Original Article

Identification of metastasis-associated genes in colorectal cancer through an integrated genomic and transcriptomic analysis

Xiaobo Li[1,2], Sihua Peng[3]

[1]Department of Computer Science and Technology, College of Engineering, Lishui University, Lishui 323000, China; [2]School of Science and Technology, Zhejiang International Studies University, Hangzhou 310012, China; [3]Department of Biological Technology, School of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China

Corresponding to: Xiaobo Li. Department of Computer Science and Technology, College of Engineering, Lishui University, Lishui 323000, China. Email: oboaixil@126.com.

Objective: Identification of colorectal cancer (CRC) metastasis genes is one of the most important issues in CRC research. For the purpose of mining CRC metastasis-associated genes, an integrated analysis of microarray data was presented, by combined with evidence acquired from comparative genomic hybridization (CGH) data.

Methods: Gene expression profile data of CRC samples were obtained at Gene Expression Omnibus (GEO) website. The 15 important chromosomal aberration sites detected by using CGH technology were used for integrated genomic and transcriptomic analysis. Significant Analysis of Microarray (SAM) was used to detect significantly differentially expressed genes across the whole genome. The overlapping genes were selected in their corresponding chromosomal aberration regions, and analyzed by using the Database for Annotation, Visualization and Integrated Discovery (DAVID). Finally, SVM-T-RFE gene selection algorithm was applied to identify metastasis-associated genes in CRC.

Results: A minimum gene set was obtained with the minimum number [14] of genes, and the highest classification accuracy (100%) in both PRI and META datasets. A fraction of selected genes are associated with CRC or its metastasis.

Conclusions: Our results demonstrated that integration analysis is an effective strategy for mining cancer-associated genes.

Keywords: Colorectal cancer metastasis; integrated analysis; comparative genomic hybridization (CGH); Significant Analysis of Microarray (SAM); Database for Annotation, Visualization and Integrated Discovery (DAVID); SVM-T-RFE gene selection algorithm

Submitted Aug 07, 2013. Accepted for publication Nov 05, 2013.
doi: 10.3978/j.issn.1000-9604.2013.11.01
Scan to your mobile device or view this article at: http://www.thecjcr.org/article/view/3069/3968

Introduction

Colorectal cancer (CRC) is one of the most common types of cancer. In 2007, it was estimated that nearly 1.2 million new cases of CRC were diagnosed throughout the world, and about 630,000 deaths were estimated to occur from CRC, accounting for 8% of all cancer deaths (1). The vast majority of CRC deaths are due to the metastasis. CRC is highly curable when it is diagnosed at an early stage. However, CRC is less likely to be curable when it is detected at an advanced stage (when distant metastasis occurs). When CRC is confined to the colon or rectum, the 5-year survival may be as high as 90%. The 5-year survival rate is 68% for CRC patients who are diagnosed at the regional stage, while for patients with metastasis, is as much lower as 11% (2). It is estimated that approximately 60% CRC patients will eventually develop with metastasis (3). Thus, identification of CRC metastasis genes is one of the most important issues in CRC research.

DNA copy number changes can have a great impact on oncogenes and tumor suppressor genes (4). DNA copy number amplification will enhance the product or activity of oncogenes, and DNA copy number deletion may lead to

inactivation of tumor suppressor genes. Numerous studies have revealed that DNA copy number changes have a direct role on gene expression values. Hyman *et al*. (5) compared DNA copy number and expression levels of the genes in breast cancer, and found that both high-level and low-level changes of DNA copy number have a great impact on gene expression values, with 44% of the highly amplified genes being overexpressed, and 10.5% of the overexpressed genes showing an increase in gene copy number. Tsafrir *et al*. (6) investigated the relationships between DNA copy number and gene expression levels in CRC, and showed that changes in expression level are correlated with alterations in DNA content across many large regions of the genome.

Comparative genomic hybridization (CGH) is a powerful method for molecular cytogenetic analysis of DNA copy number changes (7). CGH technology can detect either chromosomal gains or chromosomal losses across the whole genome. In our previous study (8), 15 important chromosomal aberration sites, including 6 most common gains of chromosomal regions of 7p, 7q11-32, 8q, 13q, 20p and 20q, and 9 most common losses of 1p13-36, 4p15, 4q33-34, 8p12-23, 15q13-14, 15q24-25, 17p, 18p and 18q, were indentified in CRC.

Due to the rapid advance of microarray techniques, large amounts of microarray data have been deposited into public databases like Gene Expression Omnibus (GEO) (9), and the analysis of these data is usually limited to one single platform at a time. However, recent years have seen a few studies to conduct integrated analysis across multiple platforms (10-13). In this study, for the purpose of mining CRC metastasis-related genes, we present an integrated analysis of microarray data, by combining with evidence acquired from CGH data.

Materials and methods

Data

Gene expression profile data and associated clinical information of patients were obtained at GEO (9) website (http://www.ncbi.nlm.nih.gov/geo/, Series GSE2109). Samples of pathological stage 0 or 1 were treated as early-stage CRC, and samples of pathological stage 4 were treated as late-stage CRC. We obtained 55 early-stage primary CRC samples (pathological stage: 0 or 1; Group 1), 56 late-stage primary CRC samples (pathological stage: 4; Group 2), and 34 colorectal metastatic cancer (Group 3). The 34 metastatic sites include liver 26, lung 4, omentum 2, peritoneum 1 and abdominal wall mass 1 (*Table S1*) (14).

Group 1 *vs*. Group 2 was combined into PRI data set, which was used to classify between early-stage primary CRC and late-stage primary CRC. Group 2 *vs*. Group 3 was used to classify between late-stage primary CRC and colorectal metastatic cancer, and was joined into META data set.

All samples from the GSE2109 database are detected by Affymetrix Human Genome U133 Plus 2.0 arrays containing 54,675 probes. The gene expression values obtained from the database site have been calculated and normalized by Microarray Suite 5.0 (MAS5.0, Affymetrix, Inc.). In order to make distributions of each sample identical, the quantile algorithm (15) was used for further normalization (14).

Integrated genomic and transcriptomic analysis

The information obtained from CGH analysis was used for integrated genomic and transcriptomic analysis (8). Since the six most important amplified chromosomal regions of 7p, 7q11-32, 8q, 13q, 20p and 20q, and the 9 most important deleted chromosomal regions of 1p13-36, 4p15, 4q33-34, 8p12-23, 15q13-14, 15q24-25, 17p, 18p and 18q were determined in our previous study, the above regions were further investigated to identify metastasis-related genes in CRC. The integrated analysis is base on the assumption that oncogenes are present in chromosomal amplification regions, while tumor suppressor genes being located in chromosomal deletion fragments (16).

Gene expression profile data were analyzed to identify metastasis-related genes in CRC. The workflow diagram of integrated analysis of genomic and transcriptomic data is shown in *Figure 1*: firstly, Significant Analysis of Microarray (SAM) (17) is used to detect significantly differentially expressed genes across the whole genome. Next, the significantly differentially expressed genes located in the important chromosomal aberration regions are selected, and we choose the genes whose expression changes and copy number changes are consistent. In the chromosomal amplification sites, the up-regulated genes are selected, and the down-regulated genes are selected in the chromosomal deletion sites. The selected genes are called as the genomic and transcriptomic overlapping genes (overlapping genes). In the third step, the overlapping gene sets are analyzed by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (18,19). Finally, a gene selection algorithm is employed to identify metastasis-associated genes in CRC.

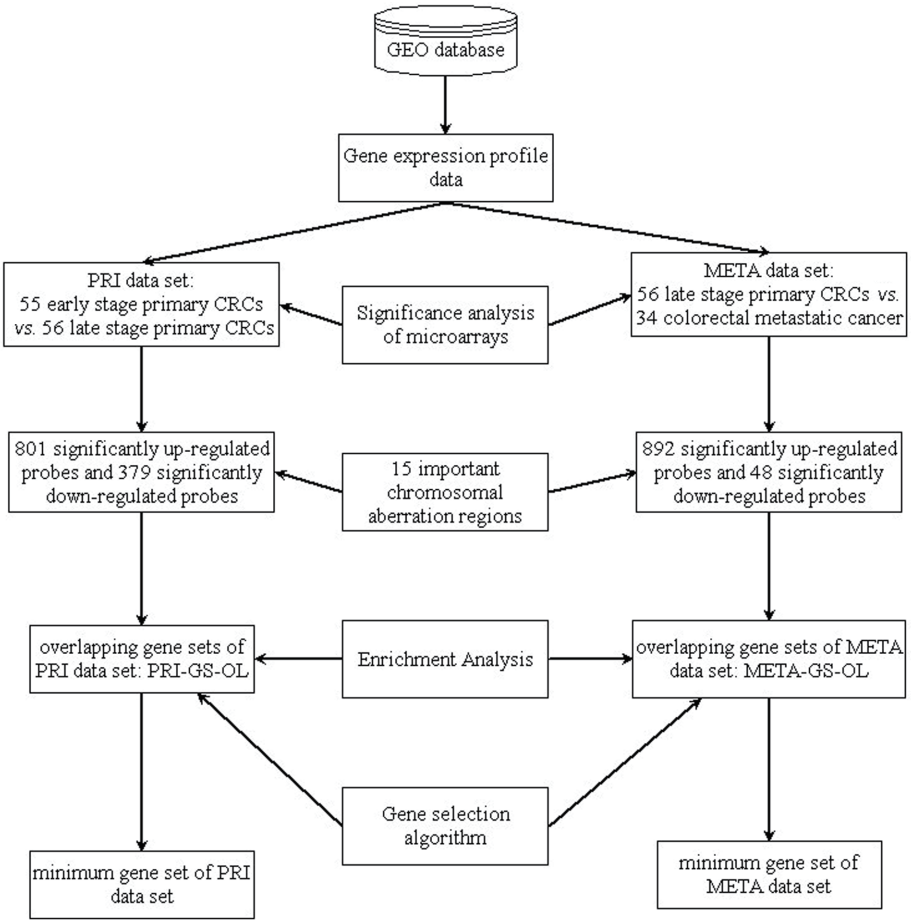

Figure 1 Workflow diagram of integrated analysis of genomic and transcriptomic data.

SAM

SAM is a statistical method proposed by Tusher *et al*. (17) in 2001, which is mainly used for determining whether the changes in gene expression are statistically significant. The SAM package is available at the following website: http://www-stat.stanford.edu/~tibs/SAM/.

SAM identifies significant genes by gene specific *t*-test, and it uses non-parametric statistical methods when the data may not follow a normal distribution. SAM uses repeated permutations of the data to determine whether the expression of any genes was significantly associated with the response variable. The use of permutation-based method can avoid the parametric assumptions of the distribution of single genes, which is the advantage of the method compared to other statistical techniques assuming the equal variance or gene independence (17).

SAM calculates the value d_i of gene i, and this value is used to measure the strength of the relationship between gene expression and the response variable. The value d_i is calculated as follows (20):

$$d_i = \frac{r_i}{s_i + s_0}, i = 1,2,\cdots n \quad [1]$$

where r_i is the linear regression coefficient of gene i, s_i is the standard error of r_i, s_0 is an exchangeability factor, and n is the number of genes.

SAM uses false discovery rate (FDR) to estimate the number of falsely significant genes. The formula is as follows (20):

$$\textit{False Discovery Rate (FDR)} = \frac{\textit{Median (90th Percentile) of \# of falsely called genes}}{\textit{\# of genes called significant}} \quad [2]$$

where the numerator of the formula is median (or 90th percentile) of the number of falsely called genes, and the denominator is the number of obtained significant genes. Generally, FDR is controlled less than 5%. SAM uses q-value to determine the lowest FDR, which is similar to the well-known P-value, but modified under multiple-testing

Table 1 Number of significantly up-regulated or down-regulated probes in the 15 important chromosomal aberration regions in PRI data set

Chromosomal aberration regions	Up-regulated probe number	Down-regulated probe number
Chromosomal amplification regions		
+7p	20	1
+7q11-32	14	4
+8q	22	0
+13q	24	0
+20p	3	1
+20q	6	4
Total	89	10
Chromosomal deletion regions		
–1p13-36	25	26
–4p15	4	4
–4q33-34	1	1
–8p12-23	13	8
–15q13-14	2	0
–15q24-25	3	5
–17p	7	21
–18p	4	0
–18q	3	7
Total	62	72

situations. The q-value measures the significance of gene *i*, and the corresponding q-value decreases as d_i increases (20).

Enrichment analysis

The DAVID (18,19) is utilized to analyze enriched gene ontology and pathway for overlapping gene sets. The Functional Annotation Chart tool is used to discover enriched annotations. The Functional Annotation Clustering tool is employed to cluster the related functional annotations into groups.

Gene selection algorithm

Gene selection procedure is conducted on the overlapping gene sets. SVM-T-RFE gene selection algorithm (14) is applied to generate ranked gene set, where the gene scores rank from high to low. The number of gene sets is reduced from original number to 1, and the leave-one-out cross-validation (LOOCV) method is used to assess performance of the classifiers. The gene set with the least number and highest accuracy is chosen as the minimum gene set.

