

# ADDIS Summary of Mediators GTExV8

Audrey Fu Lab

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	1438	0.2725550
Trans Only	3756	0.7119030
Both Cis & Trans	82	0.0155421

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

Description	Total.Unique.Genes	Percent.Of.Total
Cis Only Unique	686	0.1914062
Trans Only Unique	2816	0.7857143
Both Cis & Trans	82	0.0228795
Total	3584	1.0000000

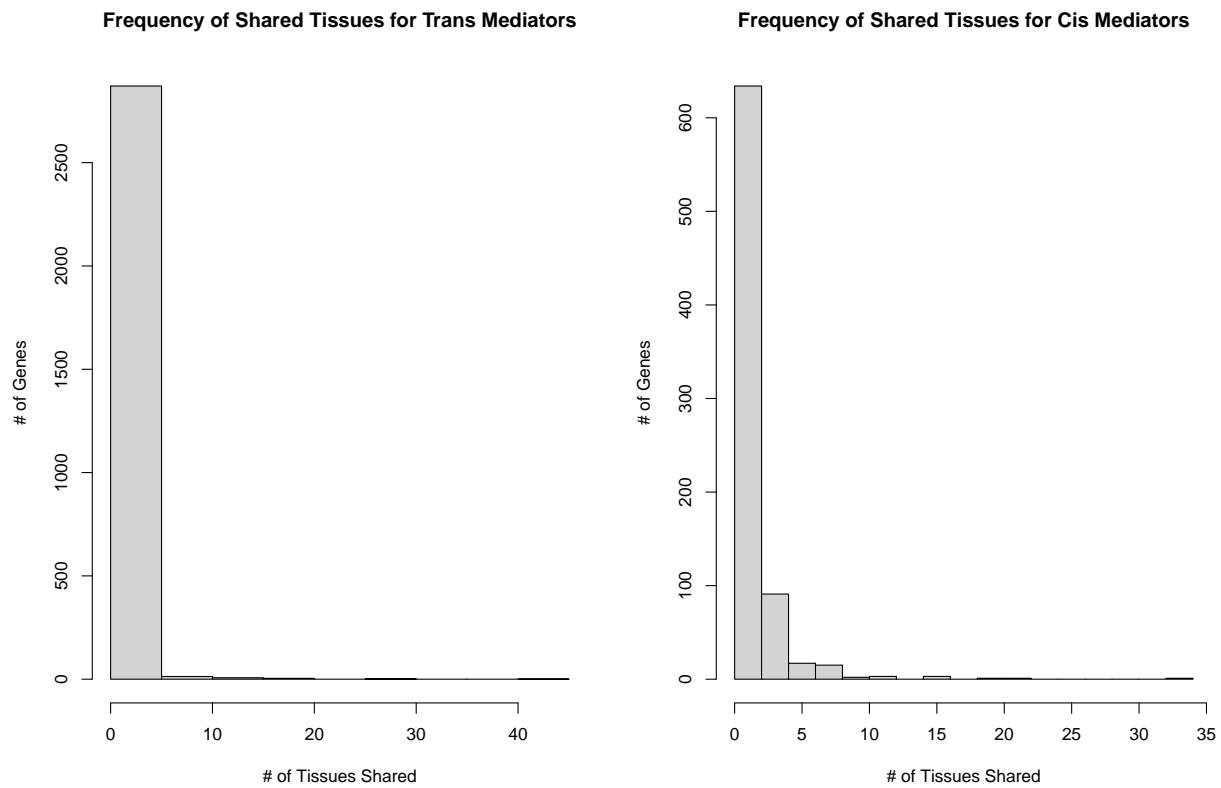


Table 3: ADDIS Trans Mediator Tissue Counts

Number of Genes	Number of Tissues Shared
2656	1
160	2
34	3
13	4
8	5
3	6
1	7
4	8
2	9
3	10
3	11
0	12
2	13
1	14
1	15
1	16
1	17
0	18
0	19

Number of Genes	Number of Tissues Shared
2	20
0	21
0	22
0	23
0	24
0	25
1	26
0	27
0	28
0	29
1	30
0	31
0	32
0	33
0	34
0	35
0	36
0	37
0	38
0	39
0	40
0	41
0	42
0	43
1	44
0	45
0	46
0	47
0	48

Table 4: ADDIS Cis Mediator Tissue Counts

Number of Genes	Number of Tissues Shared
514	1
120	2
60	3
31	4
10	5
7	6
7	7
8	8
0	9

Number of Genes	Number of Tissues Shared
2	10
2	11
1	12
0	13
0	14
2	15
1	16
0	17
0	18
0	19
1	20
1	21
0	22
0	23
0	24
0	25
0	26
0	27
0	28
0	29
0	30
0	31
0	32
0	33
1	34
0	35
0	36
0	37
0	38
0	39
0	40
0	41
0	42
0	43
0	44
0	45
0	46
0	47
0	48

