

Supplementary File 3.

Out of 21,329 genes in the arrays, 132 genes correlated positively and 203 genes correlated negatively with *KRT19* expression, using a total of 139 HCC samples. Gene Set Enrichment Analysis, a computational web based analysis method, was used to statistically analyse the overlap with previous published gene expression signatures (see tables below) ^{1,2}. Additionally, STRING analysis was performed to predict protein interactions, either direct (physical) or indirect (functional) associations (see pictures below) ³.

Table. Gene Set Enrichment Analysis of genes positively correlated with *KRT19* (n=132)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p value	Source
ONDER_CDH1_TARGETS_2_DN	Genes down-regulated in HMLE cells (immortalized nontransformed mammary epithelium) after E-cadherin (CDH1) [Gene ID=999] knockdown by RNAi.	25	0.0529	0,00E+00	Pubmed 18483246
DELYS_THYROID_CANCER_UP	Genes up-regulated in papillary thyroid carcinoma (PTC) compared to normal tissue.	24	0.0600	0,00E+00	Pubmed 17621275
WU_CELL_MIGRATION	Genes associated with migration rate of 40 human bladder cancer cells.	18	0.0968	0,00E+00	Pubmed 18724390
LEE_LIVER_CANCER_SURVIVAL_DN	Genes highly expressed in hepatocellular carcinoma with poor survival.	18	0.1452	0,00E+00	Pubmed 15349906
CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_UP	Top 200 marker genes up-regulated in the 'proliferation' subclass of hepatocellular carcinoma (HCC); characterized by increased proliferation, high levels of serum AFP [Gene ID=174], and chromosomal instability.	14	0.1022	3.77E-15	Pubmed 18701503
SMID_BREAST_CANCER_BASAL_UP	Genes up-regulated in basal subtype of breast cancer samples.	24	0.0355	9.66E-15	Pubmed 18451135
CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_DN	Genes down-regulated in luminal-like breast cancer cell lines compared to the basal-like ones.	20	0.0439	4.39E-14	Pubmed 16288205
JAEGER_METASTASIS_DN	Genes down-regulated in metastases from malignant melanoma compared to the primary tumors.	16	0.0606	1.41E-13	Pubmed 17289871
GRUETZMANN_PANCREATIC_CANCER_UP	Genes up-regulated in pancreatic ductal adenocarcinoma (PDAC) identified in a meta analysis across four independent studies.	17	0.0491	6.93E-13	Pubmed 15897887
WOO_LIVER_CANCER_RECURRENCE_UP	Genes positively correlated with recurrence free survival in patients with hepatitis B-related (HBV) hepatocellular carcinoma (HCC).	11	0.1048	2.93E-12	Pubmed 18381945
HOSHIDA_LIVER_CANCER_SUBCLASS_S1	Genes from 'subtype S1' signature of hepatocellular carcinoma (HCC): aberrant activation of the WNT signaling pathway.	14	0.0591	7.4E-12	Pubmed 19723656
MCBRYAN_PUBERTAL_BREAST_4_5WK_UP	Genes up-regulated during pubertal mammary gland development between week 4 and 5.	14	0.0502	6.55E-11	Pubmed 17486082
COLDREN_GEFITINIB_RESISTANCE_DN	Genes down-regulated in NSCLC (non-small cell lung carcinoma) cell lines resistant to gefitinib [PubChem=123631] compared to the sensitive ones.	13	0.0570	6.77E-11	Pubmed 16877703

