Published in final edited form as:
Cancer Res. 2010 January 15; 70(2): 552–559. doi:10.1158/0008-5472.CAN-09-2653.

Personalized Prediction of Germline *CDKN2A* Mutations and Cancer Risk in Hereditary Melanoma

Wenyi Wang[1,2], **Kristin B. Niendorf**[3], **Devanshi Patel**[4], **Amanda Blackford**[5], **Fabio Marroni**[6], **Arthur J. Sober**[7], **Giovanni Parmigiani**[8,9], and **Hensin Tsao**[4,7,10]

[1]Stanford Genome Technology Center, Department of Biochemistry, Stanford University, Stanford, CA, USA 94304

[2]Department of Statistics, University of California at Berkeley, Berkeley, CA, USA 94720

[3]University of Minnesota, Minneapolis, Minnesota 55455

[4]Center for Cancer Risk Assessment, Massachusetts General Hospital, Boston, MA, USA 02114

[5]Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD, USA 21205

[6]Institute of Applied Genomics, Udine, Italy 33100

[7]Department of Dermatology, Massachusetts General Hospital, Boston, MA, USA 02114

[8]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston MA, 02115

[9]Department of Biostatistics, Harvard School of Public Health, Boston MA 02115

[10]Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, MA, USA 02114

Abstract

Personalized cancer risk assessment remains an essential imperative in post-genomic cancer medicine. In hereditary melanoma, germline *CDKN2A* mutations have been reproducibly identified in melanoma-prone kindreds worldwide. However, genetic risk counseling for hereditary melanoma remains clinically challenging. To address this challenge, we developed and validated MelaPRO: an algorithm that provides germline *CDKN2A* mutation probabilities and melanoma risk to individuals from melanoma-prone families. MelaPRO builds upon comprehensive genetic information, and uses Mendelian modeling to provide fine resolution and high accuracy. In an independent validation on 195 individuals from 167 families, MelaPRO exhibited good discrimination with a concordance index (C) of 0.86 (95% CI: 0.75–0.97) and good calibration, with no significant difference between observed and predicted carriers (26; 95% CI: 20–35, as compared to 22 observed). In cross-validation, MelaPRO outperformed the existing predictive model MELPREDICT (C: 0.82; 95% CI: 0.61–0.93), with a difference of 0.05 (95% CI: 0.007 to 0.17). MelaPRO is a clinically accessible tool that can effectively provide personalized risk counseling for the hereditary melanoma family members.

Keywords

Risk prediction; Mendelian models; Cancer genetics

Corresponding author: Hensin Tsao, MD PhD. Department of Dermatology, Massachusetts General Hospital, 622 Bartlett Hall, 48 Blossom Street, Boston, MA, 02114. Tel: 617-726-9569. Fax: 617-724-2745. htsao@partners.org.

There is no potential conflict of interest with any author. A preliminary version of this work was presented at the 2006 International Genetic Epidemiology Society Meeting.

INTRODUCTION

In 2009, there will be an estimated 68,720 new cases of melanoma with 8,650 deaths (1). Despite decades of therapeutic investigation, metastatic melanoma is still considered incurable, thereby making identification of high-risk individuals with an eye towards early detection a cornerstone in the strategy for cure.

A fundamental goal of personalized medicine is to uncover germline variants that identify individuals at the greatest risk for disease. For melanoma, the first such mutations were found in *CDKN2A* over a decade ago (2). Since then, heritable alterations in *CDKN2A* (encoding two proteins: p16/Ink4a and p14/ARF) and *CDK4*, the inhibitory target of p16/Ink4a, have also been found in a significant subset of melanoma-prone families (2–8). Earlier validation of a computational tool – MELPREDICT, for estimating *CDKN2A* carrier probability, showed reasonable performance in ranking carriers higher than noncarriers among melanoma patients (9). However, MELPREDICT is based on logistic regression models and therefore cannot effectively incorporate crucial biological information embedded within the pedigree structure. Moreover, it lacks the flexibility to account for variations in *CDKN2A* mutation prevalence and penetrance across geographical regions (3,4).

To this end, we developed MelaPRO, a new model to estimate the probability of carrying a mutation in *CDKN2A* in melanoma families, using a general Mendelian risk prediction approach (10) that integrates Mendelian inheritance and Bayesian probability theories. This computational strategy effectively translates genetic information into a clinically useful algorithm for carrier probability estimation and has been successfully applied to develop BRCAPRO (11–14) for the breast and ovarian cancer syndrome, MMRpro (15) for the Lynch syndrome and PancPRO (16) for familial pancreatic cancer. In this initial validation, we show that MelaPRO exhibits strong discrimination and calibration ability, and outperforms the regression model, MELPREDICT.

METHODS

Model Development

MelaPRO translates population estimates of the mutation prevalence and penetrance of *CDKN2A* into mutation prediction for any designated family member (the counselee), given his or her family history and assuming autosomal dominant inheritance. The penetrance refers to the age-specific risk of developing cutaneous melanoma depending on *CDKN2A* carrier status and gender.

