Supporting Information File 2 for “Integrative Analysis of Genetical Genomics Data Incorporating Network Structures” by

Bin Gao[1,2†], Xu Liu[3†], Hongzhe Li[4], and Yuehua Cui[1*]

[1]*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA*

[2]*Quantitative Sciences, Janssen Research & Development, LLC, Spring House, PA 19477*

[3]*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China*

[4]*Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, PA 19014, USA*

[†]These two authors contributed equally to the work.

*Corresponding author: cuiy@msu.edu

This supporting information file contains a list of SNPs (eQTL) for each one of the top 3 genes listed in Table 4 in the main context, and the results for additional pathway analysis.

1. SNP list

Table S0: List of SNPs (eQTL) of the three selected genes in KEGG “metabolism of xenobiotics by cytochrome P450” pathway.

CYP2E1 (ch10)				**EPHX1 (ch1)**				**SULT2A1 (ch19)**			
ch	**position**	**dbSNP_rsID**	**gene symbol**	**ch**	**position**	**dbSNP_rsID**	**gene symbol**	**ch**	**position**	**dbSNP_rsID**	**gene symbol**
1	94505899	rs7535547	-	1	88671547	rs10923082	-	1	4236741	rs2411738	-
2	178678885	rs1405645	PDE11A	2	195812999	rs777728	-	1	5427658	rs12724170	-
3	127870573	rs2139747	-	2	211354884	rs4233981	-	2	172065670	rs16859370	METTL8
3	150664359	rs4456805	-	3	4486891	rs2633844	SUMF1	2	201138169	rs6722649	LOC26010
3	179064725	rs13076136	-	3	132888749	rs4854829	CPNE4	2	228754646	rs12694772	SPHKAP
3	191511082	rs1403604	CLDN1	3	138082512	rs2031029	NCK1	3	150880491	rs1946704	-
5	161035720	rs6893149	-	3	142535528	rs17787621	ZBTB38	4	38314401	rs4073968	-
6	11270212	rs6456946	-	4	12794110	rs12646501	-	4	39819270	rs6846248	-
6	143792107	rs11155293	ADAT2	4	167354019	rs2292083	TLL1	4	41164354	rs17444017	-
7	80325910	rs1636594	-	5	97244313	rs6895924	-	4	84449248	rs4693571	-
7	146110550	rs10238991	CNTNAP2	8	19217659	rs10107389	SH2D4A	5	13479104	rs4702943	-
7	148539819	rs886712	-	8	77726733	rs16939329	-	5	55615894	rs158497	-
8	23790794	rs4872203	-	9	107441555	rs6477578	-	5	143538057	rs6876352	KCTD16
9	102700439	rs2639345	-	10	15875283	rs12262809	C10orf97	5	174788017	rs2644640	-
10	21584363	rs6482170	-	10	18369762	rs12358478	SLC39A12	6	44616383	rs736259	-
10	30654280	rs10763778	PAPD1	10	24445562	rs11013905	KIAA1217	8	23790505	rs17723874	-
12	11926931	rs2739085	ETV6	10	69630097	rs4746739	MYPN	8	106377210	rs7823980	-
13	100815014	rs7986611	NALCN	10	107214664	rs2611050	-	10	15875283	rs12262809	C10orf97
13	100917657	rs4112527	ITGBL1	11	90015379	rs4753018	-	11	16702219	rs1392998	-
				12	19709769	rs10770516	-	11	79038552	rs651420	-
				12	110179096	rs1265566	CUX2	11	81286656	rs1945897	-
				13	52868079	rs9536409	-	12	91363010	rs2277364	-
				14	57547077	rs17094208	C14orf37	13	53060274	rs9536471	-
				17	24375446	rs997996	LOC100130794	14	29229449	rs11847410	PRKD1
				17	36253248	rs2469831	CPS1				

2. Additional pathway analysis

The purpose of additional pathway analysis was to check the performance of the methods from a large number of pathways. We randomly chose 65 pathways from different KEGG categories from the KEGG website (https://www.genome.jp/kegg/pathway.html). We analyzed the data with the three methods: IVGC, IV and EN. The full details about the three methods can be found in the main context. Among the 65 pathways analyzed, 4 pathways have no gene passed the 0.6 selection rate threshold. Among the rest 61 pathways, IVGC has smaller prediction error than IV and EN in 26 pathways. Pathways with the smallest prediction error by using IVGC were highlighted with yellow. Below, we list the detailed results of the 65 pathways.

Carbohydrate Metabolism

hsa00010 [xml | html | graphics]	Glycolysis / Gluconeogenesis
hsa00020 [xml | html | graphics]	Citrate cycle (TCA cycle)
hsa00030 [xml | html | graphics]	Pentose phosphate pathway
hsa00040 [xml | html | graphics]	Pentose and glucuronate interconversions
hsa00052 [xml | html | graphics]	Galactose metabolism

Glycolysis / Gluconeogenesis (hsa00010)

There are total 68 genes in this pathway, among which 62 were mapped to this dataset. So we focused the analysis on the 62 genes. Table S1 lists the top selected genes which passed the 60% stability selection rate by using any one of the three methods (IVHC, IV and EN). Overall, the top genes have higher selection rate using IVGC than the other two methods. The prediction errors of the IVGC, IV and EN are 0.4909, 0.5307 and 0.5613, respectively, indicating good performance of IVGC in selecting important genes in this pathway.

Table S1: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00010*.

Gene Symbol	IVGC	IV	EN
PDHA1	0.94	0.8	0.99
LDHAL6A	0.94	0.51	0.86
DLAT	0.94	0.32	0.71
PFKM	0.93	0.7	0.95
ALDH2	0.93	0.6	0.92
DLD	0.93	0.15	0.39
ENO3	0.92	0.23	0.62
ADH1B	0.92	0.14	0.4
LDHA	0.92	0.06	0.4
ALDOC	0.91	0.71	0.96
ADPGK	0.91	0.61	0.91
PKLR	0.91	0.07	0.17
PGM2	0.89	0.24	0.67
G6PC2	0.86	0.36	0.72
ENO2	0.86	0.28	0.65
GPI	0.86	0.13	0.34
HK1	0.84	0.11	0.28
ADH6	0.83	0.08	0.34
LDHB	0.83	0	0.07
PDHA2	0.82	0.41	0.8
ACSS1	0.79	0.31	0.73
ALDH3A2	0.79	0.27	0.66

ACSS2	0.79	0.23	0.6
ALDH1B1	0.76	0.33	0.69
ADH4	0.76	0.04	0.13
AKR1A1	0.73	0.02	0.1
LDHC	0.71	0.17	0.53
ADH7	0.7	0.06	0.19
TPI1	0.68	0.05	0.27
ALDH3B2	0.67	0.1	0.36
GALM	0.66	0.33	0.66
G6PC3	0.61	0.02	0.16
G6PC	0.26	0.22	0.61

The network connectivity plots of the 62 and the top selected genes are shown in Fig. S1. Given the tight connection information among the genes, it is not surprising to see the improved prediction performance of the IVGC method compared to others.

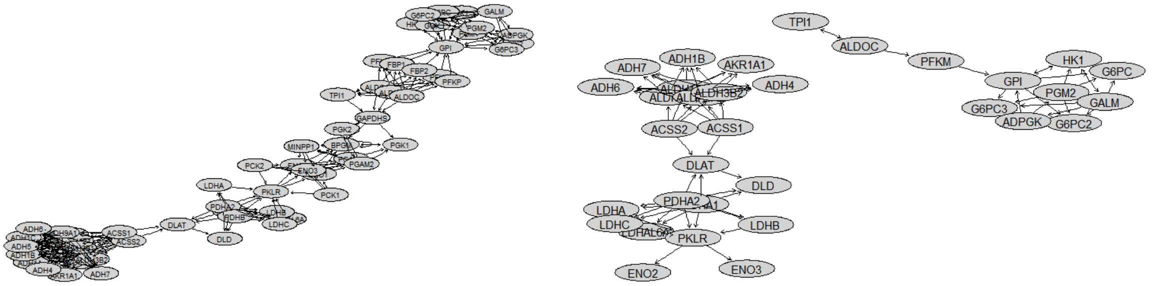

Figure S1: The network connectivity plots of the 62 genes (left) and the top selected genes (right) for pathway *hsa00010*.

Citrate cycle (TCA cycle) (hsa00020)

There are total 30 genes in this pathway and all were mapped to this dataset. Table S2 lists the top selected genes which passed the 60% stability selection rate by using any one of the three methods (IVHC, IV and EN). Overall, the top genes have higher selection rate using IVGC than the other two methods. The prediction errors of the IVGC, IV and EN are 0.4896, 0.5134 and 0.5040, respectively, indicating good performance of IVGC in selecting important genes in this pathway.

Table S2: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00020*.

Gene Symbol	IVGC	IV	EN
FH	1.00	0.92	1.00
PDHA1	0.99	0.90	0.97
MDH2	0.95	0.66	0.90
CS	0.95	0.51	0.69
PDHA2	0.93	0.47	0.73
PC	0.92	0.61	0.85
ACO1	0.91	0.51	0.79
IDH2	0.91	0.43	0.65
PCK1	0.82	0.11	0.26
SUCLG1	0.77	0.33	0.56
PCK2	0.77	0.04	0.07
ACLY	0.74	0.04	0.10
SDHC	0.70	0.01	0.05

DLAT	0.68	0.36	0.52
DLST	0.68	0.10	0.34
IDH3G	0.64	0.03	0.10
IDH1	0.62	0.04	0.15
PDHB	0.62	0.03	0.13

Figure S2 shows the network connectivity information of the 30 genes and the top selected genes listed in Table S2. Given the tight connection information among the genes, it is not surprising to see the improved prediction performance of the IVGC method compared to others.

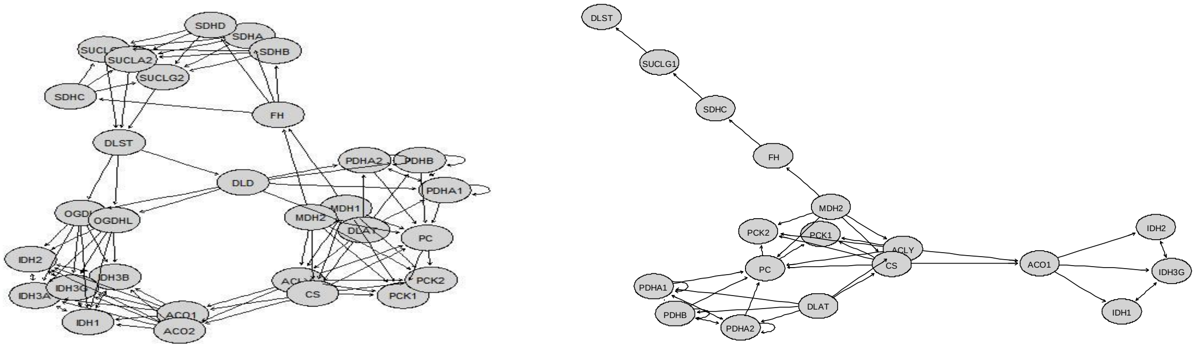

Figure S2: The network connectivity plots of the 30 genes (left) and the top selected genes (right) for pathway *hsa00020*.

Pentose phosphate pathway (hsa00030)

There are total 30 genes in this pathway, among which 28 were mapped to this dataset. So we focused the analysis on the 28 genes. Table S3 lists the top selected genes which passed the 60% stability selection rate by using any one of the three methods (IVHC, IV and EN). The prediction errors of the IVGC, IV and EN are 0.5410, 0.5388 and 0.5373, respectively. For this pathway, IVGC has slightly larger prediction error than the other two methods.

Table S3: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00030*.

Gene Symbol	**IVGC**	**IV**	**EN**
ALDOC	0.81	0.75	0.97
PFKM	0.81	0.72	0.96
GLYCTK	0.72	0.6	0.89
TKT	0.68	0.6	0.84
RBKS	0.47	0.35	0.62
PRPS2	0.46	0.41	0.77

Figure S3 shows the network connectivity information of the 28 genes and the top selected genes listed in Table S3. The connectivity information among the 6 genes are relatively high. Although IVGC has larger prediction error, the difference not very significant among the three. For this pathway, EN method gives higher selection rate for most genes compared to IVGC and IV.

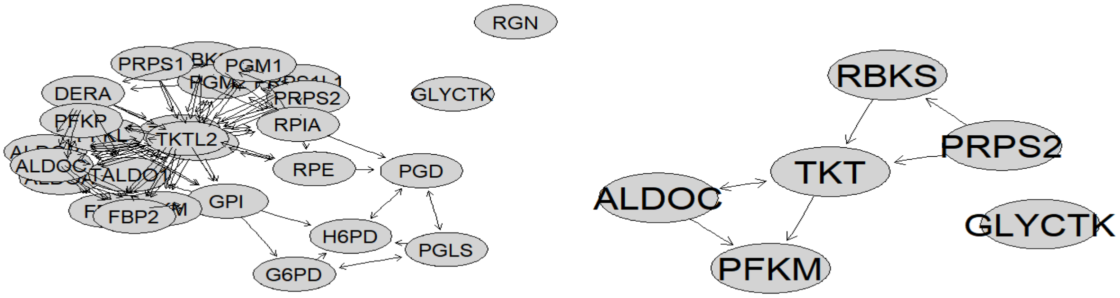

Figure S3: The network connectivity plots of the 28 genes (left) and the top selected genes (right) for pathway *hsa00030*.

Pentose and glucuronate interconversions (hsa00040)

There are total 34 genes in this pathway, among which 31 were mapped to the dataset. So we focused the analysis on the 31 genes. Table S4 lists the top selected genes. The prediction errors for IVGC, IV and EN are 0.5165, 0.5148 and 0.5159, respectively. The network connectivity information are shown in Fig. S4. From the figure, the top selected genes have no direct connection in this pathway. Thus, it is not surprising to see the similar prediction error of the three methods.

Table S4: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00040*.

Gene Symbol	**IVGC**	**IV**	**EN**
UGT1A1	0.81	0.85	0.92
AKR1B1	0.78	0.83	0.91
UGT1A9	0.64	0.58	0.82

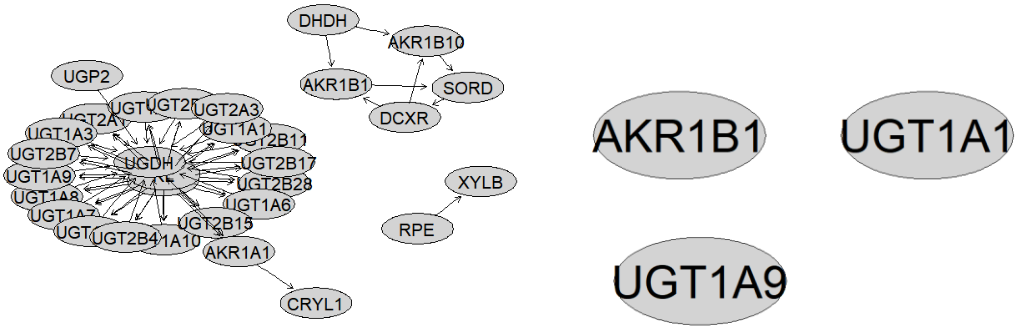

Figure S4: The network connectivity plots of the 28 genes (left) and the top selected genes (right) for pathway *hsa00040*.

Galactose metabolism (hsa00052)

There are total 31 genes in this pathway, among which 29 were mapped to this dataset. So we focused the analysis on the 29 genes. Table S5 lists the top selected genes which passed the 60% stability selection rate by using any one of the three methods (IVHC, IV and EN). The prediction errors of the IVGC, IV and EN are 0.5210, 0.5695 and 0.5357, respectively, indicating good performance of IVGC in selecting important genes in this pathway.

Table S5: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00052*.

