Section 1.

Algorithm Version 1 – Trio Specific setup

Step 1. – Preform two regressions

2.1 - using Trans Gene as a response:

$$T_j = \beta_0 + \beta_1 C_i + \beta_2 L_i + \Gamma_{ij} X_{ij} + \epsilon$$

2.2 – using the cis gene as a response:

$$C_i = \beta_0 + \beta_1 T_j + \beta_2 L_i + \Gamma_{ij} X_{ij} + \epsilon$$

Where L_i is the SNP, C_i is the cis gene, T_j is the trans gene, and X_{ij} represents the trio-specific chosen confounders/PC's

The results from the coefficient tests for β_1 , β_2 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		$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
M0.1	SNP → Cis	$= 0$	$= 0$
M0.2	SNP → Trans	$= 0$	$\neq 0$
M1.1	Cis Mediation	$\neq 0$	$= 0$
M1.2	Trans Mediation	$\neq 0$	$\neq 0$
M2.1	SNP/Trans → Cis	$\neq 0$	$\neq 0$
M2.2	SNP/Cis → Trans	$\neq 0$	$= 0$
M3	SNP → Cis; SNP → Trans	$= 0$	$\neq 0$
M4	SNP → Cis; SNP → Trans; Cis ↔ Trans	$\neq 0$	$\neq 0$

With Cis Gene as response	Description	Test for conditional association with Cis Gene	Test for conditional association with SNP
Model		$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
M0.1	SNP → Cis	$= 0$	$\neq 0$
M0.2	SNP → Trans	$= 0$	$= 0$
M1.1	Cis Mediation	$\neq 0$	$\neq 0$
M1.2	Trans Mediation	$\neq 0$	$= 0$
M2.1	SNP/Trans → Cis	$\neq 0$	$= 0$
M2.2	SNP/Cis → Trans	$\neq 0$	$\neq 0$
M3	SNP → Cis; SNP → Trans	$= 0$	$\neq 0$
M4	SNP → Cis; SNP → Trans; Cis ↔ Trans	$\neq 0$	$\neq 0$

Step 2.1 - The results of the tests are combined by converting the test results using the following cases for each test:
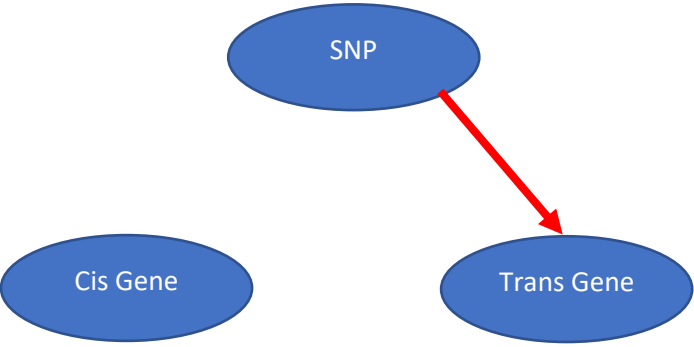
$$2 \times Pr(T_{obs} > |T|) = P = \begin{cases} 1, & if\ P\ <\ \alpha \\ 0, & else \end{cases}$$

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

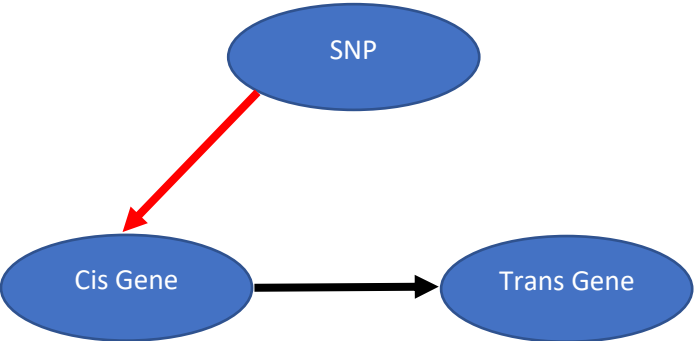| M0.1	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	0	1
Trans Response	0	0
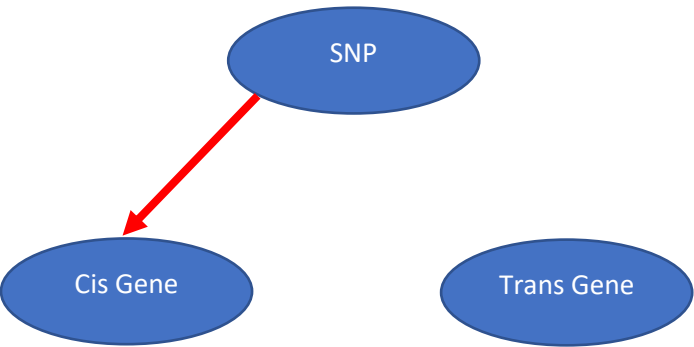
M0.2	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	0	0
Trans Response	0	1

