

Predicting *hMSH2* and *hMLH1* Mutations in Colorectal Cancer Patients

Jihong Zong, M.D., Ph.D.¹, Marsha L Frazier, Ph. D.², Carol J Etzel, Ph.D.², Bernard Levin, M.D.³, Nancy A Viscofsky, M.P.H.², Patrick M Lynch, J.D., M.D.⁴, and Christopher I Amos, Ph.D.²

¹ Section of Clinical Cancer Genetics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.

² Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.

³ Division of Cancer Prevention, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.

⁴ Department of Gastrointestinal Medicine and Nutrition, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.

Background. Individuals who carry mutations in *hMSH2* or *hMLH1* are at greatly elevated risk for colorectal cancer (CRC) and other, extracolonic cancers. Treatment of hereditary cancers should be tailored to the patient to be maximally effective. However, relatively few studies have yet been presented to identify the clinical risk factors that best predict the presence of *hMSH2* or *hMLH1* mutation, and studies in U.S. populations are particularly lacking.

Methods. We analyzed 123 probands with colorectal cancer, using classification tree modeling and logistic regression to identify the characteristics of a case that most strongly predicts the presence of a pathogenic *hMSH2* or *hMLH1* mutation. As potential clinical indicators, we evaluated sex, age at diagnosis of CRC, satisfying Amsterdam criteria I, satisfying Amsterdam criteria II, number of CRCs in first-degree relatives, presence of multiple cancers in the proband, and interaction terms. The likelihood of being an *hMSH2* or *hMLH1* mutation carrier within each terminal node based on different population risk was calculated using Bayes' theorem.

Results. In the multiple logistic regression analysis, the number of CRCs in first-degree relatives was identified as the only significant predictor, with an odds ratio (OR) of 1.7. In the tree-based analysis, at least one CRC in first-degree relatives (OR=3.3), age of diagnosis before 60 for patients having at least one CRC in first-degree relatives (OR=infinity), and fulfilling Amsterdam criteria II for patients diagnosed before the age of 60 and having at least one first-degree relative with CRC (OR=7.4) were identified as the most important predictors in the classification tree. For the logistic regression and tree-based models, the sensitivities were 34% and 59% at a specificity of 84%, and the areas under the receiver-operating-characteristic (ROC) curves were 0.651 and 0.733, respectively, indicating better performance by the tree-based method.

Conclusions. Our results showed that the tree-based method more accurately predicted carriers of mutations in these genes than logistic regression. In addition, we provide an easily implemented procedure for predicting those at highest risk of carrying mutations. Future research with information on more complete clinical features [and] including subjects from different clinical settings is needed.