corresponding metastases. *Histopathology* **72**, 814-825 (2018).

39. Manfredi L, *et al.* Highly Concordant Results Between Immunohistochemistry and Molecular Testing of Mutated V600E BRAF in Primary and Metastatic Melanoma. *Acta Derm Venereol* **96**, 630-634 (2016).

40. Boursault L, *et al.* Tumor homogeneity between primary and metastatic sites for BRAF status in metastatic melanoma determined by immunohistochemical and molecular testing. *PLoS One* **8**, e70826 (2013).

41. Uguen A, *et al.* Dual NRASQ61R and BRAFV600E mutation-specific immunohistochemistry completes molecular screening in melanoma samples in a routine practice. *Hum Pathol* **46**, 1582-1591 (2015).

42. Bruno W, *et al.* Heterogeneity and frequency of BRAF mutations in primary melanoma: Comparison between molecular methods and immunohistochemistry. *Oncotarget* **8**, 8069-8082 (2017).

43. Yu KH, *et al.* Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* **7**, 12474 (2016).

44. Mobadersany P, *et al.* Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* **115**, E2970-E2979 (2018).

45. Wich LG, *et al.* Developing a multidisciplinary prospective melanoma biospecimen repository to advance translational research. *Am J Transl Res* **1**, 35-43 (2009).

46. Piris A, Mihm MC, Jr., Hoang MP. BAP1 and BRAFV600E expression in benign and malignant melanocytic proliferations. *Hum Pathol* **46**, 239-245 (2015).

Acknowledgments: We thank Luis Chiriboga from the NYU Experimental Pathology Immunohistochemistry Core Laboratory. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. **Funding**: This research was supported, in part, by the NYU School of Medicine Orbuch-Brand Pilot Grant Program for Cancers of the Skin; by the Laura and Isaac Perlmutter Cancer Center Support Grant; NIH/NCI P30CA016087; and by the National Institutes of Health S10 Grants; NIH/ORIP S10OD01058 and S10OD018338. AT is supported by the American Cancer Society (RSG-15-189-01-RMC). **Author contributions**: Study concept and design: RHK, SN, IO, AT. Acquisition of data: RHK, SN, ZD, GJ, UM, RLS, RSB. Analysis and interpretation of data: RHK, SN, NC, GJ, JSW, NR, IO, AT; Study supervision: NC, IO, AT. **Competing interests**: JSW declares the following competing interests:

Stock or Other Ownership: Altor BioScience, Biond, CytomX Therapeutics, Protean Biodiagnostics

Honoraria: Bristol-Myers Squibb, Merck, Genentech, AbbVie, AstraZeneca, Daiichi Sankyo, GlaxoSmithKline, Eisai, Altor BioScience, Amgen, Roche, Ichor Medical Systems, Celldex, CytomX Therapeutics, Nektar, Novartis, Sellas, WindMIL, Takeda, Protean Biodiagnostics

Consulting or Advisory Role: Celldex, Ichor Medical Systems, Biond, Altor BioScience, Bristol-Myers Squibb, Merck, Genentech, Roche, Amgen, AstraZeneca, GlaxoSmithKline, Daiichi Sankyo, AbbVie, Eisai, CytomX Therapeutics, Nektar, Novartis, Sellas, WindMIL, Takeda

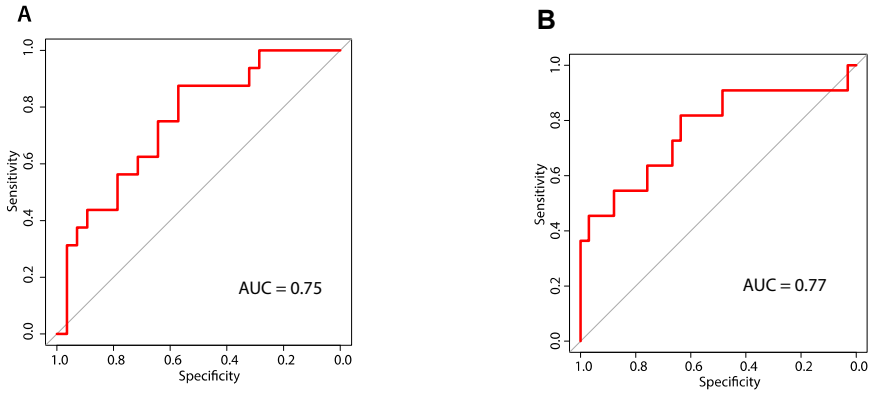
Research Funding (to the Institution): Bristol-Myers Squibb, Merck, GlaxoSmithKline, Genentech, Astellas Pharma, Incyte, Roche, Novartis

