Appendix 1

Methods

Data processing and quality control

Microarray raw data of the 8 datasets was downloaded via txt format from the corresponding platform. The data obtained for GSE3325, GSE6956, and GSE55945 was gathered by employing log2 transformation using the Limma Package (version 3.40.6) in R (http://www.bioconductor.org/packages/release/bioc/html/limma.html). While for the five datasets GSE17951, GSE32571, GSE46602, GSE69223, and GSE89194, the original data was used since these had already undergone log2 transformation. Then IQR method in the MetaDE Package (version 1.0.5) was used to summarize the multiple probes to one intensity (28). Finally, the quality control (QC) steps were performed on these datasets by using the MetaQC package (version 0.1.13) in R (28,29). The MetaQC package has two main functions, metaQC, and runQC, which function to implement the objective quality control as well as the inclusion and exclusion criteria based on 6 quantitative quality control measures: internal quality control (IQC), external quality control (EQC), accuracy quality control of different expression (DE) genes (AQCg), accuracy quality control of pathways (AQCp), consistency quality control of DE genes (CQCg), and consistency quality control of pathways (CQCp) (29). Scores of these 6 indices were calculated by MetaQC package, and a standardized mean rank (SMR) summary score based on the 6 indexes, was generated to evaluate the quality of each dataset. $0 < \text{SMR} \leq 1$ and large SMR indicates a dataset of low quality which should be filtered. While executing the metaQC function, the GSEA Biocarta v6.2 pathways was used since the pathways were cancer specific. While excuting the runQC function, the parameter "B" was set as "1e5", "nPath" was set as "50", "pvalCut" was set as "0.05" and the GSEA c2.all.v6.2 pathways was used as "fileForCQCp". Also, the PCA (principal component analysis) biplot was drawn to visualize the QC results. The 6 QC measures of each datasets was projected to the first two principal components subspace using arrows. Datasets with low quality often occur on the opposite side of arrows in the PCA biplots and have large SMR scores.
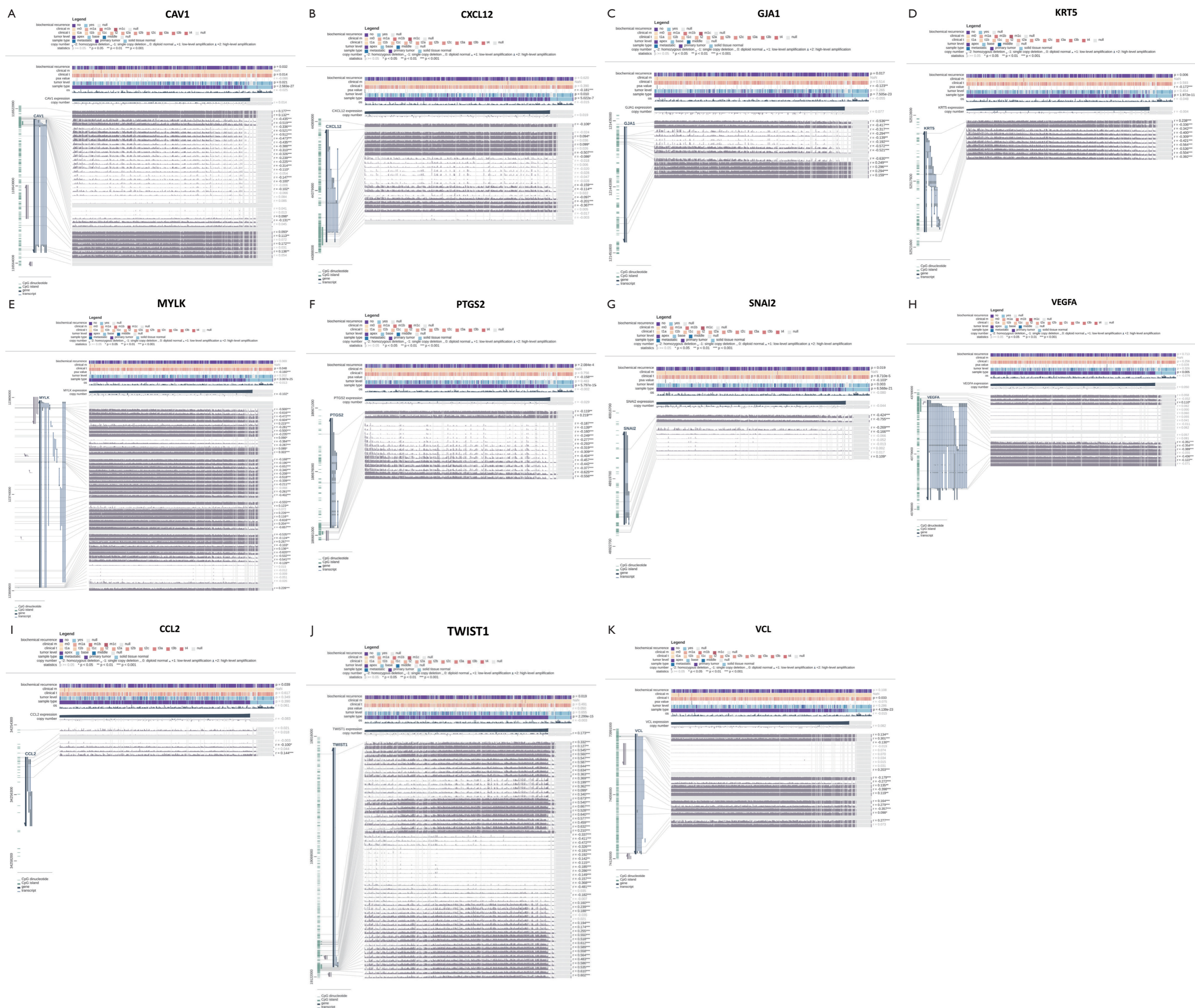
Figure S1 Association of methylation sites with expression of the 12 hub genes. (A) CAV1, (B) CXCL12, (C) GJA1, (D) KRT5, (E) MYLK, (F) PTGS2, (G) SNAI2, (H) VEGFA, (I) CCL2, (J) TWIST1 and (K) VCL. The methylation analysis in MEXPRESS showed that numerous methylation sites existed in the DNA sequences of (A) CAV1, (B) CXCL12, (C) GJA1, (D) KRT5, (E) MYLK, (F) PTGS2, (G) SNAI2 and (H) VEGFA, which were negatively correlated with the expression levels of the hub genes. On the contrary, (I) CCL2, (J) TWIST1 and (K) VCL showed positive results. The dark green line in the center of the plot represents ascending gene expression. Pearson's correlation coefficients and P values from the Wilcoxon rank-sum test for methylation sites and query gene expression are shown on the right side. The gray lines stand for Infinium 450k probes, and their heights represent the beta value for this probe. The dark blue lines at the bottom left indicate the gene and CpG islands.

