

Volkan Lab Behavioral Genetics RNA-Seq

Charlie Soeder

2021-04-05

Contents

1	Introduction	4
2	Materials, Methods, Data, Software	5
2.1	Reference Genomes	5
2.2	Reference Annotations	6
2.3	Gene Lists	12
2.3.1	Ionotropic	12
2.3.2	Derived from GO terms	12
2.3.3	Bryson's Lists	13
2.4	Sequenced Reads	14
2.4.1	Pre-Processing	15
2.5	Mapped Reads	17
2.5.1	Raw Mapsplice	17
2.5.2	Filtered Multimap	23
2.5.3	Downsampled Multimapped	25
2.5.4	Uniquely Mapped	25
2.5.5	Spliced-Only	26
2.5.6	Alignment Process Overview	26
2.6	Assigning Reads to Annotated Features	28
2.6.1	Fruitless by exon	32
2.6.2	Fruitless by splice junction	34
2.6.3	Fruitless by intron	35
2.6.4	Ambiguous Assignment Strategy Comparison	36
2.7	Expression	37
2.7.1	Focus on Fruitless	41
2.8	Differential Expression Analysis.	41
2.8.1	Differential Exon Use	43

2.8.1.1	The edgeHog	43
2.8.1.2	DEXSeq	43
2.9	Gene Ontology Enrichment	46
3	Results	47
3.1	Wildtype: Group-housed vs. Isolated	47
3.1.1	preshrunk comparison across alignment strategies	47
3.1.2	effect size: preshrunk vs shrunk	47
3.1.3	shrunk comparison across alignment strategies	47
3.1.4	differential expression overview	48
3.1.5	In relation to gene lists	49
3.1.6	Genes with top 10 most significant changes	50
3.1.7	Top 10 genes with biggest (significant) effect sizes	51
3.1.8	Top 10 highest expressed genes with significant change	51
3.1.9	rank-correllation between alignment strategies	52
3.2	Group Housed: Wildtype vs Mutants	52
3.2.1	wt vs OR47b	52
3.2.1.1	preshrunk comparison across alignment strategies	52
3.2.1.2	differential expression overview	52
3.2.1.3	Genes with top 10 most significant changes	53
3.2.1.4	Top 10 genes with biggest (significant) effect sizes	54
3.2.1.5	Top 10 highest expressed genes with significant change	54
3.2.2	wt vs 67d	55
3.2.2.1	preshrunk comparison across alignment strategies	55
3.2.2.2	differential expression overview	55
3.2.2.3	Genes with top 10 most significant changes	56
3.2.2.4	Top 10 genes with biggest (significant) effect sizes	56
3.2.2.5	Top 10 highest expressed genes with significant change	56
3.2.3	wt vs FruLexaFru440	57
3.2.3.1	preshrunk comparison across alignment strategies	57
3.2.3.2	differential expression overview	57
3.2.3.3	Genes with top 10 most significant changes	58
3.2.3.4	Top 10 genes with biggest (significant) effect sizes	59
3.2.3.5	Top 10 highest expressed genes with significant change	59
3.3	Comparing Expression Changes from Housing with Expression Changes from Genotype (Single-Factor Comparison Summary)	60
3.3.1	Housing & OR47b	60

3.3.2	Housing & 67d	63
3.3.3	Housing & Fru	65
3.3.4	Overview	68
3.4	Simultaneously Modeling Housing & Genotype.	70
3.4.1	compare wildTypeHousing results to housing results from hausWtVsMut	71
3.4.2	compare RPKM to baseMean expression	73
3.4.3	Housing Contrast (Simultaneous model)	74
3.4.3.1	Top Tens	79
3.4.3.2	Gene Ontology Enrichment	80
3.4.4	47b vs wt (Simultaneous model)	81
3.4.4.1	Top Tens	82
3.4.4.2	Gene Ontology Enrichment	83
3.4.5	67d vs wt (Simultaneous model)	84
3.4.5.1	Top Tens	85
3.4.5.2	Gene Ontology Enrichment	86
3.4.6	FruLexA/Fru440 vs wt (Simultaneous model)	88
3.4.6.1	Top Tens	89
3.4.6.2	Gene Ontology Enrichment	90
3.4.7	Transcriptional Profiles	92
3.4.7.1	ion channel activity genes	92
3.4.8	Fruitless-less	94
3.4.8.1	Perturbation to Housing Contrast	96
3.4.8.2	Perturbation to 47b1 Contrast	98
3.4.8.3	Perturbation to 67d Contrast	101
3.4.9	Reduced FruLexaFru440 Samples	104
3.4.9.1	Perturbation to Housing Contrast	108
3.4.9.2	Perturbation to 47b1 Contrast	111
3.4.9.3	Perturbation to 67d Contrast	114
3.4.9.4	Perturbation to FruLexa/Fru440 Contrast	117
3.4.9.5	Perturbation to FruLexa/Fru440 Contrast (single-factor)	120
3.5	Comparing Expression Changes from Housing with Expression Changes from Genotype	122
3.5.1	Housing & OR47b	122
3.5.2	Housing & 67d	125
3.5.3	Housing & FruLexFru440	127
3.5.4	Overview (Heatmaps)	129
3.5.4.1	Ion Channel Activity	129

3.6	Comparing Expression Changes Between Mutants	129
3.6.1	Fru & 67d	130
3.6.2	Fru & 47b	130
3.6.3	47b & 67d	130
3.7	Mutually Significant Differential Expression Overview	130
3.8	Focus on Fruitless	132
3.8.1	By Exon	133
3.8.1.1	Ambiguous Read Assignment: None	133
3.8.1.2	Ambiguous Read Assignment: All	136
3.8.2	By Exon Junction	139
3.8.3	By Intron	143
3.8.4	Overall	148
3.8.5	edgeHog	150
3.8.6	DEXSeq	152
3.9	Impact of 47b2 on 88a	158
3.9.1	Top Tens	159
3.9.2	Gene Ontology Enrichment	160
3.10	Two Days' Difference	162
3.10.1	Top Tens	164
3.10.2	Gene Ontology Enrichment	165
3.11	CantonS & amos Mutants Thereof	167
3.11.1	Impact of amos Mutant	167
3.11.1.1	Top Tens	183
3.11.1.2	Gene Ontology Enrichment	184
3.11.2	Differences in “garden Variety” genomes: w vs CantonS	186
3.11.2.1	Top Tens	222
3.11.2.2	Gene Ontology Enrichment	223
3.12	Rotund Mutants: Gene Correlation Networks	225
4	Bibliography	231

1 Introduction

The primary question being investigated here is: how does gene expression change in the fruit fly nervous system, under a) various genome mutations (in OR47b, OR67d, and Fruitless) and b) different developmental environments (raised in groups or in social isolation). Towards this antennal RNASeq has been collected from flies under various conditions.

The basic analysis outline is:

I. Sequenced reads are filtered for quality, mapped to a reference genome, and counted against a reference gene annotation.

- II. Read counts are used to estimate differential gene expression.
 - A. A specific 1-factor model (~ housing) is implemented
 - B. This is expanded to generic 1-factor models (~ condition, condition = housing, genotype...)
 - C. This implementation is generalized further to an arbitrary multi-factor model, which is applied to a 2-factor model (~ housing + genotype)
- III. Estimates for expression were compared across experimental treatments to identify genes of interest with similar behaviors.

Fruitless was examined in particular to check for differential exon use.

Along the way, two exterior concerns arose: that one of the Fru mutant samples were problematic, and that we were not certain of the sex composition of the samples. The former was examined by rerunning the models without it. This produced remarkably little change in the 1-factor models and remarkably large change in the 2-factor models. The latter concern was examined by comparing coverage on the sex chromosomes using published reads from NCBI as controls; this approach was inconclusive.

Late in analysis the base model was reverted from the 2-variable model to a collection of single variable models; the justification for this reversion was unclear:

i had no idea that you had to merge two different experiments to generate the base mean and keep it like that. I also thought that since we have all the count data each experiment should be represented separately. Otherwise what is the purpose of doing the experiment or having the sh being an experimental condition? it makes no biological sense. ... the questions one asks about the data is what constantly changes. to understand and make sense of the biology not the stats
 (P. Volkan, Slack 24 Nov 2020)

In the same sequencing run, samples supported further experiments:

- 1) the impact of the 47b2 mutation on an 88a background
- 2) the impact of two days' development (day 7 vs day 5) on 47b1 mutants

The analysis was expanded by downloading previously published reads, to determine:

- 3) The impact of mutations in Amos, controlling for different background genomes.

2 Materials, Methods, Data, Software

At a top level, the workflow compares sequenced reads to bioinformatic databases, then uses specialized statistical software to analyze the results.

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	138M
number contigs	8

2.2 Reference Annotations

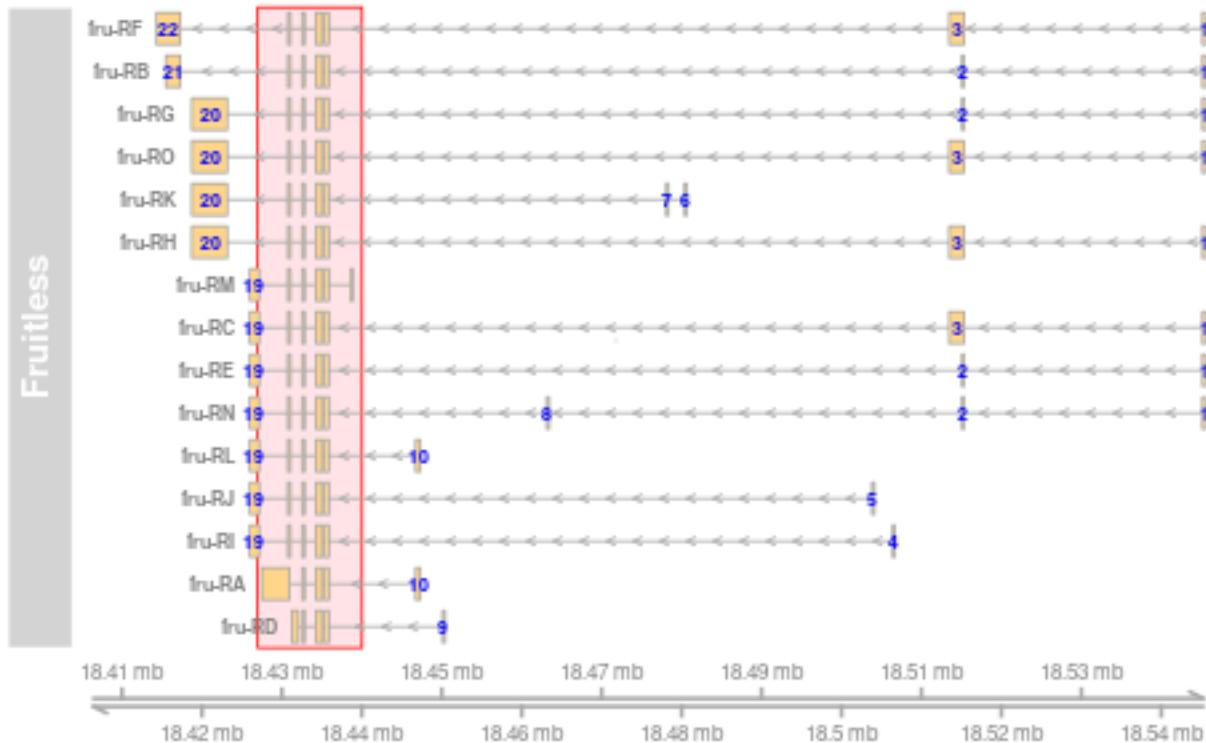
The dm6 reference annotations were used to define gene locii for differential expression analysis:

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

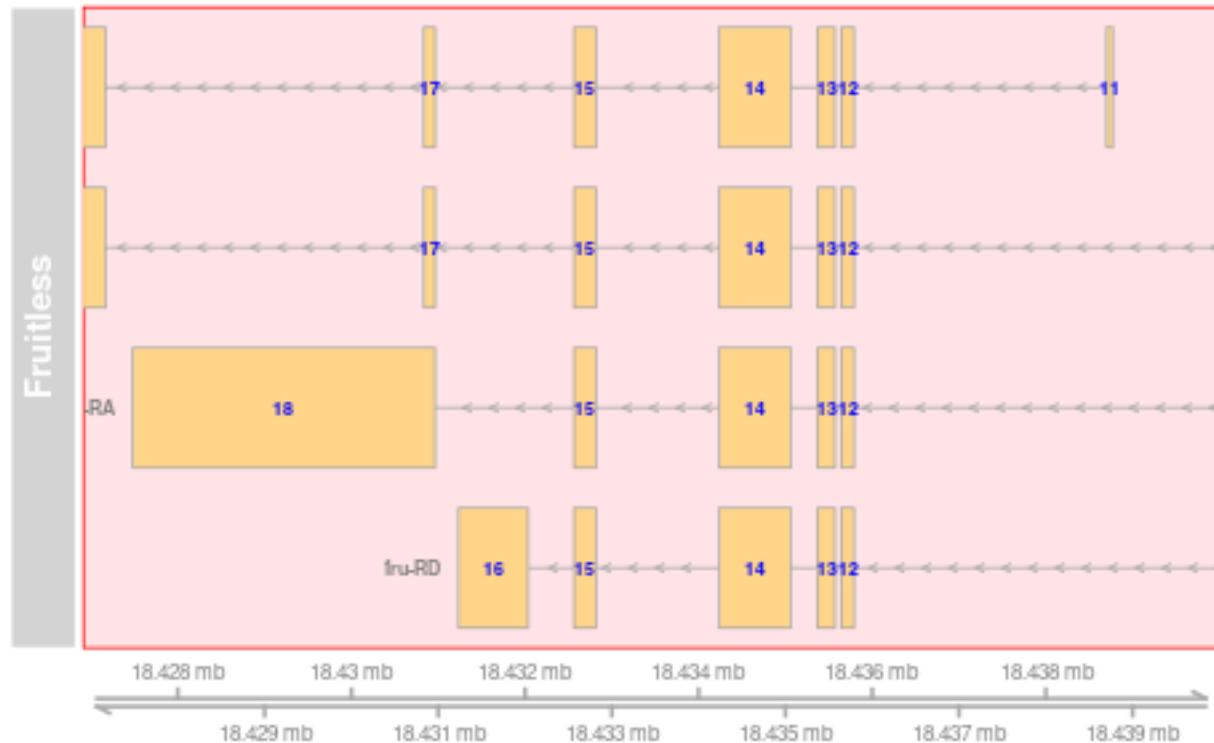
In addition to the genome as a whole, the gene *Fruitless* was given particular attention.

Figure 1. Fruitless gene model: exons and transcripts



```
## png
## 2
```

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



```
## png
## 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)**

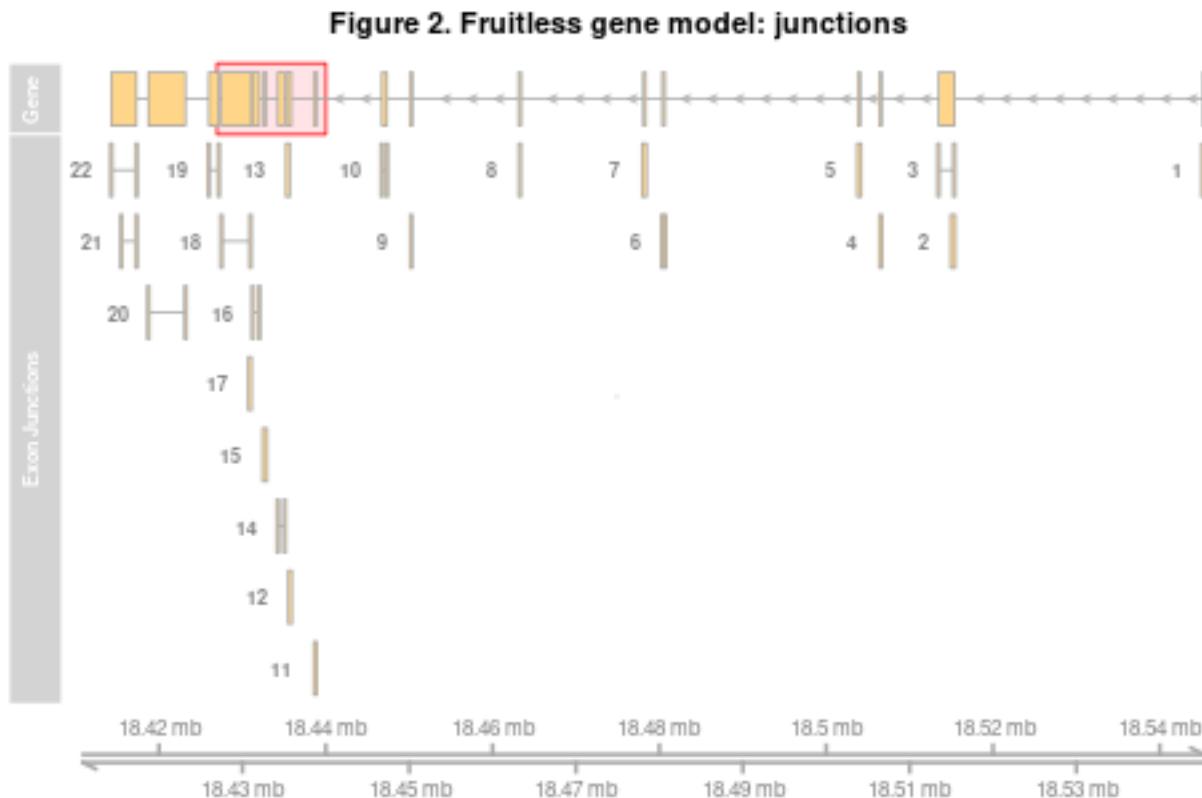
	start	stop
exon_22	18414273	18417301
exon_21	18415473	18417301
exon_20	18418716	18423183
exon_19	18425959	18427167
exon_18	18427480	18430965
exon_17	18430832	18430965
exon_16	18431233	18432035
exon_15	18432564	18432819
exon_14	18434235	18435063
exon_13	18435370	18435571
exon_12	18435643	18435791
exon_11	18438700	18438772
exon_10	18446701	18447330
exon_9	18450235	18450255
exon_8	18463267	18463282
exon_7	18478064	18478333

exon_6	18480328	18480677
exon_5	18503846	18504067
exon_4	18506494	18506563
exon_3	18513451	18515344
exon_2	18515052	18515344
exon_1	18545113	18545587

```
cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf | grep "fru"
cat /proj/cdjones_lab/Genomics_Data_Commons/annotations/drosophila_melanogaster/dmel-all-r6.13.gtf | grep "fru"
cat fru.test.gtf.exon fru.test.gtf.gene | bedtools sort > utils/annotations/fru_ex.gtf
```

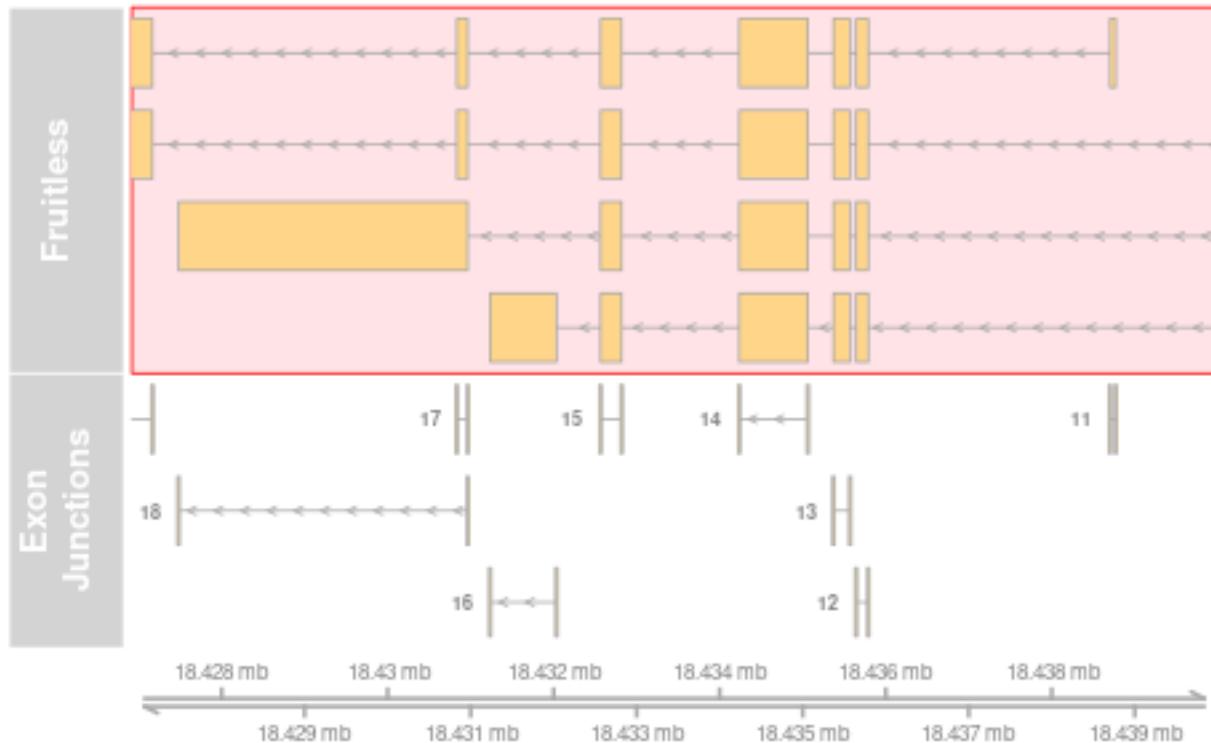
```
cat fru.test.gtf.exon | cut -f 1,4,5,7,9 | tr -d '"' | tr -d ";" | sed -e 's/gene_id //g' | awk '{print $1,$2,$3,$4,$5,$6,$7}' > utils/annotations/fru_ex.bed.tmp
cat utils/annotations/fru_ex.bed.tmp | tr "_" "\t" | awk '{print$1,$2,$3,$4_"$5,$6,$7}' | tr " " "\t" > utils/annotations/fru_ex.gtf
```

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:



```
## png
## 2
```

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

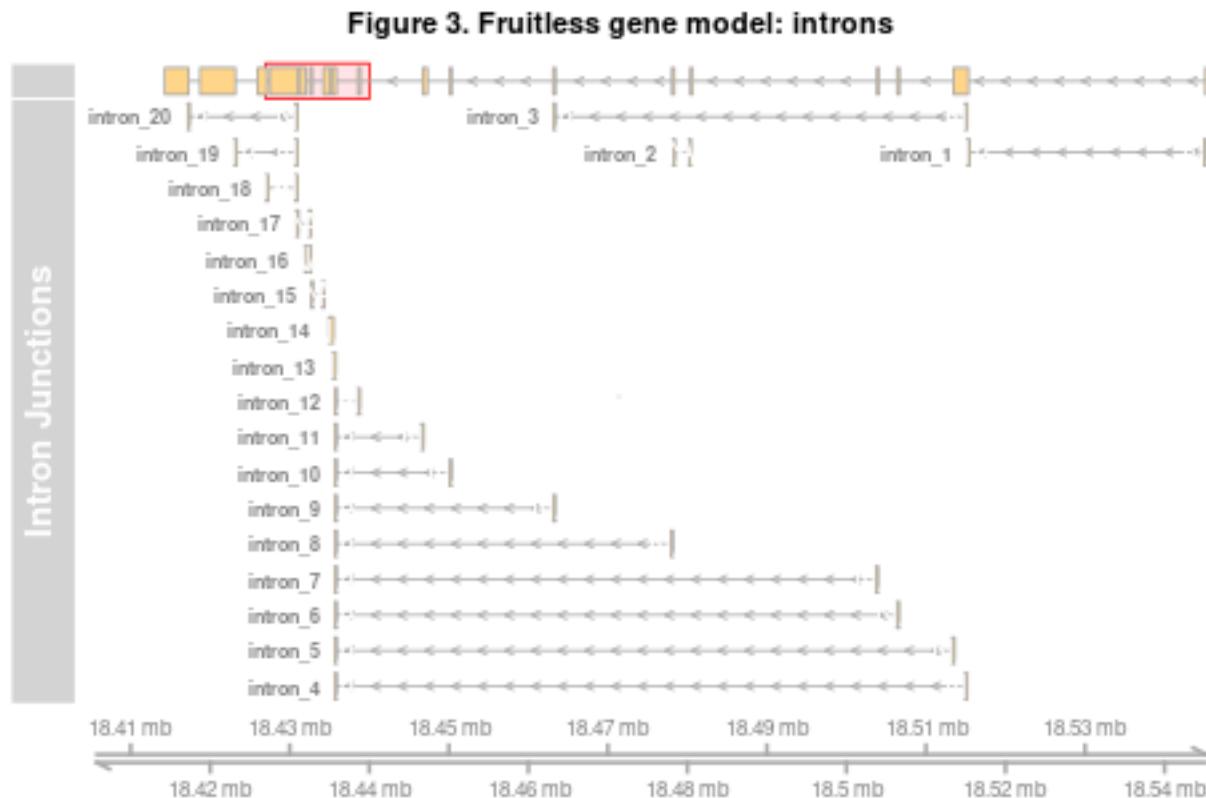


```

## png
## 2

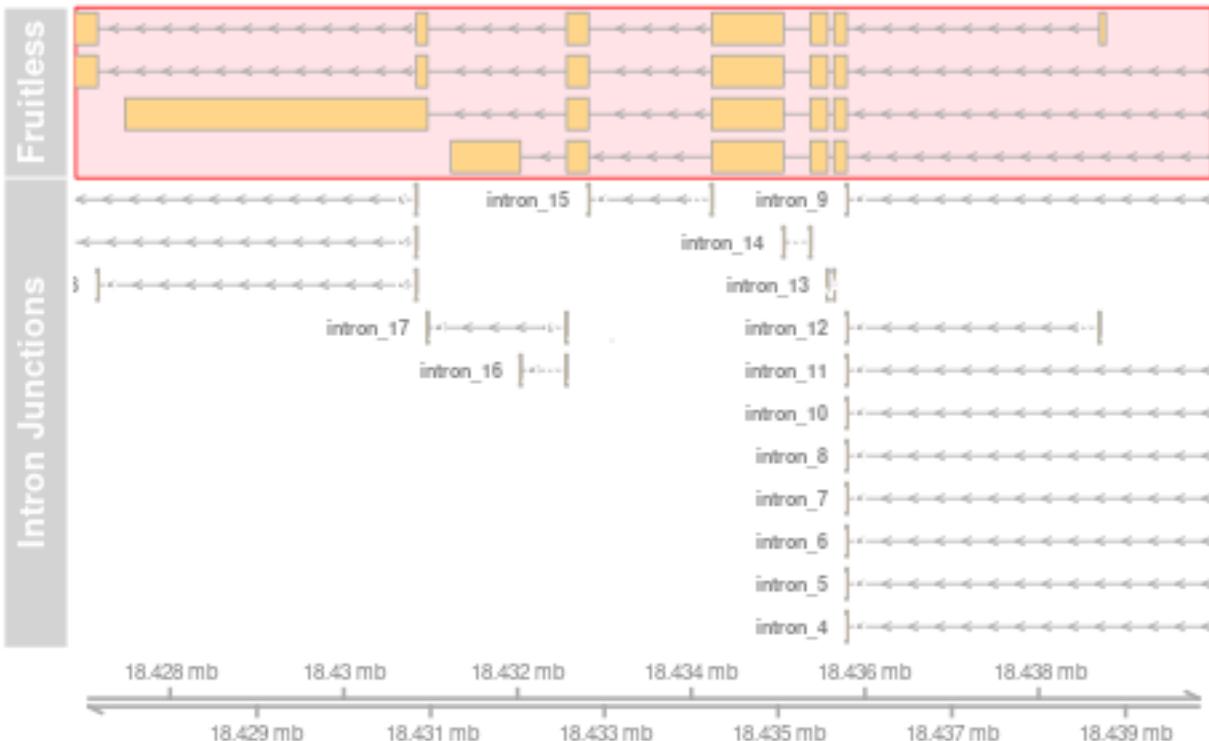
cat utils/annotations/fru_ex.gtf | grep -w gene > utils/annotations/fru_exonEdges.gtf
cat utils/annotations/fru_ex.gtf | grep -w gene | cut -f 1,2 > utils/annotations/fru_exonEdges.gtf.front
paste <(> cat utils/annotations/fru_ex.gtf | grep -w gene | cut -f 6- ) <(> cat utils/annotations/fru_ex.gtf | grep -w gene | awk '{print $1"\t$2\t$3\t$4\t$5\t$6\t$7\t$8\t$9\t$10\t$11\t$12\t$13\t$14\t$15\t$16\t$17\t$18\t$19\t$20\t$21\t$22\t$23\t$24\t$25\t$26\t$27\t$28\t$29\t$30\t$31\t$32\t$33\t$34\t$35\t$36\t$37\t$38\t$39\t$40\t$41\t$42\t$43\t$44\t$45\t$46\t$47\t$48\t$49\t$50\t$51\t$52\t$53\t$54\t$55\t$56\t$57\t$58\t$59\t$60\t$61\t$62\t$63\t$64\t$65\t$66\t$67\t$68\t$69\t$70\t$71\t$72\t$73\t$74\t$75\t$76\t$77\t$78\t$79\t$80\t$81\t$82\t$83\t$84\t$85\t$86\t$87\t$88\t$89\t$90\t$91\t$92\t$93\t$94\t$95\t$96\t$97\t$98\t$99\t$100\t$101\t$102\t$103\t$104\t$105\t$106\t$107\t$108\t$109\t$110\t$111\t$112\t$113\t$114\t$115\t$116\t$117\t$118\t$119\t$120\t$121\t$122\t$123\t$124\t$125\t$126\t$127\t$128\t$129\t$130\t$131\t$132\t$133\t$134\t$135\t$136\t$137\t$138\t$139\t$140\t$141\t$142\t$143\t$144\t$145\t$146\t$147\t$148\t$149\t$150\t$151\t$152\t$153\t$154\t$155\t$156\t$157\t$158\t$159\t$160\t$161\t$162\t$163\t$164\t$165\t$166\t$167\t$168\t$169\t$170\t$171\t$172\t$173\t$174\t$175\t$176\t$177\t$178\t$179\t$180\t$181\t$182\t$183\t$184\t$185\t$186\t$187\t$188\t$189\t$190\t$191\t$192\t$193\t$194\t$195\t$196\t$197\t$198\t$199\t$200\t$201\t$202\t$203\t$204\t$205\t$206\t$207\t$208\t$209\t$210\t$211\t$212\t$213\t$214\t$215\t$216\t$217\t$218\t$219\t$220\t$221\t$222\t$223\t$224\t$225\t$226\t$227\t$228\t$229\t$230\t$231\t$232\t$233\t$234\t$235\t$236\t$237\t$238\t$239\t$240\t$241\t$242\t$243\t$244\t$245\t$246\t$247\t$248\t$249\t$250\t$251\t$252\t$253\t$254\t$255\t$256\t$257\t$258\t$259\t$260\t$261\t$262\t$263\t$264\t$265\t$266\t$267\t$268\t$269\t$270\t$271\t$272\t$273\t$274\t$275\t$276\t$277\t$278\t$279\t$280\t$281\t$282\t$283\t$284\t$285\t$286\t$287\t$288\t$289\t$290\t$291\t$292\t$293\t$294\t$295\t$296\t$297\t$298\t$299\t$300\t$301\t$302\t$303\t$304\t$305\t$306\t$307\t$308\t$309\t$310\t$311\t$312\t$313\t$314\t$315\t$316\t$317\t$318\t$319\t$320\t$321\t$322\t$323\t$324\t$325\t$326\t$327\t$328\t$329\t$330\t$331\t$332\t$333\t$334\t$335\t$336\t$337\t$338\t$339\t$340\t$341\t$342\t$343\t$344\t$345\t$346\t$347\t$348\t$349\t$350\t$351\t$352\t$353\t$354\t$355\t$356\t$357\t$358\t$359\t$360\t$361\t$362\t$363\t$364\t$365\t$366\t$367\t$368\t$369\t$370\t$371\t$372\t$373\t$374\t$375\t$376\t$377\t$378\t$379\t$380\t$381\t$382\t$383\t$384\t$385\t$386\t$387\t$388\t$389\t$390\t$391\t$392\t$393\t$394\t$395\t$396\t$397\t$398\t$399\t$400\t$401\t$402\t$403\t$404\t$405\t$406\t$407\t$408\t$409\t$410\t$411\t$412\t$413\t$414\t$415\t$416\t$417\t$418\t$419\t$420\t$421\t$422\t$423\t$424\t$425\t$426\t$427\t$428\t$429\t$430\t$431\t$432\t$433\t$434\t$435\t$436\t$437\t$438\t$439\t$440\t$441\t$442\t$443\t$444\t$445\t$446\t$447\t$448\t$449\t$450\t$451\t$452\t$453\t$454\t$455\t$456\t$457\t$458\t$459\t$460\t$461\t$462\t$463\t$464\t$465\t$466\t$467\t$468\t$469\t$470\t$471\t$472\t$473\t$474\t$475\t$476\t$477\t$478\t$479\t$480\t$481\t$482\t$483\t$484\t$485\t$486\t$487\t$488\t$489\t$490\t$491\t$492\t$493\t$494\t$495\t$496\t$497\t$498\t$499\t$500\t$501\t$502\t$503\t$504\t$505\t$506\t$507\t$508\t$509\t$510\t$511\t$512\t$513\t$514\t$515\t$516\t$517\t$518\t$519\t$520\t$521\t$522\t$523\t$524\t$525\t$526\t$527\t$528\t$529\t$530\t$531\t$532\t$533\t$534\t$535\t$536\t$537\t$538\t$539\t$540\t$541\t$542\t$543\t$544\t$545\t$546\t$547\t$548\t$549\t$550\t$551\t$552\t$553\t$554\t$555\t$556\t$557\t$558\t$559\t$560\t$561\t$562\t$563\t$564\t$565\t$566\t$567\t$568\t$569\t$5610\t$5611\t$5612\t$5613\t$5614\t$5615\t$5616\t$5617\t$5618\t$5619\t$5620\t$5621\t$5622\t$5623\t$5624\t$5625\t$5626\t$5627\t$5628\t$5629\t$5630\t$5631\t$5632\t$5633\t$5634\t$5635\t$5636\t$5637\t$5638\t$5639\t$5640\t$5641\t$5642\t$5643\t$5644\t$5645\t$5646\t$5647\t$5648\t$5649\t$5650\t$5651\t$5652\t$5653\t$5654\t$5655\t$5656\t$5657\t$5658\t$5659\t$5660\t$5661\t$5662\t$5663\t$5664\t$5665\t$5666\t$5667\t$5668\t$5669\t$56610\t$56611\t$56612\t$56613\t$56614\t$56615\t$56616\t$56617\t$56618\t$56619\t$56620\t$56621\t$56622\t$56623\t$56624\t$56625\t$56626\t$56627\t$56628\t$56629\t$56630\t$56631\t$56632\t$56633\t$56634\t$56635\t$56636\t$56637\t$56638\t$56639\t$56640\t$56641\t$56642\t$56643\t$56644\t$56645\t$56646\t$56647\t$56648\t$56649\t$56650\t$56651\t$56652\t$56653\t$56654\t$56655\t$56656\t$56657\t$56658\t$56659\t$56660\t$56661\t$56662\t$56663\t$56664\t$56665\t$56666\t$56667\t$56668\t$56669\t$566610\t$566611\t$566612\t$566613\t$566614\t$566615\t$566616\t$566617\t$566618\t$566619\t$566620\t$566621\t$566622\t$566623\t$566624\t$566625\t$566626\t$566627\t$566628\t$566629\t$566630\t$566631\t$566632\t$566633\t$566634\t$566635\t$566636\t$566637\t$566638\t$566639\t$566640\t$566641\t$566642\t$566643\t$566644\t$566645\t$566646\t$566647\t$566648\t$566649\t$566650\t$566651\t$566652\t$566653\t$566654\t$566655\t$566656\t$566657\t$566658\t$566659\t$566660\t$566661\t$566662\t$566663\t$566664\t$566665\t$566666\t$566667\t$566668\t$566669\t$5666610\t$5666611\t$5666612\t$5666613\t$5666614\t$5666615\t$5666616\t$5666617\t$5666618\t$5666619\t$5666620\t$5666621\t$5666622\t$5666623\t$5666624\t$5666625\t$5666626\t$5666627\t$5666628\t$5666629\t$5666630\t$5666631\t$5666632\t$5666633\t$5666634\t$5666635\t$5666636\t$5666637\t$5666638\t$5666639\t$5666640\t$5666641\t$5666642\t$5666643\t$5666644\t$5666645\t$5666646\t$5666647\t$5666648\t$5666649\t$5666650\t$5666651\t$5666652\t$5666653\t$5666654\t$5666655\t$5666656\t$5666657\t$5666658\t$5666659\t$5666660\t$5666661\t$5666662\t$5666663\t$5666664\t$5666665\t$5666666\t$5666667\t$5666668\t$5666669\t$56666610\t$56666611\t$56666612\t$56666613\t$56666614\t$56666615\t$56666616\t$56666617\t$56666618\t$56666619\t$56666620\t$56666621\t$56666622\t$56666623\t$56666624\t$56666625\t$56666626\t$56666627\t$56666628\t$56666629\t$56666630\t$56666631\t$56666632\t$56666633\t$56666634\t$56666635\t$56666636\t$56666637\t$56666638\t$56666639\t$56666640\t$56666641\t$56666642\t$56666643\t$56666644\t$56666645\t$56666646\t$56666647\t$56666648\t$56666649\t$56666650\t$56666651\t$56666652\t$56666653\t$56666654\t$56666655\t$56666656\t$56666657\t$56666658\t$56666659\t$56666660\t$56666661\t$56666662\t$56666663\t$56666664\t$56666665\t$56666666\t$56666667\t$56666668\t$56666669\t$566666610\t$566666611\t$566666612\t$566666613\t$566666614\t$566666615\t$566666616\t$566666617\t$566666618\t$566666619\t$566666620\t$566666621\t$566666622\t$566666623\t$566666624\t$566666625\t$566666626\t$566666627\t$566666628\t$566666629\t$566666630\t$566666631\t$566666632\t$566666633\t$566666634\t$566666635\t$566666636\t$566666637\t$566666638\t$566666639\t$566666640\t$566666641\t$566666642\t$566666643\t$566666644\t$566666645\t$566666646\t$566666647\t$566666648\t$566666649\t$566666650\t$566666651\t$566666652\t$566666653\t$566666654\t$566666655\t$566666656\t$566666657\t$566666658\t$566666659\t$566666660\t$566666661\t$566666662\t$566666663\t$566666664\t$566666665\t$566666666\t$566666667\t$566666668\t$566666669\t$5666666610\t$5666666611\t$5666666612\t$5666666613\t$5666666614\t$5666666615\t$5666666616\t$5666666617\t$5666666618\t$5666666619\t$5666666620\t$5666666621\t$5666666622\t$5666666623\t$5666666624\t$5666666625\t$5666666626\t$5666666627\t$5666666628\t$5666666629\t$5666666630\t$5666666631\t$5666666632\t$5666666633\t$5666666634\t$5666666635\t$5666666636\t$5666666637\t$5666666638\t$5666666639\t$5666666640\t$5666666641\t$5666666642\t$5666666643\t$5666666644\t$5666666645\t$5666666646\t$5666666647\t$5666666648\t$5666666649\t$5666666650\t$5666666651\t$5666666652\t$5666666653\t$5666666654\t$5666666655\t$5666666656\t$5666666657\t$5666666658\t$5666666659\t$5666666660\t$5666666661\t$5666666662\t$5666666663\t$5666666664\t$5666666665\t$5666666666\t$5666666667\t$5666666668\t$5666666669\t$56666666610\t$56666666611\t$56666666612\t$56666666613\t$56666666614\t$56666666615\t$56666666616\t$56666666617\t$56666666618\t$56666666619\t$56666666620\t$56666666621\t$56666666622\t$56666666623\t$56666666624\t$56666666625\t$56666666626\t$56666666627\t$56666666628\t$56666666629\t$56666666630\t$56666666631\t$56666666632\t$56666666633\t$56666666634\t$56666666635\t$56666666636\t$56666666637\t$56666666638\t$56666666639\t$56666666640\t$56666666641\t$56666666642\t$56666666643\t$56666666644\t$56666666645\t$56666666646\t$56666666647\t$56666666648\t$56666666649\t$56666666650\t$56666666651\t$56666666652\t$56666666653\t$56666666654\t$56666666655\t$56666666656\t$56666666657\t$56666666658\t$56666666659\t$56666666660\t$56666666661\t$56666666662\t$56666666663\t$56666666664\t$56666666665\t$56666666666\t$56666666667\t$56666666668\t$56666666669\t$566666666610\t$566666666611\t$566666666612\t$566666666613\t$566666666614\t$566666666615\t$566666666616\t$566666666617\t$566666666618\t$566666666619\t$566666666620\t$566666666621\t$566666666622\t$566666666623\t$566666666624\t$566666666625\t$566666666626\t$566666666627\t$566666666628\t$566666666629\t$566666666630\t$566666666631\t$566666666632\t$566666666633\t$566666666634\t$566666666635\t$566666666636\t$566666666637\t$566666666638\t$566666666639\t$566666666640\t$566666666641\t$566666666642\t$566666666643\t$566666666644\t$566666666645\t$566666666646\t$566666666647\t$566666666648\t$566666666649\t$566666666650\t$566666666651\t$566666666652\t$566666666653\t$566666666654\t$566666666655\t$566666666656\t$566666666657\t$566666666658\t$566666666659\t$566666666660\t$566666666661\t$566666666662\t$566666666663\t$566666666664\t$566666666665\t$566666666666\t$566666666667\t$566666666668\t$566666666669\t$5666666666610\t$5666666666611\t$5666666666612\t$5666666666613\t$5666666666614\t$5666666666615\t$5666666666616\t$5666666666617\t$5666666666618\t$5666666666619\t$5666666666620\t$5666666666621\t$5666666666622\t$5666666666623\t$5666666666624\t$5666666666625\t$5666666666626\t$5666666666627\t$5666666666628\t$5666666666629\t$5666666666630\t$5666666666631\t$5666666666632\t$5666666666633\t$5666666666634\t$5666666666635\t$5666666666636\t$5666666666637\t$5666666666638\t$5666666666639\t$5666666666640\t$5666666666641\t$5666666666642\t$5666666666643\t$5666666666644\t$5666666666645\t$5666666666646\t$5666666666647\t$5666666666648\t$5666666666649\t$5666666666650\t$5666666666651\t$5666666666652\t$5666666666653\t$5666666666654\t$5666666666655\t$5666666666656\t$5666666666657\t$5666666666658\t$5666666666659\t$5666666666660\t$5666666666661\t$5666666666662\t$5666666666663\t$5666666666664\t$5666666666665\t$5666666666666\t$5666666666667\t$5666666666668\t$5666666666669\t$56666666666610\t$56666666666611\t$56666666666612\t$56666666666613\t$56666666666614\t$56666666666615\t$56666666666616\t$56666666666617\t$56666666666618\t$56666666666619\t$56666666666620\t$56666666666621\t$56666666666622\t$56666666666623\t$56666666666624\t$56666666666625\t$56666666666626\t$56666666666627\t$56666666666628\t$56666666666629\t$56666666666630\t$56666666666631\t$56666666666632\t$56666666666633\t$56666666666634\t$56666666666635\t$56666666666636\t$56666666666637\t$56666666666638\t$56666666666639\t$56666666666640\t$56666666666641\t$56666666666642\t$56666666666643\t$56666666666644\t$56666666666645\t$56666666666646\t$56666666666647\t$56666666666648\t$56666666666649\t$56666666666650\t$56666666666651\t$56666666666652\t$56666666666653\t$56666666666654\t$56666666666655\t$56666666666656\t$56666666666657\t$56666666666658\t$56666666666659\t$56666666666660\t$56666666666661\t$56666666666662\t$56666666666663\t$56666666666664\t$56666666666665\t$56666666666666\t$56666666666667\t$56666666666668\t$56666666666669\t$566666666666610\t$566666666666611\t$566666666666612\t$566666666666613\t$566666666666614\t$566666666666615\t$566666666666616\t$566666666666617\t$566666666666618\t$566666666666619\t$566666666666620\t$566666666666621\t$566666666666622\t$566666666666623\t$566666666666624\t$566666666666625\t$566666666666626\t$566666666666627\t$566666666666628\t$566666666666629\t$566666666666630\t$566666666666631\t$566666666666632\t$566666666666633\t$566666666666634\t$566666666666635\t$566666666666636\t$566666666666637\t$566666666666638\t$566666666666639\t$566666666666640\t$566666666666641\t$566666666666642\t$566666666666643\t$566666666666644\t$566666666666645\t$566666666666646\t$566666666666647\t$566666666666648\t$566666666666649\t$566666666666650\t$566666666666651\t$566666666666652\t$566666666666653\t$566666666666654\t$566666666666655\t$566666666666656\t$566666666666657\t$566666666666658\t$566666666666659\t$566666666666660\t$566666666666661\t$566666666666662\t$566666666666663\t$566666666666664\t$566666666666665\t$
```

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:



```
## png  
## 2
```

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



```

## png
## 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\";"}' | tr '~' ' '

```

```
cat utils/annotations/fru_introns.gtf.tmp | sed -e 's/intron_/intron~/g' |sed -e 's/intron~20/intron_1/
```

(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	ionChannel	ionotropic	mating	nervSysDev
total count	25	35	8	250	246	3	93
annotated count	54	35	8	250	246	3	90
percent of annotations	0.3%	0.2%	0.0%	1.4%	1.4%	0.0%	0.5%
total size	679.5K	3.2M	46.9K	4.0M	3.7M	5.0K	1.8M
avg size	12.6K	90.7K	5.9K	16.2K	15.2K	1.7K	19.8K
percent genome size	0.5%	2.3%	0.0%	2.9%	2.7%	0.0%	1.3%
percent annotation size	0.7%	3.1%	0.0%	4.0%	3.7%	0.0%	1.7%

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>

- Synaptic signaling - http://flybase.org/cgi-bin/cvreport.pl?rel=is_a&id=GO:0099536
- Synapse organization - <http://flybase.org/cgi-bin/cvreport.pl?id=GO%3A0050808>

(Bryson, email 24 July 2019)

- Ion Channel Activity - http://flybase.org/cgi-bin/cvreport.pl?rel=is_a&id=GO:0005216

(Bryson, email 12 May 2020)

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

Ion Channel Activity

251, all good

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
amos	group	NA	antennae	3
CantonS	group	NA	antennae	3
FruLexaFru440	group	7	antennae	3
rn	group	1	NA	8
wt	group	1	NA	8
wt	group	7	antennae	3
wt	isolated	7	antennae	3
wt,rn	group	1	NA	8

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)

day	tissue	genotype	variable	
			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	88a	FruLexaFru440	67d
group	antennae	5	3	3	3	0	0
group	antennae	7	3	0	0	3	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. This analysis was computationally problematic and ultimately inconclusive, and has been deactivated in this version.

Reads from CantonS flies, with and without mutations in the amos genes were downloaded from NCBI (PRJNA532415 ; Mohapatra and Menuz (2019)). Unlike our samples, which are from day 5 or 7, these are described as “3-5 days old” in the paper. “each sample consisted of 300-400 antennae from approximately equal numbers of males and females”. By contrast, our samples, though referred to as “wild type” or “wt”, are in fact white-eyed mutants. Another difference is that the reads on NCBI are those which have already been mapped to the reference genome (thus some of the filtering pipeline is redundant).

Table 5c. CantonS +/- amos mutation contrast

normie	white eyes	CantonS
garden variety	3	3
amos mutant	0	3

Table 5d. wildtype +/- rotund time series

age	genotype		
	wt	wt,rn	rn
1	8	8	8

2.4.1 Pre-Processing

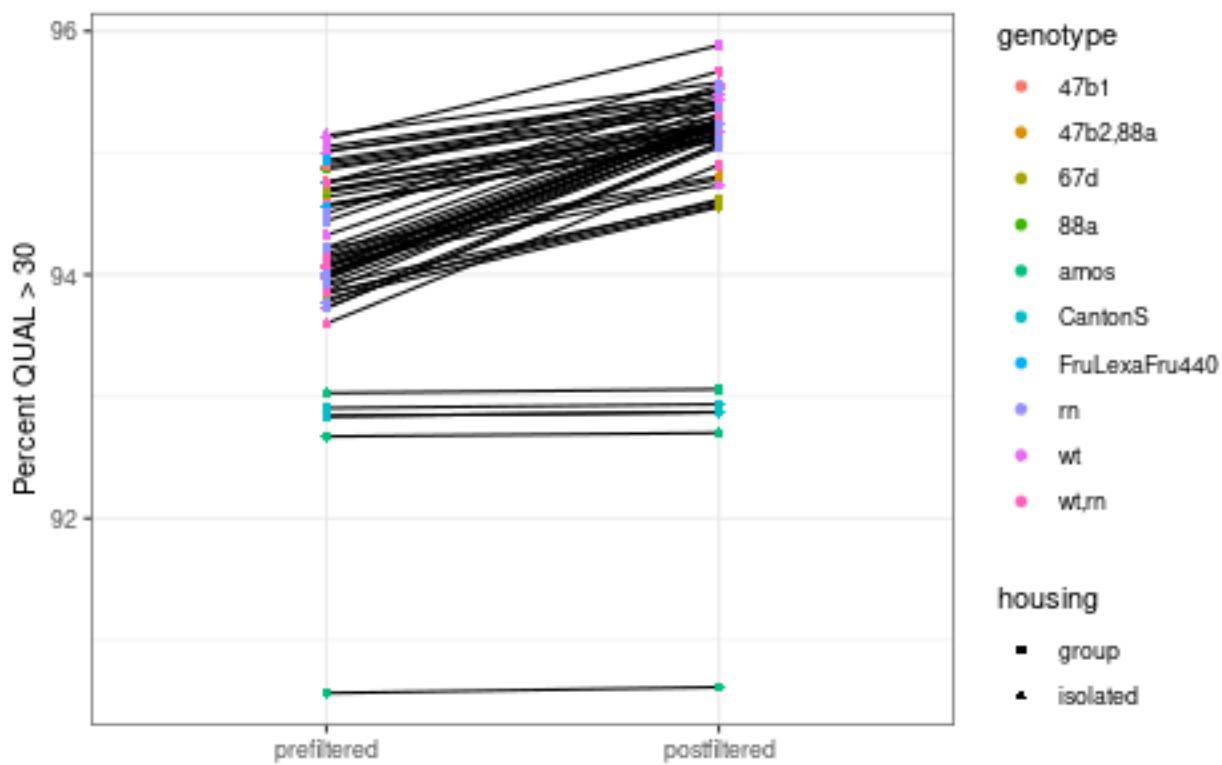
These reads were preprocessed with FASTP (S. Chen et al. 2018) for quality control and analytics.

Starting FASTQ files contained a total of 1.38G reads; after QC, this dropped to 1.35G.

Table 6. Read Retention Rate during Preprocessing

	minimum	average	maximum
prefiltered	11M	25M	56M
postfiltered	11M	25M	56M
percent retention	96	98	100

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



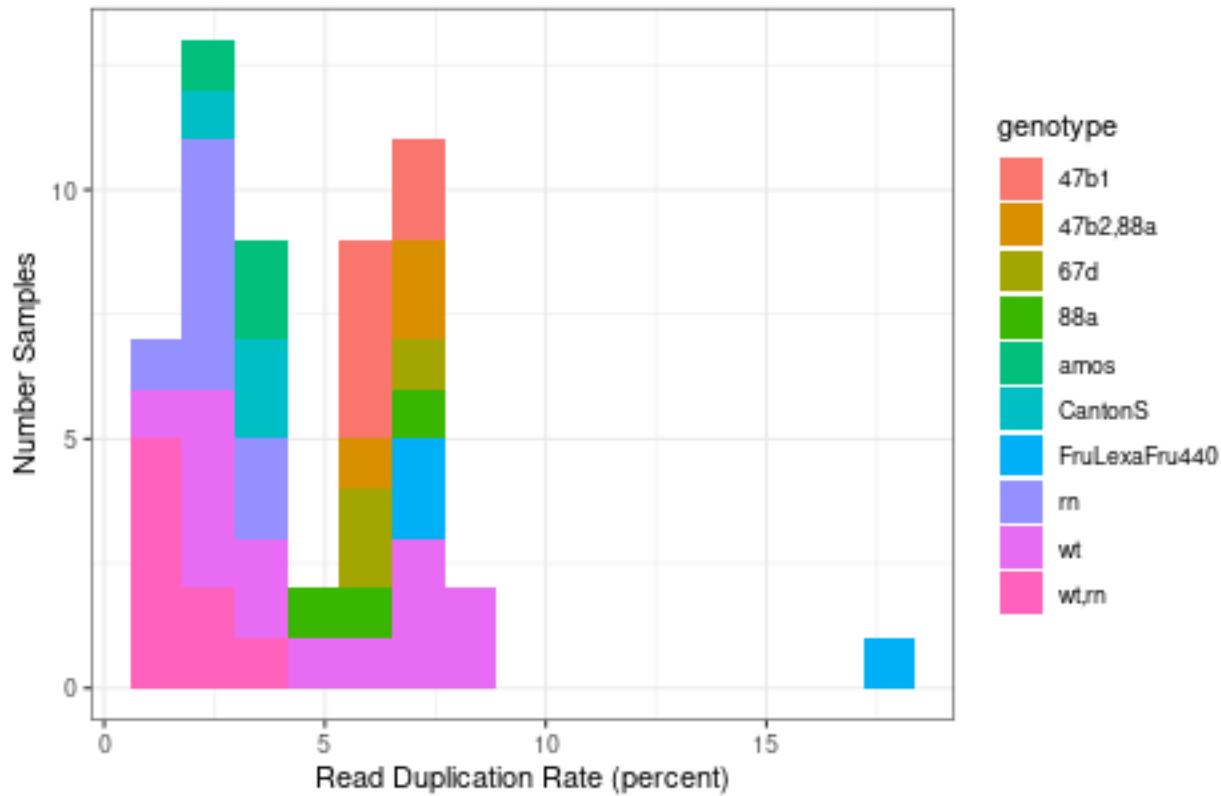
```
## png
## 2
```

Duplicate reads were also detected

Table 7. Percentage Duplication
FASTP estimate

	minimum	average	median	maximum
	1.3	4.6	3.7	17.9

Figure 5. Duplication Histogram (FASTP estimate)



```
## png
## 2
```

2.5 Mapped Reads

Reads were mapped to the reference genome using MapSplice2 (K. 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 1.35G reads, MapSplice was able to align 1.35G of them, for an overall mapping rate of 99.5695791 %.

Individual mapping rates were generally more than 98%.

Table 8. Percent of Reads Mapping
raw mapsplice output

maximum	mean	median	minimum
99.9%	99.6%	99.7%	98.5%

Table 9. Individual Mapping Rates
raw mapsplice output

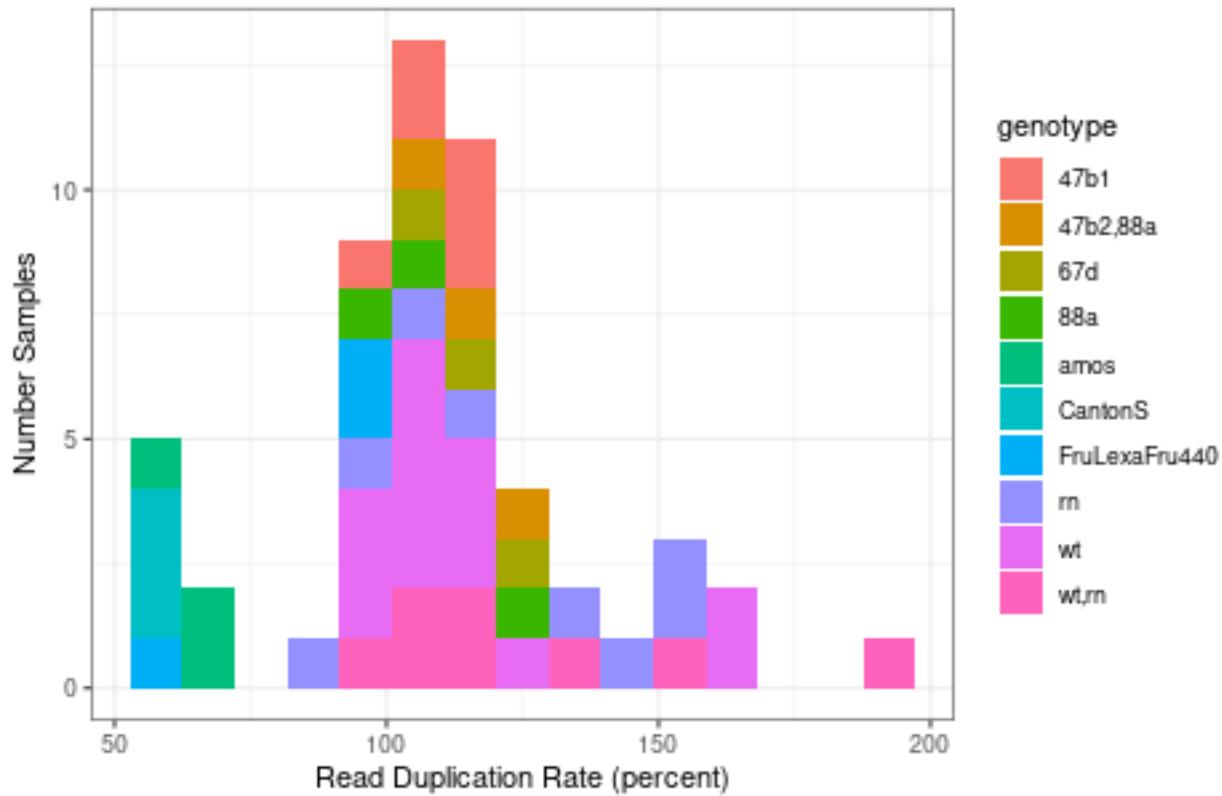
rep	day	normie	cw	total reads	reads mapped	percent mapped
group - 47b1						
1	5	NA	NA	32.0M	31.9M	99.7%
2	5	NA	NA	28.2M	28.0M	99.4%
3	5	NA	NA	24.4M	24.2M	99.0%
1	7	NA	NA	32.1M	31.9M	99.5%
2	7	NA	NA	28.9M	28.8M	99.7%
3	7	NA	NA	24.3M	24.3M	99.6%
group - 47b2,88a						
1	5	NA	NA	20.3M	20.2M	99.5%
2	5	NA	NA	31.7M	31.6M	99.5%
3	5	NA	NA	24.7M	24.5M	99.3%
group - 88a						
1	5	NA	NA	37.0M	36.8M	99.7%
2	5	NA	NA	30.4M	30.2M	99.6%
3	5	NA	NA	36.2M	36.1M	99.7%
group - 67d						
1	7	NA	NA	25.1M	25.0M	99.6%
2	7	NA	NA	31.2M	31.0M	99.5%
3	7	NA	NA	24.1M	24.0M	99.6%
group - wt						
1	7	regular	w	42.6M	42.2M	99.0%
2	7	regular	w	31.5M	31.0M	98.5%
3	7	regular	w	30.2M	29.9M	99.0%
1	1	regular	NA	18.3M	18.2M	99.8%
2	1	regular	NA	16.8M	16.8M	99.8%
1	1	regular	NA	11.5M	11.4M	99.8%
2	1	regular	NA	14.7M	14.7M	99.8%
1	1	regular	NA	13.8M	13.8M	99.8%
2	1	regular	NA	15.5M	15.4M	99.8%
1	1	regular	NA	11.1M	11.0M	99.8%
2	1	regular	NA	14.3M	14.2M	99.8%
isolated - wt						
1	7	NA	NA	30.7M	30.4M	99.2%
2	7	NA	NA	27.2M	27.1M	99.5%
3	7	NA	NA	33.8M	33.5M	99.0%
group - amos						
1	NA	mutant	C	43.3M	43.2M	99.9%
2	NA	mutant	C	40.1M	40.1M	99.9%
3	NA	mutant	C	50.1M	50.1M	99.9%
group - CantonS						
1	NA	regular	C	56.1M	56.0M	99.9%
2	NA	regular	C	44.3M	44.2M	99.9%
3	NA	regular	C	54.4M	54.4M	99.9%
group - FruLexaFru440						
1	7	NA	NA	22.0M	21.7M	98.9%

2	7	NA	NA	$30.7M$	$30.4M$	99.1%
3	7	NA	NA	$30.7M$	$30.4M$	99.1%
group - wt,rn						
1	1	mutant	NA	$19.7M$	$19.6M$	99.6%
2	1	mutant	NA	$20.5M$	$20.4M$	99.7%
1	1	mutant	NA	$13.5M$	$13.5M$	99.8%
2	1	mutant	NA	$12.3M$	$12.2M$	99.7%
1	1	mutant	NA	$13.1M$	$13.1M$	99.8%
2	1	mutant	NA	$14.0M$	$14.0M$	99.8%
1	1	mutant	NA	$13.8M$	$13.8M$	99.8%
2	1	mutant	NA	$16.2M$	$16.2M$	99.8%
group - rn						
1	1	mutant	NA	$17.3M$	$17.2M$	99.6%
2	1	mutant	NA	$16.6M$	$16.5M$	99.7%
1	1	mutant	NA	$15.2M$	$15.2M$	99.8%
2	1	mutant	NA	$12.9M$	$12.9M$	99.7%
1	1	mutant	NA	$14.0M$	$13.9M$	99.8%
2	1	mutant	NA	$14.1M$	$14.1M$	99.8%
1	1	mutant	NA	$13.5M$	$13.5M$	99.8%
2	1	mutant	NA	$11.6M$	$11.5M$	99.8%

Table 10. Percent of Duplicate Reads
raw mapsplice output

maximum	mean	median	minimum
188.1%	109.4%	107.3%	53.5%

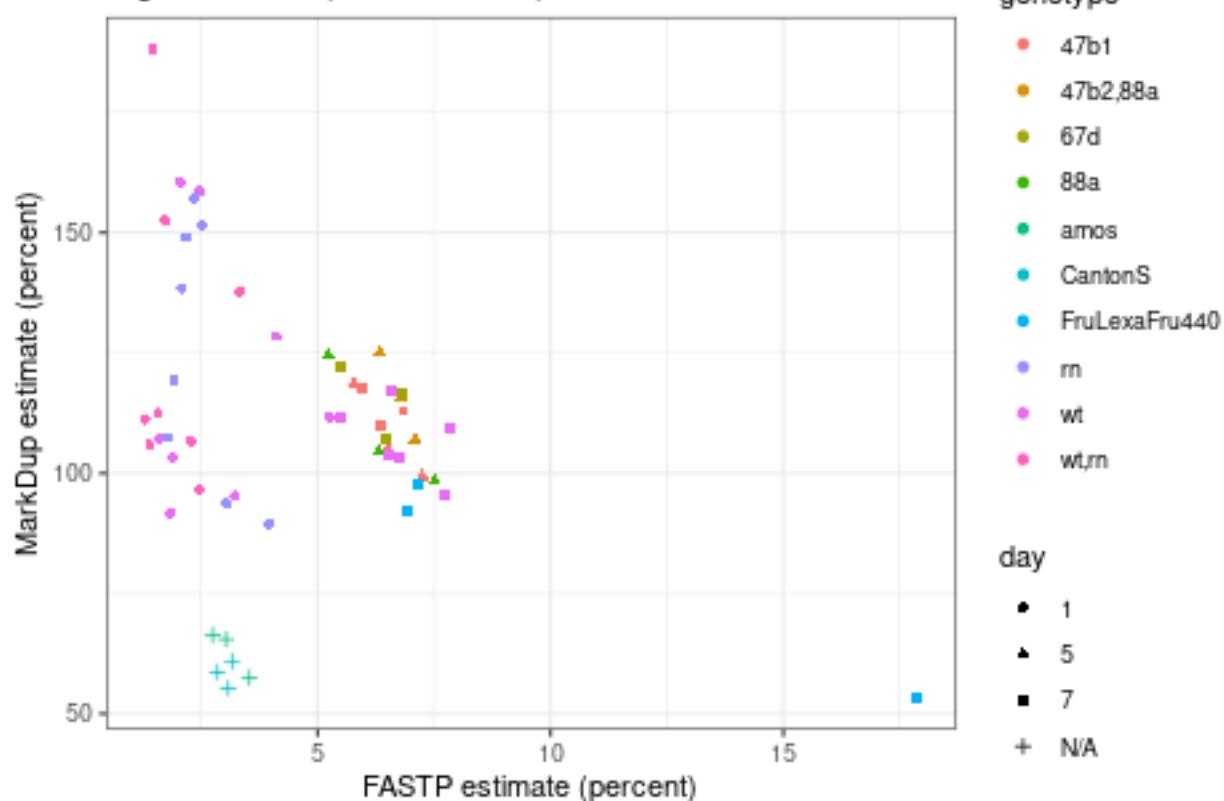
Figure 6. Duplication Histogram (Raw Mapsplice Alignment)



```
## png  
## 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.

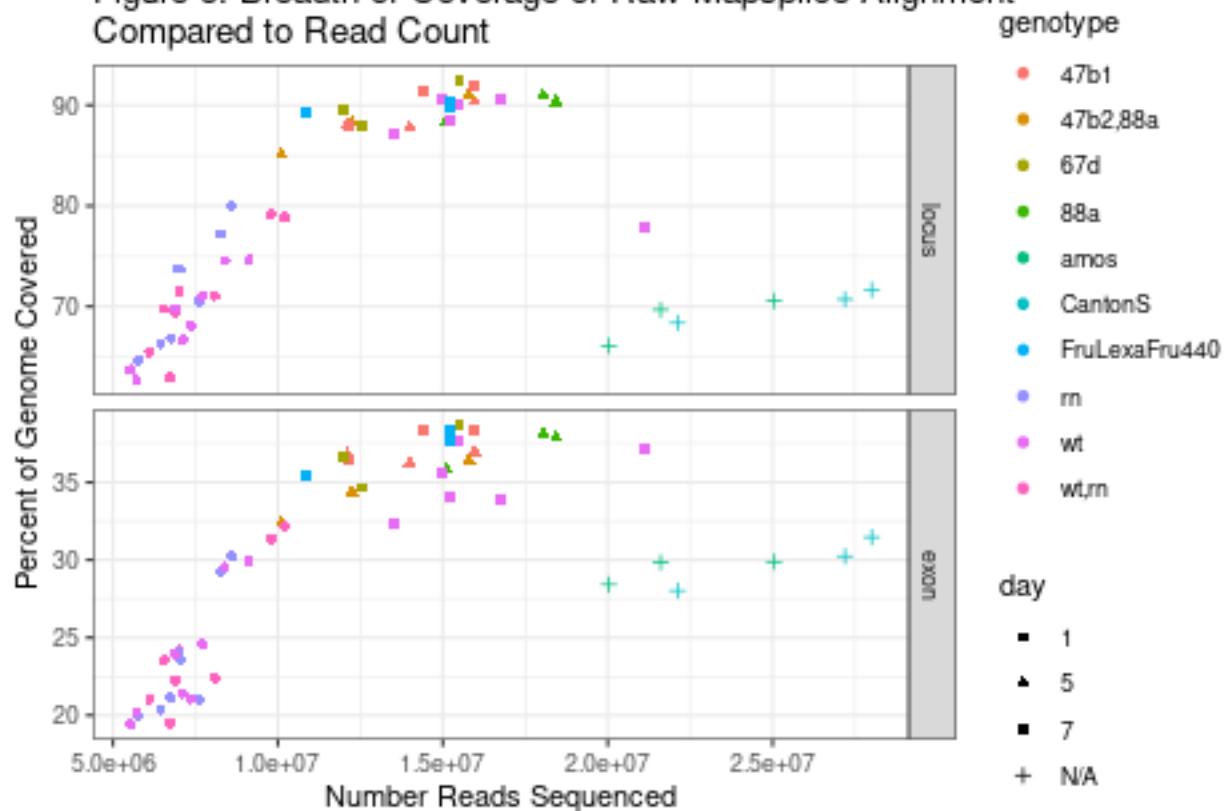
Figure 7. Comparison of Duplication Rate Estimates



```
## png
## 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 8. Breadth of Coverage of Raw Mapsplice Alignment Compared to Read Count

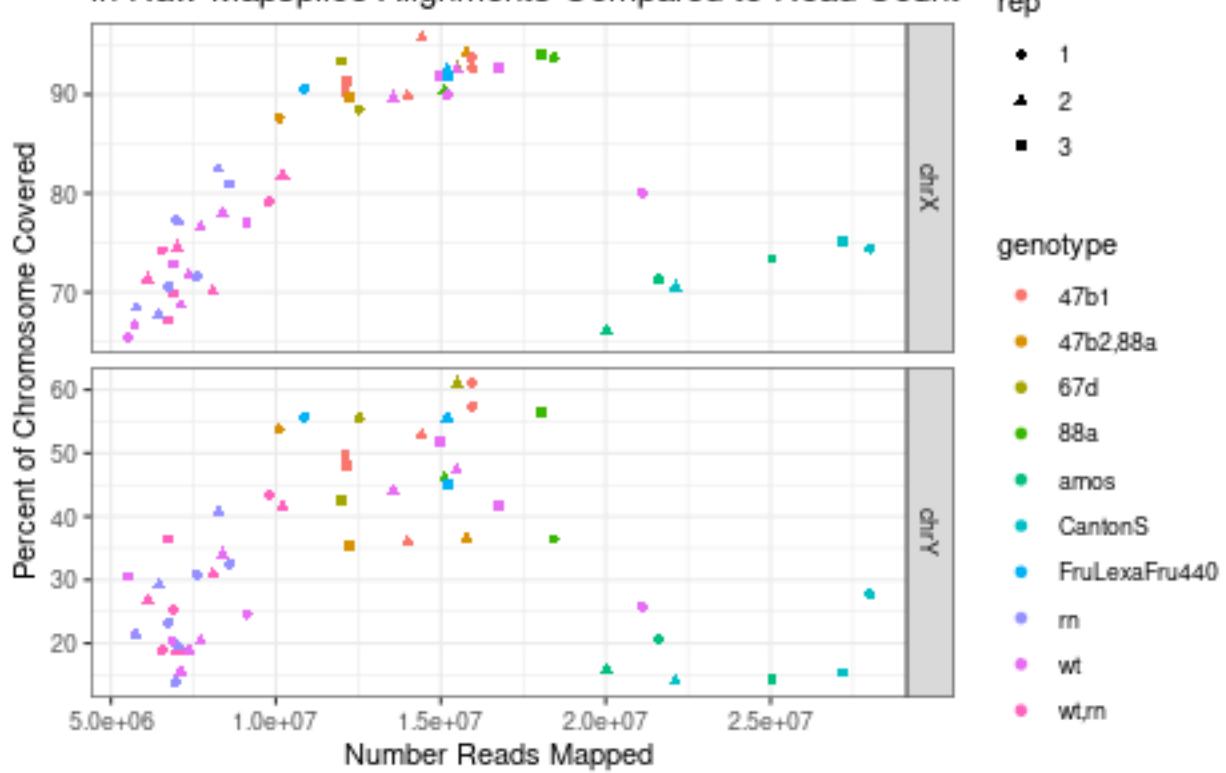


```
## png
## 2
```

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 breadth of the CantonS flies is unusually low for their depth of mapping.

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 samples from (Shiao et al. 2015) (not shown) 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



```
## png
## 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 -2.62 of $1.35G$ mapped reads, an overall mapped retention rate of 38.1772712 %.

