Human Cancer Cells Express Slug-based Epithelial-Mesenchymal Transition Gene Signature Obtained in Vivo.

Dimitris Anastassiou[2,6], Viktoria Rumjantseva[6], Wei-yi Cheng[2,6], Jianzhong Huang[6], Peter D. Canoll[3], Darrell J. Yamashiro[1,3,5], and Jessica J. Kandel[1,4,5]

[1] Institute for Cancer Genetics, [2] Center for Computational Biology and Bioinformatics, [3] Department of Pathology and Cell Biology, [4] Department of Surgery, [5] Department of Pediatrics, [6] Department of Electrical Engineering, Columbia University, New York, NY, USA

Correspondence to: Dimitris Anastassiou (anastas@ee.columbia.edu) or Jessica Kandel (jjk47@columbia.edu)

Abstract:

The biological mechanisms underlying cancer cell motility and invasiveness remain unclear, although it has been hypothesized that they involve some type of epithelial-mesenchymal transition (EMT). Here we show that human cancer cells express in vivo a precise multi-cancer invasion-associated gene expression signature characterized by the prominent presence of collagen COL11A1 and thrombospondin THBS2. The signature is present in the expression of all solid tumor datasets that we analyzed and includes numerous EMT markers, among them the transcription factor Slug, fibronectin, and α SMA. Using xenograft models of human cancer cells in immunocompromised mice and profiling the harvested tumors separately with species specific probes, we found that human, but not mouse, cells express most of the genes of the signature and Slug is the only upregulated EMT-inducing transcription factor. Taken together with the presence of the signature in many publicly available datasets, our results suggest that this Slug-based EMT signature is produced by the cancer cells themselves in multiple cancer types, including even nonepithelial cancers such as neuroblastoma. Furthermore, we found that the presence of the signature in human xenografted cells was associated with a downregulation of adipocyte markers in the mouse tissue adjacent to the invasive tumor, suggesting contextual microenvironmental interactions when the cancer cells encounter adipocytes, as previously reported. The known and consistent gene composition of this cancer EMT signature, particularly when combined with simultaneous analysis of the adjacent microenvironment, provides unique opportunities for shedding light on the underlying mechanisms of cancer invasiveness as well as identifying potential diagnostic markers and targets for metastasis-inhibiting therapeutics.

Introduction

It has been hypothesized [1-3] that cancer cells become invasive and migratory by undergoing some type of epithelial-mesenchymal transition (EMT). Each type of EMT is assumed to be orchestrated by different unknown combinations of multiple interacting transcription factors and signaling pathways. A set of genes comprising an "EMT core signature" was recently derived [4] after triggering EMTs in different ways and observing the resulting shared changes in gene expression. However, the details of the specific types of EMT associated with cancer invasion remain elusive.

A precise multi-cancer gene expression signature involving a set of many genes, most of them being EMT markers coordinately overexpressed only in a subset of malignant samples that have exceeded a particular staging threshold specific to each cancer type was recently identified [5]. The signature is present in numerous publicly available datasets from multiple cancer types from all solid tumor types that we tried, including nonepithelial cancer such as neuroblastoma and Ewing's sarcoma but not in leukemia (a few examples of heat maps are shown at www.ee.columbia.edu/~anastas/nat_prec/supp). Among the overexpressed genes are the EMT-inducing transcription factor Slug (*SNAI2*), various collagens and proteinases, α-SMA, fibronectin, fibroblast activation protein, and many extracellular matrix glycoproteins, suggesting fibroblastic nature after passing through a Slug-based EMT. The signature, however, is not of a general fibroblastic nature, but rather has its own special characteristics, one of which is that genes *COL11A1*, *THBS2* and *INHBA* have a prominent presence in all cases, and they are strongly co-expressed with a remarkably smooth continuous transition [5]. We identified collagen *COL11A1* as a reliable proxy for the signature: in each solid tumor dataset that we tried, the list of genes whose expression is most correlated with that of *COL11A1* consistently includes the other genes of the signature at the top, with the only exception of glioblastoma in which *COL11A1* is not as prominently co-expressed with the other genes, though the signature is still present otherwise. Notably, E-cadherin (*CDH1*) is not downregulated at least at the mRNA level. Furthermore, the signature contains numerous other EMT-associated genes. Table 1 shows a list of the 64 genes corresponding to the 100 most overexpressed probe sets, as we previously reported [5], of the signature. Highlighted by bold typeface among these 64 are 20 known EMT-associated genes, 17 coming from the list of 91 upregulated "EMT core signature" [4] genes (COL5A2, FAP, POSTN, COL1A2, COL3A1, FBN1, TNFAIP6, MMP2, GREM1, BGN, CDH11, SPOCK1, DCN, COPZ2, THY1, PCOLCE, PRRX1) plus the obvious EMT markers SNAI2, FN1, ACTA2. Four additional genes (underlined but not bold in Table 1), PDGFRB, SPARC, INHBA, COL6A2, have also been reported as EMT-related [6], for a total of 24 EMT factors. Even without including these additional seven genes, the *P* value of encountering 17 out of 64 genes taken from the list of the 91 EMT core signature genes is 2×10^{-22}. Therefore, this fibroblastic signature is the result, at least in part, of an EMT.

