

Volkan Lab Behavioral Genetics RNA-Seq

Charlie Soeder

2/3/2020

Contents

1	Introduction	4
2	Materials, Methods, Data, Software	4
2.1	Reference Genomes	4
2.2	Reference Annotations	4
2.3	Gene Lists	11
2.3.1	Ionotropic	11
2.3.2	Derived from GO terms	11
2.3.3	Bryson's Lists	12
2.4	Sequenced Reads	12
2.4.1	Pre-Processing	13
2.5	Mapped Reads	15
2.5.1	Raw Mapsplice	15
2.5.2	Filtered Multimap	20
2.5.3	Downsampled Multimapped	21
2.5.4	Uniquely Mapped	22
2.5.5	Spliced-Only	22
2.5.6	Alignment Process Overview	22
2.6	Assigning Reads to Annotated Features	25
2.6.1	Fruitless by exon	29
2.6.2	Fruitless by splice junction	31
2.6.3	Fruitless by intron	32
2.6.4	Ambiguous Assignment Strategy Comparison	33
2.7	Expression	34
2.8	Differential Expression Analysis.	35
2.9	Gene Ontology Enrichment	36

3 Results	37
3.1 Wildtype: Group-housed vs. Isolated	37
3.1.1 preshrunk comparison across alignment strategies	37
3.1.2 effect size: preshrunk vs shrunk	37
3.1.3 shrunk comparison across alignment strategies	38
3.1.4 differential expression overview	38
3.1.5 In relation to gene lists	38
3.1.6 Genes with top 10 most significant changes	38
3.1.7 Top 10 genes with biggest (significant) effect sizes	38
3.1.8 Top 10 highest expressed genes with significant change	38
3.1.9 rank-correllation between alignment strategies	38
3.2 Group Housed: Wildtype vs Mutants	38
3.2.1 wt vs OR47b	38
3.2.1.1 preshrunk comparison across alignment strategies	38
3.2.1.2 differential expression overview	39
3.2.1.3 Genes with top 10 most significant changes	39
3.2.1.4 Top 10 genes with biggest (significant) effect sizes	39
3.2.1.5 Top 10 highest expressed genes with significant change	39
3.2.2 wt vs 67d	39
3.2.2.1 preshrunk comparison across alignment strategies	39
3.2.2.2 differential expression overview	39
3.2.2.3 Genes with top 10 most significant changes	39
3.2.2.4 Top 10 genes with biggest (significant) effect sizes	39
3.2.2.5 Top 10 highest expressed genes with significant change	40
3.2.3 wt vs FruLexaFru440	40
3.2.3.1 preshrunk comparison across alignment strategies	40
3.2.3.2 differential expression overview	40
3.2.3.3 Genes with top 10 most significant changes	40
3.2.3.4 Top 10 genes with biggest (significant) effect sizes	40
3.2.3.5 Top 10 highest expressed genes with significant change	40
3.3 Simultaneously Modeling Housing & Genotype.	40
3.3.1 compare wildTypeHousing results to housing results from hausWtVsMut	41
3.3.2 compare RPKM to baseMean expression	43
3.3.3 Housing Contrast (Simultaneous model)	44
3.3.3.1 Top Tens	46
3.3.3.2 Gene Ontology Enrichment	47

3.3.4	47b vs wt (Simultaneous model)	48
3.3.4.1	Top Tens	49
3.3.4.2	Gene Ontology Enrichment	50
3.3.5	67d vs wt (Simultaneous model)	52
3.3.5.1	Top Tens	53
3.3.5.2	Gene Ontology Enrichment	54
3.3.6	FruLexA/Fru440 vs wt (Simultaneous model)	56
3.3.6.1	Top Tens	57
3.3.6.2	Gene Ontology Enrichment	59
3.3.7	Fruitless-less	60
3.3.7.1	Perturbation to Housing Contrast	62
3.3.7.2	Perturbation to 47b1 Contrast	64
3.3.7.3	Perturbation to 67d Contrast	66
3.4	Comparing Expression Changes from Housing with Expression Changes from Genotype	68
3.4.1	Housing & OR47b	68
3.4.2	Housing & 67d	71
3.4.3	Housing & FruLexFru440	73
3.5	Comparing Expression Changes Between Mutants	75
3.5.1	Fru & 67d	75
3.5.2	Fru & 47b	75
3.5.3	47b & 67d	75
3.6	Mutually Significant Differential Expression Overview	76
3.7	Focus on Fruitless	78
3.7.1	By Exon	78
3.7.1.1	Ambiguous Read Assignment: None	78
3.7.1.2	Ambiguous Read Assignment: All	81
3.7.2	By Exon Junction	84
3.7.3	By Intron	88
3.7.4	Overall	93
3.8	Impact of 47b2 on 88a	93
3.8.1	Top Tens	94
3.8.2	Gene Ontology Enrichment	95
3.9	Two Days' Difference	97
3.9.1	Top Tens	99
3.9.2	Gene Ontology Enrichment	101

4 Bibliography

102

1 Introduction

words words

2 Materials, Methods, Data, Software

generic overview words

2.1 Reference Genomes

The dm6.13 reference genome was used for read alignment:

Table 1. Size and Consolidation of Reference Genomes
Drosophila Melanogaster

number bases	$144M$
number contigs	$2K$

2.2 Reference Annotations

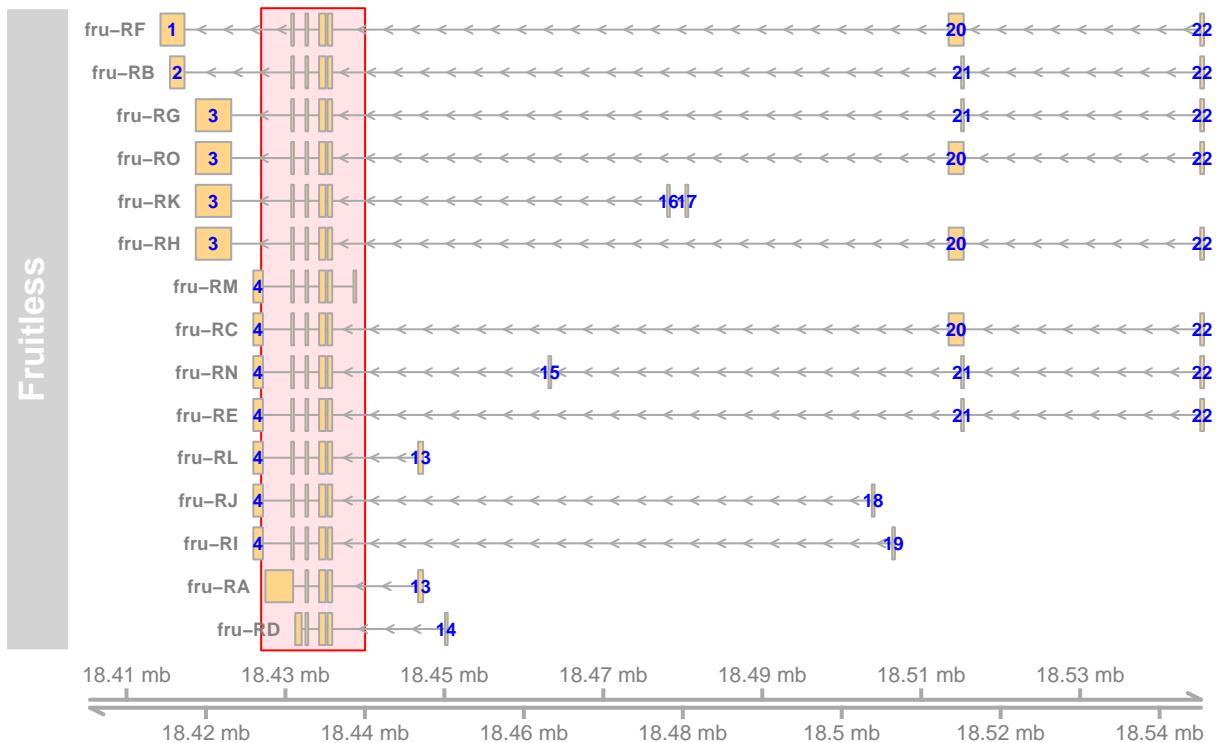
The dm6 reference annotations were used to define gene locii for differential expression analysis:

Table 2. Reference Annotations and their Sizes

annot	size (bp)		
	average	total	total count
dm6_genes	$5.8K$	$102.2M$	$17.7K$
dm6_repeats	197.1	$25.5M$	$129.4K$
fru_exons	939.3	$20.7K$	22
fru_intron	939.3	$20.7K$	22
fru_junct	939.3	$20.7K$	22

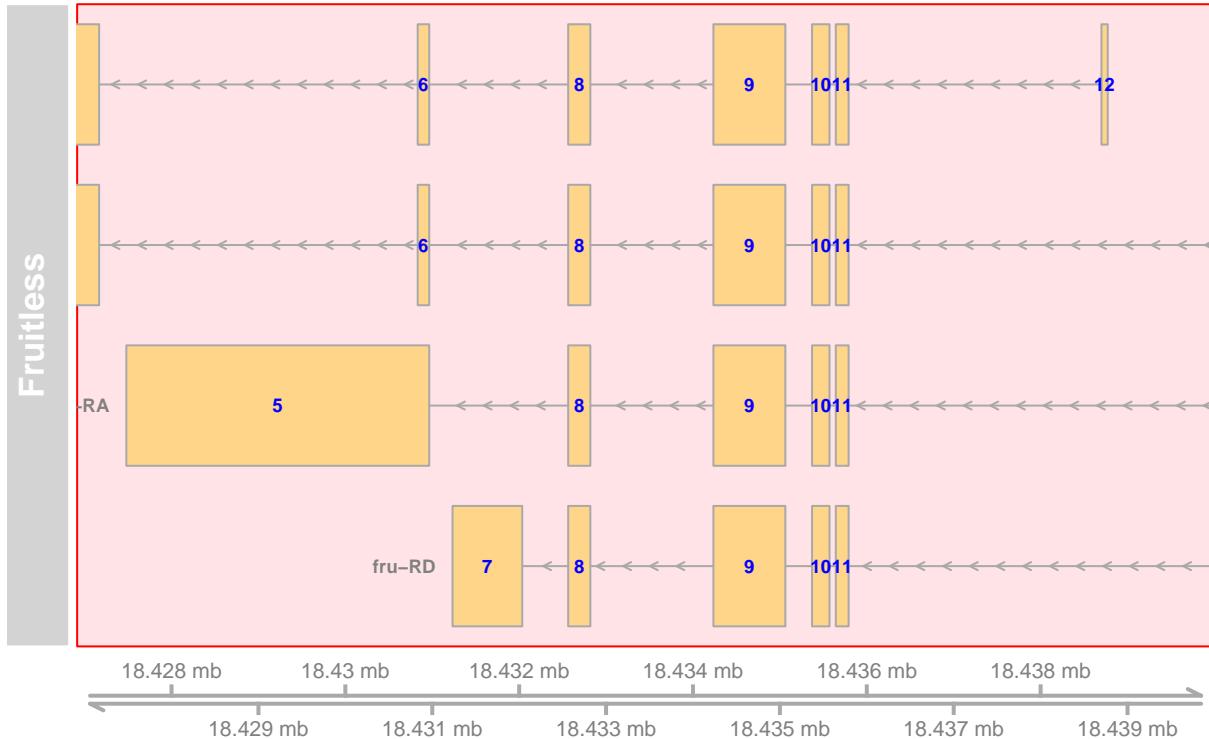
In addition to the genome as a whole, the gene Fruitless was given particular attention.

Figure 1. Fruitless gene model: exons and transcripts



```
## pdf
## 2
```

Figure 1 a. Fruitless gene model: exons and transcripts (detail)



```
## pdf
## 2
```

In order to focus on exon usage in Fru, the GTF entry was selected and decomposed into individual records per exon:

Table 3. Fru exons by Name
(chromosome 3R)

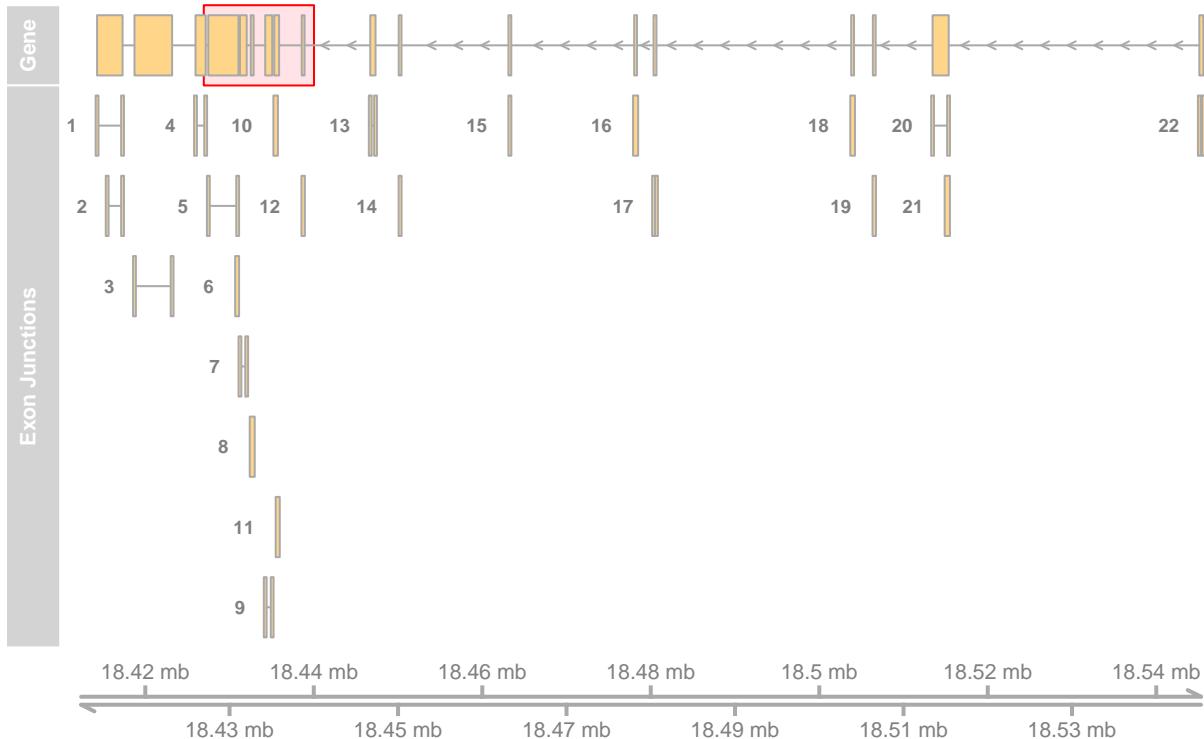
	start	stop
exon_1	18414273	18417301
exon_2	18415473	18417301
exon_3	18418716	18423183
exon_4	18425959	18427167
exon_5	18427480	18430965
exon_6	18430832	18430965
exon_7	18431233	18432035
exon_8	18432564	18432819
exon_9	18434235	18435063
exon_10	18435370	18435571
exon_11	18435643	18435791
exon_12	18438700	18438772
exon_13	18446701	18447330
exon_14	18450235	18450255
exon_15	18463267	18463282
exon_16	18478064	18478333

exon_17	18480328	18480677
exon_18	18503846	18504067
exon_19	18506494	18506563
exon_20	18513451	18515344
exon_21	18515052	18515344
exon_22	18545113	18545587

```
cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf | grep "fru.*exon"
cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf | grep "fru.*gene"
cat fru.test.gtf.exon fru.test.gtf.gene | bedtools sort > utils/annotations/fru_ex.gtf
cat fru.test.gtf.exon | cut -f 1,4,5,7,9 | tr -d '"' | tr -d ";" | sed -e 's/gene_id //g' | awk '{print $1}'
```

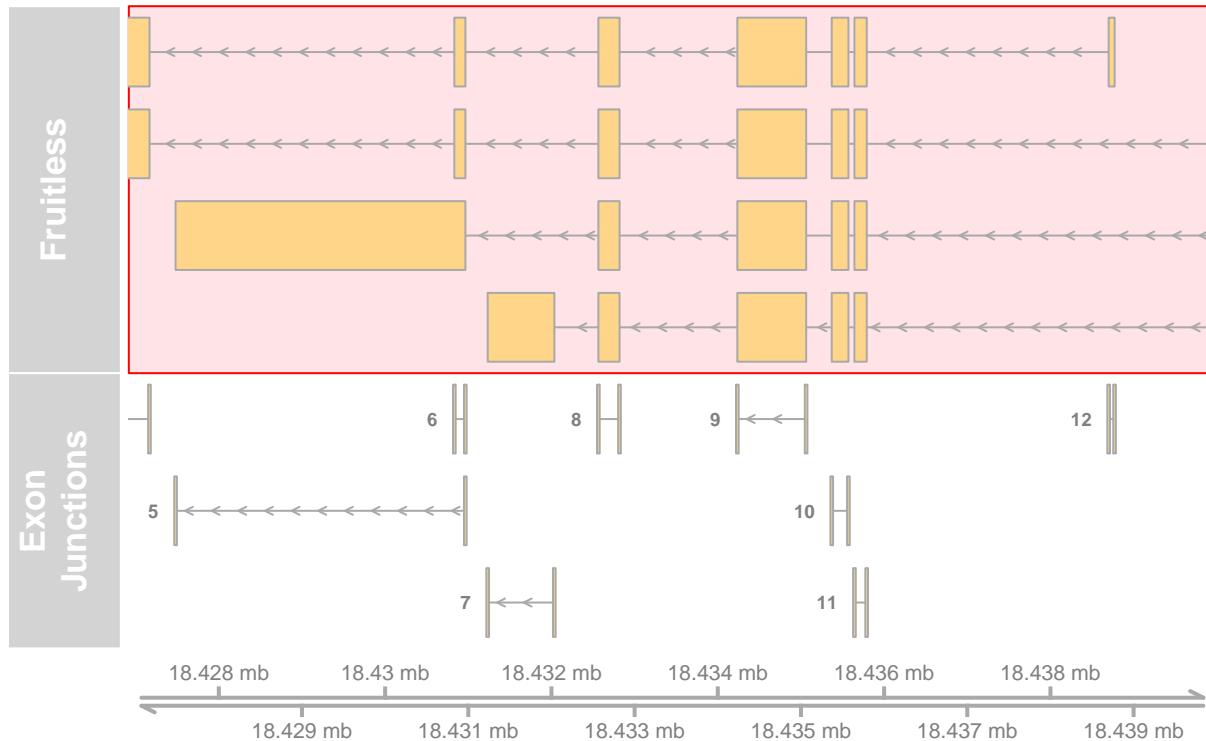
This gave the “fru_exons” annotation, to use for by-exon read counting. A further annotation, “fru_junct”, was constructed by removing all of each exon except for splice junctions, ie, the 1bp boundaries of each exon which isn’t a transcription start or stop site:

Figure 2. Fruitless gene model: junctions



```
## pdf
## 2
```

Figure 2 a. Fruitless gene model: junctions (detail)



```

## pdf
## 2

cat utils/annotations/fru_ex.gtf | grep -w gene > utils/annotations/fru_exonEdges.gtf
cat utils/annotations/fru_ex.gtf | grep -w gene | cut -f 1,2 > utils/annotations/fru_exonEdges.gtf.front
paste << cat utils/annotations/fru_ex.gtf | grep -w gene | cut -f 6- >< cat utils/annotations/fru_ex.gtf | grep -w gene | awk '{print $1}' > utils/annotations/fru_exonEdges.gtf.back
paste utils/annotations/fru_exonEdges.gtf.front <<(cat utils/annotations/fru_ex.gtf | grep -w gene | awk '{print $1}' > utils/annotations/fru_exonEdges.gtf.back)
paste utils/annotations/fru_exonEdges.gtf.back <<(cat utils/annotations/fru_ex.gtf | grep -w gene | awk '{print $1}' > utils/annotations/fru_exonEdges.gtf.back)

cat utils/annotations/fru_introns.gtf | grep -w "exon" | awk '{print"chr"$0}' | cut -f 1,4,5 | sort | uniq > edges.bed
cat utils/annotations/fru_exonEdges.gtf | grep -w "exon" | cut -f 1,4,5 | sort | uniq > edges.bed
bedtools intersect -v -a edges.bed -b introns.bed > TSS_startStop.bed

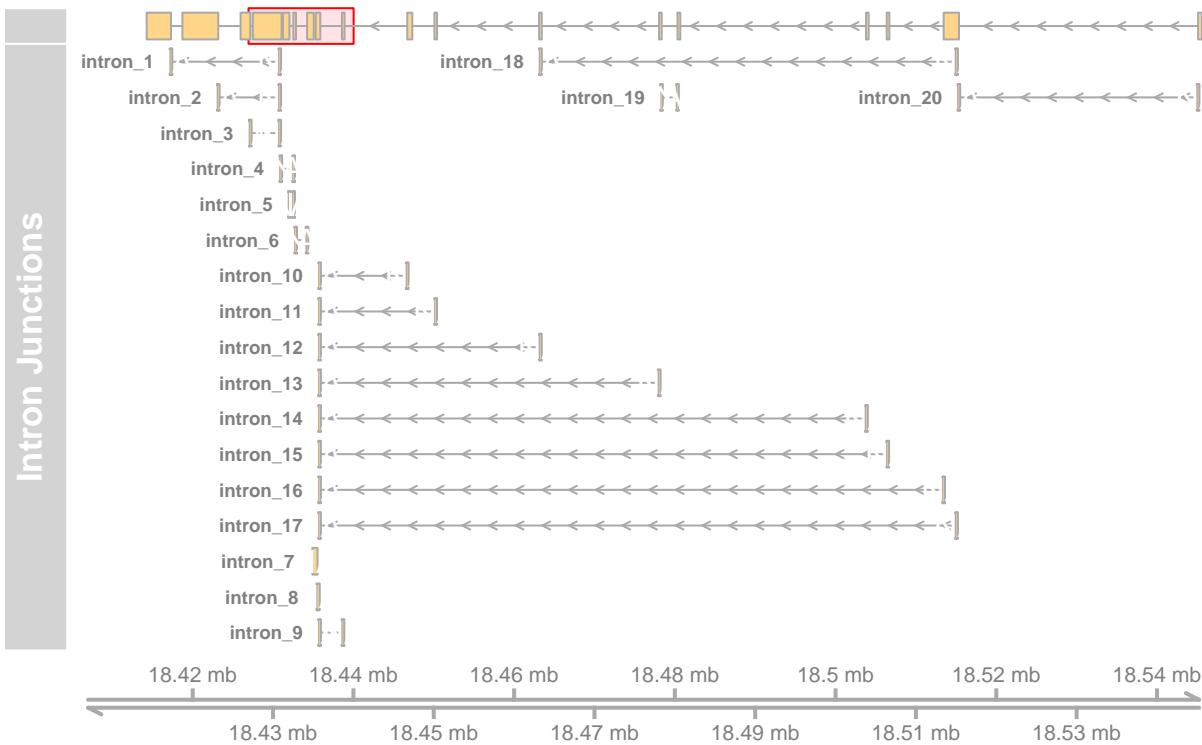
bedtools subtract -a utils/annotations/fru_exonEdges.gtf -b TSS_startStop.bed > utils/annotations/fru_exonEdges.gtf

```

Because a splice site represents two semi-independent exons but one intron, another annotation, “fru_intron”, was constructed consisting of the introns in Fruitless. The same 1-bp subintervals were used as in “fru_junct”, but in this case they were organized by the intron they bounded rather than by the exon:

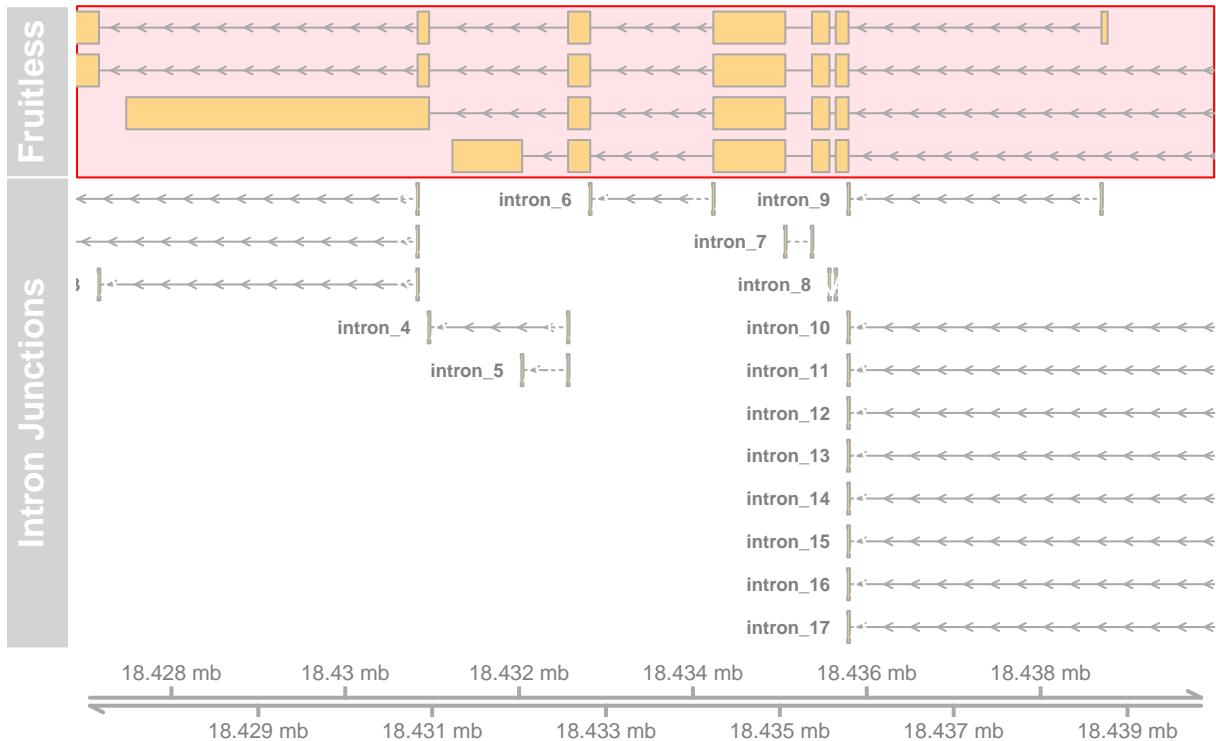
Intron Junctions

Figure 3. Fruitless gene model: introns



```
## pdf  
## 2
```

Figure 3 a. Fruitless gene model: introns (detail)



```

## pdf
## 2

rm -f coords.all
for transcript in $(cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf
echo $transcript;

cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf

head -n 1 coords.tmp | cut -f 2 | awk '{print "0\t$0}' >> coords.all
tail -n 1 coords.tmp | cut -f 2 | awk '{print $0"\t0"}' >> coords.all

paste <(cut -f 2 coords.tmp | tail -n +2 ) <(cut -f 1 coords.tmp | head -n -1 ) >> coords.all

done

cat coords.all | sort | uniq | grep -v -w 0 |awk -F'\t' 'NR>0{${0=$0}\tintron_"$NR} 1'> coords.unq

cat coords.unq | awk '{print"3R\tFlybase\tgene\t"$1"\t"$2"\t.\t-\t.\tgene_id ~"$3"~;"}' | tr '~' ' '
cat coords.unq | awk '{print"3R\tFlybase\tgene\t"$1"\t"$2"\t.\t-\t.\tgene_id ~"$3"~;"}' | tr '~' ' '

cat <(cat coords.unq | grep -v -w 0 | awk '{print"3R\tFlybase\tgene\t"$1"\t"$2"\t.\t-\t.\tgene_id ~"$3"~;"}
(pull these into an annotation-builder rule?)

fru_junct and fru_intron annotations were used with the *_SplicedOnly alignments (section ~~)

```

2.3 Gene Lists

In addition to the full annotations, subsets containing prespecified genes of interest will also be used.

Here are those subsets and their sizes:

Table 4. Predefined Subsets of Gene Annotation

measure	brysonPriority	brysonsList	histoneMod	ionotropic	mating	nervSysDev	synapseSig
total count	25	35	8	246	3	93	1
annotated count	54	35	8	246	3	90	1
percent of annotations	0.3%	0.2%	0.0%	1.4%	0.0%	0.5%	0.0%
total size	679.5K	3.2M	46.9K	3.7M	5.0K	1.8M	27.1K
avg size	12.6K	90.7K	5.9K	15.2K	1.7K	19.8K	27.1K
percent genome size	0.5%	2.2%	0.0%	2.6%	0.0%	1.2%	0.0%
percent annotation size	0.7%	3.1%	0.0%	3.7%	0.0%	1.7%	0.0%

2.3.1 Ionotropic

A list of ionotropic receptors supplied by Corbin via Flybase & George et al 2019 (email 28 May 2019). This contained 335 entries, some with multiple genes, some not unique. Once merged & uniques : 246 Annotation symbols (CGxxxxx) converted to FlyBase gene names (FBgnxxxx) using flybase ID converter (<http://flybase.org/convert/id>)

239 converted cleanly; 5 had duplicate conversions and were corrected by hand:

```
CG11430 is FBgn0041585, not FBgn0050323
CG43368 is FBgn0263111, not FBgn0041188
CG8885 is FBgn0262467, not FBgn0081377
CG9090 is FBgn0034497, not FBgn0082745
CG9126 is FBgn0045073, not FBgn0053180
```

Two were corrected to be consistent with the dm6_genes annotation:

```
CG9907 (para), is listed as FBgn0264255 not FBgn0285944
CG42345 (straw) is listed as FBgn0259247 (laccase2)
```

2.3.2 Derived from GO terms

Sub Pull out by particular GO terms?

- o Nervous system development - http://flybase.org/cgi-bin/cvreport.pl?rel=is_a&id=GO:0007399
- o Mating - http://flybase.org/cgi-bin/cvreport.pl?rel=is_a&id=GO:0007618
- o Histone modification - http://flybase.org/cgi-bin/cvreport.pl?rel=is_a&id=GO:0016570
- o Dna-binding transcription factor - <http://flybase.org/cgi-bin/cvreport.pl?id=GO%3A0003700>
- o Synaptic signaling - http://flybase.org/cgi-bin/cvreport.pl?rel=is_a&id=GO:0099536
- o Synapse organization - <http://flybase.org/cgi-bin/cvreport.pl?id=GO%3A0050808>

(Bryson, email 24 July 2019)

melanogaster-specific genes with these GO terms were retrieved using the FlyBase QueryBuilder.

Nervous System Development:

nrd, FBgn0002967, no annotated gene model
1(2)23Ab, FBgn0014978, same
aloof, FBgn0020609, same
Imp, FBgn0285926, is FBgn0262735

Mating:

Only three, but all good

synapse signalling

1 gene

Histone modification, DNA trans factor act, synapse org

MT

2.3.3 Bryson's Lists

Interest: (email, 29 Oct 2019)

Neverland: annotated as FBgn0259697, not FBgn0287185

Priority: (email, 5 Nov 2019; 7 Nov 2019)

2.4 Sequenced Reads

The sequenced reads covered three replicates each of 5 experimental conditions. The conditions included varying genotype, housing, and age (all RNA was collected from antenna tissue).

Table 5. Experimental Conditions and Replicates

genotype	housing	age (days)	tissue	# replicates
47b1	group	5	antennae	3
47b1	group	7	antennae	3
47b2,88a	group	5	antennae	3
67d	group	7	antennae	3
88a	group	5	antennae	3
FruLexaFru440	group	7	antennae	3
wt	group	7	antennae	3
wt	isolated	7	antennae	3

These samples will allow direct comparison between wild-type flies reared under group and isolated conditions, as well as comparisons between group-raised wild-type flies and two kinds of mutants (67d and 47b1) at day 7:

Table 5a. Genotype & Housing Comparison
(replicate count)

variable

day	tissue	genotype	group	isolated
7	antennae	47b1	3	0
7	antennae	67d	3	0
7	antennae	FruLexaFru440	3	0
7	antennae	wt	3	3

These samples also allow for direct comparison between mutant genotypes (47b1, 88a, and 47b2/88a) at day 5, and for a comparison between the same genotype (47b1 mutant) at two developmental stages:

Table 5b. Genotype & Time Comparison
(replicate count)

housing	tissue	day	mutant genotypes				
			47b1	47b2,88a	67d	88a	FruLexaFru440
group	antennae	5	3	3	0	3	0
group	antennae	7	3	0	3	0	3

Moreover, samples taken at the same timepoint in different genotypes allow the effect of one mutation (88a) to be studied in two different genomic backgrounds (with and without the 47b2 mutation).

In addition to the novel reads, RNA-Seq from drosophila melanogaster antennae were downloaded from NCBI (PRJNA388757; Shiao et al. (2015)), one annotated as male and the other as female. These will be compared to the unpublished samples to try to confirm the sex of the flies they came from.

2.4.1 Pre-Processing

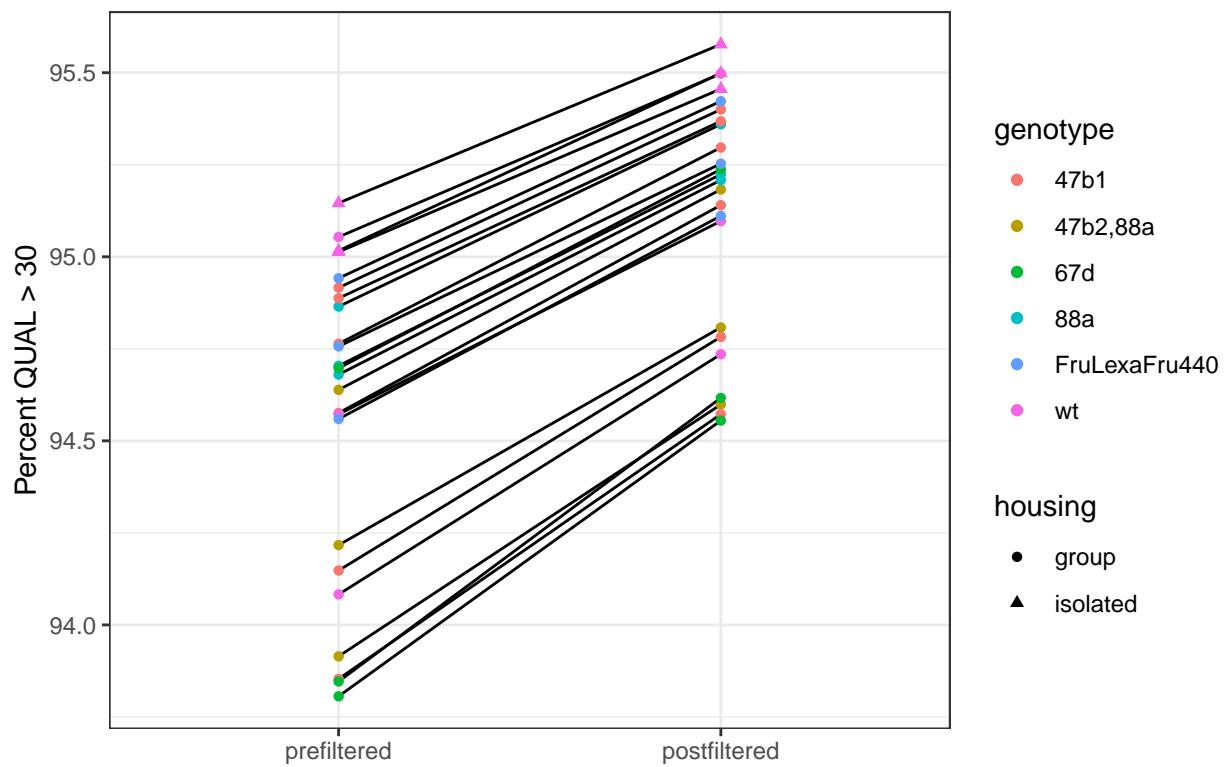
These reads were preprocessed with FASTP (S. Chen et al. 2018) for quality control and analytics.

Starting FASTQ files contained a total of 721M reads; after QC, this dropped to 710M.

