

---

# CINOEDV: a co-information based method for detecting and visualizing $n$ -order epistatic interactions – Supplementary file

Junliang Shang<sup>1,2,\*</sup>, Yingxia Sun<sup>1</sup>, Jin-Xing Liu<sup>1</sup>, Junfeng Xia<sup>3</sup>, Junying Zhang<sup>4</sup>, Chun-Hou Zheng<sup>5</sup>

<sup>1</sup>School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China, <sup>2</sup>Institute of Network Computing, Qufu Normal University, Rizhao 276826, China, <sup>3</sup>Institute of Health Sciences, Anhui University, Hefei, Anhui 230601, China, <sup>4</sup>School of Computer Science and Technology, Xidian University, Xi'an 710071, China, <sup>5</sup>College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui 230039, China

\*To whom correspondence should be addressed.

---

## NOTE 1. EVALUATION MEASURES (DETECTION POWER AND COMPUTATIONAL COMPLEXITY).

In the study, detection power and computational complexity are applied to evaluate the performance of CINOEDV.

Various forms of detection power have been proposed depending on what is desired to measure [1-9], and three types of them are introduced with constraints from conservative to modest.

Power1 is defined as the proportion of data sets in which all ground-truth SNPs are identified with no false positives where ground-truth SNPs refer to the causative SNPs that truly associated with the phenotype, in other words, the SNPs in models added into the simulation data sets. Power1 can be written as

$$\text{Power1} = \frac{1}{D} \sum_{i=1}^D x_i,$$

where  $D$  is the number of simulation data sets with the same parameter settings, and  $x_i \in \{0,1\}$  is the detection tag, i.e., if all ground-truth SNPs in data set  $i$  are detected with no false positives,  $x_i = 1$ , otherwise,  $x_i = 0$ . Though Power1 seems not practical since false positives are inevitable for any statistical tests and fewer false positives result in larger false negatives, we still introduce it because it is advantageous in practical applications, and might interest biologists due to false positives implying wasted experimental effort to validate the results.

Sometimes, allowing some small type-I error is more reasonable, thus Power2 and Power3 are introduced. Power2 and Power3 are defined as an average proportion of ground-truth SNPs in the top  $L_i$  detected SNPs, and can be respectively written as

$$\text{Power2} = \frac{1}{D} \sum_{i=1}^D \frac{y_i}{L_i},$$

$$\text{Power3} = \frac{1}{D} \sum_{i=1}^D \frac{z_i}{L_i},$$

where  $y_i$  and  $z_i$  are the numbers of ground-truth SNPs appearing in the top  $L_i$  SNPs identified in data set  $i$ . For Power2,  $L_i$  is equal to the number of ground-truth SNPs in data set  $i$ . For Power3,  $L_i$  is normally set to a value larger than that in Power2. In our experiments,  $L_i$  is set to 4.

Computational complexity is also considered. Running time in the same computational environment is measured to assess realistic applicability of compared methods.

Figure S1. GAIN RESULTS OF COMPARED DATA SETS.

