Integrating 31-Gene Expression Profiling With Clinicopathologic Features to Optimize Cutaneous Melanoma Sentinel Lymph Node Metastasis Prediction

Eric D. Whitman, MD[1]; Vadim P. Koshenkov, MD[2]; Brian R. Gastman, MD[3]; Deri Lewis, MD[4]; Eddy C. Hsueh, MD[5]; Ho Pak, MD[6]; Thomas P. Trezona, MD[7]; Robert S. Davidson, MD[8]; Michael McPhee, MD[9]; J. Michael Guenther, MD[10]; Paul Toomey, MD[11]; Franz O. Smith, MD[12]; Peter D. Beitsch, MD[13]; James M. Lewis, MD[14]; Andrew Ward, NP[14]; Shawn E. Young, MD[15]; Parth K. Shah, MD[15]; Ann P. Quick, PhD[16]; Brian J. Martin, PhD[16]; Olga Zolochevska, PhD[16]; Kyle R. Covington, PhD[16]; Federico A. Monzon, MD[16]; Matthew S. Goldberg, MD[16]; Robert W. Cook, PhD[16]; Martin D. Fleming, MD[17]; David M. Hyams, MD[18]; and John T. Vetto, MD[19]

abstract

PURPOSE National guidelines recommend sentinel lymph node biopsy (SLNB) be offered to patients with > 10% likelihood of sentinel lymph node (SLN) positivity. On the other hand, guidelines do not recommend SLNB for patients with T1a tumors without high-risk features who have < 5% likelihood of a positive SLN. However, the decision to perform SLNB is less certain for patients with higher-risk T1 melanomas in which a positive node is expected 5%-10% of the time. We hypothesized that integrating clinicopathologic features with the 31-gene expression profile (31-GEP) score using advanced artificial intelligence techniques would provide more precise SLN risk prediction.

METHODS An integrated 31-GEP (i31-GEP) neural network algorithm incorporating clinicopathologic features with the continuous 31-GEP score was developed using a previously reported patient cohort (n = 1,398) and validated using an independent cohort (n = 1,674).

RESULTS Compared with other covariates in the i31-GEP, the continuous 31-GEP score had the largest likelihood ratio ($G^2 = 91.3$, $P < .001$) for predicting SLN positivity. The i31-GEP demonstrated high concordance between predicted and observed SLN positivity rates (linear regression slope = 0.999). The i31-GEP increased the percentage of patients with T1-T4 tumors predicted to have < 5% SLN-positive likelihood from 8.5% to 27.7% with a negative predictive value of 98%. Importantly, for patients with T1 tumors originally classified with a likelihood of SLN positivity of 5%-10%, the i31-GEP reclassified 63% of cases as having < 5% or > 10% likelihood of positive SLN, for a more precise, personalized, and clinically actionable SLN-positive likelihood estimate.

CONCLUSION These data suggest the i31-GEP could reduce the number of SLNBs performed by identifying patients with likelihood under the 5% threshold for performance of SLNB and improve the yield of positive SLNBs by identifying patients more likely to have a positive SLNB.

INTRODUCTION

The American Joint Committee on Cancer stages patients according to tumor characteristics, nodal disease burden, and tumor metastasis status to estimate each patient's risk of death because of melanoma.[1] Detection of metastatic disease in one or more lymph nodes qualifies patients for adjuvant therapy and is often identified through sentinel lymph node biopsy (SLNB). However, as many as 88% of patients who undergo SLNB have a negative result and, therefore, are not eligible for adjuvant therapy but are exposed to additional costs and risks of this surgical procedure.[2]

National Comprehensive Cancer Network (NCCN) guidelines recommend that clinicians discuss and offer SLNB to patients if the likelihood of having a positive sentinel lymph node (SLN) is > 10% (T2-T4 tumors), discuss and consider SLNB if the likelihood is 5%-10% (T1a tumors with high-risk features or T1b tumors), and avoid SLNB if the likelihood is < 5% (T1a tumors without high-risk features).[3] The probability of a positive SLN rises with increasing tumor depth and other high-risk primary tumor features, such as ulceration, high mitotic rate (MR), lymphovascular invasion (LVI), uncertain microstaging, and younger age.[1,4-10] However, the tenuous association of many of

CONTEXT

Key Objective

Can integration of the continuous 31-gene expression profile (31-GEP) score with clinicopathologic features (integrated 31-GEP [i31-GEP]) accurately predict sentinel lymph node (SLN) positivity for improved personalized medicine? This study uses a neural network model to combine the 31-GEP with readily available clinicopathologic features to determine each patient's positive SLN risk.

Knowledge Generated

The 31-GEP continuous score was the most important variable contributing to the i31-GEP algorithm. In an independent validation cohort, the i31-GEP algorithm produced a personalized precise likelihood of SLN positivity, which aligned with observed SLN positivity. In patients with T1-T4 tumors, the i31-GEP increased the percentage of patients predicted to have < 5% SLN positivity risk from 8.5% using current staging guidelines to 27.7%. Specifically, for patients predicted to have 5%-10% risk by current guidelines, the i31-GEP reclassified 63% of cases as having an actual SLN positivity risk of < 5% or > 10%. Finally, the i31-GEP can identify patients with > 10% SLN positivity risk who may be offered sentinel lymph node biopsy.

Relevance

The i31-GEP can improve personalized risk estimates for SLN positivity.

these features with melanoma metastasis contributes to interstudy variability of SLN positivity rates, particularly in tumors ≤ 1.0 mm in thickness (T1).[4-8]

The subjectivity associated with identifying and measuring melanoma-associated clinicopathologic (CP) features may contribute to controversy among physicians regarding which CP features and thresholds should be used to inform decisions about SLNB,[11,12] evidenced by the wide variability in reported SLN positivity rates.[5-8,13-18] Because of the high cost and increased complication rate of SLNB (10%-16%), better identification of patients likely to have a positive SLN may be possible by integrating molecular profiling with CP features.[19,20]

Currently, a 31-gene expression profile (31-GEP) test for prognosis of melanoma recurrence risk has been developed that yields a continuous probability score between zero and one that stratifies risk of disease recurrence and is binned into one of four categories: lowest (Class 1A; 0-0.41), low (Class 1B; 0.42-0.49), increased (Class 2A; 0.50-0.58), and highest (Class 2B; 0.59-1).[21-23] The 31-GEP Class is an independent predictor of recurrence, including nodal recurrence, as demonstrated in meta-analyses and multiple prospective and retrospective studies.[23-29] Vetto et al[24] showed that patients with T1-T2 tumors, 55 years and older, and a 31-GEP Class 1A result have < 5% risk of SLN positivity and maintain > 99% melanoma-specific survival. However, to develop a more individualized SLN positivity risk assessment for each patient, we used advanced artificial intelligence techniques to integrate the continuous 31-GEP score with individual patient tumor CP features.

Previously published models for SLN positivity have used simple logistic regression and focused predominantly on intermediate-thickness tumors from a biased patient population of elevated-risk individuals who do not represent the general population seen in the clinic[30] or used metastatic tissue irrelevant for melanoma cases in which SLNB guidance is needed.[31] Some of the more commonly used models have limited validation,[30,32] incorporate outdated variables,[33,34] or do not incorporate molecular tumor data.[32-34]

To overcome the current limitations facing SLN positivity prediction, we developed and independently validated an artificial intelligence–based neural network algorithm to predict SLN positivity risk in patients with T1-T4 CM by integrating the continuous 31-GEP score with CP features (integrated 31-GEP [i31-GEP]). An objective, individualized approach to SLNB decision making could reduce unnecessary procedures and associated risks in patients with a low likelihood of having a positive SLNB. Similarly, this approach could identify previously unidentified or subjectively excluded patients with a higher likelihood (> 10% risk) of SLN positivity who may benefit most from an SLNB.

