Section 1.

**Algorithm Version 1 – Trio Specific setup**

Step 1. – Preform two regressions

2.1 - using Trans Gene as a response:

2.2 – using the cis gene as a response:

Where is the SNP, is the cis gene, is the trans gene, and represents the trio-specific chosen confounders/PC’s

The results from the coefficient tests for , determines the model of the Trio:

|  |  |  |  |
| --- | --- | --- | --- |
| With Trans Gene as response | Description | Test for conditional association with Cis Gene | Test for conditional association with SNP |
| Model |  |  |  |
| M0.1 | SNP 🡪 Cis |  |  |
| M0.2 | SNP 🡪 Trans |  |  |
| M1.1 | Cis Mediation |  |  |
| M1.2 | Trans Mediation |  |  |
| M2.1 | SNP/Trans 🡪 Cis |  |  |
| M2.2 | SNP/Cis 🡪 Trans |  |  |
| M3 | SNP 🡪 Cis; SNP 🡪 Trans |  |  |
| M4 | SNP 🡪 Cis; SNP 🡪 Trans; Cis ↔ Trans |  |  |
| With Cis Gene as response | **Description** | **Test for conditional association with Cis Gene** | **Test for conditional association with SNP** |
| Model |  |  |  |
| M0.1 | SNP 🡪 Cis |  |  |
| M0.2 | SNP 🡪 Trans |  |  |
| M1.1 | Cis Mediation |  |  |
| M1.2 | Trans Mediation |  |  |
| M2.1 | SNP/Trans 🡪 Cis |  |  |
| M2.2 | SNP/Cis 🡪 Trans |  |  |
| M3 | SNP 🡪 Cis; SNP 🡪 Trans |  |  |
| M4 | SNP 🡪 Cis; SNP 🡪 Trans; Cis ↔ Trans |  |  |

Step 2.1 - The results of the tests are combined by converting the test results using the following cases for each test:

Therefore, an indicator matrix can be constructed for each pair of tests on each regression which indicates the type of model of the trio:

|  |  |  |
| --- | --- | --- |
| **M0.1** |  |  |
| **Cis Response** | 0 | 1 |
| **Trans Response** | 0 | 0 |

|  |  |  |
| --- | --- | --- |
| **M0.2** |  |  |
| **Cis Response** | 0 | 0 |
| **Trans Response** | 0 | 1 |

|  |  |  |
| --- | --- | --- |
| **M1.1** |  |  |
| **Cis Response** | 1 | 1 |
| **Trans Response** | 1 | 0 |

|  |  |  |
| --- | --- | --- |
| **M1.2** |  |  |
| **Cis Response** | 1 | 0 |
| **Trans Response** | 1 | 1 |

|  |  |  |
| --- | --- | --- |
| **M2.1** |  |  |
| **Cis Response** | 1 | 0 |
| **Trans Response** | 1 | 1 |

|  |  |  |
| --- | --- | --- |
| **M2.2** |  |  |
| **Cis Response** | 1 | 1 |
| **Trans Response** | 1 | 0 |

|  |  |  |
| --- | --- | --- |
| **M3** |  |  |
| **Cis Response** | 0 | 1 |
| **Trans Response** | 0 | 1 |

|  |  |  |
| --- | --- | --- |
| **M4** |  |  |
| **Cis Response** | 1 | 1 |
| **Trans Response** | 1 | 1 |

Step 2.2 – Notice that there exist the following equivalencies in the indicator matrices given above:

Therefore, step 2.2 will involve breaking the equivalencies of these models using the results of the marginal associations with the SNP. The following are the expected independence relations for each model and subtype:

|  |  |  |
| --- | --- | --- |
|  | **Marginal Relations** | **Conditional Relations** |
| **M2.1** |  |  |
| **M2.2** |  |  |
| **M1.1** |  |  |
| **M1.2** |  |  |

Thus M2.1 can be distinguished from M1.2 by a marginal test between the SNP and trans gene and M2.2 can be distinguished from M1.1 by a marginal test between the SNP and cis gene

|  |  |  |  |
| --- | --- | --- | --- |
|  | SNP | Cis Gene | Trans Gene |
| SNP | 1 | Cor(SNP, Cis Gene) | Cor(SNP, Cis Gene) |
| Cis Gene |  | 1 | Cor(Cis Gene, Trans Gene) |
| Trans Gene |  |  | 1 |

Marginal and Conditional cases for all models

|  |  |  |
| --- | --- | --- |
| **Model** | **Marginal Relations** | **Conditional Relations** |
| **M2.1** |  |  |
| **M2.2** |  |  |
| **M1.1** |  |  |
| **M1.2** |  |  |
| **M0.1** |  |  |
| **M0.2** |  |  |
| **M3** |  |  |
| **M4** |  |  |

Section 2.