Table 6. Read Retention Rate during Preprocessing

	minimum	average	maximum
prefiltered	21M	30M	43M
postfiltered	20M	30M	43M
percent retention	98	98	99

Figure 4. Percent of Reads with a mean QUAL > 30



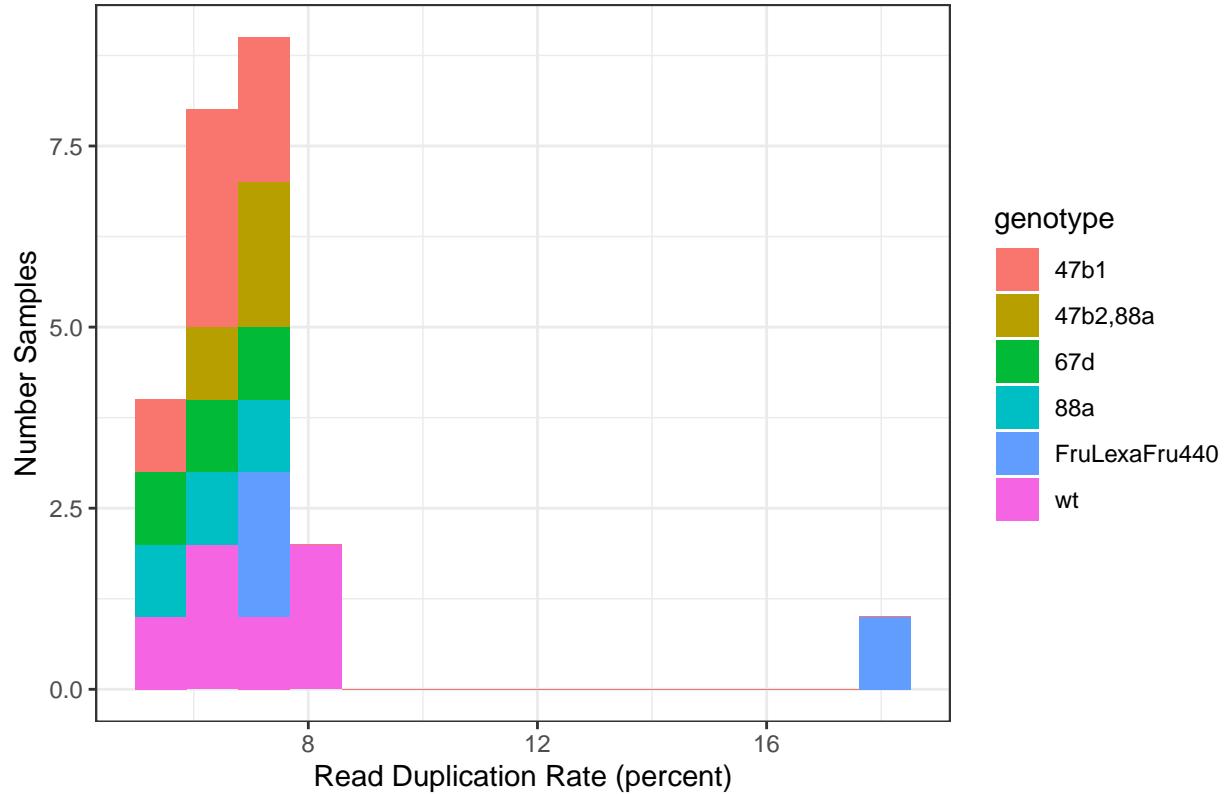
```
## pdf
## 2
```

Duplicate reads were also detected

Table 7. Percentage Duplication
FASTP estimate

	minimum	average	median	maximum
	5.2	7.1	6.7	17.9

Figure 5. Duplication Histogram (FASTP estimate)

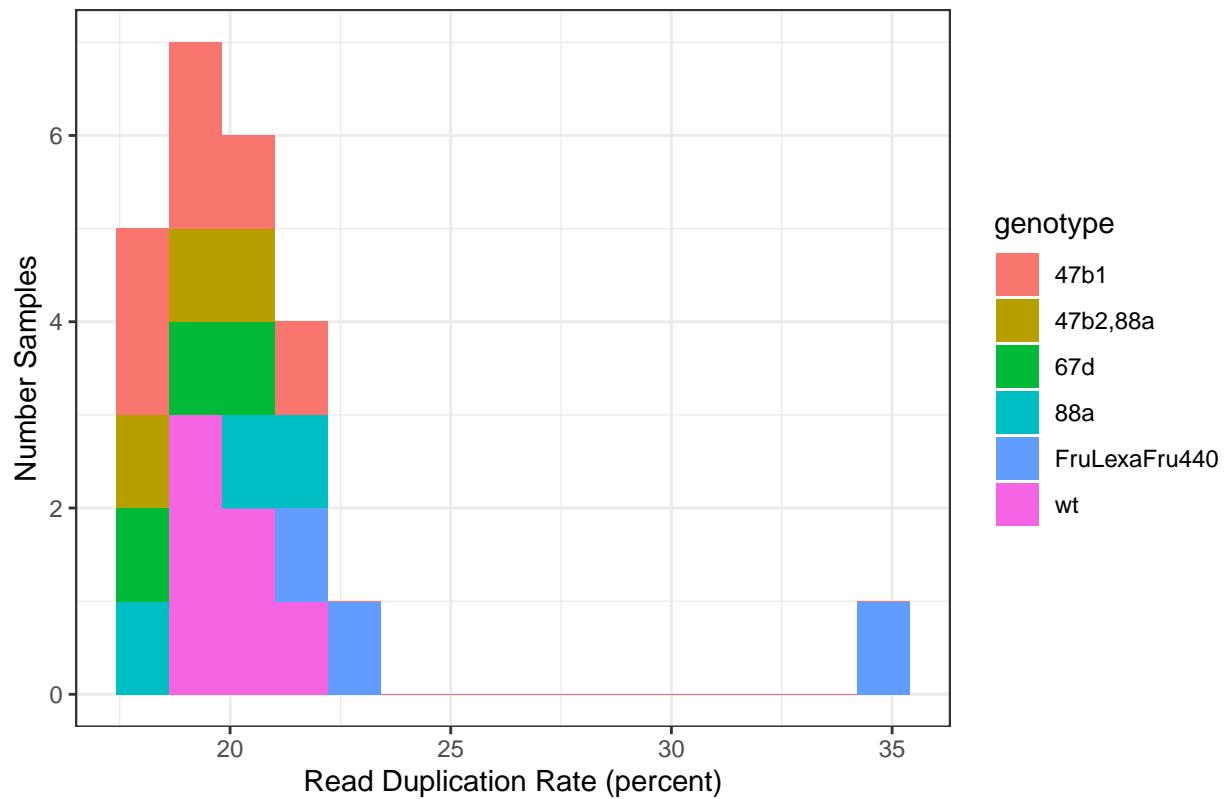


rep	day	total reads	reads mapped	percent mapped
group - 47b1				
1	5	32.0M	31.9M	99.7%
2	5	28.2M	28.0M	99.4%
3	5	24.4M	24.2M	99.0%
1	7	32.1M	31.9M	99.5%
2	7	28.9M	28.8M	99.7%
3	7	24.3M	24.3M	99.6%
group - 47b2,88a				
1	5	20.3M	20.2M	99.5%
2	5	31.7M	31.6M	99.5%
3	5	24.7M	24.5M	99.3%
group - 88a				
1	5	37.0M	36.8M	99.7%
2	5	30.4M	30.2M	99.6%
3	5	36.2M	36.1M	99.7%
group - 67d				
1	7	25.1M	25.0M	99.6%
2	7	31.2M	31.0M	99.5%
3	7	24.1M	24.0M	99.6%
group - wt				
1	7	42.6M	42.2M	99.0%
2	7	31.5M	31.0M	98.5%
3	7	30.2M	29.9M	99.0%
isolated - wt				
1	7	30.7M	30.4M	99.2%
2	7	27.2M	27.1M	99.5%
3	7	33.8M	33.5M	99.0%
group - FruLexaFru440				
1	7	22.0M	21.7M	98.9%
2	7	30.7M	30.4M	99.1%
3	7	30.7M	30.4M	99.1%

Table 10. Percent of Duplicate Reads
raw mapsplice output

maximum	mean	median	minimum
34.4%	20.5%	20.0%	17.6%

Figure 6. Duplication Histogram (Raw Mapsplice Alignment)



```
## pdf  
## 2
```

Although Samtools marks duplicates at a higher rate than FASTP, the estimates are correlated; in particular, both agree that FruLexa/Fru440 day 7 replicate 1 is a highly duplicated outlier. The NCBI reads are anomalous.

Figure 7. Comparison of Duplication Rate Estimates

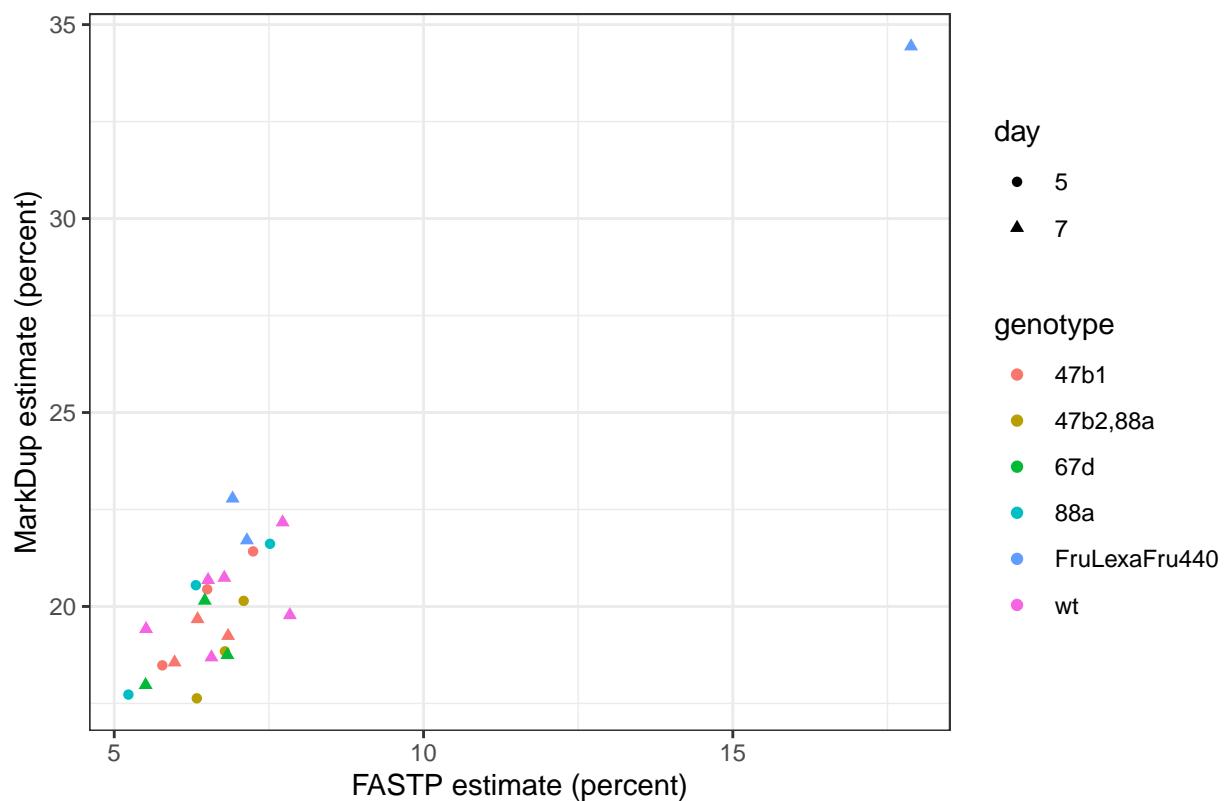
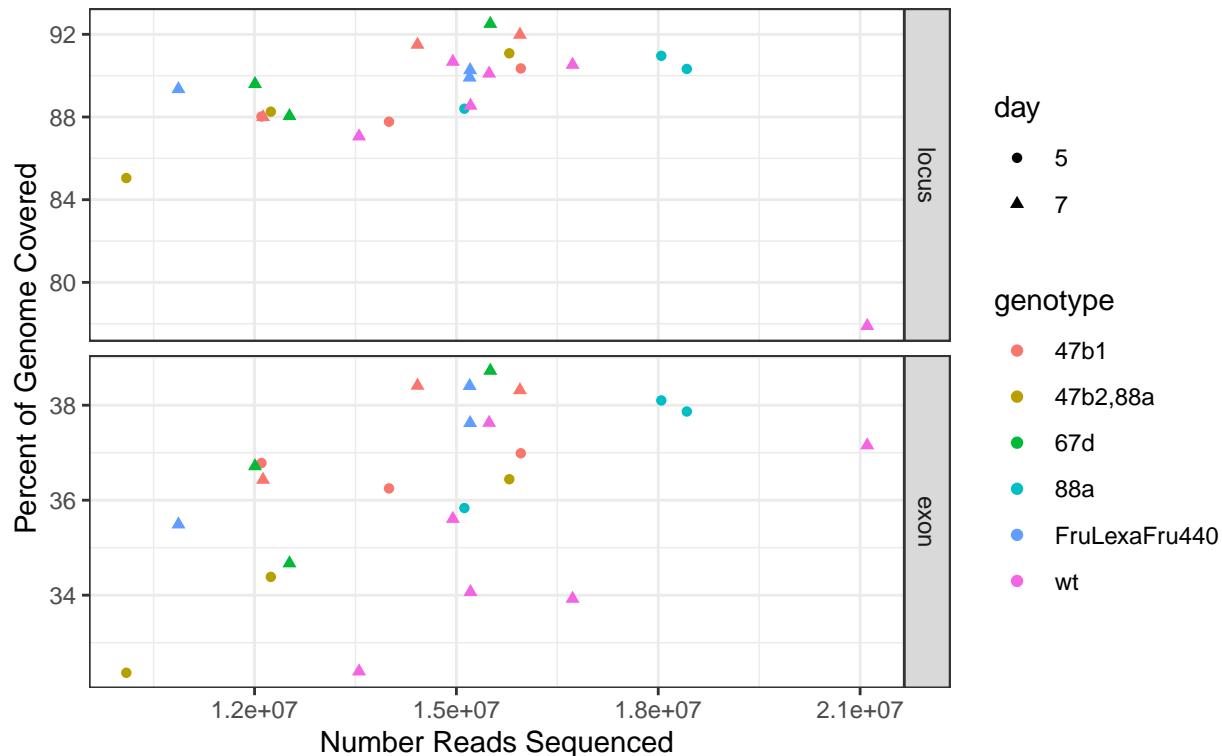


Figure 8. Breadth of Coverage of Raw Mapsplice Alignment Compared to Read Count

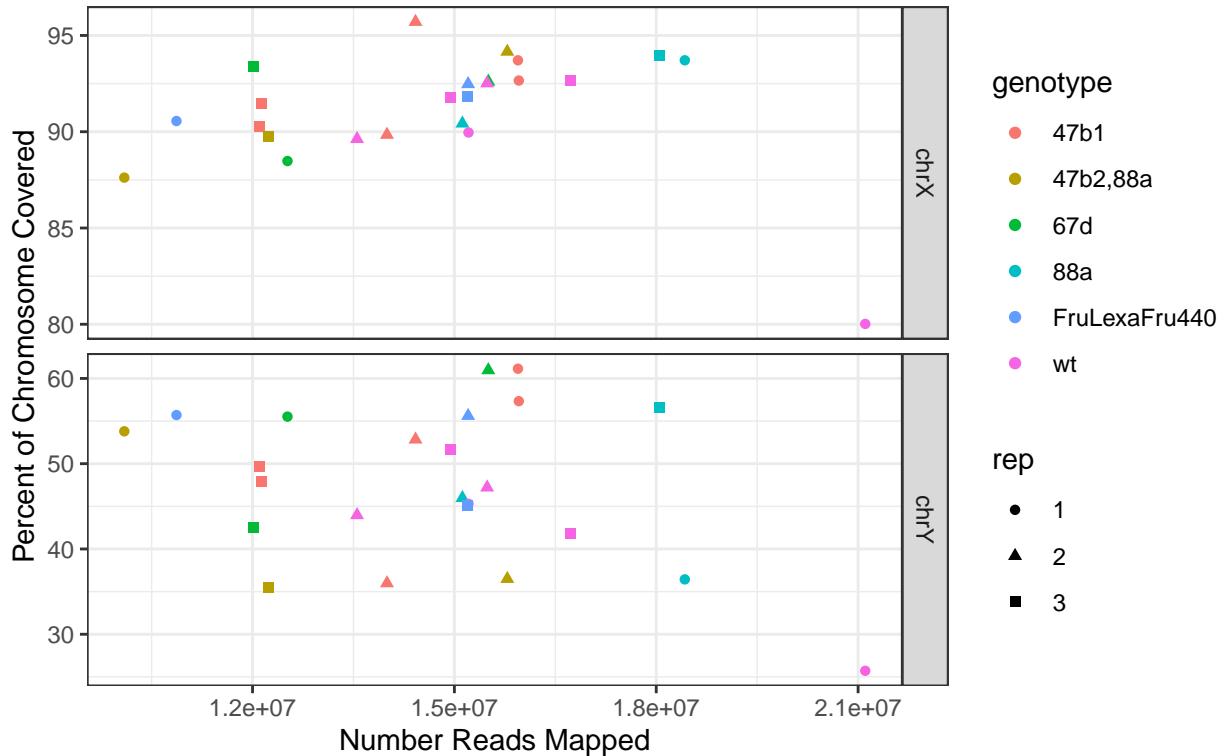


```
## pdf
## 2
```

There appears to be a slight dependence of breadth upon sequencing depth (ie, the number of reads sequenced), meaning that sequencing depth of these samples is not so great that the breadth covered is saturated. The unusually deep sequencing of the NCBI reads indicates the asymptotic behavior of this measure. When transcribed locii are considered, the breadth of the group-housed wildtype replicate 1 is unusually low given the sequencing depth.

We can also compare the breadth of coverage on the X and Y chromosomes to confirm that the flies sampled are all the same sex. The only outlier is the group-housed wildtype replicate 1, which is also anomalous genome-wide. The two NCBI samples agree well on the X chromosome, which is not unexpected, and the female-annotated sample has lower coverage on the Y, as expected. However, the difference between the NCBI controls is well within the variation of the new sequences, so this doesn't work as a decisive diagnostic.

Figure 9. Fraction of Sex Chromosome Covered in Raw Mapsplice Alignments Compared to Read Count



```
## pdf
## 2
```

2.5.2 Filtered Multimap

From the raw MapSplice output, three filtered alignments were produced. The first, mapspliceMulti, has had duplicates marked and removed, and has been filtered to require proper pairing and a minimum mapping quality (SAM flags “-q 20 -F 0x0200 -F 0x04 -f 0x0002”; markdup flags “-rS”). Thus, mapspliceMulti is a filtered alignment that retains all locii for multimapped reads.

The filtration process removed a total -1.71×10^9 mapped reads, an overall mapped retention rate of 58.5047545 %.

Table 11. Sample Read Retention Rate
percent of reads retained when filtering raw alignment

	maximum	mean	median	minimum
mapped retention	81.4%	78.5%	79.0%	64.5%

Table 12. Sample Coverage Retention Rate
percent of coverage retained when filtering raw alignment

	maximum	mean	median	minimum
spanned breadth retention	99.7%	99.6%	99.6%	99.3%
split breadth retention	97.2%	97.1%	97.1%	96.7%

Although filtration removed some (45.6009985 %) of the multimapping reads, 9.98M remain ambiguously mapped. A given read mapped, on average, to 1.10704378550129 locations. These will be kept as-is in mapsspliceMulti, but will be further filtered in other alignments.

Table 13. Mapping Uniqueness & Multiplicity
effect of filtering on multimapping reads

rep	percent of reads uniquely mapping		average per-read mapping multiplicity	
	raw	multi	raw	multi
47b1 - group - 5				
1	96.6%	96.7%	1.17	1.11
2	96.5%	96.6%	1.18	1.12
3	95.9%	96.1%	1.21	1.14
47b2,88a - group - 5				
1	96.3%	96.5%	1.21	1.13
2	96.3%	96.5%	1.20	1.13
3	96.1%	96.3%	1.21	1.14
88a - group - 5				
1	96.9%	97.1%	1.13	1.09
2	96.9%	97.3%	1.13	1.09
3	97.0%	97.3%	1.13	1.09
47b1 - group - 7				
1	96.0%	96.0%	1.19	1.13
2	95.5%	95.7%	1.20	1.14
3	95.6%	95.6%	1.19	1.12
67d - group - 7				
1	96.7%	97.0%	1.15	1.10
2	95.8%	96.0%	1.23	1.15
3	96.0%	96.3%	1.21	1.14
wt - group - 7				
1	97.6%	97.8%	1.09	1.06
2	95.8%	95.9%	1.11	1.07
3	97.4%	97.8%	1.10	1.06
wt - isolated - 7				
1	97.7%	98.0%	1.08	1.06
2	97.7%	98.1%	1.08	1.05
3	97.7%	98.0%	1.08	1.06
FruLexaFru440 - group - 7				
1	95.2%	95.0%	1.21	1.17
2	96.7%	96.8%	1.15	1.11
3	95.7%	95.6%	1.15	1.10

2.5.3 Downsampled Multimapped

mapsspliceRando is a downsampled alignment constructed by selecting at random a single location for each multimapped read, then merging the unambiguously located reads with mapsspliceUniq.

Table 14. Downsampling Retention Rate
percent of alignment retained when multimappers are downsampled

	maximum	mean	median	minimum
mapped retention	99.2%	98.2%	98.1%	97.0%
spanned breadth retention	99.4%	99.1%	99.1%	98.0%
split breadth retention	90.2%	89.8%	89.9%	89.1%

2.5.4 Uniquely Mapped

maps spliceUniq is derived from maps spliceMulti by further filtering out the multimapped reads and keeping only those which map uniquely.

Table 15. Uniquely Mapped Retention Rate
percent of alignment retained when multimappers are excluded

	maximum	mean	median	minimum
mapped retention	98.1%	96.7%	96.6%	95.0%
spanned breadth retention	99.1%	98.8%	98.8%	97.6%
split breadth retention	87.7%	87.3%	87.4%	86.5%

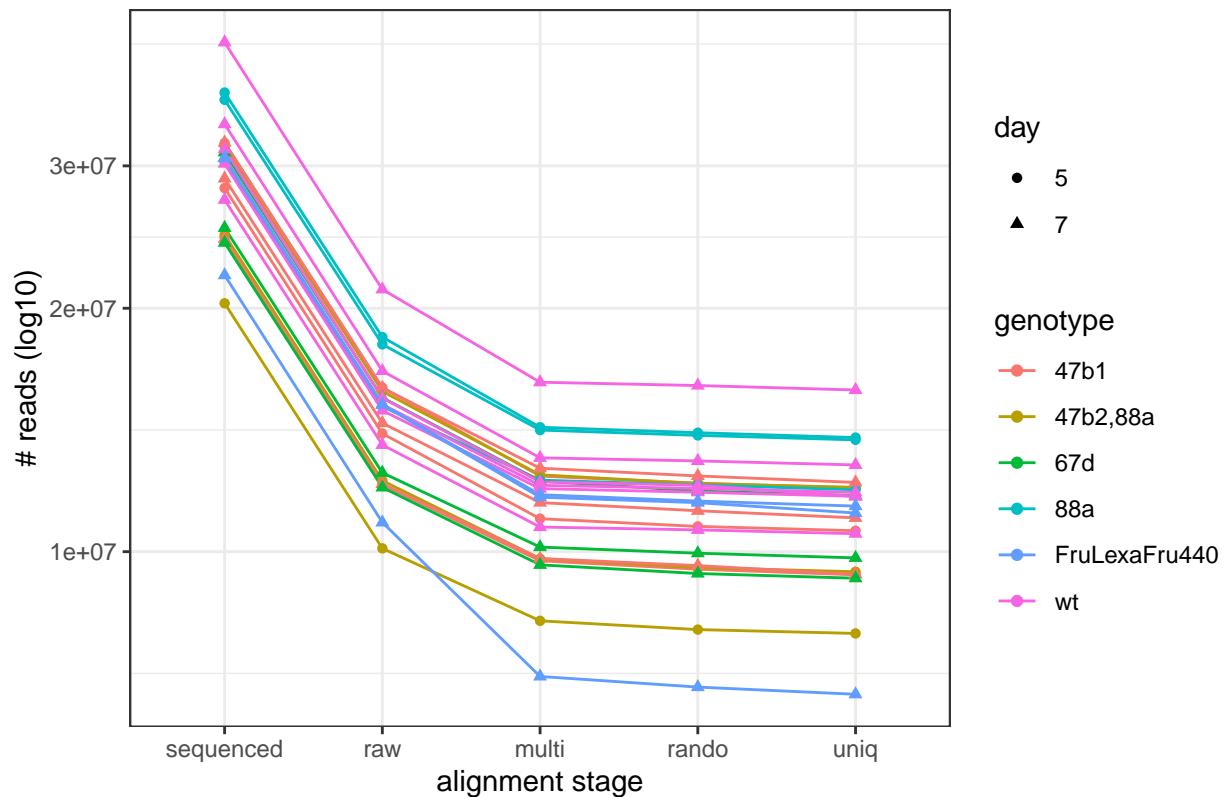
2.5.5 Spliced-Only

For each of Multi, Rando, and Uniq, a _SpliceOnly alignment was constructed by first filtering to only include spliced reads (awk ‘(\$6 ~ /N/)’), then reducing the reads to 1 bp on either side of the splice site. These are used with the fru_junct and fru_intron annotations .

2.5.6 Alignment Process Overview

Here are the number of reads per sample, from the intial sequencing to the most heavily filtered alignment:

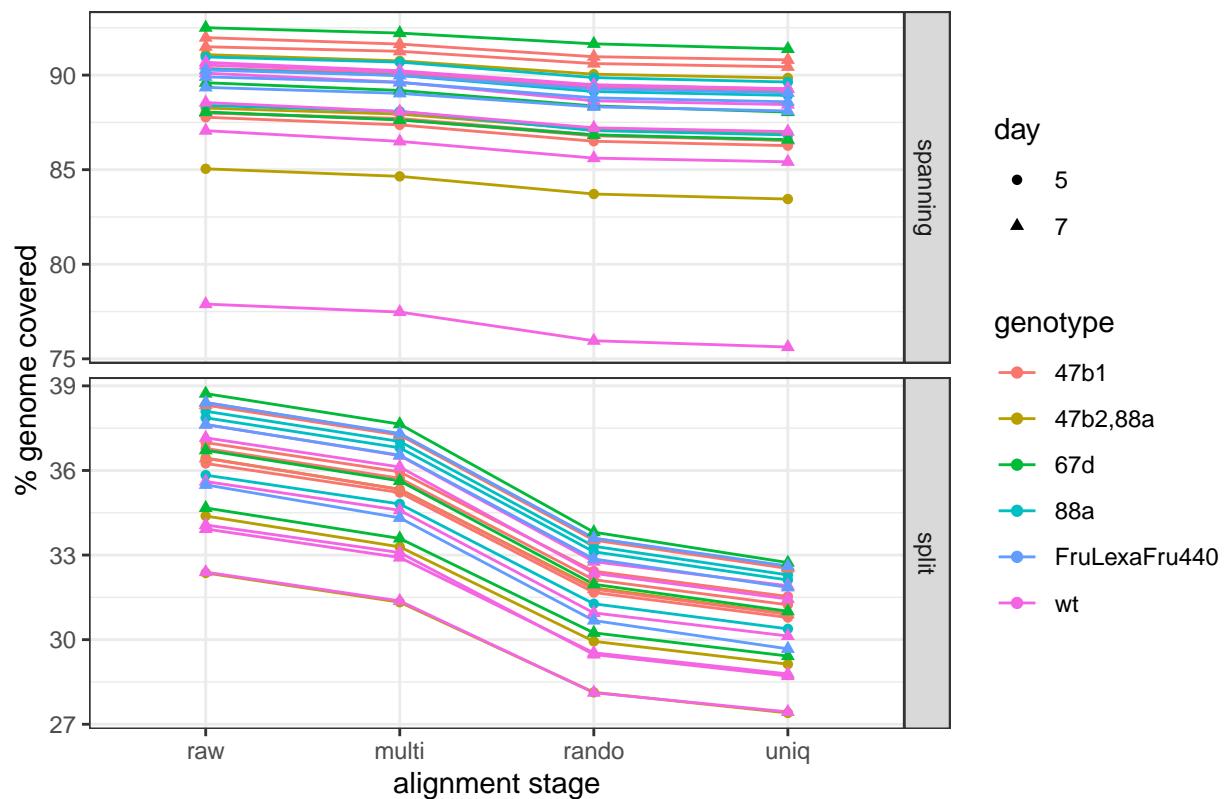
Figure 10. Read–count Dropout During Alignment Process



```
## pdf
## 2
```

The coverage dropout during the alignment filtration can be similarly tracked:

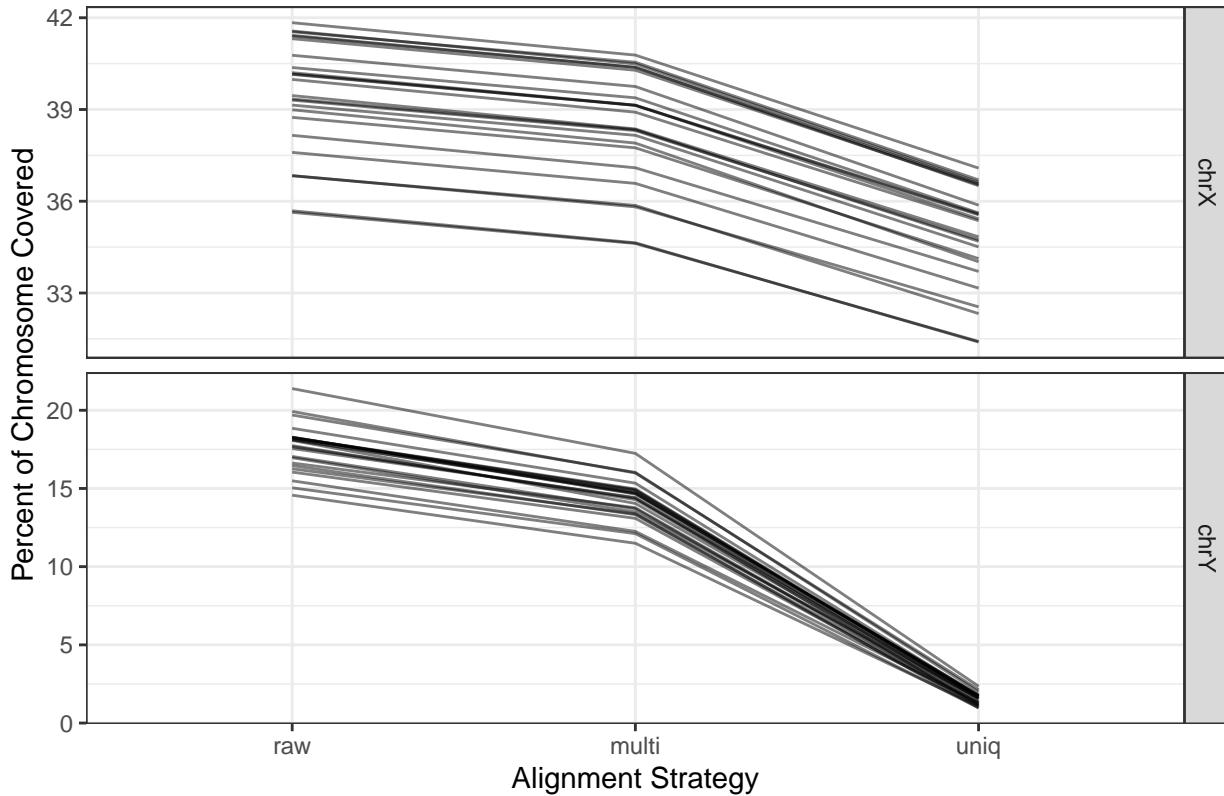
Figure 11. Coverage Loss During Alignment Process



```
## pdf
## 2
```

When restricted to the sex chromosomes, the NCBI controls were almost indistinguishable, with the difference between them much smaller than the difference between experimental samples. So, accounting for multimapping reads also doesn't make this a useful diagnostic:

Figure 12. Fraction of Sex Chromosome Covered, by alignment strategy



```
## pdf
## 2
```

2.6 Assigning Reads to Annotated Features

Mapped reads were assigned and counted using the featureCounts function from the SubRead package. (Liao, Smyth, and Shi 2014). In particular, the reads were assigned to exons in the dm6_genes GTF annotation, and these were counted towards the genes containing the exons. The two ends of paired reads were counted as separate fragments. To be counted, both ends of the paired reads must map, and map to the same chromosome. Any multimapped reads are counted at all of their mapped locations. (Command line options: “-t exon -g gene_id -M -J -p -B -C”).

By default, a read overlapping multiple genes is considered ambiguous and not counted. This makes sense when the feature being counted is a gene, but becomes problematic when counting by exon, since:

- reads which span splice junctions necessarily overlap multiple features, and thus aren’t counted
- exons which are small compared to read size will have few or no reads unspliced
- some exons are completely contained within other exons, and are precluded from having reads assigned.

Thus, some counts (filenames containing “MpBCO”) have reads assigned to all overlapping features, instead of none (filenames containing “MpBC”). featureCounts offers a third option, to assign 1/nth of a read to each of n features it overlaps; however, DESeq2 requires integer counts so this is not appropriate here.

Table 16. Percentage of Reads Assignable to Features in dm6_genes fraction of the reads which can be unambiguously counted under different alignment strategies

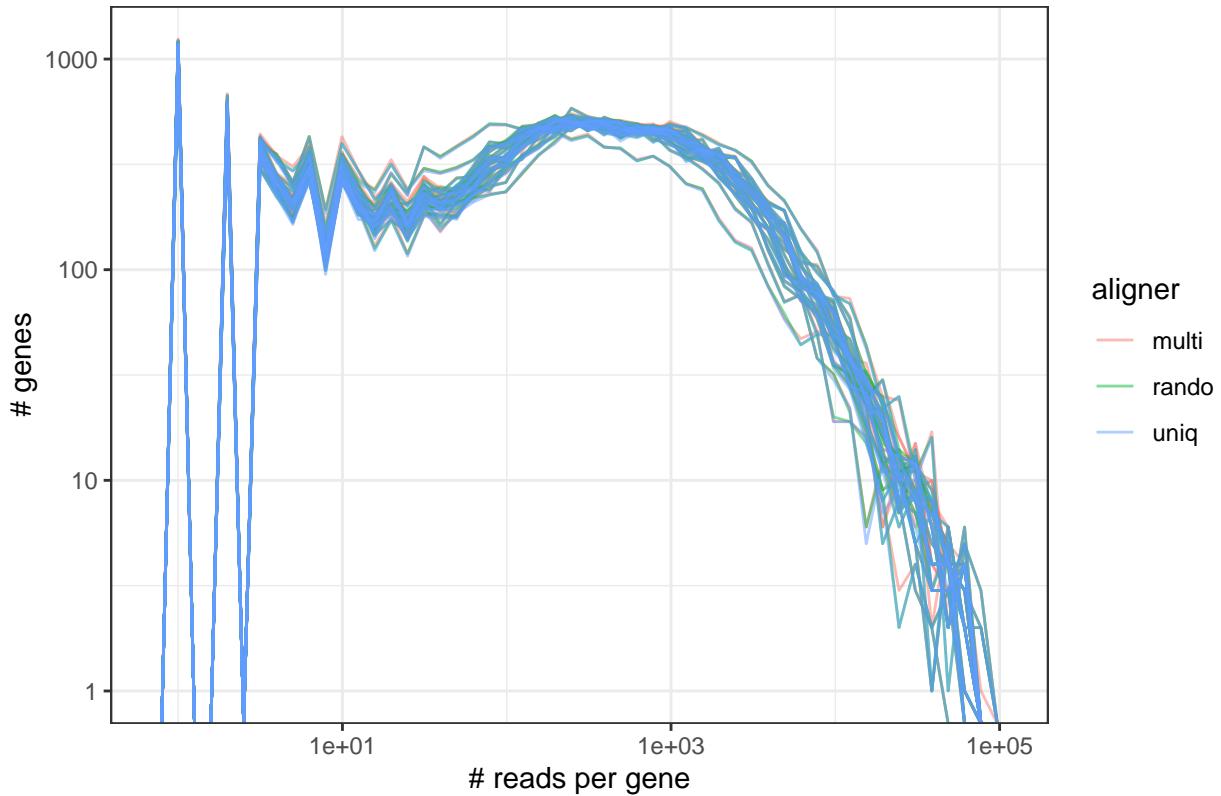
rep	mapping strategy		
	multi	rando	uniq
47b1 - group - 5			
1	90.6%	91.5%	92.1%
2	89.9%	91.1%	91.6%
3	88.7%	89.9%	90.7%
47b2,88a - group - 5			
1	89.8%	91.3%	91.7%
2	89.8%	91.4%	91.9%
3	89.7%	91.2%	91.7%
88a - group - 5			
1	90.4%	91.0%	91.7%
2	90.8%	91.2%	91.8%
3	90.6%	91.2%	91.8%
47b1 - group - 7			
1	89.7%	90.7%	91.5%
2	88.9%	90.0%	91.0%
3	89.6%	90.2%	91.5%
67d - group - 7			
1	90.8%	91.6%	92.1%
2	88.5%	90.3%	90.9%
3	89.0%	90.4%	91.0%
FruLexaFru440 - group - 7			
1	84.3%	85.9%	87.1%
2	89.8%	90.5%	91.2%
3	89.4%	89.5%	91.0%
wt - group - 7			
1	92.0%	91.9%	92.3%
2	90.3%	89.9%	91.5%
3	91.5%	91.6%	92.0%
wt - isolated - 7			
1	92.0%	91.9%	92.4%
2	92.3%	92.2%	92.6%
3	92.1%	92.1%	92.5%

Table 17. Averaged Percentage of Reads Not Assignable to Features in dm6_genes average fraction of mapped reads which were unassigned

	mapping strategy		
	multi	rando	uniq
Ambiguous	3.5%	3.5%	3.5%
No Overlap	25.3%	5.7%	5.1%

The values for “multi” are inflated because each appearance of a multi-mapped read is counted, whereas the denominator is the actual read count (FIX THIS)

Figure 13. Per–Gene Read Count Histogram (by aligner and sample)



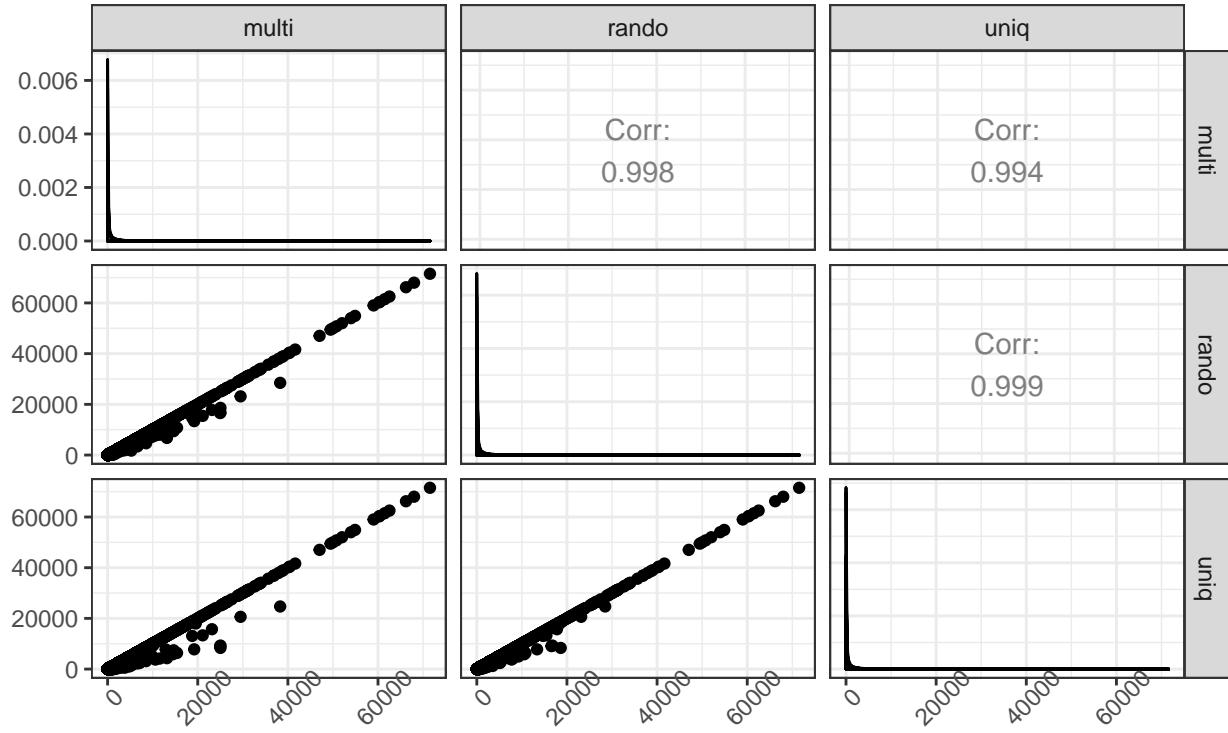
```
## pdf
## 2
```

One average, a gene had 589.479716187028 reads assigned to it, but most genes had relatively fewer, with more than a quarter having no reads assigned at all, almost half having fewer than 10 reads, and almost two thirds having fewer than 100.