Results

Overlapping gene sets of PRI and META data sets

SAM analysis was conducted on PRI data sets, unpaired two-sample *t*-test was used, permutation number was set to 100, delta=0.866, and FDR was set to 4.82%. Totally, 801 significantly up-regulated probes and 379 significantly down-regulated probes were screened out between early-stage primary CRC and late-stage primary CRC. Gene expression level changes of the significantly up-regulated or down-regulated probes in 15 important chromosomal aberration regions were analyzed (*Table 1*). It is worth noting that there were more significantly up-regulated probes in the chromosomal amplified regions (89 *vs.* 10), and more significantly down-regulated probes in the chromosomal deleted regions (72 *vs.* 62). The difference between the two groups were statistically significant as determined by the fisher's exact test (two-tailed, P=1.045e-12). The results show that an increase or decrease of genomic DNA copy number has a direct impact on the gene expression levels.

SAM analysis was also conducted on META data sets, unpaired two-sample *t*-test was used, permutation number was set to 100, delta=0.951, and FDR was set to 4.59%. Totally, 892 significantly up-regulated probes and 48 significantly down-regulated probes were detected between late-stage primary CRC and colorectal metastatic cancer. Gene expression level changes of the significantly up-regulated or down-regulated probes in 15 important chromosomal aberration regions were analyzed (*Table 2*). There was no statistically significant difference in up- or down-regulated probes between chromosomal amplified sites and deleted sites as determined by the fisher's exact test (two-tailed, P=0.319). It is speculated that other genomic alterations, such as point mutations, may have an impact on the expression value changes, and these changes cannot be detected by using CGH technology. No significantly down-regulated genes were found in certain chromosomal amplified regions, such as +13q, in both PRI and META data sets.

The overlapping gene sets were selected in which the gene expression changes and copy number changes were consistent. In other word, up-regulated genes were selected in chromosomal amplification sites, and down-regulated genes were selected in chromosomal deletion sites. The overlapping gene sets were obtained in PRI data sets (*Tables 3*,*4*) and META data sets (*Tables 5*,*6*), respectively.

Enrichment analysis

DAVID database was used to analyze the overlapping gene

Table 2 Number of significantly up-regulated or down-regulated probes in the 15 important chromosomal aberration regions in META data set

Chromosomal aberration regions	Up-regulated probe number	Down-regulated probe number
Chromosomal amplification regions		
+7p	5	0
+7q11-32	25	0
+8q	5	2
+13q	5	0
+20p	9	0
+20q	13	0
Total	62	2
Chromosomal deletion regions		
–1p13-36	49	5
–4p15	2	1
–4q33-34	1	0
–8p12-23	6	0
–15q13-14	1	1
–15q24-25	5	0
–17p	13	1
–18p	3	0
–18q	13	0
Total	93	8

sets in PRI and META data sets. The Functional Annotation Chart tool was applied to analyze enriched annotations. Thirty-seven significant gene ontology and pathway (EASE score <0.05) were identified in the overlapping gene list of PRI data sets, and a detailed list is shown in *Table S2*. The top three significant gene annotations include organelle membrane (GO:0031090), organelle part (GO:0044422), and cellular protein complex assembly (GO:0043623). In the overlapping gene list of META data sets, 79 significant gene ontology and pathway (EASE score <0.05) were identified, and a detailed list is shown in *Table S3*. The top three significant gene annotations include regulation of response to external stimulus (GO:0032101), extracellular space (GO:0005615), and regulation of signal transduction (GO:0009966).

The Functional Annotation Clustering tool was applied to cluster functionally related annotations into a group, and the default settings were used (classification stringency is medium). In the overlapping gene sets of PRI data sets, the most enriched cluster is biological process of cellular component organization and cellular component biogenesis, and a detailed list is shown in *Table S4* (enrichment score >1.5). The most enriched cluster in the overlapping gene sets of META data sets is cellular components of extracellular region, and a detailed list is shown in *Table S5* (enrichment score >1.5).

Table 3 Overlapping gene sets of PRI data set (up-regulated probes)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
201792_at	AEBP1	4.304007	2.264933	0	7p13
232266_x_at	CDK13	2.974419	1.281524	4.820638	7p13
210095_s_at	IGFBP3	3.511416	1.552975	1.608303	7p13-p12
217381_s_at	TRGV5	3.376586	1.680866	1.811866	7p14
37547_at	BBS9	3.812358	1.516202	0.864754	7p14
204051_s_at	SFRP4	4.460986	2.702903	0	7p14.1
204052_s_at	SFRP4	4.766370	3.075708	0	7p14.1
213151_s_at	SEPT7	3.386422	1.152214	1.811866	7p14.2
225631_at	EEPD1	2.988437	1.294418	4.820638	7p14.2
230728_at	FKBP14	3.047380	1.250190	4.208125	7p14.3
203106_s_at	VPS41	3.655002	1.320942	1.247959	7p14-p13
1554018_at	GPNMB	3.171425	1.721327	3.264665	7p15
203695_s_at	DFNA5	3.138812	1.476765	3.264665	7p15
210511_s_at	INHBA	3.702019	2.372578	1.120527	7p15-p13
227140_at	INHBA	4.070521	1.947805	0.548672	7p15-p13

Table 3 (*continued*)

Table 3 (continued)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
226189_at	ITGB8	3.297660	1.523012	2.449903	7p21.1
221911_at	ETV1	3.480852	1.752582	1.608303	7p21.3
229159_at	THSD7A	3.532920	1.706454	1.503265	7p21.3
225147_at	CYTH3	3.087086	1.219223	3.743346	7p22.1
229581_at	ELFN1	3.216421	1.299175	2.791487	7p22.3
213474_at	KCTD7	3.013335	1.297202	4.208125	7q11.21
235408_x_at	ZNF117	3.882888	1.492717	0.864754	7q11.21
202006_at	PTPN12	3.608639	1.282147	1.247959	7q11.23
226364_at	HIP1	3.228099	1.411253	2.791487	7q11.23
209780_at	PHTF2	3.771224	1.240836	0.864754	7q11.23-q21
1560526_at	MAGI2-IT	3.477312	1.808641	1.608303	7q21
204604_at	CDK14	4.049699	1.458664	0.548672	7q21-q22
202403_s_at	COL1A2	3.442993	1.469117	1.811866	7q22.1
202404_s_at	COL1A2	3.628397	1.523565	1.247959	7q22.1
205381_at	LRRC17	3.190571	1.458251	3.264665	7q22.1
229218_at	COL1A2	3.812675	1.809962	0.864754	7q22.1
207103_at	KCND2	3.409996	1.581182	1.811866	7q31
205003_at	DOCK4	3.231822	1.423837	2.791487	7q31.1
222421_at	UBE2H	3.382986	1.279159	1.811866	7q32
226119_at	PCMTD1	4.356758	1.264309	0	8q11.23
200949_x_at	RPS20	2.996007	1.114792	4.820638	8q12
205372_at	PLAG1	3.446751	1.490451	1.608303	8q12
212344_at	SULF1	3.501949	1.702950	1.608303	8q13.1
212353_at	SULF1	4.563269	1.991297	0	8q13.1
212354_at	SULF1	4.150777	1.851440	0	8q13.1
228438_at	LOC100132891	4.141843	1.708721	0	8q13.3
223475_at	CRISPLD1	3.131052	1.615207	3.743346	8q21.11
223614_at	MMP16	3.143758	1.305039	3.264665	8q21.3
205529_s_at	RUNX1T1	3.497952	1.559165	1.608303	8q22
227277_at	MTDH	3.015625	1.265407	4.208125	8q22.1
225681_at	CTHRC1	4.288066	2.041583	0	8q22.3
204501_at	NOV	3.978618	1.785215	0.548672	8q24.1
214321_at	NOV	3.173398	1.677539	3.264665	8q24.1
1557706_at	ZHX2	3.029406	1.298014	4.208125	8q24.13
225803_at	FBXO32	3.241878	1.540598	2.791487	8q24.13
222133_s_at	PHF20L1	4.204328	1.283748	0	8q24.22
225034_at	ST3GAL1	3.203673	1.364761	2.791487	8q24.22
227523_s_at	PHF20L1	3.212993	1.174761	2.791487	8q24.22
1559529_at	PTK2	3.499030	1.335719	1.608303	8q24.3
206188_at	ZNF623	3.338816	1.382219	2.123418	8q24.3
227902_at	ZFP41	3.199101	1.522638	2.791487	8q24.3

Table 3 (*continued*)

Table 3 (continued)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
223278_at	GJB2	3.061267	1.537424	4.208125	13q11-q12
223380_s_at	LATS2	3.777511	1.350111	0.864754	13q11-q12
227013_at	LATS2	3.700755	1.378994	1.120527	13q11-q12
204174_at	ALOX5AP	3.296789	1.608920	2.449903	13q12
226497_s_at	FLT1	3.606522	1.351586	1.247959	13q12
227401_at	IL17D	3.277236	1.362619	2.449903	13q12.11
1565765_x_at	C13orf33	3.013351	1.401512	4.208125	13q12.3
227713_at	KATNAL1	3.610553	1.408405	1.247959	13q12.3
1555778_a_at	POSTN	3.001705	1.883241	4.820638	13q13.3
210809_s_at	POSTN	3.284833	1.769054	2.449903	13q13.3
203597_s_at	WBP4	3.068590	1.250545	4.208125	13q14.11
207135_at	HTR2A	3.468428	1.342974	1.608303	13q14-q21
204084_s_at	CLN5	3.265154	1.201561	2.449903	13q21.1-q32
227261_at	KLF12	3.486387	1.385815	1.608303	13q22
229881_at	KLF12	3.433983	1.659366	1.811866	13q22
238940_at	KLF12	3.150969	1.515506	3.264665	13q22
227059_at	GPC6	3.208697	1.584515	2.791487	13q32
203640_at	MBNL2	3.246237	1.252403	2.791487	13q32.1
1557080_s_at	ITGBL1	4.510735	2.060145	0	13q33
205422_s_at	ITGBL1	4.758086	2.700174	0	13q33
214927_at	ITGBL1	4.505078	2.179526	0	13q33
231993_at	ITGBL1	3.768782	2.839518	0.864754	13q33
213728_at	LAMP1	3.296238	1.370269	2.449903	13q34
225562_at	RASA3	3.506652	1.306546	1.608330	13q34
219463_at	C20orf103	3.324177	2.106647	2.123418	20p12
206746_at	BFSP1	2.977808	1.370223	4.820638	20p12.1
226607_at	C20orf194	3.075098	1.478292	3.743346	20p13
202075_s_at	PLTP	2.970511	1.923566	4.820638	20q13.12
211892_s_at	PTGIS	3.102055	1.481988	3.743346	20q13.13
214844_s_at	DOK5	4.234107	1.477706	0	20q13.2
235616_at	TSHZ2	3.207755	1.517808	2.791487	20q13.2
238577_s_at	TSHZ2	3.301027	1.521081	2.449903	20q13.2
243940_at	TSHZ2	3.400265	1.637747	1.811866	20q13.2

Table 4 Overlapping gene sets of PRI data set (down-regulated probes)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
210440_s_at	CDC14A	–3.33285	0.694503	4.208125	1p21
222993_at	MRPL37	–3.90628	0.795444	1.811866	1p32.1
225792_at	HOOK1	–3.42743	0.759736	3.743346	1p32.1
234672_s_at	TMEM48	–3.43041	0.799813	3.743346	1p32.3

Table 4 (*continued*)

Table 4 (continued)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
233638_s_at	POMGNT1	–3.31034	0.774480	4.208125	1p34.1
225316_at	MFSD2A	–3.39515	0.641154	3.743346	1p34.2
201586_s_at	SFPQ	–3.29270	0.878785	4.208125	1p34.3
236073_at	EPHA10	–4.17836	0.670291	1.247959	1p34.3
225536_at	TMEM54	–3.61449	0.643065	2.791487	1p35-p34
202296_s_at	RER1	–3.35222	0.852227	4.208125	1p36
208002_s_at	ACOT7	–4.65798	0.700835	0	1p36
228361_at	E2F2	–3.70865	0.745922	2.123418	1p36
206499_s_at	RCC1	–3.90723	0.714794	1.811866	1p36.1
208668_x_at	HMGN2	–3.34750	0.849877	4.208125	1p36.1
202292_x_at	LYPLA2	–3.36913	0.837185	4.208125	1p36.11
236058_at	C1orf172	–3.83575	0.695138	1.811866	1p36.11
202675_at	SDHB	–3.61861	0.810530	2.791487	1p36.1-p35
222930_s_at	AGMAT	–3.81465	0.732350	1.811866	1p36.21
227617_at	TMEM201	–3.70171	0.597220	2.123418	1p36.22
229852_at	NMNAT1	–3.53730	0.776724	3.264665	1p36.22
1555360_a_at	DNAJC11	–4.70767	0.715027	0	1p36.31
215792_s_at	DNAJC11	–3.51994	0.816680	3.264665	1p36.31
202115_s_at	NOC2L	–3.55639	0.834877	2.791487	1p36.33
218580_x_at	AURKAIP1	–3.80224	0.824030	1.811866	1p36.33
225552_x_at	AURKAIP1	–3.69128	0.824732	2.449903	1p36.33
213613_s_at	NADK	–3.34778	0.743440	4.208125	1p36.33-p36.21
212311_at	SEL1L3	–3.37788	0.804773	4.208125	4p15.2
222631_at	PI4K2B	–3.65857	0.801903	2.449903	4p15.2
201386_s_at	DHX15	–3.95516	0.815145	1.608303	4p15.3
218663_at	NCAPG	–3.30746	0.733221	4.208125	4p15.33
202763_at	CASP3	–4.10295	0.806790	1.503265	4q34
208459_s_at	XPO7	–3.60312	0.771616	2.791487	8p21
209368_at	EPHX2	–3.32710	0.644251	4.208125	8p21
203110_at	PTK2B	–3.72926	0.779339	2.123418	8p21.1
219148_at	PBK	–4.25180	0.567576	1.247959	8p21.2
218777_at	REEP4	–3.63557	0.810997	2.449903	8p21.3
206797_at	NAT2	–3.94812	0.559697	1.811866	8p22
214440_at	NAT1	–3.52457	0.703290	3.264665	8p22
218096_at	AGPAT5	–3.25882	0.780746	4.820638	8p23.1
203663_s_at	COX5A	–3.77205	0.794286	2.123418	15q24.1
229426_at	COX5A	–4.71157	0.695017	0	15q24.1
209132_s_at	COMMD4	–3.34164	0.822922	4.208125	15q24.2
202069_s_at	IDH3A	–3.40210	0.788855	3.743346	15q25.1-q25.2
202070_s_at	IDH3A	–4.67006	0.675054	0	15q25.1-q25.2
200854_at	NCOR1	–3.87244	0.835912	1.811866	17p11.2

