

LOND Summary of Mediators GTE_xV8

Audrey Fu Lab

Jarred Kvamme

4/8/2021

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

	Total.Num.Genes	Percent.Of.Total
Cis Only	1806	0.4018692
Trans Only	2597	0.5778816
Both Cis & Trans	91	0.0202492

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	966	0.3309353
Trans Only Unique	1862	0.6378897
Both Cis & Trans	91	0.0311751
Total	2919	1.0000000

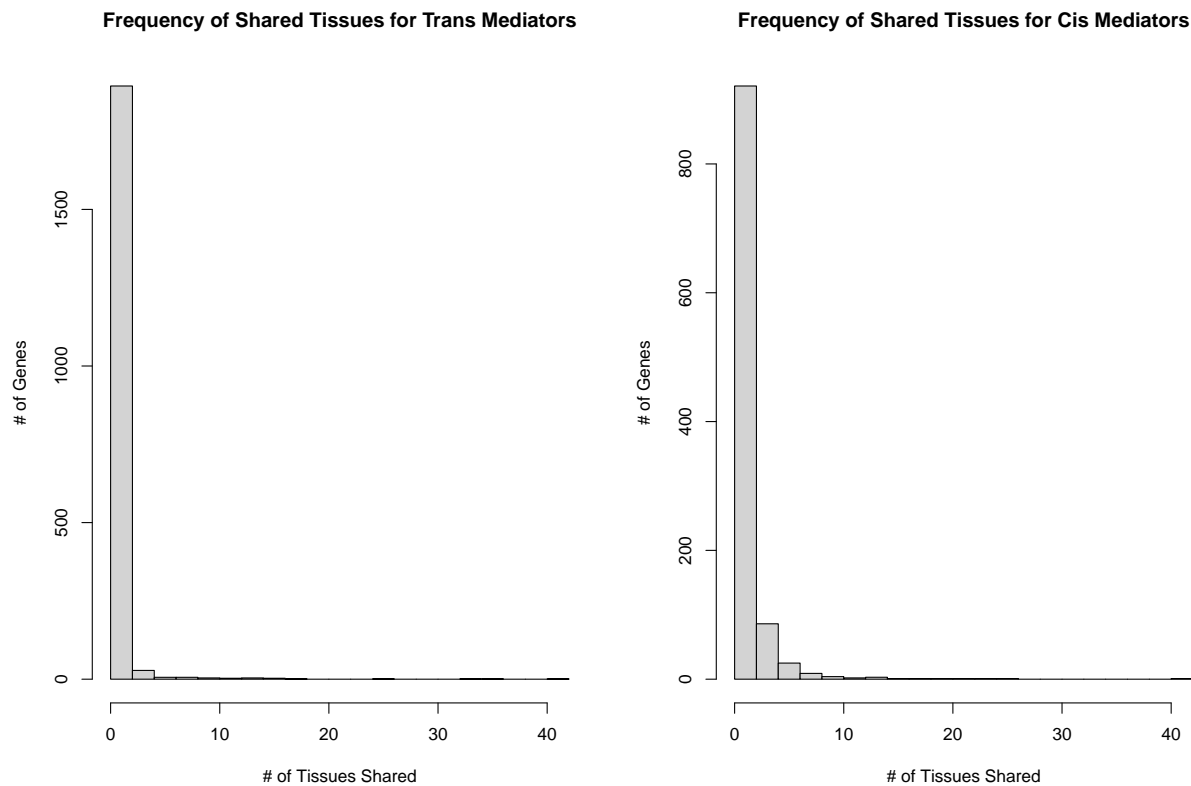


Table 3: LOND Trans Mediator Tissue Counts

Number of Genes	Number of Tissues Shared
1804	1
90	2
17	3
11	4
4	5
2	6
5	7
1	8
2	9
2	10
1	11
2	12
2	13
2	14
3	15
0	16
0	17
1	18
0	19

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

Table 4: LOND Cis Mediator Tissue Counts

Number of Genes	Number of Tissues Shared
792	1
129	2
55	3
31	4
12	5
13	6
8	7
1	8
1	9

Number of Genes	Number of Tissues Shared
3	10
2	11
0	12
2	13
1	14
1	15
0	16
0	17
1	18
0	19
1	20
1	21
0	22
0	23
1	24
1	25
0	26
0	27
0	28
0	29
0	30
0	31
0	32
0	33
0	34
0	35
0	36
0	37
0	38
0	39
0	40
1	41
0	42
0	43
0	44
0	45
0	46
0	47
0	48