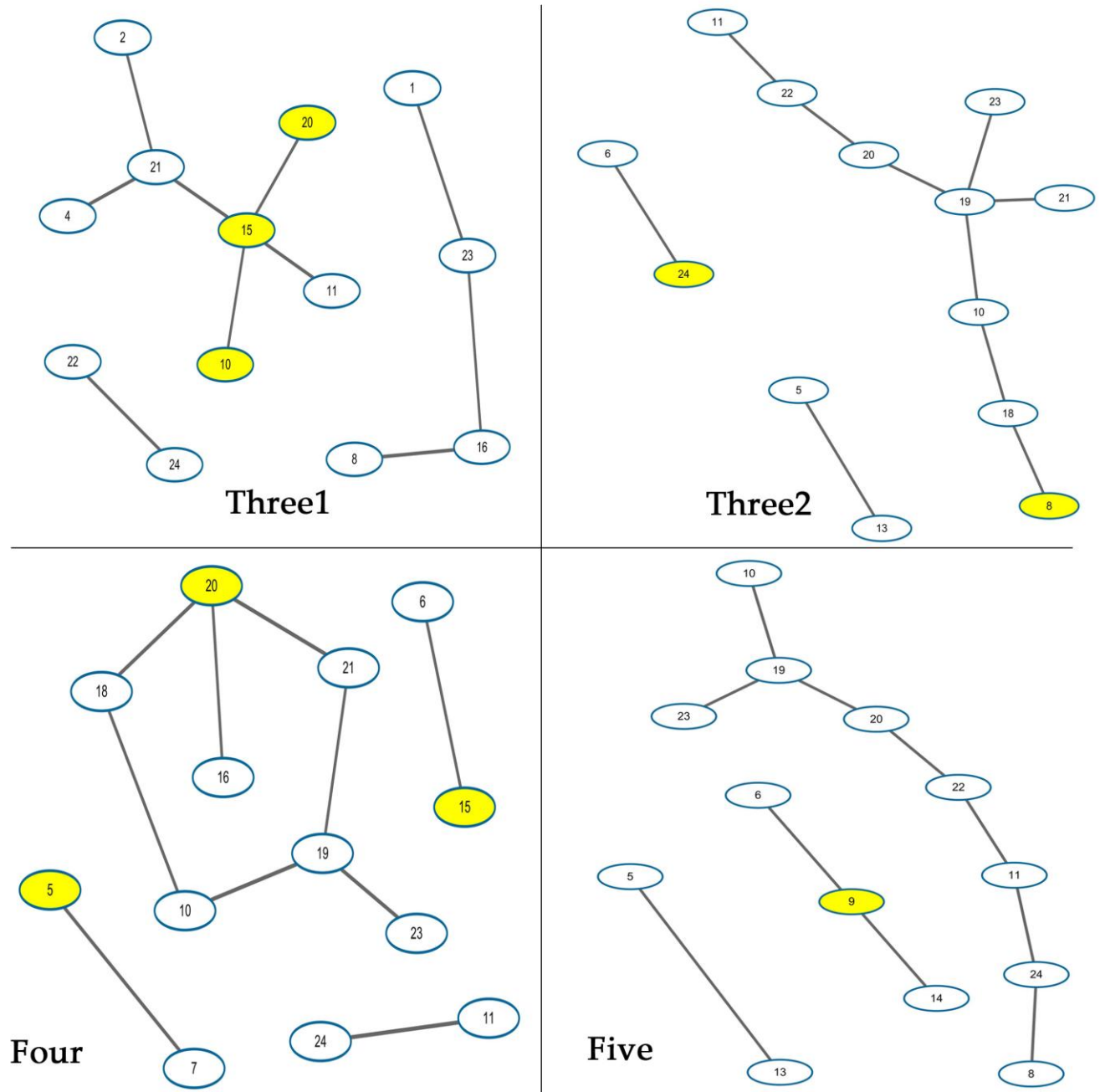


Figure S2. SEN RESULTS OF COMPARED DATA SETS.