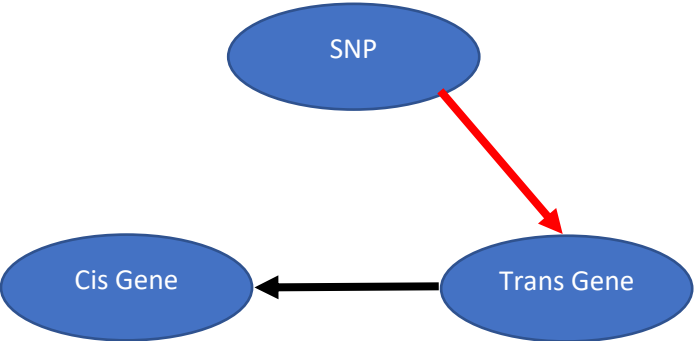| M1.1	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	1	1
Trans Response	1	0

M1.2	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	1	0
Trans Response	1	1

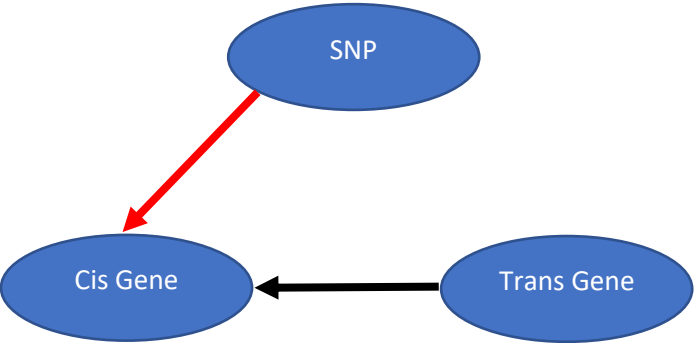| M2.1	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	1	0
Trans Response	1	1

M2.2	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	1	1
Trans Response	1	0

