Volkan Lab Behavioral Genetics RNA-Seq

Charlie Soeder

11/15/2019

Contents

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

Reference annotations were used to define gene locii for differential expression analysis:

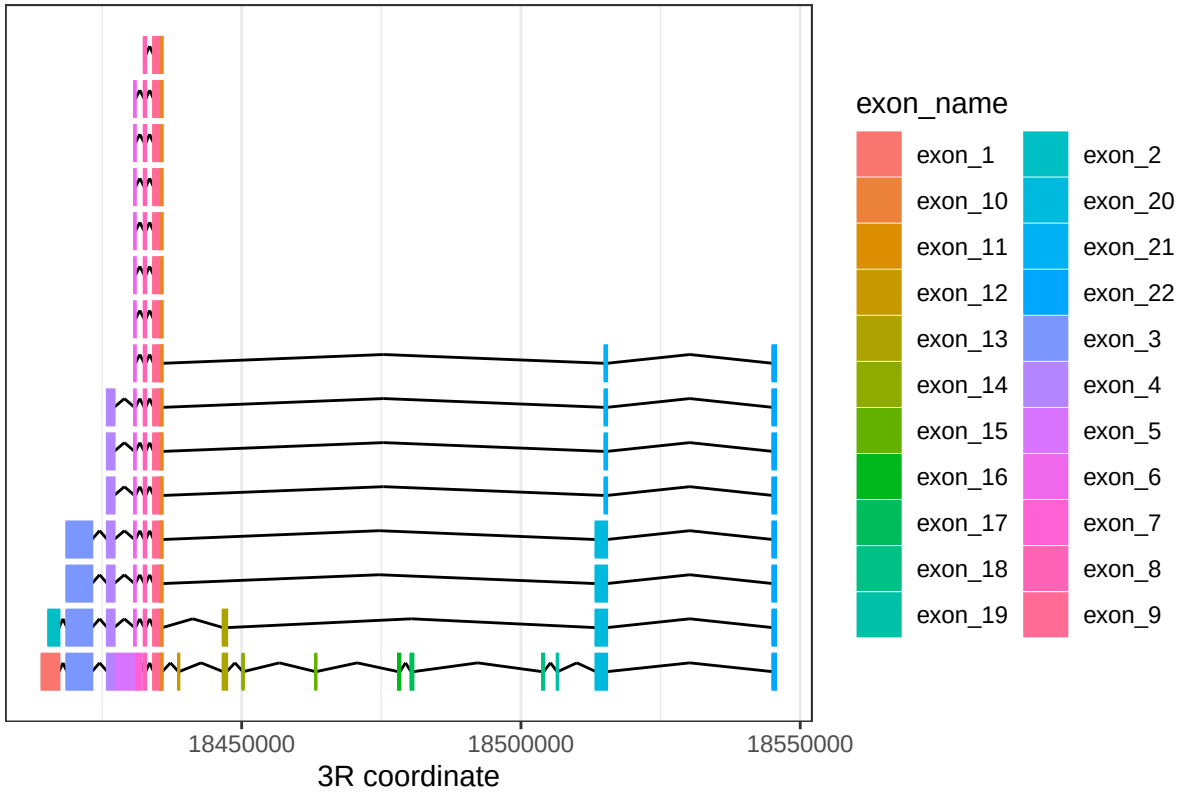
Table 2. Reference Annotations and their Sizes

	size (bp)		
annot	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

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

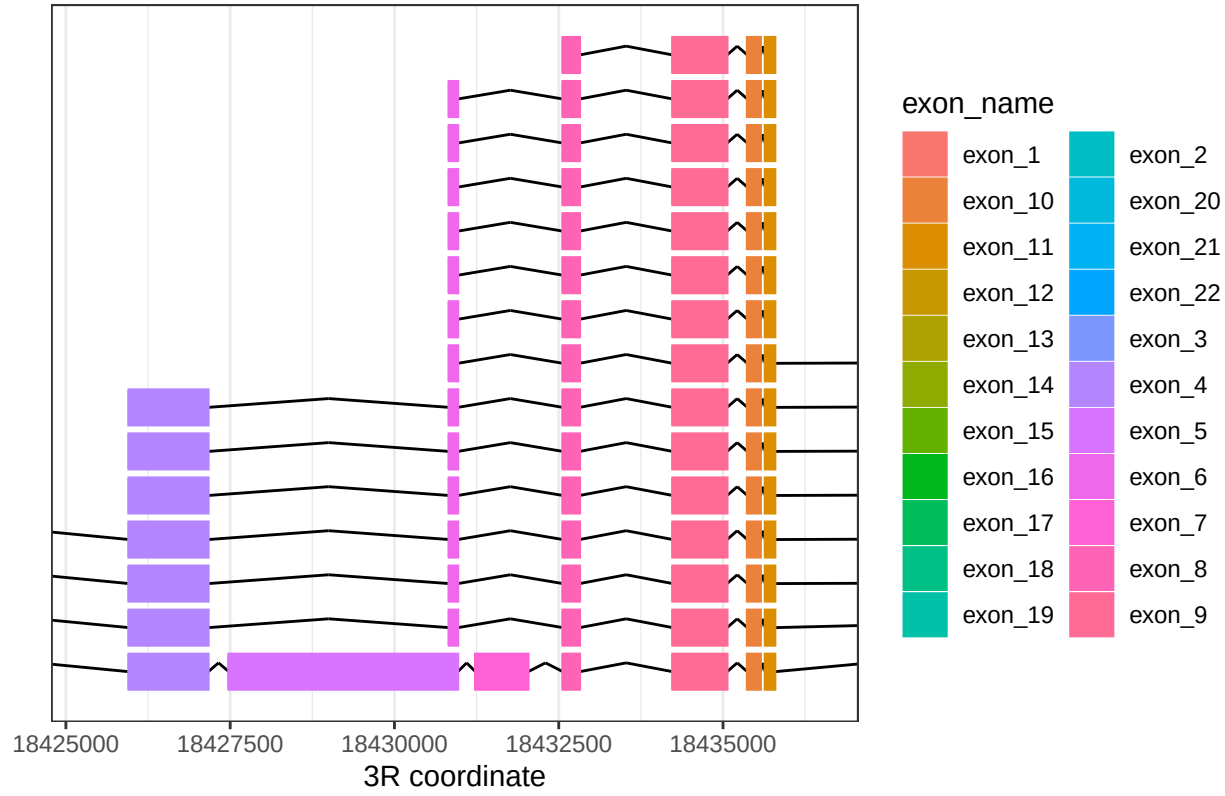
```
cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf | gr
cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf | gr
cat fru.test.gtf.exon fru.test.gtf.gene | bedtools sort > utils/annotations/fru_ex.gtf
cat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

Figure 1. Fru gene model




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

Figure 2. Fru gene model (detail)




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

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

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 mutiple genes, some not unique. Once merged & uniqued : 246 Annotation symbols (CGxxxxx) converted to FlyBase gene games (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
l(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 condi-

tions, 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)

			mutant genotypes				
housing	tissue	day	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

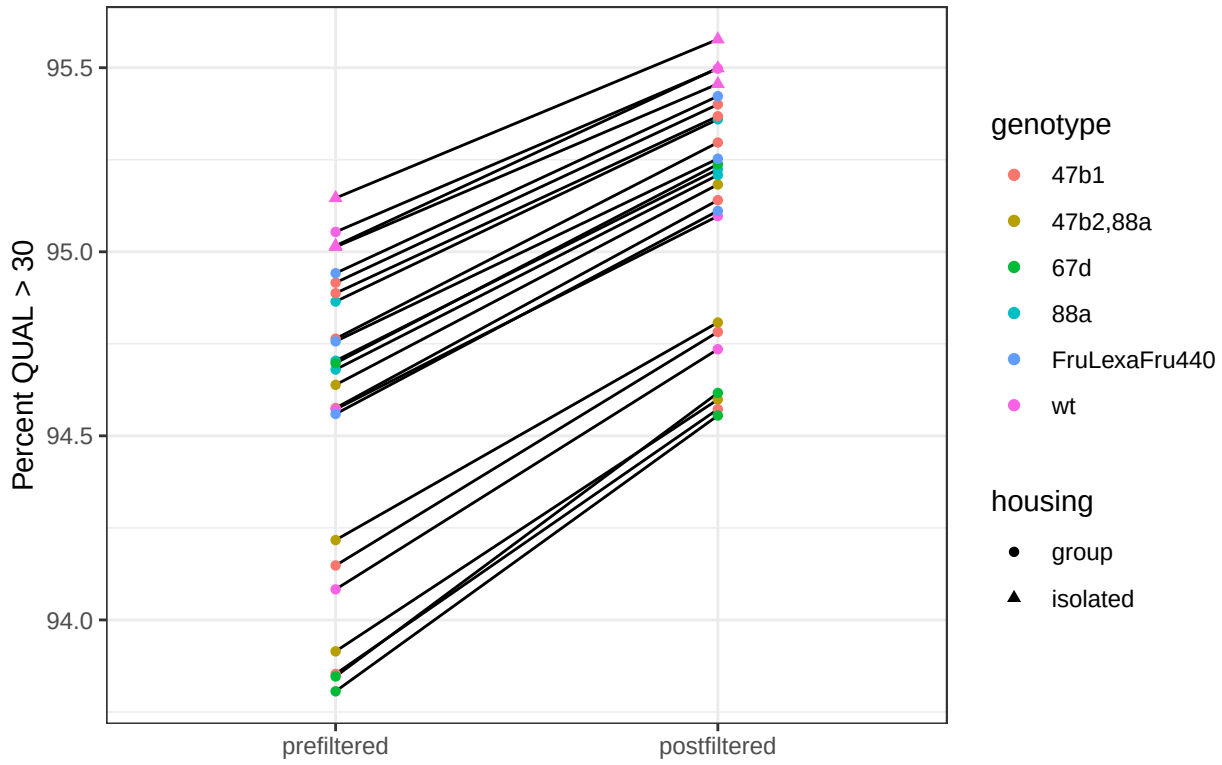
These reads were preprocessed with FASTP (S. Chen et al. 2018) for quality control and analytics.

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 3. Percent of Reads with a mean QUAL > 30




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

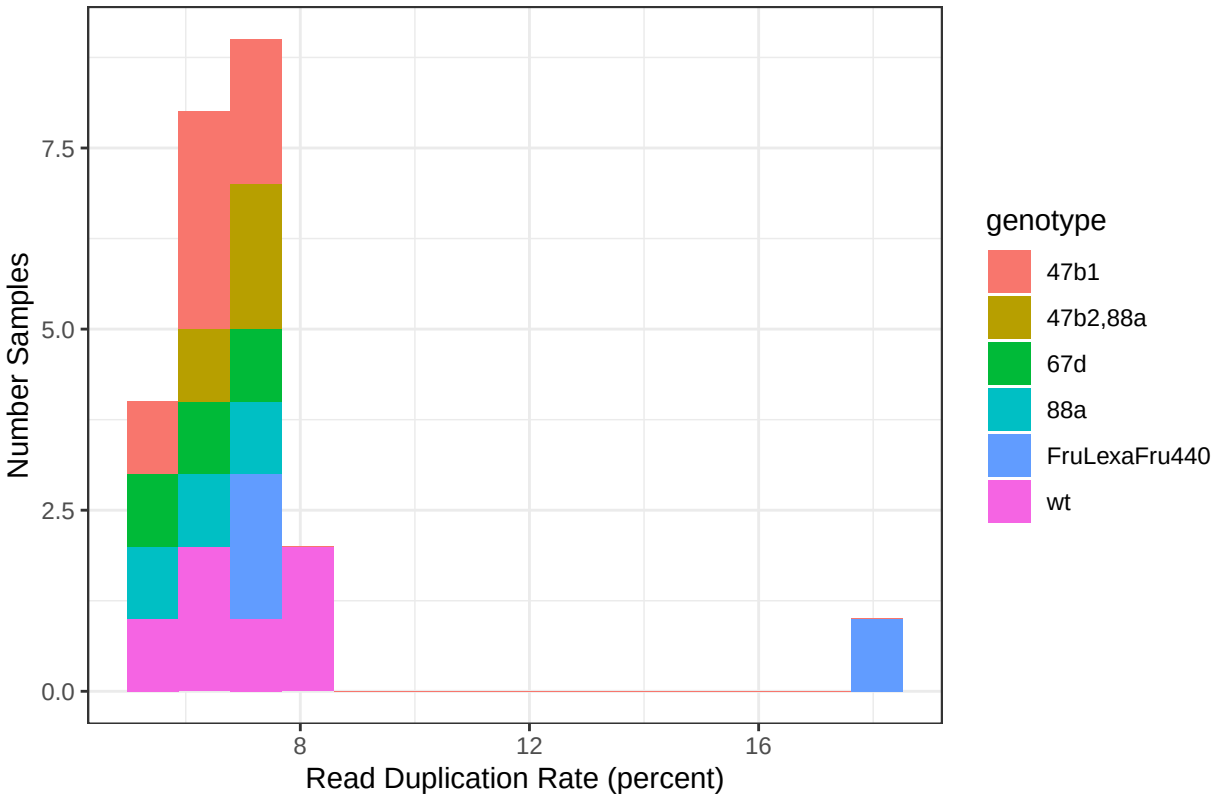
Duplicate reads were also detected

Table 7. Percentage Duplication

FASTP estimate

minimum	average	median	maximum
5.2	7.1	6.7	17.9

Figure 4. Duplication Histogram (FASTP estimate)




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

2.5 Mapped Reads

Reads were mapped to the reference genome using MapSplice2 (Wang et al. 2010). Because MapSplice is written in python2, the code was downloaded and automatically refactored using the 2to3 python utility so that it would run in the python3 snakemake environment: https://docs.python.org/2/library/2to3.html

2.5.1 Raw Mapsplice

Of the $710M$ reads, MapSplice was able to align $705M$ of them, for an overall mapping rate of 99.3370373 %.

Individual mapping rates were generally more than 98%.

Table 8. Percent of Reads Mapping
raw mapsplice output

maximum	mean	median	minimum
99.7%	99.3%	99.5%	98.5%

Table 9. Individual Mapping Rates
raw mapsplice output

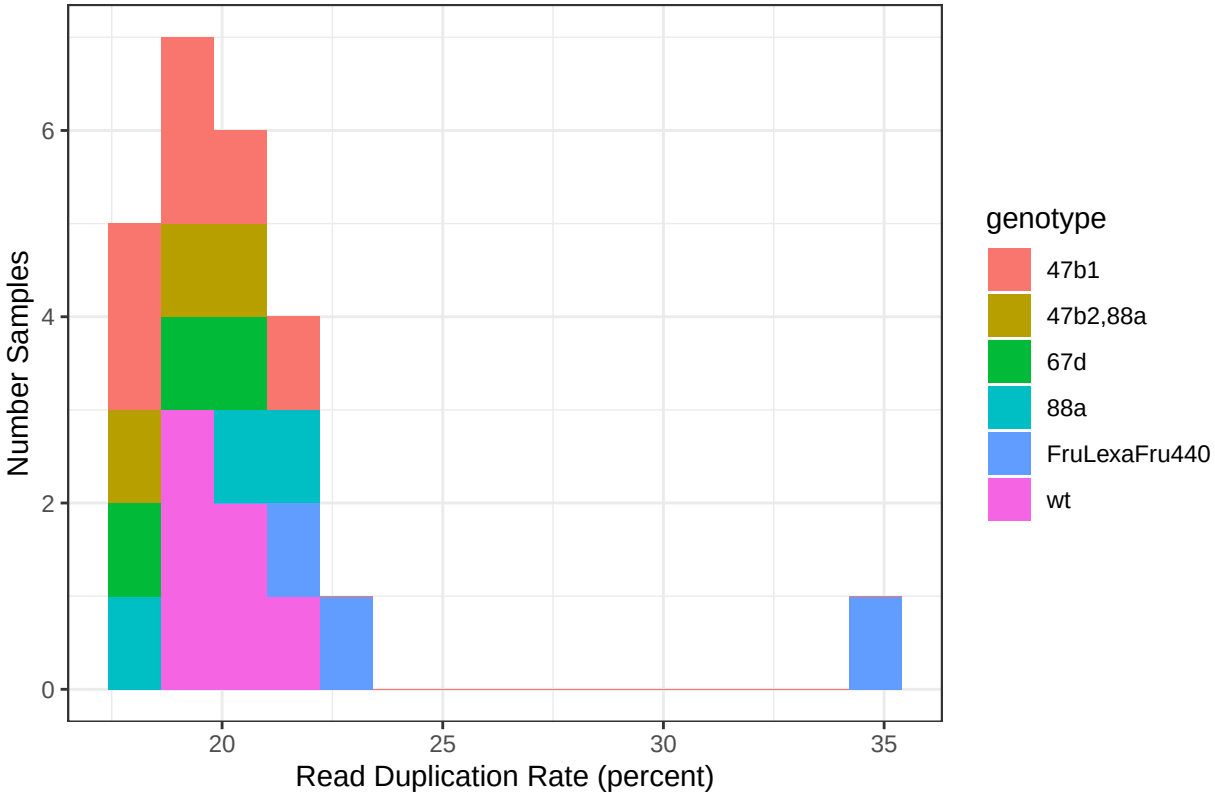
rep	day	total reads	reads mapped	percent mapped
group - 47b1				
1	5	$32.0M$	$31.9M$	99.7%
2	5	$28.2M$	$28.0M$	99.4%
3	5	$24.4M$	$24.2M$	99.0%
1	7	$32.1M$	$31.9M$	99.5%
2	7	$28.9M$	$28.8M$	99.7%
3	7	$24.3M$	$24.3M$	99.6%
group - 47b2,88a				
1	5	$20.3M$	$20.2M$	99.5%
2	5	$31.7M$	$31.6M$	99.5%
3	5	$24.7M$	$24.5M$	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%

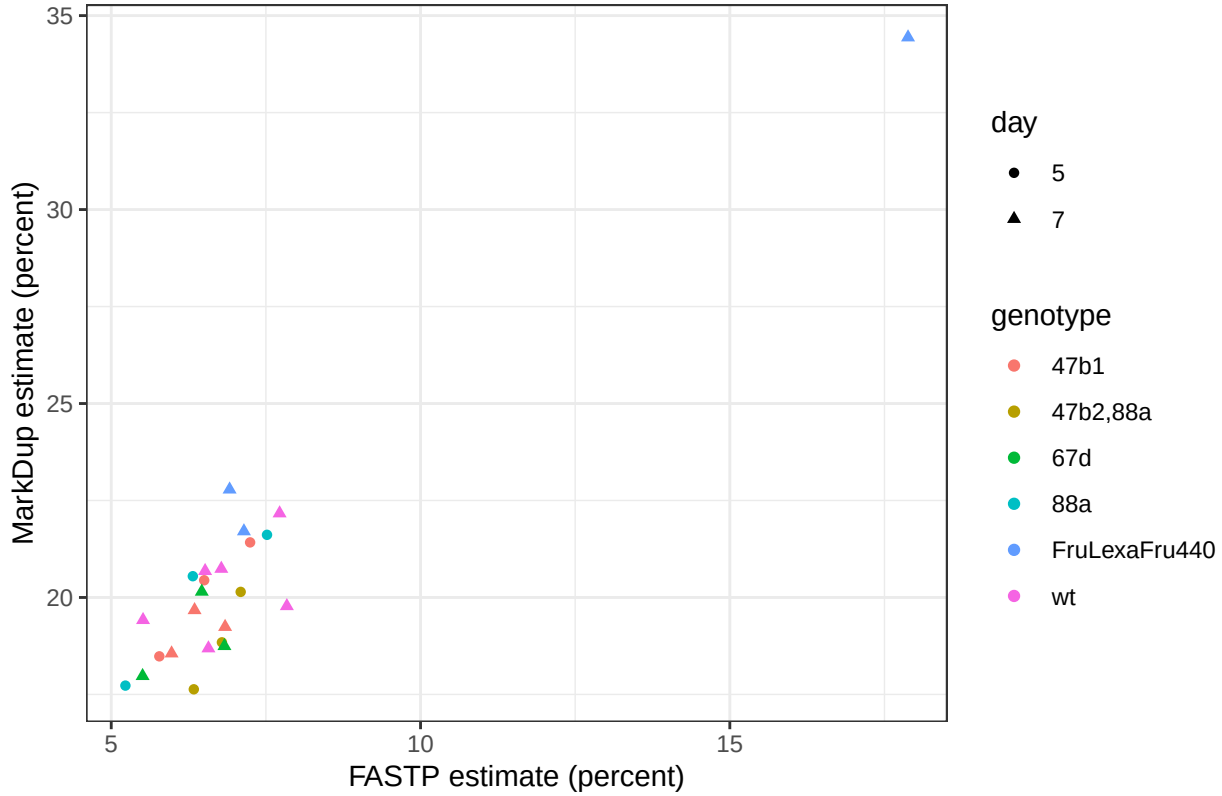
Figure 5. Duplication Histogram (Raw Mapsplice Alignment)




```
## 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.