The initiation of the signature overexpression only after reaching a particular cancer-type-specific stage suggests a biological mechanism associated with cancer invasion, but the nature of this mechanism remained unclear. Given the heterogeneity of cells in tumors, it could reflect the superposition of several mechanisms. Among other possibilities, the fibroblast-like cells producing the signature could be derived from multiple sources, such as the bone marrow, the local stroma, or the cancer cells after undergoing an EMT. A fundamental question, therefore, is

which among the genes of the signature are expressed by the cancer cells and which are expressed by the adjacent microenvironment. Relatedly, we had observed [7] a striking similarity between the set of genes in the signature and a subset of the genes that are known to be lower expressed when mouse embryonic fibroblasts are reprogrammed into induced pluripotent stem cells [8]. Because it is known [9] that a mesenchymal-epithelial transition (MET) is part of the reprogramming of mouse fibroblasts into stem cells, we reasoned that, conversely, the signature may correspond to some kind of EMT-related transition from a stem-like state to a fibroblast-like state used in early embryogenesis. Therefore, we hypothesized [7] that most of the genes in our signature are expressed from cancer stem cells (CSSs) passing through some type of EMT. Furthermore, because of the prominent presence of inhibin-A (*INHBA*) in the signature, we hypothesized that activin A (INHBA dimer) signaling may be responsible for the signature.

To test these two hypotheses, we used xenograft models implanting human cancer cell lines into NCR nude mice. We used NGP neuroblastoma cells lines, as we had had prominently found the signature in neuroblastoma publicly available datasets (as shown in the heat map at www.ee.columbia.edu/~anastas/nat_prec/supp/heatmap.neuroblastoma.pdf), even though it is a nonepithelial cancer, as confirming the signature in vivo in such tumors could shed further light on the specific EMT-like biological mechanisms. Some of the implanted cancer cell lines were in their original form, some were engineered to express INHBA, and some were engineered to express the activin antagonist follistatin (FST). Each of the resulting growing tumors was harvested and profiled for gene expression twice using human and mouse microarrays separately. Our results validated the first hypothesis, but not the second: The vast majority of the genes of the signature were found in human, but not mouse, cells, and their presence was independent of any transfections, indicating that activin signaling does not play a causal role.

The presence only in human cells of coordinately expressed genes of the same previously identified signature [5] indicates that the cancer cells themselves undergo an EMT. Furthermore, the continuity of the fibroblastic transition signature (Figure 1) suggests that it may reflect a dynamic and reversible process. Its potential reversal is consistent with the possibility that it is a requirement for all metastases, even though the signature is only observed in a subset of high-stage extracted samples.

Results and Discussion

We found very different expression levels (Figure 2) for most genes in human and mouse, suggesting that cross-species hybridization is minimal, which was confirmed with real-time PCR (see Materials and Methods). Using *COL11A1* as a proxy [5], we ranked the 18 samples accordingly and investigated if most of the EMT markers of the signature were co-expressed with *COL11A1*. We found that this was indeed the case in human cells only. For example, Figures 2A-B show color-coded scatter plots in human and mouse of the 18 samples for the expression of Slug (*SNAI2*) in terms of the expression of the main genes of the fibroblastic transition signature, *COL11A1* and *THBS2* (same genes as in the scatter plots of Figure 1) demonstrating that this co-expression is clearly present in the human cells, but absent in the mouse cells. Specifically, seven samples had high or intermediate levels of co-expressed genes in the human cells, while the remaining 11 have relatively lower levels.