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

	maximum	mean	median	minimum
mapped retention	85.3%	77.8%	78.8%	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%	99.4%	99.3%	99.0%
split breadth retention	97.5%	96.3%	96.8%	94.8%

Although filtration removed some (44.611037 %) of the multimapping reads, 16.1M remain ambiguously mapped. A given read mapped, on average, to 1.09348684519384 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
amos - group - NA				
1	98.4%	99.2%	1.01	1.01
2	98.1%	99.2%	1.01	1.01
3	98.5%	99.2%	1.01	1.01
CantonS - group - NA				
1	98.5%	99.2%	1.01	1.01
2	98.5%	99.2%	1.01	1.01
3	98.5%	99.2%	1.01	1.01
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
<hr/>				
wt,rn - group - 1				
1	94.5%	95.1%	1.21	1.14
2	95.2%	95.8%	1.20	1.14
1	95.8%	96.5%	1.18	1.12
2	96.2%	97.2%	1.15	1.10
1	97.1%	98.4%	1.08	1.05
2	97.0%	98.3%	1.09	1.05
1	95.8%	96.9%	1.16	1.11
2	96.1%	97.2%	1.14	1.10
<hr/>				
rn - group - 1				
1	93.1%	93.4%	1.32	1.21
2	93.1%	93.3%	1.41	1.28
1	96.4%	97.3%	1.15	1.10
2	96.1%	97.0%	1.17	1.11
1	96.3%	97.3%	1.11	1.07
2	96.5%	97.6%	1.10	1.06
1	96.0%	96.8%	1.17	1.12
2	96.0%	97.0%	1.16	1.11
<hr/>				
wt - group - 1				
1	96.7%	97.6%	1.10	1.07
2	96.4%	97.4%	1.11	1.08
1	96.8%	97.8%	1.10	1.07
2	96.5%	97.6%	1.11	1.07
1	97.4%	98.6%	1.06	1.04
2	97.5%	98.7%	1.06	1.04
1	95.1%	95.6%	1.16	1.12
2	96.9%	98.1%	1.08	1.06

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	100.0%	98.5%	98.4%	94.9%
spanned breadth retention	99.4%	98.7%	98.9%	97.8%
split breadth retention	95.8%	90.3%	89.8%	87.0%

2.5.4 Uniquely Mapped

mapsspliceUniq is derived from mapsspliceMulti 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

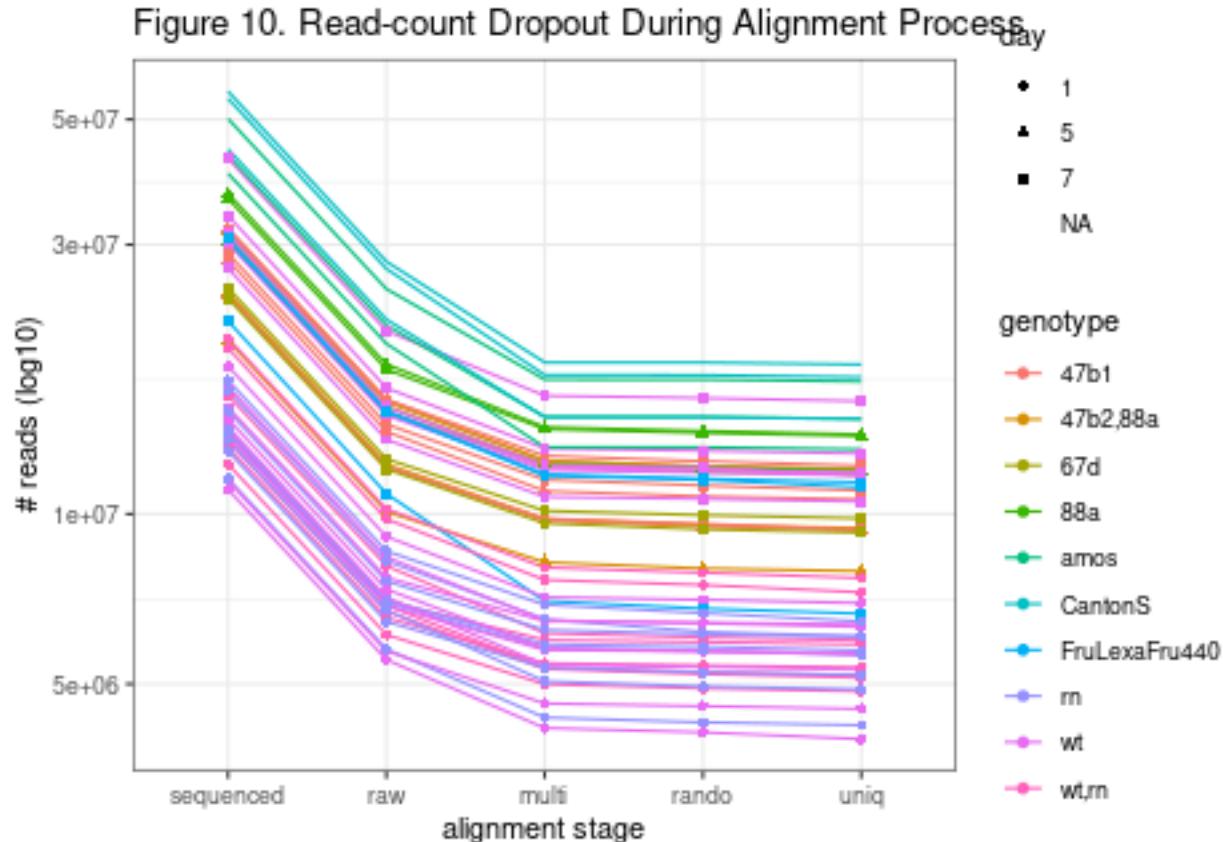
	maximum	mean	median	minimum
mapped retention	99.2%	97.1%	97.1%	93.3%
spanned breadth retention	99.1%	98.4%	98.4%	97.4%
split breadth retention	93.7%	88.1%	87.5%	85.2%

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

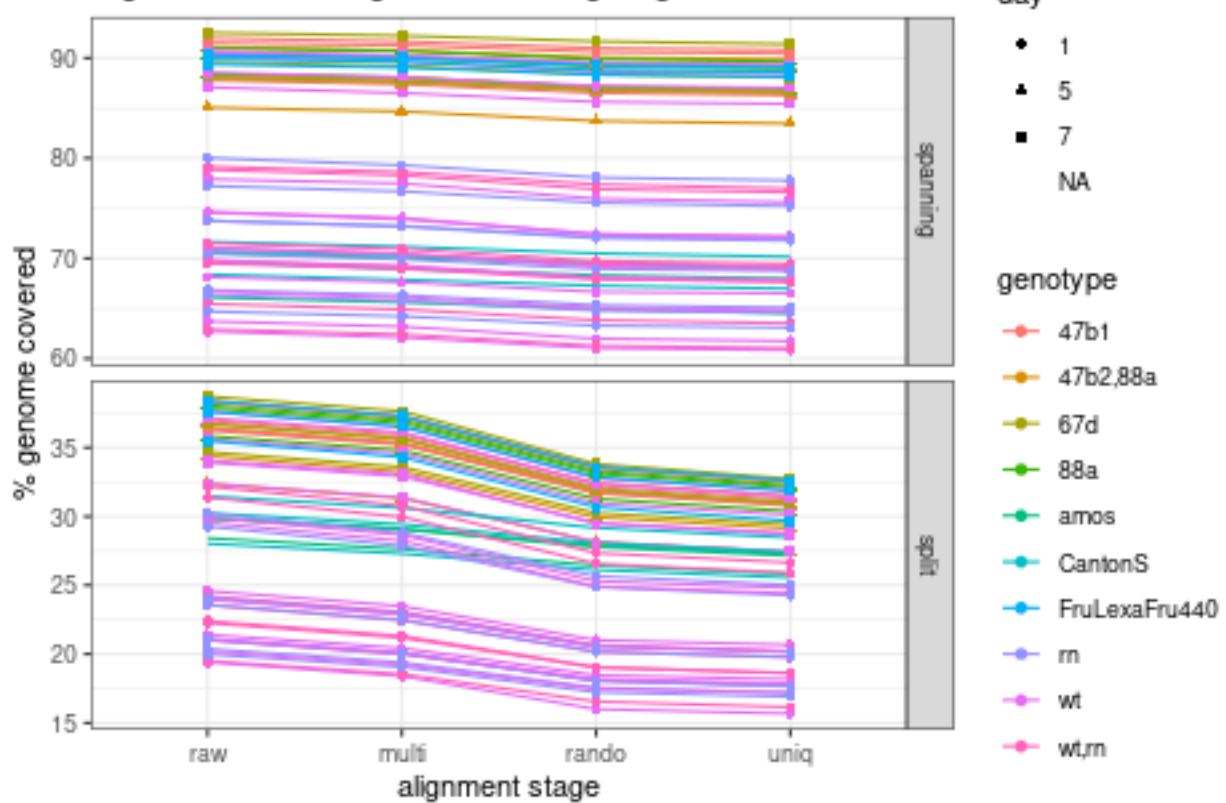
Here are the number of reads per sample, from the intial sequencing to the most heavily filtered alignment:



```
## png
## 2
```

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

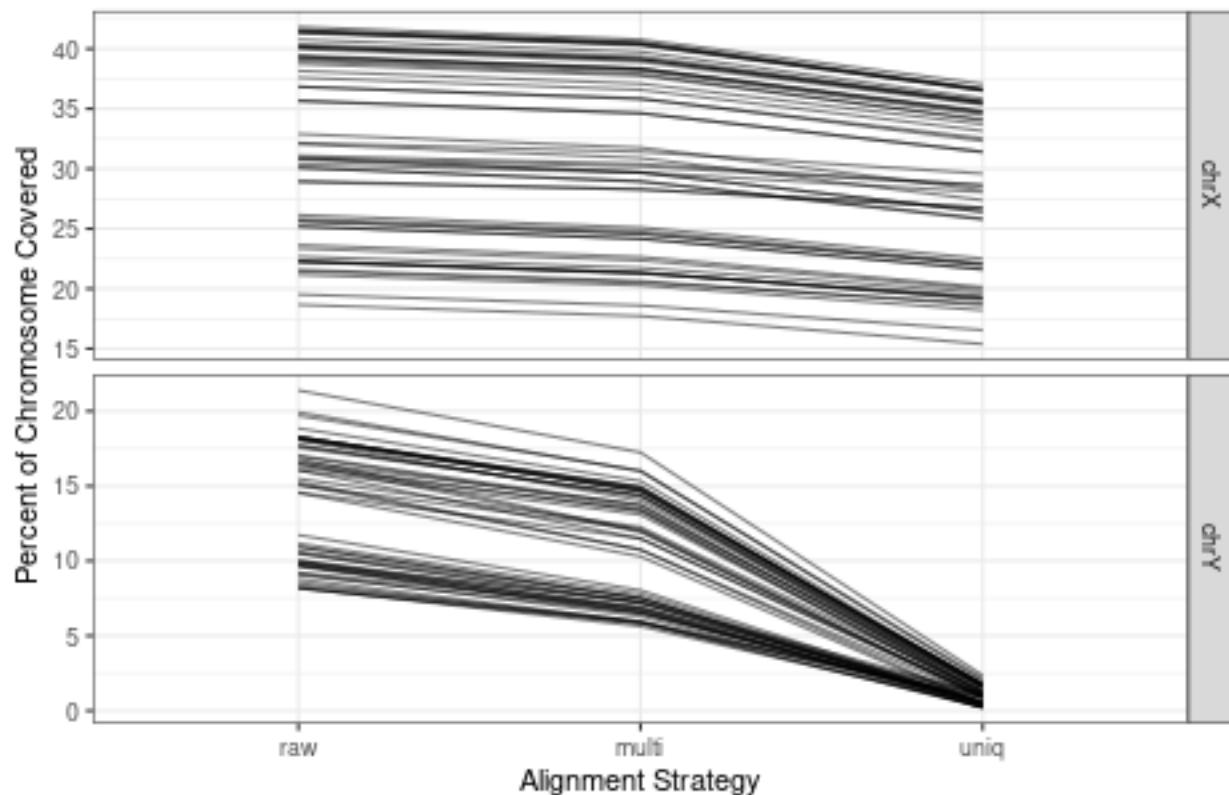
Figure 11. Coverage Loss During Alignment Process



```
## png
## 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



```
## png
## 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

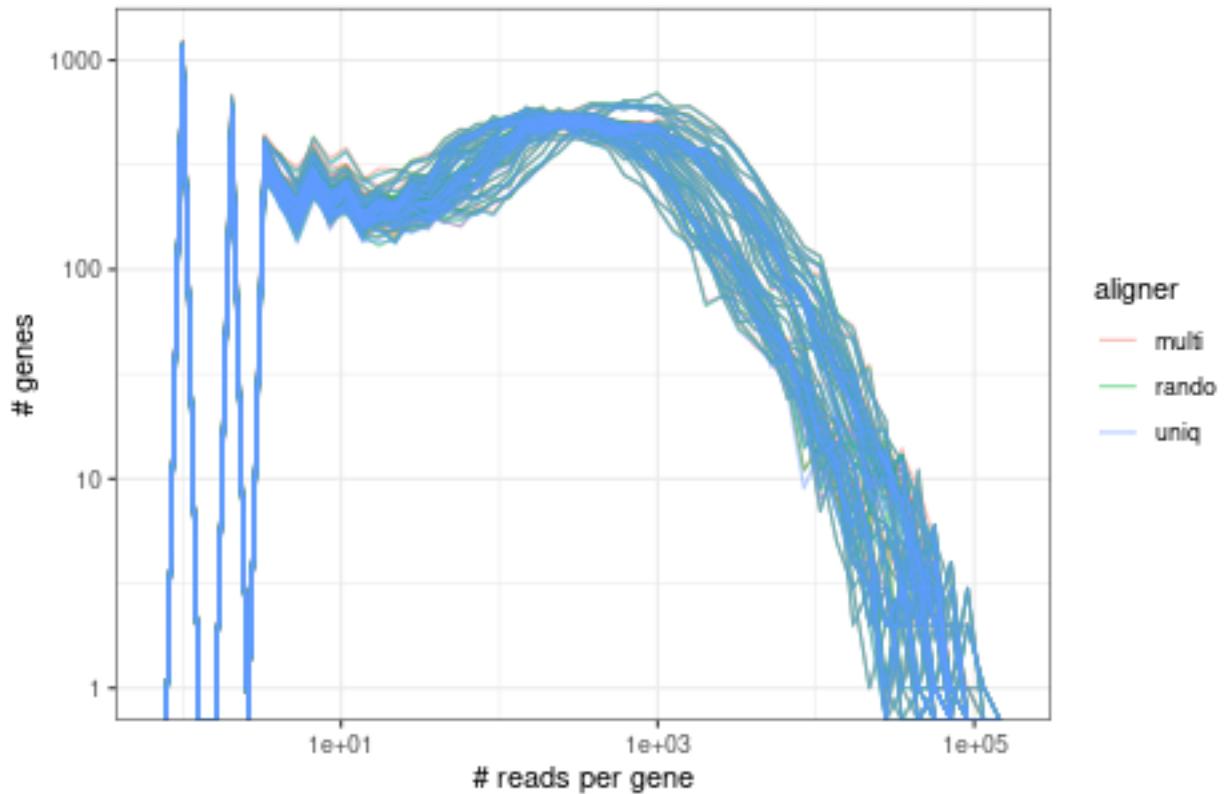
rep	mapping strategy		
	multi	rando	uniq
47b1 - group - 5			
1	90.6%	91.5%	92.1%
2	89.9%	91.1%	91.6%
3	88.7%	89.9%	90.7%
47b2,88a - group - 5			
1	89.8%	91.3%	91.7%
2	89.8%	91.4%	91.9%
3	89.7%	91.2%	91.7%
88a - group - 5			
1	90.4%	91.0%	91.7%
2	90.8%	91.2%	91.8%
3	90.6%	91.2%	91.8%
47b1 - group - 7			
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	3.3%	3.3%	3.3%
No Overlap	21.9%	5.2%	4.6%

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)



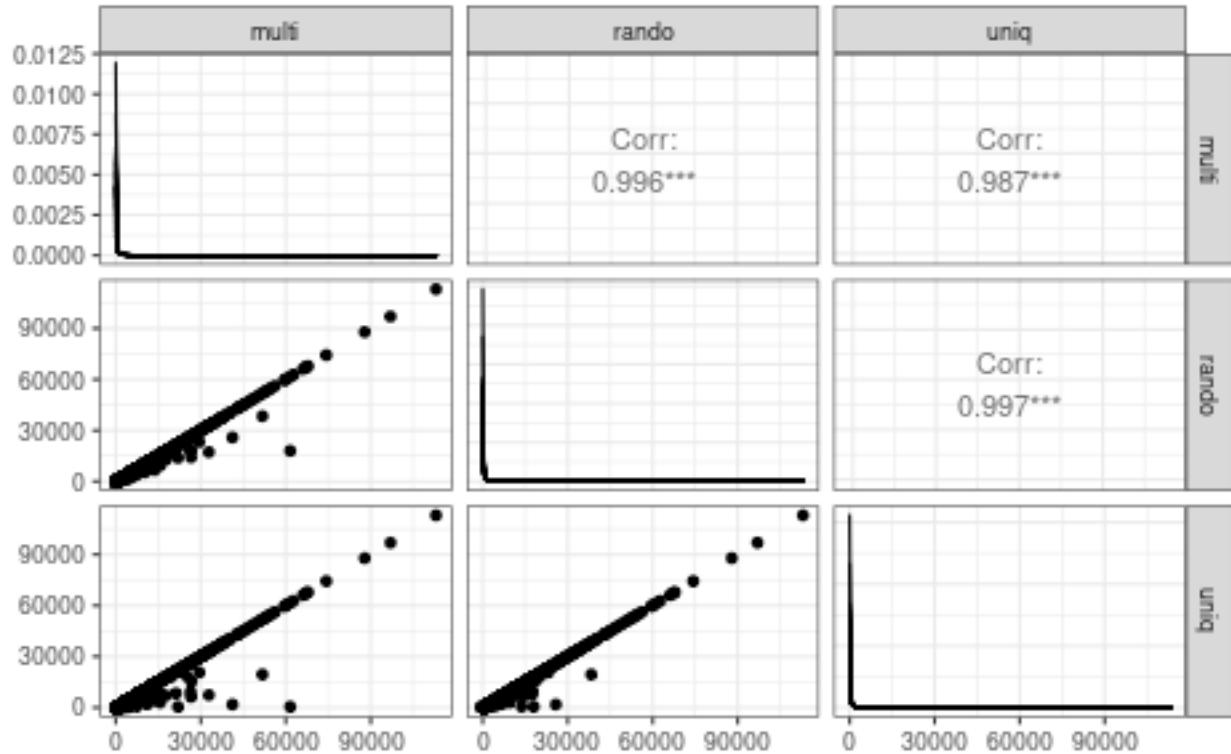
```
## png
## 2
```

One average, a gene had 491.113384014851 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	31.3%	48.2%	62.5%
rando	32.0%	48.5%	62.6%
uniq	32.6%	49.0%	62.9%

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

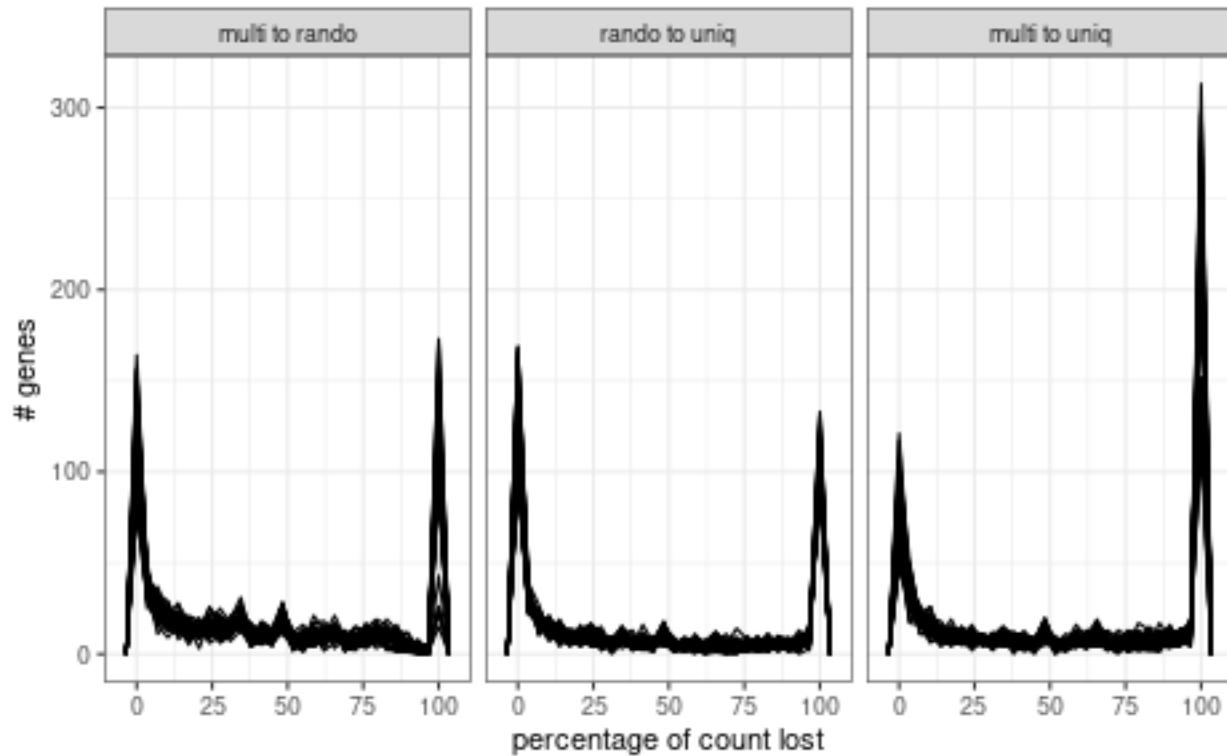


```
## png
## 2
```

The three mapping strategies generally agreed well; for 92.2355631789594 % 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 91.7461861058773 %)

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 (50; 0.00529403045126316 %), rando $>$ multi.

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



```
## png
## 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

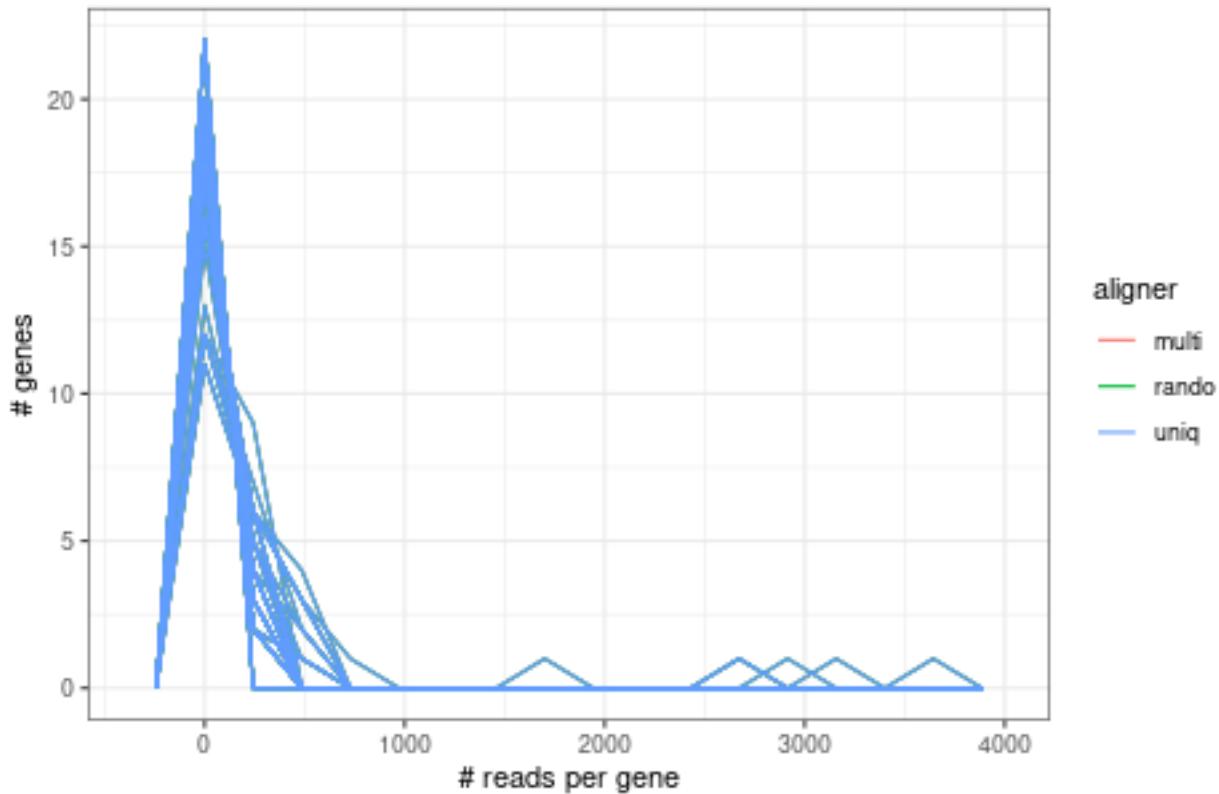
One average, a exon had 78.4876543209877 reads assigned to it, but most exons had relatively fewer, with almost a quarter having no reads assigned at all, more than a third having fewer than 10 reads, and almost two thirds having fewer than 50. These figures are for the “All” assignment strategy, and 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	25.8%	40.2%	64.4%
rando	25.8%	40.2%	64.4%
uniq	25.8%	40.2%	64.4%

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



```
## png
## 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

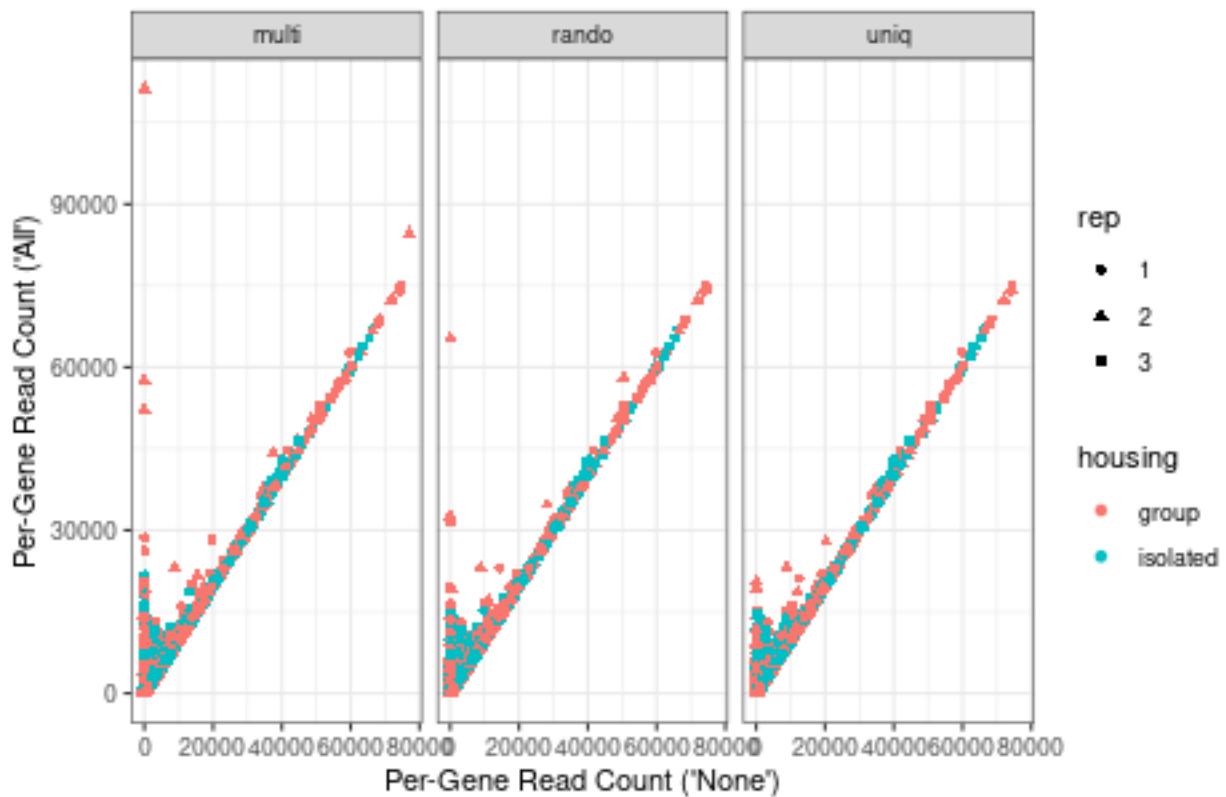
2	198	198	198
3	226	226	226
<hr/>			
47b1 - group - 7			
1	109	109	109
2	103	103	103
3	79	79	79
<hr/>			
67d - group - 7			
1	84	84	84
2	94	94	94
3	66	66	66
<hr/>			
FruLexaFru440 - group - 7			
1	25	25	25
2	76	76	76
3	70	70	70
<hr/>			
wt - group - 7			
1	137	137	137
2	136	136	136
3	103	103	103
<hr/>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



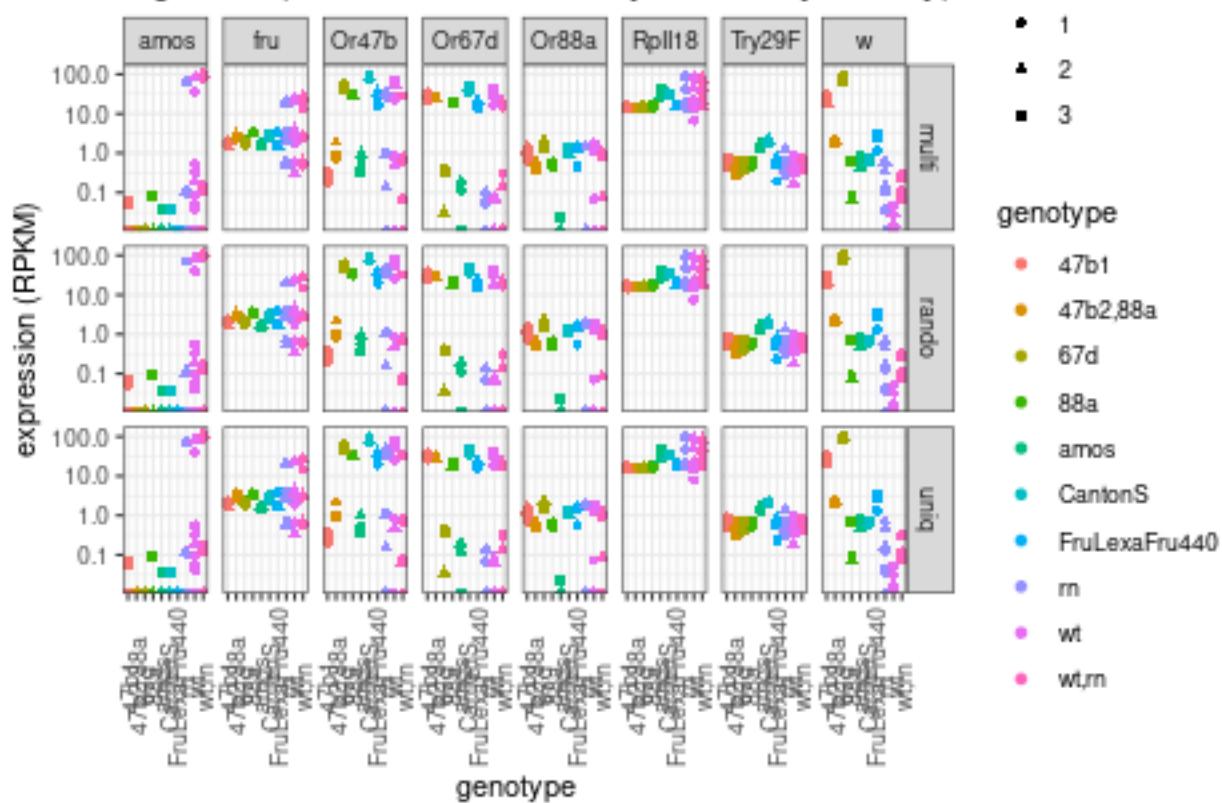
```
## png
## 2

##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## 0.00      0.00      0.00   40.91      2.00 111175.00
```

2.7 Expression

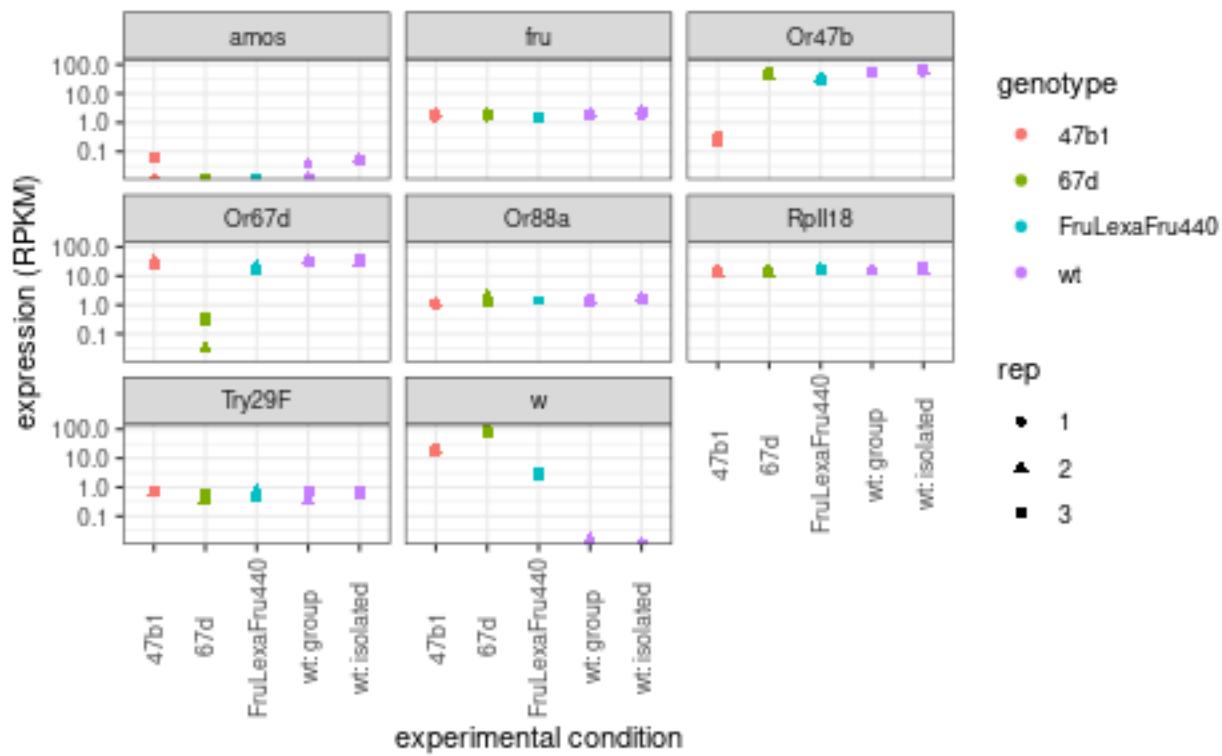
Using the per-gene read counts, the per-alignment total mapped counts, and the gene lengths, the gene expression was calculated as reads per kilobase per million mapped (RPKM). In particular, these can confirm that the knock-outs are not being expressed. This appears to be the case in the 47b and 67d mutants. The Fru440FruLexa mutants do not show any obvious reduction in expression of *Fruitless* (not knockouts - is expression expected though?) CantonS appears similar to wt in all cases. CatnonS-Amos mutants have very low expression of OR47b, OR67d, and OR88a but have typical expression values for *fru* and control genes. 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 ^{rep}



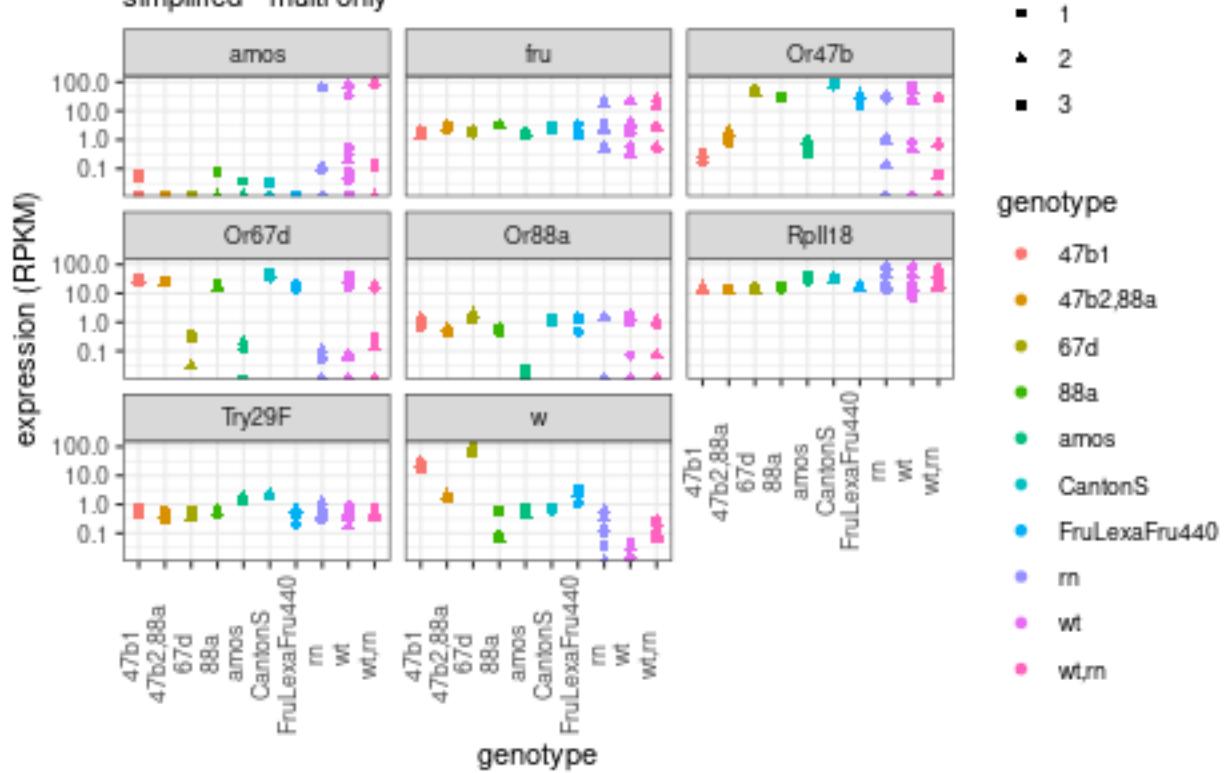
```
## png
## 2
```

Fig 18a. Expression Levels for Key Genes, by Genotype
 Subset of Key Genes, FruLexaFru440 replicate 1 excluded



```
## png
## 2
```

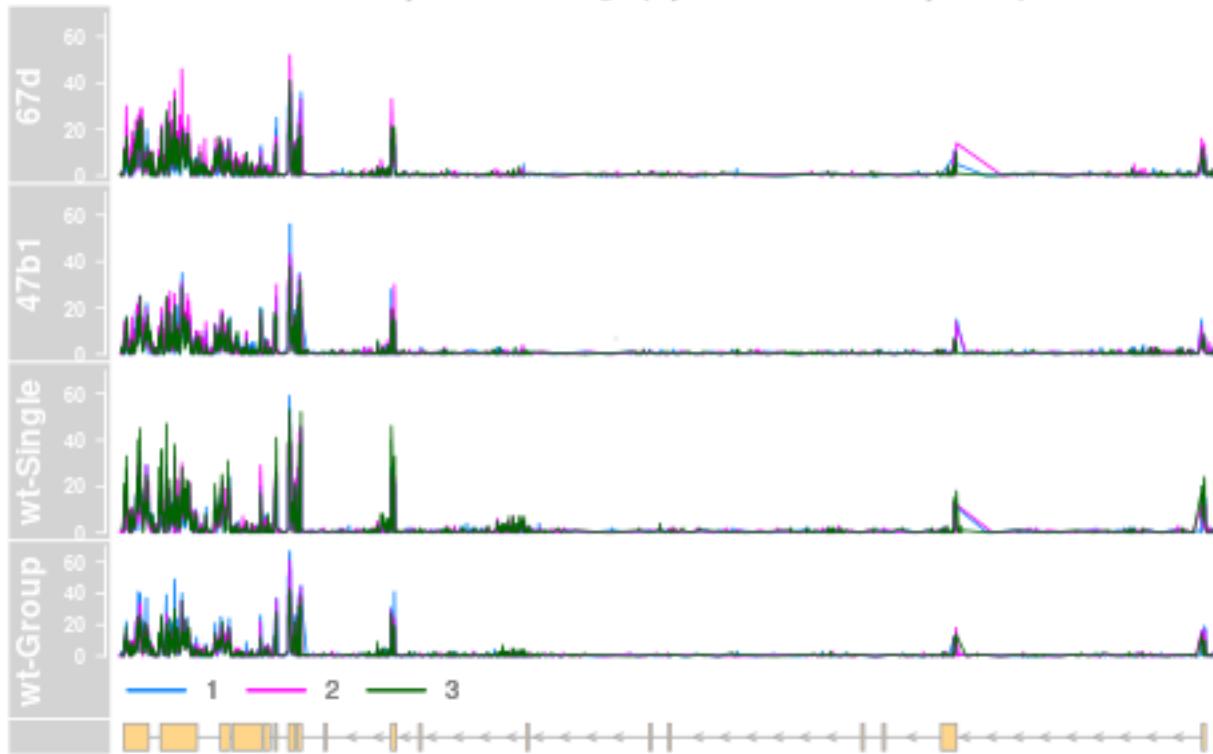
Fig 18b. Expression Levels for Key Genes, by Genotype_{rep}
simplified - multi only



```
## png
## 2
```

2.7.1 Focus on Fruitless

**Figure 19. Focus on Fruitless:
Gene-Scale Depth of Coverage (by treatment and replicate)**



```
## png
## 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.

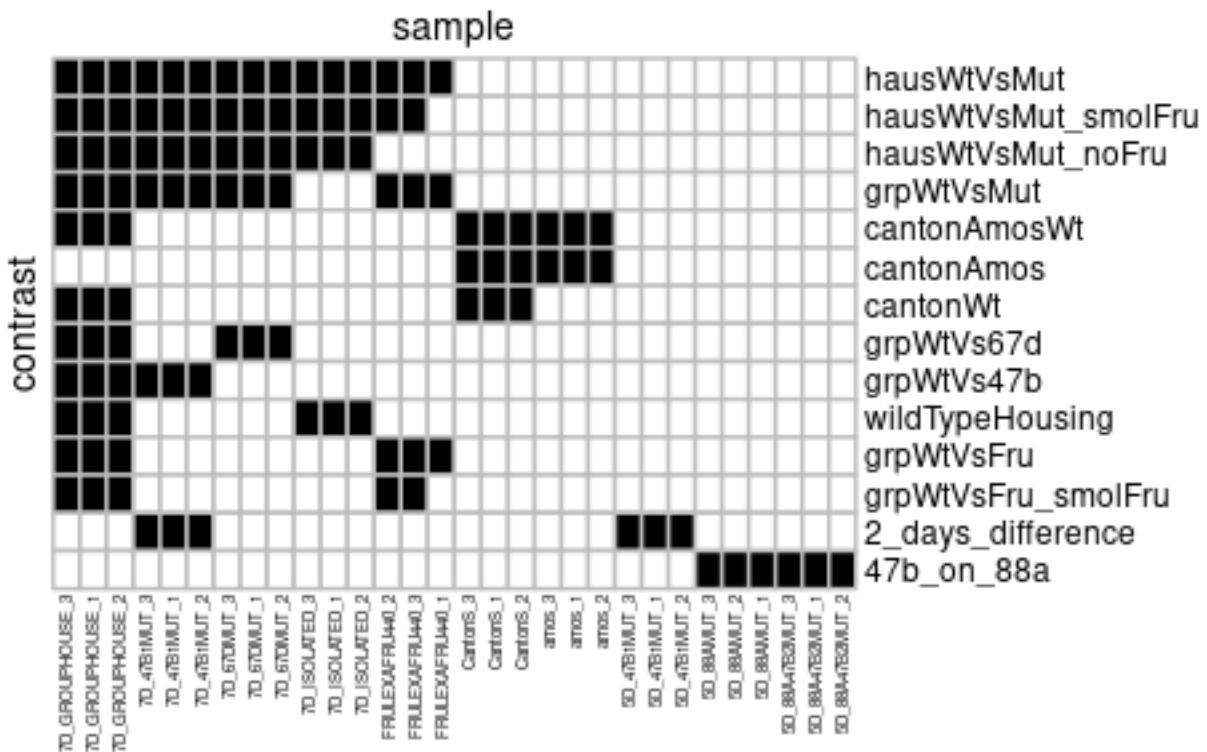
Single-factor models (wildTypeHousing, grpWtVs47b, grpWtVs67d, grpWtVsFru, grpWtVsMut) were built by specifying the axis of comparison (eg, housing) and subsetting samples to the relevant contrast (eg, wt group reps 1,2,3 and wt isolated reps 1,2,3).

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
hausWtVsMut_smolFru	~genotype + housing	genotype: wt, housing: group
grpWtVsFru_smolFru	~genotype	genotype: wt
47b_on_88a	~genotype	genotype: 88a
2_days_difference	~day	day: 5
cantonAmos	~genotype	genotype: CantonS
cantonWt	~genotype	genotype: wt
cantonAmosWt	~normie + cw	normie: regular, cw: w

Figure 20. RNASeq Samples Used in DESeq2 Contrast



```
## png
## 2
```

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 (Benjamini-Hochberg) 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/>; <https://www.biostars.org/p/248486/>)

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.

2.8.1 Differential Exon Use

For a number of reasons, estimating changes in transcript or exon usage is more challenging than estimating coarse-scaled gene expression. Some of approaches here have been to modify the reads/annotations and then analyze within a standard DESeq2 framework. Two approaches outside this schema were also explored.

There is an important distinction between exon USE, which reflects the proportion of expressed transcripts containing a given exon, and exon EXPRESSION, which is the total amount of exon RNA being transcribed (a function of both exon use and gene expression.) (Anders, Reyes, and Huber 2012) In the DESeq2-based approaches, the counts have been subset to the *Fruitless* gene, on the effects of gene expression change will be absorbed into the calculation of sizeFactors, thus any residual changes left represent changes in exon use.

2.8.1.1 The edgeHog

The Fru annotation was divided up into intervals corresponding to exons. The intervals were then subdivided further at exon boundaries, eg if two exons A and B share a 5' edge but B is longer, the interval would be divided into two adjacent intervals, AB and B. Each interval then corresponds to an element of the power set of transcripts, that is, the interval AB corresponds to the set of transcripts which include the AB interval. An interval only has one set of transcript associated, but a set of transcripts can have multiple intervals associated. For example, the set of all transcripts would be associated with the intervals of all constitutive exons.

The intervals were partitioned according to the set of transcripts associated. A new annotation of “isoids” is generated by stitching together the intervals in each partition, and the reads counted. The isoid counts for Fru were provided to DESeq; presumably differences in gene-level expression will be rolled into the size factor estimation and leave behind differences coming from exon use.

For each transcript, the p-value was collected from each isoid associated with the transcript and converted to Z-scores. Stouffer’s test was applied to estimate a significance value for an overall change in relative transcript use.

2.8.1.2 DEXSeq

To compare these results to an established tool, DEXSeq (Anders, Reyes, and Huber 2012) was used. DEXSeq repurposes the DESeq2 statistical methods but modifies the underlying annotation (in a way that doesn’t necessarily respect the exon naming/grouping I’ve used) and counts them in a somewhat idiosyncratic way:

The central data structure for our method is a table that, in the simplest case, contains for each exon of each gene the number of reads in each sample that overlap with the exon. Special attention is needed, however, if an exon’s

boundary is not the same in all transcripts. In such cases, we cut the exon in two or more parts (Fig. 1). We use the term "counting bin" to refer to exons or parts of exons derived in this manner. Note that a read that overlaps with several counting bins of the same gene is counted for each of these.

(Anders, Reyes, and Huber 2012)

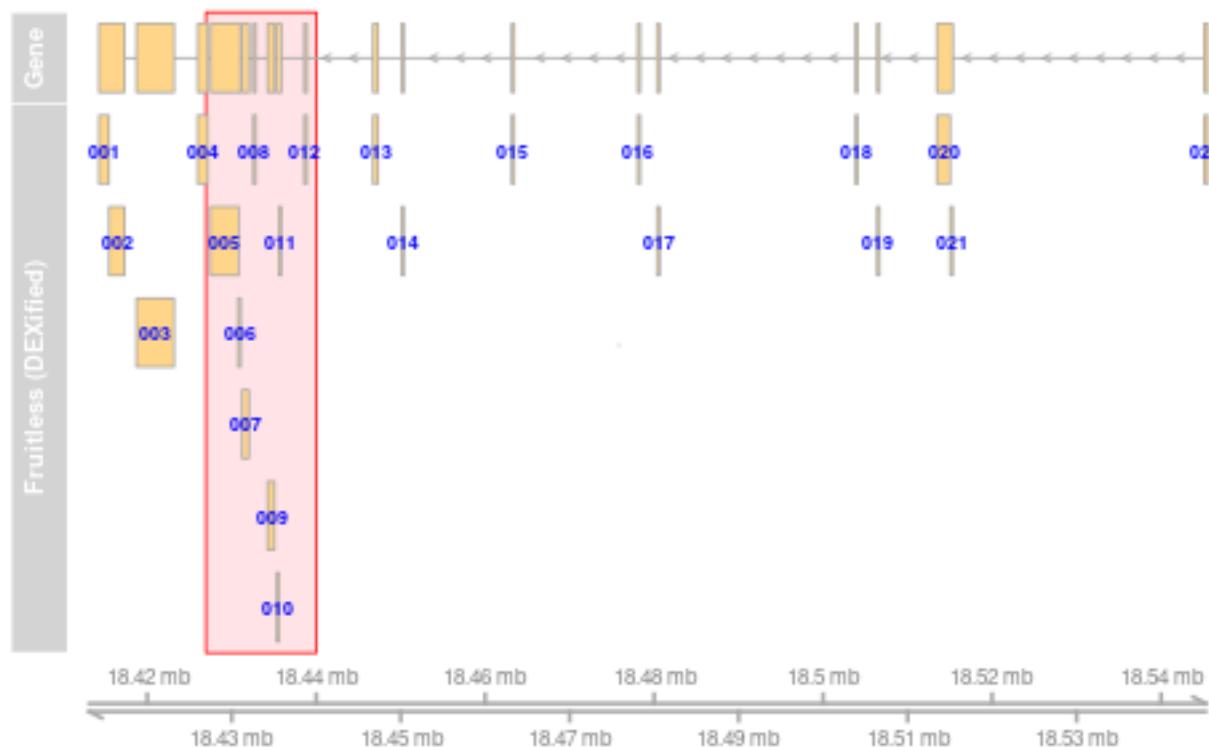
The suggested featureCounts implementation was used: https://github.com/vivekbhr/Subread_to_DEXSeq

In order for the dexseq_prepare_annotation.py script to run correctly, it had to be modified in order to account for transcriptional edge cases (eg, the polycistronic pre-mod(mdg4)-*)

```
for f in HTSeq.GFF_Reader( gtf_file ):
    if f.type != "exon":
        continue
    f.attr['gene_id'] = f.iv.chrom + '_' + f.attr['gene_id'].replace( ":", "_" ) + f.iv.strand # THIS WORKS
    #f.attr['gene_id'] = f.attr['gene_id'].replace( ":", "_" ) # THIS DOESN'T
    exons[f.iv] += ( f.attr['gene_id'], f.attr['transcript_id'] )
```

(source: <https://stat.ethz.ch/pipermail/bioconductor/2012-June/046494.html>)

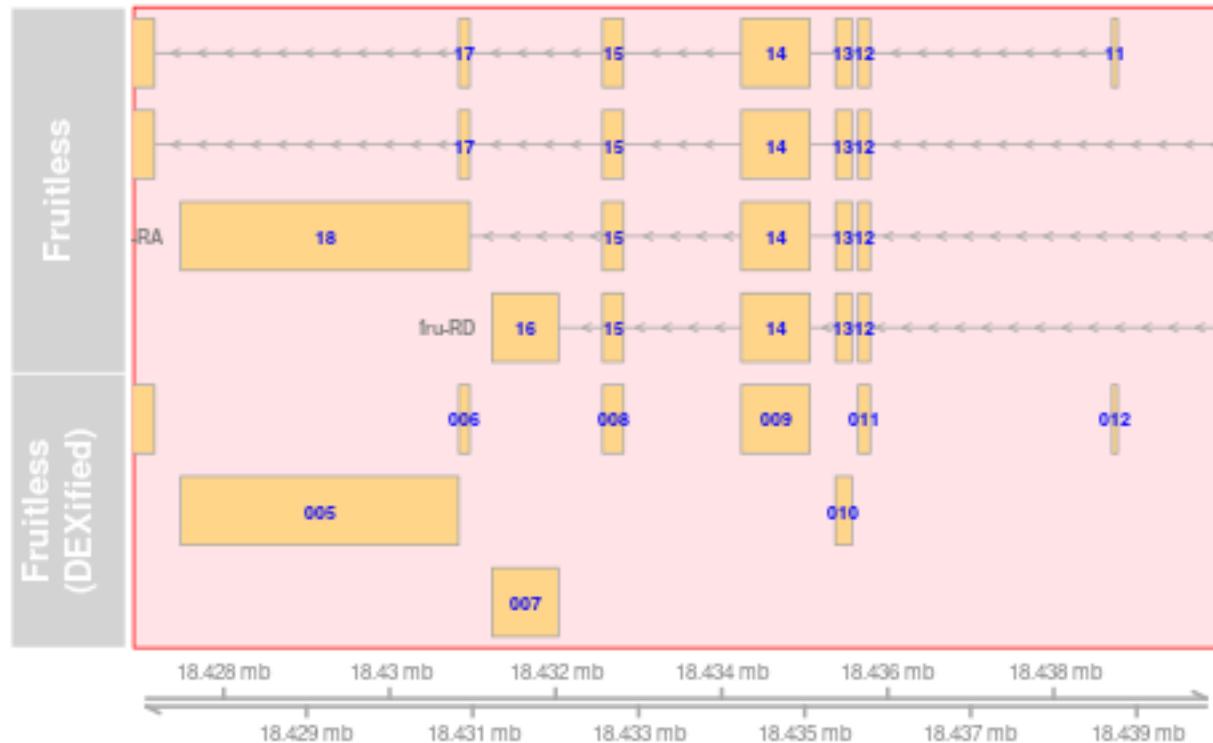
Figure 21. Fruitless gene model: DEXSeq Intervals Derived From Exons



```
## png
## 2
```

Take a careful look at the relationship between exon_18 and intervals 005 and 006.

Figure 21 a. Fruitless gene model: DEXSeq Intervals Derived From Exons (detail)



```
## png
## 2
```

In order to estimate differential exon usage, DEXSeq fits the (size & dispersion-corrected) counts with two models, one containing an interaction term and then other not, and compares the two:

Having the dispersion estimates and the size factors, we can now test for differential exon usage.
For each gene, DEXSeq fits a generalized linear model with the formula
~sample + exon + condition:exon
and compare it to the smaller model (the null model)
~ sample + exon

(Official vignette: <https://bioconductor.org/packages/devel/bioc/vignettes/DEXSeq/inst/doc/DEXSeq.html>)

Models using more than one explanatory variable are possible but more involved & haven't been explored yet

Table 24. Differential Exon Use Contrasts
all modeled as ' ~sample + exon + condition:exon'

	sample subset	reference levels
dex_wtHousing	wtHousing	housing: group
dex_grpWtVs67d	grpWtVs67d	genotype: wt
dex_grpWtVs47b	grpWtVs47b	genotype: wt
dex_grpWtVsFru	grpWtVsFru	genotype: wt

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.

```
## png  
## 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.

```
## png  
## 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.

```
## png  
## 2
```

??what's up with the outliers??

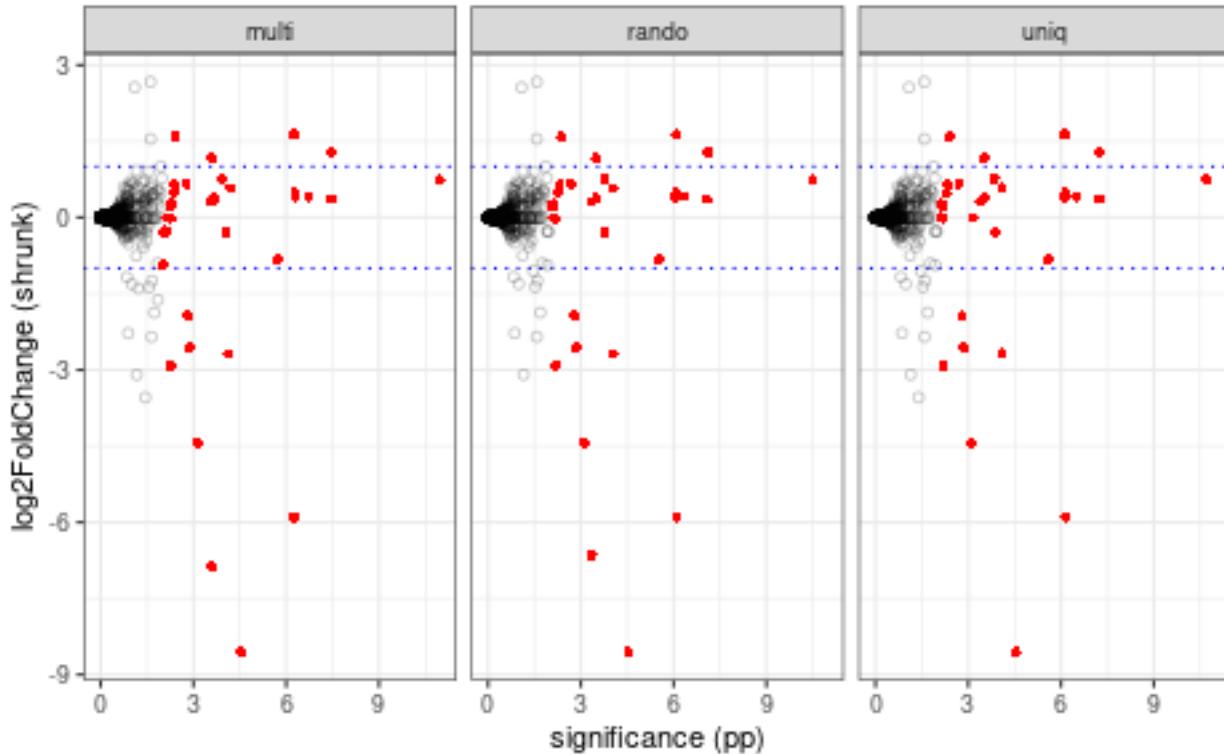
The alignment strategies also agree well when it comes to significance (shrinkage doesn't impact significance so this is the same before and after.)

```
## png  
## 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 ($\text{padj} < 0.01$) 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 depleted relative to the group-housed condition; positive fold changes are enriched

**Figure 26. Volcano Plot: Fold Change vs. Significance
(between isolated and group-housed wildtypes)**



```
## png
## 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 12 such genes, mostly shared across alignment strategy:

Table 26. Genes with Large ($2 <$ fold change), Significant ($\text{padj} < 0.01$) Changes between isolated and group-housed wildtypes

	multi	rando	uniq
TotC	yes	yes	yes
TotA	yes	yes	yes
Amy-p	yes	yes	no
MtnB	yes	yes	yes
CG11852	yes	yes	yes
Prat2	yes	yes	yes
CG15144	yes	yes	yes
CG7470	yes	yes	yes

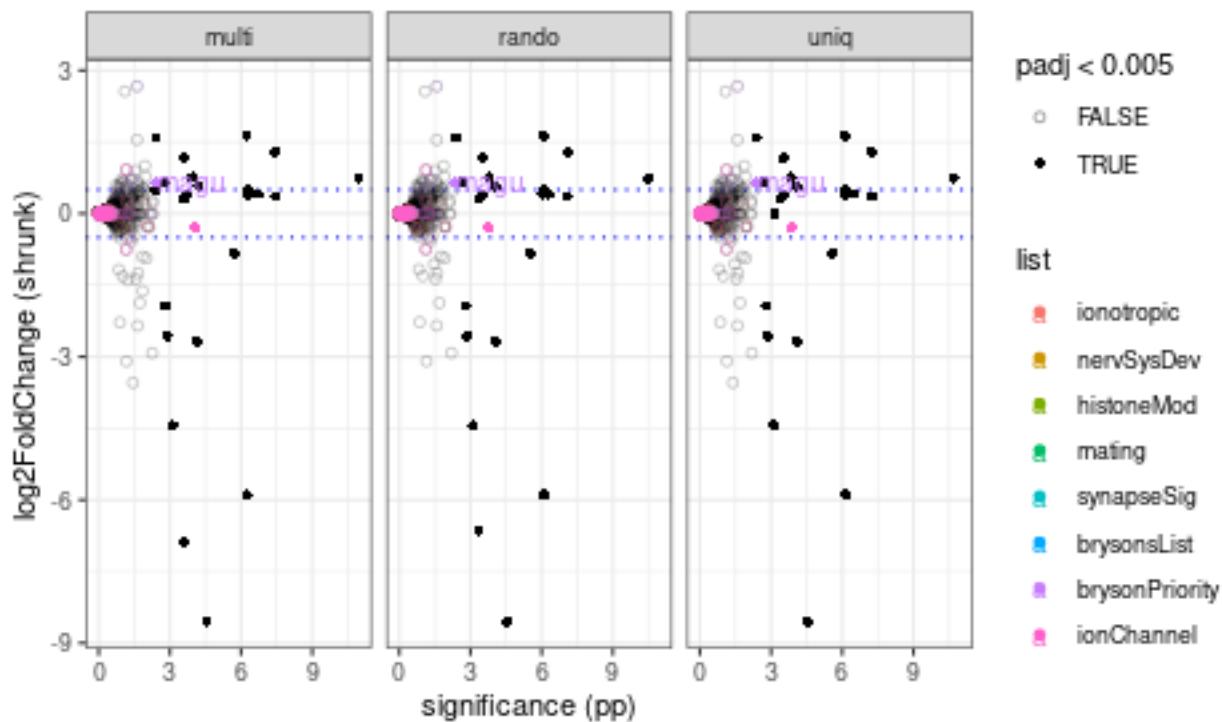
CG2736	yes	yes	yes
amd	yes	yes	yes
CG42369	yes	yes	yes
CG15822	yes	yes	yes

3.1.5 In relation to gene lists

```
## png
## 2
```

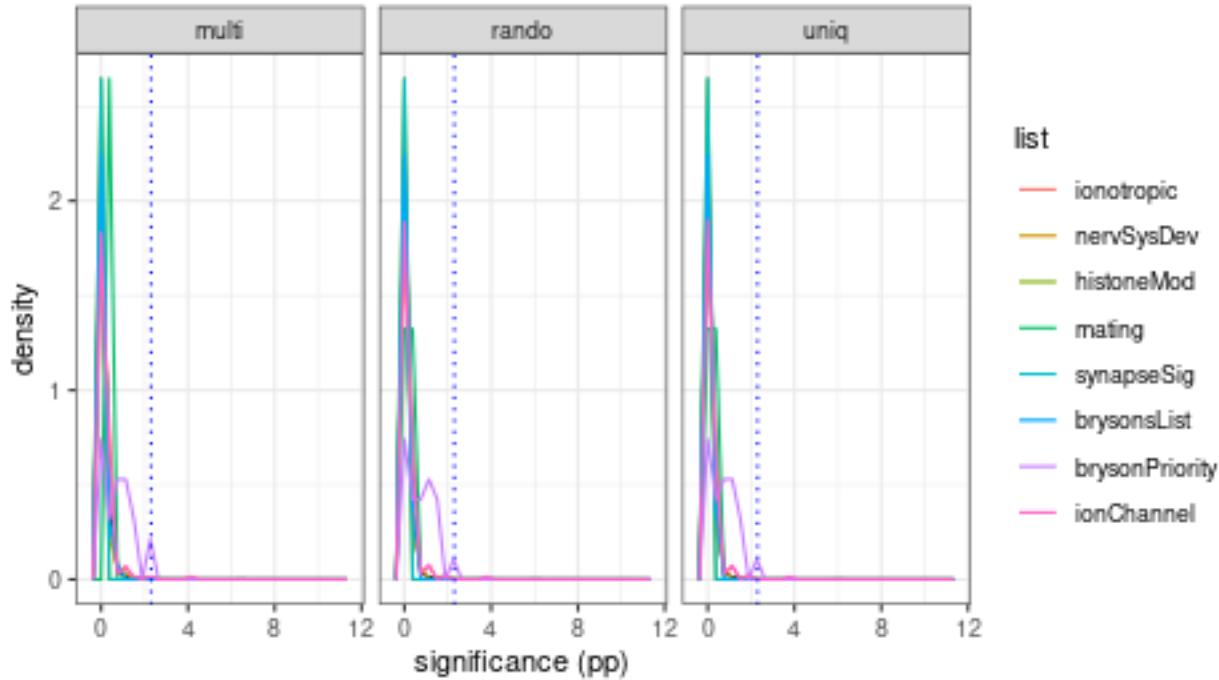
```
## png
## 2
```

Figure 27. Volcano Plot: Fold Change vs. Significance with Gene Lists (between isolated and group-housed wildtypes)
abs(lfc)<0.5 & adjusted p < 0.005 highlighted



```
## png
## 2
```

Figure 28. 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



```
## png
## 2
```

3.1.6 Genes with top 10 most significant changes

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

Table 28. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed between isolated and group-housed wildtypes

rank	multi				rando			
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	CG10050	0.41	1.10×10^{-11}	0.743	CG10050	0.41	3.16×10^{-11}	0.74
2	CG14687	3.39	3.56×10^{-8}	0.361	MtnB	0.85	7.61×10^{-8}	1.28
3	MtnB	0.85	3.56×10^{-8}	1.289	CG14687	3.39	8.39×10^{-8}	0.36
4	CG31663	0.47	1.90×10^{-7}	0.412	CG31663	0.47	4.94×10^{-7}	0.41
5	Cln3	0.87	5.14×10^{-7}	0.404	TotA	0.13	8.09×10^{-7}	-5.89
6	Obp84a	0.93	5.14×10^{-7}	0.486	CG11852	0.18	8.13×10^{-7}	1.63
7	TotA	0.13	5.54×10^{-7}	-5.899	Obp84a	0.93	8.13×10^{-7}	0.48
8	CG11852	0.18	5.67×10^{-7}	1.634	Cln3	0.87	9.32×10^{-7}	0.40
9	Dhc36C	0.03	1.95×10^{-6}	-0.828	Dhc36C	0.03	3.04×10^{-6}	-0.82
10	TotC	0.05	2.82×10^{-5}	-8.554	TotC	0.05	2.88×10^{-5}	-8.55

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 29. Top Ten Largest Magnitude Fold Changes which were Significant between isolated and group-housed wildtypes

rank	multi					rando			
	name	expression	adjusted p	log2 FoldChange		name	expression	adjusted p	log2 FoldChange
1	TotC	0.05	2.82×10^{-5}	-8.554		TotC	0.05	2.88×10^{-5}	-8.554
2	Amy-p	0.02	2.57×10^{-4}	-6.879		Amy-p	0.02	4.46×10^{-4}	-6.650
3	TotA	0.13	5.54×10^{-7}	-5.899		TotA	0.13	8.09×10^{-7}	-5.897
4	Prat2	0.01	7.54×10^{-4}	-4.442		Prat2	0.01	7.80×10^{-4}	-4.441
5	CG2736	0.01	5.55×10^{-3}	-2.921		CG2736	0.01	6.42×10^{-3}	-2.920
6	CG15144	0.01	7.54×10^{-5}	-2.682		CG15144	0.01	9.00×10^{-5}	-2.681
7	CG7470	0.01	1.33×10^{-3}	-2.564		CG7470	0.01	1.40×10^{-3}	-2.563
8	CG42369	0.03	1.50×10^{-3}	-1.936		CG42369	0.03	1.63×10^{-3}	-1.935
9	CG11852	0.18	5.67×10^{-7}	1.634		CG11852	0.18	8.13×10^{-7}	1.633
10	CG15822	0.01	3.85×10^{-3}	1.593		CG15822	0.01	4.21×10^{-3}	1.592

3.1.8 Top 10 highest expressed genes with significant change

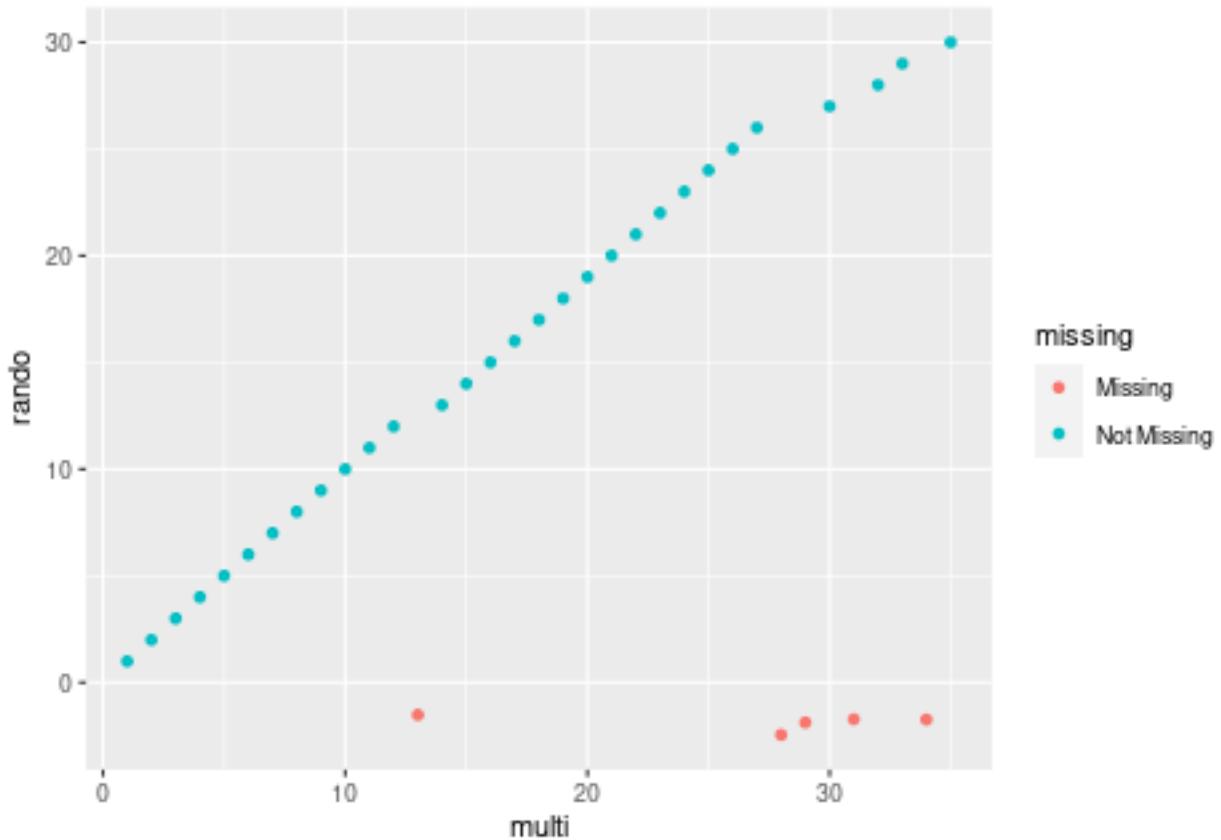
Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

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

rank	multi					rando			
	name	expression	adjusted p	log2 FoldChange		name	expression	adjusted p	log2 FoldChange
1	vir-1	6.81	5.37×10^{-3}	0.213		vir-1	6.81	8.16×10^{-3}	0.209
2	CG14687	3.39	3.56×10^{-8}	0.361		CG14687	3.39	8.39×10^{-8}	0.360
3	amd	1.76	2.57×10^{-4}	1.171		amd	1.76	3.16×10^{-4}	1.169
4	CG7888	1.55	5.37×10^{-3}	0.268		CG7888	1.55	7.80×10^{-3}	0.266
5	Obp84a	0.93	5.14×10^{-7}	0.486		Obp84a	0.93	8.13×10^{-7}	0.485
6	Cln3	0.87	5.14×10^{-7}	0.404		Cln3	0.87	9.32×10^{-7}	0.402
7	MtnB	0.85	3.56×10^{-8}	1.289		MtnB	0.85	7.61×10^{-8}	1.288
8	CG15202	0.84	2.06×10^{-4}	0.385		CG15202	0.84	3.16×10^{-4}	0.383
9	CG5745	0.78	2.57×10^{-4}	0.310		CG5745	0.78	4.46×10^{-4}	0.309
10	msps	0.48	7.33×10^{-3}	-0.288		CG31663	0.47	4.94×10^{-7}	0.412

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

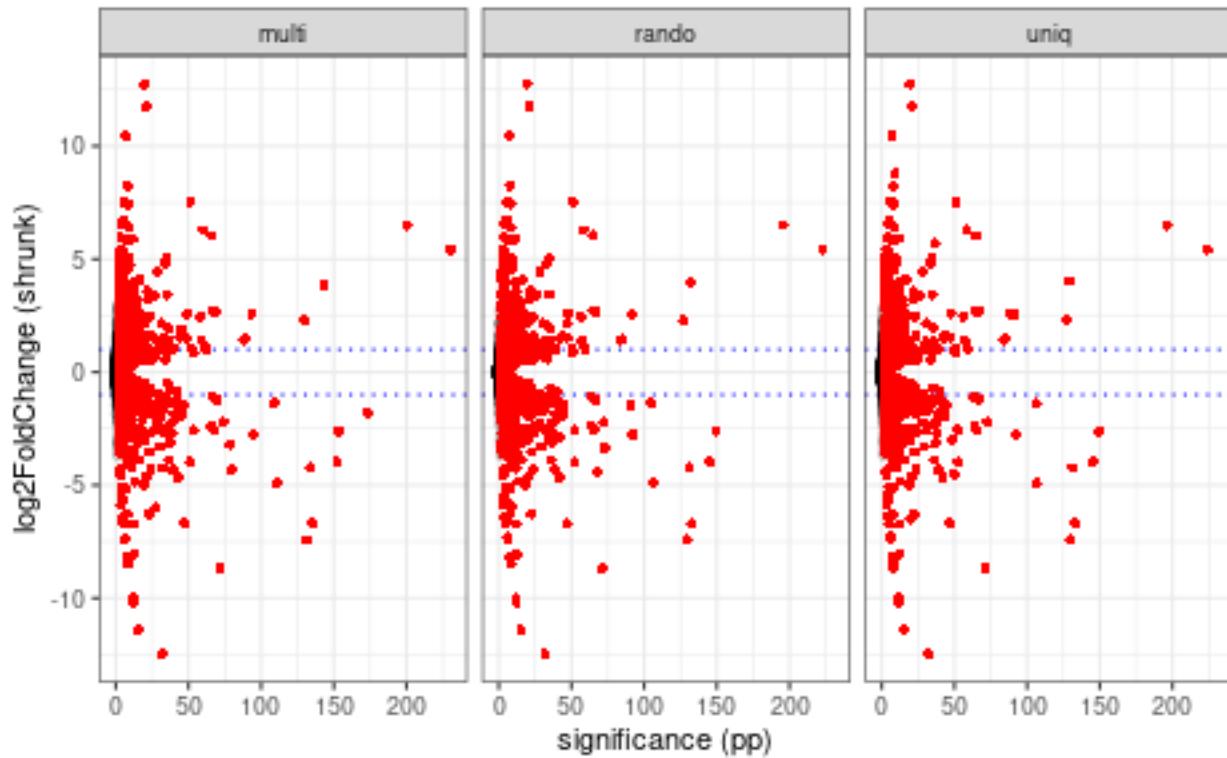
```
## png  
## 2
```

```
## png  
## 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 ($\text{padj} < 0.01$) 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 depletion relative to the group-housed condition; positive fold changes are enriched

**Figure 31. Volcano Plot: Fold Change vs. Significance
(between group-housed wildtypes and 47b mutants)**



```
## png
## 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, $\text{padj} < 0.01$) changes. There were 584 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:

Table 33. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Exp between group-housed wildtypes and 47b mutants

rank	multi					rando				
	name	expression	adjusted p	log2 FoldChange		name	expression	adjusted p	log2 FoldCh	
1	CG7900	0.97	4.10×10^{-231}	5.424		CG7900	0.97	1.09×10^{-223}		
2	CG6912	0.66	4.21×10^{-201}	6.495		CG6912	0.66	3.45×10^{-196}		
3	Unc-115a	0.83	5.42×10^{-174}	-1.785		Drip	1.15	3.82×10^{-150}		
4	Drip	1.15	5.10×10^{-154}	-2.613		Idgf2	0.44	5.90×10^{-146}		
5	Idgf2	0.44	5.30×10^{-153}	-3.968		5-HT2A	0.13	2.15×10^{-133}		
6	phr	0.19	6.29×10^{-144}	3.863		phr	0.18	1.27×10^{-132}		
7	5-HT2A	0.13	4.66×10^{-136}	-6.698		DIP-alpha	0.09	6.27×10^{-132}		

8	DIP-alpha	0.09	1.95×10^{-134}	-4.232	Or47b	0.71	2.81×10^{-130}	-
9	Or47b	0.71	6.05×10^{-132}	-7.427	Cyp9b2	2.73	6.80×10^{-128}	-
10	Cyp9b2	2.73	1.65×10^{-130}	2.299	CAH9	0.35	6.26×10^{-107}	-

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 34. Top Ten Largest Magnitude Fold Changes which were significant
between group-housed wildtypes and 47b mutants

rank	name	multi			rando		
		expression	adjusted p	log2 FoldChange	name	expression	adjusted p
1	mthl8	0.07	2.87×10^{-20}	12.745	mthl8	0.07	3.24×10^{-20}
2	CG40486	2.45	9.31×10^{-33}	-12.470	CG40486	2.45	1.07×10^{-32}
3	w	0.27	1.26×10^{-21}	11.751	w	0.27	1.33×10^{-21}
4	CG30428	0.11	3.71×10^{-16}	-11.411	CG30428	0.10	4.04×10^{-16}
5	CG43149	0.08	8.92×10^{-8}	10.451	CG43149	0.08	9.10×10^{-8}
6	ppk19	0.04	8.21×10^{-13}	-10.191	ppk19	0.04	9.11×10^{-13}
7	lncRNA:CR45502	0.07	1.59×10^{-12}	-10.008	lncRNA:CR45502	0.07	1.78×10^{-12}
8	Cyp6a17	0.46	2.11×10^{-72}	-8.681	Cyp6a17	0.46	8.16×10^{-72}
9	asRNA:CR44030	0.03	4.15×10^{-9}	-8.496	asRNA:CR44030	0.03	4.62×10^{-9}
10	lncRNA:CR44377	0.01	3.12×10^{-8}	-8.244	lncRNA:CR44377	0.01	3.40×10^{-8}

3.2.1.5 Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

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

Table 34. Top Ten Highest Expressed Genes with Significant (parametric) Difference
between group-housed wildtypes and 47b mutants

rank	name	multi			rando		
		expression	adjusted p	log2 FoldChange	name	expression	adjusted p
1	Obp83b	59.65	6.56×10^{-3}	0.294	Obp83b	59.65	7.80×10^{-3}
2	Obp83a	50.66	3.38×10^{-7}	0.404	Obp83a	50.67	5.35×10^{-7}
3	Obp28a	32.78	8.31×10^{-13}	0.505	Obp28a	32.78	2.20×10^{-12}
4	OS9	32.70	2.51×10^{-3}	0.268	OS9	32.70	3.15×10^{-3}
5	Obp19a	29.14	4.25×10^{-3}	0.201	Obp19a	29.15	5.43×10^{-3}
6	Ugt35B1	20.33	6.00×10^{-4}	0.553	Ugt35B1	20.34	7.16×10^{-4}
7	antdh	19.15	1.22×10^{-3}	0.266	antdh	19.15	1.49×10^{-3}
8	DrsI5	15.18	2.58×10^{-3}	-0.841	DrsI5	15.18	3.05×10^{-3}
9	EbpIII	14.40	8.27×10^{-7}	0.268	EbpIII	14.40	1.19×10^{-6}
10	lncRNA:noe	12.88	1.33×10^{-8}	-0.875	lncRNA:noe	12.88	2.21×10^{-8}

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:

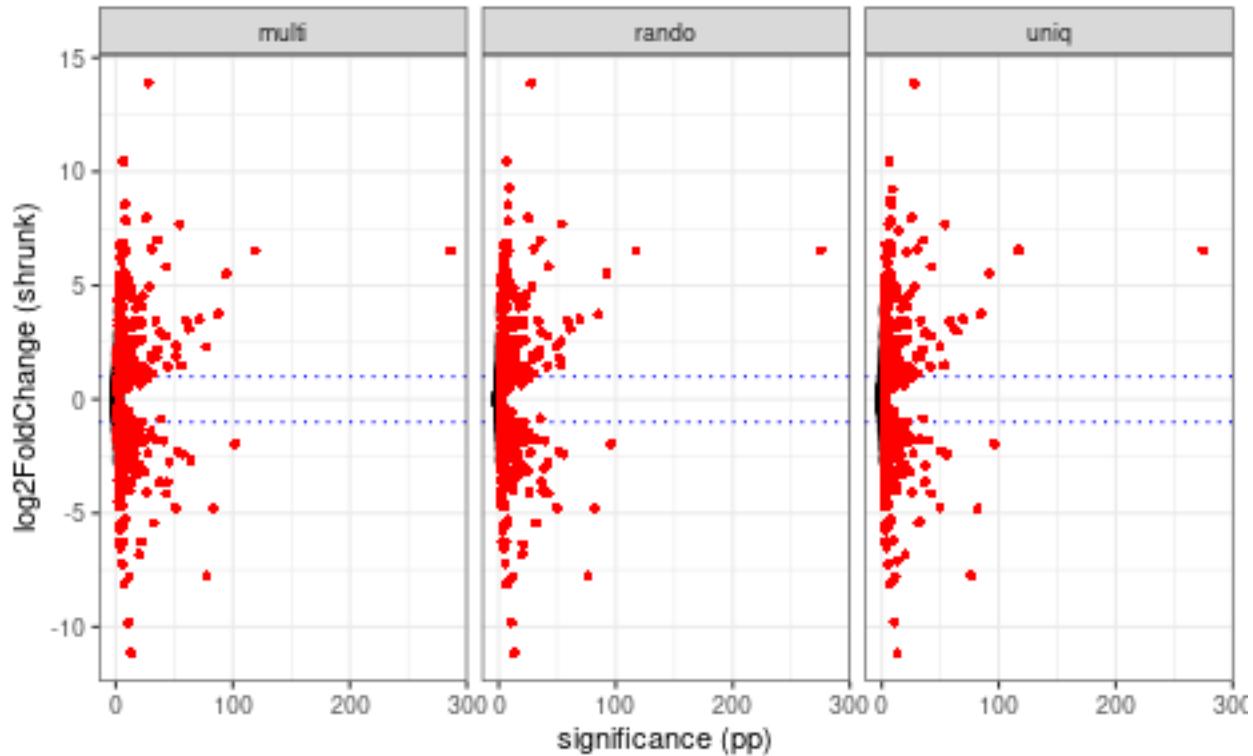
3.2.2.1 preshrunk comparison across alignment strategies

```
## png  
## 2  
  
## png  
## 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 ($\text{padj} < 0.01$) 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 depleted relative to the group-housed condition; positive fold changes are enriched

Figure 34. Volcano Plot: Fold Change vs. Significance (between group-housed wildtypes and 67d mutants)



```
## png  
## 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 546 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.

Table 38. Top Ten Most Significantly ($p\text{adj}<0.01$) Differentially Exp between group-housed wildtypes and 67d mutants

rank	multi					rando			
	name	expression	adjusted p	log2 FoldChange		name	expression	adjusted p	log2 FoldCh
1	l(2)03659	0.43	6.27×10^{-287}	6.532		l(2)03659	0.43	5.75×10^{-276}	
2	CG7900	2.03	1.47×10^{-119}	6.528		CG7900	2.03	5.71×10^{-118}	
3	NijC	1.26	2.20×10^{-102}	-1.951		NijC	1.26	4.85×10^{-97}	
4	CG6912	0.33	6.47×10^{-95}	5.521		CG6912	0.33	9.74×10^{-93}	
5	ppk25	0.31	2.06×10^{-88}	3.735		ppk25	0.31	8.49×10^{-86}	
6	DIP-alpha	0.09	3.47×10^{-84}	-4.810		DIP-alpha	0.09	1.05×10^{-82}	
7	5-HT2A	0.13	3.06×10^{-78}	-7.755		5-HT2A	0.13	5.38×10^{-77}	
8	CG32641	3.78	3.06×10^{-78}	2.309		CG10936	0.11	4.84×10^{-70}	
9	CG10936	0.11	6.28×10^{-72}	3.513		Cyp9b1	0.86	7.33×10^{-62}	
10	CR10102	0.47	1.19×10^{-64}	-2.697		CG32407	0.28	3.07×10^{-59}	

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

rank	multi					rando			
	name	expression	adjusted p	log2 FoldChange		name	expression	adjusted p	
1	w	1.13	1.57×10^{-28}	13.891		w	1.13	1.99×10^{-28}	
2	CG32437	0.05	4.10×10^{-14}	-11.144		CG32437	0.05	4.77×10^{-14}	
3	CG43149	0.08	2.97×10^{-7}	10.457		CG43149	0.08	3.09×10^{-7}	
4	lncRNA:CR44111	0.07	1.65×10^{-11}	-9.800		lncRNA:CR44111	0.07	1.95×10^{-11}	
5	ppk9	0.01	9.78×10^{-9}	8.541		CG43291	0.01	8.72×10^{-10}	
6	lncRNA:CR44377	0.01	1.42×10^{-7}	-8.112		ppk9	0.01	1.14×10^{-8}	
7	lncRNA:dntRL	0.15	1.03×10^{-26}	7.975		lncRNA:CR44377	0.01	1.64×10^{-7}	
8	Obp83g	0.08	4.39×10^{-9}	7.840		lncRNA:dntRL	0.15	2.37×10^{-26}	
9	CG9010	0.08	3.16×10^{-12}	-7.815		Obp83g	0.08	5.09×10^{-9}	
10	5-HT2A	0.13	3.06×10^{-78}	-7.755		CG9010	0.08	3.68×10^{-12}	

3.2.2.5 Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

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

Table 40. Top Ten Highest Expressed Genes with Significant ($\text{padj} < 0.01$) Difference
between group-housed wildtypes and 67d mutants

rank	multi				rando			
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	Obp69a	25.91	1.47×10^{-6}	0.678	Obp69a	25.91	2.36×10^{-6}	0.678
2	Cyp6w1	13.12	1.35×10^{-3}	0.581	Cyp6w1	13.12	1.69×10^{-3}	0.561
3	lush	13.09	7.98×10^{-6}	0.772	lush	13.09	1.16×10^{-5}	0.761
4	Obp56d	8.70	3.59×10^{-3}	0.735	Obp56d	8.70	4.29×10^{-3}	0.735
5	Cyp6a2	7.20	2.80×10^{-43}	2.777	Cyp6a2	7.20	2.04×10^{-42}	2.777
6	CG30197	6.35	1.42×10^{-7}	1.025	CG30197	6.35	2.21×10^{-7}	1.025
7	CG34220	5.68	7.23×10^{-8}	6.861	CG34220	5.68	7.70×10^{-8}	6.861
8	Cyp6a20	5.66	5.00×10^{-42}	-1.814	Cyp6a20	5.66	1.58×10^{-40}	-1.814
9	Aldh	5.03	1.02×10^{-3}	-0.528	Aldh	5.03	1.41×10^{-3}	-0.528
10	CG10357	4.95	3.58×10^{-9}	1.326	CG10357	4.95	5.48×10^{-9}	1.326

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:

3.2.3.1 preshrunk comparison across alignment strategies

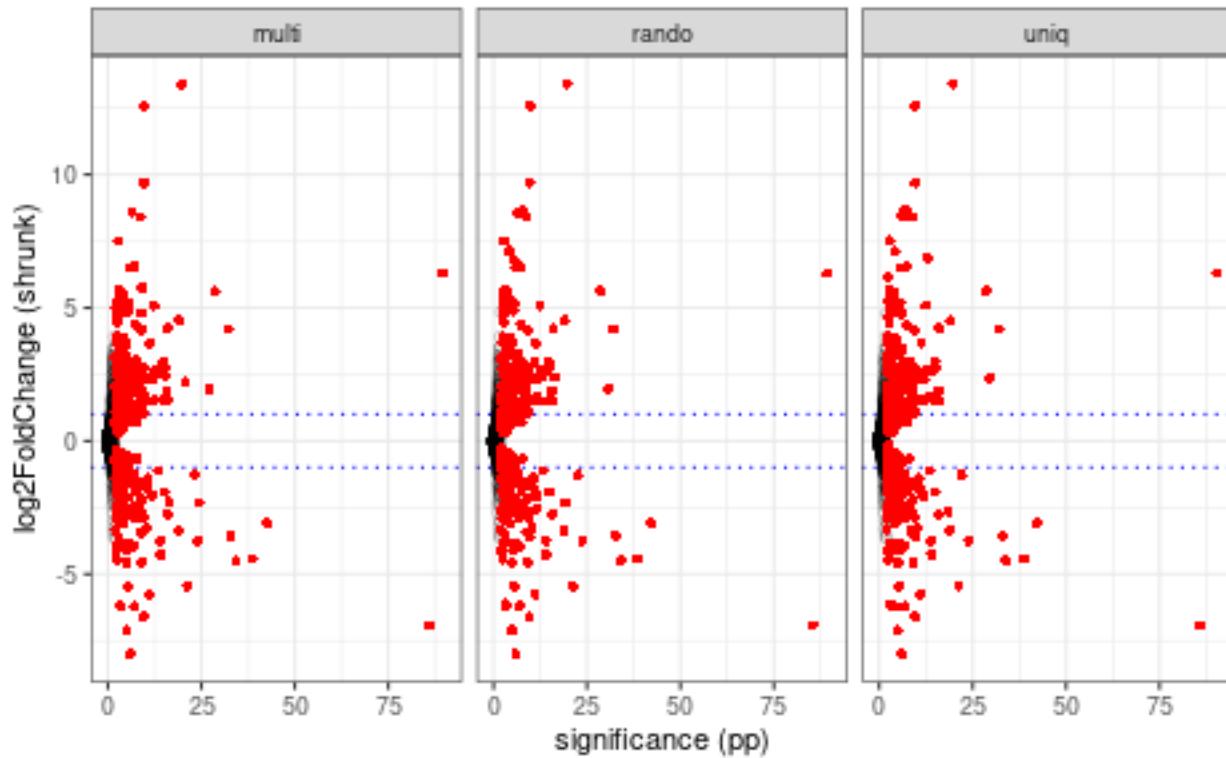
```
## png
## 2

## png
## 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 ($\text{padj} < 0.01$) 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 depleted relative to the group-housed condition; positive fold changes are enriched

**Figure 37. Volcano Plot: Fold Change vs. Significance
(between group-housed wildtypes and 67d mutants)**



```
## png
## 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 343 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 42. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed between group-housed wildtypes and Fru mutants

multi					rando			
rank	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	CG7900	1.64	3.02×10^{-90}	6.300	CG7900	1.64	5.57×10^{-90}	6
2	5-HT2A	0.12	7.96×10^{-87}	-6.925	5-HT2A	0.12	3.94×10^{-86}	-6
3	Ets21C	0.06	2.27×10^{-43}	-3.100	Ets21C	0.06	8.07×10^{-43}	-3
4	BomBc1	0.45	1.73×10^{-39}	-4.423	BomBc1	0.45	3.90×10^{-39}	-4
5	DIP-alpha	0.08	5.38×10^{-35}	-4.495	DIP-alpha	0.08	8.56×10^{-35}	-4
6	BomS1	0.90	1.39×10^{-33}	-3.565	BomS1	0.90	2.22×10^{-33}	-3
7	CG42526	0.09	4.70×10^{-33}	4.202	CG42526	0.09	7.32×10^{-33}	4

8	CG11893	0.32	2.44×10^{-29}	5.619	CG32640	1.91	1.69×10^{-31}	1
9	CG32640	2.71	6.26×10^{-28}	1.945	CG11893	0.32	2.36×10^{-29}	5
10	Cyp12d1-p	0.08	4.17×10^{-25}	-2.314	prom	0.05	1.53×10^{-24}	-3

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 43. Top Ten Largest Magnitude Fold Changes which were significant between group-housed wildtypes and Fru mutants

rank	name	multi			rando		
		expression	adjusted p	log2 FoldChange	name	expression	adjusted p
1	mthl8	0.10	1.73×10^{-20}	13.396	mthl8	0.10	2.07×10^{-20}
2	CG43149	0.26	1.84×10^{-10}	12.576	CG43149	0.26	1.75×10^{-10}
3	CG9287	0.02	2.25×10^{-10}	9.695	CG9287	0.02	2.18×10^{-10}
4	ppk27	0.01	4.09×10^{-7}	8.581	CG43291	0.01	2.28×10^{-8}
5	w	0.03	1.94×10^{-9}	8.402	ppk27	0.01	4.08×10^{-7}
6	lncRNA:CR44377	0.01	1.01×10^{-6}	-7.998	w	0.03	1.94×10^{-9}
7	lncRNA:CR44285	0.06	1.56×10^{-3}	7.509	lncRNA:CR44377	0.01	1.04×10^{-6}
8	CG43919	0.02	1.10×10^{-5}	-7.109	lncRNA:CR44285	0.06	1.56×10^{-3}
9	5-HT2A	0.12	7.96×10^{-87}	-6.925	lncRNA:CR46123	0.02	6.36×10^{-5}
10	CG32437	0.05	2.55×10^{-10}	-6.607	CG43919	0.02	1.15×10^{-5}

3.2.3.5 Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

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

Table 44. Top Ten Highest Expressed Genes with Significant (padj) Difference between group-housed wildtypes and Fru mutants

rank	name	multi			rando		
		expression	adjusted p	log2 FoldChange	name	expression	adjusted p
1	CG9449	6.51	1.40×10^{-3}	0.282	CG9449	6.51	1.75×10^{-3}
2	Cyp6a20	5.86	5.53×10^{-24}	-1.277	Cyp6a20	5.86	2.98×10^{-23}
3	CG5973	4.91	3.78×10^{-3}	0.545	CG5973	4.91	4.28×10^{-3}
4	CG10357	3.77	7.93×10^{-6}	0.883	CG10357	3.77	8.22×10^{-6}
5	CG10553	3.58	1.58×10^{-15}	1.531	CG10553	3.58	1.86×10^{-15}
6	CG32641	3.43	1.70×10^{-21}	2.215	CG33177	3.40	2.06×10^{-4}
7	CG33177	3.40	1.56×10^{-4}	-0.430	Cyp313a1	3.01	3.72×10^{-6}
8	Cyp313a1	3.01	3.18×10^{-6}	1.153	CG33970	2.78	5.99×10^{-5}
9	CG33970	2.78	4.86×10^{-5}	-0.604	CG9541	2.18	3.17×10^{-5}
10	CG32640	2.71	6.26×10^{-28}	1.945	GstE1	2.07	6.75×10^{-3}

3.3 Comparing Expression Changes from Housing with Expression Changes from Genotype (Single-Factor Comparison Summary)

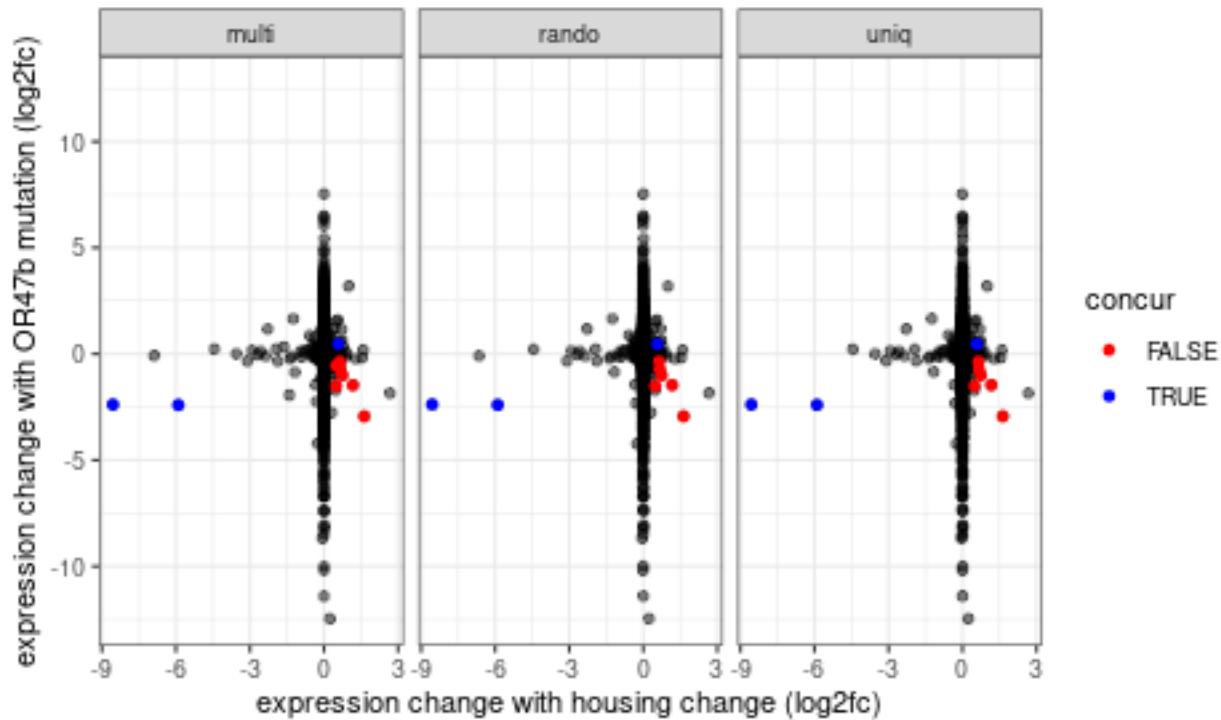
We want to see if the difference in life history creates similar changes in expression as various mutations. To do this, the differential expression data from DESeq2 are joined across pairs of contrasts. For example, the statistics from the wt-group vs wt-isolation contrast are joined by gene with the 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 enriched or both depleted) or not (ie, one enriched and the other depleted).

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 contrast vs the housing contrast (wt group & wt isolated). The upper right quadrant contains genes which are enriched in both cases; the lower left contains genes which are depleted in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

Figure 38. Scatterplot of Expression Changes in OR47b mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## png  
## 2
```

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

Table 45. 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	7	6	6

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

Table 46. Top Ten Most Significant Genes in difference expression between housing and OR47b c

rank	name	multi				name	mean expression	m
		mean expression	mean readusted p	housing l2fc	mutation l2fc			
1	TotA	0.14	7.01×10^{-5}	-5.899	-2.409	TotA	0.14	
2	CG13659	0.23	1.67×10^{-4}	0.579	0.465	TotC	0.05	
3	TotC	0.05	2.11×10^{-4}	-8.554	-2.389	CG13659	0.23	

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

rank	name	multi			rando			
		mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean re
1	TotC	-5.472	0.05	2.11×10^{-4}	TotC	-5.469	0.05	2.
2	TotA	-4.154	0.14	7.01×10^{-5}	TotA	-4.151	0.14	8.
3	CG13659	0.522	0.23	1.67×10^{-4}	CG13659	0.521	0.23	2.

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

rank	name	multi				name	mean expression	mea
		mean expression	mean readusted p	housing l2fc	OR47b l2fc			
1	Obp84a	0.71	1.68×10^{-25}	0.486	-1.515	Obp84a	0.71	
2	amd	1.21	1.95×10^{-25}	1.171	-1.468	amd	1.21	
3	CG10050	0.32	1.60×10^{-17}	0.743	-1.006	CG10050	0.32	
4	CG11852	0.11	1.25×10^{-11}	1.634	-2.938	CG11852	0.11	
5	SP1029	0.07	4.63×10^{-4}	0.663	-0.729	SP1029	0.07	
6	magu	0.25	1.05×10^{-3}	0.647	-0.359	magu	0.25	
7	Cyp9h1	0.18	3.89×10^{-3}	0.491	-0.548	NA	NA	

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 49. Top Ten Most Serious Significant Differences between difference expression between housing and OR47b contrants

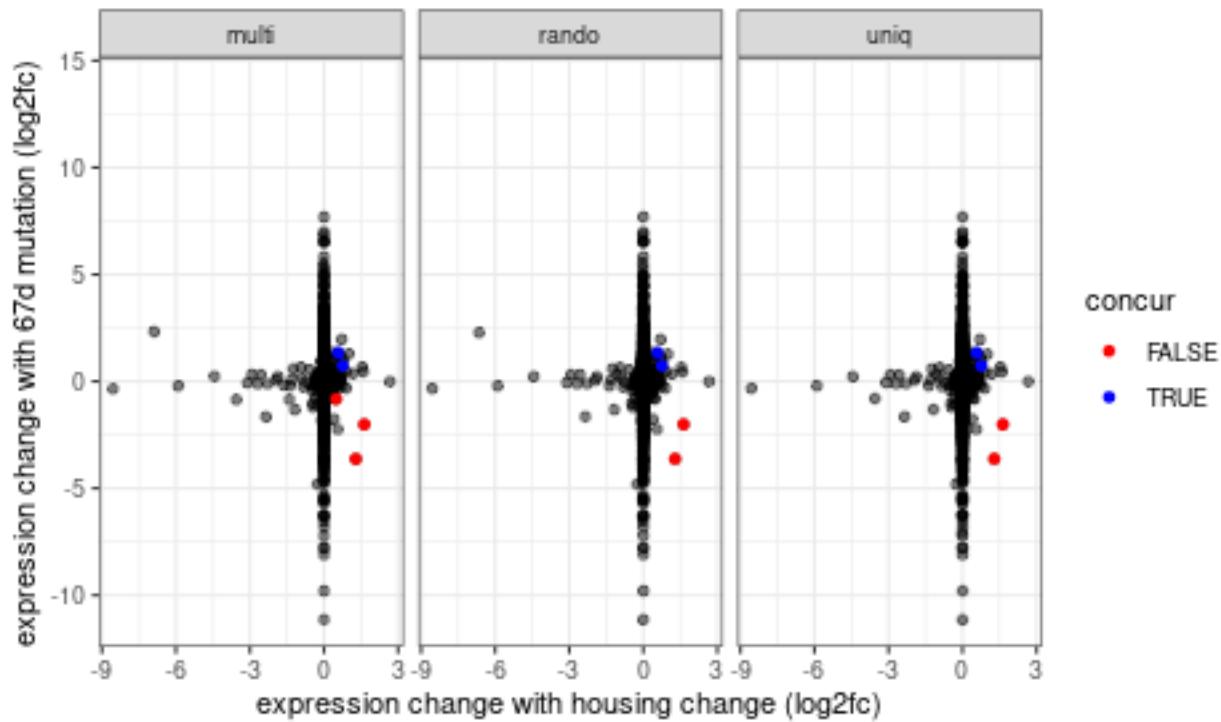
rank	multi				rando			
	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression	
1	CG11852	4.571	0.11	1.25×10^{-11}	CG11852	4.568	0.11	
2	amd	2.638	1.21	1.95×10^{-25}	amd	2.635	1.21	
3	Obp84a	2.001	0.71	1.68×10^{-25}	Obp84a	1.998	0.71	
4	CG10050	1.749	0.32	1.60×10^{-17}	CG10050	1.746	0.32	
5	SP1029	1.392	0.07	4.63×10^{-4}	SP1029	1.386	0.07	
6	Cyp9h1	1.039	0.18	3.89×10^{-3}	magu	1.002	0.25	
7	magu	1.007	0.25	1.05×10^{-3}	NA	NA	NA	

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast_and_47bContrast.multi.tsv*, *results/tables/supp/housingContrast_and_47bContrast.rando.tsv*, *results/tables/supp/housingContrast_and_47bContrast.una.tsv*.

3.3.2 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 enriched in both cases; the lower left contains genes which are depleted in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

Figure 39. Scatterplot of Expression Changes in 67d mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## png
## 2
```

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

Table 50. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs 67d

	multi	rando	uniq
Agree	2	2	2
Disagree	3	2	2

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

Table 51. Top Ten Most Significant Genes of Agr
in difference expression between housing and 67d contrants

rank	multi					rand		
	name	mean expression	mean readjusted p	housing l2fc	67d l2fc	name	mean expression	mean re
1	CG13659	0.28	2.76×10^{-7}	0.579	1.338	CG13659	0.28	4
2	CG31272	0.12	4.07×10^{-4}	0.760	0.735	CG31272	0.12	5

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

rank	multi				rando			
	name	mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean readusted p
1	CG13659	0.959	0.28	2.76×10^{-7}	CG13659	0.958	0.28	4.1
2	CG31272	0.748	0.12	4.07×10^{-4}	CG31272	0.746	0.12	5.1

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

Table 53. Top Ten Most Significant Genes of Disagreement in difference expression between housing and OR47b contrants

rank	multi				rando			
	name	mean expression	mean readusted p	housing l2fc	67d l2fc	name	mean expression	mean readusted p
1	MtnB	0.55	7.33×10^{-23}	1.289	-3.628	MtnB	0.55	2.5
2	CG11852	0.11	2.14×10^{-6}	1.634	-2.017	CG11852	0.11	3.1
3	Cyp9h1	0.18	5.31×10^{-3}	0.491	-0.823	NA	NA	NA

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 54. Top Ten Most Serious Significant Differences between housing and 67d contrants

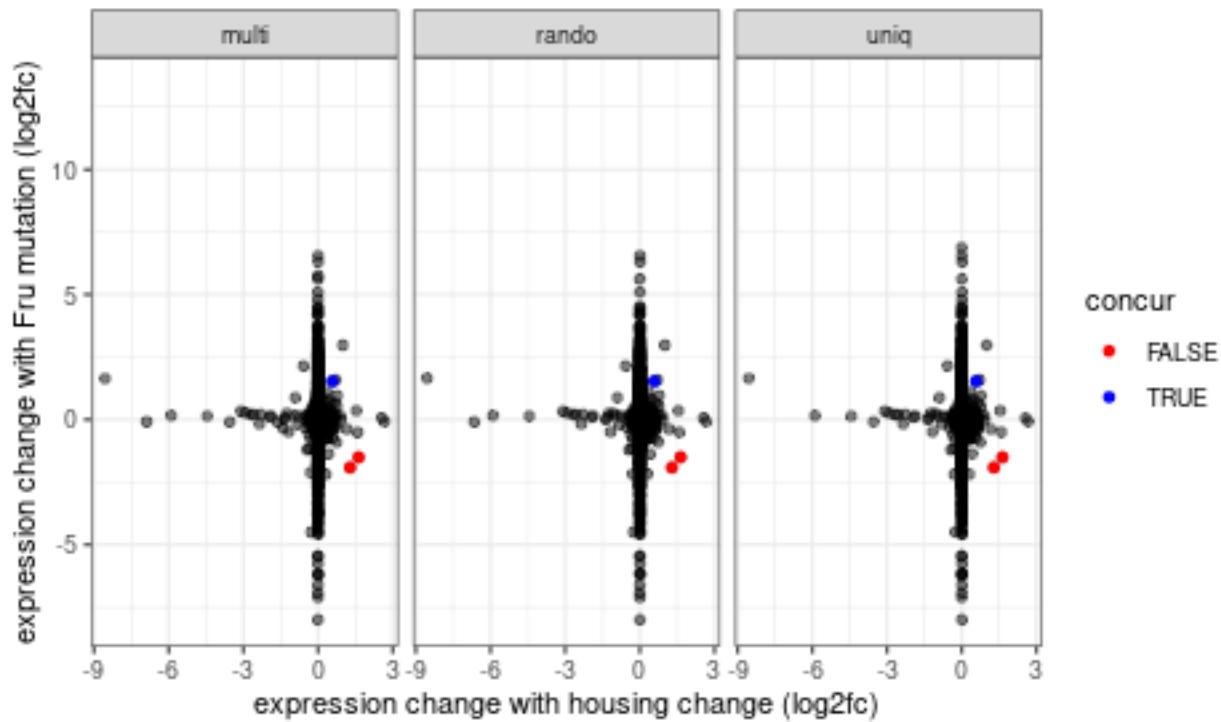
rank	multi				rando			
	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression	mean readusted p
1	MtnB	4.916	0.55	7.33×10^{-23}	MtnB	4.914	0.55	
2	CG11852	3.651	0.11	2.14×10^{-6}	CG11852	3.646	0.11	
3	Cyp9h1	1.314	0.18	5.31×10^{-3}	NA	NA	NA	

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.unadjusted.tsv*.

3.3.3 Housing & Fru

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

Figure 40. Scatterplot of Expression Changes in Fru mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## png
## 2
```

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

Table 55. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs Fru

	multi	rando	uniq
Agree	1	1	1
Disagree	2	2	2

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

Table 56. Top Ten Most Significant Genes of Agree in difference expression between housing and Fru contrants

rank	name	multi				rand			
		mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re	
1	CG13659	0.28	1.71×10^{-10}	0.579	1.532	CG13659	0.28	2.6	

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 57. Top Ten Largest Magnitude Changes In Significant Genes in difference expression between housing and Fru contrats

multi					rando				
rank	name	mean l2fc	mean expression	mean readusted p	name	mean l2fc	mean expression	mean re	
1	CG13659	1.056	0.28	1.71×10^{-10}	CG13659	1.055	0.28	2.6	

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

Table 58. Top Ten Most Significant Genes of Disagreement in difference expression between housing and Fru contrats

multi						rando			
rank	name	mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re	
1	MtnB	0.56	9.47×10^{-12}	1.289	-1.918	MtnB	0.56	1.8	
2	CG11852	0.11	2.68×10^{-5}	1.634	-1.508	CG11852	0.11	3	

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 59. Top Ten Most Serious Differences between the housing contrast and the Fru contrast

multi					rando			
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression	
1	MtnB	3.207	0.56	9.47×10^{-12}	MtnB	3.204	0.56	
2	CG11852	3.142	0.11	2.68×10^{-5}	CG11852	3.139	0.11	

Notably, many genes which have significant, high-ranking similarities in both the housing contrast and the Fru contrast ... are points of significant, high-ranking differences between the housing contrast and the 47b or 67d contrasts. In particular:

DIP-alpha
CG13659
Cpr49Ae
Dscam4
CG31288
Pop2
CG7272

As well, many genes which have significant, high-ranking difference in both the housing contrast and the Fru contrast ... are points of significant, high-ranking similarity between the housing contrast and the 47b or 67d contrasts. In particular:

MtnB

CG10050
 CG14400
 CG5895
 CG11852
 Spn47C

Full data are in the tables folder:

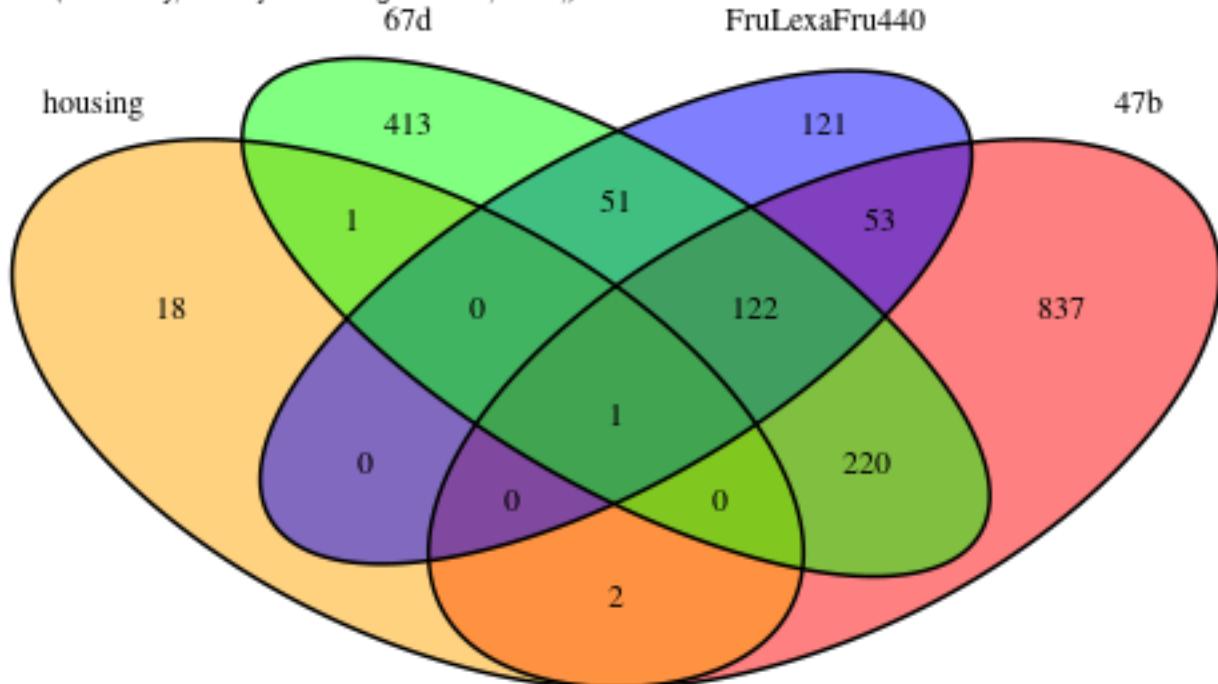
results/tables/supp/housingContrast_andFruContrast.multi.tsv *results/tables/supp/housingContrast_andFruContrast.r*
results/tables/supp/housingContrast_andFruContrast.uniq.tsv

3.3.4 Overview

How to perform multiple comparisons adjustment on a all-contrasts venn/upset plot????

Let's bonferroni-correct for n=4 comparisons, one for each single-factor model.

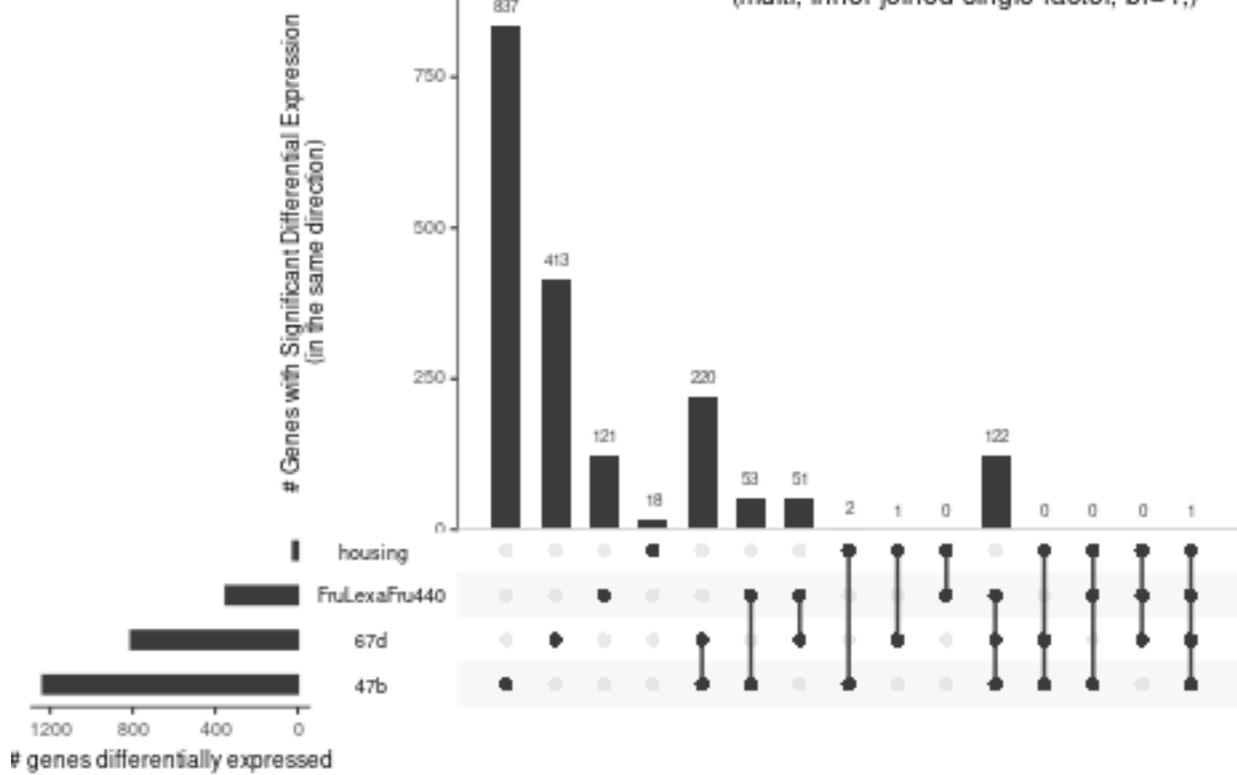
Figure 41 . Venn Diagram: # genes with shared significant change, by experimental contrast intersection (multi-only, inner-joined single-factor, bf=4.)



```
## null device
##           1

## null device
##           1
```

Figure 42 . UpSet plot: # genes with shared significant change, by experimental contrast intersection
 (multi, inner-joined single-factor, bf=4,)



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

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

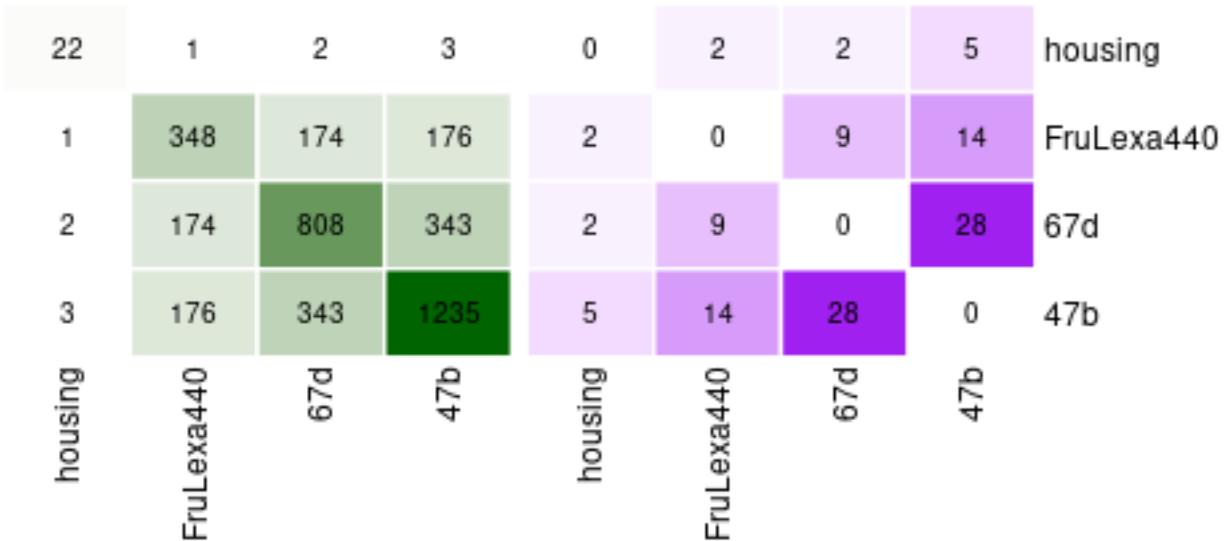
The results using full join and no correction are qualitatively similar (indeed the more sets being intersected the closer an inner join will approximate a fulljoin)

Figure 43 Heatmap of Pairwise Comparisons between Contrasts:

significant genes with the same (left)
or different (right) directions of change
(single factor models; multi only; bf=4)

a. Number Agreeing

b. Number Disagreeing



```
## png
## 2
```

3.4 Simultaneously Modeling Housing & Genotype.

gives us eye-to-eye results for all treatments

These data are in the file “results/tables/supp/hausWtVsMut.allAligners.DESeq2MpBC.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

```

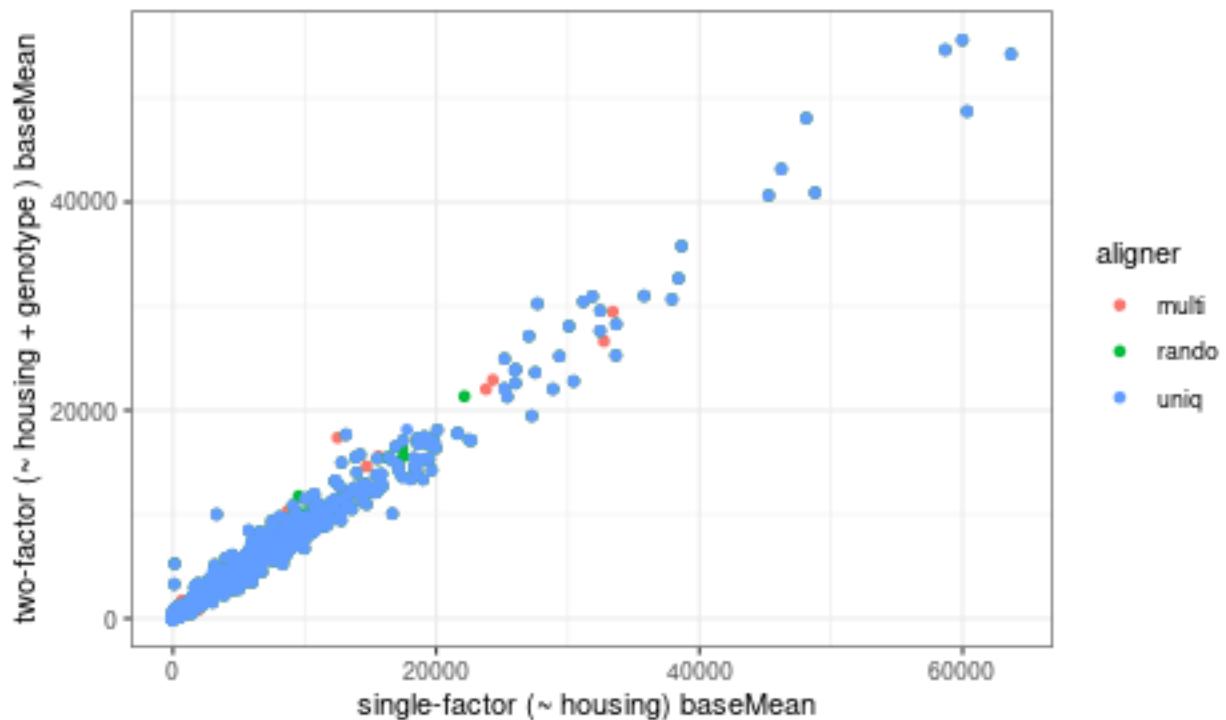
-> do % of genes available for analysis

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



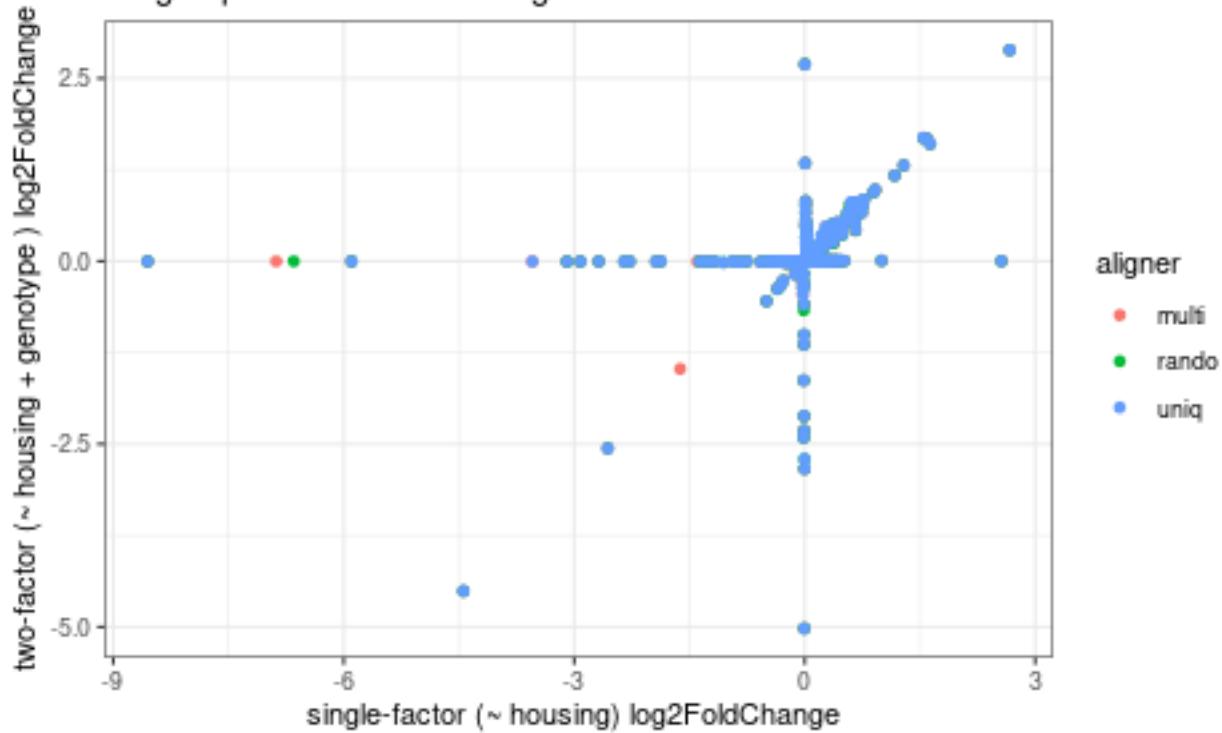
```

## png
## 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.

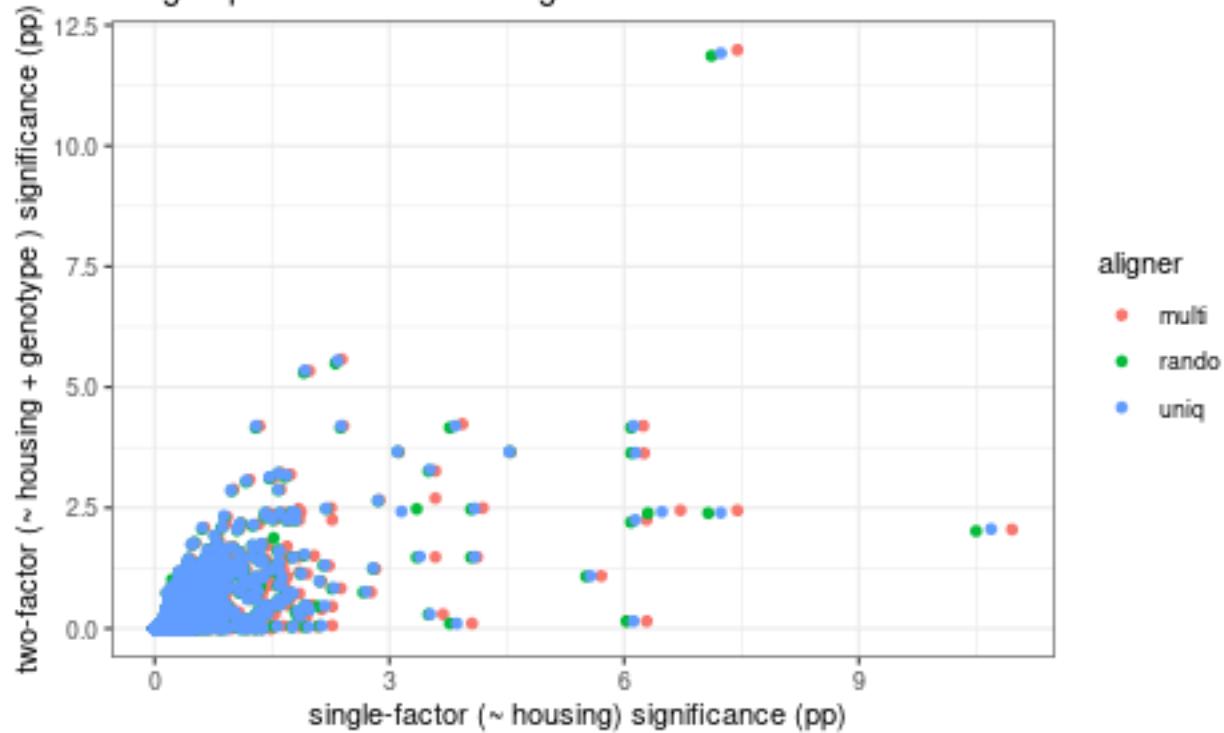
Figure 45. Scatterplot of per-gene expression difference
in full- vs. single-factor models
for group vs isolated housing treatment



```
## png
## 2
```

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

Figure 46. Scatterplot of per-gene DE significance
in full- vs. single-factor models
for group vs isolated housing treatment

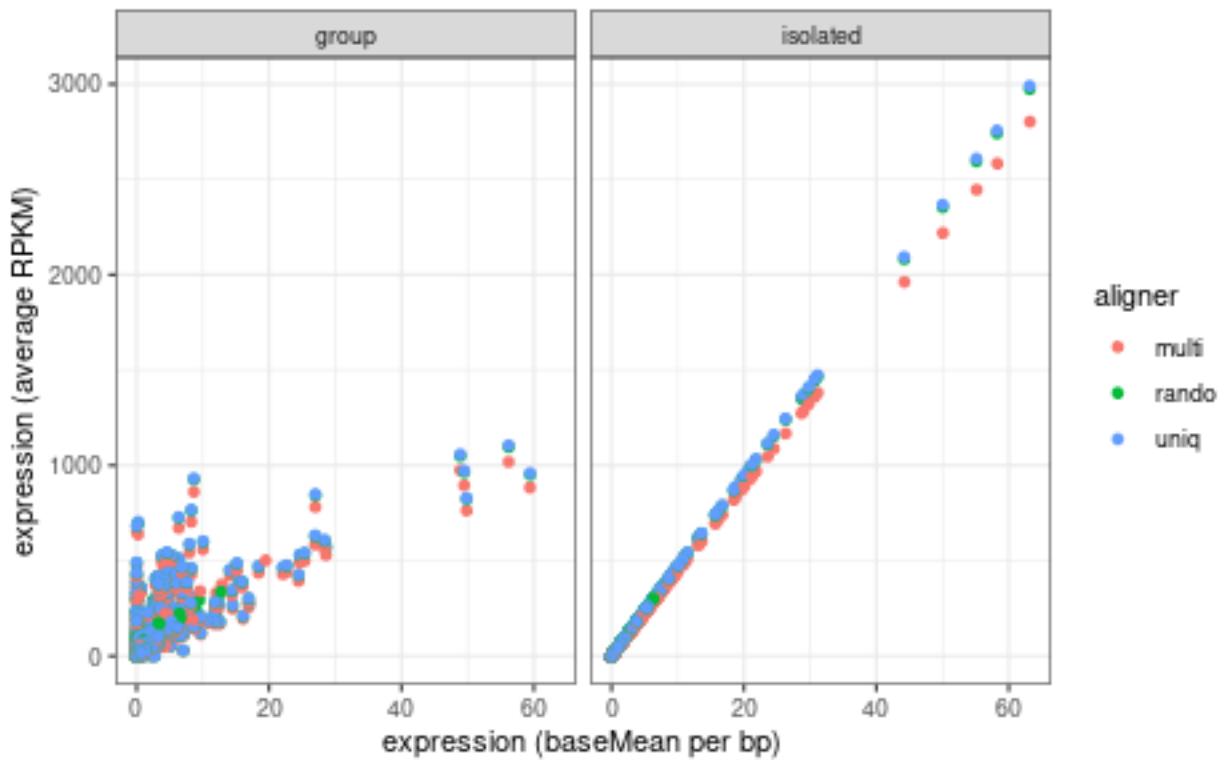


```
## png
## 2
```

3.4.2 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 47. Comparison of Expressionas Inferred by Direct Read Counting vs. DESeq2 Normalization

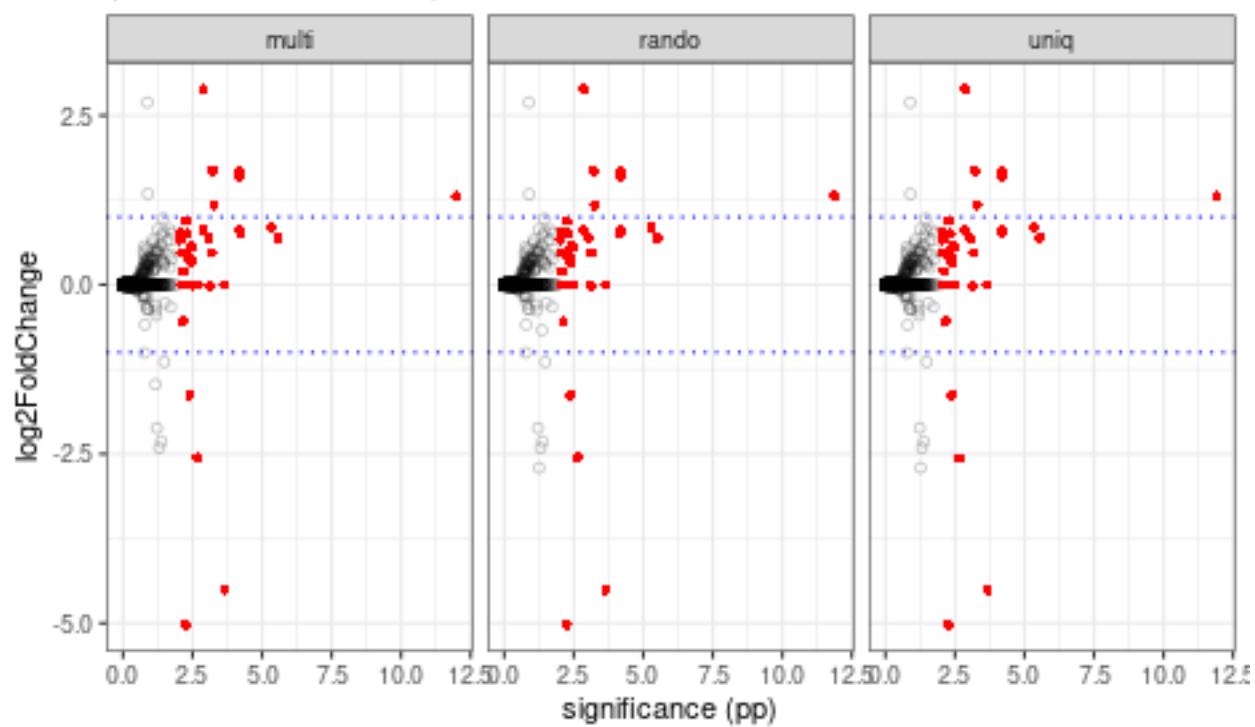


```
## png
## 2
```

3.4.3 Housing Contrast (Simultaneous model)

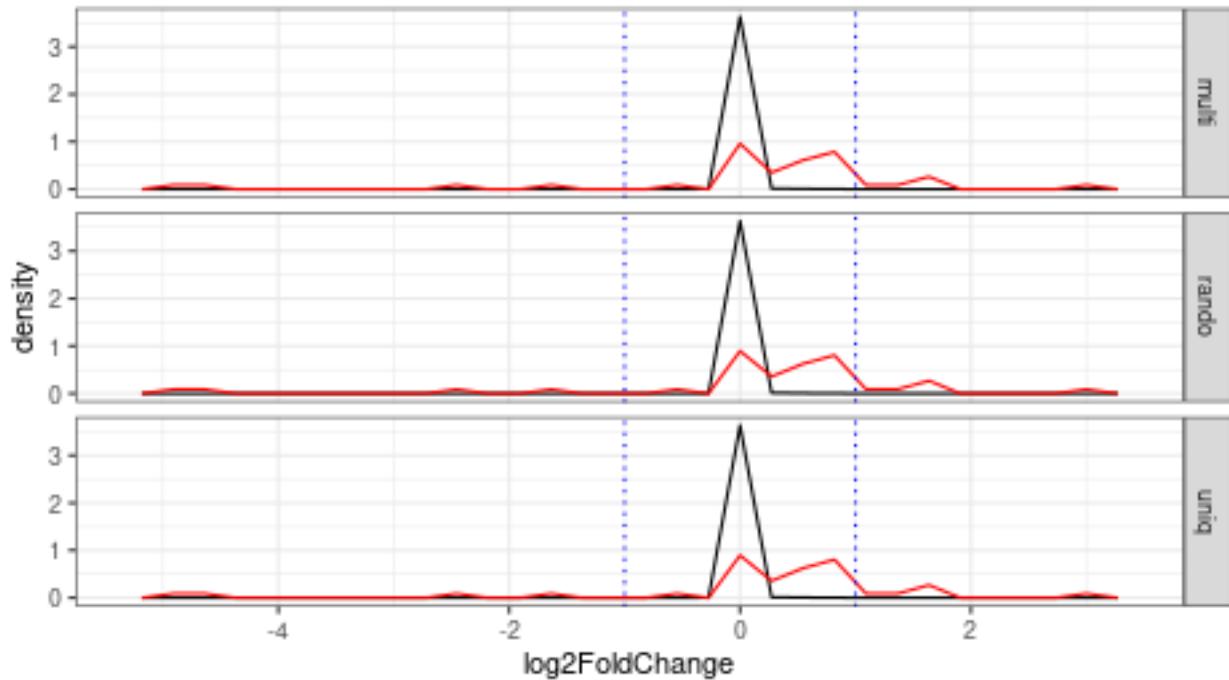
Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of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

Figure 48. Volcano Plot: Fold Change vs. Significance
(between isolated and group-housed wildtypes)
(Simultaneous Model)



```
## png
## 2
```

Figure 49. histogram of fold change
 with significant($p_{adj} < 0.01$) changes highlighted in red
 (between isolated and group-housed wildtypes)
 (Simultaneous Model)

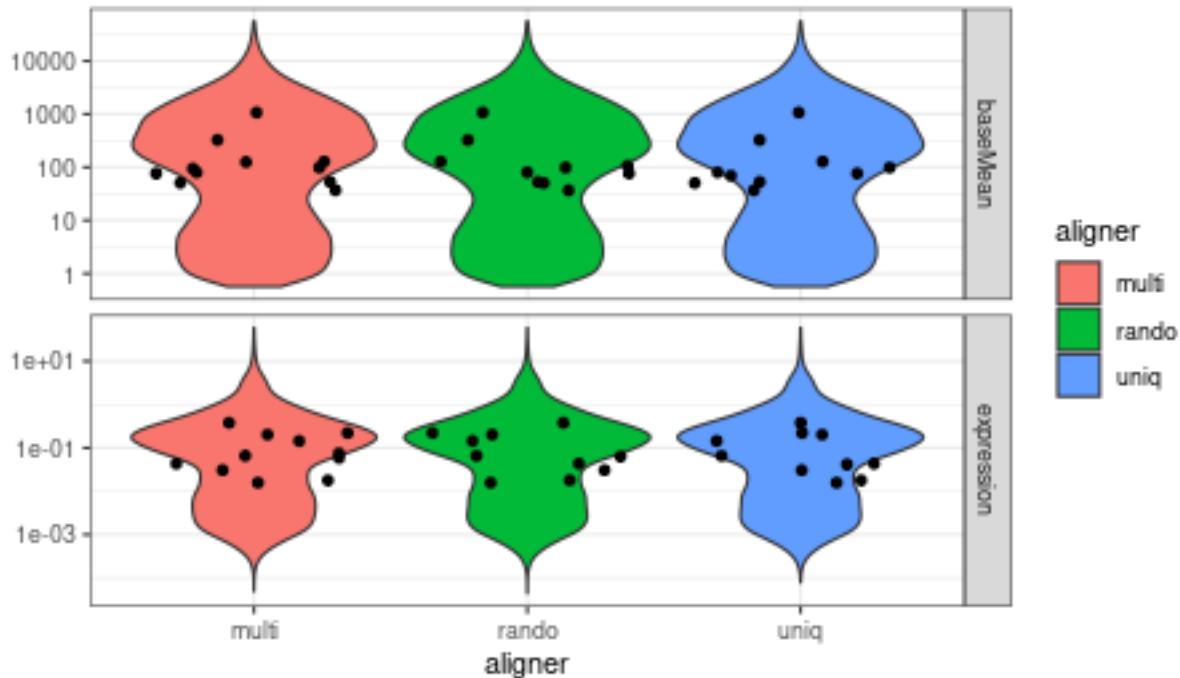


```
## png
## 2
```

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

They do not appear to be unusual in terms of read count or in terms of expression. Here background distributions are shown as violin plots, with anomalous points overplotted:

**Figure 50. High Significance, Low Effect Size Genes Do Not Have Unusual Read Counts or Expression Levels
(isolated and group-housed contrast)
(Simultaneous Model)**



```
## png
## 2
```

The gene content is not obviously skewed (eg, no tRNA genes, no rRNA genes, not overwhelmed with sketchy CGs....)

Table 60. Genes with Low Effect Size and High Significance Are Mainstream isolated and group-housed contrast; Simultaneous Model)

gene	where anomalous
Amy-d	multi
Amy-p	multi,rando,uniq
CG11400	multi,rando,uniq
CG12239	multi,rando,uniq
CG2736	multi,rando,uniq
CG4716	multi,rando,uniq
Mf	multi,rando,uniq
Mlp60A	multi,rando,uniq
PPO2	multi,rando,uniq
TotA	multi,rando,uniq
TotC	multi,rando,uniq

Here's what Mike Love has to say: (email, 13 July 2020):

So for one thing, the shrinkage tends to be more conservative than the p-value w/o shrinkage. If you do svalue=TRUE you will get s-values that correspond to this conservativeness that you see on the y-axis

The other thing is that, you probably would also lose these genes if you used lfcThreshold=x, for some x that's higher than 0.

We talk about this in the DESeq2 paper, that rejection of LFC=0 doesn't necessarily mean that that fold changes are practically meaningful, just that we have evidence that they are not equal to 0. Typically with more samples we can reject nulls when LFC is quite close to 0...

Here's Love, Huber, and Anders (2014):

Most approaches to testing for differential expression, including the default approach of DESeq2, test against the null hypothesis of zero LFC. However, if any biological processes are genuinely affected by the difference in experimental treatment, this null hypothesis implies that the gene under consideration is perfectly decoupled from these processes. Due to the high interconnectedness of cells' regulatory networks, this hypothesis is, in fact, implausible, and arguably wrong for many if not most genes. Consequently, with sufficient sample size, even genes with a very small but non-zero LFC will eventually be detected as differentially expressed. A change should therefore be of sufficient magnitude to be considered biologically significant. For small-scale experiments, statistical significance is often a much stricter requirement than biological significance, thereby relieving the researcher from the need to decide on a threshold for biological significance.

For well-powered experiments, however, a statistical test against the conventional null hypothesis of zero LFC may report genes with statistically significant changes that are so weak in effect strength that they could be considered irrelevant or distracting.

Of the 11332 genes with significance scores available, 42 have an adjusted p < 0.01 (0.3706318 %)

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 10 such genes, mostly shared across alignment strategy:

Table 61. Genes with Large (2< fold change), Significant (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
CG15822	yes	yes	yes
CG11852	yes	yes	yes
Jhe	yes	yes	yes
CG7470	yes	yes	yes
CG31324	yes	yes	yes
amd	yes	yes	yes
hgo	yes	yes	yes

3.4.3.1 Top Tens

Genes with top 10 most significant changes

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

Table 62. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Express between isolated and group-housed wildtypes; simultaneous model

multi					rando			
rank	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	MtnB	0.39	1.03×10^{-12}	1.309	MtnB	0.39	1.37×10^{-12}	1.30
2	magu	0.21	2.64×10^{-6}	0.691	magu	0.21	3.19×10^{-6}	0.68
3	CG31288	1.36	4.61×10^{-6}	0.843	CG31288	1.36	5.03×10^{-6}	0.84
4	CG31272	0.11	5.80×10^{-5}	0.769	CG31272	0.11	6.89×10^{-5}	0.76
5	CG11852	0.07	6.35×10^{-5}	1.604	CG11852	0.07	6.89×10^{-5}	1.60
6	CG10512	0.22	6.35×10^{-5}	0.799	CG10512	0.22	6.89×10^{-5}	0.79
7	CG15822	0.01	6.35×10^{-5}	1.670	CG15822	0.01	6.89×10^{-5}	1.67
8	TotC	0.14	2.22×10^{-4}	-0.002	TotC	0.14	2.19×10^{-4}	-0.00
9	Prat2	0.02	2.22×10^{-4}	-4.510	Prat2	0.02	2.19×10^{-4}	-4.51
10	TotA	0.22	2.34×10^{-4}	-0.002	TotA	0.22	2.30×10^{-4}	-0.00

Top 10 genes with biggest (significant) effect sizes

Table 63. Top Ten Largest Magnitude Fold Changes which were Signifi between isolated and group-housed wildtypes; simultaneous model

multi					rando			
rank	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	CG10799	0.04	5.59×10^{-3}	-5.021	CG10799	0.04	5.82×10^{-3}	-5.02
2	Prat2	0.02	2.22×10^{-4}	-4.510	Prat2	0.02	2.19×10^{-4}	-4.51
3	Jhe	0.26	1.30×10^{-3}	2.887	Jhe	0.26	1.36×10^{-3}	2.88
4	CG7470	0.03	2.17×10^{-3}	-2.559	CG7470	0.03	2.27×10^{-3}	-2.55
5	CG31324	0.08	5.97×10^{-4}	1.689	CG31324	0.08	6.16×10^{-4}	1.68
6	CG15822	0.01	6.35×10^{-5}	1.670	CG15822	0.01	6.89×10^{-5}	1.67
7	hgo	0.04	3.96×10^{-3}	-1.630	hgo	0.04	4.08×10^{-3}	-1.63
8	CG11852	0.07	6.35×10^{-5}	1.604	CG11852	0.07	6.89×10^{-5}	1.60
9	MtnB	0.39	1.03×10^{-12}	1.309	MtnB	0.39	1.37×10^{-12}	1.30
10	amd	1.01	5.43×10^{-4}	1.172	amd	1.01	5.50×10^{-4}	1.17

Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

Table 64. Top Ten Highest Expressed Genes with Significant ($\text{padj} <$ Difference between isolated and group-housed wildtypes; simultaneous model

multi					rando			
rank	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange

1	Fer2LCH	4.18	6.85×10^{-3}	0.200	Fer2LCH	4.18	8.40×10^{-3}	0.198
2	CG14687	2.72	3.56×10^{-3}	0.337	CG14687	2.72	4.08×10^{-3}	0.334
3	Or92a	1.57	4.57×10^{-3}	0.407	Or92a	1.57	4.96×10^{-3}	0.405
4	CG31288	1.36	4.61×10^{-6}	0.843	CG31288	1.36	5.03×10^{-6}	0.842
5	amd	1.01	5.43×10^{-4}	1.172	amd	1.01	5.50×10^{-4}	1.171
6	Nep4	0.77	4.87×10^{-3}	0.755	Nep4	0.77	4.96×10^{-3}	0.759
7	CG33056	0.75	5.49×10^{-3}	0.946	CG33056	0.75	5.60×10^{-3}	0.944
8	CG42806	0.64	1.30×10^{-3}	0.801	CG42806	0.64	1.41×10^{-3}	0.799
9	Obp84a	0.62	5.59×10^{-3}	0.451	Obp84a	0.62	6.23×10^{-3}	0.449
10	CG16743	0.58	3.32×10^{-3}	0.569	CG16743	0.58	3.81×10^{-3}	0.567

3.4.3.2 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,

```
## png
## 2
```

Table 65. Enriched GO Terms among Significantly Differentially Expressed Genes simultaneous housing contrast; multi only

GO Term	Description	p-value		
		Fisher	K-S	ontology
GO:0034644	cellular response to UV	5.60×10^{-4}	4.36×10^{-3}	BP
GO:0006584	catecholamine metabolic process	1.29×10^{-3}	2.12×10^{-3}	BP
GO:0009712	NA	1.29×10^{-3}	2.12×10^{-3}	BP
GO:0034605	cellular response to heat	5.99×10^{-3}	2.90×10^{-3}	BP
GO:0071482	cellular response to light stimulus	7.91×10^{-3}	2.30×10^{-4}	BP
GO:0005576	extracellular region	6.40×10^{-5}	2.19×10^{-3}	CC
GO:0030017	sarcomere	9.90×10^{-3}	9.10×10^{-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.

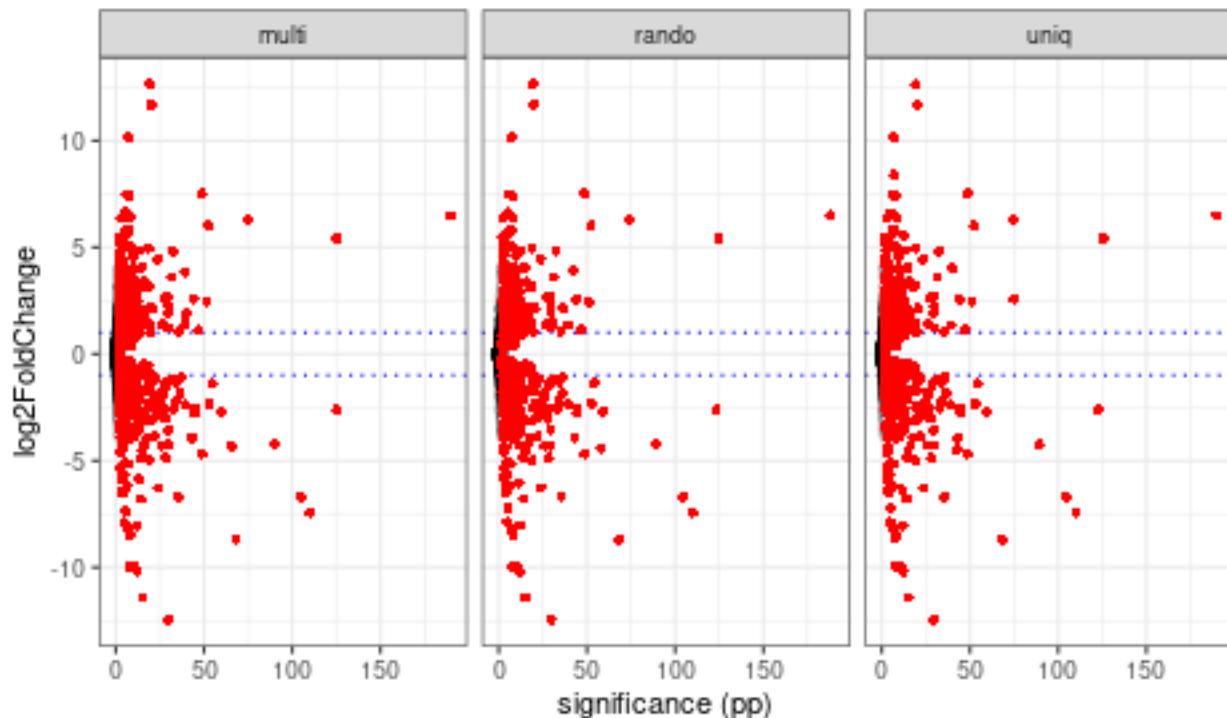
Minor differences w/aligner; see tables/supp/

3.4.4 47b vs wt (Simultaneous model)

Of the 12584 genes with significance scores available, 925 have an adjusted p < 0.01 (7.3506039 %)

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 depleted relative to the group-housed condition; positive fold changes are enriched.

**Figure 52. Volcano Plot: Fold Change vs. Significance
(between group-housed 47b mutants and wildtypes)
(simultaneous model)**



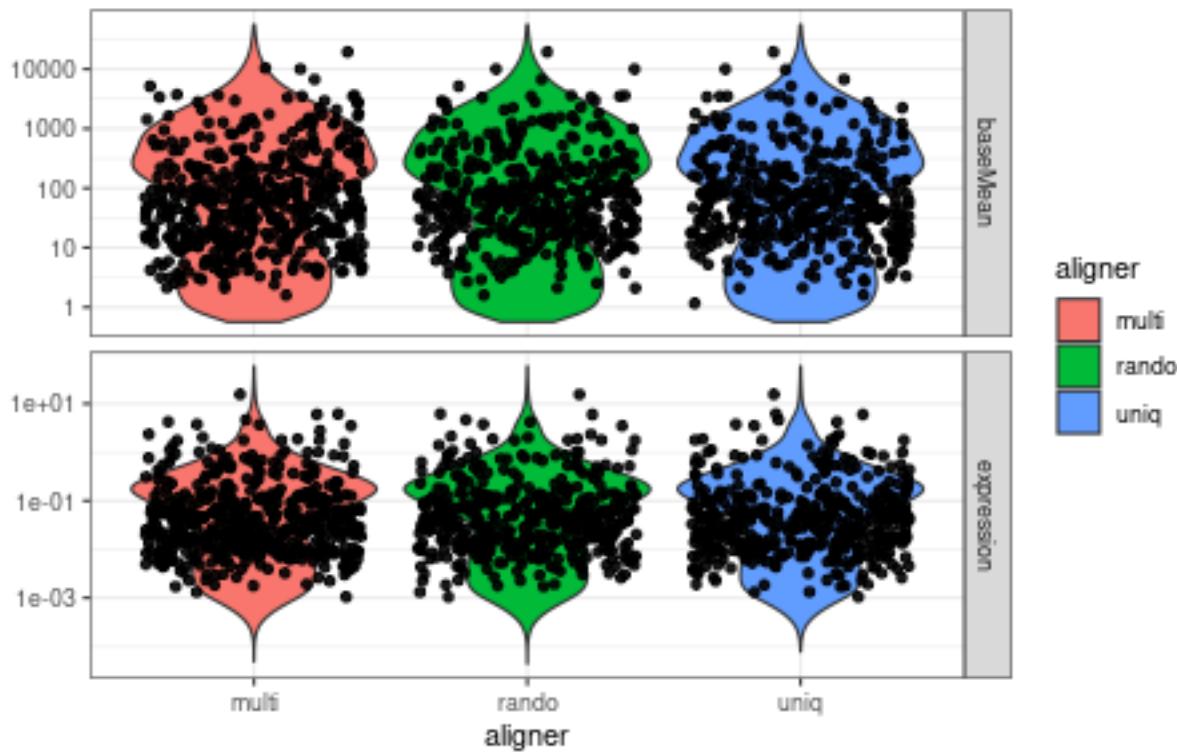
```
## png
## 2
```

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/tbl66hausWtVsMut_genotype47b_chonky.html](#)

Are “chonky” genes prone to unusually low expression?

Figure 53. 'Chonky' Gene Expression Changes Are Not Prone to Low-E₁
47b1 simultaneous contrast, adjusted p < 0.01, abs(l2fc) > 1



```
## png
## 2
```

3.4.4.1 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 66. Top Ten Most Significantly (padj<0.01) Differentially Expressed Genes between group-housed 47b mutants and wildtypes (simultaneous model)

rank	multi					rando				
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange		
1	CG6912	0.36	2.49×10^{-190}	6.486	CG6912	0.36	1.14×10^{-188}			
2	CG7900	1.76	4.95×10^{-126}	5.416	CG7900	1.76	1.61×10^{-125}			
3	Drip	1.40	4.95×10^{-126}	-2.614	Drip	1.40	5.73×10^{-124}			
4	Or47b	0.92	9.34×10^{-111}	-7.427	Or47b	0.92	1.36×10^{-110}			
5	5-HT2A	0.09	1.24×10^{-105}	-6.692	5-HT2A	0.09	4.77×10^{-105}			
6	DIP-alpha	0.06	8.56×10^{-91}	-4.228	DIP-alpha	0.06	6.88×10^{-90}			
7	CG8665	0.07	2.25×10^{-75}	6.293	CG8665	0.07	1.02×10^{-74}			
8	Cyp6a17	0.87	5.32×10^{-69}	-8.693	Cyp6a17	0.87	6.76×10^{-69}			
9	Cyp12d1-p	0.05	2.80×10^{-66}	-4.315	Or85b	0.44	5.61×10^{-60}			
10	Or85b	0.44	9.81×10^{-61}	-2.711	Cyp12d1-p	0.05	9.70×10^{-59}			

Top 10 genes with biggest (significant) effect sizes

Table 67. Top Ten Largest Magnitude Fold Changes which were observed between group-housed 47b mutants and wildtypes (simultaneous model)

rank	multi				rando			
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	
1	mthl8	0.06	6.12×10^{-20}	12.640	mthl8	0.06	6.33×10^{-20}	
2	CG40486	3.75	2.08×10^{-30}	-12.445	CG40486	3.74	2.23×10^{-30}	
3	w	0.53	9.84×10^{-21}	11.680	w	0.53	9.88×10^{-21}	
4	CG30428	0.16	6.57×10^{-16}	-11.414	CG30428	0.16	6.83×10^{-16}	
5	ppk19	0.04	8.34×10^{-13}	-10.194	ppk19	0.04	9.02×10^{-13}	
6	CG43149	0.16	8.78×10^{-8}	10.149	CG43149	0.16	8.82×10^{-8}	
7	lncRNA:CR45502	0.14	1.17×10^{-11}	-9.998	lncRNA:CR45502	0.14	1.20×10^{-11}	
8	CheA7a	0.07	1.30×10^{-8}	-9.965	CheA7a	0.07	1.32×10^{-8}	
9	Cyp6a17	0.87	5.32×10^{-69}	-8.693	Cyp6a17	0.87	6.76×10^{-69}	
10	asRNA:CR44030	0.05	4.95×10^{-9}	-8.509	asRNA:CR44030	0.05	5.06×10^{-9}	

Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

Table 68. Top Ten Highest Expressed Genes with Significant (parametric) Difference
group-housed 47b mutants and wildtypes (simultaneous model)

rank	multi				rando			
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	Drsl5	16.49	1.10×10^{-4}	-0.892	Drsl5	16.50	1.21×10^{-4}	
2	lncRNA: noe	16.02	1.72×10^{-5}	-0.857	lncRNA: noe	16.02	1.98×10^{-5}	
3	to	15.53	6.67×10^{-26}	-2.197	to	15.53	8.78×10^{-26}	
4	CG11550	10.81	1.46×10^{-3}	0.582	CG11550	10.82	1.51×10^{-3}	
5	Obp59a	8.89	3.89×10^{-7}	0.552	Obp59a	8.90	3.62×10^{-7}	
6	CG43093	6.14	2.97×10^{-9}	-0.852	CG43093	6.14	3.50×10^{-9}	
7	CG30197	6.10	5.39×10^{-8}	1.084	CG30197	6.11	5.50×10^{-8}	
8	Cyp6a2	5.98	4.30×10^{-13}	2.487	Cyp6a2	5.98	4.48×10^{-13}	
9	Cyp6a20	5.59	1.65×10^{-3}	-0.360	Cyp6a20	5.59	2.01×10^{-3}	
10	Idgf4	5.24	1.06×10^{-3}	0.677	Idgf4	5.24	1.10×10^{-3}	

3.4.4.2 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

```
## png
## 2
```

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

Table 70. 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	9.60×10^{-18}	3.90×10^{-8}
GO:0046983	protein dimerization activity	1.00×10^{-10}	1.20×10^{-4}
GO:0031492	nucleosomal DNA binding	2.90×10^{-9}	1.50×10^{-6}
GO:0004888	transmembrane signaling receptor activity	3.70×10^{-9}	2.40×10^{-6}
GO:0038023	signaling receptor activity	2.10×10^{-8}	1.40×10^{-7}
GO:0060089	NA	2.10×10^{-8}	1.40×10^{-7}
GO:0031491	nucleosome binding	4.50×10^{-8}	5.90×10^{-5}
GO:0004871	NA	1.10×10^{-7}	1.30×10^{-6}
GO:0005506	iron ion binding	1.20×10^{-7}	4.10×10^{-4}
GO:0005549	odorant binding	2.40×10^{-7}	1.30×10^{-7}
BP			
GO:0006334	nucleosome assembly	3.80×10^{-13}	9.20×10^{-7}
GO:0031497	chromatin assembly	4.50×10^{-11}	2.60×10^{-5}
GO:0007606	sensory perception of chemical stimulus	5.50×10^{-10}	5.20×10^{-12}
GO:0065004	protein-DNA complex assembly	2.90×10^{-9}	3.90×10^{-4}
GO:0034728	nucleosome organization	4.60×10^{-9}	4.50×10^{-4}
GO:0006333	chromatin assembly or disassembly	7.70×10^{-9}	6.90×10^{-4}
GO:0007600	NA	4.20×10^{-8}	1.30×10^{-9}
GO:0050907	detection of chemical stimulus involved in sensory perception	4.90×10^{-7}	1.00×10^{-5}
GO:0009593	detection of chemical stimulus	8.00×10^{-7}	7.20×10^{-6}
GO:0050906	detection of stimulus involved in sensory perception	1.20×10^{-6}	6.90×10^{-6}
CC			
GO:0000786	nucleosome	2.30×10^{-25}	9.70×10^{-14}
GO:0044815	NA	5.10×10^{-24}	1.10×10^{-12}
GO:0032993	protein-DNA complex	1.20×10^{-21}	6.90×10^{-11}
GO:0000788	nuclear nucleosome	1.80×10^{-21}	1.70×10^{-13}
GO:0016021	integral component of membrane	2.90×10^{-12}	4.00×10^{-9}
GO:0031224	intrinsic component of membrane	3.70×10^{-12}	2.60×10^{-9}
GO:0005576	extracellular region	4.00×10^{-12}	2.70×10^{-9}
GO:0044421	NA	4.40×10^{-9}	1.20×10^{-5}
GO:0044425	NA	3.40×10^{-8}	3.70×10^{-5}
GO:0005615	extracellular space	1.20×10^{-7}	2.20×10^{-4}

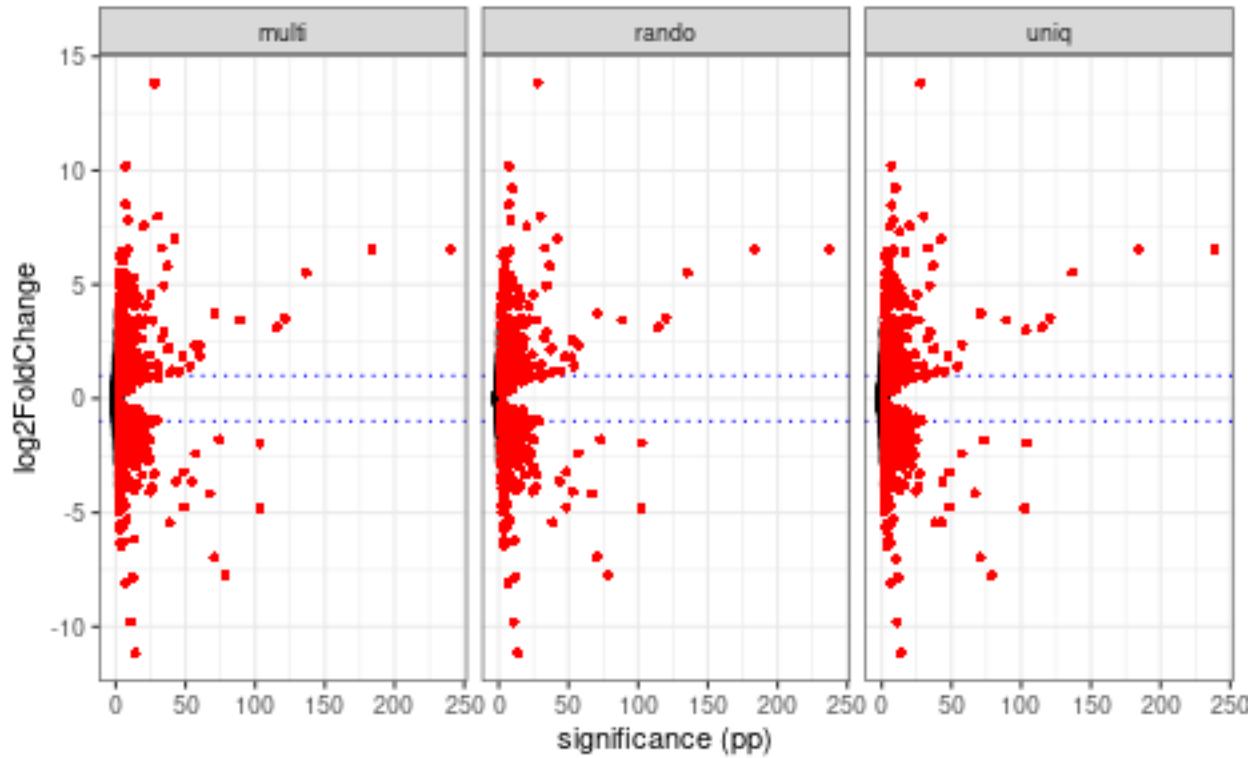
3.4.5 67d vs wt (Simultaneous model)

Of the 12103 genes with significance scores available, 1080 have an adjusted p < 0.01 (8.9234074 %)

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 \log_2 fold change of $+/-1$ (ie, a difference in expression of a factor of 2). Genes with negative \log_2 fold changes are depleted relative to the group-housed condition; positive fold changes are enriched

**Figure 55. Volcano Plot: Fold Change vs. Significance
(between group-housed 67d mutants and wildtypes, simultaneous model)**



```
## png
## 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 562 such genes, mostly shared across alignment strategy:

(Table available at *results/tables/tbl71_hausWtVsMut_genotype67d_chonky.html*)

3.4.5.1 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 72. Top Ten Most Significantly ($\text{padj} < 0.01$) Differentially Expressed Genes between group-housed 67d mutants and wildtypes (simultaneous model)

rank	multi				rando			
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	l(2)03659	0.17	5.85×10^{-241}	6.532	l(2)03659	0.17	8.29×10^{-238}	
2	CG7900	1.76	1.04×10^{-184}	6.541	CG7900	1.76	4.03×10^{-184}	
3	CG6912	0.36	3.10×10^{-137}	5.530	CG6912	0.36	4.98×10^{-136}	

4	CG10936	0.05	8.42×10^{-122}	3.525	CG10936	0.05	1.99×10^{-120}
5	Cyp9b1	0.45	2.19×10^{-116}	3.116	Cyp9b1	0.45	2.63×10^{-115}
6	DIP-alpha	0.06	4.64×10^{-104}	-4.815	NijC	1.54	2.16×10^{-103}
7	NijC	1.54	4.64×10^{-104}	-1.952	DIP-alpha	0.06	3.32×10^{-103}
8	CG32407	0.14	4.50×10^{-90}	3.439	CG32407	0.14	2.96×10^{-89}
9	5-HT2A	0.09	2.03×10^{-79}	-7.746	5-HT2A	0.09	3.41×10^{-79}
10	Cyp6a20	5.59	8.10×10^{-75}	-1.816	Cyp6a20	5.59	7.82×10^{-74}

Top 10 genes with biggest (significant) effect sizes

Table 73. Top Ten Largest Magnitude Fold Changes which were significant between group-housed 67d mutants and wildtypes (simultaneous model)

rank	name	multi			rando		
		expression	adjusted p	log2 FoldChange	name	expression	adjusted p
1	w	0.53	1.12×10^{-28}	13.839	w	0.53	1.12×10^{-28}
2	CG32437	0.04	1.57×10^{-14}	-11.155	CG32437	0.04	1.63×10^{-14}
3	CG43149	0.16	7.58×10^{-8}	10.193	CG43149	0.16	7.58×10^{-8}
4	lncRNA:CR44111	0.06	1.61×10^{-11}	-9.800	lncRNA:CR44111	0.06	1.71×10^{-11}
5	ppk9	0.00	5.84×10^{-8}	8.494	CG43291	0.01	2.23×10^{-10}
6	lncRNA:CR44377	0.01	2.22×10^{-7}	-8.094	ppk9	0.00	6.41×10^{-8}
7	lncRNA:dntRL	0.06	9.23×10^{-31}	7.995	lncRNA:CR44377	0.01	2.32×10^{-7}
8	CG9010	0.06	8.74×10^{-13}	-7.833	lncRNA:dntRL	0.06	1.03×10^{-30}
9	Obp83g	0.03	2.67×10^{-9}	7.829	CG9010	0.06	8.95×10^{-13}
10	5-HT2A	0.09	2.03×10^{-79}	-7.746	Obp83g	0.03	2.76×10^{-9}

Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

Table 74. Top Ten Highest Expressed Genes with Significant (padj < 0.05) Difference between group-housed 67d mutants and wildtypes (simultaneous model)

rank	name	multi			rando		
		expression	adjusted p	log2 FoldChange	name	expression	adjusted p
1	Obp83b	58.57	8.62×10^{-3}	0.385	Obp83b	58.58	8.76×10^{-3}
2	Obp69a	23.74	2.45×10^{-6}	0.683	Obp69a	23.75	2.90×10^{-6}
3	lush	11.92	1.86×10^{-4}	0.752	lush	11.92	1.90×10^{-4}
4	Cyp6w1	11.34	1.94×10^{-3}	0.576	Cyp6w1	11.34	1.94×10^{-3}
5	Snmp1	10.00	7.55×10^{-3}	-0.385	Snmp1	10.00	7.58×10^{-3}
6	CG1927	8.57	4.97×10^{-3}	0.335	CG1927	8.57	5.38×10^{-3}
7	Ldsdh1	6.16	9.71×10^{-3}	0.472	Ldsdh1	6.16	9.79×10^{-3}
8	CG30197	6.10	3.02×10^{-7}	1.024	CG30197	6.11	3.22×10^{-7}
9	Cyp6a2	5.98	1.67×10^{-15}	2.725	Cyp6a2	5.98	1.69×10^{-15}
10	Cyp6a20	5.59	8.10×10^{-75}	-1.816	Cyp6a20	5.59	7.82×10^{-74}

3.4.5.2 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

```
## png
## 2

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 74. Enriched GO Terms among Significantly Differentially Expressed Genes simultaneous 67d contrast; multi only; top 10 most significant per category

GO Term	Description	
MF		
GO:0005506	iron ion binding	8.20
GO:0004984	olfactory receptor activity	6.90
GO:0020037	heme binding	7.30
GO:0046906	NA	9.30
GO:0005549	odorant binding	3.80
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	1.00
GO:0016491	oxidoreductase activity	6.20
GO:0005215	transporter activity	6.70
GO:0022857	transmembrane transporter activity	1.20
GO:0048037	obsolete cofactor binding	5.60
BP		
GO:0007608	sensory perception of smell	7.40
GO:0050907	detection of chemical stimulus involved in sensory perception	3.70
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	3.90
GO:0009593	detection of chemical stimulus	5.90
GO:0050906	detection of stimulus involved in sensory perception	2.50
GO:0007600	NA	4.30
GO:0007606	sensory perception of chemical stimulus	1.80
GO:0051606	NA	1.80
GO:0055114	oxidation-reduction process	2.70
GO:0035721	intraciliary retrograde transport	5.80
CC		
GO:0016021	integral component of membrane	2.20
GO:0031224	intrinsic component of membrane	3.10
GO:0005576	extracellular region	2.30
GO:0044459	NA	7.20

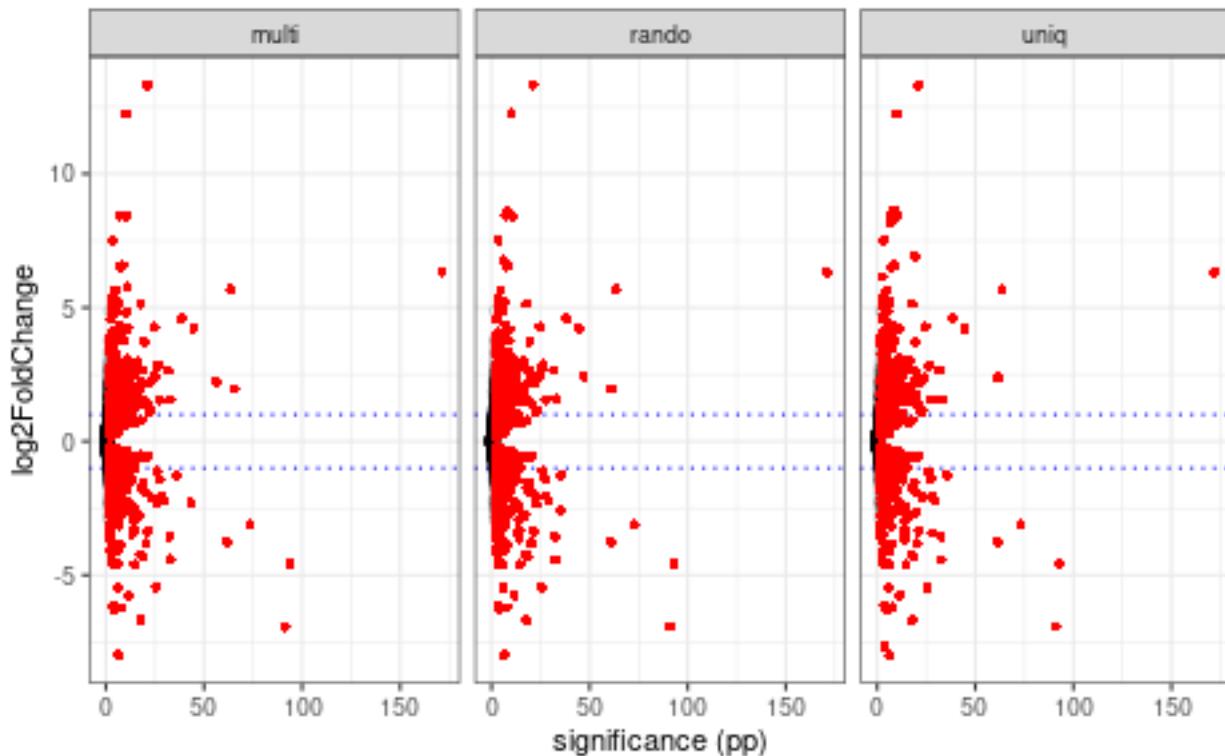
GO:0032590	dendrite membrane	2.8
GO:0044425	NA	3.0
GO:0016020	membrane	1.2
GO:0032589	neuron projection membrane	1.3
GO:0031253	NA	1.3
GO:0031256	leading edge membrane	1.0

3.4.6 FruLexA/Fru440 vs wt (Simultaneous model)

Of the 11590 genes with significance scores available, 1444 have an adjusted $p < 0.01$ (12.4590164 %)

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 depleted relative to the group-housed condition; positive fold changes are enriched

**Figure 57. Volcano Plot: Fold Change vs. Significance
(between group-housed FruLexA/Fru440 mutants and wildtypes, simultaneous)**



```
## png
## 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 517 such genes, mostly shared across alignment strategy:

(Table available at [results/tables/tbl76_hausWtVsMut_genotypeFruLexA440_chonky.html](#))

3.4.6.1 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 77. Top Ten Most Significantly (padj<0.01) Differentially Expressed between isolated and group-housed wildtypes

multi					rando				
rank	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange	
1	CG7900	1.76	4.37×10^{-172}	6.321	CG7900	1.76	1.25×10^{-171}	—	
2	DIP-alpha	0.06	1.44×10^{-94}	-4.565	DIP-alpha	0.06	1.12×10^{-93}	—	
3	5-HT2A	0.09	4.51×10^{-92}	-6.936	5-HT2A	0.09	1.25×10^{-91}	—	
4	Ets21C	0.07	3.56×10^{-74}	-3.117	Ets21C	0.07	1.41×10^{-73}	—	
5	CG32640	2.76	2.62×10^{-66}	1.963	CG11893	0.14	4.76×10^{-64}	—	
6	CG11893	0.14	1.61×10^{-64}	5.682	prom	0.04	1.19×10^{-61}	—	
7	prom	0.04	2.69×10^{-62}	-3.758	CG32640	1.88	1.19×10^{-61}	—	
8	CG32641	3.69	7.58×10^{-57}	2.242	CG32641	2.12	1.33×10^{-47}	—	
9	CG42526	0.05	1.78×10^{-45}	4.222	CG42526	0.05	3.22×10^{-45}	—	
10	Or19b	0.39	3.38×10^{-44}	-2.298	CG8665	0.07	7.21×10^{-39}	—	

Top 10 genes with biggest (significant) effect sizes

Table 78. Top Ten Largest Magnitude Fold Changes which were significant between group-housed FruLexaFru440 mutants and wildtypes

multi					rando				
rank	name	FB ID	expression	adjusted p	log2 FoldChange	name	FB ID	expression	adjusted p
1	mthl8	FBgn0052475	0.06	8.02×10^{-22}	13.321	mthl8	FBgn0052475	0.06	8.10×10^{-22}
2	CG43149	FBgn0262679	0.16	8.51×10^{-11}	12.245	CG43149	FBgn0262679	0.16	8.81×10^{-11}
3	ppk27	FBgn0035458	0.01	6.42×10^{-8}	8.431	CG43291	FBgn0262679	0.01	5.47×10^{-9}
4	w	FBgn0003996	0.53	3.81×10^{-11}	8.414	ppk27	FBgn0035458	0.53	6.62×10^{-8}
5	lncRNA:CR44377	FBgn0265527	0.01	4.86×10^{-7}	-7.986	w	FBgn0003996	0.01	3.98×10^{-11}
6	lncRNA:CR44285	FBgn0265312	0.03	3.91×10^{-4}	7.517	lncRNA:CR44377	FBgn0265527	0.03	4.86×10^{-7}
7	5-HT2A	FBgn0087012	0.09	4.51×10^{-92}	-6.936	lncRNA:CR44285	FBgn0265312	0.09	4.51×10^{-92}
8	CG32437	FBgn0052437	0.04	1.30×10^{-18}	-6.677	5-HT2A	FBgn0087012	0.04	1.30×10^{-18}
9	CR44003	FBgn0264745	0.02	3.05×10^{-9}	6.566	CR45496	FBgn0265312	0.02	3.05×10^{-9}
10	CR44383	FBgn0265533	0.08	3.42×10^{-8}	6.531	CG32437	FBgn0052437	0.08	3.42×10^{-8}

Table 78. Top Ten Largest Magnitude Fold Changes which were significant between group-housed FruLexaFru440 mutants and wildtypes, simultaneous n

multi					rando				
rank	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange	
1	mthl8	0.06	8.02×10^{-22}	13.321	mthl8	0.06	8.10×10^{-22}	—	
2	CG43149	0.16	8.51×10^{-11}	12.245	CG43149	0.16	8.81×10^{-11}	—	
3	ppk27	0.01	6.42×10^{-8}	8.431	CG43291	0.01	5.47×10^{-9}	—	
4	w	0.53	3.81×10^{-11}	8.414	ppk27	0.01	6.62×10^{-8}	—	
5	lncRNA:CR44377	0.01	4.86×10^{-7}	-7.986	w	0.01	3.98×10^{-11}	—	

6	lncRNA:CR44285	0.03	3.91×10^{-4}	7.517	lncRNA:CR44377	0.01	5.14×10^{-7}
7	5-HT2A	0.09	4.51×10^{-92}	-6.936	lncRNA:CR44285	0.03	3.88×10^{-4}
8	CG32437	0.04	1.30×10^{-18}	-6.677	5-HT2A	0.09	1.25×10^{-91}
9	CR44003	0.02	3.05×10^{-9}	6.566	CR45496	0.02	9.43×10^{-7}
10	CR44383	0.08	3.42×10^{-8}	6.531	CG32437	0.04	1.52×10^{-18}

Top 10 highest expressed genes with significant change

Ranking by DESeq2-based expression (ie, basemean scaled by gene length, in units of standard reads per base)

Table 79. Top Ten Highest Expressed Genes with Significant (padj < Difference
between group-housed FruLexaFru440 mutants and wildtypes, simultaneous model

rank	multi				rando			
	name	expression	adjusted p	log2 FoldChange	name	expression	adjusted p	log2 FoldChange
1	Obp83b	58.57	5.01×10^{-3}	0.398	Obp83b	58.58	4.93×10^{-3}	0.399
2	Obp69a	23.74	8.36×10^{-4}	0.496	Obp69a	23.75	8.71×10^{-4}	0.494
3	Orco	13.02	4.28×10^{-3}	-0.572	Orco	13.02	4.41×10^{-3}	-0.582
4	Obp56d	9.02	3.36×10^{-5}	1.168	Obp56d	9.03	3.41×10^{-5}	1.169
5	Obp59a	8.89	4.27×10^{-3}	0.334	Obp59a	8.90	4.03×10^{-3}	0.335
6	sesB	8.63	1.39×10^{-3}	0.355	sesB	8.63	1.47×10^{-3}	0.357
7	Cyp6a2	5.98	1.64×10^{-4}	1.317	Cyp6a2	5.98	1.63×10^{-4}	1.318
8	PHGPx	5.64	5.14×10^{-3}	0.480	PHGPx	5.64	5.03×10^{-3}	0.482
9	Cyp6a20	5.59	6.38×10^{-37}	-1.275	Cyp6a20	5.59	2.78×10^{-36}	-1.273
10	CG5973	4.82	1.82×10^{-9}	0.583	CG5973	4.82	2.22×10^{-9}	0.584

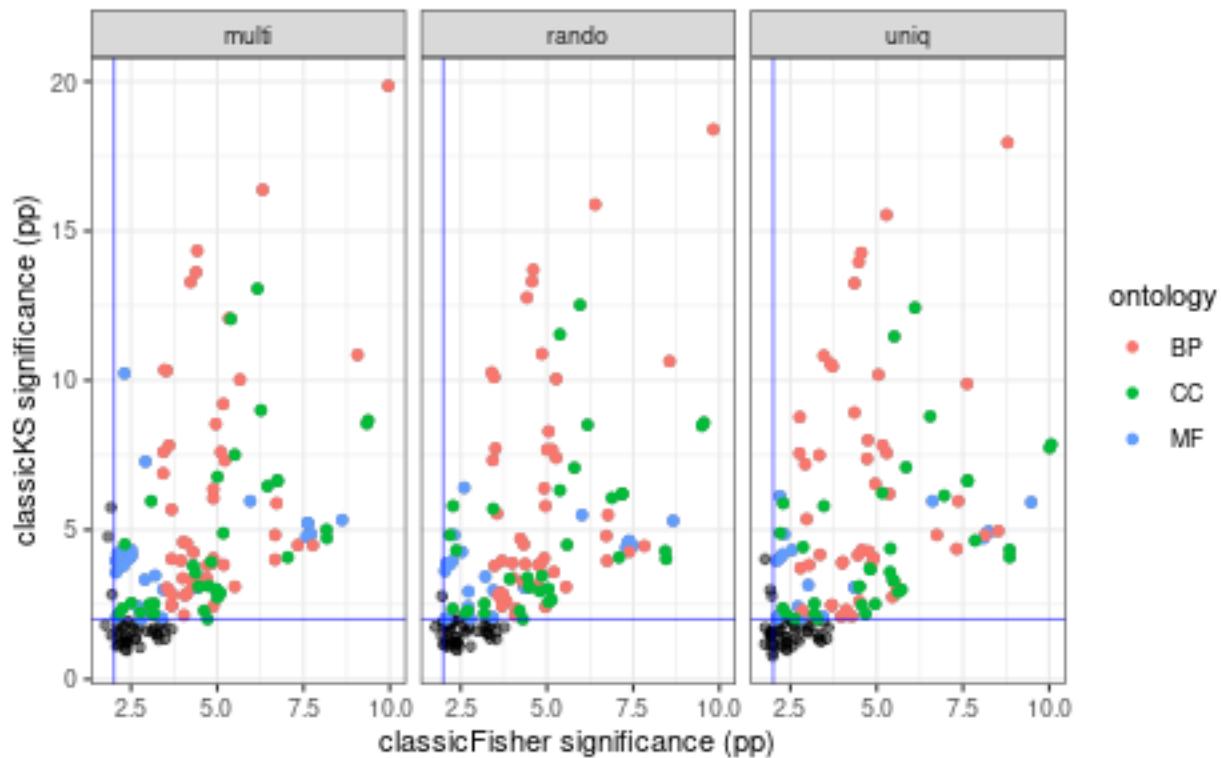
3.4.6.2 Gene Ontology Enrichment

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



```
## png
## 2

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 80. Enriched GO Terms among Significantly Differentially Expressed Genes simultaneous FruLexa440 contrast; multi only; top 10 most significant per category

GO Term	Description	p-	Fisher
MF			
GO:0004984	olfactory receptor activity		2.40×10^{-9}
GO:0004871	NA		1.90×10^{-8}

GO:0038023	signaling receptor activity	2.40×10^{-8}
GO:0060089	NA	2.40×10^{-8}
GO:0004888	transmembrane signaling receptor activity	2.60×10^{-8}
GO:0005549	odorant binding	1.10×10^{-6}
GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	4.50×10^{-5}
GO:0015924	NA	4.50×10^{-5}
GO:0051787	misfolded protein binding	3.70×10^{-4}
GO:0004930	G protein-coupled receptor activity	4.00×10^{-4}
<hr/>		
BP		
GO:0050896	response to stimulus	1.10×10^{-10}
GO:0042221	NA	8.60×10^{-10}
GO:0050907	detection of chemical stimulus involved in sensory perception	1.70×10^{-8}
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	4.60×10^{-8}
GO:0007608	sensory perception of smell	1.90×10^{-7}
GO:1905515	non-motile cilium assembly	2.10×10^{-7}
GO:0009593	detection of chemical stimulus	2.10×10^{-7}
GO:0051716	NA	4.80×10^{-7}
GO:0010033	response to organic substance	2.20×10^{-6}
GO:0007187	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	3.10×10^{-6}
<hr/>		
CC		
GO:0120025	NA	4.20×10^{-10}
GO:0042995	cell projection	4.60×10^{-10}
GO:0016021	integral component of membrane	6.60×10^{-9}
GO:0031224	intrinsic component of membrane	6.80×10^{-9}
GO:0032590	dendrite membrane	9.00×10^{-8}
GO:0044463	NA	1.80×10^{-7}
GO:0120038	NA	1.80×10^{-7}
GO:0043005	neuron projection	3.50×10^{-7}
GO:0044459	NA	5.40×10^{-7}
GO:0071944	cell periphery	6.80×10^{-7}

3.4.7 Transcriptional Profiles

Backing away from differential expression, we can also look at transcriptional profiles of treatment groups by gene.

Heatmaps were made representing expression as color intensity; for genes which did were not modeled due to low overall read count, the baseMean-derived expression was filled in as zero. Since these values range over several orders of magnitude, logarithmic scales were used; $\log_{10}(0)$ was defined as -999

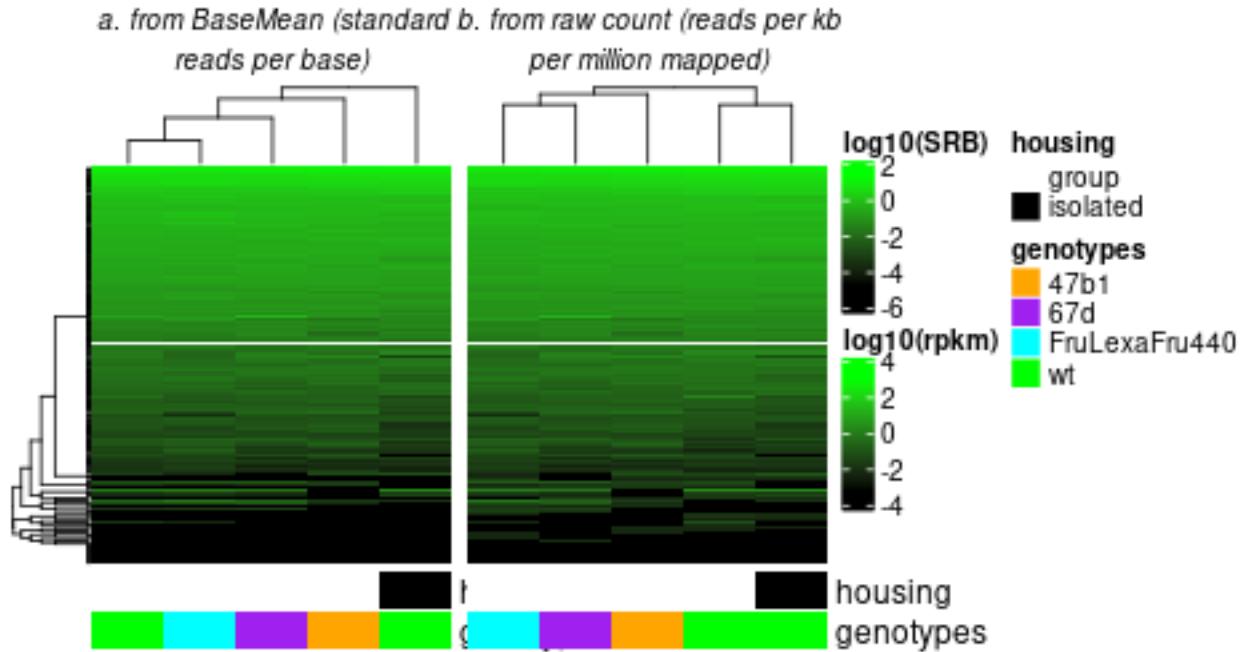
Since the RPKM-derived expression values have more clusterable genes (b/c fewer genes with all 0's), these were used to cluster both heatmaps. (This doesn't necessarily mean that the finer clustering is meaningful!)

To try to put the expression on a common scale, absolute expression values were scaled on a by-gene basis, with each gene's expression values being divided by the sum of those expressions, to calculate an expression share. Genes with a sum-expression of zero were assigned an expression share of 0 for all treatments. The relative expressions were clustered and heatplots graphed. These are susceptible to low-level noise: a single stray read is enough to mean the difference between all samples having 0% of the reads and one sample having 100% of all reads.

3.4.7.1 ion channel activity genes

Here is a heatmap of transcriptional profiles for ion-channel activity genes (GO:0005216) from the samples in the housing and genotype comparison (Table 5a).

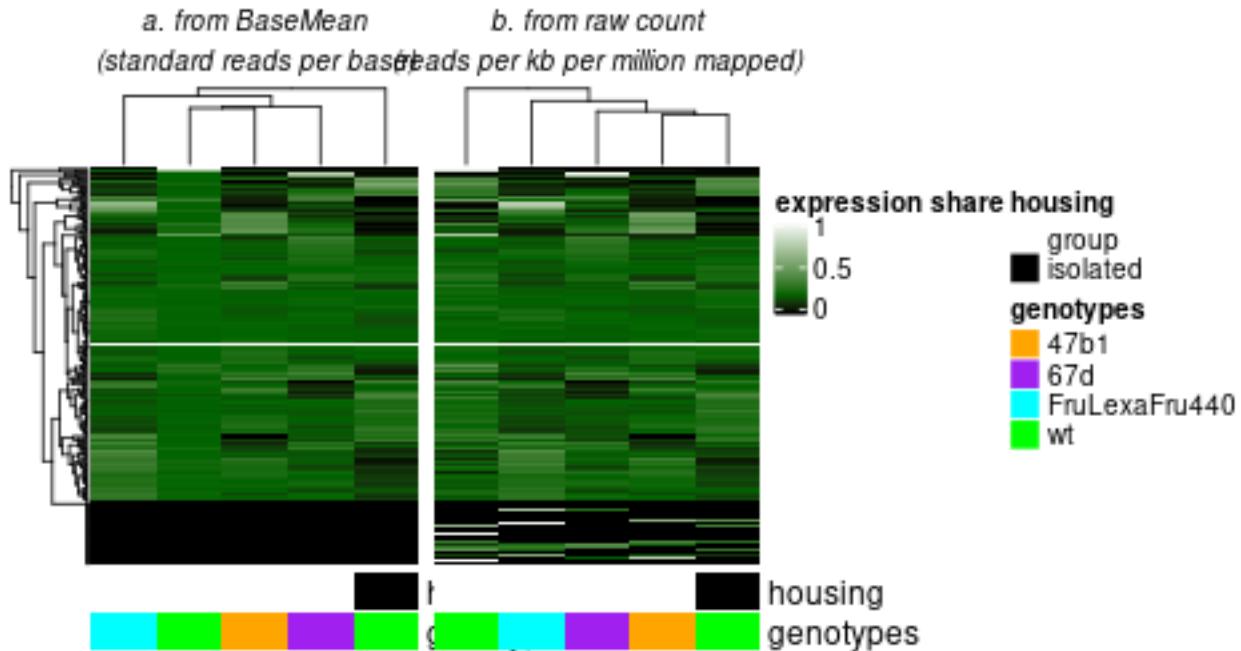
**Figure 59 . Absolute Expression Heatmap
for Ion Channel Activity Genes**
(simultaneous housing/genotype model, multi only)



```
## png
## 2
```

To try to better display relative differences between samples, a relative expression share was calculated for each gene:

**Figure 60 Relative Expression Heatmap
for Ion Channel Activity Genes
(simultaneous housing/genotype model)**



```
## png
## 2
```

3.4.8 Fruitless-less

Out of data quality concerns, the contrast was rerun with the FruLexa/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 81. 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 ($\text{padj} < 0.01$) 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 82. Changes in Differential Expression Significance when FruLexa/Fru440 samples are dropped

	change		
	gain	loss	none
47b1			
multi	387	17	12838
rando	392	18	12671
uniq	384	20	12581
67d			
multi	603	20	12619
rando	582	17	12482
uniq	562	21	12402
isolated			
multi	33	1	13208
rando	28	2	13051
uniq	28	3	12954

In some cases, the significance increase was very large:

Table 83. Top 10 Biggest Significance Changes when FruLexa/Fru440 samples are dropped

	effect size (l2fc)		adjusted p	
	with	without	with	without
47b1 - multi				
csw	0.51	0.66	0.087	1.40×10^{-10}
GstD8	-0.60	-0.67	0.013	2.62×10^{-13}
kek1	0.71	0.90	0.047	1.05×10^{-10}
CG17572	0.66	0.78	0.023	5.89×10^{-12}
Tsp42En	-0.59	-0.68	0.023	2.46×10^{-9}
CG10638	-0.30	-0.32	0.093	1.07×10^{-8}
CG13251	-0.28	-0.29	0.084	4.10×10^{-9}
Nrx-1	0.41	0.57	0.230	1.20×10^{-8}
pyd	0.51	0.61	0.046	1.86×10^{-14}
Fhos	-0.24	-0.57	0.540	2.41×10^{-10}
67d - multi				
csw	0.50	0.64	0.085	4.76×10^{-10}
Spn	-0.37	-0.41	0.060	1.69×10^{-9}
ssx	0.71	0.84	0.013	1.00×10^{-18}
RhoGAP19D	0.56	0.63	0.012	4.86×10^{-12}
CG12814	0.56	0.69	0.053	4.42×10^{-10}
Hexo1	0.70	0.81	0.018	7.13×10^{-10}
RIC-3	-0.48	-0.57	0.080	3.21×10^{-9}
CG42795	-0.65	-0.73	0.018	6.37×10^{-11}
lncRNA:CR44317	-1.30	-1.62	0.016	2.73×10^{-10}
Der-1	0.55	0.62	0.019	2.13×10^{-11}
isolated - multi				

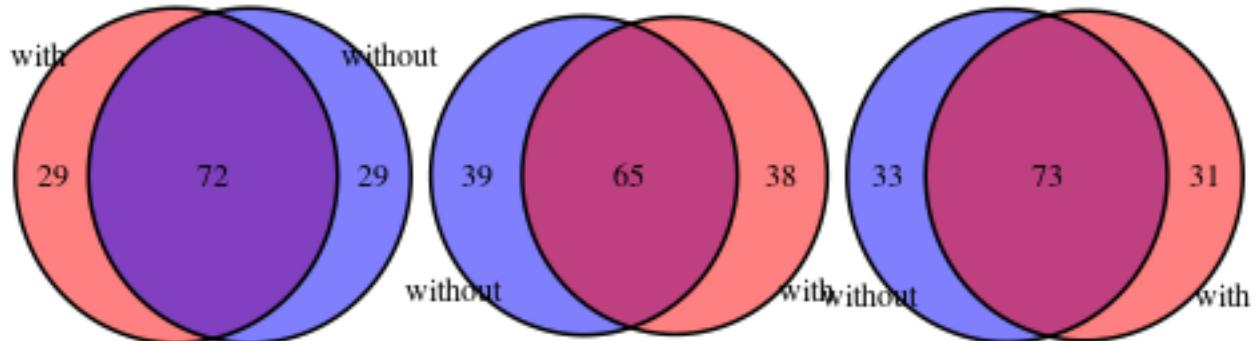
Trp1	0.01	0.27	0.507	1.30×10^{-3}
Gmap	0.38	0.41	0.018	1.14×10^{-4}
CG15202	0.01	0.39	0.508	2.94×10^{-4}
CG9498	0.53	0.71	0.134	8.26×10^{-4}
Ugt301D1	0.25	0.40	0.190	2.82×10^{-4}
Loxl2	0.01	0.43	0.342	8.32×10^{-4}
ELOVL	0.41	0.51	0.085	2.94×10^{-4}
CG9717	0.01	0.50	0.565	1.30×10^{-3}
vir-1	0.02	0.22	0.347	9.71×10^{-4}
Cda5	-0.01	-0.50	0.412	1.44×10^{-3}

3.4.8.1 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-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

Figure 62 . Similarity of Housing Contrast Top 100 Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

adjusted p:	log2FoldChange:	expression:
55% similar	46% similar	53% similar

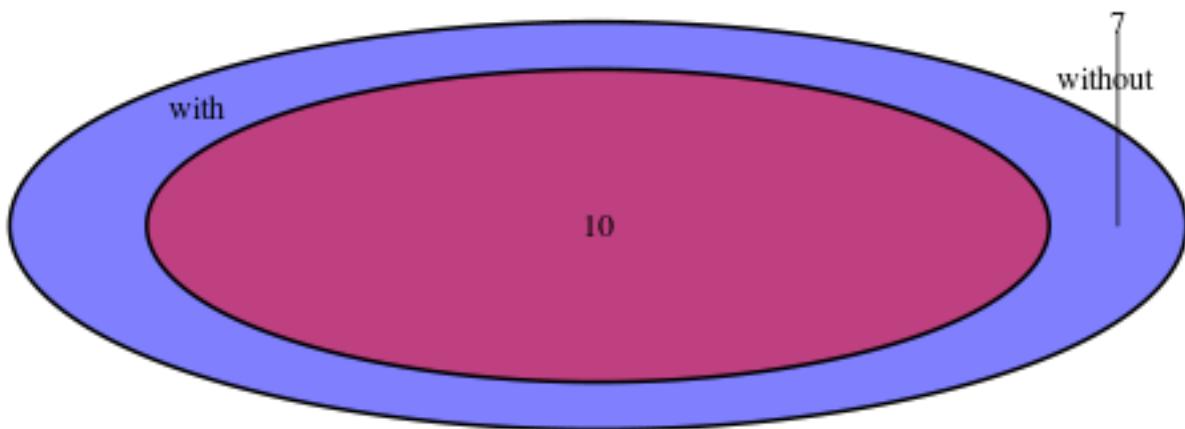


```
## png
## 2
```

Comparing the “chonky” gene lists ($\text{padj} < 0.01$, $\text{abs(l2fc)} > 1$):

Figure 63 . Similarity of Housing Contrast Chonky Lists Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

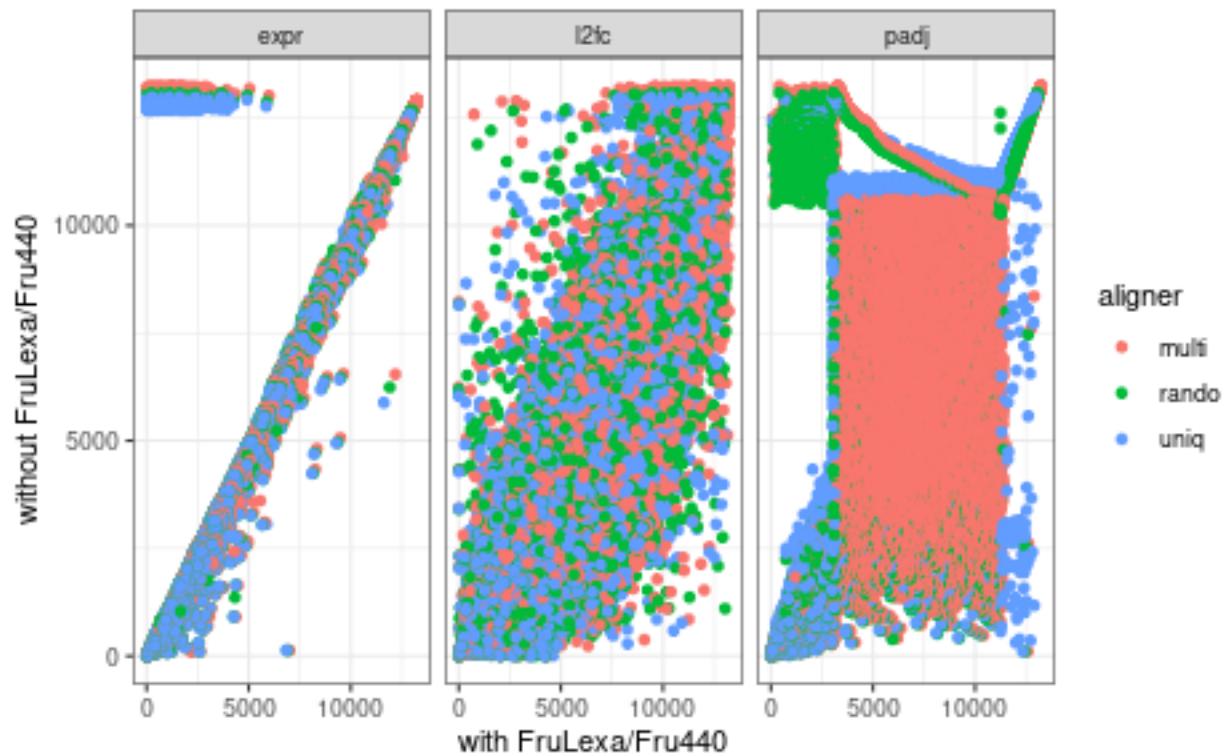
chonky:
59% similar



```
## png  
## 2
```

We can also look at the rank correlations:

Figure 64. Rank correlations of expression, effect size, and significance (housing contrasts, with/without FruLexa)



```
## png
## 2
```

3.4.8.2 Perturbation to 47b1 Contrast

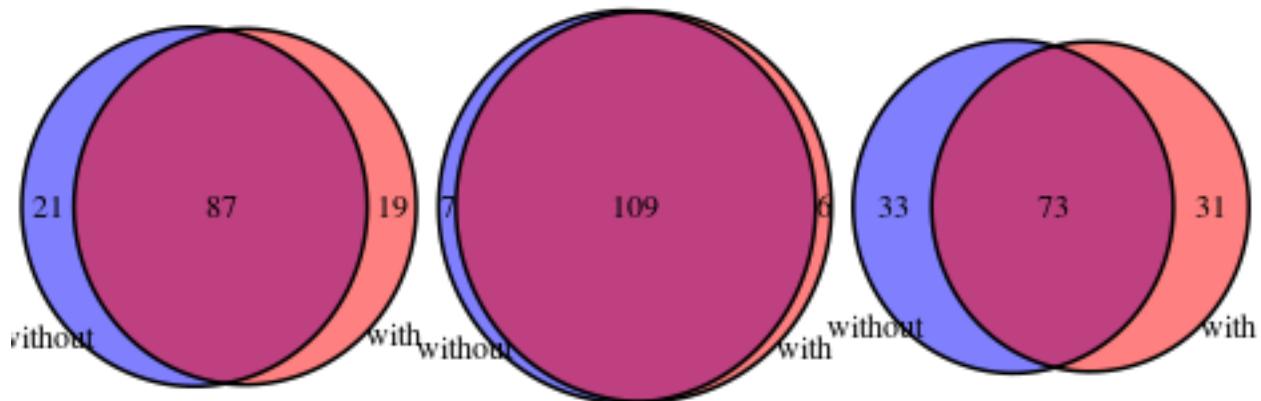
To see how much exclusion of the FruLexa/Fru440 alters the big picture results in the 47b1 vs. wt contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

Figure 65 . Similarity of 47b1 Contrast Top 10 Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

adjusted p:
69% similar

log2FoldChange:
89% similar

expression:
53% similar

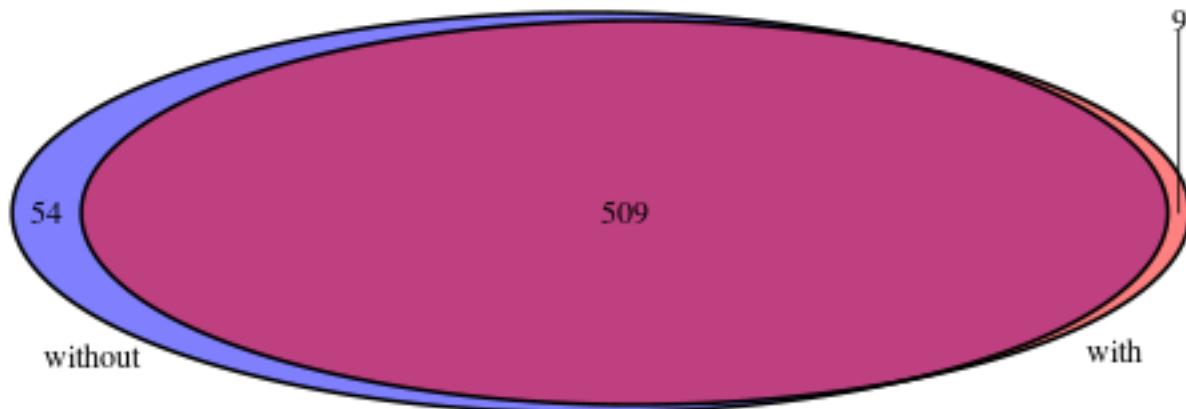


```
## png  
## 2
```

Comparing the “chonky” gene lists ($\text{padj} < 0.01$, $\text{abs(l2fc)} > 1$):

Figure 66 . Similarity of 47b Contrast Chonky Lists Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

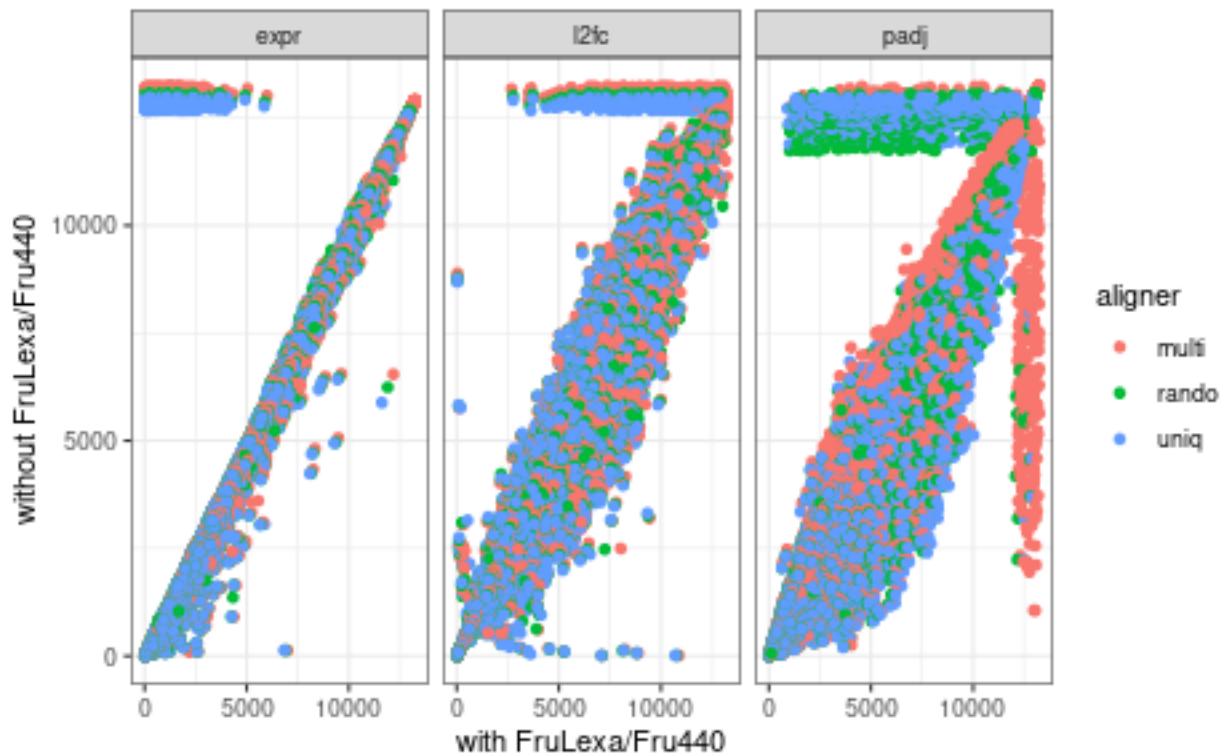
chonky:
89% similar



```
## png  
## 2
```

We can also look at the rank correlations:

Figure 67. Rank correlations of expression, effect size, and significance (47b1 contrasts, with/without FruLexa)



```
## png
## 2
```

3.4.8.3 Perturbation to 67d Contrast

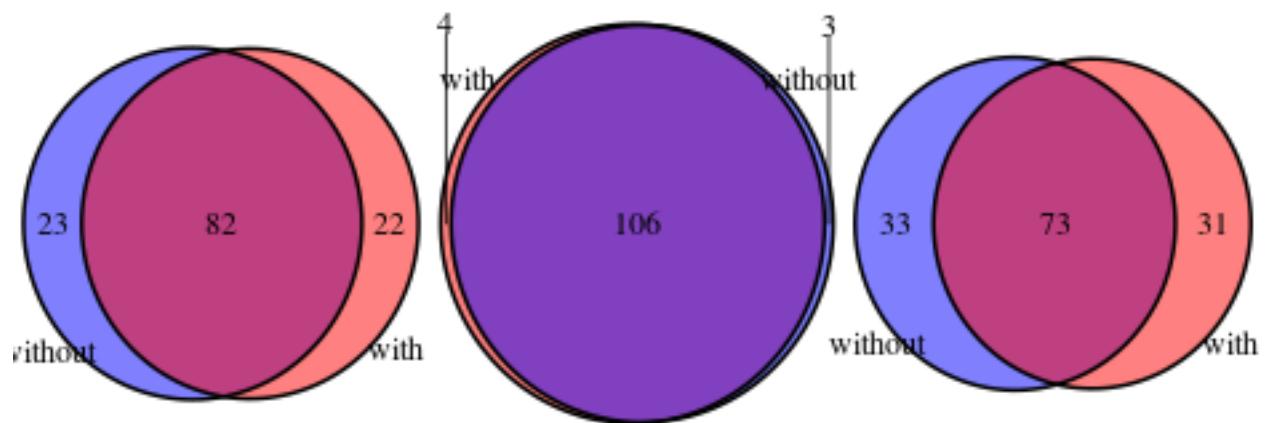
To see how much exclusion of the FruLexa/Fru440 alters the big picture results in the 67d vs. wt contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

↳ 68 . Similarity of 67d Contrast Top 10 Lists, with/without FruLexaFru440 samples (pooled alignr

adjusted p:
65% similar

log2FoldChange:
94% similar

expression:
53% similar

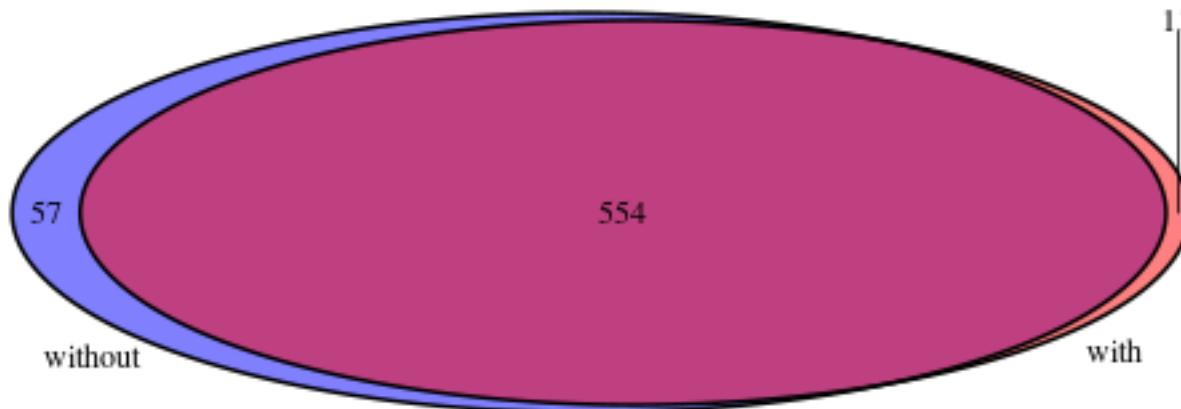


```
## png  
## 2
```

Comparing the “chonky” gene lists ($\text{padj} < 0.01$, $\text{abs(l2fc)} > 1$):

Figure 69 . Similarity of 67d Contrast Chonky Lists Lists,
with/without FruLexaFru440 samples (pooled alignment strategies)

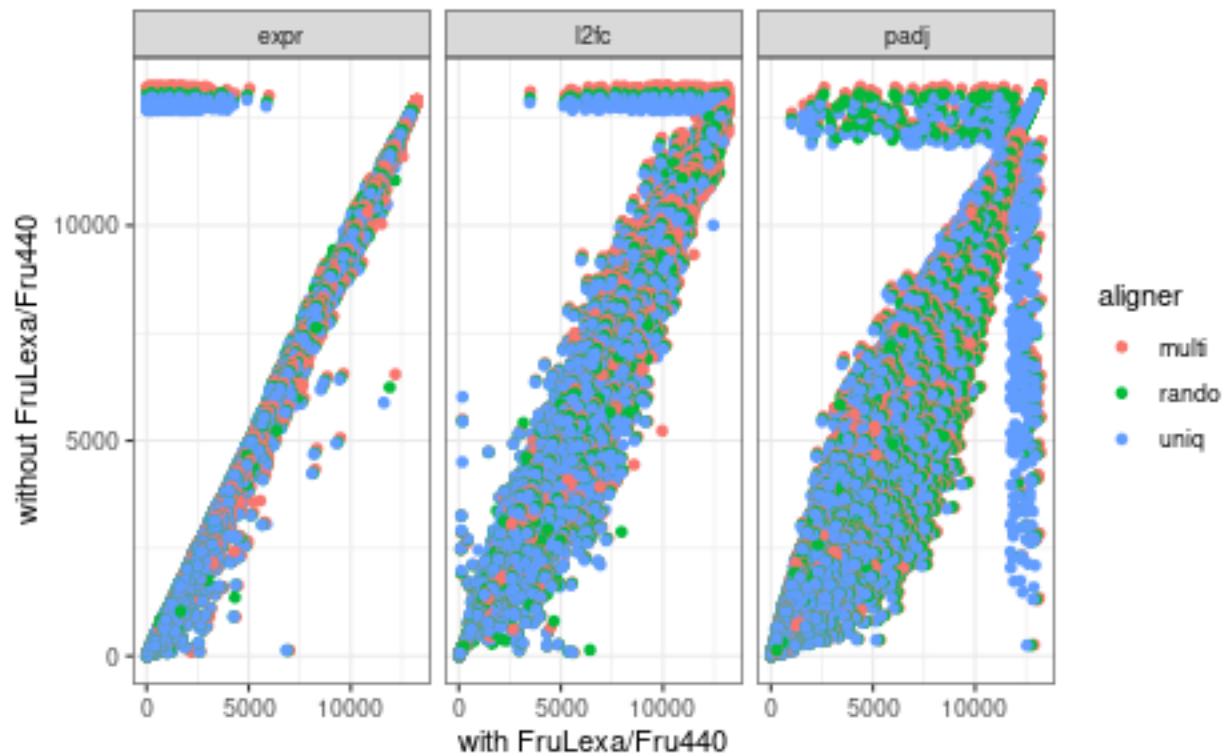
chonky:
89% similar



```
## png  
## 2
```

We can also look at the rank correlations:

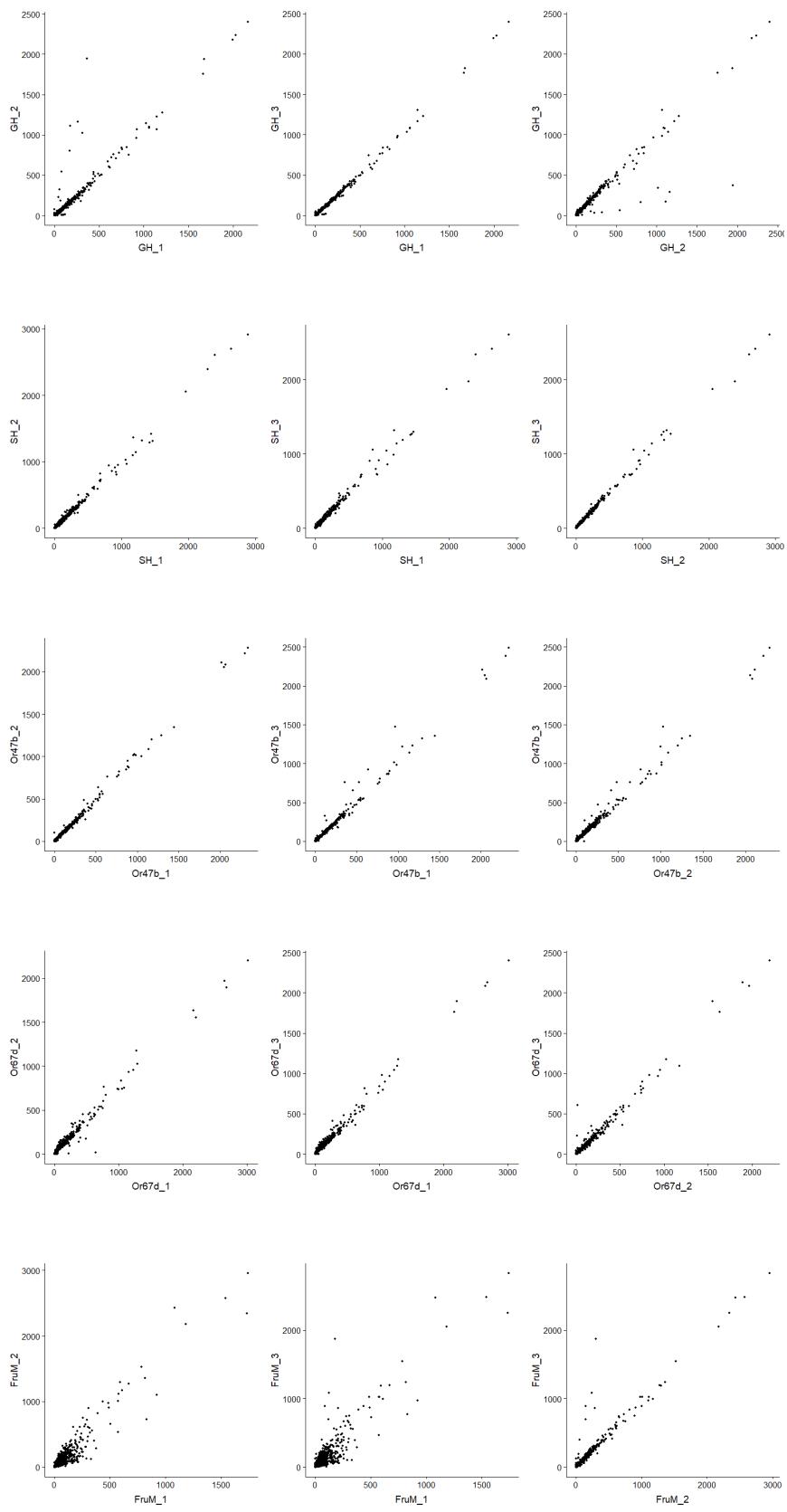
Figure 70. Rank correlations of expression, effect size, and significance (67d contrasts, with/without FruLexa)



```
## png
## 2
```

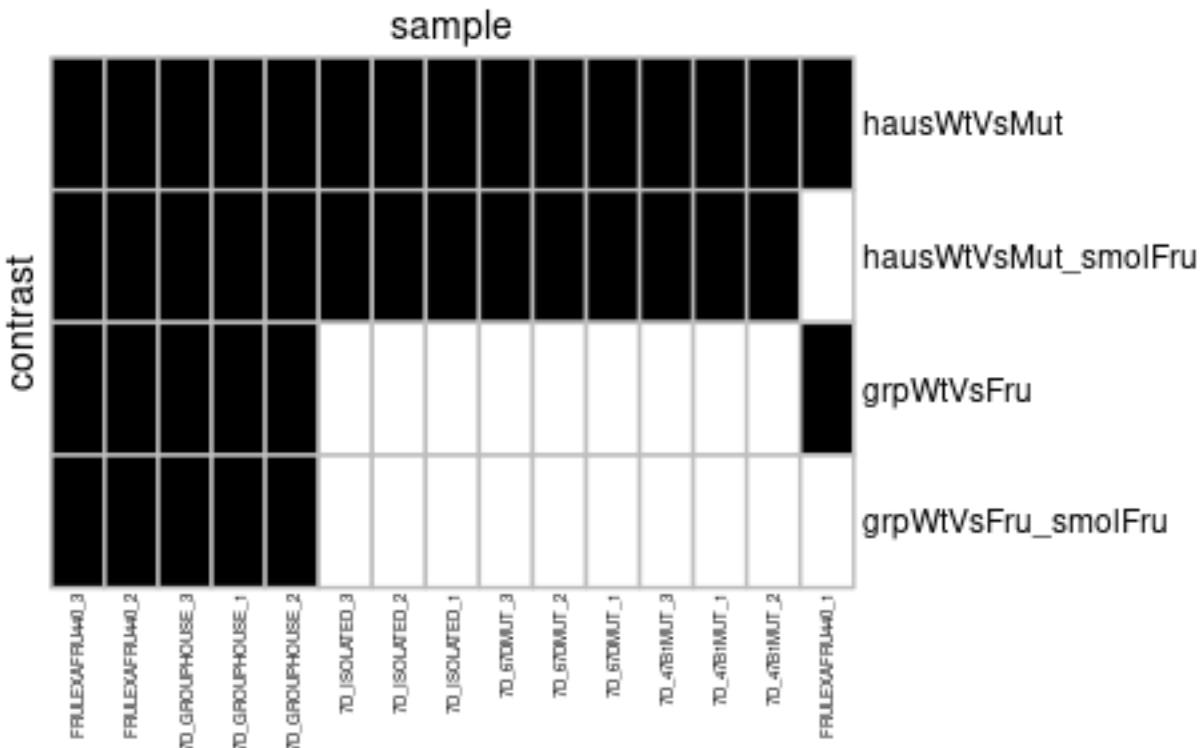
3.4.9 Reduced FruLexaFru440 Samples

FruLexaFru440 replicate 1 has been flagged as specifically problematic, possibly b/c of sex contamination.



source: qichen duan , 21 Dec

Figure 71. RNASeq Samples Used in Reduced Fruitless Comparisons



```
## png
## 2
```

The same pairwise comparisons as above are repeated on a reduced model, in which replicate 1 is excluded. The data are found here: “results/tables/supp/hausWtVsMut_smolFru.allAligners.DESeq2.MpBC.reformatted.tsv”

Every gene that can be analyzed using the Reduced counts can be analyzed using the With counts; there are 0 that cannot. On the other hand, there are 145 genes which can be analyzed using the Full counts but not the Reduced:

Table 84. Genes Lost When FruLexa/Fru440 replicate 1 Counts are Excluded
genes which no longer pass minimum count threshold

aligner	count
multi	139
rando	138
uniq	139

In 0 cases were these genes significant ($\text{padj} < 0.01$) 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 85. Changes in Differential Expression Significance
when FruLexa/Fru440 replicate 1 is dropped

	change		
	gain	loss	none
47b1			
multi	253	60	12929
rando	242	57	12782
uniq	245	49	12691
67d			
multi	203	268	12771
rando	193	278	12610
uniq	187	278	12520
FruLexaFru440			
multi	396	480	12366
rando	391	472	12218
uniq	391	476	12118
isolated			
multi	22	7	13213
rando	17	7	13057
uniq	19	7	12959

In some cases, the significance increase was very large:

Table 86. Top 10 Biggest Significance Changes
when FruLexa/Fru440 replicate 1 is dropped

	effect size (l2fc)		adjusted p	
	full	reduced	full	reduced
47b1 - multi				
csw	0.51	0.67	8.71×10^{-2}	9.63×10^{-13}
GstD8	-0.60	-0.66	1.26×10^{-2}	5.64×10^{-9}
Fer1HCH	-0.36	-0.40	6.21×10^{-2}	2.34×10^{-10}
kek1	0.71	0.91	4.71×10^{-2}	5.06×10^{-14}
CG17572	0.66	0.78	2.31×10^{-2}	3.78×10^{-9}
CG3961	-0.27	-0.33	2.96×10^{-1}	8.59×10^{-8}
Nrx-1	0.41	0.58	2.30×10^{-1}	1.32×10^{-9}
CG42541	0.71	0.88	3.76×10^{-2}	2.73×10^{-14}
dnr1	0.50	0.62	8.45×10^{-2}	1.33×10^{-8}
pyd	0.51	0.60	4.65×10^{-2}	2.33×10^{-14}
67d - multi				
csw	0.50	0.58	8.47×10^{-2}	7.90×10^{-8}
ssx	0.71	0.78	1.25×10^{-2}	3.18×10^{-13}
NT5E-2	0.68	0.90	1.34×10^{-2}	7.92×10^{-9}
RIC-3	-0.48	-0.61	8.02×10^{-2}	1.50×10^{-8}
ppk9	8.49	0.06	5.84×10^{-8}	3.04×10^{-2}
CG42541	0.78	0.79	1.98×10^{-2}	2.22×10^{-9}
CG42795	-0.65	-0.75	1.76×10^{-2}	1.38×10^{-8}

CG43902	-0.42	-0.63	1.71×10^{-1}	8.67×10^{-8}
lncRNA:CR44317	-1.30	-1.81	1.59×10^{-2}	2.50×10^{-8}
Der-1	0.55	0.65	1.89×10^{-2}	1.31×10^{-8}
<hr/>				
FruLexaFru440 - multi				
slgA	-0.19	-0.38	2.40×10^{-1}	8.60×10^{-12}
X11L	-0.12	-0.52	6.43×10^{-1}	9.00×10^{-13}
CG15270	-0.33	-0.71	1.53×10^{-1}	2.87×10^{-20}
CG15760	-0.23	-0.94	4.16×10^{-1}	4.61×10^{-11}
CG4297	-0.28	-0.63	2.08×10^{-1}	1.24×10^{-11}
CG2269	-0.14	-0.49	5.65×10^{-1}	9.41×10^{-11}
SP2353	-0.25	-0.85	3.68×10^{-1}	1.05×10^{-12}
CG16711	-0.34	-0.57	5.18×10^{-2}	7.20×10^{-12}
CG3961	-0.19	-0.50	3.69×10^{-1}	1.38×10^{-13}
nwk	-0.20	-0.51	3.50×10^{-1}	8.31×10^{-16}
<hr/>				
isolated - multi				
Fer1HCH	0.01	0.22	7.70×10^{-1}	6.18×10^{-3}
CG15270	-0.01	-0.29	7.81×10^{-1}	6.53×10^{-4}
CG15202	0.01	0.39	5.08×10^{-1}	3.06×10^{-4}
CG5171	-0.01	-1.30	1.16×10^{-1}	4.05×10^{-3}
Cyp9h1	0.37	0.49	1.47×10^{-1}	4.96×10^{-3}
LoxI2	0.01	0.42	3.42×10^{-1}	5.87×10^{-3}
CG9717	0.01	0.48	5.65×10^{-1}	8.21×10^{-3}
Npc2g	-0.00	-0.01	3.56×10^{-2}	1.13×10^{-3}
vir-1	0.02	0.22	3.47×10^{-1}	1.97×10^{-3}
Cda5	-0.01	-0.48	4.12×10^{-1}	9.63×10^{-3}

3.4.9.1 Perturbation to Housing Contrast

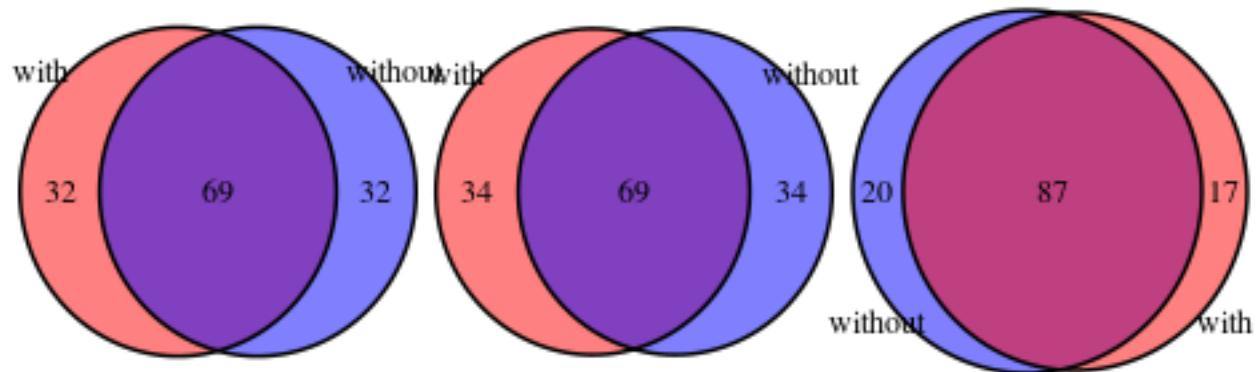
To see how much exclusion of the FruLexa/Fru440 replicate 1 alters the big picture results in the group vs. isolated contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

Figure 72 . Similarity of Housing Contrast Top 100 Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

adjusted p:
52% similar

log2FoldChange:
50% similar

expression:
70% similar

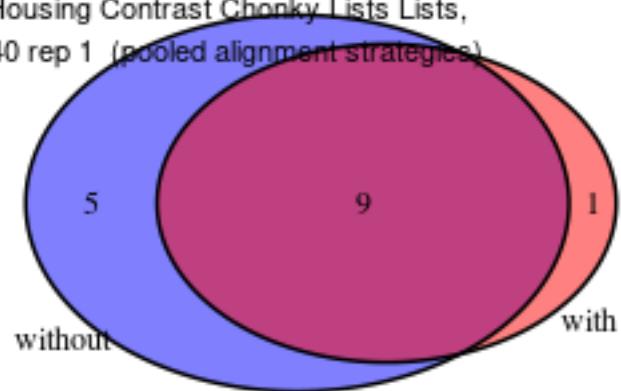


```
## png  
## 2
```

Comparing the “chonky” gene lists ($\text{padj} < 0.01$, $\text{abs(l2fc)} > 1$):

Figure 73 . Similarity of Housing Contrast Chonky Lists Lists,
with/without FruLexaFru440 rep 1 (pooled alignment strategies)

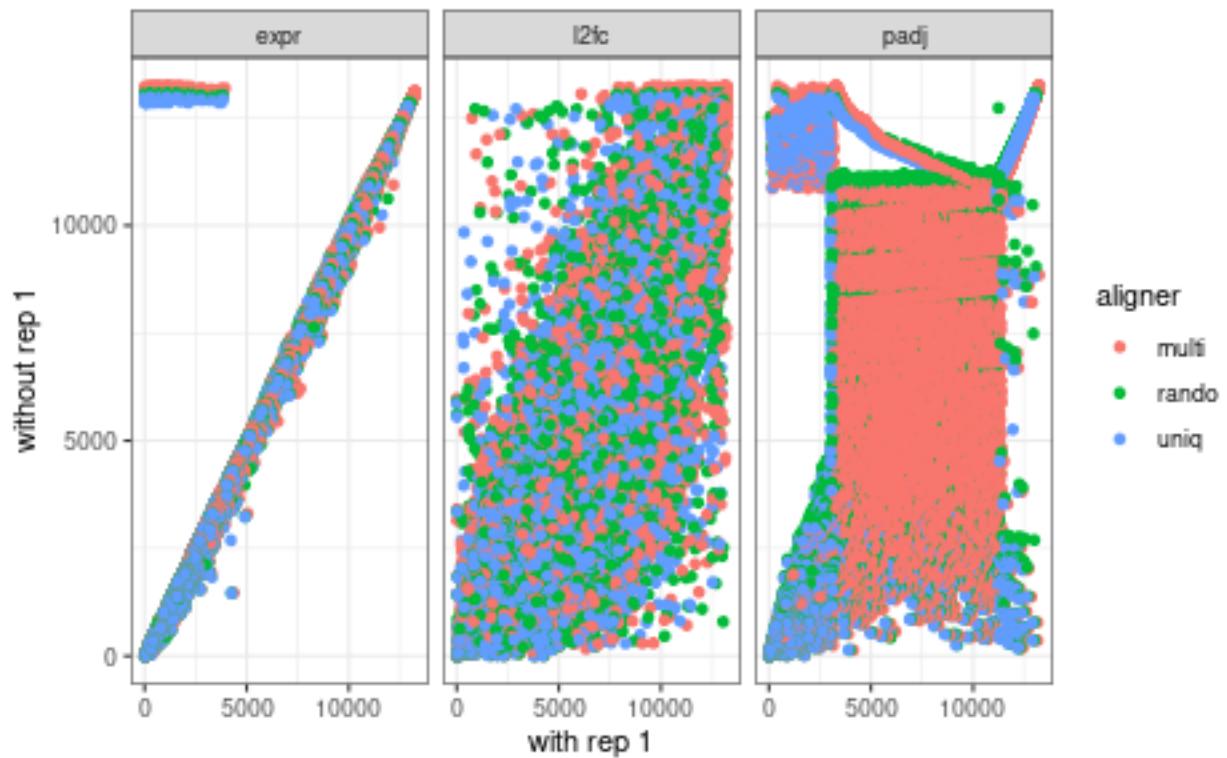
chonky:
60% similar



```
## png  
## 2
```

We can also look at the rank correlations:

Figure 74. Rank correlations of expression, effect size, and significance (housing contrasts, with/without problematic FruLexaFru440 replicate)



```
## png
## 2
```

3.4.9.2 Perturbation to 47b1 Contrast

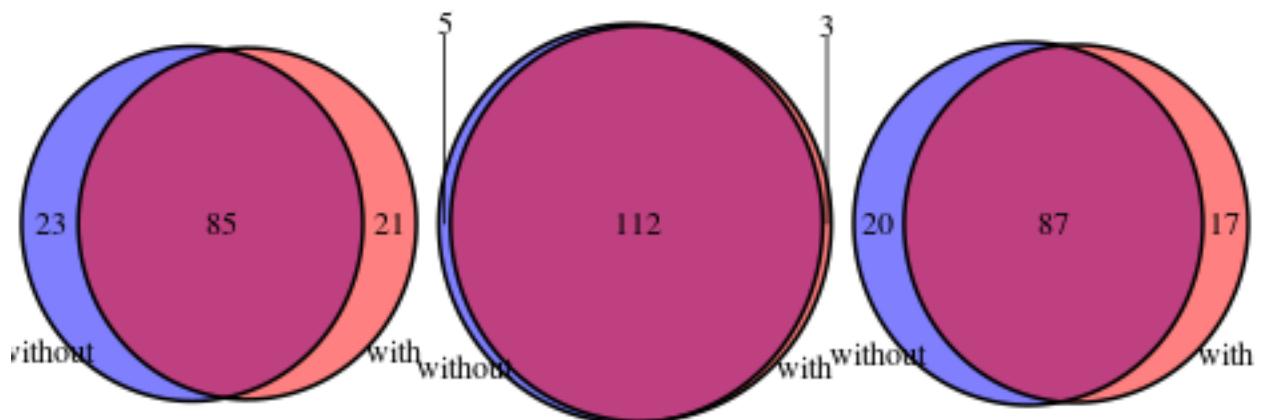
To see how much exclusion of the FruLexa/Fru440 replicate 1 alters the big picture results in the group vs. isolated contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

Figure 75 . Similarity of 47b1 Contrast Top 100 Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

adjusted p:
66% similar

log2FoldChange:
93% similar

expression:
70% similar

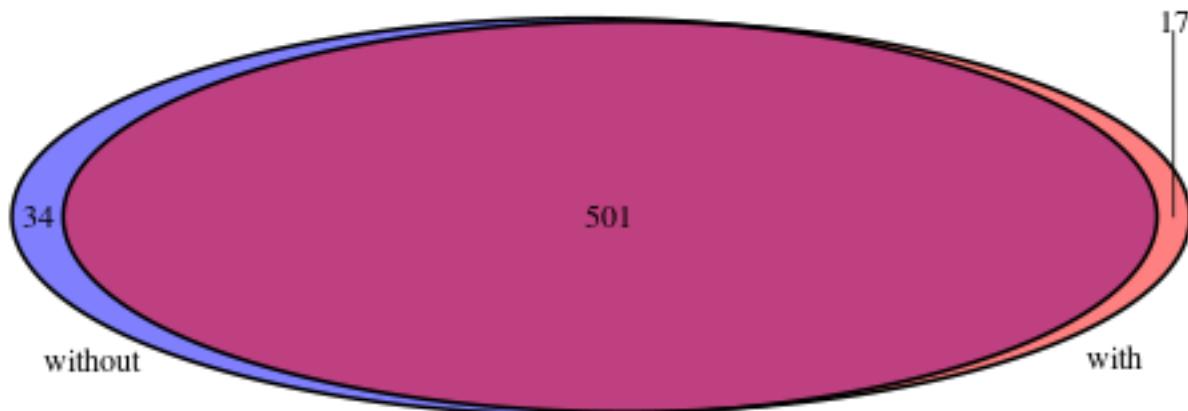


```
## png  
## 2
```

Comparing the “chonky” gene lists ($\text{padj} < 0.01$, $\text{abs(l2fc)} > 1$):

Figure 76 . Similarity of 67d Contrast Chonky Lists Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

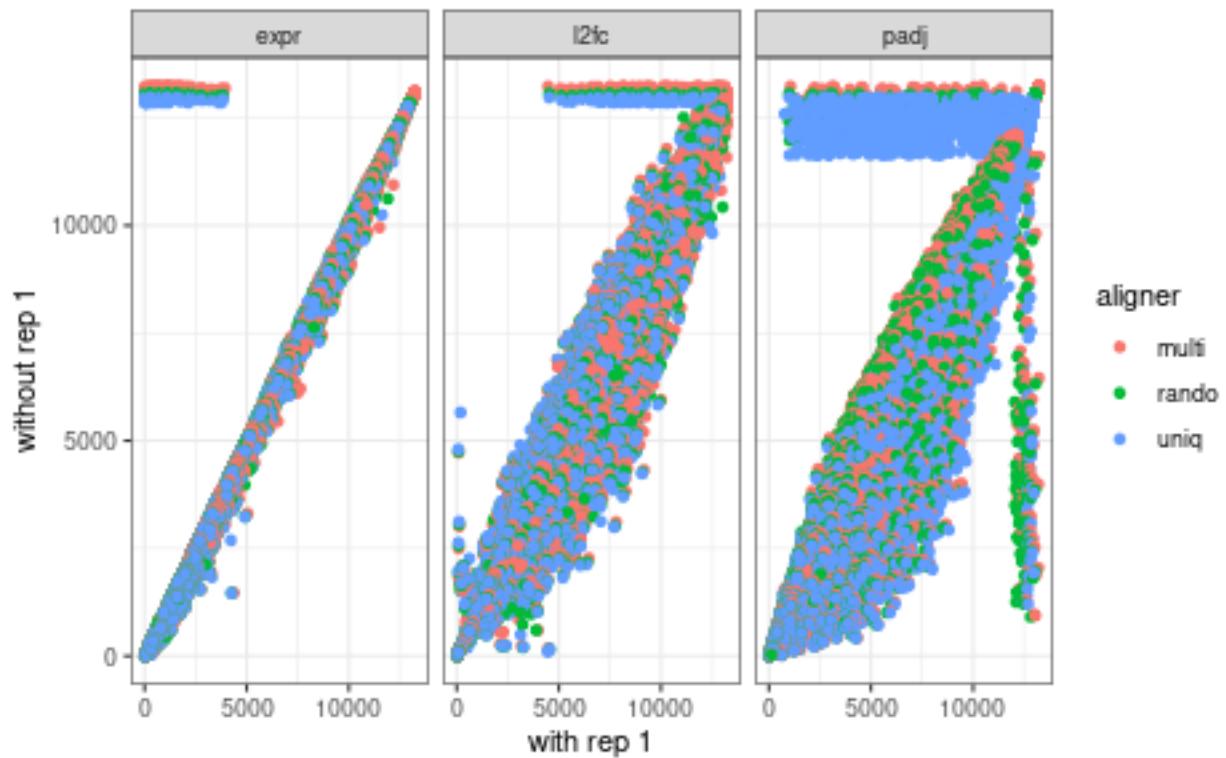
chonky:
91% similar



```
## png  
## 2
```

We can also look at the rank correlations:

Figure 77. Rank correlations of expression, effect size, and significance (47b1 contrasts, with/without problematic FruLexaFru440 replicate)



```
## png
## 2
```

3.4.9.3 Perturbation to 67d Contrast

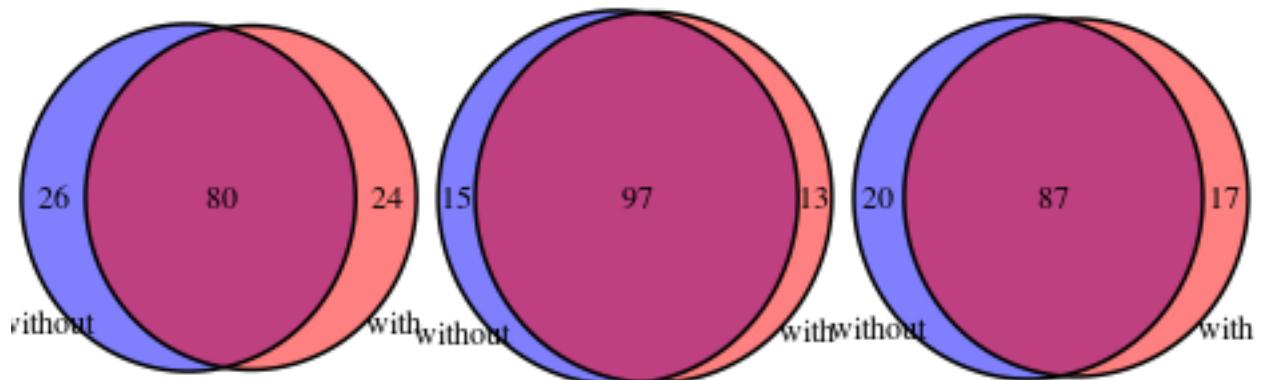
To see how much exclusion of the FruLexa/Fru440 replicate 1 alters the big picture results in the 67d vs. wt contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

Figure 78 . Similarity of 67d Contrast Top 100 Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

adjusted p:
62% similar

log2FoldChange:
78% similar

expression:
70% similar

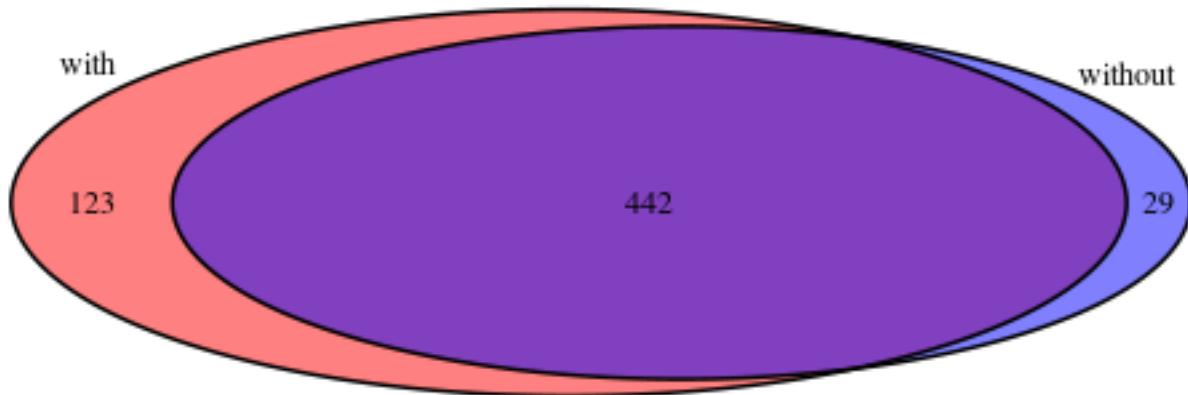


```
## png  
## 2
```

Comparing the “chonky” gene lists ($\text{padj} < 0.01$, $\text{abs(l2fc)} > 1$):

Figure 79 . Similarity of 67d Contrast Chonky Lists Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

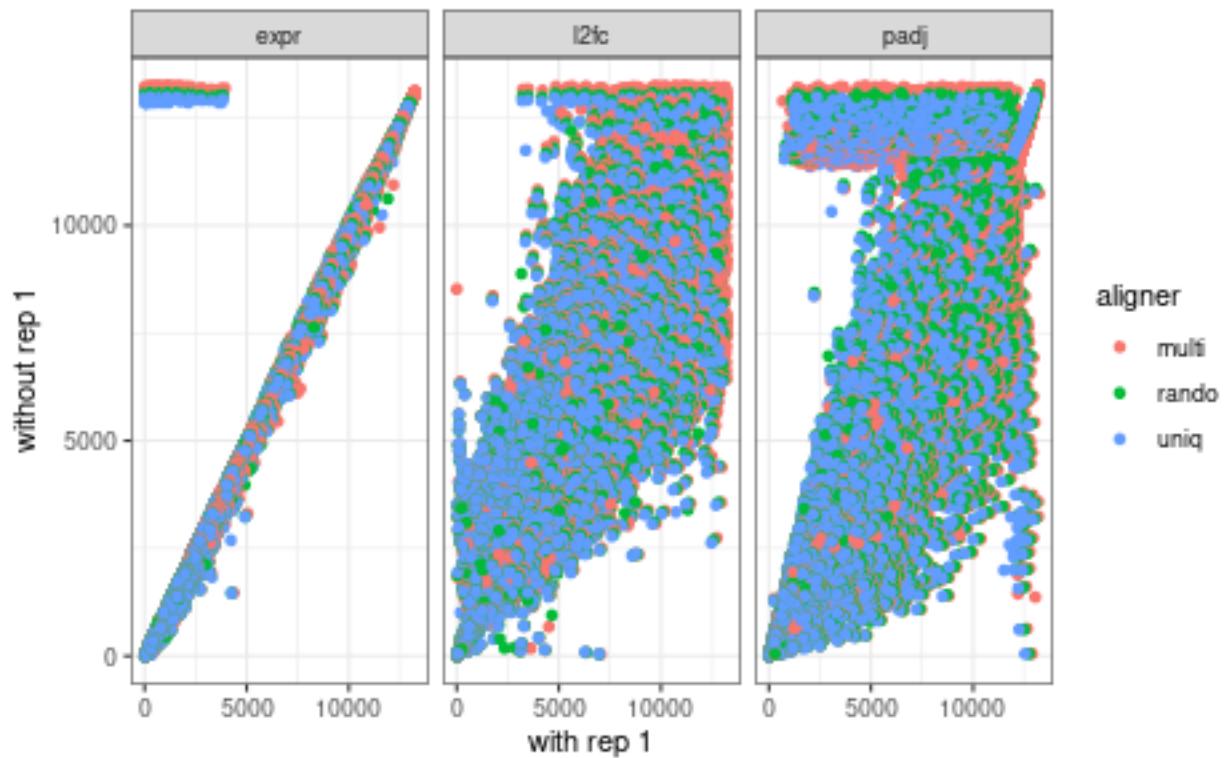
chonky:
74% similar



```
## png  
## 2
```

We can also look at the rank correlations:

Figure 80. Rank correlations of expression, effect size, and significance (67d contrasts, with/without problematic FruLexaFru440 replicate)



```
## png
## 2
```

3.4.9.4 Perturbation to FruLexa/Fru440 Contrast

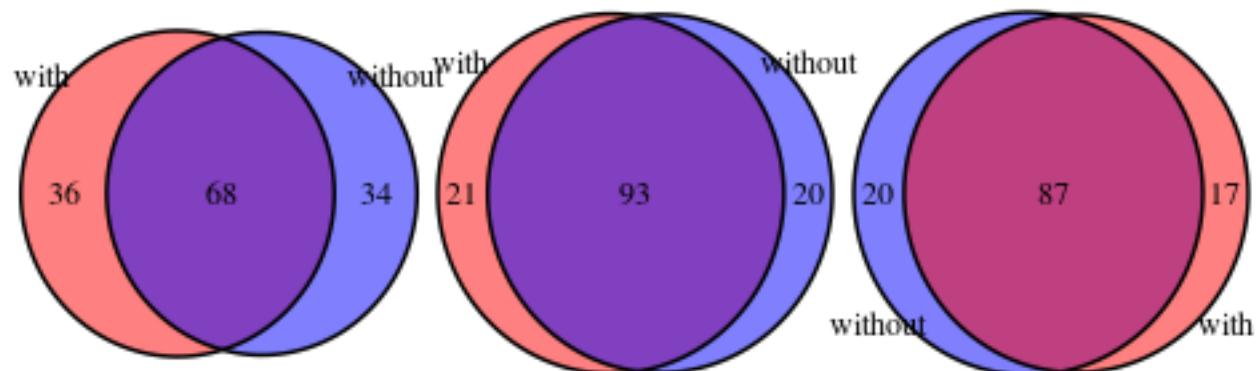
To see how much exclusion of the FruLexa/Fru440 replicate 1 alters the big picture results in the FruLexaFru440 vs. wt contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

Figure 81 . Similarity of FruLexa440 Contrast Top 100 Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

adjusted p:
49% similar

log2FoldChange:
69% similar

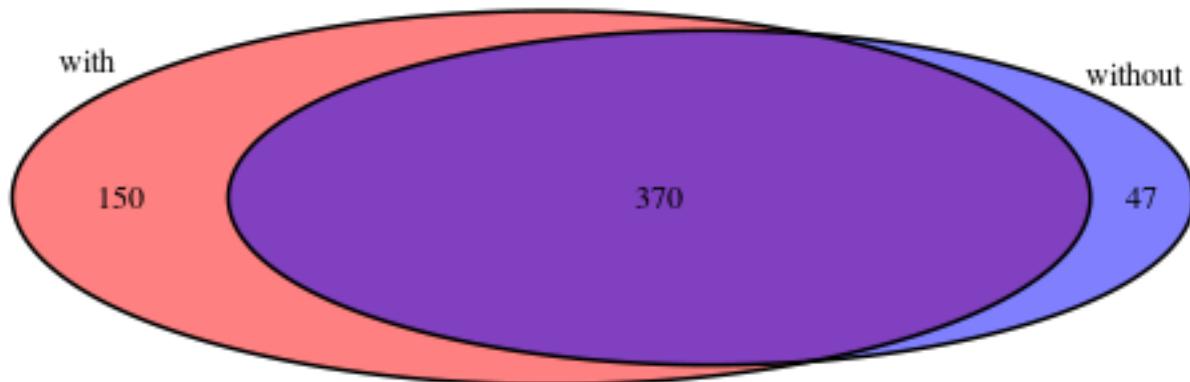
expression:
70% similar



```
## png  
## 2
```

Figure 82 . Similarity of Fru Contrast Chonky Lists Lists,
with/without FruLexaFru440 replicate 1 (pooled alignment strategies)

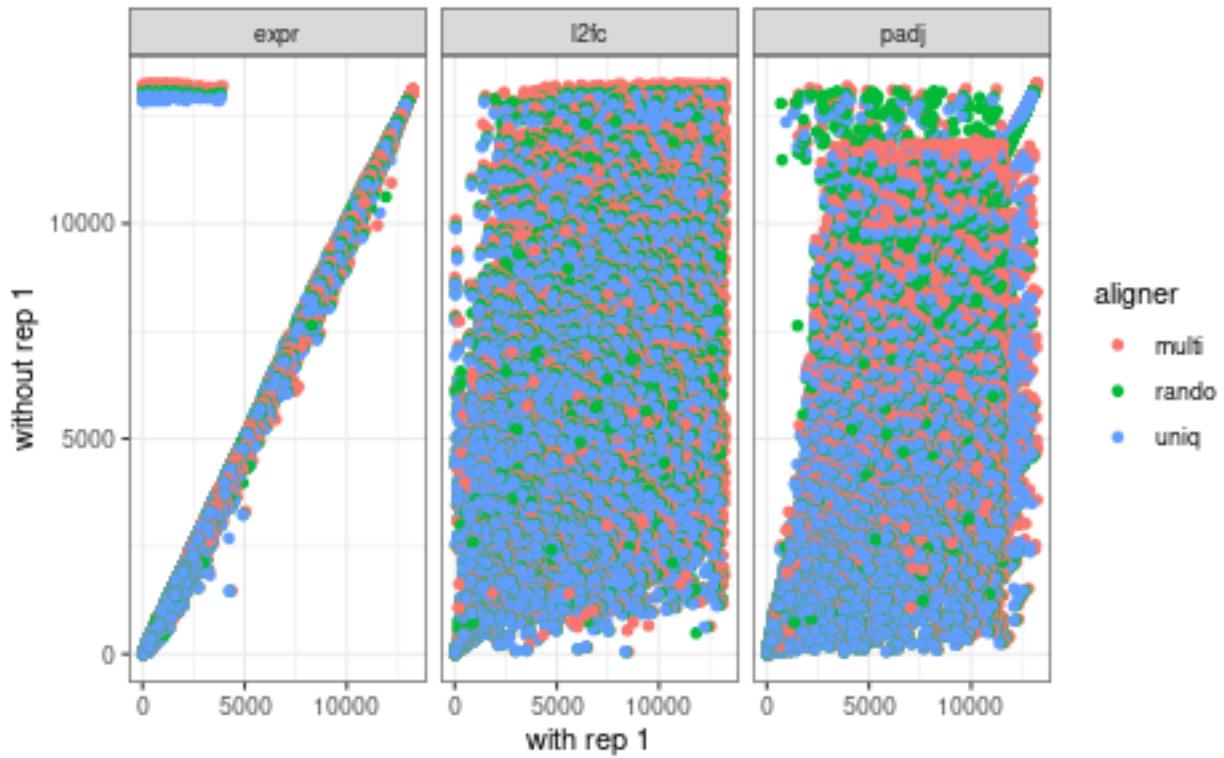
chonky:
65% similar



```
## png  
## 2
```

We can also look at the rank correlations:

Figure 83. Rank correlations of expression, effect size, and significance (67d contrasts, with/without problematic FruLexaFru440 replicate)



```
## png
## 2
```

3.4.9.5 Perturbation to FruLexa/Fru440 Contrast (single-factor)

People have objections to using the 2-factor model so if we consider single-factor models, only the FruLexaFru440 contrast actually matters (since no Fru mutant replicates are included in any other contrast, ever). Intuitively I would expect that more (presumably high-quality) samples included in the model would better buffer it to the effects of dropping replicates. Dropping replicate 1 from hausWtVsMut changes the sample size from 15 to 14, whereas in the single-factor grpWtVsFru changes it from 6 to 5.

The data are found here: “results/tables/supp/grpWtVsFru_smolFru.allAligners.DESeq2.MpBC.reformatted.tsv”

Every gene that can be analyzed using the Reduced counts can be analyzed using the With counts; there are 0 that cannot. On the other hand, there are 0 genes which can be analyzed using the Full counts but not the Reduced

In 0 cases were these genes significant ($\text{padj} < 0.01$) 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 87. Changes in Differential Expression Significance when FruLexa/Fru440 replicate 1 is dropped; single-factor

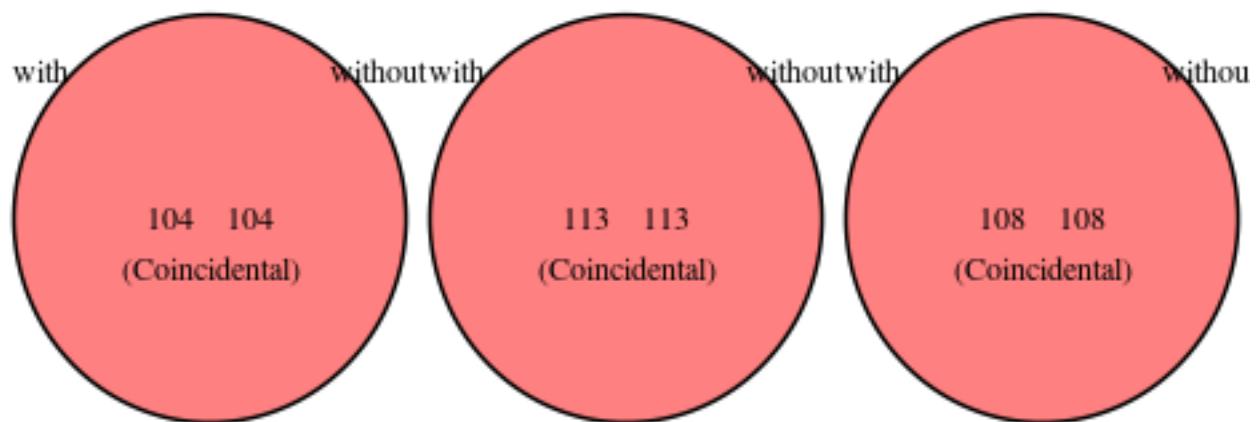
none

FruLexaFru440		
multi	12233	
rando	12103	
uniq	12011	

To see how much exclusion of the FruLexa/Fru440 replicate 1 alters the big picture results in the FruLexaFru440 vs. wt contrast, we can look at how well the top-100 lists agree (similarity is calculated as size of the intersection divided by size of the union; lists are pooled across all aligners and thus may have more than 100 unique elements)

**Figure 84 . Similarity of FruLexa440 Contrast Top 100 Lists,
with/without FruLexaFru440 replicate 1 (single-factor model)**

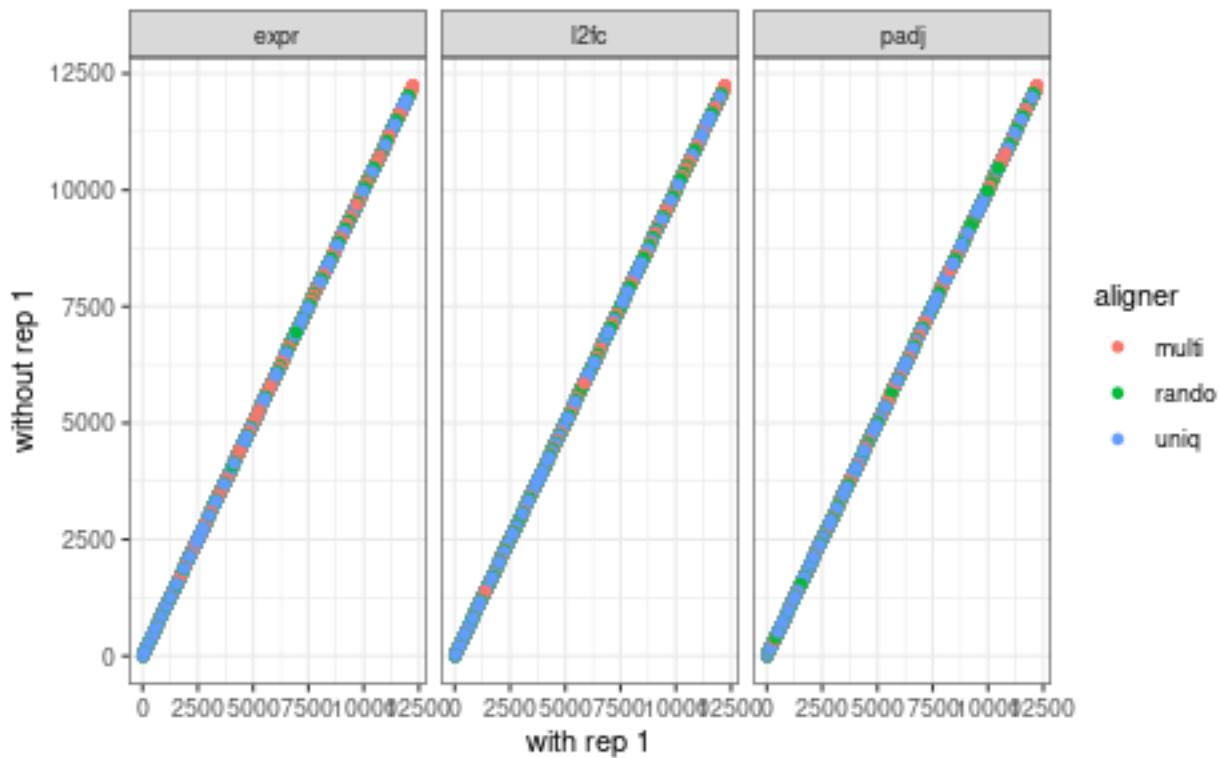
adjusted p:	log2FoldChange:	expression:
100% similar	100% similar	100% similar



```
## png
## 2
```

We can also look at the rank correlations:

Figure 85. Rank correlations of expression, effect size, and significance (single-factor FruLexaFru440 contrast, with/without problematic replicate)



```
## png
## 2
```

3.5 Comparing Expression Changes from Housing with Expression Changes from Genotype

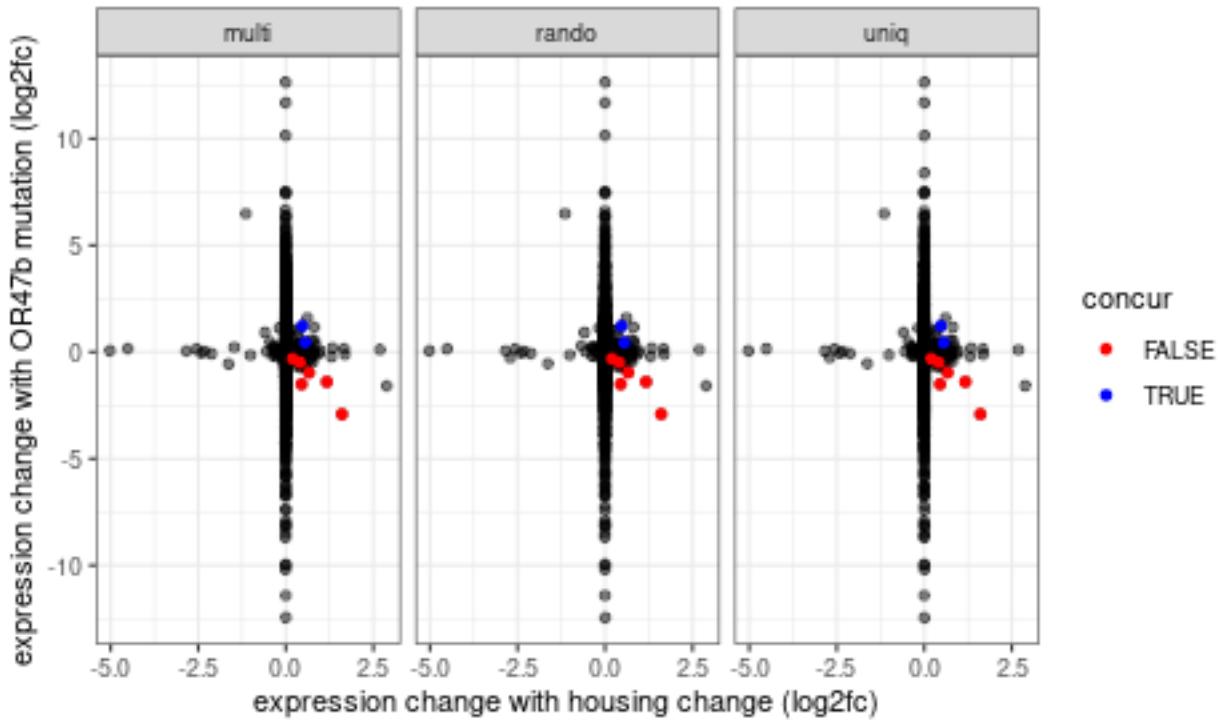
We want to see if the difference in life history creates similar changes in expression as various mutations. This was done using the differential expression data from the genotype & housing simultaneous model. This circumvents the joining step in earlier versions. Earlier versions comparing results from 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 enriched or both depleted) or not (ie, one enriched and the other depleted).

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...)

3.5.1 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 enriched in both cases; the lower left contains genes which are depleted in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

**Figure 86. Scatterplot of Expression Changes in OR47b mutants vs Expression Changes in Housing
(Significant Similarities and Differences Highlighted)**



```
## png
## 2
```

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

Table 88. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs OR47b

	multi	rando	uniq
Agree	2	2	2
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 89. Top Ten Most Significant Genes in difference expression between housing and OR47b

multi						uniq		
rank	name	mean expression	mean readjusted p	housing l2fc	mutation l2fc	name	mean expression	mean readjusted p
1	jv	0.08	3.09×10^{-17}	0.474	1.217	jv	0.08	
2	CG13659	0.31	4.69×10^{-3}	0.556	0.445	CG13659	0.31	

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 90. 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 readusted p	
1	jv	0.845	0.08	3.09×10^{-17}	jv	0.845	0.08	4.1	
2	CG13659	0.501	0.31	4.69×10^{-3}	CG13659	0.500	0.31	4.	

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

Table 91. 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 readusted p	
1	Obp84a	0.62	4.95×10^{-16}	0.451	-1.511	Obp84a	0.62	0.07	
2	CG11852	0.07	4.57×10^{-8}	1.604	-2.916	CG11852	0.07	1.01	
3	amd	1.01	1.35×10^{-5}	1.172	-1.400	amd	1.01	4.18	
4	Fer2LCH	4.18	2.12×10^{-5}	0.200	-0.322	Fer2LCH	4.18	0.24	
5	CG10050	0.24	1.46×10^{-4}	0.664	-0.968	CG10050	0.24	1.57	
6	Or92a	1.57	3.76×10^{-4}	0.407	-0.492	Or92a	1.57	0.62	

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 92. Top Ten Most Serious Significant Differences between in difference expression between housing and OR47b contrants

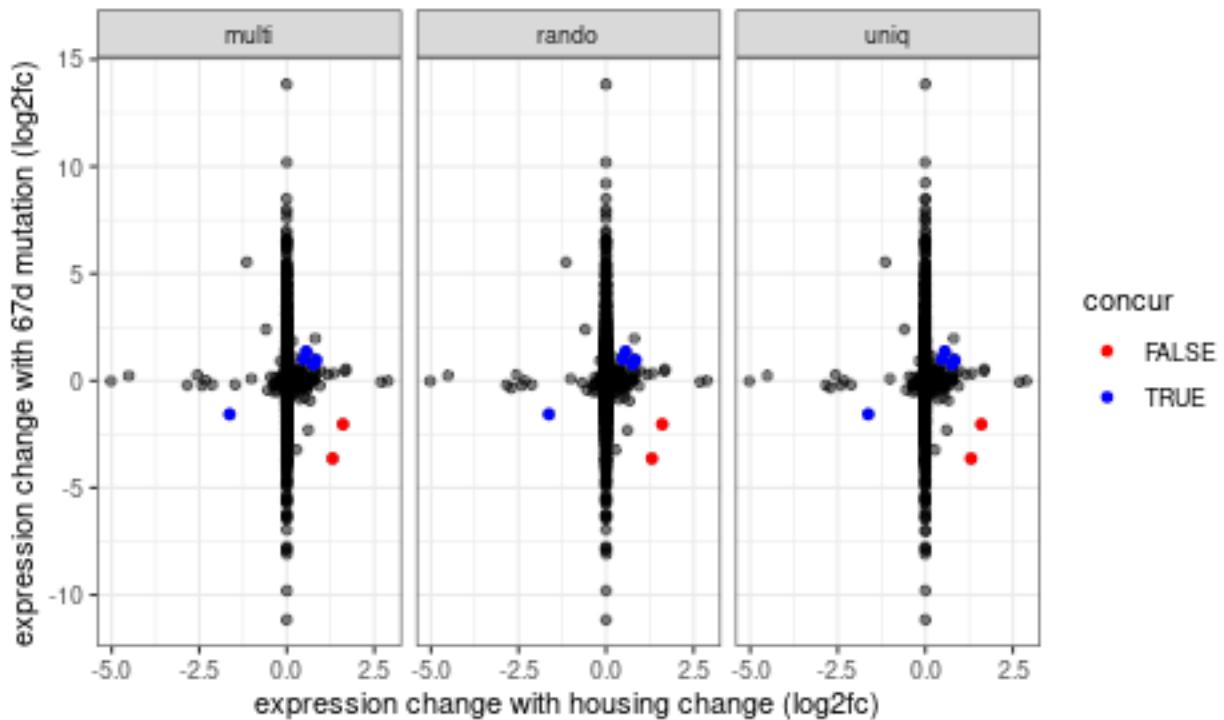
multi					rando		
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression
1	CG11852	4.520	0.07	4.57×10^{-8}	CG11852	4.514	0.07
2	amd	2.572	1.01	1.35×10^{-5}	amd	2.568	1.01
3	Obp84a	1.961	0.62	4.95×10^{-16}	Obp84a	1.958	0.62
4	CG10050	1.632	0.24	1.46×10^{-4}	CG10050	1.624	0.24
5	Or92a	0.900	1.57	3.76×10^{-4}	Or92a	0.896	1.57
6	Fer2LCH	0.522	4.18	2.12×10^{-5}	Fer2LCH	0.518	4.18

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast_and_47bContrast.multi.tsv*, *results/tables/supp/housingContrast_and_47bContrast.rando.tsv*, *results/tables/supp/housingContrast_and_47bContrast.uniq.tsv*

3.5.2 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 enriched in both cases; the lower left contains genes which are depleted in both cases. The other two quadrants contain mismatches between expression patterns. Significant changes are highlighted accordingly.

Figure 87. Scatterplot of Expression Changes in 67d mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## png
## 2
```

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

Table 93. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs 67d

	multi	rando	uniq
Agree	5	5	5
Disagree	2	2	2

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

Table 94. Top Ten Most Significant Genes of Agr
in difference expression between housing and 67d contrants

multi							rando		
rank	name	mean expression	mean readusted p	housing l2fc	67d l2fc	name	mean expression	mean re	
1	CG13659	0.31	1.57×10^{-12}	0.556	1.368	CG13659	0.31	1.8	
2	jv	0.08	5.99×10^{-12}	0.474	1.004	jv	0.08	7.7	
3	CG31288	1.36	1.08×10^{-7}	0.843	0.966	CG31288	1.36	1	
4	CG31272	0.11	2.31×10^{-5}	0.769	0.761	CG31272	0.11	2	
5	hgo	0.04	2.66×10^{-3}	-1.630	-1.567	hgo	0.04	2	

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 95. Top Ten Largest Magnitude Changes In Significant Gen
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 re
1	hgo	-1.599	0.04	2.66×10^{-3}	hgo	-1.598	0.04	2.
2	CG13659	0.962	0.31	1.57×10^{-12}	CG13659	0.962	0.31	1.8
3	CG31288	0.904	1.36	1.08×10^{-7}	CG31288	0.904	1.36	1.
4	CG31272	0.765	0.11	2.31×10^{-5}	CG31272	0.764	0.11	2.
5	jv	0.739	0.08	5.99×10^{-12}	jv	0.738	0.08	7.7

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

Table 96. Top Ten Most Significant Genes of Disa
in difference expression between housing and 67d contrasts

multi							rando		
rank	name	mean expression	mean readusted p	housing l2fc	67d l2fc	name	mean expression	mean re	
1	MtnB	0.39	2.92×10^{-28}	1.309	-3.632	MtnB	0.39	3.8	
2	CG11852	0.07	9.62×10^{-6}	1.604	-2.035	CG11852	0.07	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 97. Top Ten Most Serious Significant Differences betw
in difference expression between housing and 67d contrants

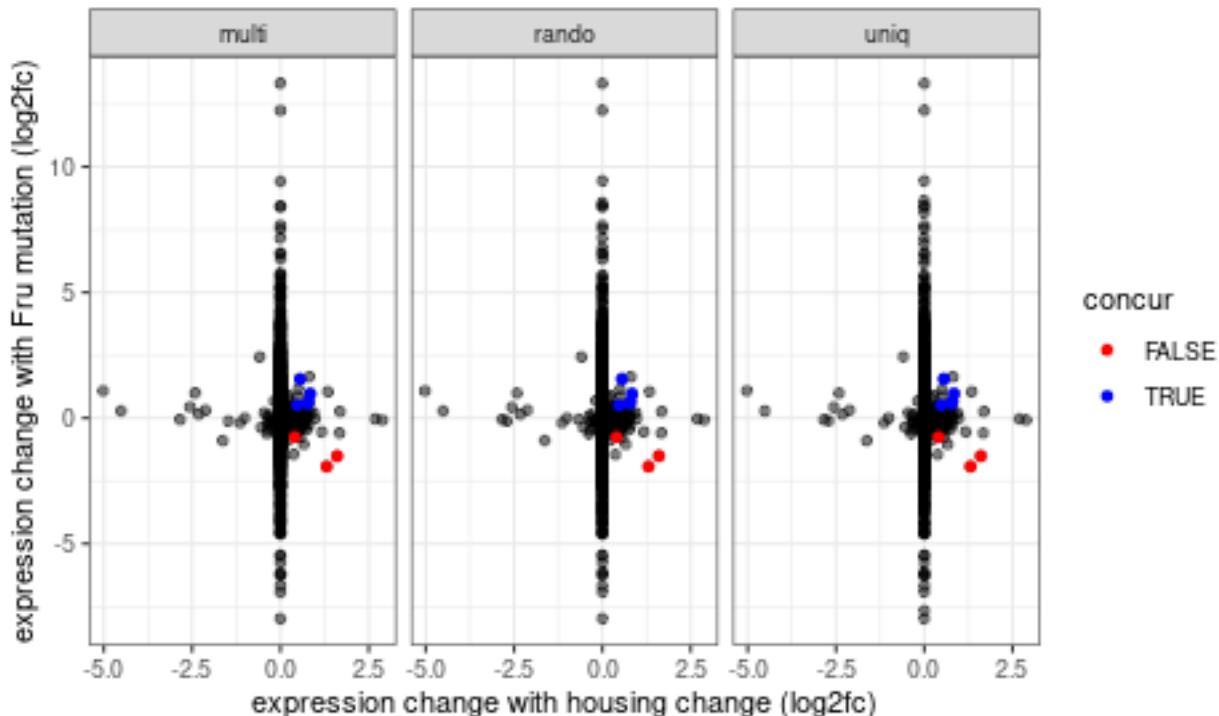
multi					rando			
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression	
1	MtnB	4.941	0.39	2.92×10^{-28}	MtnB	4.940	0.39	
2	CG11852	3.639	0.07	9.62×10^{-6}	CG11852	3.637	0.07	

The full joined comparisons can be found in the tables folder: *results/tables/supp/housingContrast_and_7dContrast.multi.tsv*, *results/tables/supp/housingContrast_and_7dContrast.rando.tsv*, *results/tables/supp/housingContrast_and_7dContrast.uniq.tsv*

3.5.3 Housing & FruLexFru440

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

Figure 88. Scatterplot of Expression Changes in Fru mutants vs Expression Changes in Housing (Significant Similarities and Differences Highlighted)



```
## png
## 2
```

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

Table 98. Number of Genes with Significant Changes in Both Contrasts, by Shared Direction of Change
change in housing vs Fru

	multi	rando	uniq
Agree	5	5	5
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 99. Top Ten Most Significant Genes of Agr in difference expression between housing and Fru contrants

multi					rando			
rank	name	mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re
1	CG13659	0.31	1.86×10^{-15}	0.556	1.552	CG13659	0.31	2.1
2	CG31288	1.36	1.28×10^{-7}	0.843	0.965	CG31288	1.36	1.
3	jv	0.08	2.05×10^{-4}	0.474	0.513	jv	0.08	2.
4	CG31272	0.11	3.93×10^{-4}	0.769	0.569	CG31272	0.11	4.
5	CG42806	0.64	2.95×10^{-3}	0.801	0.647	CG42806	0.64	3.

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 100. 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 re
1	CG13659	1.054	0.31	1.86×10^{-15}	CG13659	1.054	0.31	2.1
2	CG31288	0.904	1.36	1.28×10^{-7}	CG31288	0.904	1.36	1.
3	CG42806	0.724	0.64	2.95×10^{-3}	CG42806	0.723	0.64	3.
4	CG31272	0.669	0.11	3.93×10^{-4}	CG31272	0.669	0.11	4.
5	jv	0.493	0.08	2.05×10^{-4}	jv	0.493	0.08	2.

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

Table 101. Top Ten Most Significant Genes of Dis in difference expression between housing and Fru contrants

multi					rando			
rank	name	mean expression	mean readusted p	housing l2fc	Fru l2fc	name	mean expression	mean re
1	MtnB	0.39	1.53×10^{-16}	1.309	-1.930	MtnB	0.39	2.0
2	Or92a	1.57	8.16×10^{-7}	0.407	-0.756	Or92a	1.57	1.
3	CG11852	0.07	1.57×10^{-4}	1.604	-1.514	CG11852	0.07	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 102. Top Ten Most Serious Significant Differences betw in difference expression between housing and Fru contrasts

multi					rando			
rank	name	l2fc difference	mean expression	mean readusted p	name	l2fc difference	mean expression	mean re
1	MtnB	3.239	0.39	1.53×10^{-16}	MtnB	3.236	0.39	0.39

2	CG11852	3.118	0.07	1.57×10^{-4}	CG11852	3.113	0.07
3	Or92a	1.163	1.57	8.16×10^{-7}	Or92a	1.160	1.57

Full data are in the tables folder:

results/tables/supp/housingContrast_and_FruContrast.multi.tsv *results/tables/supp/housingContrast_and_FruContrast.results*
results/tables/supp/housingContrast_and_FruContrast.uniq.tsv

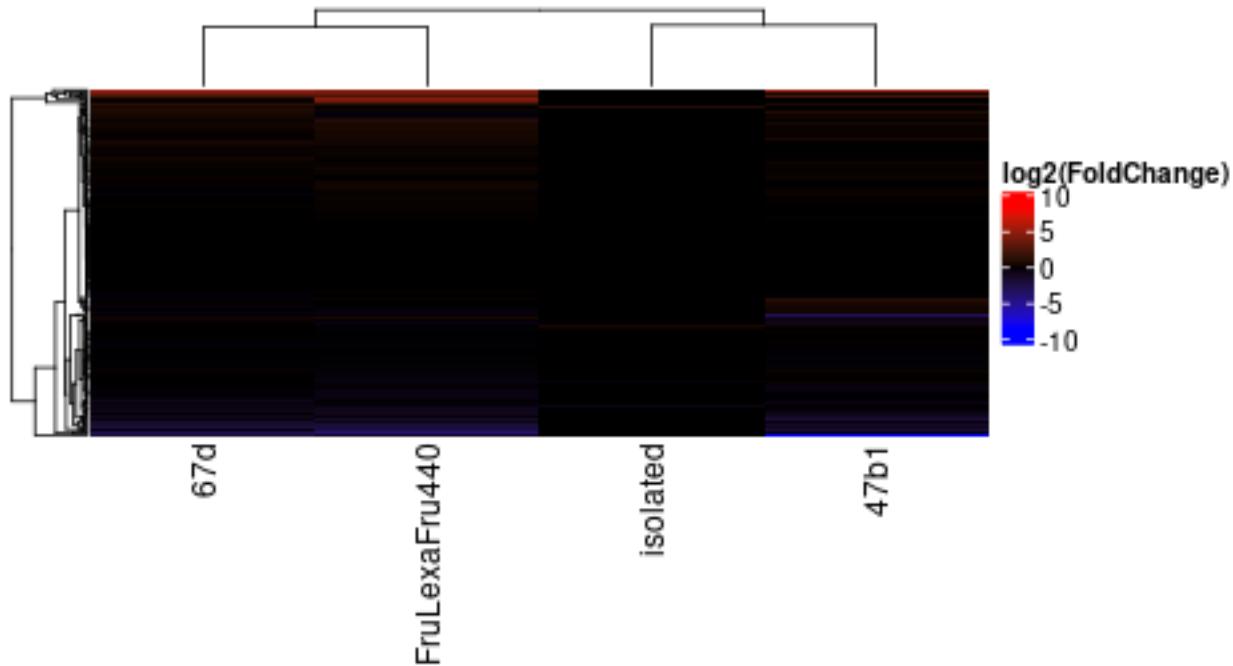
3.5.4 Overview (Heatmaps)

We can also display changes in gene expression as a heatmap. Increases in expression are show in red, and decreases in blue. Significance of change is not currently indicated.

3.5.4.1 Ion Channel Activity

Here is a heat map specific to the Ion Channel Activity genes

**Figure 89 . Heatmap Displaying Difference in Expression
in Different Experimental Contrasts
(Ion Channel Activity Genes)(multi alignment)**



```
## png
## 2
```

3.6 Comparing Expression Changes Between Mutants

do this

3.6.1 Fru & 67d

do this

3.6.2 Fru & 47b

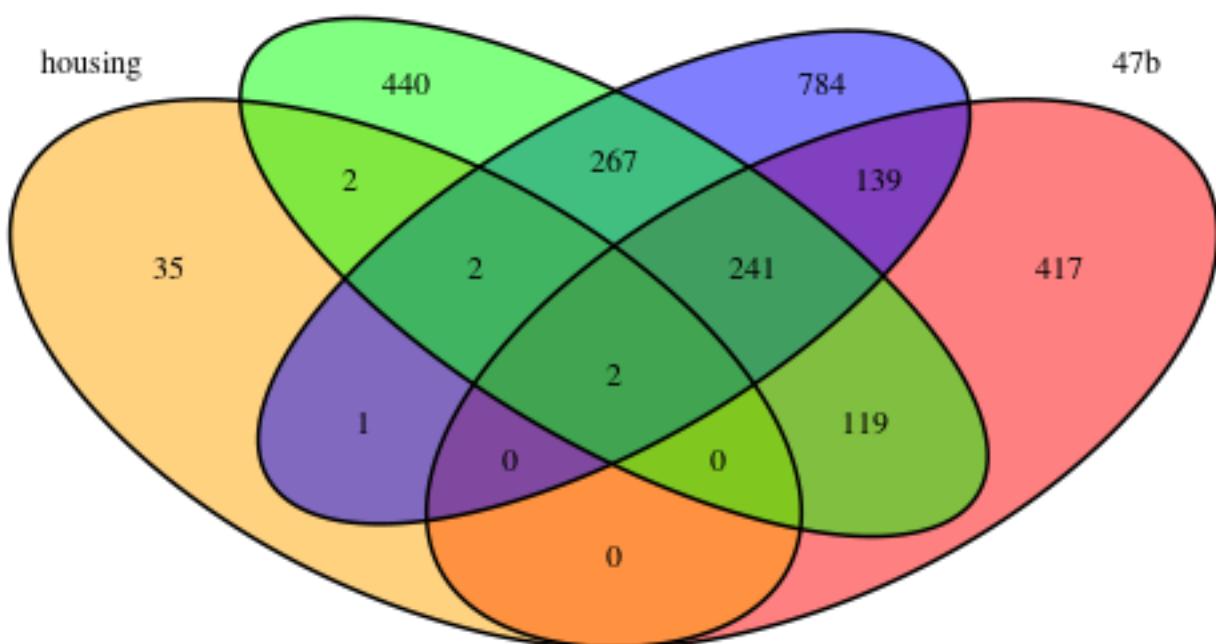
do this

3.6.3 47b & 67d

do this

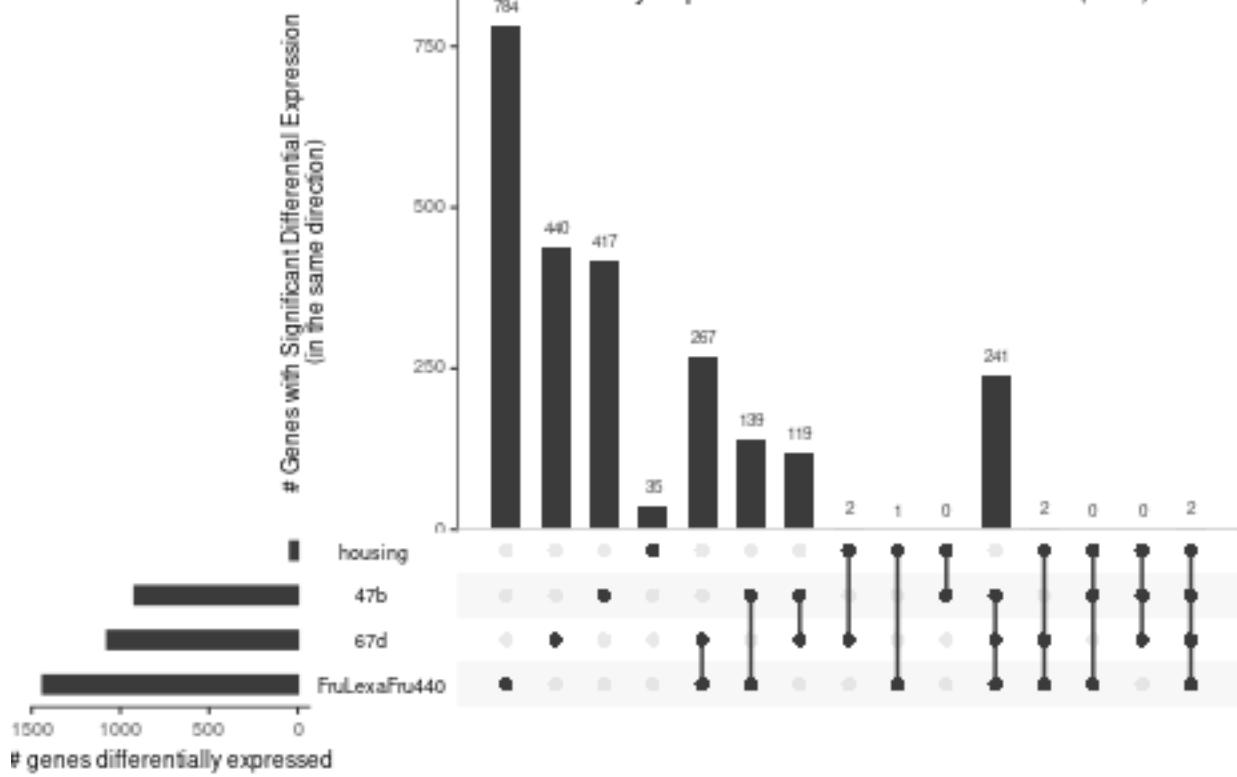
3.7 Mutually Significant Differential Expression Overview

Figure 90 . Venn Diagram: # genes with shared significant change,
 by experimental contrast intersection (multi)
 67d FruLexaFru440



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

Figure 91 . UpSet plot: # genes with shared significant change by experimental contrast intersection (multi)



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

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

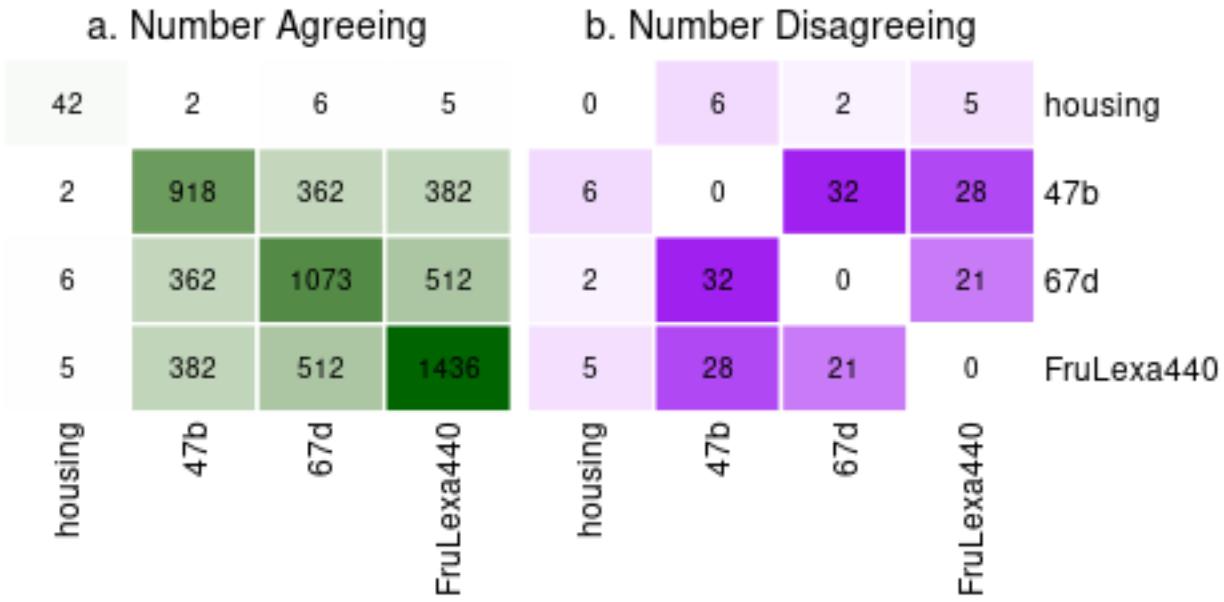
The two genes with the same behavior across all experimental contrasts are javelin and CG13659. Both are enriched in all cases:

Table 103. Genes sharing significant differential expression in all four contrasts

	47b vs wt		67d vs wt		FruLexaFru440 vs wt		iso
	log2FoldChange	padj	log2FoldChange	padj	log2FoldChange	padj	log2FoldC
multi							
CG13659	0.445	6.91×10^{-3}	1.368	1.93×10^{-22}	1.552	2.72×10^{-28}	
jv	1.217	1.49×10^{-30}	1.004	1.40×10^{-20}	0.513	1.64×10^{-5}	
rando							
CG13659	0.445	7.24×10^{-3}	1.368	2.54×10^{-22}	1.554	3.34×10^{-28}	
jv	1.217	2.31×10^{-30}	1.004	2.09×10^{-20}	0.514	1.70×10^{-5}	
uniq							
CG13659	0.444	7.35×10^{-3}	1.369	1.74×10^{-22}	1.554	2.05×10^{-28}	
jv	1.217	1.76×10^{-30}	1.004	1.55×10^{-20}	0.514	1.57×10^{-5}	

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

Figure 92 Heatmap of Pairwise Comparisons between Contrasts:
significant genes with the same (left)
or different (right) directions of change
(2-factor models)



```
## png
## 2
```

3.8 Focus on Fruitless

Table 104a. Differential Expression of Fruitless
(single factor)

	significance (p)	effect size (l2fc)
<hr/>		
housing		
multi	0.89	0.01
rando	0.90	0.01
uniq	0.89	0.01
<hr/>		
47b		
multi	0.91	0.03
rando	0.91	0.03
uniq	0.91	0.03
<hr/>		
67d		
multi	0.74	0.09
rando	0.74	0.09
uniq	0.75	0.09

Fru		
multi	0.66	0.12
rando	0.66	0.12
uniq	0.66	0.12

Table 104b. Differential Expression of Fruitless
(multifactor)

	effect size (l2fc)	significance (p)
47b1		
multi	0.02	0.99
rando	0.03	0.99
uniq	0.02	0.99
67d		
multi	0.05	0.91
rando	0.05	0.91
uniq	0.06	0.91
FruLexaFru440		
multi	0.27	0.35
rando	0.27	0.35
uniq	0.26	0.35
isolated		
multi	0.00	1.00
rando	0.00	1.00
uniq	0.00	1.00

Changes in splicing of Fruitless are of special interest, and feature counting/differential expression testing was performed on an annotation which considers all available exons separately. In this way, changes in exon use by treatment might be detected.

3.8.1 By Exon

3.8.1.1 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 105. Number of Fruitless Exons Available For Analysis
('none' counting, by aligner)

aligner	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 18 20, and 22, in the FruLexa/Fru400 contrast:

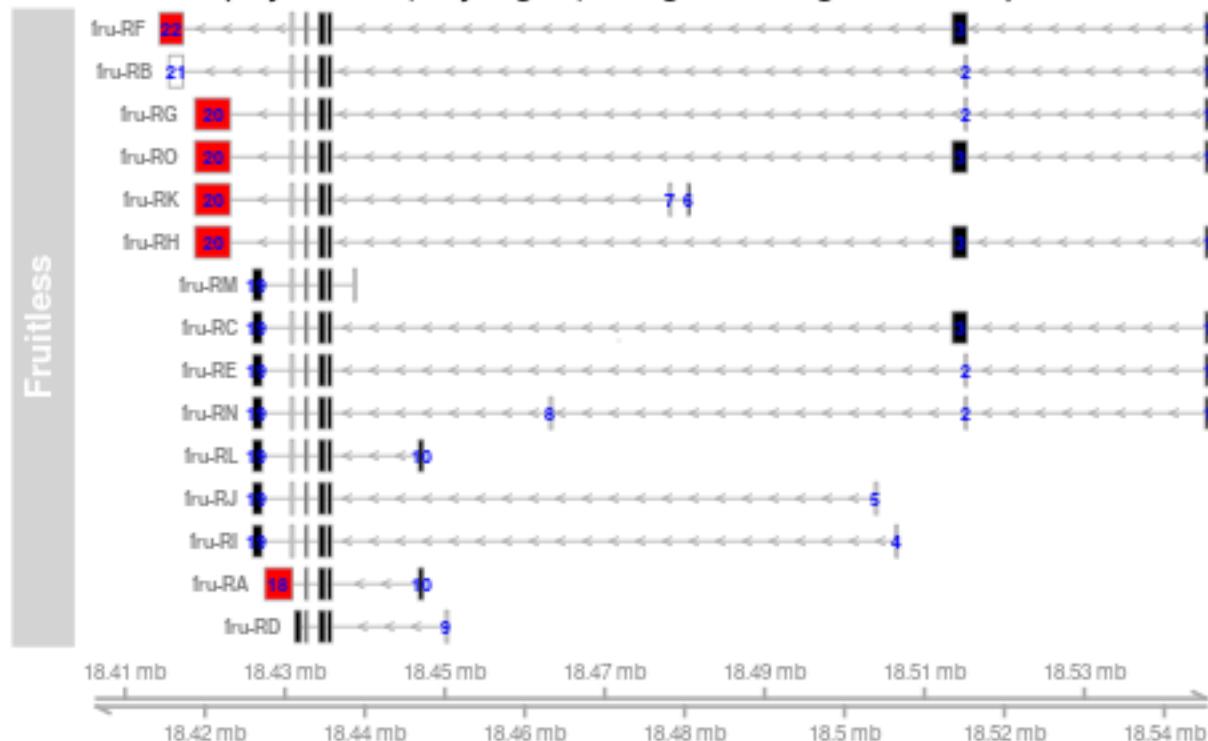
Table 106. Differential Use of Fruitless Exons, by Contrast
('none' counting, multi only)

	47b		67d		Fru		wt
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange
exon_1	-0.0000	4.61×10^{-1}	-0.0000	8.83×10^{-1}	-0.0979	9.05×10^{-1}	0.0
exon_10	-0.0000	9.47×10^{-1}	0.0000	8.83×10^{-1}	0.6673	7.98×10^{-2}	0.0
exon_12	-0.0000	9.47×10^{-1}	-0.0000	8.83×10^{-1}	-0.0565	9.05×10^{-1}	0.0
exon_13	0.0000	7.20×10^{-1}	-0.0000	8.83×10^{-1}	-0.0769	9.05×10^{-1}	0.0
exon_14	-0.0000	9.91×10^{-1}	0.0000	8.83×10^{-1}	0.0426	9.05×10^{-1}	-0.0
exon_15	0.0000	9.47×10^{-1}	-0.0000	8.83×10^{-1}	-0.0483	9.05×10^{-1}	0.0
exon_16	-0.0000	9.47×10^{-1}	-0.0000	8.83×10^{-1}	0.0330	9.05×10^{-1}	-0.0
exon_18	0.0000	4.61×10^{-1}	0.0000	4.36×10^{-1}	2.0286	2.03×10^{-3}	-0.0
exon_19	-0.0000	9.47×10^{-1}	-0.0000	8.83×10^{-1}	-0.0666	9.05×10^{-1}	0.0
exon_20	0.0000	9.47×10^{-1}	0.0000	8.83×10^{-1}	0.9622	3.16×10^{-2}	-0.0
exon_22	0.0000	9.47×10^{-1}	0.0000	8.83×10^{-1}	1.4713	2.53×10^{-2}	0.0
exon_3	-0.0000	9.47×10^{-1}	-0.0000	9.98×10^{-1}	0.1085	9.05×10^{-1}	0.0
exon_6	-0.0000	4.61×10^{-1}	-0.0000	8.83×10^{-1}	1.0387	7.98×10^{-2}	-0.0

Table 107. Fru exons with significantly ($p_{adj} < 0.05$) differential use
('none' counting)

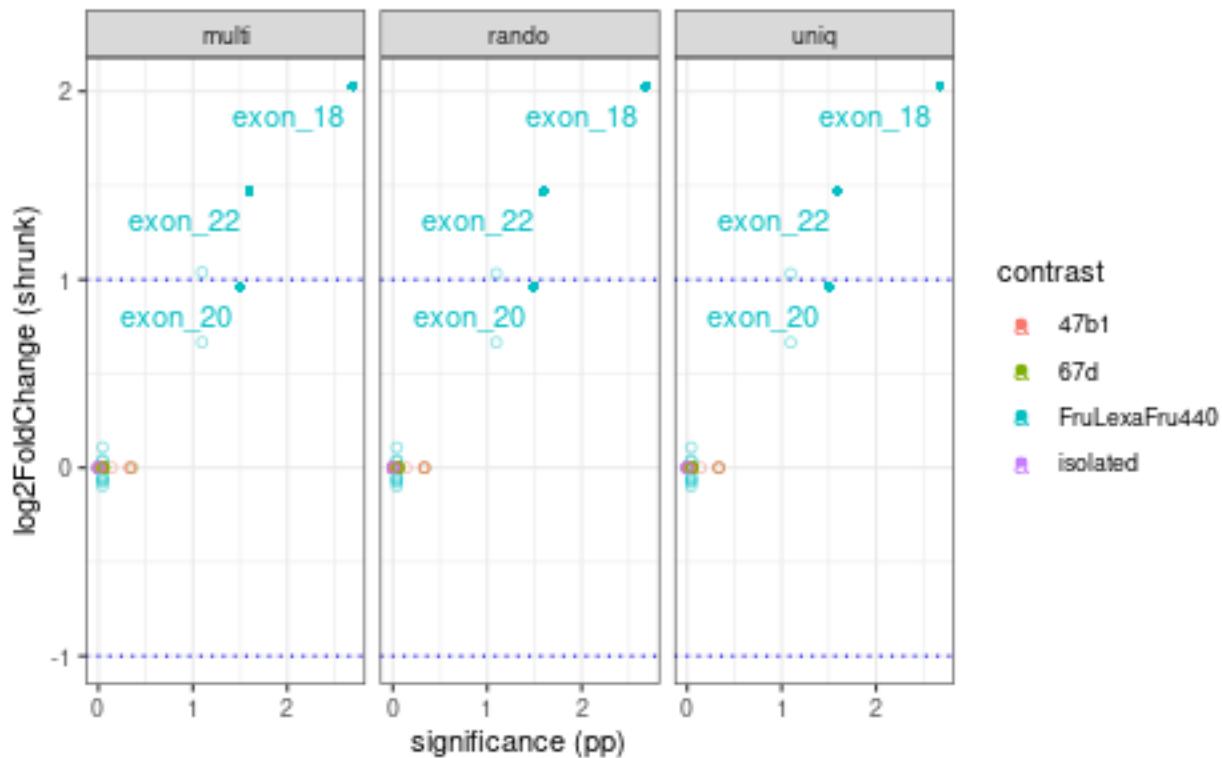
	Fru		
	<th>log2FoldChange</th> <th>adjusted p</th>	log2FoldChange	adjusted p
exon_18			
	multi	2.03	0.002
	rando	2.03	0.002
	uniq	2.03	0.002
exon_20			
	multi	0.96	0.032
	rando	0.96	0.032
	uniq	0.96	0.032
exon_22			
	multi	1.47	0.025
	rando	1.47	0.026
	uniq	1.47	0.026

**Figure 93. Fruitless gene model: exons with any significant change detected highlighted
(any contrast, any aligner, ambiguous assigned to none)**



```
## png
## 2
```

**Figure 94. Volcano Plot: Fold Change vs. Significance
(fruitless exons, 'none' counting strategy)**



```
## png
## 2
```

3.8.1.2 Ambiguous Read Assignment: All

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

**Table 108. 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

only the FruLexa/Fru440 contrast had significantly different exon use:

**Table 109. Differential Use of Fru Exons, by Contrast
('all' counting, multi only)**

	47b		67d		Fru		wt
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	
exon_1	-0.00	7.78×10^{-1}	-0.00	8.88×10^{-1}	-0.32	4.15×10^{-1}	0
exon_10	-0.00	9.67×10^{-1}	0.00	8.88×10^{-1}	0.59	7.13×10^{-2}	0
exon_12	-0.00	9.67×10^{-1}	-0.00	9.98×10^{-1}	0.02	9.91×10^{-1}	0

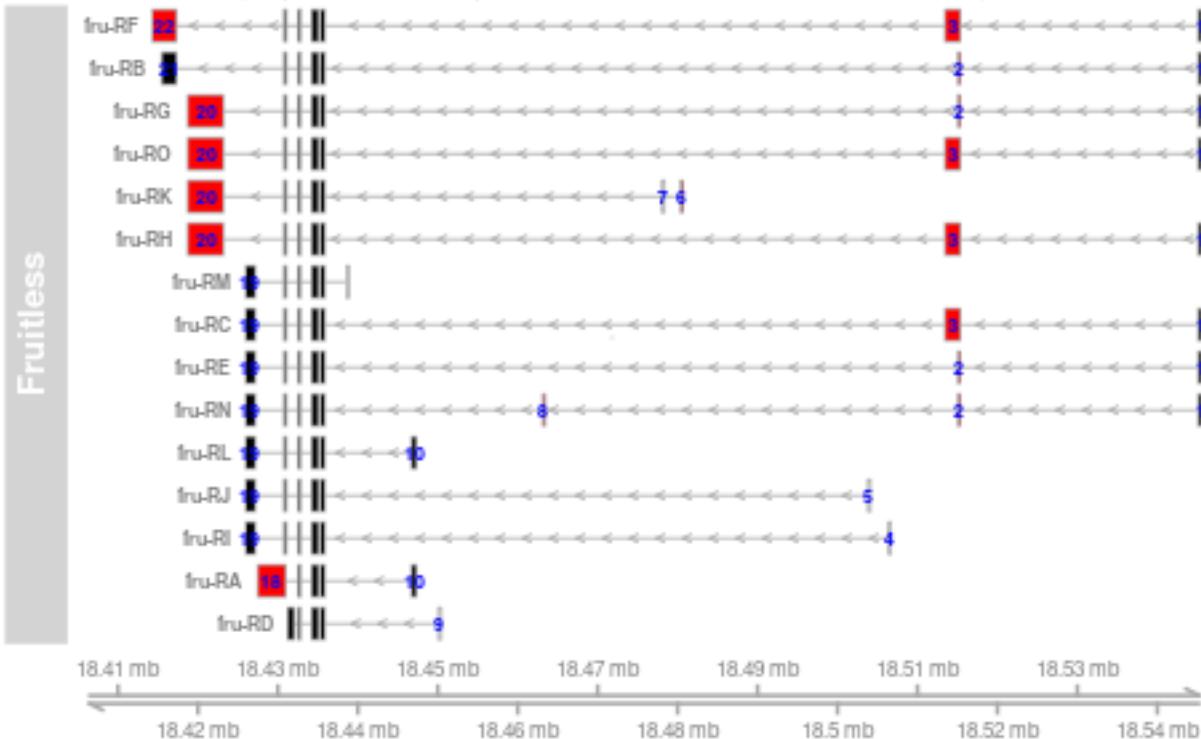
exon_13	0.00	7.78×10^{-1}	-0.00	8.88×10^{-1}	-0.16	7.04×10^{-1}	-0.00
exon_14	-0.00	9.67×10^{-1}	-0.00	8.88×10^{-1}	-0.03	9.39×10^{-1}	-0.00
exon_15	0.00	7.95×10^{-1}	-0.00	8.88×10^{-1}	-0.13	7.98×10^{-1}	0.00
exon_16	0.00	9.67×10^{-1}	-0.00	9.83×10^{-1}	0.11	9.00×10^{-1}	-0.00
exon_17	0.00	6.79×10^{-1}	-0.00	8.88×10^{-1}	-0.14	8.31×10^{-1}	0.00
exon_18	0.00	6.79×10^{-1}	0.00	8.88×10^{-1}	1.41	7.35×10^{-3}	0.00
exon_19	-0.00	9.67×10^{-1}	-0.00	8.88×10^{-1}	-0.00	9.91×10^{-1}	0.00
exon_2	-0.00	7.78×10^{-1}	-0.00	9.83×10^{-1}	-3.13	1.06×10^{-4}	0.00
exon_20	0.00	9.67×10^{-1}	0.00	8.88×10^{-1}	1.11	1.35×10^{-2}	-0.00
exon_21	-0.00	9.67×10^{-1}	0.00	9.83×10^{-1}	0.50	1.37×10^{-1}	-0.00
exon_22	-0.00	9.94×10^{-1}	0.00	9.83×10^{-1}	0.87	4.60×10^{-2}	0.00
exon_3	-0.00	7.78×10^{-1}	-0.00	9.83×10^{-1}	-1.85	1.92×10^{-4}	0.00
exon_6	-0.00	6.79×10^{-1}	-0.00	9.83×10^{-1}	1.48	4.60×10^{-2}	-0.00
exon_8	-0.00	7.78×10^{-1}	-0.00	8.88×10^{-1}	2.28	2.96×10^{-2}	0.00

Table 110. Fru exons with significantly ($\text{padj} < 0.05$) different use ('all' counting, by aligner)

aligner	Fru		
		log2FoldChange	adjusted p
exon_18			
multi		1.41	7.35×10^{-3}
rando		1.41	7.37×10^{-3}
uniq		1.41	7.37×10^{-3}
exon_2			
multi		-3.13	1.06×10^{-4}
rando		-3.13	1.06×10^{-4}
uniq		-3.13	1.06×10^{-4}
exon_20			
multi		1.11	1.35×10^{-2}
rando		1.11	1.35×10^{-2}
uniq		1.11	1.35×10^{-2}
exon_22			
multi		0.87	4.60×10^{-2}
rando		0.87	4.59×10^{-2}
uniq		0.87	4.59×10^{-2}
exon_3			
multi		-1.85	1.92×10^{-4}
rando		-1.85	1.92×10^{-4}
uniq		-1.85	1.92×10^{-4}
exon_6			
multi		1.48	4.60×10^{-2}
rando		1.48	4.59×10^{-2}
uniq		1.48	4.59×10^{-2}
exon_8			
multi		2.28	2.96×10^{-2}
rando		2.28	2.96×10^{-2}

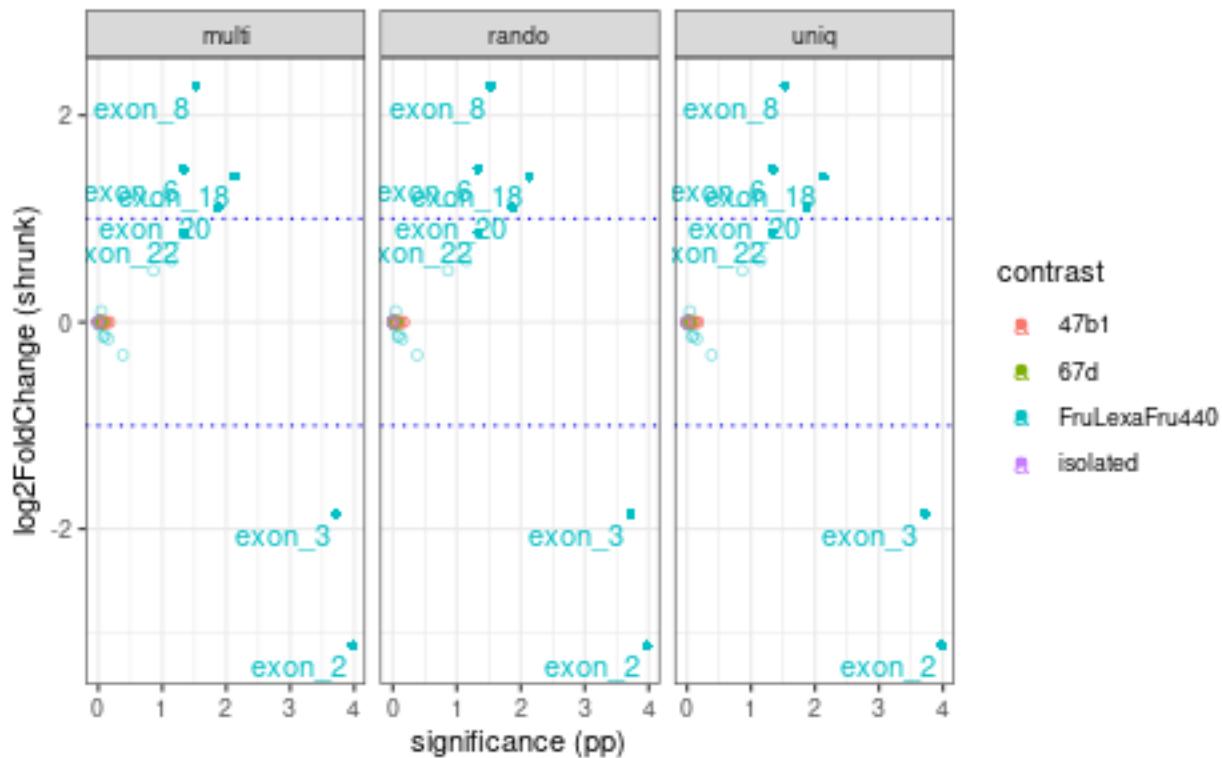
uniq 2.28 2.96×10^{-2}

Figure 95. Fruitless gene model: exons with any significant change detected highlighted (any contrast, any aligner, ambiguous assigned to all)



```
## png  
## 2
```

Figure 96. Volcano Plot: Fold Change vs. Significance
(fruitless exons, 'all' counting strategy, $p_{adj} < 0.05$)