Table 18. Averaged Percentage of Genes by Threshold Read Counts
average fraction of genes with low number of reads

aligner	read count threshold		
	< 1	< 10	< 100
multi	27.9%	45.9%	59.8%
rando	28.7%	46.2%	60.0%
uniq	29.3%	46.7%	60.3%

Figure 14. Correlations between Read Count Assigned to Gene Across Alignment Strategy (downsampled to 10%)

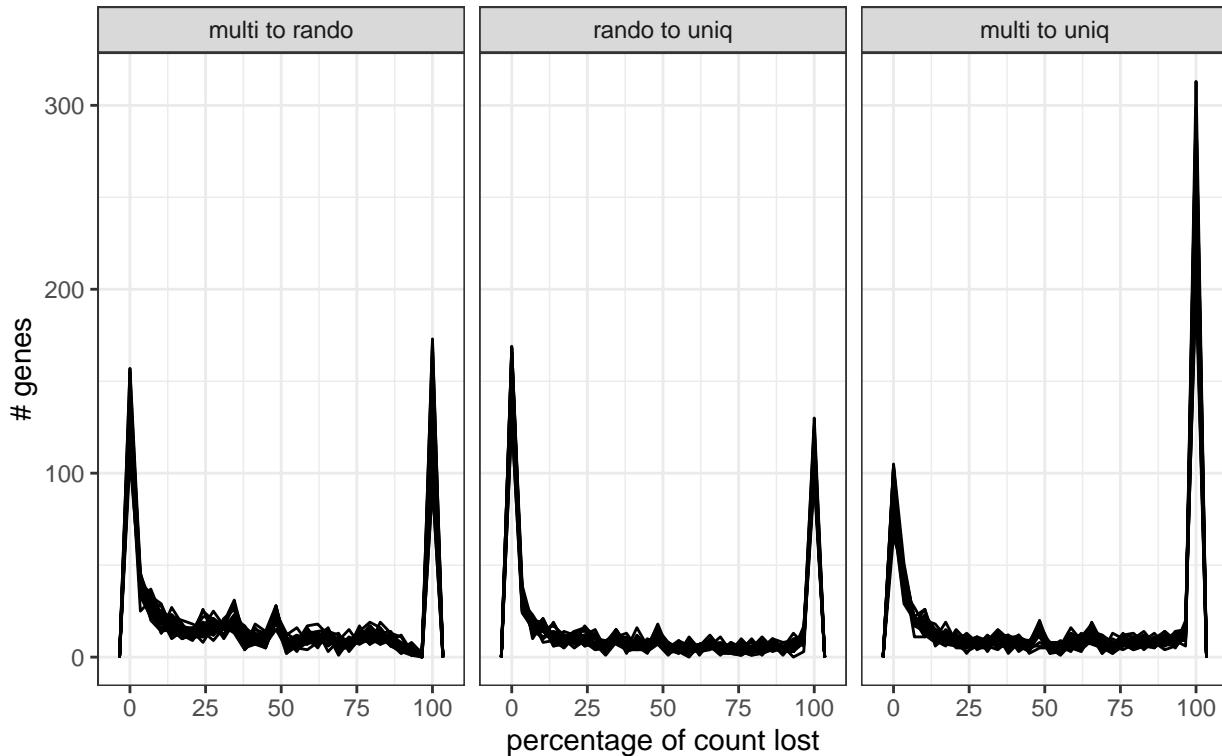


```
## pdf
## 2
```

The three mapping strategies generally agreed well; for 93.3504859919954 % of genes, the same number of reads were assigned by all three strategies in all samples. (Restricted to genes with at least one nonzero count, the proportion was 92.7421367948078 %)

By construction, the read count assigned to a gene is supposed to decrease across strategy: multi \geq rando \geq uniq. It's not clear why but for a very small number of cases (23; 0.00547932151705737 %), rando $>$ multi.

Figure 15. Percent Loss in Assigned Read Count Between Mapping Strategies (Discrepancies Only)



```
## pdf
## 2
```

2.6.1 Fruitless by exon

To study Fru on an exon-by-exon case, the existing GTF annotation was subsetted to isoforms of only this gene, and reformatted such that each exon was an individual feature to be counted. featureCounts was then run as usual on this new annotation. With many genes to study on a per-exon basis, the featureCounts -f flag might be more useful.

(Counts are so small compared to total that percentages aren't informative here)

Table 19. Number of Reads Assignable to Features in fru_exons
number of the reads which can be counted by alignment/assignment strategy

	all			none		
	multi	rando	uniq	multi	rando	uniq
47b1 - group - 5						
1	743	743	743	566	566	566
2	663	663	663	509	509	509
3	690	690	690	532	532	532
47b2,88a - group - 5						
1	785	785	785	614	614	614

2	1437	1437	1437	1118	1118	1118
3	1080	1080	1080	863	863	863
88a - group - 5						
1	1824	1824	1824	1431	1431	1431
2	1587	1587	1587	1296	1296	1296
3	1799	1799	1799	1449	1449	1449
47b1 - group - 7						
1	793	793	793	595	595	595
2	899	899	899	678	678	678
3	689	689	689	528	528	528
67d - group - 7						
1	655	654	654	496	495	495
2	971	971	971	759	759	759
3	750	750	750	568	568	568
FruLexaFru440 - group - 7						
1	978	978	978	832	832	832
2	728	728	728	580	580	580
3	739	739	739	554	554	554
wt - group - 7						
1	1126	1126	1126	818	818	818
2	935	935	935	696	696	696
3	845	845	845	638	638	638
wt - isolated - 7						
1	824	824	824	600	600	600
2	975	975	975	731	731	731
3	1095	1095	1095	831	831	831

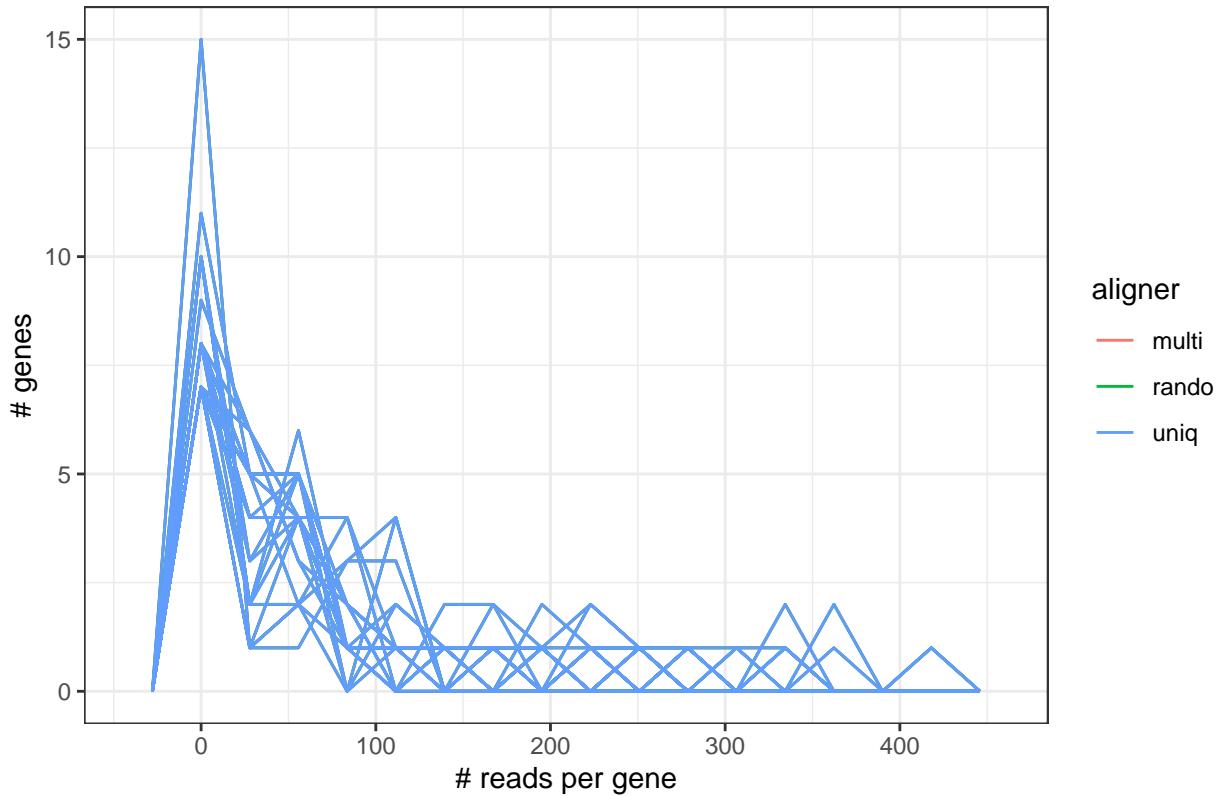
One average, a exon had 62.472222222222 reads assigned to it, but most exons had relatively fewer, with almost a quarter having no reads assigned at all, more than a third having fewer than 10 reads, and almost two thirds having fewer than 50. These figures are for the “All” assignment strategy, and are necessarily lower for the “None”.

Table 20. Averaged Percentage of Exons by Threshold Read Counts (Fruitless) average fraction of genes with low number of reads

aligner	read count threshold		
	< 1	< 10	< 50
multi	22.3%	34.5%	59.8%
rando	22.3%	34.5%	59.8%
uniq	22.3%	34.5%	59.8%

There is VERY little difference between alignment strategies when it comes to read count here. total overplot in count histogram; no point in showing other comparisons...

Figure 16. Per-Exon Read Count Histogram (by aligner and sample) for Fru



```
## pdf
## 2
```

2.6.2 Fruitless by splice junction

The fru_junct annotation is only counted under the “All” assignment strategy, since the reads being counted are spliced and thus necessarily overlap multiple exons. As well, the “SpliceOnly” version of each alignment will be used (ie, only spliced reads and only the 1bp subintervals which correspond to splice junctions)

Table 21. Number of Reads Assignable to Features in fru_junct
number of the reads which can be counted by alignment/assignment strategy

	all		
	multi	rando	uniq
47b1 - group - 5			
1	96	96	96
2	74	74	74
3	74	74	74
47b2,88a - group - 5			
1	107	107	107
2	192	192	192
3	152	152	152

88a - group - 5			
1	241	241	241
2	198	198	198
3	226	226	226
47b1 - group - 7			
1	109	109	109
2	103	103	103
3	79	79	79
67d - group - 7			
1	84	84	84
2	94	94	94
3	66	66	66
FruLexaFru440 - group - 7			
1	25	25	25
2	76	76	76
3	70	70	70
wt - group - 7			
1	137	137	137
2	136	136	136
3	103	103	103
wt - isolated - 7			
1	104	104	104
2	132	132	132
3	127	127	127

2.6.3 Fruitless by intron

The fru_intron annotation is counted under the “All” assignment strategy, using the “SpliceOnly” alignments. Because the same reads are being counted against the same intervals, the number of reads countable are identical to those in fru_junct

Table 22. Number of Reads Assignable to Features in fru_intron
number of the reads which can be counted by alignment/assignment strategy

	all		
	multi	rando	uniq
47b1 - group - 5			
1	96	96	96
2	74	74	74
3	74	74	74
47b2,88a - group - 5			
1	107	107	107
2	192	192	192
3	152	152	152
88a - group - 5			
1	241	241	241

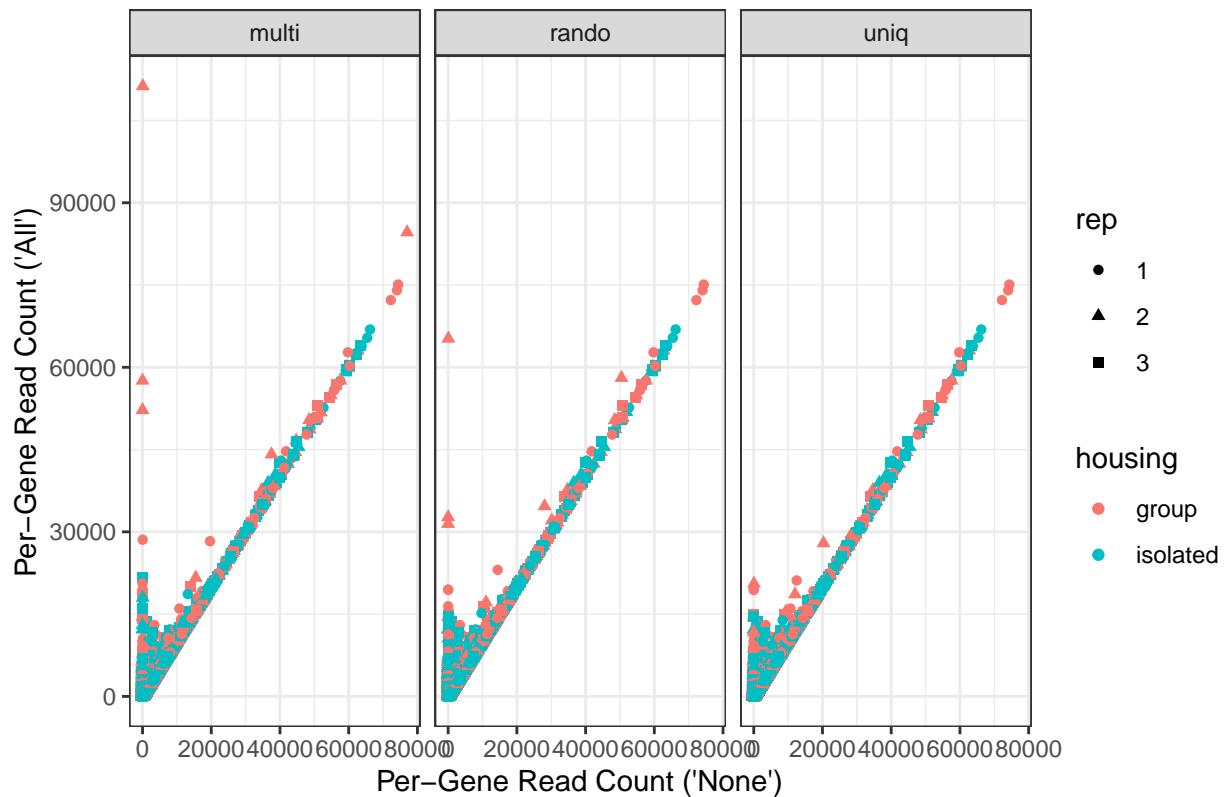
2	198	198	198
3	226	226	226
<hr/>			
47b1 - group - 7			
1	109	109	109
2	103	103	103
3	79	79	79
<hr/>			
67d - group - 7			
1	84	84	84
2	94	94	94
3	66	66	66
<hr/>			
FruLexaFru440 - group - 7			
1	25	25	25
2	76	76	76
3	70	70	70
<hr/>			
wt - group - 7			
1	137	137	137
2	136	136	136
3	103	103	103
<hr/>			
wt - isolated - 7			
1	104	104	104
2	132	132	132
3	127	127	127

2.6.4 Ambiguous Assignment Strategy Comparison

The whole gene annotation and the Fruitless exons are currently having readcounts assigned with slightly different strategies. When all genes are considered, ambiguously assigned reads (those which overlap multiple features) are simply discarded; we will call this the “None” strategy. When the exons of Fru are considered, ambiguously assigned reads count towards the tally of every exon they overlap we’ll call this “all”).

There is a big difference between these strategies at the exon level; how well do they agree on the entire dm6_genes annotation?

Fig 17. Comparison of Assignment Strategies for Ambiguous Reads



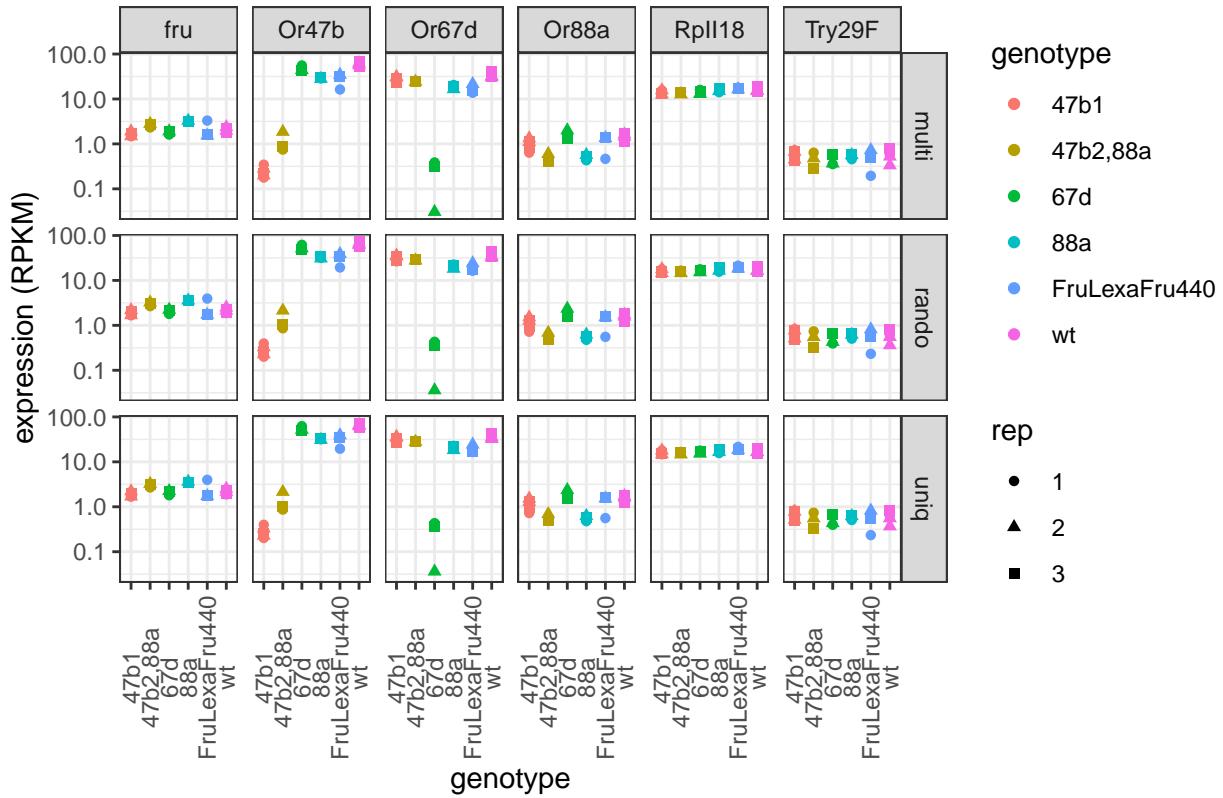
```
## pdf
## 2

##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## 0.00      0.00      0.00   61.73      3.00 111175.00
```

2.7 Expression

Using the per-gene read counts, the per-alignment total mapped counts, and the gene lengths, the gene expression was calculated as reads per kilobase per million mapped (RPKM). In particular, these can confirm that the knock-outs are not being expressed. This appears to be the case in the 47b and 67d mutants. The Fru440FruLexa mutants do not show any obvious reduction in expression of *Fruitless* (not knockouts - is expression expected though?) For context, a positive control (RNA polymerase) and a negative control (trypsin) have also been included.

Fig 18. Expression Levels for Key Genes, by Genotype



```
## pdf
## 2
```

2.8 Differential Expression Analysis.

DESeq2 (Love, Huber, and Anders 2014) was used to detect changes in expression from read-count data, following the official vignette as a guide (<http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html> ; see also <http://master.bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html>).

DESeq2 builds a statistical model in which the read counts are normalized and then fit to explanatory variables (“factors”). Each value a factor may take on is called a “level”. For example, genotype is a factor, whereas the 47b mutation is a level of the genotype factor. The model fit to the counts may contain one or more factors.

Single-factor models (wildTypeHousing, grpWtVs47b, grpWtVs67d, grpWtVsFru, grpWtVsMut) were built by specifying the axis of comparison (eg, housing) and subsetting samples to the relevant contrast (eg, wt group reps 1,2,3 and wt isolated reps 1,2,3).

Current results mostly come from a two-factor model in which both housing and genotype are considered simultaneously (hausWtVsMut).

Table 23. Differential Expression Contrasts
with model and reference levels

	fit model	reference levels
wildTypeHousing	~housing	housing: group

grpWtVs47b	~genotype	genotype: wt
grpWtVs67d	~genotype	genotype: wt
grpWtVsFru	~genotype	genotype: wt
grpWtVsMut	~genotype	genotype: wt
hausWtVsMut	~genotype + housing	genotype: wt, housing: group
hausWtVsMut_noFru	~genotype + housing	genotype: wt, housing: group
47b_on_88a	~genotype	genotype: 88a
2_days_difference	~day	day: 5

For each factor and level, DESeq2 returns two key pieces of information: an effect size and an adjusted p-value.

The effect size is reported as the base-2 logarithm of fold-change in expression between the reference level and some alternate level. Thus, if the 47b contrast for some gene G has a log2FoldChange of 1, it means that the 47b mutants express G at $2^1 = 2$ times as much as the wildtype flies. A log2FoldChange of -1 means that the 47b mutants express G at $2^{-1} = 0.5$ times as much as the wildtype flies. No change at all would be a foldchange of 1, and a log2 fold change of 0.

The p-value gives the odds that an effect size as large would be observed if there were no change in expression, just random noise. Since a p-value is estimated for each gene in the annotation, a correction for multiple comparisons is applied.

DESeq2 reports the normalized mean counts for each level; an expression level was derived from it by scaling by feature length. (More on interpretation & use of the “baseMean”: <https://support.bioconductor.org/p/75244>; <https://support.bioconductor.org/p/63567/>; <https://www.biostars.org/p/219093/>)

Counts filtered to remove genes with less than 10 reads combined across all samples. Effect-size shrinkage is currently done using apeglm; other shrinkage estimators have not yet been explored.

2.9 Gene Ontology Enrichment

Gene Ontology Enrichment was studied using topGO. <https://bioconductor.org/packages/release/bioc/vignettes/topGO/inst/doc/topGO.pdf>

For each set of DESeq data studied, the genes and their expression differences were subsetted by factor and level. Two tests were used: Fisher’s Exact, which uses counts from a discrete subset of genes (here, those with adjusted p < 0.01), and Kolmogorov-Smirnov, which uses the p-values as a quantitative score. The “classic” algorithm was used, and the top 50 nodes were collected and saved for each GO type: Molecular Function, Biological Component, Cellular Process.

topGO appears to still be plagued by an intermittent error, “There are no adj nodes for node: GO:xxxxxxxx Error in switch”, for which there is not yet a clear solution or explanation. (eg, <https://support.bioconductor.org/p/116048/>; <https://support.bioconductor.org/p/103640/>; <https://www.biostars.org/p/311104/>)

From experience, I can prevent it by masking ~30 genes. Some of these are significantly differentially expressed, however!

	flybase_gene_id	external_gene_name
1	FBgn0261268	Cul3
33	FBgn0032470	CG5142
45	FBgn0031450	Hrs
84	FBgn0051999	CG31999
88	FBgn0011828	Pxn
108	FBgn0014388	sty
142	FBgn0038358	CG4525
152	FBgn0016075	vkg

164	FBgn0050046	CG30046
169	FBgn0028573	prc
180	FBgn0033710	CG17739
184	FBgn0261800	LanB1
204	FBgn0005695	gcl
229	FBgn0020269	mspo
233	FBgn0039257	tnc
242	FBgn0262733	Src64B
296	FBgn0263930	dally
343	FBgn0040206	krz
363	FBgn0026562	SPARC
378	FBgn0041604	dlp
411	FBgn0004907	14-3-3zeta
442	FBgn0032252	loh
448	FBgn0035049	Mmp1
477	FBgn0050203	CG30203
482	FBgn0026721	fat-spondin
487	FBgn0003969	vap
505	FBgn0004390	RasGAP1
531	FBgn0031850	Tsp

Additionally, BioMart does not appear to have descriptions listed for some GO IDs; these currently need to be looked up on a case-by-case basis at <http://geneontology.org/>

Multiple comparison adjustment isn't done (see topGO vignette section 6.2)

Currently applied to the simultaneous model only.

3 Results

Earlier results were based upon the 1-factor models; these results are largely hidden in the */supp/ folders

3.1 Wildtype: Group-housed vs. Isolated

In the first contrast, wildtype flies with group-housed and isolated life histories are compared (experimental design: ~ housing). Group-housing was used as a reference level; fold changes are reported relative to it.

After filtering to remove genes with too few reads for analysis, about 11.9k of 17.7k annotated genes (67.1925771 %) remain available for testing:

3.1.1 preshrunk comparison across alignment strategies

The differential expression data were examined before shrinkage. The most discrepancy appeared between the mapspliceUniq alignment and the two which included multimappers, and in genes with small effect sizes.

3.1.2 effect size: preshrunk vs shrunk

The shrinkage step attempts to correct for the large apparent effect sizes in genes with small read counts. As expected, the shrinkage narrows the distribution around zero.

3.1.3 shrunk comparison across alignment strategies

The shrunk effect sizes agree well between alignment strategies; the “cloud” around unshrunk data at low effect size has disappeared.

??what's up with the outliers??

The alignment strategies also agree well when it comes to significance (shrinkage doesn't impact significance so this is the same before and after.)

3.1.4 differential expression overview

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than $1/2$), significant (ie, $\text{padj} < 0.01$) changes. There were 35 such genes, mostly shared across alignment strategy:

3.1.5 In relation to gene lists

3.1.6 Genes with top 10 most significant changes

Ordered in decreasing significance, the alignemnt strategies agree on the top 10 most significant changes:

3.1.7 Top 10 genes with biggest (significant) effect sizes

The alignment strategies agree on the top 10 largest fold changes (though not completely on their order):

3.1.8 Top 10 highest expressed genes with significant change

The “multi” and “rando” alignment strategies agree completely on the top 10 most expressed genes with significant changes. The “uniq” strategy differs in rank order and includes Gs2 and Msp300 instead of Calr and bun:

3.1.9 rank-correllation between alignment strategies

3.2 Group Housed: Wildtype vs Mutants

3.2.1 wt vs OR47b

After filtering to remove genes with too few reads for analysis, about $12.2k$ of $17.7k$ annotated genes (68.5937529 %) remain available for testing:

3.2.1.1 preshrunk comparison across alignment strategies

3.2.1.2 differential expression overview

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

Some of the effect sizes and p values are outrageous!!

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 524 such genes, mostly shared across alignment strategy: (see supplementary tables folder, *results/tables/supp/grpWtVs47b_chonky.html*)

3.2.1.3 Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes: rando and uniq alignment strategies agree very well; in multi, the gene “Unc-115a” has moved from off the chart to the #1 spot, bumping off “Ugt86Dd”.

3.2.1.4 Top 10 genes with biggest (significant) effect sizes

The alignment strategies agree well for the top 4, and disagree on order and content lower:

3.2.1.5 Top 10 highest expressed genes with significant change

The three alignment strategies agree well on the top 10 highest expressed genes with significant change:

3.2.2 wt vs 67d

After filtering to remove genes with too few reads for analysis, about $12.1k$ of $17.7k$ annotated genes (68.1711463 %) remain available for testing:

3.2.2.1 preshrunk comparison across alignment strategies

3.2.2.2 differential expression overview

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 553 such genes, mostly shared across alignment strategy: (see tables folder, *results/tables/supp/grpWtVs67d_chonky.html*)

3.2.2.3 Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 4 most significant changes, but disagree on the order & content after that.

3.2.2.4 Top 10 genes with biggest (significant) effect sizes

The alignment strategies agree relatively well on the genes with the top 10 largest (significant) fold changes (though not on their order):

3.2.2.5 Top 10 highest expressed genes with significant change

The alignment strategies agree well on the top 10 highest expressed genes with significant changes (though not on their order):

3.2.3 wt vs FruLexxFru440

After filtering to remove genes with too few reads for analysis, about $12.2k$ of $17.7k$ annotated genes (68.914934 %) remain available for testing:

3.2.3.1 preshrunk comparison across alignment strategies

3.2.3.2 differential expression overview

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than $1/2$), significant (ie, $\text{padj} < 0.01$) changes. There were 378 such genes, mostly shared across alignment strategy: (see tables folder, *results/tables/supp/grpWtVsFru_chonky.html*)

3.2.3.3 Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree very well on the top 10 most significant changes:

3.2.3.4 Top 10 genes with biggest (significant) effect sizes

The alignment strategies agree on the genes with the top 5 largest fold changes, less so for the next 5:

3.2.3.5 Top 10 highest expressed genes with significant change

The alignment strategies agree on the top 10 highest expressed genes with significant changes.

3.3 Simultaneously Modeling Housing & Genotype.

gives us eye-to-eye results for all treatments

These data are in the file “*results/tables/supp/hausWtVsMut.allAligners.DESeq2.MpBC.reformatted.tsv*”; columns are defined as follows:

```
external_gene_name :  
    human-readable gene symbol  
  
geneid :  
    flybase gene ID  
  
baseMean.(factor).(level) :  
    the normalized mean read count for all samples in (level) of contrast (factor).  
    Example: baseMean.genotype.wt is the normalized mean read count for wild types  
(of any housing status).
```

```

expression.(factor).(level) :
  expression level, calculated as baseMean.(factor).(level)/gene length in bp

baseMean.(factor).vs_(level).apeglm
log2FoldChange.(factor).vs_(level).apeglm
lfcSE.(factor).vs_(level).apeglm
pvalue.(factor).vs_(level).apeglm
padj.(factor).vs_(level).apeglm
expression.(factor).vs_(level).apeglm :
  equivalent to the "shrunk" data in the single-factor contrast
  for (level) compared to reference

-> do % of genes available for analysis

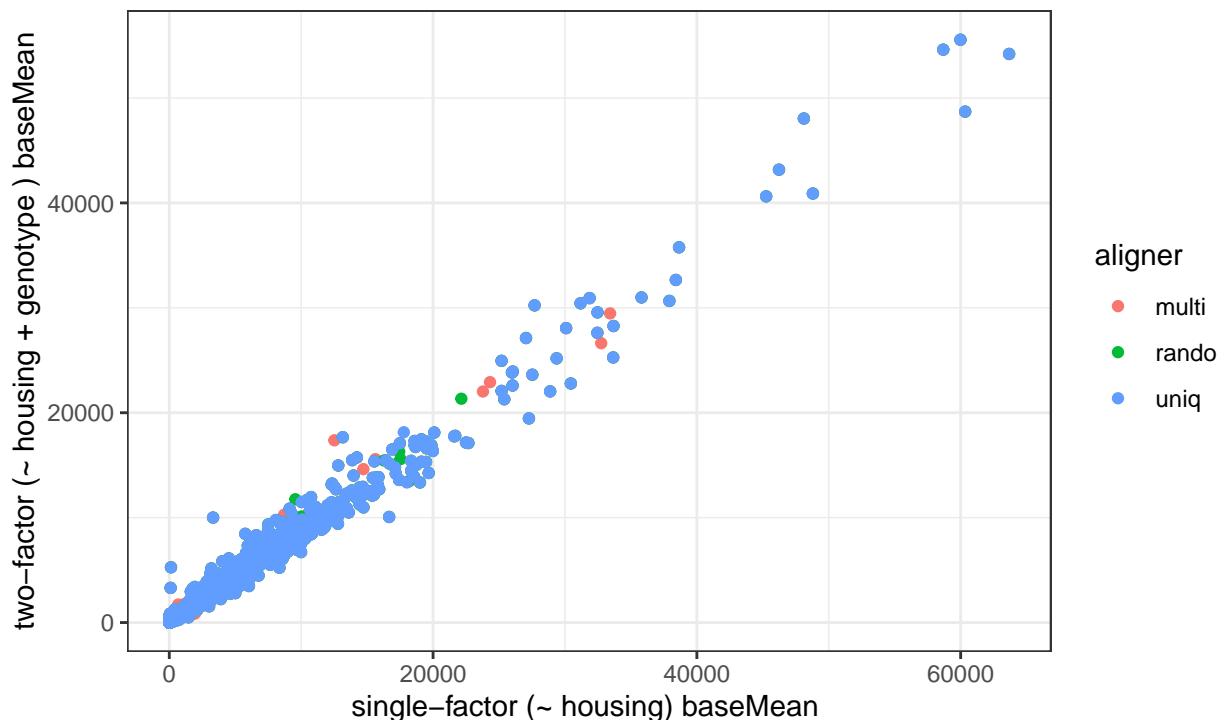
```

3.3.1 compare wildTypeHousing results to housing results from hausWtVsMut

To examine consistency with single-factor models, the two-factor model results are subsetted to the housing comparison.

Normalized mean read counts are different between the two models (which is not unexpected) but are correlated:

Figure 35. Scatterplot of per-gene normalized mean counts in full- vs. single-factor models for group vs isolated housing treatment



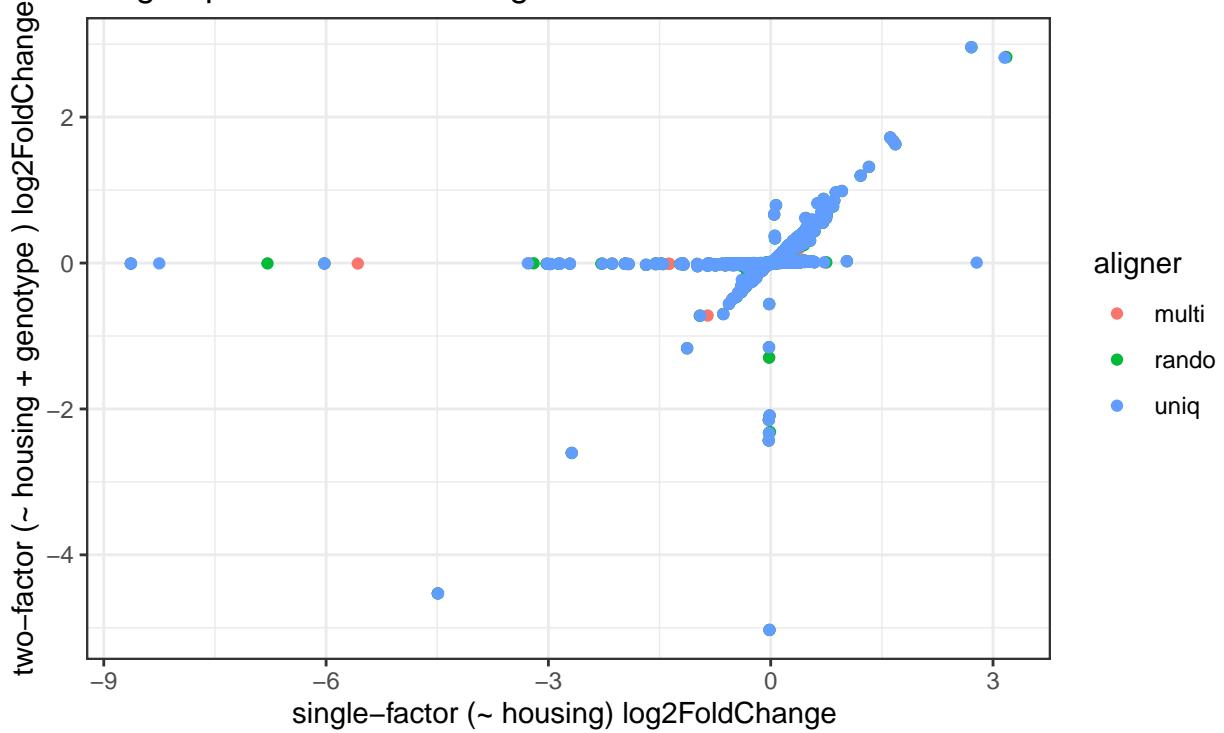
```

## pdf
## 2

```

Effect-size estimates from the two models either agree very well, or not at all. I have not had an opportunity to investigate this discrepancy.

