

Benchmark: MMLUMedical Genetics
Performance Metric: Target Greedily Decoded
Correlation Metric: Spearman

