A subset of co-expressed genes in Slug-based cancer mesenchymal transition signature remains coexpressed in normal samples in a tissue-specific manner

Wei-Yi Cheng[1] and Dimitris Anastassiou[1*]

1 Center for Computational Biology and Bioinformatics and Department of Electrical Engineering, Columbia University, New York, NY, USA

* Corresponding author email: anastas@ee.columbia.edu

Abstract

A recently identified gene expression signature of EMT markers containing the transcription factor Slug was found present in samples from many publicly available cancer gene expression datasets of multiple cancer types except leukemia. We also found many of these genes co-expressed in human cancer xenografted cells, but not in mouse stroma cells, suggesting that the signature is largely produced by cancer cells undergoing some type of EMT. Here we report that a partial signature consisting of a subset of the co-expressed genes of the full signature, including at least Slug (SNAI2), collagens COL1A1, COL1A2, COL3A1, COL6A3 and genes DCN and LUM, is also present in leukemia, in which case it is also strongly associated with the chemokine CXCL12 (aka SDF1). The same subset of co-expressed genes is also strongly present even in normal samples in a tissue-specific manner, with lowest expression in brain tissues and highest expression in reproductive system tissues. The full signature, with prominent presence of COL11A1, THBS2 and INHBA appears to be triggered in solid cancers particularly when cancer cells encounter adipocytes.

Introduction

A gene expression signature consisting of a set of genes many of which are epithelial-mesenchymal transition (EMT) markers including the EMT-inducing transcription factor Slug (SNAI2) was recently identified [1]. The signature is present in publicly available datasets of all solid cancer types that we tried. These genes are coordinately expressed at various levels, but are significantly overexpressed only in samples that have exceeded a particular stage, specific to

each cancer type. A list of the top 64 genes of the signature is shown, for reference, in Table 1. We also found [2] that many of the genes of the signature, including α-SMA, are expressed by the cancer cells themselves *in vivo*, at least in one xenograft model of neuroblastoma that we tried [3], confirming that cancer cells have undergone a mesenchymal transition (the term EMT may not be accurate in this case, as the same signature is also present in cancers, such as neuroblastoma, that are not, strictly speaking, epithelial). We refer to this multi-cancer signature as the "cancer mesenchymal transition signature." It is characterized by a prominent presence of co-expressed genes COL11A1, THBS2, INHBA.

Consistent with this hypothesis, the signature is not found present in either leukemia or normal tissues. We hypothesized, however, that a subset of these genes may still remain co-expressed with Slug even in nonsolid cancers. Such findings would shed light on the more general biological mechanism of Slug-based mesenchymal transition when it happens without the full myofibroblast-like transdifferentiation that occurs in invasive tumors (in which case the transformed cancer cells may be confused with stromal cancer associated fibroblasts).

Results

Here we report that further computational analysis in gene expression datasets of leukemia samples, as well as of normal human body multi-tissue datasets, revealed that a subset of the cancer mesenchymal transition signature, consisting at least of Slug, collagens COL1A1, COL1A2, COL3A1, COL6A3 and genes DCN and LUM remain strongly co-expressed in all cases. These results were reached by simply comparing the pairwise associations of the 64 genes of the full signature of Table 1 in various datasets. Furthermore, ranking the genes in terms of their association (such as mutual information) with SNAI2 in each data set consistently identifies the above seven genes close to the top of the list.

To demonstrate the coexpression of the above seven genes, we generated corresponding heat maps including only these genes.

The heat map of the TCGA leukemia dataset is shown in the Supplementary Figure available at www.ee.columbia.edu/~anastas/np/heatmap.leukemia.7genes.pdf We observed that the same signature is also strongly associated in leukemia with the chemokine CXCL12 (aka SDF1), providing related biological insights.

The heat map of the 504 normal samples from the "human body index" GSE7307 dataset is shown in the Supplementary Figure available at www.ee.columbia.edu/~anastas/np/heatmap.normal.7genes.pdf

Remarkably, there is strong tissue specificity associated with the presence of the Slug-based EMT signature. At one end of the spectrum where (Table 2, left side), the signature appears absent in brain tissues. At the other end of the spectrum (Table 2, right side), the signature appears present in higher amounts in reproductive system tissues. This phenomenon is fully demonstrated in the color-coded table of all 504 samples that appears as Supplementary Table available at www.ee.columbia.edu/~anastas/np/snai2tissuetype.xlsx

Because EMT is known to generate cells with properties of stem cells [4], these results are compatible with the notion that, among normal human body tissues, the brain contains highly differentiated cells with the least amount of "stemness," while the reproductive system contains cells with the highest amount of stemness. The lack of stemness in the normal state of brain cells is also consistent with our recent finding that the absence of the full Slug-based cancer mesenchymal transition signature in glioblastoma is associated with prolonged time to recurrence following treatment [5]. Glioblastoma samples reveal coexpression of most of the genes of Table 1, but COL11A1 is not as strongly coexpressed.