Gene Symbol	**IVGC**	**IV**	**EN**
PFKM	0.75	0.56	0.89

GALM	0.74	0.51	0.82
GLB1	0.7	0.47	0.79
GALK1	0.67	0.36	0.61
B4GALT2	0.66	0.53	0.82
LCT	0.66	0.34	0.65
PGM2	0.4	0.33	0.64
AKR1B1	0.31	0.47	0.83
GALT	0.23	0.32	0.61

Figure S5 shows the network connectivity information of the 29 genes and the top selected genes listed in Table S5. Although IVGC has the smallest prediction error among the three methods, it does not give high selection rate for the listed genes compared to EN.

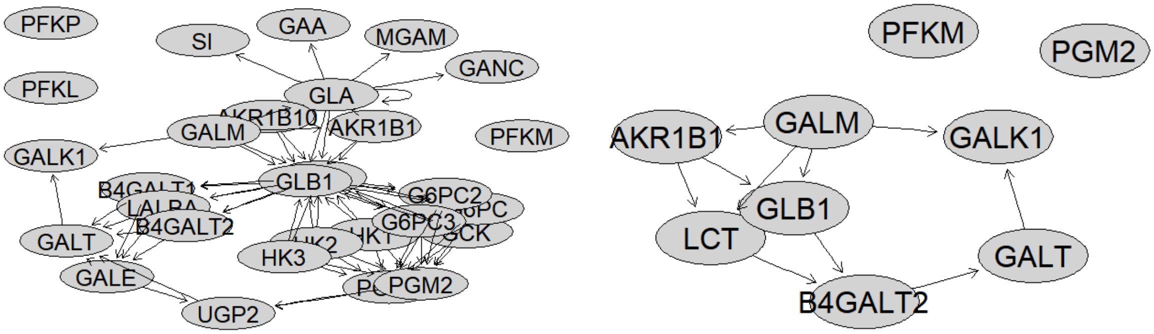

Figure S5: The network connectivity plots of the 29 genes (left) and the top selected genes (right) for pathway *hsa00052*.

Energy Metabolism

hsa00190 [xml | html | graphics] Oxidative phosphorylation
hsa00910 [xml | html | graphics] Nitrogen metabolism
hsa00920 [xml | html | graphics] Sulfur metabolism

Oxidative phosphorylation (hsa00190)

There are total 133 genes in this pathway, among which 97 were mapped to the dataset. So we focused the analysis on the 97 genes. The prediction errors of the three methods are 0.5439 0.5582 0.5470, corresponding to method IVGC, IV and EN, respectively. There is only one gene with a selection rate >60% by using EN method. Both IVGC and IV has low selection rate. By checking the gene connectivity information (Fig. S6), they are quite loose.

Table S6: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00190*.

Gene Symbol	**IVGC**	**IV**	**EN**
COX7A1	0.3	0.38	0.66

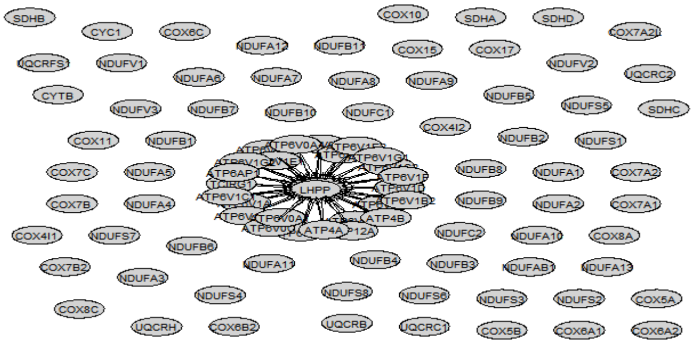

Figure S6: The network connectivity plots of the 97 analyzed genes in pathway *hsa00190*.

Nitrogen metabolism(hsa00910)

There are total 17 genes in this pathway and all were mapped to the dataset. So we focused the analysis on the 17 genes. The prediction errors of the three methods are 0.5411, 0.5359 and 0.5319, respectively corresponding to IVGC, IV and EN. Again, the gene connectivity information shown in Fig. S7 is quite loose.

Table S7: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00910*.

Gene Symbol	IVGC	IV	EN
GLUD1	0.4	0.41	0.7

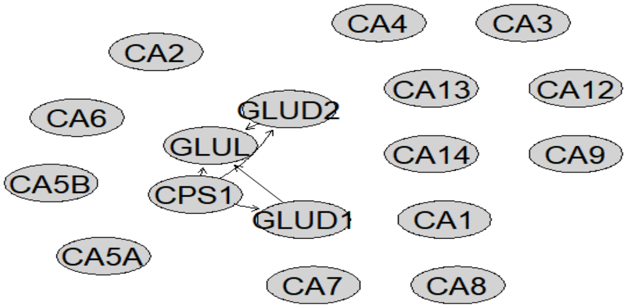

Figure S7: The network connectivity plots of the 17 analyzed genes in pathway *hsa00910*.

Sulfur metabolism(hsa00920)

There are total 9 genes in this pathway, among which 7 were mapped to the dataset. So we focused the analysis on the 7 genes. The prediction errors of the three methods are 0.5404, 0.5333 and 0.5302. IVGC does not show improved prediction performance for this pathway.

Table S8: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00920*.

Gene Symbol	IVGC	IV	EN
SUOX	0.86	0.82	0.98
MPST	0.59	0.49	0.73
TST	0.54	0.35	0.64
PAPSS1	0.48	0.41	0.66

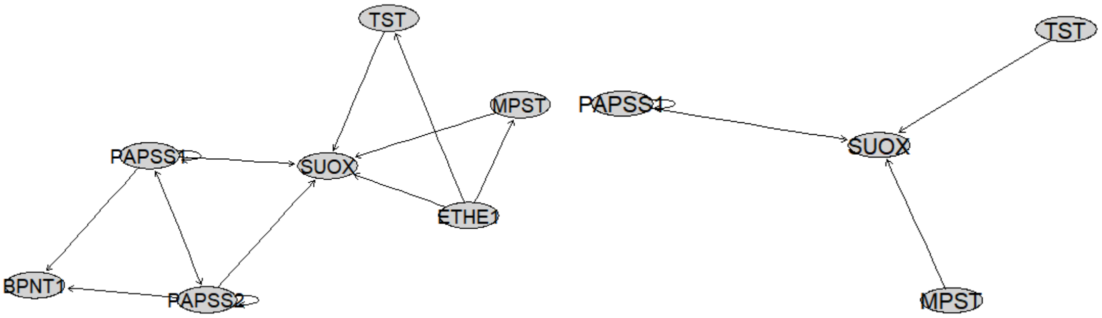

Figure S8: The network connectivity plots of the 7 genes (left) and the top selected genes (right) for pathway *hsa00920*.

Lipid Metabolism

hsa00061 [xml | html | graphics] Fatty acid biosynthesis
hsa00062 [xml | html | graphics] Fatty acid elongation in mitochondria
hsa00071 [xml | html | graphics] Fatty acid metabolism
hsa00072 [xml | html | graphics] Synthesis and degradation of ketone bodies
hsa00100 [xml | html | graphics] Biosynthesis of steroids

Fatty acid biosynthesis (hsa00061)

There are total 13 genes in this pathway, among which 9 were mapped to the dataset. So we focused the analysis on the 9 genes. The prediction errors of the three methods are 0.5395, 0.5331 and 0.5395. There is only one gene passed the 60% selection threshold by EN method.

Table S9: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00061*.

Gene Symbol	IVGC	IV	EN
ACACB	0.44	0.37	0.74

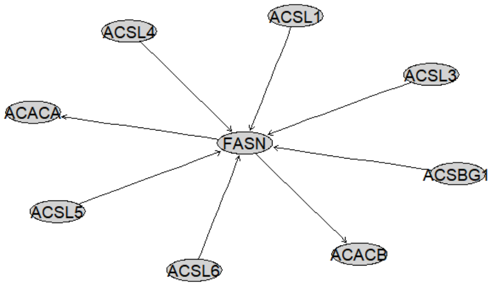

Figure S9: The network connectivity plots of the 9 analyzed genes in pathway *hsa00061*.

Fatty acid elongation in mitochondria (hsa00062)

There are total 30 genes in this pathway, among which 16 were mapped to the expression data. So we focused the analysis on the 16 genes. The prediction errors for IVGC, IV and EN are 0.5088, 0.5065 and 0.5120, respectively. For this pathway, IV has the smallest prediction error among the three, although the difference is not very significant. The connectivity plot in Fig. S10 shows that the top selected genes have no direct connection.

Table S10: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00062*.

Gene Symbol	IVGC	IV	EN
ELOVL6	0.94	0.97	0.99
ELOVL3	0.94	0.96	0.99
ACAA2	0.87	0.86	0.91
ELOVL7	0.71	0.71	0.78

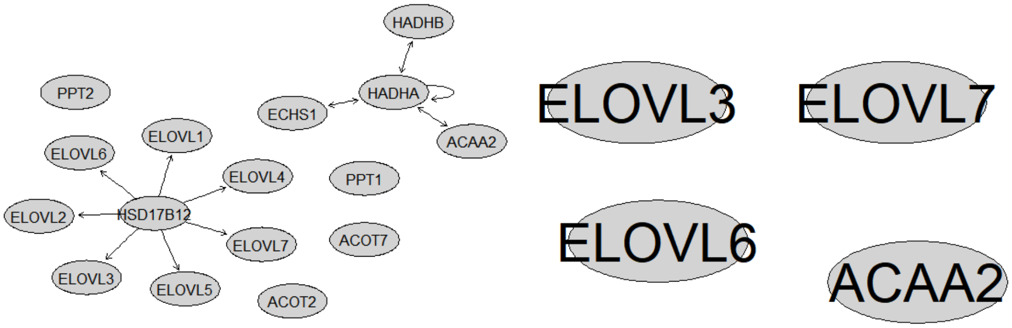

Figure S10: The network connectivity plots of the 16 genes (left) and the top selected genes (right) for pathway *hsa00062*.

Fatty acid metabolism (hsa00071)

There are total 44 genes in this pathway, among which 39 were mapped to the expression data. So we focused the analysis on the 39 genes. The prediction errors for IVGC, IV and EN are 0.4831, 0.4834 and 0.4874, respectively. The connectivity plot in Fig. S11 shows that the top selected genes have no direct connection.

Table S11: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00071*.

Gene Symbol	IVGC	IV	EN
ACAT2	1	1	1
CPT1A	0.85	0.87	0.94
ALDH1B1	0.56	0.58	0.75

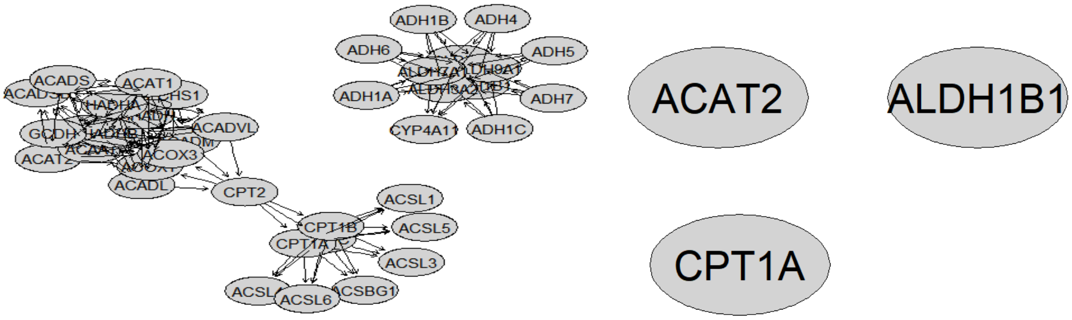

Figure S11: The network connectivity plots of the 39 genes (left) and the top selected genes (right) for pathway *hsa00071*.

Synthesis and degradation of ketone bodies (hsa00072)

There are total 10 genes in this pathway, among which 7 were mapped to the expression data. So we focused the analysis on the 7 genes. The prediction error for the three methods are 0.4801 0.4759 0.4805, respectively. There is only one gene passed the 60% selection threshold.

Table S12: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00072*.

Gene symbol	IVGC	IV	EN
ACAT2	1	1	1

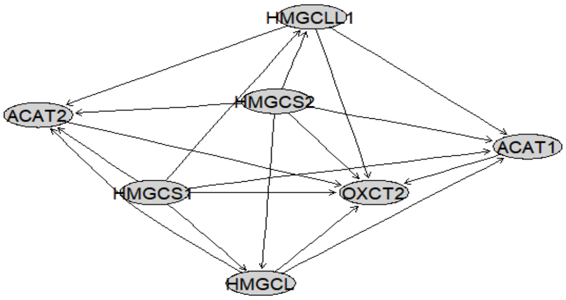

Figure S12: The network connectivity plots of the 7 analyzed genes in pathway *hsa00072*.

Biosynthesis of steroids(hsa00100)

There are total 19 genes in this pathway, among which 16 were mapped to the expression data. So we focused the analysis on the 16 genes. The prediction errors for the three methods are 0.4946, 0.4891 and 0.4932, respectively.

Table S13: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00100*.

Gene Symbol	IVGC	IV	EN
FDFT1	1	1	1
DHCR24	1	0.96	0.97
EBP	0.96	0.92	0.96
TM7SF2	0.82	0.43	0.66
SQLE	0.75	0.15	0.36
LSS	0.68	0.13	0.35
CYP51A1	0.6	0.03	0.1

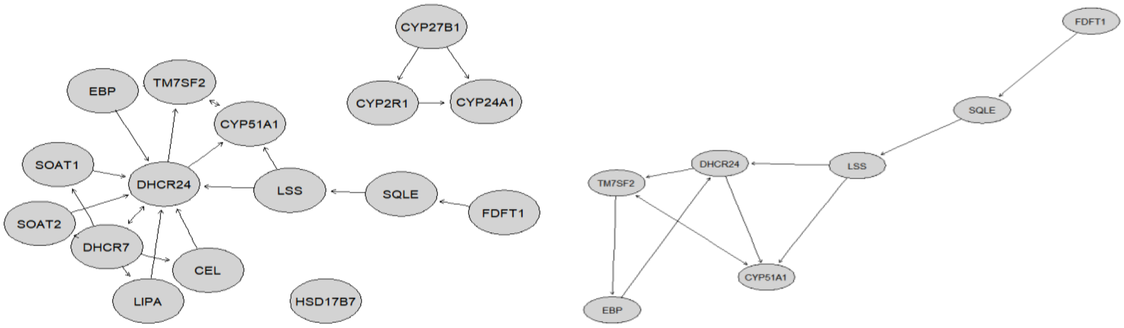

Figure S13: The network connectivity plots of the 16 genes (left) and the top selected genes (right) for pathway *hsa00100*.

Nucleotide Metabolism

hsa00240 [xml | html | graphics]   Pyrimidine metabolism

There are total 101 genes in this pathway, among which 80 were mapped to the expression data. So we focused the analysis on the 80 genes. The prediction error for the three methods are 0.5145, 0.5093 and 0.5108, respectively. For this pathway, IV has the smallest prediction error, although the difference is not very significant. The selection rates for the three methods are quite similar, indicating similar performance of the three methods. Network connectivity plot shows that the connectivity is quite loose among the top selected genes.

Table S14: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00240*.