**General Algorithm**

In this version we assume a network of size *n* nodes (not including the instrumental variable(s) meaning we have *n* genes of interest). We first narrow the number of testable edges in the network by testing all the marginal relationships between the SNP(s)/instrumental variables and the genes/other nodes.

Side Note: it is also possible to avoid preforming all *n* marginal tests by first preforming the regression outlined below and then conducting the marginal tests based on the needs of the network. See below for more details

Ex: A network with 5 Genes: n=5

Example Correlation Matrix between the nodes in a network of size *n*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **SNP** | **Gene1** | **Gene2** | **Gene3** | **Gene4** | **Gene5** | **. . .** | **Gene n** |
| **SNP** | 1 |  |  |  |  |  | **. . .** |  |
| **Gene1** |  | 1 |  |  |  |  | **. . .** |  |
| **Gene2** |  |  | 1 |  |  |  | **. . .** |  |
| **Gene3** |  |  |  | 1 |  |  | **. . .** |  |
| **Gene4** |  |  |  |  | 1 |  | **. . .** |  |
| **Gene5** |  |  |  |  |  | 1 | **. . .** |  |
| **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** |  |
| **Gene n** |  |  |  |  |  |  |  | 1 |

We thus preform the pairwise correlation tests between the instrumental variable and the other nodes and remove edges that lack a significant relationship.

Using student’s *T* distribution, the statistic for testing the marginal relationship between the instrumental variable/SNP and the *ith* gene/node is:

Ex:

We then preform *n* regressions using each gene as the response once and the remaining *n-1* genes/nodes and their confounders as explanatory variables such that:

Where **Xi,n**is the full set of confounding variables among all nodes in the graph. This produces the following table for the coefficients of interest:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Response:** | **SNP** | **Gene1** | **Gene2** | **Gene3** | **Gene4** | **Gene5** | **. . .** | **Gene n** |
| **Gene1** |  |  |  |  |  |  | **. . .** |  |
| **Gene2** |  |  |  |  |  |  | **. . .** |  |
| **Gene3** |  |  |  |  |  |  | **. . .** |  |
| **Gene4** |  |  |  |  |  |  | **. . .** |  |
| **Gene5** |  |  |  |  |  |  | **. . .** |  |
| **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** | **.**  **.**  **.** |  |
| **Gene n** |  |  |  |  |  |  | **. . .** |  |

We then formulate trios out of the nodes in the network. For a network of size *n* there are possible trios to be formed. Using the information from the marginal tests and regressions we can then match each trio to one of the cases outlined in the trio specific algorithms. Therefore, we select the 3 X 3 submatrix from the above regression matrix corresponding to the nodes in the trio:

Ex: we select the trio formed by the SNP and genes 1 and 22:

Ex. Selection of the submatrix from the regression table corresponding to the trio formed by the SNP and genes 1-2

|  |  |  |  |
| --- | --- | --- | --- |
| **Response:** | **SNP** | **Gene1** | **Gene2** |
| **Gene1** |  |  |  |
| **Gene2** |  |  |  |

Ex: we select the trio formed by the SNP and genes 3 and 4

Ex. Selection of the submatrix from the regression table corresponding to the trio formed by the SNP and genes 3-4

|  |  |  |  |
| --- | --- | --- | --- |
| **Response:** | **SNP** | **Gene3** | **Gene4** |
| **Gene3** |  |  |  |
| **Gene4** |  |  |  |

Like the specific trio algorithm given in **Section 1.** above, we formulate a 2 X 2 indicator matrix and armed with the results of the marginal tests we can classify each trio within the network and obtain an estimate of the entire network