bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

Genetic influences on translation in yeast

Frank W. Albert, Dale Muzzey, Jonathan S. Weissman, et al.

bioRxiv first posted online March 16, 2014
Access the most recent version at doi: http://dx.doi.org/10.1101/003319

Han Fang

09/15/2014

Lyon lab journal club

Outline

- Introduction
- Materials and Methods (Overview)
- Results
- Discussions
- Materials and Methods (Details)

Introduction

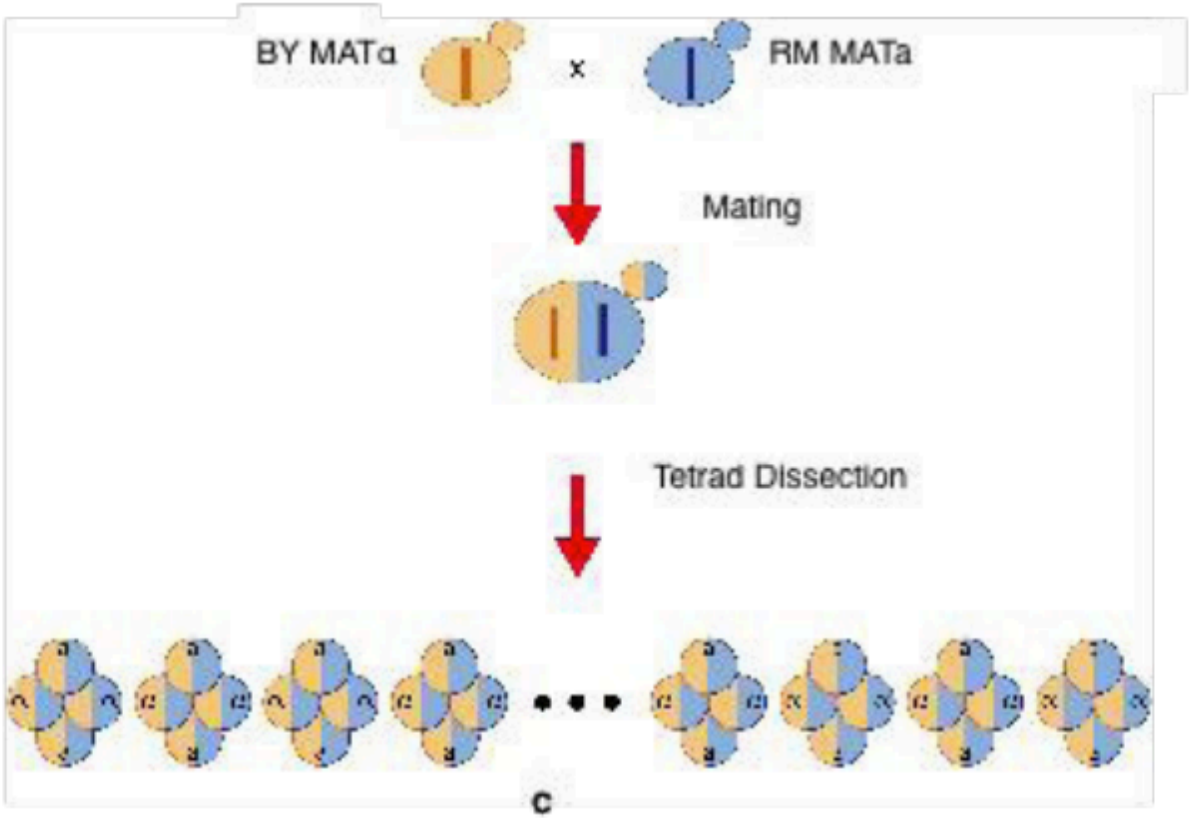
- Previous reports suggest posttranscriptional processes:

1) Mass-spectrometry -> dif. Genetics architecture between protein and mRNA levels
2) Many eQTL did not correspond to a protein QTL

- The rate of translation (a better predictor of protein levels than mRNA)
- In *cis* (physically linked, ASE) and in *trans* (both alleles)
- BY, RM, and diploid hybrid
- Ribosome footprints (FP) and mRNA levels
- Differences in FPs between BY & RM highly correlated with those in mRNA (parents and ASE in the hybrid)
- A small # genes (strong translational-specific effect) & hundreds of genes modest effects

Methods - Overview

- Yeast strains (the same as in Bloom et al.)
- prototrophic, i.e. no engineered deletions of metabolic genes.



Methods – Overview cont.

- Cycloheximide (much lower dosages)
- Ribosome profiling and sequencing

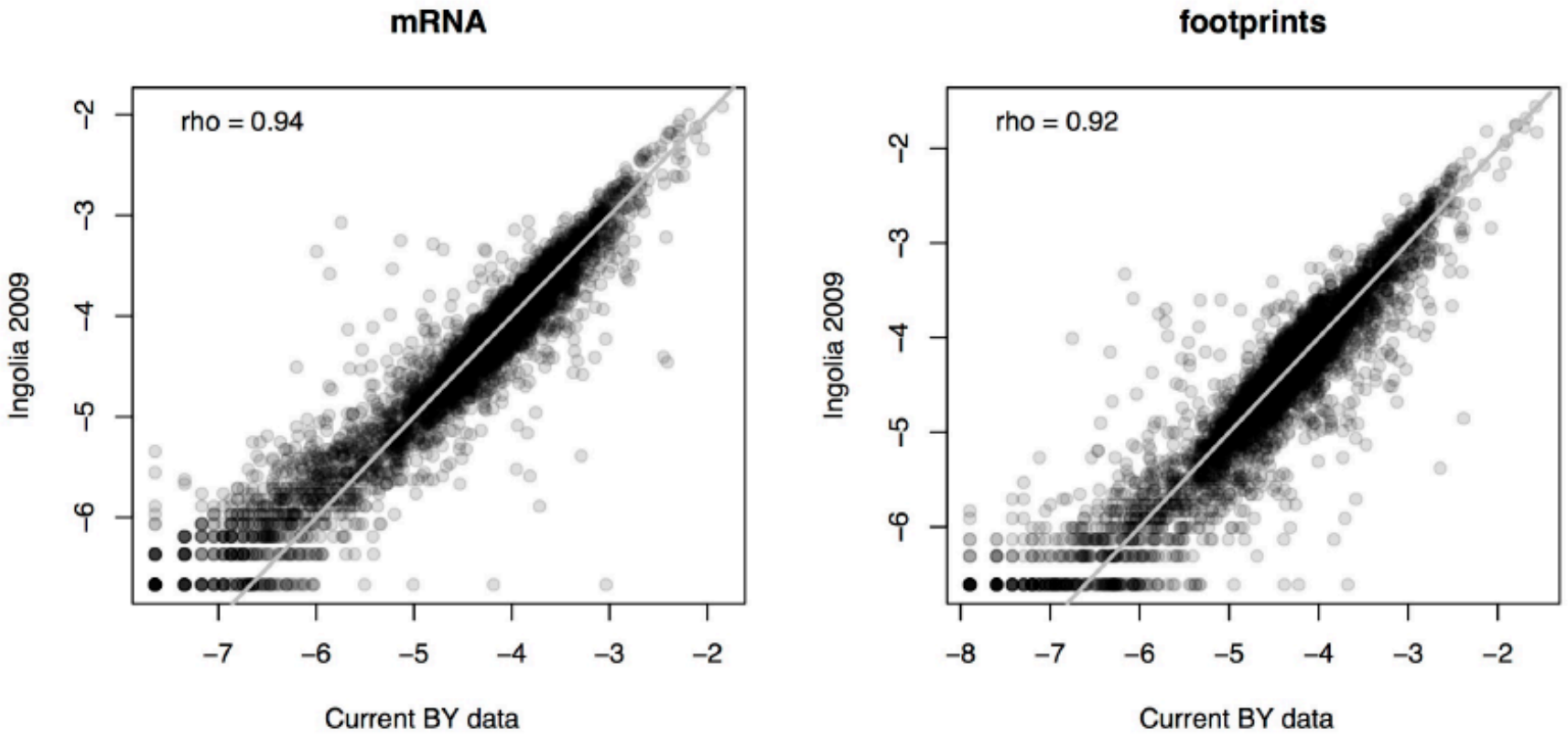
1) cells were cultured in YNB
2) the reverse-transcription
3) highly abundant rRNA species were hybridized to biotinylated oligos and subtracted

- BWA align, SAMtools SNP calling, modified reference (BY)
- Translational efficiency (TE): the ratio of FP abundance to mRNA abundance
- Statistical analysis:

1) binomial tests (main text)
2) DESeq analysis framework (more conservative Supplementary Note S2)

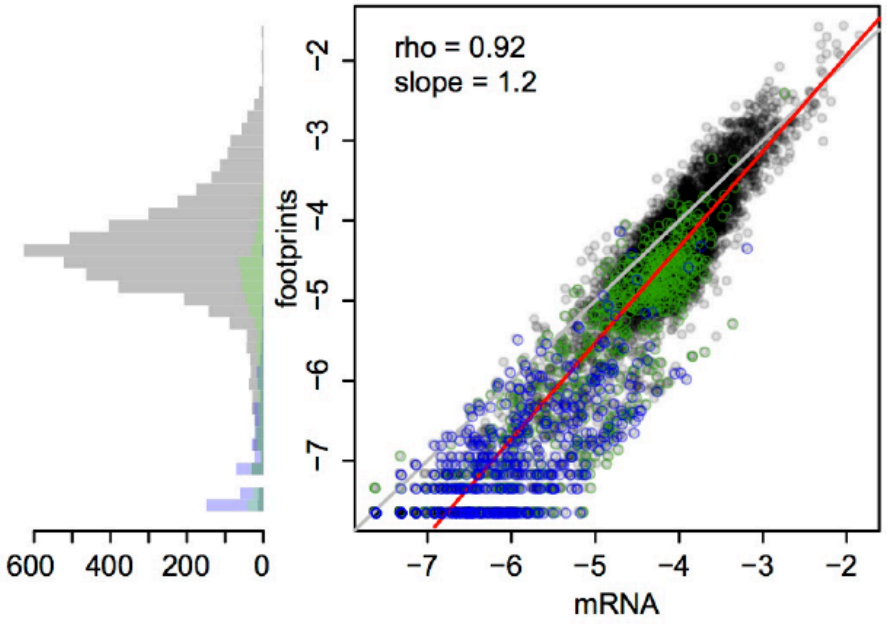
Results

Supplementary Figure S1 – Comparison to Ingolia *et al.* 2009 data



Shown are log10 transformed normalized read counts. The grey line marks identity.

Figure 1 – Global mRNA and footprint abundance

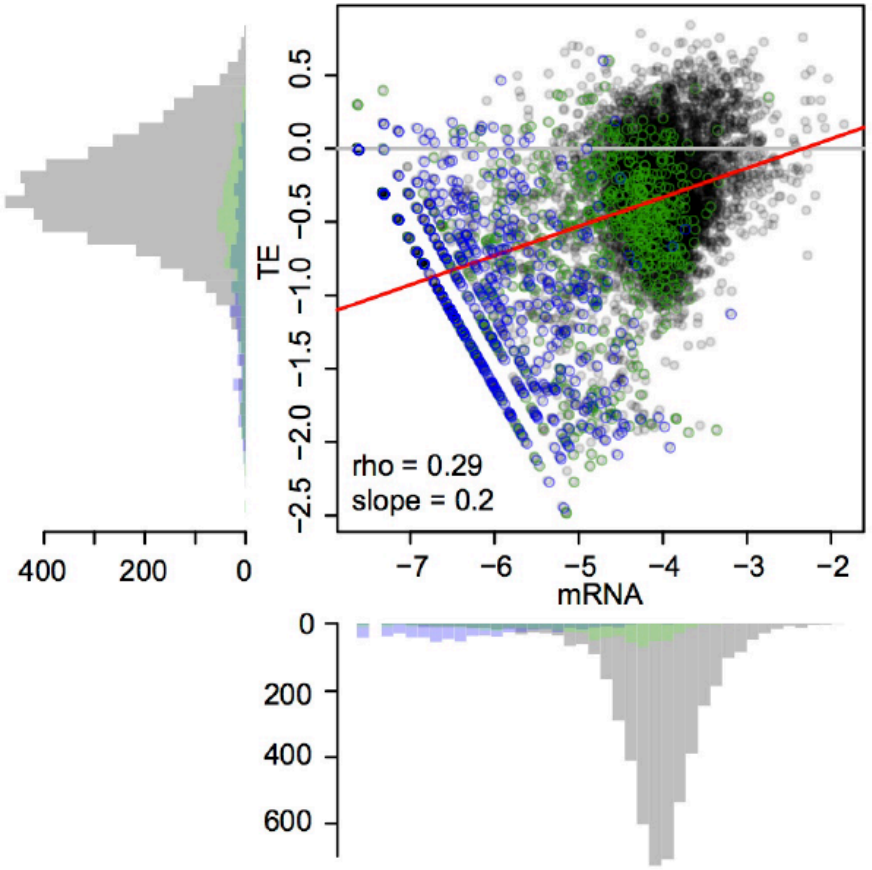


mRNA vs. footprint abundance

log10-transformed normalized read counts

The red line shows the regression of footprint on mRNA abundance.
The grey line indicates identity.

