BMC Genomics

RESEARCH ARTICLE   **Open Access**

Defining new criteria for selection of cell-based intestinal models using publicly available databases

Jon Christensen[1†], Sara El-Gebali[2†], Manuela Natoli[3†], Thierry Sengstag[4], Mauro Delorenzi[5,6], Susanne Bentz[2,7], Hanifa Bouzourene[8], Martin Rumbo[9], Armando Felsani[3], Sanna Siissalo[10], Jouni Hirvonen[10], Maya R Vila[11], Piercarlo Saletti[12], Michel Aguet[6,13] and Pascale Anderle[2,5,6,7,14*]

Abstract

Background: The criteria for choosing relevant cell lines among a vast panel of available intestinal-derived lines exhibiting a wide range of functional properties are still ill-defined. The objective of this study was, therefore, to establish objective criteria for choosing relevant cell lines to assess their appropriateness as tumor models as well as for drug absorption studies.

Results: We made use of publicly available expression signatures and cell based functional assays to delineate differences between various intestinal colon carcinoma cell lines and normal intestinal epithelium. We have compared a panel of intestinal cell lines with patient-derived normal and tumor epithelium and classified them according to traits relating to oncogenic pathway activity, epithelial-mesenchymal transition (EMT) and stemness, migratory properties, proliferative activity, transporter expression profiles and chemosensitivity. For example, SW480 represent an EMT-high, migratory phenotype and scored highest in terms of signatures associated to worse overall survival and higher risk of recurrence based on patient derived databases. On the other hand, differentiated HT29 and T84 cells showed gene expression patterns closest to tumor bulk derived cells. Regarding drug absorption, we confirmed that differentiated Caco-2 cells are the model of choice for active uptake studies in the small intestine. Regarding chemosensitivity we were unable to confirm a recently proposed association of chemo-resistance with EMT traits. However, a novel signature was identified through mining of NCI60 GI50 values that allowed to rank the panel of intestinal cell lines according to their drug responsiveness to commonly used chemotherapeutics.

Conclusions: This study presents a straightforward strategy to exploit publicly available gene expression data to guide the choice of cell-based models. While this approach does not overcome the major limitations of such models, introducing a rank order of selected features may allow selecting model cell lines that are more adapted and pertinent to the addressed biological question.

Keywords: Cell lines, Genomic profiling, Malignant traits, Epithelial-mesenchymal transition, Intestine, Colon cancer, Chemosensitivity

* Correspondence: pascale.anderle@ibmm.unibe.ch
†Equal contributors
2Institute of Biochemistry and Molecular Medicine, University of Berne, Berne, Switzerland
5Swiss Institute of Bioinformatics, Lausanne, Switzerland
Full list of author information is available at the end of the article

Background

A wide panel of intestinal cell lines is being used to study the biology of the intestine. All of these cell lines are either directly derived from primary colo-rectal cancers (CRCs) of different clinical stages and differentiation grades or from metastatic sites originated from a colon tumor. The major oncogenic pathways in colon cancer include loss of function mutations in APC, TP53 and SMAD4 (approximately 80–85% of sporadic tumors), or DNA mismatch repair genes, and activating mutations in beta-catenin [1]. As a consequence, the Wnt pathway is activated in most tumors and derived CRC cell lines, albeit to a different extent, depending on the genetic lesions [2-5]. The morphology, expression of differentiation markers, migratory characteristics and their potential to form metastases differ vastly between the cell lines [5-8].

While cancer drug discovery has mainly focused on targeting tumor cell proliferation, the outcome of a cancer depends largely on tumor invasion and dissemination [9]. Recent advances in understanding underlying mechanisms in cancer biology including cancer stem cell (CSC) properties and epithelial-mesenchymal transition (EMT) and their relation to drug susceptibility require that relevant traits are considered for choosing appropriate cell-based models. It is believed that at the invasive front, the tumor cells undergo EMT resulting in increased migratory capacity. Furthermore, EMT has recently been linked in breast cancer to stem cell like properties [10] as well as resistance to chemotherapy in different tumor types including CRC [11-14].

The variety of available human cancer cell lines reflects the genomic heterogeneity across the human cancer population at least in part and has therefore regained attention notably to predict responsiveness of anticancer drugs [15,16]. Starting with the landmark paper by Scherf et al. [17] numerous studies followed that aimed at linking drug response in terms of growth inhibition with gene expression signatures, some specifically focusing on colon cancer (e.g. [18-26]). To our knowledge, no study has specifically focused on linking expression of malignant traits (i.e. EMT, WNT activity, stemness signatures) in colon cancer cell lines to response to therapy.