METHODS

Patient Demographics

Development cohort. The development cohort has been previously described.[24] The model was trained on 1,398 patients age ≥ 18 years with primary CM tumors of known Breslow thickness (T1-T4), with a continuous 31-GEP test result, and with either clinically (287 of 1,398; 20.5%) or pathologically (1,111 of 1,398; 79.5%) known SLN status (Appendix Fig A1).[24]

Validation cohort. A total of 1,674 patients with a continuous 31-GEP test result were enrolled under one of three Western Institutional Review Board–approved studies from 25 surgical and five dermatologic centers. Eligibility criteria were the same as for the development cohort (Appendix Fig A1).

31-GEP Testing

The 31-GEP test (DecisionDx-Melanoma; Castle Biosciences Inc, Friendswood, TX) was used to analyze the expression of 27 prognostic genes (28 probes) and three control genes from primary CM tumors, as previously described.[21] All 31-GEP testing was performed in a College of American Pathologists–accredited and Clinical Laboratory Improvement Amendments–certified laboratory.

i31-GEP Development and Validation

Data collected for analysis and i31-GEP algorithm training included the continuous variables 31-GEP score, Breslow thickness, MR, and age and the categorical variables ulceration status, regression, LVI, absence versus presence of tumor-infiltrating lymphocytes (TILs), sex, microsatellites, histopathologic subtype (superficial spreading, nodular, nevoid, lentiginous, acral, desmoplastic, and others), transected bases (ie, microstaging), and tumor site (head and neck, trunk, and extremity). Models were generated in R package v3.6.3 using the caret package to generate neural networks with the nnet submodule and four times ten-fold cross-validation for hyperparameter selection. Next, multiple iterations of the model were run with the remaining features to determine which features provided the most accurate algorithm without overfitting. Nodal events were coded as 0 for negative or 1 for positive to generate a regression algorithm.

Validation of the algorithm was performed on an independent cohort of eligible patients with T1-T4 tumors as described. Patients with T1a disease with documented MR $\geq$ 2/mm^2, LVI, absence of TILs, age $<$ 40 years, microsatellites, regression, or transected base were categorized as having high-risk T1a (T1a-HR) disease. Patients with T1a tumors and none of the above-specified features were considered low-risk T1a (T1a-LR).

Complete variable selection methods and accuracy metric calculations are provided in the Appendix 1.

Statistical Analysis

Comparison of CP features between cohorts was performed using the Mann-Whitney U test, the chi-square test, or Fisher's exact test. A P value of $<$ .05 was considered statistically significant. Continuous variables are reported as median (range) and dichotomous variables are reported as n (%). Simple logistic regression was performed to show the probability of a positive SLN for each variable within the training cohort. Kaplan-Meier analysis and the log-rank test were used to compare survival outcomes. R statistical package v3.6.3 and Graphpad Prism v8.4.1 were used for statistical analysis.

Study Approval

A waiver of consent was granted by Western Institutional Review Board as the research met the requirements for waiver of consent, including that the study involves no more than minimal risk to the participants and will not adversely affect the rights and welfare of the participants.

RESULTS

Patient Demographics

Demographic analyses revealed a significant difference between the development and validation cohorts in median GEP score (0.35 [range 0-1] v 0.40 [range 0-1]; $P <$ .001), age (63 years [range 18-101] v 65 years [range 21-97]; $P <$ .001), and MR (1.0/mm^2 [range 0-74.0/mm^2] v 1.0/mm^2 [range 0-235.0/mm^2]; $P <$ .001) and the number of patients with an absence of TILs (13.3% v 10.9%; $P <$ .001), presence of microsatellites (0.4% v 1.2%; $P <$ .001), a transected base (19.5% v 34.9%; $P <$ .001), presence of regression (13.7% v 14.6%; $P <$ .001), histologic subtype ($P <$ .001), and proportion of cases in which SLNB was performed (79.5% v 75.1%; $P =$ .005). There was no significant difference between cohorts for tumor location ($P =$ .584), sex ($P =$ .468), presence of LVI ($P =$ .104), Breslow thickness ($P =$ .342), ulceration ($P =$ .427), or SLN status ($P =$ .639; Appendix Fig A1, Table 1).

i31-GEP Algorithm Development and Specification

Features that significantly contributed to the model, as described in the methods, were included in i31-GEP development and included the continuous variables 31-GEP, Breslow thickness, MR, and age and the binary variable ulceration. Variable importance and logistic regression of variables within the training cohort are shown in Appendix Figure A2 and Appendix Table A1. The 31-GEP had the highest variable importance score (100, indicating the highest importance relative to the other variables) and log-likelihood value (G^2 = 91.3; $P <$ .001), indicating that it is the best predictor of SLN positivity, followed by Breslow thickness (G^2 = 53.5; $P <$ .001).

i31-GEP Performance

Validation in an independent cohort of 1,674 patients with T1-T4 tumors demonstrated alignment between observed SLN positivity rates and i31-GEP predictions, with a slope of 0.999 demonstrated by linear regression (Fig 1). Moreover, the i31-GEP model predicted that 27.7% (464 of 1,674) of patients had a predicted SLN positivity probability of $<$ 5%, and 41.6% (696 of 1,674) had a predicted probability of $>$ 10%. In contrast, just 8.5% (142 of 1,674) of patients with T1a tumors had $<$ 5% SLN positivity risk when analyzed by T stage. Three hundred seventy-seven tumors were T1a (235 of which had one or more high-risk features), and 328 were T1b. The i31-GEP reclassified 68.5% (161 of 235) of T1a tumors with at least one high-risk feature, and 40.9% (134 of 328) of T1b as having $<$ 5% likelihood of a positive SLN, for a total of 52.4% of higher-risk T1 tumors reclassified as $<$ 5% risk. Moreover, the i31-GEP reclassified 4.7% (11 of 235) of patients with a T1a tumor and at least one high-risk feature, and 14.3% (47 of 328) of T1b tumors as having $>$ 10% likelihood of a positive

TABLE 1. Demographics and Clinical Characteristics of the Training and Validation Cohort

