

Validating Orchid's Melanoma Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Melanoma is a cancer that develops in melanocytes, the cells that produce melanin, the pigment that gives skin its color. It is more likely than other types of skin cancer to invade nearby tissue and spread to other parts of the body.¹ According to the National Cancer Institute, approximately 2.2% of people will be diagnosed with melanoma during their lifetime.²

Risk factors include fair skin that burns easily, exposure to natural or artificial sunlight, a history of blistering sunburns, and having many moles.¹ Early detection is important because most melanomas are diagnosed while still localized, and five-year relative survival is approximately 100% for localized disease but 34% for disease that has spread to distant sites.²

Genetic Risk Score

Melanoma is shaped by both environmental and genetic factors. Some rare monogenic variants such as pathogenic variants in *CDKN2A* or *CDK4* are known to substantially increase the risk of melanoma, but most cases arise from the combined effects of many genetic variants and environmental exposures.^{3,4} A minority of individuals carry a pathogenic variant in one of these genes. For these individuals, the single pathogenic variant is the most important factor in determining their risk of melanoma. For the large majority of individuals who do not carry one of these rare pathogenic variants, the risk of melanoma can be modeled by building a genetic risk score (GRS) which combines the small effects of many variants into a single score. Although not diagnostic, a GRS can indicate how likely an individual is to experience this disease compared to the population baseline risk.

The melanoma GRS was developed using data from five sources: the cutaneous melanoma case-control cohort of Landi et al. 2020, for which we computed summary statistics from the individual-level genotype data deposited in dbGaP under accession phs001868 (rather than using the published Landi et al. summary statistics, which include UKB participants),⁴ the FinnGen study,⁵ the Million Veteran Program,⁶ the All of Us Research Program, for which we generated GWAS summary statistics,⁷ and the UK Biobank (UKB).⁸ For the UKB, a subset of individuals was held out to use in the validation study, and the training portion was restricted to individuals of European ancestry. Summary statistics were generated for the UKB using the Regenie method.⁹ Summary statistics for all five studies were then meta-analyzed using METAL.¹⁰ The final GRS was trained using SBayesRC and includes over 7 million variants.¹¹

Orchid considers a GRS meaningfully predictive if individuals at approximately the 97.7th percentile have an odds ratio (OR) of at least 2. The melanoma GRS meets this criterion for individuals of European ancestry. Because the UKB contains too few melanoma cases of non-European ancestry to validate the GRS in other ancestry groups, the melanoma GRS is currently offered only to individuals of European ancestry.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's melanoma GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.⁸ Because some data from the UKB was used to train the GRS, our validation study included only the set of samples that were held out for testing and not included in the GRS training. We restricted the analysis to individuals of British ancestry and defined melanoma cases as any record of ICD-10 code C43 (malignant melanoma of skin) in the UKB cancer registry (field 40006). Individuals whose only record of melanoma was a self-report (code 1059 in field 20001) were excluded from the evaluation set rather than counted as controls. This yielded 790 cases and 80,269 controls (1.0% prevalence). We then grouped individuals by GRS percentile and compared the observed melanoma prevalence within each group to our model's predictions (Figure 1). For additional technical details, see the Supplementary Information.

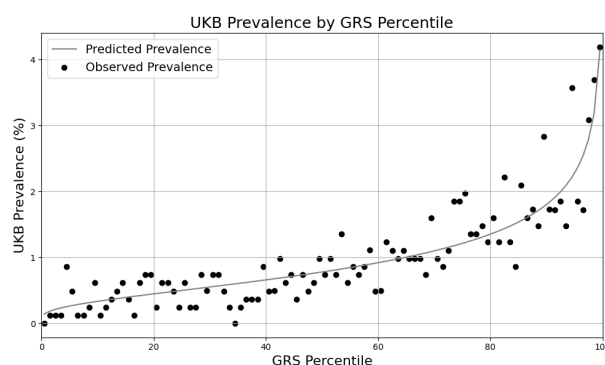


Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

Table 1 shows the melanoma observed prevalence for individuals in the UKB grouped by GRS percentile range (top 10%, 5%, and 1%), as well as how their risk compares to the base-

line risk at the 50th GRS percentile. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of melanoma, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	0.86%	1.00
Top 10%	2.49%	2.94
Top 5%	2.91%	3.44
Top 1%	4.19%	5.03

Table 1. Observed prevalence of melanoma in the UKB by GRS percentile range. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of melanoma.

