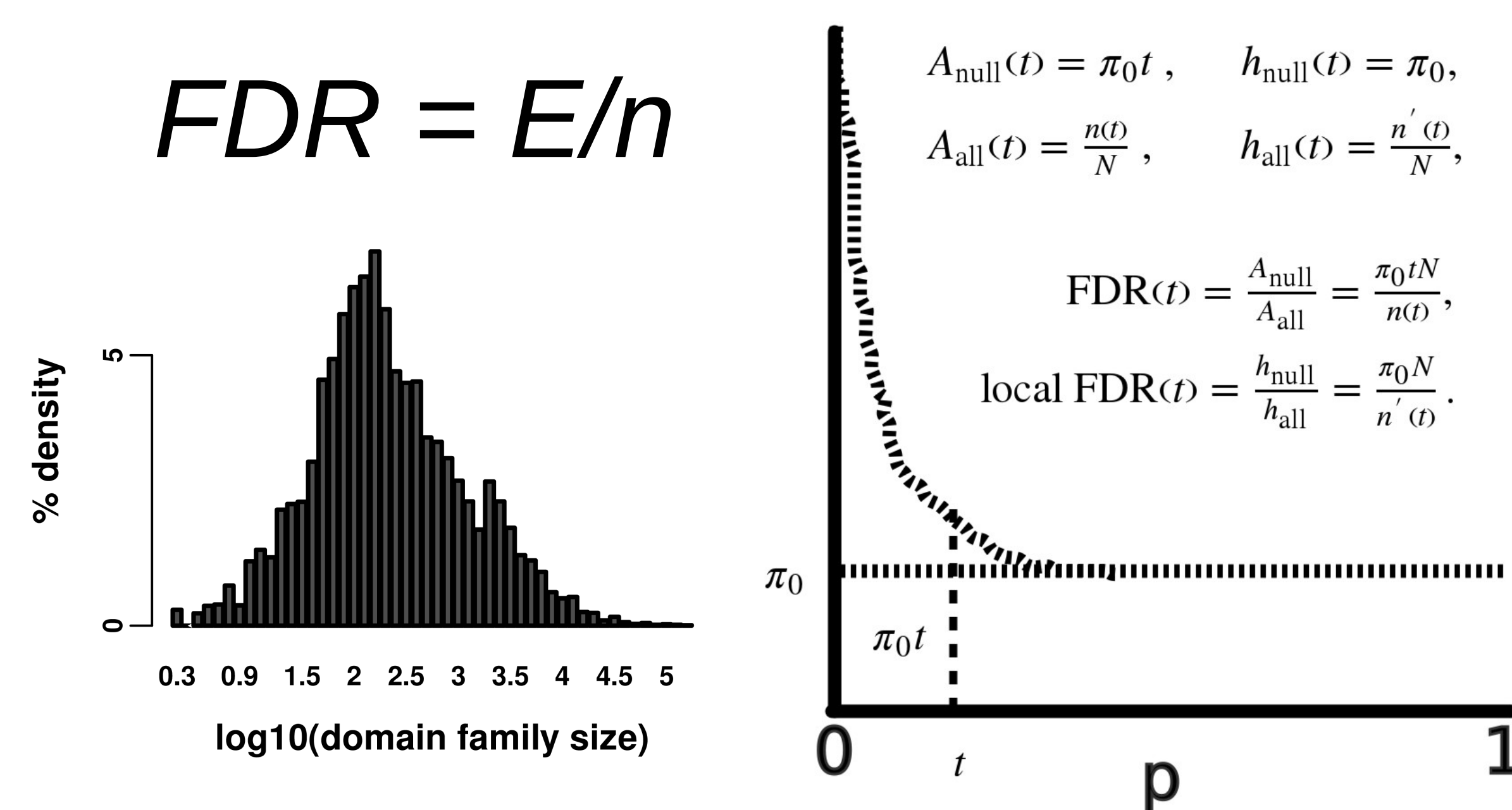


Forget the *E*-value: family-based *q*-values for protein domain prediction, and empirical error detection

