Gene expression profiling beyond Breslow thickness and ulceration for prediction of distant metastases in early-stage melanoma: the population-based Dutch Early-Stage Melanoma (D-ESMEL) study

Catherine Zhou[1], Yan-Ting Chen[2], Antien L. Mooyaart[3], Erik T. Valent[2], Lara Pozza[2], Daphne M.S. Huigh[2], Tamar E.C. Nijsten[1], Loes M. Hollestein[1,4]

[1] Department of Dermatology, Erasmus MC Cancer Institute, dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands
[2] SkylineDx B.V., Lichtenauerlaan 40, 3062 ME Rotterdam, The Netherlands
[3] Department of Pathology, Erasmus MC Cancer Institute, dr. Molewaterplein 40, 3015 GD Rotterdam, The Netherlands
[4] Department of Research and Development, Netherlands Comprehensive Cancer Organization, Rijnkade 5, 3511 CV Utrecht, The Netherlands

Corresponding author: Loes M. Hollestein, l.hollestein@erasmusmc.nl

Running head
GEP equals Breslow thickness and ulceration in melanoma distant metastases prediction

Funding sources
The project was co-funded by the PPP Allowance made available by Health~Holland (Grant Number: EMCLSH19008), Top Sector Life Sciences & Health.

Previous presentations
This study was presented in part at the European Association of Dermato-Oncology (EADO) Congress 2025 and the Society for Melanoma Research (SMR) Congress 2025.

Preprint availability statement
A preprint of this manuscript is available on medRxiv (doi: 10.1101/2025.11.10.25339906).

Authors' Disclosures of Potential Conflicts of Interest
Y.T.C, E.T.V., L.P. and D.M.S.H. are employees and option holders of SkylineDx BV. All other authors declare no competing interests.

Abstract

Purpose

Despite their central role in the current staging system, Breslow thickness and ulceration do not fully identify early-stage melanoma patients who will develop distant metastases. We assessed whether gene expression profiling (GEP) improves prediction of distant metastases beyond standard staging factors in early-stage melanoma.

Methods

Data were derived from the population-based Dutch Early-Stage Melanoma (D-ESMEL) study, including a matched discovery set of 442 stage I/II melanomas (221 case-control pairs) and a validation cohort of 308 melanomas nested within 5,815 patients. The discovery set was used to identify genes associated with distant metastases, independent of age, sex, Breslow thickness, and ulceration. The validation cohort was partitioned into model development and independent validation subsets. Candidate genes from the discovery set were used to develop and validate a GEP model, evaluated by weighted area under the curve (AUC) and concordance index (C-index).

Results

RNA sequencing succeeded for 356 melanomas in the discovery set, 200 in the model development subset, and 94 melanomas in the independent validation subset. Differential gene expression analyses and modeling identified 558 candidate genes. In the independent validation subset, the GEP model achieved a weighted AUC of 0.77 (95% CI, 0.66-0.86) and weighted C-index of 0.79 (95% CI, 0.69-0.88), comparable to the clinical model based on Breslow thickness and ulceration (weighted AUC 0.82 (95% CI, 0.73-0.90), weighted C-index 0.84 (95% CI, 0.76-0.91)). Integration of GEP with the clinical model did not improve accuracy. Gene set enrichment analyses showed enrichment of proliferative and stress-related pathways.

Conclusion

While GEP captured biologically relevant signals, its predictive accuracy for distant metastases was comparable to that of Breslow thickness and ulceration in a population-based early-stage melanoma cohort.

Introduction

Accurate prognostication in early-stage cutaneous melanoma is critical for identifying patients at risk of progression. The American Joint Committee on Cancer (AJCC) staging system, based on Breslow thickness and ulceration, has demonstrated robust prognostic utility [1]. Nevertheless, these clinicopathological parameters alone do not fully capture the biological heterogeneity that drives variability in patient outcomes, as a subset of early-stage patients still develops distant metastases despite favorable initial staging [2-4]. Approximately 41% of melanoma deaths and 60% of distant metastases occur in patients initially diagnosed with stage I/II disease [2,4]. This underscores the need for molecular biomarkers that extend beyond current staging parameters to better capture the drivers of progression in early-stage melanoma.

Gene expression profiling (GEP) has emerged as a promising approach to refine risk prediction by identifying molecular signatures [5,6]. Several GEP assays have been developed for melanoma risk stratification [7-11]. CP-GEP (commercially known as the Merlin assay, SkylineDx) was developed to identify patients at low risk of sentinel lymph node metastasis who may forgo sentinel lymph node biopsy (SLNB) and has been validated in multiple studies [12-14]. Other assays, including DecisionDx-Melanoma (Castle Biosciences, Inc.), MelaGenix (NeraCare GmbH), LMC_150 (Leeds Melanoma Cohort) and Cam_121 (Cambridge Cancer Center), have demonstrated their ability to identify patient subgroups at higher risk of progression [8-11]. Despite their potential to improve risk stratification, current clinical guidelines from the National Comprehensive Cancer Network (NCCN), the Society of Surgical Oncology (SSO) and the

79 American Academy of Dermatology (AAD) do not yet recommend routine GEP testing, awaiting results
80 from prospective validation in large datasets of unselected patients [15-17].

81 Development and validation of clinically useful GEP models require large, well-powered studies with
82 sufficient metastatic events to ensure stable, generalizable predictions and to demonstrate clinical
83 utility [5,6]. A critical concern is insufficient adjustment for established prognostic variables, which may
84 result in high-risk GEP classifications that primarily reflect known high-risk features rather than providing
85 independent prognostic information [18]. While multivariable adjustment can address this issue,
86 application to cohorts with limited events risks statistical instability, model overfitting, and wide
87 confidence intervals despite high hazard ratios, as illustrated in prior evaluations [9,19].

88 The Dutch Early-Stage Melanoma (D-ESMEL) study was designed to address these challenges by
89 evaluating the value of GEP for identifying early-stage melanoma patients at risk of distant metastasis
90 beyond well-established clinicopathological factors. We employed a matched case-control discovery set
91 to control for key prognostic variables, followed by model development and independent validation in a
92 nationwide cohort to enable robust performance assessment [20]. Using this study design, we sought to
93 develop and validate a GEP model for predicting distant metastasis risk in early-stage melanoma and
94 determine whether it provides added value over current clinicopathological staging factors.

95 Methods

96 Study design and patient population

97 This study utilized the methodological framework of the D-ESMEL study, drawing data from the
98 Netherlands Cancer Registry (NCR) and the Dutch Nationwide Pathology Databank (PALGA), as
99 previously described [20]. In brief, the discovery set comprised 221 matched case-control pairs of stage I/II
100 melanoma patients matched on age, sex, Breslow thickness, and ulceration. Cases were patients who

101 developed distant metastases (stage IV disease based on the AJCC); controls were patients who
102 remained distant metastases-free within the time period in which the matched case developed distant
103 metastases. This design enabled identification of candidate genes independent of known
104 clinicopathological factors.

105 To develop and validate a prognostic model, we established a validation cohort using a nested case-
106 control (NCC) design within a nationwide cohort of 5,815 patients diagnosed with stage I/II melanoma
107 between 2016 and 2020. This included 154 case-control sets matched on follow-up time. The validation
108 cohort was split into two-thirds for model development and one-third for independent validation. The
109 NCC design permits estimation of absolute risks [21-23].

