

# CEACAM Gene Family: A Circuitous Journey towards Metastasis in Breast Cancer

## Abstract

The ubiquitous up-regulation of CEACAM6 in colon, pancreatic, breast and lung cancer is well established. This protein is known for its invasive and metastatic properties in pancreatic adenocarcinoma as well as in breast cancer. We propose that the over-expression of CEACAM5 and CEACAM6 are a pre-requisite for invasive and metastatic behavior of breast cancer. We have conducted bioinformatics studies to compare the expression profiles of CEA gene family members in sets of RNA-seq data for MCF10A (non-tumorigenic epithelial cell line) and MCF7 (human breast cancer cell line) obtained from European Nucleotide Archives. RNA-seq data was mapped using *HISAT2* followed by alignment and abundance analysis using *Stringtie* and visualized using *ballgown* package in *R* software environment. Specifically, we observed a 4.5-fold up-regulation in CEACAM5 expression while 7-fold increase was recorded for CEACAM6 expression. We propose that the up-regulation of both these proteins in MCF7 cell line compared to MCF10A implicates their inconspicuous role in tumorigenesis, enhanced invasiveness and thus, leading to increased propensity towards breast cancer metastasis. Further studies are required in breast cancer cell lines and appropriate animal models to validate these *in silico* observations.

**Keywords:** CEACAM5; CEACAM6; Metastasis; Bioinformatics; Breast Cancer; Tumor Biomarkers

## Research Article

Volume 5 Issue 4 - 2017

**Waqas Iqbal<sup>1</sup>, Saleh Alkarim<sup>1</sup>, Hani SH Mohammed Ali<sup>1</sup> and Kulvinder S Saini<sup>1,2\*</sup>**

<sup>1</sup>Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia

<sup>2</sup>School of Biotechnology, Eternal University, Baru Sahib-173101, HP India

\*Corresponding author: Kulvinder Saini, King Abdulaziz University, Jeddah, Saudi Arabia, Tel: 966-555-880-173; Fax: 966-264-007-56; Email: sainikulvinder@gmail.com

Received: March 30, 2017 | Published: May 04, 2017

**Abbreviations:** CEA: Carcinoembryonic Antigens; Ig: Immunoglobulin; CRC: Colorectal Cancer; CEACAMs: Carcinoembryonic Antigen related Cell Adhesion Molecules; CSCs: Cancer Stem Cells; PSGs: Pregnancy-Specific Glycoproteins; ENA: European Nucleotide Archive

## Introduction

CEA gene family (CEA) belonging to immunoglobulin (Ig) supergene family was identified more than 50 years ago, comprises of 35 genes/pseudo genes (21 are protein coding) located on chromosome 19 (between q13.1-13.3), with wide range of patho-physiological functions [1,2]. Despite the over-expression of various CEA genes in very diverse cancers (breast, colon, prostate, pancreas, stomach, ovary, lung & medullary), its primary application as a serum biomarker is confined to the diagnosis & prognosis of colorectal cancers (CRC), and in the detection of liver metastasis. CEA gene family has two groups, CEACAMs (carcinoembryonic antigen related cell adhesion molecules) and PSGs (pregnancy-specific glycoproteins). The 12 CEACAMs subgroup encoded proteins exhibit one variable domain known as the N domain, with the only exception of CEACAM16 that consists of two N domains. The N domain is either followed by none or C2-like Ig domains, referred to as A or B. These extracellular domains usually act as intercellular adhesion molecules in epithelial, endothelial, dendritic and leukocytes [3,4]. CEACAM5 (CEA) comprises of one N domain followed by six C2-like domains (A1, B1, A2, B2, A3 and B3) [5-8], whereas CEACAM6 has only two C2-like domains, termed as A and B [1,9].