Travel, Accommodations, Expenses: Bristol-Myers Squibb, GlaxoSmithKline, Daiichi Sankyo, Roche, Celldex, Amgen, Merck, AstraZeneca, Genentech, Novartis, WindMIL, Takeda

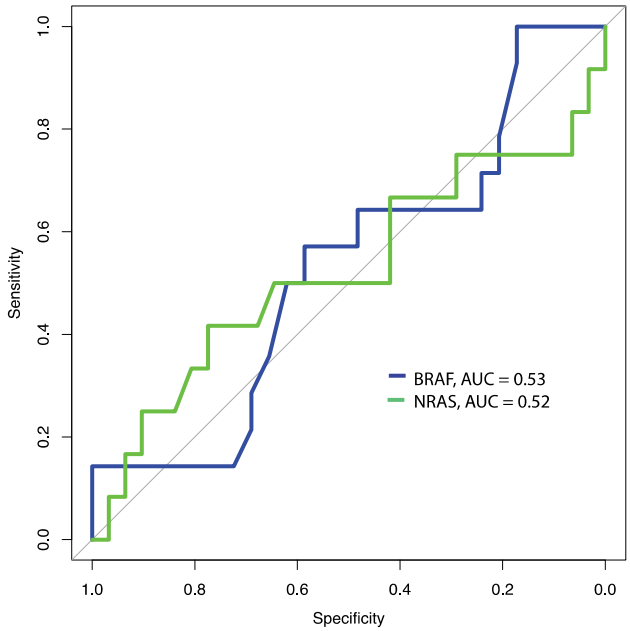
Table 1. Patient characteristics

Cohorts		Training						Validation						Independent					
Unique Patients (n)		182						43						41					
Mutation		BRAF		NRAS		WT/WT		BRAF		NRAS		WT/WT		BRAF		NRAS		WT/WT	
		n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
		56	30.8	65	35.7	61	33.5	14	32.6	12	27.9	17	39.5	21	51.2	9	22.0	11	26.8
Year of Diagnosis	<2000	2	3.6	4	6.2	1	6.2	3	21.4	0	0	1	5.9	1	4.8	0	0	0	0
	2001 - 2010	50	89.3	48	73.8	52	85.2	8	57.1	12	100	12	70.6	19	90.5	9	100	11	100
	>2010	4	7.1	13	20.0	8	13.1	3	21.4	0	0	4	23.5	1	4.8	0	0	0	0
Age, Avg. +/- Std.		59.4 +/- 16.7		61.2 +/- 17.2		65.1 +/- 14.9		53.9 +/- 19.9		70.9 +/- 16.0		68.5 +/- 14.1		59.0 +/- 14.8		59.9 +/- 14.4		65.1 +/- 12.8	
Sex	Female	23	41.1	29	44.6	24	39.3	9	64.3	7	58.3	7	41.2	6	28.6	6	66.7	2	18.2
	Male	33	58.9	36	55.4	37	60.7	5	35.7	5	41.7	10	58.8	15	71.4	3	33.3	9	81.8
Thickness	<1.01 mm	8	14.3	18	27.7	20	32.8	2	14.3	3	27.3	3	17.6	3	14.3	2	22.2	3	27.3
	1.01 - 2.0 mm	14	25.0	19	29.2	15	24.6	6	42.9	4	36.4	7	41.2	6	28.6	5	55.6	2	18.2
	2.01 - 4.0 mm	18	32.1	20	30.8	15	24.6	5	35.7	1	9.1	4	23.5	5	23.8	1	11.1	4	36.5
	> 4.0	16	28.6	8	12.3	11	18.0	1	7.1	3	27.3	3	17.6	7	33.3	1	11.1	2	18.2
Thickness, Median (IQR)		2.7 (1.7 – 4.6)		1.8 (1 – 2.7)		1.7 (0.9 – 3.0)		1.4 (1.2 – 3.3)		1.8 (1.0 - 3.9)		1.5 (1.1 – 3.0)		2.7 (1.6 – 5.5)		1.5 (1.0-1.9)		2.3 (1.1 – 3.0)	
Histologic subtype	Superficial Spreading	22	39.3	29	44.6	32	52.5	8	57.1	6	50.0	9	52.9	9	42.9	4	44.4	5	45.5
	Nodular	31	55.4	34	52.3	25	41.0	6	42.9	6	50.0	8	47.1	12	57.1	5	55.6	6	54.5