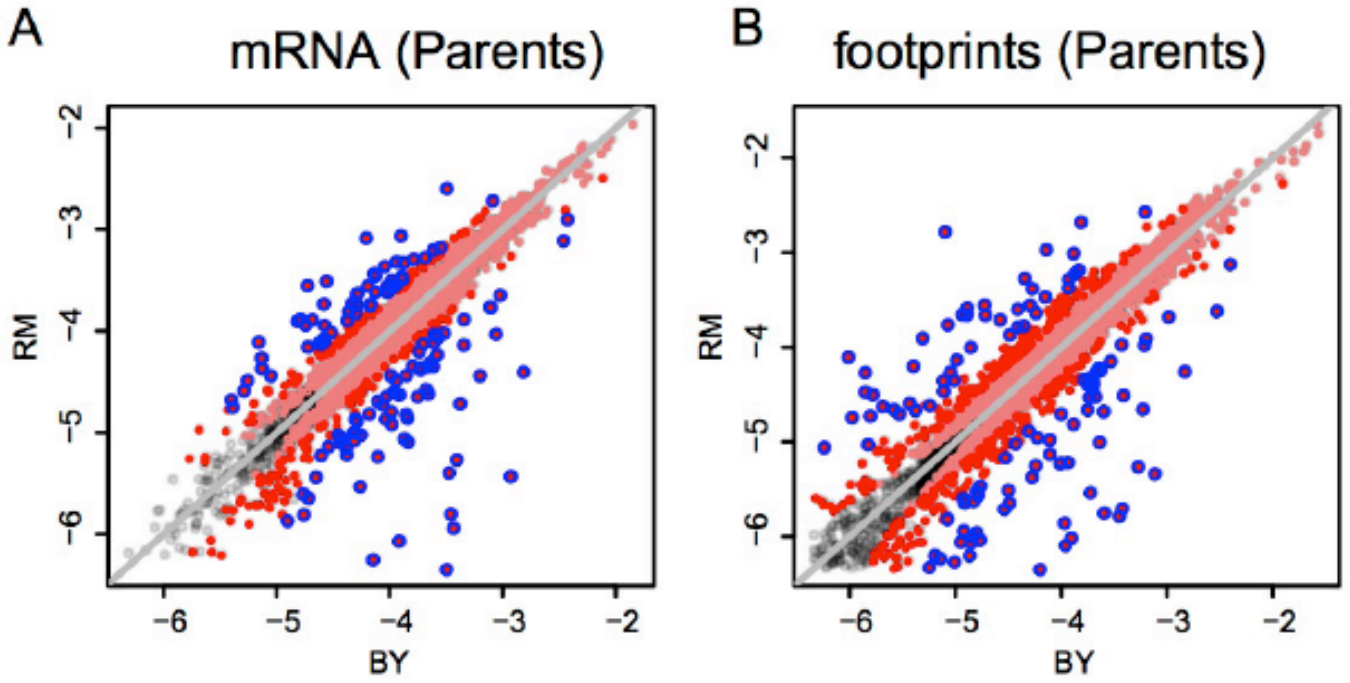
The transparent grey points are "verified" ORFs, green points are "uncharacterized" ORFs and blue points are "dubious" ORFs.



TE as a function of mRNA abundance.

The grey line denotes identity between footprint and mRNA levels (i.e. log10(TE) = 0).
The red line shows the regression of TE on mRNA abundance

Figure 2 – Expression in BY vs. RM and ASE in the hyl



Shown are log10-transformed normalized read counts based on the downsampled data. Grey diagonal lines mark identity. Light red points are genes with significant differences (Bonferroni corrected: $p < 9e\text{-}6$ in parent data and $p < 1.5\ e\text{-}5$ in hybrid data), darker red points are significant genes with a fold change ≥ 2. The blue circles denote genes that were called significant by DESeq (Benjamini-Hochberg adjusted $p < 0.05$)

Table 1 – Differential expression statistics

- 54% & 58% significant
- >90% less than 2-fold

Comparison	Reads	Data	Analyzed genes	2-fold	Binomial / G test[1]	Intersect	DESeq[2]
Parent	All	mRNA	5,316	331	2,862	314 (6%)	189
Parent	All	Footprint	5,316	514	3,057	490 (9%)	145
Parent	All	TE	5,316	135	2,228	111 (2%)	NA
Parent	SNP	mRNA	3,342	249	517	171 (5%)	75
Parent	SNP	Footprint	3,342	475	671	289 (9%)	67
Parent	SNP	TE	3,342	329	319	97 (3%)	NA
Hybrid	SNP	mRNA	3,342	100	198	40 (1%)	40
Hybrid	SNP	Footprint	3,342	194	210	67 (2%)	70
Hybrid	SNP	TE	3,342	216	106	26 (1%)	9

[1]Bonferroni corrected $p < 0.05$. Binomial tests were run on mRNA and footprint data. G-tests were used to test for differential TE.

[2]DESeq was run on all of genes where at least one sample had more than zero counts (6,457 genes for the parent comparison using all reads and 4,361 genes for the SNP-based analyses). DESeq results were corrected for multiple testing using the Benjamini-Hochberg correction.

Supplementary Table S3 – FDR-based differential expression statistics

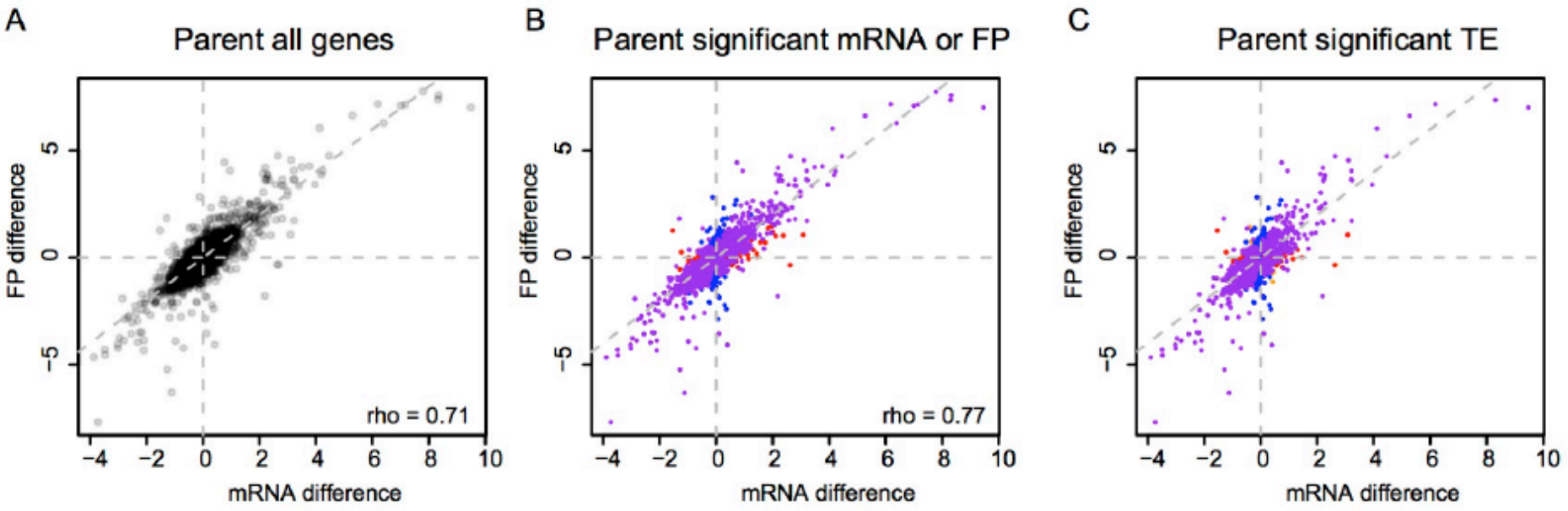
Comparison	Reads	Data	Analyzed genes	2-fold	Binomial test[1]	Intersect
Parent	All	mRNA	5,316	331	4,575	331 (6%)
Parent	All	Footprint	5,316	514	4,669	512 (10%)
Parent	All	TE	5,316	135	4,256	135 (3%)
Parent	SNP	mRNA	3,342	249	1,159	225 (7%)
Parent	SNP	Footprint	3,342	475	1,486	441 (13%)
Parent	SNP	TE	3,342	329	1,155	278 (8%)
Hybrid	SNP	mRNA	3,342	100	529	65 (2%)
Hybrid	SNP	Footprint	3,342	194	617	128 (4%)
Hybrid	SNP	TE	3,342	216	638	148 (4%)

- Similar patterns remained

[1]q-value < 0.05

Figure 3 – mRNA vs. footprint differences

- Direction agreed for 82%



Shown are log2-transformed fold changes. A-C: parents lines are the diagonal.

- Left: all genes.
- Middle: genes with a significant mRNA (red), footprint (blue) or both mRNA and footprint (purple) difference.
- Right column: genes with a significant TE difference. Red: genes with only a significant mRNA difference, blue: genes with only a significant footprint difference, purple: genes with both a significant mRNA and footprint difference, orange: genes with neither a significant mRNA nor a significant footprint difference.

Table 1 – Differential expression statistics

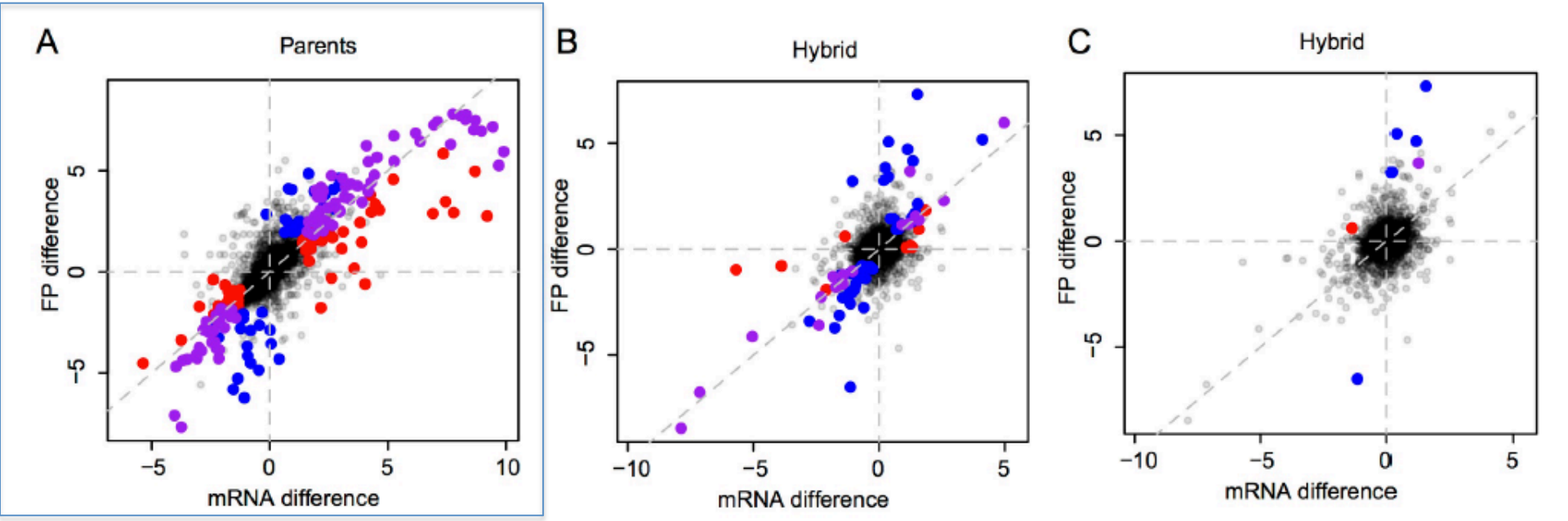
- DESeq yielded substantially fewer diff. exp. genes
- High agreement remained intact.

Comparison	Reads	Data	Analyzed genes	2-fold	Binomial / G test[1]	Intersect	DESeq[2]
Parent	All	mRNA	5,316	331	2,862	314 (6%)	189
Parent	All	Footprint	5,316	514	3,057	490 (9%)	145
Parent	All	TE	5,316	135	2,228	111 (2%)	NA
Parent	SNP	mRNA	3,342	249	517	171 (5%)	75
Parent	SNP	Footprint	3,342	475	671	289 (9%)	67
Parent	SNP	TE	3,342	329	319	97 (3%)	NA
Hybrid	SNP	mRNA	3,342	100	198	40 (1%)	40
Hybrid	SNP	Footprint	3,342	194	210	67 (2%)	70
Hybrid	SNP	TE	3,342	216	106	26 (1%)	9

[1]Bonferroni corrected $p < 0.05$. Binomial tests were run on mRNA and footprint data. G-tests were used to test for differential TE.

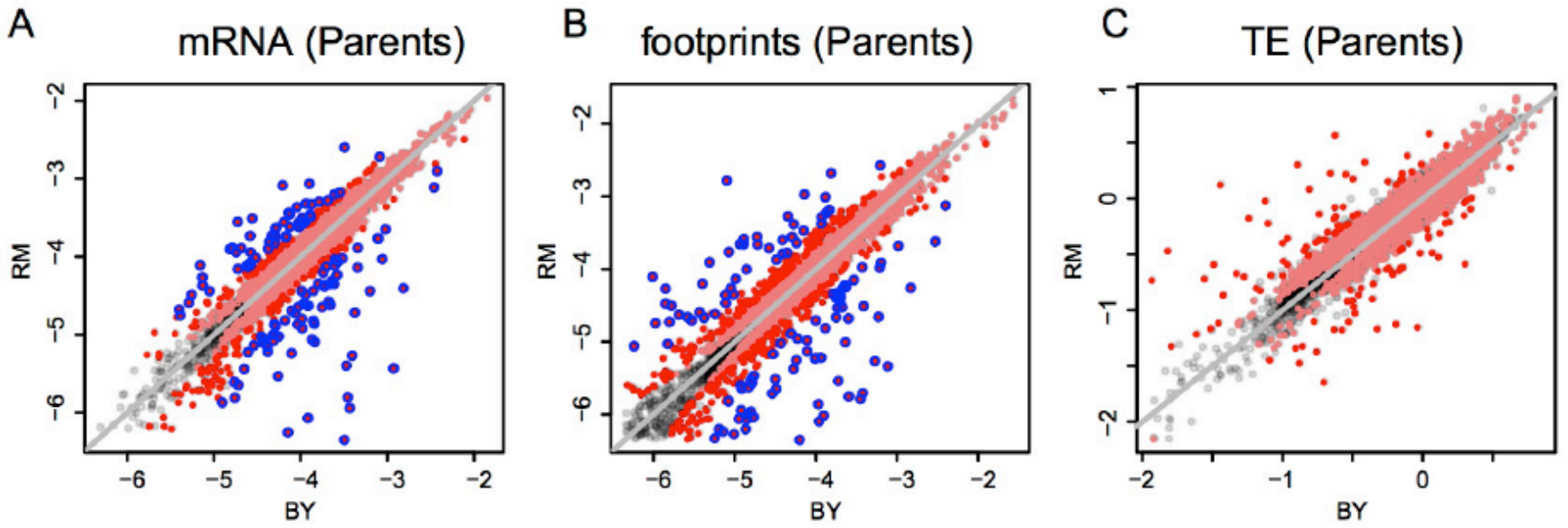
[2]DESeq was run on all of genes where at least one sample had more than zero counts (6,457 genes for the parent comparison using all reads and 4,361 genes for the SNP-based analyses). DESeq results were corrected for multiple testing using the Benjamini-Hochberg correction.

