GREAT LAKES BIOENERGY RESEARCH CENTER

Making sense of RNA-Seq data: from low-level processing to functional analysis

Oleg Moskvin, Sean McIlliwain, Irene Ong

CAMDA, Boston, July 11, 2014

U.S. DEPARTMENT OF ENERGY

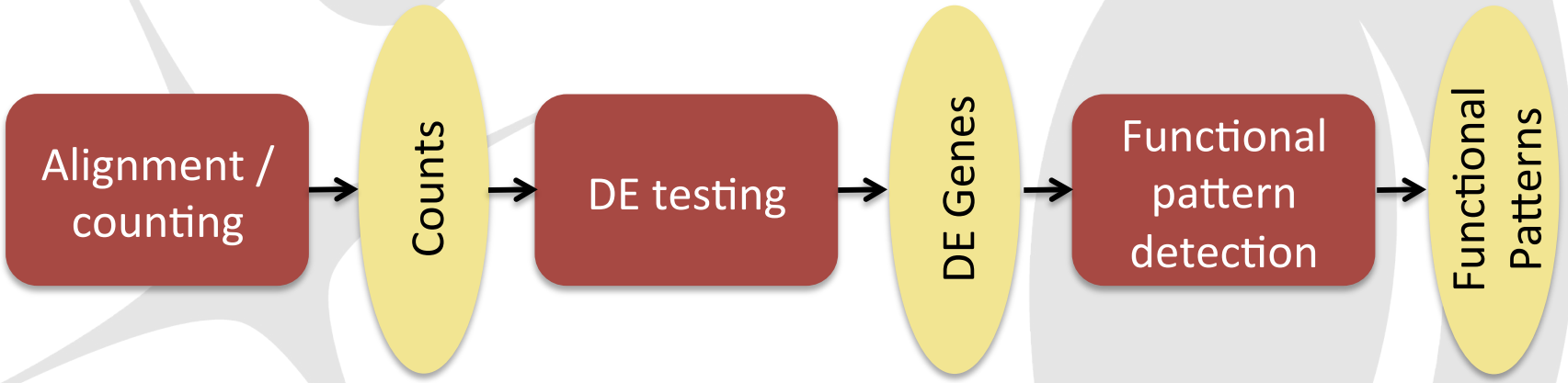
DOE Bioenergy Research Centers

Outline

- From counts to functional patterns: impact of different layers of RNA-Seq analysis
  - Low-level processing
  - Testing for differential expression
  - Functional summaries of the transcriptional responses
  - Scanning the analysis options across 3 layers:

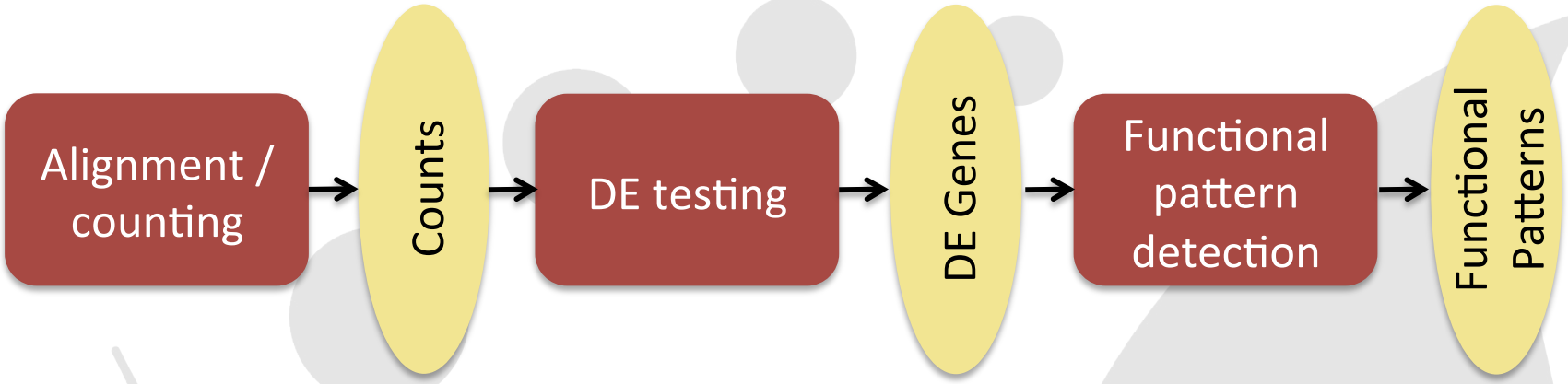
    what 3,456 100+-mer vectors can tell us
- Improving functional analysis via continuous dialogue between biology and statistics
  - Starting from differential expression: unfolding potential of a 3-condition model
  - Boosting interpretability of multifactorial experiments

Generalized RNA-Seq analysis pipeline



Omitted: count normalization, gene clustering….

Relative impact of processing stages on the final result?



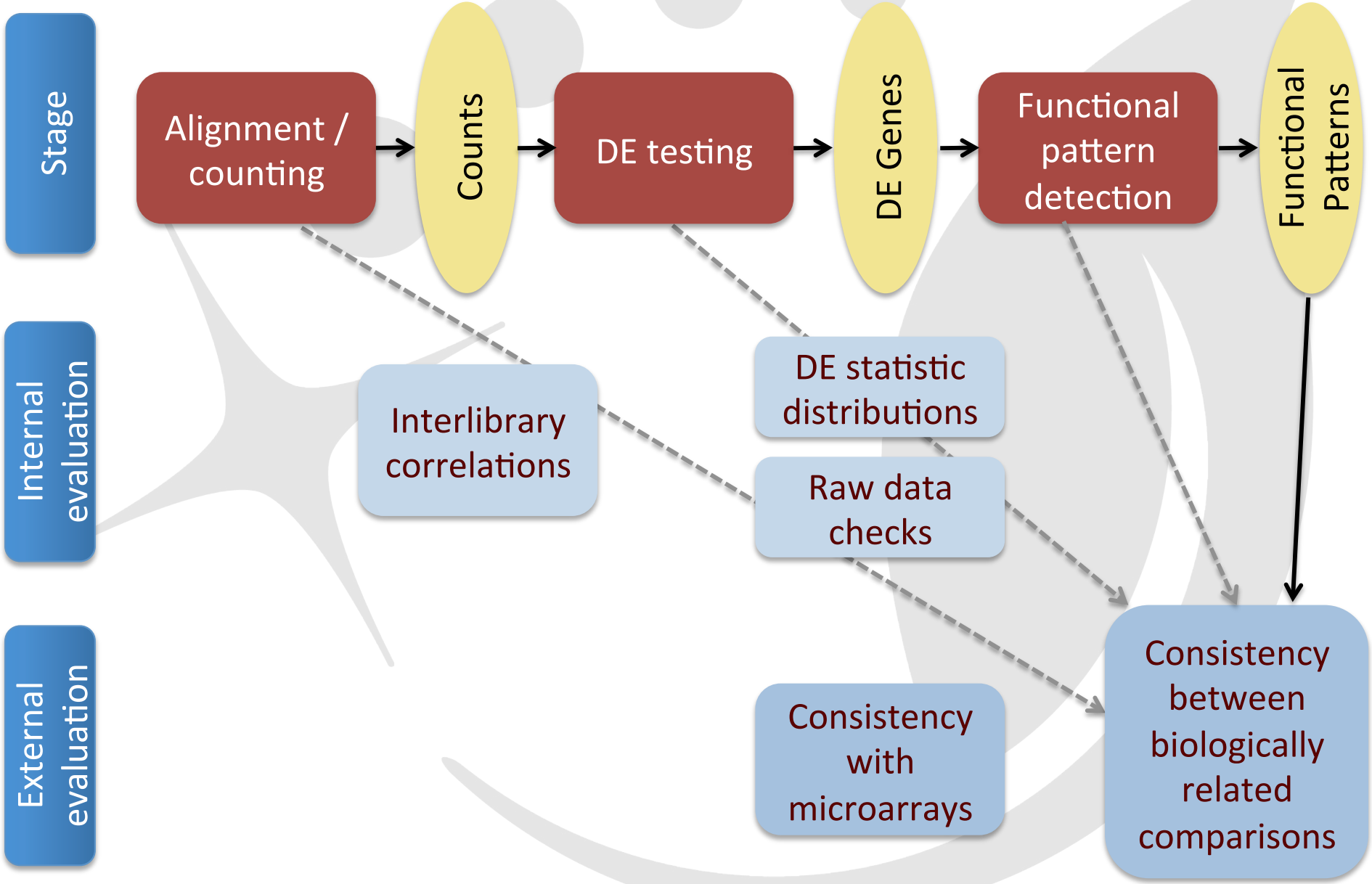
1. “Lever” Hypothesis

Image source: http://etc.usf.edu/clipart/35900/35944/lever_35944_md.gif

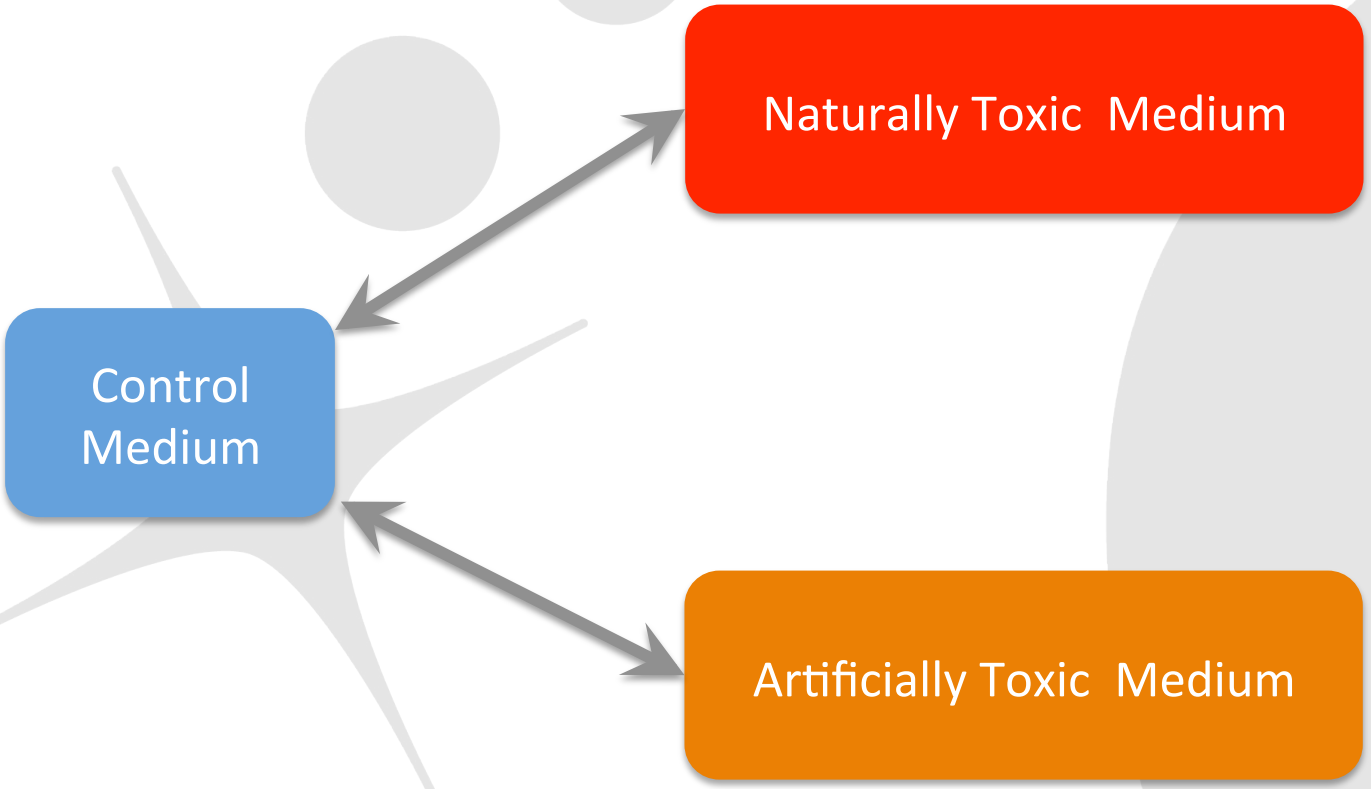
2. “Flashlight” Hypothesis

http://www.picture-newsletter.com/scuba-diving/diver-flashlight-7wi2.jpg

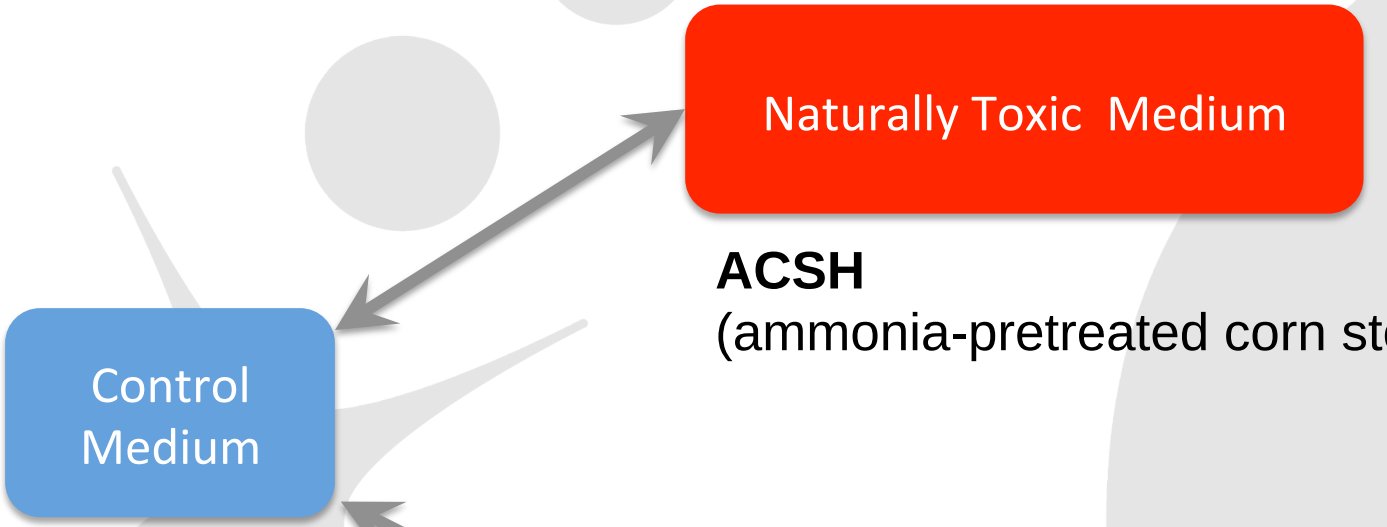
Multilayer evaluation of RNA-Seq analysis pipeline



The biological case



The biological case



ACSH
(ammonia-pretreated corn stover hydrolysate)

SynH
("synthetic hydrolysate"
without toxic compounds)