Table 4 (*continued*)

Table 4 (continued)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
200857_s_at	NCOR1	–3.50761	0.843093	3.264665	17p11.2
203858_s_at	COX10	–4.68613	0.807404	0	17p12
205891_at	ADORA2B	–3.69089	0.653336	2.449903	17p12
202790_at	CLDN7	–4.19963	0.625198	1.247959	17p13
205527_s_at	GEMIN4	–3.82507	0.730225	1.811866	17p13
217099_s_at	GEMIN4	–3.67737	0.753848	2.449903	17p13
223221_at	SCO1	–3.39535	0.785848	3.743346	17p13.1
223286_at	C17orf81	–3.46450	0.772520	3.264665	17p13.1
209208_at	MPDU1	–3.71669	0.744898	2.123418	17p13.1-p12
207567_at	SLC13A2	–3.42348	0.590773	3.743346	17p13.2
221255_s_at	TMEM93	–3.25477	0.848944	4.820638	17p13.2
236225_at	GGT6	–4.04904	0.612922	1.608303	17p13.2
202632_at	DPH1/OVCA2	–3.35084	0.840405	4.208125	17p13.3
207088_s_at	SLC25A11	–4.24295	0.781024	1.247959	17p13.3
207521_s_at	ATP2A3	–3.50373	0.521814	3.264665	17p13.3
208910_s_at	C1QBP	–4.91441	0.652626	0	17p13.3
213036_x_at	ATP2A3	–3.79962	0.553766	1.811866	17p13.3
214214_s_at	C1QBP	–4.93701	0.699116	0	17p13.3
214671_s_at	ABR	–3.42788	0.775664	3.743346	17p13.3
219868_s_at	ANKFY1	–3.51131	0.753165	3.264665	17p13.3
213738_s_at	ATP5A1	–3.52712	0.818488	3.264665	18q21
217640_x_at	SKA1	–3.28139	0.686403	4.820638	18q21.1
207017_at	RAB27B	–3.29007	0.566578	4.820638	18q21.2
200027_at	NARS	–3.82585	0.824023	1.811866	18q21.31
207037_at	TNFRSF11A	–3.37319	0.665530	4.208125	18q22.1
223180_s_at	C18orf55	–3.73005	0.780212	2.123418	18q22.3
203322_at	ADNP2	–3.42595	0.834176	3.743346	18q23

Table 5 Overlapping gene sets of META data set (up-regulated probes)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
220106_at	NPC1L1	4.05075	3.721565	0.314247	7p13
205302_at	IGFBP1	4.825902	126.2486	0	7p13-p12
1560416_at	DNAH11	3.501318	1.438866	1.424107	7p21
227170_at	ZNF316	3.303586	1.277174	2.089251	7p22.1
229581_at	ELFN1	3.475702	1.518288	1.542669	7p22.3
204608_at	ASL	3.183063	1.501959	3.423644	7cen-q11.2
208928_at	POR	3.995840	1.634573	0.314247	7q11.2
1570505_at	ABCB4	4.135380	3.835955	0.314247	7q21.1
205998_x_at	CYP3A4	3.951908	6.466223	0.540426	7q21.1
205999_x_at	CYP3A4	4.259389	40.58838	0	7q21.1

Table 5 (*continued*)

Table 5 (continued)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
207819_s_at	ABCB4	4.535347	7.479770	0	7q21.1
208367_x_at	CYP3A4	4.015195	117.8514	0.314247	7q21.1
209961_s_at	HGF	4.399540	1.551411	0	7q21.1
209994_s_at	ABCB1/ABCB4	3.142725	2.234568	3.940973	7q21.1/7q21.12
202605_at	GUSB	3.570870	1.285986	1.247747	7q21.11
203775_at	SLC25A13	4.133443	1.568941	0.314247	7q21.3
206345_s_at	PON1	3.503265	27.82819	1.424107	7q21.3
213695_at	PON3	4.238667	15.42315	0	7q21.3
229061_s_at	SLC25A13	4.597306	1.552435	0	7q21.3
229081_at	SLC25A13	3.869709	1.406888	0.70315	7q21.3
211843_x_at	CYP3A7	3.508686	11.53890	1.424107	7q21-q22.1
205923_at	RELN	4.271732	2.878173	0	7q22
207257_at	EPO	3.548880	2.434852	1.424107	7q22
207883_s_at	TFR2	3.842054	4.455354	0.70315	7q22
210215_at	TFR2	3.968742	16.19620	0.540426	7q22
209309_at	AZGP1	4.040993	4.910858	0.314247	7q22.1
217014_s_at	AZGP1	4.000525	4.639553	0.314247	7q22.1
219753_at	STAG3	3.178664	1.785313	3.423644	7q22.1
231704_at	ZNF498	4.732386	2.762958	0	7q22.1
209631_s_at	GPR37	3.915808	2.250653	0.540426	7q31
232721_at	TRIM55	4.118310	2.570882	0.314247	8q13.1
236175_at	TRIM55	3.977655	2.893115	0.314247	8q13.1
206065_s_at	DPYS	3.924076	12.30999	0.540426	8q22
239860_at	LOC100130232	3.387196	3.542149	2.023082	8q22.3
207420_at	COLEC10	4.098130	1.876468	0.314247	8q23-q24.1
206651_s_at	CPB2	4.134995	76.85469	0.314247	13q14.11
239683_at	CLYBL	3.622698	1.468668	1.05182	13q32
205620_at	F10	4.289245	3.111049	0	13q34
207300_s_at	F7	3.928524	3.119869	0.540426	13q34
208034_s_at	PROZ	3.744082	4.285206	0.957639	13q34
211416_x_at	GGTLC1	3.107466	1.219471	3.940973	20p11.1
221440_s_at	RBBP9	3.451234	1.521874	1.542669	20p11.2
235724_at	ACSS1	3.154978	1.695083	3.423644	20p11.23-p11.21
220224_at	HAO1	4.283785	35.28161	0	20p12
202896_s_at	SIRPA	3.434839	1.324024	1.759496	20p13
209237_s_at	SLC23A2	3.182192	1.661882	3.423644	20p13
211852_s_at	ATRN	3.095574	1.477758	3.940973	20p13
218145_at	TRIB3	3.638672	1.657822	1.05182	20p13-p12.2
241741_at	CRLS1	3.534782	1.507576	1.424107	20p13-p12.3
211652_s_at	LBP	4.774147	10.25493	0	20q11.23
214461_at	LBP	4.954503	21.43375	0	20q11.23

Table 5 (*continued*)

Table 5 (continued)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
201042_at	TGM2	3.433253	1.593450	1.759496	20q12
202071_at	SDC4	3.553425	1.338339	1.247747	20q12
211003_x_at	TGM2	3.356415	1.772808	2.023082	20q12
211573_x_at	TGM2	3.342880	1.545483	2.023082	20q12
236652_at	LOC149703	3.188570	6.107429	3.423644	20q13.12
220451_s_at	BIRC7	3.279162	1.674870	2.518540	20q13.3
221848_at	ZGPAT	3.119508	1.854418	3.940973	20q13.3
57539_at	ZGPAT	3.136639	1.771246	3.940973	20q13.3
208383_s_at	PCK1	3.522144	3.973031	1.424107	20q13.31
233000_x_at	DPH3P1	3.100489	1.508928	3.940973	20q13.33
233328_x_at	SLC17A9	3.721116	1.744356	0.957639	20q13.33

Table 6 Overlapping gene sets of META data set (down-regulated probes)

Probe ID	Gene symbol	Score (d)	Fold change	q-value (%)	Chromosomal regions
206458_s_at	WNT2B	–4.21352	0.523566	2.023082	1p13
224937_at	PTGFRN	–4.84643	0.621072	0.314247	1p13.1
224950_at	PTGFRN	–4.64233	0.663579	0.703150	1p13.1
243357_at	NEGR1	–4.02139	0.390821	2.828227	1p31.1
204879_at	PDPN	–4.24088	0.395762	1.759496	1p36.21
234936_s_at	CC2D2A	–4.18253	0.672379	2.023082	4p15.32
218469_at	GREM1	–4.19294	0.315323	2.023082	15q13.3
208807_s_at	CHD3	–4.21163	0.699398	2.023082	17p13.1

SVM-T-RFE gene selection algorithm

SVM-T-RFE gene selection algorithm was applied to select informative genes in PRI data set. The initial overlapping gene sets (PRI-GS-OL) of PRI data set contained 161 probes. SVM-T-RFE gene selection algorithm generates ranked gene set, where the genes rank from high to low score. The number of genes was reduced from 161 to 1, and LOOCV method was used to assess the performance of the classifier. Parameter θ in SVM-T-RFE gene selection algorithm was selected from finite set {0, 0.01, 0.02,...0.98, 0.00, 1.0}, where each element increases from 0 to 1 by a step of 0.01. When parameter θ was set to 0.51, a minimum gene set was obtained with the minimum number (14) of genes, and the highest classification accuracy (100%) (*Figure 2*).

Table 7 shows the minimum gene set of PRI data set. Secreted frizzled-related protein 4 (SFRP4) (21), and RAB27B, member RAS oncogene family (RAB27B) (22,23) were reported to be associated with CRC or its metastasis.

SVM-T-RFE gene selection algorithm was also carried out to select informative genes in META data set. The initial overlapping gene sets (META-GS-OL) contained 70 probes. When parameter θ was set to 0.36, a minimum gene set was obtained with the minimum number (14) of genes, and the highest classification accuracy (100%) (*Figure 3*).

Table 8 gives the gene list of the minimum gene set in META data set. Evidence shows that hepatocyte growth factor (HGF) (24), and cytochrome P450, family 3, subfamily A, polypeptide 4 (CYP3A4) (25) are related to CRC or its metastasis.

Discussion

In recent years, due to the rapid development of molecular biology experimental techniques, a large amount of data has been accumulated, including genome, transcriptome and proteome detection platform. Previous studies often focus on data from a single platform, and rarely address the problem of integration of data from a variety of

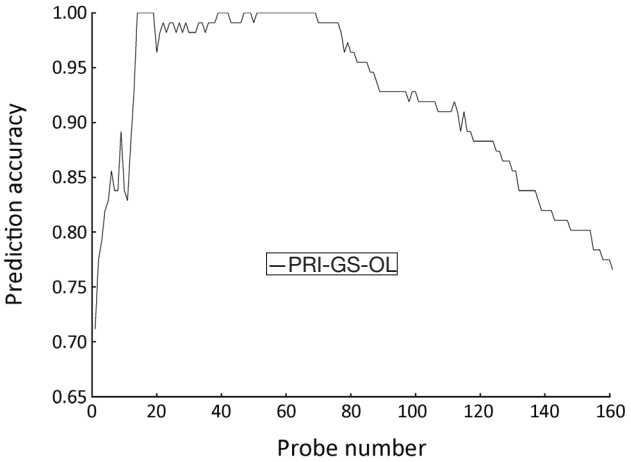

Figure 2 Prediction accuracy of the overlapping gene sets (PRI-GS-OL) obtained from PRI data set. The probe number was reduced from 161 to 1.

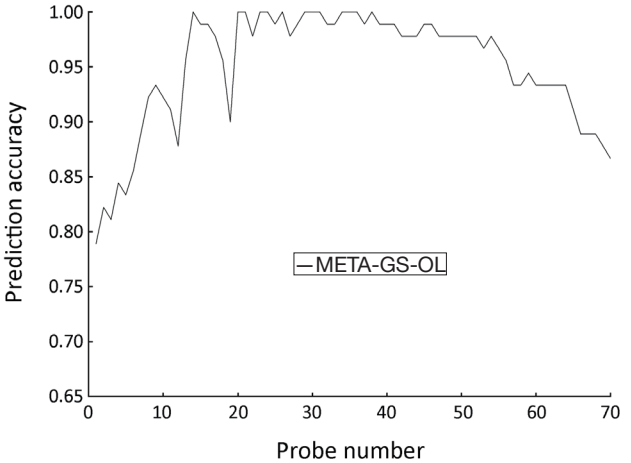

Figure 3 Prediction accuracy of the overlapping gene sets (META-GS-OL) obtained from META data set. The probe number was reduced from 70 to 1.