Supplementary Figure S2 – mRNA vs. footprint differences identified by DESeq



Shown are log2-transformed fold changes. A: parents, B & C: hybrid ASE. A & B: genes with a significant (Benjamini-Hochberg corrected $p < 0.05$) mRNA (red), footprint (blue) or both mRNA and footprint (purple) difference. C: genes with a significant TE difference. Red: genes with only a significant mRNA difference, blue: genes with only a significant footprint difference, purple: genes with both a significant mRNA and footprint difference.

Figure 2 – Expression in BY vs. RM and ASE in the hybrid



Shown are log10-transformed normalized read counts based on the downsampled data. Grey diagonal lines mark identity. Light red points are genes with significant differences (Bonferroni corrected: $p < 9e\text{-}6$ in parent data and $p < 1.5\ e\text{-}5$ in hybrid data), darker red points are significant genes with a fold change ≥ 2. The blue circles denote genes that were called significant by DESeq (Benjamini-Hochberg adjusted $p < 0.05$)

Table 1 – Differential expression statistics

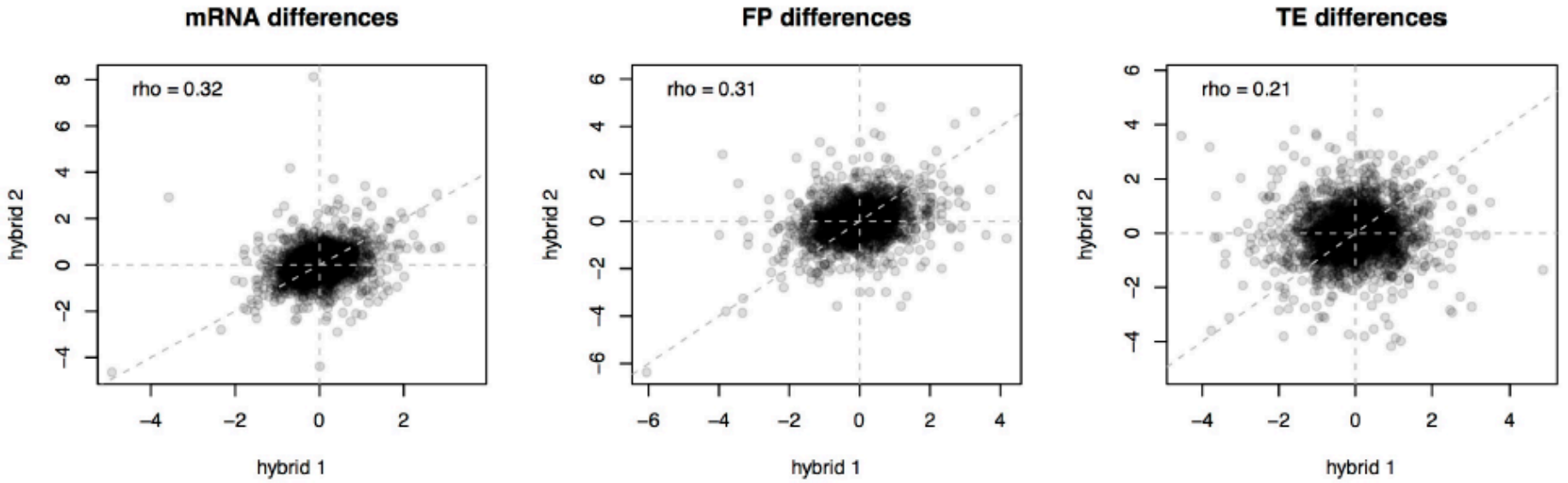
- 42% significant
- Most TE diff. were of small magnitude

Comparison	Reads	Data	Analyzed genes	2-fold	Binomial / G test[1]	Intersect	DESeq[2]
Parent	All	mRNA	5,316	331	2,862	314 (6%)	189
Parent	All	Footprint	5,316	514	3,057	490 (9%)	145
Parent	All	TE	5,316	135	2,228	111 (2%)	NA
Parent	SNP	mRNA	3,342	249	517	171 (5%)	75
Parent	SNP	Footprint	3,342	475	671	289 (9%)	67
Parent	SNP	TE	3,342	329	319	97 (3%)	NA
Hybrid	SNP	mRNA	3,342	100	198	40 (1%)	40
Hybrid	SNP	Footprint	3,342	194	210	67 (2%)	70
Hybrid	SNP	TE	3,342	216	106	26 (1%)	9

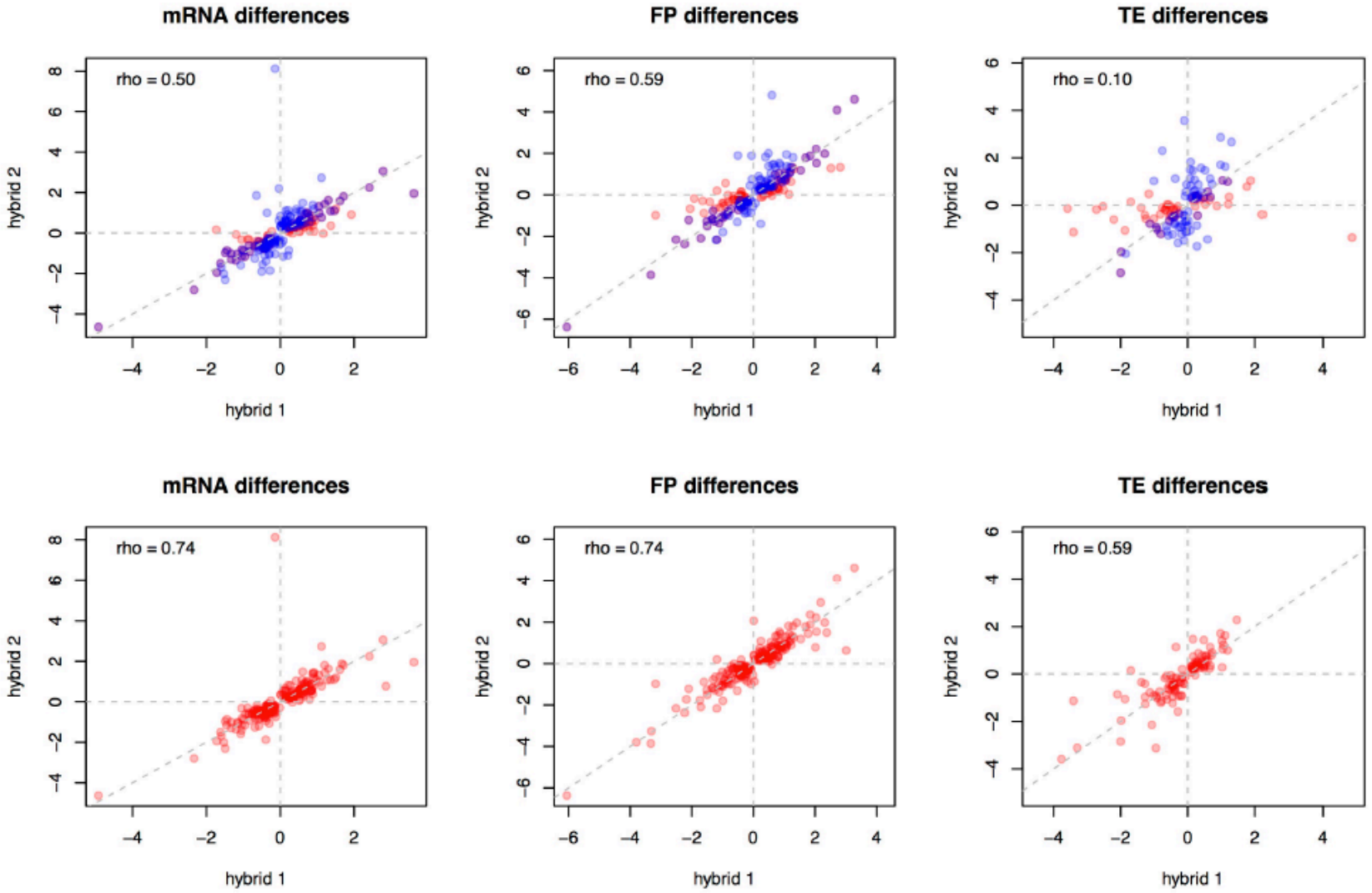
[1]Bonferroni corrected $p < 0.05$. Binomial tests were run on mRNA and footprint data. G-tests were used to test for differential TE.

[2]DESeq was run on all of genes where at least one sample had more than zero counts (6,457 genes for the parent comparison using all reads and 4,361 genes for the SNP-based analyses). DESeq results were corrected for multiple testing using the Benjamini-Hochberg correction.

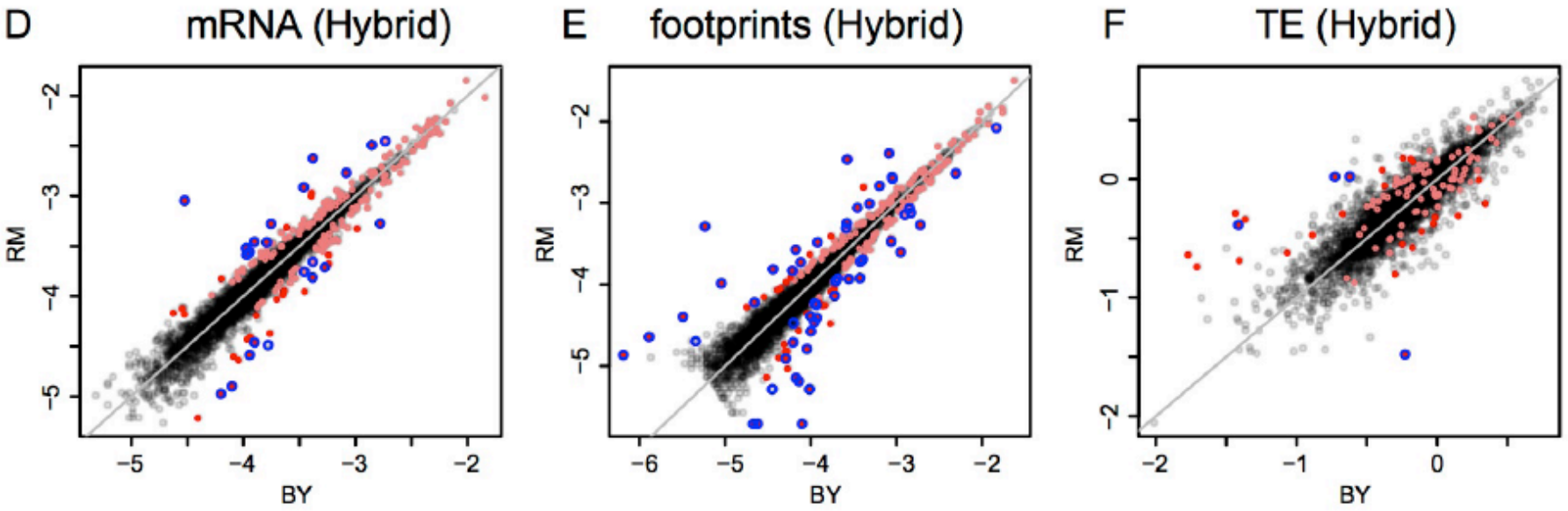
Supplementary Figure S3 – Reproducibility of hybrid measurements



Shown are log2-transformed fold changes. Grey diagonals mark identity. Top row: all genes. Middle row: significant genes in one and / or the other replicate. Bottom row: significant genes in the combined hybrid data. Spearman correlation coefficients between replicates are given in each panel.



Shown are log2-transformed fold changes. Grey diagonals mark identity. Top row: all genes. Middle row: significant genes in one and / or the other replicate. Bottom row: significant genes in the combined hybrid data. Spearman correlation coefficients between replicates are given in each panel.