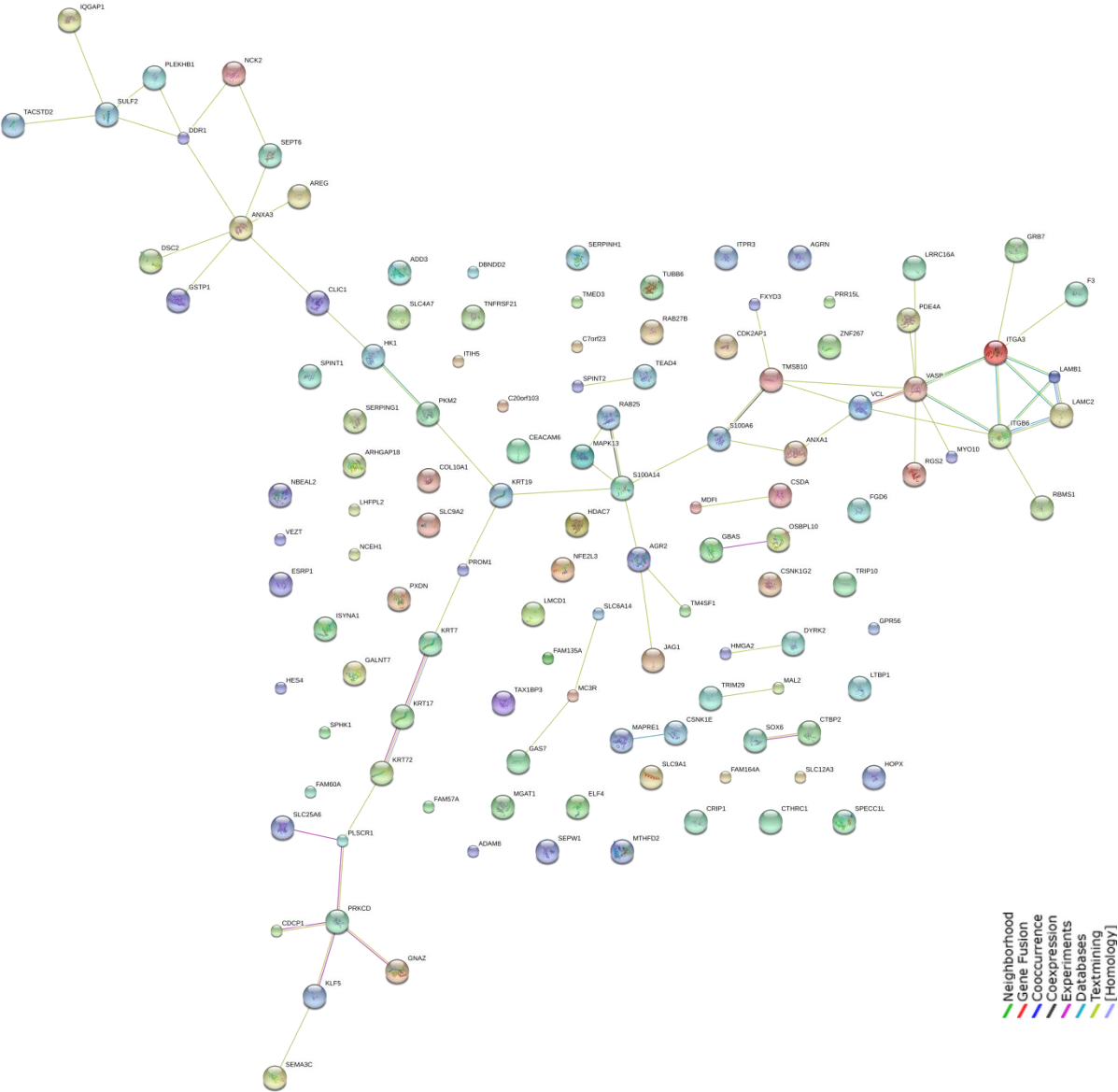
NUYTEN_NIPP1_TARGETS_DN	Genes down-regulated in PC3 cells (prostate cancer) after knockdown of NIPP1 [Gene ID=5511] by RNAi.	21	0.0270	9.56E-11	Pubmed 17724462
LIU_PROSTATE_CANCER_DN	Genes down-regulated in prostate cancer samples.	17	0.0357	1.07E-10	Pubmed 16618720
SMID_BREAST_CANCER_LUMINAL_B_DN	Genes down-regulated in the luminal B subtype of breast cancer.	18	0.0301	4.63E-10	Pubmed 18451135
FARMER_BREAST_CANCER_BASAL_VS_LUMINAL	Genes which best discriminated between two groups of breast cancer according to the status of ESR1 and AR [Gene ID=2099, 367]: basal (ESR1- AR-) and luminal (ESR1+ AR+).	14	0.0418	7.21E-10	Pubmed 15897907
CHARAFE_BREAST_CANCER_BASAL_VS_MESENCHYMAL_UP	Genes up-regulated in basal-like breast cancer cell lines as compared to the mesenchymal-like ones.	9	0.0732	7.58E-9	Pubmed 16288205
VANTVEER_BREAST_CANCER_ESR1_DN	Down-regulated genes from the optimal set of 550 markers discriminating breast cancer samples by ESR1 [Gene ID=2099] expression: ER(+) vs ER(-) tumors.	11	0.0493	9.77E-9	Pubmed 11823860
CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_UP	Genes up-regulated in luminal-like breast cancer cell lines compared to the mesenchymal-like ones.	14	0.0307	3.63E-8	Pubmed 16288205

Table. Gene Set Enrichment Analysis of genes negatively correlated with *KRT19* (n=203)