Based on this partition, we identified 398 significantly (both $Q < 0.05$ and FC > 2) up-regulated genes, 29 among which (COL11A1, THBS2, COL5A2, COL5A1, VCAN, COL1A1, COL3A1, FN1,SULF1, FBN1, ASPN, SPARC, CTSK, MMP2, BGN, LUM, LOXL2, COL6A3, TIMP3, CDH11, SERPINF1, EDNRA, ACTA2, PDGFRB, SNAI2, LGALS1, GLT8D2, NID2, PRRX1) belong to the set of 64 genes of the fibroblastic transition signature in Table 1 ($P = 10^{-27}$), as well as VIM (vimentin). The presence in this list of *SNAI2* (Slug), *ACTA2* (α-SMA), *FN1* (fibronectin), *VIM* (vimentin), together with many of the other EMT markers mentioned above, indicates that some human cancer cells underwent EMT. Other EMT-inducing transcription factors (Snail, Twist, ZEB1, ZEB2, SIP1, FOXC2) were not upregulated (Figure 2C), while the upregulation of SNAI2 (Figure 2A) was very significant ($Q < 3\times10^{-4}$ and FC = 5.22).

The heat map in Figure 3 shows the co-expression of the above 29 significantly upregulated genes. *INHBA*, the third prominent gene of the signature in addition to *COL11A1* and *INHBA*, is not included in the list, because its expression was manipulated by the transfections with consistent results. Furthermore, as shown in the heat map, the transfection of cancer cells with either INHBA (labeled I) or FST (labeled F) did not have any effect on the presence of the signature. Precisely these same 29 genes are used in the heat maps at www.ee.columbia.edu/~anastas/nat_prec/supp , suggesting that the same EMT signature may be expressed in all solid tumors, even in nonepithelial tumors, such as neuroblastoma and Ewing's sarcoma. Therefore, our results imply the employment of a mesenchymal transition process more general than what EMT is assumed to be.

We next analyzed the mouse microarray data to identify genes correlated with the presence of the cancer EMT signature in the human cells. We found 32 significantly (both $Q < 0.05$ and FC > 2) downregulated mouse genes in the presence of the human fibroblastic transition signature. Among them, the top two genes with the highest fold change (12.3 and 11.8 respectively) were the adipocyte markers adiponectin (*ADIPOQ*) and adipsin (*CFD*). We observed that these genes were strongly co-regulated with many other adipocyte markers, including fatty acid binding protein 4 (*FABP4*) aka *aP2*. This cluster of adipocyte markers whose downregulation in the mouse cells is strongly associated with the upregulation of the fibroblastic transition signature genes in human cells is also shown in the heat map of Figure 3. Many among these genes are known as adipocyte differentiation markers, and their downregulation is consistent with the finding [10] that adipocytes are dedifferentiated as they encounter adjacent invading cancer cells in a "vicious cycle" of interaction.

The quality of "stemness" in cells appears to be closely interconnected with the ability to pass through transitions to and from mesenchymal characteristics. Indeed, EMT generates cells with properties of stem cells [11] and, conversely, MET is involved in the reprogramming of fibroblasts into stem cells [9]. Therefore, we speculate [7] that the cancer EMT signature initiates from CSCs, which may even appear spontaneously [12]. It is also possible that the

adipocytes of the stroma adjacent to the tumor are dedifferentiated into a mesenchymal stem cell-like state and, together with other mesenchymal stem cells derived from the bone marrow [13] interact with the invading fibroblastic cancer cells in a manner that reactivates a particular early embryonic developmental program.

Not all fibroblastic transition genes in Table 1 are upregulated in the human cells. For example, PLAU (urokinase plasminogen activator) is slightly upregulated only in mouse cells consistently with its assumed role of being secreted by reactive stromal cells as a pro-enzyme that activates itself and other proteases in the presence of cancer cells.

Another prominent fibroblastic transition gene that is not upregulated in human cells is MMP11 (matrix metallopeptidase 11, aka stromelysin 3), in agreement with the hypothesis [10, 14, 15] that MMP11 is expressed in the adipocytes of the adjacent reactive stroma, indicating that the full version of the cancer EMT signature may be stabilized by contextual microenvironmental signals when cancer cells encounter adipocytes triggering their dedifferentiation, apoptosis, or metabolism. This process may reflect a Darwinian selection mechanism, as cancer cells may not survive the uptake of lipid droplets [14]. The expression in cancer cells of the adipocyte enhancer binding protein 1 (AEBP1), a prominent gene of the signature known to bind in the promoter region of the adipocyte fatty-acid protein FABP4 may play an important role in that respect, as may also the presence of oxidative stress and TNF signaling suggested by the presence of TNFAIP6 and C1QTNF3 in the signature.