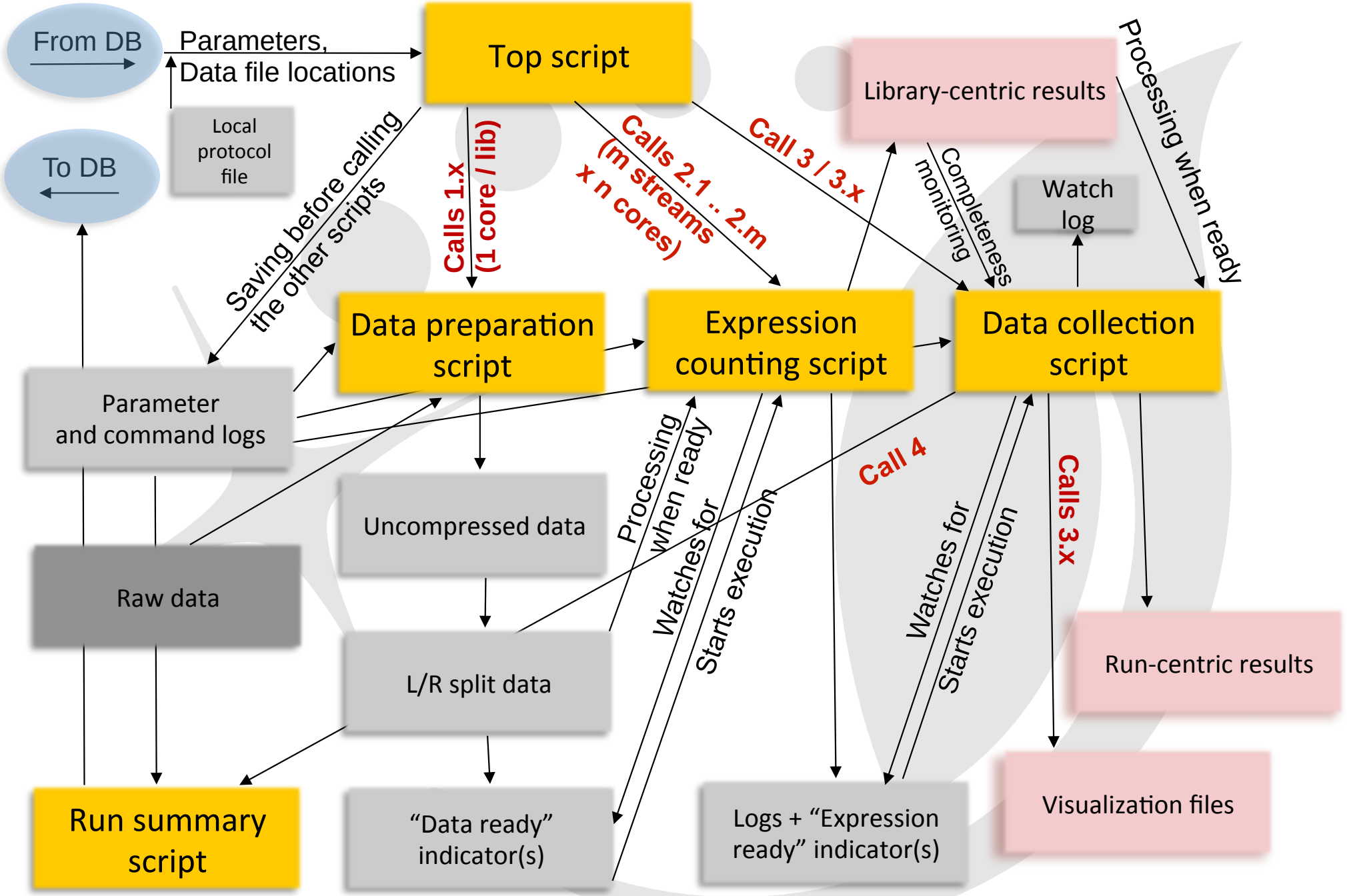
SynH-LT
("synthetic hydrolysate"
with a cocktail of toxic compounds
discovered in ACSH - "lignotoxins" – added)

Method scanning overview

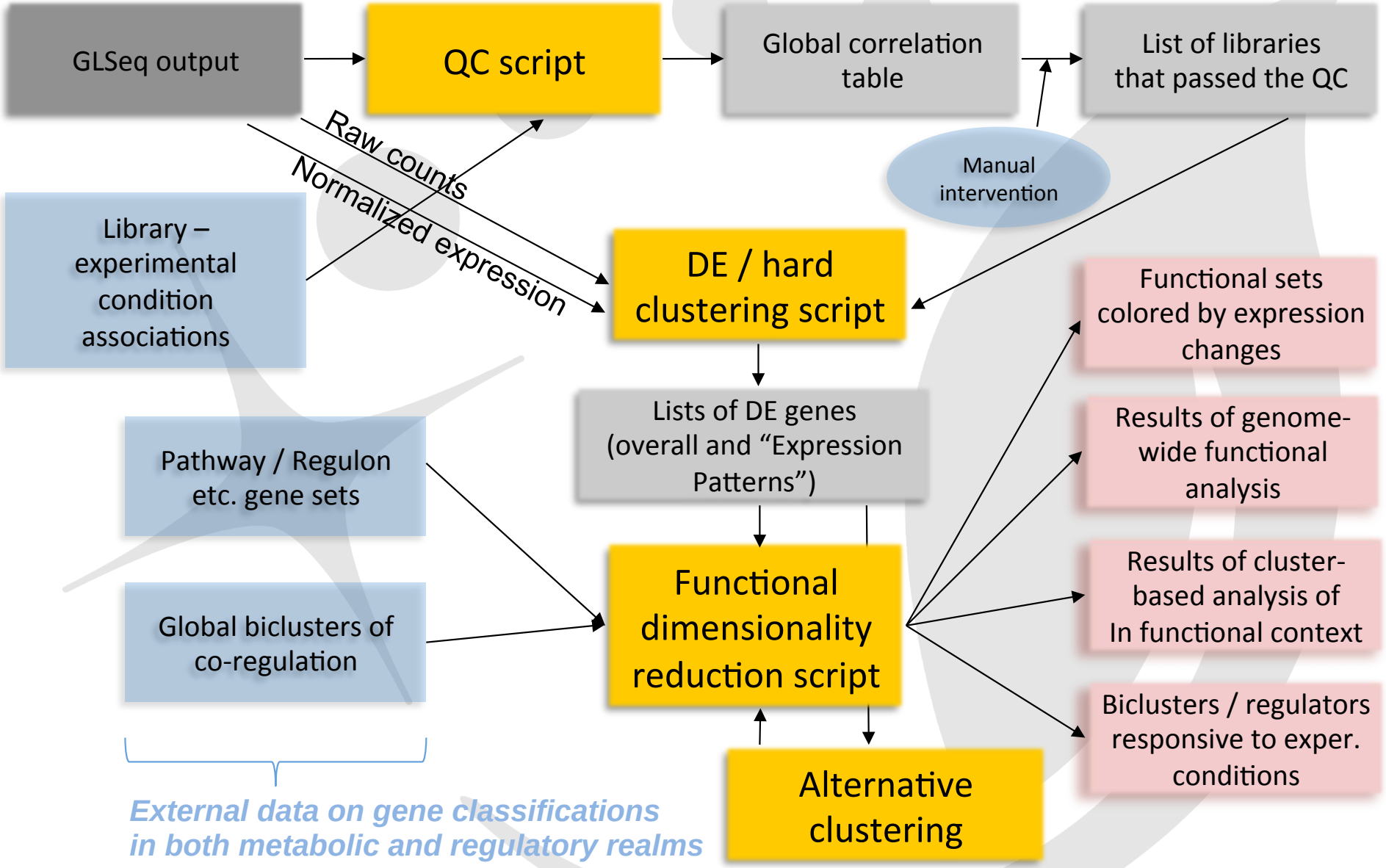
2 read pre-processing strategies
x 2 alignment / counting pipelines
x 3 count normalization methods
x 2 pairwise comparisons
x 3 time points
x 4 DE calling methods
x 2 FPKM-based filtering options
x 3 functional overrepresentation strategies
x 4 geneset types

= 6,912 ranked lists of functional enrichment results
(~1M of individual gene set enrichment p-values)
(+576 comparisons of DE lists with microarray data)

The *in silico* toolset


From DB
Parameters,
Data file locations
Top script
Library-centric results
Processing when ready
To DB
Local protocol file
Saving before calling the other scripts
Calls 1.x (1 core / lib)
Calls 2.1 .. 2.m (m streams x n cores)
Call 3 / 3.x
Completeness monitoring
Watch log
Data preparation script
Expression counting script
Data collection script
Parameter and command logs
Uncompressed data
Processing when ready
Watches for
Starts execution
Call 4
Calls 3.x
Raw data
L/R split data
Run-centric results
Run summary script
"Data ready" indicator(s)
Logs + "Expression ready" indicator(s)
Visualization files

From Genes To Blocks of Biological Response (BBR): Downstream Analysis Assistant (DAA) pipeline



The Low Level: alignments / counting

Read pre-processing options

A. Take reads as they come from the sequencing facility - “RAW” (read length=100)

A. Remove 12 nt from 5’ end and any number of nucleotides from 3’-end that have the average quality score < 30 in a 3-nt sliding window; keep the read if 36 or more nt are left – “QC”

Alignment / counting options

A. Align to genome with BWA, then estimate coverage with HTSeq

B. Align to transcriptome with Bowtie, then estimate coverage probabilistically with RSEM (Li and Dewey, 2011)

Count-based correlations between libraries

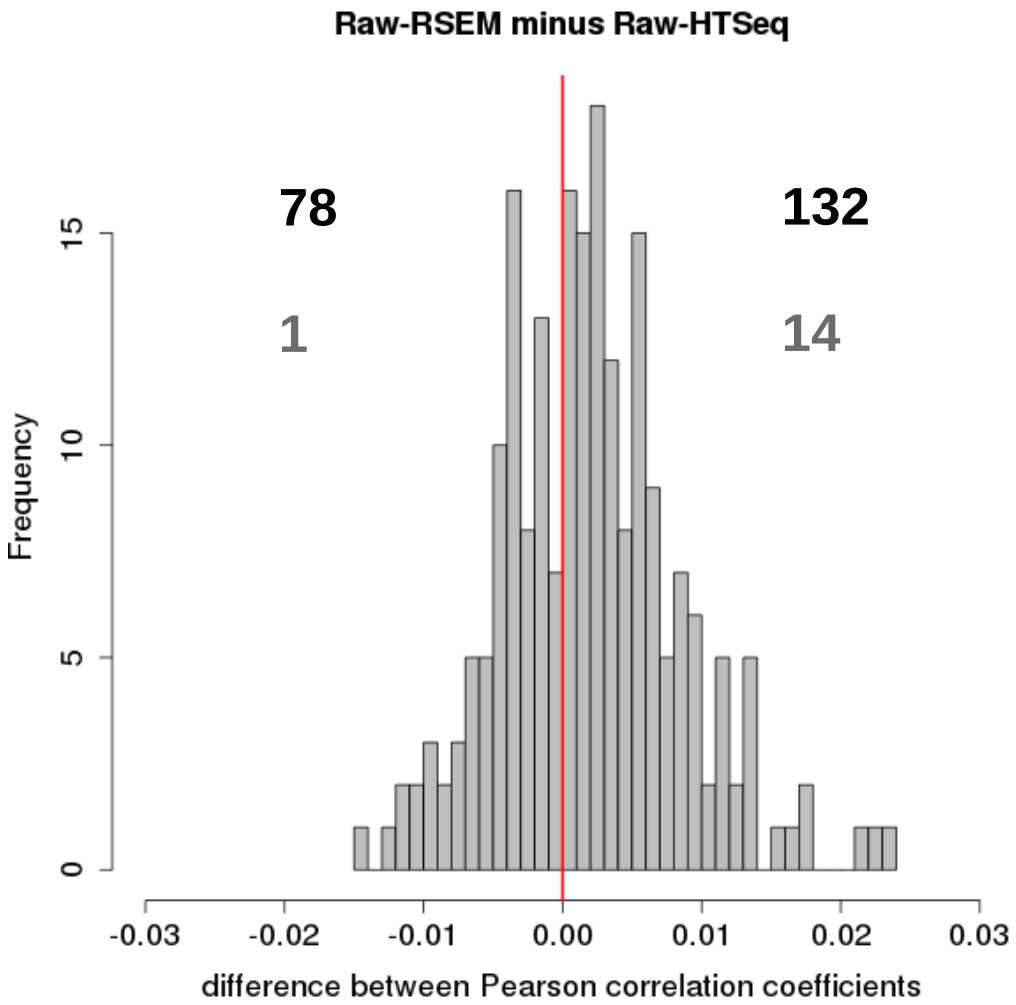
RSEM (Transcriptome alignment; probabilistic counting)

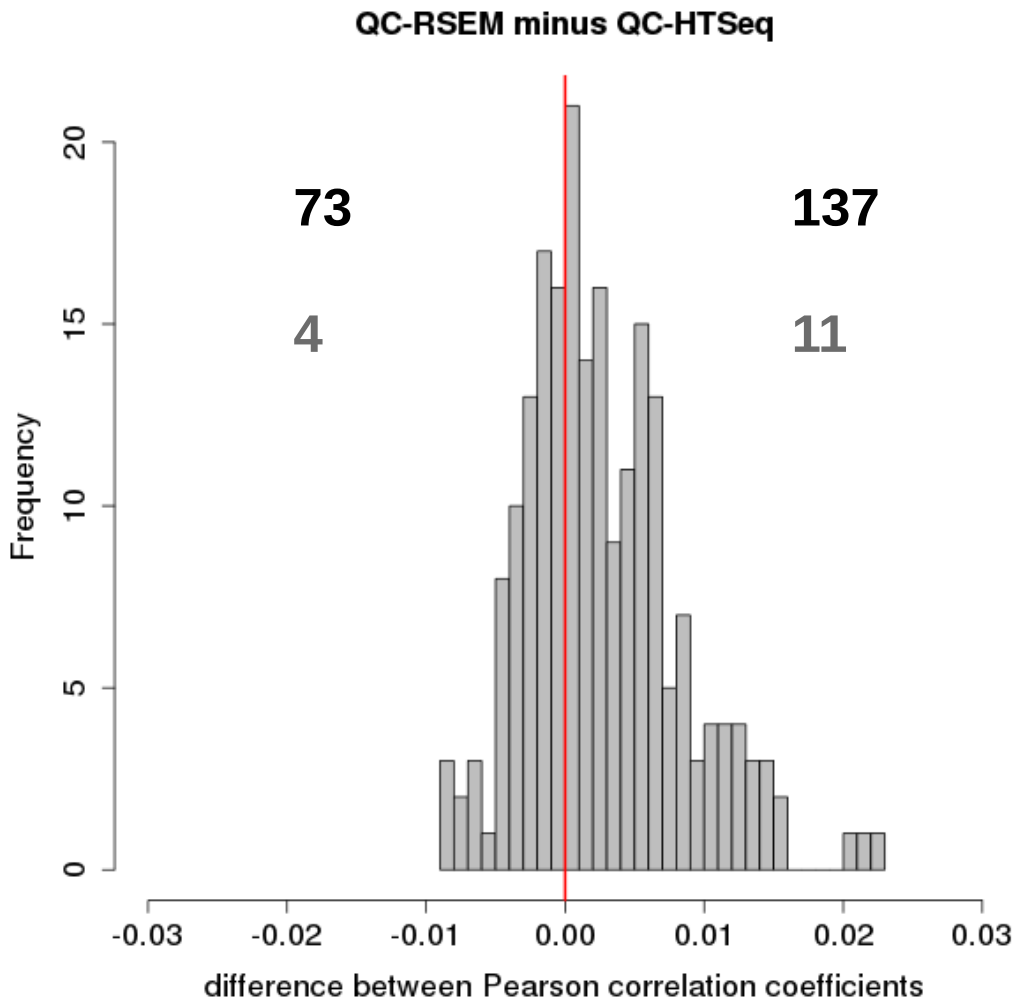
	c1.synHv3.T2	c1.synHv3.T2	c1.synHv3.T3	c1.synHv3.T3	c1.synHv3.T4	c1.synHv3.T4	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c3.ACSH.T2_	c3.ACSH.T2_	c3.ACSH.T3_	c3.ACSH.T3_	c3.ACSH.T4_	c3.ACSH.T4_
c1.synHv3.T2_BXHU	1.000																				
c1.synHv3.T2_CBYS	0.967	1.000																			
c1.synHv3.T3_CUNZ	0.951	0.973	1.000																		
c1.synHv3.T3_CUOB	0.958	0.976	0.998	1.000																	
c1.synHv3.T4_CBYN	0.934	0.950	0.914	0.910	1.000																
c1.synHv3.T4_CUOC	0.884	0.946	0.966	0.963	0.861	1.000															
c2.synHv3Toxin.T2_BXNX	0.821	0.793	0.745	0.759	0.787	0.646	1.000														
c2.synHv3Toxin.T2_CBYU	0.771	0.763	0.677	0.680	0.828	0.562	0.921	1.000													
c2.synHv3Toxin.T2_CCOG	0.783	0.788	0.705	0.706	0.829	0.600	0.908	0.989	1.000												
c2.synHv3Toxin.T3_BXNC	0.898	0.898	0.824	0.828	0.947	0.739	0.880	0.936	0.929	1.000											
c2.synHv3Toxin.T3_BXXC	0.854	0.867	0.785	0.784	0.907	0.689	0.854	0.951	0.962	0.973	1.000										
c2.synHv3Toxin.T3_CBYW	0.838	0.845	0.766	0.761	0.933	0.668	0.792	0.928	0.926	0.969	0.979	1.000									
c2.synHv3Toxin.T4_BXNB	0.884	0.858	0.803	0.804	0.961	0.705	0.796	0.865	0.840	0.953	0.906	0.951	1.000								
c2.synHv3Toxin.T4_BXNZ	0.754	0.704	0.642	0.637	0.848	0.488	0.736	0.881	0.853	0.898	0.900	0.947	0.940	1.000							
c2.synHv3Toxin.T4_CHBW	0.893	0.889	0.841	0.836	0.980	0.752	0.780	0.870	0.859	0.958	0.932	0.972	0.987	0.932	1.000						
c3.ACSH.T2_CIYS	0.705	0.601	0.564	0.589	0.666	0.447	0.784	0.663	0.605	0.712	0.605	0.611	0.770	0.683	0.683	1.000					
c3.ACSH.T2_CNFT	0.786	0.719	0.674	0.701	0.728	0.583	0.880	0.738	0.705	0.779	0.694	0.661	0.778	0.660	0.714	0.961	1.000				
c3.ACSH.T3_CIYT	0.529	0.384	0.412	0.421	0.521	0.273	0.492	0.414	0.330	0.501	0.365	0.449	0.664	0.620	0.571	0.851	0.698	1.000			
c3.ACSH.T3_CNFU	0.560	0.413	0.451	0.458	0.537	0.306	0.514	0.427	0.348	0.514	0.384	0.461	0.671	0.629	0.584	0.848	0.705	0.996	1.000		
c3.ACSH.T4_CNFW	0.454	0.303	0.379	0.377	0.434	0.238	0.359	0.283	0.207	0.372	0.251	0.346	0.556	0.531	0.482	0.699	0.530	0.957	0.965	1.000	
c3.ACSH.T4_CNGG	0.491	0.348	0.371	0.383	0.486	0.243	0.469	0.379	0.290	0.464	0.321	0.405	0.632	0.575	0.531	0.858	0.702	0.993	0.984	0.940	1.000