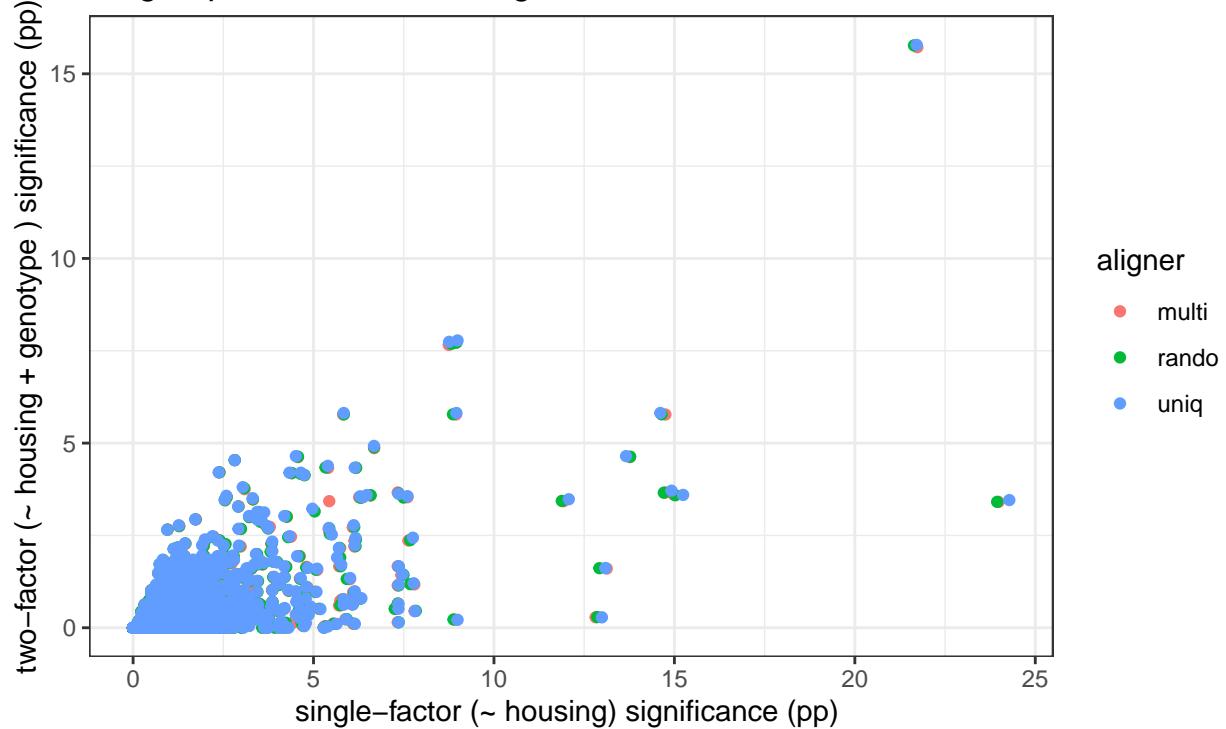
Figure 36. Scatterplot of per-gene expression difference in full- vs. single-factor models for group vs isolated housing treatment



```
## pdf  
## 2
```

Significance of differential expression estimates agree well enough, I guess:

Figure 37. Scatterplot of per-gene DE significance in full- vs. single-factor models for group vs isolated housing treatment

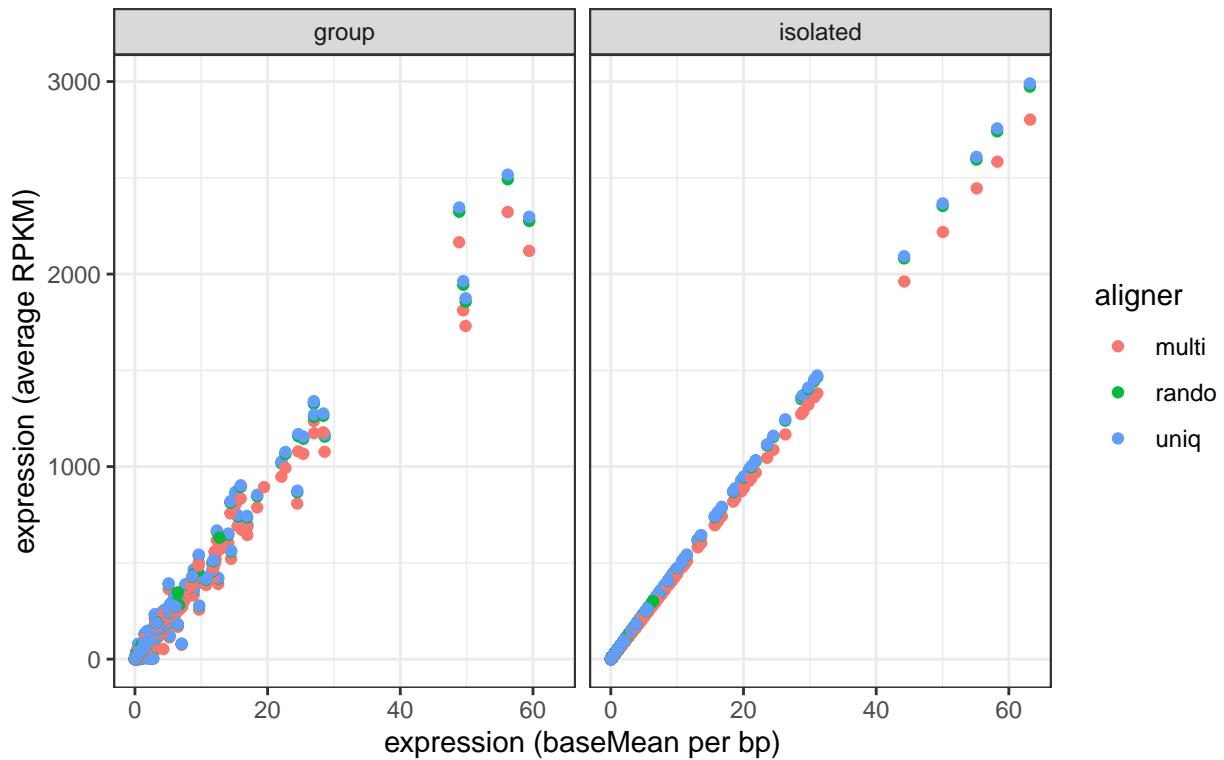


```
## pdf
## 2
```

3.3.2 compare RPKM to baseMean expression

Two estimates of gene expression have been made: one is based upon normalized mean read count from DESeq2, and the other is an RPKM value calculated from the raw counts. Let's see how they agree

Figure 38. Comparison of Expressionas Inferred by Direct Read Counting vs. DESeq2 Normalization

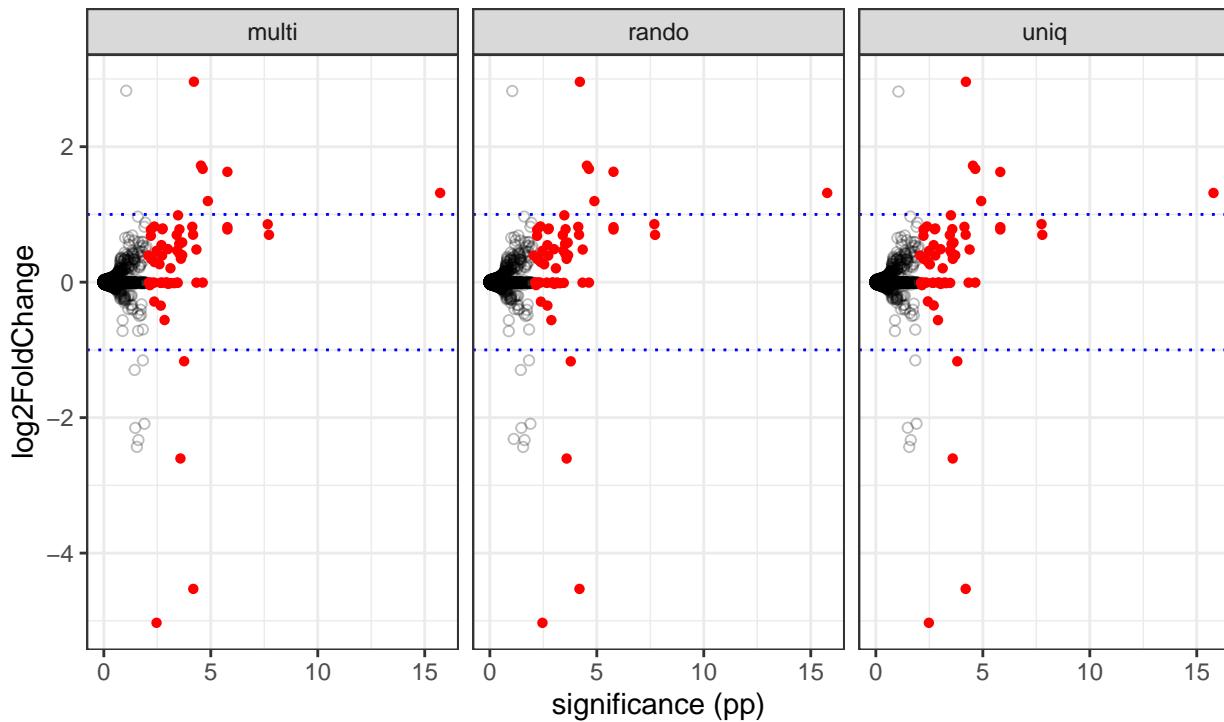


```
## pdf
## 2
```

3.3.3 Housing Contrast (Simultaneous model)

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of ± 1 (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

**Figure 39. Volcano Plot: Fold Change vs. Significance
(between isolated and group-housed wildtypes)
(Simultaneous Model)**



```
## pdf
## 2
```

I'm concerned about the "tail" of genes with very small effect sizes but high significance....

Of the 10832 genes with significance scores available, 64 have an adjusted $p < 0.01$ (0.5908419 %)

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 10 such genes, mostly shared across alignment strategy:

**Table 42. Genes with Large ($2 <$ fold change), Significant ($\text{padj} < 0.01$) Changes
between isolated and group-housed wildtypes, simultaneous model**

	multi	rando	uniq
MtnB	yes	yes	yes
Prat2	yes	yes	yes
CG10799	yes	yes	yes
CG11852	yes	yes	yes
Jhe	yes	yes	yes
CG15822	yes	yes	yes
CG31324	yes	yes	yes
CG7470	yes	yes	yes
amd	yes	yes	yes
CG11400	yes	yes	yes

3.3.3.1 Top Tens

Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes:

Table 42. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed between isolated and group-housed wildtypes; simultaneous model

multi					rando				
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p	
1	MtnB	0.39	1.317	1.90×10^{-16}	MtnB	0.39	1.317	1.70×10^{-16}	
2	magu	0.21	0.699	1.91×10^{-8}	magu	0.21	0.698	1.89×10^{-8}	
3	CG31288	1.36	0.855	2.19×10^{-8}	CG31288	1.36	0.857	2.05×10^{-8}	
4	CG31272	0.11	0.779	1.68×10^{-6}	CG31272	0.11	0.778	1.66×10^{-6}	
5	CG11852	0.07	1.629	1.68×10^{-6}	CG11852	0.07	1.629	1.66×10^{-6}	
6	CG10512	0.22	0.812	1.68×10^{-6}	CG10512	0.22	0.811	1.66×10^{-6}	
7	amd	1.01	1.197	1.37×10^{-5}	amd	1.01	1.196	1.30×10^{-5}	
8	CG15822	0.01	1.673	2.39×10^{-5}	CG15822	0.01	1.672	2.33×10^{-5}	
9	TotA	0.22	-0.006	2.39×10^{-5}	TotA	0.22	-0.006	2.33×10^{-5}	
10	CG31324	0.08	1.721	2.88×10^{-5}	CG31324	0.08	1.720	2.89×10^{-5}	

Top 10 genes with biggest (significant) effect sizes

Table 43. Top Ten Largest Magnitude Fold Changes which were Significant between isolated and group-housed wildtypes; simultaneous model

multi					rando				
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p	
1	CG10799	0.04	-5.029	3.43×10^{-3}	CG10799	0.04	-5.030	3.48×10^{-3}	
2	Prat2	0.02	-4.528	6.58×10^{-5}	Prat2	0.02	-4.529	6.46×10^{-5}	
3	Jhe	0.26	2.959	6.15×10^{-5}	Jhe	0.26	2.958	6.19×10^{-5}	
4	CG7470	0.03	-2.603	2.63×10^{-4}	CG7470	0.03	-2.604	2.58×10^{-4}	
5	CG31324	0.08	1.721	2.88×10^{-5}	CG31324	0.08	1.720	2.89×10^{-5}	
6	CG15822	0.01	1.673	2.39×10^{-5}	CG15822	0.01	1.672	2.33×10^{-5}	
7	CG11852	0.07	1.629	1.68×10^{-6}	CG11852	0.07	1.629	1.66×10^{-6}	
8	MtnB	0.39	1.317	1.90×10^{-16}	MtnB	0.39	1.317	1.70×10^{-16}	
9	amd	1.01	1.197	1.37×10^{-5}	amd	1.01	1.196	1.30×10^{-5}	
10	CG11400	0.07	-1.168	1.76×10^{-4}	CG11400	0.07	-1.169	1.66×10^{-4}	

Top 10 highest expressed genes with significant change

Table 44. Top Ten Highest Expressed Genes with Significant ($\text{padj} < 0.01$) Difference between isolated and group-housed wildtypes; simultaneous model

multi					rando				
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p	
1	Fer2LCH	4.18	0.205	7.69×10^{-4}	Fer2LCH	4.18	0.205	8.15×10^{-4}	
2	CG14687	2.72	0.346	2.48×10^{-4}	CG14687	2.72	0.344	2.58×10^{-4}	
3	CG32276	2.00	0.266	2.50×10^{-3}	CG32276	2.00	0.264	2.87×10^{-3}	

4	Or92a	1.57	0.420	2.92×10^{-4}	Or92a	1.57	0.420	2.98×10^{-4}
5	CG31288	1.36	0.855	2.19×10^{-8}	CG31288	1.36	0.857	2.05×10^{-8}
6	CG18135	1.10	0.821	4.35×10^{-3}	CG18135	1.10	0.822	4.31×10^{-3}
7	amd	1.01	1.197	1.37×10^{-5}	amd	1.01	1.196	1.30×10^{-5}
8	Nep4	0.77	0.781	2.92×10^{-4}	Nep4	0.77	0.782	2.89×10^{-4}
9	CG33056	0.75	0.987	3.38×10^{-4}	CG33056	0.75	0.987	3.30×10^{-4}
10	CG5745	0.66	0.298	4.35×10^{-3}	CG5745	0.66	0.298	4.23×10^{-3}

3.3.3.2 Gene Ontology Enrichment

Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

-> check consistency between GO terms between alignment strategies -> filter out very broad/very specific terms?

Correlation between significance values for the two tests are.... well, folks,

Figure 39. Scatterplot of GO Term Enrichment Significance for Two Statistical Tests (housing contrast; simultaneous model)

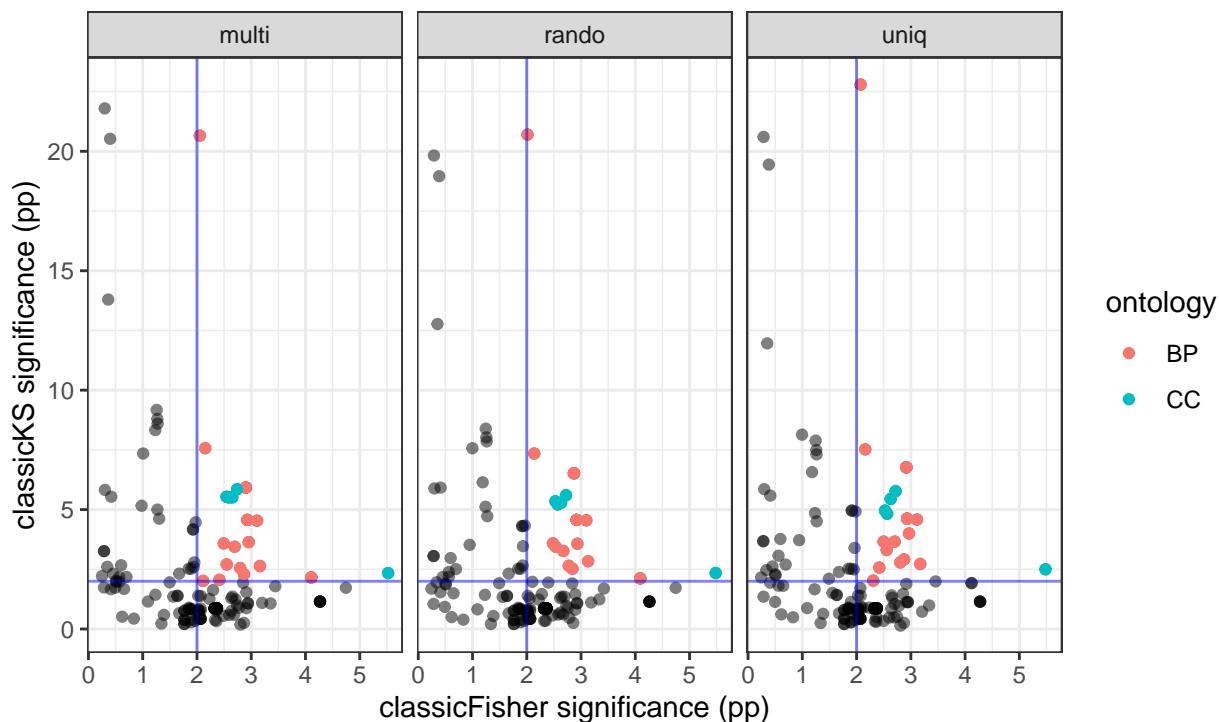


Table 45. Enriched GO Terms among Significantly Differentially Expressed Genes
simultaneous housing contrast; multi only

GO Term	Description	p-value		
		Fisher	K-S	ontology
GO:0006584	catecholamine metabolic process	7.80×10^{-5}	6.92×10^{-3}	BP
GO:0009712	NA	7.80×10^{-5}	6.92×10^{-3}	BP

GO:0009617	response to bacterium	6.90×10^{-4}	2.30×10^{-3}	BP
GO:0042381	hemolymph coagulation	7.80×10^{-4}	2.90×10^{-5}	BP
GO:0006520	cellular amino acid metabolic process	1.11×10^{-3}	2.30×10^{-4}	BP
GO:0007599	hemostasis	1.18×10^{-3}	2.70×10^{-5}	BP
GO:0050817	NA	1.18×10^{-3}	2.70×10^{-5}	BP
GO:0009607	NA	1.26×10^{-3}	1.20×10^{-6}	BP
GO:0043207	NA	1.26×10^{-3}	1.20×10^{-6}	BP
GO:0051707	NA	1.26×10^{-3}	1.20×10^{-6}	BP
GO:0045087	innate immune response	1.36×10^{-3}	5.12×10^{-3}	BP
GO:1901605	NA	1.60×10^{-3}	2.80×10^{-3}	BP
GO:0042440	pigment metabolic process	2.02×10^{-3}	3.60×10^{-4}	BP
GO:0009620	response to fungus	2.83×10^{-3}	1.98×10^{-3}	BP
GO:0050878	NA	3.21×10^{-3}	2.60×10^{-4}	BP
GO:0010035	NA	3.87×10^{-3}	8.70×10^{-3}	BP
GO:0009408	response to heat	7.05×10^{-3}	2.70×10^{-8}	BP
GO:0009411	response to UV	7.70×10^{-3}	9.66×10^{-3}	BP
GO:0042221	NA	8.85×10^{-3}	2.20×10^{-21}	BP
GO:0005576	extracellular region	3.00×10^{-6}	4.56×10^{-3}	CC
GO:0030017	sarcomere	1.82×10^{-3}	1.40×10^{-6}	CC
GO:0030016	myofibril	2.24×10^{-3}	3.10×10^{-6}	CC
GO:0044449	NA	2.60×10^{-3}	3.10×10^{-6}	CC
GO:0043292	NA	2.85×10^{-3}	2.90×10^{-6}	CC

catechol-containing compound metabolic process (GO:0009712)
 coagulation (GO:0050817)
 response to biotic stimulus (GO:0009607)
 response to external biotic stimulus (GO:0043207)
 response to other organism (GO:0051707)
 alpha-amino acid metabolic process (GO:1901605)
 regulation of body fluid levels (GO:0050878)
 response to inorganic substance (GO:0010035)
 response to chemical (GO:0042221)
 obsolete contractile fiber part (GO:0044449)
 contractile fiber (GO:0043292)

“Bare” GO terms are mostly response (stimulus, chemical) and metabolic (amino acid, catechol) processes, and contractile fiber components.

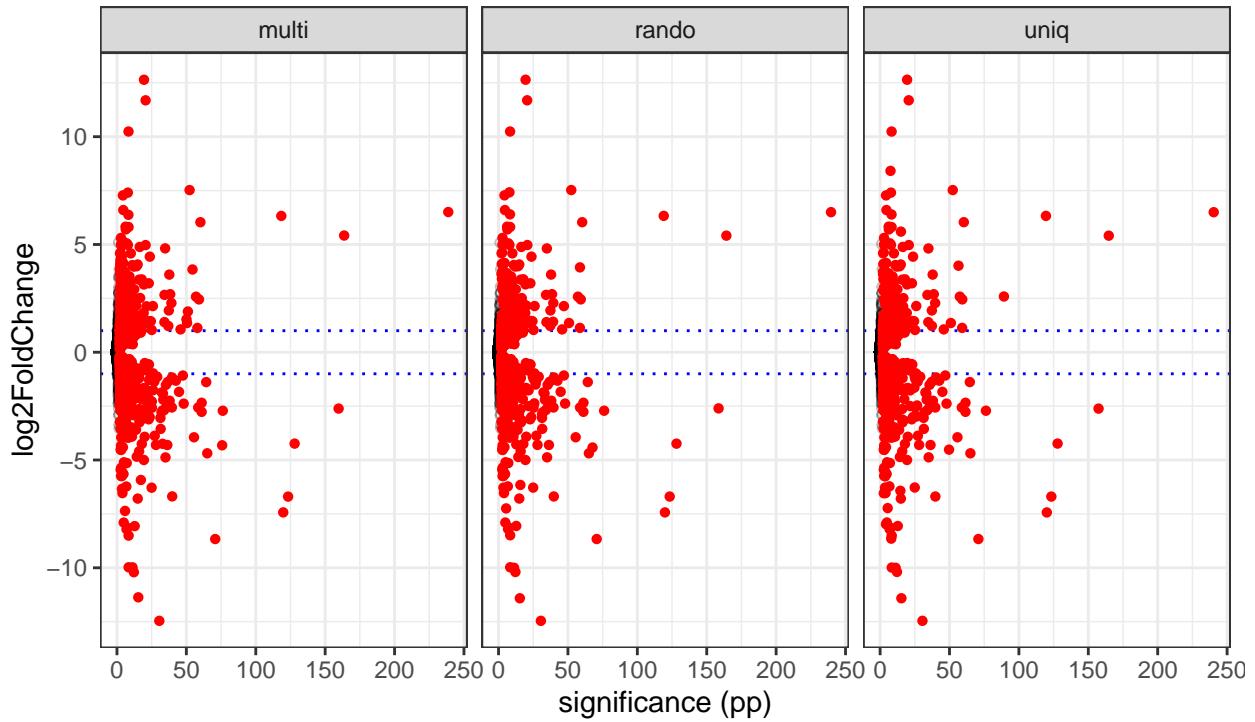
Minor differences w/aligner; see tables/supp/

3.3.4 47b vs wt (Simultaneous model)

Of the 11855 genes with significance scores available, 1039 have an adjusted p < 0.01 (8.7642345 %)

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of +/-1 (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

**Figure 40. Volcano Plot: Fold Change vs. Significance
(between isolated and group-housed wildtypes)
(simultaneous model)**



```
## pdf
## 2
```

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 518 such genes, mostly shared across alignment strategy:

(Table available at [results/tables/tbl47_hausWtVsMut_genotype47b_chonky.html](#)

3.3.4.1 Top Tens

Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes:

Table 47. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed between isolated and group-housed wildtypes

multi					rando				
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p	
1	CG6912	0.36	6.506	2.04×10^{-239}	CG6912	0.36	6.500	2.83×10^{-239}	
2	CG7900	1.76	5.414	2.09×10^{-164}	CG7900	1.76	5.410	7.21×10^{-164}	
3	Drip	1.40	-2.614	1.99×10^{-160}	Drip	1.40	-2.605	2.89×10^{-160}	
4	DIP-alpha	0.06	-4.240	9.84×10^{-129}	DIP-alpha	0.06	-4.240	6.26×10^{-129}	
5	5-HT2A	0.09	-6.697	5.00×10^{-124}	5-HT2A	0.09	-6.696	4.86×10^{-124}	
6	Or47b	0.92	-7.429	1.61×10^{-120}	Or47b	0.92	-7.431	1.33×10^{-120}	

7	CG8665	0.07	6.325	3.97×10^{-119}	CG8665	0.07	6.328	9.70×10^{-119}
8	Or85b	0.44	-2.713	5.52×10^{-77}	Or85b	0.44	-2.712	8.28×10^{-77}
9	Cyp12d1-p	0.05	-4.317	1.65×10^{-76}	Cyp6a17	0.87	-8.665	1.89×10^{-76}
10	Cyp6a17	0.87	-8.666	1.60×10^{-71}	Cyp12d1-p	0.05	-4.420	1.87×10^{-71}

Top 10 genes with biggest (significant) effect sizes

Table 48. Top Ten Largest Magnitude Fold Changes which were significant between isolated and group-housed wildtypes

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	mthl8	0.06	12.643	3.27×10^{-20}	mthl8	0.06	12.643
2	CG40486	3.75	-12.461	3.08×10^{-31}	CG40486	3.74	-12.461
3	w	0.53	11.689	2.35×10^{-21}	w	0.53	11.689
4	CG30428	0.16	-11.371	4.36×10^{-16}	CG30428	0.16	-11.371
5	CG43149	0.16	10.241	4.74×10^{-9}	CG43149	0.16	10.241
6	ppk19	0.04	-10.205	6.79×10^{-13}	ppk19	0.04	-10.205
7	lncRNA:CR45502	0.14	-9.979	6.59×10^{-12}	lncRNA:CR45502	0.14	-9.979
8	CheA7a	0.07	-9.977	3.96×10^{-9}	CheA7a	0.07	-9.977
9	Cyp6a17	0.87	-8.666	1.60×10^{-71}	Cyp6a17	0.87	-8.666
10	asRNA:CR44030	0.05	-8.507	4.12×10^{-9}	asRNA:CR44030	0.05	-8.507

Top 10 highest expressed genes with significant change

Table 49. Top Ten Highest Expressed Genes with Significant (p < 0.01) Difference between isolated and group-housed wildtypes

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	Obp28a	29.11	0.469	3.96×10^{-3}	Obp28a	29.11	0.468
2	Drsl5	16.49	-0.912	4.22×10^{-6}	Drsl5	16.50	-0.910
3	lncRNA:noe	16.02	-0.872	3.68×10^{-7}	lncRNA:noe	16.02	-0.870
4	to	15.53	-2.201	4.23×10^{-34}	to	15.53	-2.201
5	Est-6	12.02	0.398	4.27×10^{-3}	Est-6	12.02	0.399
6	lush	11.92	0.504	6.04×10^{-3}	lush	11.92	0.504
7	CG11550	10.81	0.596	1.40×10^{-4}	CG11550	10.82	0.597
8	Obp59a	8.89	0.556	4.02×10^{-9}	Obp59a	8.90	0.559
9	CG43093	6.14	-0.861	6.33×10^{-12}	CG43093	6.14	-0.860
10	CG30197	6.10	1.095	1.98×10^{-10}	CG30197	6.11	1.097

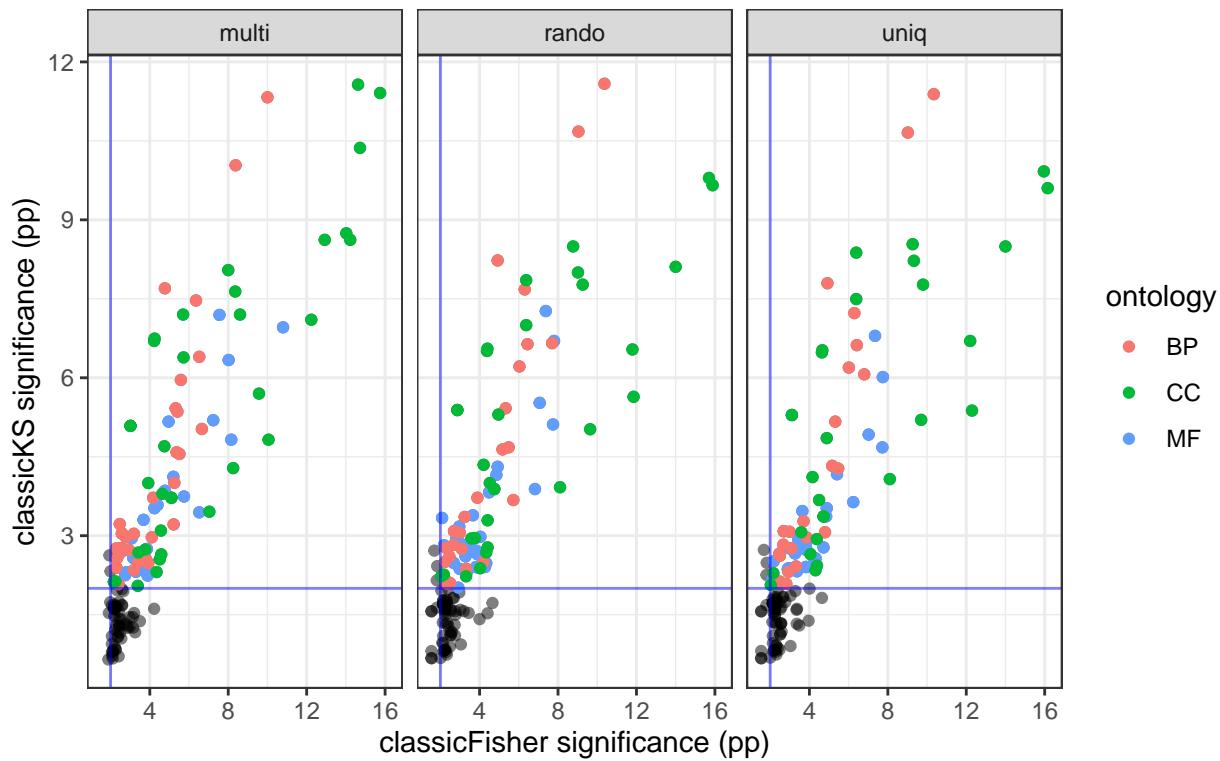
3.3.4.2 Gene Ontology Enrichment

Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

-> check consistency between GO terms between alignment strategies -> filter out very broad/very specific terms?

Correlation between significance values for the two tests

Figure 40. Scatterplot of GO Term Enrichment Significance for Two Tests (47b contrast from simultaneous model)



molecular transducer activity (GO:0060089)
 sensory perception (GO:0007600)
 system process (GO:0003008)
 DNA packaging complex (GO:0044815)
 obsolete membrane part (GO:0044425)

Table 50. Enriched GO Terms among Significantly Differentially Expressed Genes
 simultaneous 47b contrast; multi only; top 10 most significant per category

GO Term	Description	p-value	
		Fisher	K-S
MF			
GO:0046982	protein heterodimerization activity	1.60×10^{-11}	1.10×10^{-7}
GO:0004888	transmembrane signaling receptor activity	7.00×10^{-9}	1.50×10^{-5}
GO:0038023	signaling receptor activity	9.50×10^{-9}	4.60×10^{-7}
GO:0005549	odorant binding	2.80×10^{-8}	6.40×10^{-8}
GO:0060089	NA	5.80×10^{-8}	6.40×10^{-6}
GO:0046983	protein dimerization activity	3.00×10^{-7}	3.60×10^{-4}
GO:0005506	iron ion binding	1.80×10^{-6}	1.80×10^{-4}
GO:0004984	olfactory receptor activity	6.30×10^{-6}	7.60×10^{-5}
GO:0031492	nucleosomal DNA binding	1.10×10^{-5}	6.80×10^{-6}
GO:0022857	transmembrane transporter activity	1.70×10^{-5}	1.40×10^{-4}

BP

GO:0007606	sensory perception of chemical stimulus	1.00×10^{-10}	4.70×10^{-12}
GO:0007600	NA	4.20×10^{-9}	9.20×10^{-11}
GO:0006334	nucleosome assembly	2.20×10^{-7}	9.40×10^{-6}
GO:0050896	response to stimulus	3.00×10^{-7}	4.00×10^{-7}
GO:0007608	sensory perception of smell	4.40×10^{-7}	3.40×10^{-8}
GO:0003008	NA	2.60×10^{-6}	1.10×10^{-6}
GO:0009593	detection of chemical stimulus	3.10×10^{-6}	2.80×10^{-5}
GO:0050906	detection of stimulus involved in sensory perception	3.90×10^{-6}	4.40×10^{-6}
GO:0050907	detection of chemical stimulus involved in sensory perception	4.30×10^{-6}	2.60×10^{-5}
GO:0050877	nervous system process	4.80×10^{-6}	3.80×10^{-6}

CC

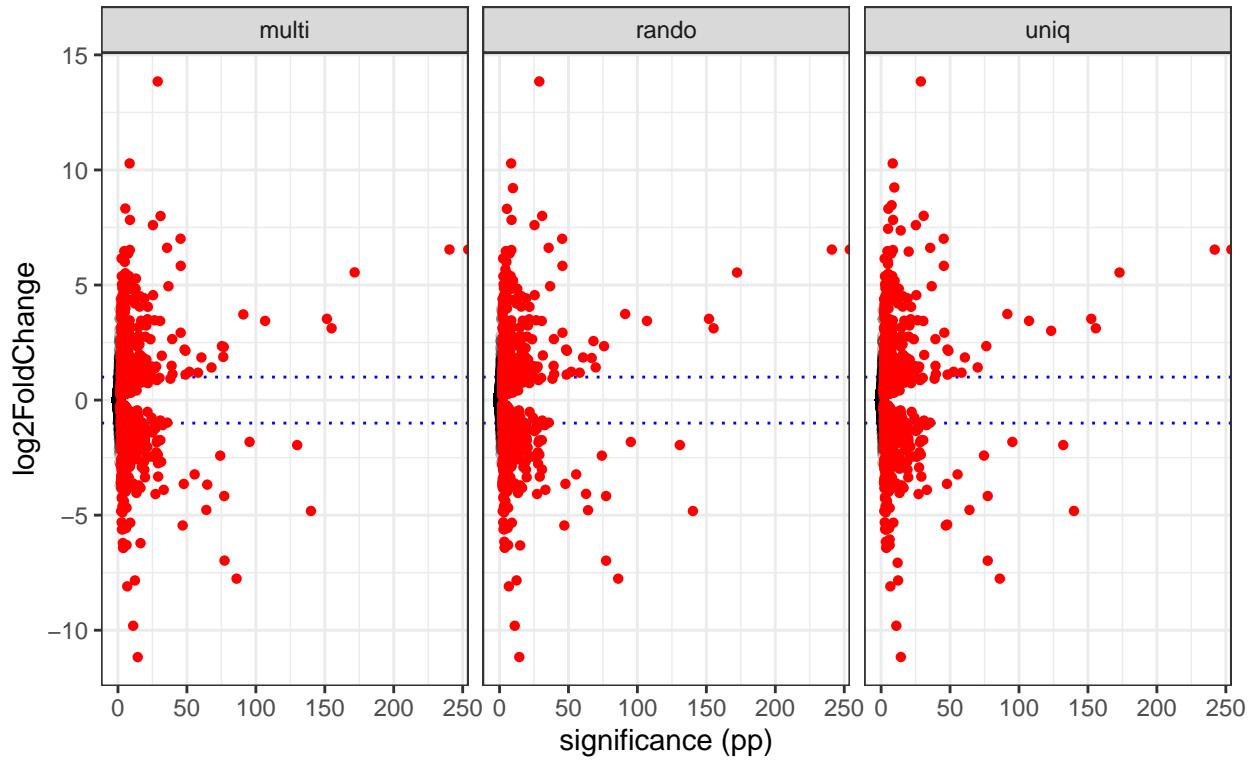
GO:0000786	nucleosome	1.80×10^{-16}	3.90×10^{-12}
GO:0044815	NA	1.90×10^{-15}	4.30×10^{-11}
GO:0000788	nuclear nucleosome	2.40×10^{-15}	2.70×10^{-12}
GO:0016021	integral component of membrane	6.00×10^{-15}	2.40×10^{-9}
GO:0031224	intrinsic component of membrane	9.50×10^{-15}	1.80×10^{-9}
GO:0032993	protein-DNA complex	1.20×10^{-13}	2.40×10^{-9}
GO:0005576	extracellular region	5.80×10^{-13}	7.90×10^{-8}
GO:0044425	NA	8.80×10^{-11}	1.50×10^{-5}
GO:0016020	membrane	2.70×10^{-10}	2.00×10^{-6}
GO:0005886	plasma membrane	2.50×10^{-9}	6.30×10^{-8}

3.3.5 67d vs wt (Simultaneous model)

Of the 11597 genes with significance scores available, 1299 have an adjusted p < 0.01 (11.2011727 %)

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of +/-1 (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

**Figure 41. Volcano Plot: Fold Change vs. Significance
(between isolated and group-housed wildtypes)**



```
## pdf
## 2
```

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 567 such genes, mostly shared across alignment strategy:

(Table available at [results/tables/tbl52_hausWtVsMut_genotype67d_chonky.html](#))

3.3.5.1 Top Tens

Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes:

Table 52. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed between isolated and group-housed wildtypes

multi					rando				
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p	
1	l(2)03659	0.17	6.542	0.00	l(2)03659	0.17	6.542		
2	CG7900	1.76	6.543	3.76×10^{-241}	CG7900	1.76	6.541	1.03×10^{-241}	
3	CG6912	0.36	5.550	2.38×10^{-172}	CG6912	0.36	5.544	5.94×10^{-172}	
4	Cyp9b1	0.45	3.120	1.18×10^{-155}	Cyp9b1	0.45	3.120	5.82×10^{-155}	
5	CG10936	0.05	3.530	2.92×10^{-152}	CG10936	0.05	3.530	1.31×10^{-152}	
6	DIP-alpha	0.06	-4.818	1.09×10^{-140}	DIP-alpha	0.06	-4.820	5.90×10^{-140}	

7	NijC	1.54	-1.955	1.07×10^{-130}	NijC	1.54	-1.953	1.88×10^{-130}
8	CG32407	0.14	3.441	2.07×10^{-107}	CG32407	0.14	3.441	1.26×10^{-107}
9	Cyp6a20	5.59	-1.815	3.61×10^{-96}	Cyp6a20	5.59	-1.818	6.38×10^{-96}
10	ppk25	0.16	3.723	1.24×10^{-91}	ppk25	0.16	3.742	7.63×10^{-91}

Top 10 genes with biggest (significant) effect sizes

Table 53. Top Ten Largest Magnitude Fold Changes which were significant between isolated and group-housed wildtypes

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	w	0.53	13.846	1.54×10^{-29}	w	0.53	13.846
2	CG32437	0.04	-11.168	4.82×10^{-15}	CG32437	0.04	-11.168
3	CG43149	0.16	10.284	3.65×10^{-9}	CG43149	0.16	10.284
4	lncRNA:CR44111	0.06	-9.806	9.61×10^{-12}	lncRNA:CR44111	0.06	-9.806
5	ppk9	0.00	8.323	4.70×10^{-6}	CG43291	0.01	9.121
6	lncRNA:CR44377	0.01	-8.093	1.82×10^{-7}	ppk9	0.00	8.323
7	lncRNA:dntRL	0.06	7.999	1.35×10^{-31}	lncRNA:CR44377	0.01	-8.093
8	CG9010	0.06	-7.836	4.77×10^{-13}	lncRNA:dntRL	0.06	8.093
9	Obp83g	0.03	7.829	2.08×10^{-9}	CG9010	0.06	-7.836
10	5-HT2A	0.09	-7.759	9.84×10^{-87}	Obp83g	0.03	7.829

Top 10 highest expressed genes with significant change

Table 54. Top Ten Highest Expressed Genes with Significant (padj) Difference between isolated and group-housed wildtypes

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	Obp83b	58.57	0.392	1.64×10^{-3}	Obp83b	58.58	0.392
2	Obp69a	23.74	0.682	3.41×10^{-8}	Obp69a	23.75	0.681
3	lush	11.92	0.768	8.71×10^{-6}	lush	11.92	0.768
4	Cyp6w1	11.34	0.583	2.01×10^{-4}	Cyp6w1	11.34	0.582
5	Snmp1	10.00	-0.385	1.38×10^{-3}	Snmp1	10.00	-0.377
6	Obp56d	9.02	0.737	3.02×10^{-3}	Obp56d	9.03	0.740
7	CG1927	8.57	0.338	8.52×10^{-4}	CG1927	8.57	0.338
8	Ldsdh1	6.16	0.485	1.79×10^{-3}	Ldsdh1	6.16	0.481
9	CG30197	6.10	1.037	1.91×10^{-9}	CG30197	6.11	1.037
10	Cyp6a2	5.98	2.752	1.12×10^{-20}	Cyp6a2	5.98	2.756

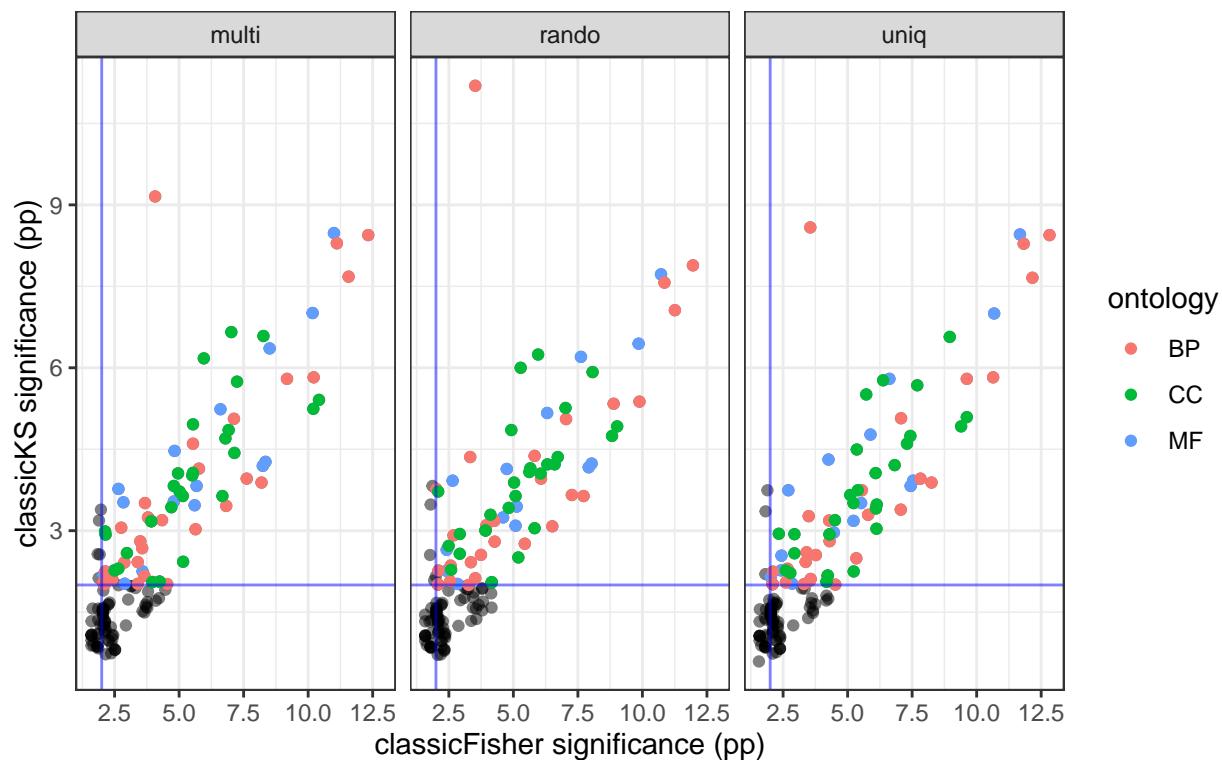
3.3.5.2 Gene Ontology Enrichment

Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

-> check consistency between GO terms between alignment strategies -> filter out very broad/very specific terms?