Table 5: LOND Unique Cis Mediator Gene Types

	Percentage
antisense	0.00104
IG_V_gene	0.00414
IG_V_pseudogene	0.00104
lncRNA	0.26812
miRNA	0.00207
polymorphic_pseudogene	0.00104
processed_pseudogene	0.04141
protein_coding	0.52484
pseudogene	0.00104
sense_intronic	0.00311
snoRNA	0.00104
TEC	0.01553
TR_C_gene	0.00104
TR_V_gene	0.00104
TR_V_pseudogene	0.00104
transcribed_processed_pseudogene	0.01863
transcribed_unitary_pseudogene	0.00932
transcribed_unprocessed_pseudogene	0.06729
translated_unprocessed_pseudogene	0.00104
unprocessed_pseudogene	0.03623

Table 6: LOND Unique Trans Mediator Gene Types

	Percentage
IG_D_gene	0.00322
IG_J_gene	0.00054
IG_V_gene	0.00591
IG_V_pseudogene	0.00859
lncRNA	0.19656
miRNA	0.06176
misc_RNA	0.05048
Mt_tRNA	0.00215
polymorphic_pseudogene	0.00054
processed_pseudogene	0.19388
protein_coding	0.26960
rRNA	0.00054
rRNA_pseudogene	0.00859
scaRNA	0.00054
snoRNA	0.02846
snRNA	0.06176

	Percentage
TEC	0.00698
TR_J_gene	0.00376
TR_J_pseudogene	0.00054
TR_V_gene	0.00054
transcribed_processed_pseudogene	0.01020
transcribed_unprocessed_pseudogene	0.02363
unprocessed_pseudogene	0.06069
NA's	0.00054

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

Gene.Type	Percent
IG_V_gene	0.01099
lncRNA	0.15385
processed_pseudogene	0.10989
protein_coding	0.46154
TEC	0.01099
transcribed_processed_pseudogene	0.02198
transcribed_unitary_pseudogene	0.01099
transcribed_unprocessed_pseudogene	0.10989
unprocessed_pseudogene	0.10989

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

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

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

Table 8: LOND Unique cis mediator gene types after grouping all gene types into 4 categories: pseudogene, protein_coding, lncRNA, and Others

Type	Count	Proportion
pseudogene	172	0.1780538
protein_coding	507	0.5248447
lncRNA	259	0.2681159
others	28	0.0289855

Table 9: LOND Unique trans mediator gene types after grouping all gene types into 4 categories: pseudogene, protein_coding, lncRNA, and Others

Type	Count	Proportion
pseudogene	571	0.3066595
protein_coding	502	0.2696026
lncRNA	366	0.1965628
others	423	0.2271751

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-LOND, 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 2×2 contingency table is given in **table 11 and 12**. The test yielded a $\chi^2_{(1)} = 50.45$, $p \approx 0$ and we therefore rejected the null hypothesis of no dependence. From **tables 11 and 12** the resulting rejection is due to an approximately 12% 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 12% enriched with pseudogenes or is the cis position is 12% 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	794	0.8219	1291	0.6933	2085
pseudo	172	0.1781	571	0.3067	743
row.total	966	1.0000	1862	1.0000	2828

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

	Value
Chi-Squared	53.64582
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 \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 $\chi^2_{(1)} = 28.72$, $p = 8.345e^{-08}$ 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 = 33.891, df = 1, p-value = 5.83e-09
```

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

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	1291	0.6933	571	0.3067
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	794	0.8219	172	0.1781
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 $\chi^2_{(1)} = 175.83$, $p \approx 0$ 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	459	0.4752	1360	0.8219	1819

	cis	%cis	trans	%trans	Col.Total
protein_coding	507	0.5248	502	0.1781	1009
row.total	966	1.0000	1862	1.0000	2828

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

	Value
Chi-Squared	179.4477
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 $\chi^2_{(1)} = 39.95$, $p = 2.596e^{-10}$. we therefore rejected the null hypothesis of equality of observed and expected proportions. **Table 16** indicates that the rejection is due to an approximately 18 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 = 27.635, df = 1, p-value = 1.465e-07
```

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

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	1360	0.7304	502	0.2696
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	459	0.4752	507	0.5248

	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 $\chi^2_{(1)} = 15.18$, $p = 9.73e^{-05}$, 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	259	0.2681	366	0.1966	625
lncRNA	707	0.7319	1496	0.8034	2203
row.total	966	1.0000	1862	1.0000	2828

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

	Value
Chi-Squared	18.501333
P	0.000017
df	1.000000

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 $\chi^2_{(1)} = 47.96$, $p = 4.351e^{-12}$. we therefore rejected the null hypothesis of equality of observed and expected proportions. **Table 19** indicates that the rejection is due to an approximately 7% 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 = 25.387, df = 1, p-value = 4.691e-07
```

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

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	366	0.1966	1496	0.8034
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	259	0.2681	707	0.7319
Genome	18843	0.2469	57471	0.7531