The hypothesis that the EMT signature is triggered by adipocytes is consistent with the facts [5] that in breast cancer the signature overexpression appears immediately upon reaching invasive stage I, while in ovarian cancer overexpression appears only when the tumor is already well into stage III as in omental metastasis, because ovarian cancer initially progresses by disseminating locally across mesothelial surfaces and probably carried by the physiological movement of peritoneal fluid to the peritoneum and omentum, a fatty structure [16].

Many among the top genes of the cancer EMT signature (Table 1) have previously been individually identified as associated with metastatic potential in cancer. Such associations can now largely be explained by the fact that these genes are expressed by the cancer cells themselves undergoing an EMT expressing a precise gene signature. The known composition of the signature, particularly when combined with its separation in a species-specific manner in xenograft models, provides multiple and unique opportunities for understanding the underlying biological mechanisms and identifying prognostic markers and potential targets for metastasis-inhibiting therapeutics.

Materials and Methods

Tumor Implantation

An inoculum of 10^6 NGP neuroblastoma cells containing FUW-Luciferase plasmid (kindly donated by Dr. Adolfo Ferrando) suspended in 0.1 mL of phosphate-buffered saline (PBS) was injected into the left kidney of 18 mice. The NGP cell line was originally obtained from Garrett

Brodeur, Children's Hospital of Philadelphia. All cells were authenticated by short tandem repeat profiling analysis. Some of the implanted cells had previously been stably co-transfected with either FST-pReceiver-Lv105 or INHBA- pReceiver-Lv105 (GeneCopoeia; Rockville, MD). Seven mice were implanted with INHBA co-expressing NGP cells, six with FST and five with control NGP cells.

Harvesting of Specimens
Mice were sacrificed when estimated tumor weight reached 1.5 g followed by collection of contralateral kidney and tumor. Tumor tissue was snap frozen for RNA isolation.

Microarrays and probes preparation
15 Ag of biotinylated cRNA prepared as recommended using Affymetrix GeneChip 3' IVT express kit (Affymetrix, Santa Clara, CA) from RNA that was isolated and was hybridized to HG-U133A 2.0 and 430 A2.0 Gene Chips (Affymetrix, Santa Clara, CA) according to Affymetrix recommendations **(GeneChip Hybridization, Wash, and Stain Kit).** The samples were scanned with Affymetrix Gene Chip Scanner 3000. After scanning, expression values for each gene were determined using Affymetrix Gene Chip software version 4.0.

RNA isolation and semiquantitative reverse transcription-PCR
Total RNA was isolated from tumors using Qiagen miRNeasy mini kit (Qiagen , Germantown, MD) followed by reverse transcription using SuperScript First-Standard synthesis System for RT-PCR from Invitrogen according to manufacture recommendations (Carlsbad, CA USA). Relative expression of human versus mouse COL11A1 (Hs00266273_m1,Ms00483387_m1) genes in tumor xenografts were examined by RT-PCR. Products were detected by Hot Start-IT Probe qPCR Master Mix from USB Affymetrix (Santa Clara, CA) according to manufacture instructions and analyzed by Stratagene Mx3005p real time PCR machine. After scanning, expression values for genes were determined using MxPro 410 (Agilent Technologies, Inc., Santa Clara, CA). To correct for sample variations in RT-PCR efficiency and errors in quantitation, analysis of human HPRT, CYC, GAPD and mouse GAPD, ACTB, GUSB, expression was used.

Microarray dataset
The data set, corresponding to 18 tumors profiled separately with human and mouse microarrays, will be available from the GEO database under accession number to be assigned. Data were RMA normalized using the Bioconductor open source software.

Differential expression analysis
We regrouped the 18 samples, according to the expression level of human *COL11A1*, into 7 samples with high or intermediate *COL11A1* expression values, and 11 with low *COL11A1* expression values. Based on this partition, we performed differential expression analysis using significance analysis of microarrays (SAM) [17], implemented in the Bioconductor package *samr*. We define the significantly differentially expressed gene as those having both a *Q*-value less than 0.05 and a fold-change greater than 2.