CEA gene family members are involved in diverse pathophysiological functions [4,10], including as receptors

for microbial pathogens [11]. They play a significant role in carcinogenesis, particularly in cancer detection, progression and metastasis [12,13]. Gold and Freedman [14], were the first to discover CEACAM5 in the blood of colon cancer patients and further research established that its Overexpression in numerous malignancies is usually correlated with poor prognosis, and increased mortality [8,14,15]. In prostate and in colorectal cancers, CEACAM 5 over-expression was documented as an excellent tumor biomarker [16,17], although it may not be useful as a standalone early screening tool for CRC [18]. Additional evidence about the Overexpression of CEACAM6 in CRC is also associated with increased invasiveness and liver metastasis [19]. CEACAM6 Overexpression has been reported in a number of different malignancies, such as-breast, pancreatic, ovarian, lung and gastric adenocarcinomas [20]. Individually and sometimes together, CEACAM5 and CEACAM6 [21] are also associated with adhesion, invasion and metastasis in pancreatic, colon and breast cancers. In this regard, another study validated the effects of three monoclonal antibodies specifically targeting and blocking two domains (NH2-terminal, A1B1 domains) of CEACAM5/CEACAM6 and A3B3 domain present solely on CEACAM5 [22]. The inhibition of these specific domains affects invasiveness, extravasation and metastases *in vitro* as well as *in vivo* [21,22].

Analysis of differential gene expression data obtained by high-throughput sequencing requires fast, reliable and accurate software tools to have meaningful clinical applications. This has led to the development of numerous open-source software tools as well as proprietary technologies. In this study, we procured, stored and mined data, from the newly developed pipeline for raw RNA-seq data analysis from open-source tools. From analyzing raw reads to visualization, *HISAT2*, *Stringtie* and *Ballgown* pipeline has

been regarded as the best “New Tuxedo package” superseding the original tuxedo package (*TopHat2-Cufflinks*) [23]. We carried out bioinformatics analysis to evaluate the up-regulation of CEACAM5 and CEACAM6 in MCF7-metastatic cell line, as compared to MCF10A-normal epithelial cell line, using these new tools. Our data corroborates and validate these earlier “wet lab” studies, that these two proteins are not just great tumor biomarkers, but also actively involved in metastatic cells’ initiation, invasion and colonies propagation at secondary malignant tissue sites.

## Materials and Methods

### Cell line Samples

Our datasets contained two breast cell lines with three replicates each. MCF10A is a, non-tumorigenic, normal epithelial cell line, whereas MCF7 is a metastatic breast cancer cell line.

### RNA-seq Data Analysis

Fastq files were downloaded from ENA (European nucleotide

archive) [24]. Using *HISAT2* [25], the fastq files were mapped to human reference genome. The SAM files obtained were sorted and converted into BAM files using *Samtools* [26]. BAM files thus obtained were aligned using a reference file, annotated, merged, and the estimation for abundance was calculated using *Stringtie* [27], followed by differential gene expression analysis using *ballgown* package in R open source programming language [23,28,29].

## Results

Raw reads obtained from ENA (Table 1) were aligned using *HISAT2* with pre-built human genome index [H. Sapiens, UCSC hg38] downloaded from their website. The output SAM files containing the transcripts analyzed using *Stringtie* and *Ballgown* package in R programming software showed a substantial differential expression of CEACAM5 (upregulated 4.5 fold) and CEACAM6 (upregulated ~7 fold) genes in MCF7 cell line compared to MCF10A, normal epithelial cell line (Table 2).

**Table 1:** GEO series and SRA raw read files. GEA series represents series accession number. GEO sample denotes sample accession number whereas run accession is the unique number given to each sample. Raw data for each sample was downloaded from ENA and analyzed.