The carrier probability is modeled via Bayes' rule as follows (10):

$$\Pr_{\text{genotype}|\text{history}} = \Pr_{\text{genotype}} X \Pr_{\text{history}|\text{genotype}} / \Pr_{\text{history}},$$

Here, Pr denotes probability, genotype denotes whether the counselee carries a deleterious mutation in *CDKN2A*, and history denotes family history (as detailed in Table 1). The $\Pr_{\text{genotype}}$ term is the mutation prevalence; the $\Pr_{\text{history}|\text{genotype}}$ is a weighted average of the probabilities of family history given each possible genotype configuration of all relatives, where the weights are the probabilities of the genotype configuration based on Mendelian transmission. This step uses the Elston-Stewart algorithm (17), as implemented in the latest version of the R package BayesMendel[1]. The probability of family history given each genotype configuration can be broken down into the product of each relative's probability of phenotype

[1]http://astor.som.jhmi.edu/BayesMendel

given genotype, assuming conditional independence. Here, each probability term is calculated as either the cumulative penetrance (age specific) for affected relatives (MPM or SPM) or 1-cumulative penetrance for unaffected relatives. The $Pr_{history}$ is the sum of terms like $Pr_{genotype}$ X $Pr_{history|genotype}$ across all possible genotypes of the counselee. Risks of developing SPM and MPM for unaffected individuals are estimated by a weighted average of the carrier's and noncarrier's penetrance, where the weights are the carrier probabilities.

Parameter settings

MelaPRO incorporates three distinct penetrance estimates. The GenoMEL (3) consortium collected high-risk families (>2 affected family members) and estimated separate penetrances for areas with high baseline incidence (HBI) and low (LBI) up to age 80 using logistic regression. Alternatively, the GEM (4) Study Group collected melanoma patients from the general population and estimated the penetrance in 5-year age intervals using the nonparametric kin-cohort method. We extrapolated the GenoMEL data and interpolated the GEM data using estimates from the SEER DevCan Software[2] as a reference, to establish the age-specific penetrance between ages 1 and 110. We then calculated the mutation prevalence indirectly from the penetrance estimates. By Bayes' rule, we have Pr(G)=Pr(G|B)XPr(B)/Pr(B|G), where G denotes being a *CDKN2A* carrier, and B denotes new cases per year between 2001 and 2005. From previous studies, we obtained Pr(G|B) = 0.0179 (4) and Pr(B) = 19.4/100000‡ for the North American population. Pr(B|G) is a weighted average of probability distribution function of penetrance for *CDKN2A* carrier, where the weights are melanoma incidence within 10-year age interval[3].

Accounting for multiple primary melanomas (MPMs) versus single primary melanoma (SPM)

We used X to indicate number of primary melanomas and G to indicate carrier status, with X=1 for SPM and X≥2 for MPM. The published penetrance estimates are P_0 = Pr(X≥1|G = 0) and P_1 = Pr(X≥1|G = 1). The relative risk of MPM for carriers and noncarriers among melanoma cases is Pr(X≥2|X≥1,G=1)/Pr(X≥2|X≥1,G=0)=1.8 (18), and the risk ratio of having MPM versus SPM for carriers is Pr(X≥2|G=1)/Pr(X=1|G=1) = 1.14 (by age 50, ref. 4). Based on these numbers we estimated the MPM and SPM specific penetrances.

Test sensitivity and specificity—For the genetic results, the default specificity was set at 1.0, because only known mutations were included in the analysis; putative polymorphisms (e.g. Ala148Thr) and variants of unknown significance (VUS) were excluded since accurate penetrance data are not available for these alterations. As such, the model does not currently calculate the probability of detecting established polymorphisms or variants of unknown significance (VUS). Other possibilities for a false positive, such as sample confusion, can be considered negligible. Since our mutational screen does not detect deep intronic mutations and large chromosomal deletions, we set our sensitivity at 0.9 presuming that these types of deleterious changes occur in no more than 10 percent of the cases. It is straightforward for users to replace these estimates with different ones.

MELPREDICT (9)—The MELPREDICT model is a multiple logistic regression, in which the estimated carrier probability of the counselee being a mutation carrier is given by $e^L/(1+e^L)$. $L = \beta0 + \beta1 \times$ (no. of counselee primaries) $+ \beta2 \times$ (no. of additional family primaries) $- \beta3 \times$ (ln(counselee age)).

[2]DevCan: Probability of Developing or Dying of Cancer Software, Version 6.1.1. Statistical Research and Applications Branch, National Cancer Institute, 2005. URL http://srab.cancer.gov/devcan.

[3]Res LAG, Melbert D, Krapcho M, et al. (eds). SEER Cancer Statistics Review, 1975–2005, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site, 2008

Validation

Study Population—We used data from the Massachusetts General Hospital Melanoma and Pigmented Lesion Center (PLC). This series was not used in the development of MelaPRO and provides an independent validation. This study was performed in accordance with a protocol approved by the MGH Institutional Review Board. From April 2001 to January 2008, all patients with invasive or in-situ melanoma evaluated at the PLC were screened for eligibility as follows: (1) ≥1 first-degree relatives with melanoma, or (2) ≥2 affected relatives with melanoma on one side of the family (first- or second-degree), or (3) ≥3 primary cutaneous melanomas irrespective of family history. The presence and number of melanomas for counselees were confirmed via pathology reports, except for a small number of cases (<10%, data not shown). Medical record confirmation of reported family histories was pursued but limited to relatives who provided prior consent to participate in the study. We excluded two families: one because it lacked counselee information and the other because the counselee is unaffected and therefore ineligible for comparison with MELPREDICT.