```
## png
## 2
```

3.8.2 By Exon Junction

When the *_SplicedOnly alignments were counted (“all” strategy) against the fru_junct annotation:

Table 111. 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 112. Differential Exon Use in Fruitless, by Contrast
Junction-based, 'all' counting (Multi only)

	47b		67d		Fru		wt
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	
exon_1	-0.01	8.07×10^{-1}					
exon_10	-0.04	4.60×10^{-1}					
exon_12	-0.06	4.60×10^{-1}					

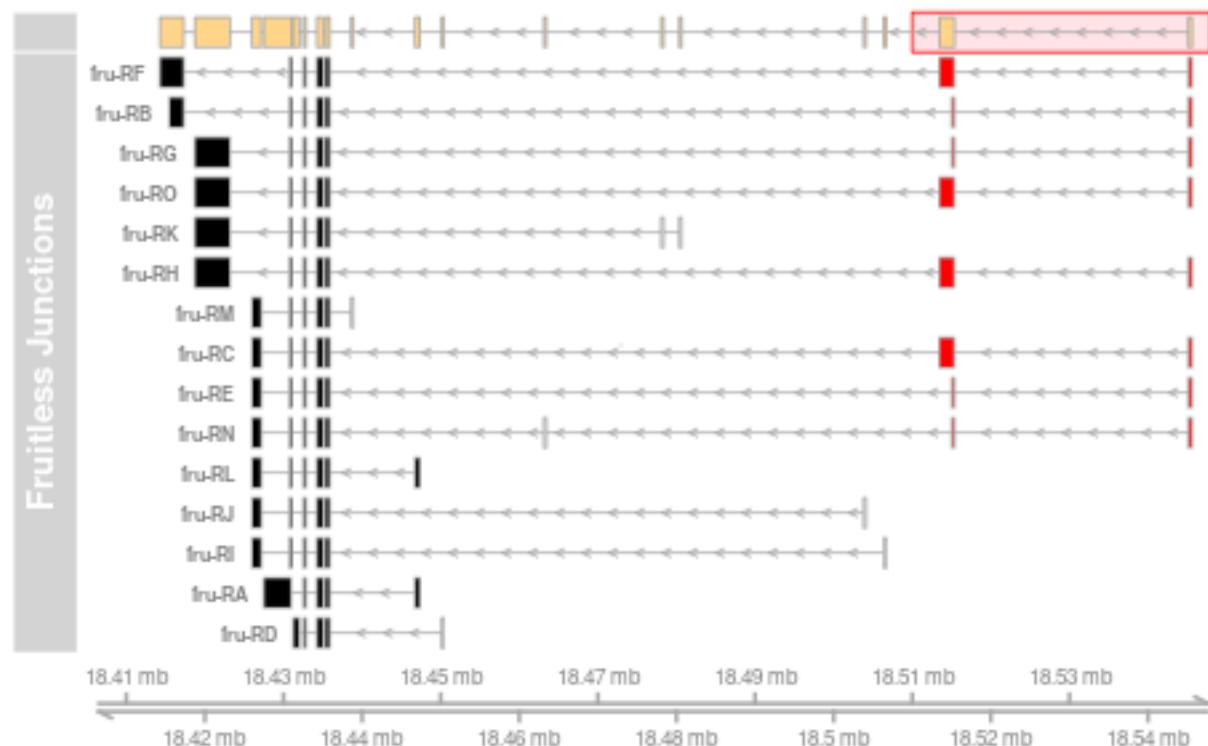
exon_13	0.01	8.30×10^{-1}	0.00	9.93×10^{-1}	0.00	9.80×10^{-1}	0
exon_14	-0.02	8.07×10^{-1}	-0.00	9.93×10^{-1}	-0.00	7.24×10^{-1}	-0
exon_15	0.02	8.07×10^{-1}	0.00	9.93×10^{-1}	0.00	9.80×10^{-1}	-0
exon_16	0.01	4.60×10^{-1}	0.00	7.10×10^{-1}	0.00	3.97×10^{-1}	0
exon_17	0.02	8.07×10^{-1}	-0.00	9.93×10^{-1}	-0.00	9.80×10^{-1}	-0
exon_18	0.06	1.20×10^{-1}	0.00	8.69×10^{-1}	0.00	7.24×10^{-1}	0
exon_19	-0.01	7.73×10^{-1}	-0.00	9.93×10^{-1}	-0.00	9.80×10^{-1}	-0
exon_2	-0.04	4.60×10^{-1}	0.00	9.93×10^{-1}	-2.94	2.17×10^{-3}	-0
exon_20	0.01	8.07×10^{-1}	0.00	9.92×10^{-1}	0.00	9.80×10^{-1}	0
exon_21	-0.00	8.07×10^{-1}	-0.00	7.10×10^{-1}	-0.00	8.75×10^{-1}	0
exon_22	-0.00	8.07×10^{-1}	-0.00	7.10×10^{-1}	-0.00	8.75×10^{-1}	0
exon_3	-0.01	8.07×10^{-1}	0.00	9.93×10^{-1}	-0.00	3.46×10^{-2}	0

Exons 1,2, and 3, the most 5' of exons, are less used in the FruLexa/Fru440 contrast; however, exons 1 and 3 bizarrely low effect sizes given their significance:

Table 113. Fru exons with significantly ($\text{padj} < 0.05$) different use

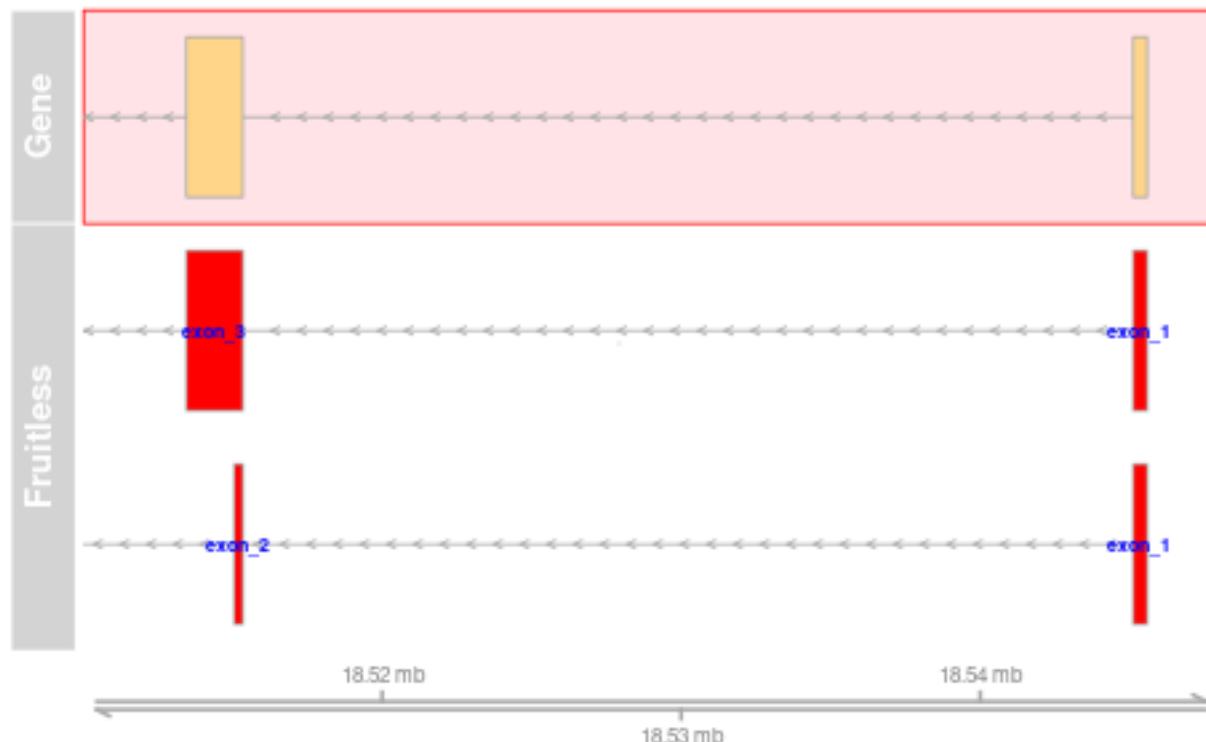
Junction-based, 'all' counting			
	Fru		
	<th>log2FoldChange</th> <th>adjusted p</th>	log2FoldChange	adjusted p
exon_1			
multi		-0.00	0.035
rando		-0.00	0.035
uniq		-0.00	0.035
exon_2			
multi		-2.94	0.002
rando		-2.94	0.002
uniq		-2.94	0.002
exon_3			
multi		-0.00	0.035
rando		-0.00	0.035
uniq		-0.00	0.035

Figure 97 Fruitless exons with significant change in use (measured by junction)



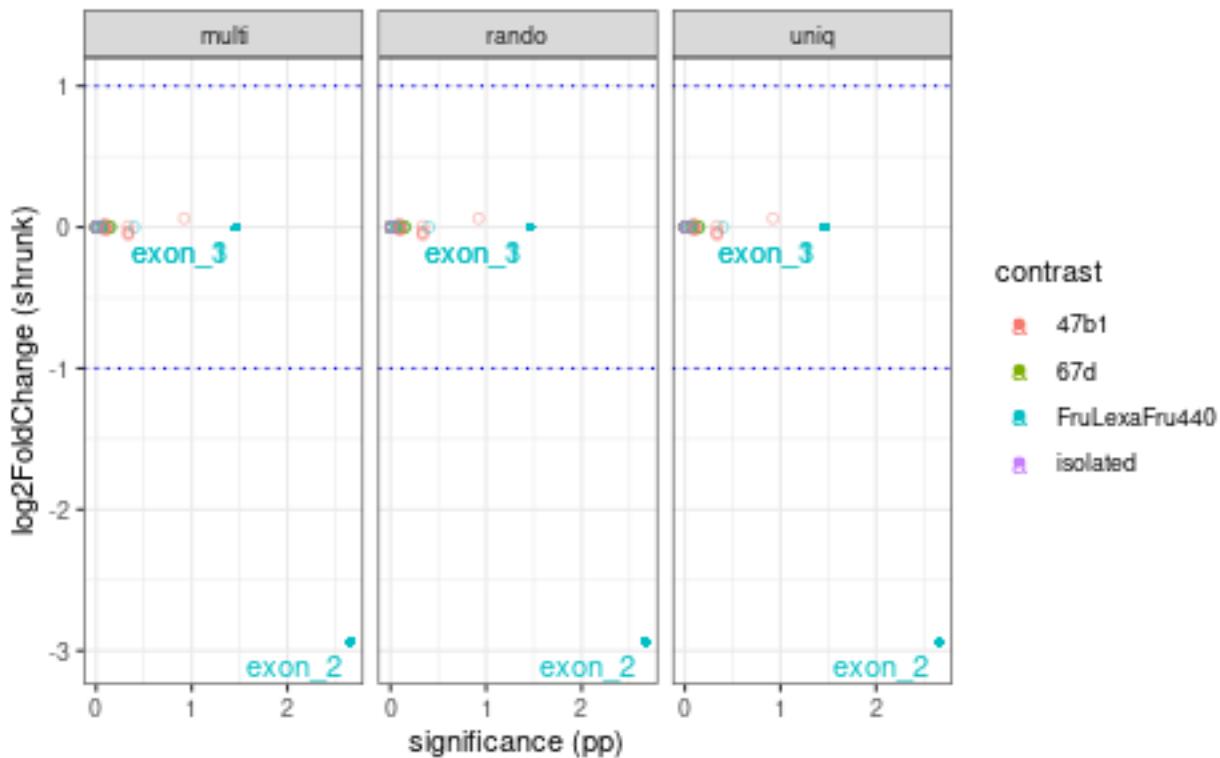
```
## png  
## 2
```

Figure 97 a. Fruitless exons with significant change in use (measured by junction) (detail)



```
## png  
## 2
```

Figure 98. Volcano Plot: Fold Change vs. Significance
(fruitless exons, 'all' counting strategy, $p_{adj} < 0.05$)



```
## png
## 2
```

3.8.3 By Intron

When the *_SplicedOnly alignments were counted ("all" strategy) against the fru_introns annotation:

Table 114. 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 115. Differential Intron Use in Fruitless, by Contrast
(Multi only)

	47b		67d		Fru		W
	log2FoldChange	adjusted p	log2FoldChange	adjusted p	log2FoldChange	adjusted p	
intron_1	0.28	5.47×10^{-1}	-0.00	9.93×10^{-1}	-0.00	4.01×10^{-2}	
intron_10	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}	
intron_11	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}	

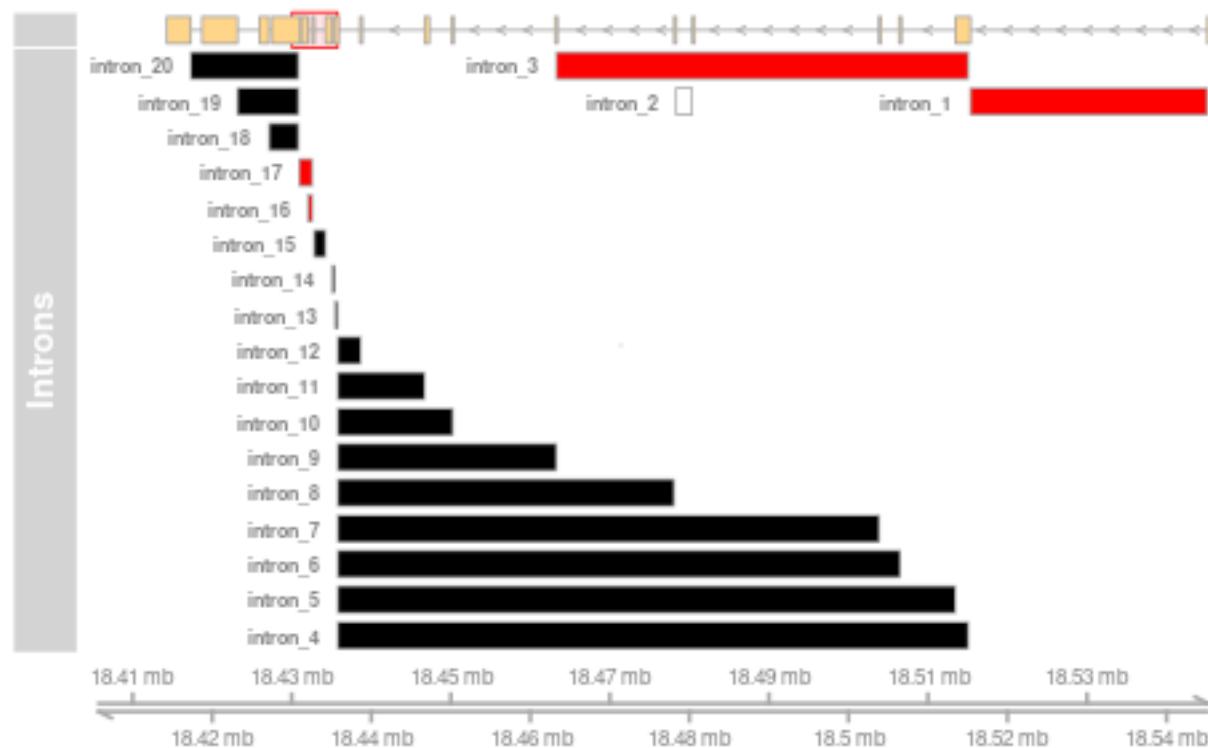
intron_12	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}
intron_13	0.62	6.95×10^{-2}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}
intron_14	0.55	2.18×10^{-1}	-0.00	9.93×10^{-1}	-0.00	9.84×10^{-1}
intron_15	0.31	5.13×10^{-1}	-0.00	9.37×10^{-1}	-0.00	9.84×10^{-1}
intron_16	2.09	5.69×10^{-6}	0.00	6.51×10^{-1}	0.00	2.14×10^{-1}
intron_17	2.09	5.69×10^{-6}	0.00	6.51×10^{-1}	0.00	2.14×10^{-1}
intron_18	0.30	5.13×10^{-1}	-0.00	9.37×10^{-1}	-0.00	9.84×10^{-1}
intron_19	0.30	5.13×10^{-1}	-0.00	9.37×10^{-1}	-0.00	9.84×10^{-1}
intron_20	0.33	5.13×10^{-1}	-0.00	9.37×10^{-1}	-0.00	9.84×10^{-1}
intron_3	-0.26	5.51×10^{-1}	-0.00	9.93×10^{-1}	-0.00	3.69×10^{-3}
intron_4	-0.03	9.31×10^{-1}	-0.00	9.93×10^{-1}	-0.00	9.84×10^{-1}
intron_5	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}
intron_6	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}
intron_7	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}
intron_8	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}
intron_9	-0.02	9.31×10^{-1}	0.00	9.93×10^{-1}	0.00	9.84×10^{-1}

Introns 1 and 3 come up significant in the FruLexa/Fru440 contrast, though they have bizarrely small effect sizes. Introns 16 and 17 come up significant in the 47b contrast.

Table 116. Fru introns with significantly ($\text{padj} < 0.05$) different use
(by aligner)

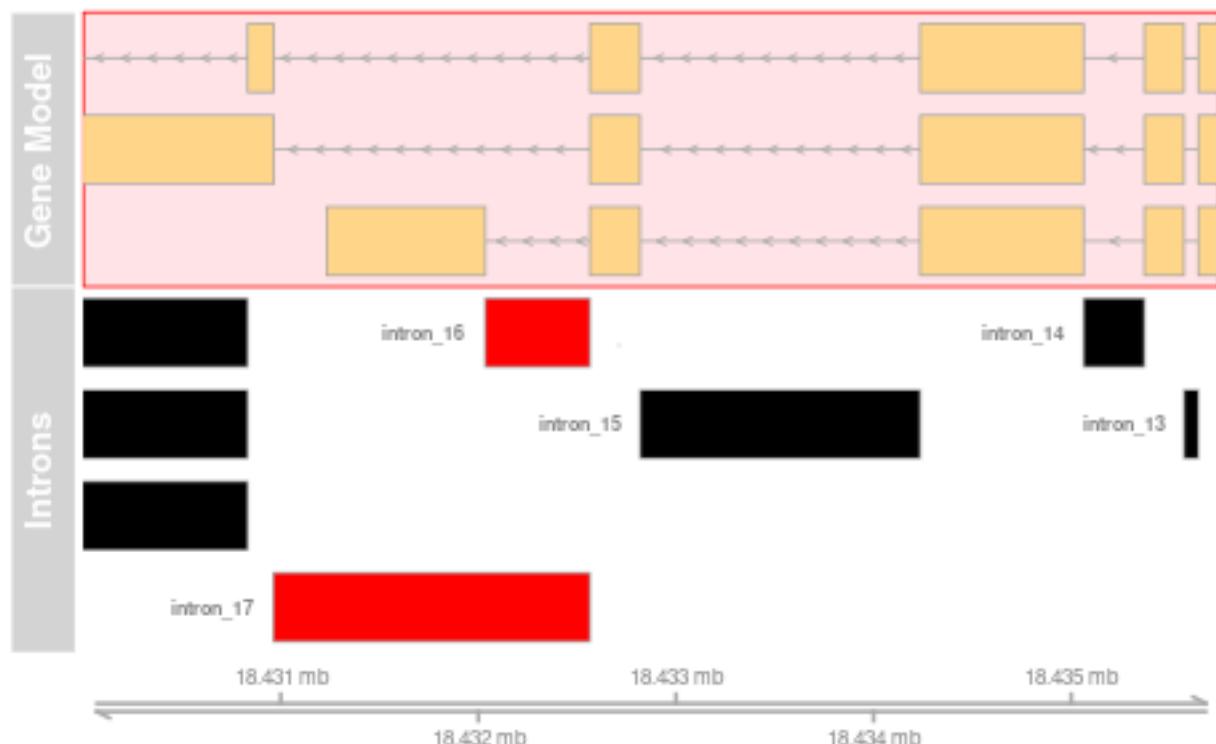
	log2FoldChange	adjusted p
<hr/>		
intron_1 - FruLexaFru440		
multi	-0.00	4.01×10^{-2}
rando	-0.00	4.01×10^{-2}
uniq	-0.00	4.01×10^{-2}
<hr/>		
intron_16 - 47b1		
multi	2.09	5.69×10^{-6}
rando	2.09	5.69×10^{-6}
uniq	2.09	5.69×10^{-6}
<hr/>		
intron_17 - 47b1		
multi	2.09	5.69×10^{-6}
rando	2.09	5.69×10^{-6}
uniq	2.09	5.69×10^{-6}
<hr/>		
intron_3 - FruLexaFru440		
multi	-0.00	3.69×10^{-3}
rando	-0.00	3.69×10^{-3}
uniq	-0.00	3.69×10^{-3}

Figure 99 Fruitless introns with significant change



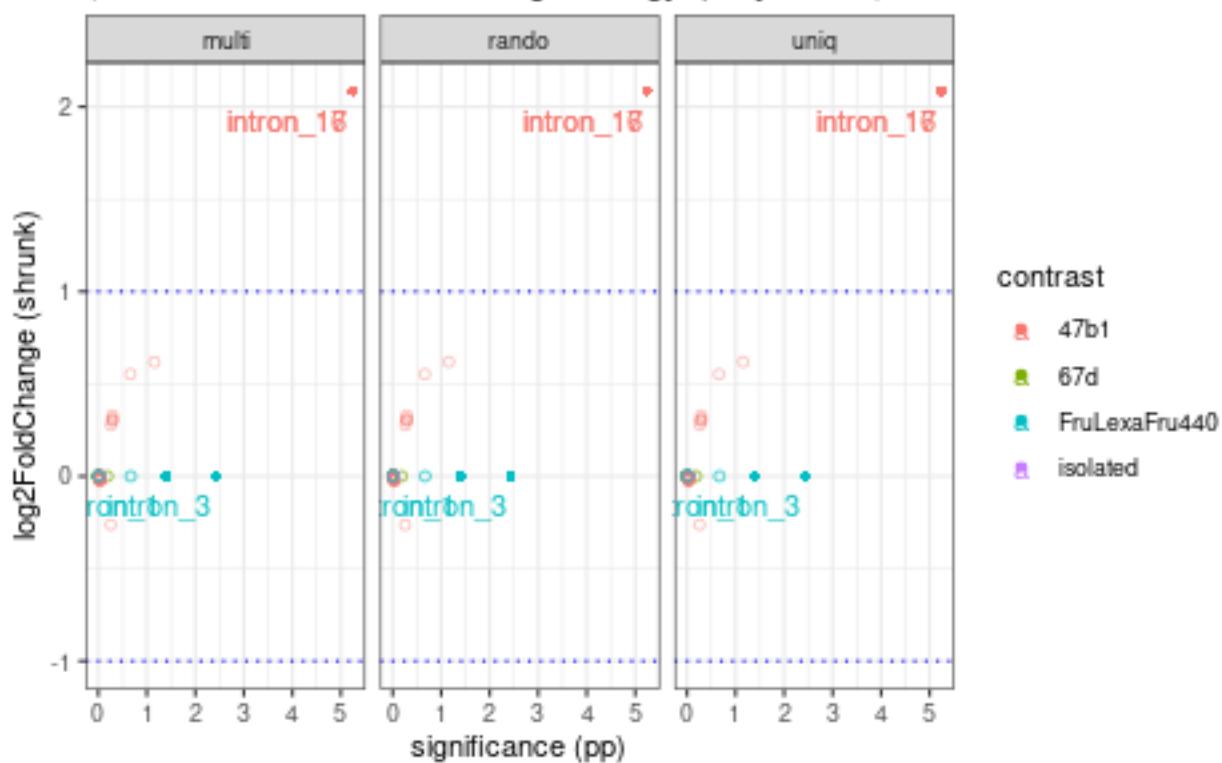
```
## png  
## 2
```

Figure 99 a. Fruitless introns with significant change (detail)



```
## png  
## 2
```

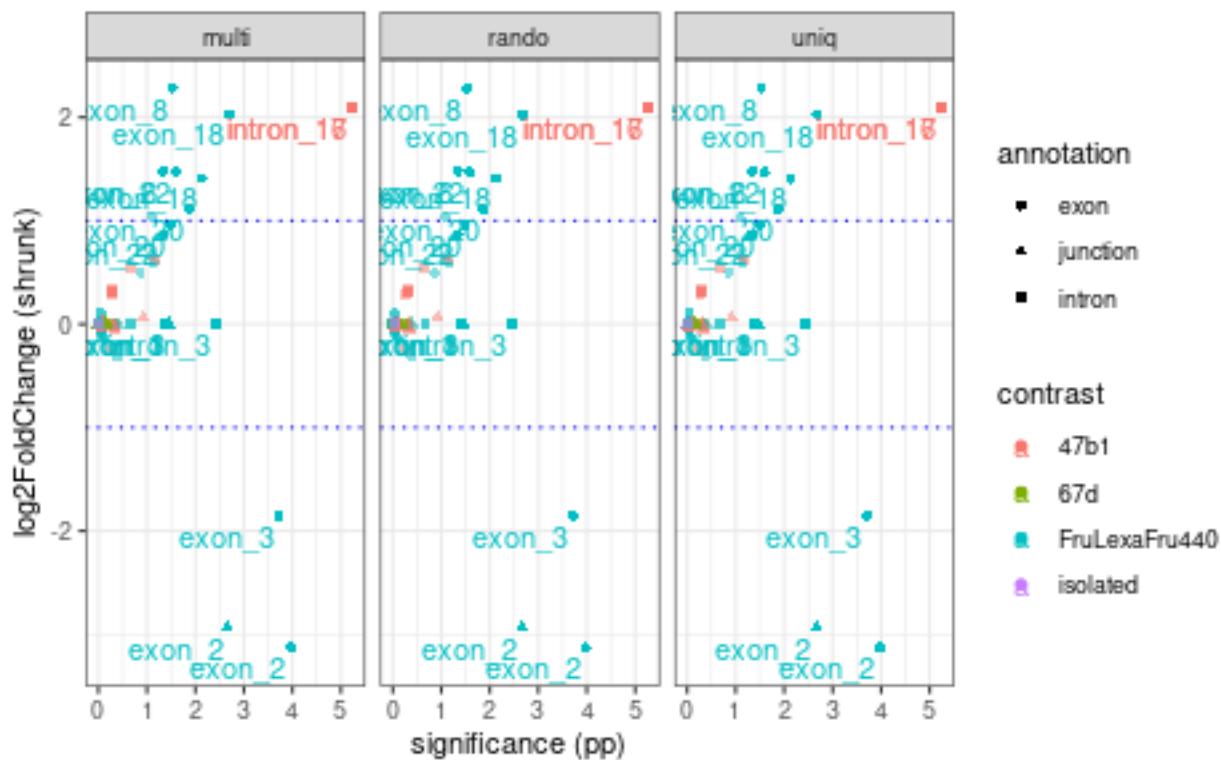
Figure 100. Volcano Plot: Fold Change vs. Significance
(fruitless introns, 'all' counting strategy, padj < 0.05)



```
## png
## 2
```

3.8.4 Overall

**Figure 101. Volcano Plot: Fold Change vs. Significance
(all fruitless breakdowns, 'all' counting strategy, $p_{adj} < 0.05$)**

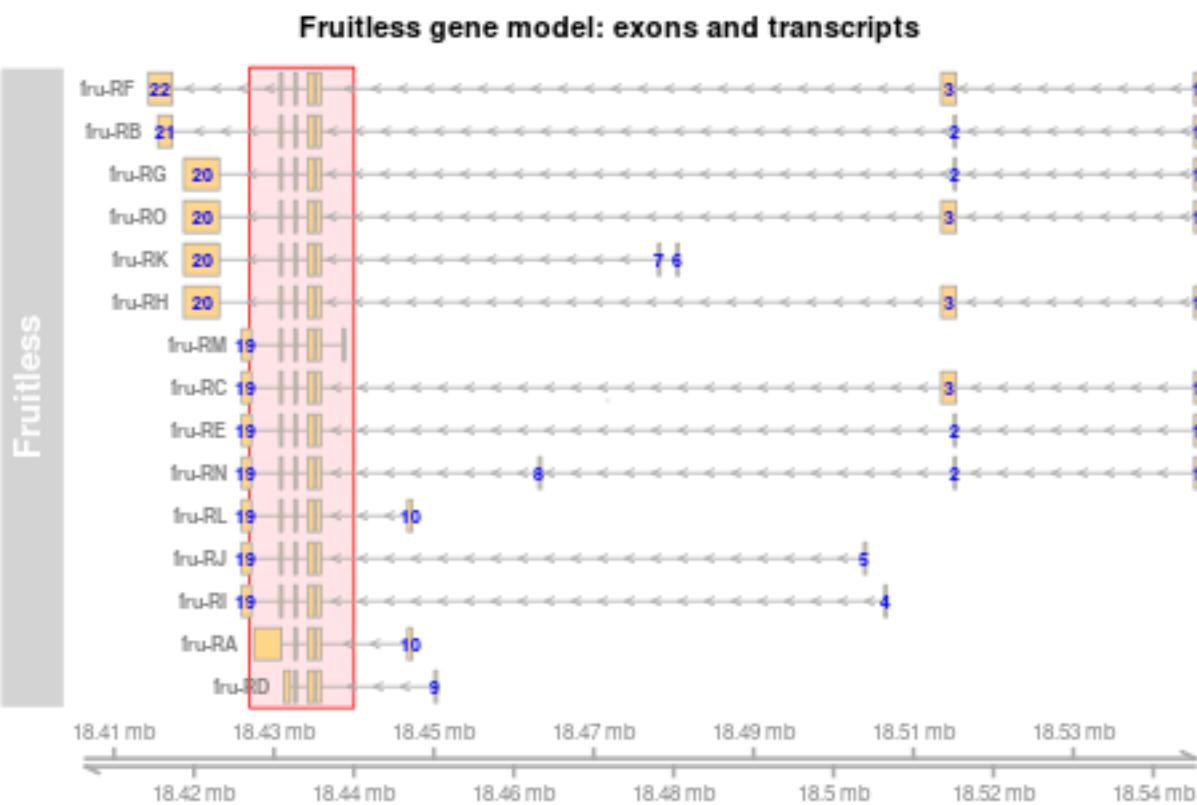


```
## png
## 2
```

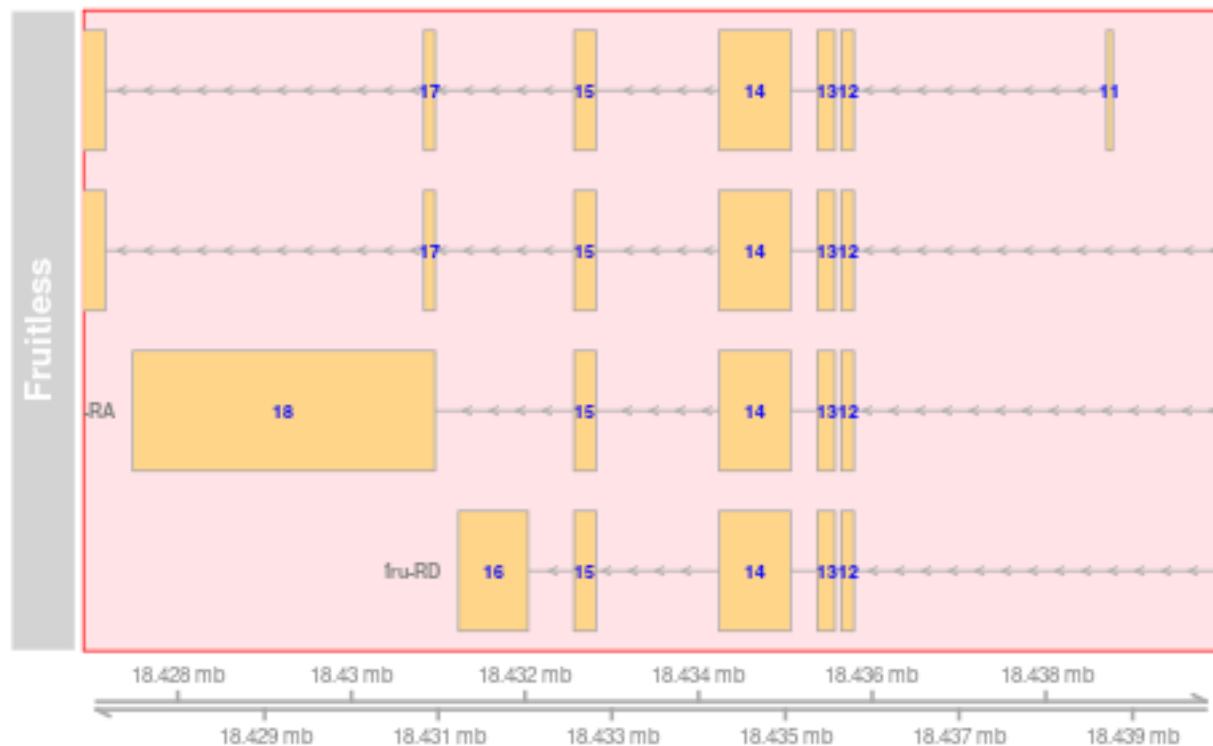
Reexamining the underlying gene models, we can try to interpret these results:

The results from both strategies for handling ambiguous assignments indicated increases in the use of 3' exons 18, 20, and 21/22, in the FruLexa/Fru440 contrast, which would mean an increase in transcripts RA, RG/RO/RK/RH, and maybe RF/RB. Measured by junction, a decrease in the use of exon 2 (and maybe exon 3) in the FruLexa/Fru440 contrast was detected. On the one hand, this is hard to reconcile with the previous observation, since all but RA and RK include either exon 2 or 3, and the RK-specific exons 6 and 7 don't show any compensating increase in use. Also, exon 19 or 16, which are the 3' exons which would be used instead of 18/20/21/22, never show a compensating decrease.

The results from the intron-based analysis technically supports the decrease in the use of the 5' exons 1 and 2 but the effect size is bizarrely low. the 47b1 contrast results are more sensible, indicating an enhanced use of the most 5' exon 1. It appears to specifically differentiate the use of exons 2 and 3, specifically finding an increase in the intron between exons 2 and 8, ie, an increased use of transcript RN.



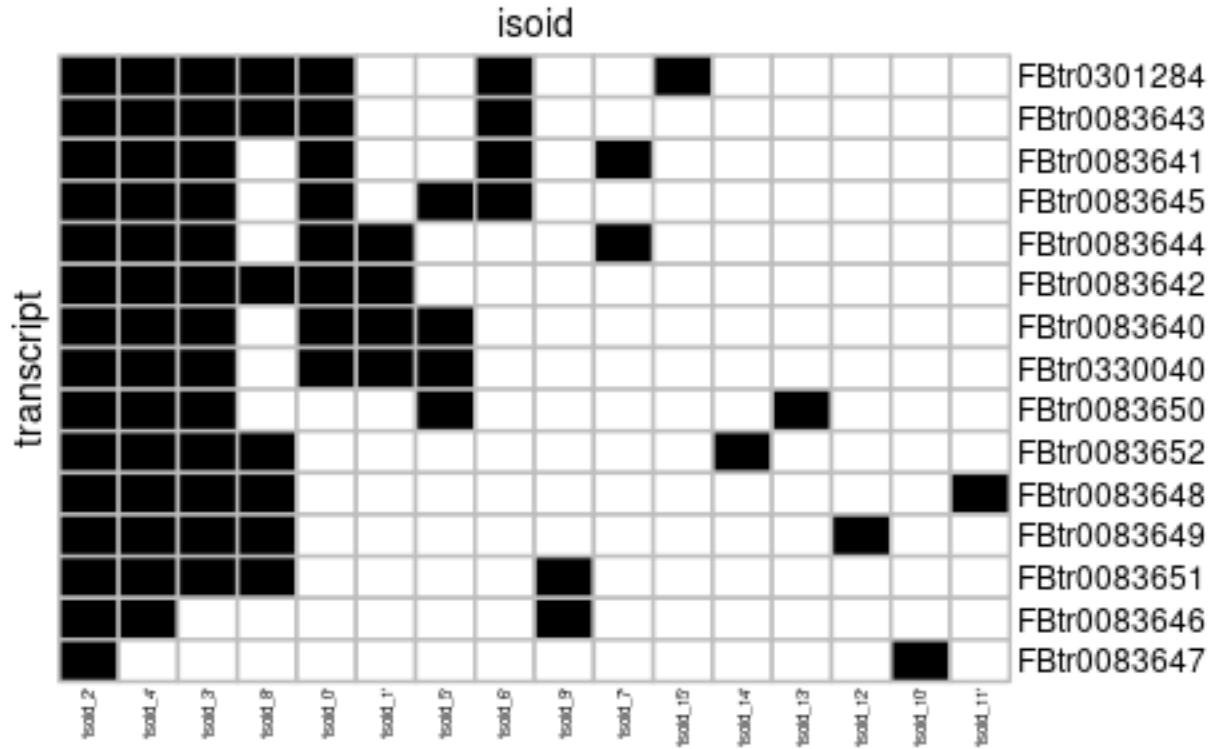
Fruitless gene model: exons and transcripts (detail)



3.8.5 edgeHog

An experimental approach here is similar to the junction/intron assignment, except the splice sites have been grouped according to which subset of transcripts contain them. For example, the constitutive exons would be assigned to a group representing all transcripts, whereas transcript-specific exon junction would contribute to a group representing only that transcript. The gene model representing a subset of transcripts is called an “isoid”.

Figure 103 . Transcript Subsets and their Isoid Representations (Fruitless)



```
## png
## 2
```

Am I handling 2-sidedness correctly? double check this

The “SplicedOnly” reads were counted against the isoids, with assignment to all annotations overlapped. These counts were used with DESeq2 and the hausWtVsMut contrast. Since ONLY the Fru counts are used, this normalizes any difference in overall expression of Fru between treatments

Table 117. Significant Changes in Fru Transcript Use
by Stouffer's Test on DESeq2 + Isoids

transcript	housing		FruLexaFru440		67d		47b	
	Z	p	Z	p	Z	p	Z	p
FBtr0083640	NA	NA	NA	NA	NA	NA	NA	NA
FBtr0083641	0.68	5.0×10^{-1}	0.97	3.3×10^{-1}	0.13	9.0×10^{-1}	-0.50	6.2×10^{-1}
FBtr0083642	NA	NA	NA	NA	NA	NA	NA	NA
FBtr0083643	0.14	8.9×10^{-1}	0.95	3.4×10^{-1}	-0.39	7.0×10^{-1}	0.03	9.7×10^{-1}
FBtr0083644	NA	NA	NA	NA	NA	NA	NA	NA
FBtr0083645	-0.04	9.7×10^{-1}	1.23	2.2×10^{-1}	-0.49	6.2×10^{-1}	-0.07	9.5×10^{-1}
FBtr0083646	0.12	9.1×10^{-1}	0.24	8.1×10^{-1}	0.91	3.7×10^{-1}	0.10	9.2×10^{-1}
FBtr0083647	-1.04	3.0×10^{-1}	-1.64	1.0×10^{-1}	-1.72	8.6×10^{-2}	-2.03	4.2×10^{-2}
FBtr0083648	NA	NA	NA	NA	NA	NA	NA	NA
FBtr0083649	NA	NA	NA	NA	NA	NA	NA	NA
FBtr0083650	NA	NA	NA	NA	NA	NA	NA	NA
FBtr0083651	0.25	8.0×10^{-1}	0.09	9.3×10^{-1}	0.88	3.8×10^{-1}	0.10	9.2×10^{-1}

FBtr0083652	NA							
FBtr0301284	NA							
FBtr0330040	NA							

The only vaguely significant change is in the FruLexaFru440 treatment, in which FBtr0083647 is depleted approximately 4 fold.

3.8.6 DEXSeq

All approaches to differential exon use detection were based around a standard featureCounts -> DESeq2 subpipeline, with modifications made to the input reads & annotations and/or downstream analysis. To compare these results to an established tool, DEXSeq (Anders, Reyes, and Huber 2012) was used. Another difference is that while the other methods have analyzed the Fruitless locus in isolation, this tool was run on the entire annotation and Fruitless results extracted later.

DEXSeq divides the exons in the annotation into non-overlapping intervals by exon start/end points:

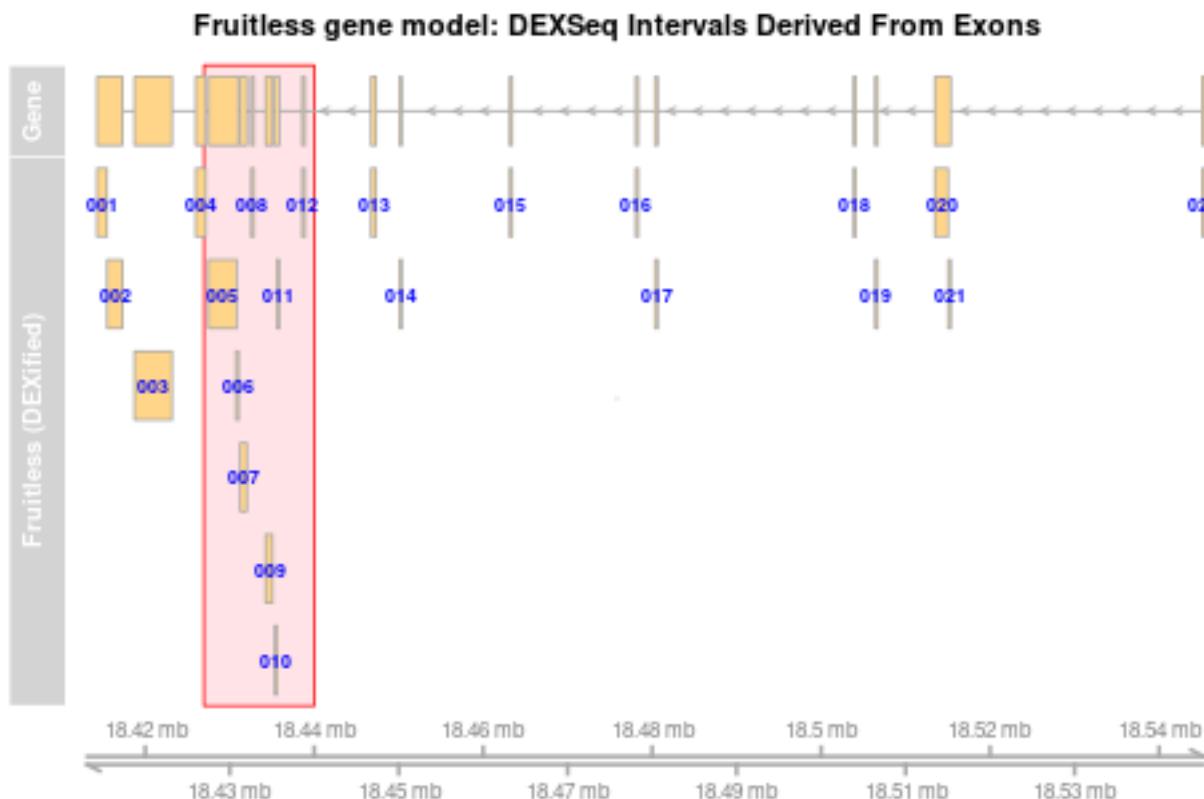
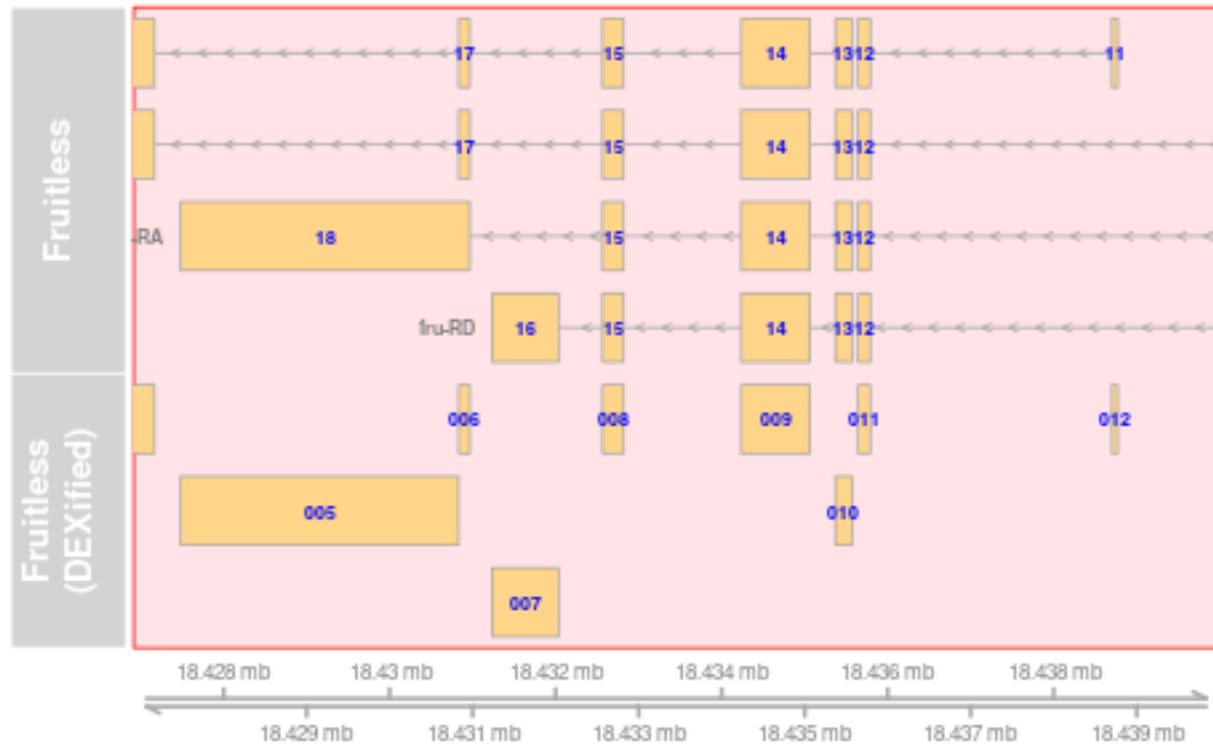


Figure 103 a. Fruitless gene model: DEXSeq Intervals Derived From Exons (detail)



**Table 118. DEXSeq Test for Differential Exon Use
Fruitless (FBgn0004652) Exons**

internal name	genomic locus	log2FoldChange	padj
grpWtVs47b			
E005	chr3R:18427480-18430831	0.97	8.17×10^{-4}
grpWtVsFru			
E021	chr3R:18515052-18515343	-3.90	1.25×10^{-10}
grpWtVs67d			
E005	chr3R:18427480-18430831	1.27	2.00×10^{-7}

The interval E005 corresponds to the 3' end unique to exon_18, and a significant decrease in its use is detected in the 47b and 67d treatments. E021 corresponds to exon_2/the shared 5' end of exon_3, and a significant increase in its use is detected in the FruLexaFru440 treatment.

Figure 104. DEXSeq Estimate of Exon Use
47b1 contrast; significant (adjusted p<0.01) Differences Circled



```
## png
## 2
```

Fruitless gene model: DEXSeq Intervals Derived From Exons

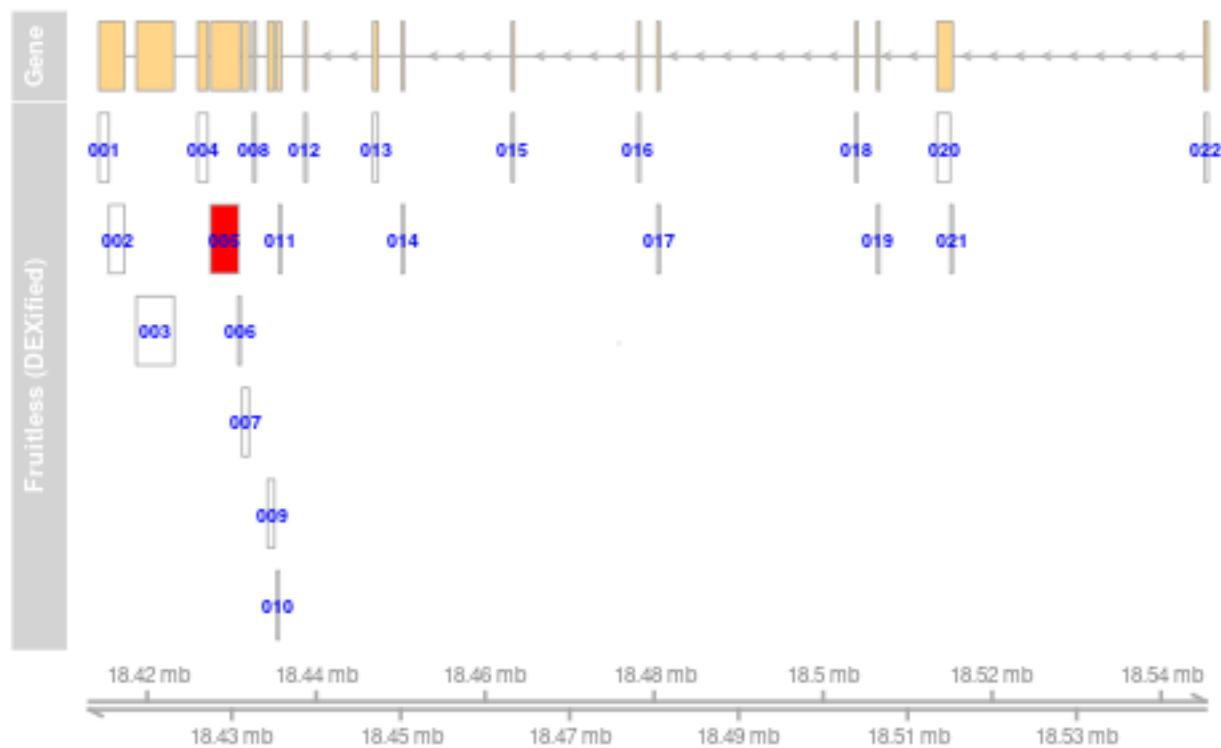
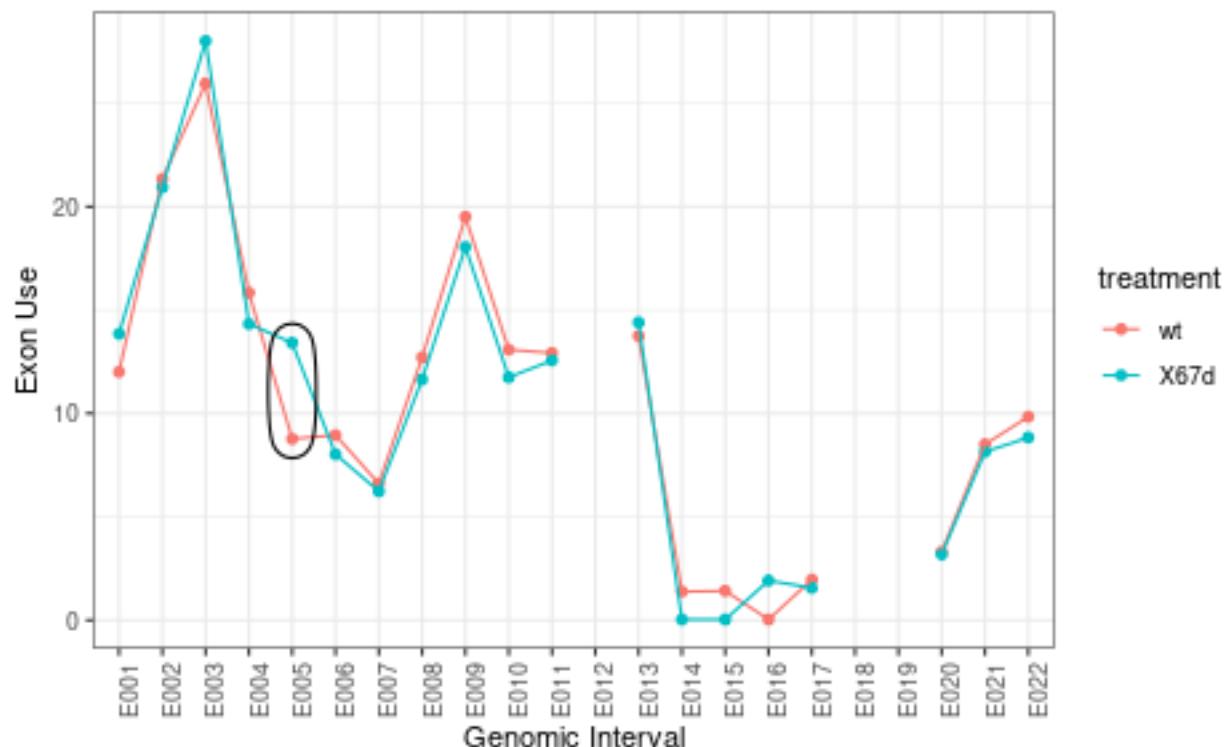
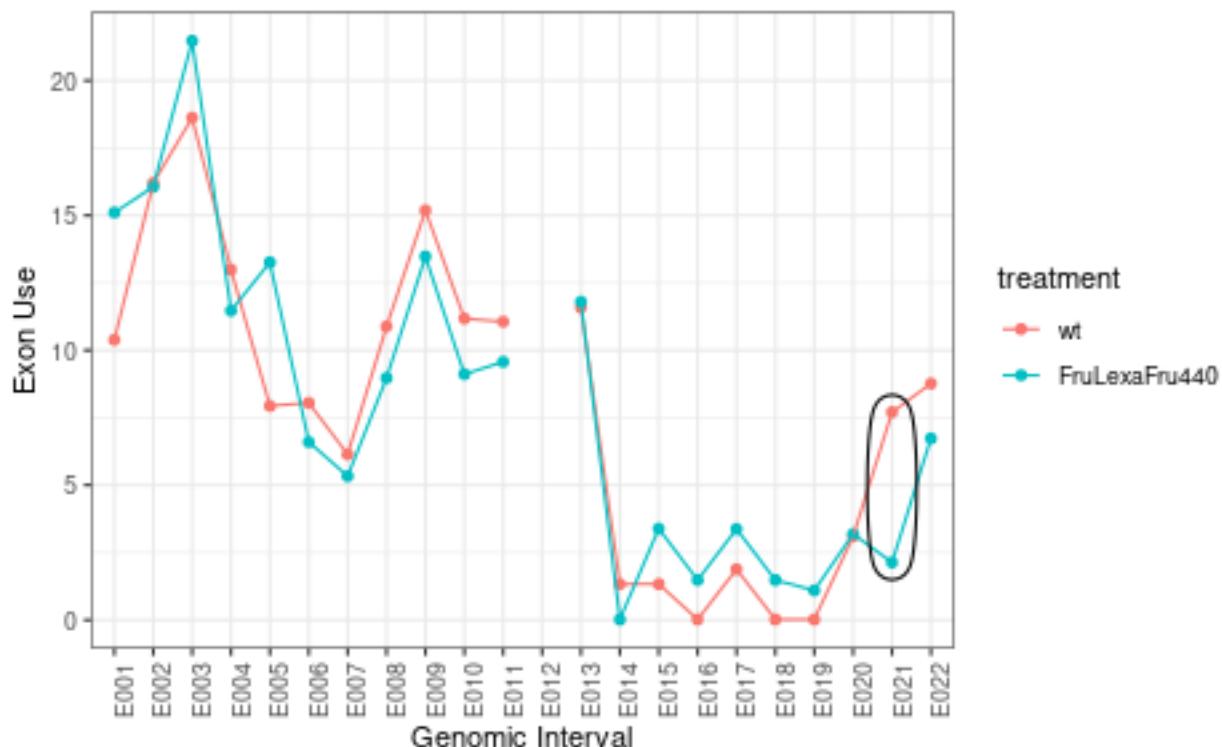


Figure 105. DEXSeq Estimate of Exon Use
67d contrast; significant (adjusted p<0.01) Differences Circled



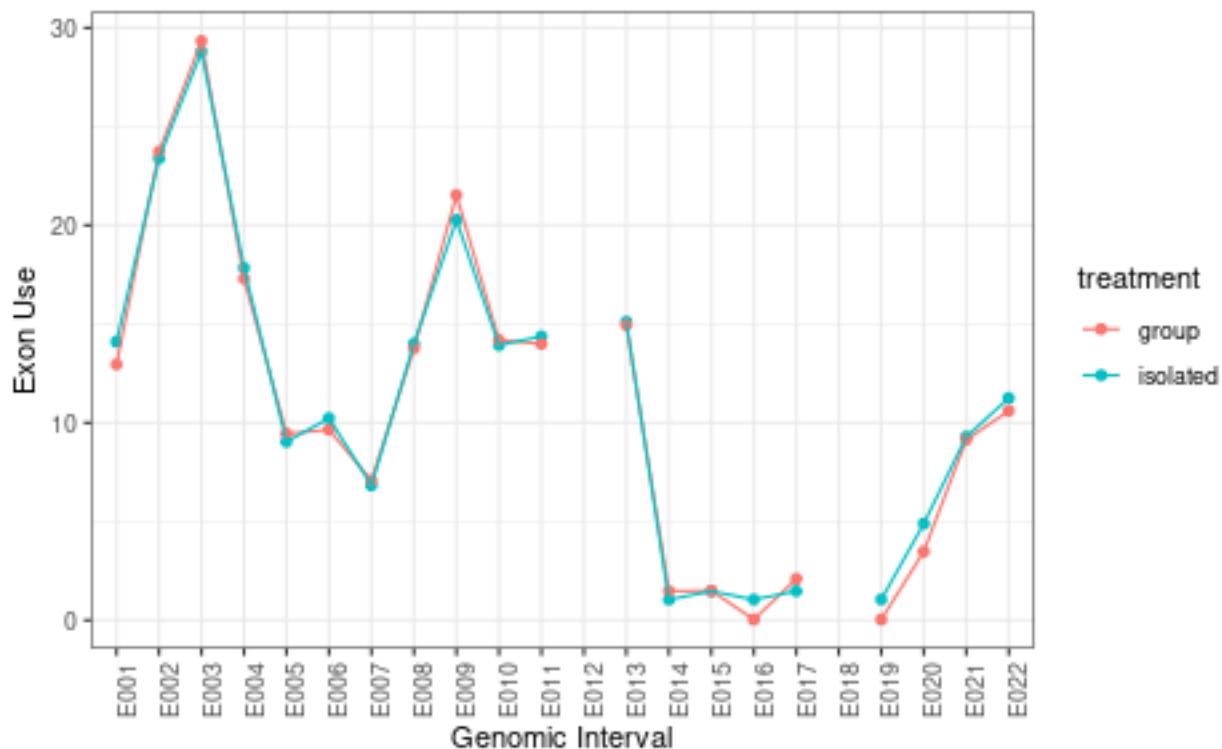
```
## png
## 2
```

Figure 106. DEXSeq Estimate of Exon Use
FruLexaFru440 contrast; significant (adjusted p<0.01) Differences Circled



```
## png
## 2
```

Figure 107. DEXSeq Estimate of Exon Use
 Housing contrast; significant (adjusted p<0.01) Differences Circled



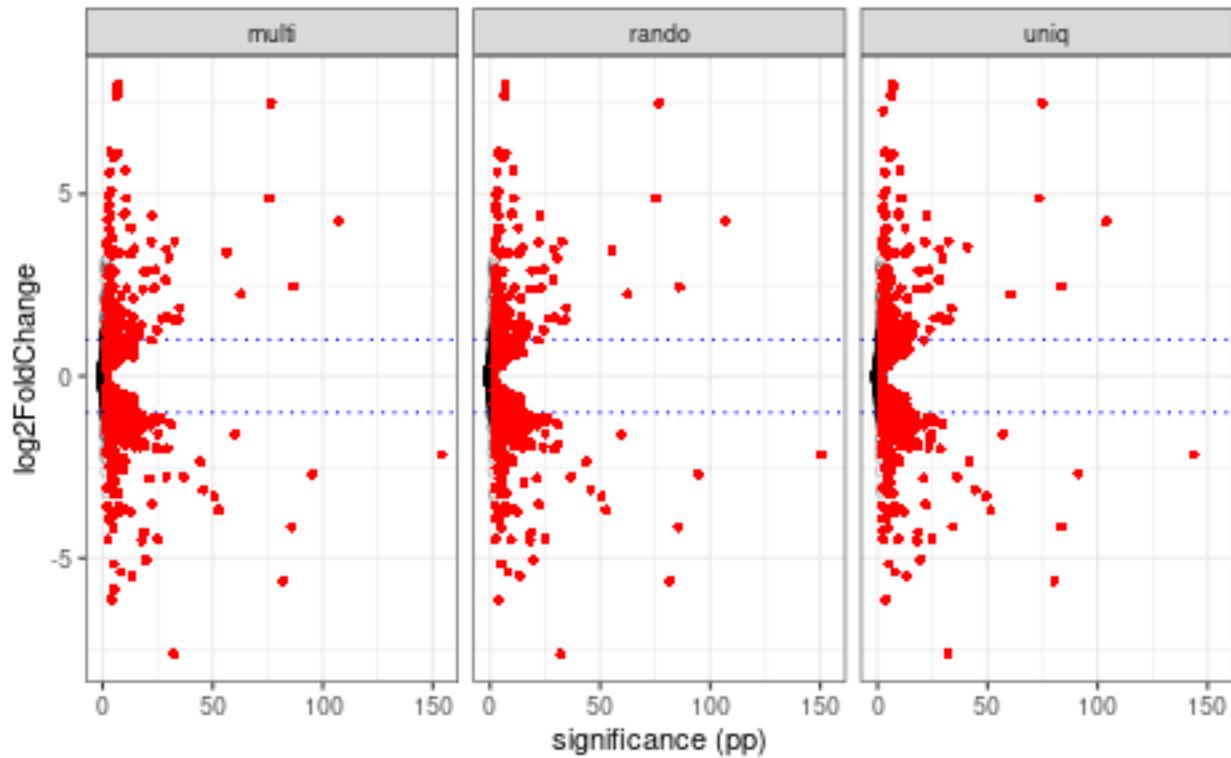
```
## png
## 2
```

3.9 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 ($\text{padj}<0.01$) 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 depleted relative to the group-housed condition; positive fold changes are enriched .

**Figure 108. Volcano Plot: Fold Change vs. Significance
(88a mutants, with/without 47b2)**



```
## png
## 2
```

Of the 11308 genes with significance scores available, 856 have an adjusted $p < 0.01$ (7.569862 %)

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 355 such genes, mostly shared across alignment strategy:

results/tables/tbl11947b2on88a_chonky.html

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 120. 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	-2.154	1.03×10^{-154}	Drip	1.45	-2.153	2.42×10^{-154}
2	CG10936	0.11	4.251	4.73×10^{-108}	CG10936	0.11	4.252	1.17×10^{-108}
3	Cyp4ac2	1.28	-2.692	4.28×10^{-96}	Cyp4ac2	1.27	-2.689	2.36×10^{-96}
4	Cyp6a2	3.91	2.454	2.76×10^{-87}	Cyp6a2	3.91	2.454	9.08×10^{-87}

5	CG14400	0.82	-4.146	1.77×10^{-86}	CG14400	0.82	-4.146	2.62×10^{-86}
6	Or85b	0.32	-5.627	1.23×10^{-82}	Or85b	0.32	-5.626	1.93×10^{-82}
7	Cyp6a17	0.53	7.487	3.16×10^{-77}	Cyp6a17	0.53	7.487	1.53×10^{-77}
8	wntD	0.25	4.880	2.84×10^{-76}	wntD	0.25	4.880	4.04×10^{-76}
9	Cyp6a8	0.46	2.250	1.38×10^{-63}	Cyp6a8	0.46	2.250	3.51×10^{-63}
10	Cyp6a20	4.64	-1.602	7.19×10^{-61}	Cyp6a20	4.64	-1.601	2.91×10^{-61}

Top 10 genes with biggest (significant) effect sizes

Table 121. 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	
1	lncRNA:CR45791	0.02	8.022	1.29×10^{-7}	lncRNA:CR45791	0.02	8.022	
2	CG14069	0.02	7.998	1.71×10^{-7}	CG14069	0.02	7.998	
3	CR43734	0.04	7.885	1.27×10^{-7}	CR43734	0.04	7.885	
4	lncRNA:CR44057	0.02	7.684	4.63×10^{-7}	lncRNA:CR44057	0.02	7.684	
5	CG10462	0.07	-7.633	6.19×10^{-33}	CG10462	0.07	-7.633	
6	Cyp6a17	0.53	7.487	3.16×10^{-77}	Cyp6a17	0.53	7.487	
7	CG30091	0.00	6.149	2.94×10^{-4}	CG30091	0.00	6.149	
8	CG42329	0.00	-6.146	1.42×10^{-4}	CG42329	0.00	-6.146	
9	CG34124	0.00	6.107	8.94×10^{-8}	CG34124	0.00	6.107	
10	CG13748	0.02	5.999	4.86×10^{-6}	CG13748	0.02	5.999	

Top 10 highest expressed genes with significant change

Table 122. Top Ten Highest Expressed Genes with Significant (padj < 0.01) Difference
88a mutants, with/without 47b2

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	CG14661	27.65	0.316	2.97×10^{-5}	CG14661	27.66	0.316	3.70×10^{-5}
2	CG6409	24.34	-0.411	5.61×10^{-3}	CG6409	24.34	-0.412	5.86×10^{-3}
3	CG9497	21.41	-0.333	8.32×10^{-6}	CG9497	21.41	-0.332	1.14×10^{-5}
4	Ugt35B1	16.88	-0.261	7.60×10^{-4}	Ugt35B1	16.88	-0.260	9.30×10^{-4}
5	antdh	16.35	-0.442	8.47×10^{-8}	antdh	16.35	-0.441	1.25×10^{-7}
6	Est-6	12.82	0.286	1.02×10^{-3}	Est-6	12.82	0.286	1.14×10^{-3}
7	Obp56d	11.55	0.667	5.23×10^{-8}	Obp56d	11.55	0.670	5.68×10^{-8}
8	Cyp6w1	9.97	-0.456	3.79×10^{-8}	Cyp6w1	9.97	-0.455	5.47×10^{-8}
9	CG5867	9.90	-0.283	9.82×10^{-3}	Cyp4e2	9.22	-0.305	7.63×10^{-4}
10	Cyp4e2	9.23	-0.308	5.94×10^{-4}	CG11550	9.13	0.400	1.18×10^{-7}

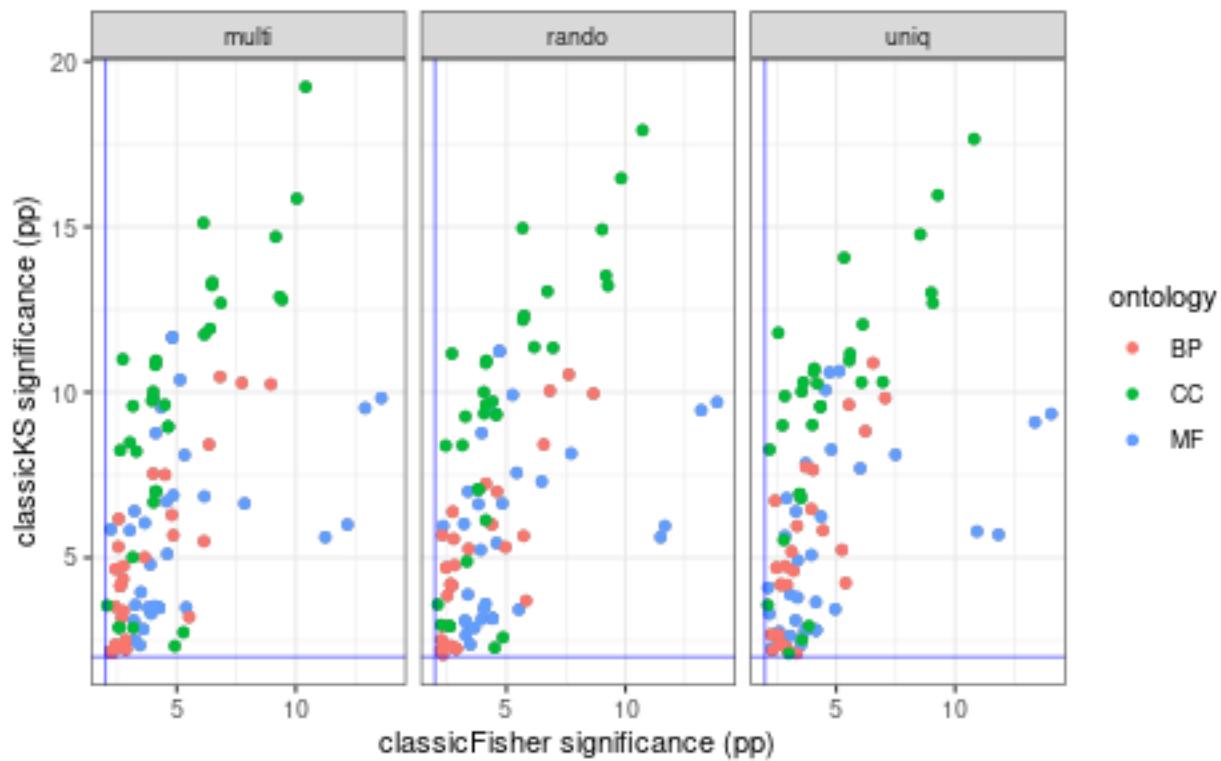
3.9.2 Gene Ontology Enrichment

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



```
## png
## 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 123. 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	2.5
GO:0020037	heme binding	1.2

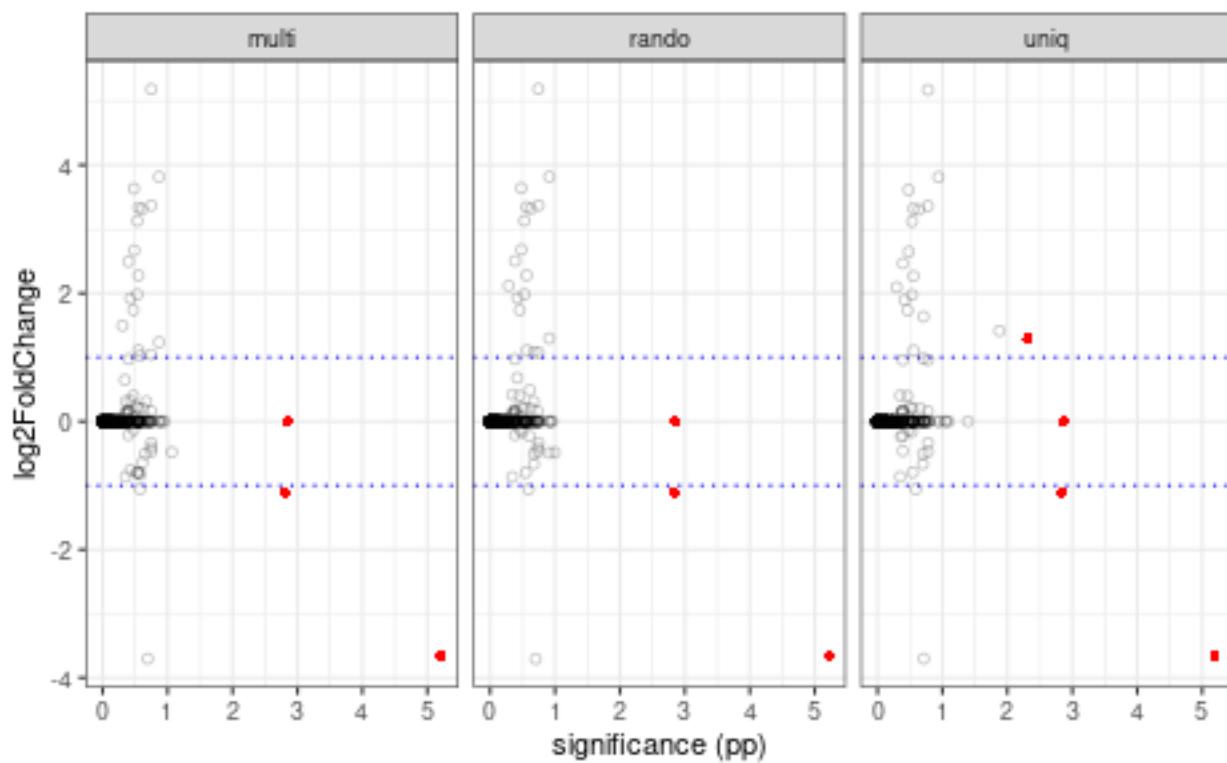
GO:0005506	iron ion binding	6.7
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	5.7
GO:0048037	obsolete cofactor binding	1.4
GO:0016491	oxidoreductase activity	6.9
GO:0015081	sodium ion transmembrane transporter activity	4.0
GO:0004984	olfactory receptor activity	4.7
GO:0005549	odorant binding	7.2
GO:0031492	nucleosomal DNA binding	1.4
<hr/>		
BP		
GO:0003008	NA	1.1
GO:0007600	NA	1.8
GO:0050877	nervous system process	1.5
GO:0007606	sensory perception of chemical stimulus	4.4
GO:0009593	detection of chemical stimulus	7.2
GO:0055114	oxidation-reduction process	3.0
GO:0050907	detection of chemical stimulus involved in sensory perception	1.4
GO:0050906	detection of stimulus involved in sensory perception	1.6
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	3.1
GO:0007608	sensory perception of smell	9.7
<hr/>		
CC		
GO:0005576	extracellular region	3.8
GO:0005886	plasma membrane	9.0
GO:0016021	integral component of membrane	3.7
GO:0031224	intrinsic component of membrane	4.7
GO:0071944	cell periphery	6.9
GO:0044421	NA	1.4
GO:0031226	intrinsic component of plasma membrane	3.2
GO:0005887	integral component of plasma membrane	3.3
GO:0016020	membrane	4.0
GO:0005615	extracellular space	6.9

3.10 Two Days' Difference

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

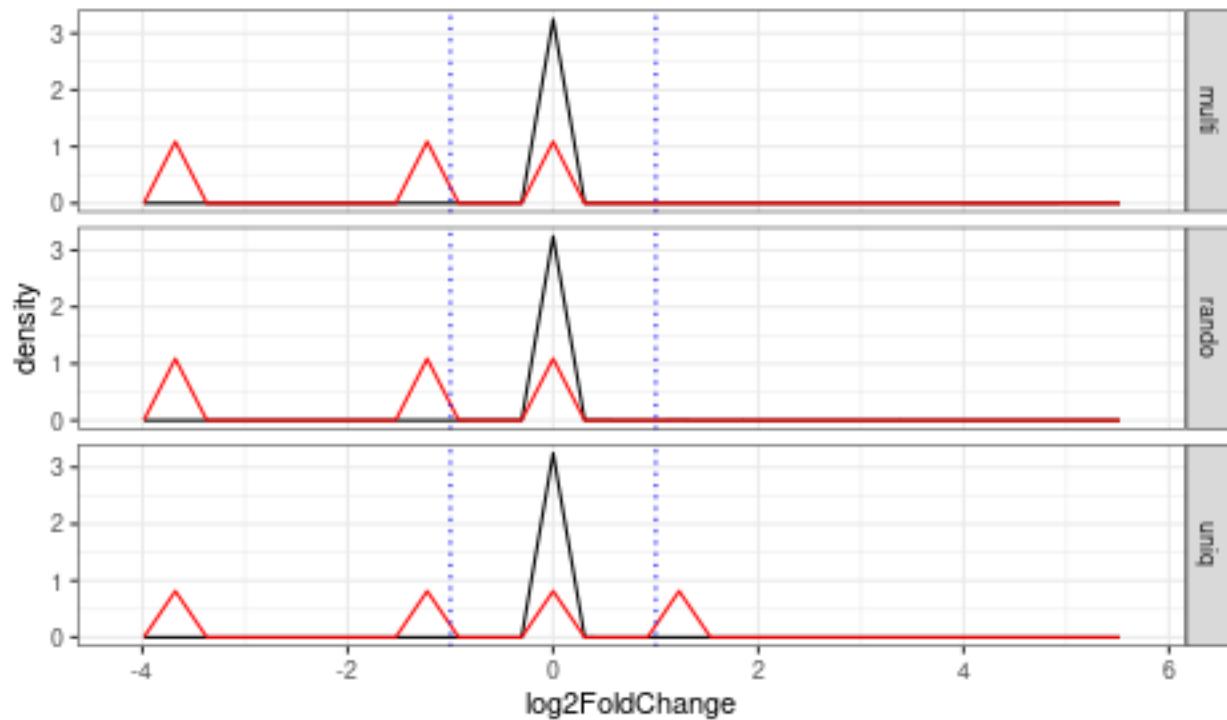
Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of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

Figure 110. Volcano Plot: Fold Change vs. Significance
(from day 5 to day 7)



```
## png
## 2
```

Figure 111. histogram of fold change
with significant($p_{adj} < 0.01$) changes highlighted in red
(from day 5 to day 7)



```
## png
## 2
```

Of the 11863 genes with significance scores available, 4 have an adjusted $p < 0.01$ (0.0337183 %)

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 3 such genes, mostly shared across alignment strategy:

Table 124. Genes with Large ($|2| > \text{fold change}$), Significant ($p_{adj} < 0.01$) Changes from day 5 to day 7

	multi	rando	uniq
CheB38c	yes	yes	yes
Cpr64Ac	yes	yes	yes
Gyc76C	no	no	yes

3.10.1 Top Tens

Genes with top 10 most significant changes

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

Table 125. Top Ten Most Significantly ($p_{adj} < 0.01$) Differentially Express 88a mutants, with/without 47b2

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	CheB38c	0.04	-3.658	6.21×10^{-6}	CheB38c	0.04	-3.659	6.04×10^{-6}
2	wntD	0.07	-0.000	1.41×10^{-3}	wntD	0.07	-0.000	1.42×10^{-3}
3	Cpr64Ac	0.57	-1.106	1.57×10^{-3}	Cpr64Ac	0.57	-1.107	1.47×10^{-3}
4	NA	NA	NA	NA	NA	NA	NA	NA

Top 10 genes with biggest (significant) effect sizes

Table 126. Top Ten Largest Magnitude Fold Changes which were Significant
47b1 mutants, from day 5 to day 7

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	CheB38c	0.04	-3.658	6.21×10^{-6}	CheB38c	0.04	-3.659	6.04×10^{-6}
2	Cpr64Ac	0.57	-1.106	1.57×10^{-3}	Cpr64Ac	0.57	-1.107	1.47×10^{-3}
3	wntD	0.07	-0.000	1.41×10^{-3}	wntD	0.07	-0.000	1.42×10^{-3}
4	NA	NA	NA	NA	NA	NA	NA	NA

