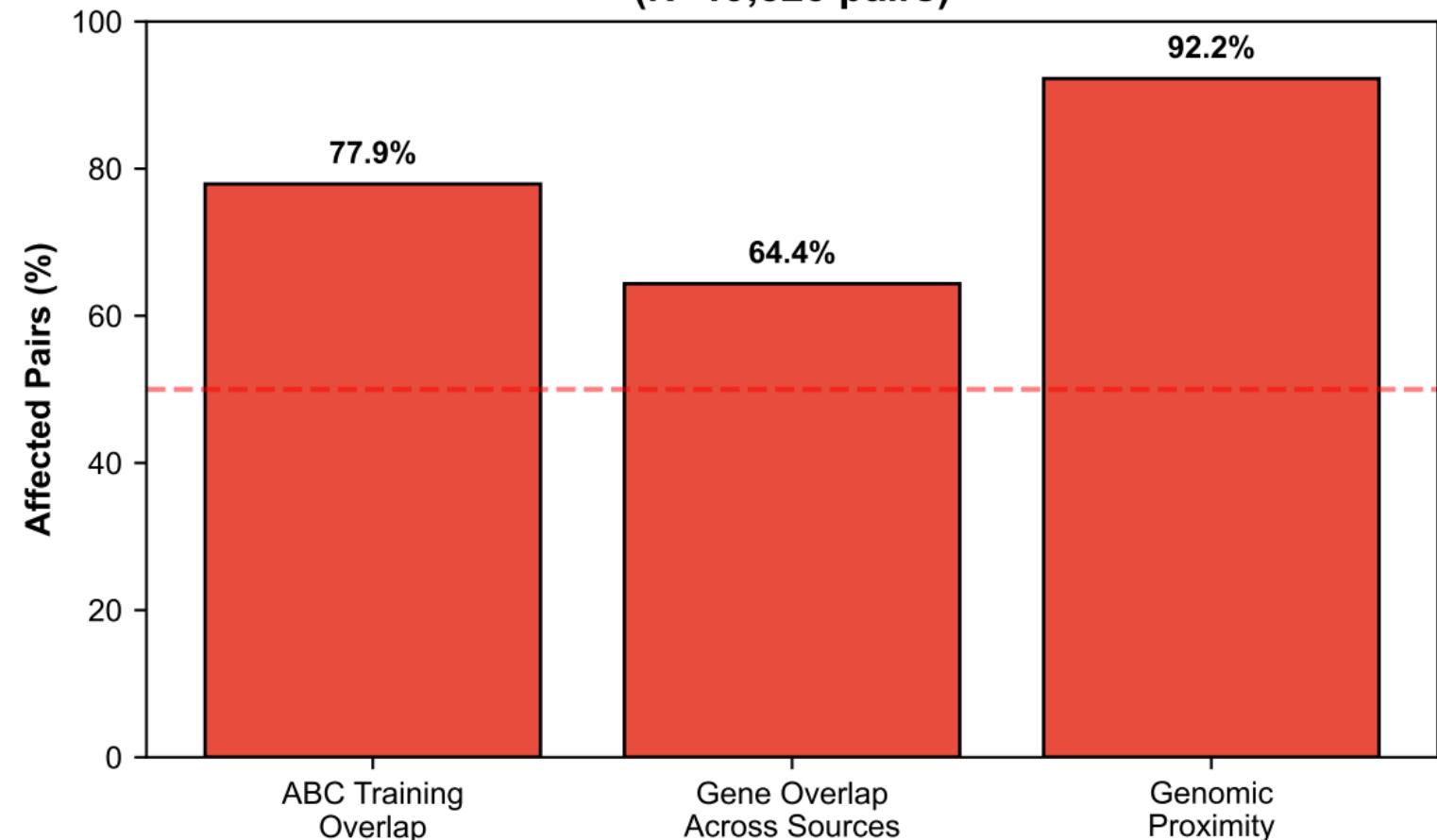


## Task B Leakage Audit Results (N=19,825 pairs)



## Leave-One-Study-Out Splits for Unbiased Evaluation

