

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	1715	0.3977273
Trans Only	2506	0.5811688
Both Cis & Trans	91	0.0211039

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

	Total.Unique.Genes	Percent.Of.Total
Cis Only Unique	1005	0.3356713
Trans Only Unique	1898	0.6339345
Both Cis & Trans	91	0.0303941

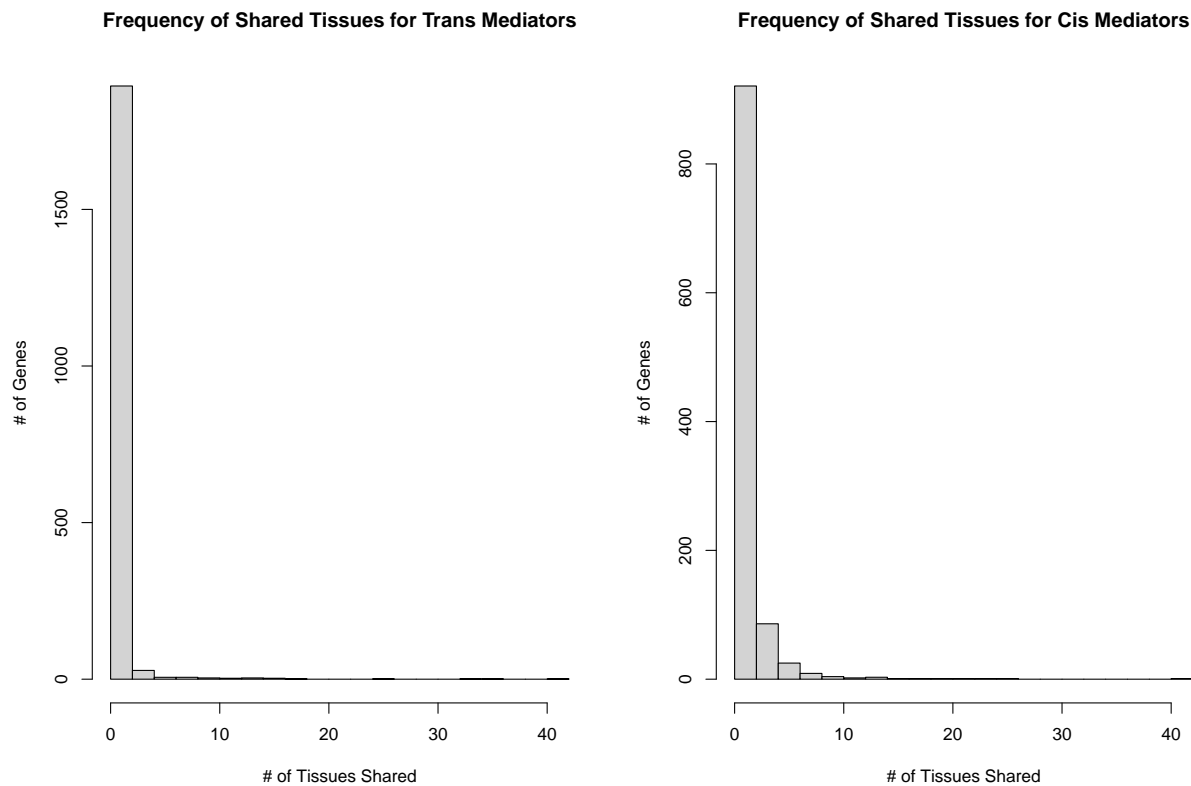


Table 3: LOND Trans Mediator Tissue Counts

Num.of.Genes	Num.of.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
1	18
1	26
1	34
1	36

Num.of.Genes	Num.of.Shared
1	41

Table 4: LOND Cis Mediator Tissue Counts

Num.of.Genes	Num.of.Shared
792	1
129	2
55	3
31	4
12	5
13	6
8	7
1	8
1	9
3	10
2	11
2	13
1	14
1	15
1	18
1	20
1	21
1	24
1	25
1	41