Top 10 highest expressed genes with significant change

Table 127. Top Ten Highest Expressed Genes with Significant (padj < Difference)
47b1 mutants, from day 5 to day 7

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	Cpr64Ac	0.57	-1.106	1.57×10^{-3}	Cpr64Ac	0.57	-1.107	1.47×10^{-3}
2	wntD	0.07	-0.000	1.41×10^{-3}	wntD	0.07	-0.000	1.42×10^{-3}
3	CheB38c	0.04	-3.658	6.21×10^{-6}	CheB38c	0.04	-3.659	6.04×10^{-6}
4	NA	NA	NA	NA	NA	NA	NA	NA

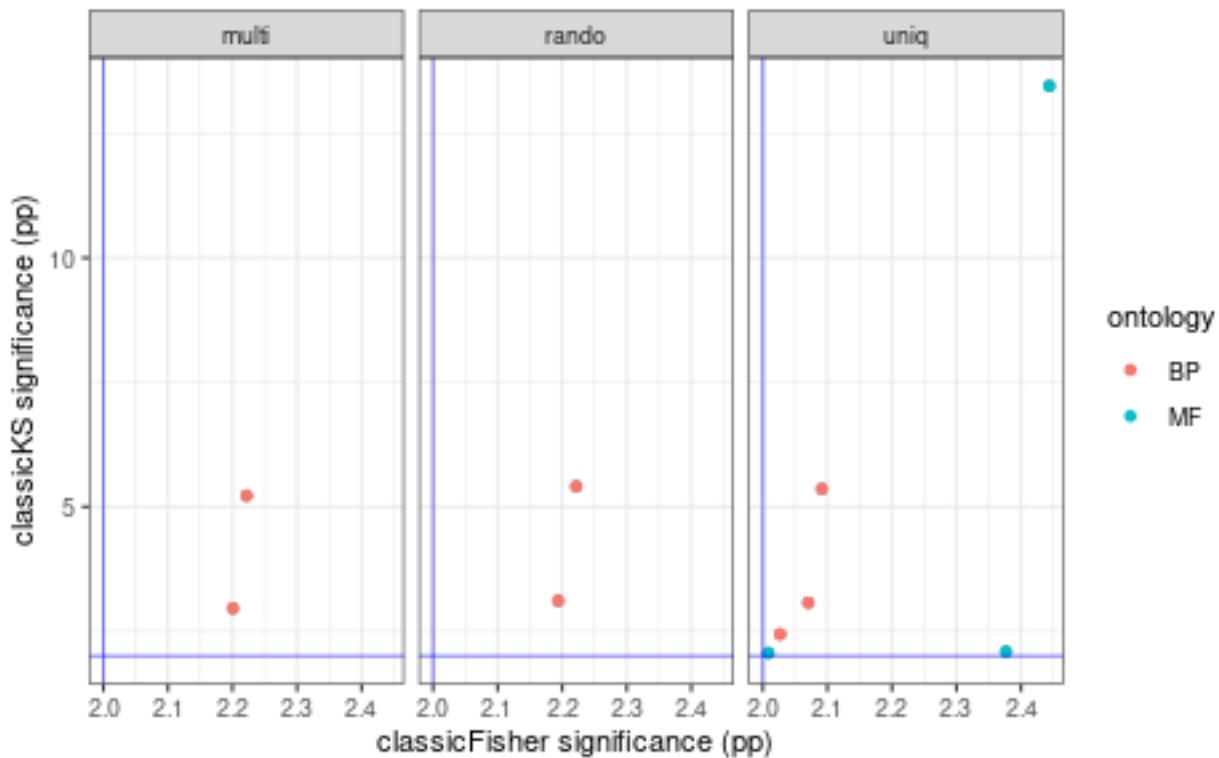
3.10.2 Gene Ontology Enrichment

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



```
## png
## 2

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

Table 128. 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	3.60×10^{-3}	3.50×10^{-14}
GO:0030215	semaphorin receptor binding	4.20×10^{-3}	8.20×10^{-3}
GO:0009975	NA	9.80×10^{-3}	8.70×10^{-3}
BP			
GO:0007370	ventral furrow formation	8.10×10^{-3}	4.40×10^{-6}
GO:0007280	pole cell migration	8.50×10^{-3}	8.50×10^{-4}
GO:0052652	NA	9.40×10^{-3}	3.66×10^{-3}

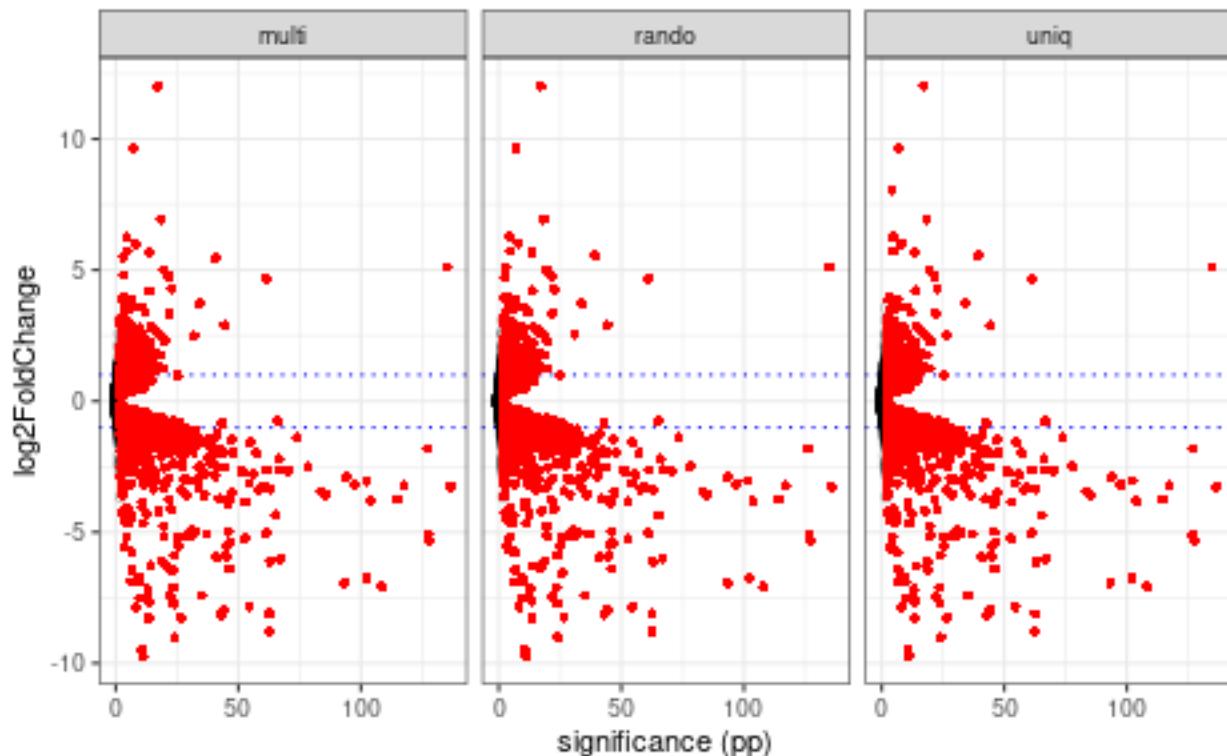
3.11 CantonS & amos Mutants Thereof

Reads from CantonS flies, with and without a mutation in amos, were downloaded from NCBI (Mohapatra and Menuz 2019). These were analyzed along with the group-housed wild-type flies from this study in the cantonAmosWt contrast. This was a 2-factor model, with one explanatory variable being the state of amos (normal vs mutant) and the other being genomic background (cantonS vs white-eyed). An earlier version used a single-factor model for each of these variables, but the results were questionable, especially the latter. Full table: results/tables/supp/cantonAmosWt.de.tsv

3.11.1 Impact of amos Mutant

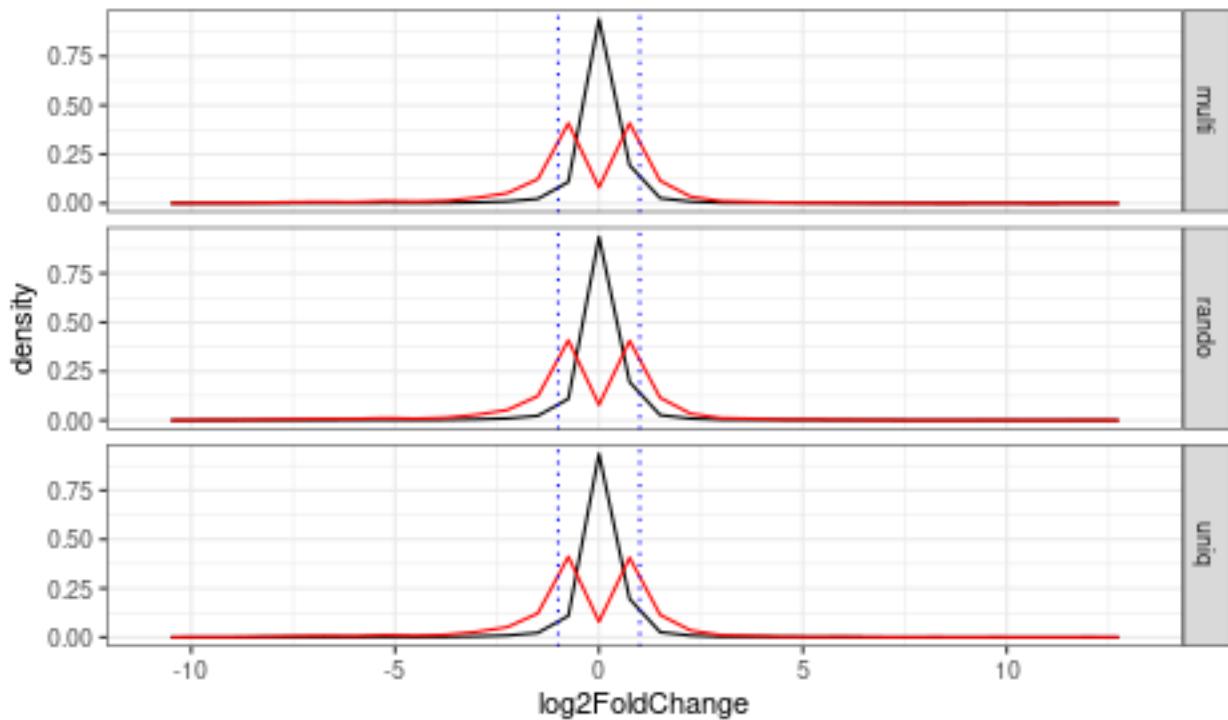
Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of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

**Figure 113. Volcano Plot: Fold Change vs. Significance
(amos mutants vs garden-variety background)**



```
## png
## 2
```

**Figure 114. histogram of fold change
with significant($p_{adj} < 0.01$) changes highlighted in red
(from garden-variety to amos mutant)**



```
## png
## 2
```

Of the 11647 genes with significance scores available, 2074 have an adjusted $p < 0.01$ (17.8071606 %)

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 804 such genes, mostly shared across alignment strategy:

Table 129. Genes with Large ($|2| <$ fold change), Significant ($p_{adj} < 0.01$) Changes from garden-variety to amos mutants

	multi	rando	uniq
CG7509	yes	yes	yes
CG43861	yes	yes	yes
CG9555	yes	yes	yes
Gr64f	yes	yes	yes
CG42284	yes	yes	yes
CG34342	yes	yes	yes
CG7896	yes	yes	yes
Or47b	yes	yes	yes
Or7a	yes	yes	yes
Galphaf	yes	yes	yes
ninaB	yes	yes	yes
CG15170	yes	yes	yes
Or85b	yes	yes	yes

CG17321	yes	yes	yes
CG43222	yes	yes	yes
Kua	yes	yes	yes
CG6938	yes	yes	yes
CG30271	yes	yes	yes
CG9541	yes	yes	yes
Gr21a	yes	yes	yes
Or67d	yes	yes	yes
Or56a	yes	yes	yes
Or59b	yes	yes	yes
Orco	yes	yes	yes
Or49b	yes	yes	yes
5-HT2B	yes	yes	yes
CG42237	yes	yes	yes
CG33177	yes	yes	yes
CG34265	yes	yes	yes
Or43b	yes	yes	yes
CG30274	yes	yes	yes
CG5079	yes	yes	yes
Gr63a	yes	yes	yes
CG43647	yes	yes	yes
CG30116	yes	yes	yes
btd	yes	yes	yes
CG31288	yes	yes	yes
OSCP1	yes	yes	yes
CG31115	yes	yes	yes
HnRNP-K	yes	yes	yes
GstD8	yes	yes	yes
mAChR-B	yes	yes	yes
lncRNA:CR45502	yes	yes	yes
Or69a	yes	yes	yes
Or2a	yes	yes	yes
Or67a	yes	yes	yes
GstE4	yes	yes	yes
Or13a	yes	yes	yes
unpg	yes	yes	yes
CG13659	yes	yes	yes
GstD11	yes	yes	yes
Sp1	yes	yes	yes
hec	yes	yes	yes
CG10874	yes	yes	yes
tut	yes	yes	yes
Or67b	yes	yes	yes
CG15522	yes	yes	yes
CG33288	yes	yes	yes
Cyp6a14	yes	yes	yes
Or85f	yes	yes	yes
slp2	yes	yes	yes
Cyp6a2	yes	yes	yes
Tsp42En	yes	yes	yes
ssp6	yes	yes	yes
CG42540	yes	yes	yes
CG4133	yes	yes	yes
amd	yes	yes	yes

Or43a	yes	yes	yes
Lgr1	yes	yes	yes
CG7786	yes	yes	yes
GABA-B-R1	yes	yes	yes
CG31689	yes	yes	yes
fid	yes	yes	yes
Tsp42Ef	yes	yes	yes
Nep1	yes	yes	yes
BBS1	yes	yes	yes
Osi7	yes	yes	yes
Snmp1	yes	yes	yes
Pde6	yes	yes	yes
Spn43Ab	yes	yes	yes
Jhedup	yes	yes	yes
Obp83ef	yes	yes	yes
Obp28a	yes	yes	yes
CG10035	yes	yes	yes
TrissinR	yes	yes	yes
beat-Ia	yes	yes	yes
CG40486	yes	yes	yes
Krn	yes	yes	yes
asRNA:CR44960	yes	yes	yes
ppk5	yes	yes	yes
acj6	yes	yes	yes
Or23a	yes	yes	yes
boss	yes	yes	yes
mthl3	yes	yes	yes
fabp	yes	yes	yes
CG4822	yes	yes	yes
CG10638	yes	yes	yes
CG10650	yes	yes	yes
eyg	yes	yes	yes
CG42766	yes	yes	yes
ATP8B	yes	yes	yes
peb	yes	yes	yes
Or67c	yes	yes	yes
CG5071	yes	yes	yes
Cluap1	yes	yes	yes
CG11852	yes	yes	yes
Or19b	yes	yes	yes
CG31760	yes	yes	yes
RSG7	yes	yes	yes
lncRNA:CR44541	yes	yes	yes
fa2h	yes	yes	yes
ppk6	yes	yes	yes
CG12911	yes	yes	yes
Or9a	yes	yes	yes
Rspf3	yes	yes	yes
Tsp42El	yes	yes	yes
CG30339	yes	yes	yes
SerT	yes	yes	yes
CG13928	yes	yes	yes
Osi8	yes	yes	yes
Or19a	yes	yes	yes

CG7342	yes	yes	yes
CG15169	yes	yes	yes
ADPS	yes	yes	yes
Or83c	yes	yes	yes
CG32668	yes	yes	yes
IM14	yes	yes	yes
Oseg2	yes	yes	yes
CG43646	yes	yes	yes
Oseg6	yes	yes	yes
SLC5A11	yes	yes	yes
Ugt49C1	yes	yes	yes
CG42806	yes	yes	yes
CG12885	yes	yes	yes
CG3556	yes	yes	yes
Gr98b	yes	yes	yes
spab	yes	yes	yes
CG43737	yes	yes	yes
CG8300	yes	yes	yes
Kif3C	yes	yes	yes
OS9	yes	yes	yes
Syngr	yes	yes	yes
CG6912	yes	yes	yes
CG12268	yes	yes	yes
Nos	yes	yes	yes
asRNA:CR45600	yes	yes	yes
CG17672	yes	yes	yes
kek1	yes	yes	yes
Obp69a	yes	yes	yes
Cpr47Ea	yes	yes	yes
asRNA:CR44065	yes	yes	yes
btv	yes	yes	yes
Sox15	yes	yes	yes
CG32547	yes	yes	yes
sns	yes	yes	yes
CG6337	yes	yes	yes
CG6495	yes	yes	yes
CG31125	yes	yes	yes
CG1695	yes	yes	yes
Or82a	yes	yes	yes
Adk1	yes	yes	yes
Dh31-R	yes	yes	yes
Lrp4	yes	yes	yes
CNT2	yes	yes	yes
CG17919	yes	yes	yes
Slob	yes	yes	yes
CG13437	yes	yes	yes
CG13606	yes	yes	yes
CG32137	yes	yes	yes
Gr93a	yes	yes	yes
gogo	yes	yes	yes
CG30356	yes	yes	yes
CG8641	yes	yes	yes
BBS4	yes	yes	yes
CG5758	yes	yes	yes

Frq1	yes	yes	yes
IFT54	yes	yes	yes
Epac	yes	yes	yes
CG13251	yes	yes	yes
Or65a	yes	yes	yes
ppk19	yes	yes	yes
CG32333	yes	yes	yes
Ugt36F1	yes	yes	yes
CG4199	yes	yes	yes
Or85a	yes	yes	yes
5-HT2A	yes	yes	yes
CG31674	yes	yes	yes
Tsp42A	yes	yes	yes
Nost	yes	yes	yes
CG15211	yes	yes	yes
lush	yes	yes	yes
lncRNA:CR43856	yes	yes	yes
Fer1	yes	yes	yes
Dh44-R1	yes	yes	yes
Or22a	yes	yes	yes
miple2	yes	yes	yes
lncRNA:CR43834	yes	yes	yes
BBS9	yes	yes	yes
Cyp6a8	yes	yes	yes
Or65b	yes	yes	yes
CG17716	yes	yes	yes
Root	yes	yes	yes
CG8170	yes	yes	yes
Mvl	yes	yes	yes
toy	yes	yes	yes
CG7236	yes	yes	yes
CG11841	yes	yes	yes
CG11000	yes	yes	yes
Cyp4p2	yes	yes	yes
Hf	yes	yes	yes
dpr9	yes	yes	yes
Tsp39D	yes	yes	yes
CG6282	yes	yes	yes
Rh4	yes	yes	yes
CG9427	yes	yes	yes
CG9689	yes	yes	yes
CG14692	yes	yes	yes
ea	yes	yes	yes
dysc	yes	yes	yes
CG34347	yes	yes	yes
Tsp29Fa	yes	yes	yes
CG13578	yes	yes	yes
Su(H)	yes	yes	yes
Agpat1	yes	yes	yes
Or98a	yes	yes	yes
Pdfr	yes	yes	yes
CR43170	yes	yes	yes
Ady43A	yes	yes	yes
CG18641	yes	yes	yes

halo	yes	yes	yes
CG17097	yes	yes	yes
bgm	yes	yes	yes
Ir60a	yes	yes	yes
Lsd-1	yes	yes	yes
CG17278	yes	yes	yes
CG10936	yes	yes	yes
ppk25	yes	yes	yes
Gld	yes	yes	yes
ZnT41F	yes	yes	yes
smog	yes	yes	yes
Or92a	yes	yes	yes
CG31345	yes	yes	yes
CG3349	yes	yes	yes
Or33a	yes	yes	yes
link	yes	yes	yes
CG1394	yes	yes	yes
bru2	yes	yes	yes
Rab26	yes	yes	yes
CG1358	yes	yes	yes
CG34309	yes	yes	yes
Sobp	yes	yes	yes
CG34445	yes	yes	yes
Vmat	yes	yes	yes
CG12672	yes	yes	yes
BBS8	yes	yes	yes
CG15209	yes	yes	yes
CG9498	yes	yes	yes
CG10336	yes	yes	yes
phyl	yes	yes	yes
CG15096	yes	yes	yes
CG8654	yes	yes	yes
CG16713	yes	yes	yes
CG32447	yes	yes	yes
CG13694	yes	yes	yes
CG1227	yes	yes	yes
mthl4	yes	yes	yes
CG6123	yes	yes	yes
Cby	yes	yes	yes
lncRNA:CR44525	yes	yes	yes
Cyp12e1	yes	yes	yes
CG6765	yes	yes	yes
NijC	yes	yes	yes
Gdap1	yes	yes	yes
ppk14	yes	yes	yes
trp	yes	yes	yes
CG14314	yes	yes	yes
IFT52	yes	yes	yes
CG40485	yes	yes	yes
Gat	yes	yes	yes
CG16798	yes	yes	yes
Cht5	yes	yes	yes
nerfin-2	yes	yes	yes
CG34219	yes	yes	yes

Pask	yes	yes	yes
CG12970	yes	yes	yes
Coop	yes	yes	yes
Acox57D-d	yes	yes	yes
dtr	yes	yes	yes
CG14153	yes	yes	yes
lncRNA:CR44317	yes	yes	yes
Mdr49	yes	yes	yes
CG10126	yes	yes	yes
CG13318	yes	yes	yes
CG6739	yes	yes	yes
pHCl-1	yes	yes	yes
CG14079	yes	yes	yes
CG30158	yes	yes	yes
N	yes	yes	yes
Or22b	yes	yes	yes
rad	yes	yes	yes
E5	yes	yes	yes
CG12896	yes	yes	no
Ugt49B1	yes	yes	yes
pyx	yes	yes	yes
Or88a	yes	yes	yes
GNBP3	yes	yes	yes
CG32206	yes	yes	yes
CG8861	yes	yes	yes
Drip	yes	yes	yes
Cyp6g1	yes	yes	yes
Tep2	yes	yes	yes
lncRNA:CR30009	yes	yes	yes
Wnt2	yes	yes	yes
tRNA:Leu-AAG-1-4	yes	yes	yes
Tsp42Er	yes	yes	yes
sNPF	yes	yes	yes
CG14247	yes	yes	yes
Or65c	yes	yes	yes
Cyp6a9	yes	yes	yes
lncRNA:CR45320	yes	yes	yes
Jheh3	yes	yes	yes
dila	yes	yes	yes
Cyp12d1-d	no	no	yes
CG14982	yes	yes	yes
ems	yes	yes	yes
H15	yes	yes	yes
CG11854	yes	yes	yes
Ude	yes	yes	yes
hmw	yes	yes	yes
krimp	yes	yes	yes
CG32809	yes	yes	yes
to	yes	yes	yes
CG43324	yes	yes	yes
blanks	yes	yes	yes
CG32104	yes	yes	yes
CG11356	yes	yes	yes
prom	yes	yes	yes

CG12910	yes	yes	yes
CG11453	yes	yes	yes
CG31221	yes	yes	yes
CG9368	yes	yes	yes
lncRNA:CR32111	yes	yes	yes
CG13558	yes	yes	yes
CAHbeta	yes	yes	yes
brk	yes	yes	yes
Lcp65Ag2	yes	yes	yes
CG11191	yes	yes	yes
TrpA1	yes	yes	yes
Kal1	yes	yes	yes
asRNA:CR46136	yes	yes	yes
Cyp9b1	yes	yes	yes
CG15186	yes	yes	yes
Cpr50Ca	yes	yes	yes
Tsp5D	yes	yes	yes
ry	yes	yes	yes
lncRNA:CR43498	yes	yes	yes
CG3746	yes	yes	yes
GstE1	yes	yes	yes
Gr64a	yes	yes	yes
CG13202	yes	yes	yes
aurB	yes	yes	yes
BBS5	yes	yes	yes
Spn42De	yes	yes	yes
lncRNA:CR44285	yes	yes	yes
CG5890	yes	yes	yes
PsGEF	yes	yes	yes
7B2	yes	yes	yes
Ccp84Ac	yes	yes	yes
lncRNA:CR46218	yes	yes	yes
CG33143	yes	yes	yes
CG10738	yes	yes	yes
CG13793	yes	yes	yes
CG18234	yes	yes	yes
Tdc1	yes	yes	yes
CG32006	yes	yes	yes
Cpr92F	yes	yes	yes
CG6845	yes	yes	yes
CG2145	yes	yes	yes
cDIP	yes	yes	yes
CG13386	yes	yes	yes
CG11550	yes	yes	yes
CG15414	yes	yes	yes
CG14693	yes	yes	yes
ham	yes	yes	yes
kek4	yes	yes	yes
Gr43a	yes	yes	yes
CG15701	yes	yes	yes
Mal-B1	yes	yes	yes
lncRNA:CR45504	yes	yes	no
Cpr64Ac	yes	yes	yes
Orct2	yes	yes	yes

CG14024	yes	yes	yes
lncRNA:CR45259	yes	yes	yes
lncRNA:CR44320	yes	yes	yes
CG9568	no	yes	yes
Unc-115b	no	no	yes
CG5535	yes	yes	yes
CG34136	yes	yes	yes
Cyp318a1	yes	yes	yes
CG8539	yes	yes	yes
CG7016	yes	yes	yes
Mur89F	yes	yes	yes
CG10116	yes	yes	yes
lncRNA:CR43700	yes	yes	yes
CG7966	yes	yes	yes
CG6071	yes	yes	yes
Idgf6	yes	yes	yes
CR43086	yes	yes	yes
CG32260	yes	yes	yes
CR43214	yes	yes	no
SIFaR	yes	yes	yes
Sema5c	yes	yes	yes
lncRNA:CR45956	yes	yes	yes
snoRNA:2R:9445410	yes	yes	yes
B9d1	yes	yes	yes
CG32444	yes	yes	yes
CarT	yes	yes	yes
Ror	yes	yes	yes
Or33b	yes	yes	yes
CG14715	yes	yes	yes
CG4194	yes	yes	yes
CG14694	yes	yes	yes
Tep4	yes	yes	yes
CG1213	yes	yes	yes
PGRP-SC1a	yes	yes	no
IFT46	yes	yes	yes
CG13250	yes	yes	yes
Obp83cd	yes	yes	yes
CG32271	yes	yes	yes
CG43689	yes	yes	yes
CG18467	yes	yes	yes
CG14321	yes	yes	yes
Acbp6	yes	yes	yes
Orct	yes	yes	yes
CG17350	yes	yes	yes
CG4409	yes	yes	yes
spirit	yes	yes	yes
CG13707	yes	yes	yes
Hsp23	yes	yes	yes
CG2065	yes	yes	yes
LanA	yes	yes	yes
dsb	yes	yes	yes
CG8492	yes	yes	yes
CG2816	yes	yes	yes
jhamt	yes	yes	yes

CG31097	yes	yes	yes
Obp8a	yes	yes	yes
E(spl)m2-BFM	yes	yes	yes
CCHa2-R	yes	yes	yes
Lip3	yes	yes	yes
CG7860	yes	yes	yes
CG1504	yes	yes	yes
TwdlT	yes	yes	yes
Cht7	yes	yes	yes
CR43215	yes	no	no
CG10359	yes	yes	yes
Cyp6a17	yes	yes	yes
alpha-Est2	yes	yes	yes
CG7848	yes	yes	yes
CG34227	yes	yes	yes
CCHa1-R	yes	yes	yes
ringer	yes	yes	yes
dy	yes	yes	yes
Spn42Dd	yes	yes	yes
CG33679	yes	yes	yes
lncRNA:CR43650	yes	yes	yes
CG43795	yes	yes	yes
fusl	yes	yes	yes
CG43267	yes	yes	yes
CG4210	yes	yes	yes
tRNA:Met-CAT-1-5	yes	yes	yes
CG14551	yes	yes	yes
PGRP-SC1b	yes	yes	no
Tsp	yes	yes	yes
Con	yes	yes	yes
asRNA:CR44137	yes	yes	yes
Osi17	yes	yes	yes
CG15201	yes	yes	yes
CG11425	yes	yes	yes
CG14301	yes	yes	yes
snoRNA:Psi28S-1153	yes	yes	yes
CG17732	yes	yes	yes
Cyp6a19	yes	yes	yes
lncRNA:CR44646	yes	yes	yes
lncRNA:CR45280	yes	yes	yes
comm3	yes	yes	yes
CG34355	yes	yes	yes
lin-28	yes	yes	yes
snoRNA:2R:9445205	yes	yes	yes
Dsk	yes	yes	yes
CG33993	yes	yes	yes
CG10730	yes	yes	yes
dati	yes	yes	yes
vvl	yes	yes	yes
Rgk2	yes	yes	yes
Marf1	yes	yes	yes
E(spl)m7-HLH	yes	yes	yes
Spn42Db	yes	yes	yes
CG12869	yes	yes	yes

asRNA:CR45924	yes	yes	yes
CG8665	yes	yes	yes
lncRNA:CR45290	yes	yes	yes
CG7201	yes	yes	yes
CG5194	no	yes	no
GstD10	yes	yes	yes
hdm	yes	yes	yes
CG43236	yes	yes	yes
lncRNA:CR46245	yes	yes	yes
lncRNA:CR42646	yes	yes	yes
Spn88Eb	yes	yes	yes
CG42329	yes	yes	yes
CG30259	yes	yes	yes
Tsf1	yes	yes	yes
CG34456	yes	yes	yes
Def	yes	yes	yes
CG18635	yes	yes	yes
CG9328	yes	yes	yes
Odc1	yes	yes	yes
yellow-e	yes	yes	yes
CG4927	yes	yes	yes
mthl15	yes	yes	yes
DIP-lambda	yes	yes	yes
snRNA:U2:14B	yes	yes	no
iotaTry	yes	yes	yes
NPFR	yes	yes	yes
jtb	yes	yes	yes
CG8248	yes	yes	yes
lncRNA:CR44043	yes	yes	yes
CecC	yes	yes	yes
lncRNA:CR44091	yes	yes	yes
Scp2	yes	yes	yes
CG31644	yes	yes	yes
CG10479	yes	yes	yes
asRNA:CR45822	yes	yes	yes
GstE7	yes	yes	yes
Kaz1-ORFB	yes	yes	yes
CG3769	yes	yes	yes
CG31189	yes	yes	yes
CG30383	yes	yes	yes
CG9519	yes	yes	yes
FMRFaR	yes	yes	yes
lncRNA:CR42715	yes	yes	yes
CG13954	yes	yes	yes
CG13616	yes	yes	yes
CG14219	yes	yes	yes
qin	yes	yes	yes
antr	yes	yes	yes
asRNA:CR43476	yes	yes	yes
ppk7	yes	yes	yes
mus301	yes	yes	yes
CG13500	yes	yes	yes
lncRNA:CR46041	yes	yes	yes
lncRNA:CR44833	yes	yes	yes

Cyp9h1	yes	yes	yes
CG9649	yes	yes	yes
Tsp42Eq	yes	yes	yes
Skeletor	yes	yes	yes
snRNA:U2:38ABA	yes	no	no
CG30265	yes	yes	yes
CG6870	yes	yes	yes
CG13101	yes	yes	yes
GstS1	yes	yes	yes
FarO	yes	yes	yes
Sgsh	yes	yes	yes
Cyp6d2	no	yes	no
CG15594	yes	yes	yes
asRNA:CR46033	yes	yes	yes
Sirup	yes	yes	yes
Ugt305A1	yes	yes	yes
Vajk4	yes	yes	yes
mmd	yes	yes	yes
Obp99d	yes	yes	yes
CG12880	yes	yes	yes
CG15529	yes	yes	yes
Cyp305a1	yes	yes	yes
Hml	yes	yes	yes
GstZ2	yes	yes	yes
lncRNA:CR45181	yes	yes	yes
asRNA:CR33945	yes	yes	yes
CG43110	yes	yes	yes
MFS1	yes	yes	yes
Ser	yes	yes	yes
Qsox2	yes	yes	yes
rdhB	yes	yes	yes
CG34166	yes	yes	yes
CG7763	yes	yes	yes
tsl	yes	yes	yes
lncRNA:CR44298	yes	yes	yes
CR33294	yes	yes	yes
Cpr62Bc	yes	yes	yes
Eglp2	yes	yes	yes
Cyp6a23	yes	yes	yes
CG15368	yes	yes	yes
Sodh-2	yes	yes	yes
nxf2	yes	yes	yes
lncRNA:CR44926	no	yes	yes
Acbp4	yes	yes	yes
CG42826	yes	yes	yes
CG14615	yes	yes	yes
SecCl	yes	yes	yes
lncRNA:CR46254	yes	yes	yes
mGluR	yes	yes	yes
Pxd	yes	yes	yes
CG3777	yes	yes	yes
CG2157	yes	yes	yes
CG12224	yes	yes	yes
Zip89B	yes	yes	yes

CG42346	yes	yes	yes
CrzR	yes	yes	yes
Hr83	yes	yes	yes
CG11378	yes	yes	yes
CG30059	no	no	yes
CG9815	yes	yes	yes
sad	yes	yes	yes
CR43364	yes	yes	yes
Lsp1beta	yes	yes	yes
CG13082	yes	yes	yes
l(2)34Fc	yes	yes	yes
Trpgamma	yes	yes	yes
CG4752	yes	yes	yes
CG11362	yes	yes	yes
alpha-Est4aPsi	yes	yes	yes
lncRNA:CR45082	yes	yes	yes
CG15731	yes	yes	yes
CG30156	yes	yes	yes
CG32040	yes	yes	yes
CG43165	yes	yes	yes
CG30090	yes	yes	yes
Lox1	yes	yes	yes
Dl	yes	yes	yes
asRNA:CR43730	yes	yes	yes
CG30203	yes	yes	yes
lncRNA:CR45985	yes	yes	yes
sage	yes	yes	yes
lncRNA:CR43627	yes	yes	yes
CG33125	yes	yes	yes
AstC-R2	yes	yes	yes
CR43482	yes	yes	no
CG42402	yes	yes	yes
CG7255	yes	yes	yes
CG14105	yes	yes	yes
CG7997	yes	yes	yes
Ir94b	yes	yes	yes
CG9593	yes	yes	yes
CG30082	yes	yes	yes
l(2)k05911	yes	yes	yes
lncRNA:CR33942	yes	yes	yes
CG34375	yes	yes	yes
Cyp313a4	yes	yes	yes
CG30076	yes	yes	yes
Ets96B	yes	yes	yes
form3	yes	yes	yes
Spn85F	yes	yes	yes
fw	yes	yes	yes
CG10352	yes	yes	yes
fd3F	no	no	yes
CG3277	yes	yes	yes
CG42822	yes	yes	yes
lncRNA:CR46061	yes	yes	yes
CG42821	yes	yes	yes
CG14052	yes	yes	yes

nerfin-1	yes	yes	yes
fln	yes	yes	yes
CG18327	yes	yes	yes
CG31445	yes	yes	yes
Gnmt	yes	yes	yes
CG30016	yes	yes	yes
Wnt10	yes	yes	yes
CG34002	yes	yes	yes
Had2	yes	yes	yes
fax	yes	yes	yes
CG41284	no	yes	no
CG4766	yes	yes	yes
ATPsynepsilonL	yes	yes	yes
CG43233	yes	yes	yes
CG7365	yes	yes	yes
CG17777	yes	yes	yes
CG32052	yes	yes	yes
lncRNA:CR45615	yes	yes	yes
CG32512	yes	yes	yes
CG5639	yes	yes	yes
CG6967	yes	yes	yes
CG18233	yes	yes	yes
mol	yes	yes	yes
lncRNA:CR45115	yes	yes	yes
lncRNA:CR44206	yes	yes	yes
CG14391	yes	yes	yes
CG4000	yes	yes	yes
CG33475	yes	yes	yes
Acp1	no	no	yes
hpRNA:CR32207	yes	yes	yes
CG3397	yes	yes	yes
CG31493	yes	yes	yes
CR43687	yes	yes	yes
CG17819	yes	yes	yes
CG13458	yes	yes	yes
Gadd45	yes	yes	yes
CG34446	yes	yes	yes
tfc	yes	yes	yes
CG31028	yes	yes	yes
LpR1	yes	yes	yes
lncRNA:CR44283	yes	yes	yes
CR43087	yes	yes	no
CR42548	yes	yes	yes
Mal-A5	yes	yes	yes
lncRNA:CR45054	yes	yes	yes
CG44140	no	yes	no
Dtg	yes	yes	yes
CG33178	yes	yes	yes
Or33c	yes	yes	yes
SPH93	yes	yes	yes
CG4783	yes	yes	yes
CG33946	yes	yes	yes
Gr98c	yes	yes	yes
Tb	yes	yes	yes

CG9733	yes	yes	yes
CG30046	yes	yes	yes
lncRNA:CR43651	yes	yes	yes
SoYb	yes	yes	yes
CG13970	yes	yes	yes
CG14400	yes	yes	yes
CG30424	yes	yes	yes
Cpr66D	yes	yes	yes
CG42747	yes	yes	yes
spd-2	no	no	yes
lncRNA:CR44514	yes	yes	yes
ninaD	yes	yes	yes
lncRNA:CR44912	yes	yes	yes
Tim17a2	yes	yes	yes
Acbp3	yes	yes	yes
CG44153	yes	yes	yes
CG7881	yes	yes	yes
CG9701	yes	yes	yes
CR14033	yes	yes	yes
CG11951	yes	yes	yes
CG9272	yes	yes	yes
CG30340	yes	yes	yes
lncRNA:CR44923	yes	yes	yes
CG3588	yes	yes	yes
CG43896	yes	yes	yes
lncRNA:CR46239	yes	yes	yes
CG13077	yes	yes	yes
Gr98d	yes	yes	yes
CG43338	yes	yes	yes
grim	yes	yes	yes
CR43696	yes	no	no
CG43120	yes	yes	yes
eIF4E6	yes	yes	yes
CG6688	yes	yes	yes
CG42711	yes	yes	yes
lncRNA:CR44779	yes	yes	yes
DNApol-epsilon58	yes	no	yes
CG1850	yes	yes	yes
Zwilch	yes	yes	yes
NimB2	yes	yes	yes
dysf	yes	yes	yes
Lkr	yes	yes	yes
asRNA:CR44600	yes	yes	yes
IFT20	yes	yes	yes
CG1673	yes	yes	yes
Csas	yes	yes	yes
CG31279	yes	yes	yes
CG7568	yes	yes	yes
CG6927	yes	yes	yes
CG32816	yes	yes	yes
CG34109	yes	yes	yes
eater	yes	yes	yes
lncRNA:CR45674	yes	yes	yes
CG4757	yes	yes	yes

CR43671	yes	yes	yes
LManI	yes	yes	yes
CG32284	yes	yes	yes
CG13921	yes	no	no
lncRNA:CR44166	yes	yes	yes
PGRP-SD	yes	yes	yes
CG14205	yes	yes	yes
laza	yes	yes	yes
RecQ4	yes	yes	yes
CG15909	yes	yes	yes
CG33914	yes	yes	yes
CG9114	yes	yes	yes
CG4998	yes	yes	yes
CG1607	yes	yes	yes
CG1545	yes	yes	yes
hgo	yes	yes	yes
pirk	yes	yes	yes
CG7330	yes	yes	yes
CG30062	yes	yes	yes
CG4860	yes	yes	yes
Ir41a	yes	yes	yes
wat	yes	yes	yes
Ugt317A1	yes	yes	yes
Sr-CI	yes	yes	yes
lncRNA:CR46015	yes	yes	yes
CG34454	yes	yes	yes
Osi24	yes	yes	yes
CG14854	yes	yes	yes
Ant2	yes	yes	yes
rn	yes	yes	yes
sstn	yes	yes	yes
CG18063	yes	yes	yes
Adh	yes	no	yes
CG13857	yes	yes	yes
CG8046	yes	no	yes

results/tables/tbl129_cantonAmosMutation_chonky.html

3.11.1.1 Top Tens

Genes with top 10 most significant changes

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

Table 130. Top Ten Most Significantly (padj<0.01) Differentially Expressed genes between garden-variety vs amos mutants

rank	multi				rando			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	CG43861	0.19	-3.261	2.46×10^{-137}	CG43861	0.19	-3.261	1.99×10^{-137}
2	CG7509	0.20	5.110	6.06×10^{-136}	CG7509	0.20	5.110	1.59×10^{-136}
3	CG9555	0.32	-5.309	1.77×10^{-128}	CG9555	0.32	-5.304	5.68×10^{-128}
4	Gr64f	0.49	-5.110	3.07×10^{-128}	Gr64f	0.49	-5.109	1.65×10^{-128}

5	CG42284	0.88	-1.815	7.06×10^{-128}	CG42284	0.88	-1.814	6.65×10^{-128}
6	CG34342	2.06	-3.229	2.60×10^{-118}	CG34342	2.06	-3.227	8.73×10^{-118}
7	CG7896	0.12	-3.755	1.25×10^{-115}	CG7896	0.12	-3.755	5.13×10^{-115}
8	Or47b	1.46	-7.076	3.61×10^{-109}	Or47b	1.46	-7.072	8.71×10^{-109}
9	Galphaf	0.40	-3.813	1.12×10^{-104}	Galphaf	0.40	-3.813	1.81×10^{-104}
10	Or7a	0.46	-6.748	4.22×10^{-103}	Or7a	0.46	-6.748	8.26×10^{-103}

Top 10 genes with biggest (significant) effect sizes

Table 131. Top Ten Largest Magnitude Fold Changes which garden-variety vs amos mutants

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	IM14	1.09	12.017	7.88×10^{-18}	IM14	1.09	12.017
2	Tsp42A	0.03	-9.722	7.66×10^{-12}	Tsp42A	0.03	-9.722
3	CG17097	0.01	9.638	9.13×10^{-8}	CG17097	0.01	9.638
4	lncRNA:CR43834	0.17	-9.478	2.14×10^{-11}	lncRNA:CR43834	0.17	-9.478
5	Or67c	0.26	-9.027	8.92×10^{-25}	Or67c	0.26	-9.027
6	Or67d	0.81	-8.791	2.84×10^{-63}	Or67d	0.81	-8.791
7	Or65a	0.56	-8.279	3.71×10^{-14}	Or65a	0.56	-8.279
8	Or23a	0.16	-8.246	2.34×10^{-27}	Or23a	0.16	-8.246
9	Or67a	0.28	-8.147	1.23×10^{-43}	Or67a	0.28	-8.147
10	Or56a	0.49	-8.107	2.53×10^{-63}	Or56a	0.49	-8.107

Top 10 highest expressed genes with significant change

Table 132. Top Ten Highest Expressed Genes with Significant (padj) Difference garden-variety vs amos mutants

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	CG6409	52.71	0.526	8.23×10^{-3}	CG6409	52.69	0.526
2	Drsl5	33.70	0.500	3.86×10^{-3}	Drsl5	33.69	0.501
3	CG14661	30.39	-0.701	1.64×10^{-3}	CG14661	30.38	-0.701
4	Obp28a	29.26	-2.920	1.10×10^{-34}	Obp28a	29.25	-2.920
5	antdh	29.01	-0.534	7.85×10^{-3}	antdh	29.00	-0.541
6	OS9	28.84	-4.120	2.75×10^{-23}	OS9	28.83	-4.120
7	Jhedup	25.44	-5.016	4.19×10^{-33}	Jhedup	25.44	-5.016
8	Obp69a	20.49	-6.119	1.59×10^{-19}	Obp69a	20.48	-6.118
9	Obp59a	20.03	0.989	2.81×10^{-9}	Obp59a	20.03	0.990
10	GstE4	19.51	-4.979	1.16×10^{-46}	GstE4	19.51	-4.979

3.11.1.2 Gene Ontology Enrichment

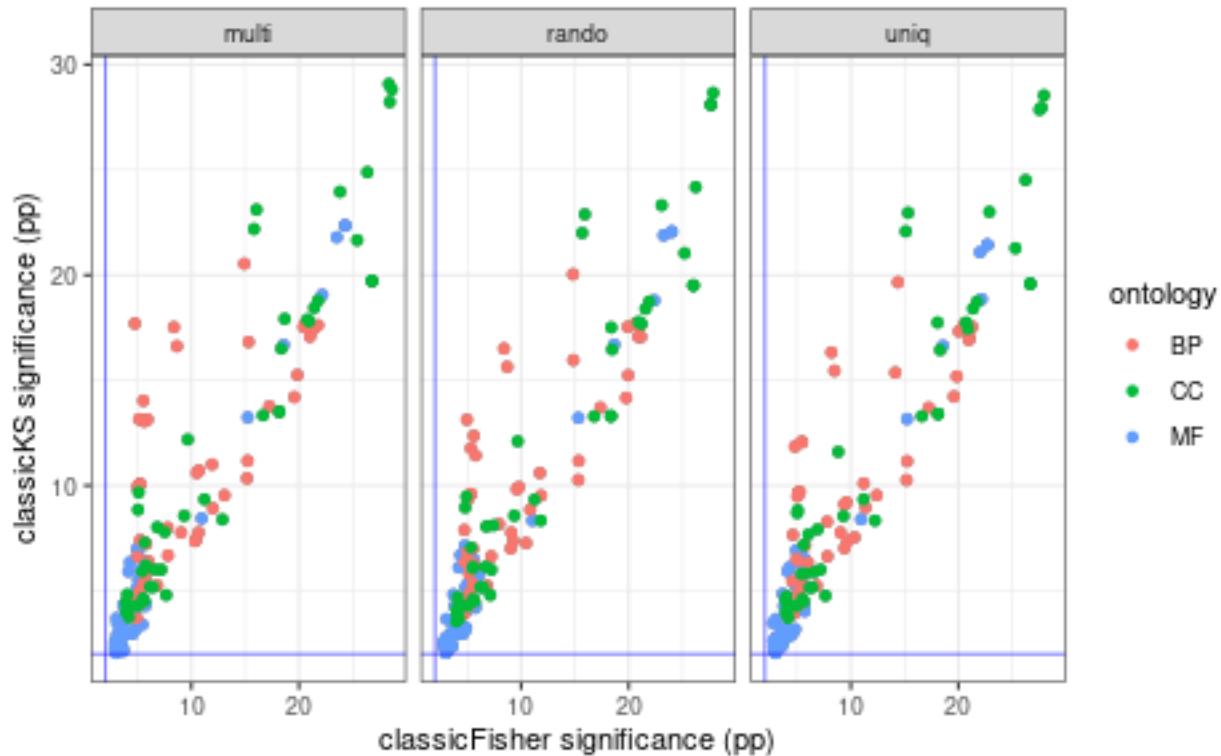
Genes were analyzed for GO Term Enrichment using topGO, using Fisher's test applied to those whose expression difference passed a significance threshold ($p < 0.01$), and applying the Kolmogorov-Smirnov test using p-values as scores.

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

terms?

Correlation between significance values for the two tests

Figure 115. Scatterplot of GO Term Enrichment Significance for Two Tests (garden-variety vs amos mutant)



```
## png  
## 2
```

```
molecular transducer activity (GO:0060089)  
tetrapyrrole binding (GO:0046906)  
sensory perception (GO:0007600)  
system process (GO:0003008)  
detection of stimulus (GO:0051606)  
obsolete cell projection part (GO:0044463)  
obsolete plasma membrane bounded cell projection part (GO:0120038)  
obsolete neuron part (GO:0097458)  
obsolete membrane part (GO:0044425)
```

Table 133. Enriched GO Terms among Significantly Differentially Expressed Genes garden-variety vs amos mutants; uniq only; top 10 most significant per category

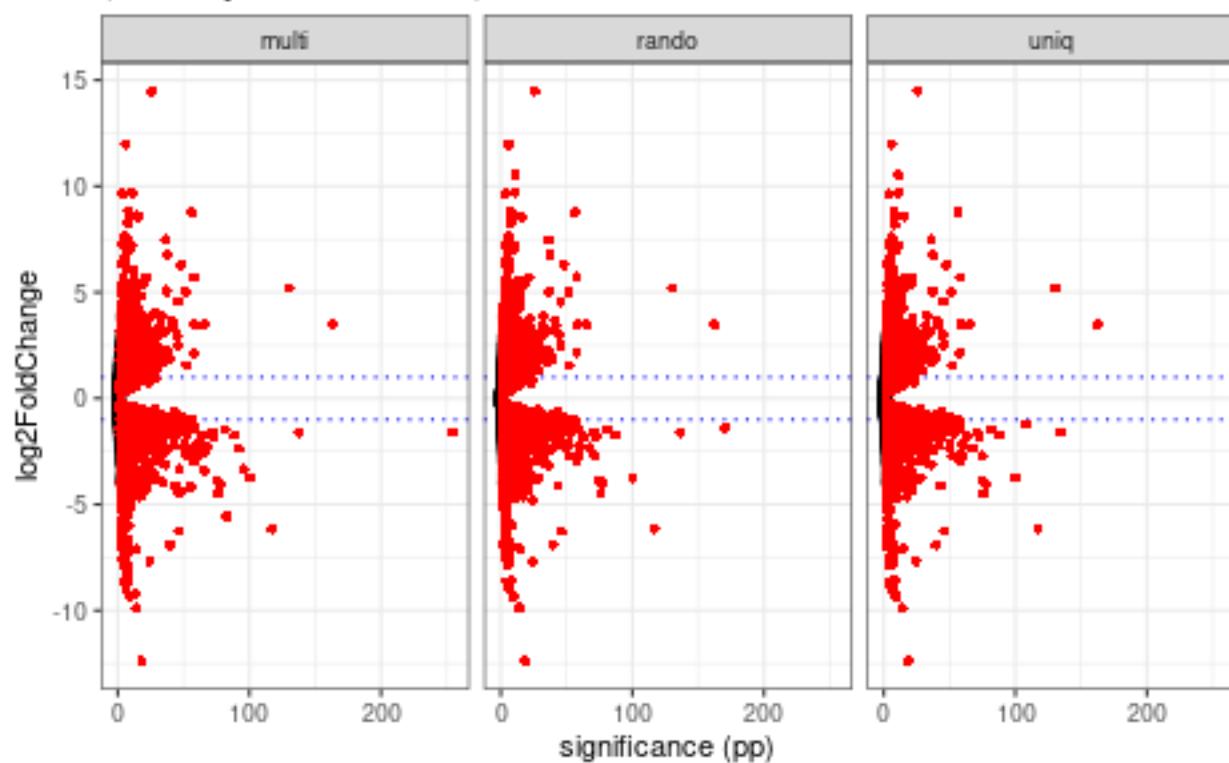
GO Term	Description	p-value	
		Fisher	K-S
MF			
GO:0038023	signaling receptor activity	2.00×10^{-23}	3.70×10^{-22}

GO:0060089	NA	2.00×10^{-23}	3.70×10^{-22}
GO:0004888	transmembrane signaling receptor activity	6.90×10^{-23}	1.40×10^{-19}
GO:0004871	NA	9.70×10^{-23}	8.20×10^{-22}
GO:0004984	olfactory receptor activity	2.70×10^{-19}	2.30×10^{-17}
GO:0005549	odorant binding	5.80×10^{-16}	6.90×10^{-14}
GO:0004930	G protein-coupled receptor activity	1.10×10^{-11}	4.00×10^{-9}
GO:0016798	hydrolase activity, acting on glycosyl bonds	2.00×10^{-6}	9.30×10^{-5}
GO:0015075	ion transmembrane transporter activity	2.10×10^{-6}	2.90×10^{-7}
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	2.30×10^{-6}	5.70×10^{-5}
<hr/>			
BP			
GO:0007600	NA	5.20×10^{-22}	2.90×10^{-18}
GO:0050907	detection of chemical stimulus involved in sensory perception	9.20×10^{-22}	9.60×10^{-18}
GO:0003008	NA	1.10×10^{-21}	1.20×10^{-17}
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	2.80×10^{-21}	2.20×10^{-18}
GO:0050877	nervous system process	8.80×10^{-21}	4.70×10^{-18}
GO:0050906	detection of stimulus involved in sensory perception	1.40×10^{-20}	6.50×10^{-16}
GO:0007606	sensory perception of chemical stimulus	2.60×10^{-20}	5.90×10^{-15}
GO:0007608	sensory perception of smell	5.60×10^{-18}	2.00×10^{-14}
GO:0009593	detection of chemical stimulus	5.80×10^{-16}	7.00×10^{-12}
GO:0051606	NA	6.60×10^{-16}	5.50×10^{-11}
<hr/>			
CC			
GO:0031224	intrinsic component of membrane	1.10×10^{-28}	2.90×10^{-29}
GO:0016021	integral component of membrane	1.80×10^{-28}	1.10×10^{-28}
GO:0016020	membrane	2.80×10^{-28}	1.40×10^{-28}
GO:0044463	NA	2.10×10^{-27}	2.60×10^{-20}
GO:0120038	NA	2.10×10^{-27}	2.60×10^{-20}
GO:0097458	NA	5.60×10^{-27}	3.10×10^{-25}
GO:0043005	neuron projection	4.90×10^{-26}	5.30×10^{-22}
GO:0044425	NA	1.30×10^{-23}	1.00×10^{-23}
GO:0032589	neuron projection membrane	1.80×10^{-22}	1.80×10^{-19}
GO:0031256	leading edge membrane	4.30×10^{-22}	3.90×10^{-19}

3.11.2 Differences in “garden Variety” genomes: w vs CantonS

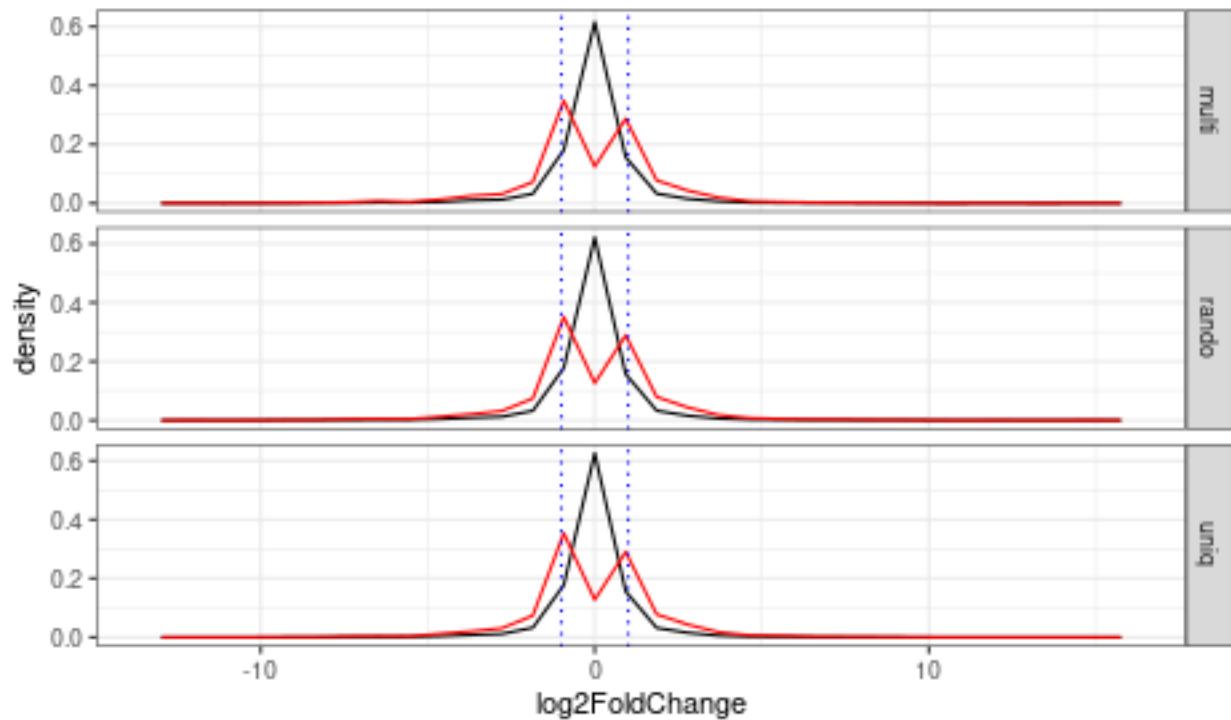
Here is a volcano plot for the three alignment strategies, with significance on the horizontal axis and log2 fold change on the vertical. Significant ($\text{padj} < 0.01$) differences are highlighted in red. Dashed blue guidelines mark a log2 fold change of $+/-1$ (ie, a difference in expression of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

Figure 116. Volcano Plot: Fold Change vs. Significance
(white-eyed vs CantonS)



```
## png
## 2
```

**Figure 117. histogram of fold change
with significant($p_{adj} < 0.01$) changes highlighted in red
(from white-eyed to CantonS)**



