topconfects

Top confident effect sizes for differential expression

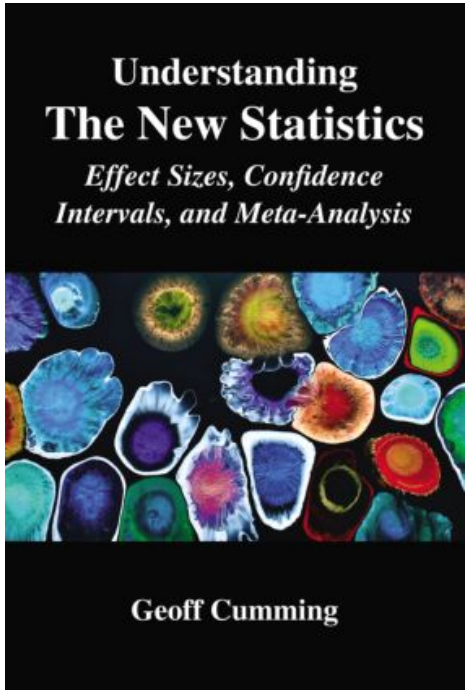
Paul Harrison @paulfharrison

Monash Bioinformatics Platform, Monash University

Widespread criticism of *p*-values

- Commonly misinterpreted.
- Only reject effect size zero, effect may not be important.
- Invalid when *selectively* reported.

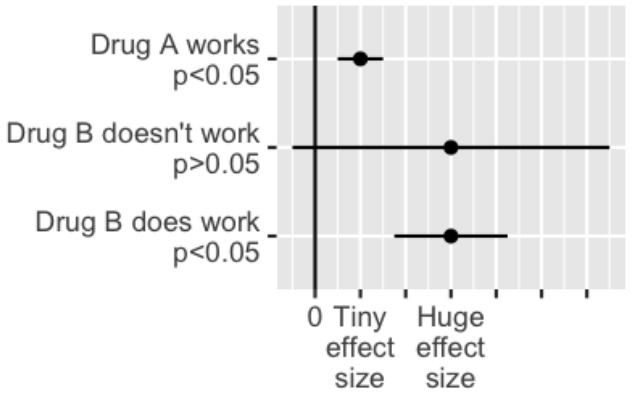
Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, 2(8):e124.



Confidence Intervals are better

~~significance~~ measurement accuracy

- Apparent paradoxes resolved.
- Same underlying theory.
- Also invalid when *selectively* reported.



Controlling the False Coverage-statement Rate (FCR)

Benjamini, Y. and Yekutieli, D. (2005). False Discovery Rate–Adjusted Multiple Confidence Intervals for Selected Parameters. *Journal of the American Statistical Association*, 100(469):71– 81.

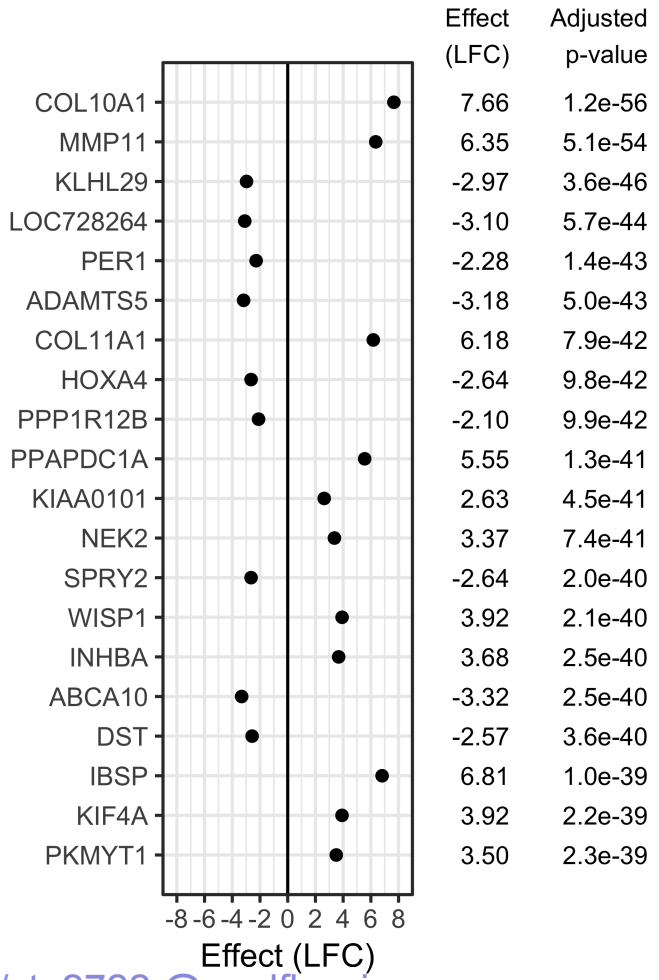
> To maintain an FCR of q
>
> after *selecting* k of n results
>
> provide $1 - \frac{k}{n}q$ confidence intervals.