Gene Symbol	IVGC	IV	EN
DCTD	0.96	0.96	0.99
ENPP1	0.91	0.91	0.98
ENTPD5	0.66	0.65	0.75
POLE2	0.59	0.53	0.71

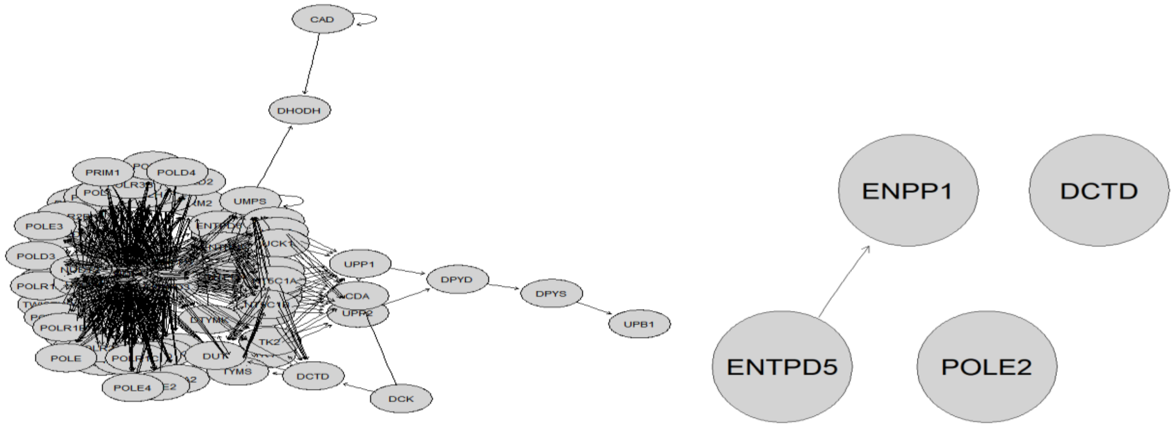

Figure S14: The network connectivity plots of the 80 genes (left) and the top selected genes (right) for pathway *hsa00240*.

Amino Acid Metabolism

hsa00260 [xml | html | graphics]   Glycine, serine and threonine metabolism
hsa00280 [xml | html | graphics]   Valine, leucine and isoleucine degradation
hsa00310 [xml | html | graphics]   Lysine degradation
hsa00330 [xml | html | graphics]   Arginine and proline metabolism
hsa00350 [xml | html | graphics]   Tyrosine metabolism

Glycine, serine and threonine metabolism(hsa00260)

There are total 40 genes in this pathway, among which 38 were mapped to the expression data. So we focused the analysis on the 38 genes. The prediction error for the three methods are 0.5303, 0.5416 and 0.5339, respectively. For this pathway, the three methods perform quite similarly with IVGC giving smaller prediction error among the three methods.

Table S15: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00260*.

Gene Symbol	IVGC	IV	EN
ALAS1	0.89	0.8	0.96
CHDH	0.87	0.76	0.94
SARDH	0.78	0.72	0.87
MAOB	0.59	0.45	0.68

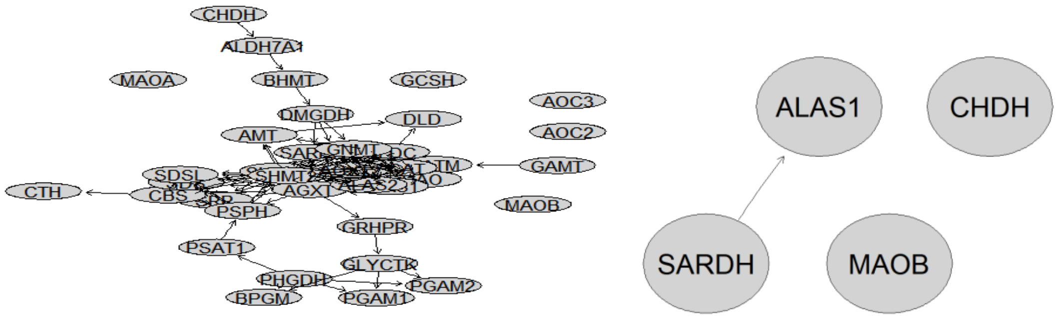

Figure S15: The network connectivity plots of the 38 genes (left) and the top selected genes (right) for pathway *hsa00260*.

Valine, leucine and isoleucine degradation(hsa00280)

There are total 48 genes in this pathway, among which 44 were mapped to the expression data. So we focused the analysis on the 44 genes. The prediction errors for the three methods are 0.4912, 0.4908 and 0.4896, respectively. Again, the selection performance of the three methods is quite similar.

Table S16: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00280*.

Gene Symbol	IVGC	IV	EN
ACAT2	1	0.99	1
AOX1	0.94	0.84	0.93
ABAT	0.9	0.78	0.88
ALDH1B1	0.62	0.47	0.62

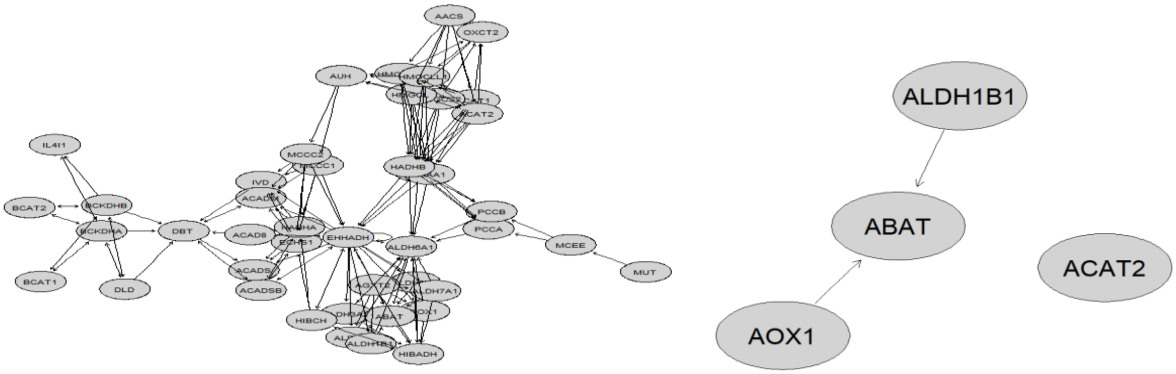

Figure S16: The network connectivity plots of the 44 genes (left) and the top selected genes (right) for pathway *hsa00280*.

Lysine degradation(hsa00310)

There are total 59 genes in this pathway, among which 36 were mapped to the expression data. So we focused the analysis on the 36 genes. The prediction errors for the three methods are 0.4874, 0.4939 and 0.5059, respectively. Although the selection rate with IVGC is not as high as EN due to loose connection in the network (Fig. S17), it does have the smallest prediction error among the three.

Table S17: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00310*.

Gene Symbol	IVGC	IV	EN
ACAT2	1	0.99	1
PRDM6	0.94	0.9	0.97
PRDM2	0.75	0.63	0.87
DOT1L	0.74	0.67	0.92
AADAT	0.67	0.61	0.83
ALDH1B1	0.6	0.51	0.79
ALDH2	0.5	0.38	0.67
EZH1	0.47	0.35	0.67

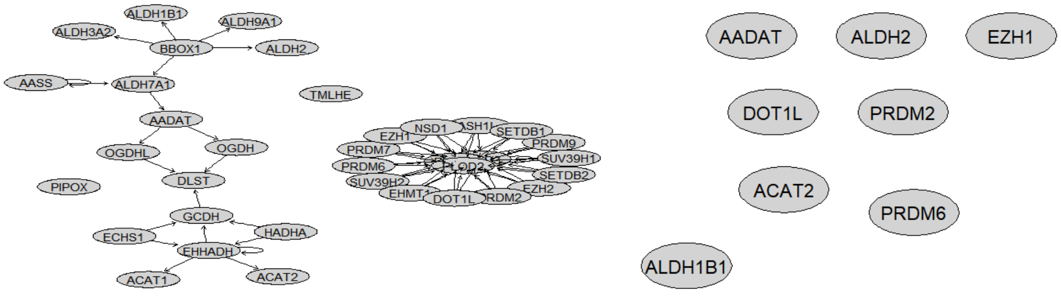

Figure S17: The network connectivity plots of the 36 genes (left) and the top selected genes (right) for pathway *hsa00310*.

Arginine and proline metabolism(hsa00330)

There are total 50 genes in this pathway, among which 40 were mapped to the expression data. So we focused the analysis on the 40 genes. The prediction errors for the three methods are 0.5361, 0.5721 and 0.5554, respectively. Overall, IVGC has better selection rate and smaller prediction error than the other two methods.

Table S18: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00330*.

Gene Symbol	IVGC	IV	EN
AOX1	0.94	0.84	0.93
ABAT	0.9	0.78	0.88
ALDH2	0.81	0.45	0.78
SMS	0.76	0.39	0.78
MAOB	0.76	0.39	0.62
CKMT2	0.76	0.34	0.77
ALDH4A1	0.76	0.27	0.64
GOT2	0.72	0.43	0.68
ARG1	0.68	0.28	0.66
ALDH3A2	0.68	0.15	0.39

OAT	0.67	0.17	0.45
ALDH1B1	0.62	0.47	0.62
SAT2	0.62	0.27	0.62
GOT1	0.6	0.21	0.57

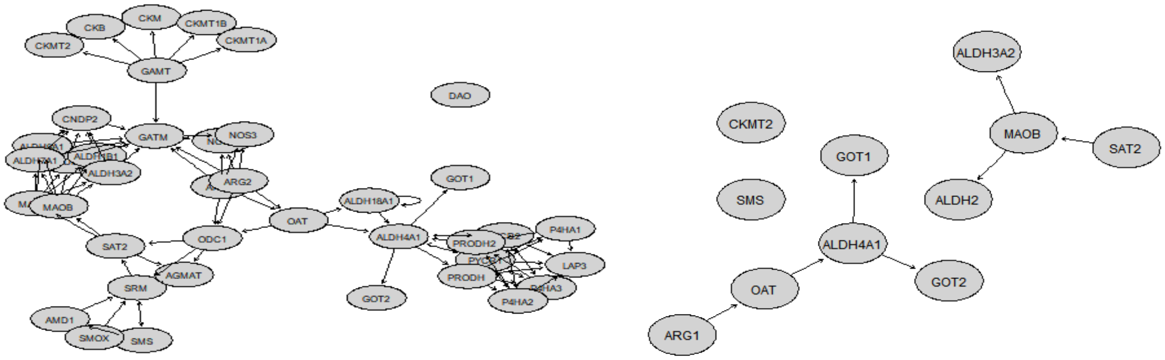

Figure S18: The network connectivity plots of the 40 genes (left) and the top selected genes (right) for pathway *hsa00330*.

Tyrosine metabolism(hsa00350)

There are total 36 genes in this pathway, among which 34 were mapped to the expression data. So we focused the analysis on the 34 genes. The prediction errors for the three methods are 0.5051, 0.5271 and 0.5221, respectively. Among the three methods, IVGC has the smallest prediction error and the largest selection rate for nearly all the genes.

Table S19: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00350*.

Gene Symbol	**IVGC**	**IV**	**EN**
AOX1	0.97	0.91	0.98
DDC	0.9	0.53	0.73
MAOB	0.88	0.43	0.7
GOT2	0.75	0.32	0.5
GOT1	0.74	0.36	0.63
HGD	0.63	0.29	0.48

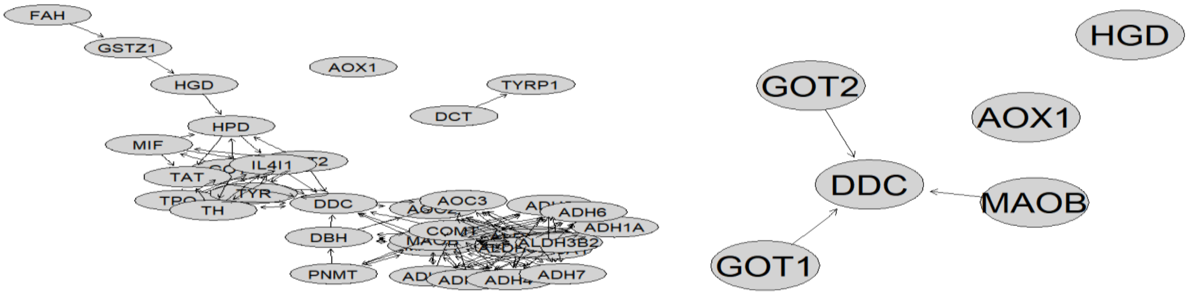

Figure S19: The network connectivity plots of the 40 genes (left) and the top selected genes (right) for pathway *hsa00350*.

Glycan Biosynthesis and Metabolism

hsa00510 [xml | html | graphics] N-Glycan biosynthesis

hsa00512 [xml | html | graphics] O-Glycan biosynthesis

hsa00532 [xml | html | graphics]   Chondroitin sulfate biosynthesis
hsa00531 [xml | html | graphics]   Glycosaminoglycan degradation

N-Glycan biosynthesis(hsa00510)

There are total 49 genes in this pathway, among which 39 were mapped to the expression data. So we focused the analysis on the 39 genes. The prediction errors for the three methods are 0.5459, 0.5393 and 0.5261, respectively.

Table S20: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00510*.

Gene Symbol	IVGC	IV	EN
ST6GAL1	0.68	0.7	0.96
MAN1B1	0.55	0.52	0.84
B4GALT2	0.53	0.49	0.79
ALG2	0.45	0.42	0.68
B4GALT3	0.41	0.37	0.61

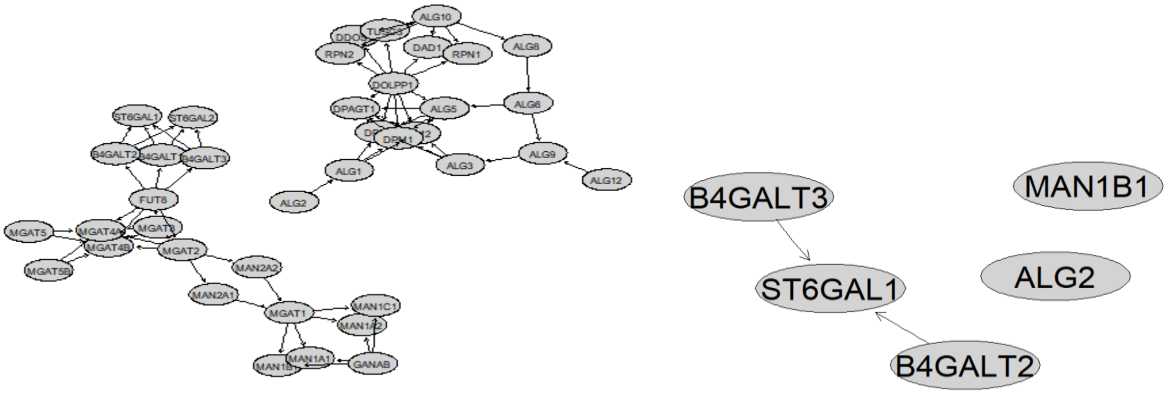

Figure S20: The network connectivity plots of the 39 genes (left) and the top selected genes (right) for pathway *hsa00510*.

O-Glycan biosynthesis(hsa00512)

There are total 31 genes in this pathway, among which 24 were mapped to the expression data. So we focused the analysis on the 24 genes. The prediction errors for the three methods are 0.5149, 0.5142 and 0.5145, respectively. The top selected genes have no direct connection (see Fig. S20).

Table S21: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00512*.

Gene Symbol	IVGC	IV	EN
GALNT3	0.97	0.89	0.98
GALNT4	0.91	0.83	0.95
GALNT6	0.85	0.68	0.83
GALNT9	0.83	0.69	0.86
GALNT8	0.65	0.3	0.56

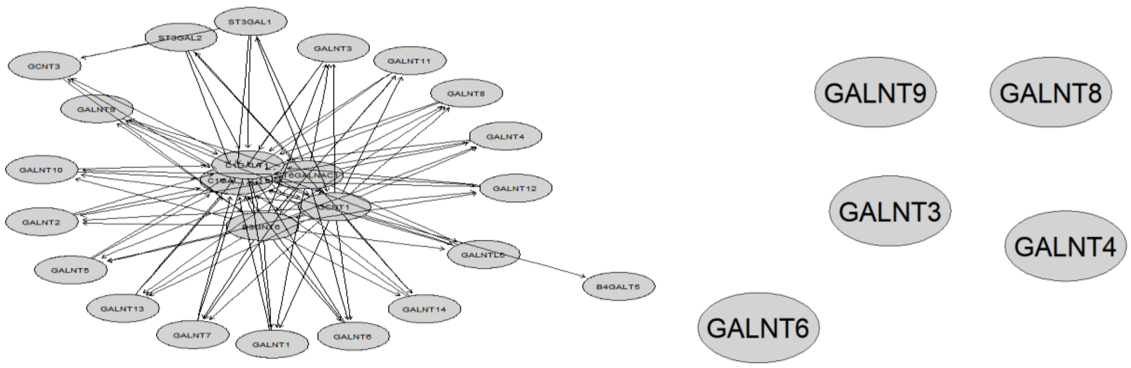

Figure S21: The network connectivity plots of the 24 genes (left) and the top selected genes (right) for pathway *hsa00512*.

