ADDIS Summary of Mediators GTExV8

Audrey Fu Lab

Jarred Kvamme

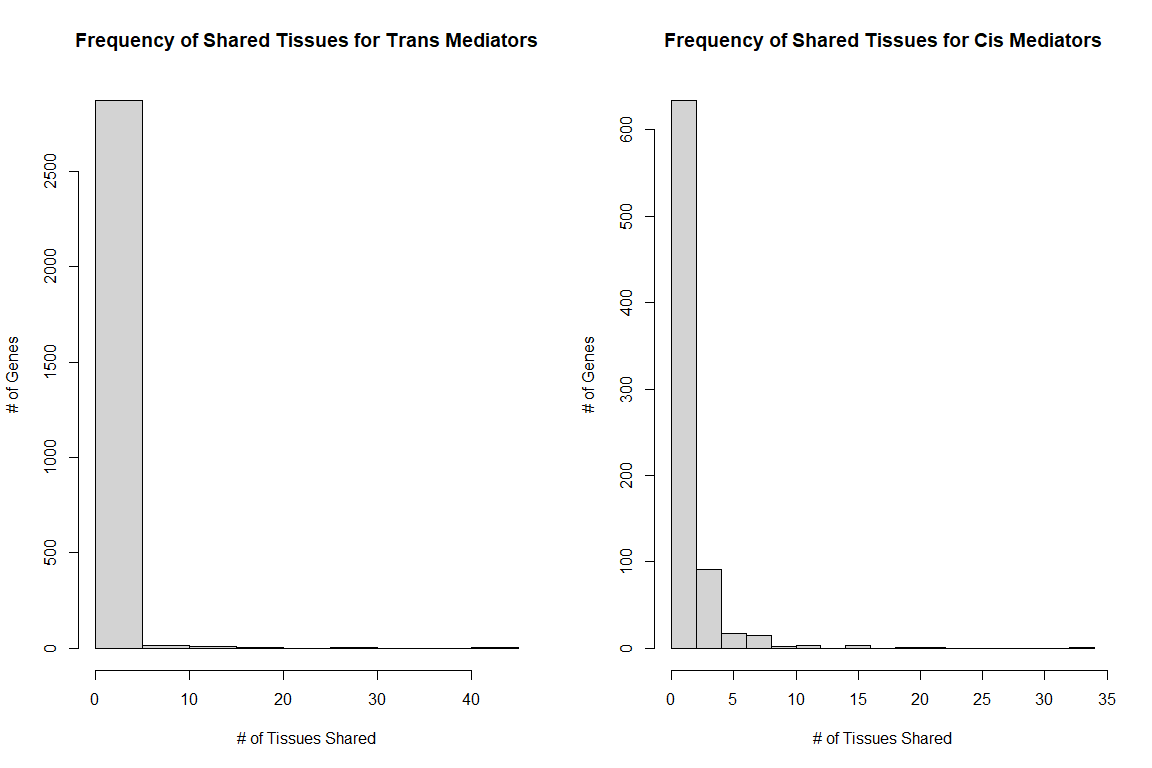
3/28/2021

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 |

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 |



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 |
| 1 | 44 |

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 |

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 |
| NA’s | 0.01096 |

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 |

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 |

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

|  |  |  |
| --- | --- | --- |
| Type | Count | Proportion |
| pseudogene | 142 | 0.1945205 |
| protein\_coding | 367 | 0.5027397 |
| lncRNA | 192 | 0.2630137 |
| others | 29 | 0.0397260 |

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 | 574 | 0.2015449 |
| others | 708 | 0.2485955 |

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 |

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:

The resulting contingency table is given in . The test yielded a and we therefore rejected the null hypothesis of no dependence. From the resulting rejection is due to an approximately 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 enriched with pseudogenes or is the cis position is depleted.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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 |

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

|  |  |
| --- | --- |
|  | Value |
| Chi-Squared | 36.54885 |
| 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 ${\bf p}$ is taken to be the observed proportion of pseudo and non-pseudogene types among trans mediator genes, and ${\bf p\_0}$ 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:

$$ H\_0: {\bf p} = {\bf p\_0} $$

$$ H\_A: {\bf p} \neq {\bf p\_0} $$

The test resulted in a and we therefore rejected the null hypothesis of equality of observed and expected proportions. indicates that the rejection is due to a roughly 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 , 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

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 | 50037 | 0.7454 | 17093 | 0.2546 |

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 (). The test yielded a and we rejected the null hypothesis of no dependence. The rejection is due to an approximately ratio of protein coding to non-protein coding gene types among the cis mediators and an approximately ratio of protein coding to non-protein coding gene types among the trans mediators.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | cis | %cis | trans | %trans | Col.Total |
| non-protein\_coding | 363 | 0.4973 | 2161 | 0.8055 | 2524 |
| protein\_coding | 367 | 0.5027 | 687 | 0.1945 | 1054 |
| row.total | 730 | 1.0000 | 2848 | 1.0000 | 3578 |

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

|  |  |
| --- | --- |
|  | Value |
| Chi-Squared | 189.9823 |
| 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 . we therefore rejected the null hypothesis of equality of observed and expected proportions. indicates that the rejection is due to an approximately 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 = 122.72, df = 1, p-value < 2.2e-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 | 363 | 0.4973 | 367 | 0.5027 |
| Genome | 44338 | 0.6605 | 22792 | 0.3395 |

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 (). The test yielded a , and we rejected the null hypothesis of no dependence. The rejection is due to an approximately ratio of non-lncRNA to lncRNA gene types among the cis mediators and an approximately ratio of non-lncRNA to lncRNA gene types among the trans mediators.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | cis | %cis | trans | %trans | Col.Total |
| non-lncRNA | 192 | 0.263 | 574 | 0.2015 | 766 |
| lncRNA | 538 | 0.737 | 2274 | 0.7985 | 2812 |
| row.total | 730 | 1.000 | 2848 | 1.0000 | 3578 |

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

|  |  |
| --- | --- |
|  | Value |
| Chi-Squared | 12.6859428 |
| P | 0.0003684 |
| 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 . we therefore rejected the null hypothesis of equality of observed and expected proportions. indicates that the rejection is due to an approximately 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 = 63.062, df = 1, p-value = 2.003e-15

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | non-lncRNA | %non-lncRNA | lncRNA | %lncRNA |
| Observed | 574 | 0.2015 | 2274 | 0.7985 |
| Genome | 17951 | 0.2674 | 49179 | 0.7326 |