Table 7 The minimum gene set selected in PRI dataset (gene scores rank from high to low)

Probe ID	Gene symbol	Gene name	Chromosomal regions
219868_s_at	ANKFY1	Ankyrin repeat and FYVE domain containing1	17p13.3
213613_s_at	NADK	NAD kinase	1p36.33-p36.21
208002_s_at	ACOT7	Acyl-coa thioesterase 7	1p36
222133_s_at	PHF20L1	PHD finger protein 20-like 1	8q24.22
203858_s_at	COX10	COX10 homolog, cytochrome c oxidase assembly protein, heme A: farnesyltransferase (yeast)	17p12
204051_s_at	SFRP4	Secreted frizzled-related protein 4	7p14.1
207567_at	SLC13A2	Solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 2	17p13.2
225803_at	FBXO32	F-box protein 32	8q24.13
205527_s_at	GEMIN4	Gem (nuclear organelle) associated protein 4	17p13
207017_at	RAB27B	RAB27B, member RAS oncogene family	18q21.2
206746_at	BFSP1	Beaded filament structural protein 1, filensin	20p12.1
217099_s_at	GEMIN4	Gem (nuclear organelle) associated protein 4	17p13
233638_s_at	POMGNT1	Protein O-linked mannose beta1,2-N-acetylglucosaminyltransferase	1p34.1
217381_s_at	TRGV5	T cell receptor gamma variable 5	7p14

platforms. In this study, we present an integrated strategy, by combining microarray data with CGH information, in an attempt to indentify CRC metastasis-related genes.

In the minimum gene sets selected from PRI data set, at least two genes were found to have direct evidence of the associations with CRC. SFRP4, located in the region of 7p14.1, is a member of the secreted frizzled related proteins (sFRPs) family. The sFRPs are able to bind and inhibit Wnt signalling pathways. SFRP4 expression was reported to be up-regulated in CRC (21), and the up-regulation level was found to be 2.7 folds in PRI data set. RAB27B, whose chromosomal location is on 18q21.2, is a member of RAS oncogene family. RAB27B was reported to be associated with poor prognosis in human breast tumors, and crucial for the invasiveness and

Table 8 The minimum gene set selected in META dataset (gene scores rank from high to low)

Probe ID	Gene symbol	Gene name	Chromosomal regions
234936_s_at	CC2D2A	Coiled-coil and C2 domain containing 2A	4p15.32
218469_at	GREM1	Gremlin 1	15q13.3
209961_s_at	HGF	Hepatocyte growth factor (hepapoietin A; scatter factor)	7q21.1
243357_at	NEGR1	Neuronal growth regulator 1	1p31.1
241741_at	CRLS1	Cardiolipin synthase 1	20p13-p12.3
207257_at	EPO	Erythropoietin	7q22
211652_s_at	LBP	Lipopolysaccharide binding protein	20q11.23
205999_x_at	CYP3A4	Cytochrome P450, family 3, subfamily A, polypeptide 4	7q21.1
224937_at	PTGFRN	Prostaglandin F2 receptor negative regulator	1p13.1
239683_at	CLYBL	Citrate lyase beta like	13q32
204879_at	PDPN	Podoplanin	1p36.21
224950_at	PTGFRN	Prostaglandin F2 receptor negative regulator	1p13.1
208383_s_at	PCK1	Phosphoenolpyruvate carboxykinase 1 (soluble)	20q13.31
1560416_at	DNAH11	Dynein, axonemal, heavy chain 11	7p21

metastasis of breast cancer cell lines (22). Dong *et al*. (23) reported that RAB27B may serve as a valuable prognostic predictor for hepatocellular carcinoma patients, and they also found a close relationship between RAB27B expression and clinicopathological characteristics and prognosis in CRC.

At least two genes in the minimum gene sets of META data set were found to be associated with CRC. HGF, located in 7q21.1, regulates cell proliferation, migration, and morphogenesis by binding to the receptor of proto-oncogene c-met. Previous studies indicated that the HGF/Met pathway plays a role in the progression of CRC, and c-met gene overexpression contributes to the metastatic phenotype of CRC (24). CYP3A4, whose chromosomal location is also on 7q21.1, is a member of the cytochrome P450 superfamily, which catalyzes many reactions involved in drug metabolism as well as the synthesis of cholesterol, steroids and other lipids. CYP3A4 was reportedly the highest frequency of strong immunoreactivity of P450 in normal colon and there was significant correlation between its immunoreactivity and CRC stage (25).

Our results demonstrated that integrated analysis is an effective strategy for mining cancer-associated genes. There are corresponding oncogenes and tumor suppressor genes in chromosomal amplification and deletion regions. CRC metastasis, as the advanced stage of colorectal tumorigenesis, is a complicated, multi-step biological process. CRC metastasis has rarely been systematically addressed by previous studies, and the molecular mechanism remains far from being completely elucidated. Although many of the selected genes need to be validated in further molecular experiments, the preliminary results show that multiple oncogenes and tumor suppressor genes participate in the process of CRC metastasis. Moreover, the gene expression profiles of our study were downloaded from one single data set (GSE2109), and heterogeneous microarray data sets can be combined to discover metastasis-related genes in CRC in the future work (26). In addition, network-based approach can be used to integrate genomic, transcriptomic, and proteomic data, in an attempt to pinpoint significant genes and highlight the molecular mechanisms underlying CRC metastasis (27). In conclusion, in-depth study of these genes will lead to a better understanding of the molecular mechanisms of CRC metastasis and provide potential biomarkers and therapeutic targets for CRC metastasis.

Acknowledgements

This study was supported by a grant from the National Natural Science Foundation of China (Grant No. 61373057), and a grant from the Zhejiang Provincial Natural Science Foundation of China (Grant No. Y1110763).
Disclosure: The authors declare no conflict of interest.

References

1. ACS. Global Cancer Facts and Figures 2007. Atlanta:

American Cancer Society; 2007.
2. ACS. Colorectal Cancer Facts and Figures 2008-2010. Atlanta: American Cancer Society; 2010.
3. Boige V, Malka D, Elias D. Hepatic arterial infusion of oxaliplatin and intravenous LV5FU2 in unresectable liver metastases from colorectal cancer after systemic chemotherapy failure. Ann Surg Oncol 2008;15:219-26.
4. Pollack JR, Sorlie T, Perou CM, et al. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 2002;99:12963-8.
5. Hyman E, Kauraniemi P, Hautaniemi S, et al. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 2002;62:6240-5.
6. Tsafrir D, Bacolod M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 2006;66:2129-37.
7. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 1992;258:818-21.
8. Li X, Chen J, Lu BJ, et al. –8p12-23 and +20q Are Predictors of Subtypes and Metastatic Pathways in Colorectal Cancer: Construction of Tree Models Using Comparative Genomic Hybridization Data. OMICS 2011;15:37-47.
9. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013;41:D991-5.
10. Natrajan R, Weigelt B, Mackay A, et al. An integrative genomic and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. Breast Cancer Res Treat 2010;121:575-89.
11. Huang N, Shah PK, Li C. Lessons from a decade of integrating cancer copy number alterations with gene expression profiles. Brief Bioinform 2012;13:305-16.
12. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346-52.
13. Rakosy Z, Ecsedi S, Toth R, et al. Integrative genomics identifies gene signature associated with melanoma ulceration. PLoS One 2013;8:e54958.
14. Li X, Peng S, Chen J, et al. SVM-T-RFE: A novel gene selection algorithm for identifying metastasis-related genes in colorectal cancer using gene expression profiles. Biochem Biophys Res Commun 2012;419:148-53.
15. Bolstad BM, Irizarry RA, Astrand M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185-93.
16. Li XB. Mathematical modeling of carcinogenesis based on chromosome aberration data. Chin J Cancer Res 2009;21:240-6.
17. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-21.
18. Dennis G Jr, Sherman BT, Hosack DA, et al. DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:p3.
19. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44-57.
20. Chu G, Jun Li, Narasimhan B, et al. SAM Significance Analysis of Microarrays-Users guide and technical document. Available online: http://www-stat.stanford.edu/~tibs/SAM/sam.pdf
21. Feng Han Q, Zhao W, Bentel J, et al. Expression of sFRP-4 and beta-catenin in human colorectal carcinoma. Cancer Lett 2006;231:129-37.
22. Hendrix A, Maynard D, Pauwels P, et al. Effect of the secretory small GTPase Rab27B on breast cancer growth, invasion, and metastasis. J Natl Cancer Inst 2010;102:866-80.
23. Dong WW, Mou Q, Chen J, et al. Differential expression of Rab27A/B correlates with clinical outcome in hepatocellular carcinoma. World J Gastroenterol 2012;18:1806-13.
24. Di Renzo MF, Olivero M, Giacomini A, et al. Overexpression and amplification of the met/HGF receptor gene during the progression of colorectal cancer. Clin Cancer Res 1995;1:147-54.
25. Kumarakulasingham M, Rooney PH, Dundas SR, et al. Cytochrome p450 profile of colorectal cancer: identification of markers of prognosis. Clin Cancer Res 2005;11:3758-65.
26. Park PJ, Kong SW, Tebaldi T, et al. Integration of heterogeneous expression data sets extends the role of the retinol pathway in diabetes and insulin resistance. Bioinformatics 2009;25:3121-7.
27. Shi M, Beauchamp RD, Zhang B. A network-based gene expression signature informs prognosis and treatment for colorectal cancer patients. PLoS One 2012;7:e41292.

Cite this article as: Li X, Peng S. Identification of metastasis-associated genes in colorectal cancer through an integrated genomic and transcriptomic analysis. Chin J Cancer Res 2013;25(6):623-636. doi: 10.3978/j.issn.1000-9604.2013.11.01

Supplementary data

Table S1 Information of 55 early-stage primary CRC, 56 late-stage primary CRC and 34 colorectal metastatic cancer

GEO accession	Pathological stage	Location
GSM53055	1 (primary)	Colon
GSM102549	1 (primary)	Colon
GSM76629	1 (primary)	Colon
GSM38075	1 (primary)	Colon
GSM53126	1 (primary)	Colon
GSM76576	1 (primary)	Colon
GSM76611	1 (primary)	Colon
GSM89044	1 (primary)	Colon
GSM102561	1 (primary)	Colon
GSM102572	1 (primary)	Colon
GSM102579	1 (primary)	Colon
GSM117656	1 (primary)	Colon
GSM117672	1 (primary)	Colon
GSM117676	1 (primary)	Colon
GSM117678	1 (primary)	Colon
GSM117681	1 (primary)	Colon
GSM137922	1 (primary)	Colon
GSM117720	1 (primary)	Colon
GSM138005	1 (primary)	Colon
GSM138022	1 (primary)	Colon
GSM138037	1 (primary)	Colon
GSM102479	1 (primary)	Rectosigmoid
GSM102517	1 (primary)	Rectosigmoid
GSM117652	1 (primary)	Rectosigmoid
GSM117699	1 (primary)	Rectosigmoid
GSM88982	1 (primary)	Rectum
GSM89074	1 (primary)	Rectum
GSM89095	1 (primary)	Rectum
GSM117577	1 (primary)	Rectum
GSM117683	1 (primary)	Rectum
GSM117707	1 (primary)	Rectum
GSM203731	1 (primary)	Colon
GSM152602	1 (primary)	Colon
GSM152632	1 (primary)	Colon
GSM152666	1 (primary)	Colon
GSM152684	1 (primary)	Colon
GSM152725	1 (primary)	Colon
GSM179820	1 (primary)	Colon
GSM179838	1 (primary)	Colon
GSM179844	1 (primary)	Colon
GSM179859	1 (primary)	Colon
GSM179889	1 (primary)	Colon
GSM203625	1 (primary)	Colon
GSM203673	1 (primary)	Colon
GSM203687	1 (primary)	Colon
GSM203723	1 (primary)	Colon
GSM231921	1 (primary)	Colon
GSM277680	1 (primary)	Colon
GSM353890	1 (primary)	Colon
GSM179894	1 (primary)	Rectosigmoid
GSM301702	1 (primary)	Rectosigmoid
GSM325825	0 (primary)	Rectosigmoid
GSM152629	1 (primary)	Rectum
GSM203788	1 (primary)	Rectum
GSM353915	1 (primary)	Rectum
GSM88994	4 (primary)	Colon
GSM38061	4 (primary)	Colon
GSM38089	4 (primary)	Colon
GSM76512	4 (primary)	Colon
GSM76529	4 (primary)	Colon
GSM76531	4 (primary)	Colon
GSM76605	4 (primary)	Colon
GSM76615	4 (primary)	Colon
GSM102568	4 (primary)	Colon
GSM102574	4 (primary)	Colon
GSM89103	4 (primary)	Colon
GSM102472	4 (primary)	Colon
GSM102513	4 (primary)	Colon
GSM102521	4 (primary)	Colon
GSM102533	4 (primary)	Colon
GSM137926	4 (primary)	Colon
GSM137929	4 (primary)	Colon
GSM117760	4 (primary)	Colon
GSM137942	4 (primary)	Colon
GSM137998	4 (primary)	Colon
GSM138036	4 (primary)	Colon
GSM138042	4 (primary)	Colon
GSM53132	4 (primary)	Rectosigmoid
GSM102535	4 (primary)	Rectosigmoid
GSM117734	4 (primary)	Rectosigmoid
GSM137985	4 (primary)	Rectosigmoid