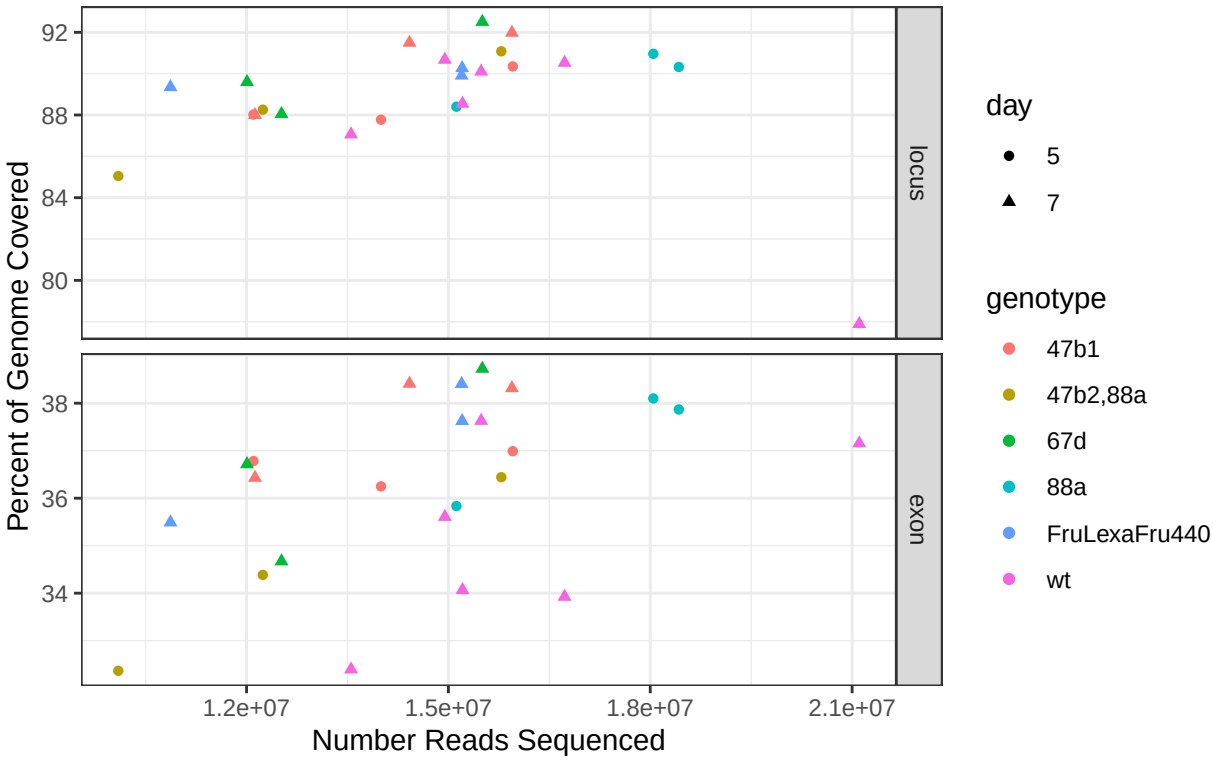
Figure 6. Comparison of Duplication Rate Estimates




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

Genome-wide depth of coverage is not very meaningful here, in the case of RNA-Seq. Breadth of coverage (the fraction of the genome which is covered by at least one read) is, but the ideal case is not 100% coverage like in a DNA-Seq; rather, we'd expect breadth to approximate the fraction of the genome which is under active transcription. Another complication is whether the reads which fall on splice junctions are treated as covering the intronic region or not (this corresponds to the distinction between the percent of the genome which is a transcribed locus vs the percent which is a transcribed exon).

Figure 7. Breadth of Coverage of Raw Mapsplice Alignment Compared to Read Count

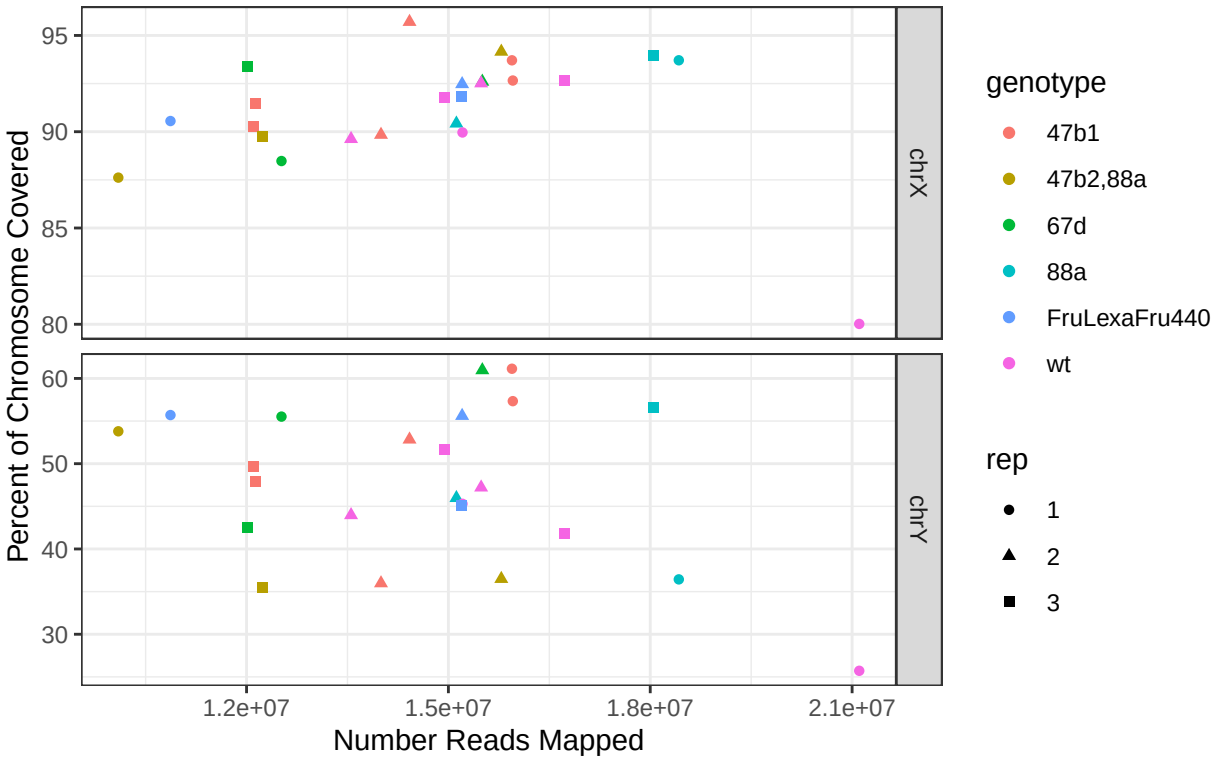



```
## pdf
##   2
```

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 unusally 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 8. 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 of $710M$ 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 mapspliceMulti, but will be further filtered in other alignments.

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

	percent of reads uniquely mapping		average per-read mapping multiplicity	
rep	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

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

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

mapspliceUniq is derived from mapspliceMulti by further filtering out the multimapped reads and keeping only those which map uniquely.

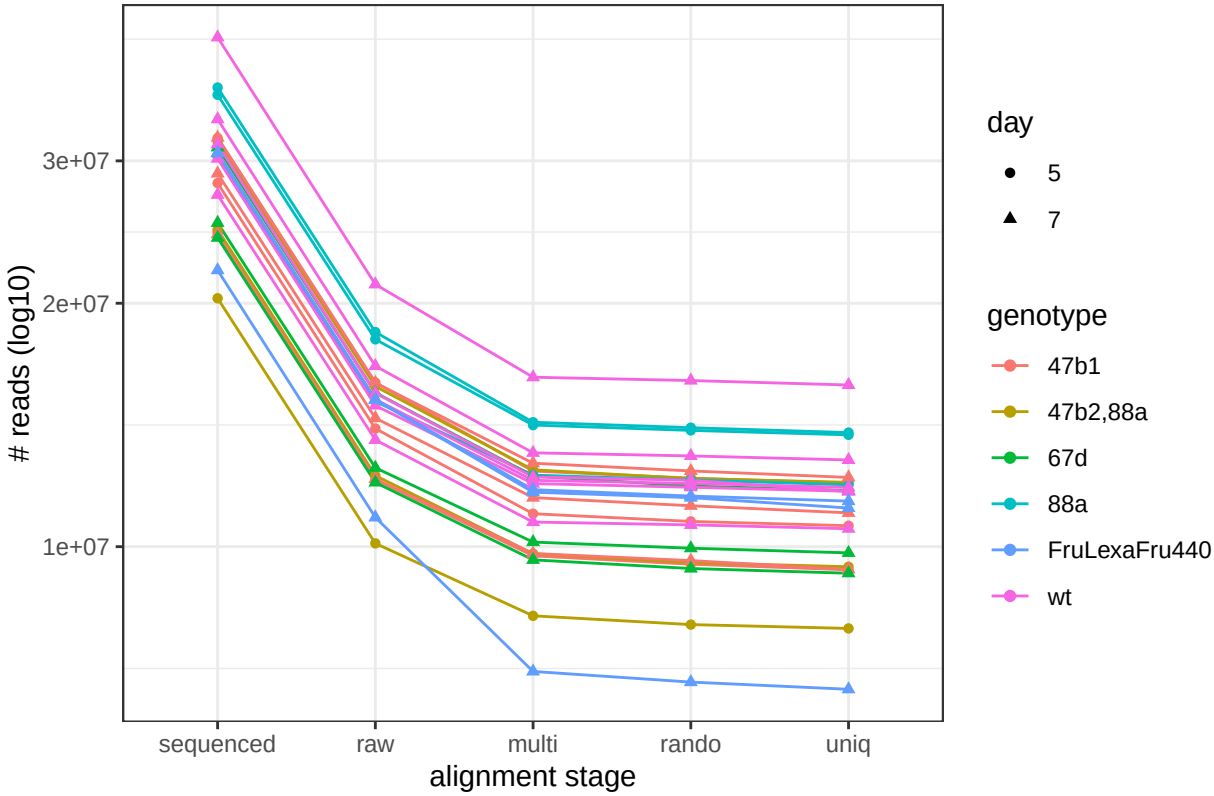
Table 15. Uniquely Mapped Retention Rate
percent of alignment retained when multimappers are excluded

	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 Alignment Process Overview

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

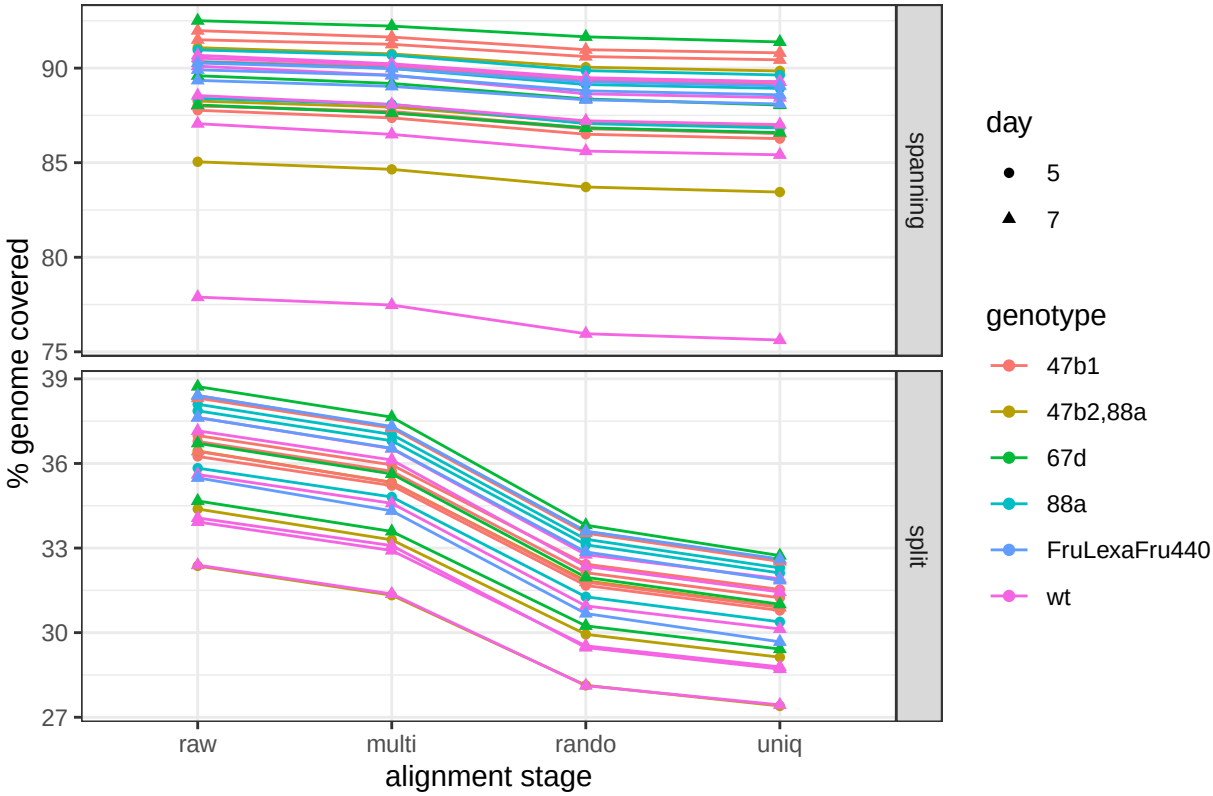
Figure 9. Read−count Dropout During Alignment Process




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

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

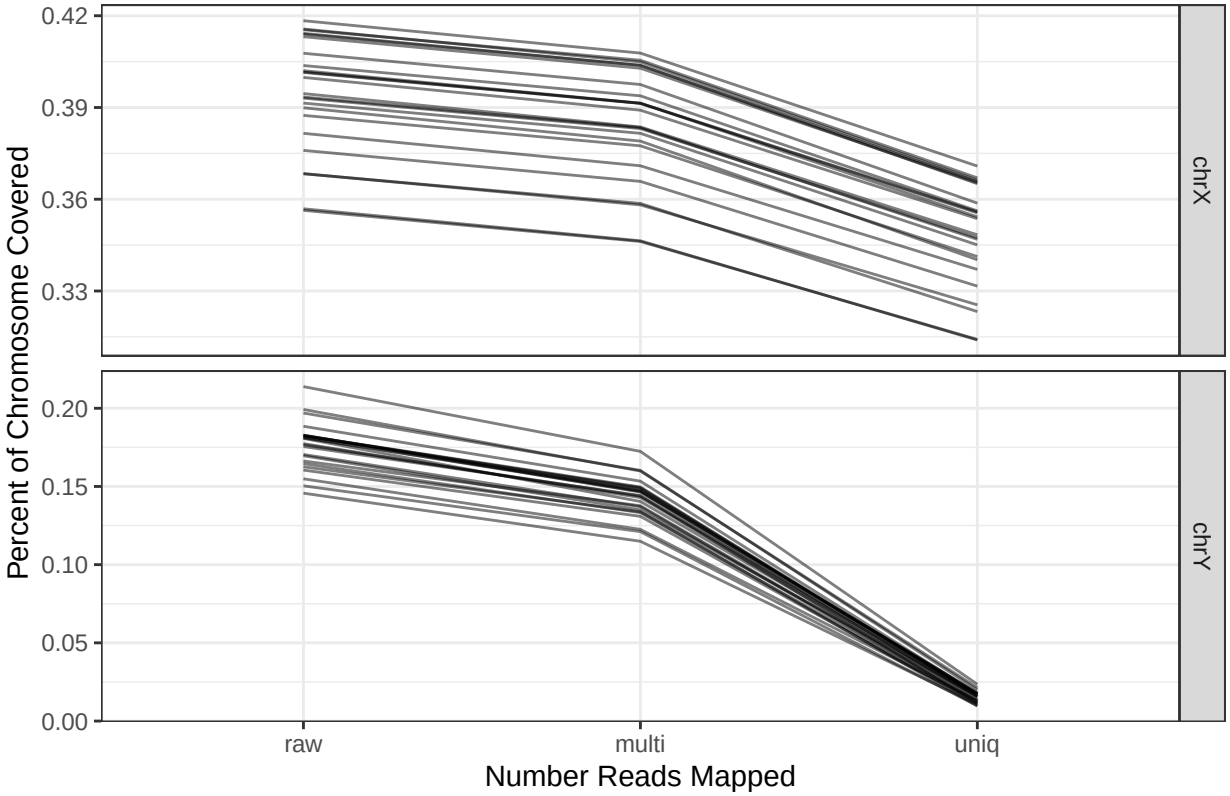
Figure 10. Coverage Loss During Alignment Process




```
## 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 11. Fraction of Sex Chromosome Covered in Raw Mapsplice Alignm




```
## 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.