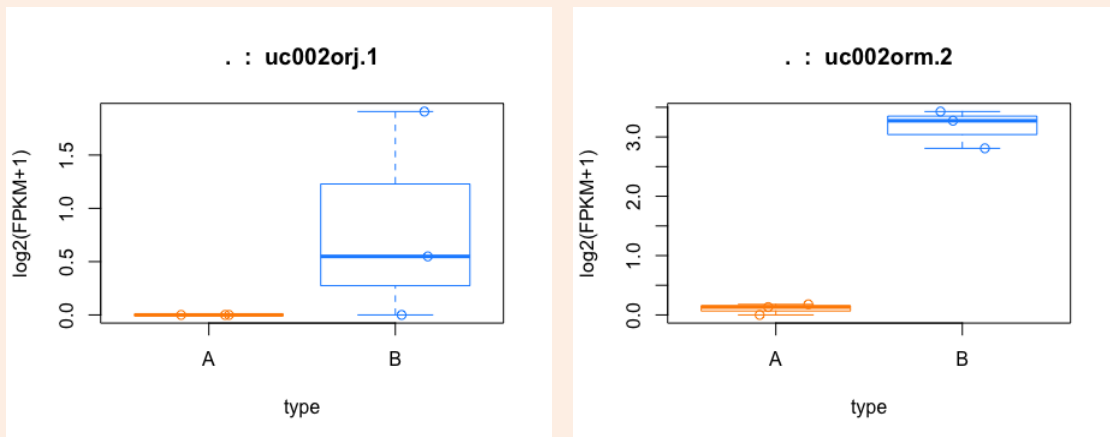
Differential Gene Expression between MCF10A and MCF7 cells			
GEO Series	GEO Sample	Run Accession	Cell Line
GSE71862	GSM1847015	SRR2149928	MCF10
	GSM1847016	SRR2149929	MCF10
	GSM1847017	SRR2149930	MCF10
	GSM1847018	SRR2149931	MCF7
	GSM1847019	SRR2149932	MCF7
	GSM1847020	SRR2149933	MCF7

**Table 2:** *Ballgown* output file in tabular format. Fastq files were analyzed using *HISAT2*, *Stringtie* and *Ballgown* pipeline. CEACAM5 & CEACAM6 expression data after comparative analyses between MCF10A & MCF7 cell lines are reported. Both CEACAM5 and CEACAM6 were upregulated in MCF7 cell line. Data obtained was considered significant at  $p$  value  $<0.05$ . Table summarizes gene names, Fc is fold change observed and denotes differential expression for both the transcripts as  $\log_2$

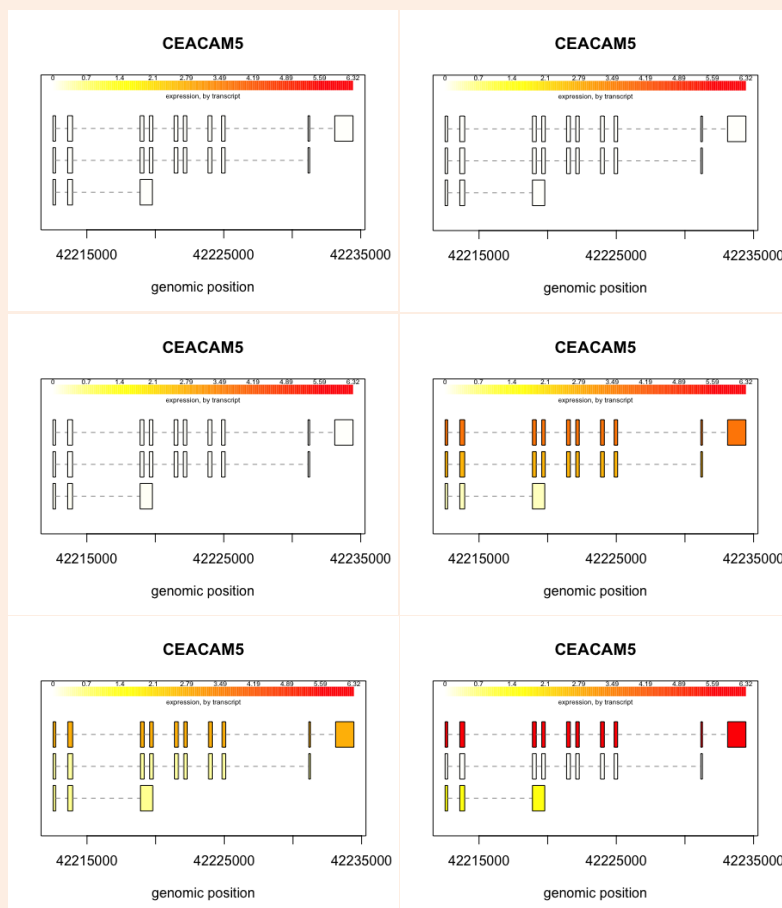
Gene_name	UCSC_id	Fc	pval	qval	de	Regulation in MCF7
CEACAM5	uc002orj.1	4.50585722	0.004114438	0.061435608	2.171801599	Up-regulated
CEACAM6	uc002orm.2	7.010235172	0.000864163	0.032530362	2.809462843	Up-regulated