- Generalization of FDR.
- **Any** selection rule may be used!

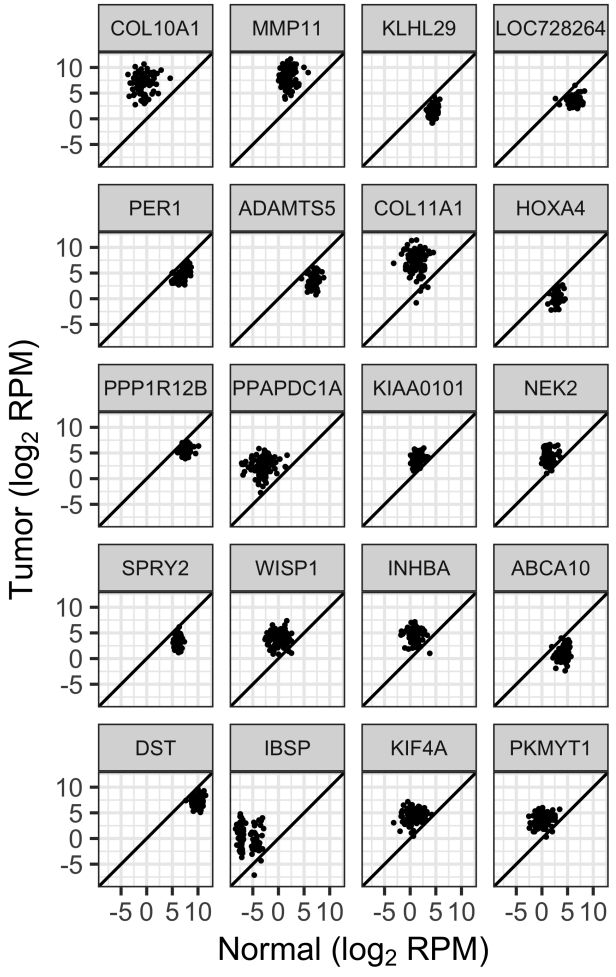
Caveats:

- Point estimates remain biassed by selection.
- The experiment without interesting results that is never published is still a problem.

Ranking by p is a common default in bioinformatics software



	Effect (LFC)	Adjusted p-value
COL10A1	7.66	1.2e-56
MMP11	6.35	5.1e-54
KLHL29	-2.97	3.6e-46
LOC728264	-3.10	5.7e-44
PER1	-2.28	1.4e-43
ADAMTS5	-3.18	5.0e-43
COL11A1	6.18	7.9e-42
HOXA4	-2.64	9.8e-42
PPP1R12B	-2.10	9.9e-42
PPAPDC1A	5.55	1.3e-41
KIAA0101	2.63	4.5e-41
NEK2	3.37	7.4e-41
SPRY2	-2.64	2.0e-40
WISP1	3.92	2.1e-40
INHBA	3.68	2.5e-40
ABCA10	-3.32	2.5e-40
DST	-2.57	3.6e-40
IBSP	6.81	1.0e-39
KIF4A	3.92	2.2e-39
PKMYT1	3.50	2.3e-39



TCGA breast cancer data:

RNA-Seq of 97 tumor-normal pairs.

`limma-voom` analysis.

Using this table:
Read down the table until desired FDR reached.

eg 13,784 DE genes at 5% FDR!

Sorting by *p*-value focusses on high signal-noise ratio.