M3	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	0	1
Trans Response	0	1

M4	$H_0: \beta_1 = 0$	$H_0: \beta_2 = 0$
Cis Response	1	1
Trans Response	1	1
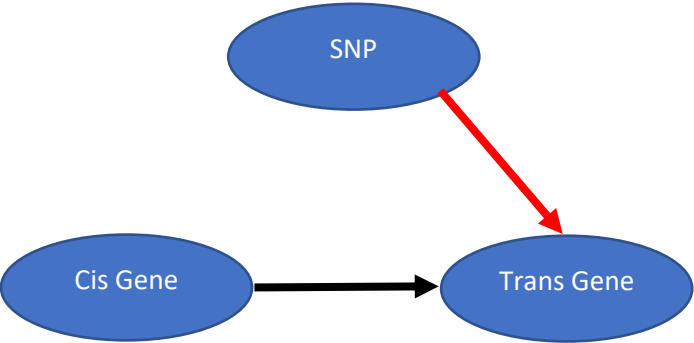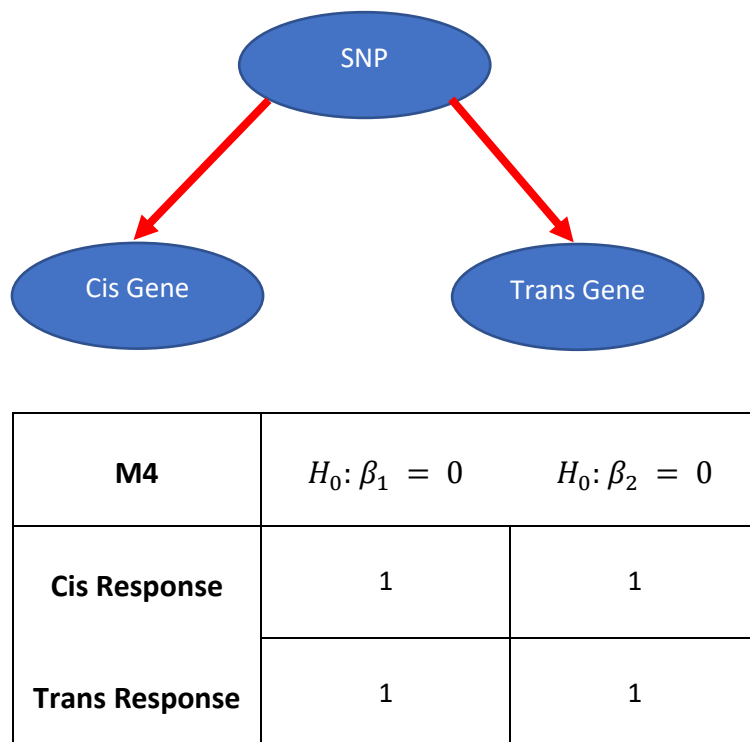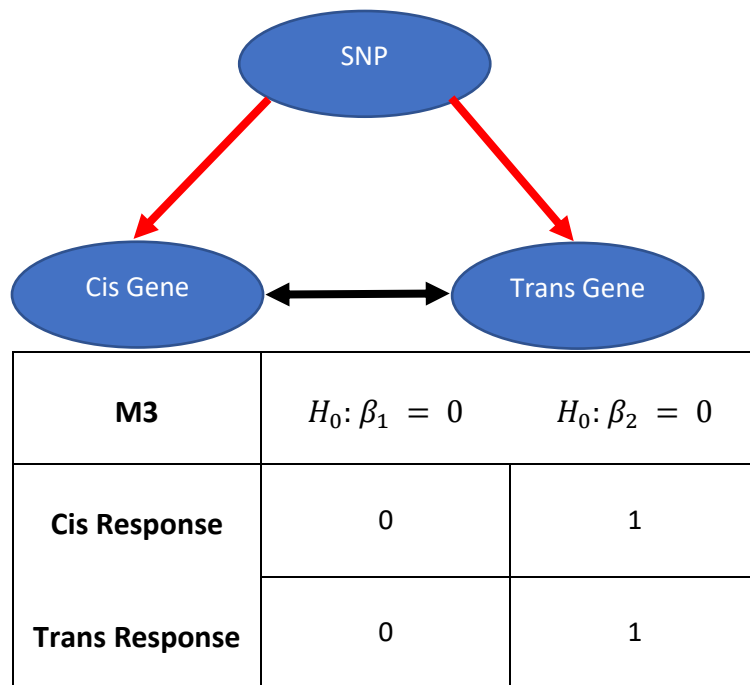
Step 2.2 – Notice that there exist the following equivalencies in the indicator matrices given above:

$$[M2.1] \equiv [M1.2]$$

$$[M2.2] \equiv [M1.1]$$

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	$SNP \not\perp Cis$; $SNP \perp Trans$; $Cis \not\perp Trans$	$SNP \not\perp Trans \mid Cis$
M2.2	$SNP \perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$SNP \not\perp Cis \mid Trans$
M1.1	$SNP \not\perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$SNP \perp Trans \mid Cis$
M1.2	$SNP \not\perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$SNP \perp Cis \mid Trans$

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)