BWA-HTSeq (Genome alignment; hard-threshold counting)

	c1.synHv3.T2	c1.synHv3.T2	c1.synHv3.T3	c1.synHv3.T3	c1.synHv3.T4	c1.synHv3.T4	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c2.synHv3To	c3.ACSH.T2_	c3.ACSH.T2_	c3.ACSH.T3_	c3.ACSH.T3_	c3.ACSH.T4_	c3.ACSH.T4_
c1.synHv3.T2_BXHU	1.000																				
c1.synHv3.T2_CBYS	0.963	1.000																			
c1.synHv3.T3_CUNZ	0.948	0.973	1.000																		
c1.synHv3.T3_CUOB	0.956	0.975	0.998	1.000																	
c1.synHv3.T4_CBYN	0.926	0.950	0.918	0.913	1.000																
c1.synHv3.T4_CUOC	0.875	0.945	0.961	0.958	0.866	1.000															
c2.synHv3Toxin.T2_BXNX	0.813	0.779	0.738	0.752	0.764	0.632	1.000														
c2.synHv3Toxin.T2_CBYU	0.769	0.759	0.680	0.683	0.814	0.561	0.921	1.000													
c2.synHv3Toxin.T2_CCOG	0.780	0.783	0.706	0.707	0.814	0.596	0.909	0.989	1.000												
c2.synHv3Toxin.T3_BXNC	0.894	0.894	0.824	0.827	0.936	0.734	0.870	0.932	0.925	1.000											
c2.synHv3Toxin.T3_BXXC	0.854	0.866	0.789	0.789	0.898	0.689	0.850	0.948	0.958	0.972	1.000										
c2.synHv3Toxin.T3_CBYW	0.837	0.849	0.775	0.770	0.931	0.675	0.778	0.920	0.917	0.966	0.976	1.000									
c2.synHv3Toxin.T4_BXNB	0.878	0.857	0.807	0.807	0.959	0.706	0.776	0.853	0.829	0.947	0.901	0.950	1.000								
c2.synHv3Toxin.T4_BXNZ	0.752	0.703	0.649	0.644	0.841	0.489	0.721	0.869	0.842	0.893	0.895	0.942	0.938	1.000							
c2.synHv3Toxin.T4_CHBW	0.886	0.888	0.844	0.838	0.978	0.754	0.758	0.857	0.847	0.951	0.926	0.972	0.985	0.929	1.000						
c3.ACSH.T2_CIYS	0.702	0.595	0.564	0.588	0.655	0.438	0.778	0.664	0.609	0.710	0.608	0.609	0.764	0.684	0.676	1.000					
c3.ACSH.T2_CNFT	0.780	0.708	0.668	0.694	0.713	0.568	0.873	0.739	0.708	0.774	0.695	0.657	0.768	0.660	0.702	0.961	1.000				
c3.ACSH.T3_CIYT	0.525	0.382	0.414	0.422	0.520	0.268	0.481	0.412	0.330	0.502	0.367	0.450	0.664	0.627	0.572	0.844	0.692	1.000			
c3.ACSH.T3_CNFU	0.557	0.412	0.454	0.461	0.537	0.300	0.507	0.429	0.351	0.517	0.388	0.464	0.670	0.637	0.585	0.844	0.701	0.995	1.000		
c3.ACSH.T4_CNFW	0.449	0.303	0.382	0.379	0.439	0.234	0.351	0.283	0.209	0.372	0.254	0.350	0.556	0.539	0.486	0.693	0.525	0.956	0.963	1.000	
c3.ACSH.T4_CNGG	0.485	0.346	0.373	0.384	0.488	0.240	0.457	0.377	0.291	0.463	0.323	0.407	0.633	0.581	0.534	0.850	0.694	0.992	0.981	0.940	1.000

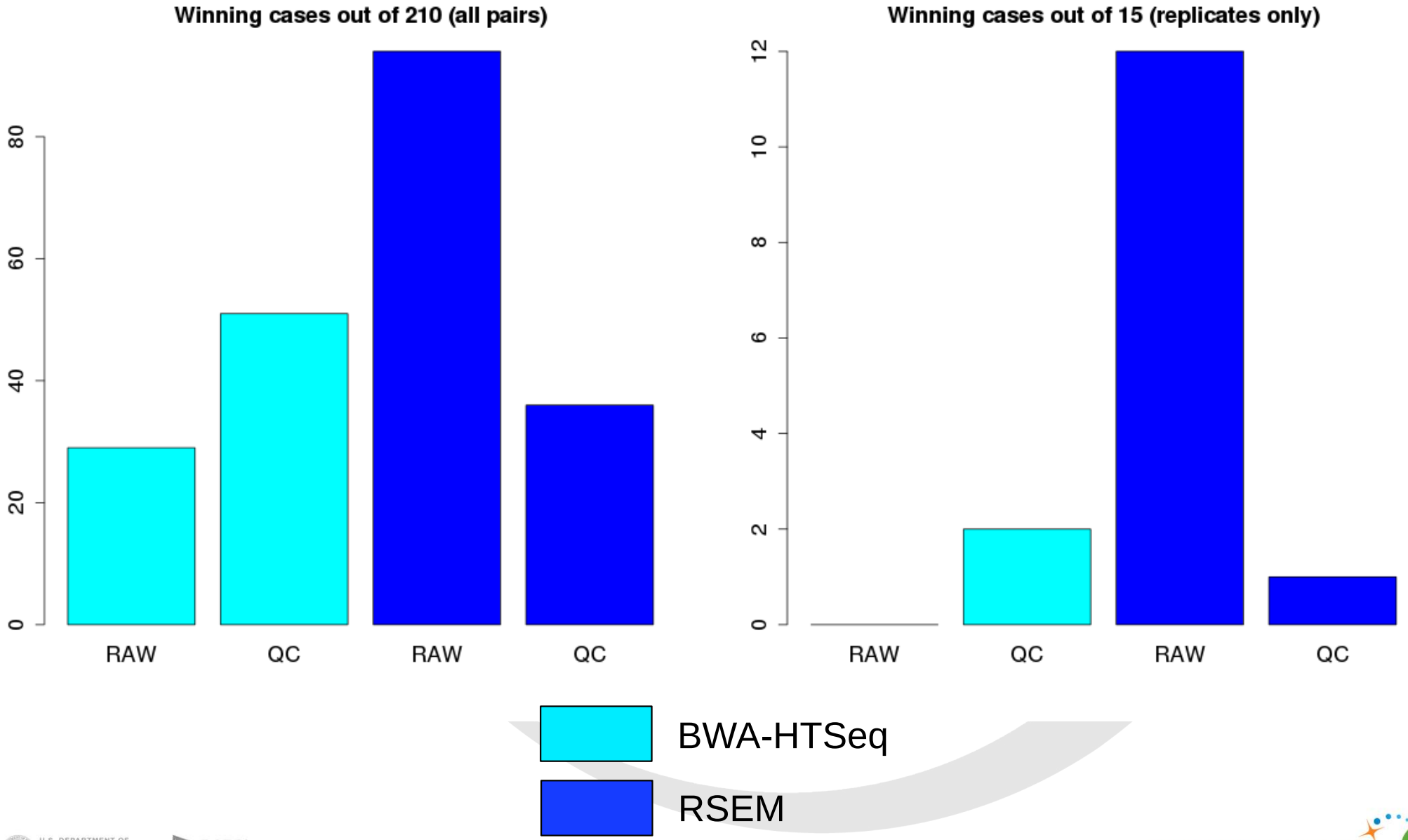
The advantage of probabilistic alignment / counting: RSEM vs. BWA-HTSeq





Black numbers – correlation pairs from overall pool
Green numbers – correlation pairs representing biological replicates

Ranking of the preprocessing – alignment / counting combinations by number of individual highest correlations

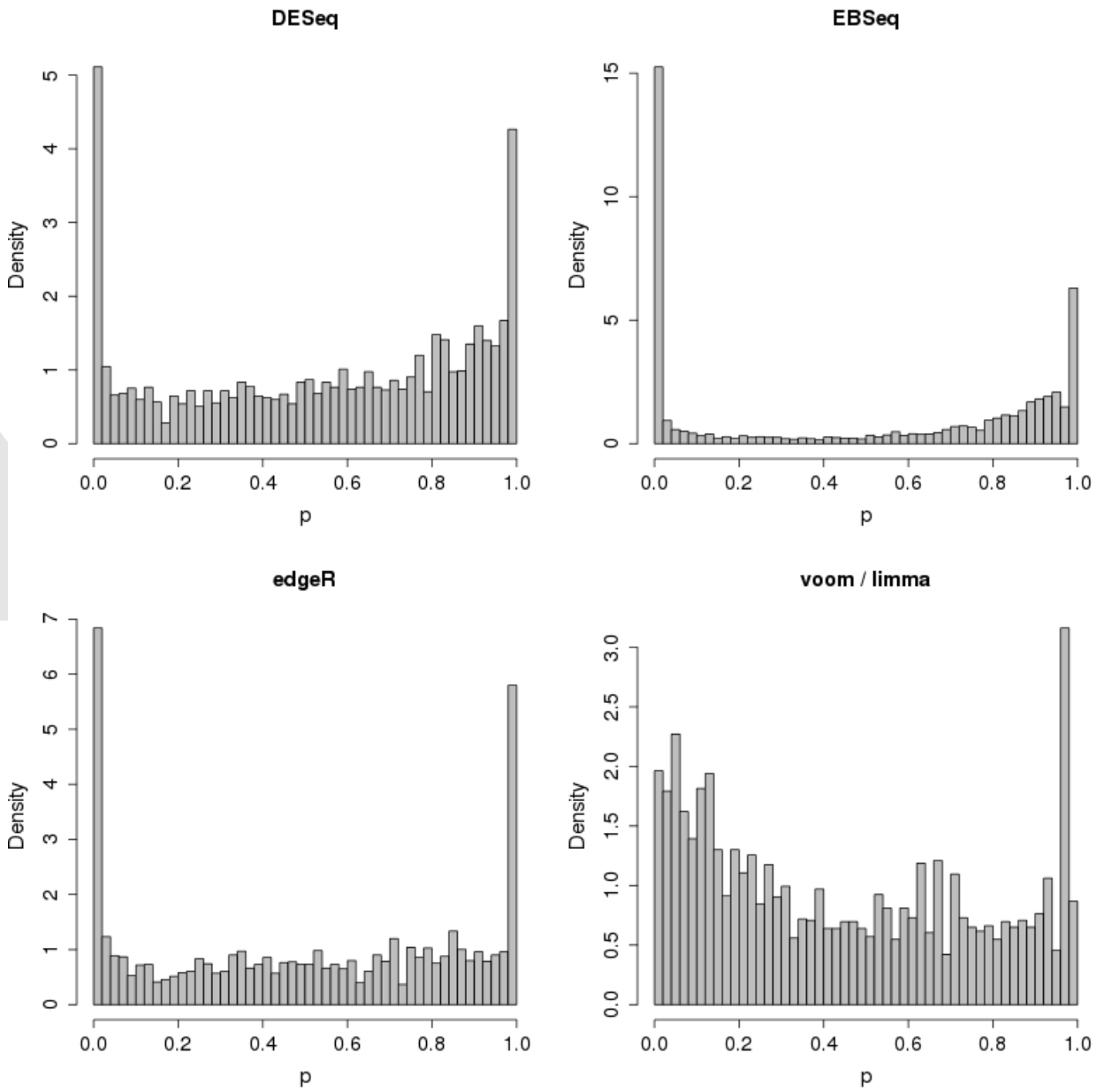


Running conclusions: low-level processing

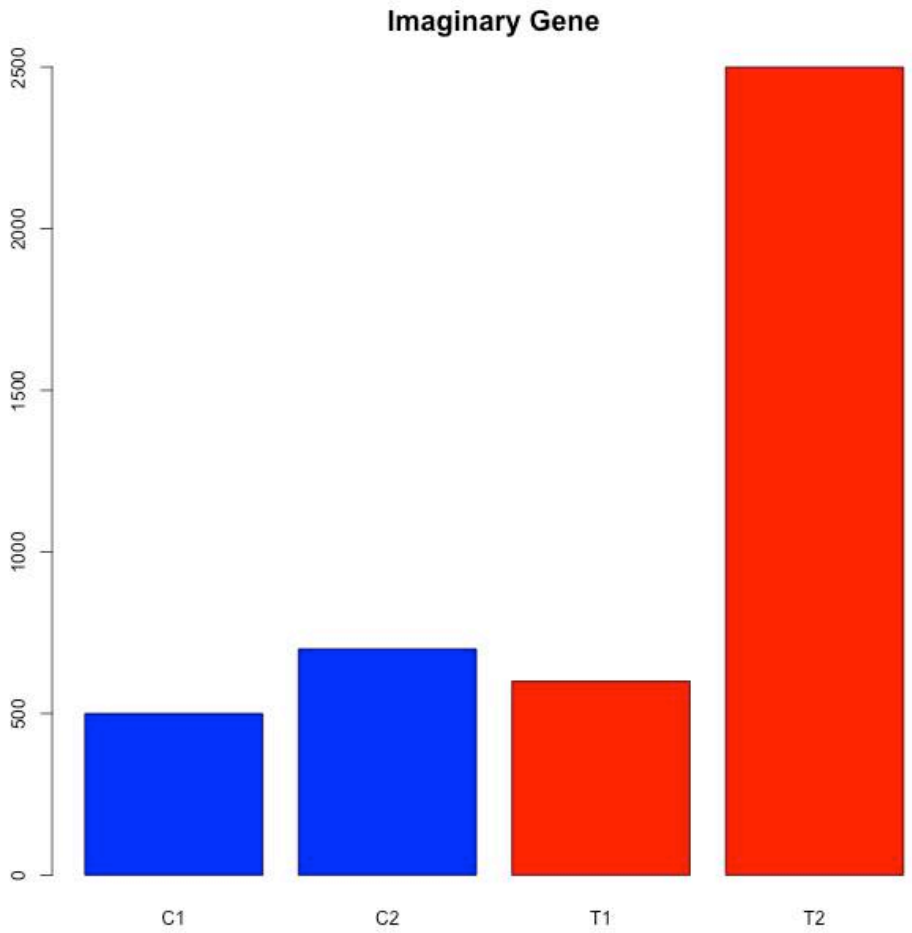
- Probabilistic counting with RSEM is more robust overall, compared to traditional counting
- Pre-processing of the reads may be either advantageous or disadvantageous, depending on the fragile balance between improving the overall nucleotide calling quality and keeping the extra sequence information