Like ranking by Cohen's *d*,
but not comparable between experiments.

TREAT solves the problem of too many results

McCarthy, D. J. and Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, 25(6):765–771.

Null hypothesis for each gene that absolute LFC is smaller than threshold e.

- Specify threshold to be a biologically significant amount.
- Table sorted by adjusted p-values, again allows reader to choose FDR.

But:

- FDR of 5% is fine, don't need to give me this choice!
- Not clear how to choose threshold.

(Compared to worst case p-value from t-tests within the interval, TREAT shrinks p-values by up to half.)

Topconfects improves TREAT usability

TOP CONfident efFECT Sizes

Harrison, P. F., Pattison, A. D., Powell, D. R., and Beilharz, T. H. (2019). Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome Biology*, 20(1):67.

An alternative presentation of TREAT.

- Fixed FDR.
- Table ranked by **"confect"** values, allow reader to select threshold e.
- Once threshold chosen, confect values are confidence bounds that control the FCR (slightly conservative).

Topconfects method

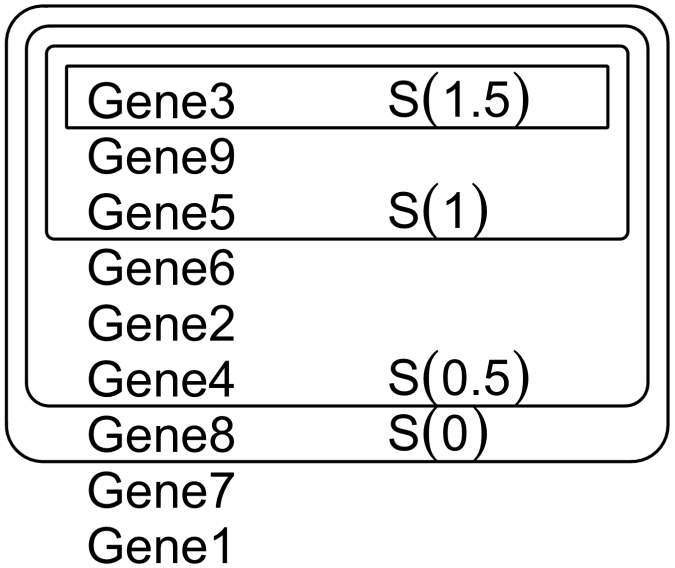
Let $S(e)$ be the set of genes selected by TREAT at threshold e, for some fixed FDR (eg 5%).

These sets nest:

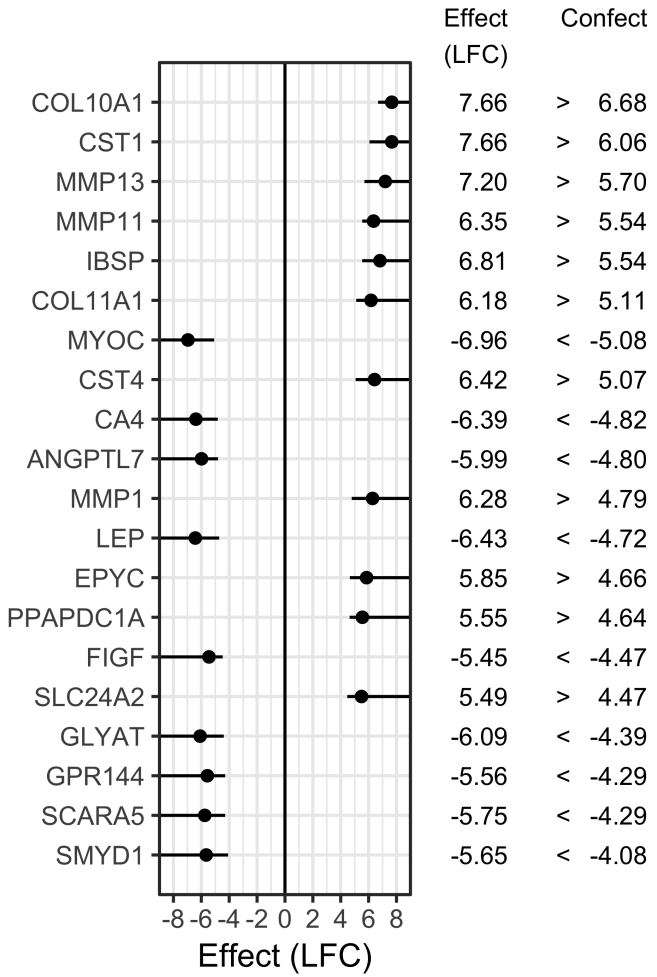
- If $e_1 > e_2$ then $S(e_1) \subseteq S(e_2)$.