In this study we compare for the first time gene expression signatures relating to a wide panel of commonly known intestinal cell lines, primary cell cultures of human cancer-associated fibroblasts and laser-dissected human colonocytes, small intestinal enterocytes and tumor cells. We delineate selection criteria for CRC derived cell lines based on genomic expression patterns related to clinical parameters, migratory capacities and proliferative activities. While some cell lines are mainly being used to study mechanisms associated with tumor biology others serve as models for normal enterocytes studying drug absorption. Thus, we used the same strategy to assess the validity of models for oral drug absorption.

Results

Most dominant gene expression differences as revealed by principal component analysis (PCA) were observed between normal epithelium and all tumor-derived cell lines, and cells with epithelial versus mesenchymal properties.

PCA was performed to assess the major gene expression differences between microdissected normal and tumor epithelium, primary cancer-associated fibroblast (CAF) cultures and the various CRC cell lines. PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. Hence, the first principal component accounts for as much of the variability in the data as possible and each succeeding component accounts for as much of the remaining variability as possible. Two breast cancer cell lines with known pronounced (MB231) or weak (MCF7) EMT traits [27] were included into the comparison. The first component was defined mainly by the difference between small intestinal enterocytes and all CRC cell lines grown to subconfluency, while the second component was delineated by the difference between CAFs and epithelial cells (Figure 1). T84, HT29 and Caco-2 cells grown to confluency for 3 weeks seemed to be most similar to the small intestinal enterocytes, and also to laser-dissected tumor cells, while SW480 and MB231 proved closest to CAFs.
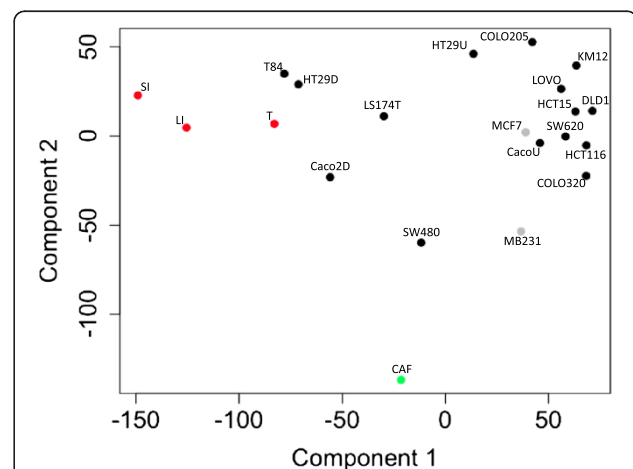

Figure 1 Most dominant gene expression differences between normal epithelium and all tumor-derived cell lines, and cells with epithelial versus mesenchymal properties. Principal component analysis (PCA) of genomic expression data from human colon carcinoma cell lines (black); laser-dissected tissues (red) of normal human colonocytes (LI), tumor cells (T) and small-intestinal enterocytes (SI); primary cell cultures (green, CAFs); breast cancer cell lines (grey). For each sample 2–3 three replicates were measured.

Selection of relevant cell lines to study tumor biology based on expression of signatures related to malignant traits

Next, the suitability of the various cell lines for tumor biology related studies was assessed. There are no standard models and selection criteria are not well defined. The activation of the Wnt pathway is a hallmark in colon carcinogenesis. Recent findings indicate that the Wnt activity defines CRC stem cells and is regulated by the microenvironment [28]. Thus, the overall Wnt signaling activity was compared in the various samples.

