analysis for Divergence of Alternative Splicing Patterns Between Paralogs in *Arabidopsis thaliana*

Filtering

The first order of business is to exclude from our dataset an pair of paralogs wherein one partner has a read count of 0 in *any* of the 3 replicates.

The complete dataset (ABC_raw_lens.tsv) has 28775 genes, of which we will discard 8110 where there are 0 reads; leaving us with 20665 genes for filtering. To be conservative, we will then filter by the variance among replicates, since genes where the number of reads varies greatly between replicates may indicate places where the molecular biology is error-prone.

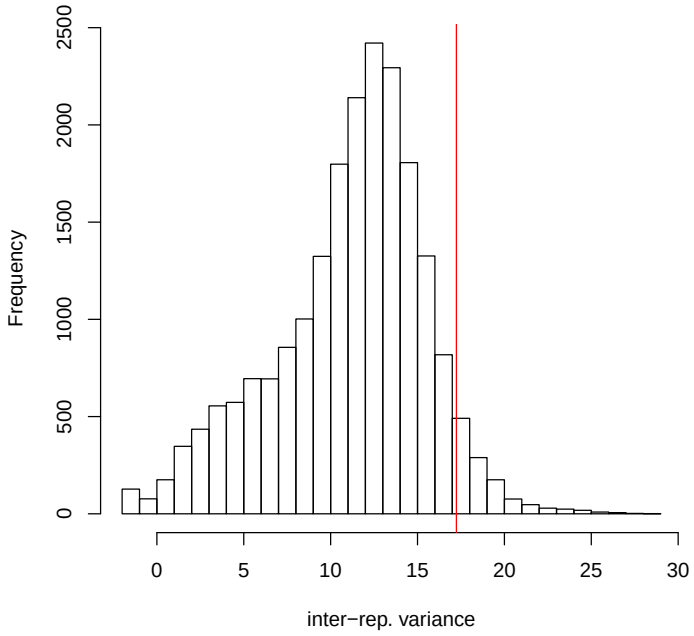

Figure 1: trimming top 5% by variance

Notably, there is an inflection point at approximately the 95% quantile and so we will truncate the dataset to remove the genes that are the 5% most variable among replicates. This strips out a further 1034 genes for a total of 19631 (fig1). *EDIT* – the inflection point is less inflect-y with the new data...I think it still

makes sense to dump the most variably-read genes though, since this will exclude rows with the least consistent no. reads.

Lastly, we will examine the data for visual outliers (fig2).

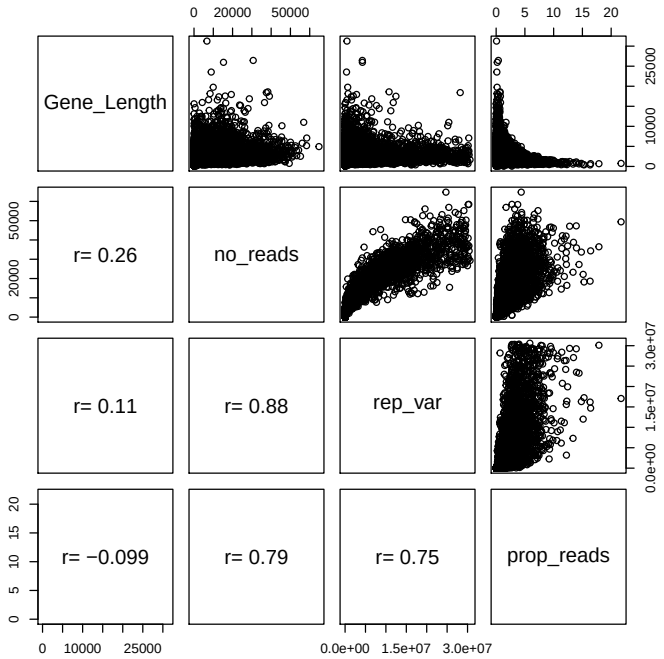

Figure 2: visual check for 'weirdos' in data

There are no genes remaining that seem obviously weird. If we decide to do more filtering, extra steps should be included here. NB: Added trimming step! I have also removed a few rows where there were alternative reads but no constitutive reads because they seemed weird. Trimming pairs from the junction datasets leaves 1443 tandem junctions and 4079 alpha junctions going into analysis.

```
[1] 3251   19
```

```
[1] 3243   19
```

```
[1] 3218   19
```

```
[1] 1052   19
```

```
[1] 1042   19
```

```
[1] 1035   19
```

```
[1] 3251   20
```

```
[1] 1052   20
```