Correlation between significance values for the two tests

Figure 41. Scatterplot of GO Term Enrichment Significance for Two Tests (67d contrast from simultaneous model)



tetrapyrrole binding (GO:0046906)
 sensory perception (GO:0007600)
 detection of stimulus (GO:0051606)
 cell projection membrane (GO:0031253)
 obsolete plasma membrane part (GO:0044459)
 leading edge membrane (GO:0031256)
 obsolete membrane part (GO:0044425)

Table 55. Enriched GO Terms among Significantly Differentially Expressed Genes
 simultaneous 47b contrast; multi only; top 10 most significant per category

GO Term	Description	
MF		
GO:0004984	olfactory receptor activity	1.00
GO:0005549	odorant binding	6.70
GO:0005506	iron ion binding	3.1
GO:0020037	heme binding	4.5
GO:0046906	NA	5.5
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	2.5
GO:0005215	transporter activity	2.1
GO:0022857	transmembrane transporter activity	2.5

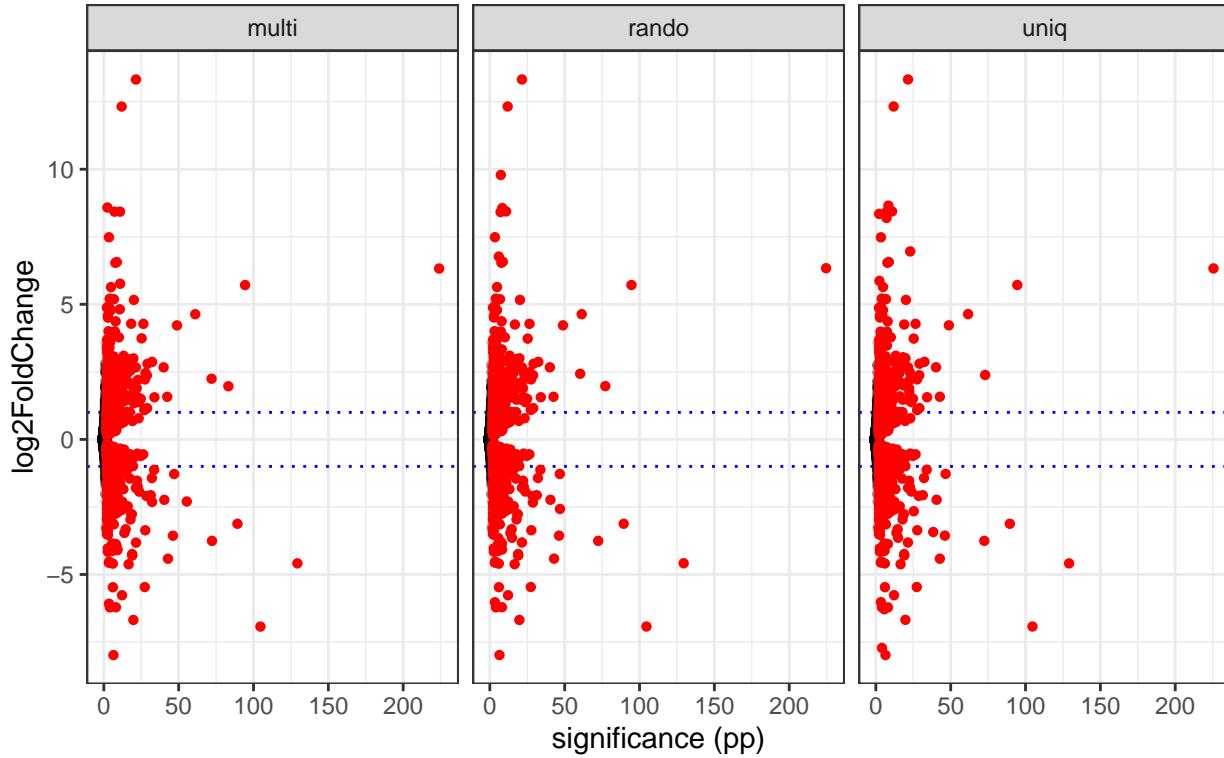
GO:0048037	cofactor binding	1.5
GO:0016491	oxidoreductase activity	1.0
<hr/>		
BP		
GO:0007608	sensory perception of smell	4.7
GO:0050907	detection of chemical stimulus involved in sensory perception	2.7
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	7.7
GO:0009593	detection of chemical stimulus	6.1
GO:0050906	detection of stimulus involved in sensory perception	6.6
GO:0007600	NA	6.4
GO:0007606	sensory perception of chemical stimulus	2.4
GO:0035721	intraciliary retrograde transport	7.5
GO:0051606	NA	1.5
GO:0060271	cilium assembly	1.7
<hr/>		
CC		
GO:0016021	integral component of membrane	3.8
GO:0031224	intrinsic component of membrane	6.3
GO:0032590	dendrite membrane	5.4
GO:0031253	NA	5.5
GO:0044459	NA	7.2
GO:0016020	membrane	9.5
GO:0032589	neuron projection membrane	1.2
GO:0031256	NA	1.0
GO:0044425	NA	2.1
GO:0071944	cell periphery	1.1

3.3.6 FruLexA/Fru440 vs wt (Simultaneous model)

Of the 11340 genes with significance scores available, 1764 have an adjusted $p < 0.01$ (15.5555556 %)

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

**Figure 42. Volcano Plot: Fold Change vs. Significance
(between isolated and group-housed wildtypes)**



```
## pdf
## 2
```

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 537 such genes, mostly shared across alignment strategy:

(Table available at [results/tables/tbl57_hausWtVsMut_genotypeFruLexa440_chonky.html](#))

3.3.6.1 Top Tens

Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes:

Table 58. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed Genes between isolated and group-housed wildtypes

multi					rando				
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p	
1	CG7900	1.76	6.325	9.47×10^{-225}	CG7900	1.76	6.335	2.18×10^{-225}	
2	DIP-alpha	0.06	-4.588	4.76×10^{-130}	DIP-alpha	0.06	-4.585	3.33×10^{-130}	
3	5-HT2A	0.09	-6.928	2.36×10^{-105}	5-HT2A	0.09	-6.927	2.36×10^{-105}	
4	CG11893	0.14	5.712	4.23×10^{-95}	CG11893	0.14	5.714	2.45×10^{-95}	
5	Ets21C	0.07	-3.124	5.26×10^{-90}	Ets21C	0.07	-3.123	3.82×10^{-90}	
6	CG32640	2.76	1.967	6.21×10^{-84}	CG32640	1.88	1.973	7.01×10^{-84}	

7	prom	0.04	-3.756	5.29×10^{-73}	prom	0.04	-3.755	3.93×10^{-73}
8	CG32641	3.69	2.246	1.05×10^{-72}	CG8665	0.07	4.635	4.22×10^{-72}
9	CG8665	0.07	4.633	8.11×10^{-62}	CG32641	2.12	2.431	4.38×10^{-62}
10	Or19b	0.39	-2.298	3.79×10^{-56}	CG42526	0.05	4.222	1.36×10^{-56}

Top 10 genes with biggest (significant) effect sizes

Table 58. Top Ten Largest Magnitude between isolated and group-housed wildtypes

rank	name	FB ID	multi			name	FB ID
			expression	log2 FoldChange	adjusted p		
1	mthl8	FBgn0052475	0.06	13.321	3.46×10^{-22}	mthl8	FBgn0052475
2	CG43149	FBgn0262679	0.16	12.320	1.27×10^{-12}	CG43149	FBgn0262679
3	CG9287	FBgn0032057	0.01	8.583	4.06×10^{-3}	CG9287	FBgn0032057
4	w	FBgn0003996	0.53	8.430	1.45×10^{-11}	CG43291	FBgn0262679
5	ppk27	FBgn0035458	0.01	8.429	6.89×10^{-8}	w	FBgn0003996
6	lncRNA:CR44377	FBgn0265527	0.01	-7.985	3.71×10^{-7}	ppk27	FBgn0035458
7	lncRNA:CR44285	FBgn0265312	0.03	7.481	3.83×10^{-4}	lncRNA:CR44377	FBgn0265527
8	5-HT2A	FBgn0087012	0.09	-6.928	2.36×10^{-105}	lncRNA:CR44285	FBgn0265312
9	CG32437	FBgn0052437	0.04	-6.684	1.88×10^{-20}	5-HT2A	FBgn0087012
10	CR44003	FBgn0264745	0.02	6.563	2.16×10^{-9}	CR45496	FBgn0264745

Table 58. Top Ten Largest Magnitude Fold Changes which between isolated and group-housed wildtypes

rank	name	multi			name	rando		
		expression	log2 FoldChange	adjusted p		expression	log2 FoldChange	adjusted p
1	mthl8	0.06	13.321	3.46×10^{-22}	mthl8	0.06	13.321	3.46×10^{-22}
2	CG43149	0.16	12.320	1.27×10^{-12}	CG43149	0.16	12.320	1.27×10^{-12}
3	CG9287	0.01	8.583	4.06×10^{-3}	CG9287	0.01	8.583	4.06×10^{-3}
4	w	0.53	8.430	1.45×10^{-11}	CG43291	0.01	8.430	1.45×10^{-11}
5	ppk27	0.01	8.429	6.89×10^{-8}	w	0.53	8.429	6.89×10^{-8}
6	lncRNA:CR44377	0.01	-7.985	3.71×10^{-7}	ppk27	0.01	8.429	6.89×10^{-8}
7	lncRNA:CR44285	0.03	7.481	3.83×10^{-4}	lncRNA:CR44377	0.01	7.481	3.83×10^{-4}
8	5-HT2A	0.09	-6.928	2.36×10^{-105}	lncRNA:CR44285	0.03	7.481	3.83×10^{-4}
9	CG32437	0.04	-6.684	1.88×10^{-20}	5-HT2A	0.09	-6.684	1.88×10^{-20}
10	CR44003	0.02	6.563	2.16×10^{-9}	CR45496	0.02	6.563	2.16×10^{-9}

Top 10 highest expressed genes with significant change

Table 60. Top Ten Highest Expressed Genes with Significant (padj < Difference) between isolated and group-housed wildtypes

rank	name	multi			name	rando		
		expression	log2 FoldChange	adjusted p		expression	log2 FoldChange	adjusted p
1	Obp83b	58.57	0.405	8.49×10^{-4}	Obp83b	58.58	0.406	7.70×10^{-4}
2	Obp19d	49.59	0.397	2.53×10^{-3}	Obp19d	49.60	0.399	2.39×10^{-3}

3	Jhedup	27.70	-0.463	4.67×10^{-3}	Jhedup	27.71	-0.461	4.68×10^{-3}
4	Obp69a	23.74	0.501	8.03×10^{-5}	Obp69a	23.75	0.502	7.73×10^{-5}
5	Orco	13.02	-0.600	6.10×10^{-4}	Orco	13.02	-0.596	6.00×10^{-4}
6	lush	11.92	0.459	9.96×10^{-3}	lush	11.92	0.460	9.46×10^{-3}
7	Obp56d	9.02	1.194	8.24×10^{-7}	Obp56d	9.03	1.196	7.80×10^{-7}
8	Obp59a	8.89	0.338	7.16×10^{-4}	Obp59a	8.90	0.336	6.16×10^{-4}
9	sesB	8.63	0.358	1.72×10^{-4}	sesB	8.63	0.359	1.67×10^{-4}
10	CG1927	8.57	0.271	8.31×10^{-3}	CG1927	8.57	0.269	7.93×10^{-3}

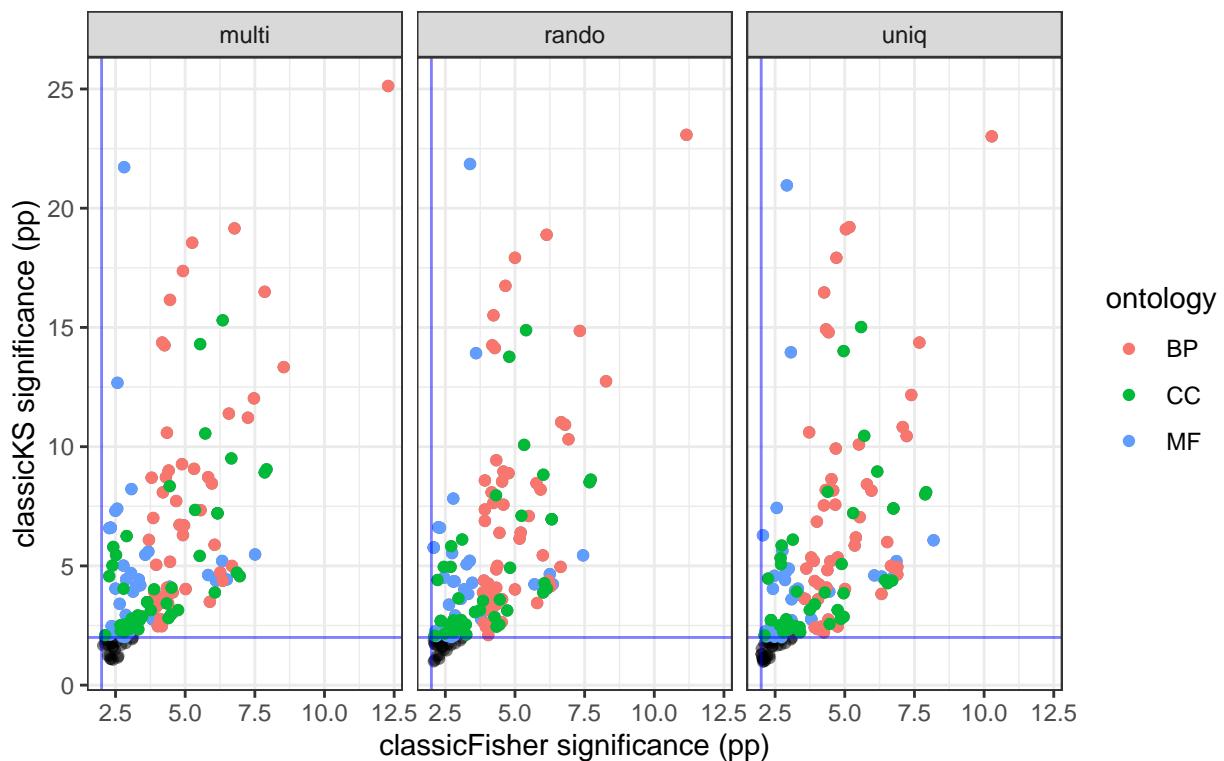
3.3.6.2 Gene Ontology Enrichment

Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

-> check consistency between GO terms between alignment strategies -> filter out very broad/very specific terms?

Correlation between significance values for the two tests

Figure 42. Scatterplot of GO Term Enrichment Significance for Two Tests (FruLexa/Fru440 contrast from simultaneous model)



molecular transducer activity (GO:0060089)
mannosyl-oligosaccharide mannosidase activity (GO:0015924)
adenyl nucleotide binding (GO:0030554)
response to chemical (GO:0042221)
transport (GO:0006810)
establishment of localization (GO:0051234)
cellular response to stimulus (GO:0051716)

plasma membrane bounded cell projection (GO:0120025)
 obsolete cell projection part (GO:0044463)
 obsolete plasma membrane bounded cell projection part (GO:0120038)
 obsolete plasma membrane part (GO:0044459)

Table 61. Enriched GO Terms among Significantly Differentially Expressed Genes simultaneous FruLex440 contrast; multi only; top 10 most significant per category

GO Term	Description	p-value	
		Fisher	K-S
MF			
GO:0004984	olfactory receptor activity	3.10×10^{-8}	3.30×10^{-6}
GO:0060089	NA	3.20×10^{-7}	3.70×10^{-5}
GO:0005549	odorant binding	4.70×10^{-7}	6.20×10^{-6}
GO:0038023	signaling receptor activity	7.90×10^{-7}	4.20×10^{-5}
GO:0004888	transmembrane signaling receptor activity	1.50×10^{-6}	2.40×10^{-5}
GO:0008092	cytoskeletal protein binding	3.60×10^{-5}	7.40×10^{-5}
GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	1.50×10^{-4}	1.71×10^{-3}
GO:0015924	NA	1.50×10^{-4}	1.71×10^{-3}
GO:0032559	adenyl ribonucleotide binding	2.10×10^{-4}	2.50×10^{-6}
GO:0030554	NA	2.70×10^{-4}	3.40×10^{-6}
BP			
GO:0050896	response to stimulus	5.20×10^{-13}	7.50×10^{-26}
GO:0042221	NA	2.90×10^{-9}	4.60×10^{-14}
GO:0051179	NA	1.40×10^{-8}	3.20×10^{-17}
GO:0006810	NA	3.40×10^{-8}	9.40×10^{-13}
GO:0051234	NA	5.60×10^{-8}	6.10×10^{-12}
GO:0051716	NA	1.70×10^{-7}	7.00×10^{-20}
GO:1905515	non-motile cilium assembly	2.10×10^{-7}	1.00×10^{-5}
GO:0010033	response to organic substance	2.70×10^{-7}	4.10×10^{-12}
GO:0050907	detection of chemical stimulus involved in sensory perception	4.50×10^{-7}	4.20×10^{-5}
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	5.50×10^{-7}	1.90×10^{-5}
CC			
GO:0120025	NA	1.20×10^{-8}	8.90×10^{-10}
GO:0042995	cell projection	1.40×10^{-8}	1.20×10^{-9}
GO:0016021	integral component of membrane	1.10×10^{-7}	2.70×10^{-5}
GO:0031224	intrinsic component of membrane	1.40×10^{-7}	1.90×10^{-5}
GO:0016020	membrane	2.20×10^{-7}	3.10×10^{-10}
GO:0071944	cell periphery	4.50×10^{-7}	5.00×10^{-16}
GO:0044463	NA	6.90×10^{-7}	6.20×10^{-8}
GO:0120038	NA	6.90×10^{-7}	6.20×10^{-8}
GO:0032590	dendrite membrane	8.60×10^{-7}	1.30×10^{-4}
GO:0044459	NA	1.90×10^{-6}	2.80×10^{-11}

3.3.7 Fruitless-less

Out of data quality concerns, the contrast was rerun with the FruLex4/Fru440 samples excluded.

Every gene that can be analyzed using the Without counts can be analyzed using the With counts; there are 0 that cannot. On the other hand, there are 343 genes which can be analyzed using the With counts but

not the Without:

Table 62. Genes Lost When FruLexa/Fru440 Counts are Excluded
genes which no longer pass minimum count threshold

aligner	count
multi	333
rando	318
uniq	318

In 0 cases were these genes significant ($p_{adj} < 0.01$) in the With tests.

A gene with significance values in both tests may gain significance when FruLexa/Fru440 samples are dropped, lose significance, maintain significance while switching direction, or remain unchanged. No switches were seen, but moderate numbers (up to ~5%) gained significance.

Table 63. Changes in Differential Expression Significance
when FruLexa/Fru440 samples are dropped

	change		
	gain	loss	none
47b1			
multi	576	11	12655
rando	573	10	12498
uniq	569	9	12407
67d			
multi	862	14	12366
rando	835	15	12231
uniq	832	14	12139
isolated			
multi	101	1	13140
rando	101	1	12979
uniq	101	1	12883

In some cases, the significance increase was very large:

Table 64. Top 10 Biggest Significance Changes
when FruLexa/Fru440 samples are dropped

	effect size (l2fc)		adjusted p	
	with	without	with	without
47b1 - multi				
csw	0.57	0.67	0.032	4.02×10^{-17}
kek1	0.77	0.91	0.015	1.51×10^{-15}
Updo	0.56	0.66	0.037	4.75×10^{-12}
Jheh3	-0.51	-0.59	0.033	1.98×10^{-11}
CG10638	-0.30	-0.32	0.037	1.59×10^{-12}
CG13251	-0.28	-0.30	0.033	1.87×10^{-12}
Nrx-1	0.46	0.58	0.123	3.96×10^{-14}

CG42541	0.75	0.87	0.011	6.24×10^{-12}
pyd	0.54	0.61	0.014	5.12×10^{-23}
Fhos	-0.30	-0.58	0.415	7.15×10^{-16}
<hr/>				
67d - multi				
csw	0.55	0.65	0.032	3.20×10^{-16}
Rab3	-0.32	-0.44	0.293	2.74×10^{-12}
GstD8	-0.46	-0.52	0.028	2.36×10^{-12}
Spn	-0.38	-0.41	0.021	1.09×10^{-14}
CG16935	-0.46	-0.58	0.107	1.99×10^{-12}
CG12814	0.60	0.69	0.018	1.03×10^{-14}
RIC-3	-0.50	-0.57	0.032	3.69×10^{-13}
Ir8a	-0.46	-0.56	0.074	1.26×10^{-13}
pyd	0.40	0.47	0.075	1.08×10^{-13}
Fhos	-0.27	-0.51	0.422	1.18×10^{-12}
<hr/>				
isolated - multi				
Trp1	0.05	0.28	0.313	6.22×10^{-6}
CG15202	0.03	0.39	0.310	3.53×10^{-6}
CG9498	0.61	0.74	0.038	1.18×10^{-6}
Ugt301D1	0.33	0.42	0.072	3.24×10^{-7}
Loxl2	0.04	0.45	0.171	8.40×10^{-6}
Cln3	0.02	0.39	0.519	1.11×10^{-4}
ELOVL	0.46	0.52	0.022	3.02×10^{-7}
CG9717	0.02	0.52	0.353	5.24×10^{-6}
vir-1	0.16	0.23	0.188	2.26×10^{-6}
Cda5	-0.02	-0.52	0.233	1.23×10^{-5}

3.3.7.1 Perturbation to Housing Contrast

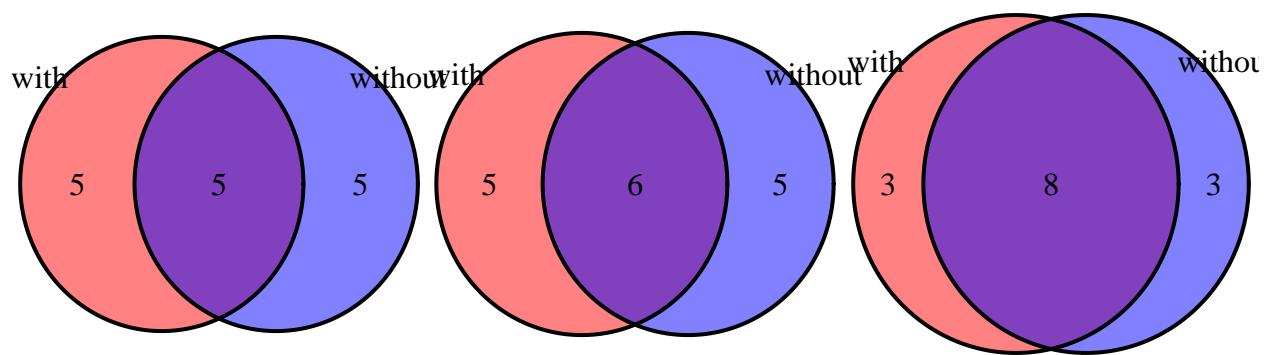
To see how much exclusion of the FruLexa/Fru440 alters the big picture results in the group vs. isolated contrast, we can look at how well the top-10 lists agree (similarity is calculated as size of the intersection divided by size of the union)

Figure 43 . Similarity of Housing Contrast Top 10 Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

adjusted p:
33% similar

log2FoldChange:
38% similar

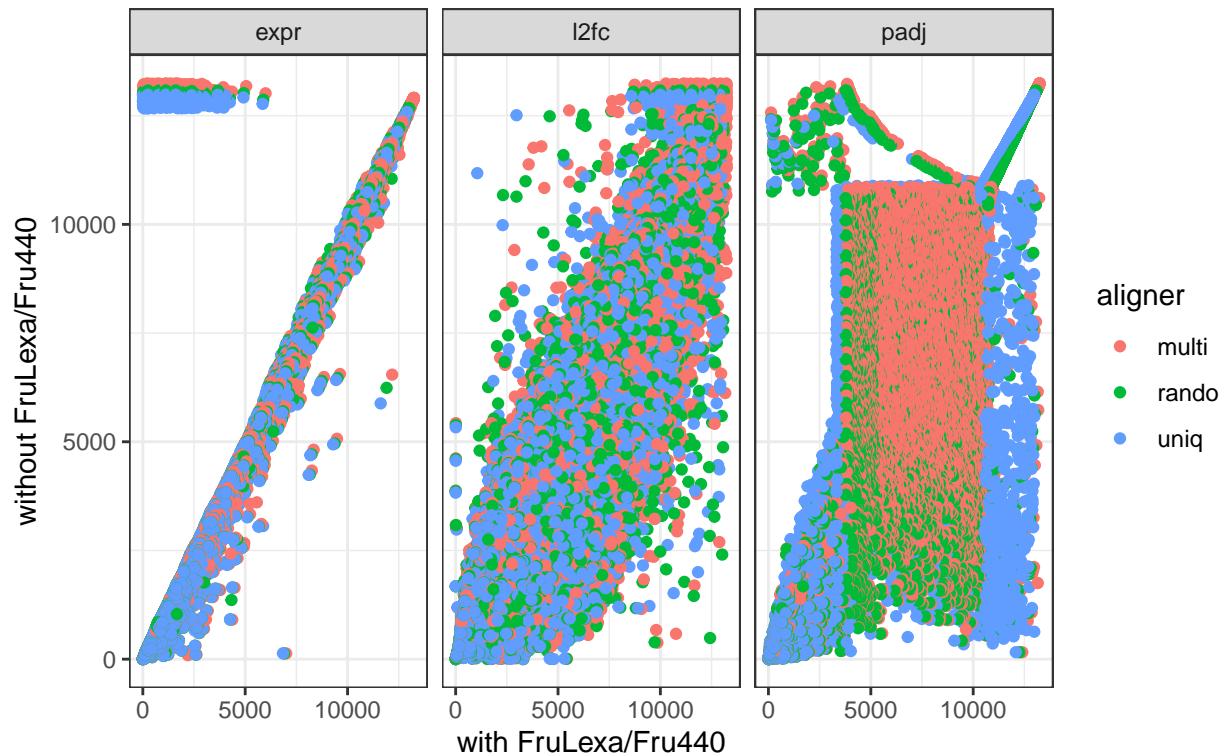
expression:
57% similar



```
## pdf  
## 2
```

We can also look at the rank correlations:

Figure 44. Rank correlations of expression, effect size, and significance (housing contrasts, with/without FruLexa)



```
## pdf
## 2
```

3.3.7.2 Perturbation to 47b1 Contrast

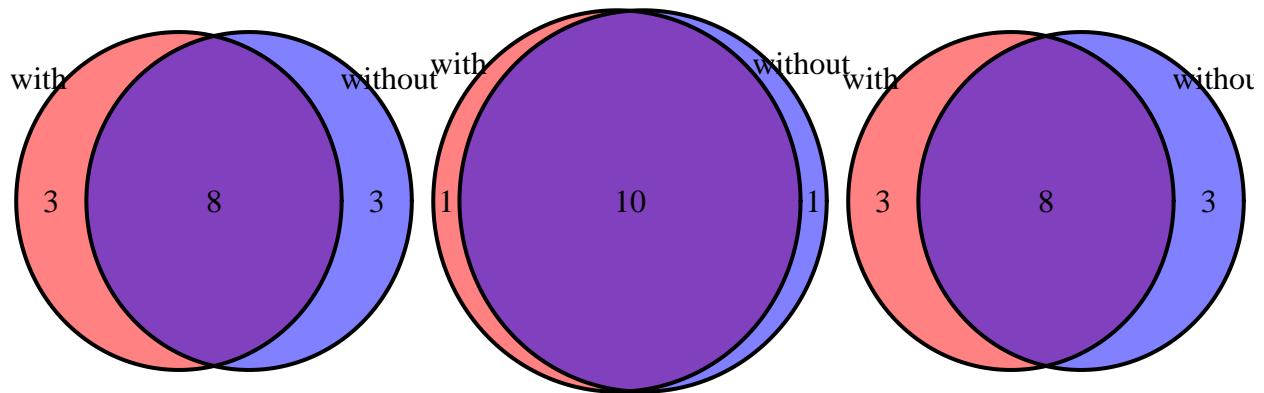
To see how much exclusion of the FruLexa/Fru440 alters the big picture results in the 47b1 vs. wt contrast, we can look at how well the top-10 lists agree (similarity is calculated as size of the intersection divided by size of the union)

Figure 45 . Similarity of 47b1 Contrast Top 10 Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

adjusted p:
57% similar

log2FoldChange:
83% similar

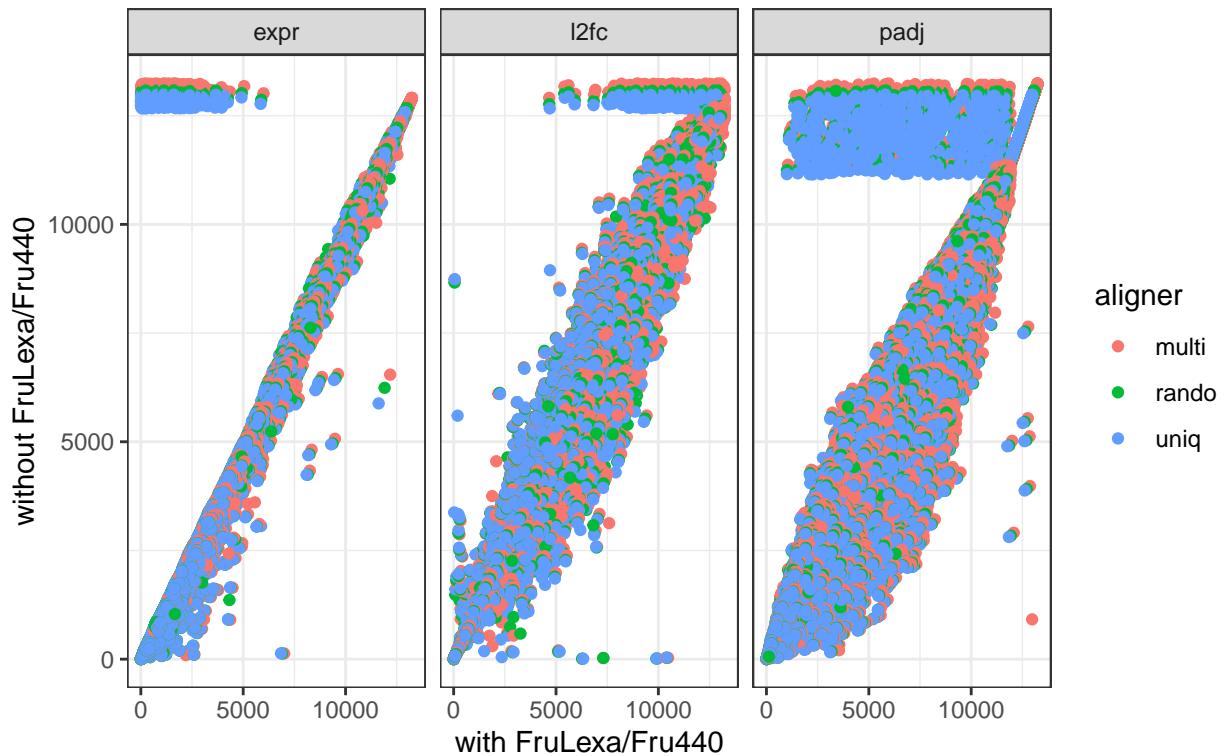
expression:
57% similar



```
## pdf  
## 2
```

We can also look at the rank correlations:

Figure 46. Rank correlations of expression, effect size, and significance (47b1 contrasts, with/without FruLexa)



```
## pdf
## 2
```

3.3.7.3 Perturbation to 67d Contrast

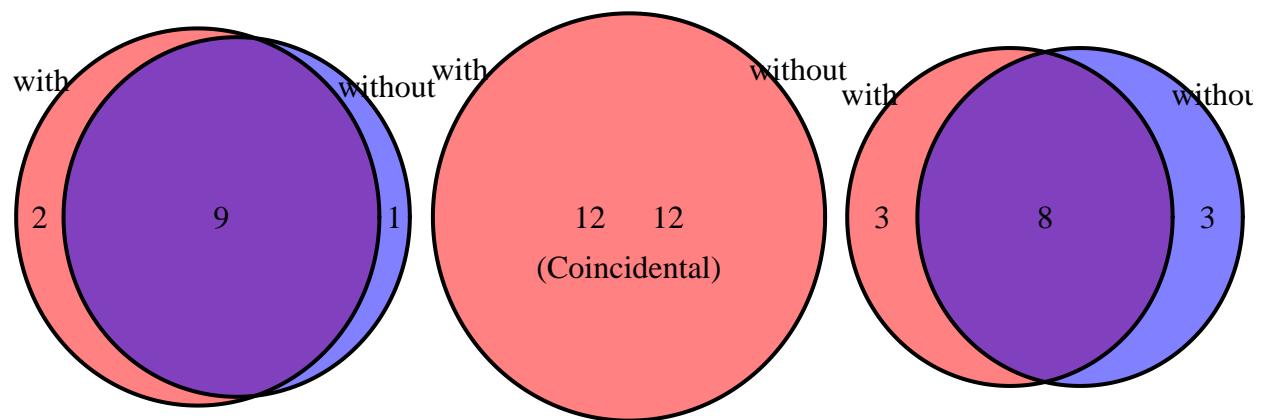
To see how much exclusion of the FruLexa/Fru440 alters the big picture results in the 67d vs. wt contrast, we can look at how well the top-10 lists agree (similarity is calculated as size of the intersection divided by size of the union)