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

	mapping strategy		
rep	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			

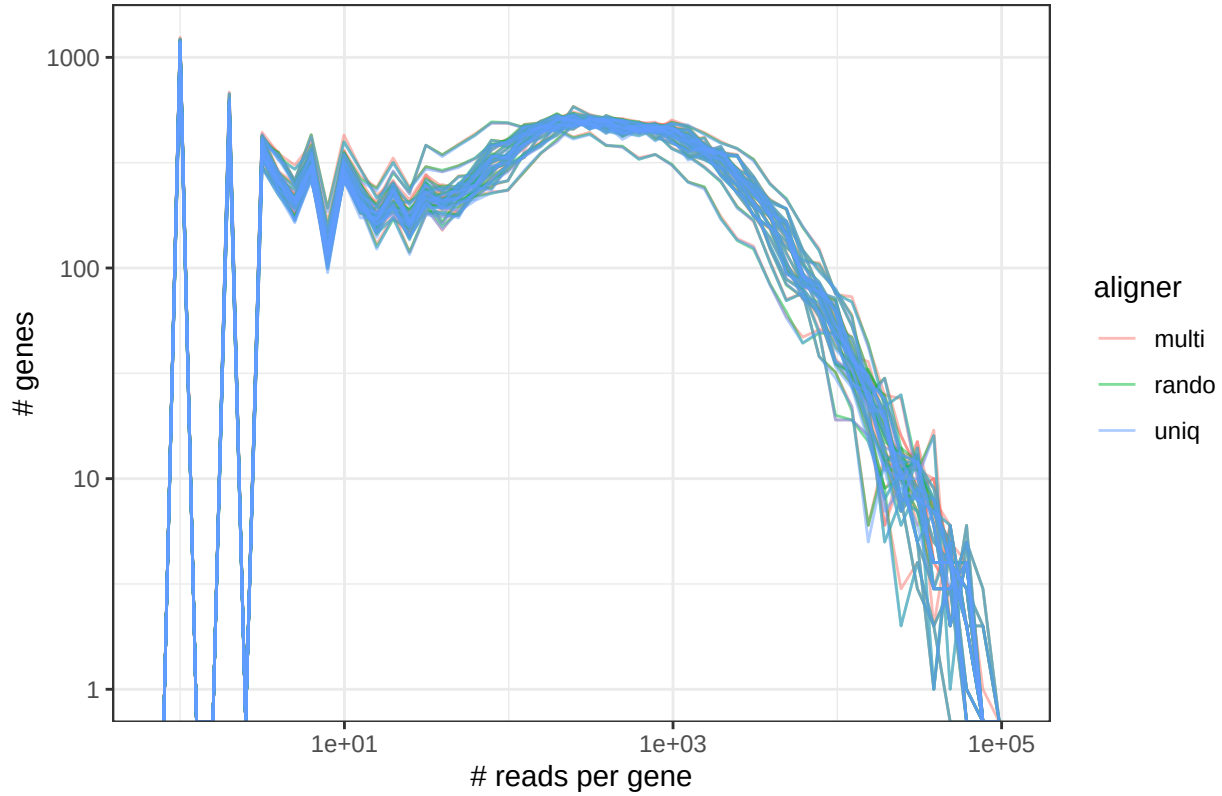
1	89.8%	91.3%	91.7%
2	89.8%	91.4%	91.9%
3	89.7%	91.2%	91.7%
88a - group - 5			
1	90.4%	91.0%	91.7%
2	90.8%	91.2%	91.8%
3	90.6%	91.2%	91.8%
47b1 - group - 7			
1	89.7%	90.7%	91.5%
2	88.9%	90.0%	91.0%
3	89.6%	90.2%	91.5%
67d - group - 7			
1	90.8%	91.6%	92.1%
2	88.5%	90.3%	90.9%
3	89.0%	90.4%	91.0%
FruLexaFru440 - 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 12. Per−Gene Read Count Histogram (by aligner and sample)



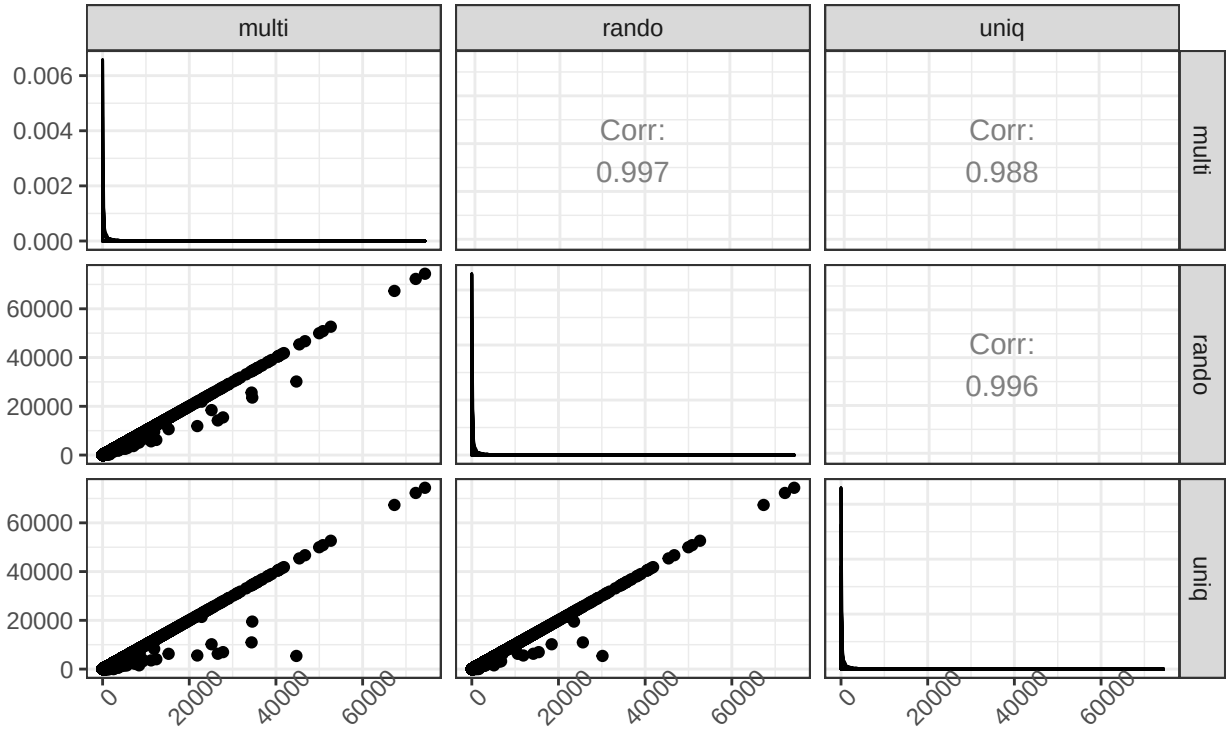

```
## pdf
##   2
```

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

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

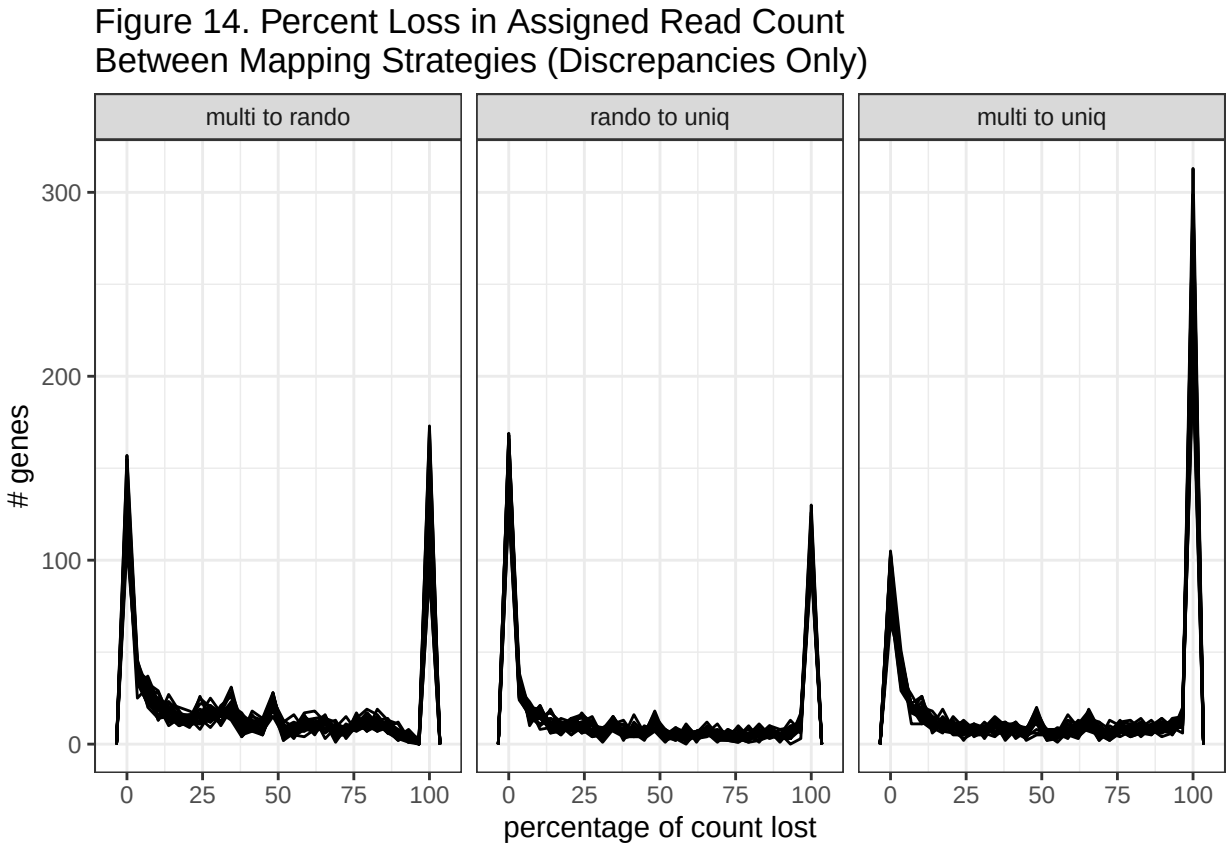
Figure 13. Correlations between Read Count Assigned to Gene Across Alignment Strategy (downsampled to 10%)




```
## 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 >= rando >= uniq. It's not clear why but for a very small number of cases (23; 0.00547932151705737 %), rando > multi.




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

2.6.1 Fru 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. (May need to restructure this..... see 3.5 below)

2.7 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

http://master.bioconductor.org/packages/release/workflows/vignettes/rnaseqGene/inst/doc/rnaseqGene.html

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

Currently, single-variable contrasts are being calculated, in which the axis of comparison is specificied (eg, housing; genotype) and available samples are subset to the relevant contrast (eg, wt group reps 1,2,3 and wt isolated reps 1,2,3; wt group reps 1,2,3 and 67d group reps 1,2,3)

Because the "baseMean" reported by DESeq2 is a replicate-normalized read count, an expression level was derived from it by scaling by feature length.

Interpretation of "baseMean": https://support.bioconductor.org/p/75244

3 Results

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:

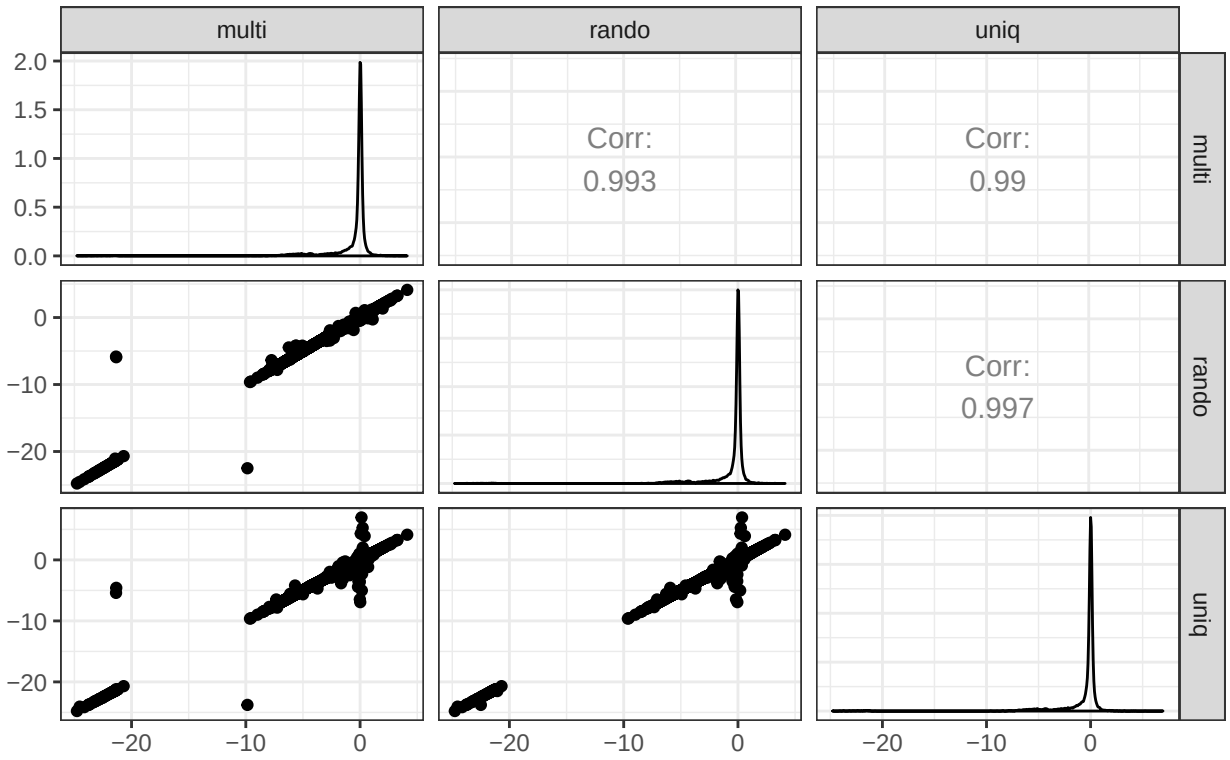
Table 19. Number Genes with Sufficient Read Count for Differential Expression Analysis
from 17747 annotations in dm6_genes

	tested	percent
multi	$12.0K$	67.8%
rando	$11.9K$	67.1%
uniq	$11.8K$	66.6%

3.1.1 preshrunk comparison across alignment strategies

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

Figure 15. Agreement on (unshrunk) Effect Size (log2 fold change) Between Alignment Strategies (isolated vs group−housed wildtypes)

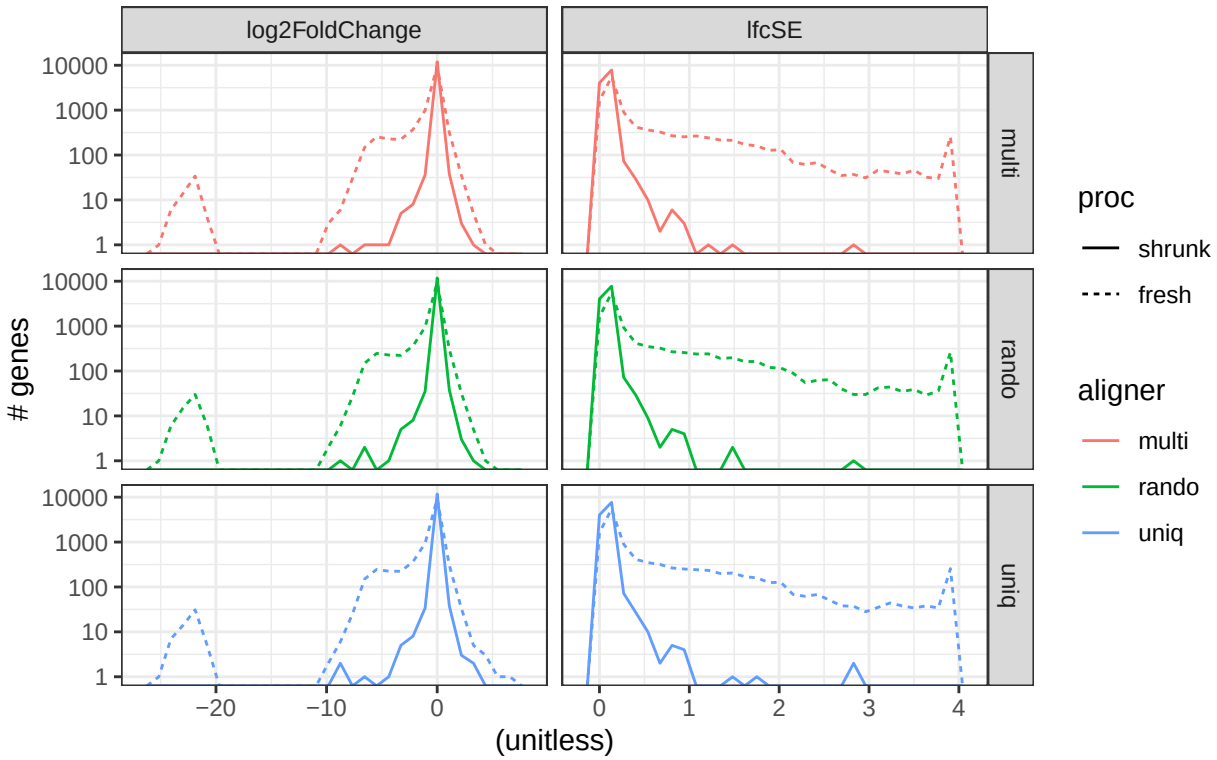



