Table S3. Statistical comparison of the best preforming 12 groups in QA2.1 mode (per-residue quality estimates assessed on a per-target basis)

		56	**397**	**78**	**426**	**369**	**490**	**273**	**80**	**324**	**308**	**119**	**367**
PconsM	**56**	X	116	116	113	114	113	116	116	106	114	116	116
ModFOLDclust2	**397**	0.01	X	116	113	114	113	116	116	106	114	116	116
IntFOLD-QA	**78**	<0.01	<0.01	X	113	114	113	116	116	106	114	116	116
MetaMQAPclust	**426**	<0.01	0.03	0.2	X	113	110	113	113	105	113	113	113
MQAPmulti	**369**	<0.01	<0.01	0.03	0.52	X	111	114	114	106	114	114	114
MULTICOM	**490**	<0.01	<0.01	<0.01	0.05	0.1	X	113	113	106	111	113	113
Pcomb	**273**	<0.01	<0.01	0.03	0.16	0.27	0.33	X	116	106	114	116	116
Multicom-construct	**80**	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.62	X	106	114	116	116
AOBA	**324**	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.39	0.68	X	106	106	106
MQAPsingle	**308**	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.21	0.1	0.78	X	114	114
Multicom-refine	**119**	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.17	0.13	0.27	0.39	X	116
ModFOLDclustQ	**367**	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	0.04	X

Results of the two-tailed paired t-tests on Pearson's correlation coefficients for per-residue estimates. The upper right part of the table contains the numbers of common targets predicted. The lower part displays the probabilities that the differences between the two correlation coefficients are due to chance. Shaded cells highlight pairs of statistically indistinguishable groups at the 10^{-2} significance level.