When comparing the average expression of direct Wnt targets (i.e. 24 Wnt target genes relevant to human colon and/or other cancers, with a proven direct transcriptional control through TCF binding sites as described in [29]), the SW620 cell line showed the strongest Wnt signal followed by SW480, laser-dissected tumor cells, COLO320, DLD1 and undifferentiated Caco-2 cells (Table 1, detailed description of all gene sets see supplementary data section "expression data analysis" and Additional file 1: Table S2). The ranking of cell lines with strong, medium and low average expression of

Table 1 Rank order of cell lines according to their signature strength

Cell Line	EMT	HuISC	WNT	Trans_BC	SLC_SI3	SLC_PM	ABC	MKI67_CO	IC50_pos	IC50_neg
CAF	1	1	8	29	24	23	30	24	6	7
MB231	2	2	23	30	30	29	27	10	10	21
COLO320	3	11	4	17	27	28	10	4	4	29
SW480	4	4	2	28	25	27	17	17	24	12
SW620	5	9	1	27	29	22	20	8	7	26
Tumor	6	3	3	20	8	24	26	23	5	20
Normal	7	10	26	12	7	14	9	27	2	18
Caco,Fl,D	8	15	14	6	6	5	6	21	8	13
Caco,Fl,U	9	7	6	5	11	6	2	13	16	1
Caco_CDX2,U	10	6	21	7	13	7	7	18	11	23
Caco_CDX2,D	11	25	25	3	3	3	4	28	29	14
JE	12	29	30	13	4	12	22	29	3	30
KM12	13	22	20	10	23	11	21	2	13	5
HCT116	14	12	24	21	26	25	18	14	9	24
Caco,Fi,D	15	26	27	2	2	2	5	26	17	19
T84	16	17	18	26	9	17	29	22	22	25
CacoReady	17	19	16	4	5	4	1	20	15	10
HT29,R	18	16	11	24	16	10	13	5	25	3
DLD1	19	13	5	9	18	15	24	3	18	2
Caco,S	20	27	15	23	22	20	11	9	28	6
HT29,S	21	18	13	18	15	13	12	1	20	4
MCF7	22	24	28	25	28	30	25	15	14	28
HT29,D	23	28	22	22	10	21	16	25	26	8
Caco,R	24	14	12	16	19	18	8	7	27	11
IL	25	30	29	1	1	1	3	30	1	27
HT29,U	26	23	19	19	14	9	23	16	30	17
LOVO	27	5	7	14	21	16	19	6	12	15
LS174T	28	8	10	15	12	26	28	19	23	16
COLO205	29	20	9	11	17	8	15	12	21	22
HCT15	30	21	17	8	20	19	14	11	19	9

EMT = EMT-related genes, HuISC = "Humanized intestinal stem cell signature", WNT = direct Wnt targets, Trans_BC = Transporters known to play important role in oral drug absorption, SLC_SI3 = solute carriers expressed at a significantly higher level in small intestinal samples versus colon samples, SLC_PM = solute carriers of the plasma membrane, ABC = ABC transporters, MKI67_CO: Set of genes correlating with MKI67 expression in colon cancer samples, IC50_pos, IC50_neg: Genes with either strong positive or negative correlation between GI50 values of chemotherapeutics and expression across 60 cell line panel (for details of signatures see Additional file 6, methods). Tumor = laser dissected tumor cells, Normal = laser dissected colonocytes, JE, IL = laser dissected enterocytes of the jejunum or ileum, respectively, D = differentiated, U = undifferentiated, Fi = grown on filters, Fl = grown in flasks, R = resistant to methotraxate, S = sensitive to methotrexate.