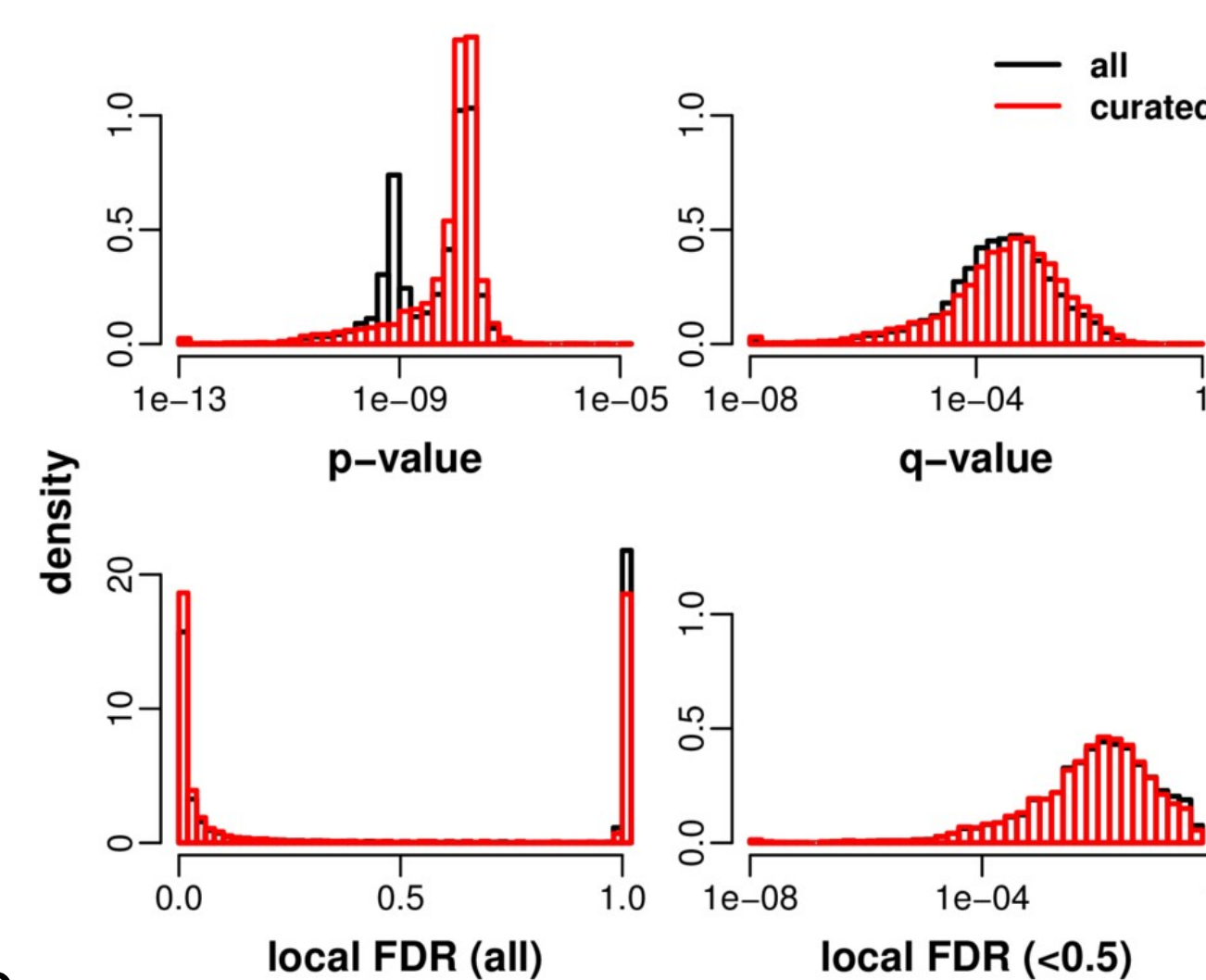
Alejandro Ochoa, Manuel Llinás, John Storey, and Mona Singh, Princeton University

The *FDR*, *local FDR*, and why they're better for domains



Pfam curated thresholds in terms of other stats

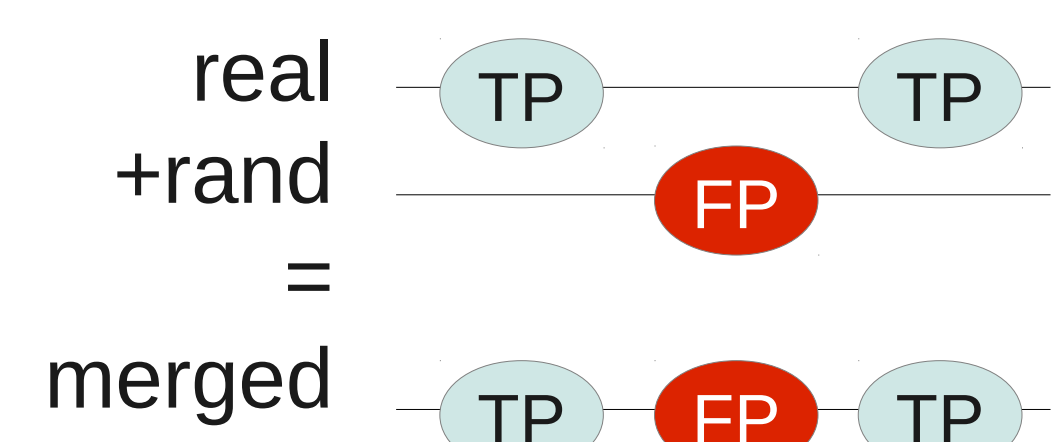
Median *q*-value = ~1/2000
Median *local FDR* = ~1/100



