**Methodology for Calculating P-Values from iHS Scores**

The script processes iHS scores to calculate p-values and adjusts them for multiple testing using a series of statistical methods. Initially, the script reads the .ihs.out file, which contains iHS scores for various single nucleotide polymorphisms (SNPs). Assuming the iHS scores follow a standard normal distribution under the null hypothesis of no selection, the script calculates two-sided p-values.

The p-value calculation is based on the cumulative distribution function (CDF) of the standard normal distribution. Specifically, for a given iHS score, the two-sided p-value is computed using the formula p=2×(1−Φ(∣iHS∣))p = 2 \times (1 - \Phi(|\text{iHS}|))p=2×(1−Φ(∣iHS∣)), where Φ\PhiΦ is the CDF of the standard normal distribution and ∣iHS∣|\text{iHS}|∣iHS∣ is the absolute value of the iHS score. This formula captures the probability of observing an absolute iHS score as extreme or more extreme under the null hypothesis.

After calculating the p-values, the script sorts them in ascending order and assigns ranks to each SNP based on their p-values. To control the false discovery rate (FDR) due to multiple hypothesis testing, the script adjusts the p-values using the Benjamini-Hochberg (BH) method. The adjusted p-value is calculated using the formula adjusted\_p=(p\_value×total\_testsrank)\text{adjusted\\_p} = \left( \frac{\text{p\\_value} \times \text{total\\_tests}}{\text{rank}} \right)adjusted\_p=(rankp\_value×total\_tests​), where total\_tests is the total number of SNPs tested, and rank is the rank of the SNP's p-value. The adjusted p-values are then clipped to a maximum value of 1.0 to ensure they do not exceed this threshold.

The methodology employed in the script is supported by statistical literature. The BH method, which controls the FDR, was introduced by Benjamini and Hochberg in their 1995 paper, "Controlling the false discovery rate: a practical and powerful approach to multiple testing" (Journal of the Royal Statistical Society: Series B, 57(1), 289-300). Additionally, the concept of using standard normal distribution properties to compute p-values for iHS scores is based on the work by Voight et al. in their 2006 paper, "A map of recent positive selection in the human genome" (PLoS Biology, 4(3), e72).