Estimating Lifetime Risk

According to the National Cancer Institute, approximately 2.2% of people will be diagnosed with melanoma during their lifetime,² higher than the observed prevalence of 1.0% in the UKB testing set. We adjust our model so that its average prevalence aligns with the National Cancer Institute estimate (Figure 2).¹² People at the high end of the GRS distribution are predicted to have an elevated risk of melanoma relative to the population (Table 2).

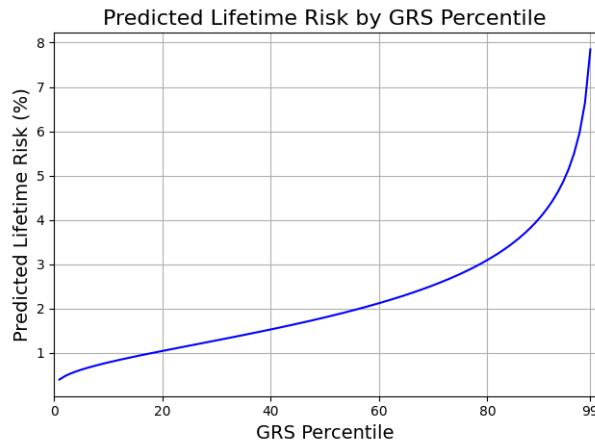


Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 2.2% estimate from the National Cancer Institute.²

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	1.80%	1.00x
95th	5.15%	2.87x
97th	5.97%	3.32x
99th	7.85%	4.37x

Table 2. Predicted lifetime prevalence of melanoma at different GRS percentiles. Individuals with the highest GRS percentiles are predicted to have an increased risk of melanoma relative to those at the 50th percentile.

Conclusion

In this study, we evaluated our melanoma GRS on data from the UKB. We found that it performed well, particularly for identifying individuals with elevated risk of melanoma relative to the population. In our embryo and couple reports, we adjust the model to predict risk consistent with the prevalence estimated by the National Cancer Institute.² The melanoma GRS model is currently only available to individuals of European ancestry, due to limited data availability for validation in other ancestry groups, although research efforts are underway to address this issue.

Acknowledgments

This research was conducted using the UK Biobank Resource under Application Number 80545 and data from the *All of Us* Research Program. We gratefully acknowledge UK Biobank and *All of Us* participants for their contributions, without whom this research would not have been possible.

References

1. U.S. National Cancer Institute. Melanoma Treatment (PDQ) — Patient Version. National Cancer Institute, 2025. URL: <https://www.cancer.gov/types/skin/patient/melanoma-treatment-pdq>.
2. U.S. National Cancer Institute. Melanoma of the Skin — Cancer Stat Facts. SEER, 2026. URL: <https://seer.cancer.gov/statfacts/html/melan.html>.
3. U.S. National Cancer Institute. Genetics of Skin Cancer (PDQ) — Health Professional Version. National Cancer Institute, 2025. URL: <https://www.cancer.gov/types/skin/hp/skin-genetics-pdq>.
4. M.T. Landi, D.T. Bishop, S. MacGregor, et al. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nature Genetics*, 52(5):494–504, 2020. doi:10.1038/s41588-020-0611-8.
5. M.I. Kurki, J. Karjalainen, P. Palta, et al. FinnGen provides genetic insights from a well-phenotyped iso-

lated population. *Nature*, 613(7944):508–518, 2023. doi:10.1038/s41586-022-05473-8.

6. A. Verma, J.E. Huffman, A. Rodriguez, et al. Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science*, 385(6706):eadj1182, 2024. doi:10.1126/science.adj1182.
7. All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature*, 627(8003):340–346, 2024. doi:10.1038/s41586-023-06957-x.
8. C. Sudlow, J. Gallacher, N. Allen, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779, 2015. doi:10.1371/journal.pmed.1001779.
9. J. Mbatchou, L. Barnard, J. Backman, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics*, 53:1097–1103, 2021. doi:10.1038/s41588-021-00870-7.
10. C.J. Willer, Y. Li, and G.R. Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, 2010. doi:10.1093/bioinformatics/btq340.
11. Z. Zheng, S. Liu, J. Sidorenko, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nature Genetics*, 56:767–777, 2024. doi:10.1038/s41588-024-01704-y.
12. N. Chatterjee, J. Shi, and M. García-Closas. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nature Reviews Genetics*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.

Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
1.80%	1.94	3.75

Table 3. OR per SD. The baseline risk for an individual with a median GRS, and the predicted OR at one and two SDs, respectively. A GRS must have a predicted OR >2 at 2 SD to be included in Orchid’s clinical reports.

UKB Prevalence	Population Prevalence	Liability R^2
1.0%	2.2%	7.79%

Table 4. Liability R^2 . The estimated liability R^2 using a population prevalence of 2.2%.

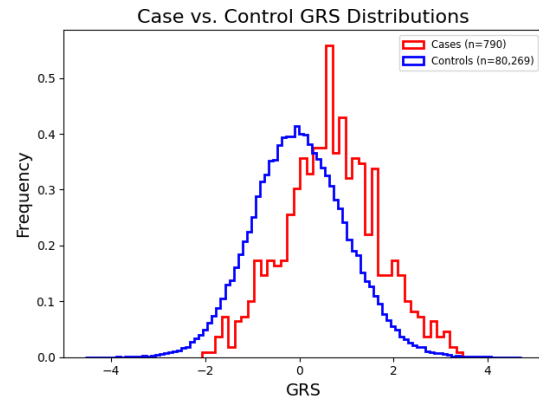


Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.

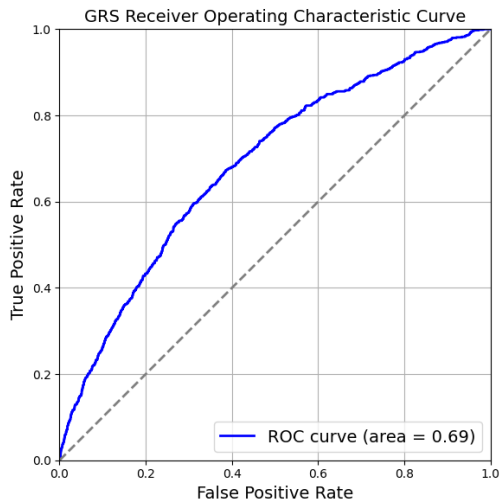


Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds. A curve closer to the top-left indicates a better model, while a diagonal line ($AUC = 0.5$) represents random guessing.

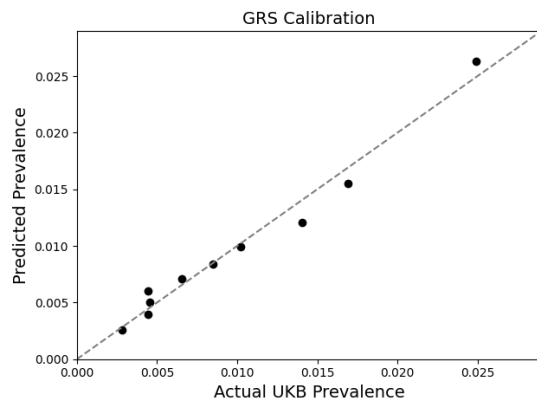


Figure 5. Calibration Curve. Calibration plot showing observed melanoma prevalence versus predicted risk across GRS deciles.