Chondroitin sulfate biosynthesis(hsa00532)

There are total 20 genes in this pathway, among which 13 were mapped to the expression data. So we focused the analysis on the 13 genes. The prediction errors for the three methods are 0.5293, 0.5409 and 0.5377, respectively.

Table S22: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00532*.

Gene Symbol	IVGC	IV	EN
CHSY1	0.55	0.6	0.82
CHST3	0.54	0.53	0.8
CHST12	0.46	0.47	0.75

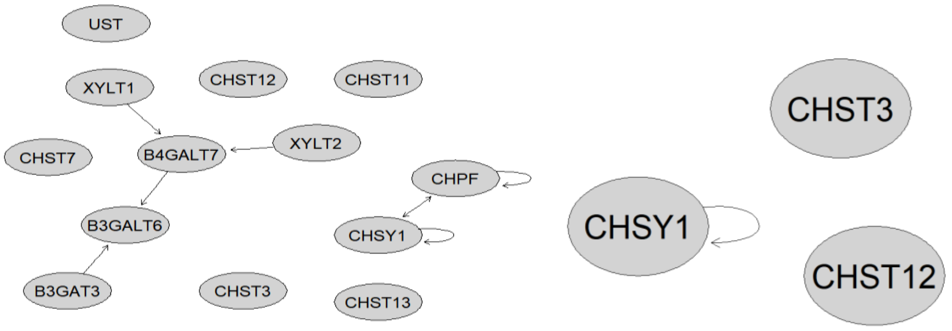

Figure S22: The network connectivity plots of the 13 genes (left) and the top selected genes (right) for pathway *hsa00532*.

Glycosaminoglycan degradation(hsa00531)

There are total 19 genes in this pathway, among which 18 were mapped to the expression data. So we focused the analysis on the 18 genes. The prediction errors for the three methods are 0.5372, 0.5446 and 0.5373, respectively.

Table S23: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00071*.

Gene Symbol	IVGC	IV	EN
GALNS	0.75	0.48	0.79
HYAL1	0.62	0.39	0.63
GLB1	0.24	0.39	0.66

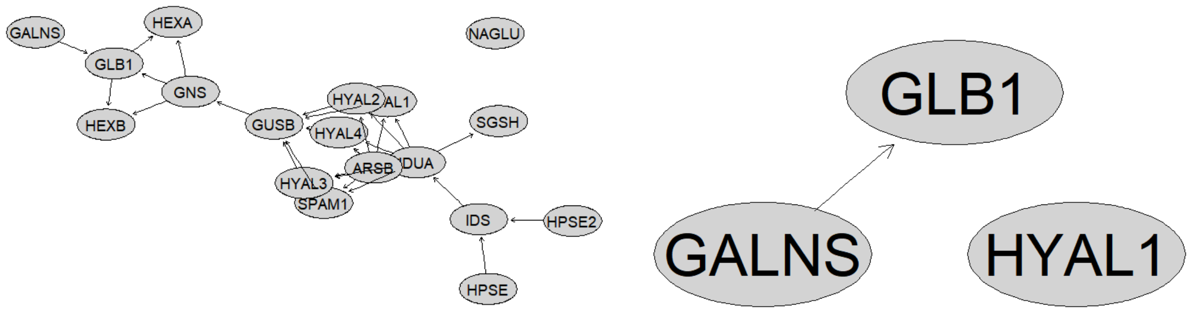

Figure S23: The network connectivity plots of the 18 genes (left) and the top selected genes (right) for pathway *hsa00531*.

Metabolism of Cofactors and Vitamins

hsa00730 [xml | html | graphics] Thiamine metabolism
hsa00740 [xml | html | graphics] Riboflavin metabolism
hsa00750 [xml | html | graphics] Vitamin B6 metabolism
hsa00760 [xml | html | graphics] Nicotinate and nicotinamide metabolism
hsa00770 [xml | html | graphics] Pantothenate and CoA biosynthesis

Thiamine metabolism(hsa00730)

There are total 16 genes in this pathway, among which 13 were mapped to the expression data. So we focused the analysis on the 13 genes. No genes have selection rate larger than 60% in this pathway.

Riboflavin metabolism(hsa00740)

There are total 8 genes in this pathway, among which 6 were mapped to the expression data. So we focused the analysis on the 6 genes. The prediction errors for the three methods are 0.5159, 0.5374 and 0.5119, respectively. For this pathway, IVGC and EN perform very similarly.

Table S24: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00740*.

Gene Symbol	IVGC	IV	EN
ENPP1	0.99	0.96	1
BLVRB	0.9	0.75	0.92
ACP2	0.85	0.52	0.8
ACP5	0.84	0.49	0.75
ACP1	0.67	0.56	0.81
ENPP3	0.49	0.35	0.63

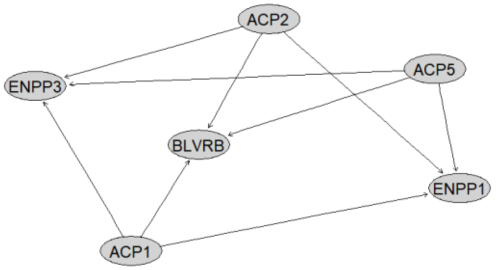

Figure S24: The network connectivity plots of the 6 genes in pathway *hsa00740*.

Vitamin B6 metabolism(hsa00750)

There are total 6 genes in this pathway, among which 6 were mapped to the expression data. So we focused the analysis on the 6 genes. The prediction errors for the three methods are 0.4979, 0.4983 and 0.5000, respectively.

Table S25: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00750*.

Gene Symbol	IVGC	IV	EN
AOX1	0.99	1	1
PHOSPHO2	0.96	0.99	1
PDXK	0.67	0.63	0.78
PNPO	0.65	0.59	0.83

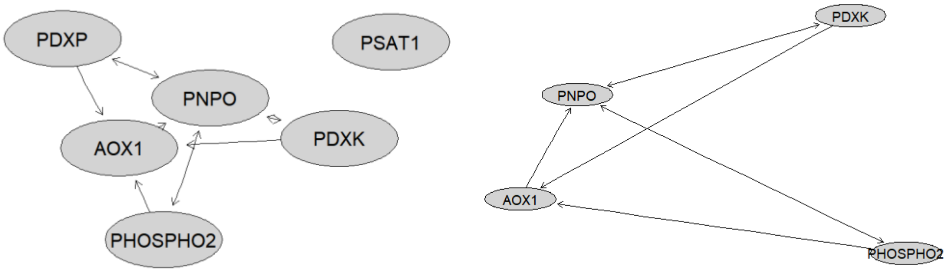

Figure S25: The network connectivity plots of the 6 genes (left) and the top selected genes (right) for pathway *hsa00750*.

Nicotinate and nicotinamide metabolism(hsa00760)

There are total 30 genes in this pathway, among which 20 were mapped to the expression data. So we focused the analysis on the 20 genes. The prediction errors for the three methods are 0.4996, 0.4976 and 0.5011, respectively.

Table S26: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00760*.

Gene Symbol	IVGC	IV	EN
AOX1	0.99	0.96	0.99
NADK	0.97	0.85	0.94
QPRT	0.96	0.85	0.92
ENPP1	0.94	0.69	0.85

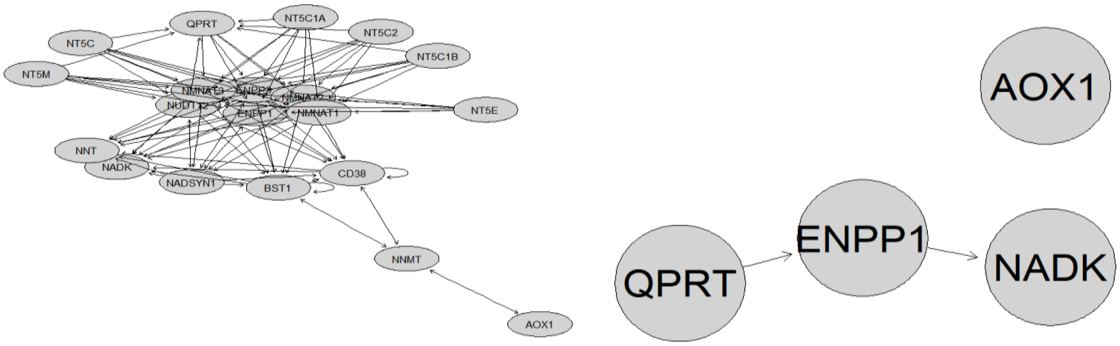

Figure S26: The network connectivity plots of the 20 genes (left) and the top selected genes (right) for pathway *hsa00760*.

Pantothenate and CoA biosynthesis(hsa00770)

There are total 19 genes in this pathway and all were mapped to the expression data. So we focused the analysis on the 19 genes. The prediction errors for the three methods are 0.5105, 0.5108 and 0.5129, respectively.

Table S27: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00770*.

Gene Symbol	IVGC	IV	EN
ENPP1	0.93	0.92	0.99
GADL1	0.85	0.83	0.93
COASY	0.85	0.81	0.94
DPYS	0.69	0.62	0.76
PPCDC	0.68	0.81	0.92
BCAT1	0.55	0.49	0.67

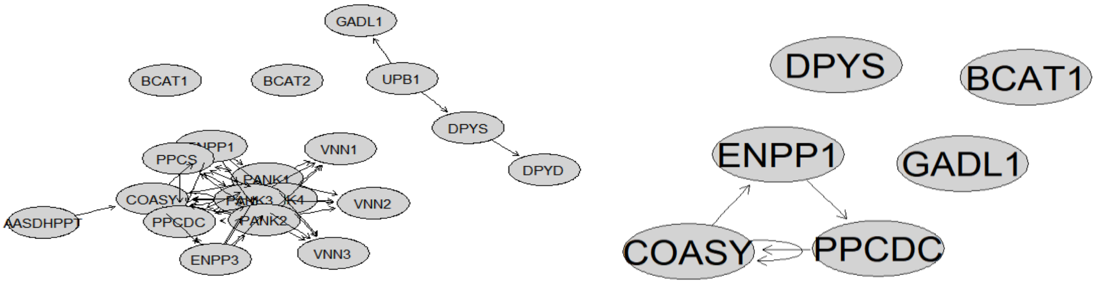

Figure S27: The network connectivity plots of the 19 genes (left) and the top selected genes (right) for pathway *hsa00770*.

Biosynthesis of Secondary Metabolites

hsa00900 [xml | html | graphics]   Terpenoid biosynthesis

There are total 22 genes in this pathway, among which 19 were mapped to the expression data. So we focused the analysis on the 19 genes. The prediction errors for the three methods are 0.4912, 0.4774 and 0.4794, respectively. Although the selection rate for the top genes is quite high for IVGC, its prediction error is quite low compared to IV and EN.

Table S28: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa00900*.

Gene Symbol	IVGC	IV	EN
ACAT2	1	1	1
HMGCS1	0.74	0.01	0.09
HMGCS2	0.69	0.09	0.15
HMGCR	0.67	0.13	0.23

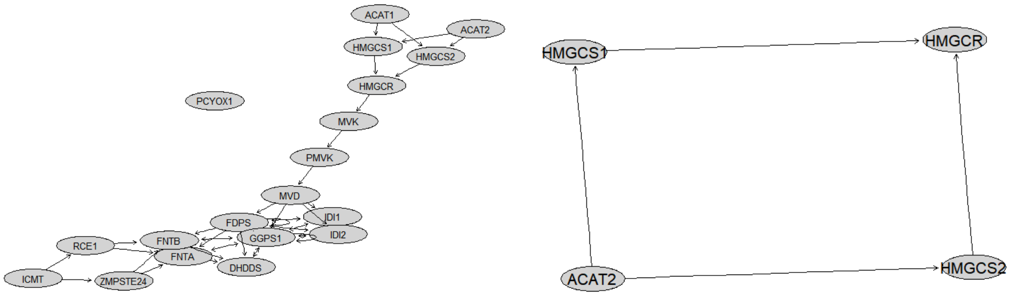

Figure S28: The network connectivity plots of the 19 genes (left) and the top selected genes (right) for pathway *hsa00900*.

Sorting and Degradation

hsa04130 [xml | html | graphics] SNARE interactions in vesicular transport

hsa04140 [xml | html | graphics] Regulation of autophagy

hsa04120 [xml | html | graphics] Ubiquitin mediated proteolysis

SNARE interactions in vesicular transport(hsa04130)

There are total 34 genes in this pathway, among which 25 were mapped to the expression data. So we focused the analysis on the 25 genes. The prediction errors for the three methods are 0.5385, 0.5473 and 0.5298, respectively. The relatively higher prediction error may be explained by the low connectivity between genes (see Fig. S29).

Table S29: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04130*.

Gene Symbol	IVGC	IV	EN
YKT6	0.61	0.55	0.81
STX11	0.54	0.49	0.77
BET1	0.52	0.52	0.82
STX1A	0.49	0.4	0.68
STX10	0.39	0.36	0.62

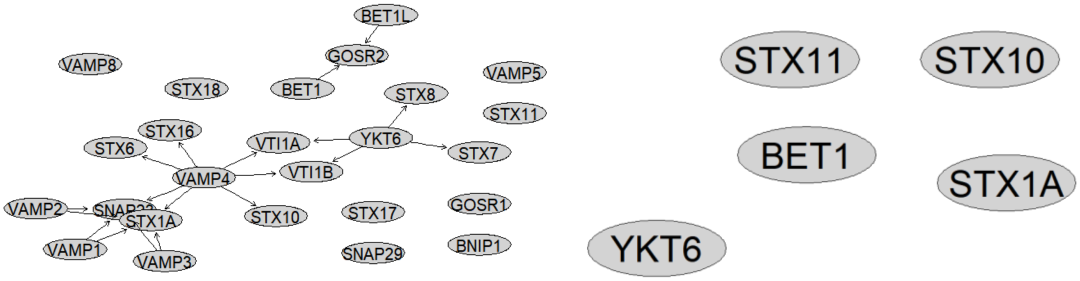

Figure S29: The network connectivity plots of the 25 genes (left) and the top selected genes (right) for pathway *hsa04130*.

Regulation of autophagy(hsa04140)

There are total 128 genes in this pathway, among which 107 were mapped to the expression data. So we focused the analysis on the 107 genes. The prediction errors for the three methods are 0.5548, 0.5420 and 0.5546, respectively. Again, the relatively large prediction error may be explained by the low connectivity between top contributing genes (see Fig. S30).

Table S30: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04140*

Gene Symbol	IVGC	IV	EN
AKT1S1	0.45	0.3	0.65
MTMR4	0.43	0.25	0.64
RAB33B	0.39	0.26	0.62

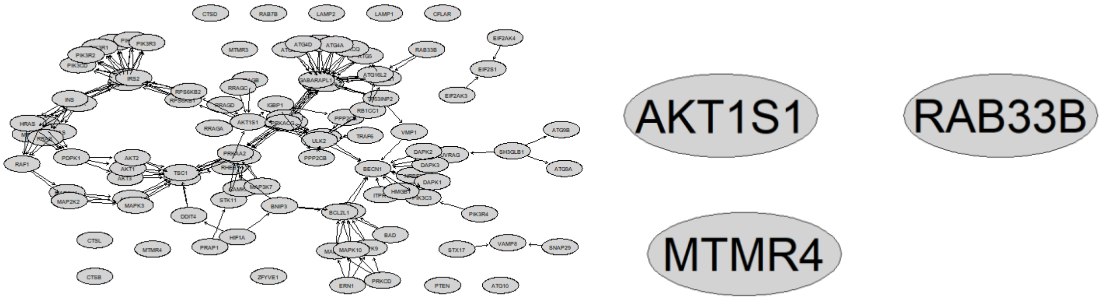

Figure S30: The network connectivity plots of the 107 genes (left) and the top selected genes (right) for pathway *hsa04140*.