Characteristic	Training Cohort (n = 1,398)	Validation Cohort (n = 1,674)	P
31-GEP,[a] a.u. (range)	0.35 (0.00-1.00)	0.40 (0.00-1.00)	<.001[b]
Breslow thickness,[a] mm (range)	1.2 (0.1-60.0)	1.2 (0.1-68.0)	.342[b]
Ulceration, No. (%)			.427[c]
Present	302 (21.6)	393 (23.5)	
Absent	1,082 (77.4)	1,267 (75.7)	
Unknown or not reported	14 (1.0)	14 (0.8)	
MR,[a] 1/mm² (range)	1.0 (0-74.0)	1.0 (0-235.0)	<.001[b]
TILs, No. (%)			<.001[c]
Present	626 (44.8)	876 (52.3)	
Absent	186 (13.3)	182 (10.9)	
Unknown or not reported	586 (41.9)	616 (36.8)	
Microsatellites, No. (%)			<.001[c]
Present	5 (0.4)	20 (1.2)	
Absent	14 (1.0)	4 (0.2)	
Unknown or not reported	1,379 (98.6)	1,650 (98.6)	
Transected base, No. (%)			<.001[c]
Present	272 (19.5)	585 (34.9)	
Absent	883 (63.2)	808 (48.3)	
Unknown or not reported	243 (17.4)	281 (16.8)	
Regression, No. (%)			<.001[c]
Present	191 (13.7)	245 (14.6)	
Absent	1,013 (72.5)	1,260 (75.3)	
Unknown or not reported	194 (13.9)	169 (10.1)	
LVI, No. (%)			.104[c]
Present	39 (2.8)	54 (3.2)	
Absent	1,315 (94.1)	1,544 (92.2)	
Unknown or not reported	44 (3.1)	76 (4.5)	
Location, No. (%)			.584[c]
Head and Neck	282 (20.2)	352 (21.0)	
Trunk	559 (40.0)	679 (40.6)	
Extremity	549 (39.3)	638 (38.1)	
Unknown or not reported	8 (0.6)	5 (0.3)	
Histologic subtype, No. (%)			<.001[c]
Superficial spreading	368 (26.3)	512 (30.6)	
Nodular	167 (11.9)	304 (18.2)	
Nevoid	31 (2.2)	36 (2.2)	
Lentigo maligna	18 (1.3)	38 (2.3)	
Desmoplastic	30 (2.1)	39 (2.3)	
Acral	8 (0.6)	15 (0.9)	
Others or unspecified	776 (55.5)	730 (43.6)	
Sex, No. (%)			.468[c]
Male	764 (54.6)	923 (55.1)	
Female	567 (40.6)	656 (39.2)	
Unknown or not reported	67 (4.8)	95 (5.7)	
Age,[a] years (range)	62.7 (18.0-101.45)	65.2 (20.6-96.6)	<.001[b]
SLN[e] status, No. (%)			.639[d]
Positive	145 (10.4)	183 (10.9)	
Negative	1,253 (89.6)	1,491 (89.1)	

(Continued in next column)

Characteristic	Training Cohort (n = 1,398)	Validation Cohort (n = 1,674)	P
SLNB performed, No. (%)	1,111 (79.5)	1,258 (75.1)	.005[d]
SLNB[e] status, No. (%)			.368[d]
Positive	143 (12.9)	179 (14.2)	
Negative	968 (87.1)	1,079 (85.8)	

Abbreviations: 31-GEP, 31-gene expression profile; a.u., arbitrary units; LVI, lymphovascular invasion; MR, mitotic rate; SLN, sentinel lymph node; SLNB, sentinel lymph node biopsy; TIL, tumor-infiltrating lymphocytes.

[a]Median continuous value.

[b]Mann-Whitney U test.

[c]Chi-square test.

[d]Fisher's exact test.

[e]SLN status includes clinically and pathologically assessed SLN, and SLNB includes only those with a biopsy performed.

SLN, reclassifying a total of 10.3% of higher-risk T1 tumors as having a predicted SLN-positive likelihood of > 10%. In sum, of the 563 of 1,542 patients with SLN positivity risk classified by T stage as 5%-10%, the i31-GEP reclassified 62.7% (353 of 563) to < 5% or > 10% likelihood of having a positive SLN (Fig 2, Table 2). Similarly, validation of cases in the T2 population demonstrated that 12.5% (52 of 416) of patients with T2a tumors and 4.2% (5 of 118) with T2b tumors had a predicted SLN-positive likelihood of < 5%. Furthermore, 44.7% (186 of 416) of T2a and 44.1% (52 of 118) of T2b cases had a 5%-10% likelihood of SLN positivity (Fig 2, Table 2).

On the other hand, although only 0.3% (1 of 303) of T3 cases had a < 5% risk prediction of SLN positivity, 10.2% (31 of 303) of cases had a risk between 5% and 10%, with the majority of T3 cases having a risk > 10% (as expected). Validation in patients with T4 tumors confirmed that, although the majority (96%) had SLN positivity predictions higher than 10%, the range was wide (9.5%-58%; Table 2, Appendix Fig A3), reinforcing the need for personalized SLN positivity likelihood prediction.

i31-GEP Accuracy

To assess the accuracy of the i31-GEP, a predicted likelihood of a positive SLN < 5% was considered a negative test and a ≥ 5% likelihood of involvement was considered a positive test. The overall sensitivity for the model was 95.1%, and the negative predictive value (NPV) was 98.1%. The T1a low-risk population had no pathologically positive SLNs, whereas the T3 population had only one negative test result and the T4 population had no negative results. In the T1a-HR-T2 population, the range of cases most likely to need additional guidance, the i31-GEP had an overall high NPV (97.4%) and sensitivity (89.8%), indicating a low false-negative rate. On the basis of the low likelihood of identifying a pathologically positive SLN with a negative i31-GEP result, the procedure reduction rate
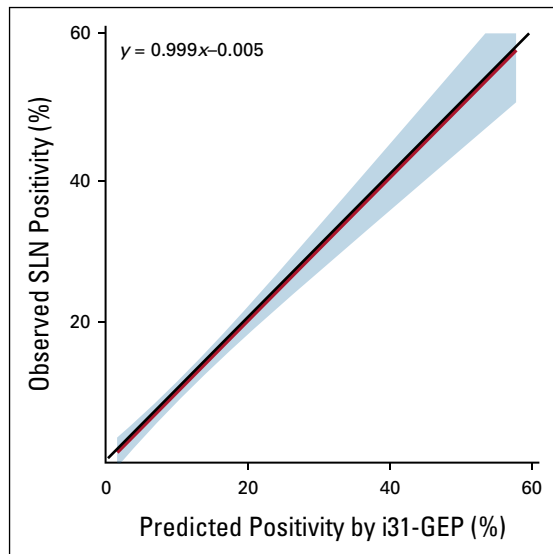
FIG 1. 31-GEP improves precision of SLN positivity predictions compared with T stage–based predictions in an independent validation cohort (n = 1,674) with T1-T4 CM. The integration of the 31-GEP score and clinicopathologic features (i31-GEP) is represented by the red line. Gray shading represents 95% CI. The solid black line represents a perfect match of predicted and observed SLN-positive rates. Linear regression shows a $y = 0.999x - 0.005$ relationship between predicted and observed positivity demonstrating the close alignment of i31-GEP–predicted risk of SLN positivity and observed SLN positivity. 31-GEP, 31-gene expression profile; i31-GEP, integrated 31-gene expression profile; SLN, sentinel lymph node.

(32.1% overall) was calculated as the proportion of negative test results for the given population. Within the T1a-HR population, a reduction rate of 68.5% was achieved with an NPV of 97.5%. Similarly, in the T1b population, there was a reduction rate of 40.9% with an NPV of 97.8% (Table 3).

i31-GEP Survival Outcomes

The study included cases from a prospective, multicenter US study that was recently published and had data on SLN status and 3.2 years median follow-up,[35] allowing for assessment of patient outcomes in the < 5% and ≥ 5% risk group described by the i31-GEP model.[27] Patients predicted by the i31-GEP to have < 5% SLN positivity risk had significantly higher recurrence-free survival (RFS: 96.8%, $P < .001$), distant metastasis–free survival (DMFS: 98.6%, $P = .002$), and overall survival (OS: 97.7%, $P = .043$) relative to patients predicted to have ≥ 5% likelihood who were node-negative (RFS: 88.3%, DMFS: 93.5%, OS: 93.3%) and patients predicted to have ≥ 5% likelihood who were node-positive (RFS: 61.8%, DMFS: 71.0%, OS: 81.5%; Fig 3). As expected, patients with ≥ 5% likelihood of a positive SLN predicted by the i31-GEP who did not have a positive SLN had worse outcomes than those with a predicted likelihood of < 5%. Furthermore, a positive SLNB in the ≥ 5% risk group negatively affected overall outcomes.