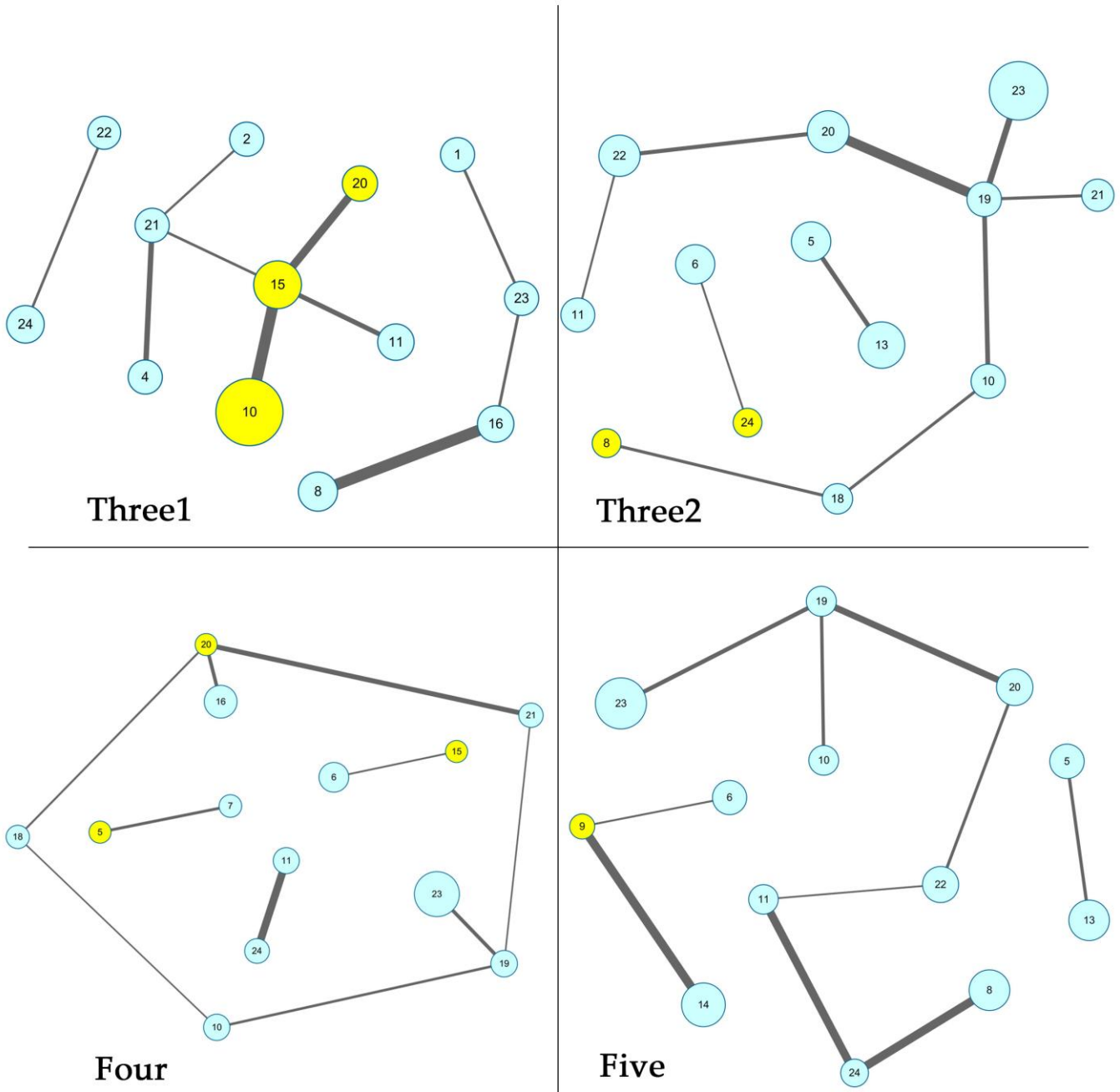


Figure S3. VISEN RESULTS OF COMPARED DATA SETS.

