(A) KRLMM workflow

1. IDAT files
2. GenomeStudio
3. GenCall output (including X raw & Y raw values)
4. crlmm R/Bioconductor package
5. KRLMM: `genotype.Illumina(..., call.method='krlmm')`
   - Normalization (between sample quantile per channel)
   - Genotyping (one-dimensional k-means clustering of M-values, using k predicted by logistic regression)
6. Genotype calls
   Call confidence scores

(B) Regression analysis to choose k

For SNP $i$:

- **$R_{ik}$**: residual sum of squares for SNP $i$ for a given $k$ (1, 2, 3)
- **$D_{ik}$**: Mahalanobis distance between k-means cluster centers and nearest consensus cluster center for SNP $i$ for a given $k$ (1, 2, 3)
- **$H_{ik}$**: deviation from Hardy-Weinberg equilibrium for SNP $i$ for a given $k$ (2, 3)

Use variables $R_{ik}$, $D_{ik}$, and $H_{ik}$ to predict $k$ for SNP $i$ using logistic regression with coefficients pre-determined using 10,000 SNPs where $k$ is known in advance (based on HapMap samples and independent calls).