Mutation analysis—*CDKN2A* exons 1α, 1β and 2 were screened for sequence variants as previously described (9).

Data Analysis—All analyses were performed in R[4]. Within each family, we assigned each *CDKN2A*-tested individual in turn as the counselee and calculated the probability of detecting a *CDKN2A* mutation using MELPREDICT (9) and all modules of MelaPRO. For the MelaPRO modules, this probability is obtained by multiplying the probability of carrying a *CDKN2A* mutation, provided by the model, by the sensitivity of the mutation analysis (default=0.9). The comparison between models required additional exclusion of four cases in which tested individuals were unaffected, as MELPREDICT does not apply. We evaluated the discrimination, calibration and accuracy performance of each model by comparing the calculated probabilities with the observed mutational status. Discrimination reflects a model's ability to differentiate individuals with positive outcomes from those with negative outcomes. It can be visualized using Receiver Operating Characteristic (ROC) curves and summarized by the underlying area, or concordance index (C). Calibration is a model's ability to make unbiased estimates of the proportion of carriers. We also used positive predictive value (PPV) and negative predictive value (NPV) to measure accuracy, and mean squared error (MSE) for an overall comparison of performance. MELPREDICT was developed based on a subset of our validation set. Therefore, we used cross-validation (leave-one-out) to obtain evaluation statistics for MELPREDICT. In our cross-validation, we fixed the covariates selected by the original MELPREDICT model, but re-estimated the coefficients in each training set. We obtained 95% confidence intervals using the bootstrap (19). We also evaluated sensitivity and specificity for the descriptive classifier (FH) defined by having at least 2 affected relatives. We present hypothetical but realistic family history scenarios for illustration.

RESULTS

MelaPRO Features

The MelaPRO model treats melanoma family history as a diagnostic test or profile, and *CDKN2A* genotype as an occult condition to be diagnosed. To use MelaPRO during a typical counseling session, the counselor collects the counselee's family history information, and enters it into MelaPRO to obtain a carrier probability and an estimate of future risk if the counselee is still free of the disease. The family history information is detailed in Table 1, and it includes family members' relationship, occurrence of cutaneous melanoma (including

[4]R Development Core Team: A language and environment for statistical computing. R Development Core, Vienna, Austria. 2006 URL http://www.R-project.org

whether single or multiple primaries were found), age of diagnosis, or age at last contact for unaffected family members, and earlier germline testing results of any family members, if available. There is no restriction to which family member can be designated the counselee and no limit to the size of the family tree that can be processed, as long as there is no inbreeding. Predictions can be obtained using any subset of the information in Table 1.

Figure S1 shows the penetrance estimates from GenoMEL (3) and GEM (4), which applied to all melanoma diagnoses combined. We estimated the allele frequency (mutation prevalence) as 0.00015 using the HBI penetrance; 0.0003 using the LBI penetrance; and 0.0004 using the GEM penetrance. Additionally, MelaPRO incorporated an estimate that 53% of diagnoses in carriers are MPM compared to 30% for noncarriers.

MelaPRO provides three modules: MelaPRO-HBI (HBI), MelaPRO-LBI (LBI) and MelaPRO-GEM (GEM) reflecting different penetrances, now adjusted to be MPM/SPM-specific. Users choose the module that best matches the population where the model is used to the characteristics of the original studies. To illustrate, in Figure 1, for scenario 1 (Table 2), MelaPRO gave a probability estimate of 0.43-HBI, 0.90-LBI and 0.85-GEM. For comparison, the probabilities without the MPM/SPM adjustment are: 0.27-HBI, 0.83-LBI and 0.74-GEM.

Users can also specify the sensitivity (default=0.9) and specificity (default=1) of the germline testing method when results are available for some family members.

Software

MelaPRO is open source and freely available as part of the BayesMendel (10) risk prediction package at http://astor.som.jhmi.edu/BayesMendel/ and the Cancer-Gene (20) counseling package at http://www4.utsouthwestern.edu/breasthealth/cagene/.

Clinical Illustration

Figure 1 illustrate how MelaPRO provides high-resolution information to support clinical counseling, by presenting carrier probability estimates for several hypothetical, but realistic, scenarios. We compared our results to the descriptive classifier FH, and to MELPREDICT (9), a logistic regression model based on number of primary melanomas in the counselee, in all other family members and the counselee's age at diagnosis.

In the pedigree, MELPREDICT estimated a carrier probability of 0.24 as compared to the MelaPRO's estimate of 0.43 (HBI, see other modules in Table 2). MelaPRO captured the two relatives' earlier disease onset (59 years in general population[5]) as additional indication of carrier status. It also responded, with considerable increase in probabilities, to modification of the father's disease history, while the total number of familial melanomas remained the same (HBI: 0.43 to 0.77 and 0.43 to 0.82). The additional scenarios demonstrate how carrier probabilities varied as the number of affected individuals and patients' relationship to counselee were changed (i.e. aunt healthy or brother affected).

External Validation

We assembled a validation set containing 167 families with an average of 29 members. There were, in total, 26 carriers, 22 of which were affected with melanoma, and 603 primary melanomas. The mean number of primary melanomas in families of carriers and noncarriers was 7.9 (95%CI: 5.5–10.3) and 3.2 (95%CI: 2.9–3.5), respectively. There were 207 genotyped individuals within the 167 families. Among these, 195 were cases, with 85 males and 110

[5]Res LAG, Melbert D, Krapcho M, et al. (eds). SEER Cancer Statistics Review, 1975–2005, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2005/, based on November 2007 SEER data submission, posted to the SEER web site, 2008.

females. The mean age at diagnosis was 46.4 years (95% CI: 43.5–49.3) for males and 41.3 years (95% CI: 38.6–43.9) for females, and it was 36.6 years (95% CI: 30.7–42.6) for affected carriers and 44.4 years (95% CI: 42.3–46.4) for affected noncarriers. The proportion of mutation carriers increased with the number of primary melanomas in the counselee, the number of affected relatives, and the number of primary melanomas in relatives (Table S1). There were a total of eight relatives from seven families affected with pancreatic cancer.

The Boston validation cohort is derived from a relatively high incidence region (3) and is familial in ascertainment. We deployed MelaPRO-HBI and predicted the presence of approximately 26 mutations (95%CI: 20, 35); the MelaPRO-GEM module predicted 41 mutations (95%CI: 31, 58), and MELPREDICT predicted 20 mutations (95%CI: 19, 24, see Observed/Expected (O/E) ratios in Table 3). Both MelaPRO-HBI and MELPREDICT showed a close correspondence with the observed 22 mutations. MelaPRO-GEM and MelaPRO-LBI predicted a substantially higher number of mutations than was observed, likely because their parameter estimates do not fit our cohort profile.

MelaPRO shows good discriminatory ability with all three modules. Figure 2 shows the ROC curves for MelaPRO and MELPREDICT, as well as the sensitivity and specificity based on the summary family history criterion FH. The corresponding AUCs are presented in Table 3. The difference between the AUC for MelaPRO-HBI and that for MELPREDICT is 0.05 (95% CI: 0.007 to 0.17). Part of this difference is attributable to the gap visible at the top right of Figure 2: MelaPRO achieved an estimated sensitivity of 90% at the cost of about 70% false positives, while MELPREDICT provided limited discrimination at this level of sensitivity. The point corresponding to the sensitivity and specificity based on FH lay below the ROC curves, with an 81% sensitivity at the cost of a >40% false positive rate, while model-based prediction achieved higher sensitivity with 10% fewer false positives.

We also investigated the accuracy of MelaPRO and MELPREDICT predictions associated with a carrier probability cutoff of 50%. The positive predictive value (PPV) was 0.70, 0.57 and 0.44 for MelaPRO-HBI, MelaPRO-GEM and MELPREDICT, respectively. The negative predictive values (NPVs) were 0.97, 0.97 and 0.90 for the same three models. The mean squared error of prediction, which evaluates the overall performance of the algorithm, was significantly better in the MelaPRO-HBI (0.06, 95%CI: 0.03, 0.08) and MelaPRO-GEM (0.08, 95%CI: 0.06, 0.11) modules than the MelaPRO-LBI (0.19, 95%CI: 0.15, 0.22) module, with the former two slightly better than MELPREDICT (0.09, 95%CI: 0.04, 0.12, see Table 3). We then considered how often MelaPRO led to a re-classification compared to MELPREDICT and FH. As shown in Table 4, the re-classification fraction ranged from 4% to 34%. MelaPRO-HBI re-classified correctly 5, 10 and 65 more individuals than MelaPRO-GEM, MELPREDICT and FH respectively. The 50% threshold was chosen for illustrative purposes only and is not based on any clinical recommendations.

DISCUSSION

One’s ability to create a personalized risk portfolio for patients with hereditary melanoma remains a formidable challenge. To this end, we have developed and successfully validated MelaPRO for individualized *CDKN2A* carrier estimation. This open-source tool delivers a useful and easily deployable instrument for cancer risk counselors who wish to frame a more informative discussion for individuals pursuing *CDKN2A* genetic testing. Our results indicate that MelaPRO provides high resolution and accurate risk assessment, discriminating between individuals with or without germline mutations in *CDKN2A*.

An ideal personalized risk model would rely on a menu of modules that best fit the clinical profile. Since geographical location and other unknown genetic factors which may co-

segregate with melanoma families appear to influence both penetrance and prevalence of *CDKN2A* mutations (4), we constructed three distinct MelaPRO modules based on separate penetrance estimates: GenoMEL-HBI, GenoMEL-LBI and GEM. This is a first step towards accounting for both genetic and environmental factors. We also derived the corresponding mutation prevalence of *CDKN2A* based on penetrance. As more data emerges, these estimates can be easily updated, so that the model will continue to operate using the best available information. For example, using allele frequency that is specific to Europe might further improve the performance of GenoMEL-LBI in this population. The prevalence estimate (0.00015) for MelaPRO-HBI matches that from Bishop et al. (3). Sensitivity analysis with variation in prevalence (between 0.00015 and 0.0004) showed similar discrimination performance in all modules, and higher O/E ratios with lower prevalence values for GEM and LBI. In the Boston validation set, the GenoMEL-HBI presented significantly better performance than others (Table 3), suggesting the utility of the geography and ascertainment specific modules.