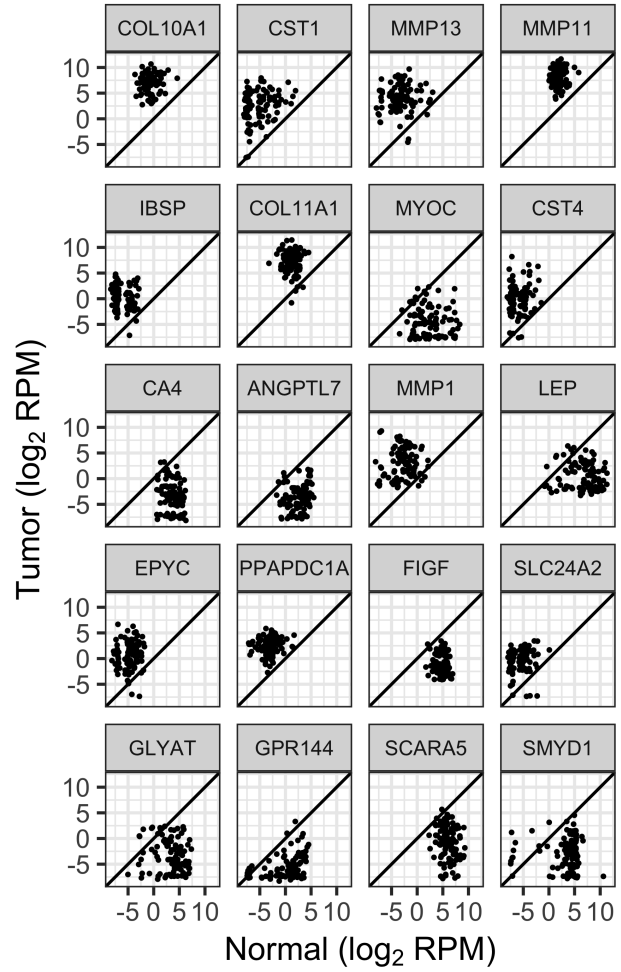
Try $e = 0, 0.01, 0.02, 0.03, \ldots$

Call the largest e for which a gene is a member of $S(e)$ its "confect", with appropriate sign.



Topconfects at 5% FDR





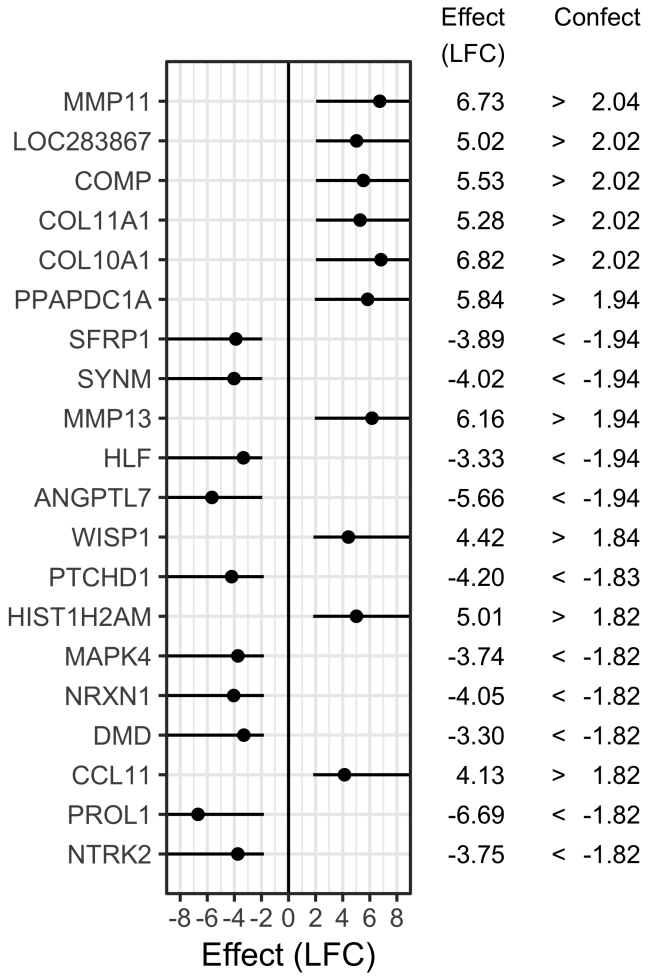
Top ranked genes have large LFCs.