References

1. Hay, E.D., *An overview of epithelio-mesenchymal transformation.* Acta Anat (Basel), 1995. **154**(1): p. 8-20.
2. Thiery, J.P., *Epithelial-mesenchymal transitions in tumour progression.* Nat Rev Cancer, 2002. **2**(6): p. 442-54.
3. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* J Clin Invest, 2009. **119**(6): p. 1420-8.
4. Taube, J.H., et al., *Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes.* Proc Natl Acad Sci U S A, 2010. **107**(35): p. 15449-54.
5. Kim, H., et al., *Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1.* BMC Med Genomics, 2010. **3**: p. 51.
6. Jechlinger, M., et al., *Expression profiling of epithelial plasticity in tumor progression.* Oncogene, 2003. **22**(46): p. 7155-69.
7. Cheng, W.Y., et al., *Cancer invasion associated gene expression signature is present in differentially expressed genes in the reprogramming of fibroblasts into stem cells. Available from Nature Precedings <http://hdl.handle.net/10101/npre.2011.5924.1 >.* 2011.
8. Boue, S., et al., *Analysis of human and mouse reprogramming of somatic cells to induced pluripotent stem cells. What is in the plate?* PLoS One, 2010. **5**(9).
9. Polo, J.M. and K. Hochedlinger, *When fibroblasts MET iPSCs.* Cell Stem Cell, 2010. **7**(1): p. 5-6.
10. Motrescu, E.R. and M.C. Rio, *Cancer cells, adipocytes and matrix metalloproteinase 11: a vicious tumor progression cycle.* Biol Chem, 2008. **389**(8): p. 1037-41.
11. Mani, S.A., et al., *The epithelial-mesenchymal transition generates cells with properties of stem cells.* Cell, 2008. **133**(4): p. 704-15.
12. Chaffer, C.L., et al., *Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state.* Proc Natl Acad Sci U S A, 2011. **108**(19): p. 7950-5.
13. Quante, M., et al., *Bone marrow-derived myofibroblasts contribute to the mesenchymal stem cell niche and promote tumor growth.* Cancer Cell, 2011. **19**(2): p. 257-72.
14. Andarawewa, K.L., et al., *Stromelysin-3 is a potent negative regulator of adipogenesis participating to cancer cell-adipocyte interaction/crosstalk at the tumor invasive front.* Cancer Res, 2005. **65**(23): p. 10862-71.
15. Dirat, B., et al., *Cancer-associated adipocytes exhibit an activated phenotype and contribute to breast cancer invasion.* Cancer Res, 2011. **71**(7): p. 2455-65.
16. Lengyel, E., *Ovarian cancer development and metastasis.* Am J Pathol, 2010. **177**(3): p. 1053-64.
17. Tusher, V.G., R. Tibshirani, and G. Chu, *Significance analysis of microarrays applied to the ionizing radiation response.* Proc Natl Acad Sci U S A, 2001. **98**(9): p. 5116-21.

Figure Legends and Figures

Figure 1: Color-coded scatter plots for the coexpression of the EMT inducing transcription factor Slug (*SNAI2*) with the main signature genes *COL11A1* and *THBS2*, indicating the strong co-expression as well as continuity of the passing of cancer cells through a Slug-based EMT in solid tumors, and the total absence of the co-expression of these genes otherwise. A, B and C, plots from three solid tumor datasets. D, plot from a leukemia dataset.

Figure 2: Color-coded scatter plots in human and mouse of the 18 samples for the expression of the EMT inducing transcription factor Slug (*SNAI2*) in terms of the expression of the main signature genes *COL11A1* and *THBS2*. A, demonstration that this co-expression is present in the xenografted human cells. B, demonstration that this co-expression is absent in the peritumoral mouse cells. C, Bar diagram indicating that other EMT inducing transcription factors are not co-expressed.

Figure 3: Heat map combining human and mouse genes. The 29 human genes include many EMT factors and were found to be significantly co-expressed in the cancer cells.