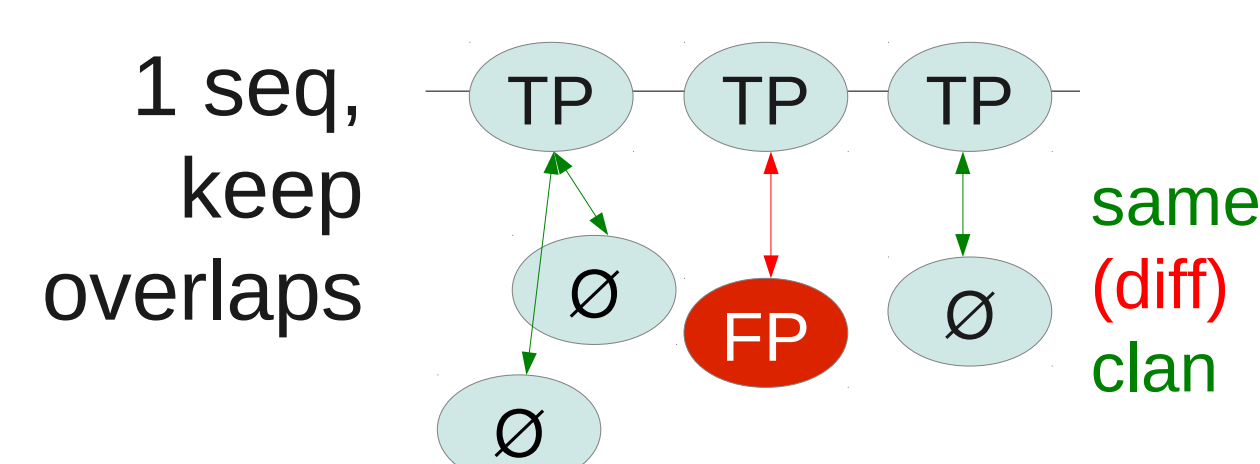
Empirical null models

$$FDR = \#FP / (\#FP + \#TP)$$

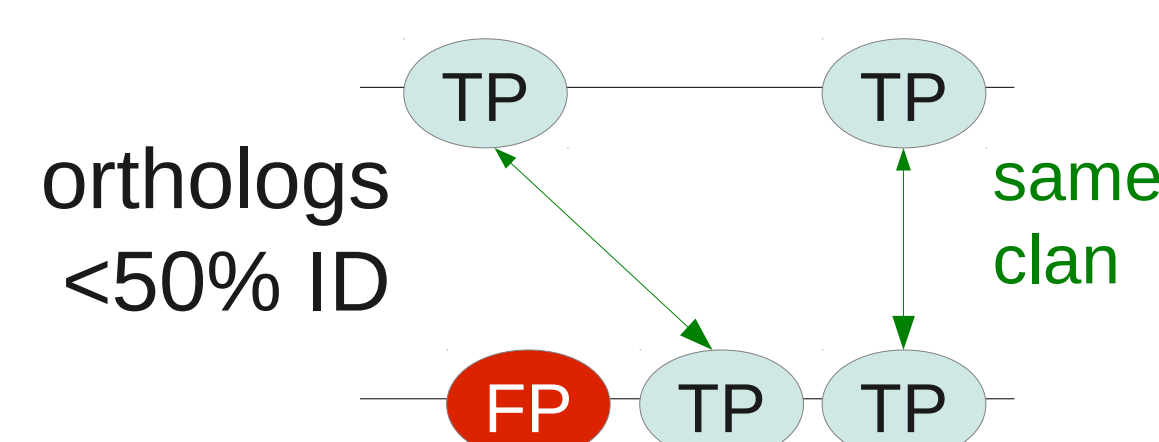
2nd order Markov Random



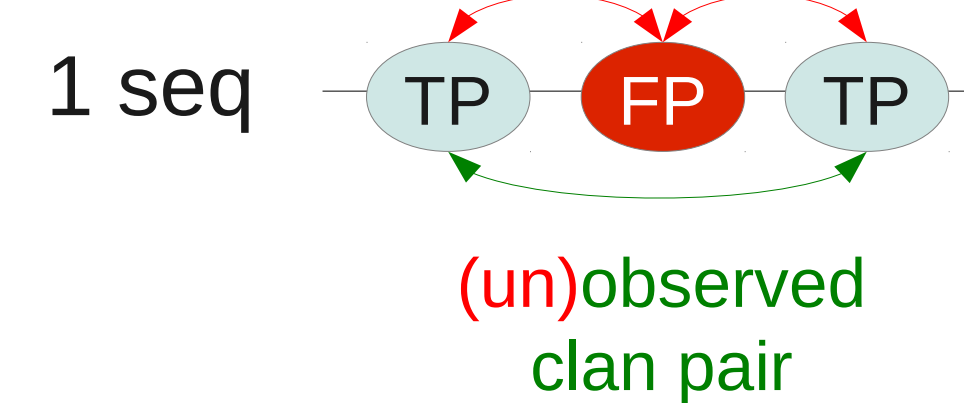
Clan Overlap



Ortholog Set Coherence



Context Coherence



Clan Overlap: Inspired/adapted from S. Eddy (p.c., 2012)

Context Coherence: Inspired/adapted from Terrapon, et al. BMC Bioinformatics 13, 67 (2012)

q > *local FDR* in benchmarks

The *q*-values are more robust than the *local FDR* when *p*-values are imperfect