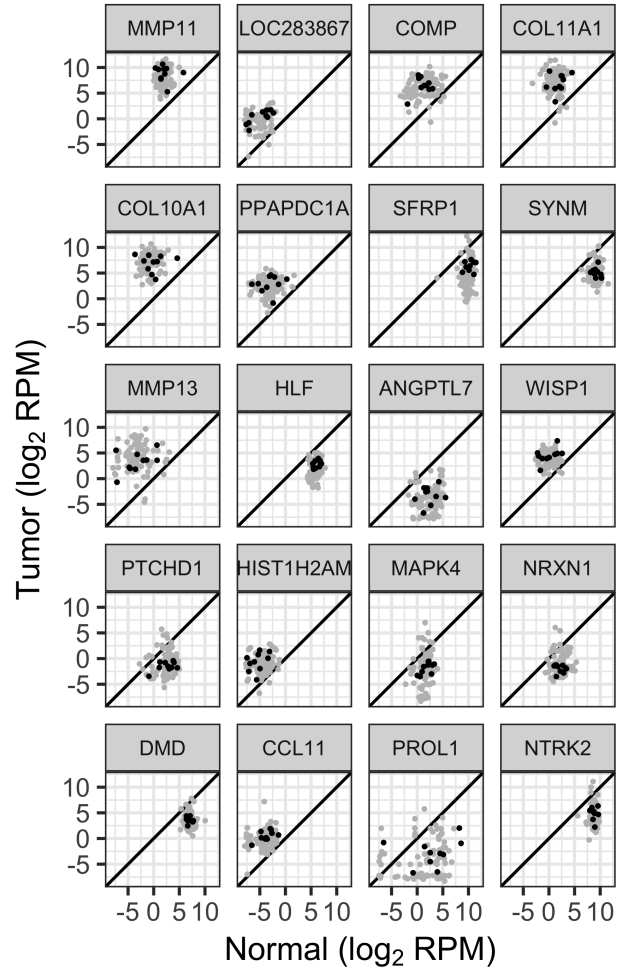
Changes are less consistent than top-ranked by *p*-value, but confidently large on average.

No ad-hoc filtering.

No threshold tuning.

Topconfects at 5% FDR, smaller number of samples





Sub-sample of 10 patients.

Confects much smaller than effects indicates lack of power.

Using `topconfects`

```
BiocManager::install("topconfects")
```

- As a further step in `limma` analysis:

```
# ...cpm(log=TRUE) or voom...
fit <- lmFit(y, design)
# ...possibly also contrasts.fit()...

limma_confects(fit, coef="thing_to_test", trend=TRUE)
```

- Also has functions to follow `edgeR` or `DESeq2` analysis (slower).

- Normal- or t-distributed errors:

```
normal_confects(estimates, standard_errors, df)
```

Thinking in terms of measuremment

What effect size best captures our interest?

How accurate is the measurement?

Top confident gene set enrichment at 5% FDR

GO BP Term	Effect (r)	Confect	n	n up	n down
chromosome organization	0.099	0.081	1045	528	291
DNA conformation change	0.087	0.075	250	157	49
system process	-0.087	-0.072	1576	384	832
DNA packaging	0.083	0.072	165	110	33
mitotic cell cycle process	0.084	0.067	769	398	226
nucleosome assembly	0.078	0.067	112	80	19
chromatin assembly	0.078	0.067	129	89	23
mitotic cell cycle	0.083	0.067	885	446	272
DNA metabolic process	0.081	0.067	924	490	242
nucleosome organization	0.077	0.066	139	97	24

(6506 genes up, 7719 down at 5% FDR. 9277 of 11218 gene sets significant at 5% FDR.)