110 This study was approved by the Erasmus MC Ethics Committee (MEC-2018-1738), the scientific
111 committee of the NCR and the Privacy Committee of PALGA. The code of conduct for responsible use of
112 human residual tissue by the Federation of Dutch Medical Scientific Societies was followed [24].

113 RNA extraction, library preparation, sequencing, and bioinformatic processing are described in detail in
114 Supplementary Methods. The number of samples yielding successful transcriptomic data is reported in
115 the Results section.

116 Differential gene expression and modelling in the discovery set

117 To identify candidate genes predictive of distant metastasis, we combined differential gene expression
118 analysis (DGEA) with penalized discriminative modeling in the discovery set. DGEA was conducted using
119 multiple statistical frameworks, and the top-ranked genes were aggregated. In parallel, least absolute
120 shrinkage and selection operator (LASSO)-penalized conditional logistic regression approach with
121 double-loop cross-validation (DLCV) was used to select and internally validate predictive genes [25].
122 Candidate genes were selected as those with significantly higher selection frequency, assessed through

123 permutation testing. Detailed methodological procedures are provided in Supplementary Figure 1 and
124 Supplementary Methods.

125 Model development

126 Following identification of candidate genes in the discovery set, model development was conducted in
127 two-thirds of the validation cohort (hereafter referred to as the model development subset) to assess
128 generalizability and predictive accuracy of the gene set.

129 First, we performed DGEA using DESeq2 [26] and limma [27] to determine how many genes identified in the
130 discovery set remained differentially expressed in the model development subset. Models were
131 conditioned for the matching variables from the discovery set and AJCC stage.

132 To assess the predictive performance of the gene set identified through DGEA and DLCV in the discovery
133 set, we constructed baseline models using logistic regression with LASSO regularization: (1)
134 clinicopathological variables (Breslow thickness and ulceration), (2) AJCC stage, (3) matching variables,
135 (4) gene expression data, and (5) an integrated model combining clinical variables and gene expression.
136 All genes identified in the discovery set were initially included as input features. Alternative feature
137 selection strategies were explored, including adaptive LASSO and ridge regression. Additionally, we
138 examined whether subsets of 558 genes with higher selection frequency across models improved
139 predictive accuracy.

140 Discriminative performance was assessed using both weighted area under the curve (AUC) and weighted
141 concordance index (C-index) with corresponding 95% confidence intervals (95% CI). Given the NCC
142 design and complete follow-up within the nationwide source cohort, sampling weights were applied to
143 generate representative performance estimates [21]. These weights were based on the distribution of
144 follow-up time, as controls were randomly sampled without matching on prognostic factors. Internal

145 validation consisted of 200 bootstrap iterations, where matched case-control pairs were kept together
146 during resampling. The .632+ estimator was used for internally validated metrics [28].

147 Independent validation

148 The best-performing and most robust models from the model development subset were applied to the
149 remaining one-third of the validation cohort (hereafter referred to as the independent validation
150 subset), which was withheld during model development, to evaluate their predictive performance.
151 Weighted AUC and C-index with corresponding 95% CI were recalculated using the methods previously
152 described. Sampling weights were applied based on the corresponding one-third of the nationwide
153 source cohort. This held-out subset enabled assessment of model stability, potential overfitting, and
154 reproducibility.

155 GEP models trained within the model development subset

156 In a complementary analysis, we trained GEP models entirely within the model development subset,
157 without relying on genes identified in the discovery set. Instead, models were built using all genes with
158 reliable expression measurements, defined as $\log_2(\text{TPM} +1) > 0.5$ in at least half of the model
159 development subset and limited to protein-coding and long non-coding RNA (n = 16,661 genes). Models
160 were trained using 200 bootstrap iterations with LASSO regularization across input sets of 50 to 1000
161 genes with the highest variances. Performance was assessed using weighted AUC and weighted C-index
162 in the model development subset and the independent validation subset.

163 Gene set enrichment analyses

164 To investigate the biological pathways associated with metastatic progression and established
165 prognostic factors, we performed gene set enrichment analysis using the MSigDB Hallmark gene sets [29]
166 and the DGEA on the model development subset of the validation cohort. Three complementary
167 analyses were performed: (1) cases versus controls conditioned on matching variables, (2) melanomas

168 with Breslow thickness >2 mm versus ≤2 mm, and (3) melanomas with versus without ulceration. Genes
169 were ranked by moderated t-statistics from limma, and enrichment scores were calculated using the
170 fgsea algorithm. Significance of enrichment was assessed based on normalized enrichment scores (NES)
171 and adjusted p-values (false discovery rate <0.05).

172 Exploratory analyses of publicly available gene sets

173 We conducted exploratory analyses using previously published gene sets as positive methodological
174 controls [7-11]. Although these gene signatures differ in endpoints, technical platforms, and intended
175 patient populations compared with our study, we hypothesized that they would still show association
176 with distant metastasis in our cohort, thereby serving as an internal validation of our methodology. We
177 examined the distribution of differentially expressed genes from each signature using volcano plots.
178 Each gene set was then used as input for logistic regression models with LASSO regularization and
179 performance evaluated using weighted AUC and weighted C-index metrics with corresponding 95% CI in
180 the model development subset. Accordingly, our analyses should be interpreted as exploratory checks
181 of methodological consistency, not as comparative assessments of clinical performance.

182 Results

183 Discovery set and validation cohort

184 The characteristics of the discovery set, the validation cohort, including the model development and the
185 independent validation subsets, and their respective source cohorts have been described previously [20].
186 For the present analysis, only samples with successful RNA sequencing (RNAseq) after quality control
187 measures were included (Supplementary Figure 2). In total, RNAseq data were successfully generated
188 for 178 case-control pairs in the discovery set, comprising 356 patients matched on age, sex, Breslow
189 thickness, and ulceration. In the validation cohort, RNAseq data were successfully generated for 147

190 case-control pairs. This cohort was divided into two subsets: the model development subset (*n* = 100

191 pairs) and the independent validation subset (*n* = 47 pairs). The corresponding source cohort sizes

192 (3,776 and 1,775 patients, respectively) were calculated based on the sampling fractions of these

193 subsets within the same nationwide cohort (n = 5,551) (Table 1).