Differential expression

Not all DE detection methods are created equal

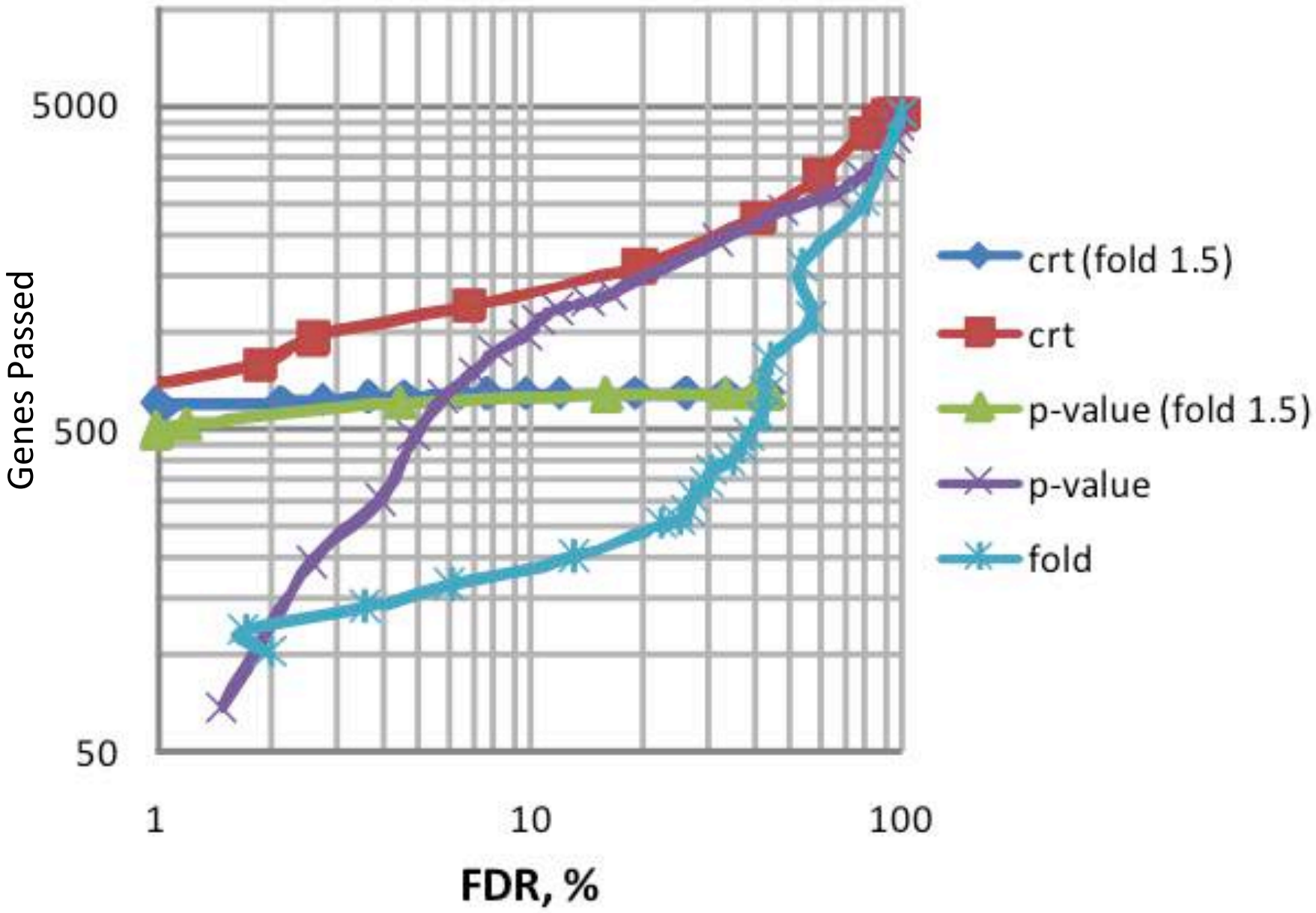


The case for “critical coefficient”

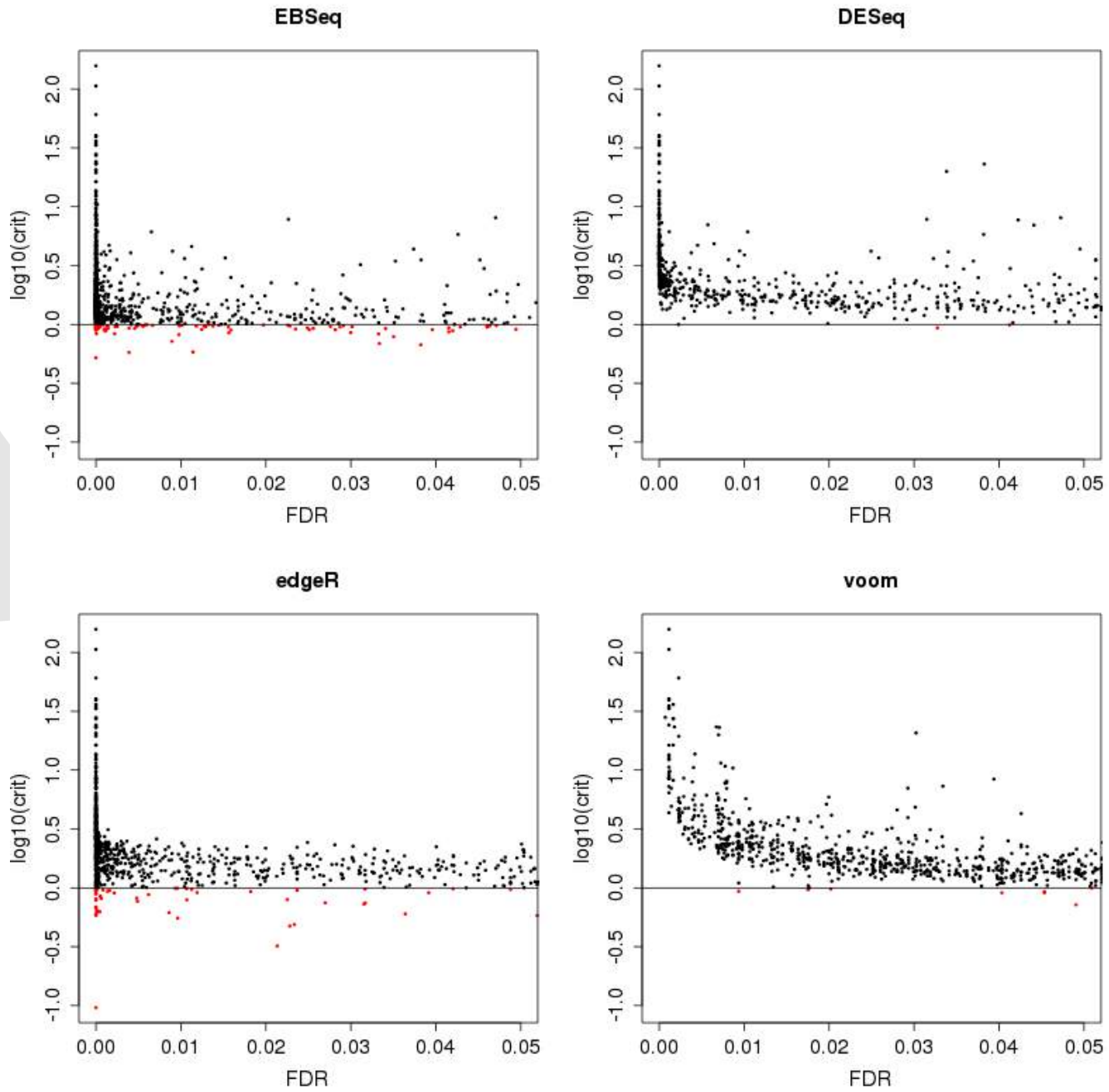


- *Would we investigate why the treatment effect looks so different with replicate 1 and replicate 2?*
- *Maybe… Next time…*

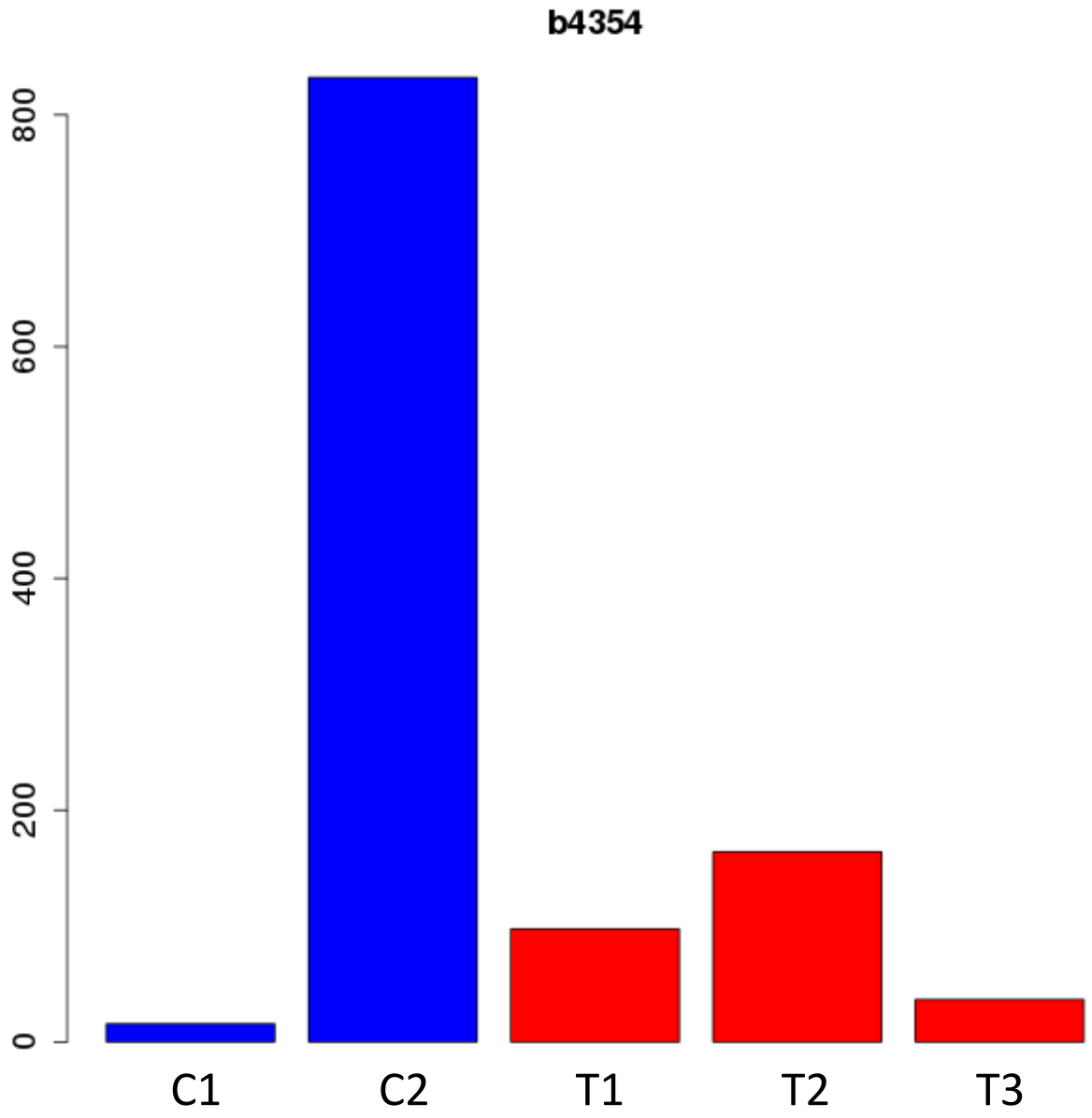
The case for “critical coefficient”



“Critical coefficient” in RNA-Seq realm



Curious preference

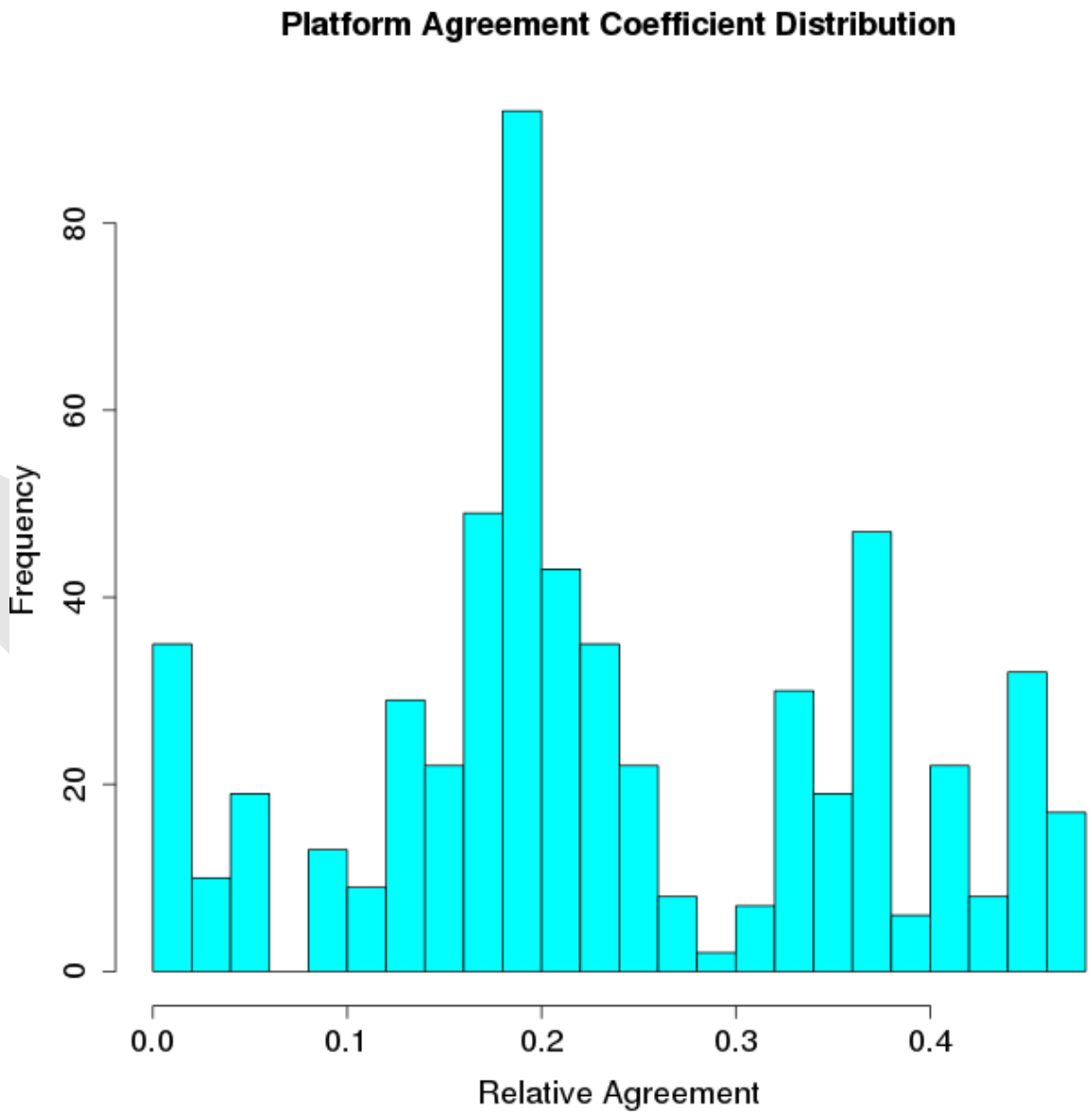


EBSeq: 0.87
DESeq: 0.60
edgeR: 0.0000082
voom: 0.97

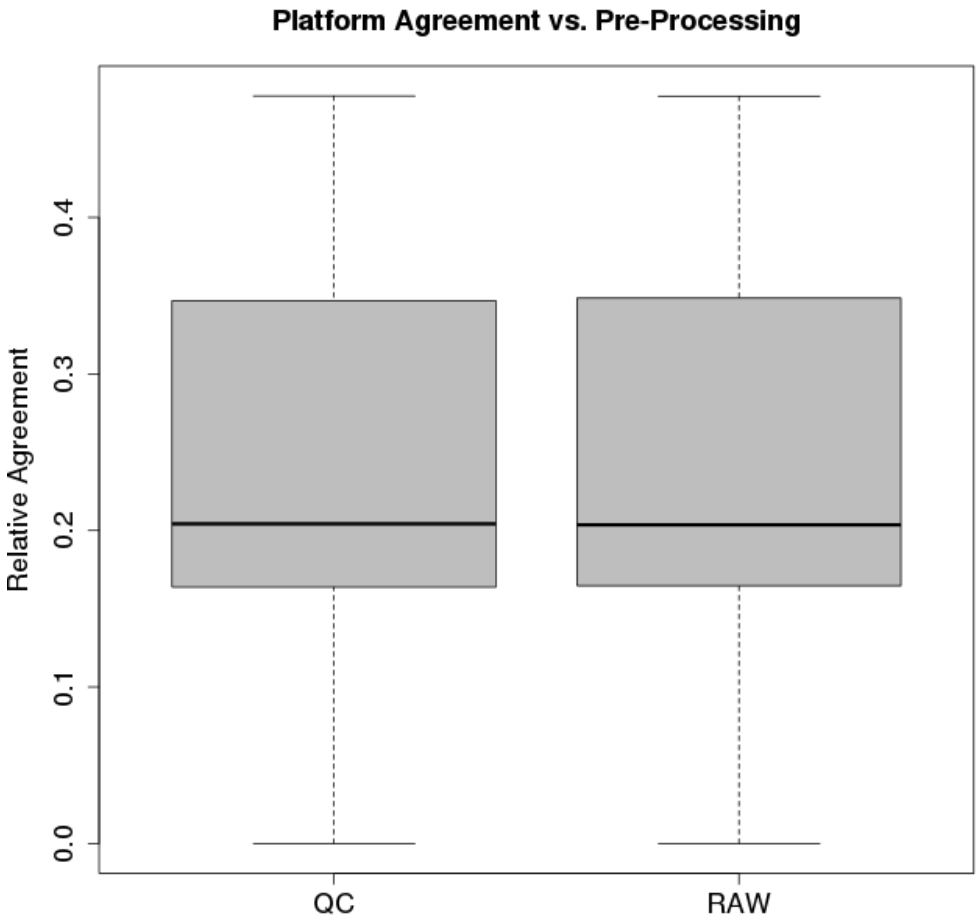
Agreement between RNA-Seq and microarray results