Effect size:
LFC correlation with gene-set indicator variable
$= \pm\sqrt{}$ variance explained.

Measurement accuracy:
Bootstrap patients to estimate standard error.

Apply `normal_confects`.

n large, not much change in order using "confect" vs "effect".

Confects give us confidence that these discoveries are real.

Discover cancer related to cell division.

Discussion

- `topconfects` confidently ranks results by what you are interested in.
- No need for parameter tuning or ad-hoc filtering.

- False Coverage-statement Rate concept is versatile and under-used.
- To test new methods (eg Bayesian methods),
  check confident/credible regions control FCR under specified selection scenarios.

Acknowledgements: Traude Beilharz, David Powell, Andrew Pattison

Extra slides

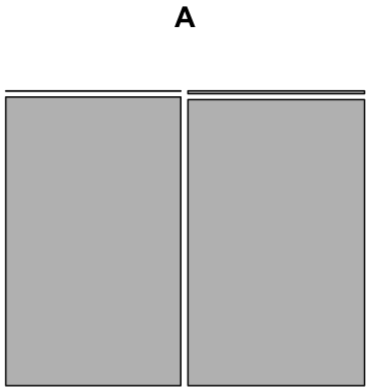
Measuring changes in ratio

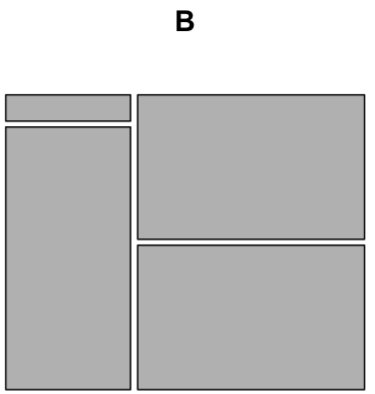
Compare:

> A. 10:10000 vs 100:10000
>
> B. 10:100 vs 100:100

- By log ratio of ratios, equally interesting.
- By difference of proportions, B is more interesting.

Would A or B have the smaller p-value,
with null hypothesis of equal ratios?