	Lentigo Maligna	0	0.0	0	0.0	1	1.6	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
	Desmoplastic	3	5.4	1	1.5	3	4.9	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
	Unknown	0	0.0	1	1.5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Ulceration	Absent	30	53.6	44	67.7	40	65.6	9	64.3	7	58.3	11	64.7	11	52.4	6	66.7	6	54.5
	Present	25	44.6	20	30.8	21	34.4	5	35.7	5	41.7	6	35.3	10	47.6	3	33.3	5	45.5
	Unknown	1	1.8	1	1.5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Mitotic Index	Absent	8	14.3	8	12.3	12	19.7	2	14.3	1	8.3	2	11.8	4	19.0	0	0.0	1	9.1
	Few	14	25.0	23	35.4	25	41.0	1	7.1	3	25.0	5	29.4	3	14.3	7	77.8	5	45.5
	Moderate	16	28.6	14	21.5	10	16.4	6	42.9	3	25.0	5	29.4	5	23.8	1	11.1	2	18.2
	Many	17	30.4	20	30.8	14	23.0	5	35.7	5	41.7	5	29.4	9	42.9	1	11.1	3	27.3
	Unknown	1	1.8	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
AJCC Stage	I	15	26.8	31	47.7	31	50.8	7	50.0	7	58.3	8	47.1	6	28.6	6	66.7	3	27.3
	II	23	41.1	21	32.3	18	29.5	1	7.1	2	16.7	6	35.3	9	42.9	2	22.2	3	27.3
	III	18	32.1	12	18.5	12	19.7	6	42.9	3	25.0	3	17.6	6	28.6	1	11.1	5	45.5
	IV	0	0.0	1	1.5	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Anatomic Site	Axial	29	51.8	21	32.3	27	44.3	3	21.4	3	25.0	7	41.2	11	52.4	4	44.4	4	36.4
	Extremity	17	30.4	34	52.3	22	36.1	10	71.4	9	75.0	9	52.9	7	33.3	5	55.6	2	18.2
	Head and Neck	10	17.9	10	15.4	12	19.7	1	7.1	0	0.0	1	5.9	3	14.3	0	0.0	5	45.5
Status	Alive	33	58.9	40	58.9	42	58.9	9	58.9	8	58.9	12	58.9	11	58.9	8	58.9	6	54.5
	Died of Melanoma	20	37.0	18	27.7	14	25.5	5	45.5	4	30.8	1	5.3	9	47.4	1	12.5	5	71.4
	Died of Other Cause	3	5.1	7	9.7	5	7.6	0	0.0	0	0.0	4	19.0	1	4.5	0	0.0	0	0.0
Recurrence	No	31	55.4	40	61.5	36	59.0	6	42.9	9	75.0	14	82.4	9	42.9	7	77.8	2	18.2
	Yes	25	44.6	25	38.5	25	41.0	8	57.1	3	25.0	3	17.6	12	57.1	2	22.2	9	81.8