direct Wnt targets was in agreement with the Wnt activity based on TOPFLASH assays reported by Rosin-Arbesfeld and co-workers [4]. Moreover, the expression of genes involved in EMT ([29]) and a set of genes, recently being described as "humanized intestinal stem cell signature" (HuISC), which identified recurrent CRC [30], were studied to rank the cell lines as models for tumor cells at the invasive front and/or tumor cells with stemness characteristics. Survival analysis studies showed that risk of recurrence was significantly higher for CRC patients with a strong EMT signature and also correlated with overall survival (Figure 2). As expected CAFs and MB231 (positive controls) had the strongest expression of the EMT signature followed by COLO320, SW480, SW620 and laser-dissected tumor epithelial cells, normal colonocytes and differentiated Caco-2 cells. Though COLO320 cells ranked highest of all intestinal epithelial cells, they also had the highest expression of CDX2, a key player in the maturation of intestinal enterocytes [31]. When the cell lines were ranked to the overall expression of stem cell related genes (HuISC), the five top ranked samples were CAF > MB231 > laser-dissected
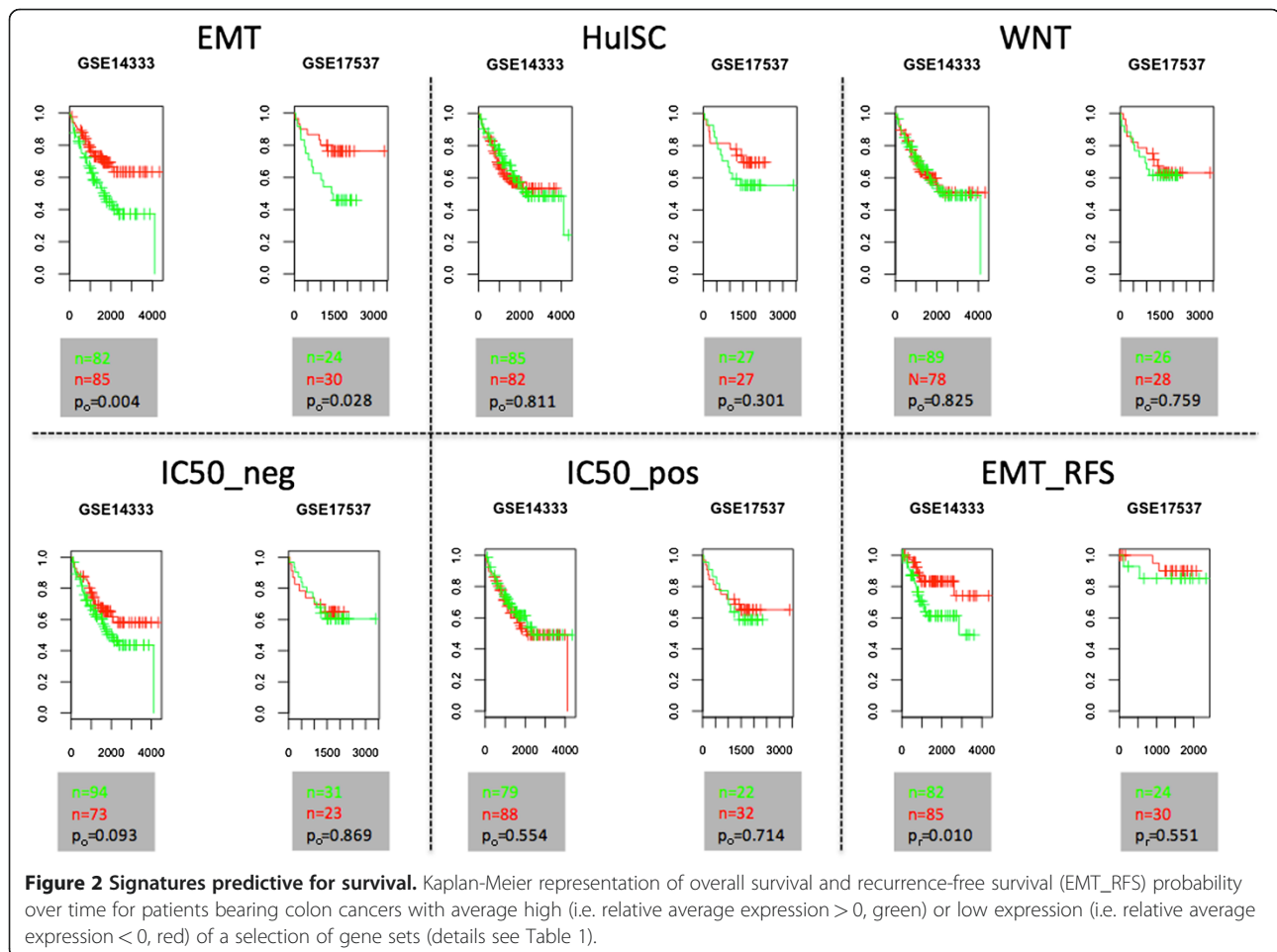
tumor cells > SW480 > LOVO (Table 1). In summary, with respect to the Wnt, EMT and HuISC signatures, SW480 proved the cell line with the strongest traits and a gene expression pattern most similar to that of micro-dissected tumor epithelium.

Gain of malignant traits and loss of differentiation signature were associated with migratory properties

Invasive tumor cells acquire higher motility through EMT [9]. Thus, we elucidated to what extent genomic expression characteristics could be linked to migratory properties. In general, the analysis revealed that higher migratory propensity was associated with strong EMT traits and weak expression of a signature derived from normal small intestinal epithelium (i.e. SLC_SI3). Accordingly, SW480 and MB231 cells showed the highest migratory capacity (Figure 3).