Improving more than domain number

q-value predictions are more informative

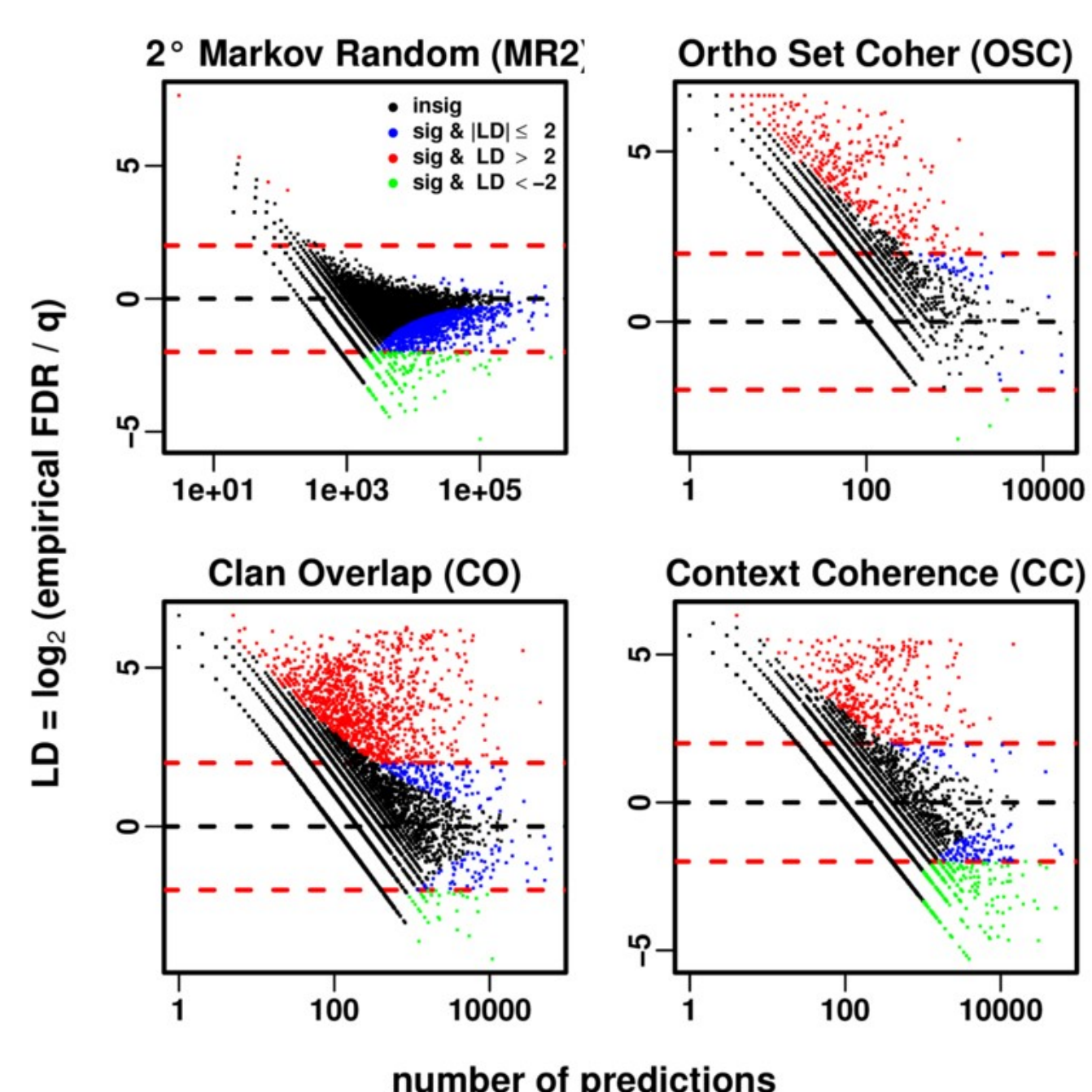
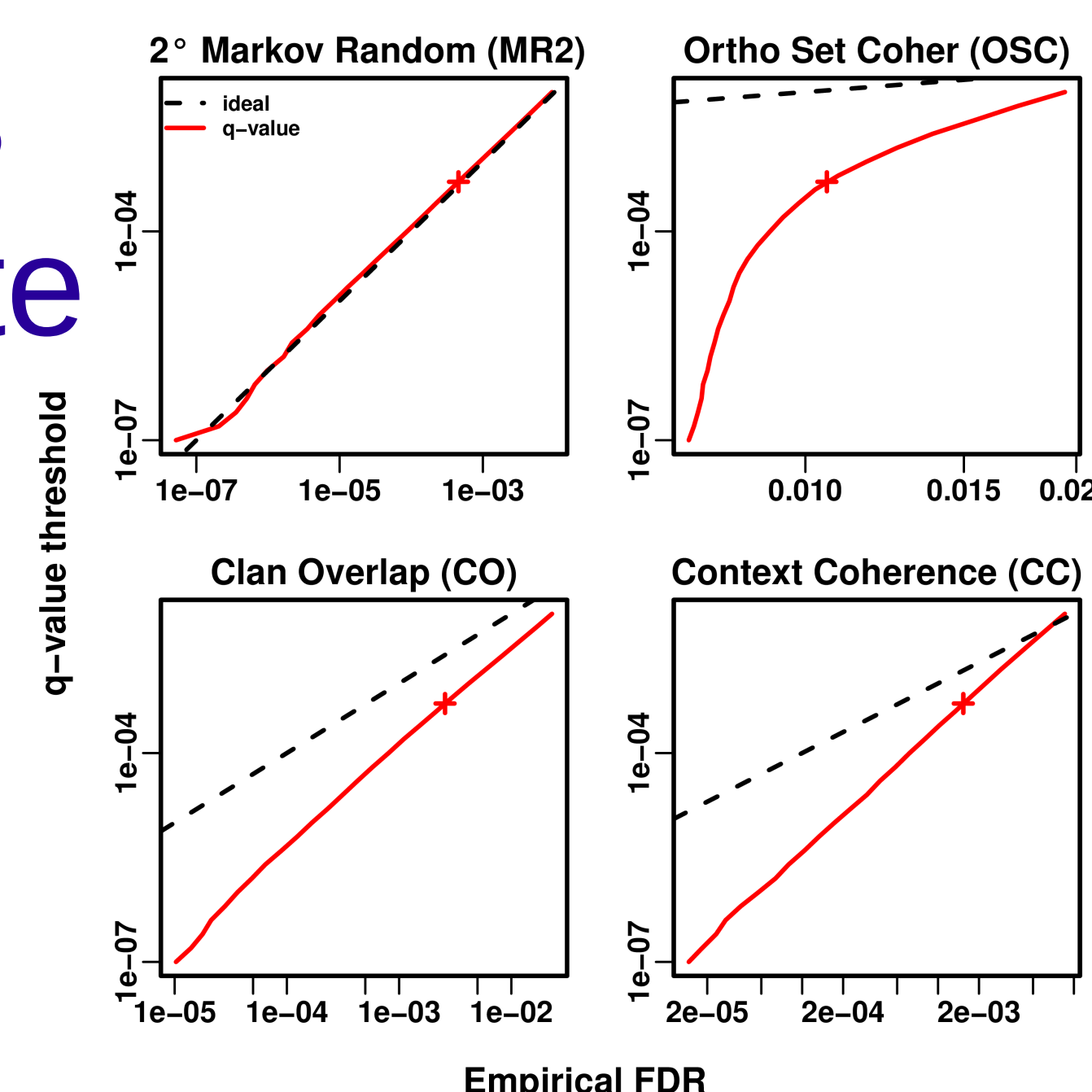
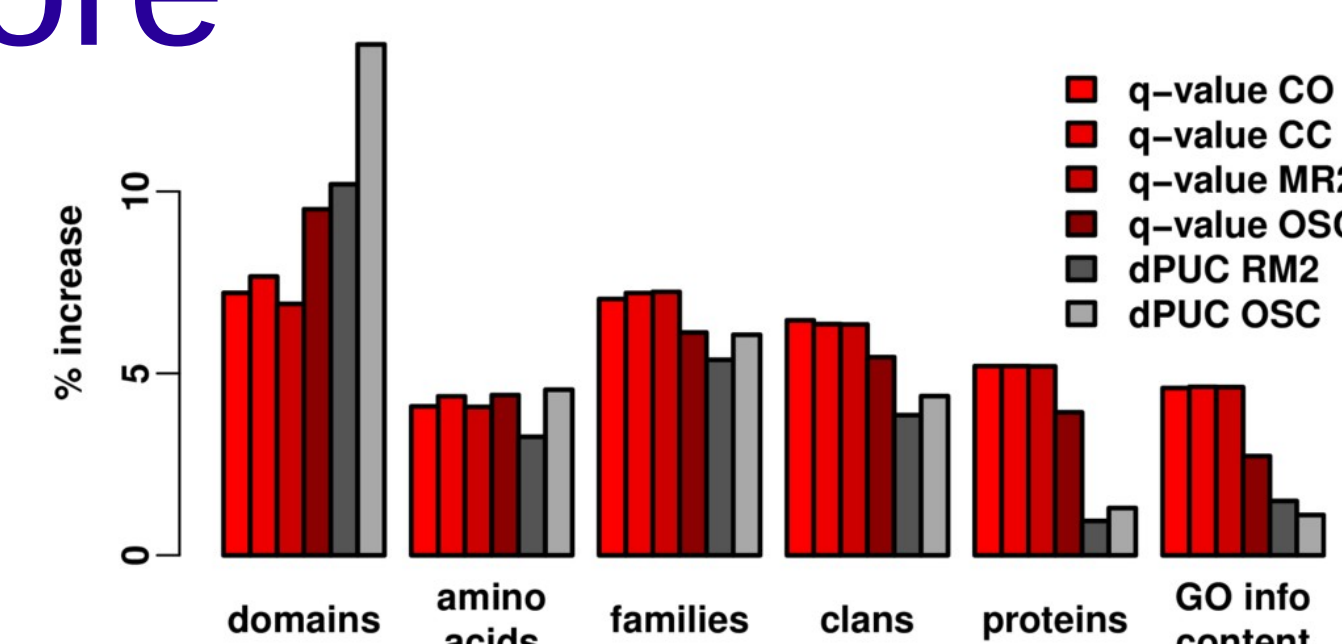