```
## png
## 2
```

Of the 12438 genes with significance scores available, 4135 have an adjusted $p < 0.01$ (33.2448947 %). This is less than when this was run as a single-factor model, but still seems like a lot for flies which are being treated as equivalent genetic backgrounds. On the other hand, the distribution of significant effect sizes in bimodal in this case, whereas before it was unimodal.

From the volcano plots, we can pull out genes with large (ie, a fold change greater than 2 or less than -2), significant (ie, $p_{adj} < 0.01$) changes. There were 1823 such genes, mostly shared across alignment strategy:

**Table 134. Genes with Large ($|2| <$ fold change), Significant ($p_{adj} < 0.01$) Changes
from white-eyed to CantonS**

	multi	rando	uniq
AGO2	yes	yes	yes
Acbp2	yes	yes	yes
CG31224	yes	yes	yes
CG7900	yes	yes	yes
DIP-alpha	yes	yes	yes
Cht10	yes	yes	yes
SdicA	yes	yes	no
Sdic4	yes	yes	no
CG10631	yes	yes	yes
Sdic1	yes	yes	no

Hipk	yes	yes	yes
lncRNA:CR43314	yes	yes	yes
sif	yes	yes	yes
prom	yes	yes	yes
Dscam4	yes	yes	yes
dtn	yes	yes	yes
axo	yes	yes	yes
Dab	yes	yes	yes
CG31663	yes	yes	yes
Sdic3	yes	yes	no
Cyp6g2	yes	yes	yes
SdicB	yes	yes	no
Ten-a	yes	yes	yes
toy	yes	yes	yes
IM33	yes	yes	yes
Shaw	yes	yes	yes
Mnt	yes	yes	yes
CG10936	yes	yes	yes
CG42795	yes	yes	yes
Cyp6a20	yes	yes	yes
nocte	yes	yes	yes
Pdfr	yes	yes	yes
CR43216	yes	yes	no
Cka	yes	yes	yes
CG30069	yes	yes	yes
Sec16	yes	yes	yes
CG43085	yes	yes	yes
stan	yes	yes	yes
Ets21C	yes	yes	yes
SelR	yes	yes	yes
CG14073	yes	yes	yes
CG2747	yes	yes	yes
Mec2	yes	yes	yes
Cyp6v1	yes	yes	yes
hiw	yes	yes	yes
CR43213	yes	yes	no
HERC2	yes	yes	yes
CG14400	yes	yes	yes
peb	yes	yes	yes
CR43211	yes	yes	no
TpnC25D	yes	yes	yes
CR33496	yes	yes	no
CG30271	yes	yes	yes
MtnA	yes	yes	yes
CG43647	yes	yes	yes
eIF4G2	yes	yes	yes
Cyp6a8	yes	yes	yes
Hers	yes	yes	yes
norpA	yes	yes	yes
CG43068	yes	yes	yes
Ncoa6	yes	yes	yes
CG13594	yes	yes	yes
Cpr64Ac	yes	yes	yes
mgl	yes	yes	yes

SdicC	yes	yes	no
Obp83ef	yes	yes	yes
CG34401	yes	yes	yes
CG3604	yes	yes	yes
rudhira	yes	yes	yes
CG11122	yes	yes	yes
HUWE1	yes	yes	yes
fs(1)h	yes	yes	yes
Ac13E	yes	yes	yes
CG16713	yes	yes	yes
GluRIIA	yes	yes	yes
zfh2	yes	yes	yes
Swim	yes	yes	yes
CG5937	yes	yes	yes
5-HT2B	yes	yes	yes
pyd3	yes	yes	yes
SLO2	yes	yes	yes
Drip	yes	yes	yes
CG43149	yes	yes	yes
ITP	yes	yes	yes
Bsg	yes	yes	yes
CG10031	yes	yes	yes
Rab3-GEF	yes	yes	yes
Nrg	yes	yes	yes
Ubr3	yes	yes	yes
E(spl)mbeta-HLH	yes	yes	yes
bon	yes	yes	yes
lncRNA:CR32218	yes	yes	yes
BomT3	yes	yes	yes
mtgo	yes	yes	yes
rno	yes	yes	yes
Unr	yes	yes	yes
egh	yes	yes	yes
Oseg6	yes	yes	yes
asRNA:CR43933	yes	yes	no
Ubr1	yes	yes	yes
jeb	yes	yes	yes
CG3434	yes	yes	yes
sli	yes	yes	yes
CR32821	yes	yes	no
Ubi-p63E	yes	yes	no
ppk25	yes	yes	yes
Dys	yes	yes	yes
Apc	yes	yes	yes
CAH9	yes	yes	yes
CG44774	yes	yes	yes
pinta	yes	yes	yes
lncRNA:noe	yes	yes	yes
CG9727	yes	yes	yes
Ctr9	yes	yes	yes
CG14762	yes	yes	yes
IA-2	yes	yes	yes
skd	yes	yes	yes
CG30116	yes	yes	yes

Mal-B1	yes	yes	yes
Cyp6a2	yes	yes	yes
CG4297	yes	yes	yes
Nost	yes	yes	yes
CG31635	yes	yes	yes
GstE1	yes	yes	yes
CG7896	yes	yes	yes
Rbcn-3A	yes	yes	yes
CG8369	yes	yes	yes
lncRNA:CR44320	yes	yes	yes
CG2157	yes	yes	yes
NijC	yes	yes	yes
Ank2	yes	yes	yes
CG18265	yes	yes	yes
pcx	yes	yes	yes
Lrrk	yes	yes	yes
Pde9	yes	yes	yes
CR43377	yes	yes	no
CG15201	yes	yes	yes
CG11873	yes	yes	yes
scrib	yes	yes	yes
CG17751	yes	yes	yes
CG9010	yes	yes	yes
CrzR	yes	yes	yes
CG12512	yes	yes	yes
PMCA	yes	yes	yes
Meltrin	yes	yes	yes
CG4049	yes	yes	yes
CG43155	yes	yes	yes
mask	yes	yes	yes
CG46302	yes	yes	yes
CG46025	yes	yes	yes
tou	yes	yes	yes
nahoda	yes	yes	yes
spn-B	yes	yes	yes
CG40228	yes	yes	yes
Msp300	yes	yes	yes
CG5707	yes	yes	yes
IM14	yes	yes	yes
CG17032	yes	yes	yes
Gas8	yes	yes	yes
Syx1A	yes	yes	yes
CG43737	yes	yes	yes
Cyp4d21	yes	yes	yes
raptor	yes	yes	yes
crb	yes	yes	yes
teq	yes	yes	yes
CR43381	yes	yes	no
CG18336	yes	yes	yes
TrpA1	yes	yes	yes
Usp32	yes	yes	yes
Teh3	yes	yes	yes
Ptp4E	yes	yes	yes
brp	yes	yes	yes

LRR	yes	yes	yes
CG9447	yes	yes	yes
CG33298	yes	yes	yes
sano	yes	yes	yes
CadN	yes	yes	yes
CG42392	yes	yes	yes
alpha-Man-IIb	yes	yes	yes
BBS1	yes	yes	yes
CG43658	yes	yes	yes
CR43215	yes	yes	no
CG42342	yes	yes	yes
Cpr47Ee	yes	yes	yes
Myo10A	yes	yes	yes
CG34456	yes	yes	yes
CG13855	yes	yes	yes
CG8858	yes	yes	yes
Uhg5	yes	yes	yes
CG6347	yes	yes	yes
dysc	yes	yes	yes
CG31140	yes	yes	yes
kmr	yes	yes	yes
asRNA:CR43426	yes	yes	yes
Appl	yes	yes	yes
CG11852	yes	yes	yes
mus201	yes	yes	yes
Vps13	yes	yes	yes
Ir21a	yes	yes	yes
Rbp6	yes	yes	yes
Ptp10D	yes	yes	yes
futsch	yes	yes	yes
Octalpha2R	yes	yes	yes
CG3009	yes	yes	yes
CR43214	yes	yes	no
Megf8	yes	yes	yes
CG12896	yes	yes	yes
CG3552	yes	yes	yes
KFase	yes	yes	yes
Nf1	yes	yes	yes
rg	yes	yes	yes
Rpb10	yes	yes	yes
CG5162	yes	yes	yes
lncRNA:CR46260	yes	yes	yes
CG7126	yes	yes	yes
lncRNA:CR43650	yes	yes	yes
CG2841	yes	yes	yes
CG10011	yes	yes	yes
lace	yes	yes	yes
Zmynd10	yes	yes	yes
sno	yes	yes	yes
CG8500	yes	yes	yes
acj6	yes	yes	yes
ct	yes	yes	yes
CG12605	yes	yes	yes
CR43379	yes	yes	no

CG11438	yes	yes	yes
CG18343	yes	yes	yes
CG41128	yes	yes	yes
Cyp9b1	yes	yes	yes
CG32032	yes	yes	yes
CASK	yes	yes	yes
CG13623	yes	yes	yes
ctrip	yes	yes	yes
Pde6	yes	yes	yes
lncRNA:CR44811	yes	yes	yes
Dhc64C	yes	yes	yes
Buffy	yes	yes	yes
CG30259	yes	yes	yes
CG43367	yes	yes	yes
AttB	yes	yes	yes
Stlk	yes	yes	yes
Yp3	yes	yes	yes
CG13185	yes	yes	yes
snu	yes	yes	yes
Arp10	yes	yes	yes
Uggg	yes	yes	yes
CG32437	yes	yes	yes
UQCR-11	yes	yes	yes
Mid1	yes	yes	yes
CG16743	yes	yes	yes
Nep4	yes	yes	yes
sturkopf	yes	yes	yes
dpy	yes	yes	yes
scrt	yes	yes	yes
NijA	yes	yes	yes
CG9717	yes	yes	yes
tx	yes	yes	yes
tw	yes	yes	yes
CG6055	yes	yes	yes
tud	yes	yes	yes
Kr-h1	yes	yes	yes
klg	yes	yes	yes
CG6472	yes	yes	yes
ds	yes	yes	yes
RNASEK	yes	yes	yes
Atx2	yes	yes	yes
Cpr97Eb	yes	yes	yes
RpL38	yes	yes	yes
unc-104	yes	yes	yes
CG7409	yes	yes	yes
GstT3	yes	yes	yes
Cyp9h1	yes	yes	yes
CG12204	yes	yes	yes
Atta	yes	yes	yes
klar	yes	yes	yes
CG7255	yes	yes	yes
GM130	yes	yes	yes
snRNA:U2:34ABb	yes	yes	no
dop	yes	yes	yes

Mical	yes	yes	yes
hec	yes	yes	yes
CG5521	yes	yes	yes
CG16711	yes	yes	yes
Lcp65Ag2	yes	yes	yes
CR43217	yes	yes	no
CG7339	yes	yes	yes
Cyp313a1	yes	yes	yes
Eip93F	yes	yes	yes
bi	yes	yes	yes
asRNA:CR45479	yes	yes	yes
mxc	yes	yes	yes
CG15356	yes	yes	yes
Herc4	yes	yes	yes
Tob	yes	yes	yes
fry	yes	yes	yes
CG6602	yes	yes	yes
Dark	yes	yes	yes
CG5267	yes	yes	yes
CG9270	yes	yes	yes
Bruce	yes	yes	yes
CG43291	yes	yes	yes
CG15203	yes	yes	yes
Cpr92A	yes	yes	yes
CG42784	yes	yes	yes
snRNA:U2:38ABa	yes	yes	no
CG9572	yes	yes	yes
CR33498	yes	yes	no
CR46274	yes	yes	no
snRNA:U2:14B	yes	yes	no
Spn47C	yes	yes	yes
Ugt49B1	yes	yes	yes
qin	yes	yes	yes
CG18789	no	no	yes
CG3700	yes	yes	yes
AGO1	yes	yes	yes
CG30339	yes	yes	yes
Rgk2	yes	yes	yes
spab	yes	yes	yes
mld	yes	yes	yes
shn	yes	yes	yes
Ir62a	yes	yes	yes
CG1894	yes	yes	yes
hyd	yes	yes	yes
Cyp6a14	yes	yes	yes
PsGEF	yes	yes	yes
lncRNA:CR44272	yes	yes	yes
InR	yes	yes	yes
CG6830	yes	yes	yes
snoRNA:Psi28S-1153	yes	yes	yes
CG14613	yes	yes	yes
ATP8B	yes	yes	yes
Hmger	yes	yes	yes
Alg-2	yes	yes	yes

FASN1	yes	yes	yes
Cyp313a4	yes	yes	yes
Eip75B	yes	yes	yes
Rbcn-3B	yes	yes	yes
Ack-like	yes	yes	yes
CycT	yes	yes	yes
Cpr49Ae	yes	yes	yes
CG9344	yes	yes	yes
CR42723	no	yes	no
lncRNA:CR45746	yes	yes	yes
slow	yes	yes	yes
Ufd4	yes	yes	yes
rut	yes	yes	yes
Tmem18	yes	yes	yes
prage	yes	yes	yes
CG34120	yes	yes	yes
CG34347	yes	yes	yes
CG43901	yes	yes	yes
CG17075	yes	yes	yes
CG15456	yes	yes	yes
CG14153	yes	yes	yes
CR42653	yes	yes	no
lili	yes	yes	yes
toc	yes	yes	yes
rempA	yes	yes	yes
Dlg5	yes	yes	yes
Spag1	yes	yes	yes
CG43066	no	yes	yes
CG4928	yes	yes	yes
Su(z)2	yes	yes	yes
RhoGAP19D	yes	yes	yes
CG33679	yes	yes	yes
hfw	yes	yes	yes
tweek	yes	yes	yes
smal	yes	yes	yes
CG30285	yes	yes	yes
lncRNA:CR46064	yes	yes	yes
CG42324	yes	yes	yes
CG15894	yes	yes	yes
Obp19c	yes	yes	yes
CG7763	yes	yes	yes
E5	yes	yes	yes
Ir75c	yes	yes	yes
ksh	yes	yes	yes
Ptp99A	yes	yes	yes
Ndae1	yes	yes	yes
lncRNA:CR44377	yes	yes	yes
ppk7	yes	yes	yes
sle	no	yes	yes
CG3437	yes	yes	yes
Slob	yes	yes	yes
Osi5	yes	yes	yes
ND-19	yes	yes	yes
CG9360	yes	yes	yes

tsl	yes	yes	yes
CG17994	yes	yes	yes
CG40439	yes	yes	no
CG13954	yes	yes	yes
htt	yes	yes	yes
CG16978	yes	yes	yes
CG9932	yes	yes	yes
CG13793	yes	yes	yes
CG13995	yes	yes	yes
spen	yes	yes	yes
CG7742	yes	yes	yes
chinmo	yes	yes	yes
CG1143	yes	yes	yes
CG4115	yes	yes	yes
poe	yes	yes	yes
CG34203	yes	yes	yes
CG6230	yes	yes	yes
CG17572	yes	yes	yes
Cad87A	yes	yes	yes
ppk6	yes	yes	yes
kst	yes	yes	yes
CG7342	yes	yes	yes
Nos	yes	yes	yes
RhoGEF64C	yes	yes	yes
CheB42a	yes	yes	yes
smp-30	yes	yes	yes
Eglp2	yes	yes	yes
CG15765	yes	yes	yes
CG33774	yes	yes	yes
gnu	yes	yes	yes
CG10472	yes	yes	yes
CG3081	yes	yes	yes
trv	yes	yes	yes
trx	yes	yes	yes
ver	yes	yes	yes
Mlp60A	yes	yes	yes
CG9815	yes	yes	yes
CG10116	yes	yes	yes
CG1090	yes	yes	yes
comm3	yes	yes	yes
Obp83cd	yes	yes	yes
CG31183	yes	yes	yes
lncRNA:roX1	yes	yes	yes
Khc-73	yes	yes	yes
Cyp6g1	yes	yes	yes
side	yes	yes	yes
Nedd4	yes	yes	yes
mwh	yes	yes	yes
CG31469	yes	yes	yes
CG18130	yes	yes	yes
TTL3A	yes	yes	yes
ems	yes	yes	yes
beat-Ib	yes	yes	yes
shot	yes	yes	yes

CG13397	yes	yes	yes
t	no	yes	no
Ptip	yes	yes	yes
sky	yes	yes	yes
B4	yes	yes	yes
btv	yes	yes	yes
tay	yes	yes	yes
lncRNA:CR31451	yes	yes	yes
ps	yes	yes	yes
tbc	yes	yes	yes
DNApol-eta	yes	yes	yes
CG7848	yes	yes	yes
eIF1	yes	yes	yes
Ir94b	yes	yes	yes
lectin-46Ca	yes	yes	yes
CG43919	yes	yes	yes
CG34034	yes	yes	yes
mv	yes	yes	yes
CG17264	yes	yes	yes
Cln7	yes	yes	yes
Or59c	yes	yes	yes
CG31809	yes	yes	yes
CG34189	yes	yes	yes
CR43212	yes	no	yes
Glut4EF	yes	yes	yes
CG9265	yes	yes	yes
CanA-14F	yes	yes	yes
kug	yes	yes	yes
CG42598	yes	yes	yes
bchs	yes	yes	yes
asRNA:CR44030	yes	yes	yes
Spn77Bc	yes	yes	yes
18SrRNA-Psi:CR41602	yes	yes	yes
X11Lbeta	yes	yes	yes
cerv	yes	yes	yes
CG9812	yes	yes	yes
puf	yes	yes	yes
CG10680	yes	yes	yes
CG3349	yes	yes	yes
CG45494	yes	yes	no
Gr39a	yes	yes	yes
disp	yes	yes	yes
CG13842	yes	yes	yes
Pfdn6	yes	yes	yes
CG45493	yes	yes	no
uzip	yes	yes	yes
CR45496	yes	yes	no
Rpn13R	yes	yes	yes
Ldh	yes	yes	yes
Dro	yes	yes	yes
jbug	yes	yes	yes
lncRNA:dntRL	yes	yes	yes
CG2064	yes	yes	yes
CG42450	yes	yes	yes

CG32694	yes	yes	yes
CG15927	yes	yes	yes
CG15365	yes	yes	yes
CG13455	yes	yes	yes
Kdm4B	yes	yes	yes
CG34250	yes	yes	yes
asRNA:CR45131	yes	yes	yes
CR43696	yes	yes	no
Rpb12	yes	yes	yes
CG34297	yes	yes	yes
SPARC	yes	yes	yes
CR41454	yes	yes	no
CG3857	yes	no	no
dikar	yes	yes	yes
CG9689	yes	yes	yes
RhoGAP15B	yes	yes	yes
CG1840	yes	yes	yes
Obp44a	yes	yes	yes
lncRNA:CR43656	yes	yes	yes
Cyp6a19	yes	yes	yes
CG43784	yes	yes	no
Obp57a	yes	yes	yes
CG34334	yes	yes	yes
Frq1	yes	yes	yes
Spn43Ab	yes	yes	yes
CG7214	yes	yes	yes
CG17669	yes	yes	yes
wts	yes	yes	yes
asRNA:CR45182	yes	yes	yes
CG9701	yes	yes	yes
ppk13	yes	yes	yes
CG10600	yes	yes	yes
CG11752	yes	yes	yes
mei-P26	yes	yes	yes
ash1	yes	yes	yes
CG8492	yes	yes	yes
Piezo	yes	yes	yes
na	yes	yes	yes
Itpr	yes	yes	yes
lncRNA:CR43870	yes	yes	yes
CG11073	yes	yes	yes
Atf3	yes	yes	yes
CG17684	yes	yes	yes
CG33310	yes	yes	yes
eca	no	no	yes
CG42808	yes	yes	yes
Ca-beta	yes	yes	yes
side-III	yes	yes	yes
CG45492	yes	yes	no
MagR	yes	yes	yes
CG5098	yes	yes	yes
cv-d	yes	yes	yes
lncRNA:CR45436	yes	yes	yes
lncRNA:CR45615	yes	yes	yes

CG45488	yes	yes	no
CG32440	yes	yes	yes
CG33217	yes	yes	yes
ppk5	yes	yes	yes
CG43689	yes	yes	yes
CG6834	yes	yes	yes
CG34353	yes	yes	yes
mRpL33	yes	yes	yes
CG17279	yes	yes	yes
CG1722	yes	yes	yes
CCDC151	yes	yes	yes
CG31690	yes	yes	yes
CG14445	yes	yes	yes
drl	yes	yes	yes
cold	yes	yes	yes
Dg	yes	yes	yes
esc	yes	yes	yes
CG6583	yes	yes	yes
CG12971	yes	yes	yes
CG13067	yes	yes	yes
pwn	yes	yes	yes
CG7016	yes	yes	yes
CG12911	yes	yes	yes
CG15725	yes	yes	yes
l(1)G0196	yes	yes	yes
CG13003	yes	yes	yes
lncRNA:CR44285	yes	yes	yes
Ada1-1	no	yes	yes
RhoGAP100F	yes	yes	yes
CR45495	yes	yes	no
CG30197	yes	yes	yes
MsrA	yes	yes	yes
faf	yes	yes	yes
Cyp12d1-d	yes	yes	yes
amn	yes	yes	yes
CG17681	yes	yes	yes
CR43382	yes	no	no
Ccp84Ac	yes	yes	yes
Cyp4p1	yes	yes	yes
CG8407	yes	yes	yes
CG31769	yes	yes	yes
CG31798	yes	yes	yes
Helz	yes	yes	yes
CG5890	yes	yes	yes
Ubi-p5E	no	no	yes
Blos1	yes	yes	yes
ninaC	yes	yes	yes
CR43383	yes	no	no
tut	yes	yes	yes
Cht2	yes	yes	yes
CG31279	yes	yes	yes
CR40190	yes	yes	yes
CG10019	yes	yes	yes
Zdhhc8	yes	yes	yes

Syx16	yes	yes	yes
OtopLa	yes	yes	yes
Oxp	yes	yes	yes
CG13931	yes	yes	yes
CG30355	yes	yes	yes
SK	yes	yes	yes
CG30286	yes	yes	yes
Or85d	yes	yes	yes
CG5013	yes	yes	yes
CR43380	yes	yes	no
Cyp6a9	yes	yes	yes
slmo	yes	yes	yes
Or19a	no	no	yes
Cpr65Au	yes	yes	yes
boil	yes	yes	yes
CG11882	yes	yes	yes
Glut1	yes	yes	yes
Tsp42Eg	yes	yes	yes
slpr	yes	yes	yes
CG33159	yes	yes	yes
CG42246	yes	yes	yes
CheB93a	yes	yes	yes
Mdr49	yes	yes	yes
CG14563	yes	yes	yes
CG3348	yes	yes	yes
Wnk	yes	yes	yes
snRNA:U2:34ABA	yes	yes	no
CG13408	yes	yes	yes
CG14105	yes	yes	yes
CG18171	yes	yes	yes
CR43364	yes	yes	yes
CG12507	yes	yes	yes
tutl	yes	yes	yes
scpr-C	yes	yes	yes
E(Pc)	yes	yes	yes
Dscam1	yes	yes	yes
tRNA:Gly-GCC-1-8	yes	yes	yes
CG10317	yes	yes	yes
lncRNA:CR45115	yes	yes	yes
CG34384	yes	yes	yes
CG45487	yes	yes	no
CG17278	yes	yes	yes
Cyp4ac1	yes	yes	yes
Itgbn	yes	yes	yes
Eip63F-1	yes	yes	yes
CG32809	yes	yes	yes
CG33939	yes	yes	yes
CR44003	yes	yes	yes
Ctr1B	yes	yes	yes
ppk10	yes	yes	yes
CG40298	yes	yes	yes
iotaTry	yes	yes	yes
Nlg4	yes	yes	yes
sut3	yes	yes	yes

CG8958	yes	yes	yes
asRNA:CR45139	yes	no	no
CG15553	yes	yes	yes
CG33110	yes	yes	yes
bru3	yes	yes	yes
Muc68Ca	yes	yes	yes
Scp2	yes	yes	yes
MESR4	yes	yes	yes
5.8SrRNA-Psi:CR45854	yes	yes	yes
CG5758	yes	yes	yes
dar1	yes	yes	yes
CG42364	yes	yes	yes
CG30172	yes	yes	yes
NC2beta	yes	yes	yes
CG11041	yes	yes	yes
CG8300	yes	yes	yes
mura	yes	yes	yes
CG32301	yes	yes	yes
mamo	yes	yes	yes
CG10339	yes	yes	yes
CG15027	yes	yes	yes
Not1	yes	yes	yes
Oatp58Dc	yes	yes	yes
Ptr	yes	yes	yes
Klp54D	yes	yes	yes
CG13751	yes	yes	yes
Toll-6	yes	yes	yes
His1:CG33816	yes	no	no
Corin	yes	yes	yes
CG11378	yes	yes	yes
CG5955	yes	yes	yes
CG3176	yes	no	no
Rdl	yes	yes	yes
Obp83g	yes	yes	yes
bol	yes	yes	yes
CG18635	yes	yes	yes
w	yes	yes	yes
CG11453	yes	yes	yes
CG10311	yes	yes	yes
Sema2b	yes	yes	yes
CG14693	yes	yes	yes
Tsp42Eh	yes	yes	yes
CG13716	yes	yes	yes
CG14505	yes	yes	yes
haf	yes	yes	yes
Yp1	yes	yes	yes
His1:CG33810	yes	no	no
Amyrel	yes	yes	yes
mthl1	yes	yes	yes
CG33977	yes	yes	yes
CG31706	yes	yes	yes
Tpi	yes	yes	yes
CG33757	yes	yes	yes
strat	yes	yes	yes

CG6739	yes	yes	yes
Tet	yes	yes	yes
cac	yes	yes	yes
d-cup	yes	yes	yes
CG44437	yes	yes	yes
CG42402	yes	yes	yes
CG16825	yes	yes	yes
lncRNA:CR44376	yes	yes	yes
His1:CG33825	yes	no	no
His1:CG33831	yes	no	no
Osi17	yes	yes	yes
Obp73a	yes	yes	yes
CG33199	yes	yes	yes
CR43384	yes	no	no
east	yes	yes	yes
Wnt10	yes	yes	yes
Cks85A	yes	yes	yes
CG10793	yes	yes	yes
Nep110	yes	yes	yes
CG13950	yes	yes	yes
CG45490	yes	no	no
CG14535	yes	yes	yes
GstE11	yes	yes	yes
CG14626	yes	yes	yes
CG15021	yes	yes	yes
Proalpha4T1	yes	yes	yes
His1:CG33840	yes	no	no
fzo	yes	yes	yes
CG3500	yes	yes	yes
Tsp42Ej	yes	yes	yes
Wnt4	yes	yes	yes
His1:CG33819	yes	no	no
CG31687	no	yes	yes
PGRP-SD	yes	yes	yes
CR45497	yes	yes	no
CR45498	yes	no	no
CG12836	yes	yes	yes
Larp4B	yes	yes	yes
CG34453	yes	yes	yes
CG1791	yes	yes	yes
HPS4	yes	yes	yes
CG6617	yes	yes	yes
Osi8	yes	yes	yes
kis	yes	yes	yes
Blimp-1	yes	yes	yes
CG34172	yes	yes	yes
CG7568	yes	yes	yes
CG30380	yes	yes	yes
Or45a	yes	yes	yes
hui	yes	yes	yes
asRNA:CR44367	yes	yes	yes
Hsc70-2	yes	yes	yes
cv-2	yes	yes	yes
mRpL36	yes	yes	yes

Thd1	yes	yes	yes
CG17065	yes	yes	yes
Hsp60B	yes	yes	yes
CG31200	yes	yes	yes
NimC1	yes	yes	yes
PIP82	yes	yes	yes
CG32436	yes	yes	yes
CG34454	yes	yes	yes
CG14325	yes	yes	yes
Calx	yes	yes	yes
CG14391	yes	yes	yes
CG31664	yes	yes	yes
kek2	yes	yes	yes
CG9068	yes	yes	yes
CG17931	yes	yes	yes
CG13442	yes	yes	yes
CG41284	yes	yes	no
His1:CG33822	yes	no	no
CG34330	yes	yes	yes
CG14921	yes	yes	yes
grass	yes	yes	yes
mtg	yes	yes	yes
Or85e	yes	yes	yes
sad	yes	yes	yes
Not11	yes	yes	yes
sosie	yes	yes	yes
CG13893	yes	yes	yes
snoRNA:2R:9445205	yes	yes	yes
Cda5	yes	yes	yes
CG1889	yes	yes	yes
CG8006	yes	yes	yes
CG42674	yes	yes	yes
CG8206	yes	yes	yes
Tep2	yes	yes	yes
Dop1R1	yes	yes	yes
asRNA:CR44065	yes	yes	yes
Spn31A	yes	yes	yes
Obp57c	yes	yes	yes
28SrRNA-Psi:CR45859	no	no	yes
CG30026	yes	yes	yes
His1:CG33837	yes	no	no
lncRNA:CR33948	yes	yes	yes
CheB38b	yes	yes	yes
CG1136	yes	yes	yes
Nepl17	yes	yes	yes
CG10205	yes	yes	yes
CG11700	yes	yes	yes
CG18234	yes	yes	yes
Hsp70Aa	no	no	yes
CG41378	yes	yes	yes
IFT20	yes	yes	yes
orb2	yes	yes	yes
CG14854	yes	yes	yes
CG43165	yes	yes	yes

CG17059	yes	yes	yes
Orct2	yes	yes	yes
His1:CG33828	yes	no	no
CG8818	yes	yes	yes
Lcp65Ag1	yes	yes	no
CG31810	yes	yes	no
CG30187	yes	yes	yes
Idgf6	yes	yes	yes
CG7365	yes	yes	yes
fred	yes	yes	yes
CAH2	yes	yes	yes
CG7296	yes	yes	yes
Gad1	yes	yes	yes
CG34417	yes	yes	yes
His1:CG33843	yes	no	no
CG7294	yes	yes	yes
CG32333	yes	yes	yes
Elk	yes	yes	yes
CG5948	yes	yes	yes
CG34025	yes	yes	yes
CG5835	yes	yes	yes
CG10352	yes	yes	yes
swi2	yes	yes	yes
CG11353	yes	yes	yes
lectin-46Cb	yes	yes	yes
CG31068	yes	yes	yes
fmt	yes	yes	yes
SPE	yes	yes	yes
Grip	yes	yes	yes
stj	yes	yes	yes
lncRNA:CR43970	yes	yes	yes
His1:CG33849	yes	no	no
Phk-3	yes	yes	yes
Cyp4p2	yes	yes	yes
lncRNA:CR45994	yes	yes	yes
asRNA:CR45682	yes	yes	yes
mab-21	yes	yes	yes
Toll-7	yes	yes	yes
CG15580	yes	yes	yes
His1:CG33807	yes	no	no
CG14507	yes	yes	yes
Pcd	yes	yes	yes
CG15772	yes	yes	yes
CG17190	yes	yes	yes
CG10804	yes	yes	yes
CG5171	yes	yes	yes
CG31955	yes	yes	yes
CadN2	yes	yes	yes
mew	yes	yes	yes
Nckx30C	yes	yes	yes
CG14797	yes	yes	yes
Lmpt	yes	yes	yes
Crg-1	yes	yes	yes
His1:CG33852	yes	no	no

Lrp4	yes	yes	yes
CG30059	no	yes	yes
CG6614	yes	yes	yes
CG32195	yes	yes	yes
enc	yes	yes	yes
CG2121	yes	yes	yes
Sccpdh1	yes	yes	yes
CG15740	yes	yes	yes
Alp8	yes	yes	yes
CG8630	yes	yes	yes
Oaz	yes	yes	yes
CG14989	yes	yes	yes
lncRNA:CR44458	yes	yes	yes
b	yes	yes	yes
lncRNA:CR46095	yes	yes	yes
Ggt-1	yes	yes	yes
Obp58b	yes	yes	yes
CG11279	yes	yes	yes
CR43086	yes	yes	yes
CG14185	yes	yes	yes
CG10996	yes	yes	yes
CG12730	yes	yes	yes
fau	yes	yes	yes
CG8677	yes	yes	yes
ana	yes	yes	yes
Ork1	yes	yes	yes
CG32260	yes	yes	yes
28SrRNA-Psi:CR41609	yes	no	yes
Lcp65Ag3	yes	yes	yes
CG10947	yes	yes	yes
iav	yes	yes	yes
CG34150	yes	yes	yes
BomBc3	yes	yes	yes
CG13018	yes	yes	yes
Or71a	yes	yes	yes
Nlg2	yes	yes	yes
CG31688	no	no	yes
CG11317	yes	yes	yes
lncRNA:CR45472	yes	yes	yes
Ssl1	yes	yes	yes
hpRNA:CR32205	yes	yes	yes
Ubc84D	yes	yes	yes
otk	yes	yes	yes
CG43675	yes	yes	yes
CG31088	yes	yes	yes
CG34430	yes	yes	yes
lncRNA:CR45183	yes	yes	yes
CG4294	yes	yes	yes
al	yes	yes	yes
28SrRNA-Psi:CR45860	yes	yes	yes
CG13739	yes	yes	yes
TLL6A	yes	yes	yes
CG30287	yes	yes	yes
Myc	yes	yes	yes

Syx8	yes	yes	yes
ChLD3	yes	yes	yes
CG32681	yes	yes	yes
rdhB	yes	yes	yes
CG3301	yes	yes	yes
CG42238	yes	yes	yes
santa-maria	yes	yes	yes
CG18368	yes	yes	yes
CG30428	yes	yes	yes
CG18063	yes	yes	yes
CG9760	yes	yes	yes
Cdc7	yes	yes	yes
ppk14	yes	yes	yes
Gld	yes	yes	yes
Uhg4	yes	yes	yes
7SLRNA:CR32864	yes	yes	yes
tty	yes	yes	yes
zfh1	yes	yes	yes
Cyp12d1-p	yes	yes	yes
asRNA:CR45924	yes	yes	yes
CG13955	yes	yes	yes
side-II	yes	yes	yes
CG4646	yes	yes	yes
CG5376	yes	yes	yes
CG6293	yes	yes	yes
TpnC73F	yes	yes	yes
CG14974	yes	yes	yes
CG42822	yes	yes	yes
lncRNA:CR45029	yes	yes	yes
Ugt37E1	yes	yes	yes
CG31313	yes	yes	yes
CG13133	yes	yes	yes
ND-49L	yes	yes	yes
CR33491	yes	no	no
CG43324	yes	yes	yes
Rspf1	yes	yes	yes
CG1265	yes	yes	yes
CheB53a	yes	yes	yes
CG14963	yes	yes	yes
CG3544	yes	yes	yes
CG3568	yes	yes	yes
Oga	yes	yes	yes
eIF4E3	yes	yes	yes
Mthfs	yes	yes	yes
Nha2	yes	yes	yes
His1:CG33804	yes	no	no
His1:CG33846	yes	no	no
CG30447	yes	yes	yes
NaPi-T	yes	yes	yes
Peritrophin-A	yes	yes	yes
CR14798	yes	yes	yes
Yp2	yes	yes	yes
Irk3	yes	yes	yes
CG18278	no	no	yes

Arr1	yes	yes	yes
CG42541	yes	yes	yes
CG12483	yes	yes	yes
His1:CG33864	yes	no	no
CG32686	yes	yes	yes
Fhos	yes	yes	yes
Psf3	yes	yes	yes
CG9733	yes	yes	yes
CR43378	yes	no	no
lncRNA:CR45517	yes	yes	yes
CG18577	yes	yes	yes
lncRNA:CR45179	yes	yes	yes
Sh	yes	yes	yes
Gr64f	yes	yes	yes
Acp24A4	yes	yes	yes
CG44158	yes	yes	yes
lncRNA:CR44222	yes	yes	yes
smog	yes	yes	yes
CG14947	yes	yes	yes
BomS4	yes	yes	yes
Cnx14D	yes	yes	yes
Vps13B	yes	yes	yes
fd3F	no	yes	yes
SmydA-9	yes	yes	yes
lncRNA:CR44795	yes	yes	yes
CG8051	yes	yes	yes
Smr	yes	yes	yes
CG34265	yes	yes	yes
CG40470	yes	yes	yes
CG12531	yes	yes	yes
CG14857	yes	yes	yes
28SrRNA-Psi:CR40741	yes	yes	yes
ppk15	yes	yes	yes
slo	yes	yes	yes
sm	yes	yes	yes
CG34165	yes	yes	yes
Sirup	yes	yes	yes
CG11686	yes	yes	yes
CG14984	yes	yes	yes
Rdh	yes	yes	yes
Tsp42Ed	yes	yes	no
Cyp4d20	yes	yes	yes
CG32655	yes	yes	yes
Glos	no	no	yes
rpr	yes	yes	yes
CG18744	yes	yes	yes
Cpr66D	yes	yes	yes
CG42526	yes	yes	yes
CG8800	yes	yes	yes
Trh	yes	yes	yes
Muc96D	yes	yes	yes
CG9902	yes	yes	yes
7SLRNA:CR42652	yes	yes	no
CG4329	yes	yes	yes

NPF	yes	yes	yes
His4:CG33903	yes	no	no
rdo	yes	yes	yes
CG33468	yes	yes	yes
CG5048	yes	yes	yes
GILT3	yes	yes	yes
Ca-Ma2d	yes	yes	yes
CG15005	yes	yes	yes
CG1227	yes	yes	yes
Efhc1.2	yes	yes	yes
CG6044	yes	yes	yes
CG7991	yes	yes	yes
CG33926	yes	yes	yes
CG4766	yes	yes	yes
CG11319	yes	yes	yes
CR43281	yes	yes	yes
Lsp1gamma	yes	yes	yes
pirk	yes	yes	yes
lncRNA:CR44441	yes	yes	yes
Ilp8	yes	yes	yes
lncRNA:CR46322	yes	yes	yes
CG31776	yes	yes	yes
CG5116	yes	yes	yes
c-cup	yes	yes	yes
28SrRNA-Psi:CR45855	yes	yes	yes
Su(Tpl)	yes	yes	yes
CG13737	yes	yes	yes
CG9436	yes	yes	yes
CR43170	yes	yes	yes
CG7213	yes	yes	yes
CG4842	yes	yes	yes
lncRNA:CR45363	yes	yes	yes
lncRNA:CR44662	yes	yes	yes
Muc30E	yes	yes	yes
Cngl	yes	yes	yes
CG4078	yes	yes	yes
CG15412	yes	yes	yes
CG4627	yes	yes	yes
CG7860	yes	yes	yes
DIP-gamma	yes	yes	yes
Tsf1	yes	yes	yes
asRNA:CR45683	yes	yes	yes
lncRNA:CR46194	yes	yes	yes
tio	yes	yes	yes
CG15878	yes	yes	yes
CG31773	yes	yes	yes
CR45470	yes	yes	no
CG3191	yes	yes	yes
ppk19	yes	yes	yes
lncRNA:CR45502	yes	yes	yes
List	yes	yes	yes
pie	yes	yes	yes
CG13032	yes	yes	yes
CG30323	yes	yes	yes

Spn	yes	yes	yes
CG33978	yes	yes	yes
Listericin	yes	yes	yes
Stacl	yes	yes	yes
CR33317	yes	yes	yes
CG43117	yes	yes	yes
plh	yes	yes	yes
CG14342	yes	yes	yes
asRNA:CR45210	yes	yes	yes
Sfp84E	yes	yes	yes
CG33296	yes	yes	yes
ham	yes	yes	yes
CG33226	yes	yes	yes
tRNA:Glu-TTC-1-3	yes	yes	yes
CG3123	yes	yes	yes
Rim	yes	yes	yes
CG8066	yes	yes	yes
CG32267	yes	yes	yes
CG34220	yes	yes	yes
Muc12Ea	yes	yes	no
CG5500	yes	yes	yes
pan	yes	yes	yes
CG15650	yes	yes	yes
CG33234	yes	yes	yes
pre-rRNA:CR45856	yes	yes	yes
CG13722	yes	yes	yes
CG14015	yes	yes	yes
asRNA:CR44690	yes	yes	yes
CG43169	yes	yes	yes
Naam	yes	yes	yes
CG14221	yes	yes	yes
NT1	yes	yes	yes
CG12123	yes	yes	yes
CG11635	yes	yes	yes
lncRNA:CR46112	yes	yes	yes
RabX2	yes	yes	yes
CG32532	yes	yes	yes
asRNA:CR45834	yes	yes	yes
His4:CG33889	yes	no	no
CG14689	yes	yes	yes
Snmp1	yes	yes	yes
CG5254	yes	yes	yes
fra	yes	yes	yes
CR43483	yes	yes	no
CG10182	yes	yes	yes
CheB42b	yes	yes	yes
Tsp42A	yes	yes	yes
lncRNA:CR45298	yes	yes	yes
CG33233	yes	yes	yes
Best3	yes	yes	yes
scpr-A	yes	yes	yes
CG43120	yes	yes	yes
CG1137	yes	yes	yes
HP1D3csd	yes	yes	yes

Gld2	yes	yes	yes
lncRNA:CR44127	yes	yes	yes
boly	yes	yes	yes
Abl	yes	yes	yes
CG32407	yes	yes	yes
Nach	yes	yes	yes
shakB	yes	yes	yes
CG4269	yes	yes	yes
Gr93b	yes	yes	yes
CG4962	yes	yes	yes
nej	yes	yes	yes
CheB42c	yes	yes	yes
CG43800	yes	yes	yes
CG30288	yes	yes	yes
CG32639	yes	yes	yes
cic	yes	yes	yes
tRNA:Gln-CTG-2-5	yes	yes	yes
CG7675	yes	yes	yes
CG32236	yes	yes	yes
MFS3	yes	yes	yes
Obp56e	yes	yes	yes
CG6511	yes	yes	yes
CG13319	yes	yes	yes
Obp56c	yes	yes	yes
Uhg1	yes	yes	yes
His4:CG33887	yes	no	no
Tsp42Er	yes	yes	yes
CG42287	yes	yes	yes
CG7084	yes	yes	yes
Drep3	yes	yes	yes
mRpL34	yes	yes	yes
Tsen34	yes	yes	yes
lncRNA:CR44768	yes	yes	yes
upSET	yes	yes	yes
CG15731	yes	yes	yes
CG33966	yes	yes	yes
CG11413	yes	yes	yes
Irk1	yes	yes	yes
CG43844	yes	yes	yes
CG2816	yes	yes	yes
CG6216	yes	yes	yes
CG32284	yes	yes	yes
CG3097	yes	yes	yes
Obp58d	yes	yes	yes
DAT	yes	yes	yes
Pp1-13C	yes	yes	yes
CG6870	yes	yes	yes
CG17127	yes	yes	yes
Mtk	yes	yes	yes
phr	no	yes	yes
CG4757	yes	yes	yes
CR33318	yes	yes	yes
Dhc16F	yes	yes	yes
CG7406	yes	yes	yes

CG32817	yes	no	no
alpha-Man-Ic	yes	yes	yes
gpp	yes	yes	yes
CG42272	yes	yes	yes
CG13255	yes	yes	yes
CG15468	yes	yes	yes
Gug	yes	yes	yes
GstE14	yes	yes	yes
Rbp	yes	yes	yes
CG7916	yes	yes	yes
wb	yes	yes	yes
Obp56f	yes	yes	yes
CG13972	yes	yes	yes
Ppm1	yes	yes	yes
CG15343	yes	yes	yes
CG43850	yes	yes	yes
Ugt35A1	yes	yes	yes
CR44391	yes	yes	yes
lncRNA:CR44582	yes	yes	yes
nan	yes	yes	yes
lncRNA:CR45215	yes	yes	yes
CG15109	yes	yes	yes
CG13965	yes	yes	yes
Rx	yes	yes	yes
CG13500	yes	yes	yes
CG43938	yes	yes	yes
CG17162	yes	yes	yes
Lsm10	yes	yes	yes
RhoGAPp190	yes	yes	yes
lncRNA:CR43495	yes	yes	yes
CG15506	yes	yes	yes
CG8888	yes	yes	yes
CG12239	yes	yes	yes
Ugt35E1	yes	yes	yes
ppk	yes	yes	yes
robo2	yes	yes	yes
CG4377	yes	yes	yes
deltaTry	yes	yes	no
CG43112	yes	yes	yes
Dsk	yes	yes	yes
CG9521	yes	yes	yes
lncRNA:CR43417	yes	yes	yes
lilli	yes	yes	yes
ppk26	yes	yes	yes
GluRIIC	yes	yes	yes
Mesh1	yes	yes	yes
Act87E	yes	yes	yes
tyf	yes	yes	yes
lncRNA:CR44283	yes	yes	yes
His4:CG33879	yes	no	no
CR33294	yes	yes	yes
CG31019	yes	yes	yes
wry	yes	yes	yes
Cpr67B	yes	yes	yes

hog	yes	yes	yes
Pdh	yes	yes	yes
CG4250	yes	yes	yes
CG42269	yes	yes	yes
CG4537	yes	yes	yes
CG15617	yes	yes	yes
IntS6	yes	yes	yes
SecCl	yes	yes	yes
lncRNA:CR44964	yes	yes	yes
lncRNA:CR44092	yes	yes	yes
CG14356	yes	yes	yes
Cpr50Cb	yes	yes	yes
CG14131	yes	yes	yes
CG8501	yes	yes	yes
CR43087	yes	no	no
lncRNA:CR43461	yes	yes	yes
Ca-alpha1T	yes	yes	yes
CG43244	yes	yes	yes
CG3513	yes	yes	yes
CG13634	yes	yes	yes
CG32107	yes	yes	yes
CG31445	yes	yes	yes
CG14490	yes	yes	yes
Snap25	yes	yes	yes
E(spl)m3-HLH	yes	yes	yes
CG16772	yes	yes	yes
Hsp67Bc	yes	yes	yes
Obp19b	yes	yes	yes
lncRNA:CR45878	yes	yes	yes
trr	yes	yes	yes
CG7333	yes	yes	yes
Osi3	yes	yes	yes
alpha-Est6	yes	yes	yes
CG32040	yes	yes	yes
Nep3	yes	yes	yes
stops	yes	yes	yes
CanA1	yes	yes	yes
CG2750	yes	yes	yes
CG33965	yes	yes	yes
CG43845	yes	yes	yes
Cyp4d14	yes	yes	yes
CG13138	yes	yes	yes
CG6660	yes	yes	yes
CG9686	yes	yes	yes
lncRNA:CR44753	yes	yes	yes
Rab27	yes	yes	yes
CG13177	yes	yes	yes
QC	yes	yes	yes
snRNA:U5:63BC	yes	yes	yes
Gr9a	yes	yes	yes
CG5984	yes	yes	yes
sug	yes	yes	yes
asRNA:CR42860	yes	yes	yes
CG11362	yes	yes	yes

l(2)efl	yes	yes	yes
CG14265	yes	yes	yes
CG17821	yes	yes	yes
CG31324	yes	yes	yes
whip	yes	yes	yes
Cyp4e3	yes	yes	yes
CG14837	yes	yes	yes
Gr33a	yes	yes	yes
CG44141	yes	yes	no
lncRNA:CR43432	yes	yes	yes
Cep89	yes	yes	yes
Muc68D	yes	yes	yes
Ndg	yes	yes	yes
CG33125	yes	yes	yes
CG42494	yes	yes	yes
CG31909	yes	yes	yes
CG17010	yes	yes	yes
lncRNA:CR45290	yes	yes	yes
CG40485	yes	yes	yes
GluRIIE	yes	yes	yes
lncRNA:CR42794	yes	yes	yes
lncRNA:CR44395	yes	yes	yes
lncRNA:CR46234	yes	yes	yes
D	yes	yes	yes
Cpr76Bc	yes	yes	yes
His4:CG33901	yes	no	no
dpr21	yes	yes	yes
CG44140	yes	yes	no
lncRNA:CR33963	yes	yes	yes
NaCP60E	yes	yes	yes
snRNA:7SK	yes	yes	yes
Cht6	yes	yes	yes
CR41379	yes	yes	yes
His4:CG33873	yes	no	no
Dcr-1	yes	yes	yes
CG11997	yes	yes	yes
rha	yes	yes	yes
CG32219	yes	yes	yes
lncRNA:CR45897	yes	yes	yes
CG42704	yes	yes	yes
pum	yes	yes	yes
LpR2	yes	yes	yes
run	yes	yes	yes
CG7714	yes	yes	yes
CG12970	yes	yes	yes
CG13026	yes	yes	yes
CG30195	yes	yes	yes
CG8329	yes	yes	yes
lncRNA:CR45674	yes	yes	yes
Cyp4s3	yes	yes	yes
Cyp6d2	yes	yes	yes
dgt3	yes	yes	yes
CG30060	yes	yes	yes
CR43671	yes	yes	yes

CG10864	yes	yes	yes
Hml	yes	yes	yes
lncRNA:CR43828	yes	yes	yes
CG7203	yes	yes	yes
asRNA:CR45009	yes	yes	yes
Pkd2	yes	yes	yes
Slc45-1	yes	yes	yes
tRNA:Thr-AGT-2-1	yes	yes	yes
CG13023	yes	yes	yes
CheB98a	yes	yes	yes
lncRNA:CR43960	yes	yes	yes
Obp99b	yes	yes	yes
CG34002	yes	yes	no
Gr77a	yes	yes	yes
CG31642	yes	yes	yes
Cib2	yes	yes	yes
CG6527	yes	yes	yes
His4:CG31611	yes	no	no
CG13250	yes	yes	yes
CG17270	yes	yes	yes
lectin-24Db	yes	yes	yes
CG16854	yes	yes	yes
Rh4	yes	yes	yes
Ca-alpha1D	yes	yes	yes
lncRNA:CR43943	yes	yes	yes
ect	yes	yes	yes
Dnai2	yes	yes	yes
His4:CG33909	yes	no	no
CG30289	yes	yes	yes
PR-Set7	yes	yes	yes
Kebab	yes	yes	yes
drm	yes	yes	yes
CG31624	no	no	yes
CG10663	yes	yes	yes
CG1545	yes	yes	yes
lncRNA:CR32385	yes	yes	yes
Sfp51E	yes	yes	yes
CG43191	yes	yes	yes
CG9514	yes	yes	yes
Mlc1	yes	yes	yes
AgmNAT	yes	yes	yes
CG13924	yes	yes	yes
CG13920	yes	yes	yes
CG14142	yes	yes	yes
CG34450	yes	yes	yes
AQP	yes	yes	yes
CG10730	yes	yes	yes
Osi19	yes	yes	yes
CG6931	yes	yes	yes
Ir94a	yes	yes	yes
side-IV	yes	yes	yes
CG34382	yes	yes	yes
Nxf3	yes	yes	yes
ninaD	yes	yes	yes

CG5002	yes	yes	yes
CG1504	no	yes	yes
CG14054	yes	yes	yes
CG2003	yes	yes	yes
rad	yes	yes	yes
lncRNA:CR44288	yes	yes	yes
CG30284	yes	yes	yes
CG31675	yes	yes	yes
His4:CG33869	yes	no	no
lncRNA:CR45217	yes	yes	yes
Obp56b	yes	yes	yes
Scp1	yes	yes	yes
asRNA:CR44375	yes	yes	yes
lncRNA:CR45177	yes	yes	yes
CG42633	yes	yes	yes
CG5612	yes	yes	yes
CG42288	yes	yes	yes
CG10384	yes	yes	yes
tn	yes	yes	yes
beat-IIb	yes	yes	yes
PDCD-5	yes	yes	yes
alrm	yes	yes	yes
CG13427	yes	yes	yes
CG10177	yes	yes	yes
lncRNA:CR44373	yes	yes	yes
Unc-115b	no	no	yes
lncRNA:CR46239	yes	yes	yes
Obp57b	yes	yes	no
lncRNA:CR44026	yes	yes	yes
CG42691	yes	yes	yes
lncRNA:CR44834	yes	yes	yes
lncRNA:CR44608	yes	yes	yes
Dop2R	yes	yes	yes
CG44303	yes	yes	yes
CG8192	yes	yes	yes
CG13841	yes	yes	yes
Doc1	yes	yes	yes
CR43975	yes	yes	yes
FipoQ	yes	yes	yes
ana3	yes	yes	yes
CG43841	yes	yes	yes
CG31882	yes	yes	yes
His4:CG33871	yes	no	no
Fmo-1	yes	yes	yes
lncRNA:CR44360	yes	yes	yes
Atg8b	yes	yes	yes
CG16727	yes	yes	yes
2mit	yes	yes	yes
Hsp70Bb	yes	yes	yes
CG45076	yes	yes	yes
CG5639	yes	yes	yes
CG4415	yes	yes	yes
l(2)k05911	yes	yes	yes
Gr43a	yes	yes	yes

CG13856	yes	yes	yes
CG6191	yes	yes	yes
Cpr62Ba	yes	yes	yes
lncRNA:CR44415	yes	yes	yes
CG30054	yes	yes	yes
NKCC	yes	yes	yes
CG34180	yes	yes	yes
lncRNA:CR33942	yes	yes	yes
CG32726	yes	yes	yes
Ddc	yes	yes	yes
CG32695	yes	yes	yes
Uros2	yes	yes	yes
Tsp42Eq	yes	yes	yes
nAChRalpha3	yes	yes	yes
Gr94a	yes	yes	yes
CG7322	yes	yes	yes
alpha-Est7	yes	yes	yes
CG4587	yes	no	no
asRNA:CR44165	yes	yes	yes
CG43339	yes	yes	yes
CG33098	yes	yes	yes
Cyp305a1	yes	yes	yes
CG31636	yes	yes	yes
lncRNA:CR45985	yes	yes	yes
CG14285	yes	yes	yes
lncRNA:CR32636	yes	yes	yes
lncRNA:CR44772	yes	yes	yes
CG10869	yes	yes	yes
Ctf4	yes	yes	yes
Dph5	yes	yes	yes
CG9650	yes	yes	yes
Frq2	yes	yes	yes
Scsalpha2	yes	yes	yes
lncRNA:CR45170	yes	yes	yes
Acp36DE	yes	yes	yes
PRL-1	yes	yes	yes
Try29F	yes	yes	yes
CG14304	yes	yes	yes
CG31826	yes	yes	yes
Ets65A	yes	yes	yes
CG14205	yes	yes	yes
CG14340	yes	yes	yes
Mlc2	yes	yes	yes
CG14085	yes	yes	yes
tRNA:Gly-GCC-1-1	yes	no	no
CG14905	yes	yes	yes
CG14913	yes	yes	yes
CG32767	yes	yes	yes
Echs1	yes	yes	yes
CG15528	yes	yes	yes
lncRNA:CR45922	yes	yes	yes
CG10183	yes	yes	yes
lncRNA:CR44506	yes	yes	yes
CR44839	no	no	yes

Ppn	yes	yes	yes
CG32249	yes	yes	yes
Prx2540-1	yes	yes	no
CG42825	yes	yes	yes
Task7	yes	yes	yes
sob	yes	yes	yes
5-HT7	yes	yes	yes
Shawl	yes	yes	yes
lncRNA:CR44396	yes	yes	yes
CG43219	yes	yes	yes
snoRNA:Psi28S-3571	yes	yes	yes
CG1850	yes	yes	yes
klu	yes	yes	yes
asRNA:CR44873	yes	yes	yes
CG13563	yes	yes	yes
lncRNA:CR44338	yes	yes	yes
CG6927	yes	yes	yes
CG12974	yes	yes	yes
asRNA:CR44053	yes	yes	yes
CG2006	yes	yes	yes
Mlp84B	yes	yes	yes
CG8155	yes	yes	yes
CG13215	yes	yes	yes
CG7460	yes	yes	yes
CG33337	yes	yes	yes
CG8646	yes	yes	yes
Cpr11B	yes	yes	yes
CG14132	yes	no	no
CG4000	yes	yes	yes
hob	yes	yes	yes
CG2926	yes	yes	yes
CG43897	yes	yes	yes
CG42816	yes	yes	yes
CG3611	yes	yes	yes
CG42266	yes	yes	yes
lncRNA:CR44614	no	yes	no
Jarid2	yes	yes	yes
wupA	yes	yes	yes
asRNA:CR44291	yes	yes	yes
CG42826	yes	yes	yes
CG9672	yes	yes	yes
lncRNA:CR44472	yes	yes	yes
CG31683	no	no	yes
dpr16	yes	yes	yes
lncRNA:CR46021	yes	yes	yes
Su(var)2-HP2	yes	yes	yes
tRNA:Arg-ACG-1-5	yes	yes	yes
alpha-Est4aPsi	yes	yes	yes
Faa	yes	yes	yes
CG18641	yes	yes	yes
Hsp70Bbb	yes	yes	yes
Ir85a	yes	yes	yes
Or1a	yes	yes	yes
asRNA:CR44981	yes	yes	yes

CG10073	yes	yes	yes
lncRNA:CR45048	yes	yes	yes
CG33293	yes	yes	yes
pb	yes	yes	yes
Uhg7	yes	yes	yes
Ripalpha	yes	yes	yes
lncRNA:CR44708	yes	yes	yes
JIL-1	yes	yes	yes
Ir51a	yes	yes	yes
grk	yes	yes	yes
CG7432	yes	yes	yes
CG14694	yes	yes	yes
CG11835	yes	yes	yes
CG32814	yes	yes	yes
Haspin	yes	yes	yes
Nep15	yes	yes	yes
Pvf2	yes	yes	yes
lncRNA:CR46041	yes	yes	yes
Dh44	yes	yes	yes
Ir56a	yes	yes	yes
ft	yes	yes	yes
CG33521	yes	yes	yes
Trpgamma	yes	yes	yes
Grik	yes	yes	yes
CG4653	yes	yes	yes
CG9072	yes	yes	yes
CG42346	yes	yes	yes
lncRNA:CR43488	yes	no	no
CG5773	yes	yes	yes
Hsp70Bc	yes	yes	yes
CR43697	yes	yes	no
CG5122	yes	yes	yes
CG13082	yes	yes	yes
CG18536	yes	yes	yes
twit	yes	yes	yes
CG8343	yes	yes	yes
CG5435	yes	yes	yes
mthl7	yes	yes	yes
CG2652	yes	yes	yes
CG10184	yes	yes	yes
CG7886	yes	yes	yes
Trf2	yes	yes	yes
CG15040	yes	yes	yes
dan	yes	yes	yes
lncRNA:CR44515	yes	yes	yes
CG32669	yes	yes	yes
CR40450	yes	yes	yes
CG14219	yes	yes	yes
lncRNA:CR44914	yes	yes	yes
CG14022	yes	yes	yes
CG15473	yes	yes	yes
Zip89B	yes	yes	yes
Cyp301a1	yes	yes	yes
alphaTub85E	yes	yes	yes

CG6749	yes	yes	yes
CG12224	yes	yes	yes
CR41509	yes	yes	no
CG31904	no	yes	no
CG3708	yes	yes	yes
CG10211	yes	yes	yes
Map205	yes	yes	yes
jing	yes	yes	yes
Def	yes	yes	yes
CG40198	yes	yes	yes
MESK4	yes	yes	no
CG17108	yes	yes	yes
CG2898	yes	yes	yes
CG31347	yes	yes	yes
ZnT33D	yes	yes	yes
tilB	yes	yes	yes
lncRNA:CR45910	yes	yes	yes
rk	yes	yes	yes
CG7560	yes	yes	yes
blanks	yes	yes	yes
asRNA:CR43683	yes	yes	yes
asRNA:CR42547	yes	yes	yes
Hpd	yes	yes	yes
CG12502	yes	yes	yes
Doc3	yes	yes	yes
lncRNA:CR44912	yes	yes	yes
CG11044	yes	yes	yes
osa	yes	yes	yes
CG5023	yes	yes	yes
Hsp70Ba	yes	yes	no
Pxd	yes	yes	yes
b6	yes	yes	yes
snoRNA:Psi28S-3305c	yes	yes	yes
CR41320	yes	no	no
CG17707	yes	yes	yes
CG33493	yes	yes	yes
CG42393	yes	yes	yes
CG42637	no	no	yes
RPA2	yes	yes	yes
lncRNA:CR31781	yes	yes	yes
CG6972	yes	yes	yes
lncRNA:CR43264	yes	yes	yes
Cyp4ac2	yes	yes	yes
CG8260	yes	yes	yes
lncRNA:CR45592	yes	no	no
CG30154	yes	yes	yes
CG12594	yes	yes	yes
beat-VI	yes	yes	yes
anal	yes	yes	yes
nAChRalpha1	yes	yes	yes
lncRNA:CR42719	yes	yes	yes
lncRNA:CR45758	yes	yes	yes
mei-218	yes	yes	yes
CG13488	yes	yes	yes

e(y)2b	yes	yes	yes
CG13749	yes	yes	yes
Cht7	yes	yes	yes
ImpL1	yes	yes	yes
CG8997	yes	yes	yes
CG12853	yes	yes	yes
slbo	yes	yes	yes
CG18540	yes	yes	yes
CG8854	yes	yes	yes
Spn43Ad	yes	yes	yes
hgo	yes	yes	yes
Dnali1	yes	yes	yes
CG34057	yes	yes	yes
DIP-iota	yes	yes	yes
Skeletor	yes	yes	yes
Strica	yes	yes	yes
CG3630	yes	yes	yes
CG34357	yes	yes	yes
Tm2	yes	yes	yes
CG34377	yes	yes	yes
twi	yes	yes	yes
JhI-21	yes	yes	yes
CG33941	yes	yes	yes
Mpcp1	yes	yes	yes
CG33062	yes	yes	yes
CG12035	yes	yes	yes
CG34054	yes	yes	yes
CG11617	yes	yes	yes
eater	yes	yes	no
SPR	yes	yes	yes
hdc	yes	yes	yes
CG15531	yes	yes	yes
CG17760	yes	yes	yes
CG17687	yes	yes	yes
CG1815	yes	yes	yes
CG42232	yes	yes	yes
yellow-b	yes	yes	yes
Adk2	yes	yes	yes
CG34446	yes	yes	yes
TwdlG	yes	yes	yes
lncRNA:CR43995	yes	yes	yes
CG15563	yes	yes	yes
nxf2	yes	yes	yes
CG17777	yes	yes	yes
CG44085	yes	yes	yes
CG12229	yes	yes	yes
Prm	yes	yes	yes
hale	yes	yes	yes
CG13407	yes	yes	yes
CG30280	yes	yes	yes
CG42249	no	yes	yes
SmydA-3	yes	yes	yes
Alk	yes	yes	yes
Cyp12c1	yes	yes	yes

Drsl4	yes	yes	yes
Act57B	yes	yes	yes
lncRNA:CR44831	yes	yes	yes
lncRNA:CR43148	yes	yes	yes
CAH13	yes	yes	yes
CG31205	yes	yes	yes
CG13202	yes	yes	yes
CR42499	yes	no	no
CG11768	yes	yes	yes
Npc2f	yes	yes	yes
Kah	yes	yes	yes
Gr66a	yes	yes	yes
CG15143	yes	no	no
CG9416	yes	yes	yes
Cpr100A	yes	yes	yes
CG32447	yes	yes	yes
Mf	yes	yes	yes
CG7329	yes	yes	yes
lncRNA:CR44779	yes	yes	yes
Elys	yes	yes	yes
CG34141	yes	yes	yes
CG14695	yes	yes	yes
GluRIB	yes	yes	yes
CG4860	yes	yes	yes
lncRNA:CR44938	yes	yes	no
lncRNA:CR44043	yes	yes	yes
LpR1	yes	yes	yes
CG42797	yes	yes	no
CG32816	yes	yes	yes
lncRNA:CR44470	yes	yes	yes
CG14964	yes	yes	yes
danr	yes	yes	yes
Ada1-2	no	no	yes
Obp58c	yes	yes	yes
CG5810	yes	yes	yes
CG16719	yes	yes	yes
bbx	yes	yes	yes
Alp5	yes	yes	yes
Or63a	yes	yes	yes
CG43206	yes	yes	yes
CG45218	yes	yes	yes
CG34155	yes	yes	yes
CG10175	yes	yes	yes
RpII215	yes	yes	yes
NimB2	yes	yes	yes
CG15529	yes	yes	yes
CG43788	yes	yes	yes
l(3)psg2	yes	yes	yes
Hsp27	yes	yes	no
lncRNA:CR43910	yes	yes	yes
TwdlT	yes	yes	yes
CG10993	yes	yes	yes
CG12825	yes	yes	yes
chif	yes	yes	yes

CG34409		yes	yes	yes
ea		yes	yes	yes
dysf		yes	yes	yes
CG15909		yes	yes	yes
CG4066		yes	yes	yes
lncRNA:sphinx		yes	yes	yes
CG3119		yes	yes	yes
CG13996		yes	yes	yes
CG7201		yes	yes	yes
CG5849		yes	yes	yes
CG34256		yes	yes	yes
Nnf1b		yes	yes	yes
CG33640		yes	yes	yes
CG14257		yes	yes	yes
lncRNA:CR44813		yes	yes	yes
CG40486		yes	yes	yes
CG11893		yes	yes	yes
Nmdar2		yes	yes	yes
Lim3		yes	yes	yes
CG14880		yes	yes	yes
CG15382		yes	yes	yes
CG1924		yes	no	no
CG13562		yes	yes	yes
TLL1B		yes	yes	yes
gudu		yes	yes	yes
CG13458		yes	yes	yes
CG31274		yes	no	no
lncRNA:CR45444		yes	yes	yes
CG33470		yes	no	no
lncRNA:CR44868		yes	yes	yes
Pop5		yes	yes	yes

results/tables/tbl134_wVsCantonS_chonky.html

3.11.2.1 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 135. Top Ten Most Significantly (padj<0.01) Differentiated garden-variety vs amos mutants

rank	multi				random			
	name	expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange	adjusted p
1	AGO2	1.66	-1.591	3.31×10^{-255}	AGO2	1.45	-1.372	5.00
2	Acbp2	4.52	3.489	5.04×10^{-164}	Acbp2	4.52	3.488	3.00
3	CG31224	0.33	-1.610	3.23×10^{-138}	CG31224	0.33	-1.611	4.00
4	CG7900	1.61	5.200	4.28×10^{-131}	CG7900	1.61	5.199	9.00
5	DIP-alpha	0.09	-6.144	1.07×10^{-117}	DIP-alpha	0.09	-6.146	1.00
6	Cht10	0.10	-3.751	3.32×10^{-101}	Cht10	0.10	-3.752	8.00
7	SdicA	0.18	-3.340	2.55×10^{-96}	CG10631	0.11	-1.733	2.00
8	Sdic4	0.25	-2.356	1.02×10^{-92}	Hipk	0.45	-1.479	1.00

9	CG10631	0.11	-1.733	2.19×10^{-89}	lncRNA:CR43314	0.05	-4.023
10	Sdic1	0.11	-5.559	1.70×10^{-83}	sif	0.01	-4.478

Top 10 genes with biggest (significant) effect sizes

Table 136. Top Ten Largest Magnitude Fold Changes which white-eyed vs CantonS

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	CG43149	2.22	14.488	1.59×10^{-26}	CG43149	2.21	14.488
2	IM14	1.09	-12.379	4.32×10^{-19}	IM14	1.09	-12.379
3	CheB42a	0.55	11.962	9.62×10^{-7}	CheB42a	0.55	11.962
4	CG32437	0.05	-9.881	6.76×10^{-15}	CG43291	0.04	10.51
5	CG1894	0.08	9.700	7.45×10^{-12}	CG32437	0.05	-9.881
6	CheB93a	0.29	9.639	2.45×10^{-4}	CG1894	0.08	9.639
7	lncRNA:CR44377	0.01	-9.337	2.93×10^{-10}	CheB93a	0.29	9.337
8	CR43217	0.11	-9.222	6.07×10^{-14}	lncRNA:CR44377	0.01	-9.222
9	CG34334	0.01	-8.994	5.21×10^{-7}	CG34334	0.01	-8.994
10	lncRNA:CR43656	0.01	-8.930	3.92×10^{-7}	lncRNA:CR43656	0.01	-8.930

Top 10 highest expressed genes with significant change

Table 137. Top Ten Highest Expressed Genes with Significant Difference
garden-variety vs amos mutants

rank	name	multi			rando		
		expression	log2 FoldChange	adjusted p	name	expression	log2 FoldChange
1	antdh	29.01	0.532	5.71×10^{-3}	antdh	29.00	0.532
2	CG8369	20.70	2.230	1.16×10^{-32}	CG8369	20.69	2.230
3	Ugt35B1	16.06	-0.701	3.11×10^{-3}	Ugt35B1	16.05	-0.701
4	Obp56d	15.56	0.751	1.17×10^{-3}	Obp56d	15.56	0.751
5	CG9336	15.30	0.577	2.94×10^{-3}	CG9336	15.30	0.577
6	CG30197	14.09	1.356	1.71×10^{-13}	CG30197	14.09	1.356
7	RpS18	13.40	0.373	2.37×10^{-3}	RpS18	13.40	0.373
8	eEF1alpha1	13.21	-0.487	4.93×10^{-3}	eEF1alpha1	13.20	-0.487
9	28SrRNA-Psi:CR41609	12.70	-3.566	6.09×10^{-7}	Idgf4	11.95	0.701
10	Idgf4	11.95	0.713	3.29×10^{-4}	RpL18A	11.88	0.373

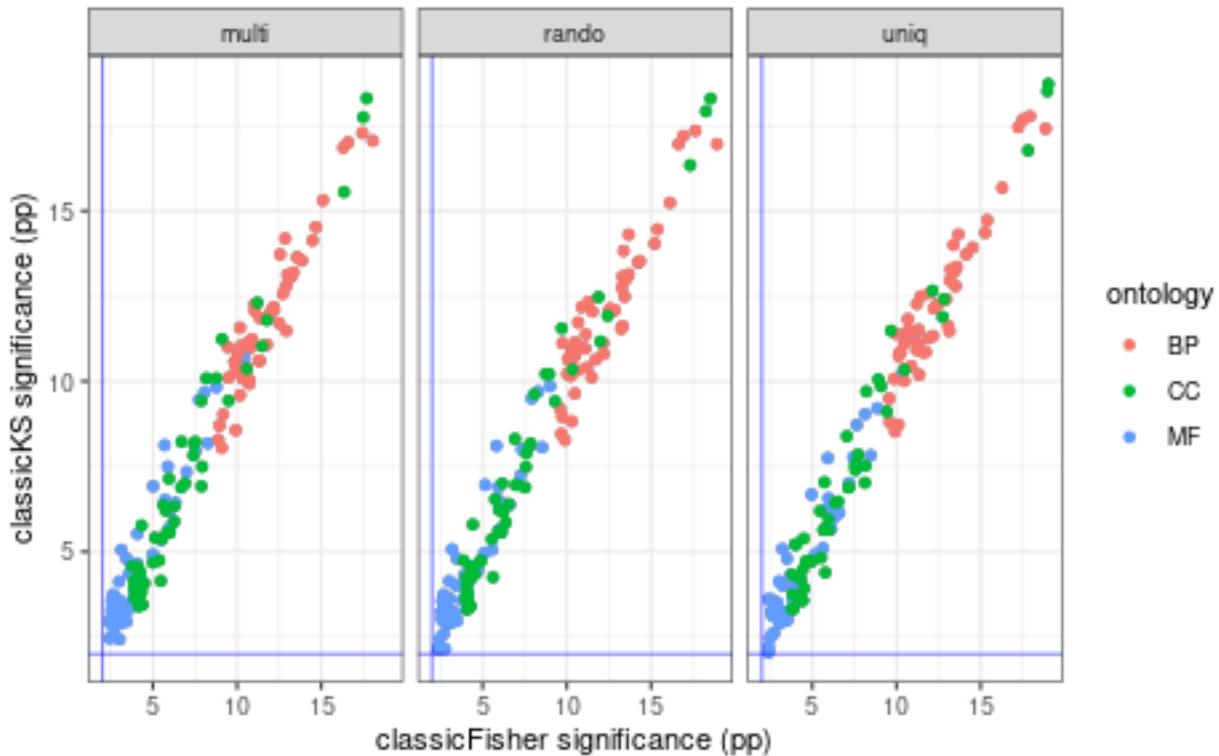
3.11.2.2 Gene Ontology Enrichment

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 118. Scatterplot of GO Term Enrichment Significance for Two Tests (white-eyed vs CantonS)



```
## png
## 2
```

```
inorganic molecular entity transmembrane transporter activity (GO:0015318)
passive transmembrane transporter activity (GO:0022803)
system development (GO:0048731)
localization (GO:0051179)
obsolete membrane part (GO:0044425)
obsolete plasma membrane part (GO:0044459)
obsolete neuron part (GO:0097458)
ribosomal subunit (GO:0044391)
membrane protein complex (GO:0098796)
plasma membrane protein complex (GO:0098797)
```

Table 138. Enriched GO Terms among Significantly Differentially Expressed Genes CantonS vs white-eye; uniq only; top 10 most significant per category

GO Term	Description	p-value	
		Fisher	K-S
MF			
GO:0003735	structural constituent of ribosome	5.00×10^{-11}	6.40×10^{-11}
GO:0008324	cation transmembrane transporter activity	1.40×10^{-9}	6.20×10^{-10}
GO:0015318	NA	3.40×10^{-9}	1.50×10^{-8}

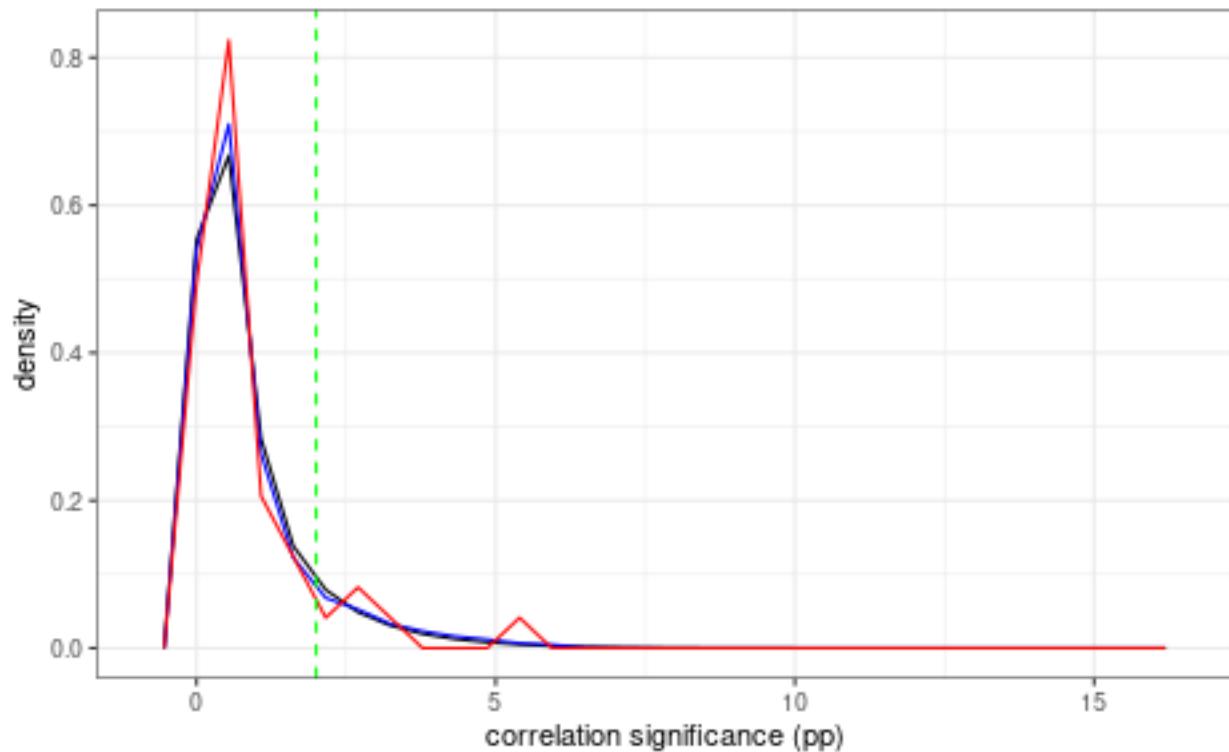
GO:0022890	inorganic cation transmembrane transporter activity	7.30×10^{-9}	9.20×10^{-10}
GO:0046873	metal ion transmembrane transporter activity	2.30×10^{-8}	1.90×10^{-9}
GO:0005509	calcium ion binding	3.70×10^{-8}	1.70×10^{-8}
GO:0015075	ion transmembrane transporter activity	6.70×10^{-8}	9.70×10^{-8}
GO:0022857	transmembrane transporter activity	2.50×10^{-7}	7.20×10^{-7}
GO:0015267	channel activity	4.70×10^{-7}	8.70×10^{-7}
GO:0022803	NA	4.70×10^{-7}	8.70×10^{-7}
<hr/>			
BP			
GO:0048731	NA	1.40×10^{-19}	3.90×10^{-18}
GO:0050896	response to stimulus	1.20×10^{-18}	1.70×10^{-18}
GO:0023052	signaling	3.10×10^{-18}	2.10×10^{-18}
GO:0007154	cell communication	5.80×10^{-18}	3.50×10^{-18}
GO:0051179	NA	5.30×10^{-17}	2.10×10^{-16}
GO:0048666	neuron development	4.10×10^{-16}	1.90×10^{-15}
GO:0030182	neuron differentiation	5.30×10^{-16}	4.40×10^{-15}
GO:0007399	nervous system development	3.00×10^{-15}	1.20×10^{-14}
GO:0031175	neuron projection development	7.20×10^{-15}	1.90×10^{-14}
GO:0006810	NA	2.10×10^{-14}	4.90×10^{-15}
<hr/>			
CC			
GO:0071944	cell periphery	9.70×10^{-20}	1.90×10^{-19}
GO:0005886	plasma membrane	1.10×10^{-19}	3.10×10^{-19}
GO:0016020	membrane	1.50×10^{-18}	1.70×10^{-17}
GO:0044425	NA	1.50×10^{-13}	3.90×10^{-13}
GO:0044459	NA	1.80×10^{-13}	1.30×10^{-12}
GO:0097458	NA	7.30×10^{-13}	2.20×10^{-13}
GO:0044391	NA	3.40×10^{-11}	4.60×10^{-11}
GO:0098796	NA	2.10×10^{-10}	3.30×10^{-12}
GO:0005840	ribosome	3.70×10^{-10}	7.80×10^{-10}
GO:0098797	NA	8.20×10^{-10}	1.40×10^{-10}

3.12 Rotund Mutants: Gene Correlation Networks

Reads were sequenced from three genotypes (wt, homozygotic rn mutants, and heterozygotes) across four developmental stages (L, P8, P40, A). These were mapped and counted using the standard pipeline, and these counts converted to RPKM (raw counts from the same sample will necessarily have a true correlation because they have the same depth of sequencing). These were correlated with the DGCA package (McKenzie et al. (2016) ; <http://htmlpreview.github.io/?https://github.com/andymckenzie/DGCA/blob/master/doc/DGCA.html>)

For a tractable example, the list of genes differentially expressed in the housing contrast were collected. The correlation statistics were collected for all correlations in the annotated genome, the sublist vs the genome, and the sublist vs. the sublist. No multiple comparison adjustment has been applied.

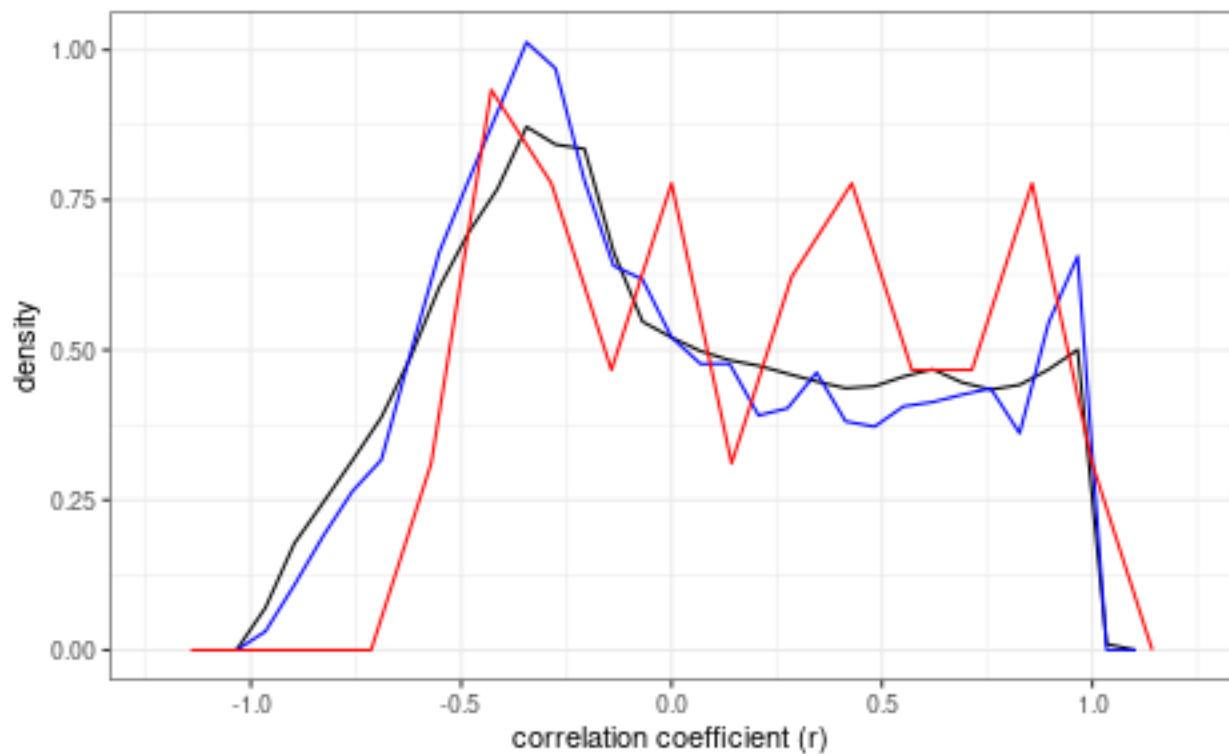
Figure 119. Significance Histogram for Gene-Pair Correlations
all gene round-robin, housing contrast vs all gene, and housing contrast round-robin



```
## png
## 2
```

Figure 120. r Histogram for Gene-Pair Correlations

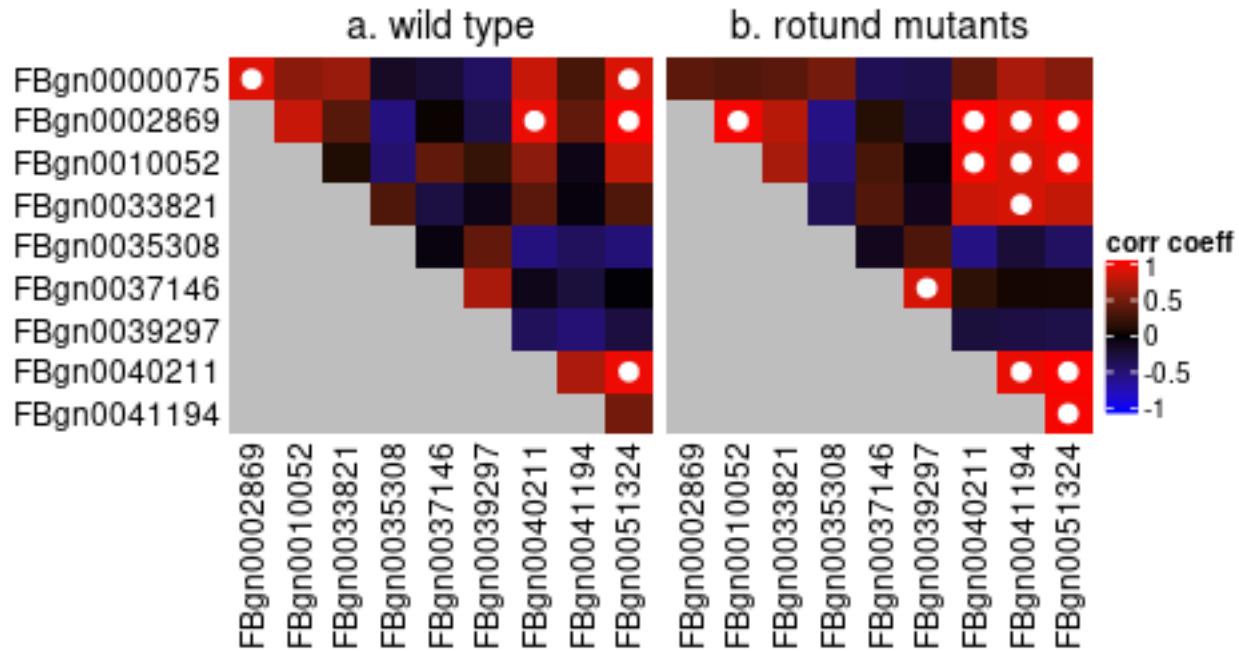
all gene round-robin, housing contrast vs all gene, and housing contrast round-robin



```
## png  
## 2
```

Correlations of the significantly differentially expressed genes in the housing contrast, vs themselves:

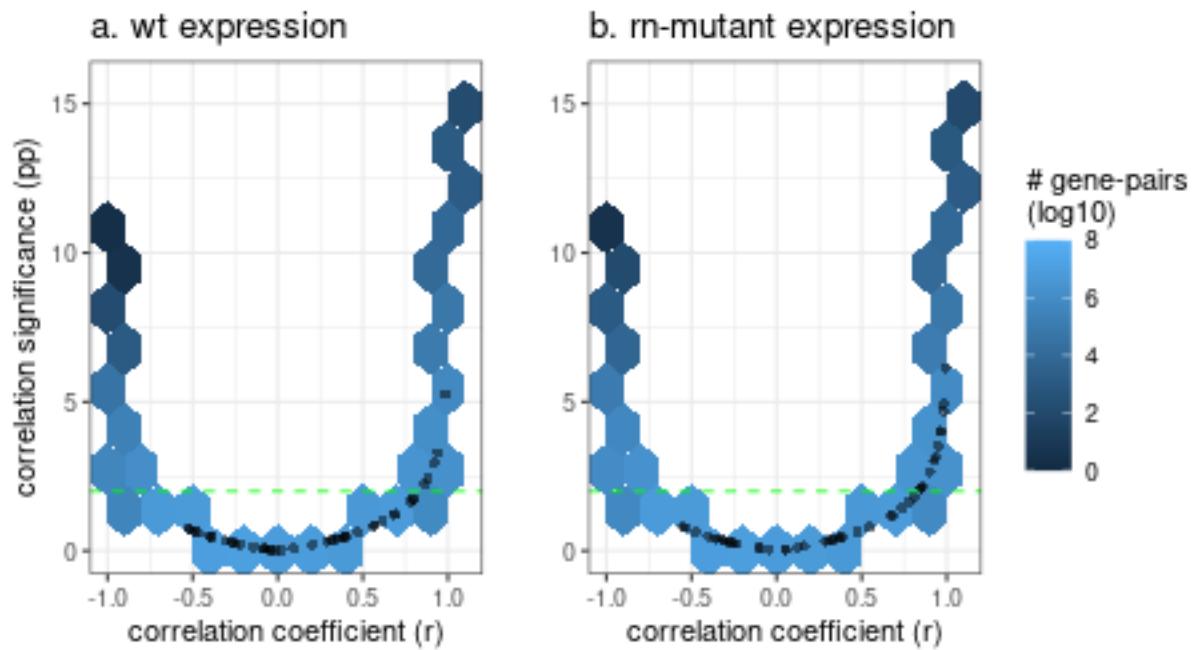
Figure 121. Correlation Strength & Significance ($p < 0.01$)
 For Pairs of Genes which were Differentially Expressed
 in the Housing Contrast



```
## png
## 2
```

Plotting the correlation statistics for the entire genome, on top of that the stats for the housing-contrast genes of interest:

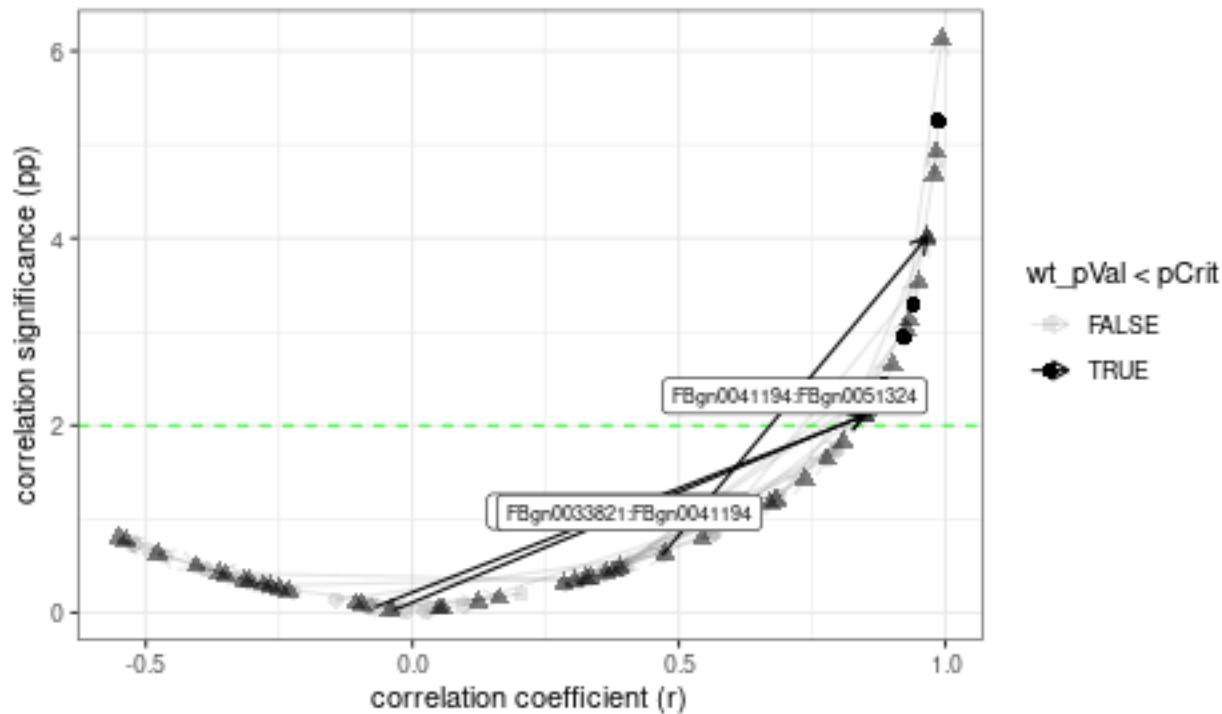
Figure 122. 'Tornado Plot':
Correlation Significance vs Strength for Gene Pair Expression
all gene round-robin background; housing contrast round-robin foreground



```
## png
## 2
```

As well as calculating gene correlations, DGCA will calculate differential gene correlation, ie a (de)coupling of gene expression patterns with treatment. 3 such pairs exist in the housing genes round-robin (difference significance threshold relaxed to 0.05 for sake of example)

**Figure 123. 'Tornado Plot':
Correlation Significance vs Strength
wt & rn mutants, housing contrast round-robin**



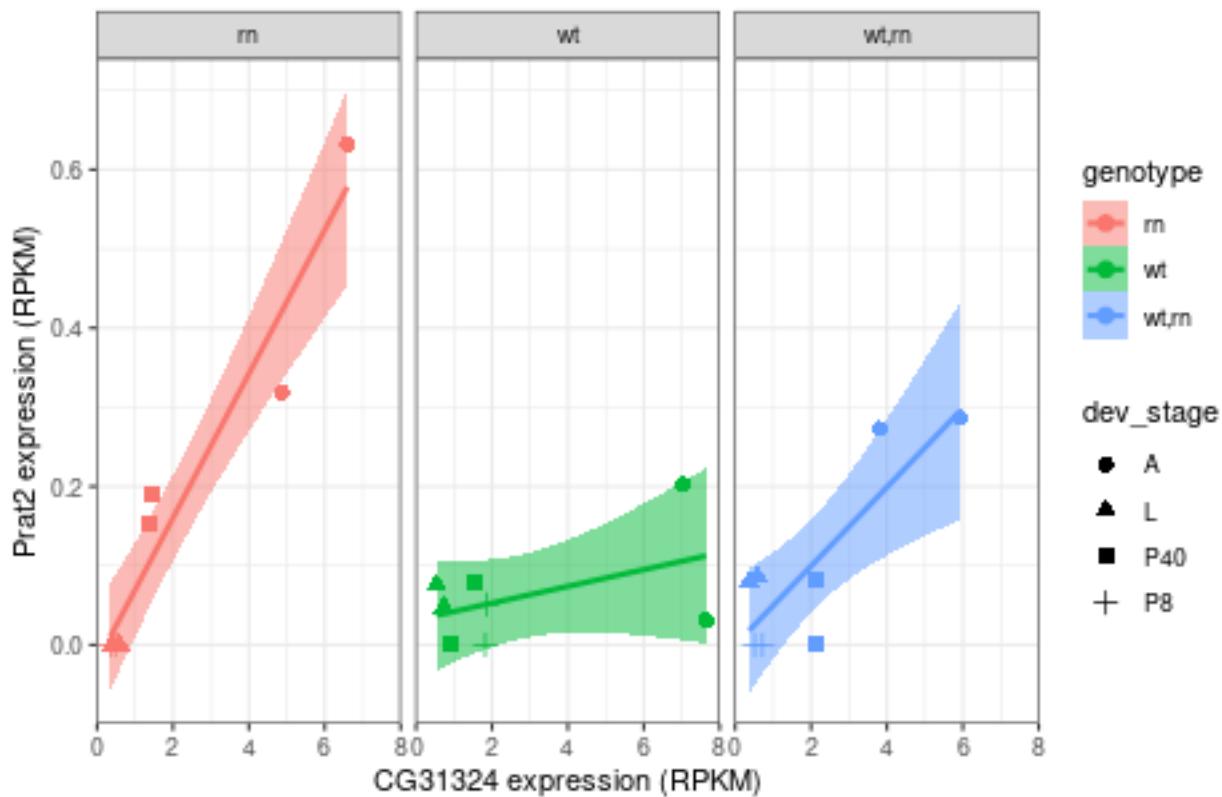
```
## png
## 2
```

Table 138. Differentially Correllated Pairs Within Genes Differentially Expressed Under Isolations
rotund mutant vs wt expression

Gene Pair		wt correlation		rn correlation		difference p-value	Classes
		r	p	r	p		
Prat2	CG31324	0.469	2.41×10^{-1}	0.966	9.72×10^{-5}	1.65×10^{-2}	0/+
Jhe	Prat2	-0.070	8.69×10^{-1}	0.848	7.88×10^{-3}	3.72×10^{-2}	0/+
CG10799	Prat2	-0.035	9.35×10^{-1}	0.849	7.73×10^{-3}	4.21×10^{-2}	0/+

The three significance differences involve couplings of CG31324, Jhe, and CG10799 with Prat2 as the latter's expression increases with rotund mutation. The most dramatic is plotted below. It's interesting that although the correlations were only computed for homozygotes, the heterozygotes are qualitatively intermediate.

Figure 124. Differential Correlations



```
## png
## 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 preserving large differences},
##   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")'.

##
## The methods within the code package can be cited as:
##
## Gu, Z. (2014) circlize implements and enhances circular visualization
## in R. Bioinformatics.
##
## A BibTeX entry for LaTeX users is
##
## @Article{,
##   title = {circlize implements and enhances circular visualization in R},
##   author = {Zuguang Gu and Lei Gu and Roland Eils and Matthias Schlesner and Benedikt Brors},
##   journal = {Bioinformatics},
##   volume = {30},
##   issue = {19},
##   pages = {2811-2812},
##   year = {2014},
## }
##
## This free open-source software implements academic research by the
## authors and co-workers. If you use it, please support the project by
## citing the appropriate journal articles.

##
## The methods within the code package can be cited as:
##
## Gu, Z. (2016) Complex heatmaps reveal patterns and correlations in
## multidimensional genomic data. Bioinformatics.
##
## A BibTeX entry for LaTeX users is
##
## @Article{,
##   title = {Complex heatmaps reveal patterns and correlations in multidimensional genomic data},
##   author = {Zuguang Gu and Roland Eils and Matthias Schlesner},
##   journal = {Bioinformatics},
##   year = {2016},
## }
##
## This free open-source software implements academic research by the
## authors and co-workers. If you use it, please support the project by
## citing the appropriate journal articles.

```

- Anders, Simon, Alejandro Reyes, and Wolfgang Huber. 2012. “Detecting differential usage of exons from RNA-seq data.” *Genome Research* 22 (10): 2008–17. doi:10.1101/gr.133744.111.
- 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.
- McKenzie, Andrew T., Igor Katsyv, Won-Min Song, Minghui Wang, and Bin Zhang. 2016. “DGCA: A comprehensive R package for Differential Gene Correlation Analysis.” *BMC Systems Biology* 10 (1). BMC Systems Biology: 106. doi:10.1186/s12918-016-0349-1.
- Mohapatra, Pratyajit, and Karen Menuz. 2019. “Molecular Profiling of the Drosophila Antenna Reveals Conserved Genes Underlying Olfaction in Insects.” *G3' Genes/Genomes/Genetics* 9 (11): 3753–71. doi:10.1534/g3.119.400669.
- 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.