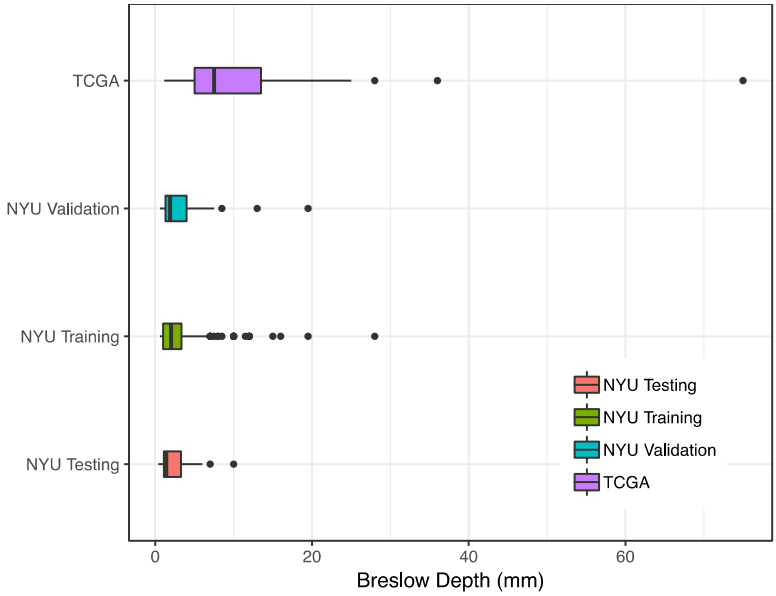
Supplemental Material



Supplemental Figure 1. Mutation prediction classifiers with manual annotation. A) Receiver Operating Characteristic curve (ROC) and AUC for BRAF vs "non-BRAF" mutation prediction on the entire independent set. **B)** Receiver Operating Characteristic curve (ROC) and AUC for NRAS vs "non-NRAS" mutation prediction on the entire independent set.

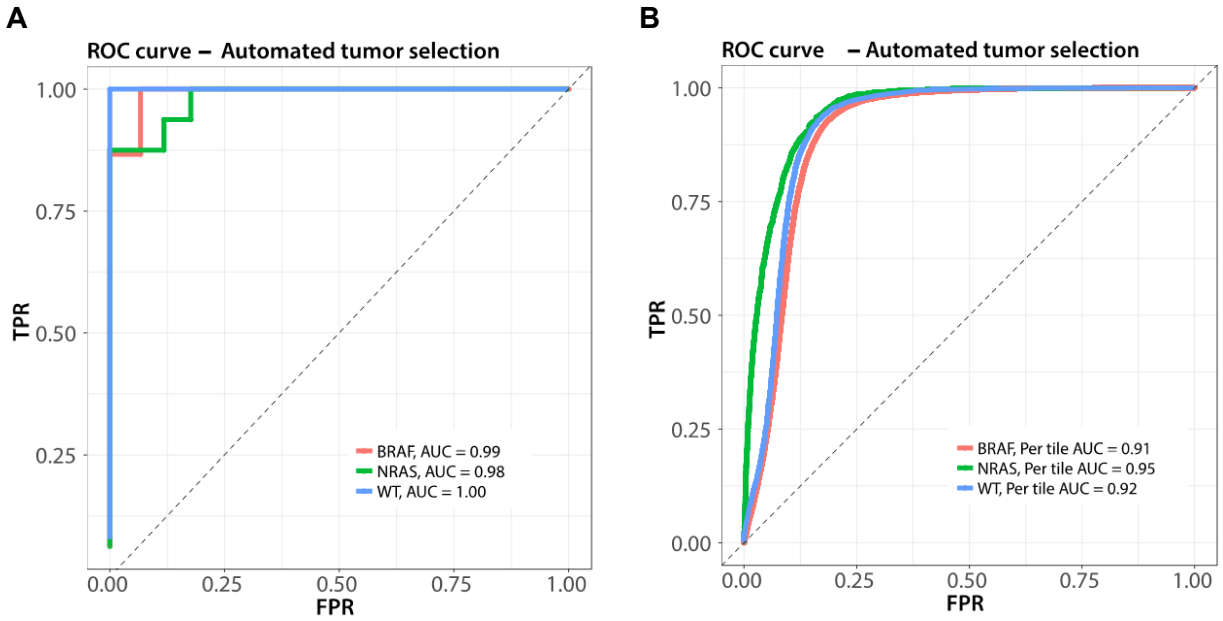


Supplemental Figure 2. ROC curves for the multivariate logistic regression model. The Breslow depth and ulceration variables are not sufficient alone to predict BRAF and NRAS mutations in melanomas, yielding random AUCs.