Marginal and Conditional cases for all models

| Model	Marginal Relations	Conditional Relations
M2.1	$SNP \not\perp Cis$; $SNP \perp Trans$; $Cis \not\perp Trans$	$SNP \not\perp Trans \mid Cis$
M2.2	$SNP \perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$SNP \not\perp Cis \mid Trans$
M1.1	$SNP \not\perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$SNP \perp Trans \mid Cis$
M1.2	$SNP \not\perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$SNP \perp Cis \mid Trans$
M0.1	$SNP \not\perp Cis$; $SNP \perp Trans$; $Cis \perp Trans$	
M0.2	$SNP \perp Cis$; $SNP \not\perp Trans$; $Cis \perp Trans$	
M3	$SNP \not\perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$Trans \perp Cis \mid SNP$
M4	$SNP \not\perp Cis$; $SNP \not\perp Trans$; $Cis \not\perp Trans$	$Trans \not\perp Cis \mid SNP$

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 <u>SNP(s)/instrumental variables</u> and the <u>genes/other nodes.</u>

<u>Side Note</u>: 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

<u>Ex: A network with 5 Genes: n=5</u>
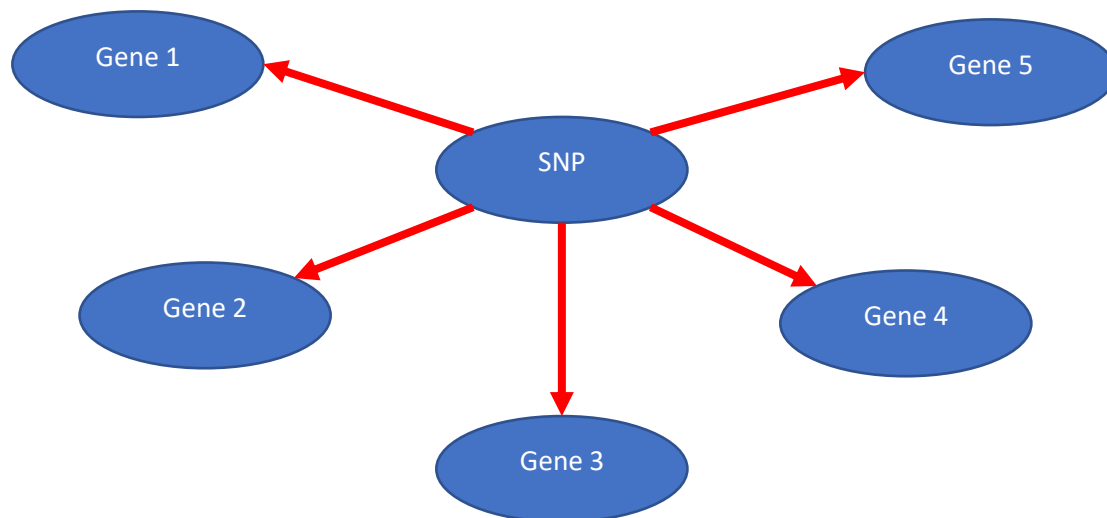
Example Correlation Matrix between the nodes in a network of size n

|	SNP	Gene1	Gene2	Gene3	Gene4	Gene5	. . .	Gene n
SNP	1	$r_{snp,1}$	$r_{snp,2}$	$r_{snp,3}$	$r_{snp,4}$	$r_{snp,5}$	. . .	$r_{snp,n}$
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:

$$t = r_{snp,i} \sqrt{\frac{n - 2}{1 - r_{snp,i}^2}}$$

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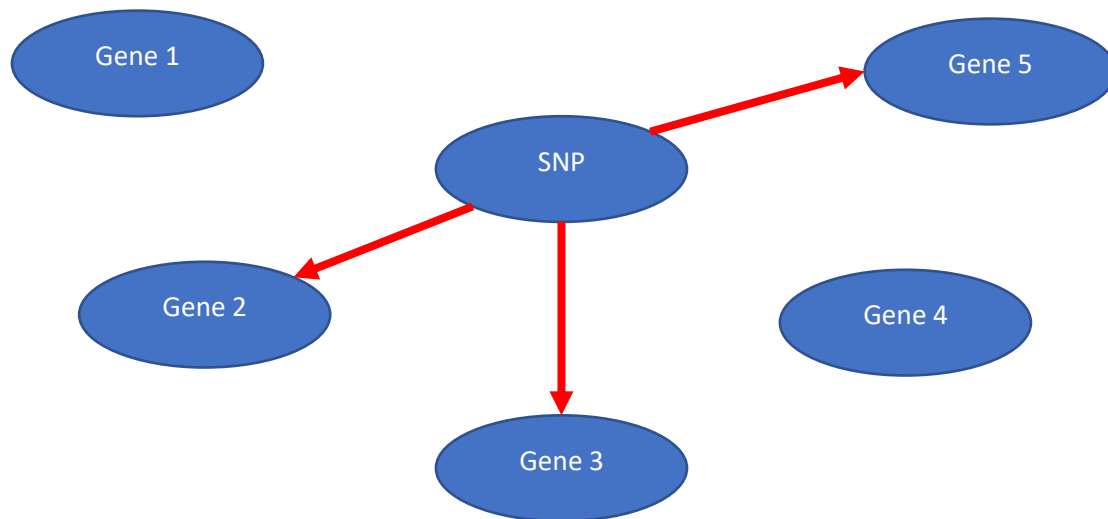
We then preform *n* regressions using each gene as the response once and the remaining *n-1* genes/nodes and their confounders as explanatory variables such that:

$$Y = \beta_0 + \beta_1 L_i + \beta_2 G_i + \beta_3 G_{i+1} + \beta_4 G_{i+2} \cdots + G_{n-1} + \Gamma X_{i,n} \quad \text{for } i \in 1:n-1$$

Where $X_{i,n}$ 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	$H_0:$ $\beta_1 = 0$		$H_0:$ $\beta_i = 0$	$H_0:$ $\beta_{i+1} = 0$	$H_0:$ $\beta_{i+2} = 0$	$H_0:$ $\beta_{i+3} = 0$	...	$H_0:$ $\beta_{n-1} = 0$
Gene2	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_i = 0$		$H_0:$ $\beta_{i+1} = 0$	$H_0:$ $\beta_{i+2} = 0$	$H_0:$ $\beta_{i+3} = 0$	...	$H_0:$ $\beta_{n-1} = 0$
Gene3	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_i = 0$	$H_0:$ $\beta_{i+1} = 0$		$H_0:$ $\beta_{i+2} = 0$	$H_0:$ $\beta_{i+3} = 0$	...	$H_0:$ $\beta_{n-1} = 0$
Gene4	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_i = 0$	$H_0:$ $\beta_{i+1} = 0$	$H_0:$ $\beta_{i+2} = 0$		$H_0:$ $\beta_{i+3} = 0$	...	$H_0:$ $\beta_{n-1} = 0$
Gene5	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_i = 0$	$H_0:$ $\beta_{i+1} = 0$	$H_0:$ $\beta_{i+2} = 0$	$H_0:$ $\beta_{i+3} = 0$		...	$H_0:$ $\beta_{n-1} = 0$
.	**.**	**.**	**.**	**.**	**.**	**.**	**.**	
Gene n	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_i = 0$	$H_0:$ $\beta_{i+1} = 0$	$H_0:$ $\beta_{i+2} = 0$	$H_0:$ $\beta_{i+3} = 0$	$H_0:$ $\beta_{i+4} = 0$	...	