Table S1 Clinical and histopathological data

Clinical variable	Values
A. GSE32571, tumor (n=59)	
Gleason score	
5, 6	5
7 (3+4)	28
7 (4+3)	12
8–10	15
Median age, years	62±7.2
B. GSE46602, tumor (n=36)	
Gleason score	
5, 6	17
7	15
8–10	4
Pathological stage	
T2a–c	19
T3a–b	17
TxN+	0
Age median (range), years	63 (46–71)
C. GSE69223, tumor (n=15)	
Gleason score	
5, 6	3
7	9
8–10	3
Pathological stage	
T2	10
T3	5
Age median (range), years	60 (47–69)
D. GSE89194, tumor (n=49)	
Gleason score	
7 (3+4)	49
Pathological stage	
T2a	14
T2c	35
Age range (years)	38–45 and 71–74

Table S2 Top GO functions (P value <0.05) relation to DEGs in network module

Category	ID	Term	Count	P value	Genes
A. Top16 GO enrichment terms of DEGs in module 1					
GOTERM_BP_DIRECT	GO:0006936	muscle contraction	3	0.000421867	CALD1, TPM2, LMOD1
GOTERM_BP_DIRECT	GO:0007015	actin filament organization	3	0.001056563	ACTC1, TPM2, LMOD1
GOTERM_BP_DIRECT	GO:0030239	myofibril assembly	2	0.005105219	LMOD1, MYL9
GOTERM_BP_DIRECT	GO:0006939	smooth muscle contraction	2	0.006559744	SMTN, MYLK
GOTERM_BP_DIRECT	GO:0070527	platelet aggregation	2	0.016330193	MYL9, VCL
GOTERM_CC_DIRECT	GO:0005829	cytosol	7	0.001961459	ACTC1, CALD1, TPM2, LMOD1, MYL9, VCL, MYLK
GOTERM_CC_DIRECT	GO:0005856	cytoskeleton	5	1.61E-05	SMTN, CALD1, TPM2, LMOD1, VCL
GOTERM_CC_DIRECT	GO:0015629	actin cytoskeleton	4	6.58E-05	SMTN, CALD1, TPM2, MYLK
GOTERM_CC_DIRECT	GO:0030016	myofibril	3	5.77E-05	CALD1, LMOD1, MYL9
GOTERM_CC_DIRECT	GO:0005884	actin filament	3	0.000435373	ACTC1, TPM2, LMOD1
GOTERM_CC_DIRECT	GO:0030017	sarcomere	2	0.015863304	ACTC1, LMOD1
GOTERM_CC_DIRECT	GO:0001725	stress fiber	2	0.025073529	MYL9, MYLK
GOTERM_MF_DIRECT	GO:0003779	actin binding	6	3.64E-08	SMTN, CALD1, TPM2, LMOD1, VCL, MYLK
GOTERM_MF_DIRECT	GO:0008307	structural constituent of muscle	3	0.00010693	SMTN, TPM2, MYL9
GOTERM_MF_DIRECT	GO:0005523	tropomyosin binding	2	0.006325393	CALD1, LMOD1
GOTERM_MF_DIRECT	GO:0017022	myosin binding	2	0.009660223	ACTC1, CALD1
B. Top14 GO enrichment terms of DEGs in module 2					
GOTERM_BP_DIRECT	GO:0006749	glutathione metabolic process	6	1.08E-09	GSTM4, GSTM3, GSTM2, GSTM1, GSTP1, GSTM5