DISCUSSION

Although NCCN guidelines recommend SLNB in patients with > 10% likelihood of identifying a positive SLN, most patients who undergo an SLNB receive a negative result, risk unnecessary adverse events from surgical intervention, and retain their initially diagnosed American Joint Committee on Cancer stage.[2] One reason for a high negative rate may be that most (60%-70%) new melanomas are diagnosed as T1, for which the indication for performing an SLNB is unclear. Furthermore, there is no consensus about which CP variables are prognostic for SLN positivity outside of staging factors, eg tumor thickness and ulceration status.[36-38] Many patients with T4 melanomas may be eligible for adjuvant therapy regardless of SLN status, and it is unclear if these patients should undergo the procedure.[36]

A test to increase the accuracy of predicting a positive SLNB that complements NCCN guidelines could help patients and physicians when considering an SLNB.[37] Our data show that integrating CP features with the continuous 31-GEP score, which provides insight into intrinsic tumor biology, accurately predicts the likelihood of having a positive SLN. Although the continuous 31-GEP score has the biggest impact on the algorithm, a major strength of the i31-GEP is that it incorporates many routinely recorded CP features, including Breslow thickness, MR, ulceration, and age into the algorithm. As a result, patients may have an SLNB recommendation changed from the consideration to the recommendation range, whereas others may have their recommendation changed to avoid SLNB.

Here, we show that the i31-GEP accurately identified a larger percentage of patients (27.7%, 464 of 1,674) with a < 5% likelihood of SLN positivity than were identified by T stage in conjunction with CP risk factors without the 31-GEP (T1a-LR, 8.5%, 142 of 1,674; Table 2). For patients in the T1a-HR group, the i31-GEP identified 43% of patients with a positive SLN while maintaining a high NPV of 97.5%. The NCCN guidelines recommend that patients with T1a-HR tumors should consider SLNB.[3] This is substantiated in our validation cohort, in which none (0%) of the patients undergoing surgery for a T1a tumor without high-risk features (n = 30) had a positive SLN, whereas 7.5% (7 of 93) of those with at least one high-risk feature had a positive SLN, and the T1b SLNB positivity rate was 6.5% (18 of 279; Appendix Table A2).[9-11,38-40]

A recent publication by Lo et al[34] using a nomogram to determine the likelihood of SLN positivity reported only 12.4% (v 27% by i31-GEP) of patients had a likelihood of SLN positivity of < 5%. The authors further predicted that only 27% (v 57% by i31-GEP) of patients with T1 tumors had a < 5% likelihood.[34] Some SLN prediction models have focused on higher-risk populations.[30] Bellomo et al[30] developed a combined CP-GEP variable evaluator using a single cohort of 754 patients from tertiary Mayo Clinic sites. In their initial study, they performed double-loop cross-

FIG 2. Distribution of SLN positivity risk predicted by i31-GEP by T stage. T1a-LR refers to patients with low risk T1a tumors with no high-risk features documented, and T1a-HR refers to those with a high risk T1a tumor who had risk factors for a positive SLN resulting in a risk between 5% and 10%. The predicted risk was truncated at 20%. T4a risk ranged from 9.5% to 50.0%, and T4b ranged from 9.5% to 58.5%. See Appendix Fig A3 for full distribution of predicted SLN positivity, including distribution for T4 tumors. i31-GEP, integrated 31-gene expression profile; SLN, sentinel lymph node; T1a-HR, high-risk T1a; T1a-LR, low-risk T1a.
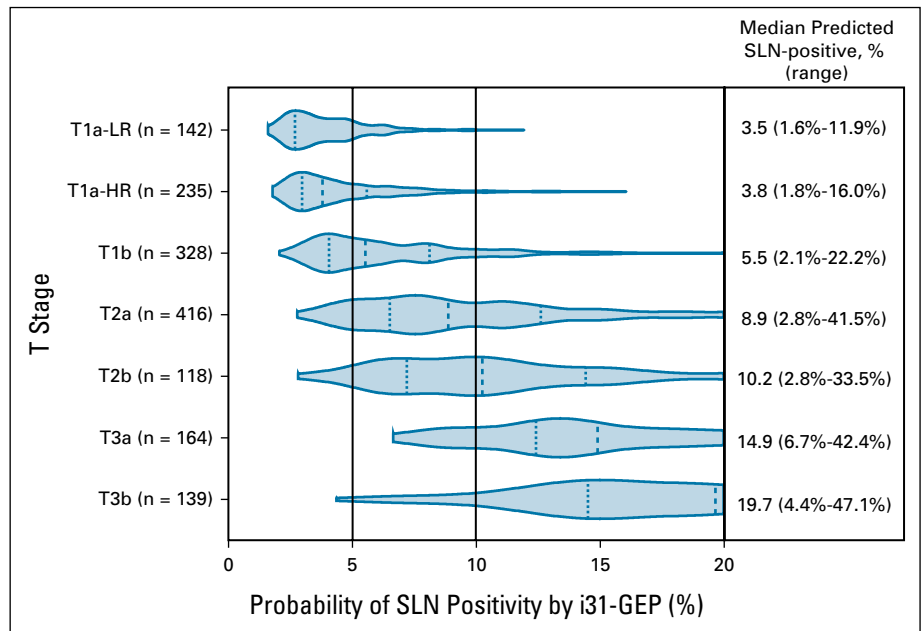
validation but did not have an independent validation. A follow-up study was performed on only 210 European patients and only had 11 T1 cases. To date, no independent validation has been performed in a large US cohort.[30,41] Moreover, although their CP-GEP test incorporates molecular tumor biology, it bins positive and negative results averaging in patients with a risk higher than 5% into the negative test result bin and is similar to current population-based methods for determining the likelihood of a positive SLN. In contrast, the i31-GEP moves toward precision medicine in which each patient receives an individualized likelihood of SLN positivity.

The populations that make up the development and validation cohorts for the i31-GEP better reflect the average patient evaluated for SLNB, with 46% of the validation population having T1 tumors that were relatively even in numbers of T1a (377) and T1b (328) tumors. Although 68.5% of T1a-HR cases have a positivity likelihood of < 5% identified by the i31-GEP, nearly 5% of the T1a-HR had > 10% likelihood of SLN positivity. These data demonstrate that the i31-GEP offers a

TABLE 2. i31-GEP Improves Precision of T Stage–Predicted Sentinel Lymph Node Positivity Risk Estimates