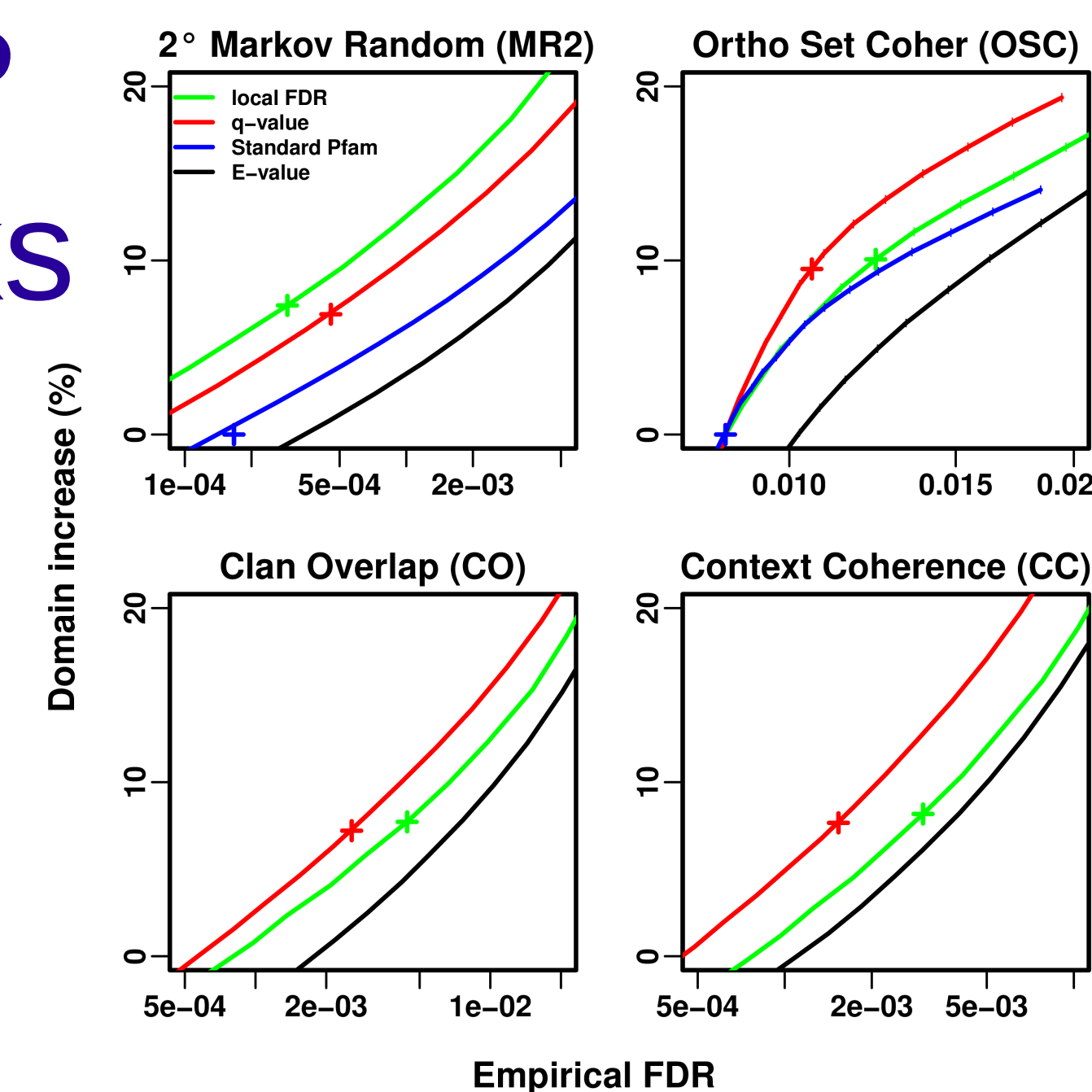
The *q*-values underestimate empirical *FDR*s