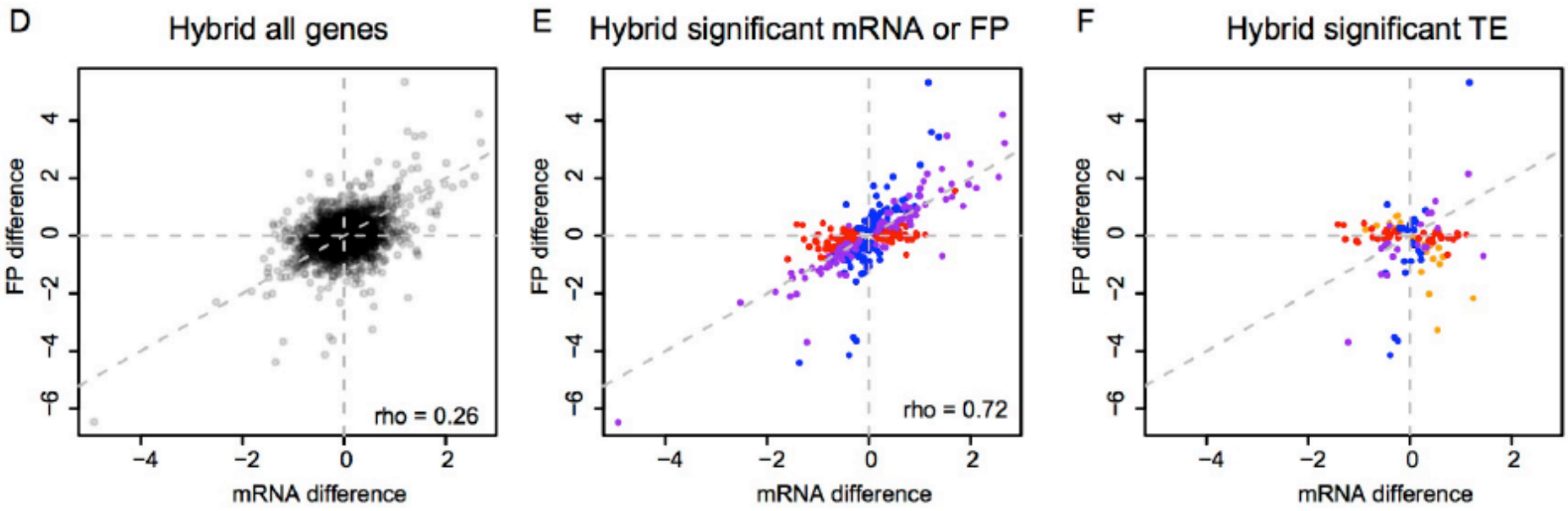
Shown are log10-transformed normalized read counts based on the downsampled data. Grey diagonal lines mark identity. Light red points are genes with significant differences (Bonferroni corrected: $p < 9e\text{-}6$ in parent data and $p < 1.5\ e\text{-}5$ in hybrid data), darker red points are significant genes with a fold change ≥ 2. The blue circles denote genes that were called significant by DESeq (Benjamini-Hochberg adjusted $p < 0.05$)

Table 1 – Differential expression statistics • Fewer significant ASE genes

Comparison	Reads	Data	Analyzed genes	2-fold	Binomial / G test[1]	Intersect	DESeq[2]
Parent	All	mRNA	5,316	331	2,862	314 (6%)	189
Parent	All	Footprint	5,316	514	3,057	490 (9%)	145
Parent	All	TE	5,316	135	2,228	111 (2%)	NA
Parent	SNP	mRNA	3,342	249	517	171 (5%)	75
Parent	SNP	Footprint	3,342	475	671	289 (9%)	67
Parent	SNP	TE	3,342	329	319	97 (3%)	NA
Hybrid	SNP	mRNA	3,342	100	198	40 (1%)	40
Hybrid	SNP	Footprint	3,342	194	210	67 (2%)	70
Hybrid	SNP	TE	3,342	216	106	26 (1%)	9

[1]Bonferroni corrected $p < 0.05$. Binomial tests were run on mRNA and footprint data. G-tests were used to test for differential TE.

[2]DESeq was run on all of genes where at least one sample had more than zero counts (6,457 genes for the parent comparison using all reads and 4,361 genes for the SNP-based analyses). DESeq results were corrected for multiple testing using the Benjamini-Hochberg correction.



Shown are log2-transformed fold changes. D-F: hybrid ASE lines are the diagonal.

- Left: all genes.
- Middle: genes with a significant mRNA (red), footprint (blue) or both mRNA and footprint (purple) difference.
- Right column: genes with a significant TE difference. Red: genes with only a significant mRNA difference, blue: genes with only a significant footprint difference, purple: genes with both a significant mRNA and footprint difference, orange: genes with neither a significant mRNA nor a significant footprint difference.

Table 1 – Differential expression statistics • Most small magnitude AS TE effects

Comparison	Reads	Data	Analyzed genes	2-fold	Binomial / G test[1]	Intersect	DESeq[2]
Parent	All	mRNA	5,316	331	2,862	314 (6%)	189
Parent	All	Footprint	5,316	514	3,057	490 (9%)	145
Parent	All	TE	5,316	135	2,228	111 (2%)	NA
Parent	SNP	mRNA	3,342	249	517	171 (5%)	75
Parent	SNP	Footprint	3,342	475	671	289 (9%)	67
Parent	SNP	TE	3,342	329	319	97 (3%)	NA
Hybrid	SNP	mRNA	3,342	100	198	40 (1%)	40
Hybrid	SNP	Footprint	3,342	194	210	67 (2%)	70
Hybrid	SNP	TE	3,342	216	106	26 (1%)	9

[1]Bonferroni corrected $p < 0.05$. Binomial tests were run on mRNA and footprint data. G-tests were used to test for differential TE.

[2]DESeq was run on all of genes where at least one sample had more than zero counts (6,457 genes for the parent comparison using all reads and 4,361 genes for the SNP-based analyses). DESeq results were corrected for multiple testing using the Benjamini-Hochberg correction.

Table 2 – Strong *cis* effects on translation

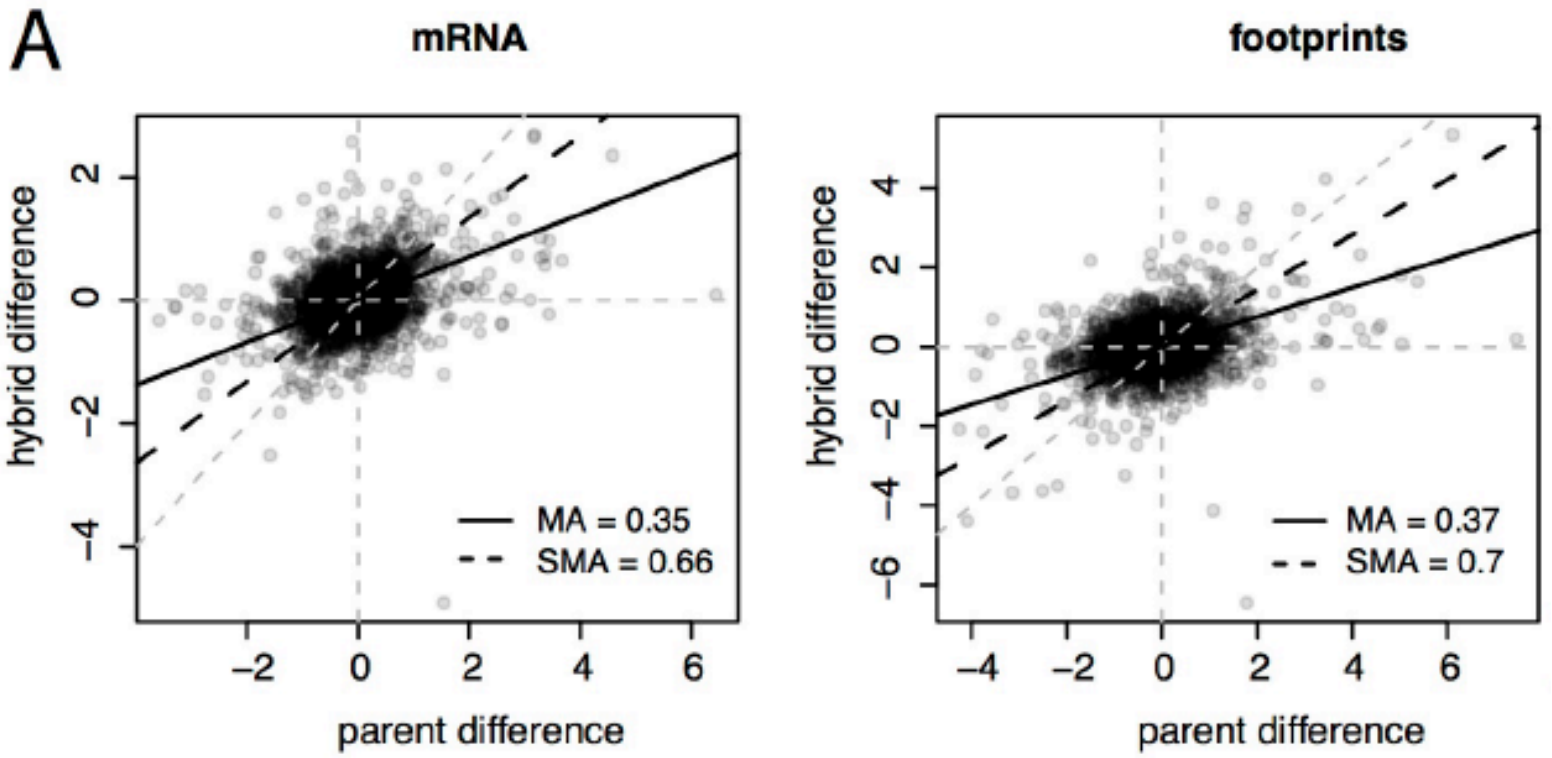
	TE		mRNA		FP		
Gene	log2(fold change)	p-value	log2(fold change)	p-value	log2(fold change)	p-value	Notes
YAL054C (*ACS1*)	-3.81	1.2E-05	0.56	1	-3.25	2.2E-04	
YBR107C (*IML3*)	1.68	2.8E-10	-1.29	5.4E-06	0.39	0.1	
YBR114W (*RAD16*)	1.10	8.3E-06	-0.86	0.01	0.25	0.2	
YDL231C (*BRE4*)	-3.41	1.8E-10	-0.23	0.02	-3.63	2.5E-08	
YDL203C (*ACK1*)	1.00	1.3E-09	-0.74	2.9E-06	0.27	6.1E-02	
YDL124W	1.01	< 2e-16	1.16	2.1E-38	2.17	1.0E-107	
YDR133C	-2.48	< 2e-16	-1.20	1.0E-152	-3.68	< 2e-16	(1)
YEL066W (*HPA3*)	1.53	4.3E-10	-0.43	0.02	1.10	1.2E-09	
YGL252C (*RTG2*)	-1.05	4.9E-06	0.23	0.5	-0.82	1.0E-05	

YGL163C (*RAD54*)	-1.39	5.9E-10	0.75	8.3E-07	-0.64	8.8E-05	
YHR195W (*NVJ1*)	-2.39	1.7E-06	0.39	0.4	-2.00	1.7E-04	
YIL165C	-3.22	< 2e-16	-0.29	0.2	-3.51	1.1E-14	(2)
YIL164C (*NIT1*)	-3.76	5.5E-07	-0.37	0.9	-4.13	3.0E-06	(2)
YJL213W	-3.41	9.4E-10	1.26	2.4E-05	-2.15	1.2E-04	
YJL132W	-1.37	3.7E-07	0.67	0.03	-0.70	3.1E-03	
YJR015W	4.16	4.2E-13	1.19	3.2E-05	5.35	1.5E-32	(3)
YJR072C (*NPA3*)	-2.14	< 2e-16	1.46	3.4E-26	-0.68	1.6E-07	
YKL163W (*PIR3*)	-1.19	1.6E-06	-0.07	0.6	-1.26	9.8E-08	
YKL095W (*YJU2*)	-1.55	2.1E-06	0.60	0.03	-0.95	1.9E-04	
YKL012W (*PRP40*)	-1.47	9.6E-06	0.24	0.9	-1.24	1.1E-04	
YLL022C (*HIF1*)	1.02	7.2E-06	-0.64	6.6E-03	0.38	0.04	
YLL007C (*LMO1*)	1.82	5.9E-08	-1.40	3.3E-08	0.42	0.03	

YLR375W (*STP3*)	-1.26	7.2E-06	0.47	0.03	-0.78	1.6E-02	
YML048W (*GSF3*)	-1.06	4.1E-08	1.13	4.7E-17	0.06	0.4	
YMR091C (*NPL6*)	1.16	2.5E-08	-1.26	8.4E-14	-0.10	0.6	
YOR304W (*ISW2*)	1.34	5.7E-12	-0.89	2.4E-08	0.45	6.1E-03	