194 Candidate prognostic genes based on the discovery set

195 We identified 288 differentially expressed genes between cases and controls using a union-based

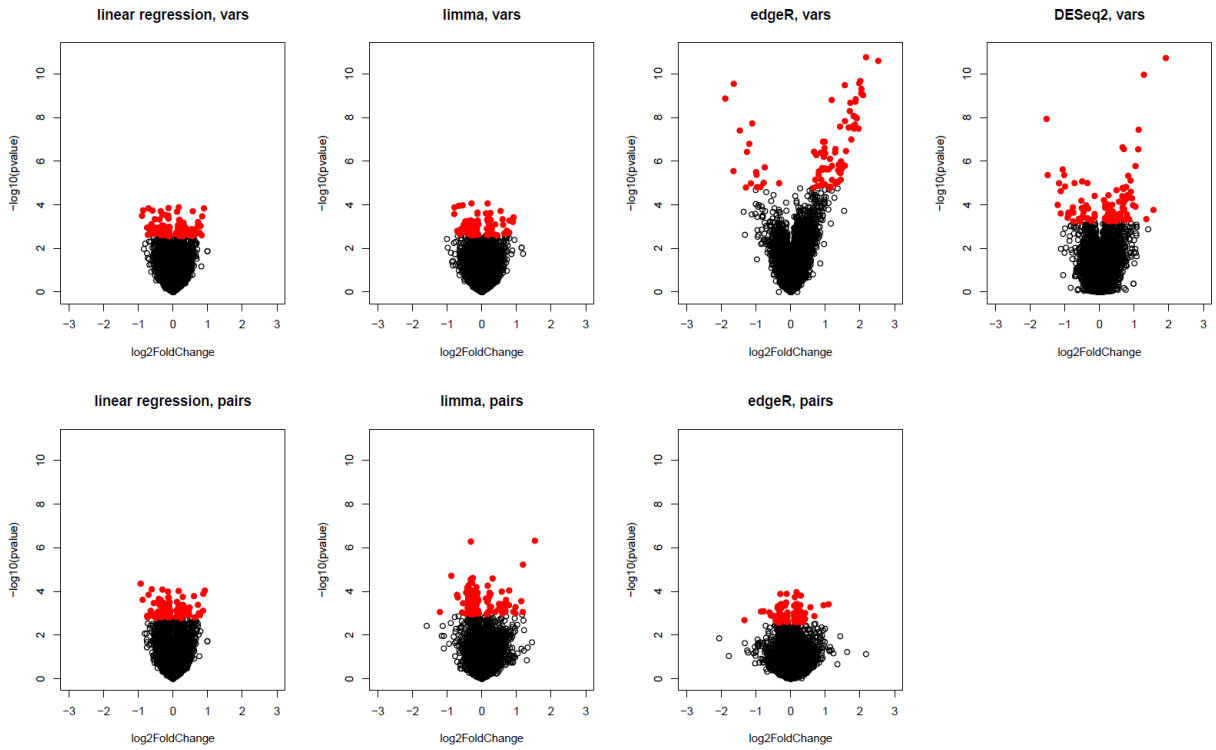
196 selection approach across multiple statistical models. The selection was based on the top 100 genes

197 with the smallest *p*-values from seven differential expression models, each conditioned on either

198 matching variables or the paired nature of the data. We ultimately did not use DESeq2 conditioned on

199 pairs due to unstable dispersion estimates. The resulting volcano plots for all models are shown in Figure

200 1.



201

202 **Fig 1.** Volcano plots from seven differential gene expression analysis models comparing cases and
203 matched controls in the discovery set. Each model was conditioned on matching variables (top) and on
204 case-control pairs (bottom). Genes highlighted in red represent the 100 genes with the smallest p-values
205 in each model.

206 LASSO conditional logistic regression with DLCV resulted in 350 genes that were most frequently
207 (defined as $\geq$ 20%) selected. Permutation testing with 100 randomized case-control label assignments
208 (Supplementary Figure 1) demonstrated that models built on original data performed better than the
209 models based on permuted data, with original data models achieving an AUC of ~0.6 with 500 or more
210 input genes (Supplementary Figure 3). Combining both approaches, we constructed a final gene set
211 comprising 558 unique genes, including 80 overlapping entries. A list of these genes is provided in
212 Supplementary Table 1.

213 Model development

214 DGEA in the model development subset without adjustment revealed 8,001 significantly differentially
215 expressed genes. When conditioning on matching variables of the discovery set, the number of
216 significantly differentially expressed genes was markedly reduced (n = 526). No significantly differentially
217 expressed genes remained when conditioning on AJCC stage (Supplementary Figure 4A-C). The 558
218 genes identified in the discovery set were not more significant than other genes in the model
219 development subset, when conditioned for matching variables (Supplementary Figure 4D and E).

220 Next, we compared the predictive performance of clinical, gene-based, and integrated models. Breslow
221 thickness combined with ulceration achieved a weighted AUC of 0.82 (95% CI, 0.78-0.86) and a weighted
222 C-index of 0.85 (95% CI, 0.72-0.88), which was comparable to models based on AJCC stage (weighted
223 AUC 0.79 [95% CI, 0.75-0.82], weighted C-index 0.83 [95% CI, 0.79-0.86]) or on the full set of matching
224 variables from the discovery set (age, sex, ulceration, and Breslow thickness) (weighted AUC 0.79, [95%

225 CI, 0.74-0.83], weighted C-index 0.83, [95% CI, 0.78-0.86]). Adding the 558 genes to Breslow thickness
226 and ulceration did not improve performance (weighted AUC 0.80 [95% CI, 0.74-0.88], weighted C-index
227 0.83 [95% CI, 0.77-0.89]) compared to Breslow thickness and ulceration alone (Figure 2A; Table 2).
228 Various gene selection via regularization methods were explored, but neither adaptive LASSO nor ridge
229 regression improved predictive accuracy. We also tested whether using subsets of genes with higher
230 selection frequency in the discovery set (*n* = 11, 80, or 117) would enhance performance, but no
231 improvement was observed (Table 2; Supplementary Figure 5).

232 Independent validation

233 In the independent validation subset, we evaluated models developed in the model development
234 subset. The clinical, gene-based, and integrated models showed similar performances; Breslow thickness
235 and ulceration achieved a weighted AUC of 0.82 (95% CI, 0.73-0.90), and weighted C-index of 0.84 (95%
236 CI, 0.76-0.91), the gene-only model showed a weighted AUC of 0.77 (95% CI, 0.66-0.86) and a weighted
237 C-index of 0.79 (95% CI, 0.69-0.88); and the combined model showed a weighted AUC of 0.79 (95% CI,
238 0.68-0.87) and a weighted C-index of 0.81 (95% CI, 0.71-0.89) (Table 2, Figure 2B).

239 Association of discovery genes with clinicopathological variables

240 To further explore the extent to which the 558 discovery genes reflected established prognostic
241 variables, we first examined their direct association with Breslow thickness and ulceration. We observed
242 that 75% of the genes were associated with ulceration and 79% with Breslow thickness in the discovery
243 set. In the model development subset of the validation cohort, 96% of genes correlated with both
244 variables. The aggregated prediction score based on the 558 genes showed a strong correlation with
245 Breslow thickness (r = 0.74) and significantly higher values in ulcerated compared with non-ulcerated
246 tumors ($p = 1.1 \times 10^{-14}$) (Supplementary Figure 6).

247 We then conducted principal component analysis (PCA) on the 558 discovery genes to further evaluate
248 their independent contribution. PC1 was strongly associated with clinical stage, indicating overlap with
249 known prognostic features (Supplementary Figure 7A). Adding clinical predictors to models based on the
250 remaining PCs (excluding PC1) did not improve predictive performance, suggesting that GEP adds limited
251 independent signal (Table 2, Figure 2). PCA-derived risk scores based on the remaining components
252 appeared largely orthogonal to the clinical model, implying that gene expression may reflect distinct
253 underlying biology (Supplementary Figure 7B).