Table 5: LOND Unique Cis Mediator Gene Types

	Percentage
IG_V_gene	0.00498
IG_V_pseudogene	0.00100
lncRNA	0.26070
miRNA	0.00199
polymorphic_pseudogene	0.00100
processed_pseudogene	0.04478
protein_coding	0.51940
snoRNA	0.00100
TEC	0.01592
TR_C_gene	0.00100
TR_V_gene	0.00100
TR_V_pseudogene	0.00100

	Percentage
transcribed_processed_pseudogene	0.01791
transcribed_unitary_pseudogene	0.00896
transcribed_unprocessed_pseudogene	0.07164
translated_unprocessed_pseudogene	0.00100
unprocessed_pseudogene	0.03781
NA's	0.00896

Table 6: LOND Unique Trans Mediator Gene Types

	Percentage
IG_D_gene	0.00316
IG_J_gene	0.00053
IG_V_gene	0.00632
IG_V_pseudogene	0.00843
lncRNA	0.19705
miRNA	0.06059
misc_RNA	0.04953
Mt_tRNA	0.00211
polymorphic_pseudogene	0.00053
processed_pseudogene	0.19125
protein_coding	0.27081
rRNA	0.00053
rRNA_pseudogene	0.00843
scaRNA	0.00053
snoRNA	0.02792
snRNA	0.06059
TEC	0.00738
TR_J_gene	0.00369
TR_J_pseudogene	0.00053
TR_V_gene	0.00053
transcribed_processed_pseudogene	0.01001
transcribed_unprocessed_pseudogene	0.02634
unprocessed_pseudogene	0.06270
NA's	0.00053

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

Gene.Type	Percent
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

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	186	0.1850746
protein_coding	522	0.5194030
lncRNA	262	0.2606965
others	35	0.0348259

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	585	0.3082192
protein_coding	514	0.2708114
lncRNA	374	0.1970495
others	425	0.2239199

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	17093	0.2546254
protein_coding	22792	0.3395203
lncRNA	17951	0.2674065
others	9294	0.1384478

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:

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

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

The resulting 2×2 contingency table is given in **table 11 and 12**. The test yielded a $\chi^2_{(1)} = 50.45$, $p \approx 0$ 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	819	0.8149	1313	0.6918	2132
pseudo	186	0.1851	585	0.3082	771
row.total	1005	1.0000	1898	1.0000	2903

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

	Value
Chi-Squared	50.45616
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 = 28.724, df = 1, p-value = 8.345e-08
```

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

	non-pseudo	%non-pseudo	pseudogene	%pseudogene
Observed	1313	0.6918	585	0.3082
Genome	50037	0.7454	17093	0.2546

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 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	483	0.4806	1384	0.8149	1867
protein_coding	522	0.5194	514	0.1851	1036
row.total	1005	1.0000	1898	1.0000	2903

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

	Value
Chi-Squared	175.839
P	0.000
df	1.000

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 = 39.957, df = 1, p-value = 2.596e-10
```

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	483	0.4806	522	0.5194
Genome	44338	0.6605	22792	0.3395

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 17: 2X2 contingency table comparing the cell counts of mediator position and gene_type:lncRNA

	cis	%cis	trans	%trans	Col.Total
non-lncRNA	262	0.2607	374	0.197	636
lncRNA	743	0.7393	1524	0.803	2267
row.total	1005	1.0000	1898	1.000	2903

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

	Value
Chi-Squared	15.1883019
P	0.0000973
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 $\chi^2_{(1)} = 47.96$, $p = 4.351e^{-12}$. 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 = 47.96, df = 1, p-value = 4.351e-12
```

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

	non-lncRNA	%non-lncRNA	lncRNA	%lncRNA
Observed	374	0.1970	1524	0.8030
Genome	17951	0.2674	49179	0.7326