T Stage (No.)	Standard System of Risk Binning,[a] % Population (No.)			Precision Risk Reclassification by i31-GEP, % Population (No.)			
	Not Recommended (< 5%)	Considered (5%-10%)	Recommended (> 10%)	Not Recommended (< 5%)	Considered (5%-10%)	Recommended (> 10%)	Percent Change[b]
T1a-LR (142)	100 (142)	—	—	78.2 (111)	21.1 (30)	0.7 (1)	21.8 (31)
T1a-HR (235)	—	100 (235)	—	68.5 (161)	26.8 (63)	4.7 (11)	73.2 (172)
T1b (328)	—	100 (328)	—	40.9 (134)	44.8 (147)	14.3 (47)	55.2 (181)
T2a (416)	—	—	100 (416)	12.5 (52)	44.7 (186)	42.8 (178)	57.2 (238)
T2b (118)	—	—	100 (118)	4.2 (5)	44.1 (52)	51.7 (61)	48.3 (57)
T3a (164)	—	—	100 (164)	0 (0)	14.6 (24)	85.4 (140)	14.6 (24)
T3b (139)	—	—	100 (139)	0.7 (1)	5.0 (7)	94.2 (131)	5.8 (8)
T4a (51)	—	—	100 (51)	0 (0)	7.8 (4)	92.2 (47)	7.8 (4)
T4b (81)	—	—	100 (81)	0 (0)	1.2 (1)	98.8 (80)	1.2 (1)

NOTE. T1a-LR (low-risk): T1a with no recorded high-risk features; T1a-HR (high-risk): T1a with one or more features that may be considered high-risk when assessing sentinel lymph node biopsy eligibility including age < 40 years, mitotic rate ≥ 2/mm², presence of regression, lymphovascular invasion, transected base, or absence of tumor-infiltrating lymphocytes.

Abbreviations: i31-GEP, integrated 31-gene expression profile; T1a-HR, high-risk T1a; T1a-LR, low-risk T1a.

[a]Classification of risk according to the National Comprehensive Cancer Network guidelines by T stage.

[b]Percent change from risk bin designated by T stage.

TABLE 3. Accuracy of the i31-GEP by T Stage

Accuracy Metric	Overall (T1-T4) (%)	T1a-HR-T2 (%)	T1a-HR (%)	T1b (%)	T2a (%)	T2b (%)
NPV	98.1	97.4	97.5	97.8	96.2	100.0
False-negative rate	1.9	2.6	2.5	2.2	3.8	0.0
Reduction rate	23.0[a]	32.1	68.5	40.9	12.5	4.2
Sensitivity	95.1	89.8	42.9	83.3	95.8	100.0
Pretest SLN positivity rate	10.9	8.0	3.0	5.5	11.5	12.7
PPV of ≥ 5% risk	14.4	10.6	4.1	7.7	12.6	13.3

NOTE. < 5.0% risk of SLN positivity was considered a negative test result, and ≥ 5% risk of SLN positivity was considered a positive test result. Overall accuracy of the model was calculated (overall, T1-T4).

Abbreviations: i31-GEP, integrated 31-gene expression profile; NPV, negative predictive value; PPV, positive predictive value; SLN, sentinel lymph node; T1a-HR, high-risk T1a.

[a]Reduction rate for T1-T4 was calculated using T1a-HR-T4 since T1a with no high-risk features would not be eligible for sentinel lymph node biopsy. Accuracy was also assessed in the T1a-HR, T1b, T2a, and T2b cases together and individually. There were no positive SLNs in the T1a group with no high-risk features. Conversely, there were no negative i31-GEP test results in the T3a, T4a, and T4b populations and only one negative test result (in a patient with a negative SLN) in the T3b population. Therefore, they were excluded from individual analysis.

more personalized risk prediction for patients with low and high likelihood of SLN metastasis relative to T stage alone. As might be expected, patients with an i31-GEP–predicted low likelihood of SLN involvement also demonstrated high RFS, DMFS, and OS (Fig 3).

Although the i31-GEP developed in this report was independently validated to refine risk assessment within the context of clinical, histologic, and molecular features, there are some limitations. The training and validation cohorts were mostly treated at surgical oncology centers, with nearly 80% undergoing SLNB. Therefore, patients not referred from a dermatology clinic may not have been included. Some T1a patients were evaluated clinically but did not have SLNB performed. Therefore, we cannot rule out the potential for occult nodal metastases in these patients. Finally, in any retrospective study from various centers, standardization of SLNB, SLN assessment, and event detection bias may influence the results.

Population	3-Year RFS, % (95% CI)	Recurrences, No. (%)[a]
< 5% n = 99	96.8 (93.3 to 100)	3 (3.0)
≥ 5%, SLN-negative n = 179	88.3 (83.5 to 93.2)	27 (15.1)
≥ 5%, SLN-positive n = 34	61.8 (46.9 to 81.6)	12 (35.3)

Population	3-Year DMFS, % (95% CI)	Distant Metastases, No. (%)[a]
< 5% n = 99	98.6 (95.9 to 100)	1 (1.0)
≥ 5%, SLN-negative n = 179	93.5 (89.8 to 97.3)	16 (8.9)
≥ 5%, SLN-positive n = 34	71.0 (56.6 to 89.1)	9 (26.5)

Population	3-Year OS, % (95% CI)	Deaths, No. (%)[a]
< 5% n = 99	97.7 (94.5 to 100)	3 (3.0)
≥ 5%, SLN-negative n = 179	93.3 (89.6 to 97.2)	15 (8.4)
≥ 5%, SLN-positive n = 34	81.5 (69.1 to 96.1)	7 (20.6)

FIG 3. Melanoma survival rates in a subset of 312 patients with long-term follow-up stratified by < 5% and ≥ 5% SLN positivity risk by i31-GEP. The blue line represents the survival of patients with an i31-GEP prediction of SLN positivity < 5%, the red line represents the survival rates of patients with ≥ 5% positivity who had a negative SLN, and the teal line represents the survival rates of patients with ≥ 5% positivity who had a positive SLN. P value on the basis of log-rank test. [a]Number of events over the full follow-up period. DMFS, distant metastasis–free survival; i31-GEP, integrated 31-gene expression profile; OS, overall survival; RFS, recurrence-free survival; SLN, sentinel lymph node.

These data demonstrate the value of advanced artificial intelligence tools for personalized risk assessment. Using the i31-GEP allows for an accurate assessment of each individual's likelihood of a positive SLN for CM of any T stage. Rather than relying on average risks based only on epidemiologic and phenotypic features, the more precise assessment of the i31-GEP provides individualized likelihood of SLN positivity. This personalized risk calculation can reduce the number of patients who would undergo SLNB and more

appropriately identify those individuals with an increased likelihood of a positive result. Consequently, morbidity, cost, and inconvenience can be reduced among patients avoiding the procedure, and the anxiety of decision making can be reduced for those whose positivity likelihood falls into the NCCN-defined indeterminant range (5%-10%). For most patients recommended for SLNB, having a high probability of a positive SLN according to the i31-GEP could confirm that the SLNB is a justifiable additional intervention.

AFFILIATIONS

[1]Carol G. Simon Cancer at Morristown Medical Center, Atlantic Health System, Morristown, NJ

[2]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ

[3]Cleveland Clinic Lerner Research Institute, Cleveland, OH

[4]Medical City Dallas Hospital, Dallas, TX

[5]Department of Surgery, St Louis University, St Louis, MO

[6]General Surgery Abington Memorial Hospital, Abington, PA

[7]Oregon Melanoma Center, Springfield, OR

[8]Morton Plant Mease Healthcare, Safety Harbor, FL

[9]Advent Health Cancer Institute, Orlando, FL

[10]St Elizabeth Physicians General & Vascular Surgery, Ft Mitchell, KY

[11]Florida State University College of Medicine, Bradenton, FL

[12]RWJ Barnabas Health, Livingston, NJ

[13]North Texas Melanoma Center, Dallas, TX

[14]University of Tennessee Graduate School of Medicine, Knoxville, TN

[15]SCL Health, Denver, CO

[16]Castle Biosciences Inc, Friendswood, TX

[17]Division of Surgical Oncology, The University of Tennessee Health Science Center, Memphis, TN

[18]Desert Surgical Oncology, Rancho Mirage, CA

[19]Oregon Health & Science University, Portland, OR

CORRESPONDING AUTHOR

Robert W. Cook, PhD, 505 S. Friendswood Drive, Ste 401, Friendswood, TX 77546; e-mail: rcook@castlebiosciences.com.