We created box plots for these two genes to observe the distribution of gene expression data for each sample in our data set. CEACAM5 had a higher expression in two of the biological replicates of MCF7 cell line whereas all the biological replicates had higher CEACAM6 expression pattern in MCF7 cell line, as compared to MCF10A (Figure 1). Next we collated and analyzed the expression of each individual transcript isoform for CEACAM5 and CEACAM6, identified in our study, to delineate the expression

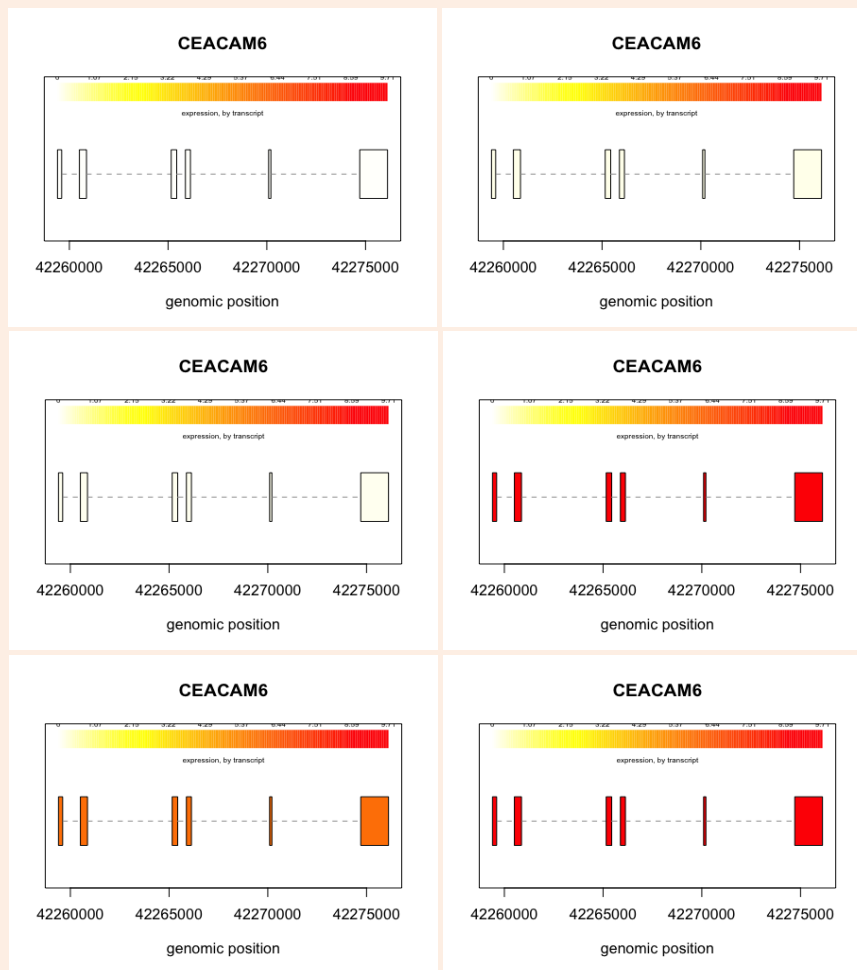
pattern of each isoform in all the six samples in our data set. The three isoforms identified for CEACAM5 were upregulated in MCF7 cell line as compared to MCF10A. However, we were able to obtain only one transcript for CEACAM6 gene that too was upregulated in MCF7 cell line (Figure 2,3). We finally plotted the mean expression patterns of transcript isoforms for both CEACAM5 and CEACAM6 from our datasets to depict the relative expression of each isoform in both groups (Figure 4).