254 GEP models trained within the model development subset

255 GEP models trained within the model development subset (without relying on the 558 genes identified
256 in the discovery set) showed comparable predictive performance in the independent validation subset.
257 The best-performing model, based on 50 input features, achieved a weighted c-index of 0.81 (95% CI,
258 0.71-0.89) (Table 2). The results were comparable to those of the clinical model based on Breslow
259 thickness and ulceration.

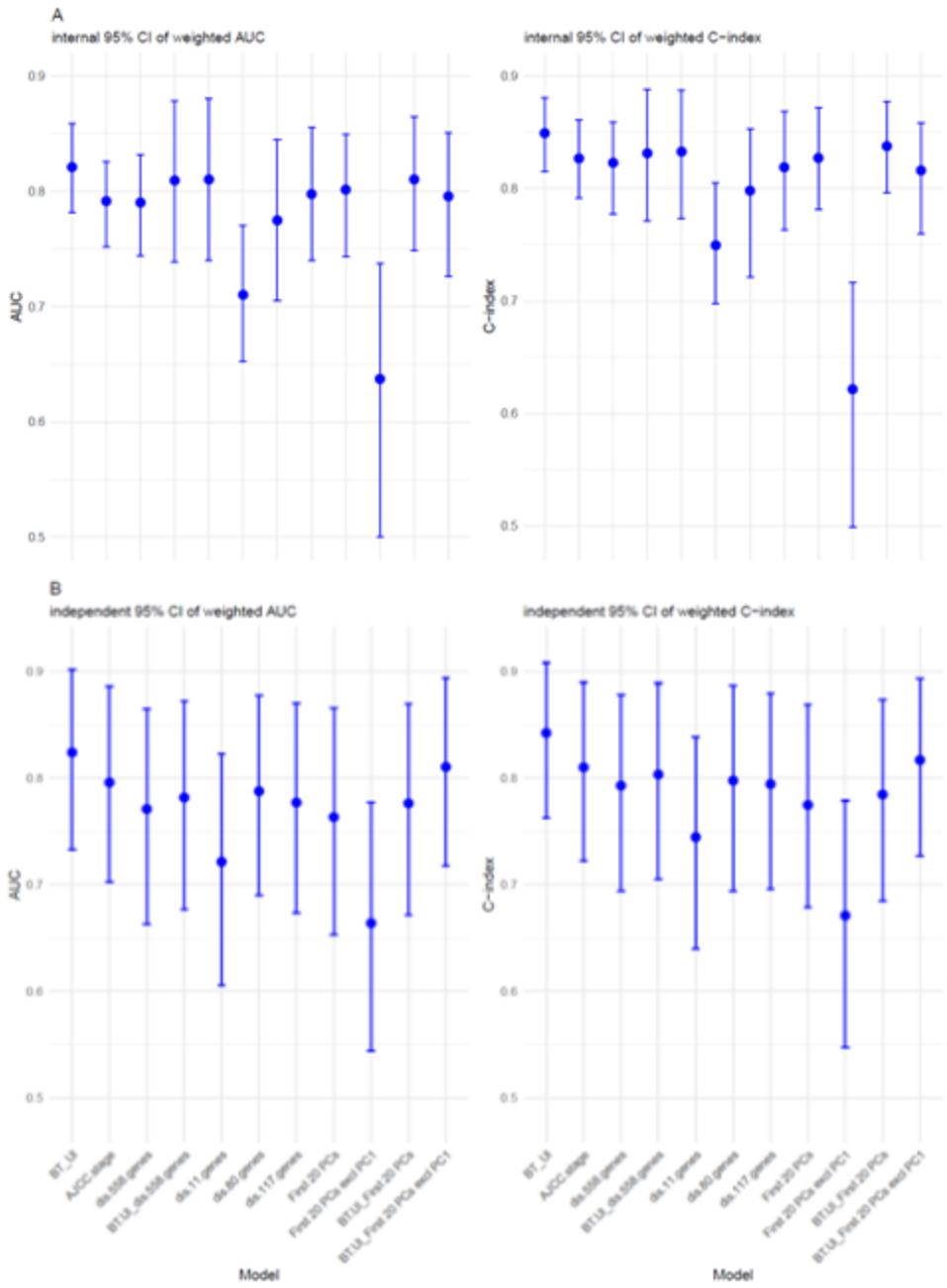


Fig 2. Weighted area under the curve (AUC) (left) and C-index (right) for all prediction models, evaluated in the model development subset (A) and independent validation subset (B). Models include: Breslow thickness plus ulceration (BT_Ul); American Joint Committee of Cancer (AJCC) stage; 558-gene signature; 558-gene signature plus BT_Ul; 11-gene, 80-gene, and 117-gene subsets; first 20 principal components (PC20); PC20 excluding PC1; PC20 plus BT_Ul; and PC20 excluding PC1 plus BT_Ul.

266 ## Gene set enrichment analyses

267 Gene set enrichment analyses were performed in the model development to explore biological
268 processes associated with metastatic progression. When conditioned on the matching variables, several
269 hallmark gene sets were significantly enriched in metastatic cases. The top enriched pathways included
270 E2F targets, MYC targets (V1), and the G2M checkpoint, all showing strong positive enrichment with
271 high normalized enrichment scores (NES). Additional enrichment was seen for hypoxia, estrogen
272 response, unfolded protein response, and both interferon gamma and alpha signaling (Figure 3A).

273 To assess whether these signals overlap with established clinicopathological risk factors, we repeated
274 the analyses stratified by Breslow thickness and ulceration. In thicker melanomas (>2 mm), the same
275 proliferative programs (MYC targets, E2F targets, G2M checkpoint) were again among the top enriched
276 gene sets, accompanied by oxidative phosphorylation, DNA repair, and mTORC1 signaling (Figure 3B).
277 Similarly, in ulcerated melanomas, a broad set of hallmark pathways was enriched, including
278 proliferative signaling (E2F, MYC, G2M), metabolic activity (glycolysis, oxidative phosphorylation, fatty
279 acid metabolism), and immune responses (inflammatory response, interferon signaling) (Figure 3C).