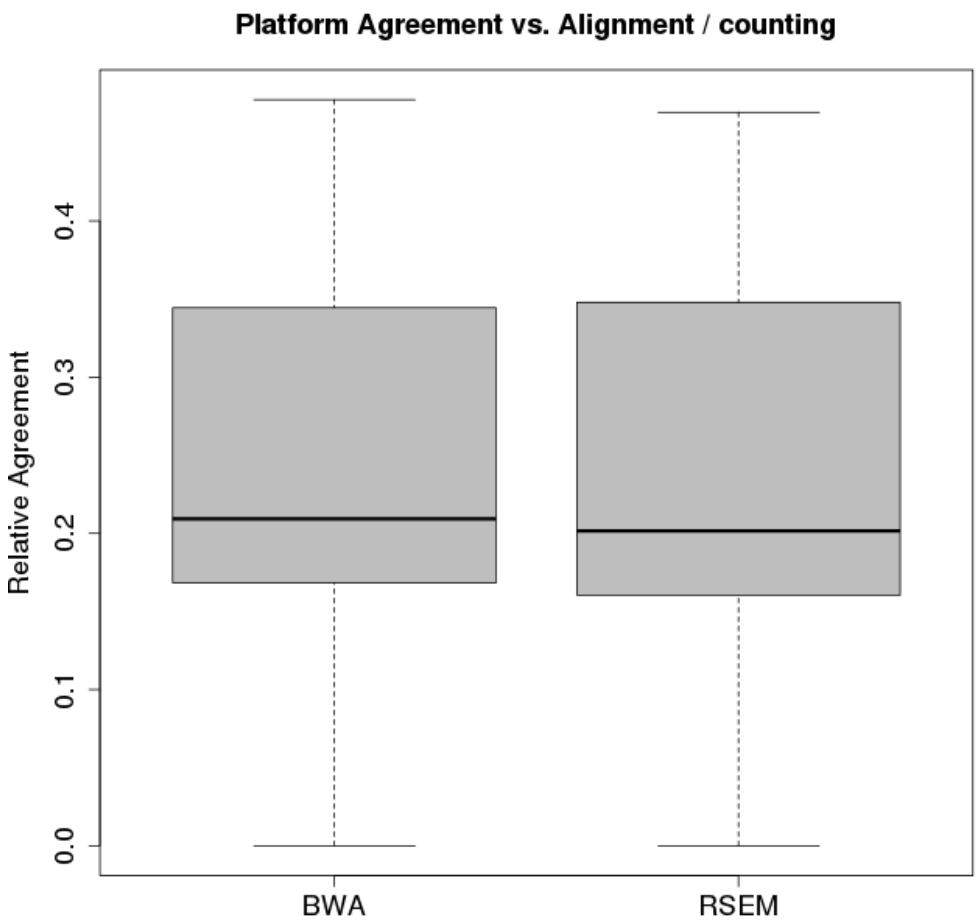
$$\textbf{Relative intersection} = \frac{\text{Number of genes called DE by both platforms (same directionality of change)}}{\text{Size of the pool of genes called DE by any platform}}$$

Agreement between RNA-Seq and microarray results: the range across 576 comparisons

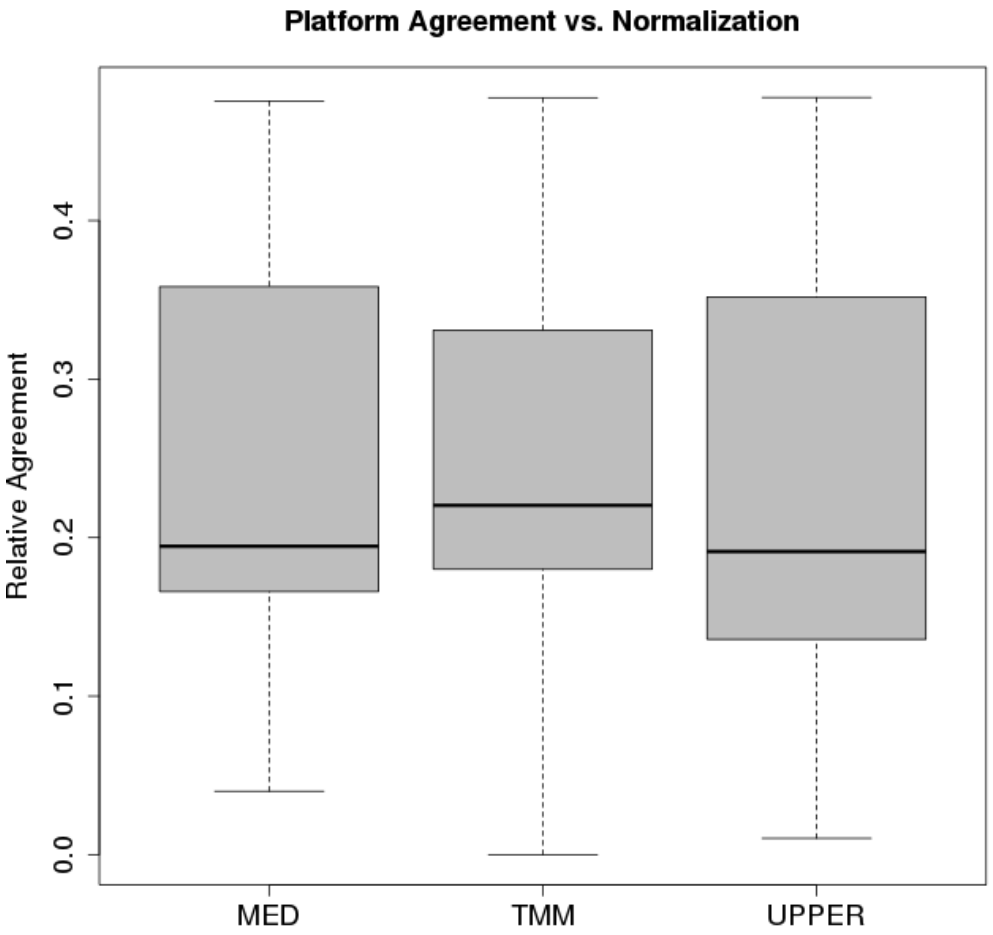


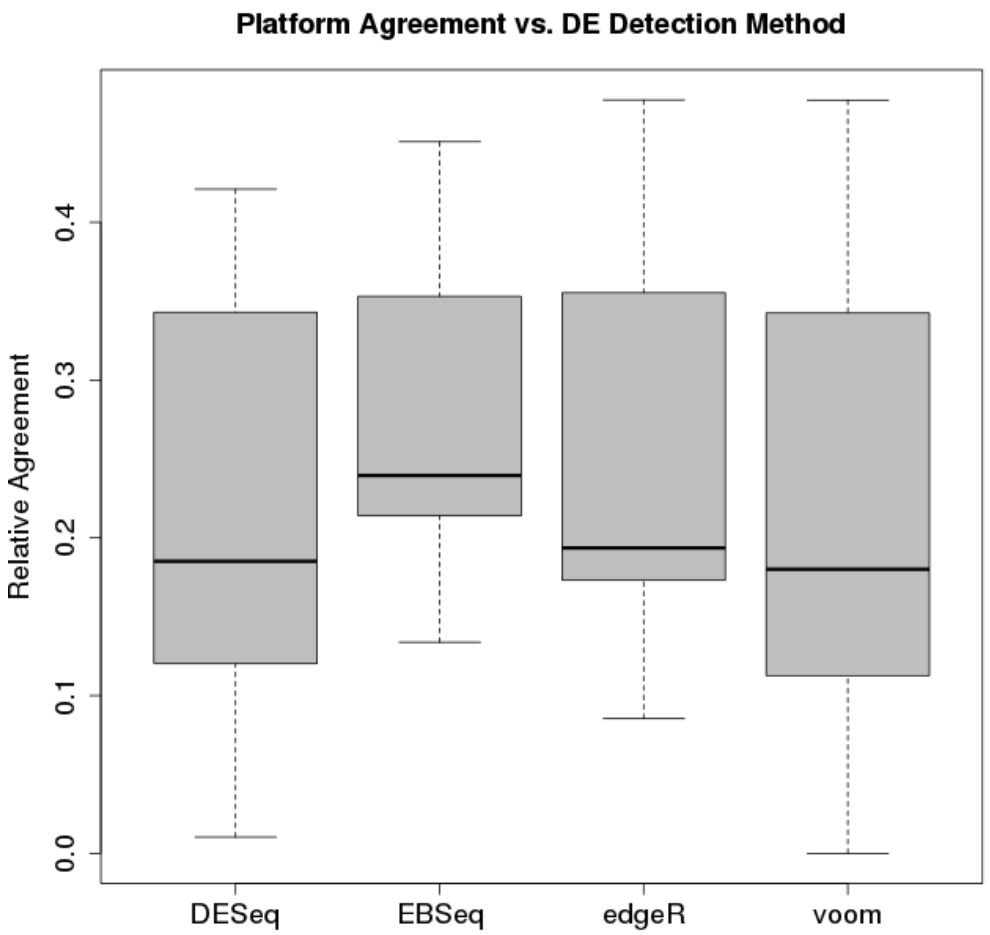
Agreement between RNA-Seq and microarray results: Influence of low-level processing



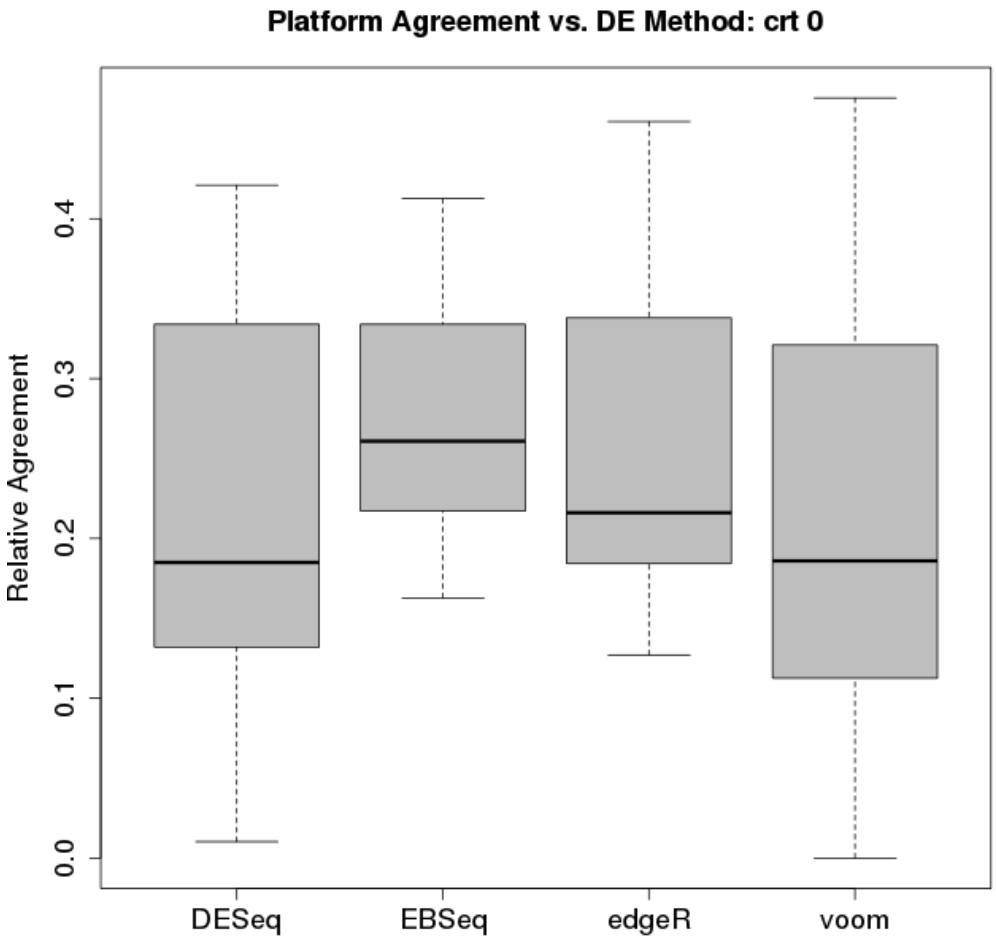


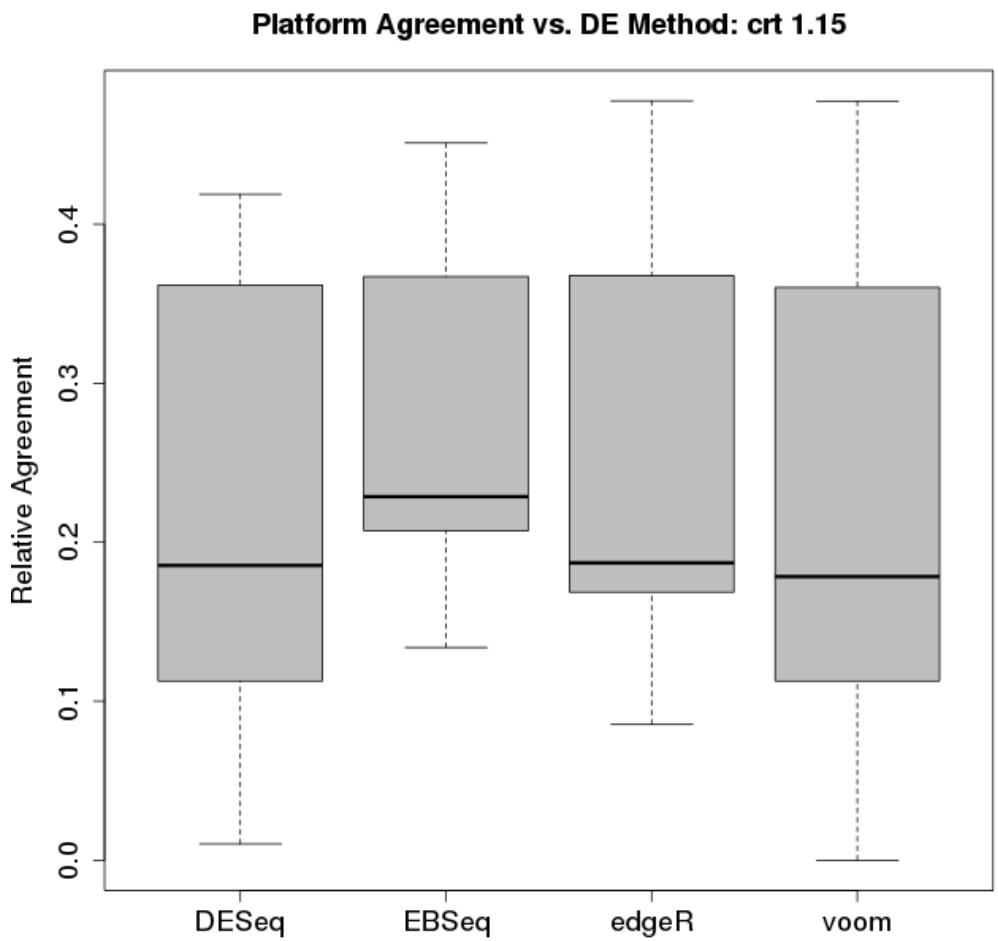
Agreement between RNA-Seq and microarray results: Influence of differential expression methodology - 1