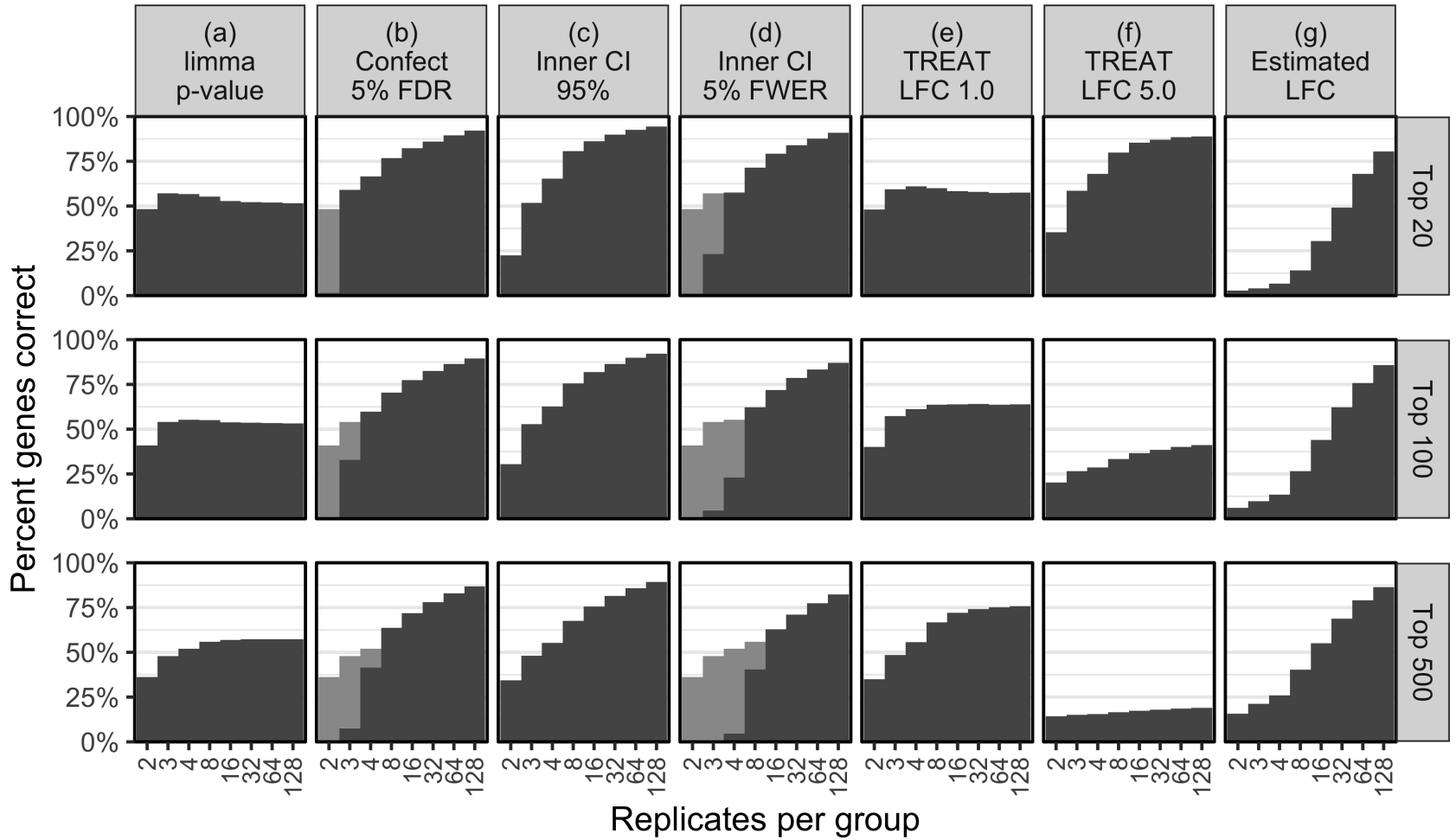
"Most significant" gene sets

Term	Effect (*r*)	FDR	n	n up	n down
response to heparin	-0.035	1.1e-280	6	1	5
negative regulation of peptide hormone secretion	-0.035	7.5e-262	39	10	22
DNA replication-independent nucleosome assembly	0.055	8.1e-259	48	40	4
protein localization to chromatin	0.030	4.0e-246	28	18	3
CENP-A containing nucleosome assembly	0.055	1.9e-243	37	33	2
positive regulation of mast cell chemotaxis	-0.026	1.7e-237	3	0	3
telomere capping	0.031	2.2e-234	47	27	14
telomere organization	0.063	3.7e-234	158	105	28
chromatin remodeling at centromere	0.055	7.1e-230	41	35	3
regulation of calcium-independent cell-cell adhesion	-0.027	7.1e-230	2	0	2

(6506 genes up, 7719 down at 5% FDR. 9277 of 11218 gene sets significant at 5% FDR.)

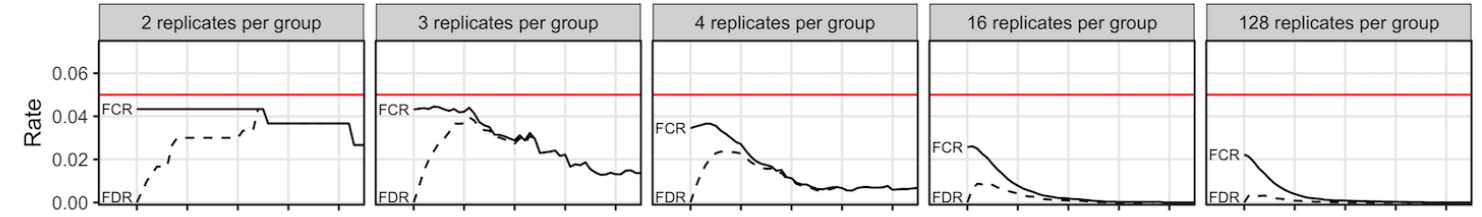
z test then ranking by "significance".