Table 5: ADDIS Unique Cis Mediator Gene Types

	Percentage
antisense	0.00146
IG_V_gene	0.00583
IG_V_pseudogene	0.00292
lncRNA	0.27259
miRNA	0.00146
polymorphic_pseudogene	0.00292
processed_pseudogene	0.04373
protein_coding	0.50875
pseudogene	0.00292
sense_intronic	0.00292
TEC	0.02041
TR_V_gene	0.00146
TR_V_pseudogene	0.00146
transcribed_processed_pseudogene	0.02187
transcribed_unitary_pseudogene	0.00729
transcribed_unprocessed_pseudogene	0.06268
translated_unprocessed_pseudogene	0.00146
unprocessed_pseudogene	0.03790

Table 6: ADDIS Unique Trans Mediator Gene Types

	Percentage
IG_C_gene	0.00107
IG_D_gene	0.00320
IG_J_gene	0.00071
IG_V_gene	0.00639
IG_V_pseudogene	0.00639
lncRNA	0.20312
miRNA	0.06889
misc_RNA	0.06037
Mt_tRNA	0.00142
polymorphic_pseudogene	0.00036
processed_pseudogene	0.18679
protein_coding	0.24041
pseudogene	0.00036
ribozyme	0.00036
rRNA	0.00284
rRNA_pseudogene	0.01385
scaRNA	0.00036
snoRNA	0.02699

	Percentage
snRNA	0.06321
TEC	0.00817
TR_C_gene	0.00036
TR_J_gene	0.00426
TR_J_pseudogene	0.00071
TR_V_gene	0.00178
TR_V_pseudogene	0.00107
transcribed_processed_pseudogene	0.00781
transcribed_unitary_pseudogene	0.00107
transcribed_unprocessed_pseudogene	0.02024
unitary_pseudogene	0.00107
unprocessed_pseudogene	0.06641

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

Gene.Type	Percent
IG_V_gene	0.01220
lncRNA	0.17073
processed_pseudogene	0.10976
protein_coding	0.43902
transcribed_processed_pseudogene	0.01220
transcribed_unitary_pseudogene	0.01220
transcribed_unprocessed_pseudogene	0.13415
unprocessed_pseudogene	0.10976

## [1] 1  
## [1] 2  
## [1] 3

## [1] 1  
## [1] 2  
## [1] 3

## [1] 1  
## [1] 2  
## [1] 3

Table 8: ADDIS Unique cis mediator gene types after grouping all gene types into 4 categories: pseudogene, protein\_coding, lncRNA, and Others

Type	Count	Proportion
pseudogene	127	0.1851312
protein_coding	349	0.5087464
lncRNA	187	0.2725948
others	23	0.0335277

Table 9: ADDIS Unique trans mediator gene types after grouping all gene types into 4 categories: pseudogene, protein\_coding, lncRNA, and Others

Type	Count	Proportion
pseudogene	862	0.3061080
protein_coding	677	0.2404119
lncRNA	572	0.2031250
others	705	0.2503551

Table 10: Gene types in the entire genome after grouping all gene types into 4 categories: pseudogene, protein\_coding, lncRNA, and Others

Type	Count	Proportion
pseudogene	18954	0.2483686
protein_coding	24935	0.3267421
lncRNA	18843	0.2469141
others	13582	0.1779752

## Analysis of Gene Type

### Detecting Pseudogene Enrichment

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

$H_0$  : Gene type is independent of mediator position

$H_A$  : Gene type is not independent of mediator position

The resulting  $2 \times 2$  contingency table is given in **table 11 and 12**. The test yielded a  $\chi^2_{(1)} = 43.77$ ,  $p = 1.489e^{-09}$  and we therefore rejected the null hypothesis of no dependence. From **tables 11 and 12** the resulting rejection is due to an approximately 10% difference in the number of pseudogenes between Cis and Trans gene positions. However, under the current test conditions, we are unable to determine if the trans position is 10% enriched with pseudogenes or is the cis position is 10% depleted.

Table 11: 2X2 contingency table comparing the cell counts of mediator position and gene\_type:pseudogene

	cis	%cis	trans	%trans	Col.Total
non-pseudo	559	0.8149	1954	0.6939	2513
pseudo	127	0.1851	862	0.3061	989
row.total	686	1.0000	2816	1.0000	3502

Table 12: Chi-Squared Test of Independence Summary:  
gene.type = pseudogene

	Value
Chi-Squared	39.2424
P	0.0000
df	1.0000

To further investigate the possibility of trans mediator pseudogene enrichment, we employed a Chi-Squared Goodness of Fit test to determine if enrichment was similar 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 13** indicates that the rejection is due to a roughly 5.4% increase of pseudogene gene type among trans mediators relative to the entire genome.



