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	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

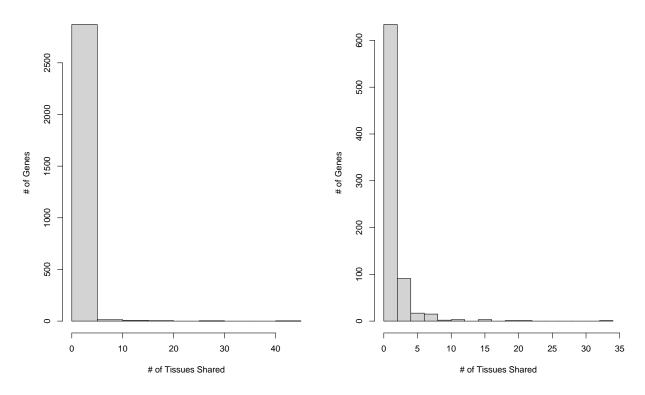|	Total.Unique.Genes	Percent.Of.Total
Cis Only Unique	730	0.1994536
Trans Only Unique	2848	0.7781421
Both Cis & Trans	82	0.0224044

Frequency of Shared Tissues for Trans Mediators **Frequency of Shared Tissues for Cis Mediators**

Table 3: ADDIS Trans Mediator Tissue Counts

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	0.00137
IG_V_gene	0.00685
IG_V_pseudogene	0.00274
lncRNA	0.26438
miRNA	0.00137
polymorphic_pseudogene	0.00274
processed_pseudogene	0.04795
protein_coding	0.50411
pseudogene	0.00274
sense_intronic	0.00274
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

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	0.20225
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	0.00281
rRNA_pseudogene	0.01369
scaRNA	0.00035
snoRNA	0.02669

	Percentage
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

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	145	0.1986301
protein_coding	368	0.5041096
lncRNA	193	0.2643836
others	24	0.0328767

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	879	0.3086376
protein_coding	687	0.2412219
lncRNA	576	0.2022472
others	706	0.2478933

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:

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

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

The resulting 2×2 contingency table is given in **table 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	585	0.8014	1969	0.6914	2554
pseudo	145	0.1986	879	0.3086	1024
row.total	730	1.0000	2848	1.0000	3578

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

	Value
Chi-Squared	33.88489
P	0.00000
df	1.00000

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 **p** is taken to be the observed proportion of pseudo and non-pseudogene types among trans mediator genes, and **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 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 = 55.415, df = 1, p-value = 9.759e-14
```

Table 13: Chi-Square GOF observed vs. expected proportions of trans gene types ADDIS: gene.type=Pseudogene

	non-pseudo	%non-pseudo	pseudogene	%pseudogene
Observed	1969	0.6914	879	0.3086
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 14: 2X2 contingency table comparing the cell counts of mediator position and gene_type:protein coding

	cis	%cis	trans	%trans	Col.Total
non-protein_coding	362	0.4959	2161	0.8014	2523
protein_coding	368	0.5041	687	0.1986	1055
row.total	730	1.0000	2848	1.0000	3578

Table 15: Chi-Squared Test of Independence Summary:
gene.type = protein coding

	Value
Chi-Squared	191.8784
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 = 94.687, df = 1, p-value < 2.2e-16
```

Table 16: Chi-Square GOF observed vs. Expected proportions of cis gene types ADDIS: Type=Protein Coding

	non-protein_coding	%non-protein_coding	protein_coding	%protein_coding
Observed	362	0.4959	368	0.5041
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 17: 2X2 contingency table comparing the cell counts of mediator position and gene_type:lncRNA

	cis	%cis	trans	%trans	Col.Total
non-lncRNA	193	0.2644	576	0.2022	769
lncRNA	537	0.7356	2272	0.7978	2809
row.total	730	1.0000	2848	1.0000	3578

Table 18: Chi-Squared Test of Independence Summary: gene.type = lncRNA

	Value
Chi-Squared	12.9301683
P	0.0003233
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 = 30.558, df = 1, p-value = 3.241e-08
```

Table 19: Chi-Square GOF observed vs. Expected proportions of gene types ADDIS: Type=lncRNA

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	576	0.2022	2272	0.7978
Genome	18843	0.2469	57471	0.7531