We have also hypothesized [2, 3] that the full Slug-based cancer mesenchymal transition signature is triggered in solid cancers following contextual microenvironmental interactions when cancer cells encounter adipocytes. In that case, COL11A1 can serve as a reliable proxy of the full signature.

References

1. **Kim H, Watkinson J, Varadan V, Anastassiou D: Multi-cancer computational analysis reveals invasion-associated variant of desmoplastic reaction involving INHBA, THBS2 and COL11A1. *BMC Med Genomics* 2010, 3:51.**
2. **Anastassiou D, Rumjantseva V, Cheng WY, Huang J, Canoll PD, Yamashiro DJ, Kandel JJ: Human Cancer Cells Express Slug-based Epithelial-Mesenchymal Transition Gene Signature Obtained in Vivo. Available from Nature Precedings <http://hdl.handle.net/10101/npre.2011.6243.1 >. 2011.**
3. **Anastassiou D, Rumjantseva V, Cheng W, Huang J, Canoll PD, Yamashiro DJ, Kandel JJ: Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo. *BMC Cancer* 2011, 11(1):529.**
4. **Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M *et al*: The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008, 133(4):704-715.**
5. **Cheng WY, Kandel J, Yamashiro DJ, Canoll P, Anastassiou D: Slug-based epithelial-mesenchymal transition gene signature is associated with prolonged time to recurrence in glioblastoma. Available from Nature Precedings <hdl.handle.net/10101/npre.2011.6544.1>. 2011.**

Table 1. Top genes coexpressed in the cancer mesenchymal transition signature.

Rank	Gene	Rank	Gene
1	COL11A1	33	LOXL2
2	THBS2	34	COL6A3
3	COL10A1	35	MXRA5
4	COL5A2	36	MFAP5
5	INHBA	37	NUAK1
6	LRRC15	38	RAB31
7	COL5A1	39	TIMP3
8	VCAN	40	CRISPLD2
9	FAP	41	ITGBL1
10	COL1A1	42	CDH11
11	MMP11	43	TMEM158
12	POSTN	44	SPOCK1
13	COL1A2	45	SFRP4
14	ADAM12	46	SERPINF1
15	COL3A1	47	DCN
16	LOX	48	C7orf10
17	FN1	49	COPZ2
18	AEBP1	50	NOX4
19	SULF1	51	EDNRA
20	FBN1	52	ACTA2
21	ASPN	53	PDGFRB
22	SPARC	54	RCN3
23	CTSK	55	SNAI2
24	TNFAIP6	56	C1QTNF3
25	HNT	57	COMP
26	EPYC	58	LGALS1
27	MMP2	59	THY1
28	PLAU	60	PCOLCE
29	GREM1	61	COL6A2
30	BGN	62	GLT8D2
31	OLFML2B	63	NID2
32	LUM	64	PRRX1

Table 2. Listing of the 20 samples at each extreme of the set of the 504 samples sorted in terms of Slug expression. At the left side of the Table the 20 samples with the minimum Slug expression are all brain tissues. At the right side of the Table, the 20 samples with the maximum Slug expression are all reproductive system tissues.

Rank	Sample	Tissue	SNAI2
1	GSM175958	midbrain	-2.39
2	GSM176371	putamen	-0.96
3	GSM176369	caudate	-0.68
4	GSM176150	putamen	-0.21
5	GSM175851	accumbens	-0.20
6	GSM176394	substantia nigra_pars compacta	0.16
7	GSM176370	accumbens	0.17
8	GSM176046	putamen	0.23
9	GSM176401	substantia nigra_pars compacta	0.41
10	GSM175848	putamen	0.83
11	GSM176073	occipital_lobe	0.86
12	GSM176454	thalamus_lateral nuclei	0.95
13	GSM176364	accumbens	0.99
14	GSM176071	ventral_tegmental_area	1.00
15	GSM176061	thalamus	1.02
16	GSM176447	gloubus pallidum_external	1.03
17	GSM176398	substantia nigra_reticulata	1.08
18	GSM176451	thalamus_subthalamic nucleus	1.11
19	GSM176379	accumbens	1.20
20	GSM175959	midbrain	1.32

Rank	Sample	Tissue	SNAI2
475	GSM176134	vagina	8.87
476	GSM176270	penis	8.94
477	GSM176043	endometrium	9.00
478	GSM176133	myometrium	9.02
479	GSM176145	myometrium	9.02
480	GSM176131	ovary	9.05
481	GSM175833	cervix	9.13
482	GSM176106	myometrium	9.15
483	GSM176141	endometrium	9.19
484	GSM176146	myometrium	9.21
485	GSM175878	vagina	9.21
486	GSM176318	ovary	9.27
487	GSM176144	myometrium	9.41
488	GSM176254	myometrium	9.44
499	GSM176230	myometrium	9.46
500	GSM176143	myometrium	9.61
501	GSM176135	cervix	9.85
502	GSM176108	myometrium	9.86
503	GSM176320	myometrium	9.87
504	GSM176102	myometrium	11.25