Supplemental figure 3

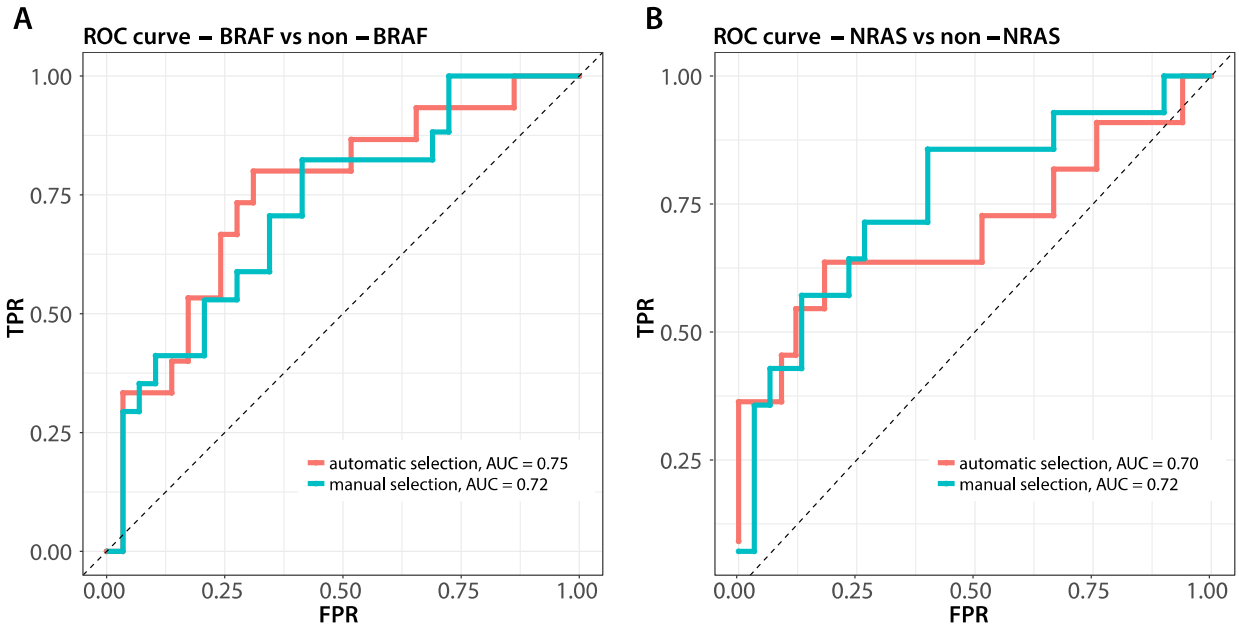
Breslow depth distribution for our NYU cohorts and the external TCGA validation cohort. It can be observed that the TCGA melanomas are overall much thicker than the ones in our cohort.



Supplemental Figure 4. Performance of automated tumor selection network.

A) AUC aggregated per slide

B) AUC per tile



Supplemental Figure 5. Performance of mutation networks after manual vs. automated tumor selection

A) BRAF mutation model

B) NRAS mutation model

Supplemental Table 1
The distribution of the number of patients and the corresponding number of slides within each cohort are shown. No patients within the training and validation cohorts overlap with those in the independent cohort.

	Number of Patients				Number of slides			
	WT	BRAF	NRAS	TOTAL	WT	BRAF	NRAS	TOTAL
Train/Validation	72	73	75	220	80	87	82	249
Independent	15	13	9	37	16	16	12	44
TOTAL	87	86	84	257	96	103	94	293

Supplemental Table 2
BRAF and *NRAS* prediction AUCs on the independent NYU test set and the TCGA FFPE cohort for different values of Breslow depth.

	AUC value for BRAF mutation prediction		AUC value for NRAS mutation prediction	
Tumor stage	NYU cohort	TCGA cohort	NYU cohort	TCGA cohort
<=1mm	0.83 95% CI[0.45-1]	-	0.73 95% CI[0.32-1]	-
<=5mm	0.74 95% CI[-0.58-0.89]	0.71 95% CI[0.35-1]	0.84 95% CI[0.70-0.98]	0.54 95% CI[0.07 - 1]
<=10mm	0.75 95% CI[0.60-0.90]	0.66 95% CI[0.41-0.91]	0.77 95% CI[0.58-0.96]	0.68 95% CI[0.32-1]
<=15mm	-	0.6 95% CI[0.37-0.83]	-	0.66 95% CI[0.40-0.92]
<=20mm	-	0.61 95% CI[0.39-0.83]	-	0.66 95% CI[0.40-0.92]

Supplemental Table 3
BRAF and *NRAS* prediction AUCs on the independent NYU test set and the TCGA FFPE cohort for slides with different ulceration status.

	AUC value for BRAF mutation prediction		AUC value for NRAS mutation prediction	
Ulceration	NYU cohort	TCGA cohort	NYU cohort	TCGA cohort
Present	0.79 95% CI[0.56-1]	0.55 95% CI[0.35-0.71]	0.45 95% CI[0.07-0.83]	0.45 95% CI[0.24-0.67]
Absent	0.71 95% CI[0.50-0.92]	0.6 95% CI[0.22-0.98]	0.92 95% CI[0.81-1]	0.89 95% CI[0.67-1]