MelaPRO treats individuals with SPM and MPMs differently, providing higher resolution and more accurate *CDKN2A* mutation risk. With the Boston validation data, the MelaPRO-HBI model without MPM adjustment gave higher probabilities to SPM families, where the counselees are often noncarriers. Overall, it gave a lower O/E ratio: 0.73, and a slightly lower C-index: 0.84. Our assumption of constant MPM/SPM risk ratios across ages can be modified as more data becomes available. Similarly, the current MelaPRO provides the basis for more refined models incorporating polygenic effects, risk modifiers and biomarkers as their role becomes clarified. Future iterations of MelaPRO will also incorporate two known risk factors: MC1R status and history of pancreatic cancer in the family.

MelaPRO captures the full pedigree data, including information on affected and unaffected family members, and is therefore able to further discriminate between individuals at higher and lower risk with the same number of affected family members. Part of the Boston validation set was used to develop MELPREDICT, specifically for choosing the covariates in the final model. Although the cross-validation should correct for part of the optimism that is associated with internal validations, it does not account for variability across studies. Therefore the gap between MelaPRO's and MELPREDICT's performances would likely be wider in a new independent set of families. The improvement of 0.05 in the concordance index C corresponds to real advances clinically at a personal level, as evidenced by the PPV/NPV and reclassification results. Lastly, MELPREDICT is not applicable to unaffected individuals in melanoma-prone families. MelaPRO is more powerful as a clinical instrument because it is applicable to the entire family and calculates pre-disease estimates of carrier probability and melanoma risk.

In the current study, we built and evaluated MelaPRO on known deleterious mutations while excluding known polymorphisms (e.g. Ala148Thr). However, MelaPRO quantifies degree of genetic segregation in melanoma families and may give high probabilities to carriers of variants of unknown significance (VUS) that have similar effects to the known variants in *CDKN2A*. Going forward, MelaPRO can further accommodate errors in classification of variants as deleterious or polymorphic, by changing the sensitivity and specificity accordingly. In broader terms, what is critically needed is a robust biochemical or genetic assay for p16/Ink4a and p14/ARF functionality, which will fundamentally improve the accuracy of risk predictions.

From the clinical perspective, MelaPRO can be easily incorporated into any genetic counseling session. Most melanoma clinicians appreciate the importance of family history in a qualitative, but not necessarily quantitative sense. However, in clinics, individuals are frequently referred for genetic counseling without regard for pedigree structure or counselee affection status - all features that can be captured for better estimation through MelaPRO. The current consensus

is that melanoma patients who have either an affected first-degree relative or more than one affected relative on one side of the family, or unaffected individuals with two or more cases of melanoma in close relatives may benefit from a genetic risk assessment. In some situations, MPM patients without family history may also consider counseling. In contrast, unaffected relatives from single-case kindreds who present to their physicians for routine mole checks comprise the largest group of at-risk individuals with a "family history" of melanoma; but, these individuals in general do not need genetic consultation since the likelihood of harboring a germline *CDKN2A* mutation is likely to be close to 1% (18). Likewise, a melanoma patient with a single, distant family history of melanoma, especially if not substantiated by a medical record, would not routinely need genetic risk counseling unless special circumstances exist. Beyond the valuable exercise of counseling, the decision to undergo *CDKN2A* germline testing should be made in conjunction with a trained professional who can integrate the genetic, psychological and social implications of genetic testing.

Our ability to assess calibration and discrimination is limited by the 22 *CDKN2A* mutation carriers found in the validation set. We excluded four carriers, as they were ineligible for MELPREDICT analysis. We also evaluated the performance of MelaPRO alone, using all carriers and obtained similar results. Although suboptimal for analysis, the 11% mutation rate among all individuals with a family history of melanoma is probably appropriate for general clinical use (3,4,21). In addition, since most counselees in the validation set were melanoma patients, we could not properly evaluate the model on unaffected individuals. Finally, MelaPRO does not explicitly account for germline *CDK4* variants. However, since *CDK4* mutations are thought to be exclusive to *CDKN2A* mutations and since extant data suggest that the *CDK4*-mutation phenotype is identical to the *CDKN2A*-mutation phenotype (9), MelaPRO assumed *CDKN2A* and *CDK4* as a single genetic unit without resorting to a two-locus model. There were no *CDK4* kindreds among our families.

In summary, we have developed and validated a risk prediction model, MelaPRO, whose central goal is to enhance melanoma risk counseling by providing accurate pre-test assessment of *CDKN2A* carrier probability. MelaPRO's architecture is of such flexibility that when data become available through ongoing and proposed gene/gene and gene/environment studies, such biological information can be readily assimilated into the model, achieving higher resolution and higher accuracy in risk assessment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by American Cancer Society grant RSG MGO-112970 (to H.T. and G.P.), NCI Grant No's P50 CA-93683 (to H.T.), R01CA105090-01A1 (to G.P.) and the generous philanthropic donors to the Massachusetts General Hospital.