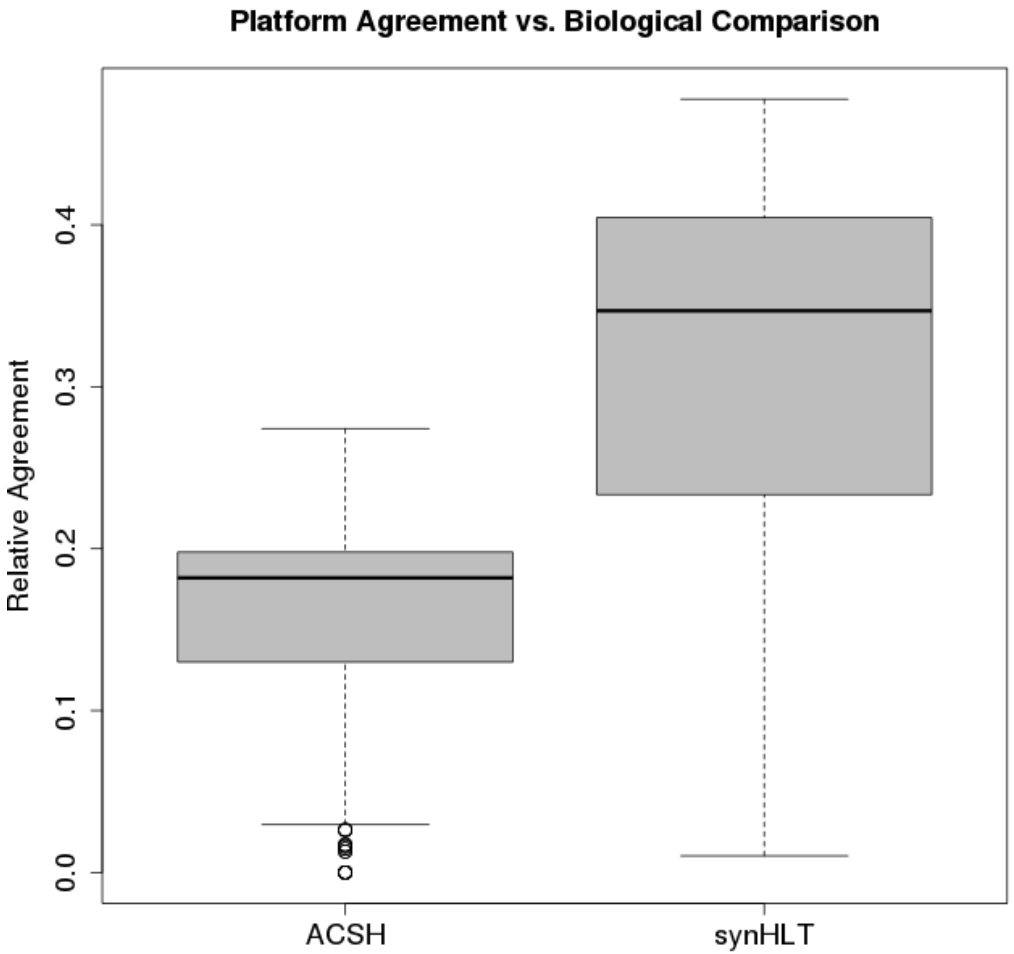


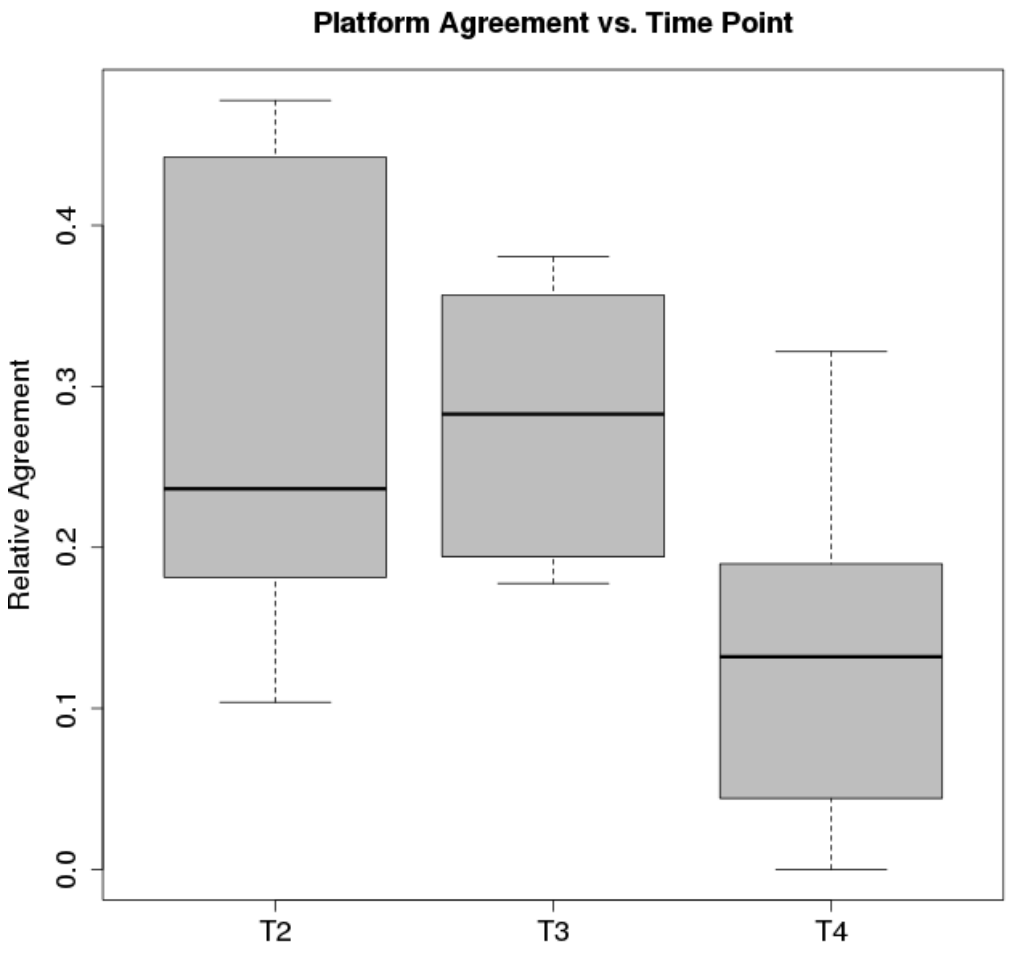
Agreement between RNA-Seq and microarray results: Influence of differential expression methodology - 2





Agreement between RNA-Seq and microarray results: Influence of biological phenomena





Agreement between RNA-Seq and microarray results: generalized model

relInt ~ QC + align + norm + comp + time + DE + crit

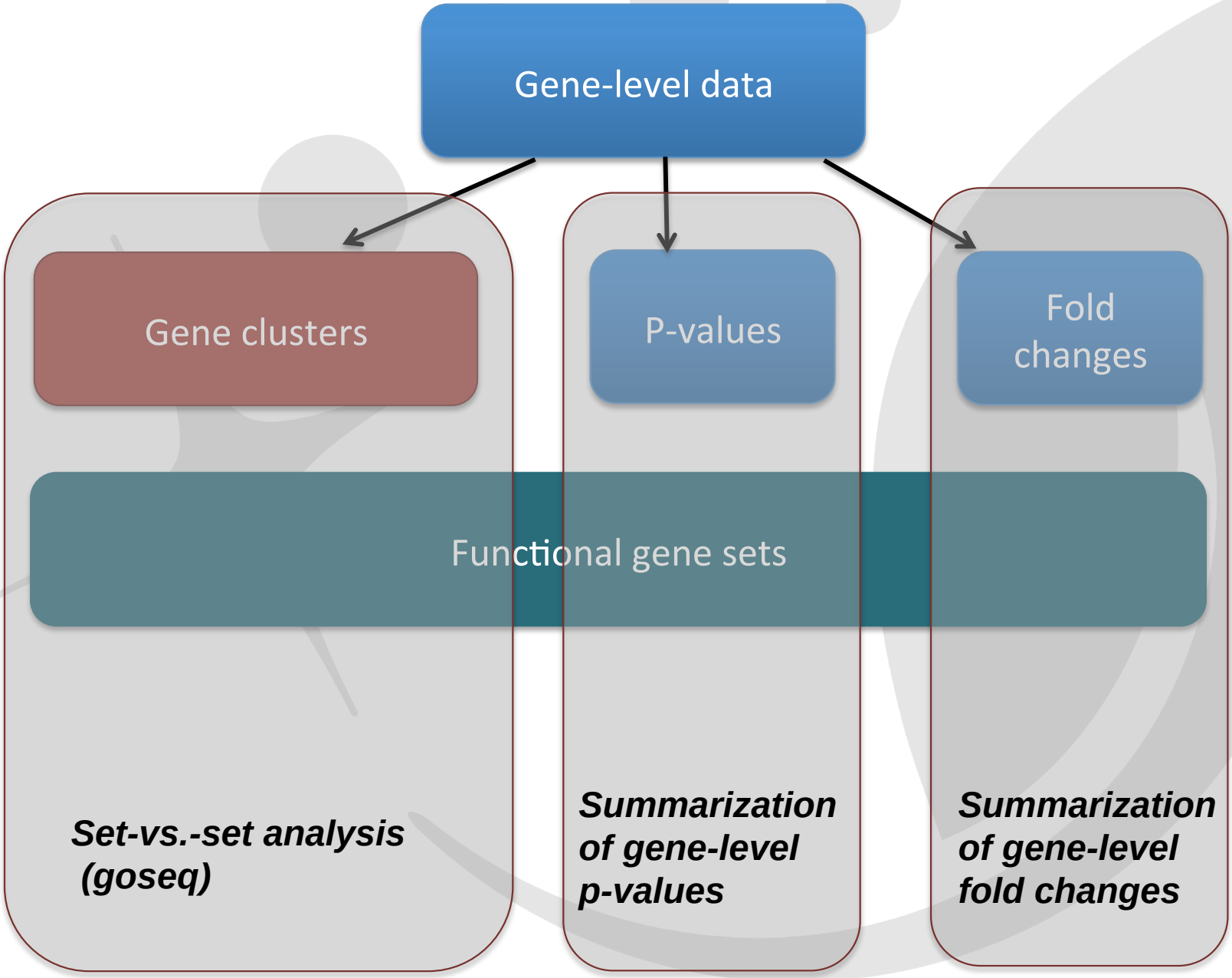
	Estimate	Std. Error	t value	Pr(>|t|)	
normUPPER	-0.016931	0.006150	-2.753	0.0061	**
compsynHLT	0.150192	0.005022	29.909	< 2e-16	***
timeT3	-0.024140	0.006150	-3.925	9.74e-05	***
timeT4	-0.172025	0.006150	-27.971	< 2e-16	***
DEEBSeq	0.068720	0.007102	9.677	< 2e-16	***
DEedgeR	0.045171	0.007102	6.361	4.15e-10	***

Functional pattern detection and the overall “method impact” model

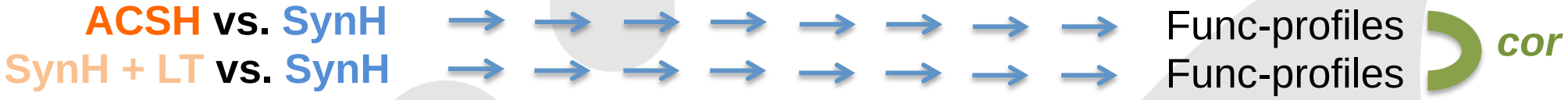
Functional pattern detection:
Types of knowledge-based gene sets

- KEGG Pathways
- Species-specific pathways
- Regulons
- Transporters
- Global biclusters

Functional enrichment reports: the democracy advantage



Looking through the analytical layers: impact on consistency between functionality profiles

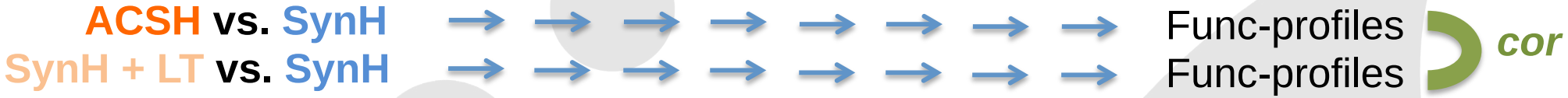


Looking through the analytical layers: impact on consistency between functionality profiles

ACSH vs. SynH → → → → → → → → Func-profiles
SynH + LT vs. SynH → → → → → → → → Func-profiles
cor

Cor(Func-prof) ~ readQC + Counting + Norm + DE + FuncEnrich + TimePoint

Looking through the analytical layers: impact on consistency between functionality profiles



Cor(Func-prof) ~ readQC + Counting + Norm + DE + FuncEnrich + TimePoint

Low-Level Processing

DE

Func. Dimens. Reduction

Biological Context

Looking through the analytical layers: impact on consistency between functionality profiles

ACSH vs. SynH → → → → → → → → Func-profiles
SynH + LT vs. SynH → → → → → → → → Func-profiles
cor

Cor(Func-prof) ~ readQC + Counting + Norm + DE + **FuncEnrich** + **TimePoint**

Low-Level Processing | DE | Func. Dimens. Reduction | Biological Context

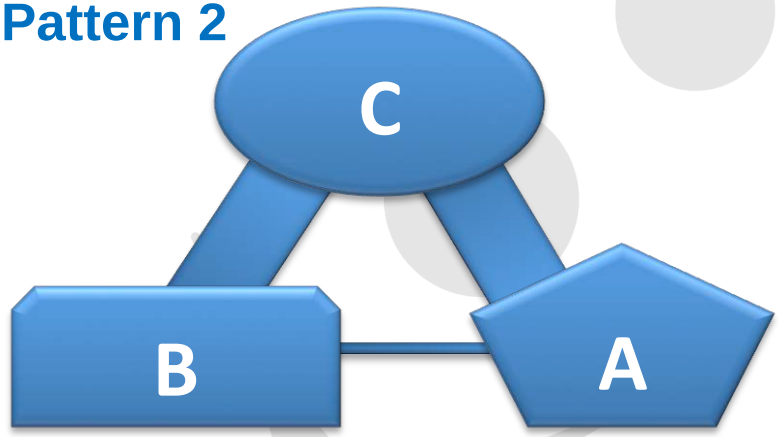
Running conclusions: overall method impact

- With all the within-level differences and peculiarities, low-level processing and DE options have surprisingly low impact on the final functionality profiles; main attention should be devoted to optimization of the functional enrichment stage

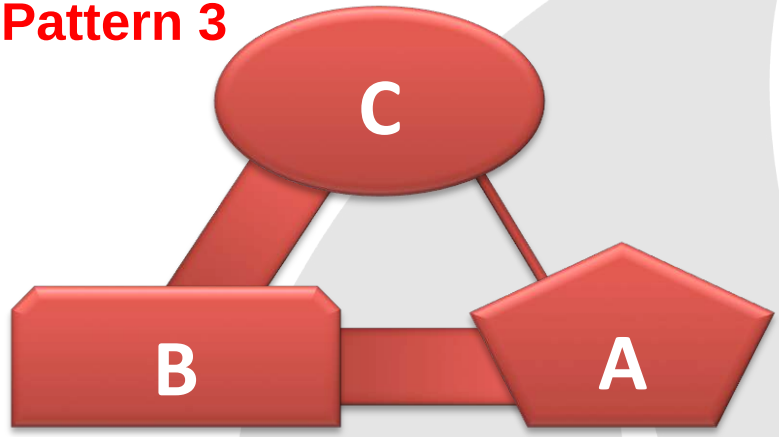
Multicondition designs and functionality representations: the dialogue between biology and statistics