Table S1 (continued)

GEO accession	Pathological stage	Location
GSM46819	4 (primary)	Rectum
GSM88976	4 (primary)	Rectum
GSM76571	4 (primary)	Rectum
GSM152692	4 (primary)	Colon
GSM179778	4 (primary)	Colon
GSM179805	4 (primary)	Colon
GSM179867	4 (primary)	Colon
GSM179925	4 (primary)	Colon
GSM179928	4 (primary)	Colon
GSM179948	4 (primary)	Colon
GSM203657	4 (primary)	Colon
GSM203688	4 (primary)	Colon
GSM203691	4 (primary)	Rectosigmoid
GSM203724	4 (primary)	Colon
GSM203743	4 (primary)	Colon
GSM203763	4 (primary)	Colon
GSM231904	4 (primary)	Colon
GSM231959	4 (primary)	Colon
GSM231964	4 (primary)	Colon
GSM277731	4 (primary)	Colon
GSM325829	4 (primary)	Colon
GSM353895	4 (primary)	Appendix
GSM152710	4 (primary)	Rectosigmoid
GSM152750	4 (primary)	Rectosigmoid
GSM179833	4 (primary)	Rectum
GSM231934	4 (primary)	Rectum
GSM325822	4 (primary)	Rectum
GSM353938	4 (primary)	Colon
GSM353922	4 (primary)	Colon
GSM353934	4 (primary)	Colon
GSM88946	4 (metastatic)	Liver (primary: colon)
GSM53088	4 (metastatic)	Liver (primary: colon)
GSM53090	4 (metastatic)	Liver (primary: colon)
GSM46970	4 (metastatic)	Liver (primary: rectosigmoid)
GSM46971	4 (metastatic)	Liver (primary: colon)
GSM89030	4 (metastatic)	Liver (primary: colon)
GSM137897	4 (metastatic)	Liver (primary: rectosigmoid)
GSM102531	4 (metastatic)	Liver (primary: colon)
GSM137921	4 (metastatic)	Liver (primary: rectum)
GSM137933	4 (metastatic)	Liver (primary: colon)
GSM137935	4 (metastatic)	Liver (primary: rectum)
GSM138043	4 (metastatic)	Liver (primary: colon)
GSM46828	4 (metastatic)	Liver or omentum (primary: colon)
GSM76540	4 (metastatic)	Omentum (primary: colon)
GSM137952	4 (metastatic)	Peritoneum (primary: colon)
GSM76625	4 (metastatic)	Abdominal mass (primary: rectum)
GSM76574	4 (metastatic)	Lung (primary: colon)
GSM138047	4 (metastatic)	Lung (primary: colon)
GSM152584	4 (metastatic)	Liver (primary: colon)
GSM152612	4 (metastatic)	Liver (primary: colon)
GSM152626	4 (metastatic)	Liver (primary: colon)
GSM152708	4 (metastatic)	Liver (primary: colon)
GSM152760	4 (metastatic)	Liver (primary: colon)
GSM152765	4 (metastatic)	Liver (primary: colon)
GSM179840	4 (metastatic)	Liver (primary: colon)
GSM179865	4 (metastatic)	Liver (primary: colon)
GSM203704	4 (metastatic)	Liver (primary: colon)
GSM203775	4 (metastatic)	Liver (primary: colon)
GSM231883	4 (metastatic)	Liver (primary: rectum)
GSM277687	4 (metastatic)	Liver (primary: colon)
GSM325819	4 (metastatic)	Liver (primary: rectosigmoid)
GSM152592	4 (metastatic)	Lung (primary: colon)
GSM353897	4 (metastatic)	Lung (primary: colon)
GSM301654	4 (metastatic)	Omentum (primary: rectum)

Table S1 (*continued*)

Table S2 Significant gene ontology and pathway identified in overlapping gene list of PRI data sets

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_CC_ALL	GO:0031090~organelle membrane	20	14.81481	6.00E-04	MTDH, COX10, RER1, VPS41, COX5A, GJB2, HOOK1, ST3GAL1, SLC25A11, SDHB, POMGNT1, ALOX5AP, DNAJC11, MPDU1, ATP5A1, RAB27B, ANKFY1, CLN5, HIP1, SCO1	122	1,096	15,908	2.379442	0.129449	0.129449	0.766965
GOTERM_CC_ALL	GO:0044422~organelle part	49	36.29630	0.001267	E2F2, HMGN2, CDC14A, COX10, COX5A, WBP4, LATS2, ST3GAL1, HOOK1, POMGNT1, CASP3, PTGIS, DNAJC11, MRPL37, RAB27B, SNHG3-RCC1, KCND2, LOC729687, LOC729505, VPS41, NOC2L, C1QBP, FKBP14, ANKFY1, LOC648822, CLN5, MTDH, BBS9, RER1, RCC1, NCAPG, PTK2B, ALOX5AP, MPDU1, DHX15, RPS20, SKA1, GEMIN4, HIP1, SCO1, BFSP1, IDH3A, DOCK4, GJB2, SLC25A11, SDHB, TMEM48, SFPQ, SULF1, ATP5A1, XPO7, KATNAL1, NCOR1	122	4,251	15,908	1.503006	0.253855	0.136203	1.613195
GOTERM_BP_ALL	GO:0043623~cellular protein complex assembly	7	5.185185	0.00152	PTK2, TMEM48, COX10, PTK2B, XPO7, SCO1, HIP1	109	162	14,116	5.595877	0.826941	0.826941	2.417597
GOTERM_CC_ALL	GO:0044446~intracellular organelle part	48	35.555560	0.002054	E2F2, HMGN2, CDC14A, COX10, COX5A, WBP4, LATS2, HOOK1, ST3GAL1, CASP3, POMGNT1, PTGIS, DNAJC11, MRPL37, RAB27B, SNHG3-RCC1, KCND2, LOC729687, LOC729505, VPS41, NOC2L, C1QBP, FKBP14, ANKFY1, CLN5, LOC648822, MTDH, BBS9, RER1, RCC1, NCAPG, PTK2B, ALOX5AP, MPDU1, DHX15, RPS20, SKA1, GEMIN4, SCO1, HIP1, BFSP1, IDH3A, GJB2, SLC25A11, SDHB, TMEM48, SFPQ, SULF1, ATP5A1, KATNAL1, XPO7, NCOR1	122	4,225	15,908	1.481393	0.378152	0.146452	2.603889

Table S2 (*continued*)

Table S2 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0009987~cellular process	93	68.88889	0.005836	CTHRC1, HMGN2, POSTN, ST3GAL1, HOOK1, PTGIS, PHTF2, MRPL37, KCND2, ZHX2, LYPLA2, VPS41, EEPD1, COL1A2, PCMTD1, LOC648822, NMNAT1, TSHZ2, ADORA2B, BBS9, RER1, EPHA10, RCC1, AGMAT, ITGBL1, PTK2, ITGB8, RPS20, SKA1, HIP1, PLAG1, ZNF623, ABR, KLF12, EPHX2, C18ORF55, IDH3A, PTPN12, SLC25A11, LAMP1, GGT6, ATP2A3, SFRP4, SLC13A2, FBXO32, ZNF117, XPO7, NCOR1, E2F2, CLDN7, AEBP1, CDC14A, ZFP41, NARS, COX10, MFSD2A, WBP4, LATS2, POMGNT1, CASP3, TNFRSF11A, AGPAT5, LYPLA2P1, SNHG3-RCC1, CDK14, CDK13, LOC729687, RUNX1T1, LOC729505, UBE2H, PBK, INHBA, FKBP14, CLN5, CYTH3, PR47, NCAPG, PTK2B, ALOX5AP, MPDU1, DHX15, ETV1, GPNMB, GEMIN4, SCO1, LOC100131909, DFNA5, FLT1, C17ORF81, NADK, GJB2, SDHB, ADNP2, TMEM48, SFPQ, SULF1, ATP5A1, IGFBP3, HTR2A	109	10,541	14,116	1.142579	0.998827	0.96575	8.985215
GOTERM_BP_ALL	GO:0034622~cellular macromolecular complex assembly	8	5.925926	0.011076	PTK2, TMEM48, COX10, PTK2B, XPO7, GEMIN4, SCO1, HIP1	109	318	14,116	3.257977	0.999997	0.986167	16.40343
GOTERM_CC_ALL	GO:0044444~cytoplasmic part	50	37.03704	0.014368	CDC14A, NARS, COX10, COX5A, LATS2, ST3GAL1, HOOK1, CASP3, POMGNT1, ACOT7, PTGIS, AGPAT5, DNAJC11, PHTF2, MRPL37, RAB27B, CDK14, VPS41, C1QBP, FKBP14, ANKFY1, CLN5, MTDH, BBS9, RER1, AGMAT, PTK2, PTK2B, ALOX5AP, MPDU1, RPS20, GPNMB, GEMIN4, HIP1, SCO1, ABR, NAT1, NAT2, EPHX2, NADK, C18ORF55, IDH3A, PTPN12, GJB2, SLC25A11, LAMP1, SDHB, ATP2A3, SULF1, ATP5A1	122	4,895	15,908	1.331904	0.964675	0.56647	16.94552
GOTERM_BP_ALL	GO:0006461~protein complex assembly	10	7.407407	0.015299	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, SCO1, HIP1	109	505	14,116	2.564447	1	0.988252	21.96471

Table S2 (*continued*)

Table S2 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0070271~protein complex biogenesis	10	7.407407	0.015299	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, SCO1, HIP1	109	505	14,116	2.564447	1	0.988252	21.96471
GOTERM_BP_ALL	GO:0019674~NAD metabolic process	3	2.222222	0.015514	NADK, IDH3A, NMNAT1	109	25	14,116	15.54055	1	0.972827	22.23794
GOTERM_BP_ALL	GO:0006793~phosphorus metabolic process	15	11.11111	0.016218	LOC100131909, FLT1, ADORA2B, CDC14A, COX10, NADK, EPHA10, PBK, LATS2, PTPN12, PTK2, PTK2B, ATP5A1, IGFBP3, CDK14, CDK13	109	973	14,116	1.996474	1	0.956810	23.12810
GOTERM_BP_ALL	GO:0006796~phosphate metabolic process	15	11.11111	0.016218	LOC100131909, FLT1, ADORA2B, CDC14A, COX10, NADK, EPHA10, PBK, LATS2, PTPN12, PTK2, PTK2B, ATP5A1, IGFBP3, CDK14, CDK13	109	973	14,116	1.996474	1	0.956810	23.12810
GOTERM_BP_ALL	GO:0051186~cofactor metabolic process	6	4.444444	0.016961	SDHB, GGT6, COX10, NADK, IDH3A, NMNAT1	109	195	14,116	3.984757	1	0.940259	24.05688
GOTERM_BP_ALL	GO:0044237~cellular metabolic process	63	46.66667	0.018978	E2F2, AEBP1, NARS, ZFP41, CDC14A, COX10, WBP4, LATS2, ST3GAL1, CASP3, POMGNT1, PTGIS, AGPAT5, LYPLA2P1, PHTF2, MRPL37, CDK14, CDK13, RUNX1T1, ZHX2, LYPLA2, PBK, UBE2H, EEPD1, INHBA, FKBP14, PCMTD1, NMNAT1, TSHZ2, ADORA2B, EPHA10, AGMAT, PTK2, ITGB8, PTK2B, ALOX5AP, MPDU1, DHX15, ETV1, RPS20, GEMIN4, SCO1, HIP1, PLAG1, LOC100131909, ZNF623, FLT1, KLF12, C17ORF81, EPHX2, NADK, PTPN12, IDH3A, SDHB, LAMP1, ADNP2, GGT6, ATP2A3, SFPQ, SULF1, FBXO32, ATP5A1, ZNF117, NCOR1, IGFBP3	109	6,636	14,116	1.229474	1	0.936805	26.52489
GOTERM_BP_ALL	GO:0016310~phosphorylation	13	9.62963	0.01912	LOC100131909, FLT1, ADORA2B, COX10, NADK, EPHA10, PBK, LATS2, PTK2, PTK2B, ATP5A1, IGFBP3, CDK14, CDK13	109	800	14,116	2.10445	1	0.915686	26.69555
GOTERM_BP_ALL	GO:0034621~cellular macromolecular complex subunit organization	8	5.925926	0.019726	PTK2, TMEM48, COX10, PTK2B, XPO7, GEMIN4, SCO1, HIP1	109	357	14,116	2.902064	1	0.899458	27.42109
GOTERM_CC_ALL	GO:0031967~organelle envelope	11	8.148148	0.019855	SLC25A11, SDHB, TMEM48, MTDH, COX10, DNAJC11, ALOX5AP, ATP5A1, XPO7, COX5A, SCO1	122	620	15,908	2.313432	0.990271	0.604075	22.68551

Table S2 (*continued*)