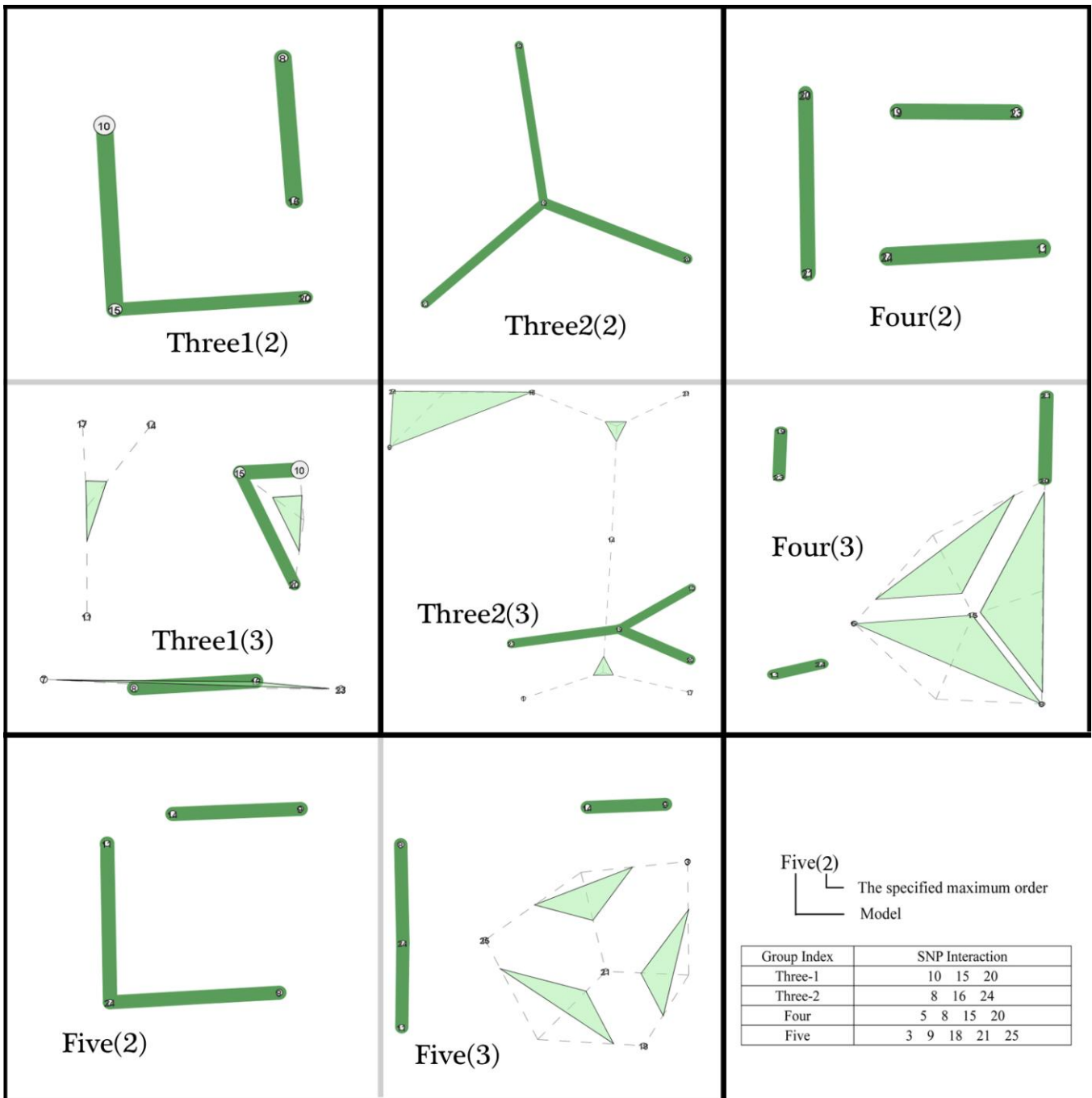
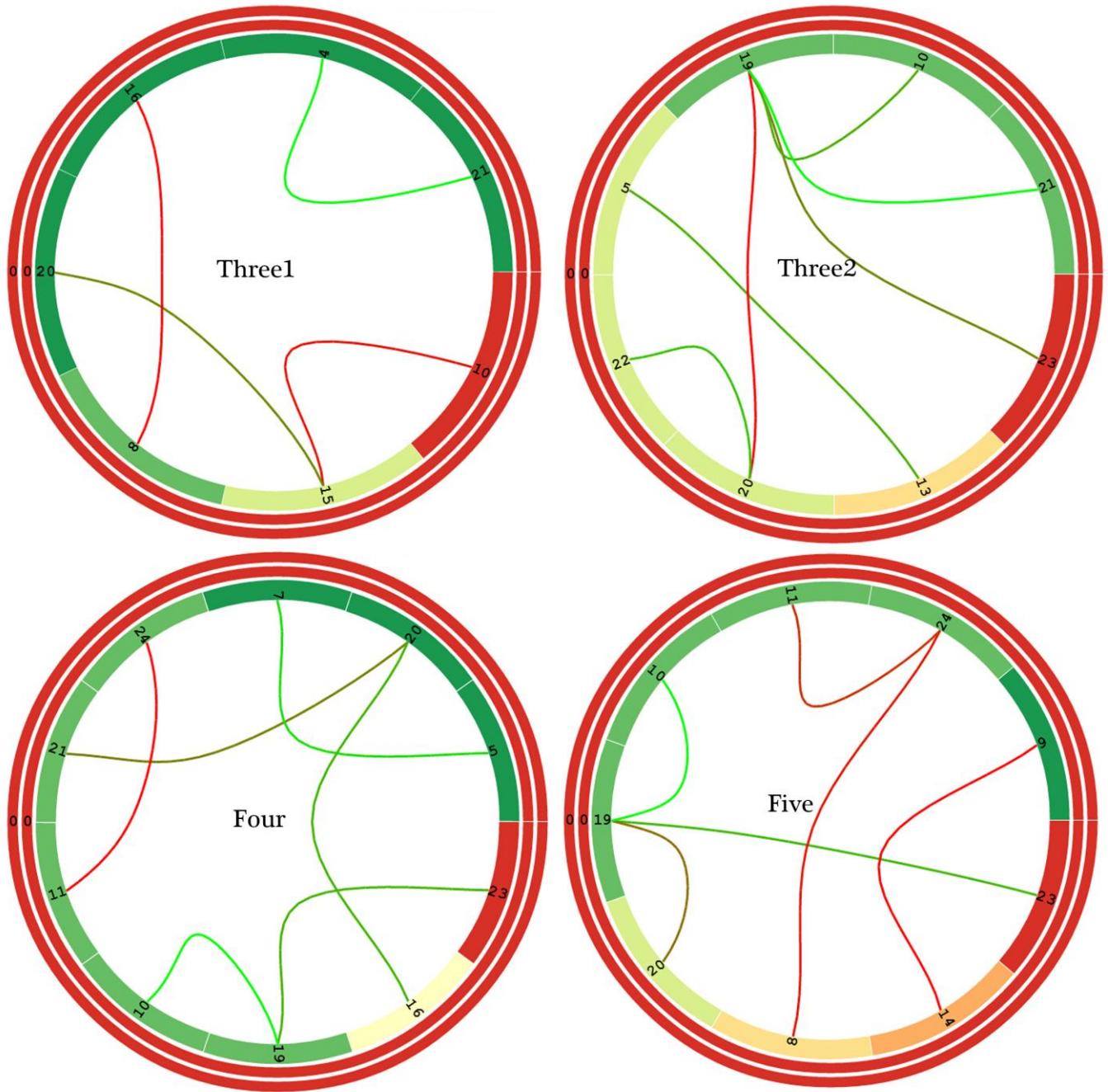


Figure S4. EINVIS RESULTS OF COMPARED DATA SETS.