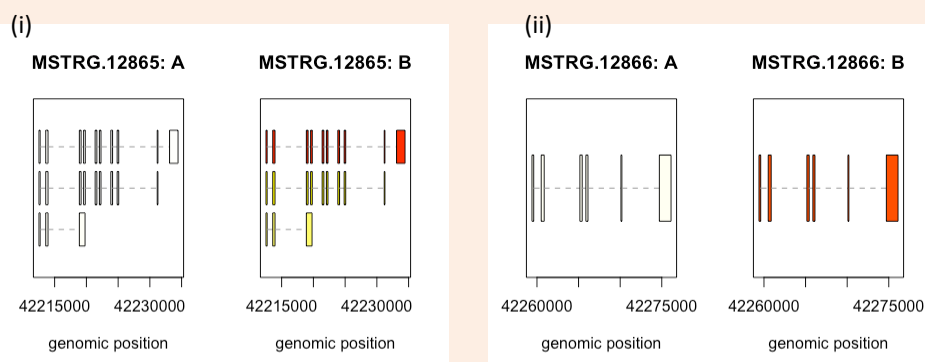
**Figure 1:** Distribution of *FPKM* values. Box plots depicts the distribution of *FPKM* (Fragments Per Kilobase of transcripts per Million mapped reads) values in both MCF10A and MCF7 samples for transcripts uc002orj.1 and uc002orm.2 from CEACAM5 and CEACAM6 genes respectively. Here type represents MCF10A (A) & MCF7 (B).



**Figure 2:** Expression levels of isoforms. CEACAM5 transcripts in MCF10A (a-c) and MCF7 (d-f). The structure and levels of expression of three isoforms of CEACAM5 gene in all six samples are shown individually. Color intensities depict expression levels where lighter shade represents lower expression while darker shade denotes higher expression. Highest expression was observed for the first isoform in MCF7 cell line, indicated by darker shade (d).



**Figure 3:** Expression levels of isoforms. CEACAM6 transcripts in MCF10A (a-c) and MCF7 (d-f). Structure and expression levels of one isoform of CEACAM6 gene in all the six samples are shown. Color intensities depict expression levels, where lighter shade represents lower expression while darker shade denotes higher expression. Highest expression was observed in sample d and f in MCF7 cell line, indicated by darker shade.



**Figure 4:** Plots depicting mean expression patterns. CEACAM5 & CEACAM6 expression for all the transcripts between the two groups.

- i. MSTRG.12865:A and MSTRG.12865:B represents CEACAM5 in MCF10A and MCF7 respectively while
- ii. MSTRG.12866:A and MSTRG.12866:B represents CEACAM6 in MCF10A and MCF7 respectively. Color intensities depict expression levels where lighter shade represents lower expression while darker shade denotes higher expression. Highest expression was observed in the first isoform of CEACAM5 and CEACAM6 in MCF7 cell line as indicated by darker shades.

## Discussion

During the initiation of liver metastasis, CEACAM5 (CEA) exerts its action by binding to its receptor (CEAr)-a protein related to the hnRNP M family of RNA binding proteins. CEA-CEAr interactions lead to the activation and production of pro- and anti-inflammatory cytokines, primarily IL-1, IL-6, IL-10 and TNF- $\alpha$  [30]. Taken together, these cytokines modify the micro-environment of hepatocytes & Kupffer cells, and their cell-cell interactions with the hepatic sinusoids. These interactions not only affect the tumor cells, or other liver cells, but also seem to promote the survival of CSCs and other circulating tumor cells in the blood stream. As proposed by Thomas, et al. [30], down-regulating these cytokines, particularly IL-6 and IL-10, in hepatic sinusoids prior to curative surgery for colorectal cancers has added benefit of causing reduced relapse in certain patients?

Among the CEA gene family members, CEACAM5 & CEACAM6 are overexpressed in many cancers, and have been found to be unique mediators during tumor cell adhesion and metastasis [3,4,22]. In this study where we evaluated the expression pattern of CEACAM5 and CEACAM6 in metastatic breast cancer cell line in comparison to a normal epithelial cell line, both these genes were upregulated in MCF7 breast cancer cell line, as observed by others [20,31]. We further assessed the expression at the transcript level, observing the up-regulation of different isoforms identified in this study. All the three isoforms for CEACAM5 and one-isoform for CEACAM6 were over-expressed in MCF7 cell line. Moreover the transcript level expression of CEACAM6 gene was higher than that of CEACAM5 as reported by Blumenthal, et al. [20]. Increased expression of CEAs in various malignancies implicates their role in epithelial malignancies. Nevertheless higher expression of CEACAM5 and CEACAM6 distorts normal tissue architecture [32,33] and might lead to alterations in epithelial-mesenchymal-transition, thereby setting up the stage for the initiation of metastasis. Other possible explanation could be that increased expression of these 2 CEACAMs might exacerbates metastasis in colon cancer by inhibiting immune cells' response against colon cancer cells [34]. Taken together, therapeutic approaches aimed at down-regulating CEACAM5/CEACAM6 will help us restrain the metastatic process.