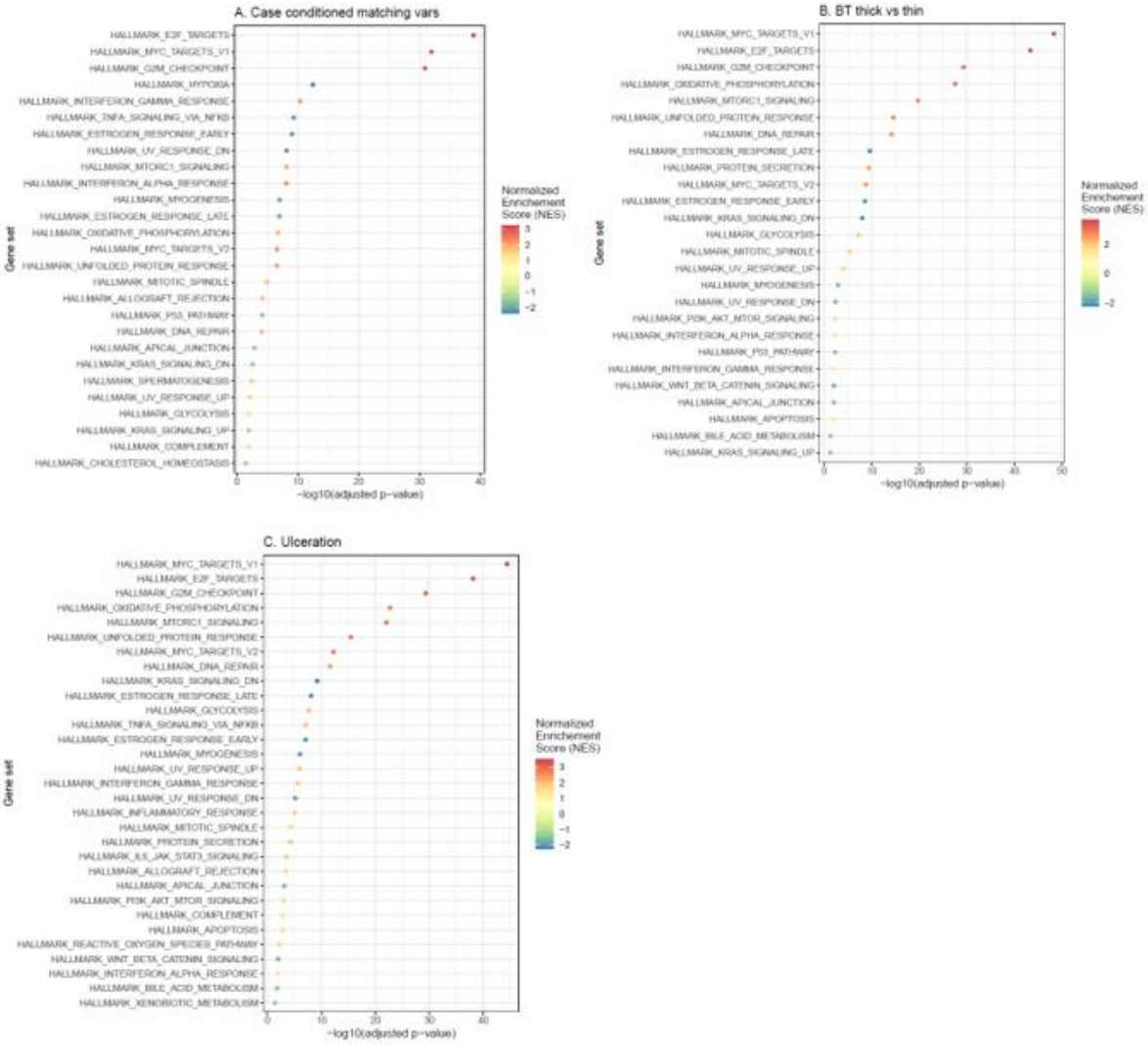
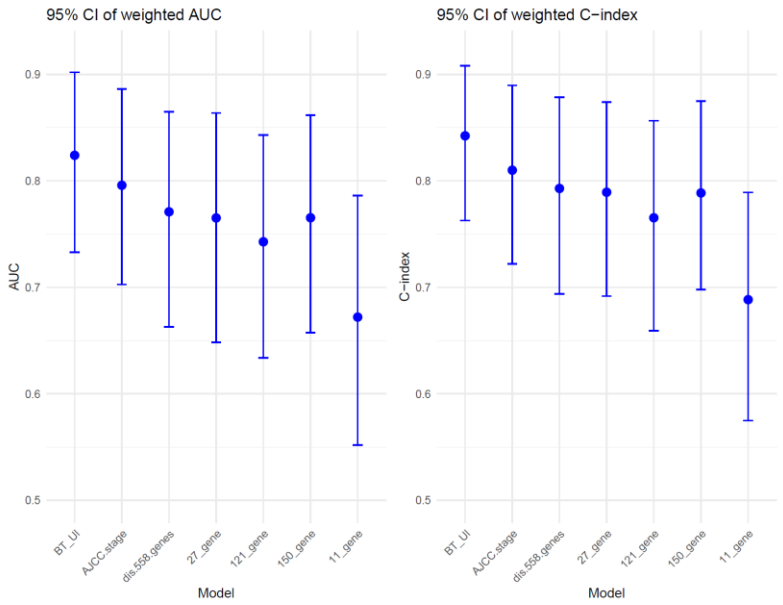


Fig 3. Gene set enrichment analyses (GSEA) of hallmark gene set in the model development subset of the validation cohort. A. Case-control analysis conditioned on matching variables (age, sex, Breslow thickness, ulceration). B. GSEA following differential expression analysis by Breslow thickness (>2 mm vs ≤2 mm). C. GSEA following differential expression analysis by ulceration status (present vs absent).

Exploratory analyses of publicly available gene sets

We conducted exploratory analyses of previously published gene sets within the model development subset of the validation cohort [7-11]. None of these sets showed a distinct differential expression pattern

288 in volcano plots (Supplementary Figure 8). Predictive models built from these published gene sets
289 demonstrated discriminative performance similar to that of our 558-gene set, and comparable to
290 models based on Breslow thickness with ulceration or on AJCC stage (Figure 4).



291

292 **Fig 4.** Weighted area under the curve (AUC (left) and weighted C-index (right) with 95% confidence
293 intervals based on 1,000 bootstrap iterations (B = 1000) are shown for models based on Breslow
294 thickness and ulceration (BT_UI), AJCC stage, the 558 discovery genes, and previously published gene
295 sets.

296 Discussion

297 Accurately identifying early-stage melanoma patients at risk of distant metastases remains clinically
298 challenging. Using data from the D-ESMEL study, we developed and validated GEP models for predicting
299 distant metastases in stage I/II melanoma. While the GEP revealed biologically relevant transcriptional
300 differences, its predictive performance for distant metastases was comparable to that of Breslow
301 thickness and ulceration. These findings suggest that the prognostic information captured by gene
302 expression largely overlaps with, and may partly reflect, established clinicopathological variables.

303 GEP assays have demonstrated significant promise in melanoma risk stratification. A recent systematic
304 review and meta-analyses of the DecisionDx-Melanoma (Castle Biosciences Inc.) with >14,000 patients
305 have demonstrated consistent risk stratification with significantly different survival outcomes [30]. CP-GEP
306 (SkylineDx) has been validated in multiple independent cohorts, with a recent systematic review and
307 meta-analysis involving >1,000 patients demonstrating that CP-GEP performs well in identifying
308 melanoma patients at low risk of SLN positivity, showing high negative predictive value (95%),
309 supporting its value in SLNB decision-making [31]. With ongoing prospective trials like NivoMela
310 (NCT04309409), the utility of Melagenix (NeraCare GmbH) in guiding adjuvant therapy decisions is being
311 evaluated. At the same time, translating GEP promise into consistent clinical utility faces methodological
312 challenges, particularly in studies with relatively few outcome events. For example, population-based
313 SEER data reported high hazard ratios for DecisionDx-Melanoma classifications, but with only 68/4,687
314 melanoma-specific deaths, the 95% CI's were wide (HR 28.25 [95% CI: 12.69–62.89] univariable; HR 7.00
315 [95% CI: 2.70–18.00] multivariable) [19]. Similarly, the LMC_150 signature was associated with melanoma-
316 specific survival in stage I patients, but effect estimates were less precise after adjustment for
317 clinicopathological variables (e.g., HR 9.8 [95% CI: 1.1–86.2]) [9]. The Cam_121 signature, developed in
318 stage II disease, showed minimal discrimination in the earlier-stage Leeds Melanoma Cohort,
319 highlighting potential limitations in generalizability across stages (five-year melanoma-specific mortality
320 9% vs. 8%) [10]. These examples highlight the importance of study design, outcome definitions, and
321 patient selection when evaluating the added value of GEPs.