The Bayesian Model Involving 3 Conditions: DE Patterns

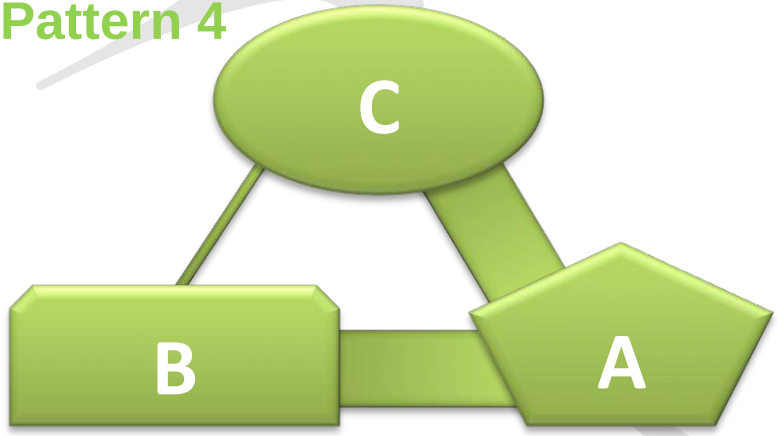
Pattern 2



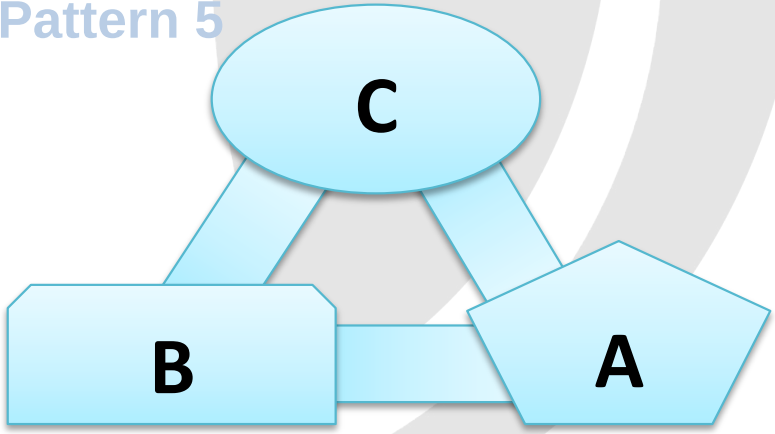
Pattern 3



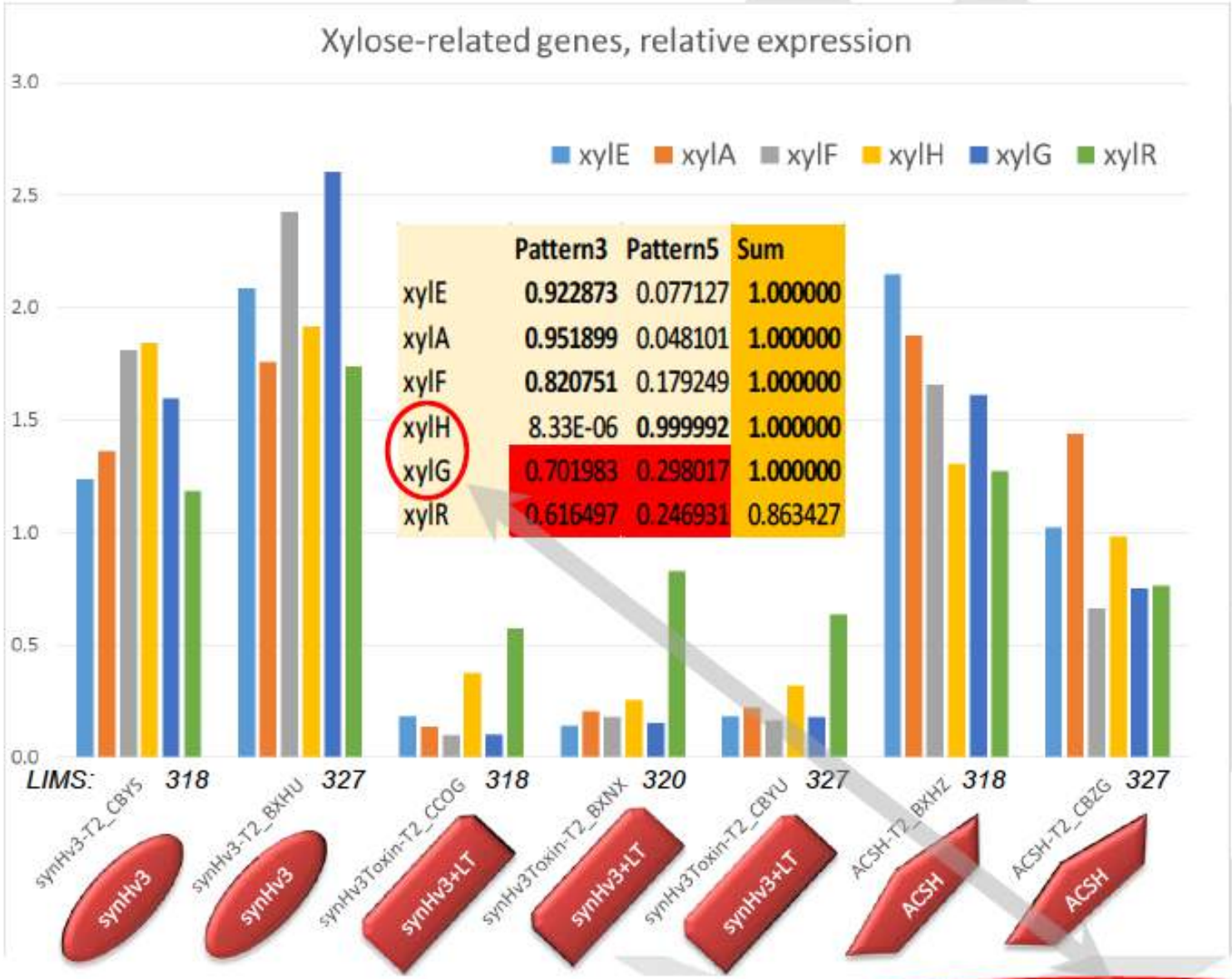
Pattern 4



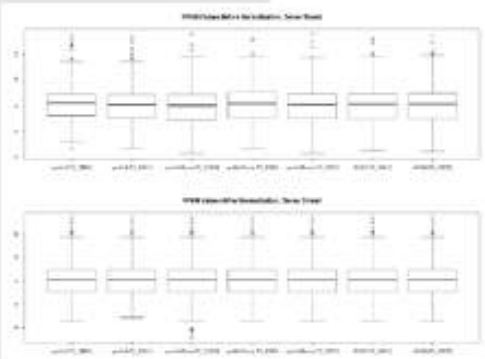
Pattern 5



CEBF case



	Pattern3	Pattern5	Sum
xylE	0.922873	0.077127	1.000000
xylA	0.951899	0.048101	1.000000
xylF	0.820751	0.179249	1.000000
xylH	8.33E-06	0.999992	1.000000
xylG	0.701983	0.298017	1.000000
xylR	0.616497	0.246931	0.863427



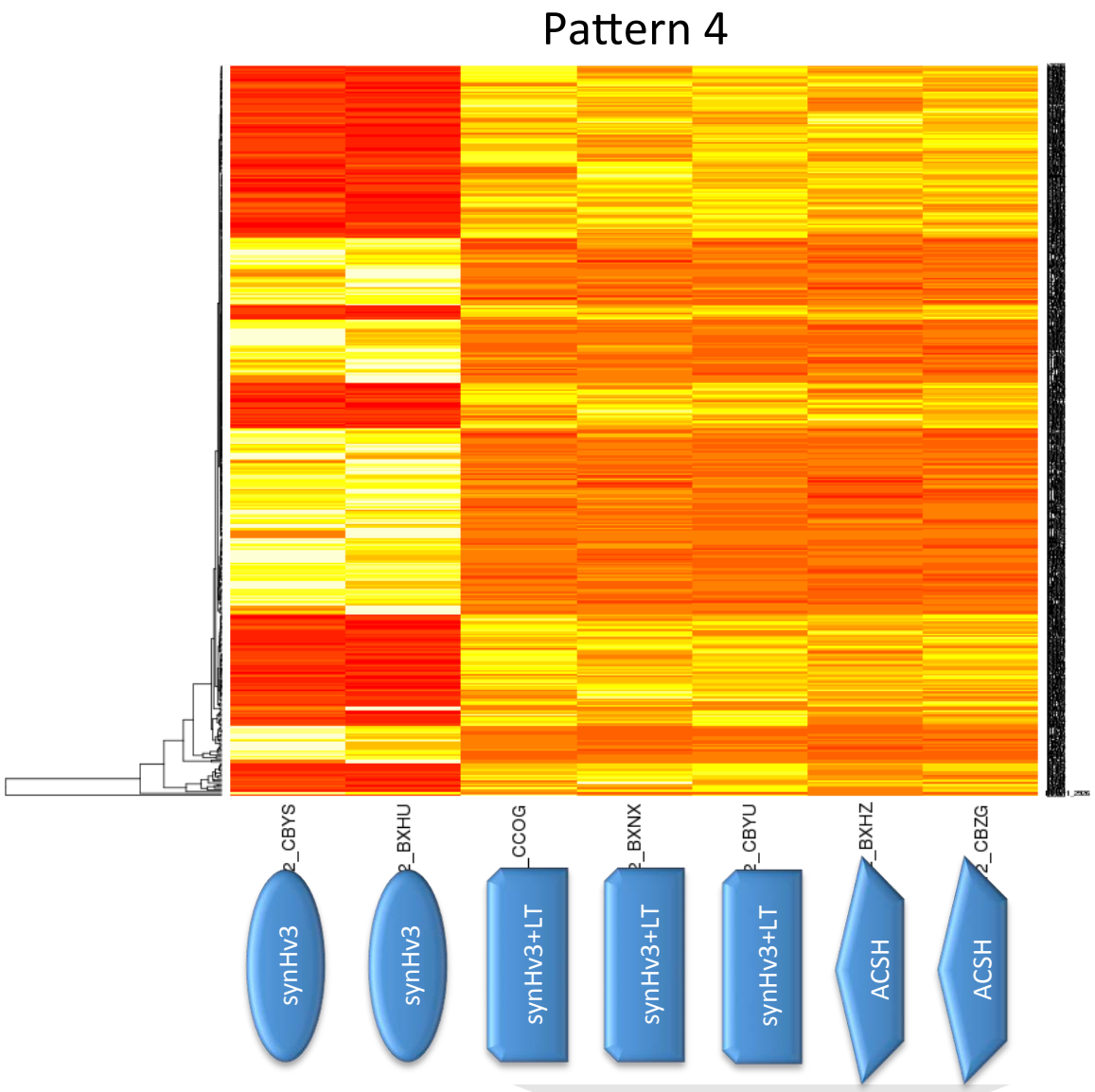
CEBF case

Introducing Co-Exclusivity-Based Filter

CEBF = *max(P(Pattern2)..P(Pattern4)) + P(Pattern5)*

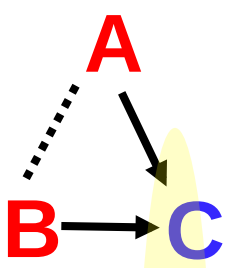
Gene ID	Pattern2	Pattern3	Pattern4	Pattern5	CEBF	fold 1	fold 2	fold 3	
cysD	**0.614651**	3.91E-33	1.85E-43	0.385349	1.000000	83.1	42.5	-2.0	sulfate adenylyltransferase, subunit 2
cysN	**0.692917**	1.17E-33	1.47E-44	0.307083	1.000000	56.0	31.2	-1.8	sulfate adenylyltransferase, subunit 1
yiaY	**0.672595**	4.43E-41	7.85E-35	0.327405	1.000000	10.1	11.9	1.2	predicted alcohol dehydrogenase
metK	2.43E-19	**0.632514**	1.06E-10	0.367486	1.000000	-8.5	2.4	20.2	S-adenosylmethionine synthetase
nhaR	0.007088	**0.568368**	9.00E-08	0.424544	0.992912	2.4	1.3	-1.8	DNA-binding transcriptional activator
xylG	1.01E-09	**0.701983**	5.89E-15	0.298017	1.000000	-14.4	-1.8	8.1	fused D-xylose transporter subunits
bioD	0.000107	**0.629857**	0.016256	0.353721	0.983578	5.0	-2.4	-12.2	dethiobiotin synthetase
bioF	3.05E-07	**0.682312**	0.000315	0.317373	0.999685	6.6	-2.5	-16.4	8-amino-7-oxononanoate synthase
bioB	1.10E-07	**0.715556**	0.000115	0.284329	0.999885	6.9	-2.4	-16.8	biotin synthase
copA	4.67E-12	0.000107	**0.696925**	0.302968	0.999893	-2.0	3.4	6.7	copper transporter
araB	4.21E-118	6.69E-89	**0.593357**	0.406643	1.000000	2.0	483.3	242.5	L-ribulokinase
puuR	0.007515	1.92E-06	**0.532285**	0.460121	0.992406	-1.2	-2.2	-1.8	**transcriptional repressor**
lsrR	2.65E-06	5.91E-11	**0.546829**	0.453169	0.999997	-1.1	-2.4	-2.1	**transcriptional repressor**
mall	0.003267	5.80E-06	**0.692639**	0.303909	0.996548	-1.2	-2.5	-2.1	**transcriptional repressor**
cusF	1.02E-12	4.37E-07	0.477945	**0.522054**	1.000000	-2.0	44.2	89.6	copper- and silver-binding protein
nirC	2.79E-19	1.94E-20	0.44172	**0.55828**	1.000000	2.0	20.4	10.4	nitrite transporter

Quality-Quantity Dilemma Solved: Detecting Massive Clean Response

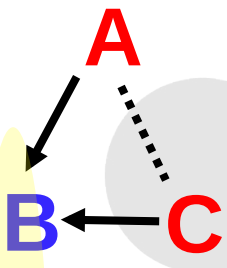


CEBF >= 0.95
crt >= 1.15

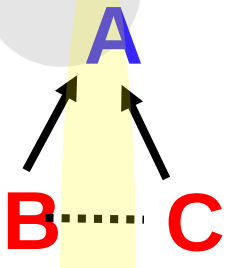
12 directional expression patterns ("hard clusters") based on EBSeq results



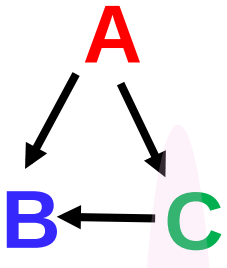
Pattern 2.1



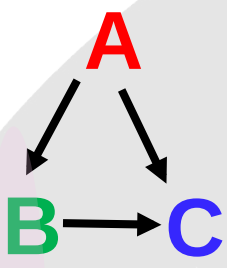
Pattern 3.1



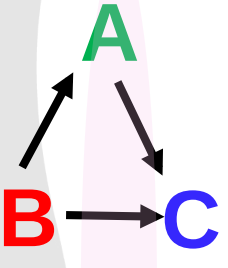
Pattern 4.1



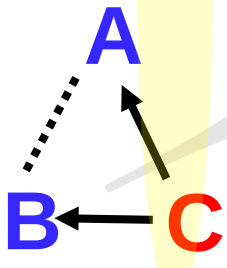
Pattern 5.1



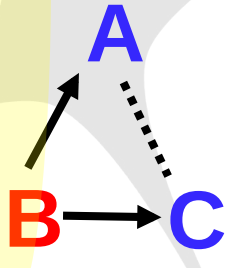
Pattern 5.3



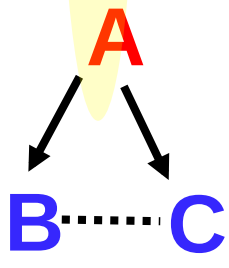
Pattern 5.5



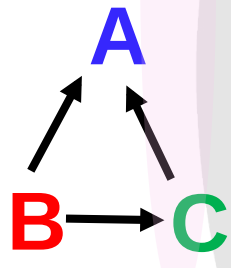
Pattern 2.2



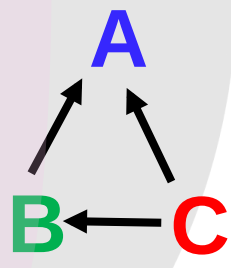
Pattern 3.2



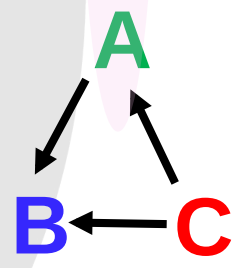
Pattern 4.2



Pattern 5.2



Pattern 5.4



Pattern 5.6

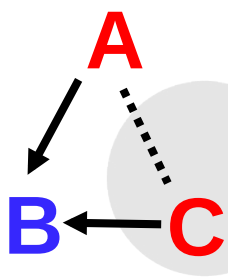
Specific condition in patterns 2-4

Transitional condition in pattern 5

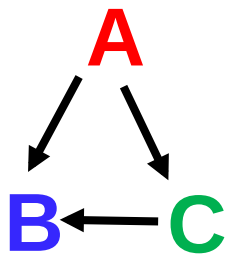
Pattern population alone may already tell the story

Pattern	Assigned genes
2.1 *ACSH* < (*synHv3* | *synHv3+LT*)	18
2.2 *ACSH* > (*synHv3* | *synHv3+LT*)	64
3.1 *synHv3+LT* < (*synHv3* | *ACSH*)	68
3.2 *synHv3+LT* > (*synHv3* | *ACSH*)	65
4.1 *synHv3* < (*synHv3+LT* | *ACSH*)	263
4.2 *synHv3* > (*synHv3+LT* | *ACSH*)	249
5.1 *synHv3* > *ACSH* > *synHv3+LT*	10
5.2 *synHv3+LT* > *ACSH* > *synHv3*	20
5.3 *synHv3* > *synHv3+LT* > *ACSH*	17
5.4 *ACSH* > *synHv3+LT* > *synHv3*	35
5.5 *synHv3+LT* > *synHv3* > *ACSH*	4
5.6 *ACSH* > *synHv3* > *synHv3+LT*	15

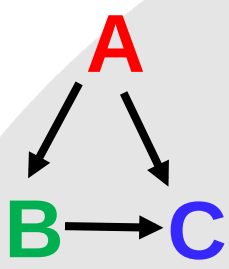
Next-Generation Pairwise Comparisons: B vs. A, Down



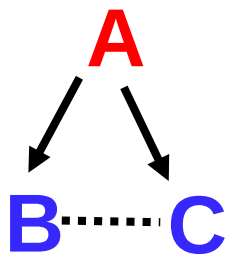
Pattern 3.1



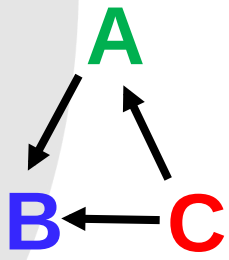
Pattern 5.1



Pattern 5.3

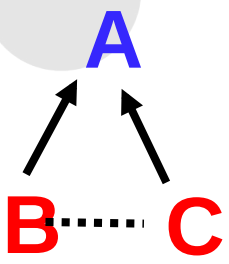


Pattern 4.2

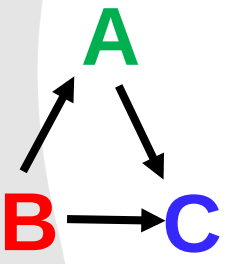


Pattern 5.6

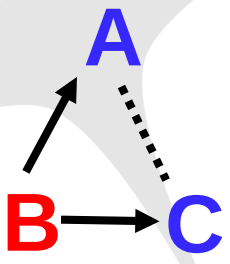
Next-Generation Pairwise Comparisons: B vs. A, Up