DATA SHARING STATEMENT

The i31-GEP algorithm uses a proprietary neural network algorithm to incorporate data from the continuous output of the 31-gene expression profile test—which assesses 28 discriminating gene targets and three control genes via qRT-PCR using a radial basis machine algorithm as previously described (Gerami et al Clin Cancer Res, 2015)—and clinicopathologic variables including continuous Breslow thickness, mitotic rate, age, and binary ulceration status. Data will not be made available publicly.

AUTHOR CONTRIBUTIONS

Conception and design: Eric D. Whitman, Brian R. Gastman, Kyle R. Covington, Federico A. Monzon, Matthew S. Goldberg, Robert W. Cook, David M. Hyams

Provision of study materials or patients: Deri Lewis, Eddy C. Hsueh, Thomas P. Trezona, J. Michael Guenther, Franz O. Smith, Peter D. Beitsch, Martin D. Fleming, David M. Hyams

Collection and assembly of data: Eric D. Whitman, Vadim P. Koshenkov, Brian R. Gastman, Deri Lewis, Eddy C. Hsueh, Ho Pak, Thomas P. Trezona, Robert S. Davidson, Michael McPhee, J. Michael Guenther, Paul Toomey, Franz O. Smith, Peter D. Beitsch, James M. Lewis, Martin D. Fleming, David M. Hyams, John T. Vetto

Data analysis and interpretation: Robert W. Cook

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

ACKNOWLEDGMENT

REFERENCES

1. Gershenwald JE, Scolyer RA, Hess KR, et al: Melanoma staging: Evidence-based changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin 67:472-492, 2017

2. Chen J, Xu Y, Zhou Y, et al: Prognostic role of sentinel lymph node biopsy for patients with cutaneous melanoma: A retrospective study of surveillance, epidemiology, and end-result population-based data. Oncotarget 7:45671-45677, 2016

3. Swetter SM, Thompson JA, Coit DG, et al: NCCN Clinical Practice Guidelines in Oncology. Cutaneous Melanoma. Version 3.2020. Plymouth Meeting, PA, National Comprehensive Cancer Network, 2020

4. Sinnamon AJ, Neuwirth MG, Yalamanchi P, et al: Association between patient age and lymph node positivity in thin melanoma. JAMA Dermatol 153:866-873, 2017

5. Friedman C, Lyon M, Torphy RJ, et al: A nomogram to predict node positivity in patients with thin melanomas helps inform shared patient decision making. J Surg Oncol 120:1276-1283, 2019

6. Kocsis A, Karsko L, Kurgyis ZS, et al: Is it necessary to perform sentinel lymph node biopsy in thin melanoma? A retrospective single center analysis. Pathol Oncol Res 26:1861-1868, 2020

7. Conic RRZ, Ko J, Damiani G, et al: Predictors of sentinel lymph node positivity in thin melanoma using the National Cancer Database. J Am Acad Dermatol 80:441-447, 2019

8. Kesmodel SB, Karakousis GC, Botbyl JD, et al: Mitotic rate as a predictor of sentinel lymph node positivity in patients with thin melanomas. Ann Surg Oncol 12:449-458, 2005

9. Gershenwald JE, Scolyer RA: Melanoma staging: American Joint Committee on Cancer (AJCC) 8th edition and beyond. Ann Surg Oncol 25:2105-2110, 2018

10. Cordeiro E, Gervais M-K, Shah PS, et al: Sentinel lymph node biopsy in thin cutaneous melanoma: A systematic review and meta-analysis. Ann Surg Oncol 23:4178-4188, 2016

11. Andtbacka RH, Gershenwald JE: Role of sentinel lymph node biopsy in patients with thin melanoma. J Natl Compr Canc Netw 7:308-317, 2009

12. Elmore JG, Barnhill RL, Elder DE, et al: Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: Observer accuracy and reproducibility study. BMJ 357:j2813, 2017

13. Tejera-Vaquerizo A, Ribero S, Puig S, et al: Survival analysis and sentinel lymph node status in thin cutaneous melanoma: A multicenter observational study. Cancer Med 8:4235-4244, 2019

14. Kruper LL, Spitz FR, Czerniecki BJ, et al: Predicting sentinel node status in AJCC stage I/II primary cutaneous melanoma. Cancer 107:2436-2445, 2006

15. Wheless L, Isom CA, Hooks MA, et al: Mitotic rate is associated with positive lymph nodes in thin melanomas. J Am Acad Dermatol 78:935-941, 2018

16. Wright BE, Scheri RP, Ye X, et al: Importance of sentinel lymph node biopsy in patients with thin melanoma. Arch Surg 143:892-899, 2008; discussion 899-900

17. Paek SC, Griffith KA, Johnson TM, et al: The impact of factors beyond Breslow depth on predicting sentinel lymph node positivity in melanoma. Cancer 109:100-108, 2007

18. Teixeira V, Vieira R, Coutinho I, et al: Prediction of sentinel node status and clinical outcome in a melanoma centre. J Skin Cancer 2013:1-7, 2013

19. Grob J-J, Garbe C, Ascierto P, et al: Adjuvant melanoma therapy with new drugs: Should physicians continue to focus on metastatic disease or use it earlier in primary melanoma? Lancet Oncol 19:e720-e725, 2018

20. Banting S, Milne D, Thorpe T, et al: Negative sentinel lymph node biopsy in patients with melanoma: The patient's perspective. Ann Surg Oncol 26:2263-2267, 2019

21. Gerami P, Cook RW, Wilkinson J, et al: Development of a prognostic genetic signature to predict the metastatic risk associated with cutaneous melanoma. Clin Cancer Res 21:175-183, 2015

22. Gerami P, Cook RW, Russell MC, et al: Gene expression profiling for molecular staging of cutaneous melanoma in patients undergoing sentinel lymph node biopsy. J Am Acad Dermatol 72:780-785.e3, 2015

23. Zager JS, Gastman BR, Leachman S, et al: Performance of a prognostic 31-gene expression profile in an independent cohort of 523 cutaneous melanoma patients. BMC Cancer 18:130, 2018

24. Vetto JT, Hsueh EC, Gastman BR, et al: Guidance of sentinel lymph node biopsy decisions in patients with T1-T2 melanoma using gene expression profiling. Future Oncol 15:1207-1217, 2019

25. Keller J, Schwartz TL, Lizalek JM, et al: Prospective validation of the prognostic 31-gene expression profiling test in primary cutaneous melanoma. Cancer Med 8:2205-2212, 2019

26. Dillon LD, Gadzia JE, Davidson RS, et al: Prospective, multicenter clinical impact evaluation of a 31-gene expression profile test for management of melanoma patients. SKIN J Cutan Med 2:111-121, 2018

27. Hsueh EC, DeBloom JR, Lee J, et al: Interim analysis of survival in a prospective, multi-center registry cohort of cutaneous melanoma tested with a prognostic 31-gene expression profile test. J Hematol Oncol 10:152, 2017