322 Our study design differs from prior GEP investigations by explicitly focusing on distant metastases as the
323 primary endpoint, excluding patients with only local or regional relapse. This approach emphasizes the
324 most clinically relevant outcome and leverages our population-based cohort with extended follow-up.
325 As outlined in methodological guidance, prognostic models should be developed using adequately
326 powered datasets with sufficient events per variable, include internal validation to quantify optimism,

327 undergo external validation to assess generalizability, and demonstrate incremental value beyond
328 established prognostic factors [18,32]. The D-ESMEL study was designed to meet these criteria, including a
329 matched discovery design to isolate signals beyond age, sex, Breslow thickness, and ulceration; and a
330 large, population-based validation cohort. As an internal validation of our methodology, we further
331 evaluated previously published melanoma gene sets. Although technical and clinical differences
332 preclude direct performance comparisons, the concordant results we observed underscore the
333 robustness of our dataset and analyses.

334 Several factors may explain why the discovery genes did not translate into stronger predictive models.
335 Despite tight matching on Breslow thickness and ulceration in the discovery set, many genes remained
336 strongly correlated with these factors. Matching balances variables at the cohort level but does not
337 remove their intrinsic biological links with gene expression. In contrast, the validation cohort reflected a
338 natural clinical distribution with more thin melanomas among controls. However, stratified analyses by
339 AJCC stage showed comparable model performance across stage subgroups. Differences in SLNB uptake
340 may also have played a role. Only 32% of discovery cases underwent SLNB, compared with 47% in the
341 validation cohort, reflecting temporal shifts in clinical practice after the introduction of adjuvant therapy
342 in 2019. Lower SLNB uptake may have resulted in under detection of nodal metastases. In contrast,
343 higher SLNB rates in the validation cohort, likely improved staging accuracy, reducing the proportion of
344 occult stage III cases. These differences in case composition, combined with the strong correlation
345 between gene expression and clinicopathological features, likely account for their comparable predictive
346 performance. Despite this overlap, our GEP model captured biologically relevant information. In the
347 independent validation subset, the GEP models demonstrated solid performance (weighted C-index =
348 0.79).

349 Gene set enrichment analyses revealed consistent enrichment of proliferative programs, as well as
350 hypoxia, estrogen response, unfolded protein response, and interferon signaling: pathways known to

351 drive tumor aggressiveness [33-37]. Importantly, the same pathways were enriched in thicker and ulcerated
352 tumors, reinforcing the notion that the biology captured by GEP mirrors established high-risk
353 clinicopathological features. In line with prior data, thicker tumors showed upregulation of DNA repair
354 and cell-cycle genes [38], while ulcerated primaries exhibited proliferative, vascular, and inflammatory
355 profiles [39]. These findings reaffirm that the prognostic signal captured by GEP substantially overlaps with
356 the biology already reflected by Breslow thickness and ulceration. Nonetheless, the biological insights
357 remain valuable. GEP may support biological stratification in research settings, integration into multi-
358 modal models combining molecular, spatial, or immune features, or provide supportive context in
359 histologically ambiguous cases.

360 In summary, the D-ESMEL study provides a population-based, methodologically rigorous evaluation of
361 GEP for predicting distant metastases in early-stage melanoma. Our findings reaffirm the dominant role
362 of Breslow thickness and ulceration, show that GEP captures overlapping biology, and highlight key
363 considerations for translating molecular data into clinically useful risk stratification.

364 Acknowledgments

365 We thank all dedicated registrars of the NCR and Palga for their work on the data collection. We thank
366 the Erasmus MC Pathology Research and Trial Service (PARTS) for sectioning the tissue material. The
367 project was co-funded by the PPP Allowance made available by Health∼Holland (Grant Number:
368 EMCLSH19008), Top Sector Life Sciences & Health.

369 Data availability

370 The clinical data used in this study were obtained from the Netherlands Cancer Registry (NCR) and are
371 available through the Netherlands Comprehensive Cancer Organization. These data are not publicly
372 available, and restrictions apply to their use. Access may be granted upon reasonable request and with

373 permission from the Netherlands Comprehensive Cancer Organization (request numbers K19.037 and
374 K21.277). The list of 558 candidate genes is provided in Supplementary Table 1. Gene expression data
375 (counts per gene), including sample ID, pair ID, and case-control status, will be made available via a
376 public repository (pending).

377 References

378 1. Gershenwald JE, Scolyer RA, Hess KR, et al: AJCC cancer staging manual. Switzerland:
379 Springer:563-89, 2017
380 2. Zhou C, Louwman M, Wakkee M, et al: Primary Melanoma Characteristics of Metastatic
381 Disease: A Nationwide Cancer Registry Study. Cancers (Basel) 13, 2021
382 3. Leeneman B, Franken MG, Coupé VMH, et al: Stage-specific disease recurrence and
383 survival in localized and regionally advanced cutaneous melanoma. Eur. J. Surg. Oncol. 45:825-831, 2019
384 4. Enninga EAL, Moser JC, Weaver AL, et al: Survival of cutaneous melanoma based on sex,
385 age, and stage in the United States, 1992-2011. Cancer Med. 6:2203-2212, 2017
386 5. Grossman D, Okwundu N, Bartlett EK, et al: Prognostic Gene Expression Profiling in
387 Cutaneous Melanoma: Identifying the Knowledge Gaps and Assessing the Clinical Benefit. JAMA
388 Dermatol 156:1004-1011, 2020
389 6. Sun J, Karasaki KM, Farma JM: The Use of Gene Expression Profiling and Biomarkers in
390 Melanoma Diagnosis and Predicting Recurrence: Implications for Surveillance and Treatment. Cancers
391 (Basel) 16, 2024
392 7. Meves A, Nikolova E, Heim JB, et al: Tumor Cell Adhesion As a Risk Factor for Sentinel
393 Lymph Node Metastasis in Primary Cutaneous Melanoma. J. Clin. Oncol. 33:2509-15, 2015
394 8. Gerami P, Cook RW, Wilkinson J, et al: Development of a prognostic genetic signature to
395 predict the metastatic risk associated with cutaneous melanoma. Clin. Cancer. Res. 21:175-83, 2015