Ubiquitin mediated proteolysis(hsa04120)

There are total 137 genes in this pathway, among which 112 were mapped to the expression data. So we focused the analysis on the 112 genes. The prediction errors for the three methods are 0.5343, 0.5409 and 0.5577, respectively. Only EN has genes with selection rate larger than 60%. The prediction error of IVGC and IV is quite similar. Since there is no direct connection between genes (Fig. S31), IVGC and IV should perform similar.

Table S31: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04120*.

Gene Symbol	IVGC	IV	EN
ITCH	0.5	0.56	0.83
CBL	0.33	0.39	0.63

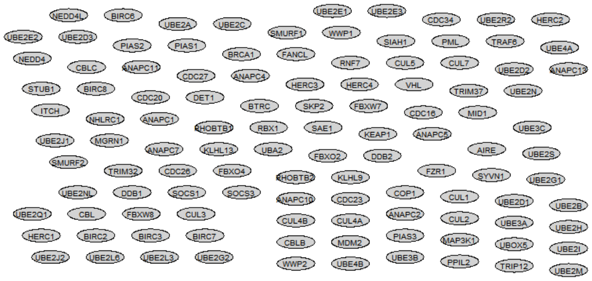

Figure S31: The network connectivity plots of the 112 genes in pathway *hsa04120.*

Signal Transduction

hsa04010 [xml | html | graphics]   MAPK signaling pathway
hsa04310 [xml | html | graphics]   Wnt signaling pathway
hsa04350 [xml | html | graphics]   TGF-beta signaling pathway
hsa04370 [xml | html | graphics]   VEGF signaling pathway
hsa04630 [xml | html | graphics]   Jak-STAT signaling pathway

MAPK signaling pathway (hsa04010)

There are total 265 genes in this pathway, among which 251 were mapped to the expression data. So we focused the analysis on the 251 genes. The prediction errors of the three methods are 0.5563, 0.5562 and 0.5272, respectively. IVGC has relatively larger prediction error than other two methods, due to loose connectivity between top contributing genes (see Fig. S32).

Table S32: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04010*.

Gene Symbol	IVGC	IV	EN
FLNA	0.65	0.44	0.86
TGFB2	0.6	0.44	0.72
CACNA1D	0.58	0.37	0.81
FGFR2	0.56	0.39	0.7
ACVR1C	0.52	0.35	0.72
FOS	0.49	0.37	0.75
FGFR4	0.48	0.31	0.65

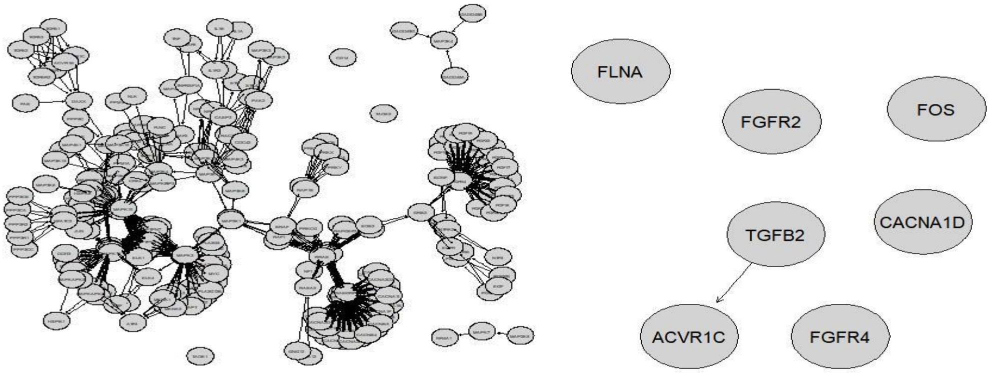

Figure S32: The network connectivity plots of the 251 genes (left) and the top selected genes (right) for pathway *hsa04140*.

TGF-beta signaling pathway (hsa04350)

There are total 89 genes in this pathway, among which 85 were mapped to the expression data. So we focused the analysis on the 85 genes. The prediction errors of the three methods are 0.5403, 0.5228 and 0.5403, respectively. For this pathway, IV has smaller prediction error, but has quite low selection rate for genes listed in Table S33.

Table S33: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04350*.

Gene Symbol	IVGC	IV	EN
ID4	0.94	0.86	1.00
ACVR1C	0.83	0.58	0.88
THBS1	0.78	0.56	0.71
ID3	0.78	0.50	0.82
TGFB2	0.68	0.34	0.62
THBS4	0.57	0.34	0.71
SMAD1	0.42	0.34	0.61

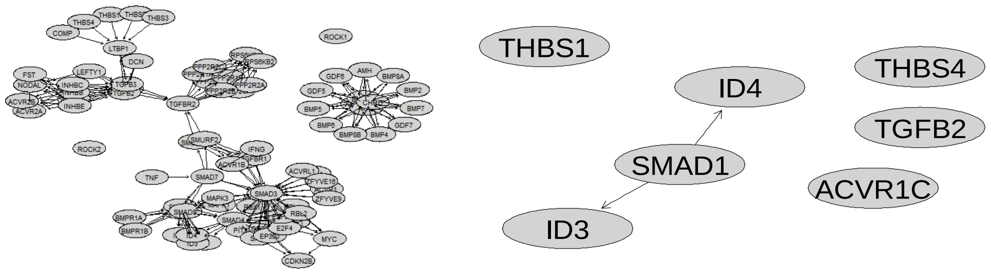

Figure S33: The network connectivity plots of the 85 genes (left) and the top selected genes (right) for pathway *hsa04350*.

Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathway (hsa04630)

There are total 155 genes in this pathway, among which 146 were mapped to the expression data. So we focused the analysis on the 146 genes. The prediction errors of the three methods are 0.5132, 0.5359 and 0.5193, respectively. For this pathway, the prediction error of IVGC is lower than both IV and EN. A full list of the top selected genes is given in Table S34.

Table S34: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04630*.

Gene Symbol	IVGC	IV	EN
IL6	0.74	0.65	0.91
IL10	0.72	0.54	0.8
PIAS3	0.71	0.53	0.79
CBL	0.7	0.55	0.78
IFNA2	0.69	0.52	0.81
IFNW1	0.62	0.45	0.74
CNTFR	0.61	0.31	0.51
MYC	0.48	0.45	0.68

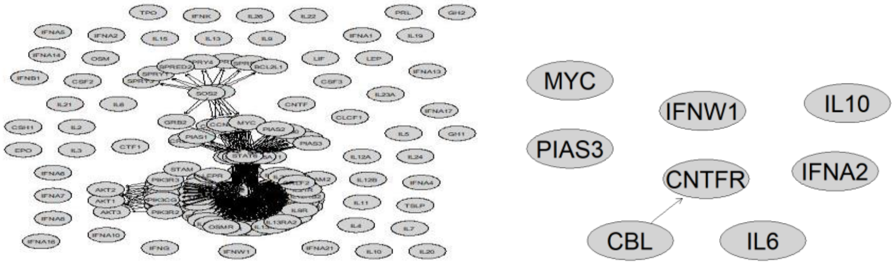

Figure S34: The network connectivity plots of the 146 genes (left) and the top selected genes (right) for pathway *hsa04630*.

Wnt signaling pathway (hsa04310)

There are total 148 genes in this pathway, among which 140 were mapped to the expression data. So we focused the analysis on the 140 genes. The prediction errors are 0.5119, 0.5548 and 0.5924. This pathway clearly demonstrates the advantage of the IVGC compared to IV and EN without incorporating graph information. For example, gene DVL2, DVL1 and DVL3 have zero selection rate with IV and EN, but more than 60% selection rate with IVGC. In checking the connectivity information of the top contributing genes listed in Table S35, these genes do have connections with other top contributing gens such as NKD2, PRICKLE2 and PRICKLE1. Without considering the graph information, these genes will not be selected.

Table S35: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04310*.

Gene Symbol	**IVGC**	**IV**	**EN**
NKD2	0.76	0.38	0.72
PRICKLE2	0.75	0.22	0.59
PPP2R2B	0.74	0.23	0.47
PRICKLE1	0.73	0.18	0.42
TBL1XR1	0.72	0.23	0.54
AXIN1	0.71	0.21	0.44
WNT5A	0.7	0.2	0.48
FZD8	0.68	0.11	0.37
DVL2	0.68	0	0
SENP2	0.67	0.21	0.45
DVL1	0.67	0	0
DVL3	0.67	0	0
FZD4	0.61	0.11	0.3
PPP2R2A	0.61	0.03	0.07
FZD1	0.6	0.02	0.17

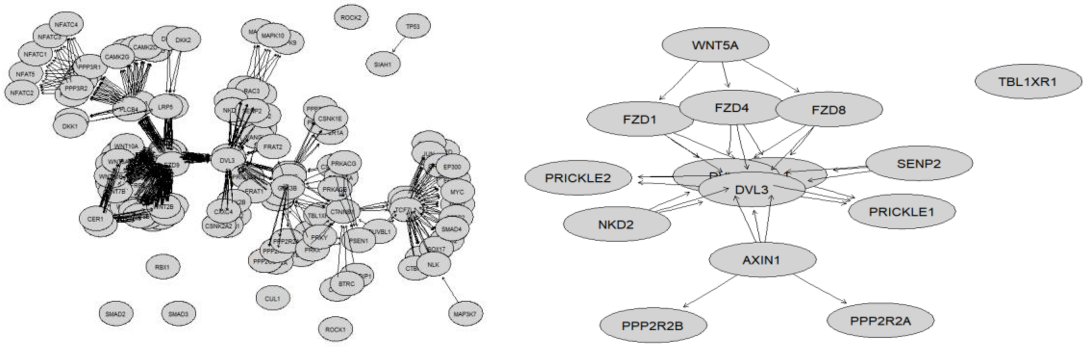

Figure S35: The network connectivity plots of the 140 genes (left) and the top selected genes (right) for pathway *hsa04310*.

VEGF signaling pathway (hsa04370)

There are total 73 genes in this pathway, among which 67 were mapped to the expression data. So we focused the analysis on the 67 genes. The prediction errors for the three methods are 0.5515, 0.5660 and 0.5502. The three top genes shows no direct connection in the pathway. Thus, it is not surprise to see the prediction error for the three methods are quite similar.

Table S36: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04370*.

Gene Symbol	**IVGC**	**IV**	**EN**
SRC	0.65	0.58	0.82
SPHK1	0.50	0.42	0.66
NFATC3	0.43	0.34	0.63

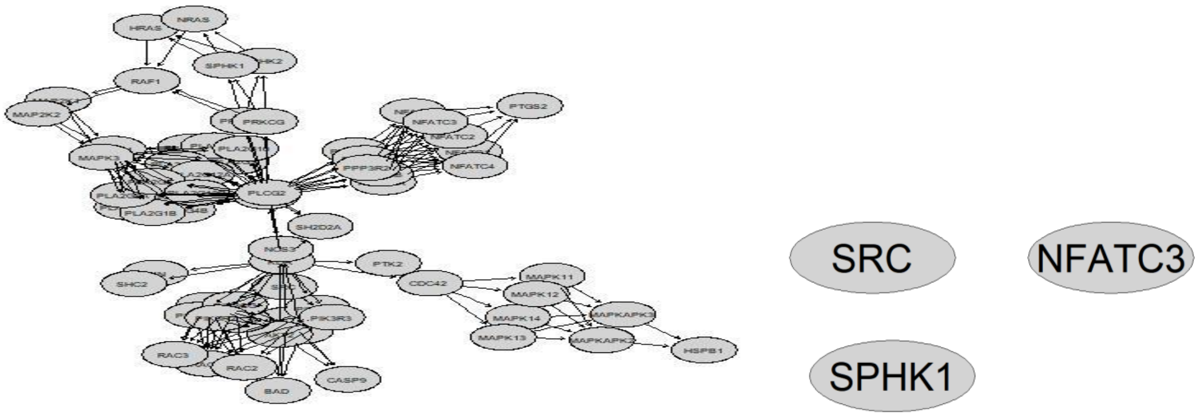

Figure S36: The network connectivity plots of the 67 genes (left) and the top selected genes (right) for pathway *hsa04370*.

Signaling Molecules and Interaction

hsa04080 [xml | html | graphics] Neuroactive ligand-receptor interaction
hsa04060 [xml | html | graphics] Cytokine-cytokine receptor interaction
hsa04512 [xml | html | graphics] ECM-receptor interaction
hsa04514 [xml | html | graphics] Cell adhesion molecules (CAMs)

Neuroactive ligand-receptor interaction(hsa04080)

There are total 277 genes in this pathway, among which 238 were mapped to the expression data. So we focused the analysis on the 238 genes. The prediction errors for the three methods are 0.5107, 0.5098 and 0.5072, respectively. The prediction error difference between the three methods is not significant.

Table S37: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04080*.

Gene Symbol	IVGC	IV	EN
ADRA2C	0.68	0.82	0.93
GRM1	0.55	0.65	0.8
PTGIR	0.51	0.54	0.78
ADRA2A	0.5	0.6	0.77
PTGER2	0.48	0.55	0.78
TBXA2R	0.48	0.52	0.74
FPR1	0.47	0.57	0.76
F2	0.34	0.47	0.66

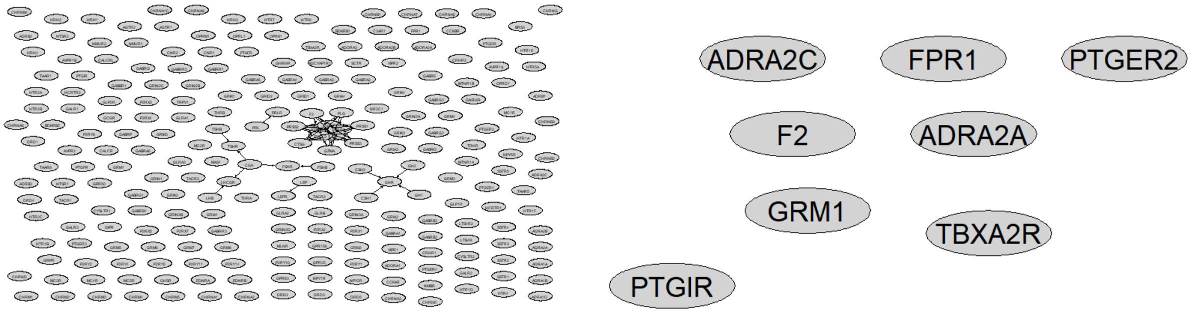

Figure S37: The network connectivity plots of the 238 genes (left) and the top selected genes (right) for pathway *hsa04080*.

Cytokine-cytokine receptor interaction(hsa04060)

There are total 294 genes in this pathway, among which 260 were mapped to the expression data. So we focused the analysis on the 260 genes. The prediction error for the three methods are 0.5366, 0.5358 and 0.5394.

Table S38: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04060*.

Gene Symbol	IVGC	IV	EN
BMP15	0.65	0.62	0.88
IL31RA	0.61	0.57	0.84
ACVR1C	0.55	0.46	0.75
IL10	0.53	0.38	0.7
TGFB2	0.47	0.42	0.59
IFNA2	0.47	0.36	0.65
IL6	0.44	0.35	0.64

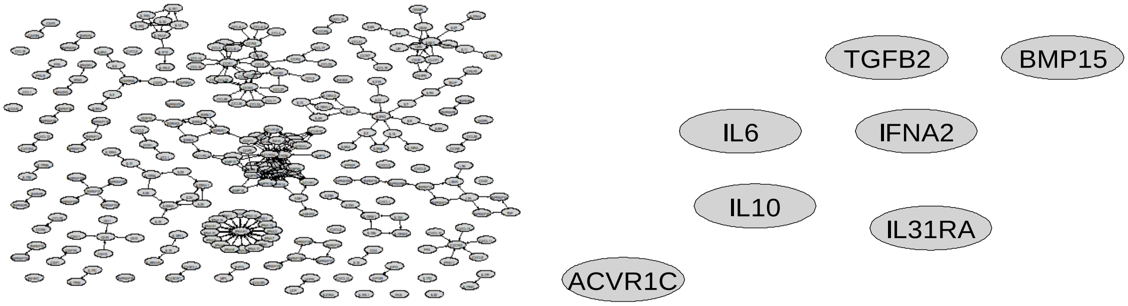

Figure S38: The network connectivity plots of the 260 genes (left) and the top selected genes (right) for pathway *hsa04060*.