GOTERM_BP_DIRECT	GO:0042178	xenobiotic catabolic process	5	6.24E-09	GSTM4, GSTM3, GSTM2, GSTM1, CYP3A5
GOTERM_BP_DIRECT	GO:0007165	signal transduction	5	0.034151714	GJA1, CXCL12, PENK, CCL2, CHGB
GOTERM_BP_DIRECT	GO:0018916	nitrobenzene metabolic process	4	3.89E-09	GSTM4, GSTM3, GSTM2, GSTM1
GOTERM_BP_DIRECT	GO:0098869	cellular oxidant detoxification	4	6.78E-05	GSTM2, GPX3, GSTP1, PTGS2
GOTERM_CC_DIRECT	GO:0005576	extracellular region	10	8.67E-05	TF, CXCL12, GPX3, GSTP1, PENK, CCL2, CHRDL1, F5, CHGB, VEGFA
GOTERM_CC_DIRECT	GO:0005615	extracellular space	8	0.001709106	TF, GOLM1, GPX3, GSTP1, CCL2, F5, CHGB, VEGFA
GOTERM_CC_DIRECT	GO:0005788	endoplasmic reticulum lumen	7	3.84E-07	TF, GOLM1, PENK, PTGS2, CHRDL1, F5, CHGB
GOTERM_CC_DIRECT	GO:0045171	intercellular bridge	5	8.18E-07	GSTM4, GSTM3, GSTM2, GSTM1, GSTM5
GOTERM_MF_DIRECT	GO:0004364	glutathione transferase activity	6	4.78E-11	GSTM4, GSTM3, GSTM2, GSTM1, GSTP1, GSTM5
GOTERM_MF_DIRECT	GO:0019899	enzyme binding	6	3.13E-05	GSTM4, GSTM3, GSTM2, GSTM1, CAV1, PTGS2
GOTERM_MF_DIRECT	GO:0042803	protein homodimerization activity	6	0.000553149	GSTM4, GSTM3, GSTM2, GSTM1, PTGS2, VEGFA
GOTERM_MF_DIRECT	GO:0005102	receptor binding	5	0.000600132	GSTM2, GJA1, CXCL12, CAV1, CCL2
GOTERM_MF_DIRECT	GO:0043295	glutathione binding	4	1.44E-07	GSTM4, GSTM3, GSTM2, GSTM1

Table S3 KEGG enrichment analysis of genes in the top 2 modules

Modules	Term	Count	P value	Genes
Module 1	Vascular smooth muscle contraction	3	2.77E-03	CALD1, MYLK, MYL9
	Focal adhesion	3	8.41E-03	MYLK, MYL9, VCL
	Regulation of actin cytoskeleton	3	8.73E-03	MYLK, MYL9, VCL
Module 2	Chemical carcinogenesis	9	2.79E-12	GSTM1, GSTM2, CYP3A5, GSTM3, GSTM4, PTGS2, ALDH3B2, GSTM5, GSTP1
	Drug metabolism - cytochrome P450	8	7.17E-11	GSTM1, GSTM2, CYP3A5, GSTM3, GSTM4, ALDH3B2, GSTM5, GSTP1
	Metabolism of xenobiotics by cytochrome P450	8	1.32E-10	GSTM1, GSTM2, CYP3A5, GSTM3, GSTM4, ALDH3B2, GSTM5, GSTP1
	Glutathione metabolism	7	9.28E-10	GSTM1, GSTM2, GSTM3, GSTM4, GPX3, GSTM5, GSTP1
	Rheumatoid arthritis	3	1.73E-02	CCL2, VEGFA, CXCL12