47 . Similarity of 67d Contrast Top 10 Lists, with/without FruLexaFru440 samples (pooled align)

adjusted p:
75% similar

log2FoldChange:
100% similar

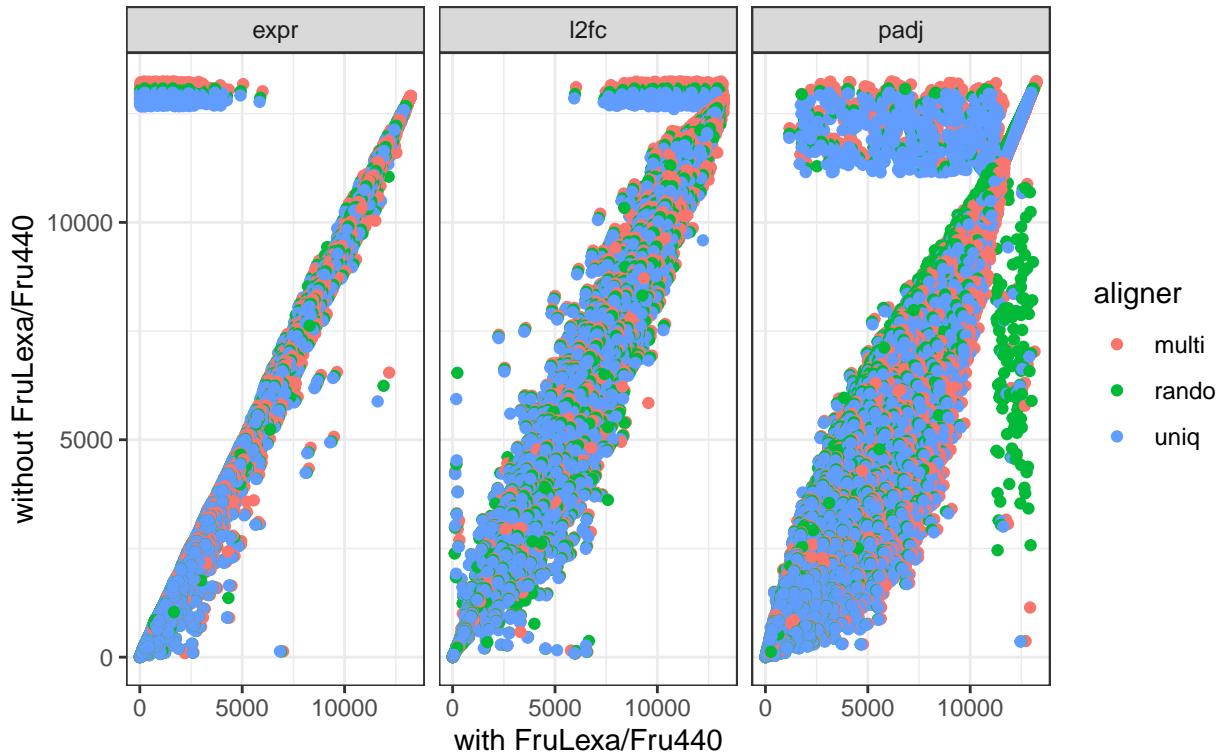
expression:
57% similar



```
## pdf  
## 2
```

We can also look at the rank correlations:

Figure 48. Rank correlations of expression, effect size, and significance (67d contrasts, with/without FruLexa)



```
## pdf
## 2
```

3.4 Comparing Expression Changes from Housing with Expression Changes from Genotype

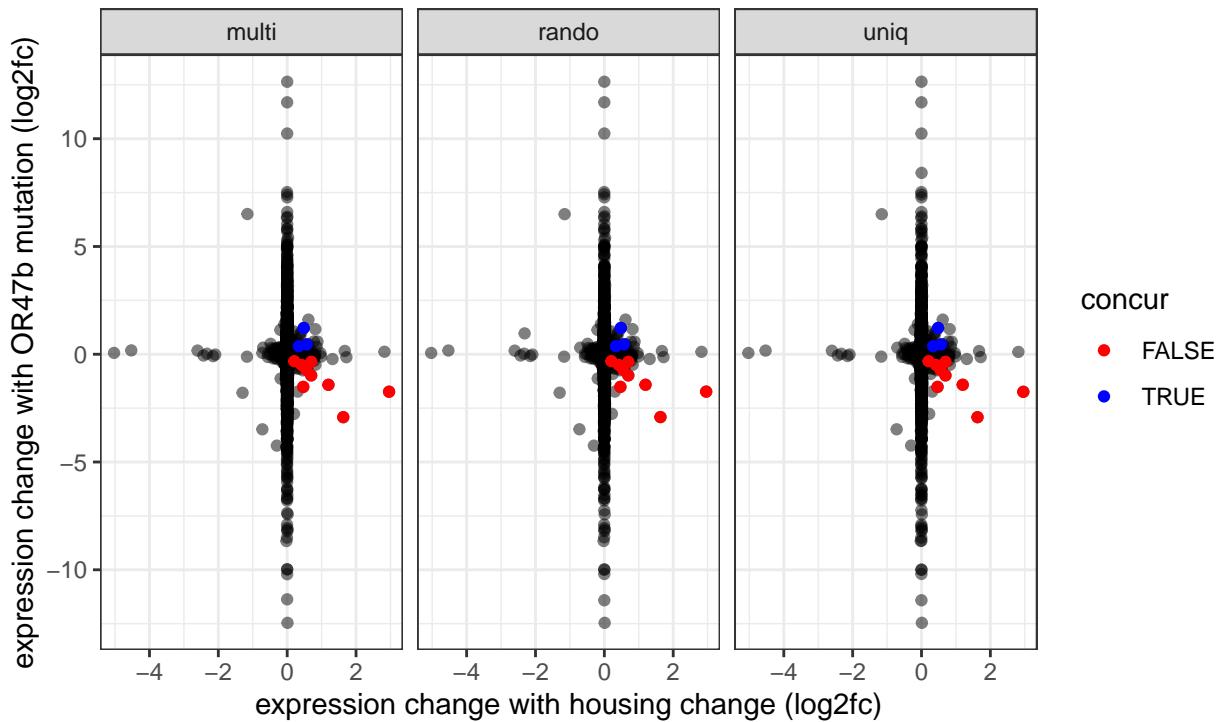
We want to see if the difference in life history creates similar changes in expression as various mutations. This was done using the differential expression data from the genotype & housing simultaneous model. This circumvents the joining step in earlier versions. Earlier versions comparing results from two distinct models readjusted the p-values with a Bonferroni correction using $n=2$; in the current iteration in which both p-values are coming from the same model, this step is skipped. Candidate genes of interest are then collected by filtering this joint comparison for genes which show a significant change in both contrasts. These candidates are further classified as to whether the expression changes are in the same direction (ie, both upregulated or both downregulated) or not (ie, one upregulated and the other down).

Average significance for gene is currently computed as $\exp((\ln(p1)+\ln(p2))/2)$. (Better to apply stouffer's?) look at NAs in fulljoin (gene dropout may be interesting...)

3.4.1 Housing & OR47b

Here is a scatterplot of the log2 fold change of the 47b & wt contrast vs the housing contrast (wt group & wt isolated). The upper right quadrant contains genes which are upregulated in both cases; the lower left contains genes which are downregulated in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

**Figure 49. Scatterplot of Expression Changes in OR47b mutants vs Expression Changes in Housing
(Significant Similarities and Differences Highlighted)**



```
## pdf
## 2
```

Of the mutually significant genes, fewer have the same direction of change than not:

Table 65. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs OR47b

	multi	rando	uniq
Agree	3	3	3
Disagree	9	9	9

Of those mutually significant genes with the same direction of change, the top 10 most significant agree well across alignment strategy:

Table 66. Top Ten Most Significant Genes in difference expression between housing and OR47b

rank	name	multi			name	mean expression
		mean expression	mean readusted p	housing l2fc		
1	jv	0.08	2.22×10^{-21}	0.482	1.219	jv
2	CG13659	0.31	7.56×10^{-4}	0.570	0.453	CG13659

3	PGRP-LB	0.39	1.44×10^{-3}	0.338	0.373	PGRP-LB	0.39
---	---------	------	-----------------------	-------	-------	---------	------

When mutually significant genes with the same direction of change are ranked by the magnitude of their mean log2FoldChange, the top 10 agree well across alignment strategy:

Table 67. Top Ten Largest Magnitude Changes In Significant Genes in difference expression between housing and OR47b contrants

multi					rando				
rank	name	mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean	
1	jv	0.851	0.08	2.22×10^{-21}	jv	0.851	0.08		
2	CG13659	0.511	0.31	7.56×10^{-4}	CG13659	0.511	0.31		
3	PGRP-LB	0.356	0.39	1.44×10^{-3}	PGRP-LB	0.356	0.39		

Of those mutually significant genes with different directions of change, the top 10 most significant agree well across alignment strategy.

Table 68. Top Ten Most Significant Genes of in difference expression between housing and OR47b contrants

multi						rando		
rank	name	mean expression	mean readusted p	housing l2fc	OR47b l2fc	name	mean expression	mean
1	Obp84a	0.62	4.48×10^{-20}	0.467	-1.515	Obp84a	0.62	
2	CG11852	0.07	1.32×10^{-9}	1.629	-2.919	CG11852	0.07	
3	amd	1.01	1.67×10^{-7}	1.197	-1.418	amd	1.01	
4	Fer2LCH	4.18	8.37×10^{-7}	0.205	-0.322	Fer2LCH	4.18	
5	CG10050	0.24	2.52×10^{-6}	0.697	-0.981	CG10050	0.24	
6	magu	0.21	1.16×10^{-5}	0.699	-0.350	magu	0.21	
7	Or92a	1.57	1.89×10^{-5}	0.420	-0.498	Or92a	1.57	
8	CG13332	0.28	1.18×10^{-4}	0.547	-0.710	CG13332	0.28	
9	Jhe	0.26	5.73×10^{-4}	2.959	-1.738	Jhe	0.26	

When mutually significant genes with different directions of change are ranked by the magnitude of their difference in log2FoldChange, the top 10 genes agree well across alignment strategy:

Table 69. Top Ten Most Serious Significant Differences between housing and OR47b contrants

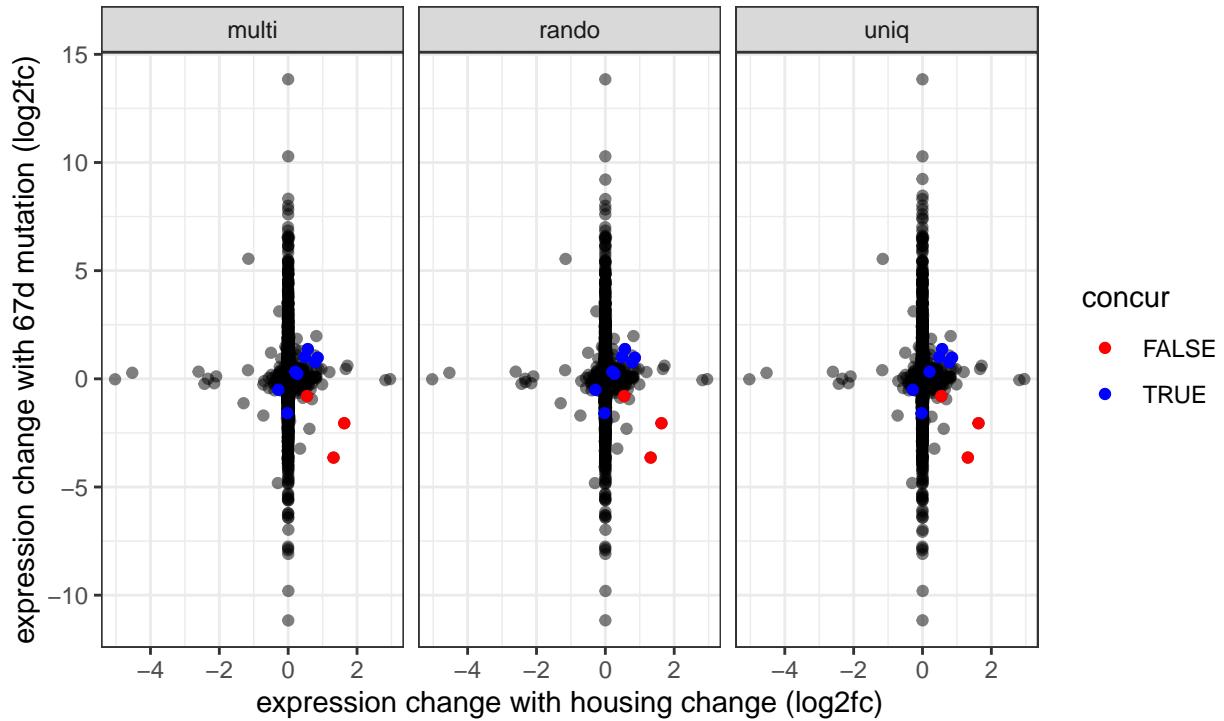
multi					rando		
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression
1	Jhe	4.698	0.26	5.73×10^{-4}	Jhe	4.693	0.26
2	CG11852	4.548	0.07	1.32×10^{-9}	CG11852	4.544	0.07
3	amd	2.615	1.01	1.67×10^{-7}	amd	2.612	1.01
4	Obp84a	1.981	0.62	4.48×10^{-20}	Obp84a	1.980	0.62
5	CG10050	1.678	0.24	2.52×10^{-6}	CG10050	1.676	0.24
6	CG13332	1.257	0.28	1.18×10^{-4}	CG13332	1.254	0.28
7	magu	1.049	0.21	1.16×10^{-5}	magu	1.046	0.21
8	Or92a	0.918	1.57	1.89×10^{-5}	Or92a	0.916	1.57
9	Fer2LCH	0.527	4.18	8.37×10^{-7}	Fer2LCH	0.525	4.18

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast_and_7bContrast.multi.tsv*
results/tables/supp/housingContrast_and_7bContrast.rando.tsv
results/tables/supp/housingContrast_and_7bContrast.uniq.tsv

3.4.2 Housing & 67d

Here is a scatterplot of the log2 fold change of the 67d & wt contrast vs the housing contrast (wt group & wt isolated). The upper right quadrant contains genes which are upregulated in both cases; the lower left contains genes which are downregulated in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

Figure 50. Scatterplot of Expression Changes in 67d mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## pdf
## 2
```

Of the mutually significant genes, slightly more have the same direction of change than not:

Table 70. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs 67d

	multi	rando	uniq
Agree	8	8	7
Disagree	3	3	3

Of those mutually significant genes with the same direction of change, the top 10 most significant agree well across alignment strategy:

Table 71. Top Ten Most Significant Genes of Agr in difference expression between housing and 67d contrants

multi							rand			
rank	name	mean expression	mean readusted p	housing l2fc	67d l2fc	name	mean expression	mean re		
1	CG13659	0.31	1.62×10^{-15}	0.570	1.374	CG13659	0.31	1.4		
2	jv	0.08	7.23×10^{-15}	0.482	1.007	jv	0.08	6.2		
3	CG31288	1.36	2.79×10^{-10}	0.855	0.975	CG31288	1.36	2.5		
4	Fer2LCH	4.18	8.90×10^{-7}	0.205	0.327	Fer2LCH	4.18	9		
5	CG31272	0.11	9.22×10^{-7}	0.779	0.769	CG31272	0.11	9		
6	CG3764	0.21	3.94×10^{-6}	-0.286	-0.501	CG3764	0.21	3		
7	hgo	0.04	7.87×10^{-4}	-0.025	-1.589	hgo	0.04	7		
8	CG32276	2.00	6.84×10^{-3}	0.266	0.232	CG32276	2.00	7		

When mutually significant genes with the same direction of change are ranked by the magnitude of their mean log2FoldChange, the top 10 agree relatively well across alignment strategy:

Table 72. Top Ten Largest Magnitude Changes In Significant Gen in difference expression between housing and 67d contrants

multi					rand				
rank	name	mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean re	
1	CG13659	0.972	0.31	1.62×10^{-15}	CG13659	0.972	0.31	1.4	
2	CG31288	0.915	1.36	2.79×10^{-10}	CG31288	0.916	1.36	2.5	
3	hgo	-0.807	0.04	7.87×10^{-4}	hgo	-0.807	0.04	7	
4	CG31272	0.774	0.11	9.22×10^{-7}	CG31272	0.773	0.11	9	
5	jv	0.744	0.08	7.23×10^{-15}	jv	0.744	0.08	6.2	
6	CG3764	-0.394	0.21	3.94×10^{-6}	CG3764	-0.396	0.21	3	
7	Fer2LCH	0.266	4.18	8.90×10^{-7}	Fer2LCH	0.266	4.18	9	
8	CG32276	0.249	2.00	6.84×10^{-3}	CG32276	0.248	2.00	7	

Of those mutually significant genes with different directions of change, the top 10 most significant agree well across alignment strategy.

Table 73. Top Ten Most Significant Genes of Dis in difference expression between housing and OR47b contrants

multi							rand		
rank	name	mean expression	mean readusted p	housing l2fc	67d l2fc	name	mean expression	mean re	
1	MtnB	0.39	3.92×10^{-32}	1.317	-3.642	MtnB	0.39	3.4	
2	CG11852	0.07	3.95×10^{-7}	1.629	-2.051	CG11852	0.07	3	
3	CG13332	0.28	5.80×10^{-5}	0.547	-0.796	CG13332	0.28	5	

When mutually significant genes with different directions of change are ranked by the magnitude of their difference in log2FoldChange, the top 10 genes agree well across alignment strategy:

Table 74. Top Ten Most Serious Significant Differences between difference expression between housing and 67d contrants

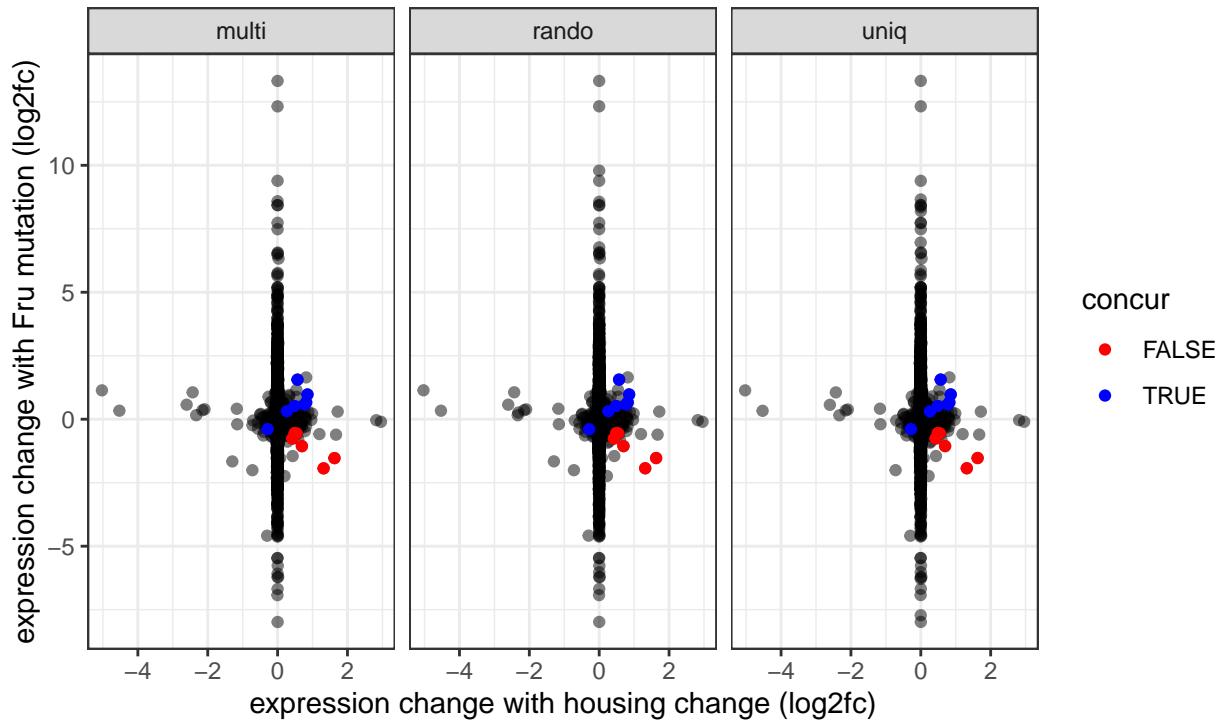
rank	name	multi			rando		
		l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression
1	MtnB	4.959	0.39	3.92×10^{-32}	MtnB	4.957	0.39
2	CG11852	3.680	0.07	3.95×10^{-7}	CG11852	3.681	0.07
3	CG13332	1.343	0.28	5.80×10^{-5}	CG13332	1.342	0.28

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast_and_67dContrast.multi.tsv*, *results/tables/supp/housingContrast_and_67dContrast.rando.tsv*, *results/tables/supp/housingContrast_and_67dContrast.uniq.tsv*.

3.4.3 Housing & FruLexFru440

Here is a scatterplot of the log2 fold change of the Fru & wt contast vs the housing contrast (wt group & wt isolated). The upper right quadrant contains genes which are upregulated in both cases; the lower left contains genes which are downregulated in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

Figure 51. Scatterplot of Expression Changes in Fru mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## pdf
## 2
```

Of the mutually significant genes, about the same number have the same direction of change as not:

Table 75. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs Fru

	multi	rando	uniq
Agree	7	7	7
Disagree	6	6	6

Of those mutually significant genes with the same direction of change, the top 10 most significant agree well across alignment strategy:

Table 76. Top Ten Most Significant Genes of Agree in difference expression between housing and Fru contrants

rank	name	multi				rando			
		mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re	
1	CG13659	0.31	4.24×10^{-19}	0.570	1.558	CG13659	0.31	3.2	
2	CG31288	1.36	3.41×10^{-10}	0.855	0.974	CG31288	1.36	2.9	
3	jv	0.08	1.50×10^{-5}	0.482	0.516	jv	0.08	1.	
4	CG31272	0.11	3.05×10^{-5}	0.779	0.580	CG31272	0.11	2.	
5	CG3764	0.21	2.45×10^{-4}	-0.286	-0.393	CG3764	0.21	2.	
6	CG42806	0.64	2.83×10^{-4}	0.818	0.661	CG42806	0.64	2.	
7	CG32276	2.00	6.74×10^{-4}	0.266	0.313	CG32276	2.00	7.	

When mutually significant genes with the same direction of change are ranked by the magnitude of their mean log2FoldChange, the top 10 agree well across alignment strategy.

Table 77. Top Ten Largest Magnitude Changes In Significant Genes in difference expression between housing and Fru contrants

rank	name	multi			rando			
		mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean re
1	CG13659	1.064	0.31	4.24×10^{-19}	CG13659	1.065	0.31	3.2
2	CG31288	0.915	1.36	3.41×10^{-10}	CG31288	0.916	1.36	2.9
3	CG42806	0.739	0.64	2.83×10^{-4}	CG42806	0.740	0.64	2.
4	CG31272	0.679	0.11	3.05×10^{-5}	CG31272	0.679	0.11	2.
5	jv	0.499	0.08	1.50×10^{-5}	jv	0.500	0.08	1.
6	CG3764	-0.340	0.21	2.45×10^{-4}	CG3764	-0.340	0.21	2.
7	CG32276	0.290	2.00	6.74×10^{-4}	CG32276	0.289	2.00	7.

Of those mutually significant genes with different directions of change, the top 10 most significant agree well across alignment strategy.

Table 78. Top Ten Most Significant Genes of Disagree in difference expression between housing and Fru contrants

rank	name	multi				rando			
		mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re	
1	MtnB	0.39	3.91×10^{-20}	1.317	-1.934	MtnB	0.39	3.0	

2	Or92a	1.57	6.23×10^{-9}	0.420	-0.763	Or92a	1.57	6.23
3	CG10050	0.24	1.64×10^{-6}	0.697	-1.061	CG10050	0.24	1.64
4	CG11852	0.07	9.90×10^{-6}	1.629	-1.531	CG11852	0.07	9.90
5	T48	0.21	4.65×10^{-4}	0.488	-0.543	T48	0.21	4.65
6	CG13332	0.28	1.93×10^{-3}	0.547	-0.561	CG13332	0.28	1.93

When mutually significant genes with different directions of change are ranked by the magnitude of their difference in log2FoldChange, the top 10 genes agree well across alignment strategy.

Table 79. Top Ten Most Serious Differences between housing and Fru contrasts

rank	name	multi			rando		
		l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression
1	MtnB	3.252	0.39	3.91×10^{-20}	MtnB	3.248	0.39
2	CG11852	3.160	0.07	9.90×10^{-6}	CG11852	3.159	0.07
3	CG10050	1.758	0.24	1.64×10^{-6}	CG10050	1.755	0.24
4	Or92a	1.184	1.57	6.23×10^{-9}	Or92a	1.181	1.57
5	CG13332	1.108	0.28	1.93×10^{-3}	CG13332	1.105	0.28
6	T48	1.031	0.21	4.65×10^{-4}	T48	1.029	0.21

Full data are in the tables folder:

results/tables/supp/housingContrast_andFruContrast.multi.tsv *results/tables/supp/housingContrast_andFruContrast.rando.tsv*
results/tables/supp/housingContrast_andFruContrast.uniq.tsv

3.5 Comparing Expression Changes Between Mutants

do this

3.5.1 Fru & 67d

do this

3.5.2 Fru & 47b

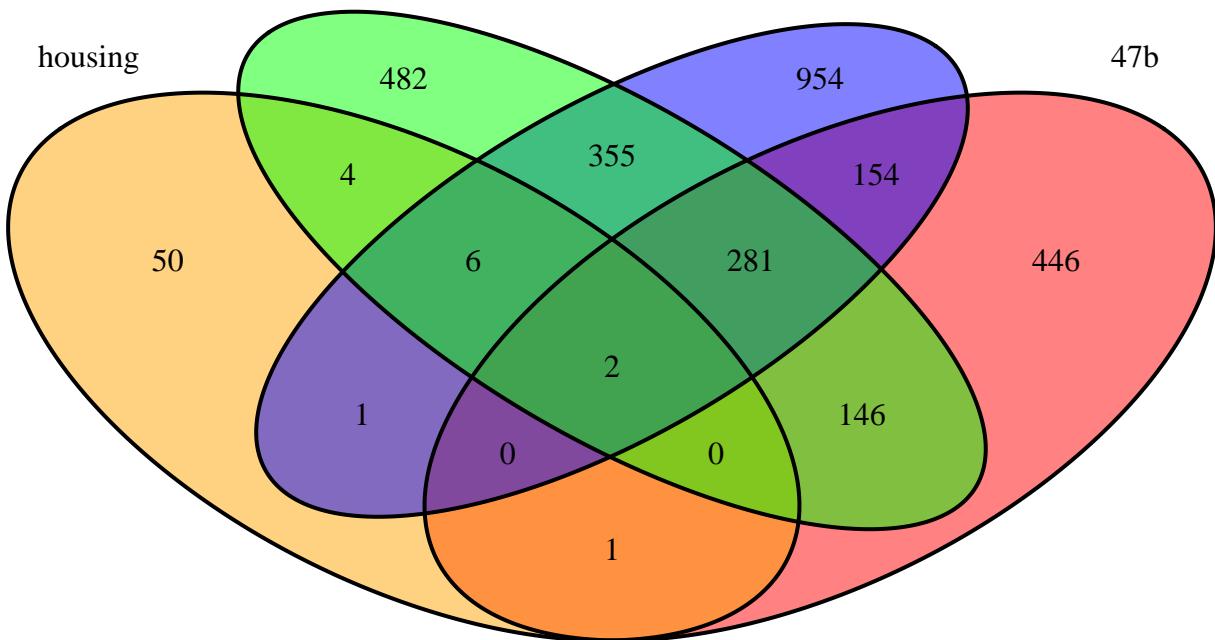
do this

3.5.3 47b & 67d

do this

3.6 Mutually Significant Differential Expression Overview

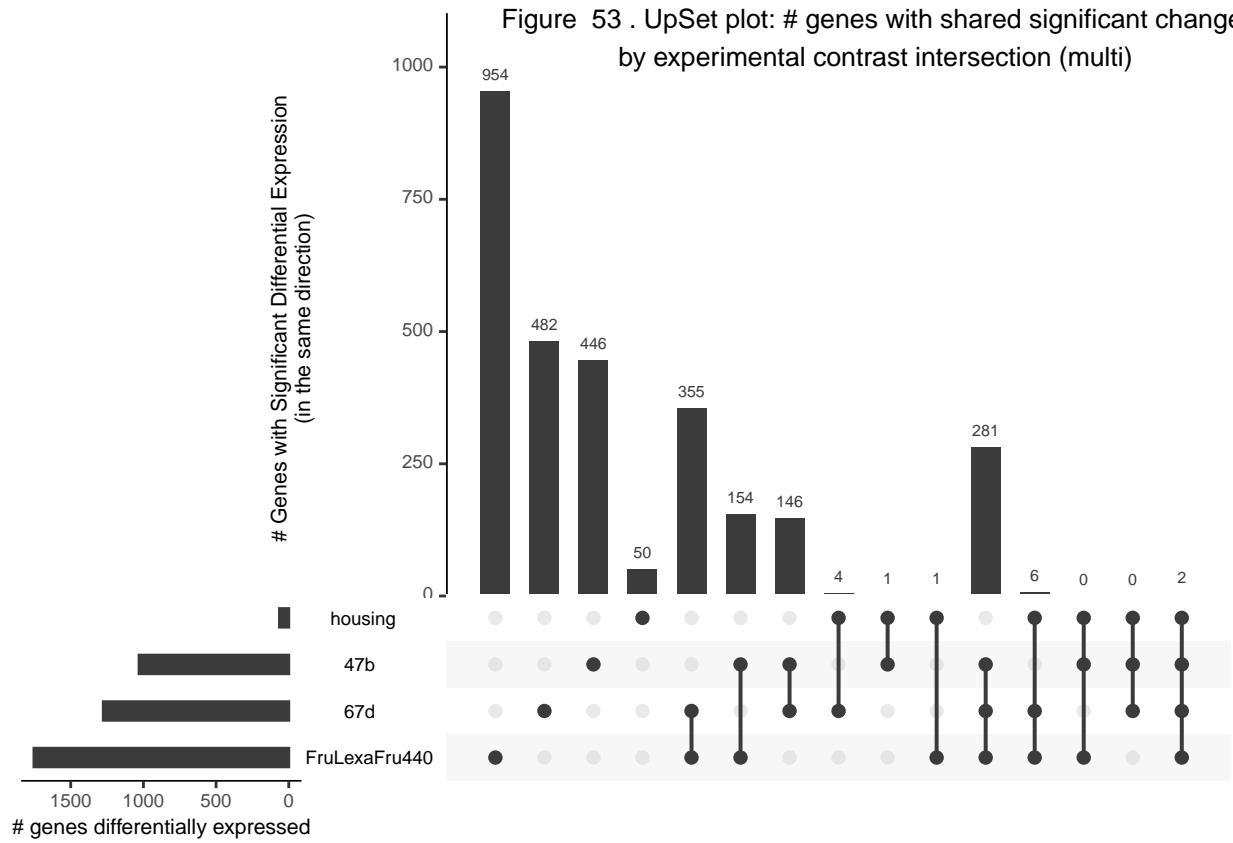
Figure 52 . Venn Diagram: # genes with shared significant change,
by experimental contrast intersection (multi)
67d FruLexaFru440



```
## null device
##      1

## null device
##      1
```

Figure 53 . UpSet plot: # genes with shared significant change, by experimental contrast intersection (multi)



```
## null device
##      1
```

```
## null device
##      1
```

The two genes with the same behavior across all experimental contrasts are javelin and CG13659. Both are upregulated in all cases:

Table 79. Genes sharing significant differential expression in all four contrasts

	47b vs wt		67d vs wt		FruLexaFru440 vs wt		log2FoldChange
	log2FoldChange	padj	log2FoldChange	padj	log2FoldChange	padj	
multi							
CG13659	0.453	1.96×10^{-3}	1.374	2.24×10^{-27}	1.558	1.54×10^{-34}	
jv	1.219	1.05×10^{-37}	1.007	2.79×10^{-25}	0.516	1.20×10^{-6}	
rando							
CG13659	0.453	1.89×10^{-3}	1.374	1.65×10^{-27}	1.560	8.96×10^{-35}	
jv	1.220	5.97×10^{-38}	1.007	2.14×10^{-25}	0.518	1.05×10^{-6}	
uniq							
CG13659	0.452	1.88×10^{-3}	1.375	1.27×10^{-27}	1.561	6.04×10^{-35}	
jv	1.220	4.92×10^{-38}	1.007	1.73×10^{-25}	0.518	9.78×10^{-7}	

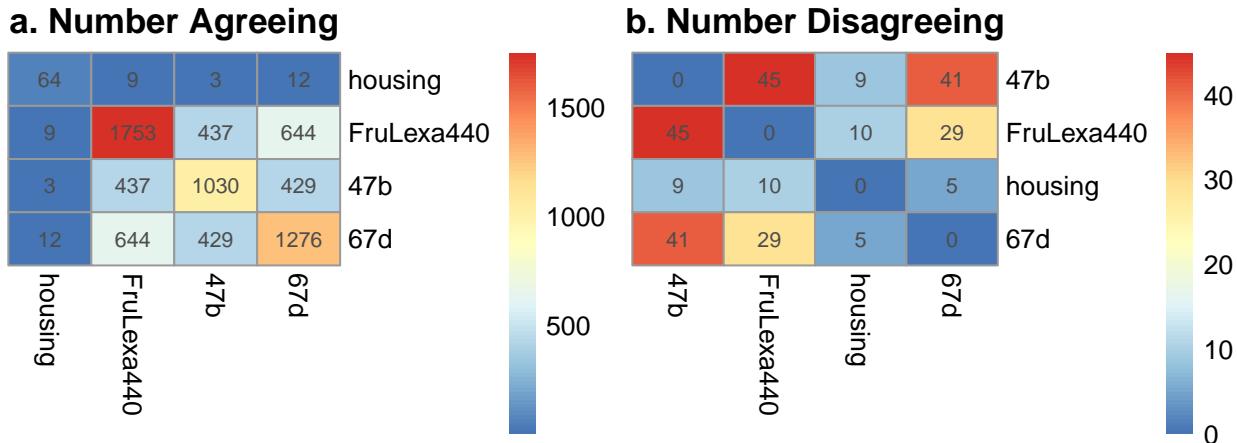


Figure 54 . Heatmap of Pairwise Comparisons between Contrasts:
significant genes with the same (left) or different (right) directions of change

```
## null device
##           1
```

```
## null device
##           1
```

results shown are for multi only; very similar across aligner strategies

3.7 Focus on Fruitless

The behavior of fruitless is of special interest, and feature counting/differential expression testing was performed on an annotation which considers all available exons separately.

3.7.1 By Exon

3.7.1.1 Ambiguous Read Assignment: None

The default featureCounts settings ignore ambiguously assigned reads. Because some exons overlap and because junction-spanning reads will be considered ambiguous in this context, some relevant reads might be being ignored and deflating the power in these tests. Several exons were filtered out entirely based on low read count number. Here are the results from this assignment strategy.