ECM-receptor interaction(hsa04512)

There are total 82 genes in this pathway, among which 80 were mapped to the expression data. So we focused the analysis on the 80 genes. The prediction errors for the three methods are 0.5282, 0.5336 and 0.5185.

Table S39: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04512*.

Gene Symbol	IVGC	IV	EN
ITGA4	0.81	0.7	0.96
LAMC3	0.73	0.66	0.92
TNC	0.72	0.64	0.93
ITGA6	0.69	0.53	0.9
ITGB1	0.66	0.37	0.64
COL9A3	0.65	0.55	0.88
DAG1	0.63	0.72	0.99
THBS1	0.63	0.52	0.71
ITGAV	0.62	0.31	0.66
ITGB8	0.58	0.4	0.73
ITGA5	0.52	0.53	0.83
THBS4	0.51	0.47	0.86

COL9A1	0.5	0.44	0.8
VWF	0.49	0.42	0.65
LAMB2	0.47	0.4	0.74
COL4A3	0.44	0.38	0.7

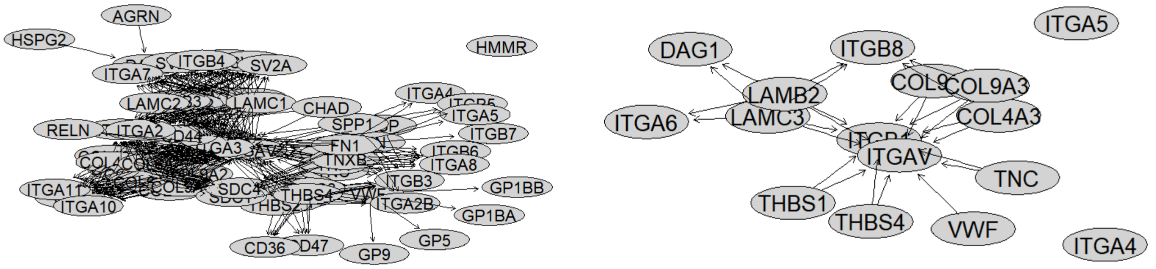

Figure S39: The network connectivity plots of the 80 genes (left) and the top selected genes (right) for pathway *hsa04512*.

Cell adhesion molecules (CAMs)(hsa04514)

There are total 144 genes in this pathway, among which 129 were mapped to the expression data. So we focused the analysis on the 129 genes. The prediction errors for the three methods are 0.5460, 0.5381 and 0.5307.

Table S40: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04514*.

Gene Symbol	IVGC	IV	EN
CLDN4	0.9	0.86	0.97
NEO1	0.8	0.75	0.94
ITGA4	0.66	0.64	0.88
NRXN2	0.52	0.53	0.82
CLDN8	0.52	0.49	0.76
CLDN10	0.45	0.4	0.63
CLDN2	0.42	0.34	0.63

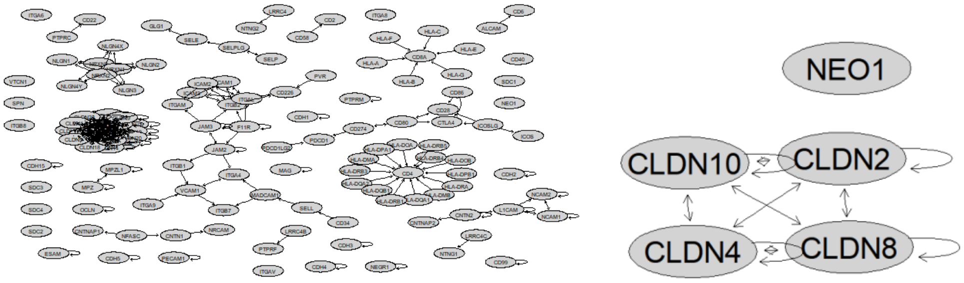

Figure S40: The network connectivity plots of the 129 genes (left) and the top selected genes (right) for pathway *hsa04514*.

Cell Growth and Death

hsa04110 [xml | html | graphics] Cell cycle
hsa04210 [xml | html | graphics] Apoptosis

Cell cycle (hsa04110)

There are total 115 genes in this pathway, among which 101 were mapped to the expression data. So we focused the analysis on the 101 genes. The prediction errors of the three methods are 0.5344, 0.5502 and 0.5334, respectively. After the stability selection, there is only one gene with selection rate >60%. The selection results indicate that this pathway does no play an important role in the activity of CYP2E1. Both IVGC and EN picked up gene TGFB2 with a selection rate >60%.

Table S41: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04110*.

Gene Symbol	IVGC	IV	EN
TGFB2	0.62	0.46	0.75

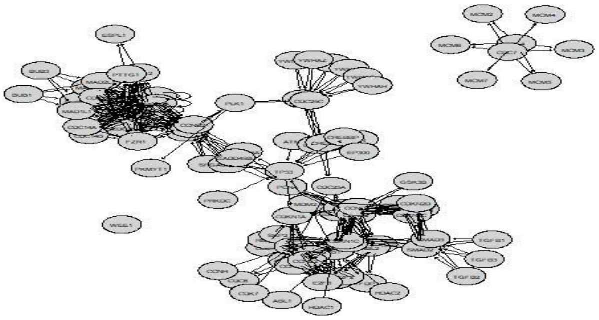

Figure S41: The network connectivity plots of the 101 genes in pathway *hsa04110*.

Apoptosis (hsa04210)

There are total 85 genes in this pathway, among which 78 were mapped to the expression data. So we focused the analysis on the 78 genes. The prediction errors are 0.5133, 0.5155 and 0.5194. Although IVGC has smaller prediction error, but the difference among the three methods is not remarkable. The selection rate for all top genes between IVGC and EN is quite similar.

Table S42: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04210*.

Gene Symbol	IVGC	IV	EN
ITGA4	0.88	0.74	0.93
PPP1CB	0.88	0.63	0.86
SLC9A1	0.87	0.8	0.95
FGFR2	0.87	0.67	0.89
MYL9	0.86	0.64	0.87
FGFR4	0.85	0.72	0.91
SRC	0.85	0.68	0.87
PPP1R12A	0.78	0.63	0.77
F2	0.76	0.62	0.81
WASF1	0.75	0.37	0.73
ITGA5	0.72	0.37	0.66
PFN2	0.69	0.47	0.64
ROCK2	0.67	0.22	0.34
BDKRB2	0.64	0.43	0.63
PDGFRA	0.64	0.35	0.55
CYFIP2	0.63	0.31	0.65
FGF3	0.58	0.34	0.6

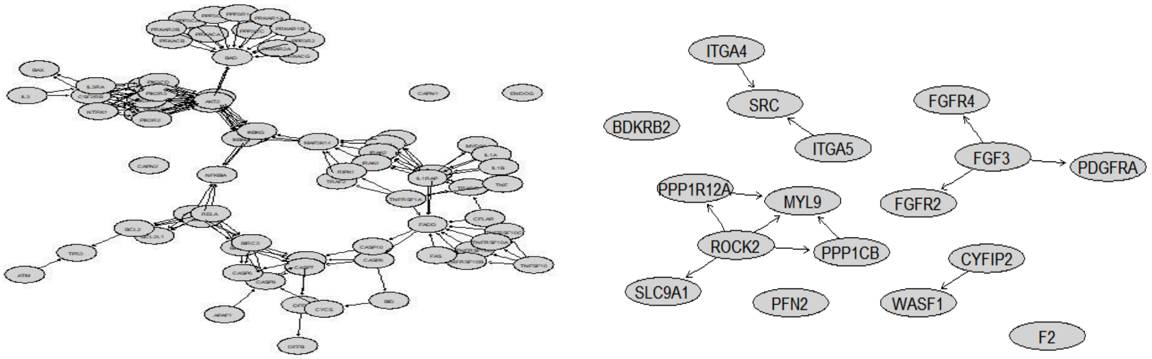

Figure S42: The network connectivity plots of the 78 genes (left) and the top selected genes (right) for pathway *hsa04210*.

Cell Communication

hsa04510 [xml | html | graphics] Focal adhesion

hsa04520 [xml | html | graphics] Adherens junction

hsa04540 [xml | html | graphics] Gap junction

Focal adhesion(hsa04510)

There are total 199 genes in this pathway, among which 183 were mapped to the expression data. So we focused the analysis on the 183 genes. The prediction errors for the three methods are 0.5336, 0.5383 and 0.5233. The selection rate between IVGC and EN is quite similar.

Table S43: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04510*.

Gene Symbol	IVGC	IV	EN
FLNA	0.85	0.53	0.9
SRC	0.8	0.49	0.84
THBS1	0.78	0.51	0.68
TNC	0.76	0.39	0.73
MYL9	0.72	0.47	0.79
LAMC3	0.72	0.4	0.7
ITGA4	0.71	0.5	0.86
PPP1R12A	0.7	0.36	0.69
ITGA5	0.69	0.29	0.65
PPP1CB	0.67	0.2	0.57
COL9A3	0.65	0.3	0.61
THBS4	0.64	0.25	0.61

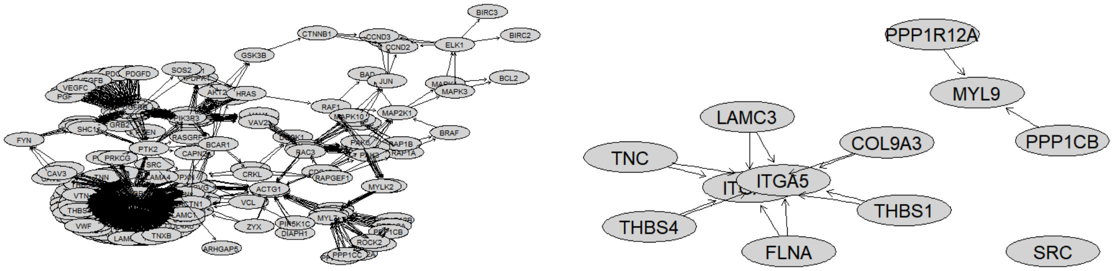

Figure S43: The network connectivity plots of the 183 genes (left) and the top selected genes (right) for pathway *hsa04510*.

Adherens junction(hsa04520)

There are total 72 genes in this pathway, among which 65 were mapped to the expression data. So we focused the analysis on the 65 genes. The prediction errors for the three methods are 0.4950, 0.4938 and 0.4959. For this pathway, the prediction error among the three methods is quite similar. The selection rate for the top contributing genes between IVGC and IV is also quite similar, which can be explained by the non-connectivity between the two top genes.

Table S44: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04520*.

Gene Symbol	IVGC	IV	EN
TJP1	0.97	0.99	1
SRC	0.84	0.84	0.95

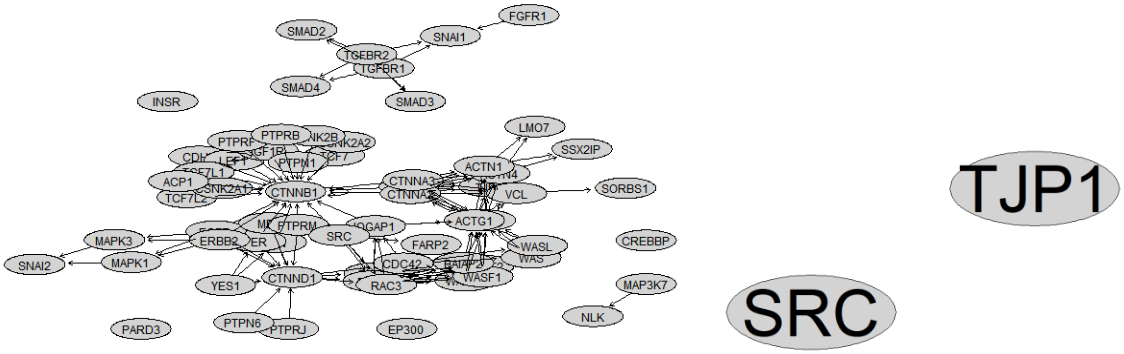

Figure S44: The network connectivity plots of the 65 genes (left) and the top selected genes (right) for pathway *hsa04520*.

Gap junction(hsa04540)

There are total 88 genes in this pathway, among which 70 were mapped to the expression data. So we focused the analysis on the 70 genes. The prediction errors for the three methods are 0.5034, 0.5059 and 0.5065. The prediction errors for the three methods are quite similar. In viewing the network connectivity of the top contributing genes in Fig. S45, it is not surprise that the three methods perform similar to each other.

Table S45: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04540*.

Gene Symbol	IVGC	IV	EN
TJP1	0.96	0.97	0.99
GRM1	0.9	0.84	0.91
SRC	0.88	0.86	0.94
ITPR3	0.71	0.58	0.7
ITPR2	0.64	0.48	0.63
TUBB2B	0.62	0.51	0.61
GRM5	0.58	0.47	0.63

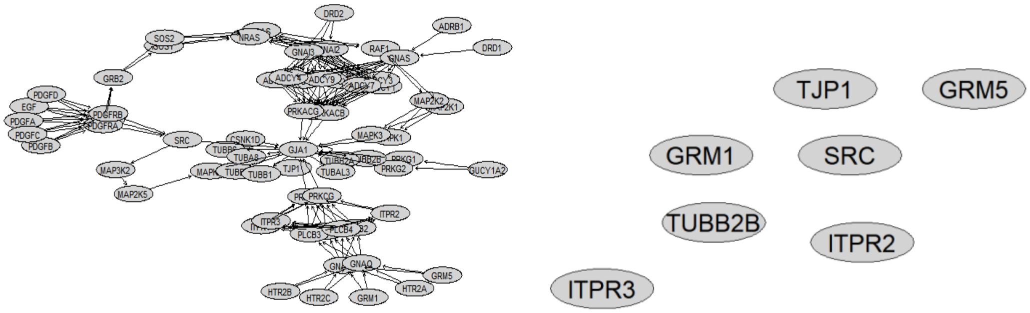

Figure S45: The network connectivity plots of the 70 genes (left) and the top selected genes (right) for pathway *hsa04540*.

Endocrine System

hsa04910 [xml | html | graphics] Insulin signaling pathway
hsa04920 [xml | html | graphics] Adipocytokine signaling pathway
hsa03320 [xml | html | graphics] PPAR signaling pathway
hsa04912 [xml | html | graphics] GnRH signaling pathway

Insulin signaling pathway(hsa04910)

There are total 137 genes in this pathway, among which 128 were mapped to the expression data. So we focused the analysis on the 128 genes. The prediction errors are 0.5324, 0.5343 and 0.5224 for the three methods. The selection rates for most genes between IVGC and IV are quite similar and are systematically lower than EN. This may explain the relatively larger prediction error of the two methods compared to EN.

Table S46: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04910*.