Qualitative analysis

The first question to address is simple; when one paralog has alternatively spliced reads, how often does the other paralog in the pair also have alternatively spliced reads?

```
 [1] "gene_name"   "species"     "junct"       "class"       "A_alt"
 [6] "A_nrm"       "B_alt"       "B_nrm"       "C_alt"       "C_nrm"
[11] "unique_code"
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
     3.0     11.0     23.0    140.7     62.0 721500.0
```

```
  1%   2% 2.5%   3%   4%   5%  10%  15%
   3    4    4    4    4    5    6    8
```

```
  class min_read_alts quant_1 quant_2 quant_2.5 quant_3 quant_4 quant_5
1  ALTA             3    3.00       3         4       4       4       4
2  ALTD             3    3.00       3         4       4       4       4
3  ALTP             3    3.00       3         3       3       4       4
4  CPLX             3    4.00       4         4       4       4       4
5  CREX             3    3.15       4         4       4       4       5
6  CRIN             3    3.00       3         3       3       4       4
7    IR             3    3.00       4         4       4       5       5
8  SKAD             3    3.38       4         4       4       4       4
9  SKIP             3    3.00       4         4       4       4       4
  quant_10 quant_15
1        5        6
2        6        7
3        4        5
4        5        6
5        5        6
6        5        6
7        7        9
8        5        6
9        6        7
```

First thing to note is that there are no cases where *neither* paralog has zero alternatively spliced reads.

First interesting finding: percentage of events which paralogs have the same event at the equivalent junction.

```
[1] "alphas"

both  one
0.31 0.69

[1] "tandems"

both  one
0.34 0.66
```

	conserved	not conserved	%age conservation
IR	881	1500	37.0
ALTA	37	346	9.7
ALTD	12	220	5.2
ALTP	0	35	0.0
total	930	2101	30.7

Table 1: alphas

	conserved	not conserved	%age conservation
IR	279	411	40.4
ALTA	22	119	15.6
ALTD	8	62	11.4
ALTP	2	17	10.5
total	311	609	33.8

Table 2: tandems

```
both  one
 930 2101

both  one
 930 2101

both  one
 930 2100

both  one
 311  609

both  one
 311  608

both  one
 311  607
```

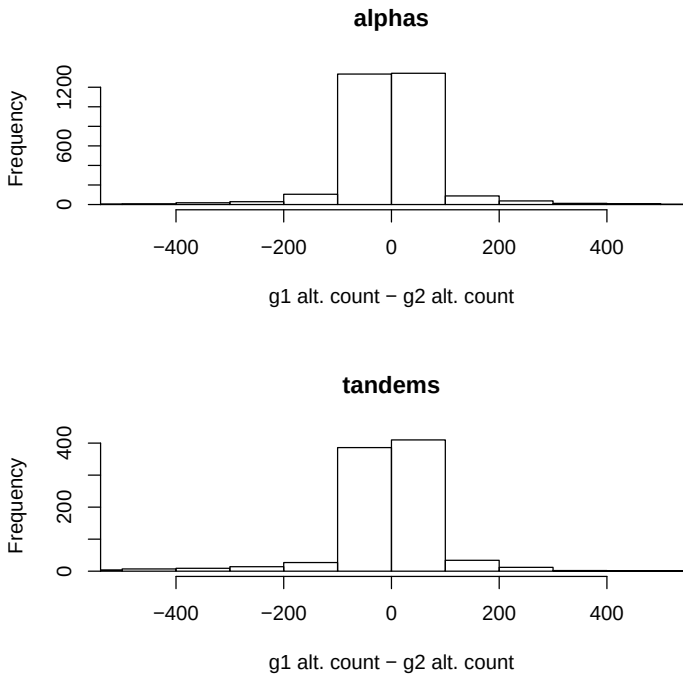

Figure 3: Difference in alternative splice read count between paralogs

The patterns are pretty similar between tandems and alphas, in that only 29–32% of junctions show alternative splicing at *both* paralogs. Another way to look at this question would be to look at the difference in the number of alternative reads (fig3).

I've trimmed the x-axis of these plots because there are a few very large differences that, if included, make the plot look like 1 column in the middle of a field. The differences max out at 6950 in the alphas and at 6172 in the tandems. Of course, this does not take account of how much each paralog is being expressed, so as an alternative I'm going to standardize the counts of alternative splicing events by the respective counts of constitutive splicing events. Seeing as the paralogs 'g1' & 'g2' are arbitrary (I assume), I'm also going to express this proportional difference in absolute terms (fig4).

This graph needs less trimming, but the point to take away is that in those pairs where both paralogs are alternatively splicing, both partners appear to be alternatively splicing at similar levels in the vast majority of cases. Another way of expressing this would be to note the maximum and median levels of difference in counts; for the alphas these are max= 6950 and median= 18, and for the tandems these are max= 6172 and median= 19, i.e. the 'typical' level