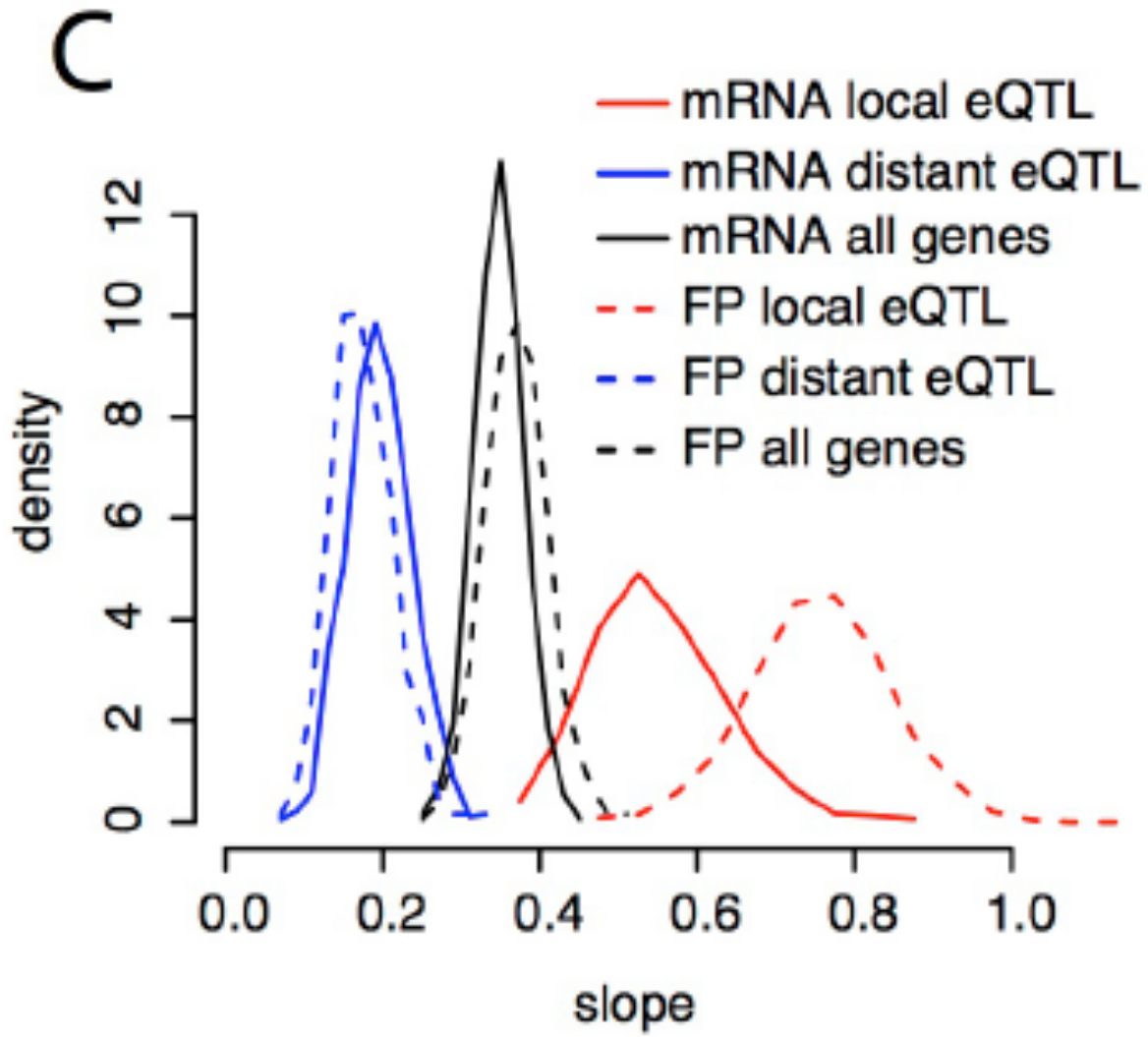
For consistency with the figures, the fold changes are log2-transformed so that zero indicates no difference and one indicates a two-fold difference. Positive values indicate higher abundance in BY compared to RM. FP: footprints. NS: neither mRNA nor footprint difference was significant. (1) "Dubious" ORF, footprint data shows translated region only partially overlaps with annotation. The TE difference is due to a nonsense SNP in BY that results in early termination compared to RM. (2) YIL165C is a "dubious" ORF immediately downstream of YIL164C (*NIT1*); in RM, these two ORFs form a single, consistently translated ORF (Figure 7C). (3) Putative protein with frameshift in RM that leads to premature termination. Note that "dubious" ORFs were not included in our analyses of nonsense SNPs so that YDR133C and YJR015W were not included in those analyses.

Figure 4 – *Cis* and *trans* effects

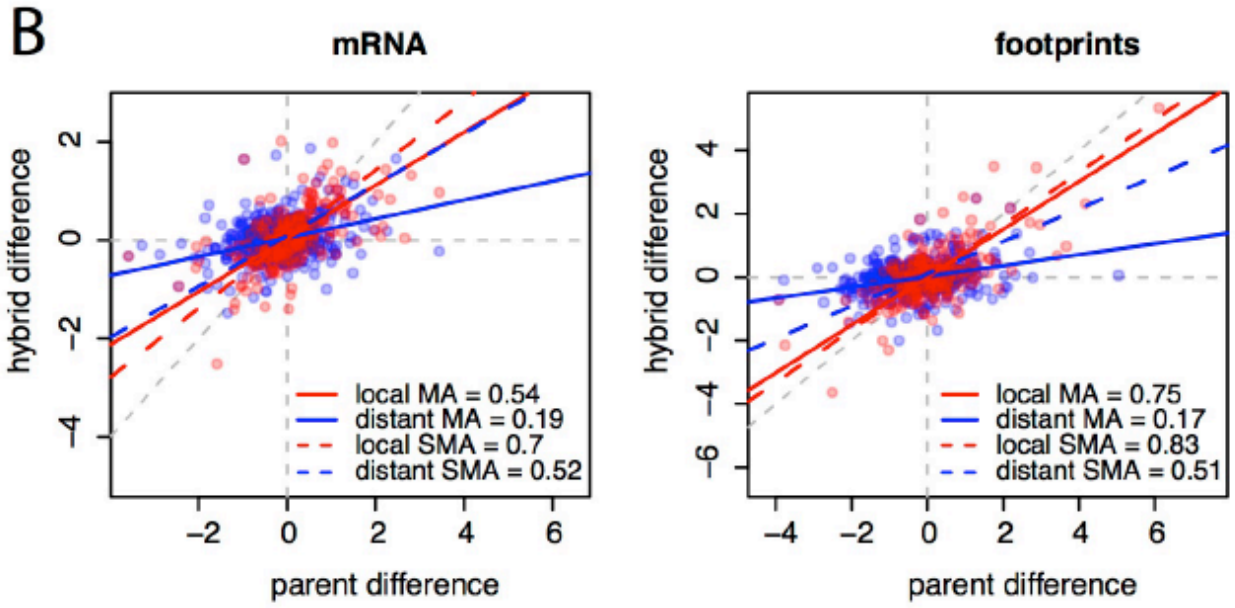


A. Parental differences (estimated based on SNP allele counts) on the x-axes, and hybrid differences on the y-axes, for all genes. Black lines show the slope of the relationship between hybrid and parental differences. The legends indicate the values of these slopes.

MA: major axis estimate; SMA: standardized major axis estimate.



C. bootstrapped distributions of MA slope estimates. Results from SMA were qualitatively similar. MA: major axis estimate; SMA: standardized major axis estimate.



B. as in A), but only for genes with eQTL in [29]. Red: genes with a local but no distant eQTL, blue: genes with a distant but no local eQTL, purple: genes with both a local and a distant eQTL. Colored lines show the respective regressions of hybrid on parental differences.

C. bootstrapped distributions of MA slope estimates. Results from SMA were qualitatively similar. MA: major axis estimate; SMA: standardized major axis estimate.

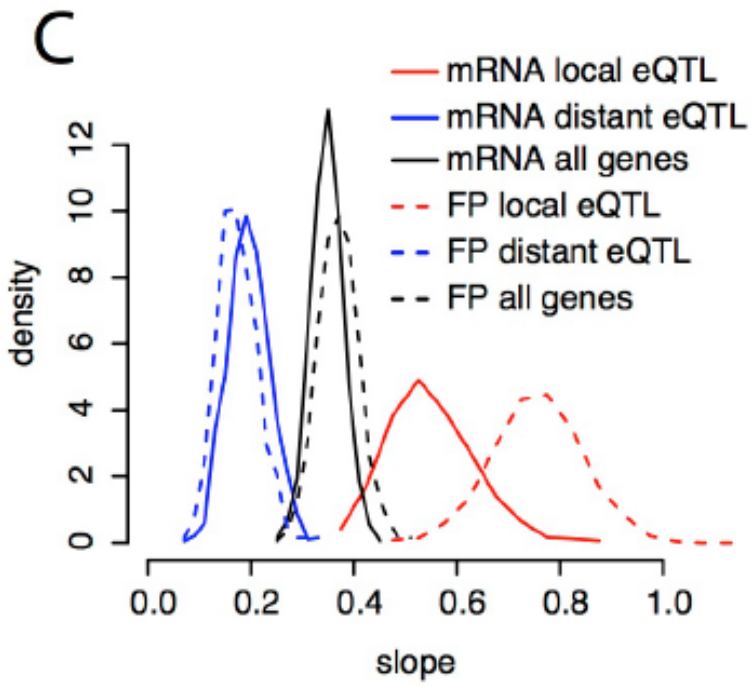
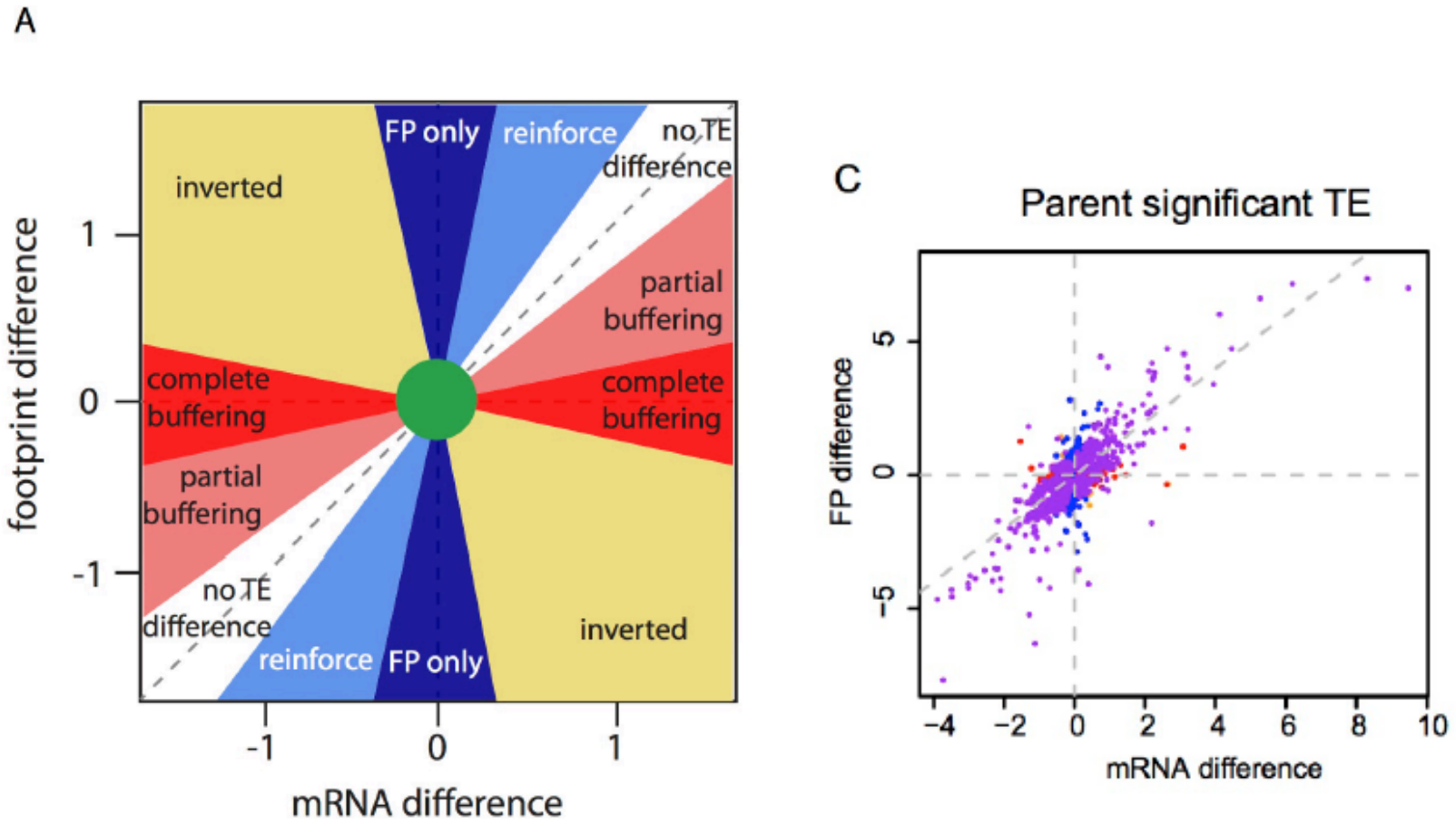


Figure 5 – Relationship between mRNA differences and footprint differences within and between species



A.Schematic representation of the possible relationships between mRNA differences and footprint differences.

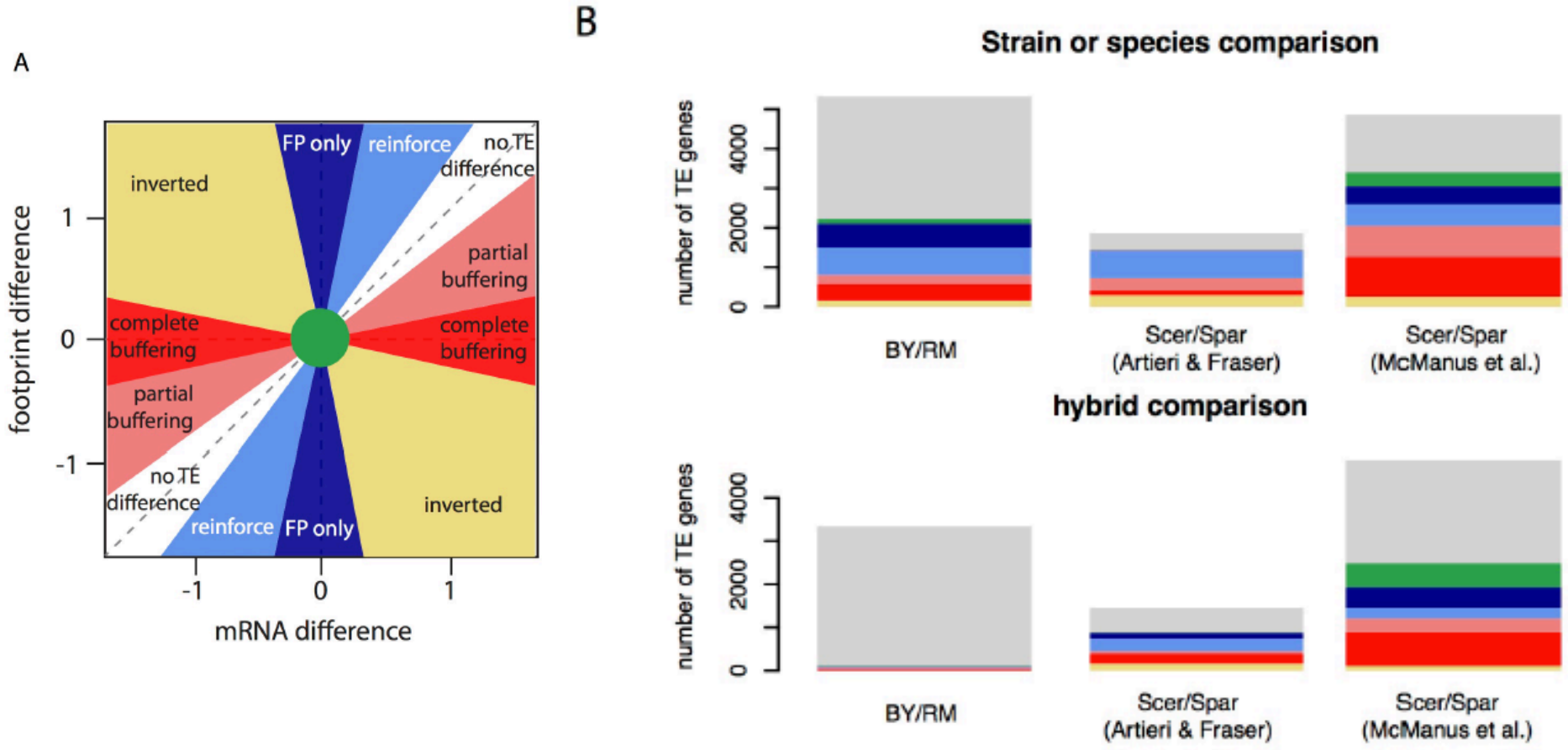
C. genes with a significant TE difference. Red: genes with only a significant mRNA difference, blue: genes with only a significant footprint difference, purple: genes with both a significant mRNA and footprint difference, orange: genes with neither a significant mRNA nor a significant footprint difference.