**Table S1. TOP 20 SNPS WITH HIGH MAIN EFFECTS.**

For the columns “Index” in Table S2, Table S3 and Table S4, bold and underlined numbers represent that their corresponding epistatic interactions are the top 20 ones with the high *CCI* values. For the columns “SNP” in Table S2, and “Epistatic Interactions” in Table S3 and Table S4, bold SNPs are those in the largest maximal connected subgraph of epistasis hypergraph Figure S1 of main paper, and underlined SNPs are those in the second largest maximal connected subgraph of epistasis hypergraph Figure S1 of main paper.

Index	SNP	Gene	Chromosome	<i>NCI</i> Value
1	rs380390	<i>CFH</i>	Chr1	0.06472
2	rs1329428	<i>CFH</i>	Chr1	0.05857
3	rs3775652	<i>INPP4B</i>	Chr4	0.05696
4	rs1930022	<i>N/A</i>	Chr9	0.04858
5	<b>rs1394608</b>	<i>SGCD</i>	Chr5	0.04562
6	rs725518	<i>RRM1</i>	Chr11	0.04496
7	rs10511467	<i>N/A</i>	Chr9	0.04485
8	rs1740752	<i>N/A</i>	Chr10	0.04434
9	rs10507949	<i>N/A</i>	Chr13	0.04345
10	<u>rs9294603</u>	<i>N/A</i>	Chr6	0.04184
11	rs6967345	<i>N/A</i>	Chr7	0.04181
12	rs3913094	<i>N/A</i>	Chr12	0.04076
13	<b>rs7006908</b>	<i>N/A</i>	Chr8	0.04025
14	rs10254116	<i>BBS9</i>	Chr7	0.04019
15	rs3914244	<i>N/A</i>	Chr12	0.04018
16	rs7104698	<i>N/A</i>	Chr11	0.03961
17	rs10501439	<i>TENM4</i>	Chr11	0.03958
18	rs7863587	<i>N/A</i>	Chr9	0.03882
19	rs10512174	<i>ISCA1</i>	Chr9	0.03870
20	rs1363688	<i>N/A</i>	Chr5	0.03867

**Table S2. TOP 20 EPISTATIC INTERACTIONS WITH HIGH 2-ORDER INTERACTION EFFECTS.**

For the columns “Index” in Table S2, Table S3 and Table S4, bold and underlined numbers represent that their corresponding epistatic interactions are the top 20 ones with the high *CCI* values. For the columns “SNP” in Table S2, and “Epistatic Interactions” in Table S3 and Table S4, bold SNPs are those in the largest maximal connected subgraph of epistasis hypergraph Figure S1 of main paper, and underlined SNPs are those in the second largest maximal connected subgraph of epistasis hypergraph Figure S1 of main paper.