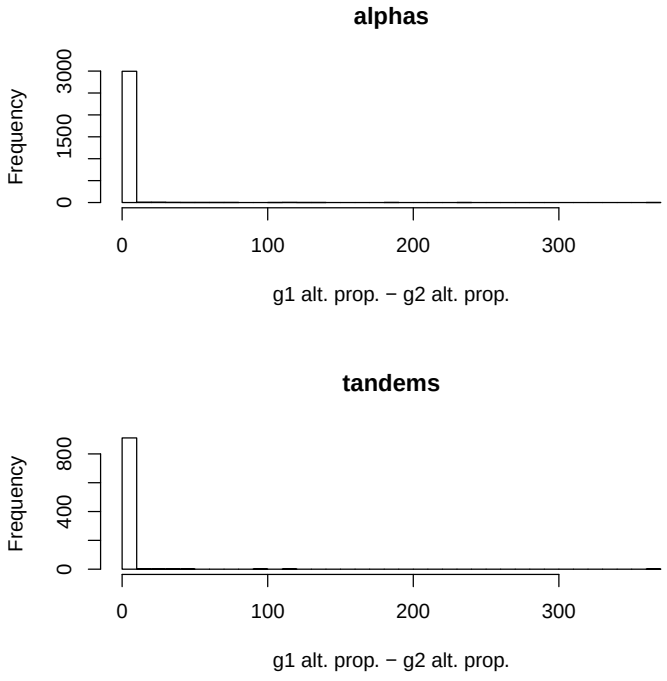

Figure 4: Difference in alternative splice read proportion between paralogs

of difference is comparatively small.

And finally, while we're working qualitatively, I'm going to try and address the question of whether event class/category is at all predictive of (lack of) conservation. Looking at the figures (fig5 & fig6) there don't appear to be any striking differences in the means, though some classes are more variable than others. When I run linear models to fit the difference in alternative reads as a function of event class I find that in the alphas there is statistical support for 'IR' having smaller differences than 'ALTA', but this model is probably overpowered due to the big sample size, since the adjusted R-squared is approx. 0.01 (i.e. the model accounts for 1% of the variation). In the tandems, there is a different picture, with the only statistically supported diffrence being that 'ALTP' junctions have larger differences than the other types. (model still accounting for only 4% variance though) Long story short; there is no compelling evidence that class is particulary informative about qualitative conservation.

```
[1] "alphas"

Call:
lm(formula = abs(junction_a$g1_all_alts - junction_a$g2_all_alts) ~
```

```
    junction_a$class2)

Residuals:
   Min     1Q Median     3Q    Max
-121.1  -44.9  -34.9  -10.9 6828.9

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            105.30      11.91   8.838  < 2e-16 ***
junction_a$class2ALTD   15.78      19.40   0.813    0.416
junction_a$class2ALTP  -41.50      41.17  -1.008    0.314
junction_a$class2IR    -55.44      12.84  -4.319 1.62e-05 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 233.2 on 3027 degrees of freedom
Multiple R-squared:  0.01128,        Adjusted R-squared:  0.0103
F-statistic: 11.52 on 3 and 3027 DF,  p-value: 1.665e-07

ALTA ALTD ALTP   IR
 383  232   35 2381

        Pearson's Chi-squared test with simulated p-value (based on 2000
        replicates)

data:  tab_mat_a[, 1:2]
X-squared = 210.7568, df = NA, p-value = 0.0004998

          [,1]       [,2]
[1,] 150.43913 -150.43913
[2,] -80.51567   80.51567
[3,] -59.18443   59.18443
[4,] -10.73903   10.73903
[5,]   0.00000    0.00000

[1] "IR"   "ALTP" "ALTD" "ALTA"

[1] "tandems"

Call:
lm(formula = abs(junction_t$g1_all_alts - junction_t$g2_all_alts) ~
    junction_t$class2)

Residuals:
   Min     1Q Median     3Q    Max
-474.5  -51.1  -39.1  -10.1 5697.5
```

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)              46.77      21.41   2.185   0.0291 *
junction_t$class2ALTD    46.37      37.17   1.248   0.2125
junction_t$class2ALTP   427.75      62.12   6.886 1.07e-11 ***
junction_t$class2IR      12.31      23.49   0.524   0.6003
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 254.2 on 916 degrees of freedom
Multiple R-squared:  0.05285,        Adjusted R-squared:  0.04975
F-statistic: 17.04 on 3 and 916 DF,  p-value: 8.966e-11

ALTA ALTD ALTP   IR
 141   70   19  690

        Pearson's Chi-squared test with simulated p-value (based on 2000
        replicates)

data:  tab_mat_t[, 1:2]
X-squared = 54.6943, df = NA, p-value = 0.0004998

           [,1]       [,2]
[1,]  45.750000 -45.750000
[2,] -25.664130  25.664130
[3,] -15.663043  15.663043
[4,]  -4.422826   4.422826
[5,]   0.000000   0.000000

[1] "IR"   "ALTP" "ALTD" "ALTA"
```

1 Quantitative analyses

In this section, I'm going to take the subset of the data where both paralogs are splicing alternatively and examine patterns quantitatively. To do this I'm fitting a logistic regression to the constitutive/alternative counts from each pair. Once again, the alphas and the tandems show similar levels of difference, with statistical support for quantitative differences in 58–68% of pairs (where both paralogs are splicing alternatively). The effects sizes have a very skewed distribution (fig7), with many small differences and a few larger ones, with only a very few very large differences. These few oddballs (far off to the right in fig7) would be one set of potentially 'interesting' candidates for biological story-telling.

Now I'm going to make some stacked bar charts These at least should go some way to illustrating the low level of event conservation that we're dealing with.

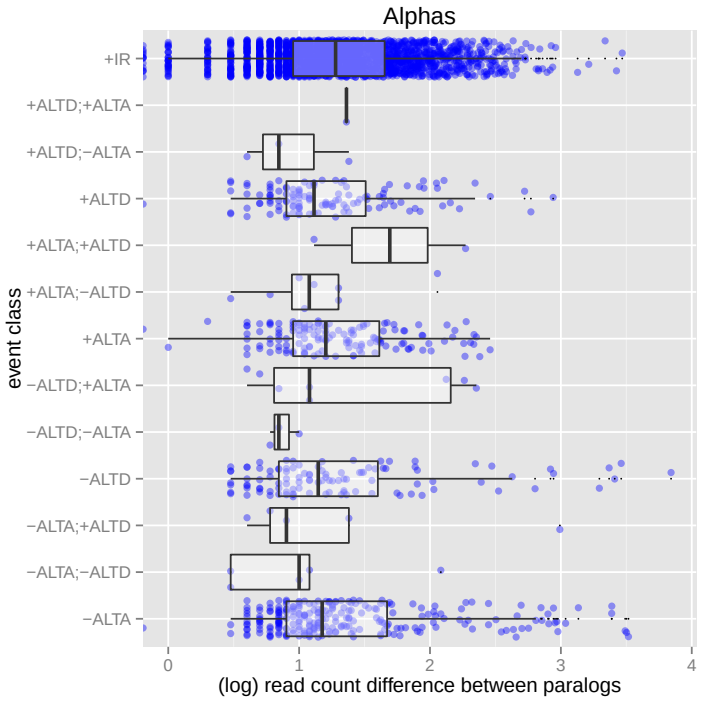

Figure 5: Difference in alternative splice read no.s by event class (alphas)

Nonsense Mediated Decay

The next question to address is whether singletons or duplicates are more likely to experience NMD. I shall test this with a Chi-squared – note that the observed and expected numbers are not *radically* different, and that p-values are large and the effect size coefficients are correspondingly small.

```
[1] "observed values"
```

```
       w/o with
sing  2868  195
dup  22592 1724
```

```
[1] "expected values"
```

```
           w/o      with
sing  2848.314  214.6863
dup  22611.686 1704.3137
```

```
[1] "statistics"
```

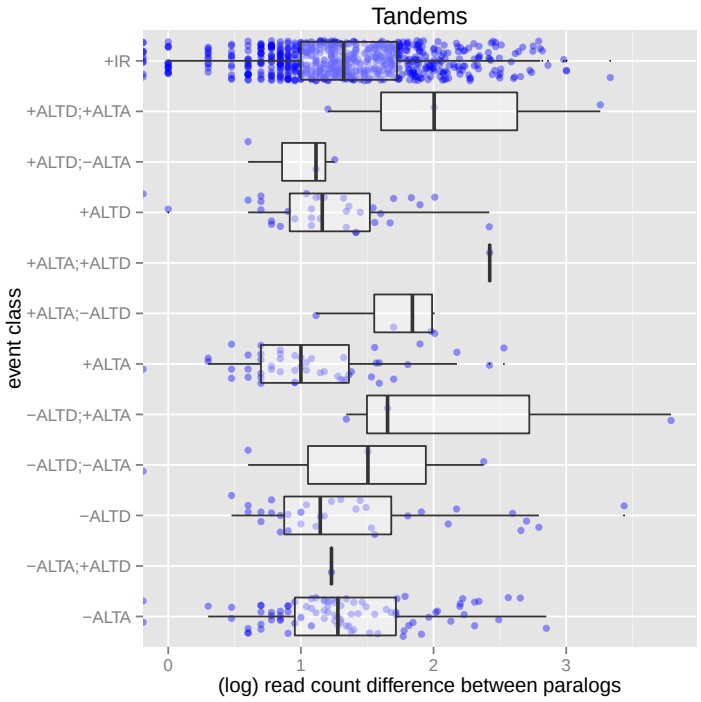

Figure 6: Difference in alternative splice read no.s by event class (tandems)

```
                   X^2 df P(> X^2)
Likelihood Ratio 2.2425  1  0.13427
Pearson          2.1858  1  0.13929

Phi-Coefficient   : 0.009
Contingency Coeff.: 0.009
Cramer's V        : 0.009
```

Next I'll use another chi-squared to test for a pattern in NMD between paralog pairs. Here there appears to be something interesting going on. Note that the off-diagonal values (where only 1 partner experiences NMD) are somewhat smaller than expected, and that the number of pairs where both paralogs experience NMD is *3 times* the expected value. Paralog pairs seem to be more likely to experience NMD – note tiny p-values and comparatively large effect size coefficients.

```
[1] "observed values"
```

```
     N  Y
N 1317 90
```

	conserved	not conserved	%age conservation
IR	279	602	31.7
ALTA	14	23	37.8
ALTD	2	10	16.7
ALTP	0	0	
total	295	635	31.7

Table 3: alphas - quantitative conservation

	conserved	not conserved	%age conservation
IR	116	163	41.6
ALTA	8	14	36.4
ALTD	4	4	50.0
ALTP	2	0	100.0
total	130	181	41.8

Table 4: tandems - quantitative conservation

```
Y   98 34

[1] "expected values"

          N         Y
N 1293.6355 113.36452
Y  121.3645  10.63548

[1] "statistics"

                   X^2 df   P(> X^2)
Likelihood Ratio 42.721  1 6.3129e-11
Pearson          61.064  1 5.5511e-15

Phi-Coefficient   : 0.199
Contingency Coeff.: 0.195
Cramer's V        : 0.199
```

Now I'm going to read in the 'NMD_out_–' files and combine them with the junciton data in order to ask some more quantitative questions about NMD.

```
[1] "do some kinds of junctions get more nmd?"

[1] "alphas"

                                    Estimate Std. Error     z value  Pr(>|z|)
(Intercept)                     -0.027028672 0.08220700 -0.32878797 0.7423160
junction_a_nmd$class-ALTA;-ALTD  0.027028672 0.58317349  0.04634757 0.9630332
junction_a_nmd$class-ALTA;+ALTD -0.309443564 0.59128257 -0.52334295 0.6007356
```

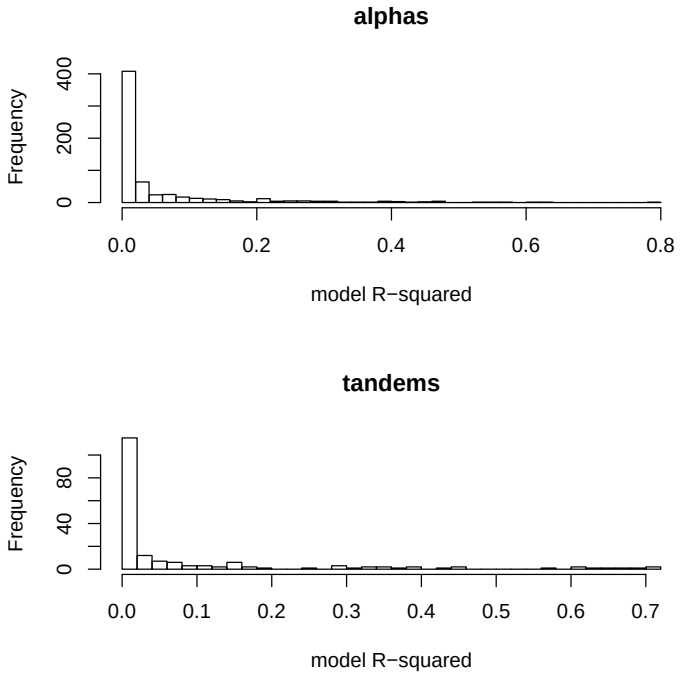

Figure 7: Distribution of effect sizes for those pairs with significantly different patterns of alternative vs. constitutive splicing

```
junction_a_nmd$class-ALTD        0.011403354 0.14961265  0.07621919 0.9392447
junction_a_nmd$class-ALTD;-ALTA -0.196114879 0.67583871 -0.29018000 0.7716785
junction_a_nmd$class-ALTD;+ALTA -0.078331843 0.46676450 -0.16781877 0.8667259
junction_a_nmd$class+ALTA        0.011028331 0.13200508  0.08354475 0.9334184
junction_a_nmd$class+ALTA;-ALTD -0.090754363 0.49281751 -0.18415410 0.8538926
junction_a_nmd$class+ALTA;+ALTD  0.314710745 0.76817398  0.40968681 0.6820357
junction_a_nmd$class+ALTD       -0.009119842 0.13714999 -0.06649539 0.9469834
junction_a_nmd$class+ALTD;-ALTA  0.027028672 0.71186936  0.03796858 0.9697127
junction_a_nmd$class+ALTD;+ALTA  0.720175853 1.22750001  0.58670130 0.5574043
junction_a_nmd$class+IR          0.016260442 0.08646321  0.18806197 0.8508281

[1] "tandems"

                                    Estimate Std. Error      z value  Pr(>|z|)
(Intercept)                     -0.009852296  0.1403742 -0.070185957 0.9440457
junction_t_nmd$class-ALTA;+ALTD  0.009852296  1.4211632  0.006932558 0.9944687
junction_t_nmd$class-ALTD        0.085838203  0.2653506  0.323489756 0.7463244
junction_t_nmd$class-ALTD;-ALTA  0.009852296  0.7209056  0.013666555 0.9890960
```

	conserved	not conserved	%age conservation
IR	279	2102	11.7
ALTA	14	369	3.7
ALTD	2	230	0.9
ALTP	0	35	0.0
total	295	2736	9.7

Table 5: alphas - overall conservation

	conserved	not conserved	%age conservation
IR	116	574	16.8
ALTA	8	133	5.7
ALTD	4	66	5.7
ALTP	2	17	10.5
total	130	790	14.1

Table 6: tandems - overall conservation

```
junction_t_nmd$class-ALTD;+ALTA  0.297534369  0.7765553  0.383146388 0.7016112
junction_t_nmd$class+ALTA        0.150434247  0.2344489  0.641650518 0.5211001
junction_t_nmd$class+ALTA;-ALTD  0.569468084  0.6423098  0.886594057 0.3752975
junction_t_nmd$class+ALTA;+ALTD  0.009852296  1.4211632  0.006932558 0.9944687
junction_t_nmd$class+ALTD        0.055314671  0.2553095  0.216657302 0.8284754
junction_t_nmd$class+ALTD;-ALTA  0.009852296  0.8284755  0.011892080 0.9905117
junction_t_nmd$class+ALTD;+ALTA -0.277829776  0.7765553 -0.357772030 0.7205139
junction_t_nmd$class+IR          0.059406111  0.1493790  0.397687262 0.6908607
```

```
[1] "NO <- there do not appear to be differences between junction classes"
```

```
[1] "do qualitatively conserved junctions get less nmd?"
```

```
[1] "alphas"
```

```
                                     Estimate Std. Error    z value
(Intercept)                       -0.007346222 0.04285525 -0.1714194
junction_a_nmd$one_gene_alts_onlyone -0.010128408 0.05129243 -0.1974640
                                      Pr(>|z|)
(Intercept)                       0.8638940
junction_a_nmd$one_gene_alts_onlyone 0.8434645
```

```
[1] "tandems"
```

```
                                        Estimate Std. Error    z value
(Intercept)                          0.06640245 0.07601560  0.8735373
junction_t_nmd$one_gene_alts_onlyone -0.02253577 0.09329391 -0.2415567
                                      Pr(>|z|)
(Intercept)                       0.3823703
junction_t_nmd$one_gene_alts_onlyone 0.8091237
```

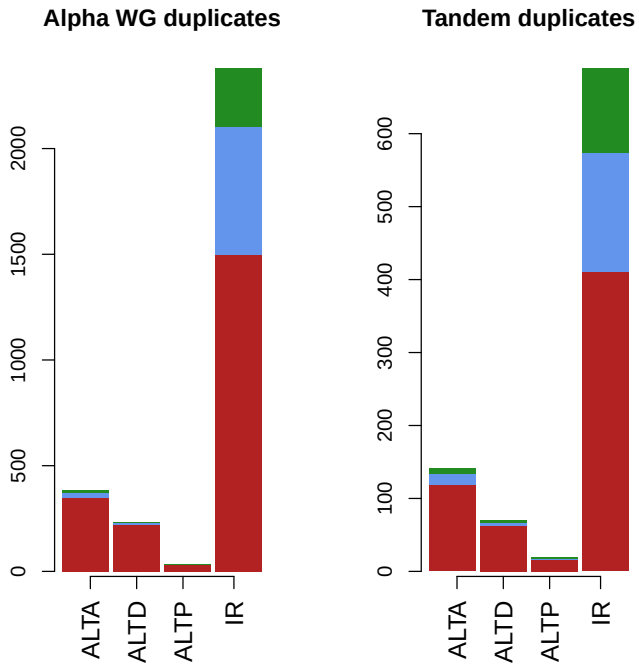

Figure 8:

```
[1] "NO  <- there appears to be no difference"

[1] "do quantitative conserved junctions get less nmd?"

[1] "alphas"

                              Estimate Std. Error      z value  Pr(>|z|)
(Intercept)                -0.0143931439 0.02474864 -0.581573042 0.5608543
junction_a_nmd$quant_conY -0.0002484061 0.08043211 -0.003088395 0.9975358

[1] "tandems"

                             Estimate Std. Error    z value  Pr(>|z|)
(Intercept)                0.04972776 0.04755296 1.04573416 0.2956838
junction_t_nmd$quant_conY 0.01214764 0.12657073 0.09597515 0.9235403

[1] "NO  <- there still appears to be no difference"

      N    Y
  N 2088  393
  Y  341  209
```

```
     N   Y
  N 724  59
  Y 112  25
```

General expression stuffs!

Question <- how close in general expression are paralog pairs? In order to answer this question we'll need go back to the *ABC_raw_lens.tsv* file. I'll look up the overall expression for each member of each pair, and calculate a proportional difference.

```
[1] "Frequency of AS"

                     Estimate Std. Error   t value     Pr(>|t|)
(Intercept)         13.616962  0.1661440 81.958801 0.0000000000
filtered$sing_dupeS -2.067086  0.5709823 -3.620228 0.0002946121

[1] "there is more AS in duplicates"

[1] "mean standardized AS per rep."

filtered$sing_dupe: D
[1] 4.538987
------------------------------------------------------------
filtered$sing_dupe: S
[1] 3.849959

[1] "variety of AS events"

              Estimate Std. Error    t value     Pr(>|t|)
(Intercept)  3.8151034 0.03000564 127.146192 0.000000e+00
sing_dupeS  -0.4085509 0.09793529  -4.171641 3.041989e-05

[1] "there is less variety of AS events among singletons..."

[1] "...but only a small difference; mean no. event types"

no_events$sing_dupe: D
[1] 1.271701
------------------------------------------------------------
no_events$sing_dupe: S
[1] 1.135517

[1] "general expression vs. AS"

               Estimate Std. Error  t value      Pr(>|t|)
(Intercept)   39.908898 1.59877284 24.96221 9.412406e-137
filtered$RKPM  3.769827 0.09462842 39.83821  0.000000e+00
```

```
[1] "there is more AS in genes which are expressed at a higher level"

Call:
lm(formula = perc_AS ~ sing_dupe, data = perc_AS)

Residuals:
   Min     1Q Median     3Q    Max
  -188   -184   -178   -162 643812

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 187.8668   205.5779   0.914    0.361
sing_dupe2   -0.1645   215.9581  -0.001    0.999

Residual standard error: 7192 on 13044 degrees of freedom
Multiple R-squared:  4.446e-11,        Adjusted R-squared:  -7.666e-05
F-statistic: 5.799e-07 on 1 and 13044 DF,  p-value: 0.9994

Call:
lm(formula = m_RKPM ~ sing_dupe, data = perc_AS)

Residuals:
    Min      1Q  Median      3Q     Max
-11.168  -7.472  -4.408   1.831 196.127

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.2003     0.3662  30.585  < 2e-16 ***
sing_dupe2   -1.4825     0.3847  -3.854 0.000117 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 12.81 on 13044 degrees of freedom
Multiple R-squared:  0.001137,        Adjusted R-squared:  0.001061
F-statistic: 14.85 on 1 and 13044 DF,  p-value: 0.0001169

Call:
lm(formula = perc_AS ~ m_RKPM, data = perc_AS)

Residuals:
   Min     1Q Median     3Q    Max
  -196   -185   -177   -158 643819

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 197.3340    79.4342   2.484    0.013 *
m_RKPM       -0.9756     4.9125  -0.199    0.843
```

```
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 7192 on 13044 degrees of freedom
Multiple R-squared:  3.024e-06,        Adjusted R-squared:  -7.364e-05
F-statistic: 0.03944 on 1 and 13044 DF,  p-value: 0.8426

[1] -0.001738838
```

Is there a relationship between NMD status and expression?

```
Call:
lm(formula = RKPM ~ nmd, data = filtered_nmd)

Residuals:
    Min      1Q  Median      3Q     Max
-12.625  -8.159  -4.166   3.927  96.254

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  11.9426     0.1293  92.335  < 2e-16 ***
nmdY          0.8237     0.3082   2.672  0.00754 **
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 12.41 on 11175 degrees of freedom
Multiple R-squared:  0.0006387,        Adjusted R-squared:  0.0005493
F-statistic: 7.142 on 1 and 11175 DF,  p-value: 0.007542
```

From this model we see a weak positive association between expression and NMD, i.e. NMD is slightly more likely to occur in genes that are more highly expressed. However, once again note that the R^2 value is tiny, indicating that this effect is unlikely to be biologically important.

Alpha vs. Tandem comparisons

This is the model that shows the difference in effect size (fold change in AS between paralogs) between alphas and tandems. The median fold change in alphas is

$$,while in tandems the value is$$

. There is *a lot* of variation within each group, which is the reason that the R^2 value appears small, but this is a solid result statistically

```
Call:
lm(formula = new$fold_change ~ new$a_t)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
 -164.5   -48.6   -46.4   -36.3 15194.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   48.997      8.463   5.790 7.59e-09 ***
new$a_tt     115.533     17.537   6.588 5.05e-11 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1

Residual standard error: 465.9 on 3949 degrees of freedom
Multiple R-squared:  0.01087,        Adjusted R-squared:  0.01062
F-statistic:  43.4 on 1 and 3949 DF,  p-value: 5.048e-11

new$a_t: a
[1] 4.088984
------------------------------------------------------------
new$a_t: t
[1] 10.22108
```

What other ways might A's and T's differ? Firstly, I could use chi-squared tests to ask whether there's more or less conservation...

```
        Pearson's Chi-squared test with simulated p-value (based on 2000
        replicates)

data:  quant
X-squared = 10.5166, df = NA, p-value = 0.002499

        con       not
A -23.49315  23.49315
T  23.49315 -23.49315

Call:
glm(formula = both_qual$sig ~ both_qual$a_t, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.5154  -1.3208   0.8736   0.8736   1.0405

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.76665    0.07046  10.881  < 2e-16 ***
both_qual$a_tt -0.43569    0.13484  -3.231  0.00123 **
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.1 âĂŸ âĂŹ 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1595.1  on 1240  degrees of freedom
Residual deviance: 1584.8  on 1239  degrees of freedom
AIC: 1588.8

Number of Fisher Scoring iterations: 4
```

Another hypothesis test: a relationship between alternative splicing frequency/variation and status as coding or untranslated region?

Evidence that there are larger fold changes in untranslated regions, but only in the alpha whole genome duplicates...

Finally, I ran a GO enrichment analysis using the GOrilla tool: Eran Eden, Roy Navon, Israel Steinfeld, Doron Lipson and Zohar Yakhini. "GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists", BMC Bioinformatics 2009, 10:48.

...aaaaaand finally-finally, extra tables for Tack with a couple of alternative thresholds.

Tables by Threshold

Qualitative Analyses

	conserved	not conserved	%age conservation
IR	881	1500	37.0
ALTA	37	346	9.7
ALTD	12	220	5.2
ALTP	0	35	0.0
total	930	2101	30.7

Table 7: alphas

	conserved	not conserved	%age conservation
IR	881	1500	37.0
ALTA	37	346	9.7
ALTD	12	220	5.2
ALTP	0	35	0.0
total	930	2101	30.7

Table 8: alphas - qualitative - threshold > 5

Quantitative Analyses

	conserved	not conserved	%age conservation
IR	881	1499	37.0
ALTA	37	346	9.7
ALTD	12	220	5.2
ALTP	0	35	0.0
total	930	2100	30.7

Table 9: alphas - qualitative - threshold > 8

	conserved	not conserved	%age conservation
IR	279	411	40.4
ALTA	22	119	15.6
ALTD	8	62	11.4
ALTP	2	17	10.5
total	311	609	33.8

Table 10: tandems

	conserved	not conserved	%age conservation
IR	279	410	40.5
ALTA	22	119	15.6
ALTD	8	62	11.4
ALTP	2	17	10.5
total	311	608	33.8

Table 11: tandems - qualitative - threshold > 5

	conserved	not conserved	%age conservation
IR	279	409	40.6
ALTA	22	119	15.6
ALTD	8	62	11.4
ALTP	2	17	10.5
total	311	607	33.9

Table 12: tandems - qualitative - threshold > 8

	conserved	not conserved	%age conservation
IR	279	602	31.7
ALTA	14	23	37.8
ALTD	2	10	16.7
ALTP	0	0	
total	295	635	31.7

Table 13: alphas - quantitative conservation

	conserved	not conserved	%age conservation
IR	279	602	31.7
ALTA	14	23	37.8
ALTD	2	10	16.7
ALTP	0	0	
total	295	635	31.7

Table 14: alpha - quantitative conservation - threshold >5

	conserved	not conserved	%age conservation
IR	279	602	31.7
ALTA	14	23	37.8
ALTD	2	10	16.7
ALTP	0	0	
total	295	635	31.7

Table 15: alpha - quantitative conservation - threshold >8

	conserved	not conserved	%age conservation
IR	116	163	41.6
ALTA	8	14	36.4
ALTD	4	4	50.0
ALTP	2	0	100.0
total	130	181	41.8

Table 16: tandems - quantitative conservation

	conserved	not conserved	%age conservation
IR	116	163	41.6
ALTA	8	14	36.4
ALTD	4	4	50.0
ALTP	2	0	100.0
total	130	181	41.8

Table 17: tandem - quantitative conservation - threshold >5

	conserved	not conserved	%age conservation
IR	116	163	41.6
ALTA	8	14	36.4
ALTD	4	4	50.0
ALTP	2	0	100.0
total	130	181	41.8

Table 18: tandem - quantitative conservation - threshold >8

	conserved	not conserved	%age conservation
IR	279	2102	11.7
ALTA	14	369	3.7
ALTD	2	230	0.9
ALTP	0	35	0.0
total	295	2736	9.7

Table 19: alphas - overall conservation

	conserved	not conserved	%age conservation
IR	279	2102	11.7
ALTA	14	369	3.7
ALTD	2	230	0.9
ALTP	0	35	0.0
total	295	2736	9.7

Table 20: alphas - overall conservation âĂŞ threshold >5

	conserved	not conserved	%age conservation
IR	279	2101	11.7
ALTA	14	369	3.7
ALTD	2	230	0.9
ALTP	0	35	0.0
total	295	2735	9.7

Table 21: alphas - overall conservation âĂŞ threshold >8

	conserved	not conserved	%age conservation
IR	116	574	16.8
ALTA	8	133	5.7
ALTD	4	66	5.7
ALTP	2	17	10.5
total	130	790	14.1

Table 22: tandems - overall conservation

	conserved	not conserved	%age conservation
IR	116	573	16.8
ALTA	8	133	5.7
ALTD	4	66	5.7
ALTP	2	17	10.5
total	130	789	14.1

Table 23: tandems - overall conservation âĂŞ threshold >5

	conserved	not conserved	%age conservation
IR	116	572	16.9
ALTA	8	133	5.7
ALTD	4	66	5.7
ALTP	2	17	10.5
total	130	788	14.2

Table 24: tandems - overall conservation âĂŞ threshold >8