Simulation


(a) limma p-value
(b) Confect 5% FDR
(c) Inner CI 95%
(d) Inner CI 5% FWER
(e) TREAT LFC 1.0
(f) TREAT LFC 5.0
(g) Estimated LFC
Top 20
Top 100
Top 500
Percent genes correct
100%
75%
50%
25%
0%
2 3 4 8 16 32 64 128
Replicates per group

A

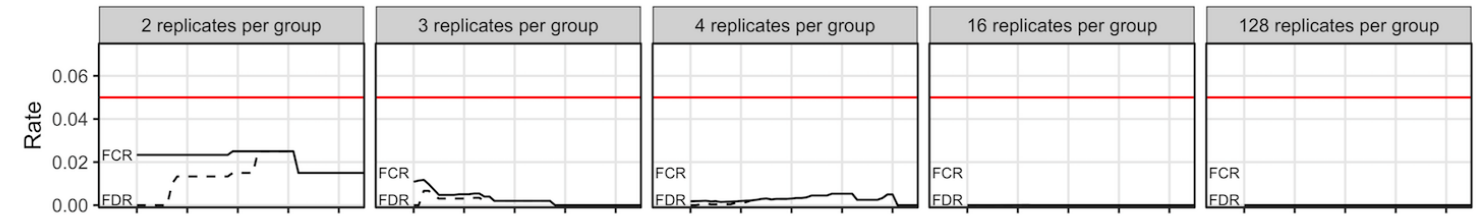
Confects at 5% FDR



	2 replicates per group					3 replicates per group					4 replicates per group					16 replicates per group					128 replicates per group				
LFC threshold	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
True genes	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61
Correct coverage	0.33	0.33	0.27	0.25	0.23	39	18	8.0	4.2	2.6	300	70	25	12	6.5	3702	481	126	51	26	9123	1155	262	93	44
False coverage	0.08	0.08	0.08	0.06	0.06	1.8	0.85	0.29	0.15	0.07	11	2.0	0.31	0.10	0.04	98	3.8	0.24	0.01	0.00	208	4.6	0.24	0.03	0.00

B

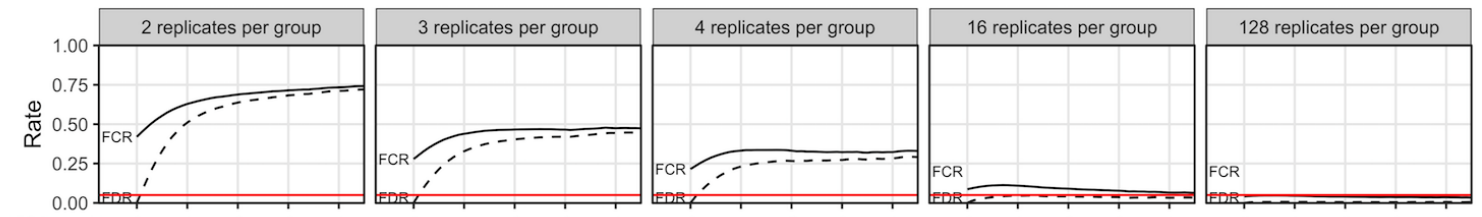
Inner bound of 5% FWER adjusted intervals



	2 replicates per group					3 replicates per group					4 replicates per group					16 replicates per group					128 replicates per group				
LFC threshold	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
True genes	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61
Correct coverage	0.22	0.18	0.16	0.16	0.14	4.8	3.3	2.3	1.7	1.3	24	13	7.7	5.1	3.6	963	218	80	38	20	5383	799	214	83	40
False coverage	0.03	0.03	0.03	0.03	0.02	0.06	0.03	0.01	0.00	0.00	0.05	0.03	0.02	0.02	0.01	0.03	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

C

Inner bound of 95% CI (note different y scale)



	2 replicates per group					3 replicates per group					4 replicates per group					16 replicates per group					128 replicates per group				
LFC threshold	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
True genes	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61	15000	2091	422	140	61
Correct coverage	842	182	66	31	18	1458	282	93	43	22	1972	355	111	48	25	4825	726	194	76	38	9348	1287	297	107	49
False coverage	608	308	146	79	49	563	220	81	37	20	540	177	55	23	12	460	89	19	6.4	2.7	408	61	12	4.1	1.8