Index	Epistatic Interaction	Gene	Chromosome	<i>NCI</i> Value
<u><b>1</b></u>	<b>rs994542</b> <b>rs9298846</b>	<i>N/A</i> <i>N/A</i>	Chr6 Chr9	0.25762
<u><b>2</b></u>	rs2359907 rs9326913	<i>N/A</i> <i>N/A</i>	Chr1 Chr5	0.22561
<u><b>3</b></u>	rs2957935 rs10519745	<i>SKAP1</i> <i>N/A</i>	Chr17 Chr5	0.22267
<u><b>4</b></u>	rs10231824 rs571361	<i>PLXNA4</i> <i>N/A</i>	Chr7 Chr1	0.20243
<u><b>5</b></u>	rs4856145 rs3828057	<i>N/A</i> <i>RORC</i>	Chr3 Chr1	0.19749
<u><b>6</b></u>	rs1436532 rs10507601	<i>N/A</i> <i>N/A</i>	Chr4 Chr13	0.19140
<u><b>7</b></u>	rs305728 rs10497071	<i>N/A</i> <i>N/A</i>	Chr5 Chr2	0.19057
<u><b>8</b></u>	<b>rs2298809</b> <b>rs7688514</b>	<i>N/A</i> <i>VEGFC</i>	Chr11 Chr4	0.18960
<u><b>9</b></u>	<b>rs2278031</b> <b>rs446815</b>	<i>APIG1</i> <i>N/A</i>	Chr16 Chr11	0.18154
<u><b>10</b></u>	rs756308 rs2011542	<i>N/A</i> <i>N/A</i>	Chr10 Chr7	0.18122
<u><b>11</b></u>	rs10516495 rs2011542	<i>SLC9B1</i> <i>N/A</i>	Chr4 Chr7	0.18122
<u><b>12</b></u>	rs339884 rs195937	<i>N/A</i> <i>N/A</i>	Chr18 Chr4	0.18119
<u><b>13</b></u>	<b>rs1408120</b> <b>rs10506115</b>	<i>PTPRD</i> <i>N/A</i>	Chr9 Chr12	0.17685
<u><b>14</b></u>	rs595113 rs9300104	<i>N/A</i> <i>RNF141</i>	Chr1 Chr11	0.16972
<u><b>15</b></u>	<b>rs10518739</b> <b>rs931210</b>	<i>MAP2K5</i> <i>N/A</i>	Chr15 Chr5	0.16603
<u><b>16</b></u>	<b>rs2386906</b> <b>rs9313446</b>	<i>N/A</i> <i>SLIT3</i>	Chr10 Chr5	0.16543
<u><b>17</b></u>	<b>rs9313446</b> <b>rs293379</b>	<i>SLIT3</i> <i>ABHD2</i>	Chr5 Chr15	0.16349
18	rs1458402 rs4901408	<i>N/A</i> <i>N/A</i>	Chr6 Chr14	0.15851
19	rs6483628 rs10504003	<i>NAV2</i> <i>N/A</i>	Chr11 Chr8	0.15589
20	<b>rs3844556</b> <b>rs10521129</b>	<i>PKD2L1</i> <i>ZZEF1</i>	Chr10 Chr17	0.15405

**Table S3. TOP 20 EPISTATIC INTERACTIONS WITH HIGH 3-ORDER INTERACTION EFFECTS.**

For the columns “Index” in Table S2, Table S3 and Table S4, bold and underlined numbers represent that their corresponding epistatic interactions are the top 20 ones with the high *CCI* values. For the columns “SNP” in Table S2, and “Epistatic Interactions” in Table S3 and Table S4, bold SNPs are those in the largest maximal connected subgraph of epistasis hypergraph Figure S1 of main paper, and underlined SNPs are those in the second largest maximal connected subgraph of epistasis hypergraph Figure S1 of main paper.

Index	Epistatic Interaction	Gene	Chromosome	<i>NCI</i> Value
<b><u>1</u></b>	<b>rs4723261</b>	<i>PTHB1</i>	Chr7	0.17507
	<b>rs1036995</b>	<i>PCDH9</i>	Chr13	
	<b>rs2298809</b>	<i>N/A</i>	Chr11	
<b><u>2</u></b>	rs595113	<i>N/A</i>	Chr1	0.16096
	<u>rs2031175</u>	<i>ANKRD15</i>	Chr9	
	<u>rs9300104</u>	<i>RNF141</i>	Chr11	
<b><u>3</u></b>	rs10517546	<i>N/A</i>	Chr4	0.15934
	rs10500875	<i>N/A</i>	Chr11	
	rs10505107	<i>ANGPT1</i>	Chr8	
4	<b>rs2386906</b>	<i>N/A</i>	Chr10	0.14875
	<b>rs1476623</b>	<i>NXPH1</i>	Chr7	
	<b>rs10518739</b>	<i>MAP2K5</i>	Chr15	
5	<b>rs293379</b>	<i>ABHD2</i>	Chr15	0.14681
	<b>rs1476623</b>	<i>NXPH1</i>	Chr7	
	<b>rs10518739</b>	<i>MAP2K5</i>	Chr15	
6	<u>rs1436532</u>	<i>N/A</i>	Chr4	0.14445
	<u>rs9294603</u>	<i>N/A</i>	Chr6	
	<u>rs595113</u>	<i>N/A</i>	Chr1	
7	<b>rs1394608</b>	<i>SGCD</i>	Chr5	0.14439
	<b>rs9323782</b>	<i>N/A</i>	Chr14	
	<b>rs931210</b>	<i>N/A</i>	Chr5	
8	<u>rs1374431</u>	<i>N/A</i>	Chr2	0.14414
	<u>rs1436532</u>	<i>N/A</i>	Chr4	
	<u>rs1978419</u>	<i>N/A</i>	Chr5	
9	rs3922799	<i>N/A</i>	Chr2	0.13967
	rs10517546	<i>N/A</i>	Chr4	
	rs10500875	<i>N/A</i>	Chr11	
10	<b>rs994542</b>	<i>N/A</i>	Chr6	0.13863
	<b>rs3781868</b>	<i>NPAT</i>	Chr11	
	<b>rs2207768</b>	<i>NRG3</i>	Chr10	
11	<b>rs1394608</b>	<i>SGCD</i>	Chr5	0.13813
	<b>rs1408120</b>	<i>PTPRD</i>	Chr9	
	<b>rs10495668</b>	<i>SMC6</i>	Chr2	
12	<b>rs1926489</b>	<i>GPC5</i>	Chr13	0.13730
	<b>rs1408120</b>	<i>PTPRD</i>	Chr9	
	<b>rs718309</b>	<i>PCDH15</i>	Chr10	
13	<u>rs10486519</u>	<i>PTHB1</i>	Chr7	0.13669
	<u>rs1978419</u>	<i>N/A</i>	Chr5	
	<u>rs10507601</u>	<i>N/A</i>	Chr13	
14	<b>rs7006908</b>	<i>N/A</i>	Chr8	0.13636
	<b>rs10518739</b>	<i>MAP2K5</i>	Chr15	
	<b>rs10495668</b>	<i>SMC6</i>	Chr2	
15	<b>rs293379</b>	<i>ABHD2</i>	Chr15	0.13526
	<b>rs9287251</b>	<i>RGS7</i>	Chr1	
	<b>rs446815</b>	<i>N/A</i>	Chr11	
16	<b>rs2182277</b>	<i>N/A</i>	Chr6	0.13441
	<b>rs1036995</b>	<i>PCDH9</i>	Chr13	