396 9. Thakur R, Laye JP, Lauss M, et al: Transcriptomic Analysis Reveals Prognostic Molecular
397 Signatures of Stage I Melanoma. Clin. Cancer Res. 25:7424-7435, 2019
398 10. Garg M, Couturier DL, Nsengimana J, et al: Tumour gene expression signature in primary
399 melanoma predicts long-term outcomes. Nat. Commun. 12:1137, 2021
400 11. Brunner G, Heinecke A, Falk TM, et al: A Prognostic Gene Signature Expressed in Primary
401 Cutaneous Melanoma: Synergism With Conventional Staging. JNCI Cancer Spectr 2:pky032, 2018
402 12. Bellomo D, Arias-Mejias SM, Ramana C, et al: Model Combining Tumor Molecular and
403 Clinicopathologic Risk Factors Predicts Sentinel Lymph Node Metastasis in Primary Cutaneous
404 Melanoma. JCO Precis. Oncol. 4:319-334, 2020
405 13. Mulder E, Dwarkasing JT, Tempel D, et al: Validation of a clinicopathological and gene
406 expression profile model for sentinel lymph node metastasis in primary cutaneous melanoma. Br J
407 Dermatol 184:944-951, 2021
408 14. Stassen RC, Mulder E, Mooyaart AL, et al: Clinical evaluation of the clinicopathologic and
409 gene expression profile (CP-GEP) in patients with melanoma eligible for sentinel lymph node biopsy: A
410 multicenter prospective Dutch study. Eur J Surg Oncol 49:107249, 2023
411 15. National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology
412 Melanoma: Version 2.2025. , 2025
413 16. Bartlett EK, O'Donoghue C, Boland G, et al: Society of Surgical Oncology Consensus
414 Statement: Assessing the Evidence for and Utility of Gene Expression Profiling of Primary Cutaneous
415 Melanoma. Ann Surg Oncol 32:1429-1442, 2025
416 17. Swetter SM, Tsao H, Bichakjian CK, et al: Guidelines of care for the management of
417 primary cutaneous melanoma. J. Am. Acad. Dermatol. 80:208-250, 2019
418 18. Steyerberg E: Clinical Prediction Models: A Practical Approach to Development,
419 Validation, and Updating, Springer, 2019

420 19. Bailey CN, Martin BJ, Petkov VI, et al: 31-Gene Expression Profile Testing in Cutaneous
421 Melanoma and Survival Outcomes in a Population-Based Analysis: A SEER Collaboration. JCO Precis
422 Oncol 7:e2300044, 2023
423 20. Zhou C, Mooyaart AL, Kerkour T, et al: The Dutch Early-Stage Melanoma (D-ESMEL)
424 study: a discovery set and validation cohort to predict the absolute risk of distant metastases in stage I/II
425 cutaneous melanoma. Eur J Epidemiol 40:27-42, 2025
426 21. Rentroia-Pacheco B, Bellomo D, Lakeman IMM, et al: Weighted metrics are required
427 when evaluating the performance of prediction models in nested case-control studies. BMC Med. Res.
428 Methodol. 24:115, 2024
429 22. Noordzij M, van Diepen M, Caskey FC, Jager KJ: Relative risk versus absolute risk: one
430 cannot be interpreted without the other. Nephrol. Dial. Transplant 32:ii13-ii18, 2017
431 23. Rentroia-Pacheco B, Steijlen O, Bellomo D, et al: Nested Case-Control and Case-cohort:
432 efficient study designs to develop biomarker-based prediction models for rare outcomes. J Invest
433 Dermatol, Submitted
434 24. Netherlands Code of Conduct for Research Integrity, The Association of Universities in
435 the Netherlands, 2018
436 25. Tibshirani R: Regression Shrinkage and Selection via the Lasso. Journal of the Royal
437 Statistical Society Series B (Methodological) 58:267-288, 1996
438 26. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for
439 RNA-seq data with DESeq2. Genome Biol 15:550, 2014
440 27. Ritchie ME, Phipson B, Wu D, et al: limma powers differential expression analyses for
441 RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47, 2015
442 28. Efron B, Tibshirani R: Improvements on Cross-Validation: The .632+ Bootstrap Method.

443 29. Liberzon A, Birger C, Thorvaldsdóttir H, et al: The Molecular Signatures Database
444 (MSigDB) hallmark gene set collection. Cell Syst 1:417-425, 2015
445 30. Durgham RA, Nassar SI, Gun R, et al: The Prognostic Value of the 31-Gene Expression
446 Profile Test in Cutaneous Melanoma: A Systematic Review and Meta-Analysis. Cancers (Basel) 16, 2024
447 31. Wong T, Ch'ng S, Ferguson P, et al: Predictive performance of the clinicopathologic gene
448 expression profile (CP-GEP) in identifying cutaneous melanoma patients for whom sentinel lymph node
449 biopsy is unnecessary: A systematic review and meta-analysis. Crit Rev Oncol Hematol 214:104816, 2025
450 32. Collins GS, Reitsma JB, Altman DG, Moons KG: Transparent reporting of a multivariable
451 prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj 350:g7594,
452 2015
453 33. Alla V, Engelmann D, Niemetz A, et al: E2F1 in melanoma progression and metastasis. J
454 Natl Cancer Inst 102:127-33, 2010
455 34. Ross DA, Wilson GD: Expression of c-myc oncoprotein represents a new prognostic
456 marker in cutaneous melanoma. Br J Surg 85:46-51, 1998
457 35. Cubillos-Ruiz JR, Bettigole SE, Glimcher LH: Tumorigenic and Immunosuppressive Effects
458 of Endoplasmic Reticulum Stress in Cancer. Cell 168:692-706, 2017
459 36. Di Conza G, Ho PC, Cubillos-Ruiz JR, Huang SC: Control of immune cell function by the
460 unfolded protein response. Nat Rev Immunol 23:546-562, 2023
461 37. Grasso CS, Tsoi J, Onyshchenko M, et al: Conserved Interferon-γ Signaling Drives Clinical
462 Response to Immune Checkpoint Blockade Therapy in Melanoma. Cancer Cell 38:500-515 e3, 2020
463 38. Fang S, Vaysse A, Brossard M, et al: Melanoma Expression Genes Identified through
464 Genome-Wide Association Study of Breslow Tumor Thickness. J Invest Dermatol 137:253-257, 2017
465 39. Jewell R, Elliott F, Laye J, et al: The clinicopathological and gene expression patterns
466 associated with ulceration of primary melanoma. Pigment Cell Melanoma Res 28:94-104, 2015

467

468 *Table 1. Clinical characteristics of the discovery set, validation cohorts and its nationwide source cohort. The validation cohort was randomly split into two-*
469 *thirds for model development and one-third for independent model validation.*