## Conclusion and way forward

Reverse-transcriptase PCR (RT-PCR) assays have been developed to detect CEA from circulating tumor cells in blood and detailed application of this technology on CSCs and metastatic cancer stem cells is imminent. Single-cell sequencing, next-generation sequencing and stage-specific gene expression analyses for both RNA-miRNA-transcriptomes, could lead to a better understanding of contextual genetic cues promoting interactions of various tissue cell types, e.g., liver cell types (hepatocytes, Kupffer cells, sinusoids, endothelial cells, etc) with the metastatic tumor cells and metastatic CSCs. Similarly, the role of CEA gene family, particularly CEACAM-5&6, during the initiation, progression and invasiveness at secondary tissue sites emanating from the spread of breast cancer metastasis require additional molecular analyses using appropriate transgenic mouse models. A better understanding of these two CEACAMs

will undoubtedly will give us better therapeutic and monitoring tools for the management of metastatic process, which remains a challenging "black box" for the cancer researchers and oncologists.

## Acknowledgement

Computational analyses using *HISAT2* and *Stringtie* software were performed with "Aziz Supercomputer" at King Abdulaziz University (KAU) High Performance Computing Center (<http://hpc.kau.edu.sa>). We are grateful to the Dr. Rashid Mehmood-Prof. of Big Data Systems, Deanship of Scientific Research (DSR) and Dean Graduate Studies (DGS) at KAU for their support of this project.

## References

1. Hammarstrom S (1999) The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. *Semin Cancer Biol* 9(2): 67-81.
2. Pavlopoulou A, Scorilas A (2014) A Comprehensive Phylogenetic and Structural Analysis of the Carcinoembryonic Antigen (CEA) Gene Family. *Genome Biol Evol* 6(6): 1314-1326.
3. Beauchemin N, Draber P, Dveksler G, Gold P, Gray-Owen S, et al. (1999) Redefined nomenclature for members of the carcinoembryonic antigen family. *Exp Cell Res* 252(2): 243-249.
4. Obrink B (1997) CEA adhesion molecules: Multifunctional proteins with signal-regulatory properties. *Curr Opin Cell Biol* 9(5): 616-626.
5. Zimmermann W, Ortlieb B, Friedrich R, Von Kleist S (1987) Isolation and characterization of cDNA clones encoding the human carcinoembryonic antigen reveal a highly conserved repeating structure. *Proc Nat Acad Sci* 84(9): 2960-2964.
6. Thompson JA, Pande H, Paxton RJ, Shively L, Padma A, et al. (1987) Molecular cloning of a gene belonging to the carcinoembryonic antigen gene family and discussion of a domain model. *Proc Nat Acad Sci* 84(9): 2965-2969.
7. Beauchemin N, Benchimol S, Cournoyer D, Fuks A, Stanners C P (1987) Isolation and characterization of full-length functional cDNA clones for human carcinoembryonic antigen. *Mol Cell Biol* 7(9): 3221-3230.
8. Thomas P, Toth CA, Saini KS, Jessup JM, Steele G (1990) The structure, metabolism and function of the carcinoembryonic antigen-gene family. *Biochim Biophys Acta* 1032: 177-189.
9. Zhou H, Fuks A, Alcaraz G, Bolling TJ, Stanners CP (1993) Homophilic adhesion between Ig superfamily carcinoembryonic antigen molecules involves double reciprocal bonds. *The J Cell Biol* 122(4): 951-960.
10. Kuespert K, Pils S, Hauck CR (2006) CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol* 18: 565-571.
11. Bos MP, Hogan D, Belland RJ (1999) Homologue scanning mutagenesis reveals CD66 receptor residues required for neisserial Opa protein binding. *J Exp Med* 190: 331-340.
12. Beauchemin N, Arabzadeh A (2013) Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev* 32: 643-71.
13. Michaelidou K, Tzouvaras A, Missitzis I, Ardavanis A, Scorilas A (2013) The expression of the CEACAM19 gene, a novel member of the CEA family, is associated with breast cancer progression. *Int J Oncol* 42: 1770-1777.