28. Greenhaw BN, Covington KR, Kurley SJ, et al: Molecular risk prediction in cutaneous melanoma: A meta-analysis of the 31-gene expression profile prognostic test in 1,479 patients. J Am Acad Dermatol 83:745-753, 2020

29. Litchman GH, Prado G, Teplitz RW, et al: A systematic review and meta-analysis of gene expression profiling for primary cutaneous melanoma prognosis. SKIN J Cutan Med 4:221-237, 2020

30. Bellomo D, Arias-Mejias SM, Ramana C, et al: Model combining tumor molecular and clinicopathologic risk factors predicts sentinel lymph node metastasis in primary cutaneous melanoma. JCO Precis Oncol 4:319-334, 2020

31. Kim Y-J, Kim K, Lee KH, et al: Immune expression signatures as candidate prognostic biomarkers of age and gender survival differences in cutaneous melanoma. Sci Rep 10:12322, 2020

32. Mulder EEAP, Dwarkasing JT, Tempel D, et al: Validation of a clinicopathological and gene expression profile model for sentinel lymph node metastasis in primary cutaneous melanoma. Br J Dermatol 184:944-951, 2021

33. Wong SL, Kattan MW, McMasters KM, et al: A nomogram that predicts the presence of sentinel node metastasis in melanoma with better discrimination than the American Joint Committee on Cancer staging system. Ann Surg Oncol 12:282-288, 2005

34. Lo SN, Ma J, Scolyer RA, et al: Improved risk prediction calculator for sentinel node positivity in patients with melanoma: The Melanoma Institute Australia Nomogram. J Clin Oncol 38:2719-2727, 2020

35. Hsueh EC, DeBloom JR, Lee JH, et al: Long-term outcomes in a multicenter, prospective cohort evaluating the prognostic 31-gene expression profile for cutaneous melanoma. JCO Precis Oncol, 2021 10.1200/PO.20.00119

36. Han D, Han G, Duque MT, et al: Sentinel lymph node biopsy is prognostic in thickest melanoma cases and should be performed for thick melanomas. Ann Surg Oncol 28:1007-1016, 2020

37. Morrison S, Han D: Re-evaluation of sentinel lymph node biopsy for melanoma. Curr Treat Options Oncol 22:22, 2021

38. Wong SL, Faries MB, Kennedy EB, et al: Sentinel lymph node biopsy and management of regional lymph nodes in melanoma: American Society of Clinical Oncology and Society of Surgical Oncology clinical practice guideline update. Ann Surg Oncol 25:356-377, 2018

39. Han D, Zager JS, Shyr Y, et al: Clinicopathologic predictors of sentinel lymph node metastasis in thin melanoma. J Clin Oncol 31:4387-4393, 2013

40. Murali R, Haydu LE, Quinn MJ, et al: Sentinel lymph node biopsy in patients with thin primary cutaneous melanoma. Ann Surg 255:128-133, 2012

41. Cawley GC, Talbot NLC: On over-fitting in model selection and subsequent selection bias in performance evaluation. J Machine Learn Res 11:2079-2107, 2010

APPENDIX 1. SUPPLEMENTAL METHODS

Algorithm Development

Variables considered for inclusion in the algorithm were Breslow thickness, ulceration, 31-gene expression profile (31-GEP), regression, mitotic rate (MR), microsatellites, tumor-infiltrating lymphocytes, transected base, age, sex, histologic subtype, and tumor location. Regression, MR, microsatellites, and ulceration were imputed to absent if not reported in the patient records. Furthermore, if MR was reported qualitatively as present, MR was set to one. If it was reported with a < or > symbol preceding the number of mitoses, it was decreased or increased by one, respectively. If MR or Breslow thickness exceeded 10/mm^2 or 10.0 mm, respectively, they were imputed to 10.

Because neural network algorithms are subject to overfitting with the inclusion of excess variables that do not contribute to the algorithm, variable selection is an essential aspect of neural network development. Therefore, variables occurring in < 5% (microsatellites and lymphovascular invasion) of cases and those with insufficient completeness because of nonstandardized variable reporting (tumor-infiltrating lymphocytes) were excluded as variables for algorithm training.

Iterations of the algorithm using histologic subtype, transected base, tumor location, and sex did not improve model fit.