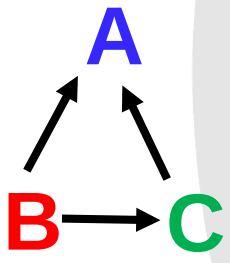
Pattern 4.1



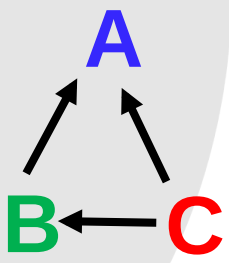
Pattern 5.5



Pattern 3.2



Pattern 5.2



Pattern 5.4

Next Generation Pairwise Comparison for generating gene-level statistics from a 3-condition EBSeq model

NGPC for geneset vs geneset (goseq) analysis was described earlier.
Edition of NGPC for genome-wide enrichment analysis with gene-level statistics follows:

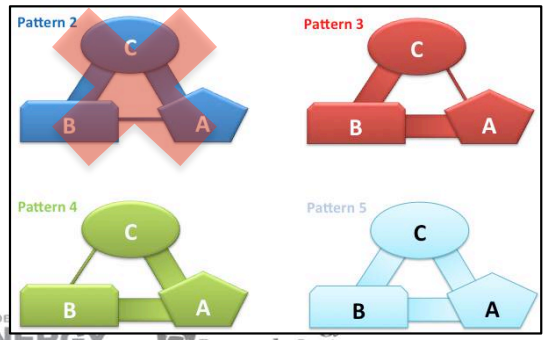
P(EE) – posterior probability of Equal Expression (Pattern 1)
P(EP) – posterior probability of assignment to a pattern that implies equal expression in the current pairwise comparison (one of P(Pattern2), P(Pattern3), P(Pattern4))

pNG = P(EE) + P(EP)

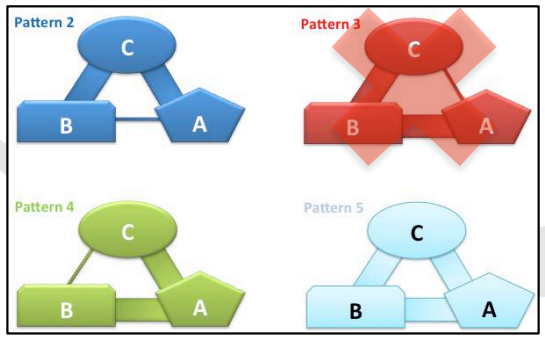
Example:

For B vs. A, Pattern2 implies no DE, Patterns 3-5 imply DE.
P(DE_AB) = P(Pattern3)+P(Pattern4)+P(Pattern5)
P(Pattern1)+P(Pattern2)+P(Pattern3)+P(Pattern4)+P(Pattern5) = 1
p = 1-PP(DE_AB) = P(Pattern1) + P(Pattern2) = P(EE) + P(EP) = pNG

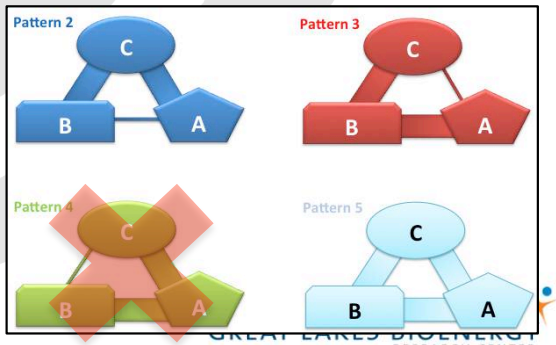
B vs. A



C vs. A



C vs. B



Revised functional analysis for *E.coli* LT experiment ($-\log_{10}(FDR)$)

Regulons - UP

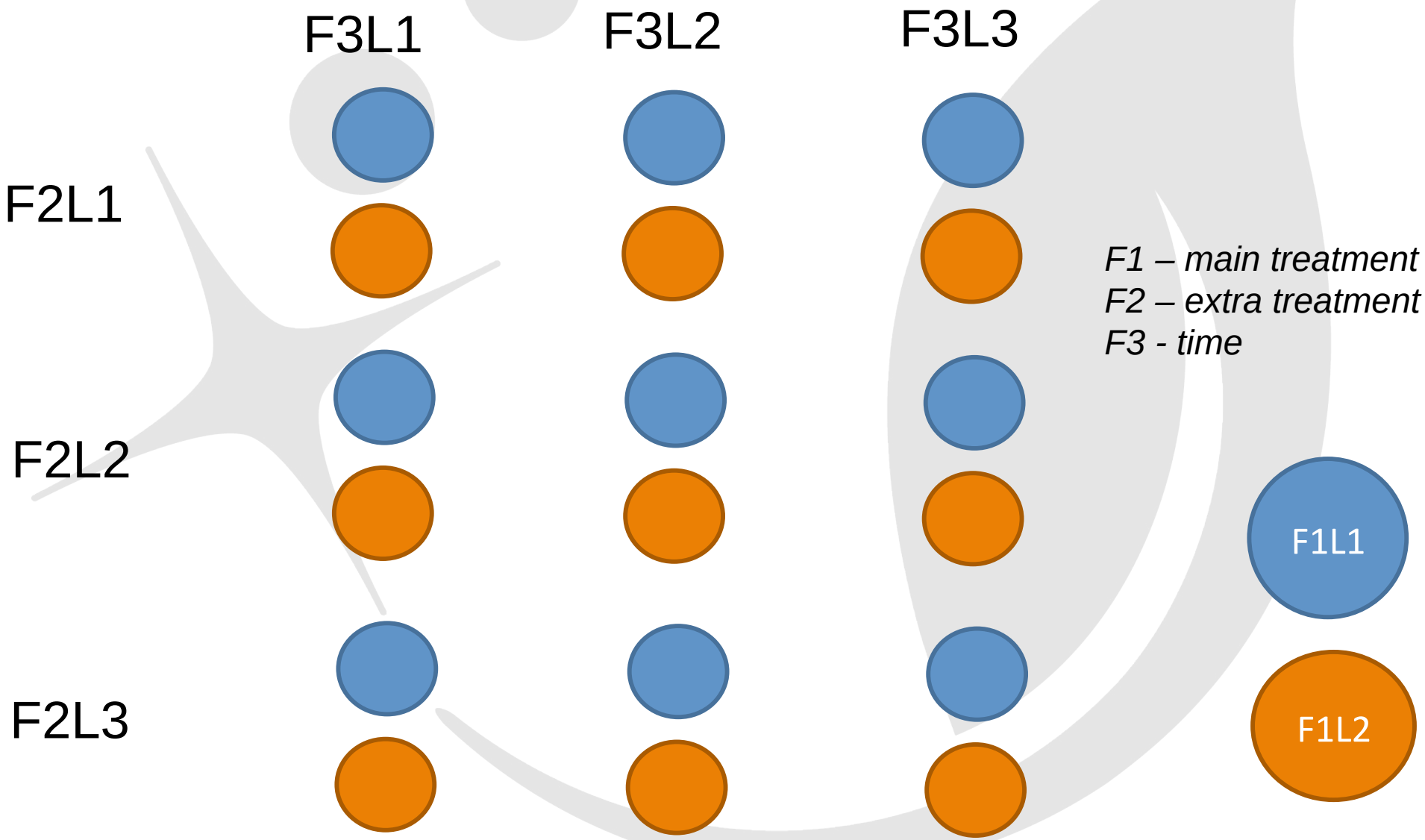
	ACSH vs. SynH, T2	ACSH vs. SynH, T3	ACSH vs. SynH, T4	SynH+LT vs. SynH, T2	SynH+LT vs. SynH, T3	SynH+LT vs. SynH, T4
PD00406_SoxS transcriptional activator_b4062	3.63	3.33	2.85	3.45	2.85	3.15
PD00365_MarA transcriptional activator_b1531	3.63	2.02	2.07	3.45	2.18	2.29
PD04418_Rob transcriptional activator_b4396	3.63	1.55	2.85	3.45	1.11	3.15
PC00040_CysB transcriptional dual regulator_b1275	3.63	0.00	0.00	3.45	0.00	0.00
PD00196_Fis transcriptional dual regulator_b3261	3.63	0.00	0.00	3.25	0.00	0.00
PD00364_MarR transcriptional repressor_b1530	2.19	0.40	1.85	2.12	0.18	1.93
PD01196_DeoR transcriptional repressor_b0840	1.66	0.00	0.00	1.67	0.00	0.00
PC00010_LexA transcriptional repressor_b4043	0.00	1.55	0.00	0.00	1.84	0.00

Regulons - DOWN

	ACSH vs. SynH, T2	ACSH vs. SynH, T3	ACSH vs. SynH, T4	SynH+LT vs. SynH, T2	SynH+LT vs. SynH, T3	SynH+LT vs. SynH, T4
EG20253-MONOMER_XylR transcriptional activator_b3569	3.15	3.75	0.91	2.85	3.55	0.76
PD00967_AppY transcriptional activator_b0564	3.15	0.00	0.00	3.15	0.00	0.00
EG12344-MONOMER_PspF transcriptional activator_b1303	2.85	3.75	0.00	3.15	3.50	0.22
PC00084_RcsB transcriptional activator_b2217	2.61	0.00	0.20	2.55	0.00	0.14
PD00353_Lrp transcriptional dual regulator_b0889	1.10	3.75	0.00	1.32	3.55	0.00
PD00288_H-NS transcriptional dual regulator_b1237	0.60	1.73	0.00	0.58	2.24	0.00
PC00063_GatR transcriptional repressor_b2090	0.25	3.75	0.48	0.17	3.50	0.39
PC00027_IHF transcriptional dual regulator_b1712	0.17	1.68	2.63	0.00	1.59	2.25
PC00040_CysB transcriptional dual regulator_b1275	0.00	3.75	0.00	0.00	3.55	0.04
PD00413_TyrR transcriptional dual regulator_b1323	0.00	3.75	0.00	0.00	3.50	0.00
PD01896_Mlc transcriptional repressor_b1594	0.00	3.75	0.00	1.87	2.64	0.00
PD04032_MetJ transcriptional repressor_b3938	0.00	3.75	0.00	0.00	3.37	0.00
PC00007_TrpR transcriptional repressor_b4393	0.00	3.20	0.00	0.00	3.50	0.00
PC00004_CRP transcriptional dual regulator_b3357	0.00	2.85	3.55	0.00	3.55	3.25
PD00237_MalT transcriptional activator_b3418	0.00	2.85	0.00	0.00	3.55	0.00
PD00197_FNR transcriptional dual regulator_b1334	0.00	2.67	0.08	0.00	1.79	0.00
EG11519-MONOMER_nickel-responsive transcriptional regulator_b3481	0.00	2.53	2.42	0.00	1.79	2.10
PD00763_LacI transcriptional repressor_b0345	0.00	2.47	0.00	0.00	1.79	0.00
EG12278-MONOMER_YiaJ transcriptional repressor_b3574	0.00	0.00	3.55	0.00	0.00	3.25
G7630-MONOMER_AgaR transcriptional regulator_b3131	0.00	0.00	3.55	0.00	0.00	3.25
PD02936_FhlA transcriptional activator_b2731	0.00	0.00	3.55	0.00	0.00	3.25
PHOB-MONOMER_PhoB transcriptional dual regulator_b0399	0.00	0.00	3.55	0.00	0.00	3.25
G7308-MONOMER_HyfR transcriptional activator_b2491	0.00	0.00	2.32	0.00	0.00	2.07
PC00009_Fur transcriptional dual regulator_b0683	0.00	0.00	2.04	0.00	0.00	2.07
G6201-MONOMER_MhpR transcriptional regulator_b0346	0.00	0.00	2.02	0.00	0.00	1.75

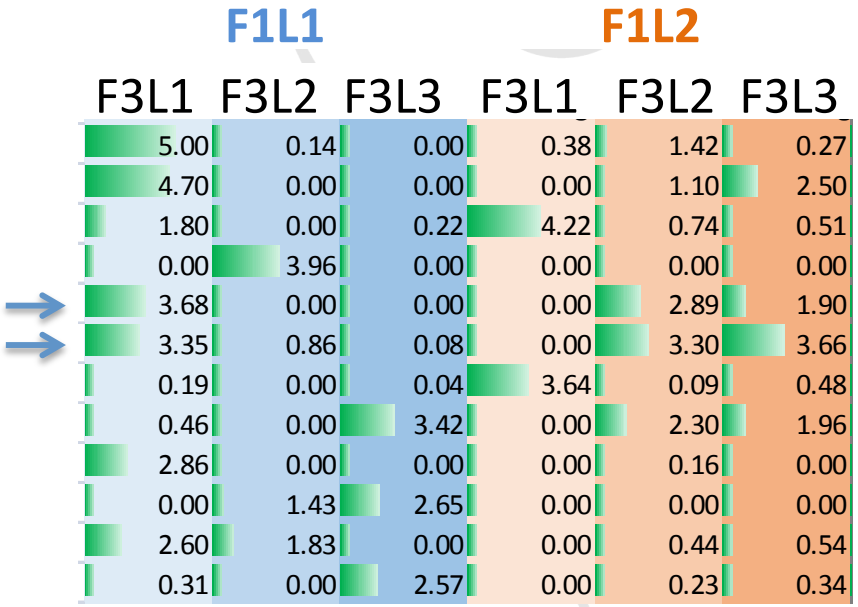
The trend of consistent changes between ACSH vs. synH and synH+LT vs. synH holds true for pathways and transporters (not shown)

Our approach to biological interpretation of multicondition RNA-Seq experiments: an example multifactorial design



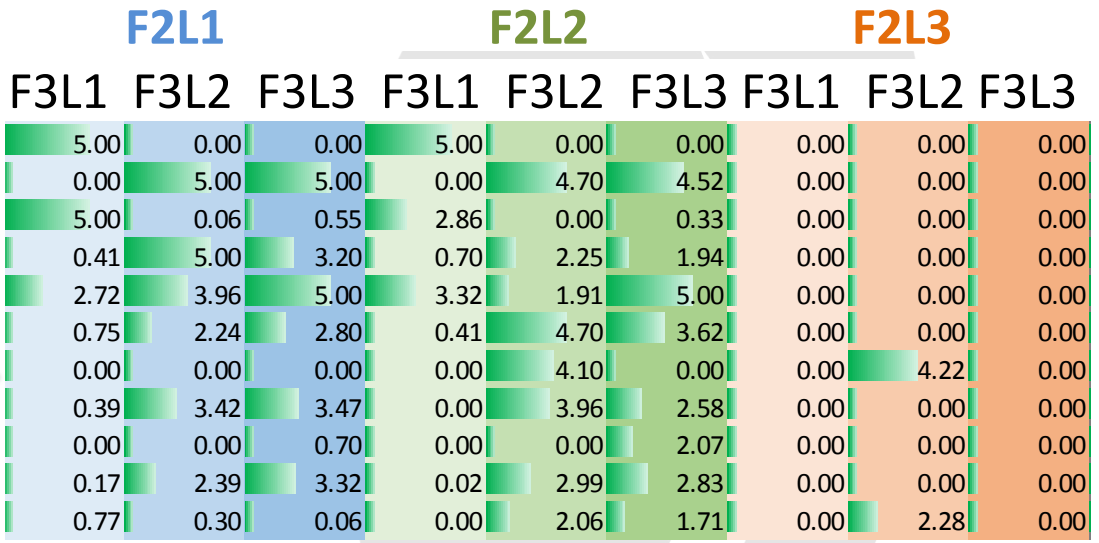
Our approach to biological interpretation of multiconditional RNA-Seq experiments: squeezing the functional analysis between subsets of experimental factors

Functional View 1: 3-way F2-centric models



(dynamics off involvement of functional entities into formation of a particular F2-centric expression pattern for each of the 2 levels of F1)

Functional View 2: 2-way F1-centric models



(explicit effects of F1 at functional level at every F2 + F3 combination)

Conclusions

- Later data processing stages have more impact on the conclusions about functional biological responses than earlier / upstream stages
- Running functional analysis over subsets of factors in multifactorial designs improves the interpretability of the results
- Look at the raw data for inspiration!

Acknowledgements

GLBRC

- Robert Landick
- Yaoping Zhang
- Yury Bukhman
- David Benton
- Branden Timm

UW-Madison – Biostatistics and Medical Informatics

- Christina Kendziorski
- Ning Leng
- Colin Dewey
- Bo Li

University of Wyoming– Molecular Biology

- Mark Gomelsky