It is important to note that, because the test statistic in both tests is asymptotically  $\chi^2$ , 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 = 50.289, df = 1, p-value = 1.327e-12
```

Table 13: Chi-Square GOF observed vs. Expected proportions of trans gene Type=pseudogene

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	1954	0.6939	862	0.3061
Genome	57360	0.7516	18954	0.2484

Table 14: Chi-Square GOF observed vs. Expected proportions of cis gene: Type=pseudogene

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	559	0.8149	127	0.1851
Genome	57360	0.7516	18954	0.2484

## Detecting Protein Coding Gene Enrichment

Similar to the above tests, we also explored the dependence among gene position and gene type with types consisting of protein coding or non-protein coding. Again, we conducted a Chi-square test of independence (**tables 14 and 15**). The test yielded a  $\chi^2_{(1)} = 189.98$ ,  $p \approx 0$  and we rejected the null hypothesis of no dependence. The rejection is due to an approximately 1 : 2 ratio of protein coding to non-protein coding gene types among the cis mediators and an approximately 1 : 5 ratio of protein coding to non-protein coding gene types among the trans mediators.

Table 15: 2X2 contingency table comparing the cell counts of mediator position and gene\_type:protein coding

	cis	%cis	trans	%trans	Col.Total
non-protein_coding	337	0.4913	2139	0.7596	2476

	cis	%cis	trans	%trans	Col.Total
protein_coding	349	0.5087	677	0.2404	1026
row.total	686	1.0000	2816	1.0000	3502

Table 16: Chi-Squared Test of Independence Summary:  
gene.type = protein coding

	Value
Chi-Squared	190.4534
P	0.0000
df	1.0000

We further explored protein coding gene type enrichment among cis mediators by conducting a Chi-Square GOF test to compare the relative proportion of protein-coding genes among cis mediators to the observed proportion of protein coding genes in the entire genome. The test resulted in a  $\chi^2_{(1)} = 122.72$ ,  $p \approx 0$ . we therefore rejected the null hypothesis of equality of observed and expected proportions. **Table 16** indicates that the rejection is due to an approximately 16 difference in the proportion of protein coding gene type among cis mediators relative to the entire genome.

```
##
## Chi-squared test for given probabilities
##
## data: summary(trans.types)
## X-squared = 95.405, df = 1, p-value < 2.2e-16
```

Table 17: Chi-Square GOF observed vs. Expected proportions of trans gene Type=protein\_coding

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	2139	0.7596	677	0.2404
Genome	51379	0.6733	24935	0.3267

Table 18: Chi-Square GOF observed vs. Expected proportions of cis gene: Type=protein\_coding

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	337	0.4913	349	0.5087

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Genome	51379	0.6733	24935	0.3267

## Detecting lncRNA Enrichment

Finally, we explored the dependence among gene position and gene type with type consisting of lncRNA or non-lncRNA using a Chi-square test of independence (**tables 17 and 18**). The test yielded a  $\chi^2_{(1)} = 12.68$ ,  $p \approx 0.00037$ , and we rejected the null hypothesis of no dependence. The rejection is due to an approximately 1 : 4 ratio of non-lncRNA to lncRNA gene types among the cis mediators and an approximately 1 : 5 ratio of non-lncRNA to lncRNA gene types among the trans mediators.

Table 19: 2X2 contingency table comparing the cell counts of mediator position and gene\_type:lncRNA

	cis	%cis	trans	%trans	Col.Total
non-lncRNA	187	0.2726	572	0.2031	759
lncRNA	499	0.7274	2244	0.7969	2743
row.total	686	1.0000	2816	1.0000	3502

Table 20: Chi-Squared Test of Independence Summary:  
gene.type = lncRNA

	Value
Chi-Squared	15.2752793
P	0.0000929
df	1.0000000

To Explore lncRNA gene type enrichment among trans mediators we conducted a Chi-Square GOF test to compare the relative proportion of lncRNA genes among trans mediators to the observed proportion of lncRNA genes in the entire genome. The test resulted in a  $\chi^2_{(1)} = 63.06$ ,  $p \approx 0$ . we therefore rejected the null hypothesis of equality of observed and expected proportions. **Table 19** indicates that the rejection is due to an approximately 5.5% difference in the proportion of lncRNA gene type among trans mediators relative to the entire genome.

```
##
## Chi-squared test for given probabilities
##
## data: summary(trans.types)
## X-squared = 29.038, df = 1, p-value = 7.096e-08
```

Table 21: Chi-Square GOF observed vs. Expected proportions of trans gene Type=lncRNA

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	572	0.2031	2244	0.7969
Genome	18843	0.2469	57471	0.7531

Table 22: Chi-Square GOF observed vs. Expected proportions of cis gene: Type=lncRNA

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	187	0.2726	499	0.7274
Genome	18843	0.2469	57471	0.7531