Leveraging tumor characteristics to predict germline variant pathogenicity in mismatch repair genes

Shuwei Li, Dajun Qian, Bryony A. Thompson, Stephanie Gutierrez, Tina Pesaran, Holly LaDuca, Hsiao-Mei Lu, Elizabeth C. Chao, Mary Helen Black

Pathogenic variants (PV) in mismatch repair (MMR) genes (*MLH1, MSH2, MSH6, PMS2*) increase cancer risks. Microsatellite instability (MSI) and protein immunohistochemical (IHC) changes due to MMR deficiency are often assessed to clinically evaluate cancer risk. It has been shown that these tumor characteristics may be useful in predicting germline MMR variant pathogenicity. In a large cohort of patients who underwent germline panel testing (GPT; n=4827) or paired germline/somatic testing (GST; n=1429), we derived likelihood ratios (LR=%PV carriers/%non-carriers) for those with abnormal vs. normal MSI/IHC status. For each variant, we computed the tumor characteristic LR (TCLR) as the product of carriers' MSI/IHC LRs. For GST samples, we additionally estimated MLH1-hypermehtylation $(hm)^{+/-}$ LRs among samples with abnormal MSI/IHC (n=583). The MSI/IHC LRs were combined with MLH1-$hm^{+/-}$ LRs to compute overall and tumor-specific (colorectal, endometrial, ovarian, breast) TCLRs. We then compared the prediction performance of 3 approaches: TCLR alone, *in silico* priors with TCLR, and integrating TCLR evidence with *in silico* and other evidence into a Bayesian multifactorial model. Comparing carriers of germline PV vs. non-carriers, GPT samples had LRs of 2.98 for abnormal and 0.10 for normal MSI/IHC, while GST LRs were 1.57 and 0.04 for abnormal and normal MSI/IHC, respectively. The LR in favor of pathogenicity for samples with abnormal MSI/IHC was 4.47 for MLH1-hm^{-} and 0.03 for MLH1-hm^{+}. MSI/IHC LRs also varied by cancer type; LRs were highest in colorectal, whereas MLH1-hm^{-} LRs were significantly higher in endometrial than colorectal cancers. TCLR alone accurately predicted 67.2% of 177 classified germline missense variants. Combining TCLR with *in silico* priors accurately classified >73.4% variants, depending on the *in silico* scores used for analysis. Integrating TCLR into a multifactorial model increased the proportion of accurately classified variants to 97.2%. Applying these models to the prediction of 221 variants of unknown significance (VUS), TCLR alone predicted 73% as (likely) pathogenic or (likely) benign, while the multifactorial model with TCLR reclassified 64%. However, TCLR alone yielded 5 false positives/negatives, while the multifactorial model resulted in 0 false predictions. These data highlight the importance of incorporating somatic sequencing in germline variant assessment and demonstrate the utility of TCLR evidence for VUS reclassification.