Table 3 – Effects of translation in genes with significant TE

Significant difference	Direction of differences	Magnitude of differences	Parent	Hybrid
Footprint only	–	–	611 (27%)	27 (25%)
mRNA and footprint	same	Footprint > mRNA	690 (31%)	10 (9%)
mRNA and footprint	same	mRNA > footprint	229 (10%)	4 (4%)
mRNA only	–	–	420 (19%)	37 (35%)
mRNA and footprint	opposite	–	159 (7%)	5 (5%)
neither	–	–	119 (5%)	23 (22%)
Sum	–	–	2,228	106

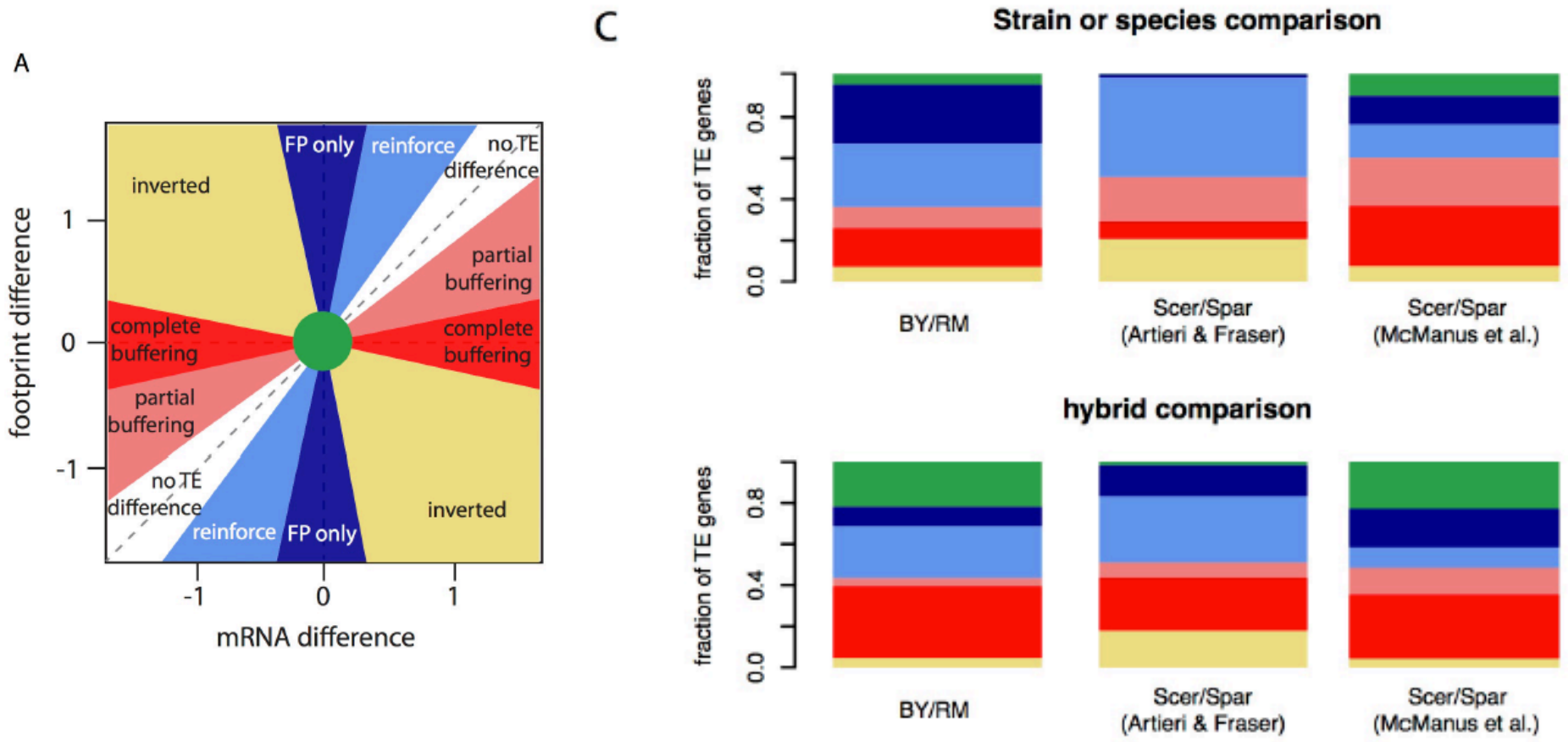
Numbers shown in this table are based on a Bonferroni-corrected significance threshold.

Figure 5 – Relationship between mRNA differences and footprint differences within and between species



B. Observed distribution of all analyzed genes in three data sets.
The color scheme is the same is in A), with light grey indicating genes without a TE difference. For BY / RM, significance was determined using a Bonferroni-corrected pvalue of < 0.05. Scer: Saccharomyces cerevisiae. Spar: Saccharomyces paradoxus. The interspecies data were analyzed from published datasets ([35] and [34])

Figure 5 – Relationship between mRNA differences and footprint differences within and between species



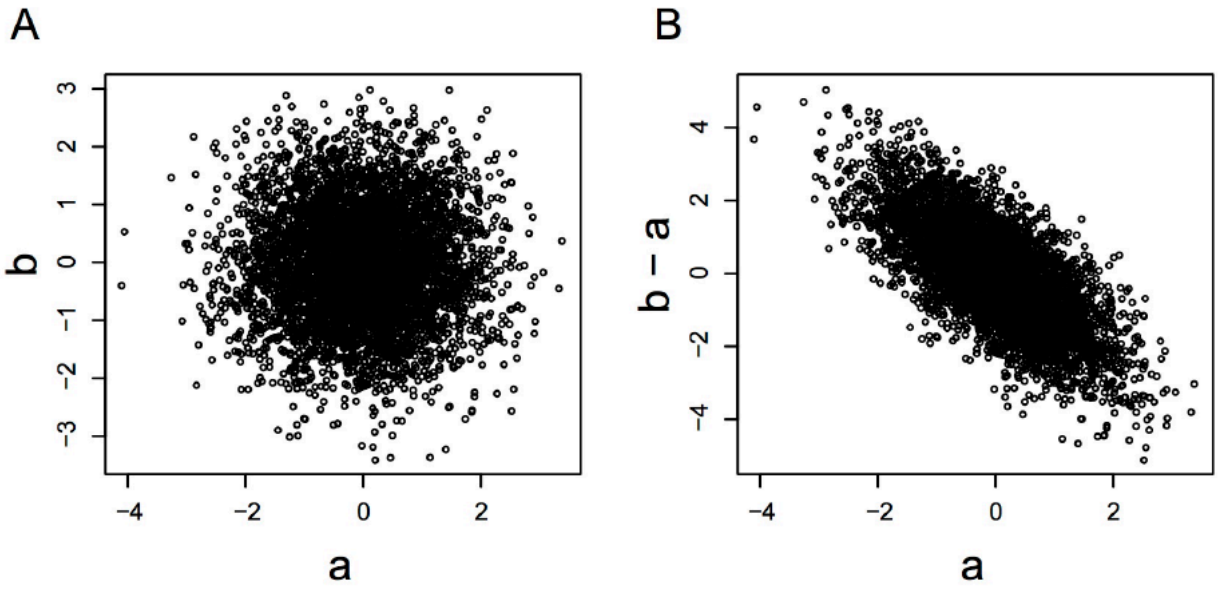
C) As in B), but showing the fraction of genes with a certain relationship among genes with a significant TE difference.

Table 4 – Tests for directional preferences in the effects of translational differences in different data sets

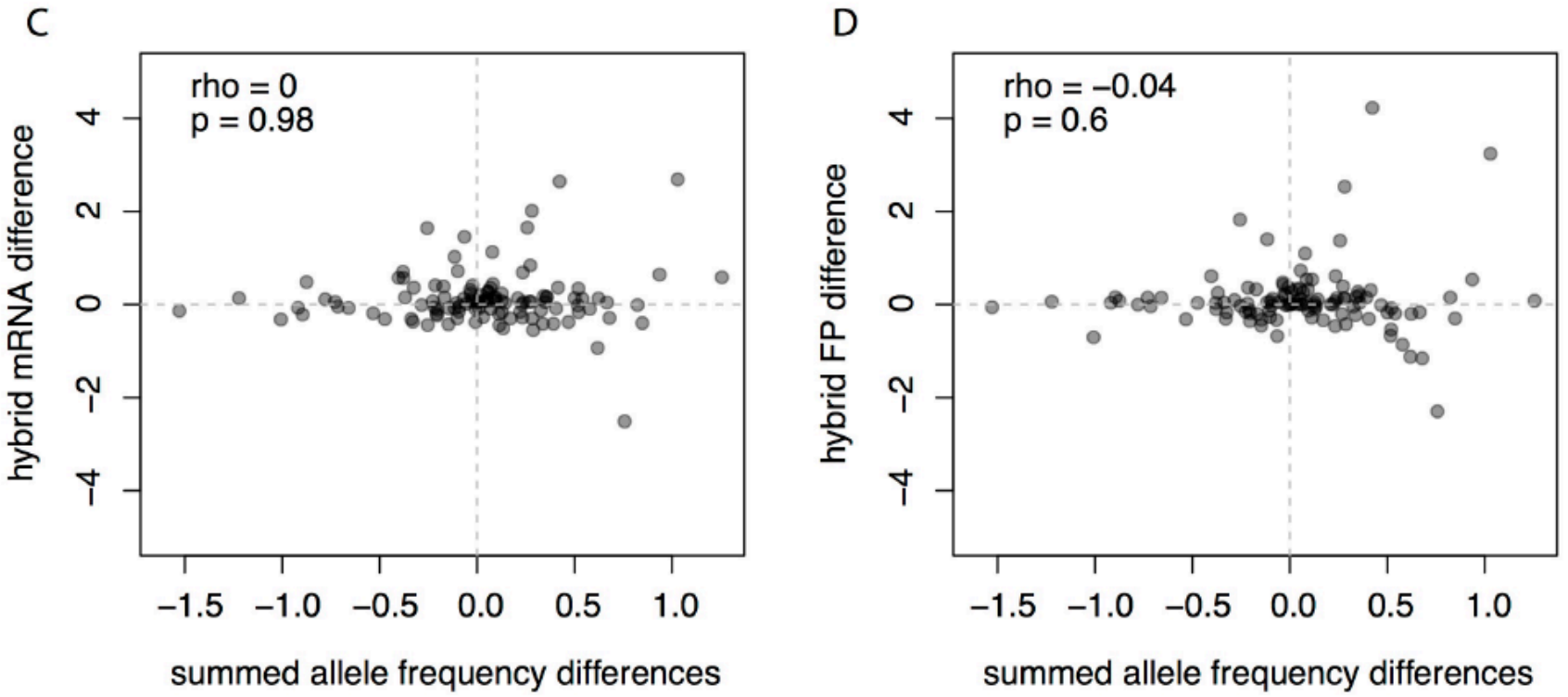
Compared groups of TE genes	BY / RM (Bonferroni)		BY / RM (FDR)		Scer / Spar (Artieri & Fraser)		Scer / Spar (McManus et al.)	
	Parents	Hybrid	Parents	Hybrid	Parents	Hybrid	Parents	Hybrid
Reinforced vs. buffered	1.1 (0.3)	0.7 (0.09)	1.3 (8e-10)	0.2 (1e-18)	1.6 (3e-13)	1.0 (0.7)	0.3 (3e-145)	0.2 (3e-118)
Reinforced vs. buffered & inverted	0.9 (0.02)	0.6 (0.03)	0.8 (3e-13)	0.2 (1e-22)	0.9 (0.2)	0.6 (9e-10)	0.3 (4e-190)	0.2 (1e-138)
Reinforced & FP only vs. buffered	2.0 (2e-49)	0.9 (0.7)	1.6 (2e-45)	1.4 (0.001)	1.6 (3e-15)	1.4 (3e-6)	0.6 (8e-52)	0.7 (9e-19)
Reinforced & FP only vs. buffered & inverted	1.6 (7e-27)	0.8 (0.3)	1.0 (0.4)	1.2 (0.03)	1.0 (0.4)	0.9 (0.2)	0.5 (7e-82)	0.6 (2e-28)

Each cell shows the ratio of TE genes where translation increases a gene expression difference (with or without "FP only" TE genes) vs. genes where translation opposes the gene expression difference (with or without inverted genes). Values greater than one indicate more genes where translation increases the gene expression difference, and values less than one indicate more genes with opposing effects. In parentheses are the p-values of a chi-squared test of the hypothesis that the two compared groups are observed at the same frequency.