Algorithm Accuracy

Accuracy metrics were calculated by assigning integrated 31-GEP predictions of sentinel lymph node tumor involvement of < 5% as a negative and ≥ 5% as a positive result. Sentinel lymph node biopsy reduction rate was calculated by dividing the number of negative test results by the full population. The percent yield was calculated as the proportion of true positive test results among all test results (positive predictive value).
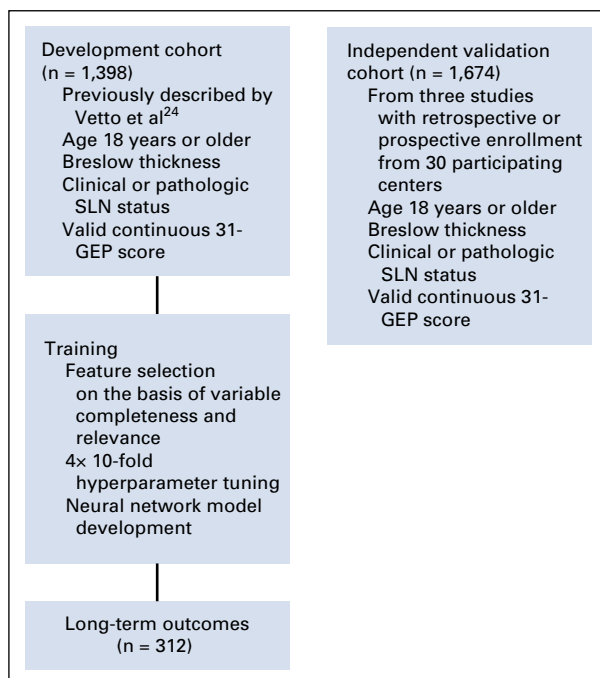
FIG A1. Training and validation cohorts. 31-GEP, 31-gene expression profile; SLN, sentinel lymph node.
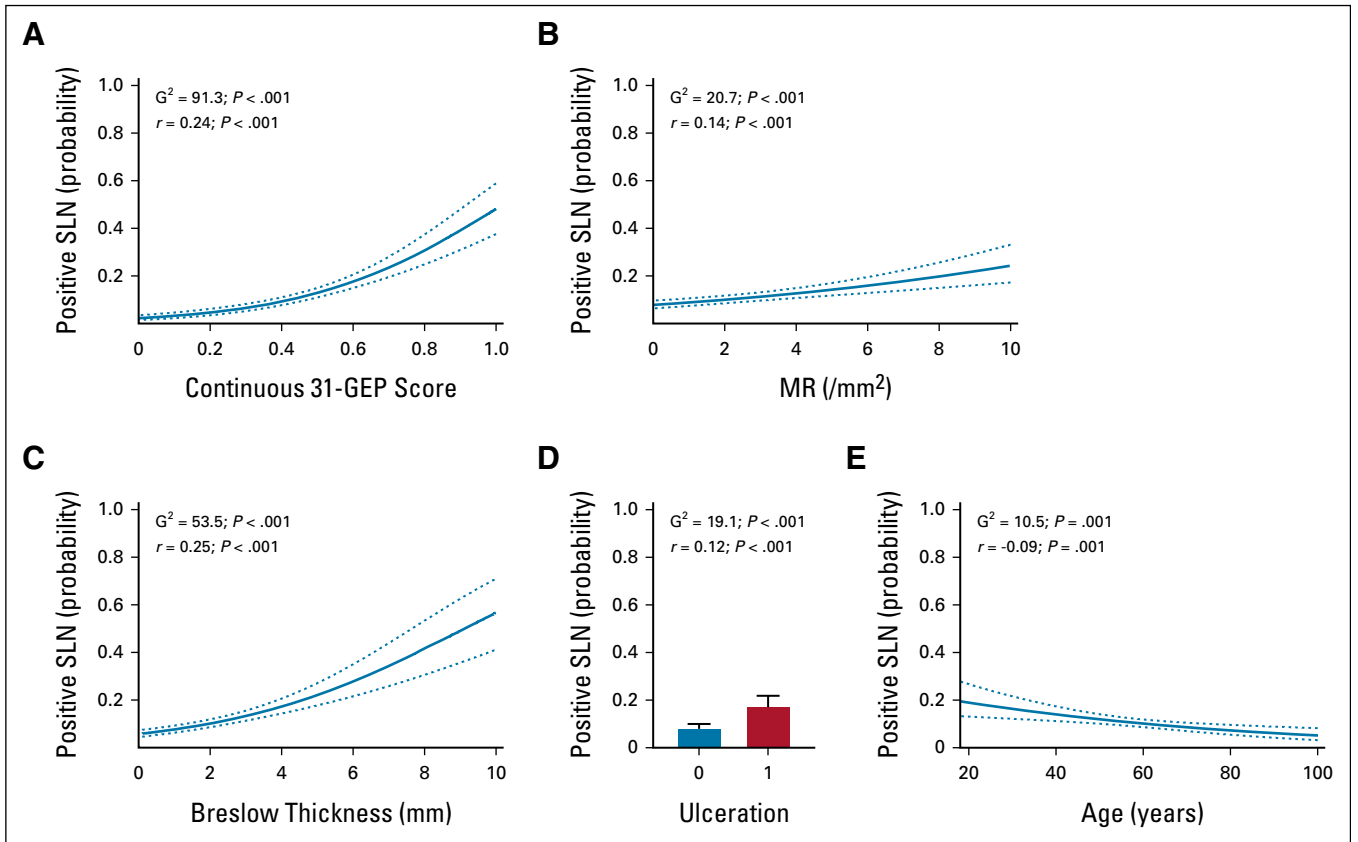
FIG A2. Correlation of individual variables score used in i31-GEP training. Correlation of the (A) continuous 31-GEP score, (B) continuous mitotic rate, (C) continuous Breslow thickness, (D) binary ulceration, and (E) continuous age with SLN positivity. Spearman's correlation (*r*) and log-likelihood ratios (G^2 values) demonstrate a significant correlation between all variables used in training. The GEP continuous score had the highest log-likelihood value and, therefore, had the best fit of all the variables. 31-GEP, 31-gene expression profile; i31-GEP, integrated 31-gene expression profile; MR, mitotic rate; SLN, sentinel lymph node.
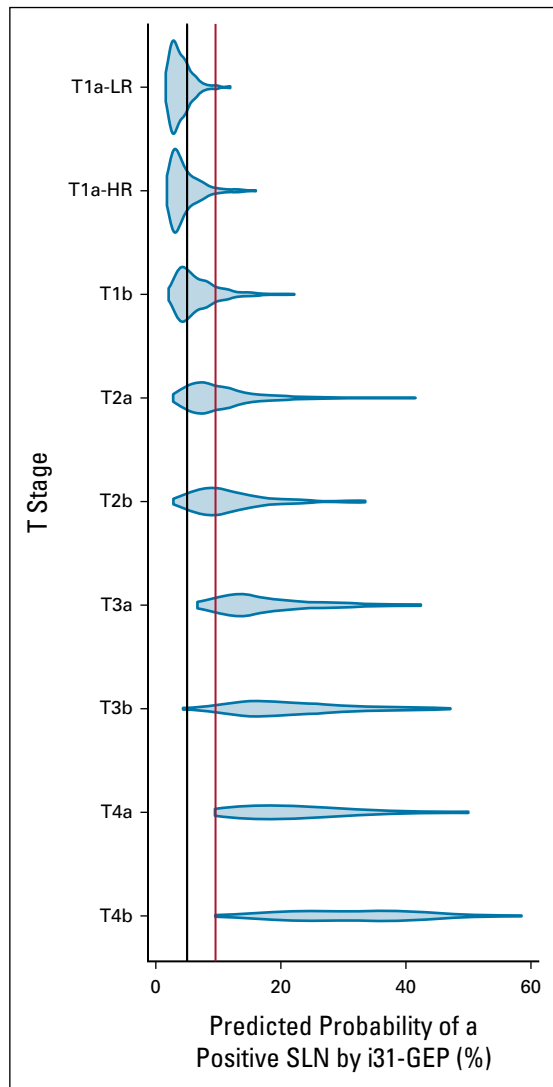
FIG A3. Full distribution of SLN positivity risk predicted by i31-GEP by T stage in T1-T4 CM. The black line is 5% and the red line is 10% predicted probability of a positive SLN. i31-GEP, integrated 31-gene expression profile; SLN, sentinel lymph node; T1a-HR, high-risk T1a; T1a-LR, low-risk T1a.

TABLE A1. Variable Importance in Sentinel Lymph Node Positivity Prediction

Variable	Variable Importance Score[a]	Log-Likelihood Value (G^2)[b]	Spearman Correlation
31-GEP score (continuous)	100	$G^2 = 91.3$; $P < .001$	$r = 0.24$; $P < .001$
Breslow thickness (continuous)	56	$G^2 = 53.5$; $P < .001$	$r = 0.25$; $P < .001$
Ulceration (categorical)	83	$G^2 = 19.1$; $P < .001$	$r = 0.12$; $P < .001$
MR (continuous)	25	$G^2 = 20.7$; $P < .001$	$r = 0.14$; $P < .001$
Age (continuous)	0	$G^2 = 10.5$; $P = .001$	$r = -0.09$; $P = .001$

Abbreviations: 31-GEP, 31-gene expression profile; MR, mitotic rate.
[a]Scale of 0-100 with 100 having the highest importance.
[b]Highest G^2 value corresponds to the best explanatory variable.

TABLE A2. Pretest SLN Positivity Rates by T Stage in 1,674 Patients With T1-T4 CM

T Stage	SLNB Assessed		
	n/N	% SLN-Positive	95% CI
T1-T4	179/1,258	14.2	12.3-16.3
T1a-LR	0/30	0	0-11.6
T1a-HR	7/93	7.5	3.1-14.9
T1b	18/279	6.5	3.9-10.0
T2a	48/378	12.7	9.5-16.5
T2b	15/106	14.2	8.1-22.3
T3a	32/147	21.8	15.4-29.3
T3b	30/119	25.2	17.7-34.0
T4a	8/42	19.0	8.6-34.1
T4b	21/64	32.8	21.6-45.7

NOTE. T1a-LR: T1a with no documented high-risk feature; T1a-HR, T1a with one or more high-risk clinicopathologic features.
Abbreviations: SLN, sentinel lymph node; SLNB, sentinel lymph node biopsy; T1a-HR, high-risk T1a; T1a-LR, low-risk T1a.