Figure 1

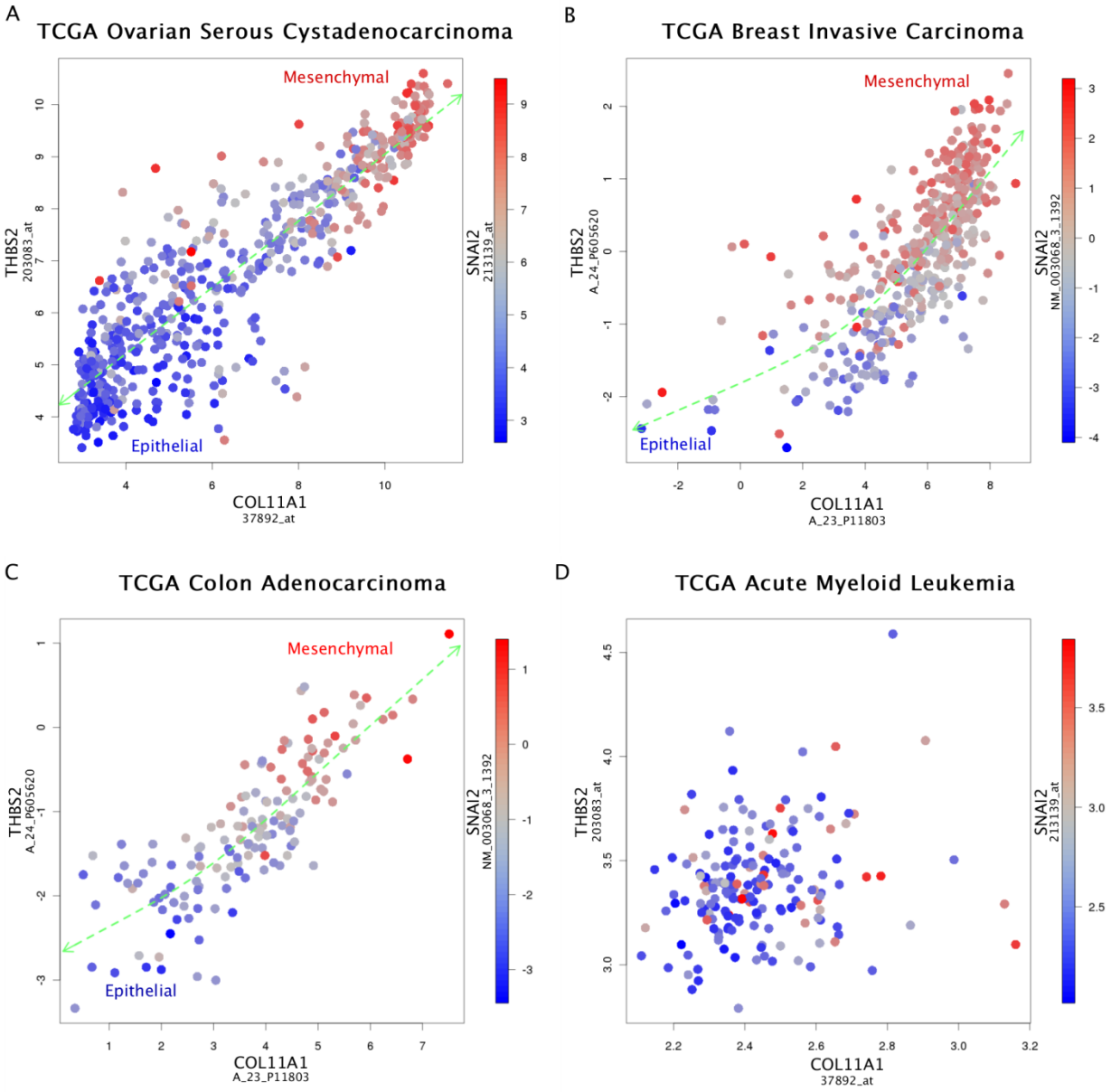


Figure 2