Supplementary Figure S4 – Spurious correlations induced by correlations between a log ratio and its denominator

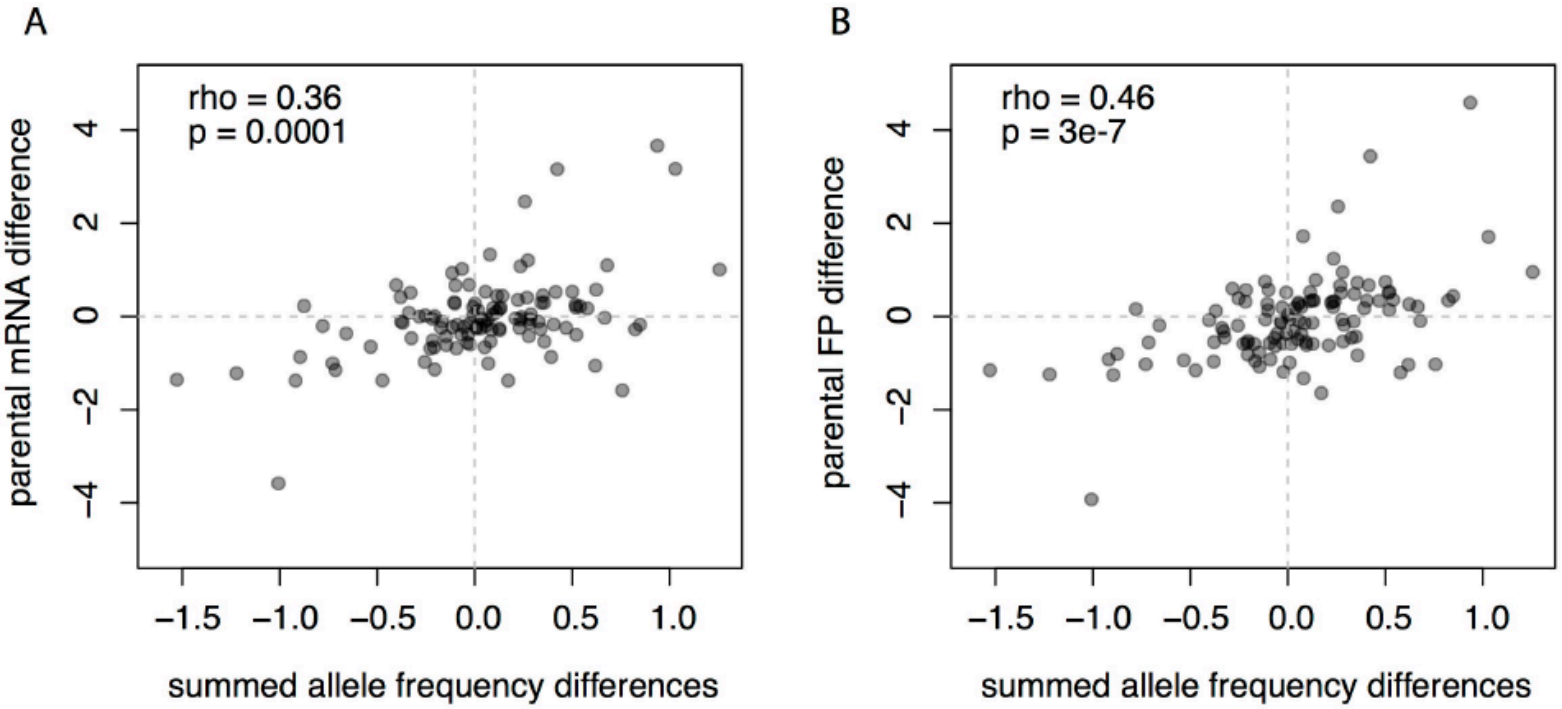


A. A scatterplot of two random samples *a* and *b* of size 5,000 from a standard normal distribution with mean = 0 and standard deviation = 1. Note that *a* and *b* are entirely uncorrelated. B. The correlation between the quantity *b* - *a* and *a* is negative and highly significant because of regression to the mean. For example, when *a* happens to be large by chance, the corresponding value of *b* will usually be closer to the mean than *a* because it is unlikely that a large value is sampled two times by chance. Therefore, the quantity *b* - *a* is systematically more likely to be less than zero for $a > 0$. If *a* and *b* are interpreted as the logarithms of mRNA and footprint differences, *b* - *a* is equivalent to the corresponding TE differences. A negative correlation between TE differences and mRNA differences is thus not by itself sufficient to infer translational buffering.



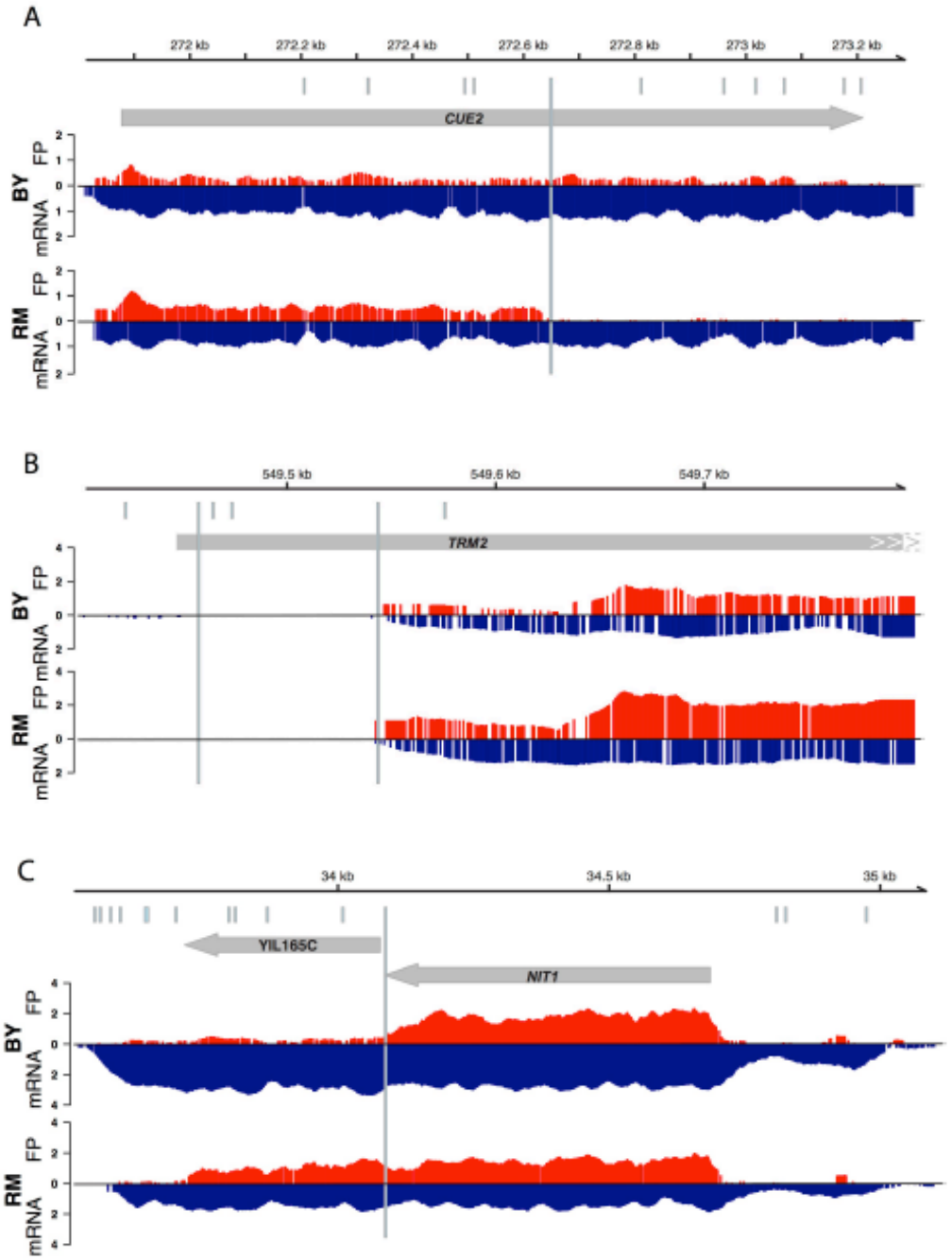
For each gene, the pQTL effects are shown as the sum of allele frequency differences at all pQTL identified by a bulk segregant approach [8]. The pQTL effect sizes shown on the x axis are identical in all four panels and are compared to A) parental mRNA differences, B) parental footprint differences, C) hybrid allele-specific mRNA differences and D) hybrid allele-specific footprint differences.

Figure 6 – Comparison of mRNA and footprint differences to pQTL effects



For each gene, the pQTL effects are shown as the sum of allele frequency differences at all pQTL identified by a bulk segregant approach [8]. The pQTL effect sizes shown on the x axis are identical in all four panels and are compared to A) parental mRNA differences, B) parental footprint differences, C) hybrid allele-specific mRNA differences and D) hybrid allele-specific footprint differences.

Figure 7 – Examples of patterns of translation at putative premature stop codons

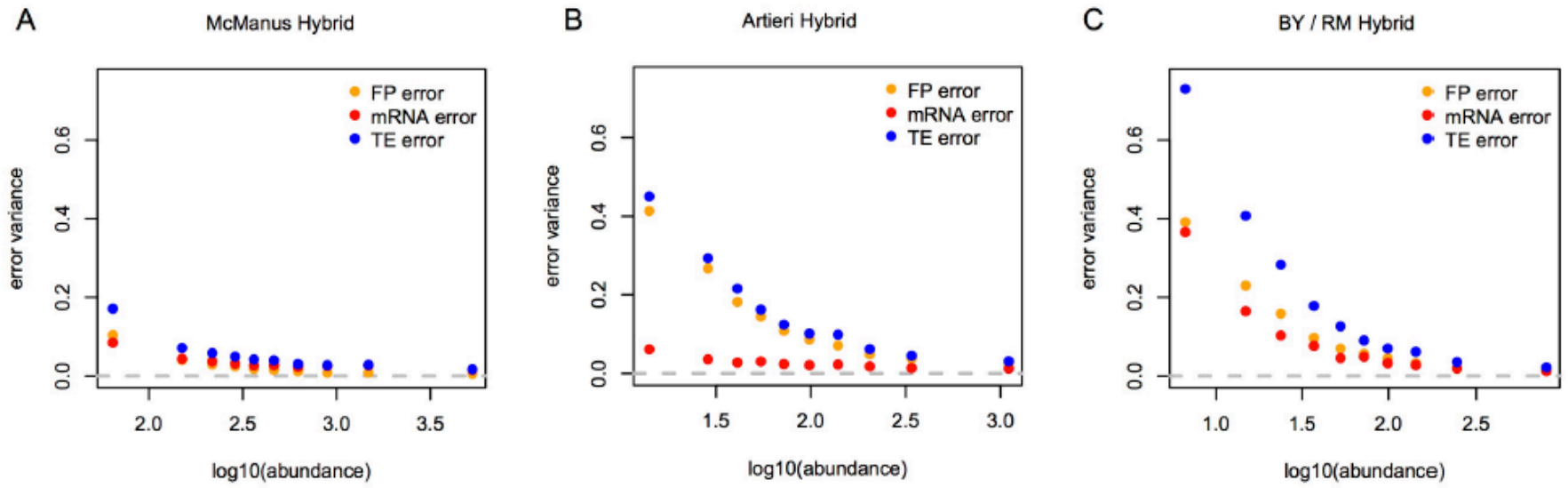


9 lead to premature termination
60%
6 close to 3’

4 upstream

2 multi-base substitutions, other amino acids change, rather than nonsense mutation

Supplementary Figure S5 – Replicate noise in different hybrid datasets



For each gene in each dataset, we calculated the log2 fold change between the alleles in the hybrid for mRNA, footprint, and TE separately for each of the two replicate datasets. The genes in each dataset were divided into 10 bins of increasing mRNA expression level. Within each bin, the average variance in the log2 fold change that is due to measurement error was calculated using the meas.est() function in the R smatr package [33] and plotted as a function of the mean abundance of the genes in the given bin. A & B: data from the published interspecies hybrid comparisons in McManus et al. [34] (A) and Artieri & Fraser [35] (B). C: Data from the BY / RM hybrid. In all datatsets, error is higher for genes with lower abundance. Footprints typically have higher error than mRNA. However, the degree to which these two data types differ varies between datasets (e.g. compare A to B). The error variance in TE is the sum of the errors in mRNA and footprints.

Discussions

- “most genes with significant differences in mRNA levels had footprint differences in the same direction”
- “most of these effects subtly modulate rather than override mRNA differences. Genetic variants that induce strong, specific effects on translation appear to be infrequent in BY and RM.”
- “most genes do not carry cis-acting variants that have large, specific influences on translation.”
- “the relative contribution of cis- vs. trans-acting variants on footprint levels was similar to that on mRNA levels.”
- “Our ribosome profiling data provides little evidence that genetic effects on translation might be responsible for these discrepancies.”
- “our results here suggest that they are more likely caused by genetic influences on protein degradation rather than on translation.”
- “genes with strain differences in TE between BY and RM tend to more often have footprint differences larger than the corresponding mRNA differences”
- “We are thus hesitant to draw strong conclusions about the relative importance of opposing / buffering or reinforcing / increasing effects of translation within and between yeast species”
- “However, the results serve as a reminder to exercise caution when interpreting the potential functional impact of variants”

Supplementary Table S1 – Sequencing and alignment statistics

Strain	Data type	Raw reads	Parent comparison: unique alignments		ASE analyses: unique & no mismatch		ASE analyses: Unique, no mismatch & spans a SNP	
Reference genome			BY	Edited BY[1]	BY	RM	BY	RM
BY parent	Footprint	189	82	-	74	-	3.7	-
BY parent	mRNA	146	53	-	46	-	2.4	-
RM parent	Footprint	222	-	151	-	129	-	6.7
RM parent	mRNA	129	-	52	-	46	-	2.5
BY/RM diploid 1	Footprint	103	-	-	33	32	0.9	0.9
BY/RM diploid 1	mRNA	98	-	-	27	26	0.7	0.7
BY/RM diploid 2	Footprint	108	-	-	44	42	1.2	1.2
BY/RM diploid 2	mRNA	113	-	-	28	28	0.8	0.8

Numbers are given in millions of reads. [1]A version of the BY reference genome with all known single nucleotide differences set to the RM allele.