Table 80. Number of Fruitless Exons Available For Analysis
(by aligner)

<th>count</th> <th>frac</th> <th>total</th>	count	frac	total
multi	13	59.1%	22
rando	13	59.1%	22
uniq	13	59.1%	22

The only exons with even marginally significant differential expression in any contrast are 1,3, and 5, in the FruLexa/Fru400 contrast:

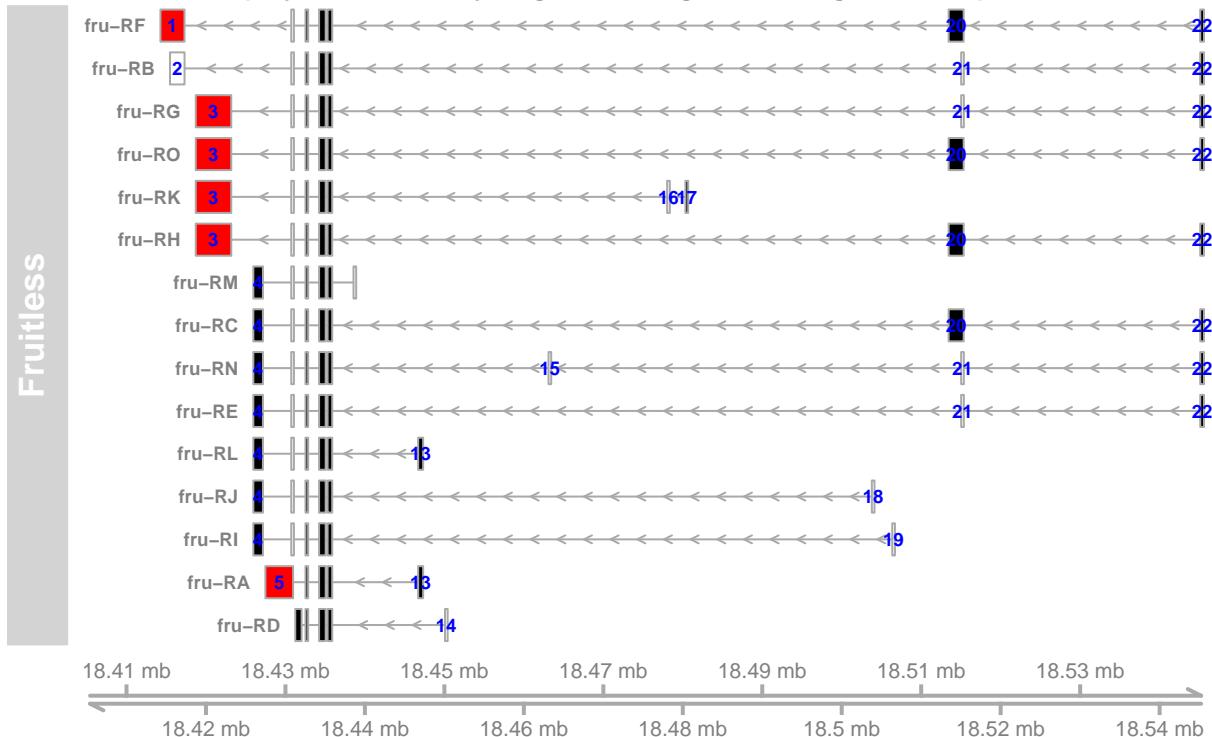
Table 81. Differential Expression in Fruitless exons, by Contrast
(Multi only)

	47b		67d		Fru		wt %
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	
exon_1	0.0000	9.46×10^{-1}	0.0000	8.84×10^{-1}	1.4767	2.43×10^{-2}	0.0
exon_10	0.0000	7.26×10^{-1}	-0.0000	8.84×10^{-1}	-0.0770	9.07×10^{-1}	0.0
exon_11	-0.0000	9.46×10^{-1}	-0.0000	8.84×10^{-1}	-0.0565	9.07×10^{-1}	0.0
exon_13	-0.0000	9.46×10^{-1}	0.0000	8.84×10^{-1}	0.6672	8.26×10^{-2}	0.0
exon_17	-0.0000	4.55×10^{-1}	-0.0000	8.84×10^{-1}	0.9129	8.26×10^{-2}	-0.0
exon_20	-0.0000	9.46×10^{-1}	-0.0000	9.98×10^{-1}	0.1063	9.07×10^{-1}	0.0
exon_22	-0.0000	4.55×10^{-1}	-0.0000	8.84×10^{-1}	-0.0974	9.07×10^{-1}	0.0
exon_3	0.0000	9.46×10^{-1}	0.0000	8.84×10^{-1}	0.9637	3.08×10^{-2}	-0.0
exon_4	-0.0000	9.46×10^{-1}	-0.0000	8.84×10^{-1}	-0.0664	9.07×10^{-1}	0.0
exon_5	0.0000	4.55×10^{-1}	0.0000	4.27×10^{-1}	2.0340	1.91×10^{-3}	-0.0
exon_7	-0.0000	9.46×10^{-1}	-0.0000	8.84×10^{-1}	0.0324	9.07×10^{-1}	-0.0
exon_8	0.0000	9.46×10^{-1}	-0.0000	8.84×10^{-1}	-0.0486	9.07×10^{-1}	0.0
exon_9	-0.0000	9.91×10^{-1}	-0.0000	8.84×10^{-1}	0.0427	9.07×10^{-1}	-0.0

Table 82. Fru exons with significant ($\text{padj} < 0.05$) differential expression
(by aligner)

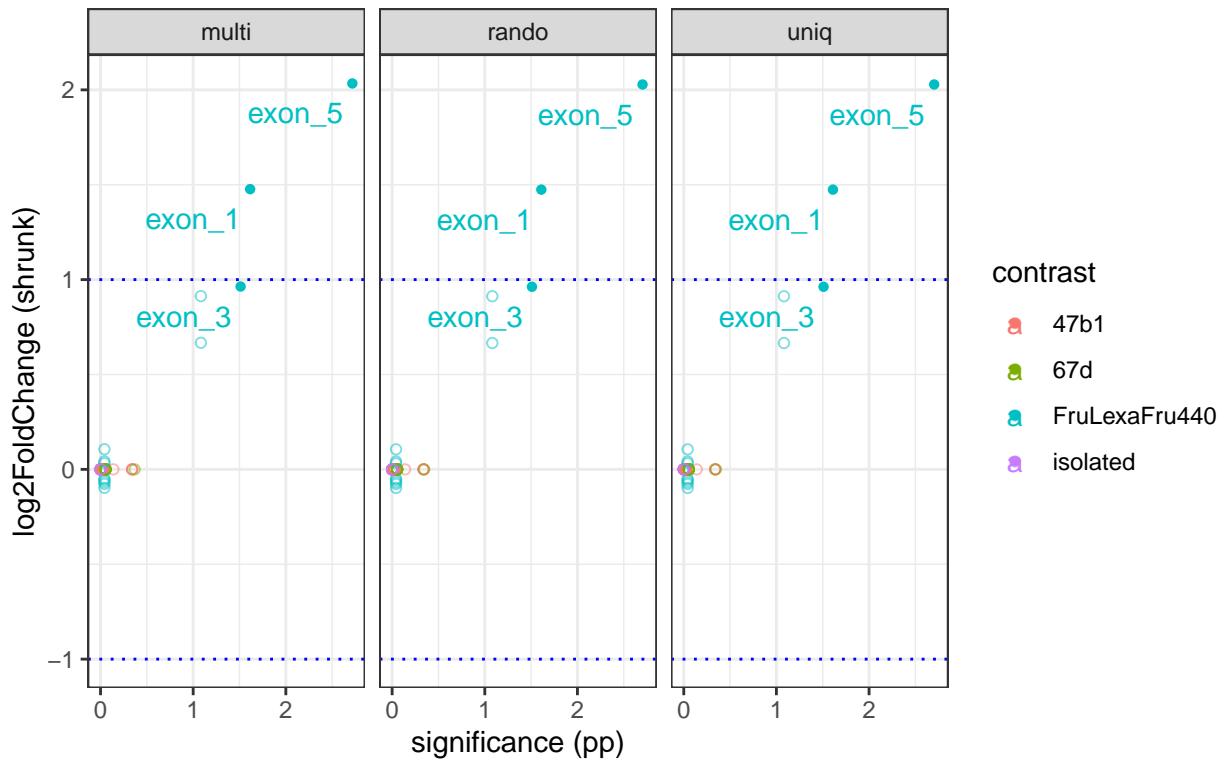
Fru		
<th>log2FoldChange</th> <th>adjusted p</th>	log2FoldChange	adjusted p
exon_1		
multi	1.48	0.024
rando	1.47	0.025
uniq	1.47	0.025
exon_3		
multi	0.96	0.031
rando	0.96	0.031
uniq	0.96	0.031
exon_5		
multi	2.03	0.002
rando	2.03	0.002
uniq	2.03	0.002

**Figure 55. Fruitless gene model: exons with any significant change detected highlighted
(any contrast, any aligner, ambiguous assigned to all)**



```
## pdf
## 2
```

Figure 56. Volcano Plot: Fold Change vs. Significance
(fruitless exons, 'none' counting strategy)



```
## pdf
## 2
```

3.7.1.2 Ambiguous Read Assignment: All

Here, ambiguous reads have been assigned to every feature they overlap, rather than none.

Table 83. Number of Fru Exons Available For Analysis
(by aligner)

aligner	count	frac	total
multi	17	77.3%	22
rando	17	77.3%	22
uniq	17	77.3%	22

The only exons 3 and 5 had even marginally significant differential expression, and only in the FruLexa/Fru440 contrast:

Table 84. Differential Expression in Fru exons, by Contrast
(Multi only)

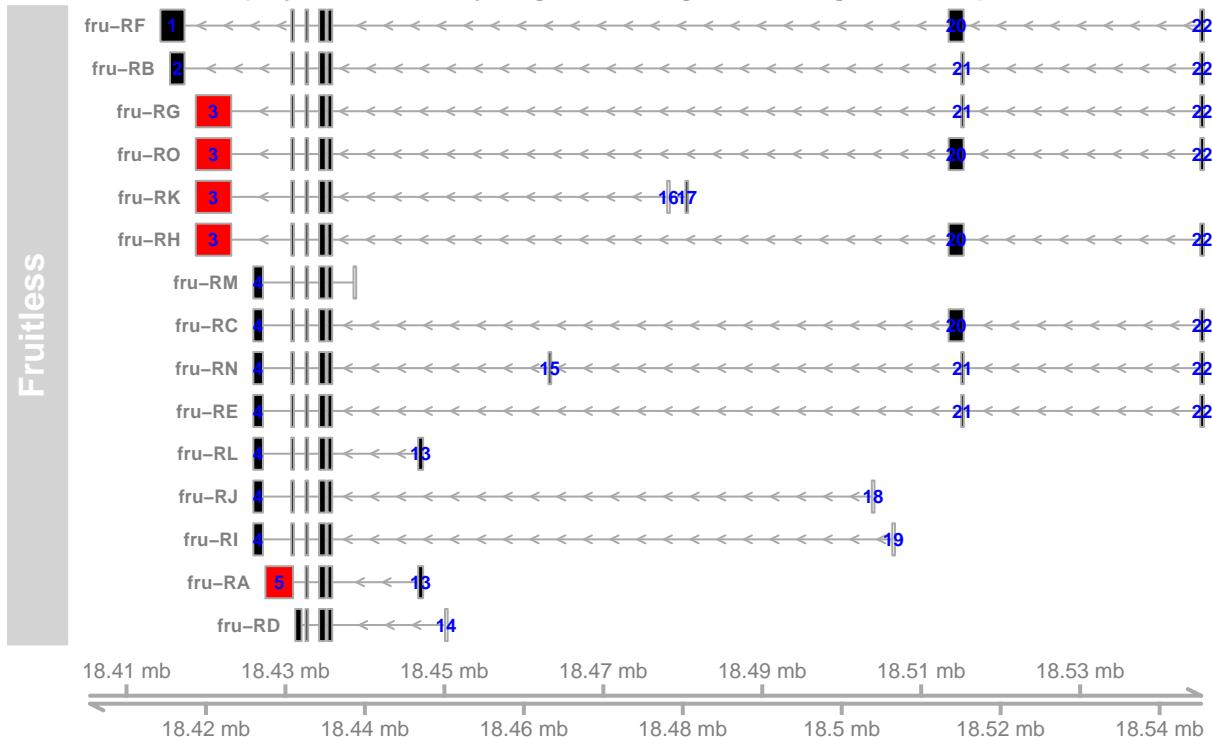
	47b		67d		Fru		wt isol
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange
exon_1	-6.76×10^{-9}	0.99	4.94×10^{-4}	0.99	4.76×10^{-1}	0.17	2.32×10^{-7}
exon_10	9.97×10^{-6}	0.77	-9.14×10^{-3}	0.94	-1.03×10^{-1}	0.73	-8.40×10^{-7}

exon_11	-2.50×10^{-6}	0.99	2.14×10^{-4}	0.99	3.30×10^{-2}	0.91	5.18×10^{-6}
exon_13	-5.39×10^{-7}	0.99	2.10×10^{-3}	0.94	4.14×10^{-1}	0.17	6.25×10^{-7}
exon_15	-1.19×10^{-7}	0.99	-1.39×10^{-4}	0.99	1.31×10^{-2}	0.73	6.69×10^{-10}
exon_17	-1.28×10^{-7}	0.99	-8.65×10^{-5}	0.99	1.05×10^{-2}	0.73	-4.39×10^{-8}
exon_2	-4.39×10^{-7}	0.99	3.83×10^{-4}	0.99	3.67×10^{-1}	0.17	-3.27×10^{-8}
exon_20	-2.11×10^{-7}	0.99	-7.47×10^{-5}	0.99	-8.90×10^{-2}	0.39	9.63×10^{-8}
exon_21	-3.72×10^{-9}	0.99	7.01×10^{-6}	0.99	-9.73×10^{-3}	0.73	5.22×10^{-7}
exon_22	-4.69×10^{-6}	0.86	-3.57×10^{-3}	0.94	-1.98×10^{-1}	0.54	3.70×10^{-6}
exon_3	1.12×10^{-6}	0.99	1.91×10^{-3}	0.94	9.36×10^{-1}	0.03	-2.40×10^{-7}
exon_4	-1.02×10^{-6}	0.99	-3.17×10^{-3}	0.94	1.32×10^{-3}	0.99	1.91×10^{-6}
exon_5	2.84×10^{-6}	0.77	2.82×10^{-3}	0.94	1.23	0.03	2.15×10^{-7}
exon_6	5.54×10^{-6}	0.77	-2.31×10^{-3}	0.94	-7.75×10^{-2}	0.73	2.03×10^{-6}
exon_7	1.43×10^{-7}	0.99	-4.45×10^{-6}	0.99	1.07×10^{-4}	0.99	3.33×10^{-8}
exon_8	7.41×10^{-6}	0.86	-6.53×10^{-3}	0.94	-8.09×10^{-2}	0.73	3.76×10^{-6}
exon_9	-3.40×10^{-6}	0.99	-1.23×10^{-2}	0.94	-6.29×10^{-2}	0.73	-1.33×10^{-5}

Table 85. Fru exons with significant ($\text{padj} < 0.05$) differential expression
(by aligner)

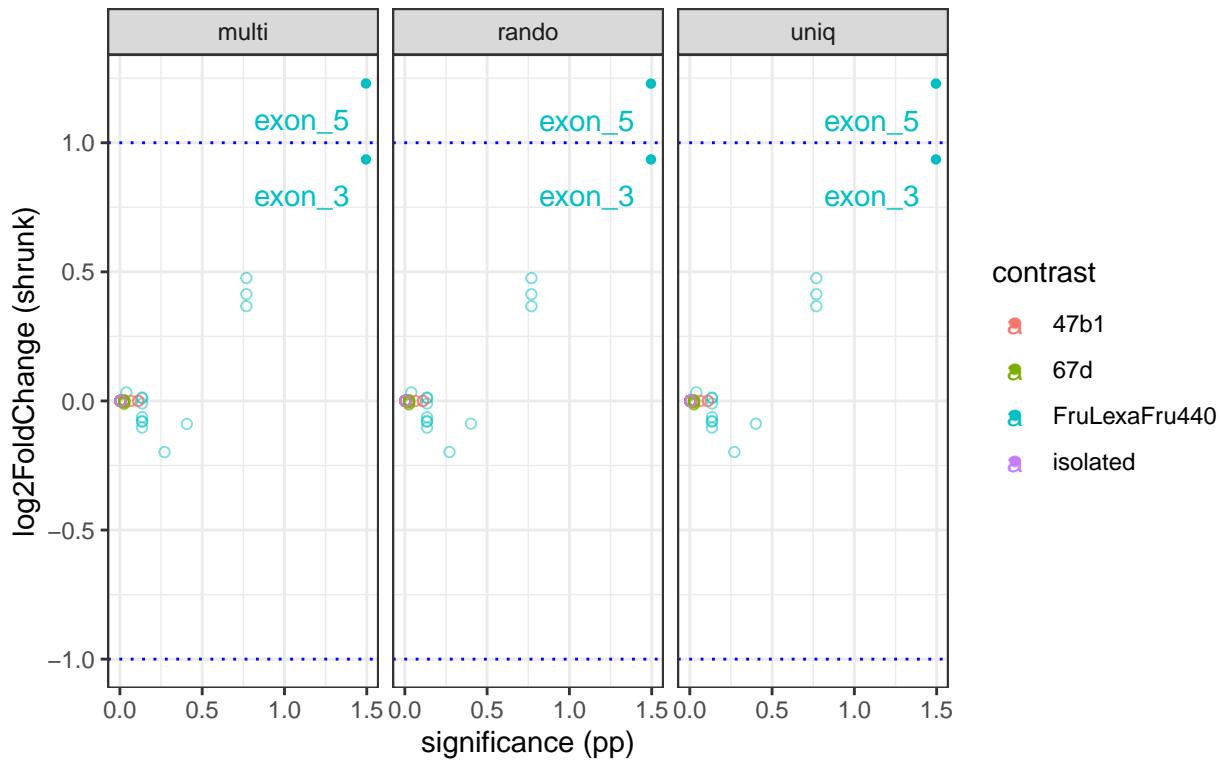
aligner	Fru		
	log2FoldChange	adjusted p	
exon_3			
multi	0.94	0.032	
rando	0.94	0.032	
uniq	0.94	0.032	
exon_5			
multi	1.23	0.032	
rando	1.23	0.032	
uniq	1.23	0.032	

**Figure 57. Fruitless gene model: exons with any significant change detected highlighted
(any contrast, any aligner, ambiguous assigned to all)**



```
## pdf
## 2
```

Figure 58. Volcano Plot: Fold Change vs. Significance
 (fruitless exons, 'all' counting strategy, $p_{adj} < 0.05$)



```
## pdf
## 2
```

3.7.2 By Exon Junction

When the *_SplicedOnly alignments were counted against the fru_junct annotation:

Table 86. Number of Fru Exons Available For Analysis
 (spliced reads counted by splice junction)

	count	fraction
multi	15	68.2%
rando	15	68.2%
uniq	15	68.2%

Table 87. Differential Expression in Fru exons, by Contrast
 (Multi only)

	47b		67d		Fru		wt
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	
exon_1	-0.01	7.55×10^{-1}	-0.00	6.35×10^{-1}	-0.00	7.99×10^{-1}	0
exon_10	0.01	8.44×10^{-1}	0.00	9.99×10^{-1}	0.00	9.88×10^{-1}	0
exon_11	-0.04	6.57×10^{-1}	0.00	9.99×10^{-1}	-0.00	9.88×10^{-1}	0
exon_13	-0.04	5.67×10^{-1}	0.00	9.99×10^{-1}	0.00	9.88×10^{-1}	0

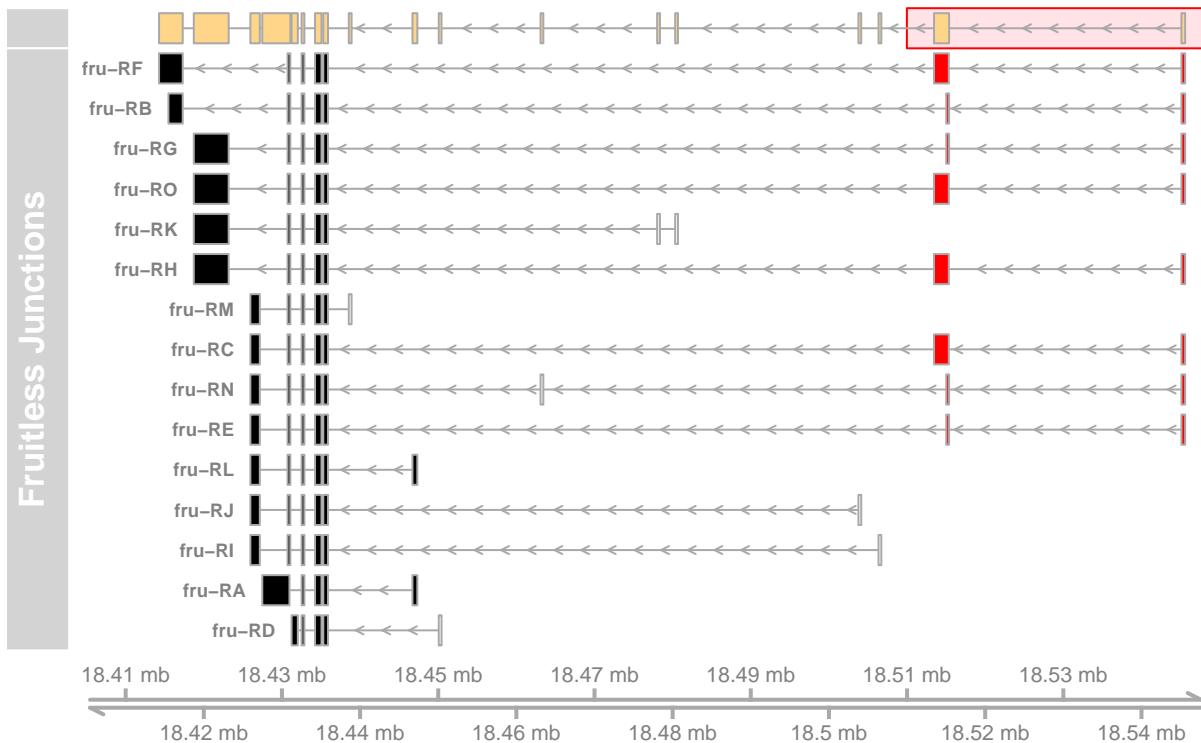
exon_2	-0.01	7.55×10^{-1}	-0.00	6.35×10^{-1}	-0.00	7.99×10^{-1}	0
exon_20	-0.01	7.55×10^{-1}	0.00	9.99×10^{-1}	-0.00	1.83×10^{-2}	0
exon_21	-0.04	5.67×10^{-1}	0.00	9.99×10^{-1}	-2.95	1.83×10^{-3}	-0
exon_22	-0.01	7.55×10^{-1}	0.00	9.99×10^{-1}	-0.00	1.83×10^{-2}	0
exon_3	0.01	7.55×10^{-1}	0.00	9.50×10^{-1}	0.00	9.88×10^{-1}	0
exon_4	-0.02	7.02×10^{-1}	-0.00	9.99×10^{-1}	-0.00	9.88×10^{-1}	-0
exon_5	0.77	8.73×10^{-2}	0.00	8.09×10^{-1}	0.00	7.99×10^{-1}	0
exon_6	0.02	7.55×10^{-1}	-0.00	9.99×10^{-1}	-0.00	9.88×10^{-1}	-0
exon_7	0.01	5.67×10^{-1}	0.00	6.35×10^{-1}	0.00	5.66×10^{-1}	0
exon_8	0.02	7.55×10^{-1}	0.00	9.99×10^{-1}	0.00	9.88×10^{-1}	-0
exon_9	-0.02	7.55×10^{-1}	-0.00	9.99×10^{-1}	-0.00	7.99×10^{-1}	-0

Exons 20,21, and 22, the most 5' of exons, are downregulated in the FruLexa/Fru440 contrast; however, exons 20 and 22 have bizarrely low effect sizes given their significance:

Table 88. Fru exons with significant ($\text{padj} < 0.05$) differential expression
(by aligner)

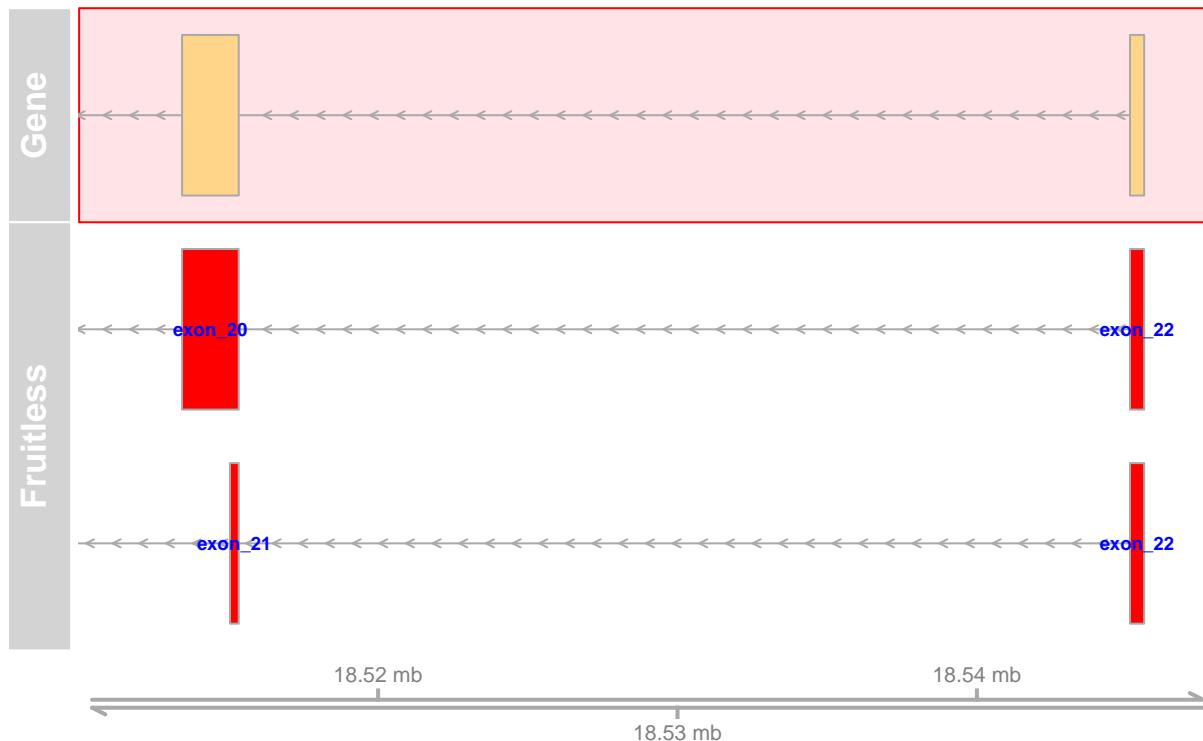
aligner	Fru	
	log2FoldChange	adjusted p
exon_20		
multi	-0.00	0.018
rando	-0.00	0.018
uniq	-0.00	0.018
exon_21		
multi	-2.95	0.002
rando	-2.95	0.002
uniq	-2.95	0.002
exon_22		
multi	-0.00	0.018
rando	-0.00	0.018
uniq	-0.00	0.018

Figure 59 Fruitless exons with significant change (measured by junction)



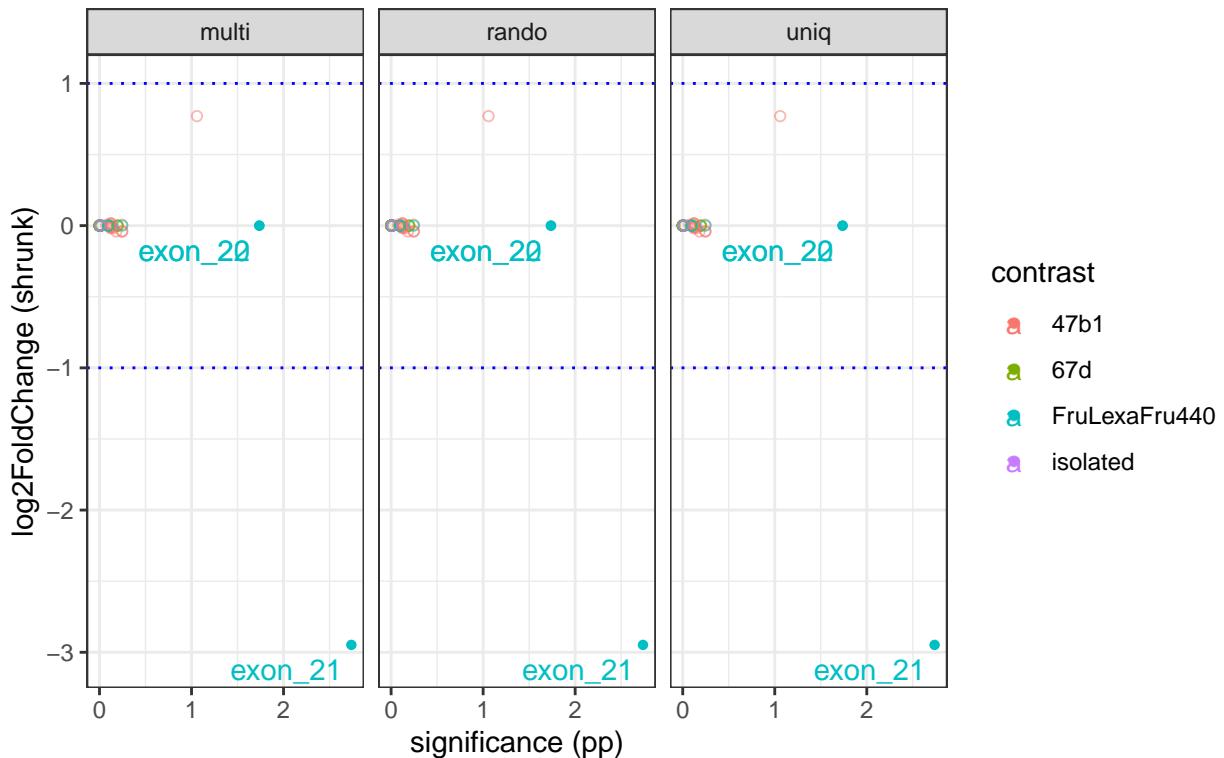
```
## pdf
## 2
```

Figure 60 a. Fruitless exons with significant change (measured by junction) (detail)



```
## pdf  
## 2
```

Figure 61. Volcano Plot: Fold Change vs. Significance
(fruitless exons, 'all' counting strategy, $p_{adj} < 0.05$)



```
## pdf
## 2
```

3.7.3 By Intron

When the *_SplicedOnly alignments were counted against the fru_introns annotation:

Table 89. Number of Fru Introns Available For Analysis
(spliced reads counted by intron boundaries)

	count	fraction
multi	19	86.4%
rando	19	86.4%
uniq	19	86.4%

Table 90. Differential Expression in Fru introns, by Contrast
(Multi only)

	47b		67d		Fru		w
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	
intron_1	3.02×10^{-1}	5.13×10^{-1}	-2.00×10^{-6}	9.87×10^{-1}	-1.02×10^{-6}	9.84×10^{-1}	$-5.13 \times$
intron_10	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.39 \times$
intron_11	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.39 \times$

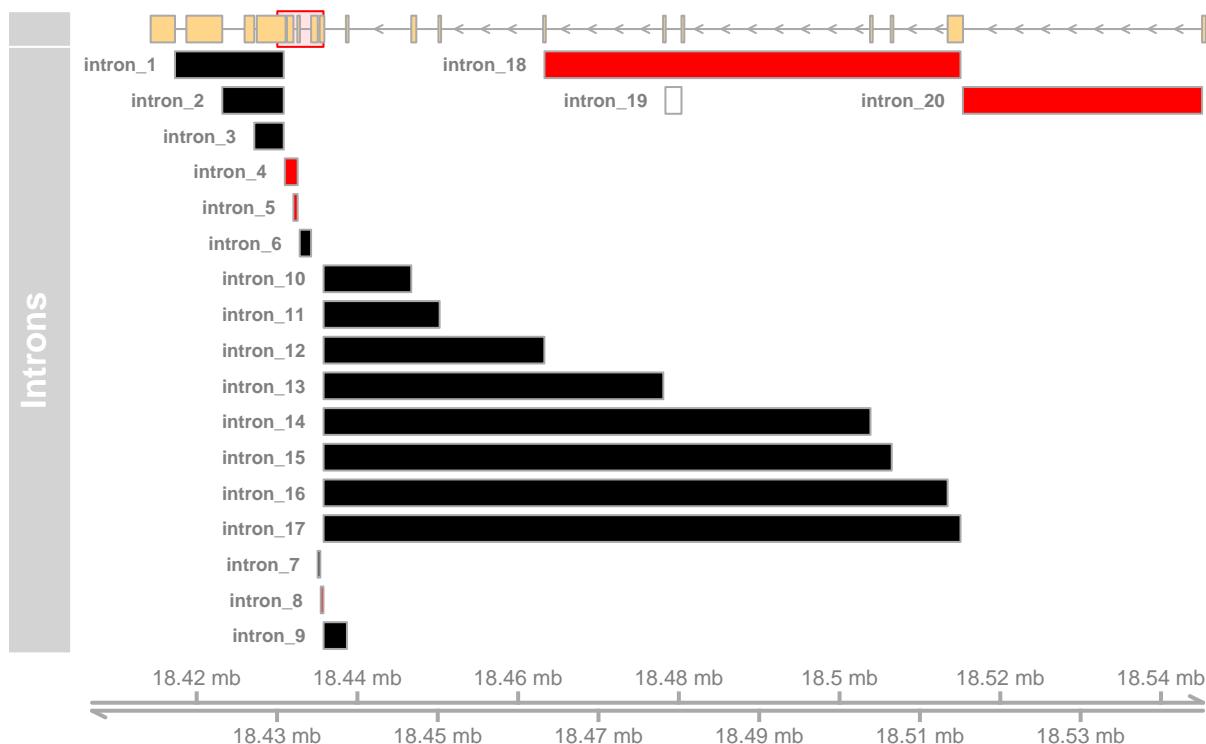
intron_12	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.39 \times$
intron_13	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.39 \times$
intron_14	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.39 \times$
intron_15	-1.87×10^{-2}	9.29×10^{-1}	1.19×10^{-6}	9.87×10^{-1}	5.38×10^{-7}	9.84×10^{-1}	$2.37 \times$
intron_16	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.39 \times$
intron_17	-3.06×10^{-2}	9.29×10^{-1}	-3.12×10^{-7}	9.87×10^{-1}	-5.53×10^{-7}	9.84×10^{-1}	$-4.23 \times$
intron_18	-2.90×10^{-1}	5.13×10^{-1}	-3.38×10^{-7}	9.87×10^{-1}	-6.70×10^{-6}	3.18×10^{-3}	$-2.07 \times$
intron_2	2.77×10^{-1}	5.13×10^{-1}	-2.01×10^{-6}	9.87×10^{-1}	-1.02×10^{-6}	9.84×10^{-1}	$-5.17 \times$
intron_20	3.06×10^{-1}	5.13×10^{-1}	-7.36×10^{-8}	9.87×10^{-1}	-6.64×10^{-6}	3.43×10^{-2}	$-4.79 \times$
intron_3	2.77×10^{-1}	5.13×10^{-1}	-2.01×10^{-6}	9.87×10^{-1}	-1.02×10^{-6}	9.84×10^{-1}	$-5.17 \times$
intron_4	2.09	1.89×10^{-6}	4.12×10^{-6}	5.51×10^{-1}	4.09×10^{-6}	1.76×10^{-1}	$2.49 \times$
intron_5	2.09	1.89×10^{-6}	4.12×10^{-6}	5.51×10^{-1}	4.09×10^{-6}	1.76×10^{-1}	$2.49 \times$
intron_6	3.35×10^{-1}	4.19×10^{-1}	-4.58×10^{-6}	9.87×10^{-1}	-1.47×10^{-6}	9.84×10^{-1}	$-7.59 \times$
intron_7	5.10×10^{-1}	3.41×10^{-1}	-1.85×10^{-7}	9.87×10^{-1}	-1.00×10^{-6}	9.84×10^{-1}	$-4.39 \times$
intron_8	6.51×10^{-1}	2.02×10^{-2}	1.40×10^{-7}	9.87×10^{-1}	3.73×10^{-6}	9.84×10^{-1}	$7.74 \times$
intron_9	-1.94×10^{-2}	9.29×10^{-1}	6.87×10^{-8}	9.87×10^{-1}	7.81×10^{-8}	9.84×10^{-1}	$2.38 \times$

Introns 18 and 20 come up significant in the FruLexa/Fru440 contrast, though they have bizarrely small effect sizes. Introns 4,5, and 8 come up significant in the 47b contrast.