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p value	Source
FLECHNER_BIOPSY_KIDNEY_TRANSPLANT_REJECTED_VS_OK_DN	Genes differentially expressed in kidney biopsies from patients with acute transplant rejection compared to the biopsies from patients with well functioning kidneys more than 1-year post transplant.	40	0.0718	0,00E+00	Pubmed 15307835
CAIRO_LIVER_DEVELOPMENT_DN	Genes down-regulated at early fetal liver stage (embryonic days E11.5 - E12.5) compared to the late fetal liver stage (embryonic days E14.5 - E16.5).	41	0.1806	0,00E+00	Pubmed 19061838
LEE_LIVER_CANCER_DENA_DN	Genes down-regulated in hepatocellular carcinoma (HCC) induced by diethylnitrosamine (DENA) [PubChem=5921].	15	0.2083	0,00E+00	Pubmed 15565109
YAMASHITA_LIVER_CANCER_STEM_CELL_DN	Genes down-regulated in hepatocellular carcinoma (HCC) cells with hepatic stem cell properties.	18	0.2400	0,00E+00	Pubmed 19150350
SU_LIVER	Genes up-regulated specifically in human liver tissue.	14	0.2414	0,00E+00	Pubmed 11904358
CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_DN	Top 200 marker genes down-regulated in the 'proliferation' subclass of hepatocellular carcinoma (HCC); characterized by increased proliferation, high levels of serum AFP [Gene ID=174], and chromosomal instability.	38	0.2452	0,00E+00	Pubmed 18701503
HOSHIDA_LIVER_CANCER_SUBCLASS_S3	Genes from 'subtype S3' signature of hepatocellular carcinoma (HCC): hepatocyte differentiation.	66	0.2481	0,00E+00	Pubmed 19723656
WOO_LIVER_CANCER_RECURRENCE_DN	Genes negatively correlated with recurrence free survival in patients with hepatitis B-related (HBV) hepatocellular carcinoma (HCC).	21	0.2625	0,00E+00	Pubmed 18381945