Should be equal, discrepancy is due to failure of theoretical *p*-values (null model)

Per-family analysis of noise

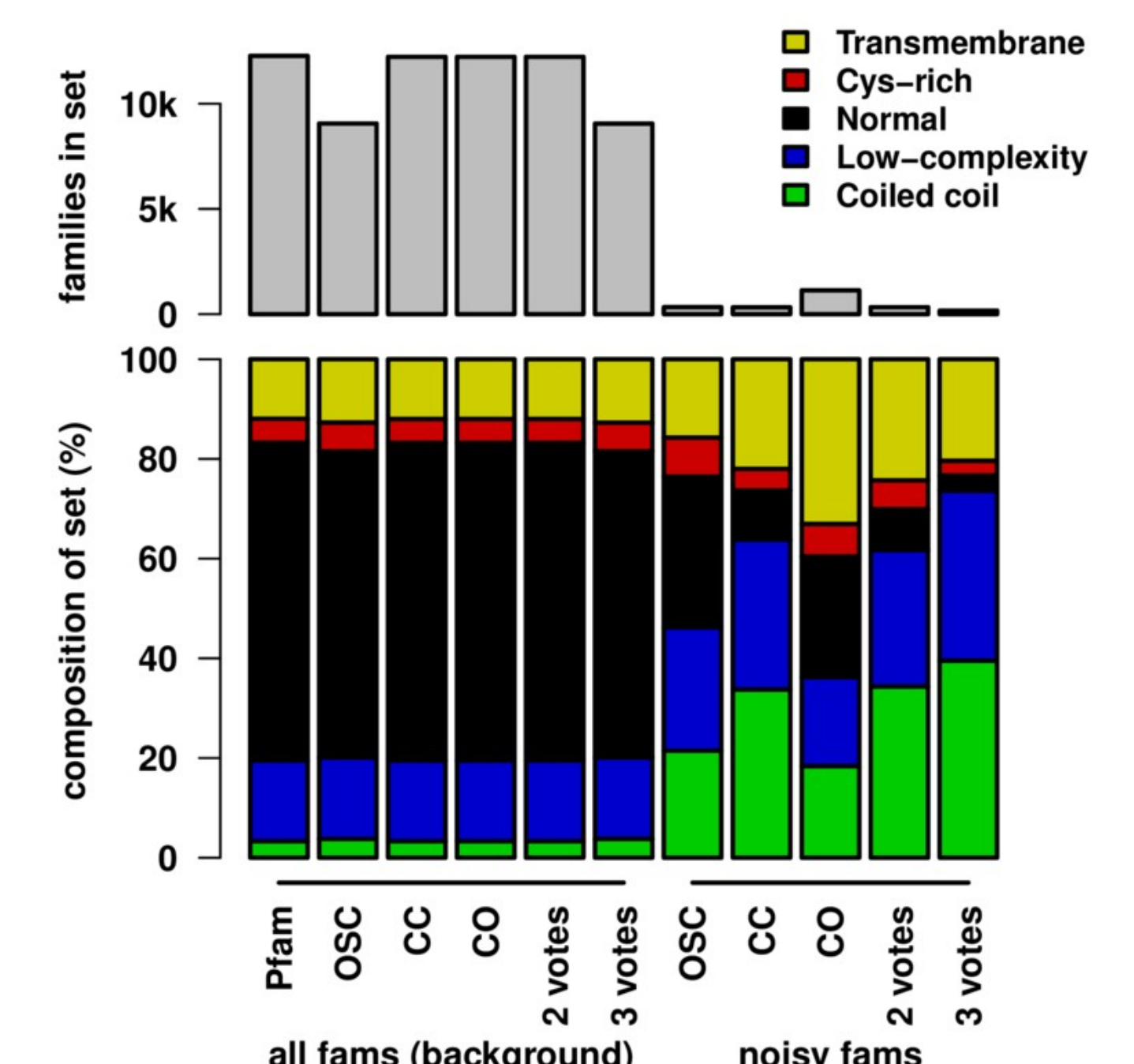
Want to identify noisy families (red), but need to be careful with low sampling of domain families (statistical significance of deviation from expectation)

Only 329 fams (2.7% of Pfam) are noisy in at least two null models.