14. Gold P, Freedman SO (1965) Specific carcinoembryonic antigens of the human digestive system. *The J Exp Med* 122: 467-481.
15. Chevinsky AH (1991) CEA in tumors of other than colorectal origin. *Semin Surg Oncol* 7: 162-166.
16. Stark KA, Weaver A, Hoffman HM, Krauss R, Valenzuela DB, et al. (2001) Cell Adhesion-Mediating Proteins and Polynucleotides encoding them. United States Patent and Trademark Office (application no. 60/289, 179; filed on May 07, 2001) and Patent Cooperation Treaty (PCT) (International Publication Date: November 14, 2002; International Publication Number: WO 02/090508 A2).
17. Tiernan JP, Perry SL, Verghese ET, West NP, Yeluri S, et al. (2013) Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting. *Br J Can* 108(3): 662-667.
18. Thomas DS, Fourkala EO, Apostolidou S, Gunu R, Ryan A, et al. (2015) Evaluation of serum CEA, CYFRA21-1 and CA125 for the early detection of colorectal cancer using longitudinal preclinical samples. *Br J Cancer* 113(2): 268-274.
19. Kim KS, Kim JT, Lee SJ, Kang MA, Choe IS, et al. (2013) Overexpression and clinical significance of carcinoembryonic antigen-related cell adhesion molecule 6 in colorectal cancer. *Clin Chim Acta* 415: 12-19.
20. Blumenthal RD, Leon E, Hansen HJ, Goldenberg DM (2007) Expression patterns of CEACAM5 and CEACAM6 in primary and metastatic cancers. *BMC Cancer* 7: 2.
21. Lewis-Wambia JS, Cunliffe HE, Kima HR, Willis AL, Jordan VC (2008) Overexpression of CEACAM6 promotes migration and invasion of oestrogen deprived breast cancer cells. *Eur J Cancer* 44(12): 1770-1779.
22. Blumenthal RD, Hansen HJ, Goldenberg DM (2005) Inhibition of Adhesion, Invasion, and Metastasis by Antibodies Targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). *Cancer Res* 65(19): 8809-17.
23. Perteau M, Kim D, Perteau GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11: 1650-1667.
24. Barutcu AR, Lajoie BR, McCord RP, Tye CE, Hong D, et al. (2015) Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* 16: 214.
25. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12: 357-360.
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25: 2078-2079.
27. Perteau M, Perteau GM, Antonescu CM, Chang TC, Mendell JT, et al. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33: 290-295.
28. Frazee AC, Perteau G, Jaffe AE, Langmead B, Salzberg SL et al. (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nature Biotechnology* 33: 243-246.
29. Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL (2015) Informatics for RNA-seq: A web resource for analysis on the cloud. *PLoS Comp Biol* 11(8): e1004393.
30. Thomas P, Forse RA, Bajenova O (2011) Carcinoembryonic antigen (CEA) and its receptor hnRNP M are mediators of metastasis and the inflammatory response in the liver. *Clin Exp Metastasis* 28: 923-932.
31. Kokkonen N, Ulibarri IF, Kauppila A, Luosujärvi H, Rivinoja A, et al. (2007) Hypoxia upregulates carcinoembryonic antigen expression in cancer cells. *Int J Cancer* 121: 2443-2450.
32. Ordóñez C, Screatón RA, Ilantzis C, Stanners CP (2000) Human carcinoembryonic antigen functions as a general inhibitor of anoikis. *Cancer Res* 60: 3419-3424.
33. Ilantzis C, DeMarte L, Screatón RA, Stanners CP (2002) Deregulated expression of the human tumor marker CEA and CEA family member CEACAM6 disrupts tissue architecture and blocks colonocyte differentiation. *Neoplasia* 4: 151-163.
34. Kammerer R, Von Kleist S (1994) The carcinoembryonic antigen (CEA) modulates effector-target cell interaction by binding to activated lymphocytes. *Int J Cancer* 68(4): 457-463.