REFERENCE

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics. CA Cancer J Clin 2009;59:225–249. [PubMed: 19474385]
2. Hussussian CJ, Struewing JP, Goldstein AM, et al. Germline p16 mutations in familial melanoma. Nat Genet 1994;8:15–21. [PubMed: 7987387]
3. Bishop DT, Demenais F, Goldstein AM, et al. Geographical variation in the penetrance of *CDKN2A* mutations for melanoma. J Natl Cancer Inst 2002;94:894–903. [PubMed: 12072543]
4. Begg CB, Orlow I, Hummer AJ, et al. Lifetime risk of melanoma in *CDKN2A* mutation carriers in a population-based sample. J Natl Cancer Inst 2005;97:1507–1515. [PubMed: 16234564]

5. Box NF, Duffy DL, Chen W, et al. MC1R genotype modifies risk of melanoma in families segregating CDKN2A mutations. Am J Hum Genet 2001;69:765–773. [PubMed: 11500805]
6. Easton DF, Cox GM, MacDonald AM, Ponder BA. Genetic susceptibility to naevi– a twin study. Br J Cancer 1991;64:1164–1167. [PubMed: 1764382]
7. Bartsch DK, Sina Frey M, Lang S, et al. CDKN2A germline mutations in familial pancreatic cancer. Ann Surg 2002;236:730–737. [PubMed: 12454511]
8. Lynch HT, Shaw TG, Lynch JF. Inherited predisposition to cancer: a historical overview. Am J Med Genet C Semin Med Genet 2004;129:5–22. [PubMed: 15264268]
9. Niendorf KB, Goggins W, Yang G, et al. MELPREDICT: a logistic regression model to estimate *CDKN2A* carrier probability. J Med Genet 2006;43:501–506. [PubMed: 16169933]
10. Chen S, Wang W, Broman KW, Katki HA, Parmigiani G. BayesMendel: an R environment for Mendelian risk prediction. Stat Appl Genet Mol Biol 2004;3 Article 21.
11. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. Am J Hum Genet 1998;62:145–158. [PubMed: 9443863]
12. Berry DA, Parmigiani G, Sanchez J, Schildkraut J, Winer E. Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. J Natl Cancer Inst 1997;89:227–238. [PubMed: 9017003]
13. Antoniou AC, Gayther SA, Stratton JF, Ponder BA, Easton DF. Risk models for familial ovarian and breast cancer. Genet Epidemiol 2000;18:173–190. [PubMed: 10642429]
14. Nanda R, Schumm LP, Cummings S, et al. Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. JAMA 2005;294:1925–1933. [PubMed: 16234499]
15. Chen S, Wang W, Lee S, et al. Prediction of germline mutations and cancer risk in the lynch syndrome. JAMA 2006;296:1479–1487. [PubMed: 17003396]
16. Wang W, Chen S, Brune KA, Hruban RH, Parmigiani G, Klein AP. PancPRO: risk assessment for individuals with a family history of pancreatic cancer. J Clin Oncol 2007;25:1417–1422. [PubMed: 17416862]
17. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. Hum Hered 1971;21:523–542. [PubMed: 5149961]
18. Berwick M, Orlow I, Hummer AJ, et al. The prevalence of *CDKN2A* germ-line mutations and relative risk for cutaneous malignant melanoma: an international population-based study. Cancer Epidemiol Biomarkers Prev 2006;15:1520–1525. [PubMed: 16896043]
19. Efron B. The bootstrap and modern statistics. J Am Stat Assoc 2000;95:1293–1296.
20. Euhus D. Risk modeling in breast cancer. Breast J 2004;10 Suppl 1:S10–S12. [PubMed: 14984482]
21. Goldstein AM, Struewing JP, Fraser MC, et al. Prospective risk of cancer in CDKN2A germline mutation carriers. J Med Genet 2004;41:421–424. [PubMed: 15173226]