```
## pdf
##   2
```

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.

Figure 16. Log2 Fold Change and Standard Error
(isolated vs group−housed wildtypes)

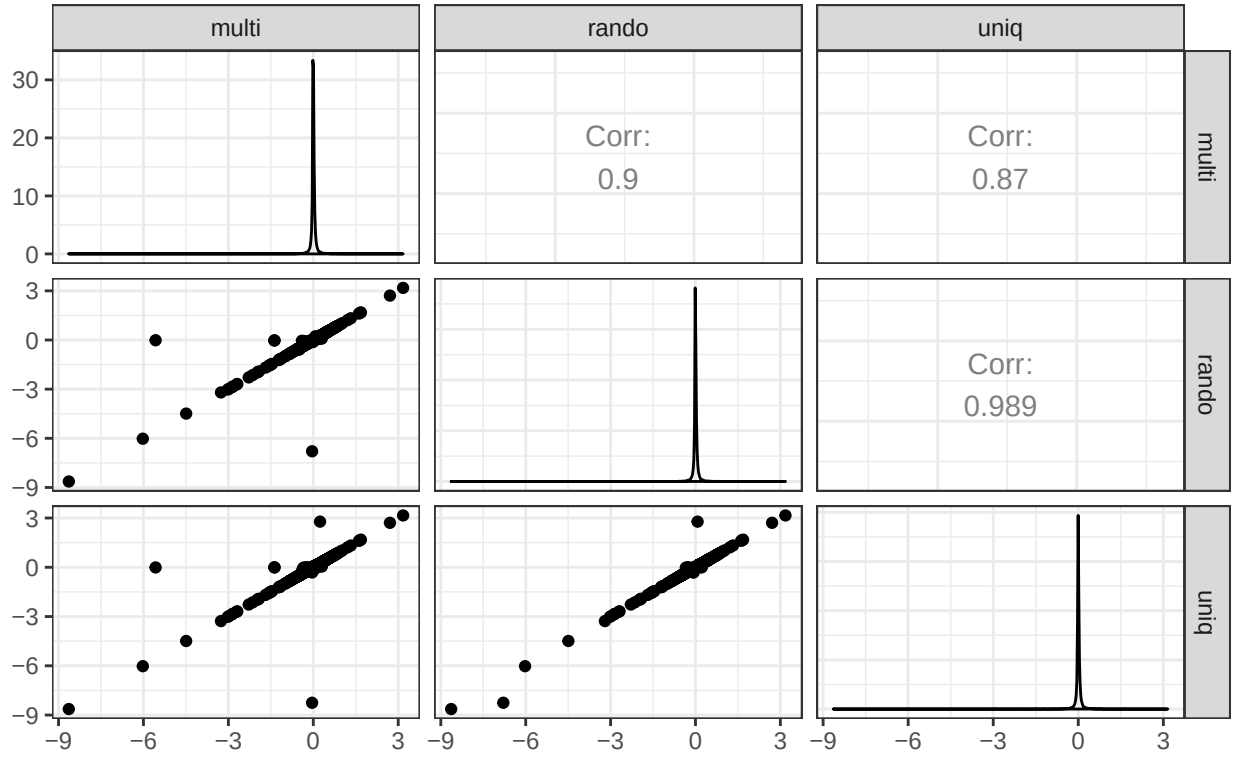



```
## pdf
##   2
```

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.

Figure 17. Agreement on (shrunk) Effect Size (log2 fold change) Between Alignment Strategies (isolated vs group−housed wildtypes)

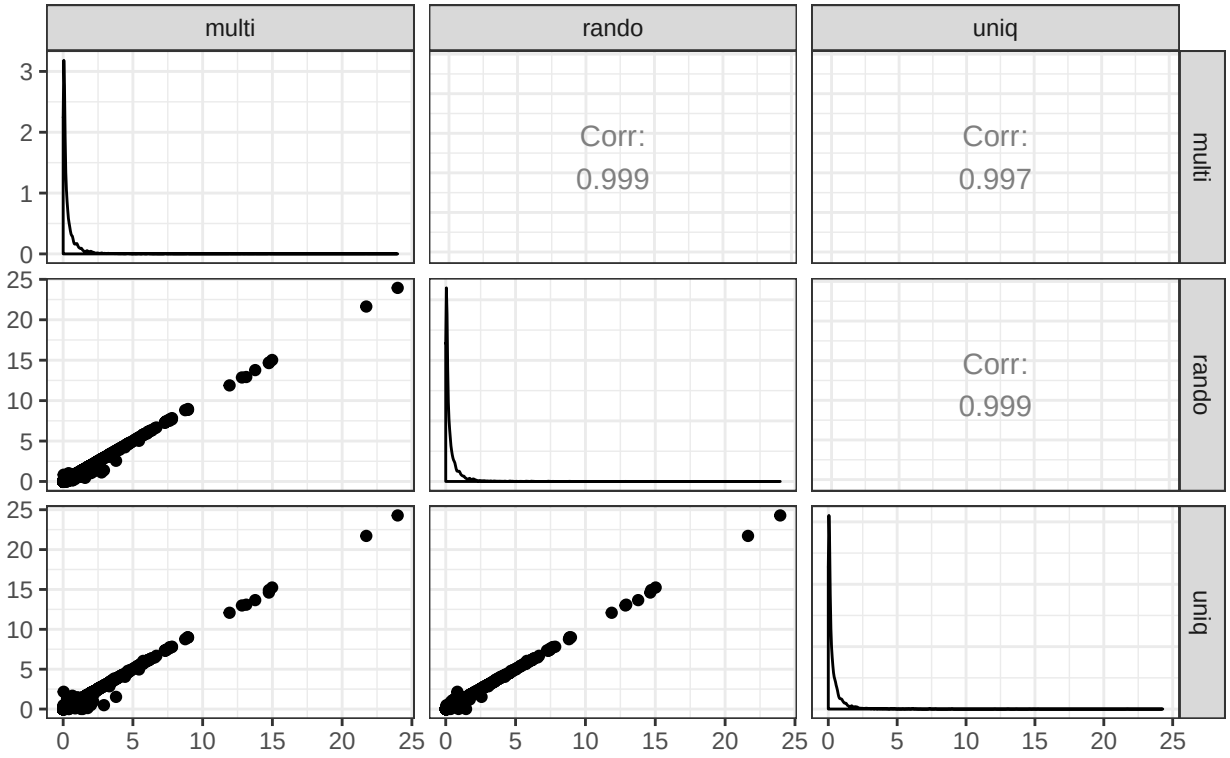



```
## pdf
##   2
```

??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.)

Figure 18. Agreement on (shrunk) Significance (−log10 adjusted p) Between Alignment Strategies (isolated vs group−housed wildtypes)

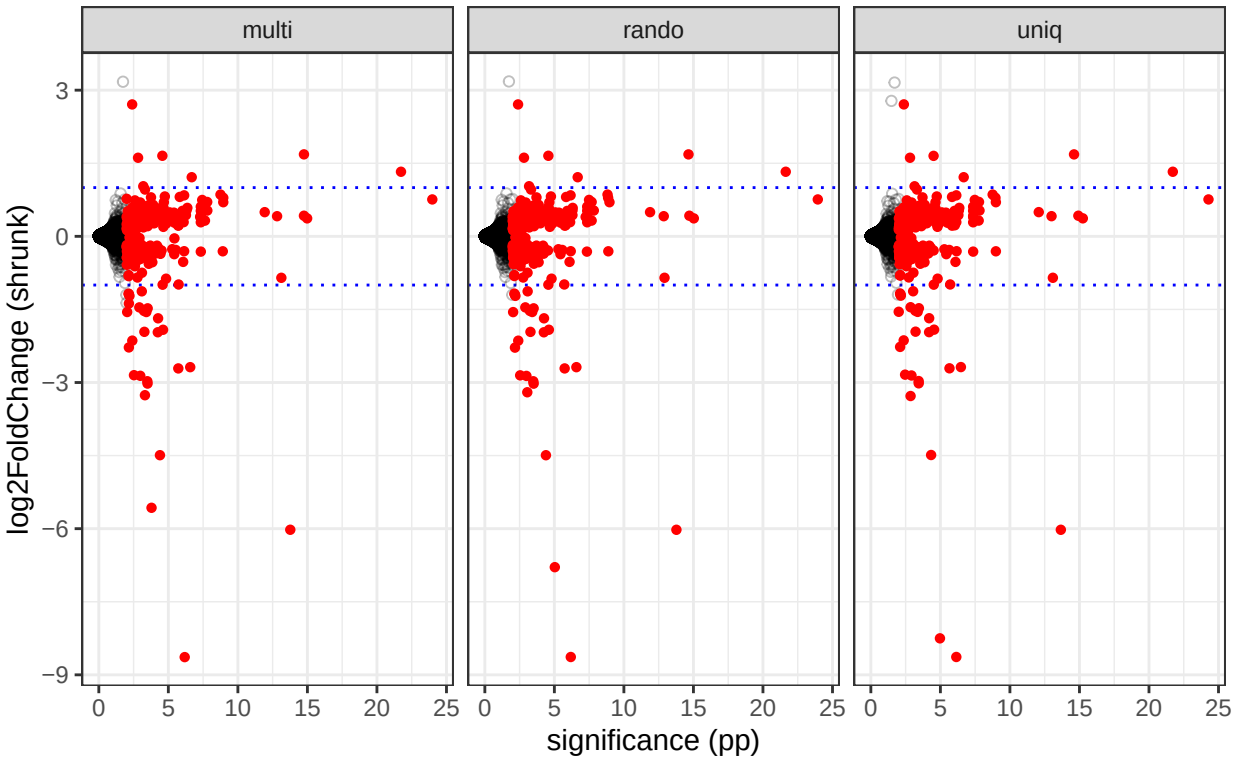



```
## pdf
##   2
```

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 (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 19. 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, padj < 0.01) changes. There were 35 such genes, mostly shared across alignment strategy:

Table 20. Genes with Large (2< fold change), Significant (padj < 0.01) Changes
between isolated and group-housed wildtypes

	multi	rando	uniq
MtnB	yes	yes	yes
TotA	yes	yes	yes
CG11852	yes	yes	yes
TotC	yes	yes	yes
Amy-p	no	yes	yes
Amy-d	yes	no	no
CG7470	yes	yes	yes
Prat2	yes	yes	yes
CG15144	yes	yes	yes
amd	yes	yes	yes
Muc68D	yes	yes	yes
LUBEL	yes	yes	yes
PPO2	yes	yes	yes
CG2736	yes	yes	yes
CG15822	yes	yes	yes

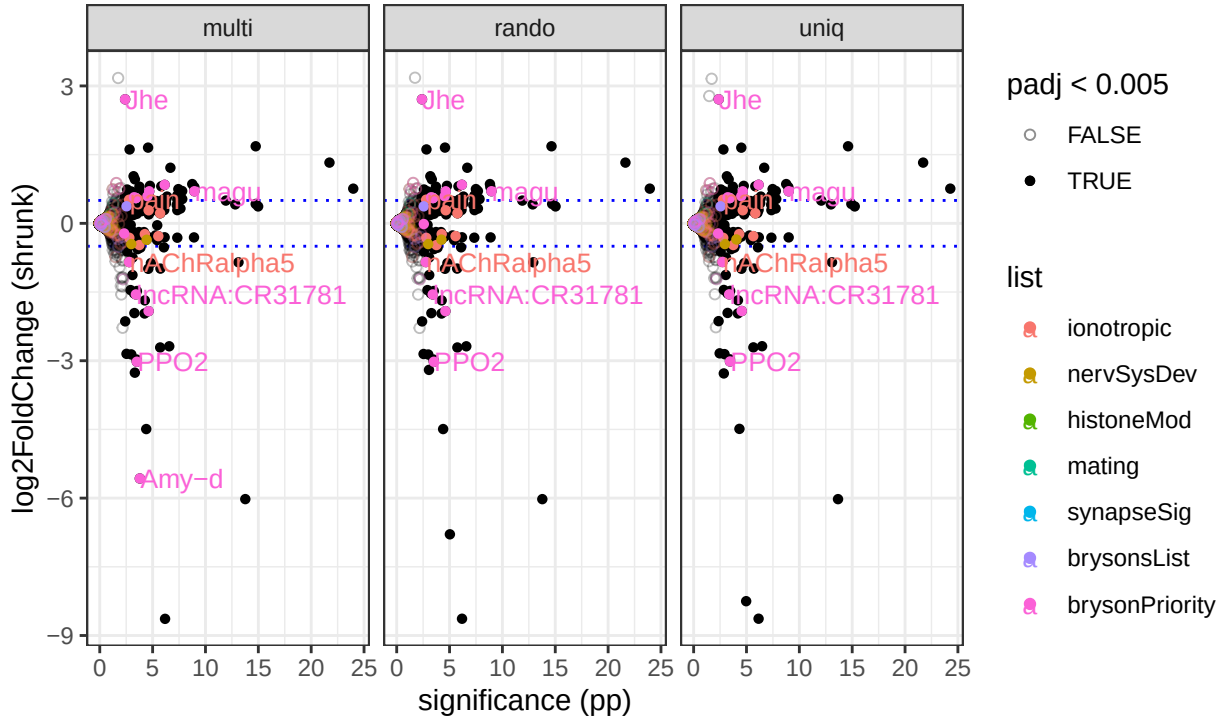
CG42369	yes	yes	yes
hgo	yes	yes	yes
Mlp60A	yes	yes	yes
Npc2g	yes	yes	yes
CG8745	yes	yes	yes
Jhe	yes	yes	yes
CG15067	yes	yes	yes
lncRNA:CR31781	yes	yes	yes
CG9572	yes	yes	yes
Lst	yes	yes	yes
CG16926	yes	yes	yes
CG31324	yes	yes	yes
lectin-28C	yes	yes	yes
Cpr64Ac	yes	yes	yes
CG11400	yes	yes	yes
Mal-A5	yes	yes	yes
CG32820	yes	no	no
Gbp2	yes	yes	yes
CG34220	yes	yes	yes
CG1146	yes	yes	yes

3.1.5 In relation to gene lists

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

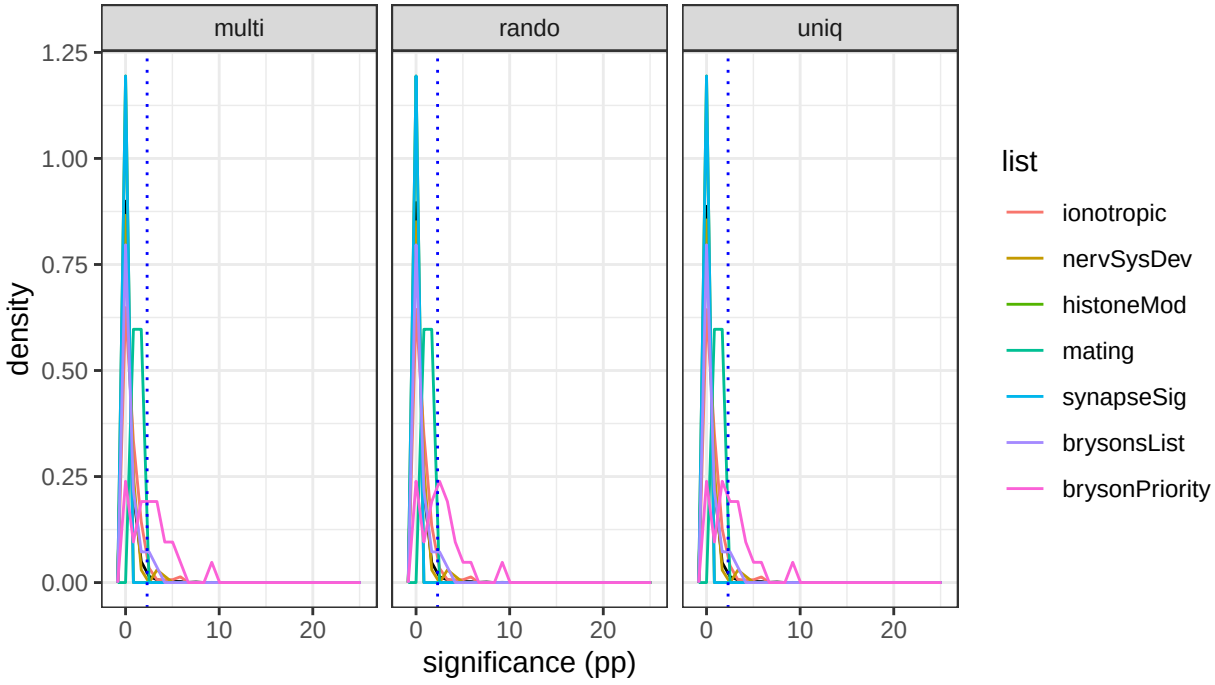
```
## pdf
##   2
```

Figure 20. Volcano Plot: Fold Change vs. Significance with Gene Lists (between isolated and group−housed wildtypes)
abs(lfc)<0.5 & adjusted p < 0.005 highlighted




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

Figure 21. p−value Distribution: Distribution of Fold−Change Significance in General and in Genes of Greatest inTerest (between isolated and group−housed wildtypes) adjusted p < 0.005 highlighted




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

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:

Table 21. Top Ten Most Significantly (padj<0.01) Differentially Expres
between isolated and group-housed wildtypes

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted
1	CG10050	0.41	0.757	1.05×10^{-24}	CG10050	0.41	0.757	1.14×10^{-2}
2	MtnB	0.85	1.326	1.86×10^{-22}	MtnB	0.85	1.326	2.30×10^{-2}
3	CG14687	3.39	0.369	1.05×10^{-15}	CG14687	3.39	0.369	9.42×10^{-1}
4	CG31663	0.47	0.423	1.77×10^{-15}	CG31663	0.47	0.423	1.96×10^{-1}
5	CG11852	0.18	1.682	1.77×10^{-15}	CG11852	0.18	1.682	2.27×10^{-1}
6	TotA	0.13	−6.022	1.71×10^{-14}	TotA	0.13	−6.023	1.67×10^{-1}
7	Dhc36C	0.03	−0.851	7.51×10^{-14}	Dhc36C	0.03	−0.851	1.20×10^{-1}
8	Cln3	0.87	0.413	1.52×10^{-13}	Cln3	0.87	0.413	1.37×10^{-1}
9	Obp84a	0.93	0.496	1.18×10^{-12}	Obp84a	0.93	0.497	1.32×10^{-1}
10	magu	0.30	0.700	1.12×10^{-9}	magu	0.30	0.701	$1.09 \times 10^{-}$

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):

Table 22. Top Ten Largest Magnitude Fold Changes which were Signifi
between isolated and group-housed wildtypes

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted
1	TotC	0.05	−8.636	6.67×10^{-7}	TotC	0.05	−8.634	$6.61 \times 10^{-}$
2	TotA	0.13	−6.022	1.71×10^{-14}	Amy-p	0.02	−6.790	$9.30 \times 10^{-}$
3	Amy-d	0.02	−5.572	1.62×10^{-4}	TotA	0.13	−6.023	1.67×10^{-1}
4	Prat2	0.01	−4.492	4.02×10^{-5}	Prat2	0.01	−4.493	$4.05 \times 10^{-}$
5	Muc68D	0.01	−3.261	4.81×10^{-4}	Muc68D	0.01	−3.200	$8.85 \times 10^{-}$
6	PPO2	0.03	−3.022	3.14×10^{-4}	PPO2	0.03	−3.023	$3.10 \times 10^{-}$
7	CG2736	0.01	−2.974	3.26×10^{-4}	CG2736	0.01	−2.975	$3.29 \times 10^{-}$
8	Mlp60A	0.01	−2.863	1.04×10^{-3}	Mlp60A	0.01	−2.865	$1.03 \times 10^{-}$
9	Npc2g	0.04	−2.850	2.96×10^{-3}	Npc2g	0.04	−2.853	$2.88 \times 10^{-}$
10	CG15144	0.01	−2.711	1.92×10^{-6}	CG15144	0.01	−2.711	$1.82 \times 10^{-}$

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:

Table 23. Top Ten Highest Expressed Genes with Significant (padj <
Difference
between isolated and group-housed wildtypes

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	Obp28a	32.47	0.311	4.22×10^{-4}	Obp28a	32.47	0.312	3.81×10^{-4}
2	a5	27.93	0.257	2.19×10^{-4}	a5	27.94	0.263	1.94×10^{-4}
3	CG9691	14.83	0.178	9.73×10^{-3}	CG9691	14.83	0.178	9.66×10^{-3}
4	CG11550	11.60	0.308	1.59×10^{-3}	CG11550	11.60	0.307	1.54×10^{-3}
5	Obp59a	9.74	0.253	1.66×10^{-3}	Obp59a	9.74	0.253	1.60×10^{-3}
6	RpL41	9.54	0.217	1.52×10^{-3}	RpL41	9.54	0.218	1.49×10^{-3}
7	Cyt-b5	7.25	0.188	3.24×10^{-3}	Cyt-b5	7.25	0.188	3.17×10^{-3}
8	vir-1	6.81	0.227	1.82×10^{-6}	vir-1	6.81	0.228	1.55×10^{-6}
9	Ldsdh1	6.80	0.245	3.25×10^{-3}	Ldsdh1	6.80	0.246	3.16×10^{-3}
10	PHGPx	6.01	0.194	2.30×10^{-5}	PHGPx	6.01	0.195	2.12×10^{-5}

3.1.9 rank-correllation between alignment strategies

3.1.10 Compare to Gene Lists?

3.1.11 Gene Ontology?

3.2 Group Housed: Wildtype vs Mutants

3.2.1 wt vs OR47b

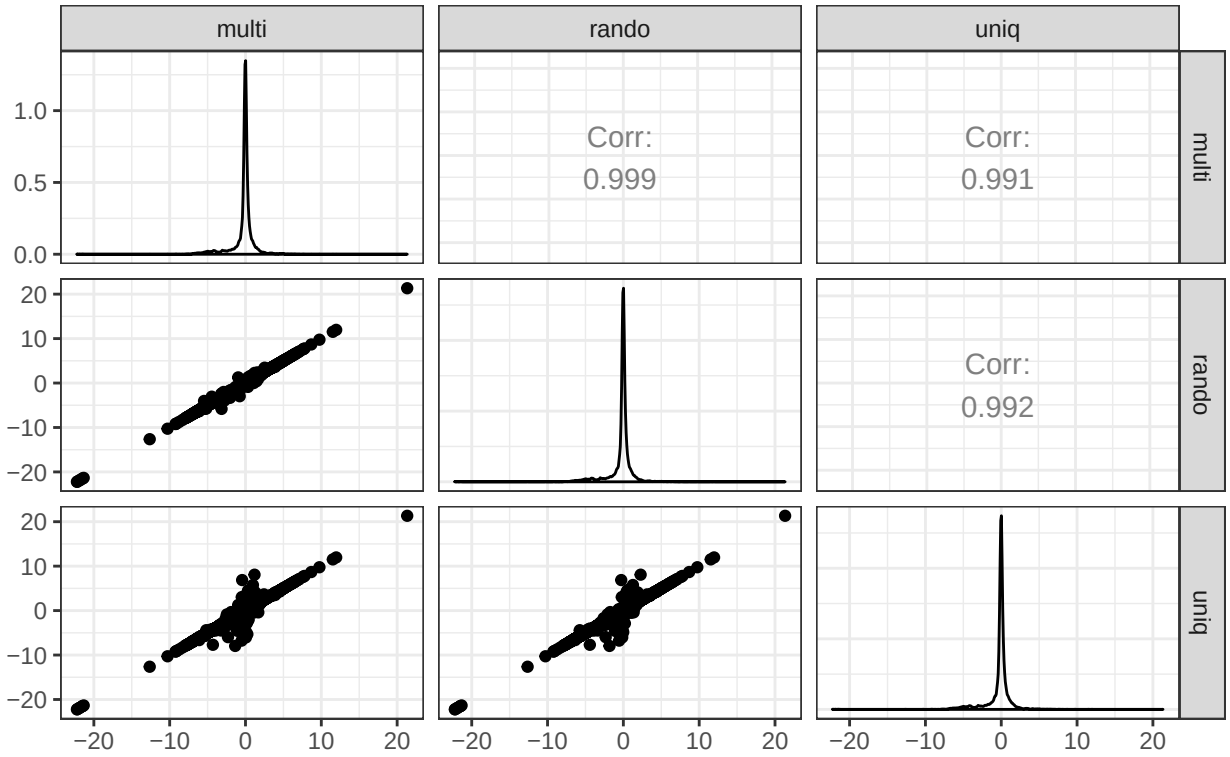
After filtering to remove genes with too few reads for analysis, about $12.2k$ of $17.7k$ annotated genes (68.5937529 %) remain available for testing:

Table 25. Number Genes with Sufficient Read Count for Differential Expression Analysis (group wt vs 47b mutant)
from 17747 annotations in dm6_genes

	tested	percent
multi	$12.3K$	69.3%
rando	$12.2K$	68.5%
uniq	$12.1K$	68.0%

3.2.1.1 preshrunk comparison across alignment strategies

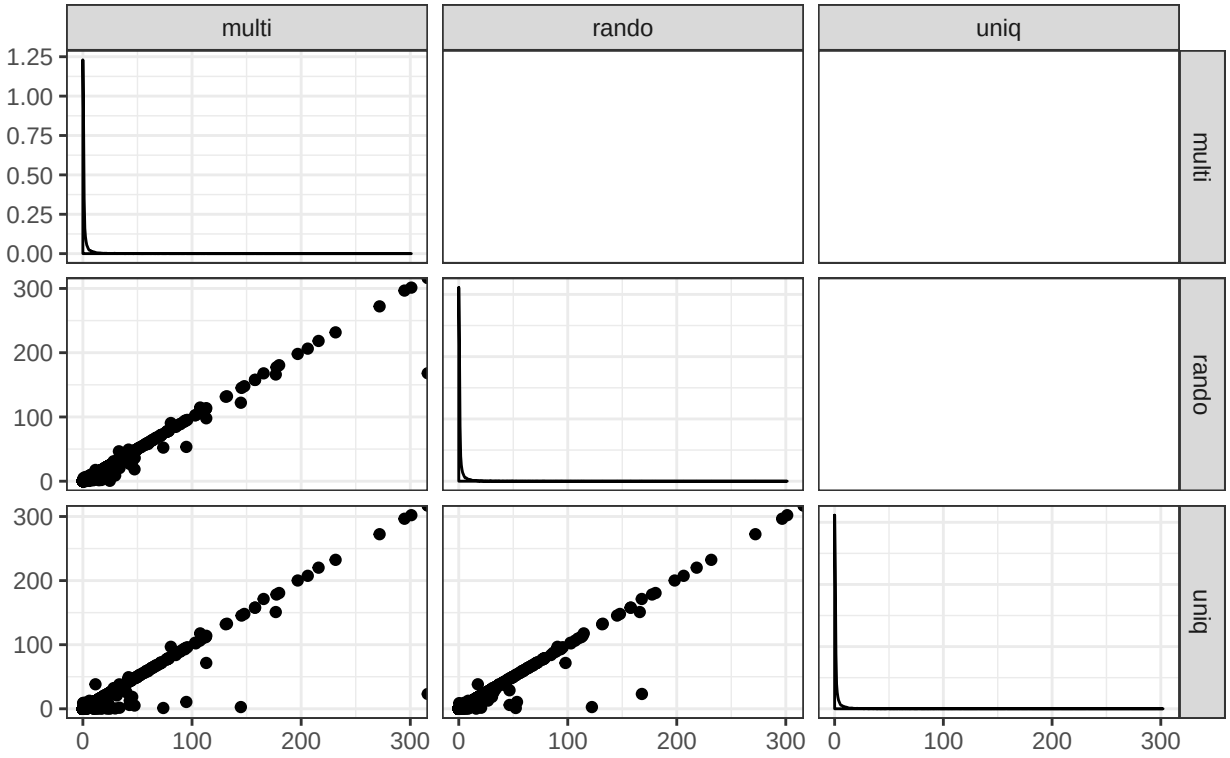
Figure 22. Agreement on (unshrunk) Effect Size (log2 fold change) Between Alignment Strategies (group−housed wildtypes vs 47b mutants)




```
## pdf
```

```
##   2
```

Figure 23. Agreement on (shrunk) Significance (−log10 adjusted p) Between Alignment Strategies (group−housed wildtypes vs 47b mutants)

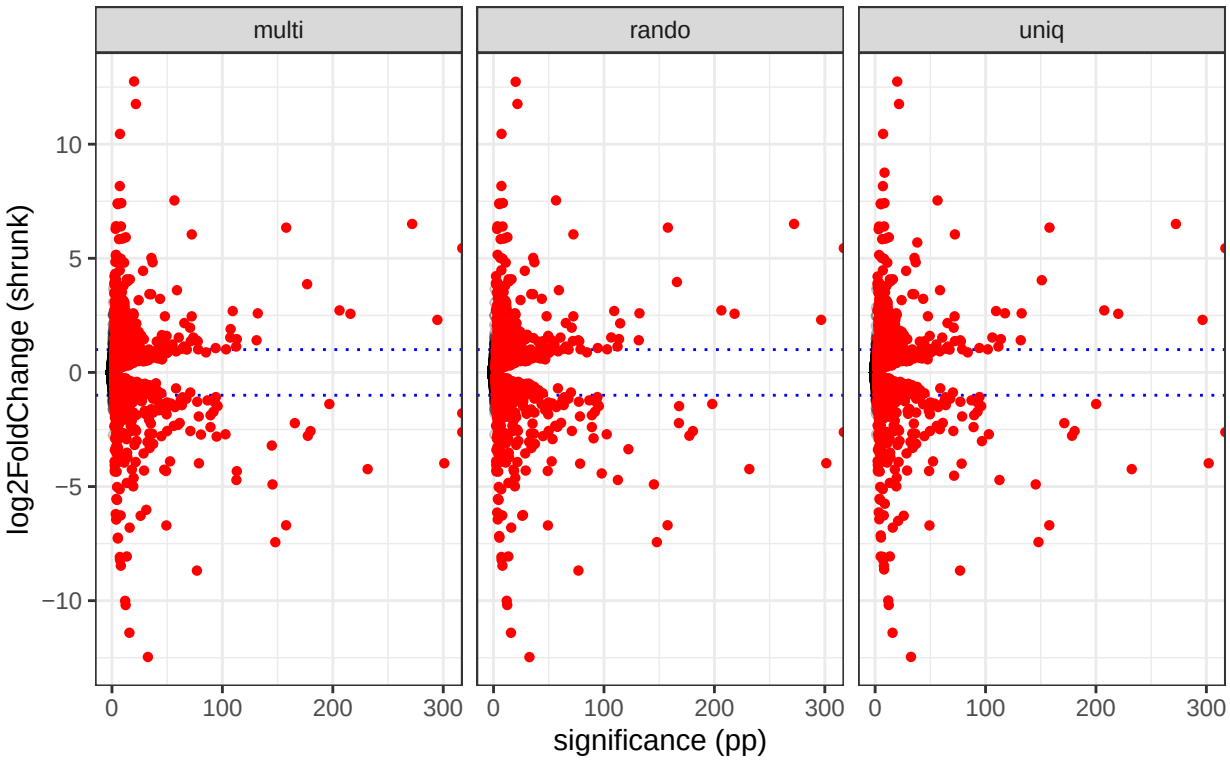



```
## pdf
##   2
```

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 (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 24. Volcano Plot: Fold Change vs. Significance
(between group−housed wildtypes and 47b mutants)




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

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, padj < 0.01) changes. There were 524 such genes, mostly shared across alignment strategy: (see supplementary tables folder, *results/tables/supp/grpWtVs47b$_c$honky.html*)

3.2.1.3 Genes with top 10 most significant changes

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

Table 25. Top Ten Most Significantly (padj<0.01) Differentially Exp
between group-housed wildtypes and 47b mutants

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjus
1	Unc-115a	0.83	−1.787	0.00	CG7900	0.97	5.444	
2	CG7900	0.97	5.443	0.00	Drip	1.15	−2.614	
3	Drip	1.15	−2.616	0.00	Idgf2	0.44	−3.978	4.96×1
4	Idgf2	0.44	−3.980	1.20×10^{-301}	Cyp9b2	2.73	2.304	2.07×1
5	Cyp9b2	2.73	2.302	2.19×10^{-295}	CG6912	0.66	6.506	7.24×1
6	CG6912	0.66	6.506	1.44×10^{-272}	DIP-alpha	0.09	−4.236	2.31×1
7	DIP-alpha	0.09	−4.237	2.96×10^{-232}	Cyp6a2	6.47	2.570	5.29×1

8	Cyp6a2	6.47	2.568	1.12×10^{-216}	Osi8	1.19	2.718	4.71×1
9	Osi8	1.19	2.716	9.11×10^{-207}	Tina-1	3.82	−1.383	8.65×1
10	Tina-1	3.82	−1.385	1.41×10^{-197}	Ugt86Dd	0.88	−2.570	3.09×1

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:

Table 26. Top Ten Largest Magnitude Fold Changes which v
between group-housed wildtypes and 47b mutants

	multi				rando		
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChan
1	mthl8	0.07	12.750	9.44×10^{-21}	mthl8	0.07	12.
2	CG40486	2.45	−12.467	2.89×10^{-33}	CG40486	2.45	−12.
3	w	0.27	11.760	2.27×10^{-22}	w	0.27	11.
4	CG30428	0.11	−11.406	1.62×10^{-16}	CG30428	0.10	−11.
5	CG43149	0.08	10.449	5.80×10^{-8}	CG43149	0.08	10.
6	ppk19	0.04	−10.192	4.25×10^{-13}	ppk19	0.04	−10.
7	lncRNA:CR45502	0.07	−10.007	1.18×10^{-12}	lncRNA:CR45502	0.07	−10.
8	Cyp6a17	0.46	−8.683	1.18×10^{-77}	Cyp6a17	0.46	−8.
9	asRNA:CR44030	0.03	−8.468	9.16×10^{-9}	asRNA:CR44030	0.03	−8.
10	lncRNA:CR44377	0.01	−8.214	5.13×10^{-8}	lncRNA:CR44377	0.01	−8.

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:

Table 27. Top Ten Highest Expressed Genes with Significant (padj
Difference
between group-housed wildtypes and 47b mutants

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted
1	Obp83b	59.65	0.303	1.62×10^{-4}	Obp83b	59.65	0.306	$1.01 \times 10^{-}$
2	Obp19d	52.17	0.344	1.11×10^{-3}	Obp19d	52.18	0.351	$6.44 \times 10^{-}$
3	Obp83a	50.66	0.407	1.72×10^{-8}	Obp83a	50.67	0.408	$3.12 \times 10^{-}$
4	Obp28a	32.78	0.507	8.00×10^{-20}	Obp28a	32.78	0.497	1.71×10^{-2}
5	OS9	32.70	0.272	5.87×10^{-5}	OS9	32.70	0.273	$3.68 \times 10^{-}$
6	Obp19a	29.14	0.203	2.70×10^{-4}	Obp19a	29.15	0.205	$1.91 \times 10^{-}$
7	Obp69a	22.72	0.302	2.18×10^{-4}	Obp69a	22.72	0.304	$1.78 \times 10^{-}$
8	GstE4	22.43	0.191	1.10×10^{-3}	GstE4	22.43	0.194	$8.09 \times 10^{-}$
9	Ugt35b	20.33	0.575	1.50×10^{-7}	Ugt35b	20.34	0.571	$7.49 \times 10^{-}$
10	CG11391	19.38	0.298	3.82×10^{-3}	CG11391	19.38	0.303	$2.88 \times 10^{-}$

3.2.2 wt vs 67d

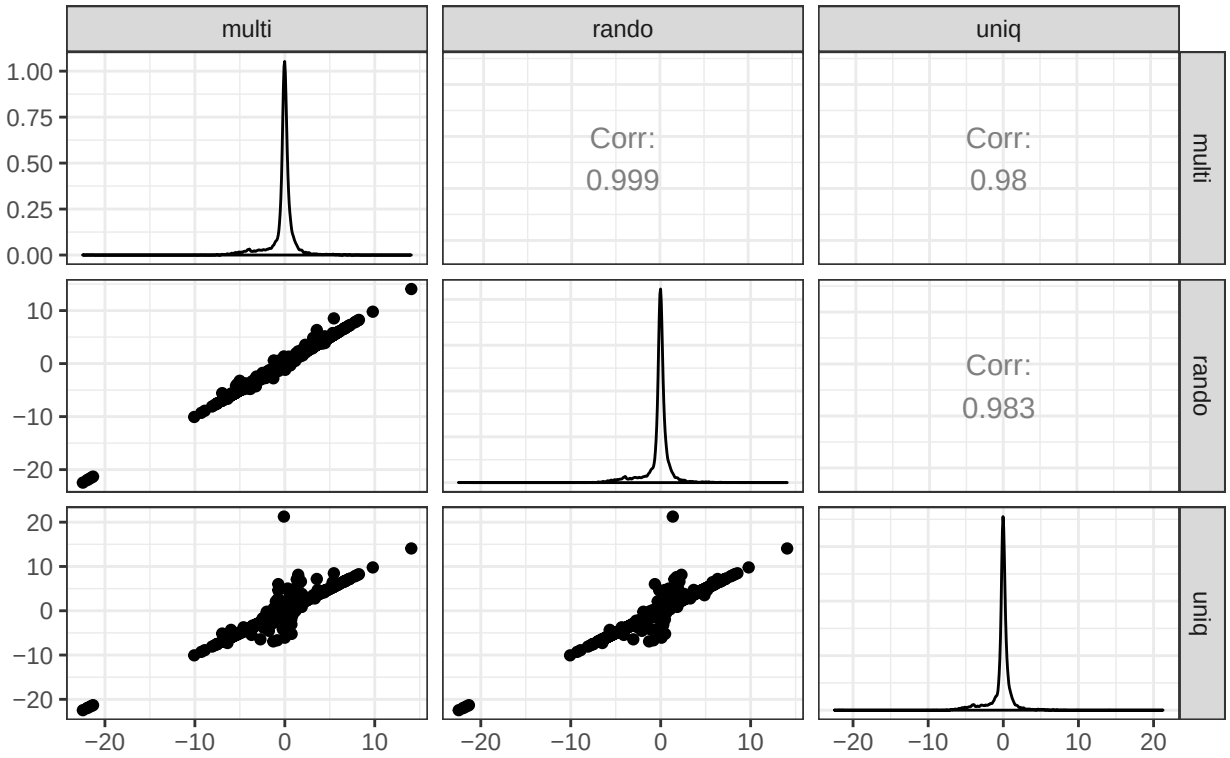
After filtering to remove genes with too few reads for analysis, about $12.1k$ of $17.7k$ annotated genes (68.1711463 %) remain available for testing:

Table 29. Number Genes with Sufficient Read Count for Differential Expression Analysis (group wt vs 67d mutant)
from 17747 annotations in dm6_genes

	tested	percent
multi	$12.2K$	68.8%
rando	$12.1K$	68.1%
uniq	$12.0K$	67.6%

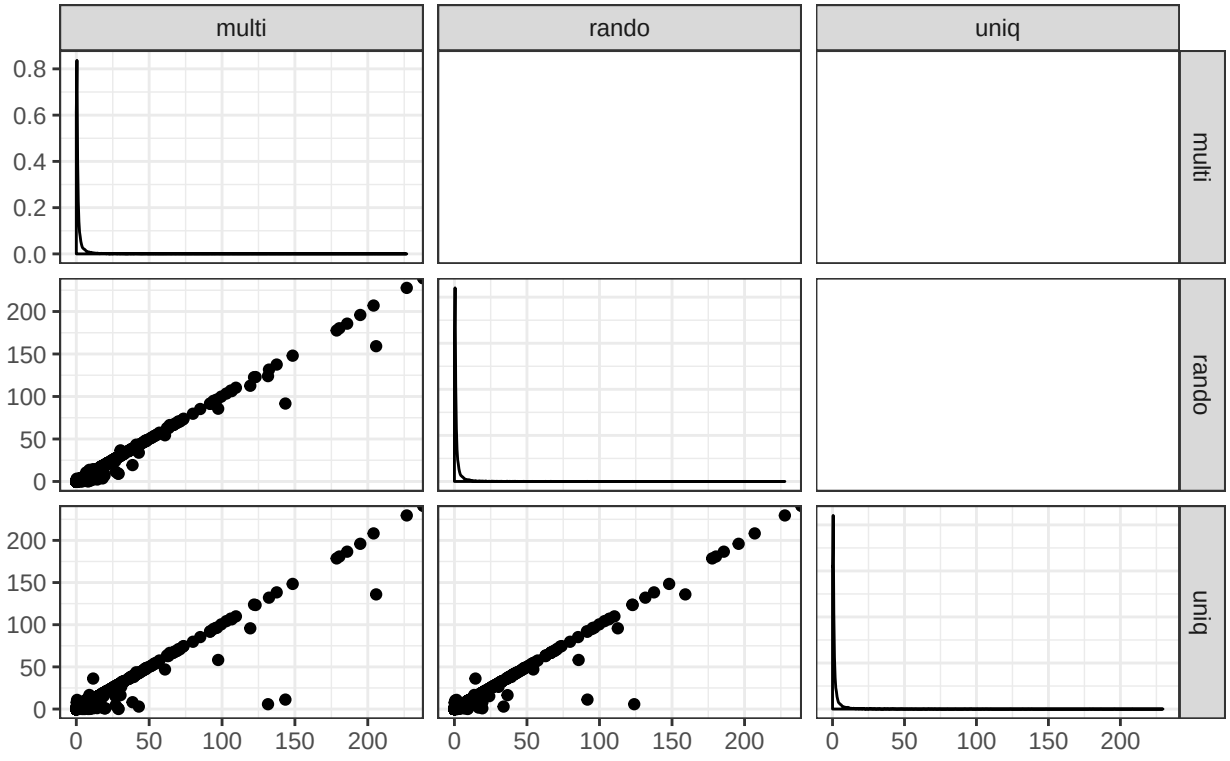
3.2.2.1 preshrunk comparison across alignment strategies

Figure 25. Agreement on (unshrunk) Effect Size (log2 fold change) Between Alignment Strategies (group−housed wildtypes vs 67d mutants)




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

Figure 26. Agreement on (shrunk) Significance (−log10 adjusted p) Between Alignment Strategies (group−housed wildtypes vs 67d mutants)

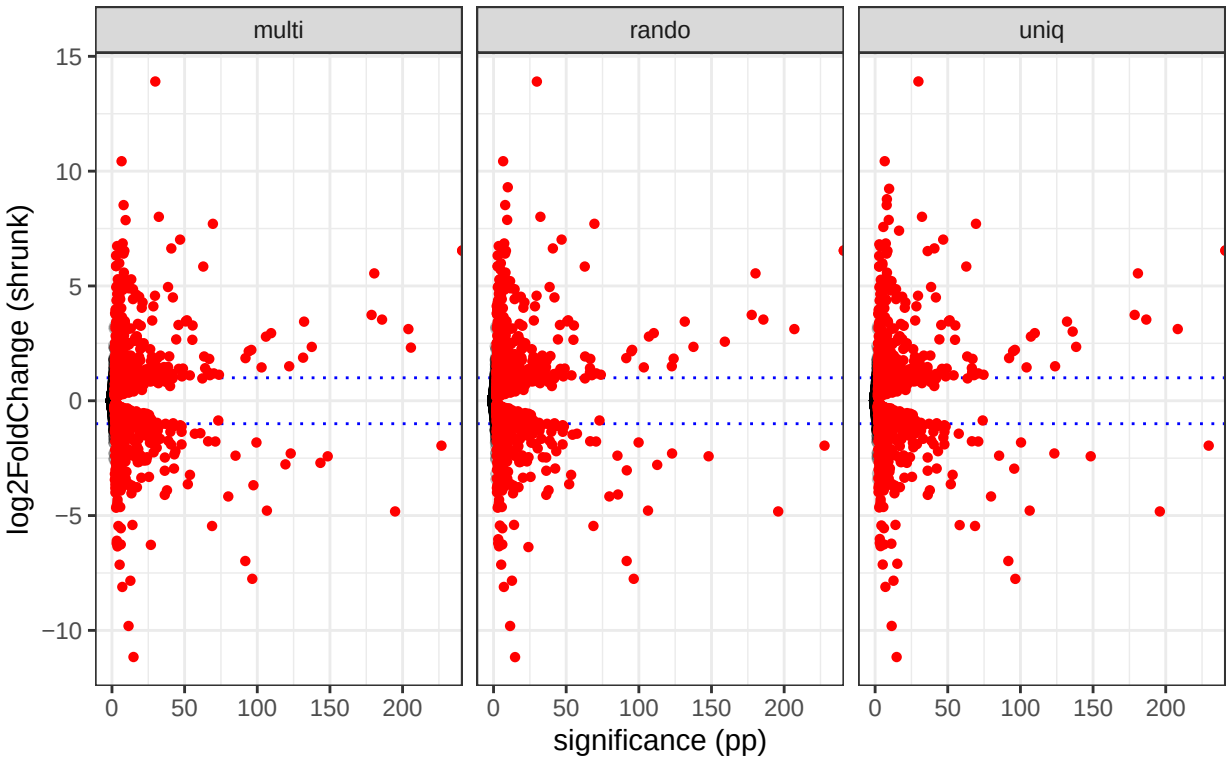



```
## pdf
##   2
```

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 (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 27. Volcano Plot: Fold Change vs. Significance
(between group−housed wildtypes and 67d mutants)




```
## 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, padj < 0.01) changes. There were 553 such genes, mostly shared across alignment strategy: (see tables folder, $results/tables/grpWtVs67d_c honky.html$)

3.2.2.3 Genes with top 10 most significant changes

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

Table 29. Top Ten Most Significantly (padj<0.01) Differentially Exp
between group-housed wildtypes and 67d mutants

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjus
1	CG7900	2.03	6.549	0.00	CG7900	2.03	6.550	
2	l(2)03659	0.43	6.531	0.00	l(2)03659	0.43	6.532	
3	NijC	1.26	−1.956	1.78×10^{-227}	NijC	1.26	−1.956	1.66×1
4	CG32641	3.78	2.315	1.83×10^{-206}	Cyp9b1	0.86	3.122	9.73×1
5	Cyp9b1	0.86	3.121	1.02×10^{-204}	DIP-alpha	0.09	−4.820	1.27×1
6	DIP-alpha	0.09	−4.821	1.43×10^{-195}	CG10936	0.11	3.536	2.00×1
7	CG10936	0.11	3.535	1.40×10^{-186}	CG6912	0.33	5.547	5.20×1

8	CG6912	0.33	5.547	3.07×10^{-181}	ppk25	0.31	3.739	2.08×1
9	ppk25	0.31	3.738	2.29×10^{-179}	CG32641	2.19	2.572	6.72×1
10	CG9447	1.54	−2.419	3.43×10^{-149}	CG9447	1.54	−2.419	9.81×1

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):

Table 30. Top Ten Largest Magnitude Fold Changes which v
between group-housed wildtypes and 47b mutants

	multi				rando		
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChan
1	w	1.13	13.905	1.56×10^{-30}	w	1.13	13.9
2	CG32437	0.05	−11.160	1.51×10^{-15}	CG32437	0.05	−11.1
3	CG43149	0.08	10.432	2.36×10^{-7}	CG43149	0.08	10.4
4	lncRNA:CR44111	0.07	−9.809	3.88×10^{-12}	lncRNA:CR44111	0.07	−9.8
5	ppk9	0.01	8.522	9.82×10^{-9}	CG43291	0.01	9.2
6	lncRNA:CR44377	0.01	−8.109	7.50×10^{-8}	ppk9	0.01	8.5
7	lncRNA:CR45923	0.15	8.011	5.58×10^{-33}	lncRNA:CR44377	0.01	−8.1
8	Obp83g	0.08	7.866	4.14×10^{-10}	lncRNA:CR45923	0.15	8.0
9	CG9010	0.08	−7.839	2.08×10^{-13}	Obp83g	0.08	7.8
10	5-HT2A	0.13	−7.756	2.80×10^{-97}	CG9010	0.08	−7.8

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):

Table 31. Top Ten Highest Expressed Genes with Significant (pa
Difference
between group-housed wildtypes and 67d mutants

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adj
1	Obp83b	60.24	0.394	2.32×10^{-3}	Obp83b	60.24	0.391	2.08
2	Obp83a	51.34	0.501	3.55×10^{-5}	Obp83a	51.34	0.503	2.83
3	Obp69a	25.91	0.696	9.33×10^{-15}	Obp69a	25.91	0.697	6.43 ×
4	lncRNA:noe	18.84	0.366	1.76×10^{-3}	lncRNA:noe	18.85	0.363	1.56
5	Drsl5	17.04	−0.403	2.19×10^{-5}	Drsl5	17.04	−0.402	2.16
6	GstE4	16.49	−0.641	4.48×10^{-6}	GstE4	16.49	−0.642	4.19
7	EbpIII	14.56	0.355	8.79×10^{-3}	EbpIII	14.57	0.357	8.23
8	Cyp6w1	13.12	0.597	1.68×10^{-7}	Cyp6w1	13.12	0.597	1.26
9	lush	13.09	0.794	1.09×10^{-13}	lush	13.09	0.795	8.60 ×
10	Snmp1	10.72	−0.380	1.37×10^{-4}	Snmp1	10.72	−0.380	1.29

3.2.3 wt vs FruLexaFru440

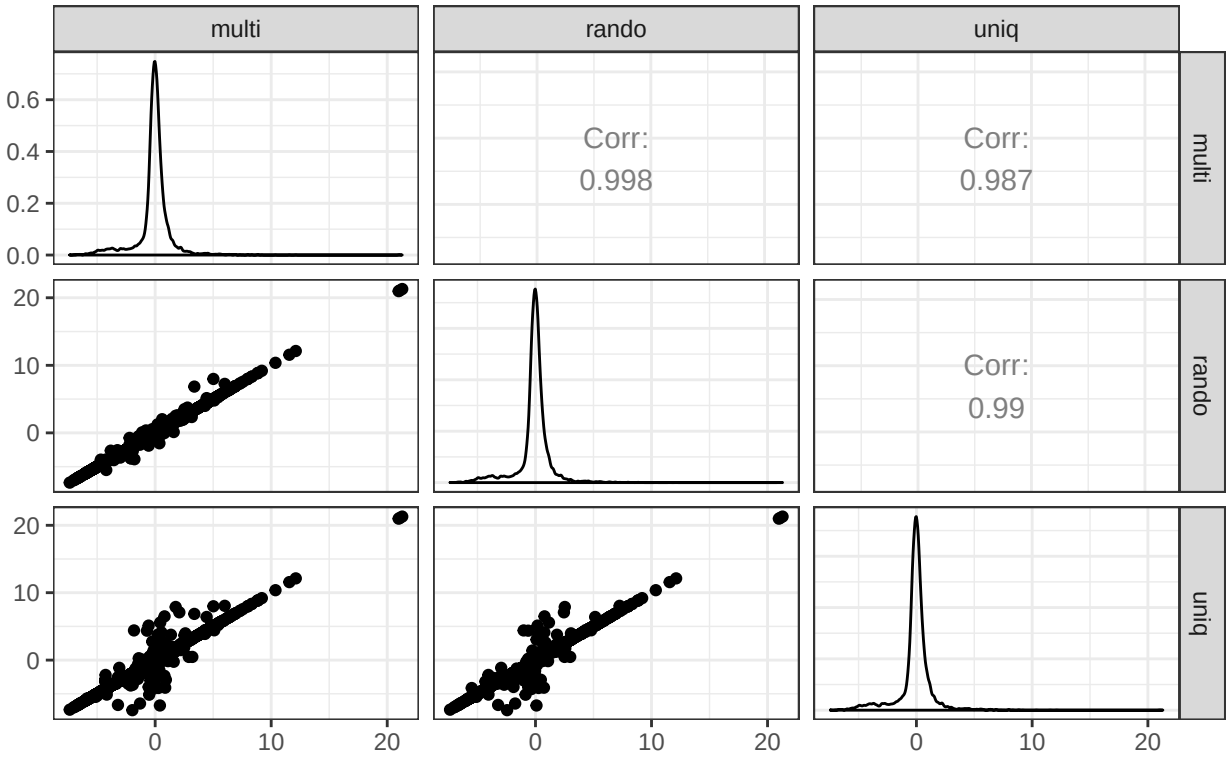
After filtering to remove genes with too few reads for analysis, about $12.2k$ of $17.7k$ annotated genes (68.914934 %) remain available for testing:

Table 33. Number Genes with Sufficient Read Count for Differential Expression Analysis (group wt vs Fru mutant)
from 17747 annotations in dm6_genes

	tested	percent
multi	$12.4K$	69.6%
rando	$12.2K$	68.8%
uniq	$12.1K$	68.3%

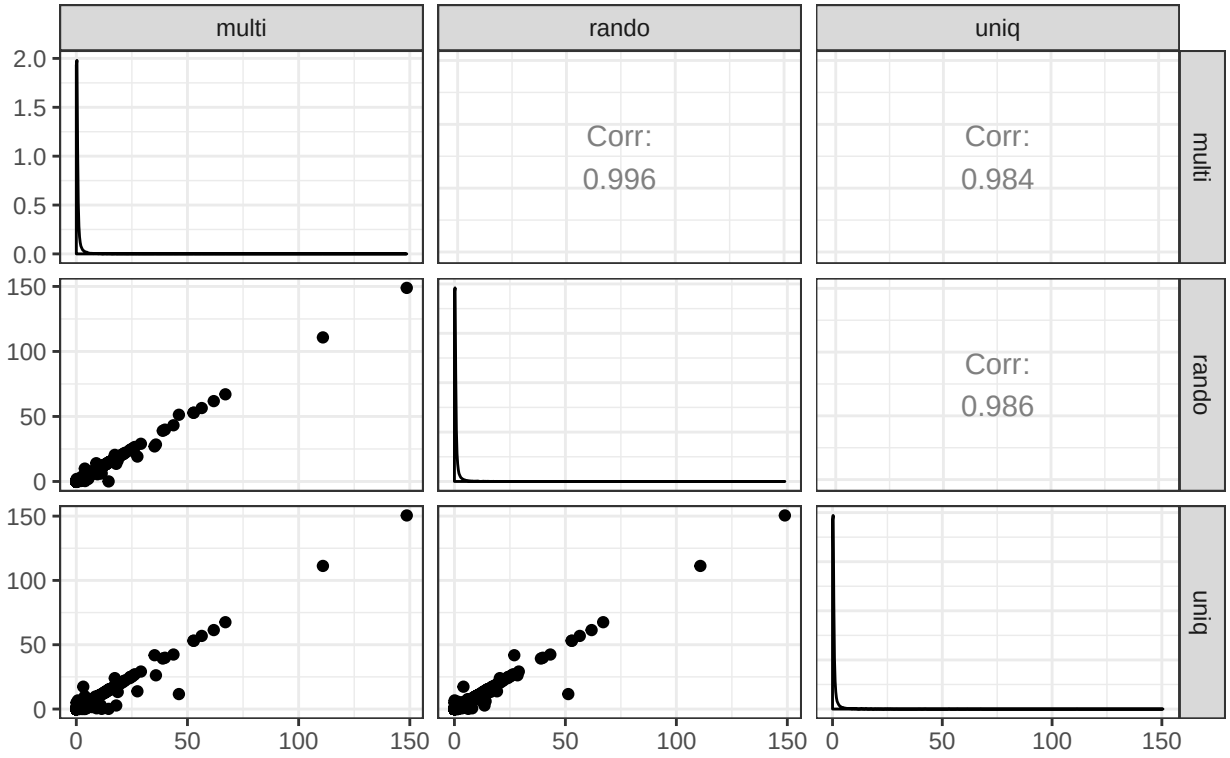
3.2.3.1 preshrunk comparison across alignment strategies

Figure 28. Agreement on (unshrunk) Effect Size (log2 fold change) Between Alignment Strategies (group−housed wildtypes vs Fru mutants)




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

Figure 29. Agreement on (shrunk) Significance (−log10 adjusted p) Between Alignment Strategies (group−housed wildtypes vs Fru mutants)

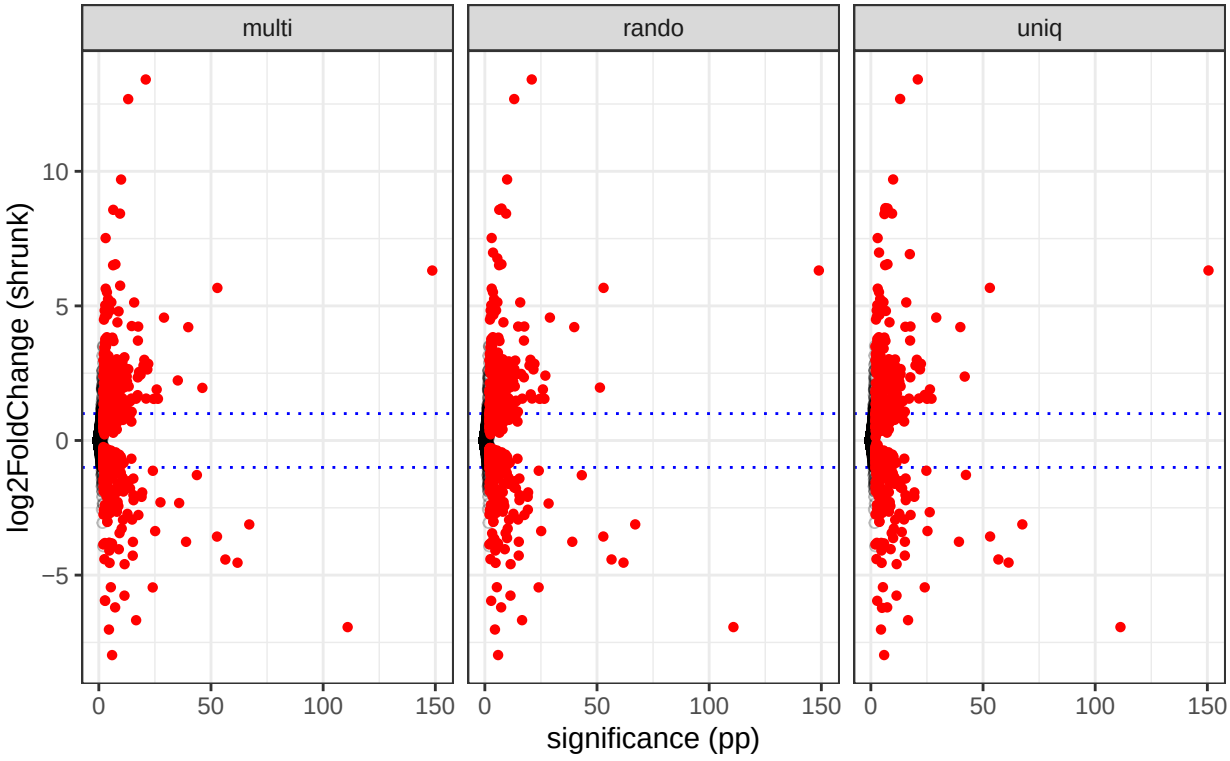



```
## pdf
##   2
```

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 (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 30. Volcano Plot: Fold Change vs. Significance
(between group−housed wildtypes and 67d mutants)




```
## 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, 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:

Table 33. Top Ten Most Significantly (padj<0.01) Differentially Exp
between group-housed wildtypes and Fru mutants

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjus
1	CG7900	1.64	6.314	2.33×10^{-149}	CG7900	1.64	6.316	1.30×1
2	5-HT2A	0.12	−6.933	1.21×10^{-111}	5-HT2A	0.12	−6.931	1.57×1
3	Ets21C	0.06	−3.118	8.71×10^{-68}	Ets21C	0.06	−3.117	$9.36 \times$
4	DIP-alpha	0.08	−4.540	1.49×10^{-62}	DIP-alpha	0.08	−4.538	$1.56 \times$
5	IM23	0.45	−4.421	3.95×10^{-57}	IM23	0.45	−4.420	$3.37 \times$
6	CG11893	0.32	5.668	1.44×10^{-53}	CG11893	0.32	5.669	$1.21 \times$
7	IM1	0.90	−3.569	2.08×10^{-53}	IM1	0.90	−3.567	$1.54 \times$

8	CG32640	2.71	1.955	7.07×10^{-47}	CG32640	1.91	1.966	$4.98 \times$
9	Cyp6a20	5.86	−1.283	2.00×10^{-44}	Cyp6a20	5.86	−1.288	$6.04 \times$
10	CG42526	0.09	4.210	1.35×10^{-40}	CG42526	0.09	4.212	$1.39 \times$

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:

Table 34. Top Ten Largest Magnitude Fold Changes which
between group-housed wildtypes and Fru mutants

	multi				rando		
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldCha
1	mthl8	0.10	13.411	1.20×10^{-21}	mthl8	0.10	13
2	CG43149	0.26	12.682	8.13×10^{-14}	CG43149	0.26	12
3	CG9287	0.02	9.694	1.20×10^{-10}	CG9287	0.02	9
4	ppk27	0.01	8.568	3.68×10^{-7}	CG43291	0.01	8
5	w	0.03	8.428	3.61×10^{-10}	ppk27	0.01	8
6	lncRNA:CR44377	0.01	−7.973	1.23×10^{-6}	w	0.03	8
7	lncRNA:CR44285	0.06	7.520	9.13×10^{-4}	lncRNA:CR44377	0.01	−7
8	CG43919	0.02	−7.024	3.09×10^{-5}	lncRNA:CR44285	0.06	7
9	5-HT2A	0.12	−6.933	1.21×10^{-111}	CG43919	0.02	−7
10	CG32437	0.05	−6.676	2.38×10^{-17}	lncRNA:CR46123	0.02	6

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.

Table 36. Top Ten Highest Expressed Genes with Significant (pac
Difference
between group-housed wildtypes and Fru mutants

	multi				rando			
rank	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjus
1	Obp19d	48.59	0.396	2.85×10^{-3}	Obp19d	48.61	0.397	$2.86 \times$
2	Obp69a	22.58	0.477	5.23×10^{-3}	Obp69a	22.59	0.478	$5.22 \times$
3	Obp56d	10.02	1.121	7.74×10^{-4}	Obp56d	10.03	1.123	$7.69 \times$
4	CG6908	7.53	0.505	1.43×10^{-3}	CG6908	7.53	0.506	$1.36 \times$
5	CG9449	6.51	0.288	3.68×10^{-7}	CG9449	6.51	0.289	$3.15 \times$
6	Cyp6a20	5.86	−1.283	2.00×10^{-44}	Cyp6a20	5.86	−1.288	$6.04 \times$
7	CG5973	4.91	0.565	1.74×10^{-5}	CG5973	4.91	0.567	$1.89 \times$
8	CG8369	4.46	0.716	9.56×10^{-3}	CG8369	4.46	0.717	$9.43 \times$
9	Vha55	4.32	−0.283	2.92×10^{-3}	Vha55	4.32	−0.283	$2.91 \times$
10	ND-MLRQ	3.94	0.435	3.83×10^{-4}	ND-MLRQ	3.94	0.436	$3.60 \times$

3.3 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. To do this, the differenctial expression data from DESeq2 are joined across pairs of contrasts. For example, the

statistics from the wt-group vs wt-isolation contrast are joined by gene with the statistics from the wt-group vs 67d-group contrast. The p-values were readjusted with a Bonferroni correction using n=2 to reflect this new comparison. 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).

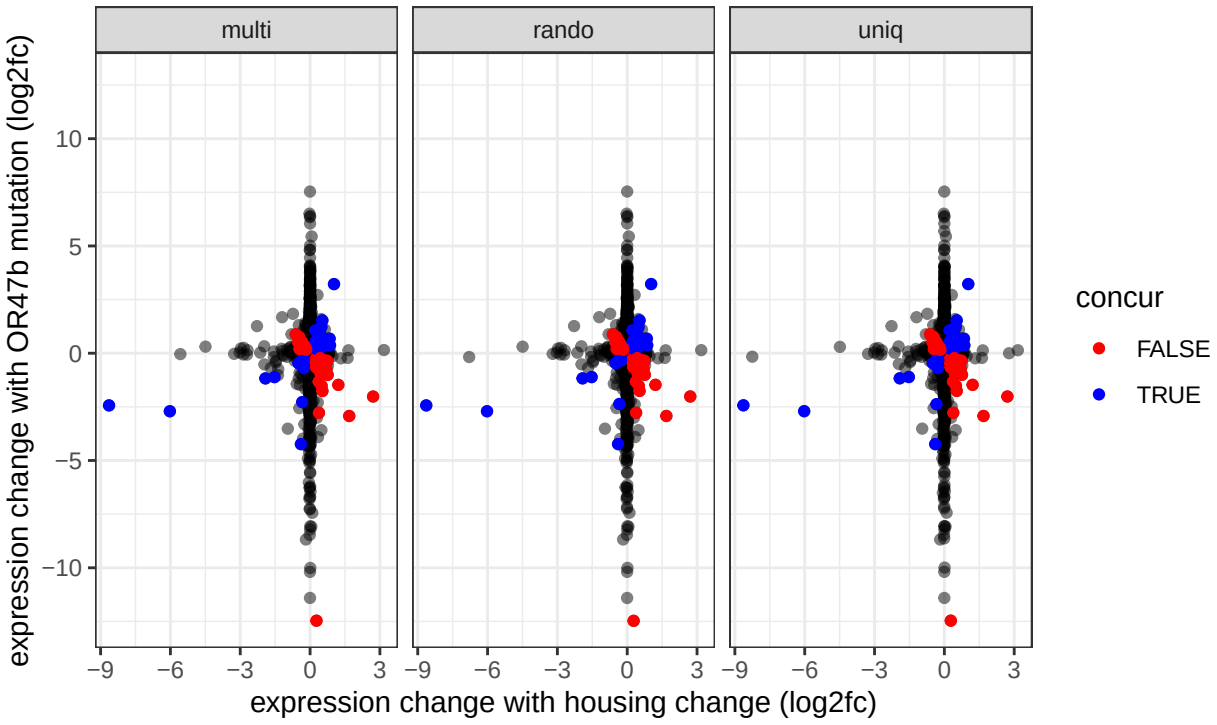
Average significance for gene is currently computed as exp((ln(p1)+ln(p2))/2). (Better to apply stouffer's?)

look at NAs in fulljoin (gene dropout may be interesting. . .)

3.3.1 Housing & OR47b

Here is a scatterplot of the log2 fold change of the 47b & 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 31. Scatterplot of Expression Changes in OR47b mutants vs Expression Changes in Housing
(Significant Similarities and Differences Highlighted)




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

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

Table 37. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs OR47b

	multi	rando	uniq
Agree	44	43	44
Disagree	49	48	48

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

Table 38. Top Ten Most Significant Gene
in difference expression between housing and OR47b

	multi						
rank	name	mean expression	mean readusted p	housing l2fc	mutation l2fc	name	mean expression
1	DIP-alpha	0.13	6.64×10^{-118}	−0.392	−4.237	DIP-alpha	0.13
2	CG9717	0.89	7.13×10^{-58}	0.530	1.530	CG9717	0.89
3	CG7272	1.80	6.08×10^{-49}	0.224	1.067	CG7272	1.80
4	jv	0.08	1.58×10^{-30}	0.486	1.224	jv	0.08
5	NA	0.23	2.19×10^{-27}	0.414	0.851	NA	0.23
6	Obp59a	10.03	1.72×10^{-24}	0.253	0.571	Obp59a	10.03
7	Cpr64Ac	0.14	8.16×10^{-24}	1.028	3.223	Cpr64Ac	0.14
8	CG3940	0.54	2.08×10^{-19}	0.357	0.708	CG3940	0.54
9	CG9498	4.11	2.69×10^{-13}	0.747	0.578	CG9498	4.11
10	Obp28a	32.62	1.16×10^{-11}	0.311	0.507	Obp28a	32.62

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 39. Top Ten Largest Magnitude Changes In Significant Ge
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	TotC	−5.534	0.05	1.63×10^{-5}	TotC	−5.533	0.05	
2	TotA	−4.363	0.14	1.71×10^{-10}	TotA	−4.363	0.14	
3	DIP-alpha	−2.315	0.13	6.64×10^{-118}	DIP-alpha	−2.314	0.13	5
4	Cpr64Ac	2.126	0.14	8.16×10^{-24}	Cpr64Ac	2.127	0.14	
5	LUBEL	−1.545	0.01	5.76×10^{-4}	LUBEL	−1.544	0.01	
6	CG9572	−1.321	0.06	1.40×10^{-3}	Dscam4	−1.351	0.06	
7	Dscam4	−1.306	0.06	1.39×10^{-5}	CG9572	−1.320	0.06	
8	CG9717	1.030	0.89	7.13×10^{-58}	CG9717	1.031	0.89	
9	jv	0.855	0.08	1.58×10^{-30}	jv	0.856	0.08	
10	dmGlut	0.759	0.31	5.48×10^{-9}	dmGlut	0.760	0.31	

Of those mutually significant genes with different directions of change, the top 10 most significant agree well across alignment strategy. (“NA” is trol, “terribly reduced optic lobes”, FBgn0267911/FBgn0284408)

Table 40. Top Ten Most Significant Genes of
in difference expression between housing and OR47b contr

	multi							mea
rank	name	mean expression	mean readusted p	housing l2fc	OR47b l2fc	name	mean expression	
1	CG14400	1.41	2.20×10^{-91}	0.382	−2.775	CG14400	1.41	
2	amd	1.21	1.35×10^{-51}	1.213	−1.475	amd	1.21	
3	SPARC	3.27	4.65×10^{-50}	0.369	−1.324	SPARC	3.27	
4	Obp84a	0.71	6.19×10^{-35}	0.496	−1.518	Obp84a	0.71	
5	CG10050	0.32	7.13×10^{-27}	0.757	−1.008	CG10050	0.32	
6	MtnA	0.85	3.15×10^{-22}	0.530	−1.760	MtnA	0.85	
7	CG40486	4.12	4.78×10^{-19}	0.276	−12.467	CG40486	4.12	
8	Loxl2	0.48	2.58×10^{-17}	0.449	−0.957	Jheh3	1.27	
9	Jheh3	1.27	2.91×10^{-17}	0.291	−0.600	Loxl2	0.48	
10	Fer1	1.20	9.73×10^{-17}	0.263	−0.664	Fer1	1.20	

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, with minor disagreements about their order:

Table 41. Top Ten Most Serious Significant Differences betw
in difference expression between housing and OR47b contrants

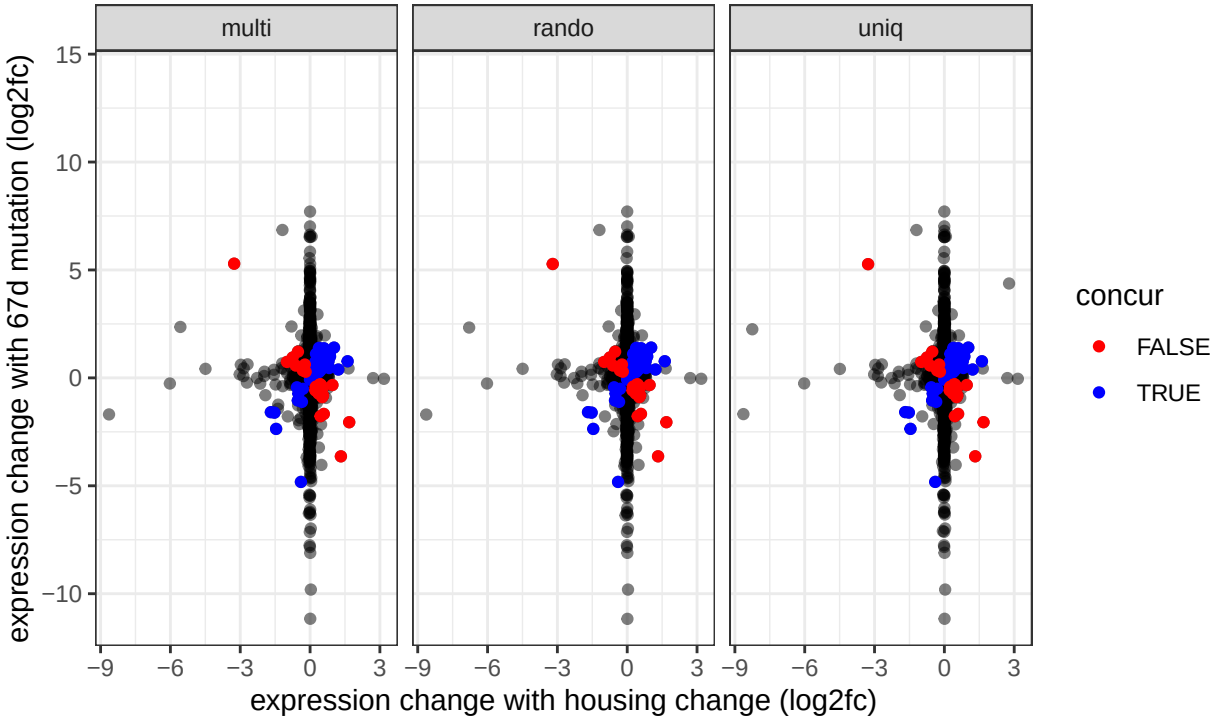
	multi				rando		
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression
1	CG40486	12.743	4.12	4.78×10^{-19}	CG40486	12.749	4.12
2	Jhe	4.725	0.35	1.94×10^{-5}	Jhe	4.724	0.35
3	CG11852	4.607	0.11	1.89×10^{-14}	CG11852	4.606	0.11
4	CG14400	3.157	1.41	2.20×10^{-91}	CG14400	3.155	1.41
5	amd	2.688	1.21	1.35×10^{-51}	amd	2.686	1.21
6	MtnA	2.290	0.85	3.15×10^{-22}	MtnA	2.288	0.85
7	Obp84a	2.014	0.71	6.19×10^{-35}	Obp84a	2.013	0.71
8	CG10050	1.765	0.32	7.13×10^{-27}	CG10050	1.763	0.32
9	SPARC	1.693	3.27	4.65×10^{-50}	SPARC	1.691	3.27
10	Spn47C	−1.494	0.04	8.34×10^{-7}	Spn47C	−1.495	0.04

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast$_a$nd$_4$7bContrast.multi.ts*
results/tables/supp/housingContrast$_a$nd$_4$7bContrast.rando.tsv results/tables/supp/housingContrast$_a$nd$_4$7bContrast.un

3.3.2 Housing & 67d

Here is a scatterplot of the log2 fold change of the 67d & 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 32. 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 fewer have the same direction of change as not:

Table 42. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs 67d

	multi	rando	uniq
Agree	44	44	44
Disagree	31	31	31

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

Table 43. Top Ten Most Significant Genes of Ag
in difference expression between housing and 67d contrants

	multi					ran		
rank	name	mean expression	mean readusted p	housing l2fc	67d l2fc	name	mean expression	mean
1	DIP-alpha	0.12	1.46×10^{-99}	−0.392	−4.821	DIP-alpha	0.12	4