Table S2 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_CC_ALL	GO:0031975~envelope	11	8.148148	0.020253	SLC25A11, SDHB, TMEM48, MTDH, COX10, DNAJC11, ALOX5AP, ATP5A1, XPO7, COX5A, SCO1	122	622	15,908	2.305993	0.991143	0.545135	23.08753
GOTERM_MF_ALL	GO:0004060~arylamine N-acetyltransferase activity	2	1.481481	0.022807	NAT1, NAT2	117	3	15,143	86.28490	0.999716	0.999716	27.09909
GOTERM_BP_ALL	GO:0045786~negative regulation of cell cycle	4	2.962963	0.024155	INHBA, CASP3, AURKAIP1, LATS2	109	81	14,116	6.395288	1	0.922921	32.51926
GOTERM_BP_ALL	GO:0008152~metabolic process	70	51.85185	0.025642	E2F2, AEBP1, NARS, ZFP41, CDC14A, COX10, COX5A, WBP4, LATS2, ST3GAL1, ACOT7, CASP3, POMGNT1, PTGIS, AGPAT5, LYPLA2P1, PHTF2, MRPL37, CDK14, CDK13, ZHX2, RUNX1T1, LYPLA2, MMP16, PBK, UBE2H, EEPD1, INHBA, PCMTD1, FKBP14, CLN5, NMNAT1, TSHZ2, ADORA2B, EPHA10, AGMAT, PTK2, ITGB8, PTK2B, ALOX5AP, MPDU1, DHX15, ETV1, RPS20, PLTP, GEMIN4, SCO1, HIP1, PLAG1, LOC100131909, ZNF623, FLT1, KLF12, C17ORF81, NAT1, NAT2, EPHX2, NADK, PTPN12, IDH3A, SDHB, LAMP1, ADNP2, GGT6, ATP2A3, SFPQ, SULF1, FBXO32, ATP5A1, ZNF117, NCOR1, IGFBP3	109	7,647	14,116	1.185474	1	0.917576	34.15462

Table S2 (*continued*)

Table S2 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_MF_ALL	GO:0005488~binding	105	77.77778	0.026588	HMGN2, AURKAIP1, POSTN, PI4K2B, COX5A, NOV, HOOK1, PTGIS, PHTF2, RAB27B, KCND2, ZHX2, VPS41, EEPD1, COL1A2, ANKFY1, LOC648822, NMNAT1, TSHZ2, EPHA10, AGMAT, RCC1, ITGBL1, IL17D, PTK2, ITGB8, SKA1, RPS20, PLTP, HIP1, PLAG1, ZNF623, ABR, KLF12, BFSP1, EPHX2, PTPN12, IDH3A, KCTD7, DOCK4, SLC25A11, ATP2A3, SFRP4, SLC13A2, FBXO32, XPO7, ZNF117, NCOR1, E2F2, CLDN7, AEBP1, NARS, ZFP41, CDC14A, LRRC17, WBP4, LATS2, ACOT7, CASP3, POMGNT1, TNFRSF11A, GPC6, DNAJC11, SNHG3-RCC1, CDK14, CDK13, LOC729687, RUNX1T1, LOC729505, SEL1L3, MMP16, MBNL2, PBK, UBE2H, NOC2L, INHBA, C1QBP, DOK5, FKBP14, CLN5, COMMD4, MTDH, CYTH3, PR47, PTK2B, NCAPG, ALOX5AP, MPDU1, DHX15, ETV1, PHF20L1, RASA3, GPNMB, GEMIN4, SCO1, DFNA5, FLT1, C17ORF81, NADK, GJB2, ELFN1, SDHB, ADNP2, SFPQ, SULF1, ATP5A1, KATNAL1, IGFBP3, HTR2A	117	12,531	15,143	1.084500	0.999928	0.991517	30.86992
GOTERM_BP_ALL	GO:0009060~aerobic respiration	3	2.222222	0.029287	SDHB, COX10, IDH3A	109	35	14,116	11.100390	1	0.928374	38.00702
GOTERM_BP_ALL	GO:0006732~coenzyme metabolic process	5	3.703704	0.029874	SDHB, GGT6, NADK, IDH3A, NMNAT1	109	153	14,116	4.232176	1	0.917735	38.60738
GOTERM_BP_ALL	GO:0065003~macromolecular complex assembly	11	8.148148	0.030977	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, GEMIN4, SCO1, HIP1	109	665	14,116	2.142181	1	0.910970	39.72136
GOTERM_BP_ALL	GO:0048285~organelle fission	6	4.444444	0.031218	COX10, NCAPG, SKA1, PBK, RCC1, SNHG3-RCC1, LATS2	109	229	14116	3.393133	1	0.898280	39.96193
GOTERM_BP_ALL	GO:0001501~skeletal system development	7	5.185185	0.03574	INHBA, AEBP1, TNFRSF11A, COL1A2, POSTN, GPNMB, IGFBP3	109	319	14,116	2.841793	1	0.915280	44.31463
GOTERM_CC_ALL	GO:0005739~mitochondrion	15	11.11111	0.036535	COX10, AGMAT, COX5A, C18ORF55, IDH3A, SLC25A11, SDHB, ACOT7, CASP3, AGPAT5, C1QBP, DNAJC11, MRPL37, ATP5A1, SCO1	122	1,087	15,908	1.799358	0.999815	0.707189	37.96641

Table S2 (*continued*)

Table S2 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0046496~nicotinamide nucleotide metabolic process	3	2.222222	0.037460	NADK, IDH3A, NMNAT1	109	40	14,116	9.712844	1	0.91333	45.89132
GOTERM_BP_ALL	GO:0006769~nicotinamide metabolic process	3	2.222222	0.037460	NADK, IDH3A, NMNAT1	109	40	14,116	9.712844	1	0.91333	45.89132
GOTERM_CC_ALL	GO:0044429~mitochondrial part	10	7.407407	0.037762	SLC25A11, SDHB, C1QBP, COX10, DNAJC11, MRPL37, ATP5A1, COX5A, IDH3A, SCO1	122	595	15,908	2.191486	0.999863	0.670935	38.97241
GOTERM_BP_ALL	GO:0009820~alkaloid metabolic process	3	2.222222	0.039189	NADK, IDH3A, NMNAT1	109	41	14,116	9.475945	1	0.911609	47.43319
GOTERM_BP_ALL	GO:0019362~pyridine nucleotide metabolic process	3	2.222222	0.040947	NADK, IDH3A, NMNAT1	109	42	14,116	9.250328	1	0.910209	48.95936
KEGG_PATHWAY	hsa05010: Alzheimer's disease	5	3.703704	0.041059	SDHB, CASP3, ATP2A3, ATP5A1, COX5A	42	163	5,085	3.713848	0.944585	0.944585	35.11945
GOTERM_BP_ALL	GO:0051726~regulation of cell cycle	7	5.185185	0.041571	E2F2, INHBA, CASP3, AURKAIP1, RCC1, SNHG3-RCC1, LATS2, CDK13	109	331	14,116	2.738768	1	0.902827	49.49135
GOTERM_BP_ALL	GO:0043933~macromolecular complex subunit organization	11	8.148148	0.045163	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, GEMIN4, SCO1, HIP1	109	710	14,116	2.006409	1	0.911261	52.45137
GOTERM_MF_ALL	GO:0019838~growth factor binding	4	2.962963	0.046779	NOV, FLT1, COL1A2, IGFBP3	117	105	15,143	4.930566	1	0.996494	48.12558

Table S3 Significant gene ontology and pathway identified in overlapping gene list of META data sets

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0032101~regulation of response to external stimulus	6	10.71429	1.97E-04	PDPN, TGM2, F7, LBP, GREM1, CPB2	49	159	14,116	10.871010	0.153495	0.153495	0.304425
GOTERM_CC_ALL	GO:0005615~extracellular space	10	17.85714	2.90E-04	ATRN, PON1, RELN, IGFBP1, LBP, GREM1, CPB2, PON3, EPO, WNT2B	52	685	15,908	4.466030	0.041415	0.041415	0.343750
GOTERM_BP_ALL	GO:0009966~regulation of signal transduction	11	19.64286	6.18E-04	ZGPAT, F10, BIRC7, TGM2, TRIB3, RELN, HGF, F7, LBP, GREM1, EPO	49	878	14,116	3.609223	0.407067	0.229978	0.951741
GOTERM_BP_ALL	GO:0009605~response to external stimulus	11	19.64286	8.45E-04	F10, PDPN, PROZ, ATRN, PON1, F7, IGFBP1, LBP, ASL, PON3, EPO	49	914	14,116	3.467066	0.510532	0.211912	1.298649
KEGG_PATHWAY	hsa00983: drug metabolism	4	7.142857	8.87E-04	CYP3A4, CYP3A7, GUSB, DPYS	24	43	5,085	19.709300	0.037428	0.037428	0.822167

Table S3 (*continued*)

Table S3 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_MF_ALL	GO:0016787~hydrolase activity	18	32.14286	8.87E-04	DNAH11, F10, GUSB, TFR2, DPYS, ABCB1, F7, HGF, RBBP9, ABCB4, AZGP1P1, AZGP1, PROZ, PON1, TGM2, RELN, CPB2, PON3, CHD3	52	2,283	15,143	2.296017	0.188912	0.188912	1.135601
GOTERM_BP_ALL	GO:0042221~response to chemical stimulus	13	23.21429	9.43E-04	CYP3A4, PDPN, ATRN, ABCB1, ASL, ABCB4, PCK1, SLC25A13, SLC23A2, PON1, IGFBP1, LBP, EPO	49	1,281	14,116	2.923545	0.549594	0.180779	1.448729
GOTERM_CC_ALL	GO:0005792~microsome	6	10.71429	9.46E-04	CYP3A4, F10, CYP3A7, PON1, POR, PON3	52	237	15,908	7.744888	0.129023	0.066739	1.11829
GOTERM_CC_ALL	GO:0042598~vesicular fraction	6	10.71429	0.001077	CYP3A4, F10, CYP3A7, PON1, POR, PON3	52	244	15,908	7.522699	0.14557	0.051089	1.272576
GOTERM_CC_ALL	GO:0005576~extracellular region	16	28.57143	0.001176	F10, ATRN, COLEC10, F7, HGF, GREM1, WNT2B, AZGP1P1, AZGP1, PROZ, PON1, RELN, IGFBP1, LBP, CPB2, PON3, EPO	52	2,010	15,908	2.435209	0.157857	0.042042	1.388926
GOTERM_BP_ALL	GO:0009611~response to wounding	8	14.28571	0.001956	F10, PDPN, PROZ, ATRN, F7, IGFBP1, LBP, EPO	49	530	14,116	4.348402	0.808812	0.281723	2.981764
GOTERM_MF_ALL	GO:0003824~catalytic activity	29	51.78571	0.002011	CYP3A4, DNAH11, CYP3A7, TFR2, DPYS, TRIB3, ASL, CLYBL, AZGP1P1, AZGP1, ACSS1, PROZ, TGM2, CHD3, CRLS1, F10, GUSB, GGTLC1, ABCB1, F7, HGF, RBBP9, POR, PCK1, ABCB4, HAO1, PON1, RELN, CPB2, PON3	52	5,198	15,143	1.624689	0.378231	0.211477	2.558699
GOTERM_CC_ALL	GO:0000267~cell fraction	11	19.64286	0.00202	CYP3A4, F10, CYP3A7, SLC23A2, PON1, DPYS, ABCB1, POR, PON3, PCK1, ABCB4	52	1,083	15,908	3.107252	0.255592	0.057324	2.374291
GOTERM_BP_ALL	GO:0010646~regulation of cell communication	11	19.64286	0.002212	ZGPAT, F10, BIRC7, TGM2, TRIB3, RELN, HGF, F7, LBP, GREM1, EPO	49	1,038	14,116	3.052888	0.846102	0.267954	3.366137
GOTERM_BP_ALL	GO:0043436~oxoacid metabolic process	8	14.28571	0.002563	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	556	14,116	4.145059	0.885687	0.266429	3.890425
GOTERM_BP_ALL	GO:0019752~carboxylic acid metabolic process	8	14.28571	0.002563	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	556	14,116	4.145059	0.885687	0.266429	3.890425
GOTERM_BP_ALL	GO:0006082~organic acid metabolic process	8	14.28571	0.002668	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	560	14,116	4.115452	0.895426	0.245901	4.04688
GOTERM_BP_ALL	GO:0042180~cellular ketone metabolic process	8	14.28571	0.00286	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	567	14,116	4.064644	0.911122	0.235813	4.331966
GOTERM_MF_ALL	GO:0008236~serine-type peptidase activity	5	8.928571	0.002999	F10, PROZ, RELN, HGF, F7	52	178	15,143	8.180099	0.507787	0.210442	3.792744

Table S3 (*continued*)