Table S4 Top25 hub genes

Betweenness	BottleNeck	Closeness	Degree	DMNC	EcCentricity	EPC	MCC	MNC	Radiality	Stress
VEGFA	VEGFA	VEGFA	VEGFA	SMTN	WT1	VEGFA	VCL	VEGFA	VEGFA	VEGFA
VCL	VCL	VCL	VCL	LMOD1	PROM1	CAV1	ACTC1	CAV1	CAV1	VCL
AMACR	AMACR	CAV1	CAV1	GSTP1	SNAI2	CXCL12	TPM2	VCL	VCL	TWIST1
TWIST1	CAV1	CXCL12	CCL2	GSTM1	VEGFA	VCL	CALD1	KRT5	CXCL12	CAV1
CAV1	MYLK	GJA1	CXCL12	ALDH3B2	NDRG2	CCL2	MYLK	CCL2	GJA1	AMACR
PTN	KRT5	PTGS2	KRT5	GPX3	DUOX1	GJA1	MYL9	CXCL12	SNAI2	SNAI2
SNAI2	TWIST1	CCL2	PTGS2	CYP3A5	ETV5	ACTC1	LMOD1	ACTC1	TWIST1	KRT5
CRYAB	CXCL12	SNAI2	AMACR	GSTM5	TMEM37	PTGS2	SMTN	KRT14	PTGS2	PTN
CLU	PTN	TWIST1	CALD1	GSTM4	MB	CALD1	GSTM3	GJA1	PROM1	CRYAB
KRT5	FOLH1	MYLK	GJA1	GSTM2	PTP4A3	SNAI2	GSTM2	PTGS2	CCL2	CLU
RRM2	ITGB4	PROM1	ACTC1	GSTM3	SPRED1	TPM2	GSTM5	CALD1	WT1	PTGS2
GSTP1	CRYAB	WT1	SNAI2	CHRDL1	GAS1	MYLK	GSTM4	SNAI2	MYLK	GJA1
GJA1	SNAI2	AMACR	CLU	MYL9	SERPINB5	TWIST1	GSTP1	COL2A1	HSPB1	CXCL12
WT1	LMOD1	CLU	TWIST1	GOLM1	JAZF1	PROM1	GSTM1	TPM2	SERPINB5	ITGB4
MYLK	CLU	HSPB1	GSTP1	TF	B3GAT1	KRT5	CYP3A5	TWIST1	TIMP3	GSTP1
PTGS2	RRM2	TIMP3	KRT14	MAP1B	FOXD1	FLNC	ALDH3B2	MYLK	CLU	CCL2
ITGB4	PTGS2	KRT5	FLNC	SDPR	PDPN	CALM1	GPX3	FLNC	S100A4	KRT14
FOLH1	GSTP1	CRYAB	ITGB4	EHD2	SEMA6D	ITGB4	VEGFA	TGFB3	ID1	PROM1
CXCL12	F5	CALD1	TPM2	KRT23	SCUBE2	MYL9	CXCL12	CALM1	CRYAB	WT1
OLFM4	CALD1	SERPINB5	CALM1	COL13A1	HSPB1	MME	CAV1	HSPB1	AMACR	MYLK
TIMP3	WT1	TGFB3	MYLK	LEPREL1	NPR2	TIMP3	CCL2	PROM1	F5	TIMP3
LMOD1	GJA1	S100A4	CRYAB	KRT13	SCGB1A1	HSPB1	PTGS2	MME	KRT5	MAP1B
PROM1	ADAMTS5	F5	MME	PTRF	SEMA3E	ANXA2	GJA1	PROM1	TGFB3	LMOD1
CCL2	SPON1	ACTC1	TIMP3	MYLK	CSRP2	PTGS1	F5	PENK	PDPN	CALD1
SCUBE2	CLDN3	COL2A1	COL2A1	CHGB	ENAH	S100A4	PENK	GSTM5	FOLH1	OLFM4