Supplementary Table S2 – Genomic sources of mRNA and footprint reads in the BY parent

	mRNA		Ribosomal footprints	
	unique	repetitive	unique	repetitive
CDS	44.4M (84%)	5.8M (6.6%)	79.2M (97%)	12.2M (12%)
UTRs[1]	8.9M (17%)	66k (0.1%)	4.7M (5.7%)	145k (0.1%)
5'UTR	3.1M (35%)[1]	34k (52%)[1]	3.2M (68%)[1]	69k (48%)[1]
3'UTR	5.9M (66%)[1]	32k (48%)[1]	1.5M (32%)[1]	75k (52%)[1]
rRNA	246k (0.5%)	79M (90%)	950k (1.2%)	88M (85%)
tRNA	73k (0.1%)	466k (0.5%)	228k (0.3%)	1.2M (1.2%)
Other noncoding: snoRNA, snRNA, ncRNA	468k (0.9%)	727	217k (0.3%)	264
Total	53M	88M	82M	103M

Percentages can sum to more than 100 due to overlapping annotations

[1]percent of all UTRs

Supplementary Table S4 – Strong *cis* effects on translation identified by DESeq

	TE	mRNA		FP		Identified by binomial test?	Notes
	p-value	Log2(fold change)	p-value	log2(fold change)	p-value		
YBR012C	2.6E-06	1.12	1.4E-03	4.69	3.0E-07		
YDL231C (*BRE4*)	1.6E-05	0.16	0.6	3.23	3.1E-06	Yes	
YDR133C	9.9E-07	1.22	1.8E-06	3.66	1.8E-15	Yes	(1)
YJL108C (*PRM10*)	3.7E-05	0.35	0.5	5.05	1.2E-05		(2)
YJR015W	1.2E-11	-1.16	3.1E-03	-6.53	5.1E-23	Yes	(3)
YJR072C (*NPA3*)	2.3E-08	-1.38	1.9E-06	0.59	0.03	Yes	
YNL020C (*ARK1*)	7.5E-11	-0.69	0.1	-Inf[1]	1.5E-15		
YNR065C	8.3E-05	0.92	0.06	Inf[1]	2.3E-03		(4)
YPR192W (*AQI1*)	2.9E-06	1.51	0.2	7.30	6.1E-23		

[1]Infinite fold changes indicate that there were zero counts in one of the groups. Such genes were excluded from the binomial tests reported in the main text. Genes not identified by the binomial test all had counts below the inclusion criteria for binomial testing. (1) "Dubious" ORF, footprint data shows translated region only partially overlaps with annotation. The TE difference is due to a nonsense SNP in BY that results in early termination compared to RM. (2) Based on the parental read data, YJL108C forms one ORF in RM with its upstream neighbor YJL107C. The combined ORF in RM is interrupted by a stop mutation in BY, resulting in two separate gene annotations. (3) Putative protein with frameshift in RM that leads to premature termination. Note that "dubious" ORFs were not included in our analyses of nonsense SNPs so that YDR133C and YJR015W were not included in those analyses. (4) Similar to (1), and uncharacterized ORF that in RM forms one ORF with the upstream "uncharacterized" YNR066C.

Supplementary Table S5 – Effects of translation in TE genes derived from alternative significance criteria

Significant difference	Direction of differences	Magnitude of differences	Parent FDR	Hybrid FDR	Hybrid DESeq
mRNA and footprint	same	Footprint > mRNA	1,638 (38%)	40 (6%)	1
Footprint only	–	–	497 (12%)	193 (30%)	6
mRNA and footprint	same	mRNA > footprint	904 (21%)	20 (3%)	0
mRNA only	–	–	401 (9%)	147 (23%)	1
mRNA and footprint	opposite	–	778 (18%)	21 (3%)	0
neither	–	–	38 (1%)	217 (34%)	1
Sum	–	–	4,256	638	9

FDR: genes with q-values < 0.05 in the binomial tests.

Supplementary Table S6 – Effects of translation in TE genes in published interspecies comparisons

Significant difference	Direction of differences	Magnitude of differences	McManus Parent	McManus Hybrid	Artieri Parent	Artieri Hybrid
Footprint only	–	–	443	471	22	132
mRNA and footprint	same	Footprint > mRNA	552	249	669	287
mRNA and footprint	same	mRNA > footprint	794	319	307	66
mRNA only	–	–	1,001	778	120	229
mRNA and footprint	opposite	–	258	108	293	159
neither	–	–	357	567	4	15
Sum	–	–	3,405	2,492	1,415	888

Methods - Details

Cycloheximide resistance

- BY and RM **differ** in cycloheximide resistance at a dose **several orders of magnitude lower** than those used in the ribosome profiling protocol.
- Q: **Equally sensitive** to the high cycloheximide dose used here to block translation?
- A:

1) Grow them at 30°C in triplicates in liquid yeast nitrogen base (YNB) medium with a range of cycloheximide concentrations centered on the dose used in the ribosome profiling protocol.
2) Growth was normal in negative controls without cycloheximide, no growth within 24 hours in any of the cycloheximide doses tested.

Ribosome profiling and sequencing

- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. Science (New York, NY) 324: 218–223. doi:10.1126/science.1168978
- exceptions:

1) cells were cultured in YNB
2) the reverse-transcription was primed by ligating miRNA Cloning Linker 1 (IDT) onto the RNA fragments
3) highly abundant rRNA species were hybridized to biotinylated oligos and subtracted using streptavidin-coated DynaBeads (Invitrogen)

- Illumina HiSeq 2000 platform

SNP set for allele specific quantification

- used BWA to align (> 50X) 94 bp PE WGS data from BY & RM to sacCer3, remove PCR duplicates
- used Samtools to extract a preliminary set of SNPs with an alternative allele (AA) frequency of 1. (43,154)
- retained only biallelic SNPs where RM carries an AA and BY carries the genome reference allele.
- Restrict to SNPs where short reads can be aligned to unique positions in both reference genomes
- For each SNP, extracted 30 bp up- and downstream sequence from sacCer3, from both + & - strand. The SNP allele itself was set to the RM allele. The resulting 61 bp sequences were aligned to the RM reference.
- removed SNPs whose flanking sequences mapped to more than one position in the RM genome and SNPs where multiple SNPs mapped to the same position in the RM genome (38,706)

SNP set for allele specific quantification cont.

- remove SNPs with alignment biases
- trimmed the hybrid DNA reads to 30 bp single end, aligned them to BY & RM reference
- counted the number of reads that overlapped the BY or RM reference alleles at each SNP
- a population of SNPs with hybrid DNA allelic ratio centered at ~1/3, i.e. a 2:1 bias towards the RM genome. sequencing coverage in RM parent was twice as high as that in BY parent.
- situated at chromosome ends and likely reflect segmental duplications of distal regions in the RM strain compared to the BY reference genome.
- excluded any regions with evidence for segmental duplications in the RM but not the BY parent
- retained only the 23,412 SNPs in ORFs

Read processing and alignments

- the reference yeast genome is based on a strain with the BY background -> read (32bp) alignments from an RM sample more difficult
- uses “personalized” genome references for the BY and the RM strain to allow unbiased read mapping.
- Removed adapter, removed the first
- Reads from the BY strain were mapped to the BY reference genome (version sacCer3)
- Reads from the RM strain were mapped to a modified version of the BY reference (43,154 SNPs RM allele)

Read processing and alignments cont.

- ASE analyses, only interested in reads that span a SNP between the BY and RM.
- short reads produced in ribosome profiling are heavily biased against mapping RM reads to the BY reference
- mapped all reads to both BY & RM reference
- considered only reads mapped to one of these two reference uniquely and without mismatch
- counted the number of reads that mapped to the correct strand of the BY or the RM genome.

Quantification of mRNA and footprint abundance

- Use htseq-count to determine the genomic source of reads in the libraries (ORFs, UTRs, ncRNAs, etc.) and for the comparison between the parent strains
- ASE analysis, added the allele counts for all SNPs in a gene
- For the hybrid, we summed the counts from the two replicates
- all statistical analyses were performed directly on count data
- Translation efficiency (TE) for a gene =

log10(mRNA fraction) - log10(footprint fraction)

Statistical analyses

- Correlations were calculated as nonparametric Spearman's rank correlations to avoid making assumptions about the distributions of the data.
- two different count-based approaches to gauge statistical significance:

1) binomial tests (main text, more conservative)
2) DESeq analysis framework (Supplementary Note S2)

- the parental footprint libraries had 30% - 70% more reads than the parental mRNA libraries -> removed these differences in total read counts by downsampling

Statistical analyses - test for differential TE

	Footprints	mRNA
BY	$y_{i,\text{ BY footprints}}$	$y_{i,\text{ BY mRNA}}$
RM	$y_{i,\text{ RM footprints}}$	$y_{i,\text{ RM mRNA}}$

where $y_{i,strain}$ is the number of downsampled counts for gene *i* in *strain* (BY or RM).

- test if the ratio of footprint and mRNA counts differed between strains.
- DESeq: models the counts using a negative binomial distribution
- asks if, for a given gene, the observed mean difference between strains > expected given the variance for a gene of the given abundance
- DESeq takes into account the fact that more highly expressed genes have higher counts and higher power than less abundant genes.

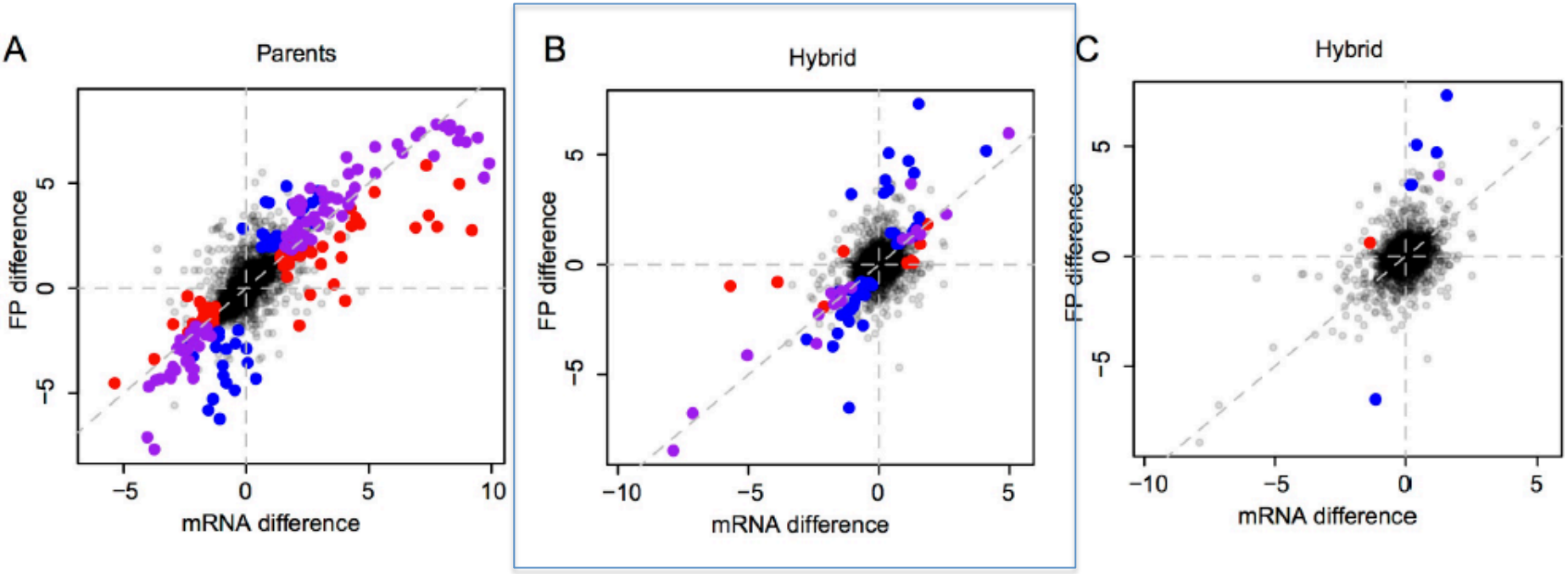
Comparison of mRNA and footprint differences to pQTL effects

- many proteins is influenced by multiple loci that segregate between the BY and the RM isolates - Albert et al.
- overlap for 114 proteins when considering only genes that can be analyzed in the hybrid
- rough expectation for the aggregate effect that the multiple pQTL -> added their effects
- used the observed difference in allele frequency at the pQTL location as a measure of effect size.

DESeq

- DESeq: models the counts using a negative binomial distribution
- asks if, for a given gene, the observed mean difference between strains > expected given the variance for a gene of the given abundance
- DESeq takes into account the fact that more highly expressed genes have higher counts and higher power than less abundant genes.

Supplementary Figure S2 – mRNA vs. footprint differences identified by DESeq



Shown are log2-transformed fold changes. A: parents, B & C: hybrid ASE. A & B: genes with a significant (Benjamini-Hochberg corrected $p < 0.05$) mRNA (red), footprint (blue) or both mRNA and footprint (purple) difference. C: genes with a significant TE difference. Red: genes with only a significant mRNA difference, blue: genes with only a significant footprint difference, purple: genes with both a significant mRNA and footprint difference.