Table S3 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_MF_ALL	GO:0017171~serine hydrolase activity	5	8.928571	0.003122	F10, PROZ, RELN, HGF, F7	52	180	15,143	8.089209	0.521896	0.168465	3.945245
GOTERM_CC_ALL	GO:0044421~extracellular region part	10	17.85714	0.00314	ATRN, PON1, RELN, IGFBP1, LBP, GREM1, CPB2, PON3, EPO, WNT2B	52	960	15,908	3.186699	0.368208	0.073677	3.669279
GOTERM_BP_ALL	GO:0009967~positive regulation of signal transduction	6	10.71429	0.003164	F10, TGM2, RELN, F7, LBP, EPO	49	295	14,116	5.859287	0.931279	0.234916	4.781126
GOTERM_CC_ALL	GO:0005624~membrane fraction	9	16.07143	0.003942	CYP3A4, F10, CYP3A7, SLC23A2, PON1, ABCB1, POR, PON3, ABCB4	52	809	15,908	3.403347	0.438256	0.079085	4.58646
GOTERM_BP_ALL	GO:0048583~regulation of response to stimulus	7	12.50000	0.004701	PDPN, BIRC7, TGM2, F7, LBP, GREM1, CPB2	49	465	14,116	4.336713	0.981343	0.303688	7.02573
GOTERM_BP_ALL	GO:0042325~regulation of phosphorylation	7	12.50000	0.00475	ZGPAT, BIRC7, TRIB3, RELN, HGF, SDC4, EPO	49	466	14,116	4.327406	0.982106	0.284857	7.096713
GOTERM_CC_ALL	GO:0005626~insoluble fraction	9	16.07143	0.004909	CYP3A4, F10, CYP3A7, SLC23A2, PON1, ABCB1, POR, PON3, ABCB4	52	839	15,908	3.281654	0.512474	0.085887	5.680823
GOTERM_BP_ALL	GO:0010647~positive regulation of cell communication	6	10.71429	0.005027	F10, TGM2, RELN, F7, LBP, EPO	49	329	14,116	5.253768	0.985858	0.279339	7.495873
GOTERM_BP_ALL	GO:0006629~lipid metabolic process	9	16.07143	0.005673	CYP3A4, HAO1, AZGP1P1, AZGP1, CRLS1, PDPN, PON1, NPC1L1, PCK1, ABCB4	49	813	14,116	3.189096	0.99183	0.290632	8.419917
GOTERM_BP_ALL	GO:0019220~regulation of phosphate metabolic process	7	12.50000	0.005758	ZGPAT, BIRC7, TRIB3, RELN, HGF, SDC4, EPO	49	485	14,116	4.157879	0.992403	0.27771	8.541496
GOTERM_BP_ALL	GO:0051174~regulation of phosphorus metabolic process	7	12.50000	0.005758	ZGPAT, BIRC7, TRIB3, RELN, HGF, SDC4, EPO	49	485	14,116	4.157879	0.992403	0.27771	8.541496
GOTERM_BP_ALL	GO:0045859~regulation of protein kinase activity	6	10.71429	0.006131	ZGPAT, BIRC7, TRIB3, RELN, HGF, SDC4	49	345	14,116	5.010115	0.994463	0.277309	9.069492
GOTERM_BP_ALL	GO:0046395~carboxylic acid catabolic process	4	7.142857	0.00633	HAO1, PON1, ASL, PON3	49	111	14,116	10.38132	0.995326	0.27067	9.350749
GOTERM_BP_ALL	GO:0016054~organic acid catabolic process	4	7.142857	0.00633	HAO1, PON1, ASL, PON3	49	111	14,116	10.38132	0.995326	0.27067	9.350749
GOTERM_BP_ALL	GO:0043549~regulation of kinase activity	6	10.71429	0.007064	ZGPAT, BIRC7, TRIB3, RELN, HGF, SDC4	49	357	14,116	4.841708	0.997496	0.283071	10.38029
GOTERM_BP_ALL	GO:0050896~response to stimulus	21	37.50000	0.007551	CYP3A4, F10, PDPN, ATRN, BIRC7, TRIB3, ABCB1, F7, ASL, PCK1, ABCB4, AZGP1P1, AZGP1, SLC25A13, SLC23A2, PROZ, PON1, RELN, LBP, IGFBP1, PON3, EPO	49	3,502	14,116	1.727503	0.998346	0.286145	11.05728

Table S3 (*continued*)

Table S3 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0051338~regulation of transferase activity	6	10.71429	0.008365	ZGPAT, BIRC7, TRIB3, RELN, HGF, SDC4	49	372	14,116	4.646478	0.999173	0.298764	12.17915
GOTERM_BP_ALL	GO:0065008~regulation of biological quality	12	21.42857	0.009005	F10, PDPN, TFR2, PROZ, TGM2, NPC1L1, RELN, F7, IGFBP1, GREM1, PCK1, EPO	49	1,469	14,116	2.353288	0.999521	0.305104	13.05151
GOTERM_BP_ALL	GO:0006950~response to stress	13	23.21429	0.009225	F10, PDPN, ATRN, BIRC7, TRIB3, F7, ASL, SLC23A2, PROZ, RELN, IGFBP1, LBP, EPO	49	1,685	14,116	2.222588	0.999603	0.299511	13.3495
GOTERM_BP_ALL	GO:0010627~regulation of protein kinase cascade	5	8.928571	0.010015	F10, BIRC7, TGM2, F7, EPO	49	249	14,116	5.784772	0.999798	0.309138	14.41188
GOTERM_MF_ALL	GO:0004063~aryldialkylphosphatase activity	2	3.571429	0.01007	PON1, PON3	52	3	15,143	194.141	0.908247	0.37981	12.21627
GOTERM_MF_ALL	GO:0008559~xenobiotic-transporting ATPase activity	2	3.571429	0.01007	ABCB1, ABCB4	52	3	15,143	194.141	0.908247	0.37981	12.21627
GOTERM_MF_ALL	GO:0015239~multidrug transporter activity	2	3.571429	0.01007	ABCB1, ABCB4	52	3	15,143	194.141	0.908247	0.37981	12.21627
GOTERM_BP_ALL	GO:0006725~cellular aromatic compound metabolic process	4	7.142857	0.01103	PON1, DPYS, CLYBL, PON3	49	136	14,116	8.472989	0.999915	0.323286	15.75795
GOTERM_MF_ALL	GO:0008233~peptidase activity	7	12.50000	0.012243	F10, TFR2, PROZ, RELN, HGF, F7, CPB2	52	574	15,143	3.55136	0.945365	0.384003	14.66392
GOTERM_MF_ALL	GO:0042910~xenobiotic transporter activity	2	3.571429	0.013405	ABCB1, ABCB4	52	4	15,143	145.6058	0.958621	0.36555	15.94771
GOTERM_MF_ALL	GO:0004064~arylesterase activity	2	3.571429	0.013405	PON1, PON3	52	4	15,143	145.6058	0.958621	0.36555	15.94771
GOTERM_BP_ALL	GO:0080134~regulation of response to stress	5	8.928571	0.013839	PDPN, BIRC7, TGM2, LBP, CPB2	49	274	14,116	5.256964	0.999992	0.375645	19.38261
GOTERM_BP_ALL	GO:0044093~positive regulation of molecular function	7	12.50000	0.013920	BIRC7, PON1, TGM2, DPH3B, RELN, HGF, DPH3, SDC4	49	586	14,116	3.441248	0.999993	0.365921	19.48461
GOTERM_MF_ALL	GO:0004252~serine-type endopeptidase activity	4	7.142857	0.015041	F10, PROZ, HGF, F7	52	154	15,143	7.563936	0.972033	0.360514	17.72478
GOTERM_BP_ALL	GO:0065009~regulation of molecular function	9	16.07143	0.015548	ZGPAT, BIRC7, PON1, TGM2, DPH3B, TRIB3, RELN, HGF, DPH3, SDC4	49	969	14,116	2.675681	0.999998	0.387636	21.51572
GOTERM_BP_ALL	GO:0007598~blood coagulation, extrinsic pathway	2	3.571429	0.016889	F10, F7	49	5	14,116	115.2327	0.999999	0.401927	23.15213
GOTERM_BP_ALL	GO:0010740~positive regulation of protein kinase cascade	4	7.142857	0.019064	F10, TGM2, F7, EPO	49	167	14,116	6.900159	1	0.429273	25.73869
GOTERM_BP_ALL	GO:0030334~regulation of cell migration	4	7.142857	0.019669	F10, PDPN, F7, GREM1	49	169	14,116	6.8185	1	0.428517	26.44338

Table S3 (*continued*)

Table S3 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0032103~positive regulation of response to external stimulus	3	5.357143	0.019965	TGM2, F7, LBP	49	64	14,116	13.50383	1	0.422883	26.78629
GOTERM_CC_ALL	GO:0009986~cell surface	5	8.928571	0.025118	CYP3A4, F10, ABCB1, GREM1, SDC4	52	348	15,908	4.395447	0.975622	0.338122	26.09285
GOTERM_BP_ALL	GO:0042060~wound healing	4	7.142857	0.027021	F10, PROZ, F7, IGFBP1	49	191	14,116	6.033123	1	0.514867	34.52473
GOTERM_BP_ALL	GO:0040012~regulation of locomotion	4	7.142857	0.027385	F10, PDPN, F7, GREM1	49	192	14,116	6.001701	1	0.508853	34.90311
GOTERM_BP_ALL	GO:0051270~regulation of cell motion	4	7.142857	0.027753	F10, PDPN, F7, GREM1	49	193	14,116	5.970604	1	0.50316	35.28216
GOTERM_BP_ALL	GO:0006916~anti-apoptosis	4	7.142857	0.032768	BIRC7, TGM2, HGF, F7	49	206	14,116	5.593818	1	0.552631	40.25566
GOTERM_MF_ALL	GO:0022804~active transmembrane transporter activity	5	8.928571	0.033564	SLC25A13, PDPN, SLC23A2, ABCB1, ABCB4	52	363	15,143	4.011178	0.999683	0.591489	35.56371
GOTERM_MF_ALL	GO:0070011~peptidase activity, acting on L-amino acid peptides	6	10.71429	0.036755	F10, PROZ, RELN, HGF, F7, CPB2	52	549	15,143	3.18264	0.999855	0.586774	38.24937
GOTERM_BP_ALL	GO:0030335~positive regulation of cell migration	3	5.357143	0.036762	F10, PDPN, F7	49	89	14,116	9.710617	1	0.584861	43.95765
GOTERM_BP_ALL	GO:0051897~positive regulation of protein kinase B signaling cascade	2	3.571429	0.036788	F10, F7	49	11	14,116	52.37848	1	0.575139	43.98085
GOTERM_BP_ALL	GO:0002688~regulation of leukocyte chemotaxis	2	3.571429	0.036788	F7, GREM1	49	11	14,116	52.37848	1	0.575139	43.98085
KEGG_PATHWAY	hsa04610: complement and coagulation cascades	3	5.357143	0.038219	F10, F7, CPB2	24	69	5,085	9.211957	0.81281	0.567345	30.41609
GOTERM_CC_ALL	GO:0031226~intrinsic to plasma membrane	9	16.07143	0.038235	F10, SLC25A13, GPR37, PDPN, SLC23A2, TFR2, ATRN, SDC4, ABCB4	52	1,215	15,908	2.266097	0.996626	0.434009	37.08374
GOTERM_MF_ALL	GO:0009055~electron carrier activity	4	7.142857	0.038258	CYP3A4, HAO1, CYP3A7, POR	52	221	15,143	5.270797	0.9999	0.566955	39.47803
GOTERM_MF_ALL	GO:0051183~vitamin transporter activity	2	3.571429	0.039689	PDPN, SLC23A2	52	12	15,143	48.53526	0.999929	0.549075	40.62703
GOTERM_MF_ALL	GO:0010181~FMN binding	2	3.571429	0.039689	HAO1, POR	52	12	15,143	48.53526	0.999929	0.549075	40.62703
GOTERM_BP_ALL	GO:0045860~positive regulation of protein kinase activity	4	7.142857	0.039995	BIRC7, RELN, HGF, SDC4	49	223	14,116	5.167384	1	0.596527	46.79669
GOTERM_BP_ALL	GO:0055085~transmembrane transport	6	10.71429	0.043102	SLC25A13, PDPN, SLC23A2, ABCB1, SLC17A9, ABCB4	49	569	14,116	3.037768	1	0.61503	49.39667
GOTERM_BP_ALL	GO:0033674~positive regulation of kinase activity	4	7.142857	0.043655	BIRC7, RELN, HGF, SDC4	49	231	14,116	4.988427	1	0.610523	49.8474

Table S3 (*continued*)

Table S3 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0040017~positive regulation of locomotion	3	5.357143	0.043772	F10, PDPN, F7	49	98	14,116	8.818825	1	0.602464	49.94211
GOTERM_BP_ALL	GO:0051272~positive regulation of cell motion	3	5.357143	0.043772	F10, PDPN, F7	49	98	14,116	8.818825	1	0.602464	49.94211
GOTERM_BP_ALL	GO:0009308~amine metabolic process	5	8.928571	0.046497	GUSB, PON1, DPYS, RELN, ASL	49	400	14,116	3.60102	1	0.616309	52.10247
GOTERM_BP_ALL	GO:0050817~coagulation	3	5.357143	0.047036	F10, PROZ, F7	49	102	14,116	8.472989	1	0.611999	52.51919
GOTERM_BP_ALL	GO:0007596~blood coagulation	3	5.357143	0.047036	F10, PROZ, F7	49	102	14,116	8.472989	1	0.611999	52.51919
GOTERM_MF_ALL	GO:0043169~cation binding	21	37.50000	0.047704	CYP3A4, F10, CYP3A7, GUSB, COLEC10, BIRC7, DPYS, F7, DPH3, CLYBL, POR, PCK1, TRIM55, ZGPAT, SLC25A13, SLC23A2, PROZ, TGM2, DPH3B, RELN, CPB2, CHD3	52	4,179	15,143	1.463375	0.99999	0.588253	46.6997
GOTERM_BP_ALL	GO:0051347~positive regulation of transferase activity	4	7.142857	0.047967	BIRC7, RELN, HGF, SDC4	49	240	14,116	4.801361	1	0.610937	53.23178

