



**Figure S4.** Best predictive models for test-retest variability using stepwise linear regression, using false positive rate  $\leq 15\%$  as the reliability cutoff. A total of 78 additional patients out of 4,044 were excluded with this criterion. Top panel (A-C): regression coefficients for (A) “base model” selected from global and reliability indices only, (B) “base + total deviation (TD) model” selected from base parameters as well as TD values at 52 locations, and (C) “base + archetype (AT) model” selected from base parameters as well as archetypes. Bottom panel (D-F): increase in Bayesian information criterion (BIC) when each parameter is removed from the respective models. *Blue* = base parameters; *orange* = TD values; *red* = archetypes. MD = mean deviation; PSD = pattern standard deviation; GHT = glaucoma hemifield test; FNR = false negative rate; FPR = false positive rate.