Gene Symbol	IVGC	IV	EN
PHKA2	0.88	0.89	0.98
PYGL	0.82	0.83	0.95
PRKCZ	0.54	0.52	0.77
CBL	0.53	0.57	0.79
PYGB	0.52	0.59	0.78
PPP1CB	0.5	0.51	0.75
CALM2	0.49	0.49	0.78
EIF4E2	0.42	0.49	0.68

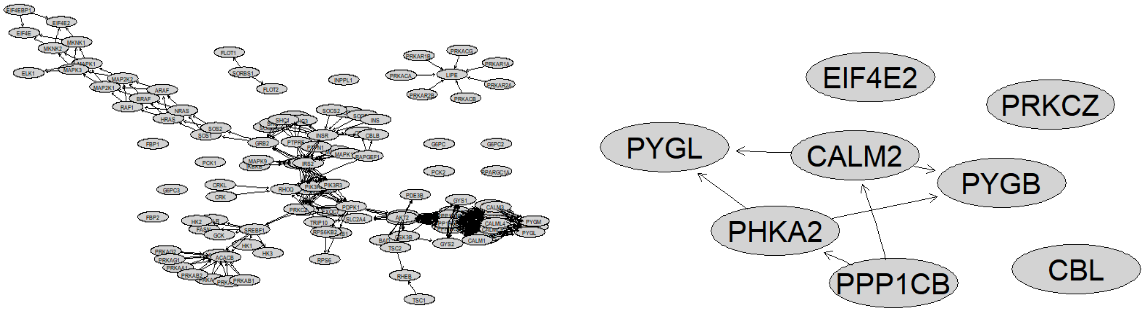

Figure S46: The network connectivity plots of the 128 genes (left) and the top selected genes (right) for pathway *hsa04910*.

Adipocytokine signaling pathway(hsa04920)

There are total 69 genes in this pathway, among which 67 were mapped to the expression data. So we focused the analysis on the 67 genes. The prediction errors are 0.5279, 0.5172 and 0.5207 for the three methods.

Table S47: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04920*.

Gene Symbol	IVGC	IV	EN
PPARA	0.63	0.66	0.87
CPT1A	0.57	0.62	0.8

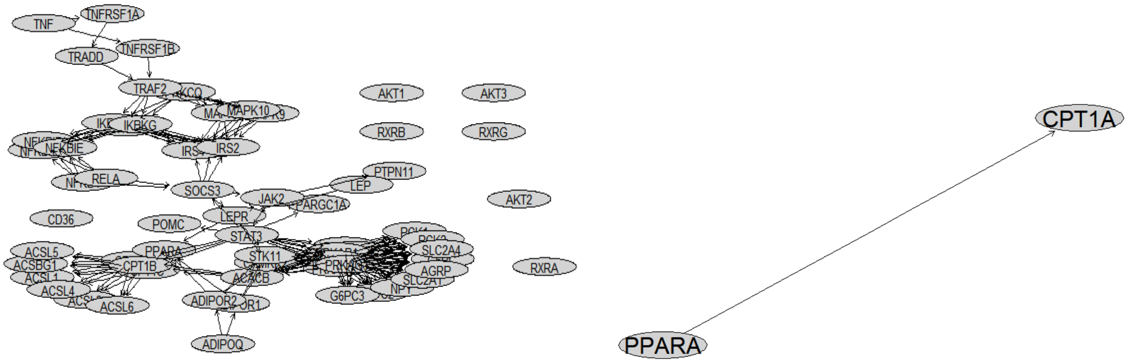

Figure S47: The network connectivity plots of the 67 genes (left) and the top selected genes (right) for pathway *hsa04920*.

PPAR signaling pathway(hsa03320)

There are total 74 genes in this pathway, among which 68 were mapped to the expression data. So we focused the analysis on the 68 genes. The prediction errors are 0.4812, 0.4898 and 0.4820 for IVGC, IV and EN, respectively. From Table S48, we can see that IVGC has higher selection rate for the top contributing genes compared to IV and EN.

Table S48: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa03320*.

Gene Symbol	IVGC	IV	EN
DBI	0.99	0.97	0.98
FABP1	0.99	0.94	0.97
PPARA	0.98	0.95	0.99
CPT1A	0.98	0.86	0.93
MMP1	0.9	0.83	0.89
SLC27A5	0.88	0.79	0.89
APOA5	0.78	0.24	0.43
FABP6	0.72	0.34	0.57
APOA2	0.71	0.37	0.58
FADS2	0.7	0.17	0.46
CYP8B1	0.69	0.23	0.44
CYP27A1	0.66	0.27	0.46
RXRG	0.57	0.45	0.68

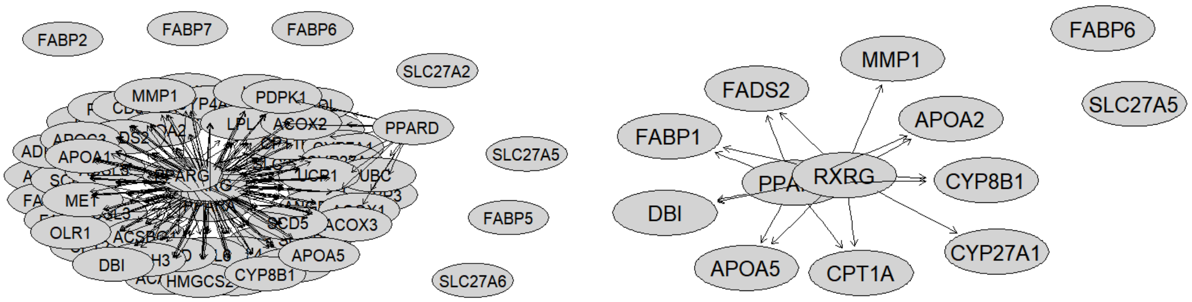

Figure S48: The network connectivity plots of the 68 genes (left) and the top selected genes (right) for pathway *hsa03320*.

GnRH signaling pathway(hsa04912)

There are total 93 genes in this pathway, among which 89 were mapped to the expression data. So we focused the analysis on the 89 genes. The prediction errors are 0.5071, 0.5547 and 0.5321 for IVGC, IV and EN, respectively. Of particular interest is the gene HBEGF which has a selection rate of 0.68 with IVGC, compared to 0.03 and 0.14 with IV and EN. In checking the network connectivity information in Fig. S49, we can see that gene HBEGF is connected with gene MMP2 and MMP14. Both genes have a high selection rate with EN.

Table S49: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04912*.

Gene Symbol	IVGC	IV	EN
MMP14	0.83	0.43	0.78
CACNA1D	0.82	0.42	0.8
ITPR2	0.8	0.34	0.74
MAP2K7	0.8	0.33	0.75
CALM2	0.8	0.32	0.75
MMP2	0.79	0.32	0.62
ADCY3	0.78	0.32	0.63
SRC	0.78	0.29	0.78
ITPR3	0.77	0.37	0.71
MAP3K4	0.7	0.34	0.71
JUN	0.69	0.22	0.57
HBEGF	0.68	0.03	0.14
FSHB	0.67	0.24	0.6
MAPK14	0.62	0.11	0.39

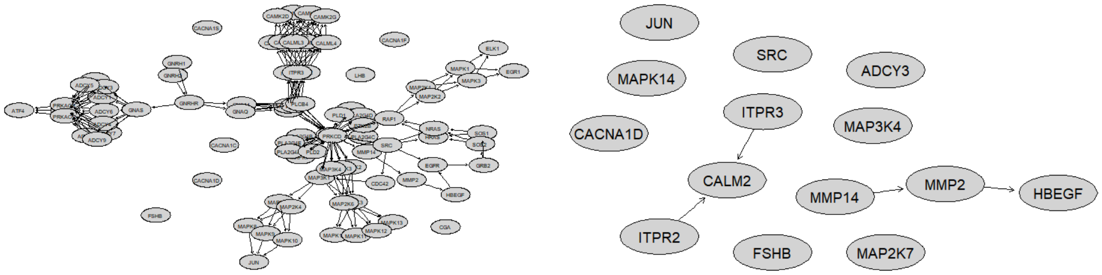

Figure S49: The network connectivity plots of the 89 genes (left) and the top selected genes (right) for pathway *hsa04912*.

Immune System

hsa04610 [xml | html | graphics]   Complement and coagulation cascades
hsa04620 [xml | html | graphics]   Toll-like receptor signaling pathway
hsa04650 [xml | html | graphics]   Natural killer cell mediated cytotoxicity
hsa04612 [xml | html | graphics]   Antigen processing and presentation
hsa04660 [xml | html | graphics]   T cell receptor signaling pathway

Complement and coagulation cascades (hsa04610)

There are total 79 genes in this pathway, among which 71 were mapped to the expression data. So we focused the analysis on the 71 genes. The prediction errors are 0.5427, 0.5592 and 0.5558 for IVGC, IV and EN, respectively.

Table S50: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04610*.

Gene Symbol	IVGC	IV	EN
C6	0.84	0.72	0.92
VWF	0.54	0.37	0.64
C9	0.49	0.39	0.67
F7	0.45	0.45	0.72

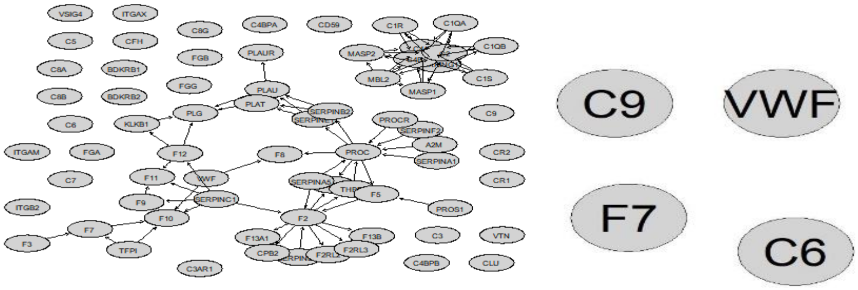

Figure S50: The network connectivity plots of the 71 genes (left) and the top selected genes (right) for pathway *hsa04610*.

Toll-like receptor signaling pathway(hsa04620)

There are total 104 genes in this pathway, among which 100 were mapped to the expression data. So we focused the analysis on the 100 genes. The prediction errors are 0.5482, 0.5468 and 0.5389 for IVGC, IV and EN, respectively. Only one gene passed the 60% threshold with EN.

Table S51: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04620*.

Gene Symbol	IVGC	IV	EN
IL6	0.53	0.46	0.75

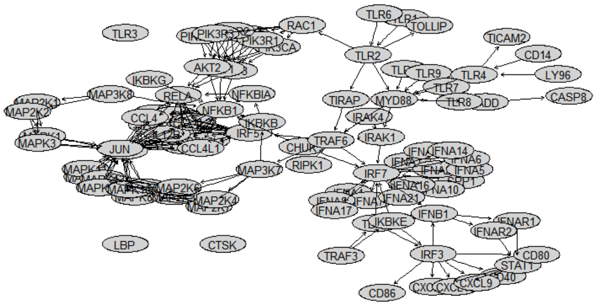

Figure S51: The network connectivity plots of the 100 genes in pathway *hsa04620*.

Natural killer cell mediated cytotoxicity(hsa04650)

There are total 131 genes in this pathway, among which 124 were mapped to the expression data. So we focused the analysis on the 124 genes. After the stability selection, no single gene has a selection rate larger than 60%, indicating the less important role of this pathway in the expression activity of enzyme CYP2E1.

Antigen processing and presentation(hsa04612)

There are total 77 genes in this pathway, among which 74 were mapped to the expression data. So we focused the analysis on the 74 genes. After the stability selection, the selection rates for all genes are less than 50% with all the three methods, indicating the less important role of this pathway in the expression activity of enzyme CYP2E1.

T cell receptor signaling pathway(hsa04660)

There are total 101 genes in this pathway, among which 94 were mapped to the expression data. So we focused the analysis on the 94 genes. The prediction errors are 0.5562, 0.5577 and 0.5786 for IVGC, IV and EN, respectively. Although the prediction errors of IVGC and IV are smaller than EN, the selection rates of all genes with IVGC and IV are all less than 60%. EN did picked up two genes with the selection rate passing the 0.6 threshold.

Table S52: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04660*.

Gene Symbol	IVGC	IV	EN
IL10	0.42	0.38	0.68
FOS	0.41	0.37	0.66

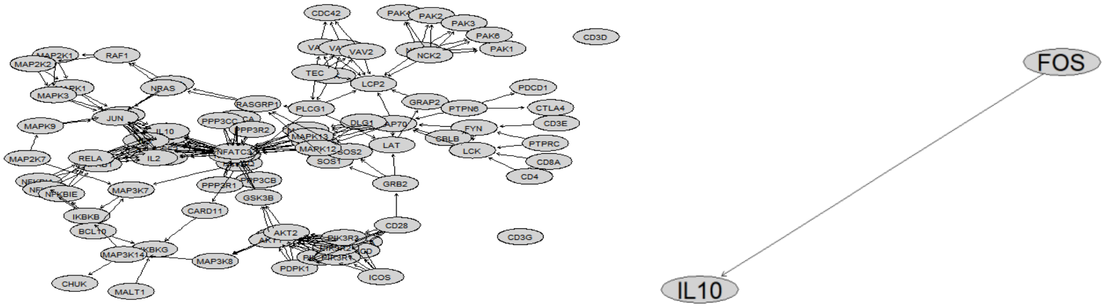

Figure S52: The network connectivity plots of the 94 genes (left) and the top selected genes (right) for pathway *hsa04660*.

Nervous System

hsa04720 [xml | html | graphics]   Long-term potentiation
hsa04730 [xml | html | graphics]   Long-term depression

Long-term potentiation(hsa04720)

There are total 67 genes in this pathway, among which 64 were mapped to the expression data. So we focused the analysis on the 64 genes. The prediction errors are 0.5441, 0.5487 and 0.5471 for IVGC, IV and EN, respectively. IV has overall much lower selection rate for all genes than IVGC or EN does. The network connectivity for the top contributing genes shown in Fig. S53 indicates that gene CALM2 has a few connections with a few other genes. From Table S53, we can tell that those genes having a connection with CALM2 have relatively high selection rates with IVGC than those with EN. This pathway analysis clearly shows the advantage of the IVGC.

Table S53: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04720*.

Gene Symbol	IVGC	IV	EN
GRM1	0.87	0.64	0.92
PPP1CB	0.87	0.61	0.9
ITPR3	0.86	0.62	0.88
ITPR2	0.82	0.53	0.8
CALM2	0.78	0.5	0.81
GRM5	0.72	0.38	0.72
PPP3CC	0.71	0.34	0.65
ITPR1	0.71	0.24	0.6
PPP3R1	0.66	0.26	0.55
RPS6KA1	0.66	0.22	0.6
GRIA2	0.64	0.32	0.64
GRIN2A	0.64	0.23	0.58
PPP3CB	0.62	0.18	0.49
RAP1B	0.62	0.11	0.41
PPP1CA	0.61	0.17	0.56
PPP3R2	0.6	0.17	0.5
NRAS	0.6	0.15	0.51

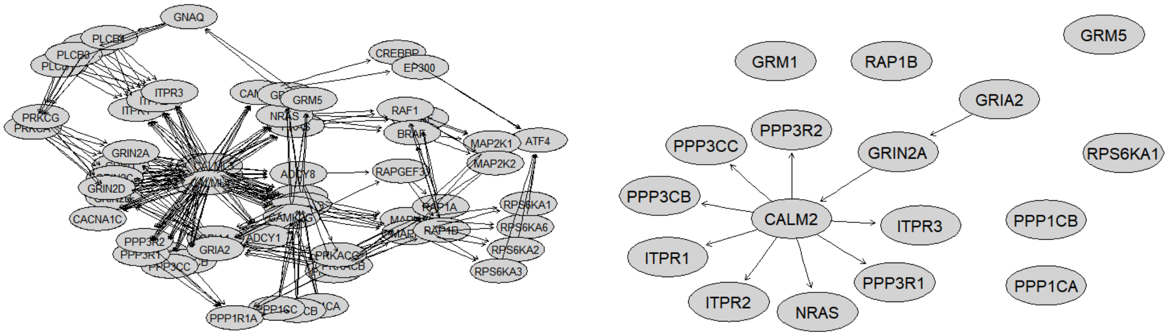

Figure S53: The network connectivity plots of the 64 genes (left) and the top selected genes (right) for pathway *hsa04720*.

Long-term depression(hsa04730)

There are total 60 genes in this pathway, among which 54 were mapped to the expression data. So we focused the analysis on the 54 genes. The prediction errors are 0.5347, 0.5311 and 0.5300 for

IVGC, IV and EN, respectively. For this pathway, EN has higher selection rates for most genes than IVGC or IV does.