2	Pop2	0.96	1.63×10^{-36}	0.255	1.113	Pop2	0.96	

3	CG14400	2.33	1.10×10^{-25}	0.382	1.389	CG14400	2.33
4	dmGlut	0.34	1.50×10^{-19}	0.841	1.142	dmGlut	0.34
5	CG9717	0.80	1.08×10^{-18}	0.530	1.207	CG9717	0.80
6	Cda5	0.09	9.89×10^{-17}	−0.524	−1.047	Cda5	0.09
7	CG31288	1.36	1.47×10^{-16}	0.857	0.990	CG31288	1.36
8	CG13659	0.28	2.37×10^{-16}	0.597	1.372	CG13659	0.28
9	jv	0.07	3.28×10^{-16}	0.486	1.008	jv	0.07
10	CG7272	1.64	5.85×10^{-12}	0.224	0.737	CG7272	1.64

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, with differences in the placement of Amy-d and Amy-p and the inclusion of CG13332.

Table 44. Top Ten Largest Magnitude Changes In Significant Ge
in difference expression between housing and 67d contrants

	multi				rando			
rank	name	mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean
1	DIP-alpha	−2.606	0.12	1.46×10^{-99}	DIP-alpha	−2.606	0.12	4
2	lectin-28C	−1.911	0.02	2.95×10^{-5}	lectin-28C	−1.911	0.02	
3	hgo	−1.636	0.05	1.41×10^{-4}	hgo	−1.636	0.05	
4	CG9572	−1.570	0.06	4.30×10^{-4}	CG9572	−1.570	0.06	
5	Cpr64Ac	1.216	0.07	4.57×10^{-4}	Cpr64Ac	1.216	0.07	
6	CG31324	1.191	0.10	6.19×10^{-4}	CG31324	1.192	0.10	
7	dmGlut	0.991	0.34	1.50×10^{-19}	dmGlut	0.992	0.34	
8	CG13659	0.985	0.28	2.37×10^{-16}	CG13659	0.985	0.28	
9	CG31288	0.923	1.36	1.47×10^{-16}	CG31288	0.924	1.36	
10	CG14400	0.886	2.33	1.10×10^{-25}	CG14400	0.886	2.33	

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

Table 45. Top Ten Most Significant Genes of Disa
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	Loxl2	0.45	6.66×10^{-39}	0.449	−1.775	Loxl2	0.45	1.2
2	MtnB	0.55	2.14×10^{-37}	1.326	−3.635	MtnB	0.55	2.8
3	CG5895	0.71	5.10×10^{-12}	0.342	−0.671	CG5895	0.71	4.8
4	CG11852	0.11	6.41×10^{-12}	1.682	−2.055	CG11852	0.11	7.0
5	CG13937	0.86	2.10×10^{-11}	0.227	−0.582	CG13937	0.86	1.9
6	CG31769	0.26	6.02×10^{-11}	−0.521	1.214	CG31769	0.26	6.6
7	CG11425	0.20	9.45×10^{-11}	0.590	−1.671	CG11425	0.20	1.0
8	CG42237	0.55	7.28×10^{-9}	0.458	−0.522	CG42237	0.55	7.
9	CG13332	0.38	2.02×10^{-8}	0.558	−0.813	CG13332	0.38	2.
10	Cyp9h1	0.18	4.48×10^{-7}	0.522	−0.893	Cyp9h1	0.18	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, with minor disagreements

about their order:

Table 46. Top Ten Most Serious Significant Differences betw
in difference expression between housing and 67d contrants

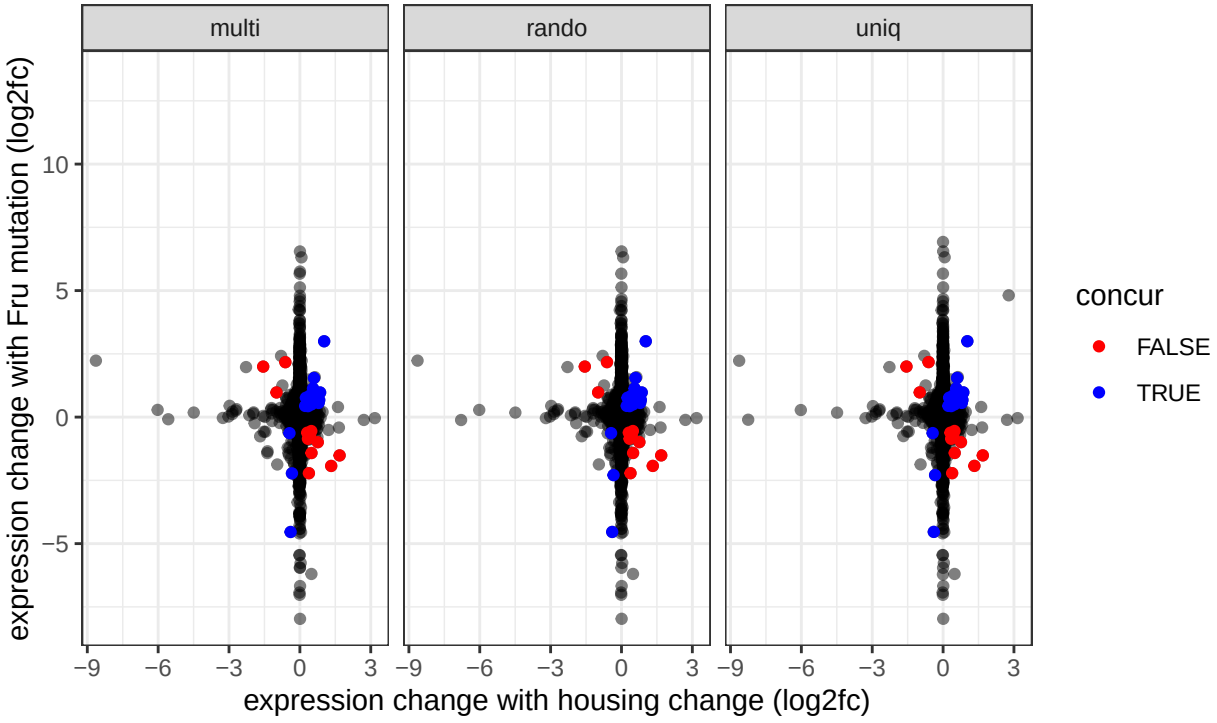
	multi				rando		
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression
1	Muc68D	−8.552	0.21	4.23×10^{-4}	Muc68D	−8.477	0.19
2	MtnB	4.961	0.55	2.14×10^{-37}	MtnB	4.960	0.55
3	CG11852	3.737	0.11	6.41×10^{-12}	CG11852	3.737	0.11
4	CG11425	2.261	0.20	9.45×10^{-11}	CG11425	2.259	0.20
5	Loxl2	2.224	0.45	6.66×10^{-39}	Loxl2	2.224	0.45
6	CG31769	−1.734	0.26	6.02×10^{-11}	CG31769	−1.734	0.26
7	Mal-B2	−1.731	0.06	1.54×10^{-4}	Mal-B2	−1.731	0.06
8	LanA	−1.691	0.02	2.93×10^{-4}	LanA	−1.691	0.02
9	Cyp9h1	1.415	0.18	4.48×10^{-7}	Cyp9h1	1.413	0.18
10	CG13332	1.371	0.38	2.02×10^{-8}	CG13332	1.370	0.38

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast*$_a$*nd*$_6$*7dContrast.multi.ts*
results/tables/supp/housingContrast$_a$*nd*$_6$*7dContrast.rando.tsv results/tables/supp/housingContrast*$_a$*nd*$_6$*7dContrast.un*

3.3.3 Housing & Fru

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 33. Scatterplot of Expression Changes in Fru 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 as not:

Table 47. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs Fru

	multi	rando	uniq
Agree	14	14	14
Disagree	12	12	12

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

Table 48. Top Ten Most Significant Genes of Ag
in difference expression between housing and Fru contrants

	multi							ran
rank	name	mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean
1	DIP-alpha	0.12	4.71×10^{-33}	−0.392	−4.540	DIP-alpha	0.12	
2	CG13659	0.28	1.87×10^{-16}	0.597	1.546	CG13659	0.28	

3	Cpr64Ac	0.12	3.64×10^{-12}	1.028	2.997	Cpr64Ac	0.12
4	CG31288	1.32	9.40×10^{-12}	0.857	0.974	CG31288	1.32
5	Pop2	0.86	3.41×10^{-8}	0.255	0.765	Pop2	0.86
6	CG31272	0.12	3.04×10^{-6}	0.791	0.558	CG31272	0.12
7	CG10178	1.29	7.46×10^{-6}	0.417	0.843	CG10178	1.29
8	CG42806	0.68	8.96×10^{-6}	0.818	0.674	CG42806	0.68
9	CG7272	1.49	2.27×10^{-5}	0.224	0.443	CG7272	1.49
10	Cpr49Ae	0.19	2.70×10^{-5}	0.555	1.131	Cpr49Ae	0.19

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 49. Top Ten Largest Magnitude Changes In Significant Ge
in difference expression between housing and Fru contrants

	multi				rando			
rank	name	mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean
1	DIP-alpha	−2.466	0.12	4.71×10^{-33}	DIP-alpha	−2.465	0.12	4
2	Cpr64Ac	2.013	0.12	3.64×10^{-12}	Cpr64Ac	2.013	0.12	3
3	Dscam4	−1.278	0.06	3.59×10^{-5}	Dscam4	−1.313	0.06	
4	CG13659	1.071	0.28	1.87×10^{-16}	CG13659	1.072	0.28	2
5	CG31288	0.915	1.32	9.40×10^{-12}	CG31288	0.916	1.32	8
6	Cpr49Ae	0.843	0.19	2.70×10^{-5}	Cpr49Ae	0.844	0.19	
7	CG42806	0.746	0.68	8.96×10^{-6}	CG42806	0.747	0.68	
8	CG31272	0.675	0.12	3.04×10^{-6}	CG31272	0.675	0.12	
9	CG10178	0.630	1.29	7.46×10^{-6}	CG10178	0.631	1.29	
10	CG43373	−0.537	0.04	3.81×10^{-4}	CG43373	−0.536	0.04	

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

Table 50. Top Ten Most Significant Genes of Disg
in difference expression between housing and Fru contrants

	multi					rand		
rank	name	mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re
1	MtnB	0.56	5.89×10^{-21}	1.326	−1.930	MtnB	0.56	6.9
2	CG10050	0.30	8.12×10^{-14}	0.757	−0.991	CG10050	0.30	8.5
3	CG14400	1.39	1.94×10^{-10}	0.382	−2.218	CG14400	1.39	2.0
4	CG11852	0.11	1.01×10^{-9}	1.682	−1.514	CG11852	0.11	1.
5	Or92a	1.91	3.74×10^{-8}	0.435	−0.756	Or92a	1.91	4.
6	Spn47C	0.07	6.76×10^{-8}	−0.614	2.171	Spn47C	0.07	6.
7	CG14275	0.49	4.55×10^{-6}	0.491	−1.418	CG14275	0.49	4.
8	CG5895	0.68	8.53×10^{-5}	0.342	−0.858	CG5895	0.68	8.
9	T48	0.24	1.18×10^{-4}	0.483	−0.550	T48	0.24	1.
10	Gbs-70E	0.10	1.80×10^{-4}	−0.989	0.975	Gbs-70E	0.10	1.

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 51. Top Ten Most Serious Significant Differ
in difference expression between housing and Fru contrasts

	multi				rand	
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	m
1	lncRNA:CR31781	−3.549	0.13	2.14×10^{-4}	lncRNA:CR31781	−3.550	
2	MtnB	3.255	0.56	5.89×10^{-21}	MtnB	3.254	
3	CG11852	3.196	0.11	1.01×10^{-9}	CG11852	3.194	
4	Spn47C	−2.785	0.07	6.76×10^{-8}	Spn47C	−2.786	
5	CG14400	2.599	1.39	1.94×10^{-10}	CG14400	2.598	
6	Gbs-70E	−1.963	0.10	1.80×10^{-4}	Gbs-70E	−1.964	
7	CG14275	1.910	0.49	4.55×10^{-6}	CG14275	1.908	
8	CG10050	1.748	0.30	8.12×10^{-14}	CG10050	1.746	
9	CG5895	1.200	0.68	8.53×10^{-5}	CG5895	1.199	
10	Or92a	1.191	1.91	3.74×10^{-8}	Or92a	1.190	

Full data are in the tables folder:

results/tables/supp/housingContrast$_a$*nd*$_F$*ruContrast.multi.tsv results/tables/supp/housingContrast*$_a$*nd*$_F$*ruContrast.ra*
results/tables/supp/housingContrast$_a$*nd*$_F$*ruContrast.uniq.tsv*

3.4 Comparing Expression Changes Between Mutants

do this

3.4.1 Fru & 67d

do this

3.4.2 Fru & 47b

do this

3.4.3 47b & 67d

do this

3.5 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.

The current featureCounts settings ignore ambiguously assigned reads. Because some exons overlap and because junction-spanning reads will be considered ambuiguous 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. (Fix this?)

Table 52. Number of Fru Exons Available For Analysis
(by contrast & aligner)

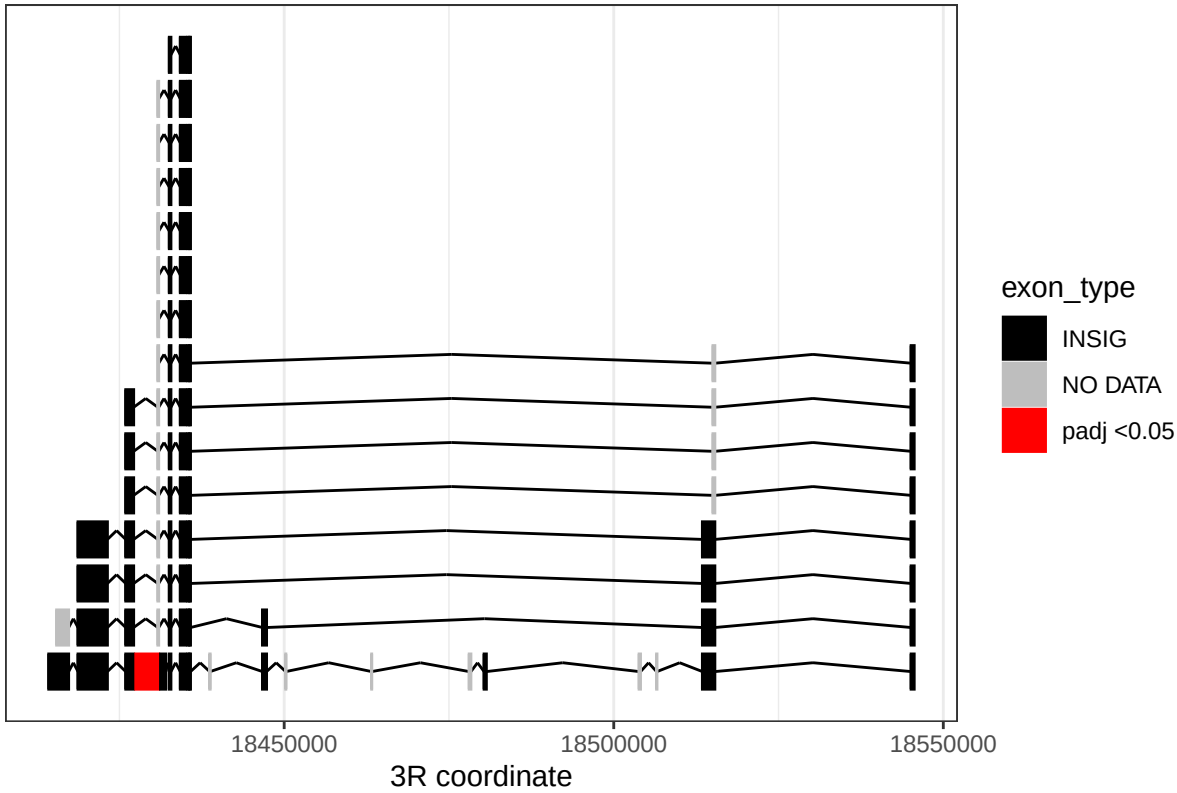
	67d		Fru		47b		SH	
	count	frac	count	frac	count	frac	count	frac
multi	12	54.5%	13	59.1%	12	54.5%	12	54.5%
rando	12	54.5%	13	59.1%	12	54.5%	12	54.5%
uniq	12	54.5%	13	59.1%	12	54.5%	12	54.5%

The only exon with even marginally significant differential expression was exon_5:

Table 53. Fru exons with Significant (p<0.05) Differences, by Contrast
(Multi only)

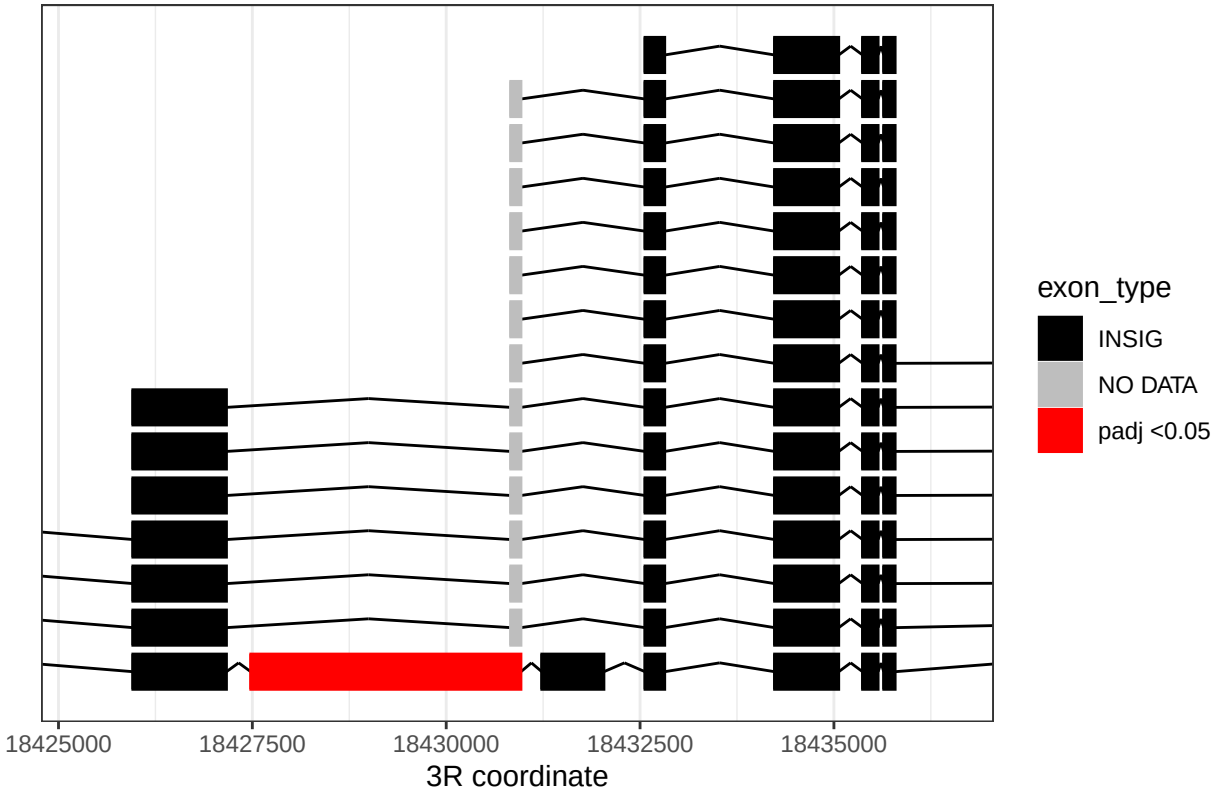
	67d		Fru		47b	
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p
exon_5	1.26	1.85×10^{-4}	1.29	4.32×10^{-2}	0.88	1.57×10^{-3}

Figure 34. Fru gene model, exon5 highlighted




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

Figure 35. Fru gene model (detail)




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

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 preserv
##     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)'.
```

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.