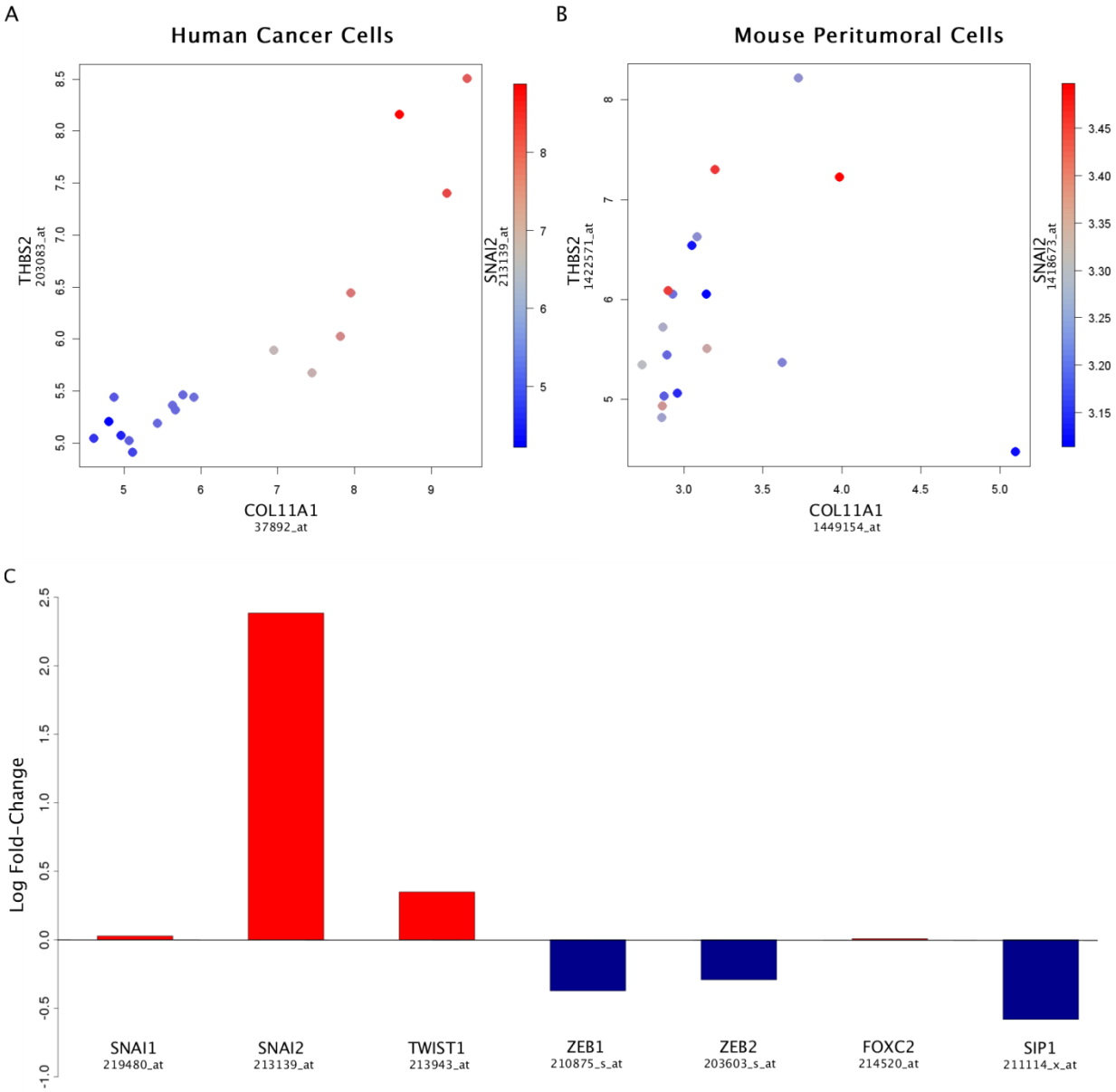
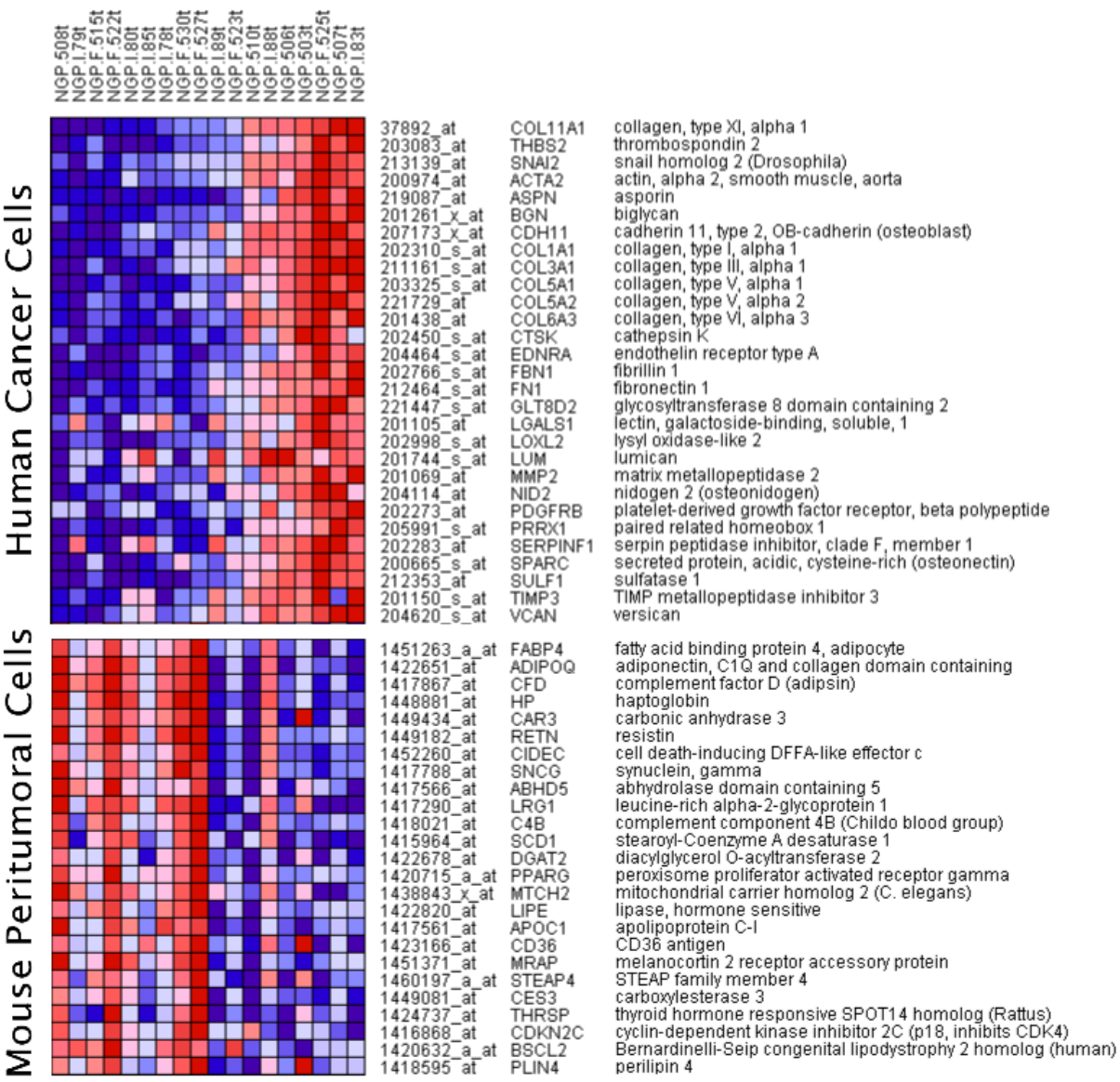