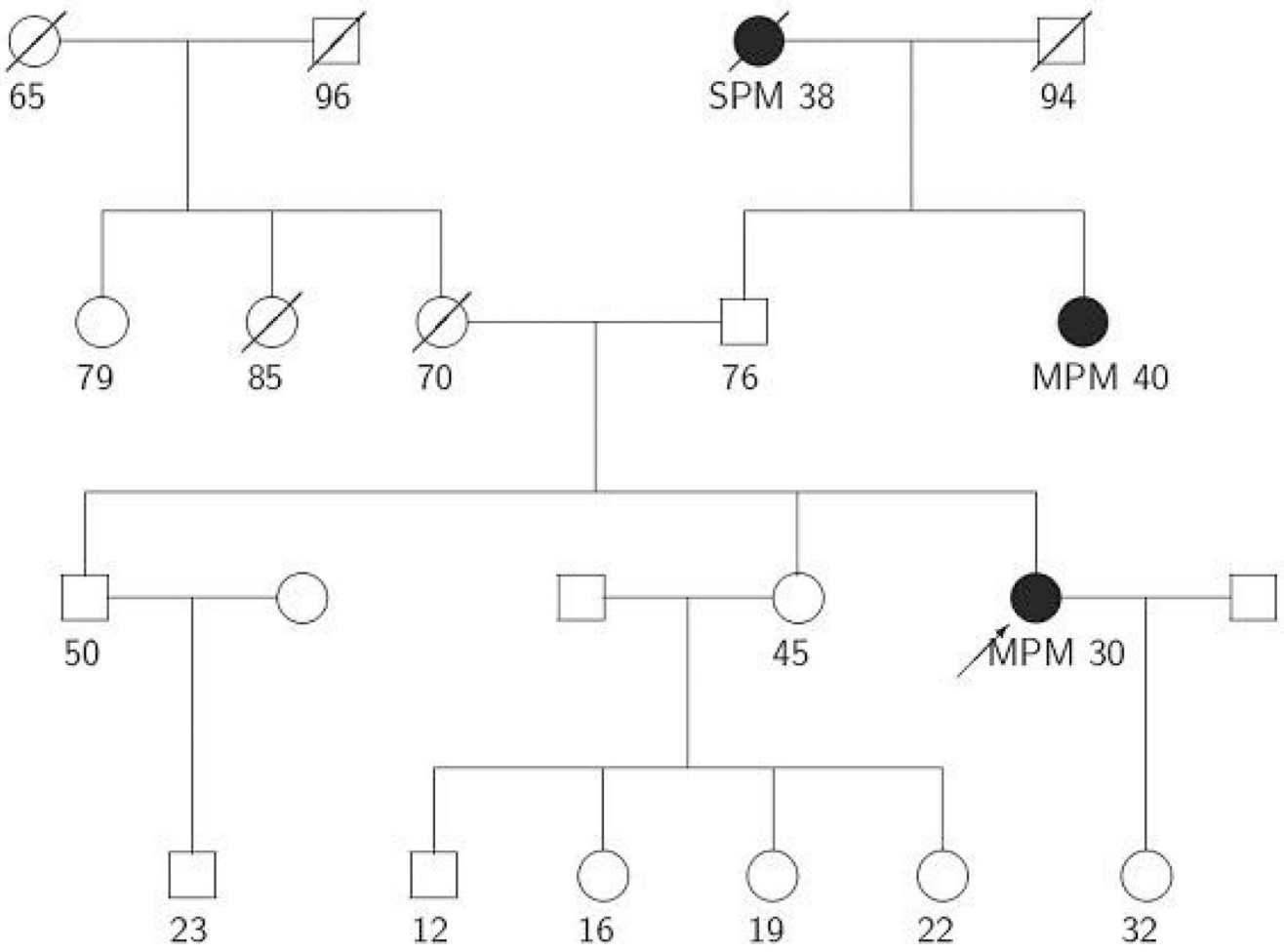

Figure 1.
A hypothetical family pedigree for comparison of MelaPRO carrier probability estimates to those provided by MELPREDICT and FH. The arrow points to the counselee, for whom the carrier probability is calculated given her family history. The probability estimates are presented in Table 2.

