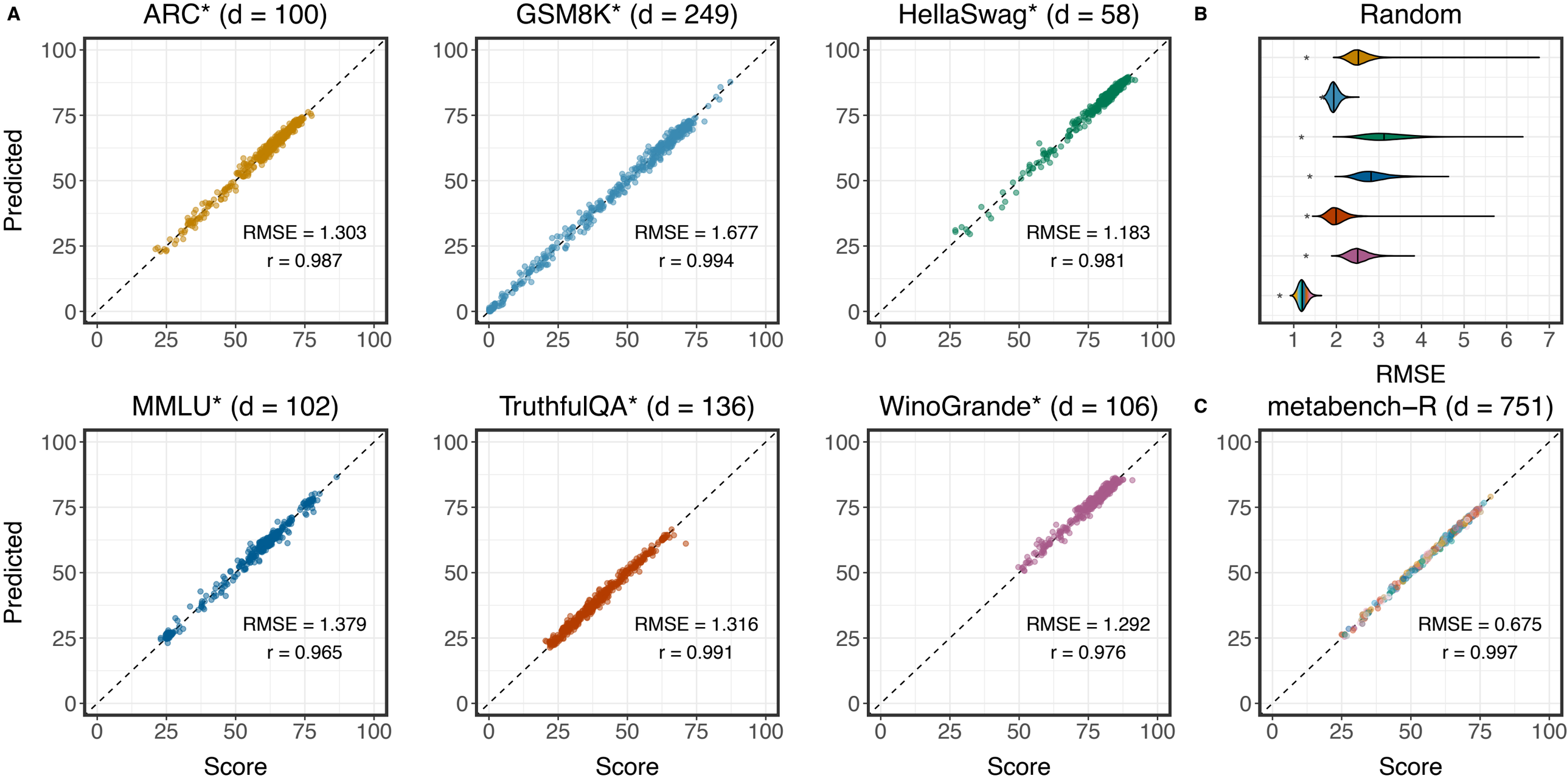
A
ARC* (d = 100)
RMSE = 1.303
r = 0.987
GSM8K* (d = 249)
RMSE = 1.677
r = 0.994
HellaSwag* (d = 58)
RMSE = 1.183
r = 0.981
B
Random
RMSE
MMLU* (d = 102)
RMSE = 1.379
r = 0.965
TruthfulQA* (d = 136)
RMSE = 1.316
r = 0.991
WinoGrande* (d = 106)
RMSE = 1.292
r = 0.976
C
metabench–R (d = 751)
RMSE = 0.675
r = 0.997
Predicted
Score