	<b>rs2207768</b>	<i>NRG3</i>	Chr10	
	<b>rs3844556</b>	<i>PKD2L1</i>	Chr10	
17	<b>rs1036995</b>	<i>PCDH9</i>	Chr13	0.13419
	<b>rs4880042</b>	<i>PAX5</i>	Chr9	
	<b>rs2386906</b>	<i>N/A</i>	Chr10	
18	<b>rs9287251</b>	<i>RGS7</i>	Chr1	0.13324
	<b>rs446815</b>	<i>N/A</i>	Chr11	
	<b>rs3743175</b>	<i>ZNF291</i>	Chr15	
19	<b>rs9323782</b>	<i>N/A</i>	Chr14	0.13317
	<b>rs4880042</b>	<i>PAX5</i>	Chr9	
	<b>rs2828155</b>	<i>N/A</i>	Chr21	
20	<b>rs9323782</b>	<i>N/A</i>	Chr14	0.13317
	<b>rs4880042</b>	<i>PAX5</i>	Chr9	

**Table S4. DETECTED SNPS AND EPISTATIC INTERACTIONS OF OTHER COMPETING METHODS.**

In the table, p-values are calculated by the *chi*-squared test. *N/A* means no gene being available. Detected SNPs and epistatic interactions of both BEAM and *epi*MODE are directly cited from their respective papers.