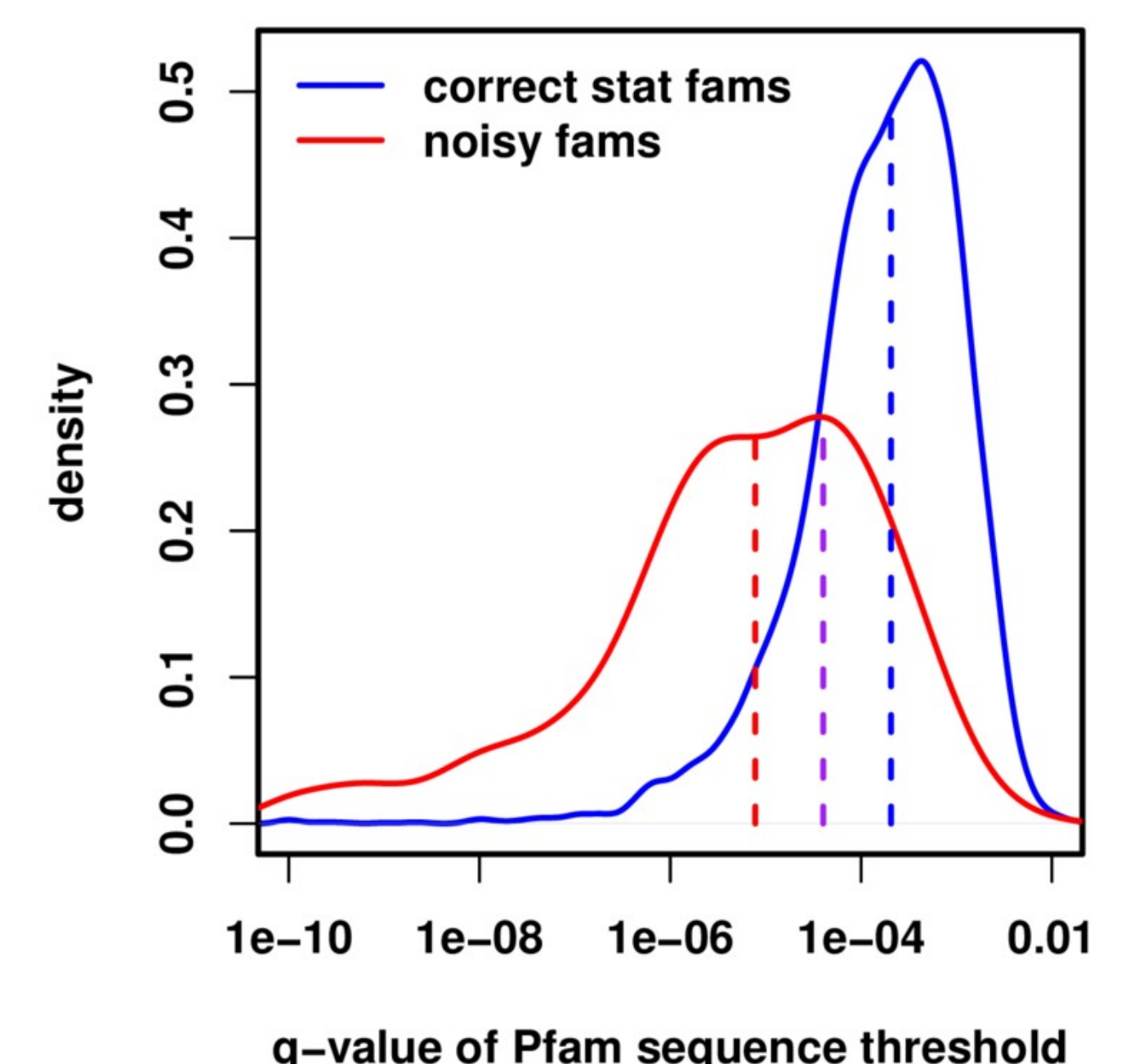
Classes enriched in noisy families

Noisy families have in common repeating patterns that make them similar although they're known to be non-homologous (in the theoretical null model similarity=homology)



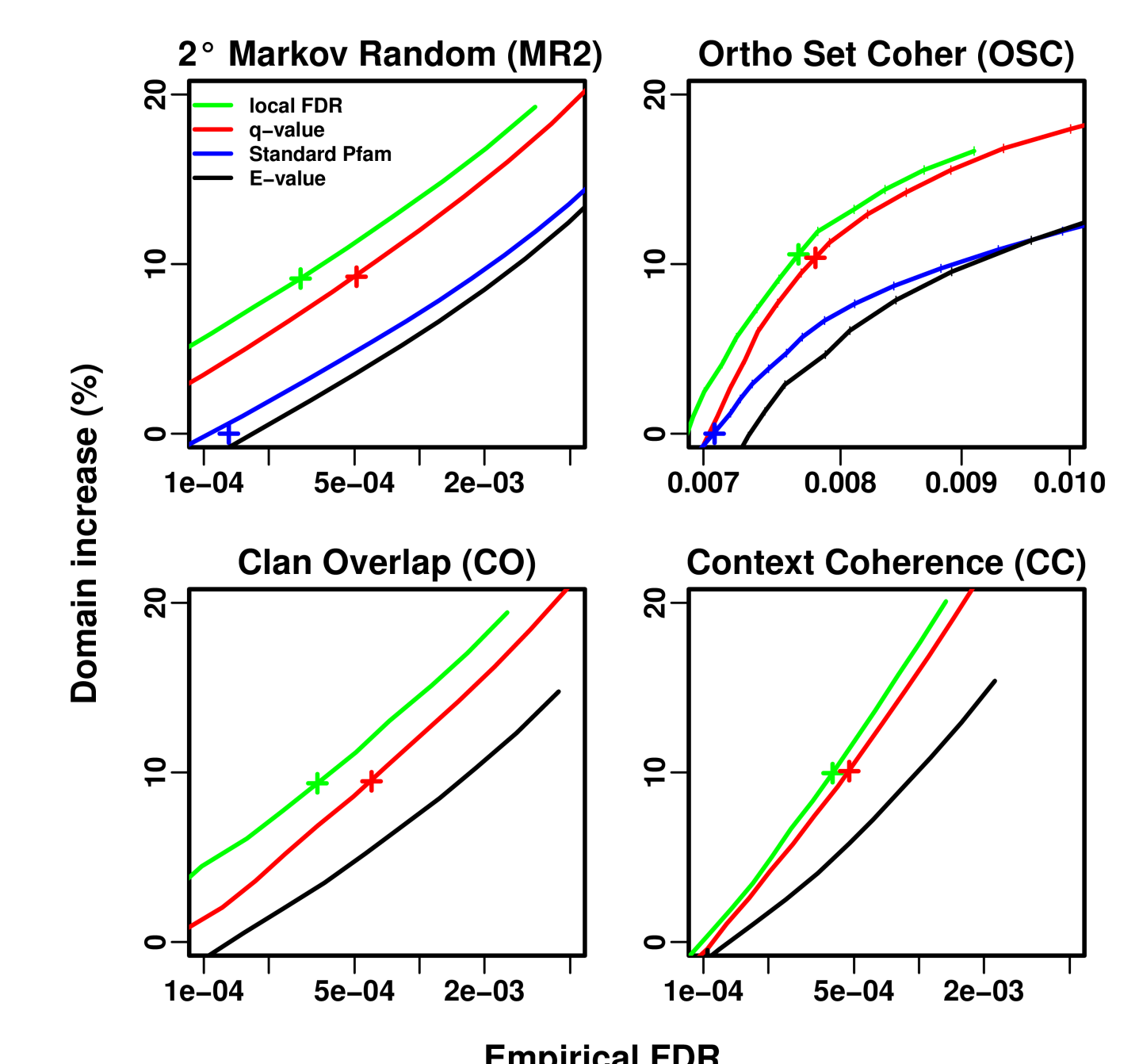
Pfam somewhat compensates on noisy fams

Noisy families have more stringent curated thresholds than correct-stat families in the median, but 31% of noisy families remain noisy in Pfam



Local FDR > *q* in fams with correct stats

The *local FDR* may become more useful in the future, when the underlying theoretical null model improves!



Contact: ochoa@princeton.edu