Table S54: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04730*.

Gene Symbol	IVGC	IV	EN
GRM1	0.67	0.57	0.81
ITPR3	0.64	0.55	0.74
ITPR2	0.52	0.43	0.61

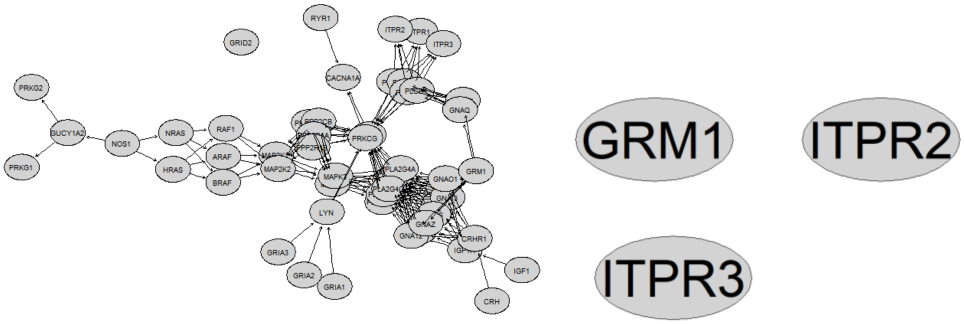

Figure S54: The network connectivity plots of the 54 genes (left) and the top selected genes (right) for pathway *hsa04730*.

Sensory System

hsa04740 [xml | html | graphics]   Olfactory transduction

hsa04742 [xml | html | graphics]   Taste transduction

Olfactory transduction(hsa04740)

There are total 448 genes in this pathway, among which 364 were mapped to the expression data. So we focused the analysis on the 364 genes. The prediction errors are 0.5265, 0.5045 and 0.5177 for IVGC, IV and EN, respectively. For this pathway, IV has smaller prediction error than IVGC or EN does. Although gene *ARRB2* is a hub gene in the network, it may not be significantly associated with CYP2E1 enzyme activity. Thus, the connectivity information or the correlation information of this gene with other genes may inflate the selection rates of genes related with it. But still IVGC has smaller selection rate (0.58) for *ARRB2* than EN does (0.65).

Table S55: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04740.*

Gene Symbol	IVGC	IV	EN
OR10G4	0.92	0.79	0.95
OR8S1	0.9	0.64	0.96
OR6B1	0.87	0.6	0.92
OR13A1	0.86	0.67	0.91
OR52K1	0.86	0.64	0.93
OR10G7	0.85	0.59	0.87
ADCY3	0.79	0.51	0.84
OR52E5	0.79	0.42	0.77
CNGA2	0.78	0.55	0.81
OR6C2	0.75	0.37	0.7
OR2T12	0.74	0.43	0.77

OR4C6	0.74	0.4	0.77
CALM2	0.73	0.45	0.79
OR5AR1	0.68	0.35	0.71
OR4F15	0.67	0.33	0.68
OR11H4	0.62	0.39	0.71
OR5K3	0.61	0.29	0.64
ARRB2	0.58	0.29	0.65

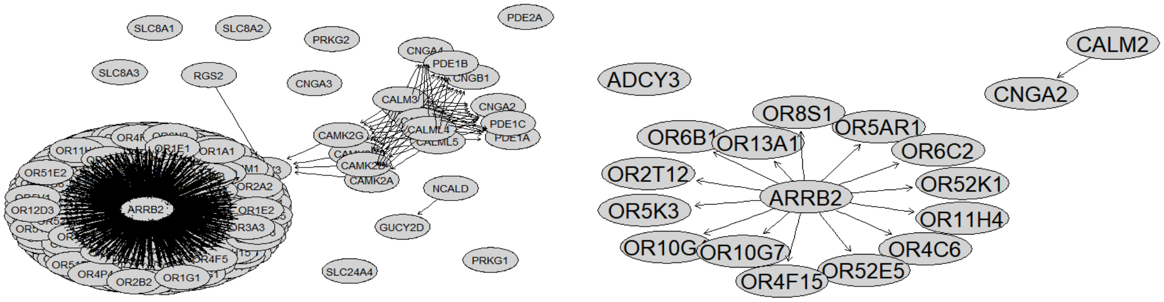

Figure S55: The network connectivity plots of the 364 genes (left) and the top selected genes (right) for pathway *hsa04740*.

Taste transduction(hsa04742)

There are total 83 genes in this pathway, among which 74 were mapped to the expression data. So we focused the analysis on the 74 genes. The prediction errors are 0.5242, 0.5379 and 0.5298 for IVGC, IV and EN, respectively. The overall network connectivity of genes in this pathway is not very strong. Still, IVGC has the smallest prediction error among the three. It picked up two genes *GRM1* and *ITPR3* as IV did, while EN picked up two more genes. The extra two more genes by EN may be due to actual correlation of the expression measures between the two genes. Since the four genes have no direct connection in the network, the selection results between IVGC and IV are quite similar to each other. This again shows the robustness of the IVGC method.

Table S56: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04742*.

Gene Symbol	IVGC	IV	EN
GRM1	0.73	0.74	0.89
ITPR3	0.7	0.72	0.85
GNB3	0.36	0.42	0.61
SCNN1G	0.33	0.41	0.65

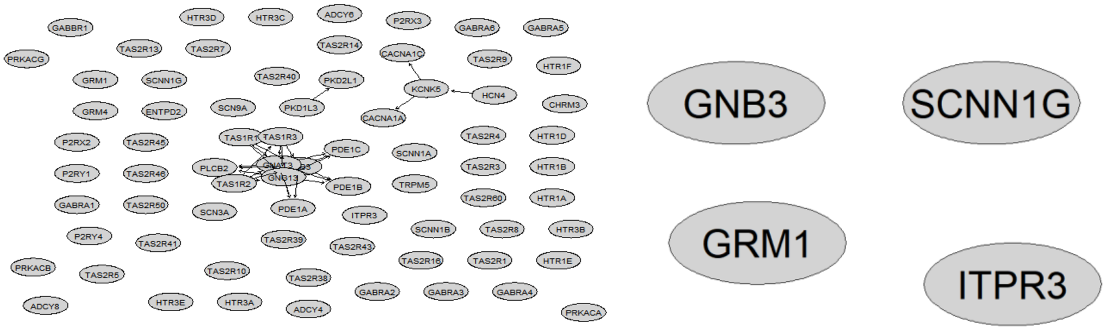

Figure S56: The network connectivity plots of the 74 genes (left) and the top selected genes (right) for pathway *hsa04742*.

Development

hsa04360 [xml | html | graphics]   Axon guidance

There are total 175 genes in this pathway, among which 163 were mapped to the expression data. So we focused the analysis on the 163 genes. There is no single gene with a selection rate larger than 0.6 by all three methods, indicating the less important role of this pathway in expressing CYP2E1 enzyme activity.

Behavior

hsa04710 [xml | html | graphics]   Circadian rhythm

There are total 31 genes in this pathway, among which 26 were mapped to the expression data. So we focused the analysis on the 26 genes. The prediction errors are 0.5351, 0.5408 and 0.5443 for IVGC, IV and EN, respectively. However, no genes have selection rate larger than 0.6 by IVGC and IV. Although EN has one gene (*RORA*) with a selection rate larger than 0.6, there is no literature report shown that this gene is associated with CYP2E1 enzyme activity. Thus, this gene could be potentially a false positive due to pseudo correlation between gene expressions.

Table S57: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa04710*.

Gene Symbol	**IVGC**	**IV**	**EN**
RORA	0.5	0.5	0.72

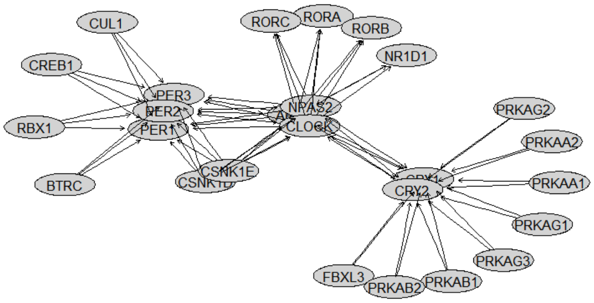

Figure S57: The network connectivity plots of the 26 genes in pathway *hsa04710*.

Infectious Diseases

hsa05120 [xml | html | graphics]   Epithelial cell signaling in Helicobacter pylori infection

There are total 68 genes in this pathway, among which 61 were mapped to the expression data. So we focused the analysis on the 61 genes. The prediction errors are 0.5035, 0.4984 and 0.4996 for IVGC, IV and EN, respectively. Two genes have a selection rate larger than 0.6 and the selection rate between IVGC and IV is quite similar, due to loose connectivity of top contributing genes in this pathway.

Table S58: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa05120*.

Gene Symbol	**IVGC**	**IV**	**EN**
TJP1	0.95	0.97	0.99
SRC	0.81	0.82	0.95

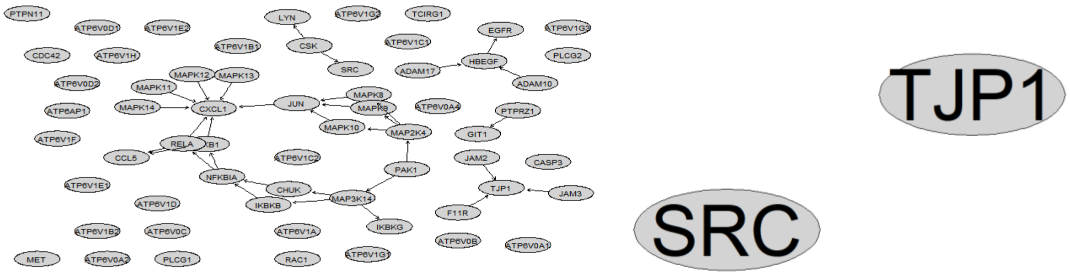

Figure S58: The network connectivity plots of the 61 genes (left) and the top selected genes (right) for pathway *hsa05120*.

Neurodegenerative Disorders

hsa05010 [xml | html | graphics]   Alzheimer's disease

hsa05020 [xml | html | graphics]   Parkinson's disease

hsa05030 [xml | html | graphics]   Amyotrophic lateral sclerosis (ALS)

Alzheimer's disease(hsa05010)

There are total 171 genes in this pathway, among which 140 were mapped to the expression data. So we focused the analysis on the 140 genes. The prediction errors are 0.5342, 0.5735 and 0.5341 for IVGC, IV and EN, respectively. For this pathway, the prediction error between IVGC and EN is quite similar. However, the selection rates of all top contributing genes between IVGC and IV are quite similar, due to loose connectivity of genes in this pathway. EN may over select genes in this pathway than IVGC or IV does.

Table S59: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa05010*.

Gene Symbol	IVGC	IV	EN
BACE2	0.81	0.8	0.96
CACNA1D	0.55	0.5	0.82
ATP2A3	0.5	0.33	0.74
CALM2	0.38	0.26	0.7
NDUFS2	0.35	0.28	0.71
NDUFS6	0.35	0.28	0.68
IL1B	0.32	0.26	0.65
ADAM10	0.29	0.24	0.7
NDUFB7	0.26	0.14	0.62

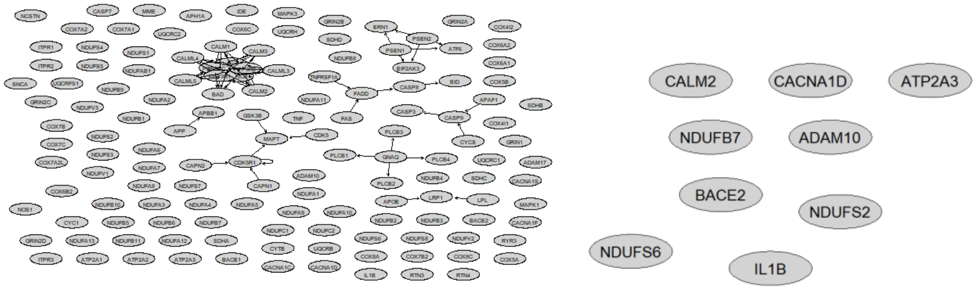

Figure S59: The network connectivity plots of the 140 genes (left) and the top selected genes (right) for pathway *hsa05010*.

Parkinson's disease(hsa05020)

There are total 35 genes in this pathway, among which 33 were mapped to the expression data. So we focused the analysis on the 33 genes. The prediction errors are 0.5144, 0.5242 and 0.5063 for IVGC, IV and EN, respectively.

Table S60: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa05020*.

Gene Symbol	IVGC	IV	EN
C6	0.96	0.99	1
IL6	0.86	0.78	0.94
LAMC1	0.85	0.75	0.84
C9	0.79	0.71	0.92
C8B	0.73	0.61	0.85
C7	0.66	0.46	0.73
SOD1	0.63	0.91	0.99
PRNP	0.6	0.41	0.64
IL1A	0.45	0.35	0.69

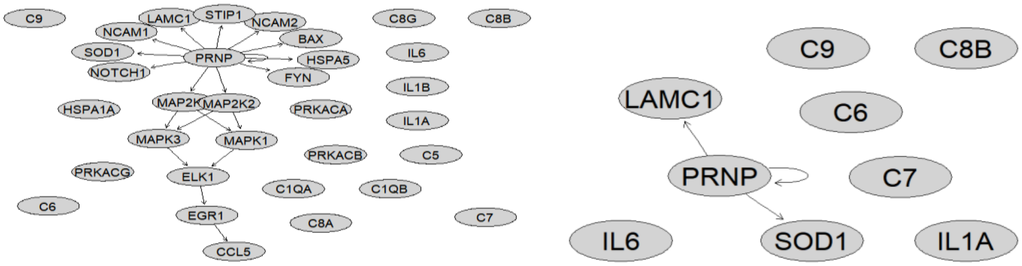

Figure S60: The network connectivity plots of the 33 genes (left) and the top selected genes (right) for pathway *hsa05020*.

Amyotrophic lateral sclerosis (ALS)(hsa05030)

There are total 49 genes in this pathway, among which 48 were mapped to the expression data. So we focused the analysis on the 48 genes. The prediction errors are 0.5343, 0.5752 and 0.5455 for IVGC, IV and EN, respectively. Although the selection rates of the top contribution genes in this pathway are higher with EN than both IVGC and IV, IVGC does have smallest prediction error among the three.

Table S61: List of top selected genes with a stability selection rate >60% by using any one of the three methods (IVGC, IV and EN) for pathway *hsa05030*.

Gene Symbol	IVGC	IV	EN
DDC	0.75	0.58	0.91
CREB3L1	0.75	0.55	0.92
MAOB	0.69	0.42	0.86
GRIN2D	0.66	0.4	0.76
JUN	0.65	0.38	0.76
CREB1	0.53	0.28	0.64
RELA	0.37	0.27	0.61

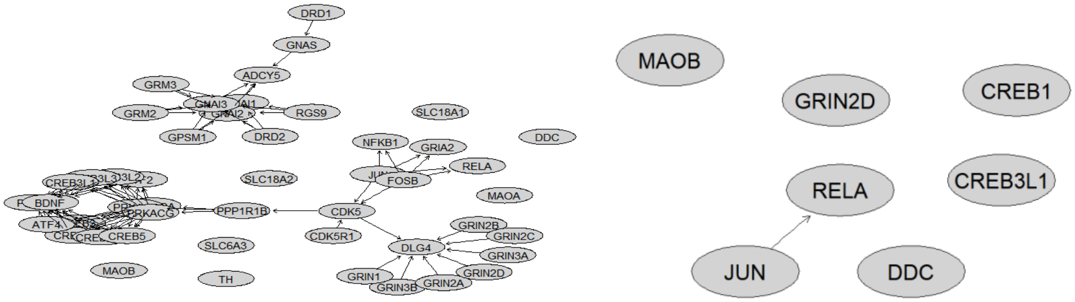

Figure S61: The network connectivity plots of the 48 genes (left) and the top selected genes (right) for pathway *hsa05030*.