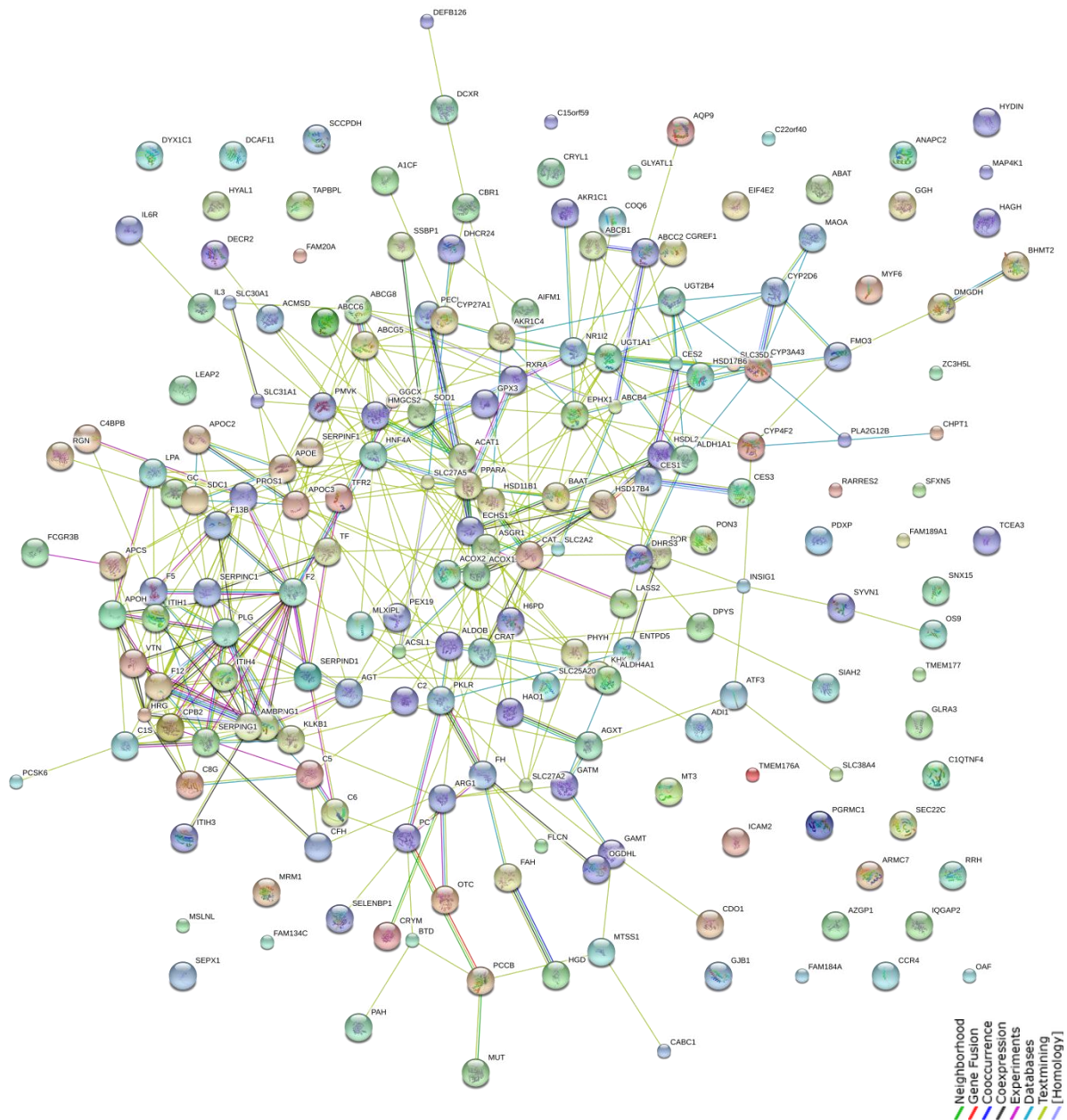
BOYALT_LIVER_CANCER_SUBCLASS_G3_DN	Down-regulated genes in hepatocellular carcinoma (HCC) subclass G3, defined by unsupervised clustering.	16	0.3019	0,00E+00	Pubmed 17187432
HSIAO_LIVER_SPECIFIC_GENES	Liver selective genes	77	0.3020	0,00E+00	Pubmed 11773596
LEE_LIVER_CANCER_SURVIVAL_UP	Genes highly expressed in hepatocellular carcinoma with good survival.	59	0.4504	0,00E+00	Pubmed 15349906
CAIRO_HEPATOBLASTOMA_DN	Genes down-regulated in hepatoblastoma samples compared to normal liver tissue.	23	0.0846	1.11E-16	Pubmed 19061838
ACEVEDO_LIVER_CANCER_DN	Genes down-regulated in hepatocellular carcinoma (HCC) compared to normal liver samples.	28	0.0519	7.11E-15	Pubmed 18413731
LEE_LIVER_CANCER_ACOX1_DN	Genes down-regulated in hepatocellular carcinoma of ACOX1 [Gene ID=51] knockout mice.	12	0.1875	1.5E-13	Pubmed 15565109
CAIRO_HEPATOBLASTOMA_CLASSES_DN	Genes down-regulated in robust Cluster 2 (rC2) of hepatoblastoma samples compared to those in the robust Cluster 1 (rC1).	17	0.0798	2.33E-12	Pubmed 19061838
LEE_LIVER_CANCER_MYC_TGFA_DN	Genes down-regulated in hepatocellular carcinoma (HCC) tissue of MYC and TGFA [Gene ID=4609, 7039] double transgenic mice.	11	0.1774	3.03E-12	Pubmed 15565109
IIZUKA_LIVER_CANCER_PROGRESSION_G2_G3_UP	Genes up-regulated during transition from G2 (moderately differentiated tumor, infected with HCV) to G3 (poorly differentiated tumor, infected with HCV) in the development of hepatocellular carcinoma.	8	0.3810	3.33E-12	Pubmed 15710396
MOOTHA_MITOCHONDRIA	Mitochondrial genes	22	0.0484	2.8E-11	Pubmed 12808457
LEE_LIVER_CANCER_E2F1_DN	Genes down-regulated in hepatocellular carcinoma (HCC) induced by overexpression of E2F1 [Gene ID=1869].	10	0.1724	4.09E-11	Pubmed 15565109
GRUETZMANN_PANCREATIC_CANCER_DN	Genes down-regulated in pancreatic ductal adenocarcinoma (PDAC) identified in a meta analysis across four independent studies.	15	0.0765	8.7E-11	Pubmed 15897887

Fig. STRING analysis of genes positively correlated with *KRT19* (n=132)



This is the **evidence view**. Different line colors represent the types of evidence for the association.

Fig. STRING analysis of genes negatively correlated with *KRT19* (n=203)



This is the **evidence view**. Different line colors represent the types of evidence for the association.

1. Subramanian, A., *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
2. Mootha, V.K., *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).
3. Szklarczyk, D., *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* **39**, D561-568 (2011).