The Kernel Based Adaptive Cluster (KBAC) test was originally developed for detecting associations with rare variants in case control studies [1]. In KBAC, the coding function \( K_0 (\tilde{X}_i, Y^*_i) \) depends on both the multi-site genotype \( \tilde{X}_i \) and the disease status \( Y^*_i \). Specifically, multi-site genotypes that are more enriched in cases are assigned higher weights, such that potentially causal variants can be distinguished from non-causal variants. The weights are then incorporated into a logistic regression model. Association testing can be performed using score tests and p-values need to be evaluated empirically through permutations. It was shown in Liu and Leal [1], that KBAC can be more powerful than alternative methods in the presence of non-causal variants, or gene interactions.

In order to generalize the KBAC statistics to analyze quantitative traits, it is necessary to extend the kernel weight function \( K_0 (\tilde{X}_i, Y^*_i) \), which was originally only defined for binary traits. Two binary auxiliary traits are defined, i.e. \( Y_{i,H}^* = \delta(Y_i > Y^H) \), and \( Y_{i,L}^* = \delta(Y_i < Y^L) \), where \( \delta \) is an indicator function, and \( Y_{i,H}^* \) and \( Y_{i,L}^* \) are trait cutoffs with \( Y_{i,H}^* > Y_{i,L}^* \). When selected samples are used, \( Y_{i,H}^* \) and \( Y_{i,L}^* \) are set to be equal to the trait thresholds employed for sample ascertainment. When population based random samples are used, as a default, \( Y_{i,H}^* \) and \( Y_{i,L}^* \) are set to be the 75th and 25th percentiles of the sample quantitative trait values.

In order to test for one-sided hypothesis, e.g. rare causal variants are more enriched in samples with high (or low) trait values, the genotype coding \( K (\tilde{X}_i, Y_i) = K_0 (\tilde{X}_i, Y_{i,H}^*) \) (or \( K (\tilde{X}_i, Y_i) = K_0 (\tilde{X}_i, Y_{i,L}^*) \)) is used. Score statistics \( T_H \) (or \( T_L \)) are calculated based upon equation 10, where full quantitative trait is analyzed. If there is no prior information on which extreme
enriches rare causal variants, two-sided hypothesis should be tested and the statistic

\[ T = \max \left\{ \left( T^H \right)^2, \left( T^L \right)^2 \right\} \]

is used. Standard permutation algorithms can be used to obtain p-values empirically.

Reference: