Summary M1 Mediators ADDIS
STAT 550

Jarred Kvamme

3/28/2021

Table 1: Total number of Cis and Trans genes identified as Mediators under MRPC-ADDIS

	Total.Num.Genes	Percent.Of.Total
Cis Only	1356	0.2652582
Trans Only	3674	0.7187011
Both Cis & Trans	82	0.0160407

Table 2: Total number of Unqiue Cis and Trans genes identified as Mediators under MRPC-ADDIS

	Total.Unique.Genes	Percent.Of.Total
Cis Only Unique	730	0.1994536
Trans Only Unique	2848	0.7781421
Both Cis & Trans	82	0.0224044

Frequency of Shared Tissues for Trans Mediators   **Frequency of Shared Tissues for Cis Mediators**
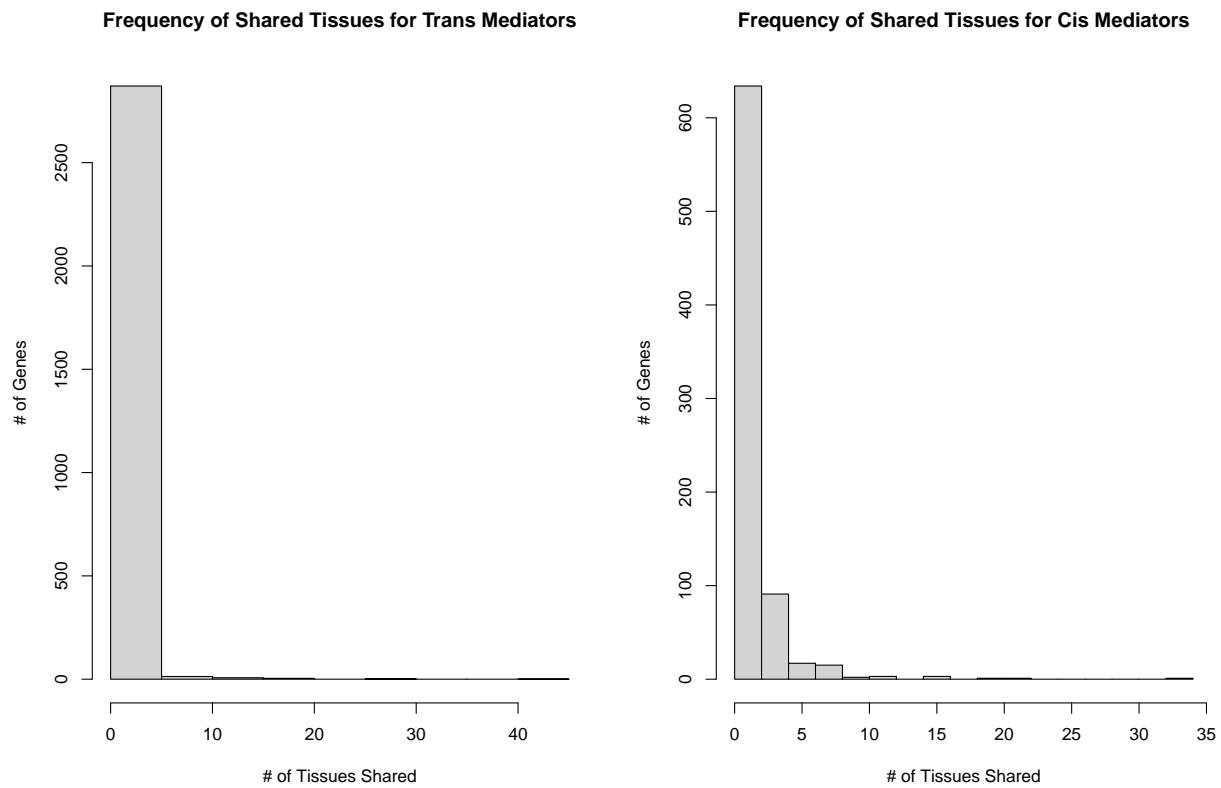
Table 3: ADDIS Trans Mediator Tissue Counts

Num.of.Genes	Num.of.Shared
2656	1
160	2
34	3
13	4
8	5
3	6
1	7
4	8
2	9
3	10
3	11
2	13
1	14
1	15
1	16
1	17
2	20
1	26
1	30

Num.of.Genes	Num.of.Shared
1	44

Table 4: ADDIS Cis Mediator Tissue Counts

Num.of.Genes	Num.of.Shared
514	1
120	2
60	3
31	4
10	5
7	6
7	7
8	8
2	10
2	11
1	12
2	15
1	16
1	20
1	21
1	34

Table 5: ADDIS Unique Cis Mediator Gene Types

	Percentage
IG_V_gene	0.00685
IG_V_pseudogene	0.00274
lncRNA	0.26301
miRNA	0.00137
polymorphic_pseudogene	0.00274
processed_pseudogene	0.04658
protein_coding	0.50274
TEC	0.01918
TR_V_gene	0.00137
TR_V_pseudogene	0.00137
transcribed_processed_pseudogene	0.02055
transcribed_unitary_pseudogene	0.00685
transcribed_unprocessed_pseudogene	0.07260
translated_unprocessed_pseudogene	0.00137
unprocessed_pseudogene	0.03973

	Percentage
NA's	0.01096

Table 6: ADDIS Unique Trans Mediator Gene Types

	Percentage
IG_C_gene	0.00105
IG_D_gene	0.00316
IG_J_gene	0.00070
IG_V_gene	0.00667
IG_V_pseudogene	0.00632
lncRNA	0.20154
miRNA	0.06812
misc_RNA	0.05969
Mt_tRNA	0.00140
polymorphic_pseudogene	0.00035
processed_pseudogene	0.18645
protein_coding	0.24122
pseudogene	0.00035
ribozyme	0.00035
rRNA	0.00211
rRNA_pseudogene	0.01369
scaRNA	0.00035
snoRNA	0.02669
snRNA	0.06250
TEC	0.00808
TR_C_gene	0.00035
TR_J_gene	0.00421
TR_J_pseudogene	0.00070
TR_V_gene	0.00176
TR_V_pseudogene	0.00105
transcribed_processed_pseudogene	0.00772
transcribed_unitary_pseudogene	0.00105
transcribed_unprocessed_pseudogene	0.02247
unitary_pseudogene	0.00105
unprocessed_pseudogene	0.06742
NA's	0.00140

Table 7: ADDIS gene types for genes found as both Cis
and Trans Mediators

Cis.Gene.Type	Percent.Cis	Trans.Gene.Type	Percent.Trans
lncRNA	0.18293	IG_V_gene	0.01220
processed_pseudogene	0.04878	lncRNA	0.17073
protein_coding	0.59756	processed_pseudogene	0.10976
transcribed_processed_pseudogene	0.03659	protein_coding	0.43902
transcribed_unprocessed_pseudogene	0.08537	transcribed_processed_pseudogene	0.01220
unprocessed_pseudogene	0.03659	transcribed_unitary_pseudogene	0.01220
NA's	0.0122	transcribed_unprocessed_pseudogene	0.13415
–	–	unprocessed_pseudogene	0.10976

Analysis of Gene Type

To analyze the mediator genes classified under MRPC-ADDIS, we employed hypothesis
testing to determine if there was dependence among a mediator gene's position (2 levels:
Cis or Trans) and its gene type (2 levels: Pseudogene or Non-pseudogene). We conducted a
Chi-Squared test of Independence with the null and alternative hypotheses:

$$H_0 : \text{Gene type is independent of mediator position}$$

$$H_A : \text{Gene type is not independent of mediator position}$$

The resulting 2×2 contingency table is given in **table 8**. The test yielded a $\chi^2_{(1)} = 43.77$, $p = 1.489e^{-09}$ and we therefore rejected the null hypothesis of no dependence. From **table 8** the
resulting rejection is due to an approximately 10% enrichment of pseudogene types among
trans mediators.

```
## 
##  Pearson's Chi-squared test with Yates' continuity correction
## 
## data:  as.matrix(cont.table)
## X-squared = 36.549, df = 1, p-value = 1.489e-09
```

Table 8: 2X2 contingency table of mediator gene type

	cis	%cis	trans	%trans	Col.Total
non-pseudo	588	0.8055	1969	0.6914	2557
pseudo	142	0.1945	879	0.3086	1021
row.total	730	1.0000	2848	1.0000	3578

To further investigate the possibility of trans mediator pseudogene enrichment, we employed a Chi-Squared Goodness of Fit test to determine if enrichment was comparable to the proportion of pseudogenes present in the whole genome. In this case, the vector of probabilities $\mathbf{p}$ is taken to be the observed proportion of pseudo and non-pseudogene types among trans mediator genes and $\mathbf{p_0}$ is the expected proportions given by the proportion of pseudo/non-pseudogene types in the entire genome. This leads to the null and alternative hypotheses:

$$H_0 : \mathbf{p} = \mathbf{p_0}$$

$$H_A : \mathbf{p} \neq \mathbf{p_0}$$

The test resulted in a $\chi^2_{(1)} = 43.77$, $p = 3.68e^{-11}$ and we therefore rejected the null hypothesis of equality of observed and expected proportions. **Table 9** indicates that the rejection is due to a roughly 5.4% increase of pseudogene gene types among trans mediators relative to the entire genome.

It is important to note that, because the test statistic in both tests is asymptotically χ^2, the large sample sizes translate into high power and therefore a small difference is translated into a null hypothesis rejection.

```
##
##  Chi-squared test for given probabilities
##
## data:  summary(trans.types)
## X-squared = 43.777, df = 1, p-value = 3.68e-11
```

Table 9: Chi-Square GOF observed vs. Expected proportions ADDIS

	non-pseudo	%non-pseudo	pseudogene	%pseudogene
observed	1969	0.6914	879	0.3086
expected	50037	0.7454	17093	0.2546

BackCheck Chi-Square with LOND

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  as.matrix(Lcont.table)
## X-squared = 50.456, df = 1, p-value = 1.219e-12
```

Table 10: 2X2 contingency table of mediator gene type

	cis	trans	margins
non-pseudo	819	1313	2132
pseudo	186	585	771
margins	1005	1898	2903

```
##
##  Chi-squared test for given probabilities
##
## data:  summary(L.trans.types)
## X-squared = 28.724, df = 1, p-value = 8.345e-08
```

Table 11: Chi-Square GOF observed vs. Expected proportions LOND

	non-pseudo	pseudogene
observed	1313	585
expected	50037	17093