	Discovery set (*n* = 356)		**Nationwide validation cohort (*n* = 5,551)**					
			Model development subset (2/3)			**Independent validation subset (1/3)**		
Characteristic	Cases	Controls	Source cohort	All cases from source cohort	Randomly selected controls	Source cohort	All cases from source cohort	Randomly selected controls
	(*n* = 178)	(*n* = 178)	(*n* = 3,776)	(*n* = 100)	(*n* = 100)	(*n* = 1,775)	(*n* = 47)	(*n* = 47)
Sex (*n* (%))								
Male	111 (62)	111 (62)	1851 (49)	52 (52)	42 (42)	858 (48)	27 (57)	20 (43)
Female	67 (38)	67 (38)	1925 (51)	48 (48)	58 (58)	917 (52)	20 (43)	27 (57)
Age, years (median (IQR))	67 (50,71)	63 (52, 71)	63 (51, 73)	67 (54, 76)	63 (53, 71)	62 (51, 73)	69 (60, 77)	64 (51, 72)
Breslow thickness, mm (median (IQR))	1.9 (1.0, 3.4)	1.8 (1.0, 3.3)	0.8 (0.5, 1.4)	2.4 (1.3, 4.2)	0.8 (0.5-1.2)	0.8 (0.5, 1.4)	3.8 (1.3, 5.0)	0.8 (0.5, 1.8)
Ulcerated (*n* (%))	60 (34)	60 (34)	355 (9)	46 (46)	7 (7)	158 (9)	19 (40)	4 (9)
Stage (*n* (%))								
IA	33 (19)	33 (19)	2347 (62)	15 (15)	60 (60)	1101 (62)	4 (9)	25 (53)
IB	52 (29)	52 (29)	767 (20)	22 (22)	25 (25)	370 (21)	13 (28)	13 (28)
IIA	30 (17)	30 (17)	322 (9)	21 (21)	11 (11)	140 (8)	7 (15)	2 (4)
IIB	46 (26)	46 (26)	212 (6)	19 (19)	3 (3)	112 (6)	10 (21)	7 (15)
IIC	17 (10)	17 (10)	128 (3)	23 (23)	1 (1)	52 (3)	13 (28)	0 (0)
Body site (*n* (%))								
Face/neck	11 (6)	9 (5)	296 (8)	7 (7)	3 (3)	137 (8)	6 (13)	3 (6)
Scalp	20 (11)	11 (6)	194 (5)	9 (9)	5 (5)	114 (6)	5 (11)	2 (4)
Upper extremities	30 (17)	41 (23)	872 (23)	21 (21)	23 (23)	393 (22)	6 (13)	10 (21)

Trunk	86 (48)	77 (43)	1512 (40)	38 (38)	42 (42)	689 (39)	17 (36)	19 (40)
Lower extremities	31 (17)	40 (22)	899 (24)	25 (25)	27 (27)	441 (25)	13 (28)	13 (28)
Unknown	0 (0)	0 (0)	3 (0)	0 (0)	0 (0)	1 (0)	0 (0)	0 (0)
Morphological subtype (*n* (%))								
Superficial spreading	118 (66)	122 (69)	3042 (81)	68 (68)	89 (89)	1383 (78)	27 (57)	36 (77)
Nodular	40 (22)	45 (25)	276 (70)	24 (24)	6 (6)	153 (9)	16 (34)	7 (15)
Acral lentiginous	1 (1)	0 (0)	38 (1)	2 (2)	0 (0)	14 (1)	1 (2)	0 (0)
Lentigo maligna melanoma	2 (1)	1 (1)	194 (5)	2 (2)	1 (1)	107 (6)	2 (4)	1 (2)
Amelanotic melanoma	1 (1)	0 (0)	4 (0)	0 (0)	1 (1)	2 (0)	0 (0)	1 (2)
Desmoplastic melanoma	1 (1)	0 (0)	24 (1)	0 (0)	1 (1)	10 (1)	0 (0)	0 (0)
Unspecified	15 (8)	10 (6)	198 (5)	4 (4)	2 (2)	106 (6)	1 (2)	2 (4)
Sentinel lymph node biopsy performed (*n*(%))	56 (32)	71 (40)	1185 (31)	47 (47)	30 (30)	598 (34)	21 (45)	16 (34)
Time until stage IV recurrence, years (median (IQR))	3.5 (1.7, 5.3)	n/a	n/a	2.0 (1.1, 3.4)	n/a	n/a	1.6 (1.0, 3.2)	n/a
Time until stage IV recurrence (*n*(%))								
<2 years	57 (32)	n/a	n/a	51 (51)	n/a	n/a	29 (62)	n/a
2-5 years	67 (38)			46 (46)			18 (38)	
5-8 years	42 (24)			3 (3)			0 (0)	
>8 years	2 (7)			0 (0)			0 (0)	
Alive survival status at last follow-up (*n*(%))	49 (28)	137 (77)	3446 (91)	30 (30)	93 (93)	1628 (92)	12 (26)	43 (92)
Follow-up duration until censoring or death, years (median (IQR))	5.8 (3.4, 8.9)	9.7 (7.2, 12.1)	3.4 (2.2, 4.7)	3.7 (2.1, 5.5)	6.1 (5.0, 7.2)	3.4 (2.2, 4.6)	3.4 (1.8, 5.1)	6.1 (4.5, 7.3)

Follow-up duration until censoring or death, years (*n*(%))								
<2 years	20 (11)	6 (3)	763 (20)	24 (24)	1 (1)	392 (22)	15 (32)	1 (2)
2-5 years	52 (29)	9 (5)	2251 (20)	44 (44)	24 (24)	1052 (59)	20 (43)	14 (30)
5-8 years	53 (30)	42 (24)	762 (20)	32 (32)	75 (75)	331 (19)	12 (26)	31 (66)
>8 years	52 (29)	121 (68)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)
Disease progression after initial follow-up period (*n* (%))								
Stage III	n/a	22 (12)	n/a	n/a	4 (4)	n/a	n/a	1 (1)
Stage IV	n/a	18 (10)	n/a	n/a	1 (1)	n/a	n/a	0 (0)

*Table 2. Performance of prognostic models during model development and independent validation. Weighted AUC and C-index are shown for each model, using 0.632+ correction in the model development subset.*variable was included as continuous variable. AJCC = American Joint Committee of Cancer. GEP = gene expression profile.*

		Model development subset		**Independent validation subset**	
Model	**Features**	**Weighted AUC**	**Weighted C-index**	**Weighted AUC**	**Weighted C-index**
Clinicopathological	Breslow thickness*, ulceration	0.82 (95% CI 0.78-0.86)	0.85 (95% CI 0.72-0.88)	0.82 (95% CI 0.73-0.90)	0.84 (95% CI 0.76-0.91)
AJCC stage	Stage IA, IB, IIA, IIB, IIC	0.79 (95% CI 0.75-0.82)	0.83 (95% CI 0.79-0.86)	0.80 (95% CI 0.70-0.89)	0.81 (95% CI 0.72-0.89)
Matching variables	Age*, sex, Breslow thickness*, and ulceration	0.79 (95% CI 0.74-0.83)	0.83 (95% CI 0.78-0.86)	0.81 (95% CI 0.72-0.89)	0.83 (95% CI 0.75-0.90)
GEP	558 discovery genes all tumors	0.79 (95% CI 0.74-0.88)	0.82 (95% CI 0.77-0.89)	0.77 (95% CI 0.66-0.86)	0.79 (95% CI 0.69-0.88)
	558 discovery genes in stage I	0.72 (95% CI 0.54-0.81)	0.72 (95% CI 0.54-0.82)	0.69 (95% CI 0.52-0.84)	0.64 (95% CI 0.46-0.80)

	558 discovery genes in stage II	0.47 (95% CI 0.20-0.75)	0.54 (95% CI 0.27-0.77)	0.46 (95% CI 0.25-0.70)	0.48 (95% CI 0.26-0.71)
Clinicopathological + GEP	Breslow thickness*, ulceration, 558 genes from discovery	0.80 (95% CI 0.74-0.88)	0.83 (95% CI 0.77-0.89)	0.79 (95% CI 0.68-0.87)	0.81 (95% CI 0.71-0.89)
High selection frequency GEP	11 genes	0.71 (95% CI 0.65-0.77)	0.75 (95% CI 0.70-0.81)	0.72 (95% CI 0.61-0.82)	0.75 (95% CI 0.64-0.84)

474