Table S4 The most enriched cluster in overlapping gene sets of PRI data sets

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
Annotation Cluster 1	Enrichment score: 1.5497317826033175											
GOTERM_BP_ALL	GO:0043623~cellular protein complex assembly	7	5.185185	0.00152	PTK2, TMEM48, COX10, PTK2B, XPO7, SCO1, HIP1	109	162	14,116	5.595877	0.826941	0.826941	2.417597
GOTERM_BP_ALL	GO:0034622~cellular macromolecular complex assembly	8	5.925926	0.011076	PTK2, TMEM48, COX10, PTK2B, XPO7, GEMIN4, SCO1, HIP1	109	318	14,116	3.257977	0.999997	0.986167	16.40343
GOTERM_BP_ALL	GO:0006461~protein complex assembly	10	7.407407	0.015299	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, SCO1, HIP1	109	505	14,116	2.564447	1	0.988252	21.96471
GOTERM_BP_ALL	GO:0070271~protein complex biogenesis	10	7.407407	0.015299	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, SCO1, HIP1	109	505	14,116	2.564447	1	0.988252	21.96471
GOTERM_BP_ALL	GO:0034621~cellular macromolecular complex subunit organization	8	5.925926	0.019726	PTK2, TMEM48, COX10, PTK2B, XPO7, GEMIN4, SCO1, HIP1	109	357	14,116	2.902064	1	0.899458	27.42109
GOTERM_BP_ALL	GO:0065003~macromolecular complex assembly	11	8.148148	0.030977	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, GEMIN4, SCO1, HIP1	109	665	14,116	2.142181	1	0.91097	39.72136

Table S4 (*continued*)

Table S4 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
GOTERM_BP_ALL	GO:0022607~cellular component assembly	12	8.888889	0.076557	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, NCOR1, GEMIN4, SCO1, HIP1	109	887	14,116	1.752035	1	0.953162	72.2299
GOTERM_BP_ALL	GO:0044085~cellular component biogenesis	12	8.888889	0.14301	E2F2, PTK2, TMEM48, KCND2, COX10, PTK2B, ALOX5AP, XPO7, NCOR1, GEMIN4, SCO1, HIP1	109	1,001	14,116	1.552503	1	0.975452	91.64727
GOTERM_BP_ALL	GO:0016043~cellular component organization	23	17.03704	0.2672	E2F2, HMGN2, COX10, POSTN, RCC1, LATS2, HOOK1, PTK2, CASP3, PTK2B, NCAPG, ALOX5AP, SKA1, SNHG3-RCC1, GEMIN4, SCO1, HIP1, KCND2, LOC729687, LOC729505, PBK, TMEM48, COL1A2, XPO7, NCOR1, CLN5, LOC648822	109	2,498	14,116	1.192396	1	0.975161	99.32684

Table S5 The most enriched clusters in overlapping gene sets of META data sets

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
Annotation Cluster 1	Enrichment score: 2.9902367536607226											
GOTERM_CC_ALL	GO:0005615~extracellular space	10	17.85714	2.90E-04	ATRN, PON1, RELN, IGFBP1, LBP, GREM1, CPB2, PON3, EPO, WNT2B	52	685	15,908	4.46603	0.041415	0.041415	0.34375
GOTERM_CC_ALL	GO:0005576~extracellular region	16	28.57143	0.001176	F10, ATRN, COLEC10, F7, HGF, GREM1, WNT2B, AZGP1P1, AZGP1, PROZ, PON1, RELN, IGFBP1, LBP, CPB2, PON3, EPO	52	2,010	15,908	2.435209	0.157857	0.042042	1.388926
GOTERM_CC_ALL	GO:0044421~extracellular region part	10	17.85714	0.003140	ATRN, PON1, RELN, IGFBP1, LBP, GREM1, CPB2, PON3, EPO, WNT2B	52	960	15,908	3.186699	0.368208	0.073677	3.669279
Annotation Cluster 2	Enrichment score: 2.680013396917869											
GOTERM_CC_ALL	GO:0005792~microsome	6	10.71429	9.46E-04	CYP3A4, F10, CYP3A7, PON1, POR, PON3	52	237	15,908	7.744888	0.129023	0.066739	1.11829
GOTERM_CC_ALL	GO:0042598~vesicular fraction	6	10.71429	0.001077	CYP3A4, F10, CYP3A7, PON1, POR, PON3	52	244	15,908	7.522699	0.14557	0.051089	1.272576
GOTERM_CC_ALL	GO:0000267~cell fraction	11	19.64286	0.002020	CYP3A4, F10, CYP3A7, SLC23A2, PON1, DPYS, ABCB1, POR, PON3, PCK1, ABCB4	52	1,083	15,908	3.107252	0.255592	0.057324	2.374291
GOTERM_CC_ALL	GO:0005624~membrane fraction	9	16.07143	0.003942	CYP3A4, F10, CYP3A7, SLC23A2, PON1, ABCB1, POR, PON3, ABCB4	52	809	15,908	3.403347	0.438256	0.079085	4.58646
GOTERM_CC_ALL	GO:0005626~insoluble fraction	9	16.07143	0.004909	CYP3A4, F10, CYP3A7, SLC23A2, PON1, ABCB1, POR, PON3, ABCB4	52	839	15,908	3.281654	0.512474	0.085887	5.680823
Annotation Cluster 3	Enrichment score: 2.0785593471594135											
GOTERM_BP_ALL	GO:0032101~regulation of response to external stimulus	6	10.71429	1.97E-04	PDPN, TGM2, F7, LBP, GREM1, CPB2	49	159	14,116	10.87101	0.153495	0.153495	0.304425

Table S5 (*continued*)

Table S5 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
Annotation Cluster 3	Enrichment score: 2.0785593471594135											
GOTERM_BP_ALL	GO:0048583~regulation of response to stimulus	7	12.50000	0.004701	PDPN, BIRC7, TGM2, F7, LBP, GREM1, CPB2	49	465	14,116	4.336713	0.981343	0.303688	7.02573
GOTERM_BP_ALL	GO:0080134~regulation of response to stress	5	8.928571	0.013839	PDPN, BIRC7, TGM2, LBP, CPB2	49	274	14,116	5.256964	0.999992	0.375645	19.38261
GOTERM_BP_ALL	GO:0002682~regulation of immune system process	3	5.357143	0.378091	F7, LBP, GREM1	49	385	14,116	2.244792	1	0.989545	99.9353
Annotation Cluster 4	Enrichment score: 1.784332005122128											
GOTERM_BP_ALL	GO:0009605~response to external stimulus	11	19.64286	8.45E-04	F10, PDPN, PROZ, ATRN, PON1, F7, IGFBP1, LBP, ASL, PON3, EPO	49	914	14,116	3.467066	0.510532	0.211912	1.298649
GOTERM_BP_ALL	GO:0009611~response to wounding	8	14.28571	0.001956	F10, PDPN, PROZ, ATRN, F7, IGFBP1, LBP, EPO	49	530	14,116	4.348402	0.808812	0.281723	2.981764
GOTERM_BP_ALL	GO:0050896~response to stimulus	21	37.50000	0.007551	CYP3A4, F10, PDPN, ATRN, BIRC7, TRIB3, ABCB1, F7, ASL, PCK1, ABCB4, AZGP1P1, AZGP1, SLC25A13, SLC23A2, PROZ, PON1, RELN, LBP, IGFBP1, PON3, EPO	49	3,502	14,116	1.727503	0.998346	0.286145	11.05728
GOTERM_BP_ALL	GO:0006950~response to stress	13	23.21429	0.009225	F10, PDPN, ATRN, BIRC7, TRIB3, F7, ASL, SLC23A2, PROZ, RELN, IGFBP1, LBP, EPO	49	1,685	14,116	2.222588	0.999603	0.299511	13.3495
GOTERM_BP_ALL	GO:0048518~positive regulation of biological process	9	16.07143	0.386911	F10, PDPN, PON1, TGM2, RELN, F7, IGFBP1, LBP, EPO	49	2,033	14,116	1.275324	1	0.990391	99.94812
GOTERM_BP_ALL	GO:0048522~positive regulation of cellular process	8	14.28571	0.441705	F10, PDPN, TGM2, RELN, F7, IGFBP1, LBP, EPO	49	1,847	14,116	1.247782	1	0.994988	99.9878
Annotation Cluster 5	Enrichment score: 1.7119825191812028											
GOTERM_BP_ALL	GO:0032101~regulation of response to external stimulus	6	10.71429	1.97E-04	PDPN, TGM2, F7, LBP, GREM1, CPB2	49	159	14,116	10.87101	0.153495	0.153495	0.304425
GOTERM_BP_ALL	GO:0009967~positive regulation of signal transduction	6	10.71429	0.003164	F10, TGM2, RELN, F7, LBP, EPO	49	295	14,116	5.859287	0.931279	0.234916	4.781126
GOTERM_BP_ALL	GO:0010647~positive regulation of cell communication	6	10.71429	0.005027	F10, TGM2, RELN, F7, LBP, EPO	49	329	14,116	5.253768	0.985858	0.279339	7.495873
GOTERM_BP_ALL	GO:0010627~regulation of protein kinase cascade	5	8.928571	0.010015	F10, BIRC7, TGM2, F7, EPO	49	249	14,116	5.784772	0.999798	0.309138	14.41188

Table S5 (*continued*)

Table S5 (continued)

Category	Term	Count	%	P value	Genes	List total	Pop hits	Pop total	Fold enrichment	Bonferroni	Benjamini	FDR
Annotation Cluster 5	Enrichment score: 1.7119825191812028											
GOTERM_BP_ALL	GO:0010740~positive regulation of protein kinase cascade	4	7.142857	0.019064	F10, TGM2, F7, EPO	49	167	14,116	6.900159	1	0.440591	25.73869
GOTERM_BP_ALL	GO:0032103~positive regulation of response to external stimulus	3	5.357143	0.019965	TGM2, F7, LBP	49	64	14,116	13.50383	1	0.433361	26.78629
GOTERM_BP_ALL	GO:0048584~positive regulation of response to stimulus	3	5.357143	0.191360	TGM2, F7, LBP	49	236	14,116	3.662055	1	0.942091	96.25146
GOTERM_BP_ALL	GO:0048518~positive regulation of biological process	9	16.07143	0.386911	F10, PDPN, PON1, TGM2, RELN, F7, IGFBP1, LBP, EPO	49	2,033	14,116	1.275324	1	0.990391	99.94812
GOTERM_BP_ALL	GO:0048522~positive regulation of cellular process	8	14.28571	0.441705	F10, PDPN, TGM2, RELN, F7, IGFBP1, LBP, EPO	49	1,847	14,116	1.247782	1	0.994988	99.9878
Annotation Cluster 6	Enrichment score: 1.543855481247302											
GOTERM_BP_ALL	GO:0019752~carboxylic acid metabolic process	8	14.28571	0.002563	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	556	14,116	4.145059	0.885687	0.266429	3.890425
GOTERM_BP_ALL	GO:0043436~oxoacid metabolic process	8	14.28571	0.002563	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	556	14,116	4.145059	0.885687	0.266429	3.890425
GOTERM_BP_ALL	GO:0006082~organic acid metabolic process	8	14.28571	0.002668	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	560	14,116	4.115452	0.895426	0.245901	4.04688
GOTERM_BP_ALL	GO:0042180~cellular ketone metabolic process	8	14.28571	0.002860	HAO1, PDPN, SLC23A2, PON1, DPYS, ASL, PON3, PCK1	49	567	14,116	4.064644	0.911122	0.235813	4.331966
GOTERM_BP_ALL	GO:0046395~carboxylic acid catabolic process	4	7.142857	0.006330	HAO1, PON1, ASL, PON3	49	111	14,116	10.38132	0.995326	0.27067	9.350749
GOTERM_BP_ALL	GO:0016054~organic acid catabolic process	4	7.142857	0.006330	HAO1, PON1, ASL, PON3	49	111	14,116	10.38132	0.995326	0.27067	9.350749
GOTERM_BP_ALL	GO:0009308~amine metabolic process	5	8.928571	0.046497	GUSB, PON1, DPYS, RELN, ASL	49	400	14,116	3.60102	1	0.634253	52.10247
GOTERM_BP_ALL	GO:0044255~cellular lipid metabolic process	5	8.928571	0.102836	HAO1, CRLS1, PDPN, PON1, PCK1	49	526	14,116	2.738419	1	0.822737	81.31979
GOTERM_BP_ALL	GO:0009056~catabolic process	8	14.28571	0.129516	HAO1, AZGP1P1, AZGP1, GUSB, BIRC7, PON1, DPYS, ASL, PON3	49	1,253	14,116	1.839308	1	0.862832	88.28667
GOTERM_BP_ALL	GO:0044248~cellular catabolic process	6	10.71429	0.267123	HAO1, BIRC7, PON1, DPYS, ASL, PON3	49	1,024	14,116	1.687978	1	0.969846	99.18092
GOTERM_BP_ALL	GO:0032787~monocarboxylic acid metabolic process	3	5.357143	0.272974	HAO1, PDPN, PCK1	49	301	14,116	2.871246	1	0.971116	99.27637
GOTERM_BP_ALL	GO:0044106~cellular amine metabolic process	3	5.357143	0.278039	PON1, DPYS, ASL	49	305	14,116	2.83359	1	0.971989	99.35052
GOTERM_BP_ALL	GO:0006519~cellular amino acid and derivative metabolic process	3	5.357143	0.337262	PON1, DPYS, ASL	49	352	14,116	2.455241	1	0.984049	99.82707