Figure 3



Tables

Table 1: Top genes overexpressed in the fibroblastic transition signature (see text for designation of genes with boldface and underline)

Rank	Probe set	Gene	Rank	Probe set	Gene
1	37892_at	COL11A1	33	202998_s_at	LOXL2
2	203083_at	THBS2	34	201438_at	COL6A3
3	217428_s_at	COL10A1	35	209596_at	MXRA5
4	221729_at	**COL5A2**	36	213764_s_at	MFAP5
5	210511_s_at	INHBA	37	204589_at	NUAK1
6	213909_at	LRRC15	38	217762_s_at	RAB31
7	212488_at	COL5A1	39	201150_s_at	TIMP3
8	204619_s_at	VCAN	40	221541_at	CRISPLD2
9	209955_s_at	**FAP**	41	205422_s_at	ITGBL1
10	202311_s_at	COL1A1	42	207173_x_at	**CDH11**
11	203878_s_at	MMP11	43	213338_at	TMEM158
12	210809_s_at	**POSTN**	44	202363_at	**SPOCK1**
13	202404_s_at	**COL1A2**	45	204051_s_at	SFRP4
14	202952_s_at	ADAM12	46	202283_at	SERPINF1
15	215076_s_at	**COL3A1**	47	209335_at	**DCN**
16	215446_s_at	LOX	48	219655_at	C7orf10
17	210495_x_at	**FN1**	49	219561_at	**COPZ2**
18	201792_at	AEBP1	50	219773_at	NOX4
19	212353_at	SULF1	51	204464_s_at	EDNRA
20	202766_s_at	**FBN1**	52	200974_at	**ACTA2**
21	219087_at	ASPN	53	202273_at	PDGFRB
22	200665_s_at	SPARC	54	61734_at	RCN3
23	202450_s_at	CTSK	55	213139_at	**SNAI2**
24	206026_s_at	**TNFAIP6**	56	220988_s_at	C1QTNF3
25	222020_s_at	HNT	57	205713_s_at	COMP
26	206439_at	EPYC	58	201105_at	LGALS1
27	201069_at	**MMP2**	59	213869_x_at	**THY1**
28	205479_s_at	PLAU	60	202465_at	**PCOLCE**
29	218469_at	**GREM1**	61	209156_s_at	COL6A2
30	201261_x_at	**BGN**	62	221447_s_at	GLT8D2
31	213125_at	OLFML2B	63	204114_at	NID2
32	201744_s_at	LUM	64	205991_s_at	**PRRX1**