We then formulate trios out of the nodes in the network. For a network of size *n* there are $\binom{n}{2}$ 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:
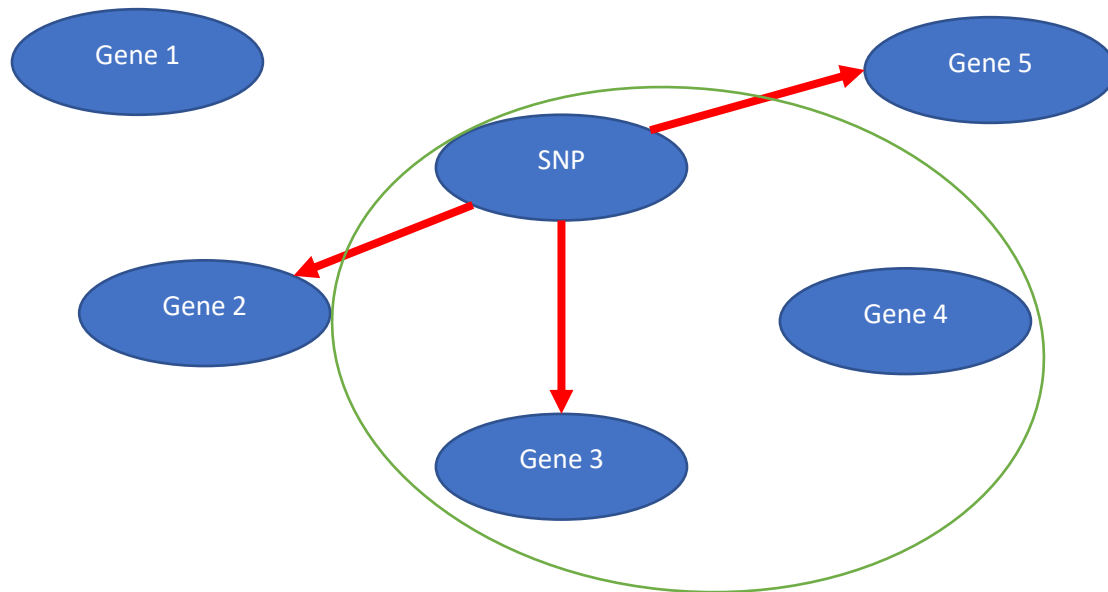
Ex: we select the trio formed by the SNP and genes 1 and 2
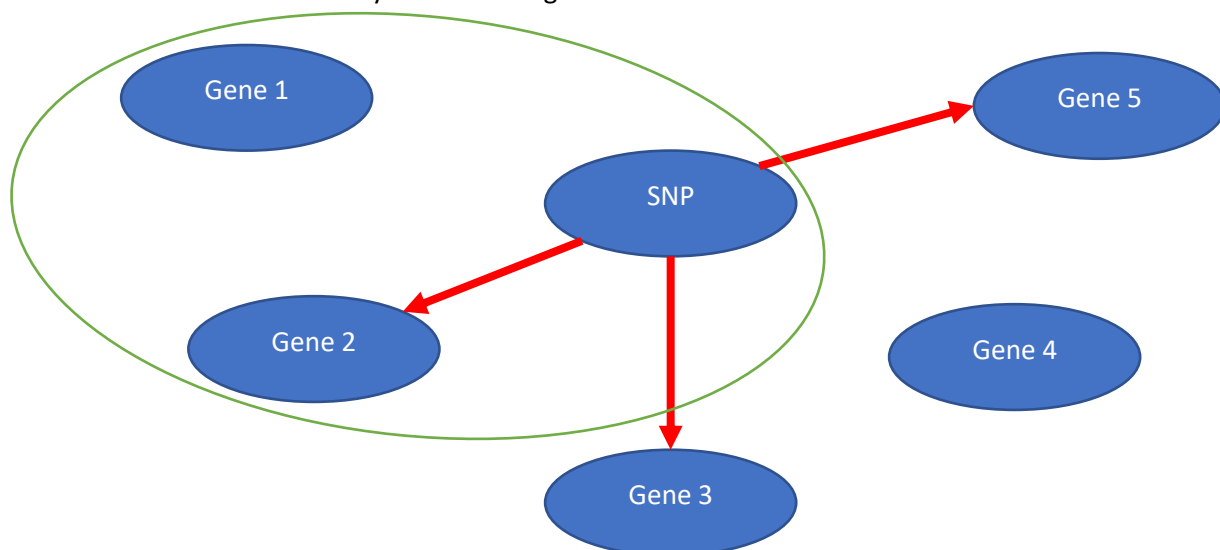
2:

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	$H_0:$ $\beta_1 = 0$		$H_0:$ $\beta_i = 0$
Gene2	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_i = 0$	

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	$H_0:$ $\beta_1 = 0$		$H_0:$ $\beta_{i+2} = 0$
Gene4	$H_0:$ $\beta_1 = 0$	$H_0:$ $\beta_{i+2} = 0$	

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 are able to classify each trio within the network and obtain an estimate of the entire network