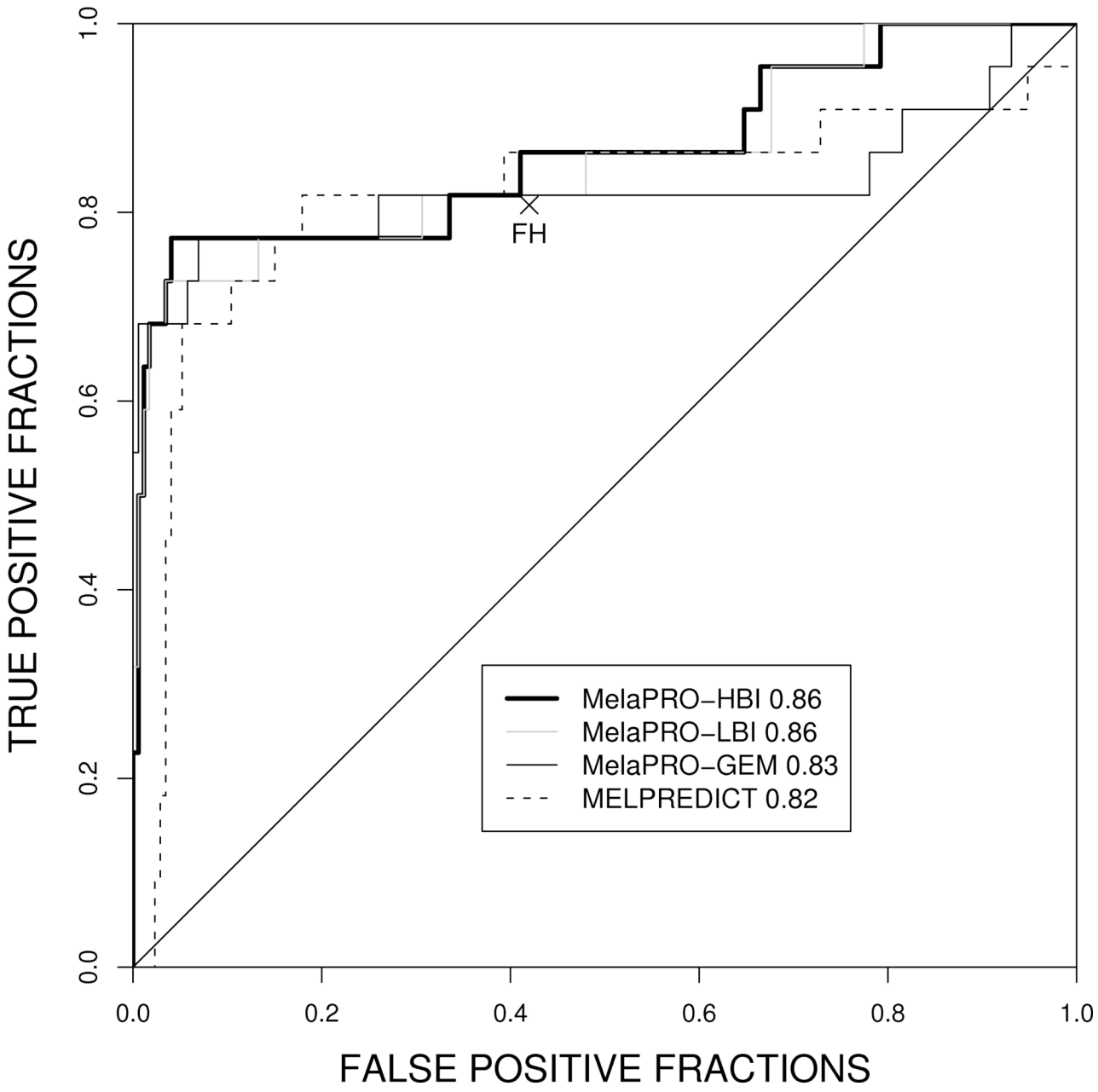

Figure 2.
Receiver operating characteristic (ROC) curves for MelaPRO-HBI, MelaPRO-LBI, MelaPRO-GEM and MELPREDICT (using leave-one-out cross-validation) on the Boston validation set. Also shown are the true positive and false positive fractions associated with the summary family history criterion: FH.

Table 1

Family history as input for MelaPRO and resulting output [*], [†]

	Features
Input (for the counselee and each relative)	Gender
	Parent information (IDs)
	Melanoma status, 0 if unaffected, 1 if single primary melanoma, 2 if multiple primary melanomas
	Age at first diagnosis (in years) of melanoma if affected
	Current age or age at last follow-up (in years) if unaffected
	Result of previous germline testing of *CDKN2A* in any family member, 0 if missing, 1 if positive, 2 if not found
Output (for the counselee)	Probability that the counselee carries a deleterious mutation of *CDKN2A*
	Probability that the counselee, if asymptomatic, will develop single or multiple primary melanomas in the future, in yearly intervals

[*] Any input can be N/A if not available, with the exception of the parental IDs that relate each family member to the counselee.

[†] Personalized prediction can be made for each family member by changing the designation of the counselee.

Table 2

Comparison of MelaPRO carrier probability estimates to those provided by MELPREDICT and FH. The resulting probabilities (for a counselee indicated by the arrow in Figure 1) correspond to five different variations of the pedigree in Figure 1, with each variation corresponding to a row in the table. The FH column is yes when the counselee has at least 2 affected relatives.

Scenario	MelaPRO			MELPREDICT	FH
	GenoMEL-HBI	GenoMEL-LBI	GEM		
1. As shown	0.43	0.90	0.85	0.24	Yes
2. As shown but the paternal aunt is healthy at 40	0.02	0.24	0.18	0.07	No
3. As shown but the father's status is unknown	0.77	0.93	0.87	0.24	Yes
4. As shown but the paternal aunt is healthy at 76, the father is MPM at 40	0.82	0.97	0.94	0.24	Yes
5. As shown but the brother is also SPM at 50	0.91	0.98	0.96	0.40	Yes

Table 3

Summary of validation results.

	Concordance index C*	Observed/Expected O/E ratio†	Mean Squared Error
MelaPRO-HBI	0.86 (0.75, 0.97)	0.85 (0.62, 1.08)	0.06 (0.03, 0.08)
MelaPRO-LBI	0.86 (0.74, 0.97)	0.31 (0.20, 0.42)	0.19 (0.15, 0.22)
MelaPRO-GEM	0.83 (0.67, 0.98)	0.54 (0.38, 0.72)	0.08 (0.06, 0.11)
MELPREDICT‡	0.82 (0.61, 0.93)	1.10 (0.92, 1.17)	0.09 (0.04, 0.12)
Difference Between MelaPRO-HBI and MelaPRO-LBI	0.007 (−0.01,0.02)	Not Applicable§	−0.13 (−0.16, −0.10)
Difference Between MelaPRO-HBI and MelaPRO-GEM	0.04 (−0.008, 0.09)	Not Applicable	−0.03 (−0.04, −0.02)
Difference Between MelaPRO-HBI and MELPREDICT	0.05 (0.007, 0.17)	Not Applicable	−0.03 (−0.06, 0.002)
	Sensitivity	Specificity	
FH:≥2 affected relatives	0.81	0.58	

*The concordance index C is equal to the AUC (area under the ROC curve)

†The ratio between the observed number of carriers and the total number of predicted carriers

‡Leave-one-out cross-validation on Boston data using fixed covariates for MELPREDICT

§Computing the difference in this case is not appropriate, as each ratio should be compared to the reference value of one.

Table 4

Comparison of MelaPRO probabilities with MELPREDICT probabilities and FH by reclassification rate. *

	Carriers		Noncarriers		
	MelaPRO-HBI	**MelaPRO-HBI**	**MelaPRO-HBI**	**MelaPRO-HBI**	**Reclassication**
	Pr ≥.5	**Pr < .5**	**Pr ≥ .5**	**Pr < .5**	**Rate****
MelaPRO-GEM					
≥.5	16	1‡	7	6†	
<.5	0†	5	0‡	160	4%
MELPREDICT					
≥.5	4	0‡	3	2†	
< .5	12†	6	4‡	164	9%
FH					
+	16	1‡	7	66†	
−	0†	5	0‡	100	34%

*MelaPRO-LBI is not included, as the study population does not apply.

†Denotes a category where MelaPRO-HBI correctly classifies as compared to other methods.

‡The grey shading denotes a category where MelaPRO-HBI classifies differently as compared to other methods.

**Reclassification rate is defined as the percentage of individuals that were classified differently by MelaPRO-HBI (categories † and ‡) as compared to the other models