Table 91. Fru introns with significant ($\text{padj} < 0.05$) differential expression
(by aligner)

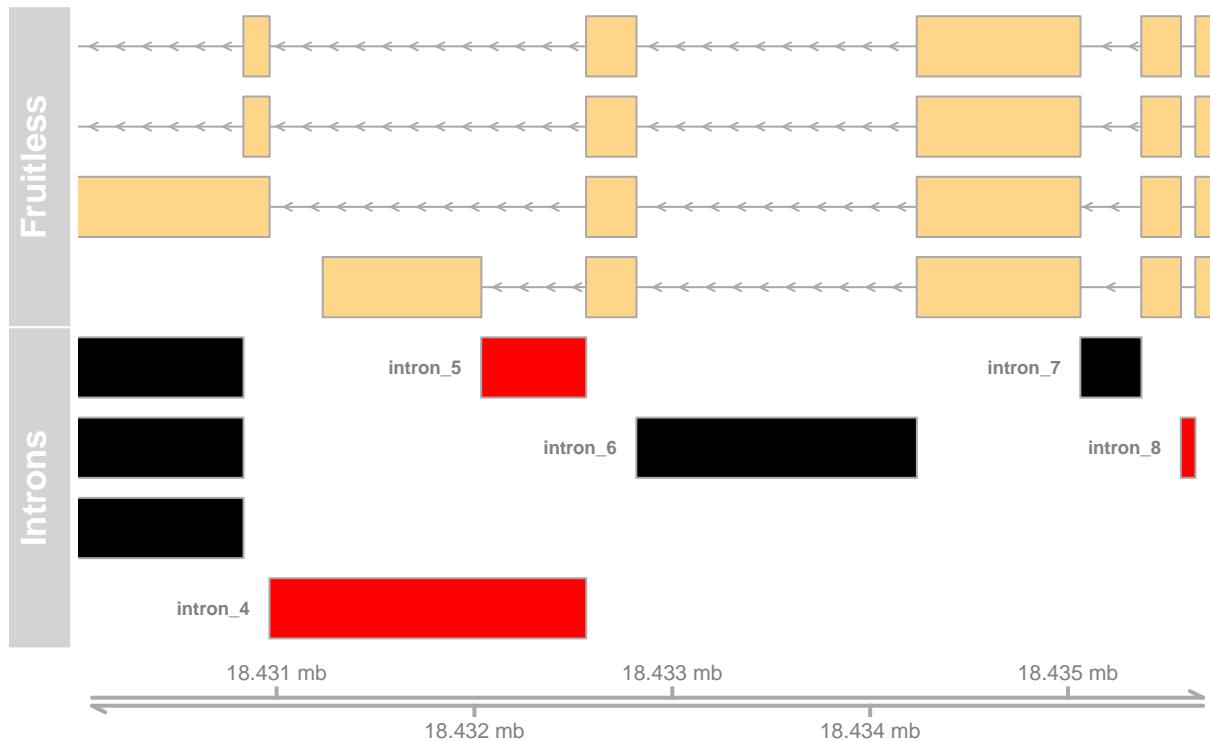
		log2FoldChange	adjusted p
intron_18 - FruLexaFru440			
multi		-6.70×10^{-6}	3.177×10^{-3}
rando		-6.70×10^{-6}	3.177×10^{-3}
uniq		-6.70×10^{-6}	3.177×10^{-3}
intron_20 - FruLexaFru440			
multi		-6.64×10^{-6}	3.429×10^{-2}
rando		-6.64×10^{-6}	3.429×10^{-2}
uniq		-6.64×10^{-6}	3.429×10^{-2}
intron_4 - 47b1			
multi		2.09	1.887×10^{-6}
rando		2.09	1.887×10^{-6}
uniq		2.09	1.887×10^{-6}
intron_5 - 47b1			
multi		2.09	1.887×10^{-6}
rando		2.09	1.887×10^{-6}
uniq		2.09	1.887×10^{-6}
intron_8 - 47b1			
multi		6.51×10^{-1}	2.023×10^{-2}
rando		6.51×10^{-1}	2.023×10^{-2}
uniq		6.51×10^{-1}	2.023×10^{-2}

Figure 62 Fruitless introns with significant change



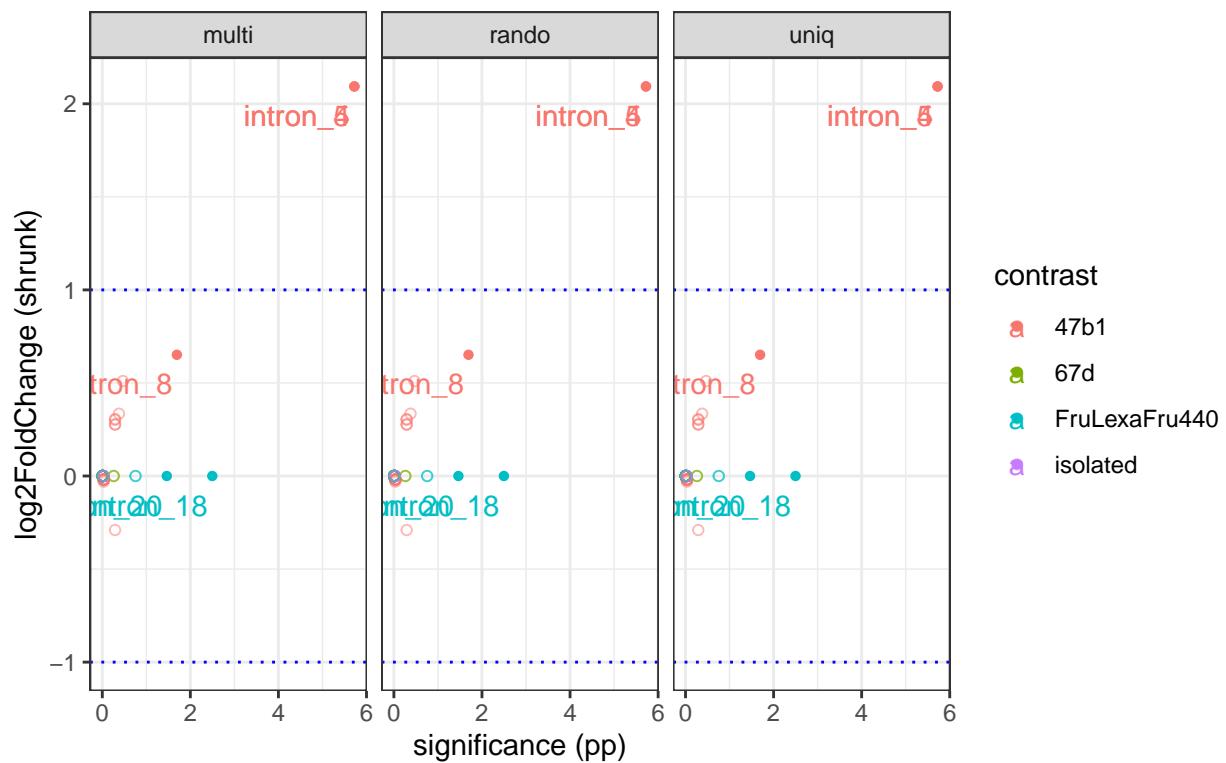
```
## pdf  
## 2
```

Figure 63 a. Fruitless introns with significant change (detail)



```
## pdf  
## 2
```

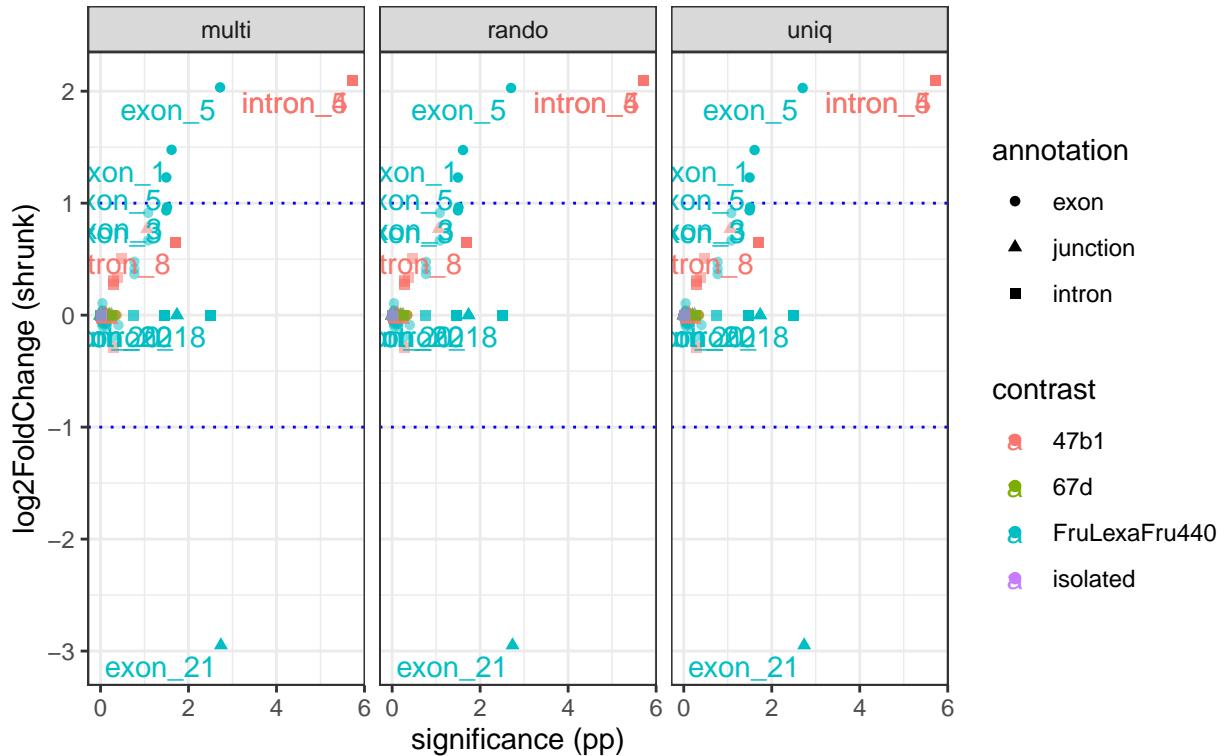
Figure 64. Volcano Plot: Fold Change vs. Significance
 (fruitless introns, 'all' counting strategy, $p_{adj} < 0.05$)



```
## pdf
## 2
```

3.7.4 Overall

Figure 65. Volcano Plot: Fold Change vs. Significance
(all fruitless breakdowns, 'all' counting strategy, $p_{adj} < 0.05$)



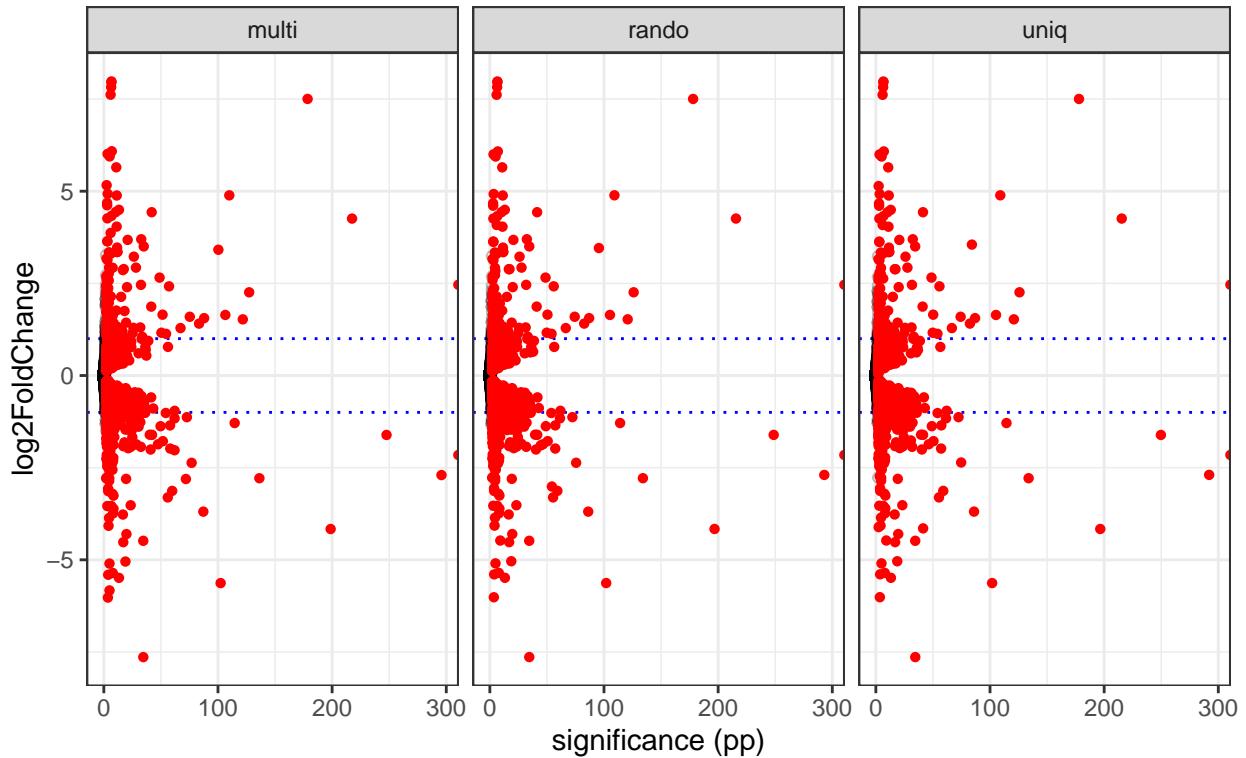
```
## pdf
## 2
```

3.8 Impact of 47b2 on 88a

full results: results/tables/supp/47b_on_88a.de.tsv

Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($p_{adj}<0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of ± 1 (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

**Figure 66. Volcano Plot: Fold Change vs. Significance
(88a mutants, with/without 47b2)**



```
## pdf
## 2
```

Of the 10617 genes with significance scores available, 1385 have an adjusted $p < 0.01$ (13.0451163 %)

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $\text{padj} < 0.01$) changes. There were 322 such genes, mostly shared across alignment strategy:

results/tables/tbl92_47b2on88a_chonky.html

3.8.1 Top Tens

Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes:

Table 92. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed 88a mutants, with/without 47b2

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	Drip	1.45	-2.156	0.00	Drip	1.45	-2.155	0.00
2	Cyp6a2	3.91	2.463	0.00	Cyp6a2	3.91	2.463	0.00
3	Cyp4ac2	1.28	-2.700	1.63×10^{-296}	Cyp4ac2	1.27	-2.697	1.01×10^{-296}
4	Cyp6a20	4.64	-1.612	2.23×10^{-248}	Cyp6a20	4.64	-1.611	2.19×10^{-248}

5	CG10936	0.11	4.257	3.53×10^{-218}	CG10936	0.11	4.258	$2.46 \times 10^{-\infty}$
6	CG14400	0.82	-4.164	2.01×10^{-199}	CG14400	0.82	-4.163	$1.46 \times 10^{-\infty}$
7	Cyp6a17	0.53	7.503	3.08×10^{-179}	Cyp6a17	0.53	7.503	$8.29 \times 10^{-\infty}$
8	wb	0.07	-2.786	5.34×10^{-137}	wb	0.07	-2.786	$8.05 \times 10^{-\infty}$
9	Cyp6a8	0.46	2.257	4.59×10^{-128}	Cyp6a8	0.46	2.258	$1.02 \times 10^{-\infty}$
10	CG3104	1.02	1.524	1.83×10^{-122}	CG3104	1.02	1.524	$1.82 \times 10^{-\infty}$

Top 10 genes with biggest (significant) effect sizes

Table 93. Top Ten Largest Magnitude Fold Changes which 88a mutants, with/without 47b2

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	lncRNA:CR45791	0.02	7.978	2.08×10^{-7}	lncRNA:CR45791	0.02	7	
2	CG14069	0.02	7.954	2.36×10^{-7}	CG14069	0.02	7	
3	CR43734	0.04	7.823	3.81×10^{-7}	CR43734	0.04	7	
4	CG10462	0.07	-7.639	2.65×10^{-35}	CG10462	0.07	-7	
5	lncRNA:CR44057	0.02	7.619	1.10×10^{-6}	lncRNA:CR44057	0.02	7	
6	Cyp6a17	0.53	7.503	3.08×10^{-179}	Cyp6a17	0.53	7	
7	CG34124	0.00	6.084	9.80×10^{-8}	CG34124	0.00	6	
8	CG42329	0.00	-6.025	2.80×10^{-4}	CG42329	0.00	-6	
9	CG30091	0.00	6.012	5.22×10^{-4}	CG30091	0.00	6	
10	CG13748	0.02	5.943	8.15×10^{-6}	CG13748	0.02	5	

Top 10 highest expressed genes with significant change

Table 94. Top Ten Highest Expressed Genes with Significant (padj) Difference
88a mutants, with/without 47b2

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	a10	67.06	0.291	3.26×10^{-6}	a10	67.07	0.291	$3.18 \times 10^{-\infty}$
2	Os-C	55.39	0.382	3.06×10^{-12}	Os-C	55.39	0.382	2.72×10^{-1}
3	Obp19d	53.29	0.168	9.51×10^{-3}	Obp19d	53.29	0.169	$9.10 \times 10^{-\infty}$
4	Obp83b	52.99	0.330	7.75×10^{-9}	Obp83b	52.99	0.330	$7.16 \times 10^{-\infty}$
5	Obp83a	44.52	0.290	2.68×10^{-10}	Obp83a	44.52	0.289	2.04×10^{-1}
6	Obp19a	35.53	0.297	1.33×10^{-8}	Obp19a	35.54	0.298	$1.18 \times 10^{-\infty}$
7	CG14661	27.65	0.326	1.20×10^{-17}	CG14661	27.66	0.327	5.61×10^{-1}
8	CG6409	24.34	-0.463	4.65×10^{-15}	CG6409	24.34	-0.462	4.91×10^{-1}
9	Obp69a	22.73	0.325	4.30×10^{-7}	Obp69a	22.73	0.326	$4.29 \times 10^{-\infty}$
10	CG9497	21.41	-0.343	6.51×10^{-20}	CG9497	21.41	-0.343	3.31×10^{-2}

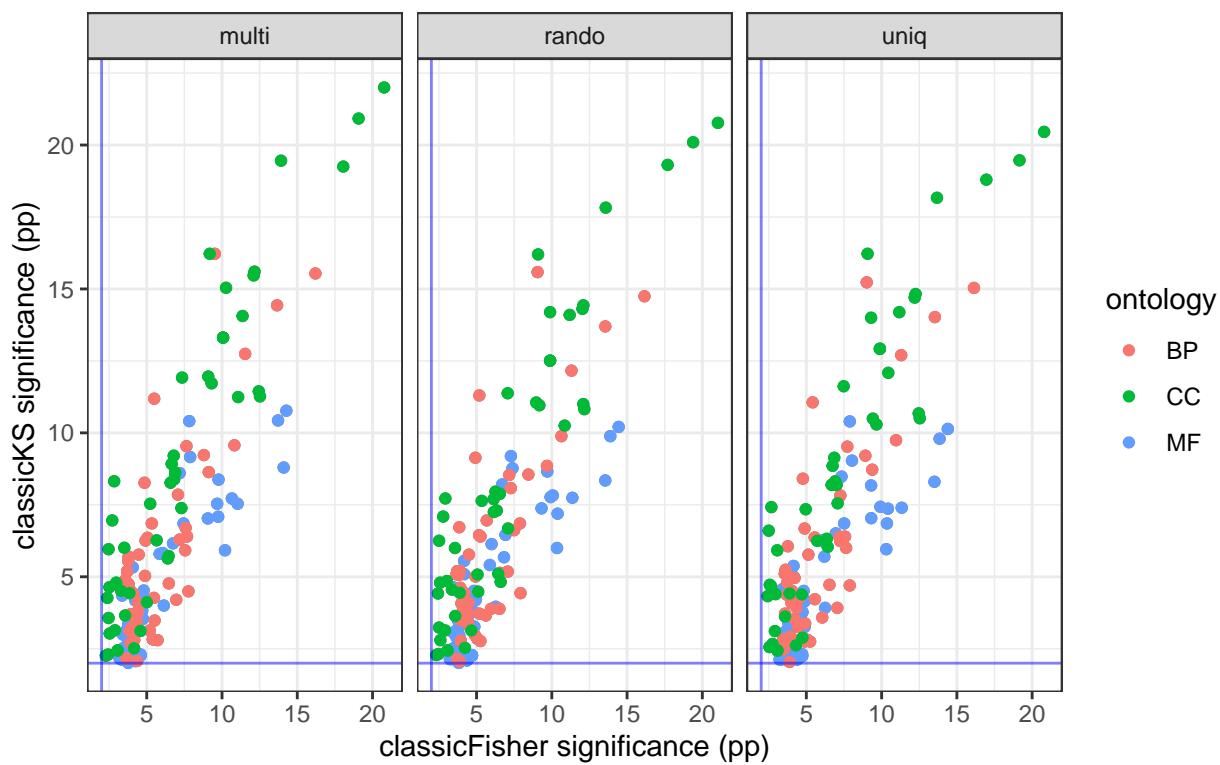
3.8.2 Gene Ontology Enrichment

Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

-> check consistency between GO terms between alignment strategies -> filter out very broad/very specific terms?

Correlation between significance values for the two tests

Figure 67. Scatterplot of GO Term Enrichment Significance for Two Tests (88a mutants, with/without 47b2)



```
## pdf
## 2

tetrapyrrole binding (GO:0046906)
inorganic molecular entity transmembrane transporter activity (GO:0015318)
system process (GO:0003008)
sensory perception (GO:0007600)
response to chemical (GO:0042221)
obsolete plasma membrane part (GO:0044459)
plasma membrane bounded cell projection (GO:0120025)
```

Table 96. Enriched GO Terms among Significantly Differentially Expressed Genes 88a mutants, with/without 47b2; multi only; top 10 most significant per category

GO Term	Description	
MF		
GO:0046906	NA	5.3
GO:0005506	iron ion binding	8.0

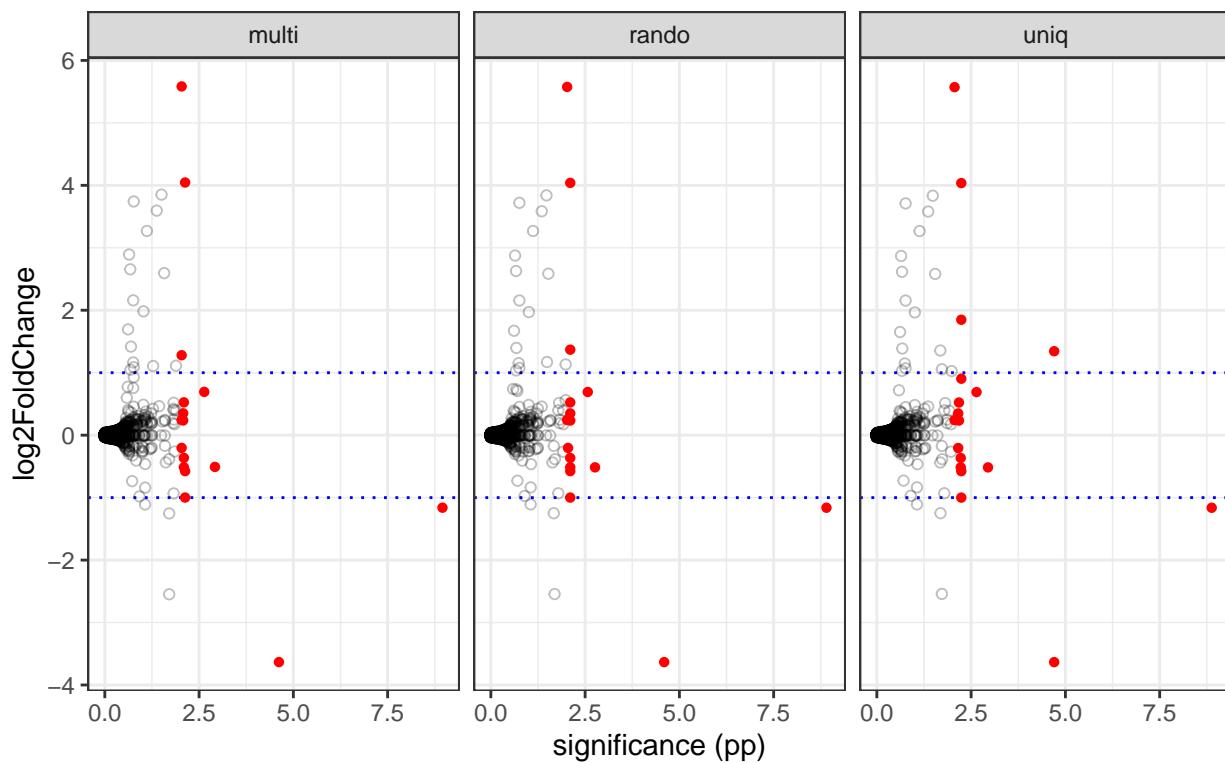
GO:0020037	heme binding	1.90
GO:0022857	transmembrane transporter activity	9.20
GO:0048037	cofactor binding	2.20
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	6.30
GO:0016491	oxidoreductase activity	1.70
GO:0005215	transporter activity	1.80
GO:0015318	NA	2.10
GO:0015075	ion transmembrane transporter activity	8.70
<hr/>		
BP		
GO:0003008	NA	6.20
GO:0050877	nervous system process	2.20
GO:0007600	NA	2.90
GO:0042221	NA	1.50
GO:0050896	response to stimulus	3.00
GO:0006811	ion transport	7.60
GO:0007606	sensory perception of chemical stimulus	1.00
GO:0042493	response to drug	1.70
GO:0055114	oxidation-reduction process	2.10
GO:0007608	sensory perception of smell	2.30
<hr/>		
CC		
GO:0005886	plasma membrane	1.70
GO:0071944	cell periphery	8.40
GO:0016020	membrane	8.80
GO:0044459	NA	1.20
GO:0016021	integral component of membrane	3.00
GO:0031224	intrinsic component of membrane	3.60
GO:0120025	NA	6.90
GO:0042995	cell projection	7.90
GO:0043005	neuron projection	4.30
GO:0005576	extracellular region	8.50

3.9 Two Days' Difference

Full table: results/tables/supp/two_days_difference.de.tsv

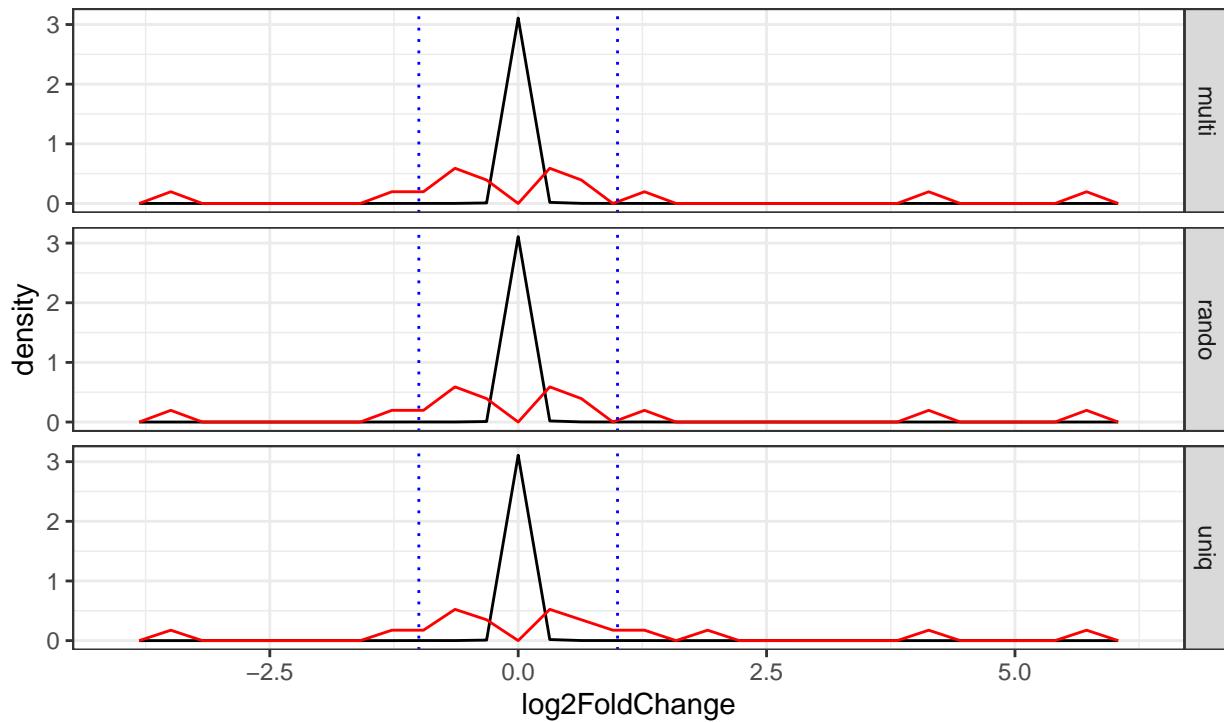
Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative log2 fold changes are downregulated relative to the group-housed condition; positive fold changes are upregulated.

Figure 68. Volcano Plot: Fold Change vs. Significance
(from day 5 to day 7)



```
## pdf
## 2
```

**Figure 69. histogram of fold change
with significant($p_{adj} < 0.01$) changes highlighted in red
(from day 5 to day 7)**



```
## pdf
## 2
```

Of the 10942 genes with significance scores available, 19 have an adjusted $p < 0.01$ (0.1736428 %)

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than 1/2), significant (ie, $p_{adj} < 0.01$) changes. There were 7 such genes, mostly shared across alignment strategy:

**Table 97. Genes with Large ($|2| > \text{fold change}$), Significant ($p_{adj} < 0.01$) Changes
from day 5 to day 7**

	multi	rando	uniq
Cpr64Ac	yes	yes	yes
CheB38c	yes	yes	yes
CG12239	yes	yes	yes
alarm	yes	yes	yes
Gyc76C	no	no	yes
CG17715	no	no	yes
28SrRNA-Psi:CR41609	yes	yes	no

3.9.1 Top Tens

Genes with top 10 most significant changes

Ordered in decreasing significance, the alignment strategies agree on the top 10 most significant changes:

Table 98. Top Ten Most Significantly (padj<0.01) Differentia
88a mutants, with/without 47b2

multi					rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	
1	Cpr64Ac	0.57	-1.161	1.11×10^{-9}	Cpr64Ac	0.57	-1.162	
2	CheB38c	0.04	-3.633	2.41×10^{-5}	CheB38c	0.04	-3.634	
3	CG7084	0.15	-0.508	1.20×10^{-3}	CG7084	0.14	-0.515	
4	HERC2	0.38	0.693	2.31×10^{-3}	HERC2	0.38	0.691	
5	alrm	0.02	4.046	7.43×10^{-3}	28SrRNA-Psi:CR41609	12.63	1.369	
6	wntD	0.07	-0.999	7.43×10^{-3}	mim	0.12	0.349	
7	CG9689	0.30	-0.576	7.43×10^{-3}	alrm	0.02	4.039	
8	Tsp	0.07	-0.514	8.01×10^{-3}	wntD	0.07	-0.999	
9	CG11122	0.13	0.526	8.01×10^{-3}	Tsp	0.07	-0.515	
10	amd	0.36	-0.364	8.01×10^{-3}	CG11122	0.13	0.524	

Top 10 genes with biggest (significant) effect sizes

Table 99. Top Ten Largest Magnitude Fold Changes w
47b1 mutants, from day 5 to day 7

multi					rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2	
1	CG12239	0.01	5.584	9.18×10^{-3}	CG12239	0.01		
2	alrm	0.02	4.046	7.43×10^{-3}	alrm	0.02		
3	CheB38c	0.04	-3.633	2.41×10^{-5}	CheB38c	0.04		
4	28SrRNA-Psi:CR41609	19.62	1.280	9.18×10^{-3}	28SrRNA-Psi:CR41609	12.63		
5	Cpr64Ac	0.57	-1.161	1.11×10^{-9}	Cpr64Ac	0.57		
6	wntD	0.07	-0.999	7.43×10^{-3}	wntD	0.07		
7	HERC2	0.38	0.693	2.31×10^{-3}	HERC2	0.38		
8	CG9689	0.30	-0.576	7.43×10^{-3}	CG9689	0.30		
9	CG11122	0.13	0.526	8.01×10^{-3}	CG11122	0.13		
10	Tsp	0.07	-0.514	8.01×10^{-3}	CG7084	0.14		

Top 10 highest expressed genes with significant change

Table 100. Top Ten Highest Expressed Genes with S
Difference
47b1 mutants, from day 5 to day 7

multi					rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2	
1	28SrRNA-Psi:CR41609	19.62	1.280	9.18×10^{-3}	28SrRNA-Psi:CR41609	12.63		
2	Obp56d	8.88	-0.204	9.18×10^{-3}	Obp56d	8.88		
3	Cpr64Ac	0.57	-1.161	1.11×10^{-9}	Cpr64Ac	0.57		
4	scrib	0.41	0.243	9.18×10^{-3}	scrib	0.41		
5	CG32809	0.39	0.236	8.37×10^{-3}	CG32809	0.39		
6	HERC2	0.38	0.693	2.31×10^{-3}	HERC2	0.38		
7	amd	0.36	-0.364	8.01×10^{-3}	amd	0.36		
8	CG9689	0.30	-0.576	7.43×10^{-3}	CG9689	0.30		

9	CG7084	0.15	-0.508	1.20×10^{-3}	CG7084	0.14
10	CG11122	0.13	0.526	8.01×10^{-3}	CG11122	0.13

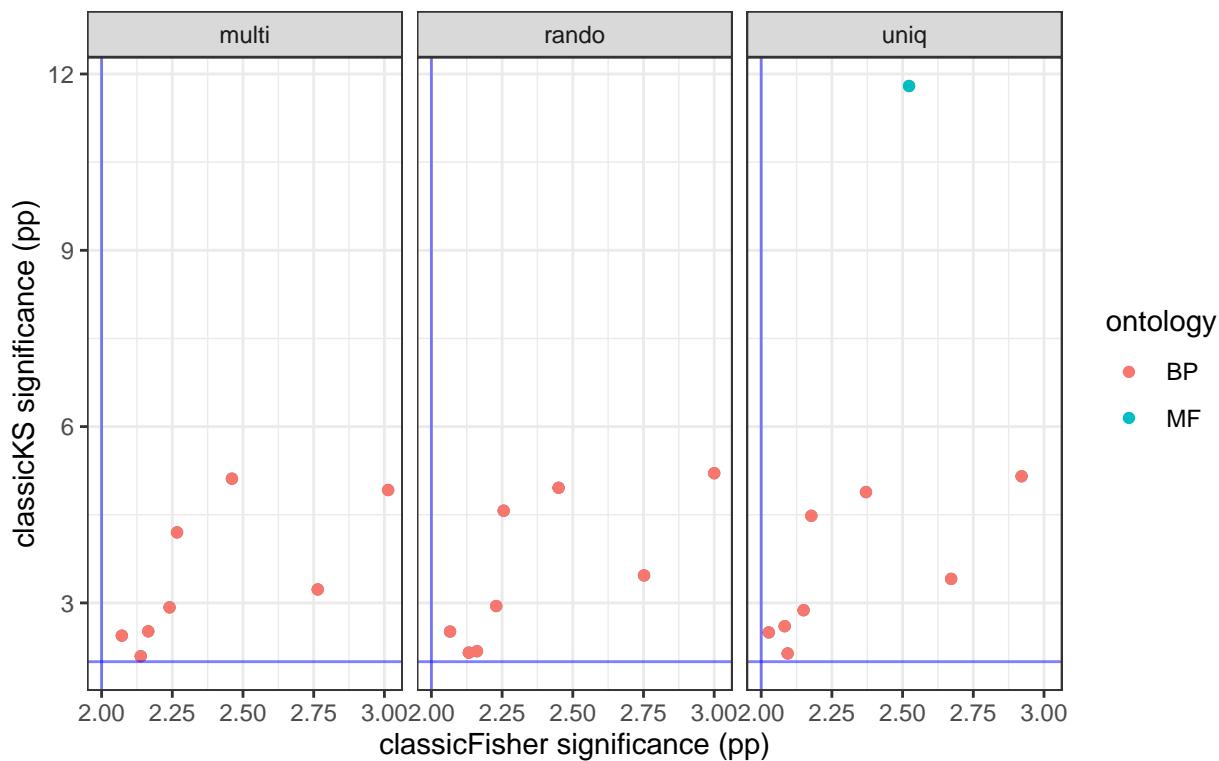
3.9.2 Gene Ontology Enrichment

Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

-> check consistency between GO terms between alignment strategies -> filter out very broad/very specific terms?

Correlation between significance values for the two tests

Figure 70. Scatterplot of GO Term Enrichment Significance for Two Tests (47b1 mutants, from day 5 to day 7)



```
## pdf
## 2

regulation of vesicle-mediated transport (GO:0060627)
```

Table 101. Enriched GO Terms among Significantly Differentially Expressed Genes 47b1 mutants, from day 5 to day 7; uniq only; top 10 most significant per category

GO Term	Description	p-value	
		Fisher	K-S

MF

GO:0005102	signaling receptor binding	3.00×10^{-3}	1.60×10^{-12}
BP			
GO:0008354	germ cell migration	1.20×10^{-3}	7.00×10^{-6}
GO:0030100	regulation of endocytosis	2.13×10^{-3}	3.90×10^{-4}
GO:0042048	olfactory behavior	4.26×10^{-3}	1.30×10^{-5}
GO:0060627	NA	6.65×10^{-3}	3.30×10^{-5}
GO:0007635	chemosensory behavior	7.08×10^{-3}	1.33×10^{-3}
GO:0016336	establishment or maintenance of polarity of larval imaginal disc epithelium	8.06×10^{-3}	7.24×10^{-3}
GO:0040003	chitin-based cuticle development	8.26×10^{-3}	2.49×10^{-3}
GO:0016335	morphogenesis of larval imaginal disc epithelium	9.40×10^{-3}	3.18×10^{-3}

4 Bibliography

```
##  
## To cite ggplot2 in publications, please use:  
##  
## H. Wickham. ggplot2: Elegant Graphics for Data Analysis.  
## Springer-Verlag New York, 2016.  
##  
## A BibTeX entry for LaTeX users is  
##  
## @Book{,  
##   author = {Hadley Wickham},  
##   title = {ggplot2: Elegant Graphics for Data Analysis},  
##   publisher = {Springer-Verlag New York},  
##   year = {2016},  
##   isbn = {978-3-319-24277-4},  
##   url = {https://ggplot2.tidyverse.org},  
## }  
  
##  
## Zhu, A., Ibrahim, J.G., Love, M.I. Heavy-tailed prior distributions  
## for sequence count data: removing the noise and preserving large  
## differences Bioinformatics (2018)  
##  
## A BibTeX entry for LaTeX users is  
##  
## @Article{,  
##   title = {Heavy-tailed prior distributions for sequence count data: removing the noise and preser-  
##   author = {Anqi Zhu and Joseph G. Ibrahim and Michael I. Love},  
##   year = {2018},  
##   journal = {Bioinformatics},  
##   doi = {10.1093/bioinformatics/bty895},  
## }  
  
##  
## Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change  
## and dispersion for RNA-seq data with DESeq2 Genome Biology 15(12):550  
## (2014)
```

```

## A BibTeX entry for LaTeX users is
##
## @Article{,
##   title = {Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2},
##   author = {Michael I. Love and Wolfgang Huber and Simon Anders},
##   year = {2014},
##   journal = {Genome Biology},
##   doi = {10.1186/s13059-014-0550-8},
##   volume = {15},
##   issue = {12},
##   pages = {550},
## }
## To cite the biomaRt package in publications use:
##
## Mapping identifiers for the integration of genomic datasets with the
## R/Bioconductor package biomaRt. Steffen Durinck, Paul T. Spellman,
## Ewan Birney and Wolfgang Huber, Nature Protocols 4, 1184-1191 (2009).
##
## BioMart and Bioconductor: a powerful link between biological
## databases and microarray data analysis. Steffen Durinck, Yves Moreau,
## Arek Kasprzyk, Sean Davis, Bart De Moor, Alvis Brazma and Wolfgang
## Huber, Bioinformatics 21, 3439-3440 (2005).
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.

## To cite package 'topGO' in publications use:
##
## Adrian Alexa and Jorg Rahnenfuhrer (2018). topGO: Enrichment Analysis
## for Gene Ontology. R package version 2.34.0.
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {topGO: Enrichment Analysis for Gene Ontology},
##   author = {Adrian Alexa and Jorg Rahnenfuhrer},
##   year = {2018},
##   note = {R package version 2.34.0},
## }
## ATTENTION: This citation information has been auto-generated from the
## package DESCRIPTION file and may need manual editing, see
## 'help("citation")'.

```

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An ultra-fast all-in-one FASTQ preprocessor.” *Bioinformatics* 34 (17): i884–i890. doi:10.1093/bioinformatics/bty560.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features.” *Bioinformatics* 30 (7): 923–30.

doi:10.1093/bioinformatics/btt656.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology* 15 (12): 1–21. doi:10.1186/s13059-014-0550-8.

Shiao, Meng Shin, Jia Ming Chang, Wen Lang Fan, Mei Yeh Jade Lu, Cedric Notredame, Shu Fang, Rumi Kondo, and Wen Hsiung Li. 2015. “Expression divergence of chemosensory genes between *Drosophila sechellia* and its sibling species and its implications for host shift.” *Genome Biology and Evolution* 7 (10): 2843–58. doi:10.1093/gbe/evv183.

Wang, Kai, Darshan Singh, Zheng Zeng, Stephen J Coleman, Yan Huang, Gleb L Savich, Xiaping He, et al. 2010. “MapSplice: accurate mapping of RNA-seq reads for splice junction discovery.” *Nucleic Acids Research* 38 (18): e178. doi:10.1093/nar/gkq622.