Method	SNP (Epistatic Interaction)	P-value	Gene	Chromosome
SNPRuler	rs10499664	1.099e-05	<i>N/A</i>	Chr7
	rs10510608		<i>MGC61571</i>	Chr3
	rs10493240	3.117e-05	<i>N/A</i>	Chr1
	rs10512484		<i>N/A</i>	Chr17
	rs10485289	3.535e-05	<i>N/A</i>	Chr6
	rs10513713		<i>N/A</i>	Chr3
	rs727113	4.038e-05	<i>N/A</i>	Chr13
	rs9284239		<i>N/A</i>	Chr13
	rs1331977	4.461e-05	<i>APBA1</i>	Chr9
	rs10506115		<i>N/A</i>	Chr12
	rs2135488	4.468e-05	<i>N/A</i>	Chr13
	rs6772323		<i>KCNAB1</i>	Chr3
	rs727922	9.632e-05	<i>CAMK4</i>	Chr5
	rs2549513		<i>N/A</i>	Chr16
	rs4674305	1.013e-04	<i>VILI</i>	Chr2
	rs10513027		<i>N/A</i>	Chr5
	rs3798425	1.750e-04	<i>MYO6</i>	Chr6
	rs692897		<i>COP1</i>	Chr11
	rs10487321	2.066e-04	<i>N/A</i>	Chr7
	rs1331977		<i>APBA1</i>	Chr9
	<b>rs380390</b>	1.752e-06	<i>CFH</i>	Chr1
	rs10511901	6.829e-04	<i>N/A</i>	Chr9
	rs10488343	1.100e-03	<i>N/A</i>	Chr7
	s10505093	2.800e-03	<i>N/A</i>	Chr8
AntEpiSeeker	rs3844556	1.473e-06	<i>PKD2L1</i>	Chr10
	rs10507949		<i>N/A</i>	Chr13
	rs10492225	4.250e-05	<i>N/A</i>	Chr12
	rs176880		<i>ZNF37A</i>	Chr10
	rs724820	5.901e-05	<i>C20orf133</i>	Chr20
	rs2143881		<i>N/A</i>	Chr6
	<b>rs380390</b>	6.500e-05	<i>CFH</i>	Chr1
	rs944838		<i>GPC6</i>	Chr13
	rs10484309	9.678e-05	<i>N/A</i>	Chr6
	rs10254116		<i>PTHB1</i>	Chr7
BEAM	<b>rs380390</b>	1.752e-06	<i>CFH</i>	Chr1
	<b>rs1329428</b>	3.643e-06	<i>CFH</i>	Chr1
<i>epi</i> MODE	rs3743175	7.398e-07	<i>ZNF291</i>	Chr15
	<b>rs1394608</b>		<i>SGCD</i>	Chr1
	rs3743175	1.341e-05	<i>ZNF291</i>	Chr15
	rs2828155		<i>N/A</i>	Chr21
	rs10507949		<i>N/A</i>	Chr13
	<b>rs1394608</b>		<i>SGCD</i>	Chr1
	rs3886593	3.698e-05	<i>MSI2</i>	Chr17
	rs10491243		<i>N/A</i>	Chr5
	<b>rs1329428</b>		<i>CFH</i>	Chr1
	<b>rs380390</b>		<i>CFH</i>	Chr1
BOOST	<b>rs380390</b>	1.752e-06	<i>CFH</i>	Chr1
	<b>rs1329428</b>	3.643e-06	<i>CFH</i>	Chr1
	rs3775652	4.152e-05	<i>INPP4B</i>	Chr4
	<b>rs1394608</b>	8.806e-05	<i>SGCD</i>	Chr1

rs1930022	2.194e-04	<i>N/A</i>	Chr9
rs994542	<b>5.068e-09</b>	<i>N/A</i>	Chr6
rs9298846		<i>N/A</i>	Chr9
rs2359907	1.242e-06	<i>N/A</i>	Chr1
rs9326913		<i>N/A</i>	Chr5
rs2957935	1.627e-06	<i>SKAP1</i>	Chr17
rs10519745		<i>N/A</i>	Chr5
rs2386906	1.915e-05	<i>N/A</i>	Chr10
rs9313446		<i>SLIT3</i>	Chr5
rs293379	2.017e-05	<i>ABHD2</i>	Chr15
rs9313446		<i>SLIT3</i>	Chr5

## REFERENCES

1. Zhang Y, Liu JS: **Bayesian inference of epistatic interactions in case-control studies.** *Nature Genetics* 2007, **39**(9):1167-1173.
2. Shang J, Zhang J, Sun Y, Zhang Y: **EpiMiner: A three-stage co-information based method for detecting and visualizing epistatic interactions.** *Digital Signal Processing* 2014, **24**:1-13.
3. Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NLS, Yu W: **BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies.** *American Journal of Human Genetics* 2010, **87**(3):325.
4. Wan X, Yang C, Yang Q, Xue H, Tang NL, Yu W: **Predictive rule inference for epistatic interaction detection in genome-wide association studies.** *Bioinformatics* 2010, **26**(1):30-37.
5. Wang Y, Liu X, Robbins K, Rekaya R: **AntEpiSeeker: detecting epistatic interactions for case-control studies using a two-stage ant colony optimization algorithm.** *BMC Research Notes* 2010, **3**(1):117.
6. Tang W, Wu X, Jiang R, Li Y: **Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy.** *PLoS genetics* 2009, **5**(5):e1000464.
7. Shang J, Zhang J, Sun Y, Liu D, Ye D, Yin Y: **Performance analysis of novel methods for detecting epistasis.** *BMC Bioinformatics* 2011, **12**(1):475.
8. Shang J, Zhang J, Lei X, Zhang Y, Chen B: **Incorporating heuristic information into ant colony optimization for epistasis detection.** *Genes & Genomics* 2012, **34**(3):321-327.
9. Aflakparast M, Salimi H, Gerami A, Dub éM, Visweswaran S, Masoudi-Nejad A: **Cuckoo search epistasis: a new method for exploring significant genetic interactions.** *Heredity* 2014, **112**(6):666-674.