Transporter profile defines absorption model

The model of choice to evaluate drug absorption are Caco-2 cells as they undergo in vitro spontaneous enterocytic differentiation by developing an apical brush
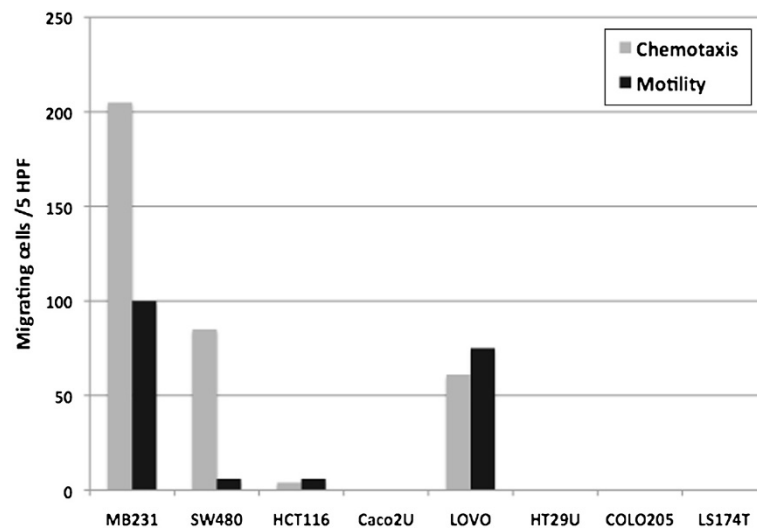

Figure 2 Signatures predictive for survival. Kaplan-Meier representation of overall survival and recurrence-free survival (EMT_RFS) probability over time for patients bearing colon cancers with average high (i.e. relative average expression > 0, green) or low expression (i.e. relative average expression < 0, red) of a selection of gene sets (details see Table 1).

Figure 3 Gain of malignant traits and loss of differentiation signature are associated with migratory properties. Migratory properties of intestinal cell lines across membrane in the absence (motility) or presence (chemotaxis) of a FCS gradient (details see "Material and Methods" section). Cell lines are ordered in dependence of rank orders of EMT and SLC_SI3 signatures as shown in Table 1.

border membrane endowed with hydrolases such as sucrase-isomaltase and aminopeptidase N [7,32-36]. On the other hand, comparison of permeability coefficients across different laboratories is still an issue [37]. Various studies have compared the genomic profiles of Caco-2 cells to large and small intestinal tissue [5,38-40]. None of these has systematically compared these profiles to the ones of small intestinal enterocytes in comparison to other intestinal cell lines, and, in particular, to profiles of the genes relevant for active transport.

Consequently, we made use of our objective approach to assess the suitability of the Caco-2 cells as model for oral drug absorption as compared to other CRC derived cell lines, taking inter-laboratory differences into account. As described above, PCA suggested that T84, HT29 and Caco-2 cells grown to confluency for 3 weeks were most similar to small intestinal enterocytes. Yet, when performing PCA filtering on solute carriers, transporters mostly responsible for active uptake, differentiated Caco-2 cells emerged as most akin to small intestinal enterocytes (Figure 4). When the analysis was extended to include expression of drug transporters known to be involved in uptake and secretion in the small intestine (Trans_BC [41]), differentiated Caco-2 cells clustered with normal colonocytes as well as small intestinal enterocytes (Figure 5). However, when focusing only on transporters potentially relevant for uptake (i.e. solute carriers expressed in plasma membrane according to Gene Ontology, SLC_PM), differentiated Caco-2 cells clustered only with small intestinal enterocytes (Additional file 2: Figure S1). When focusing on ABC transporters, differentiated Caco-2 cells clustered with HT29 and T84 cells (Additional file 3: Figure S2), suggesting that

these cell lines could also serve as a model for active export. Focusing on the expression of SLCs specific for the small intestine (i.e. SLC_SI3: SLCs expressed $\geq$ two fold in small intestine versus normal pool samples), differentiated Caco-2 cells, however, ranked highest independent of culture conditions and origins (Table 1).

Selection criteria for chemo-sensitivity studies

Recent reports suggest that chemotherapy may lead to a selection of tumor cells with EMT and stem cell properties and increased chemo-resistance [12]. For example, Gupta et al. [11] suggested that treatment with paclitaxel as compared to salinomycin significantly enriched breast
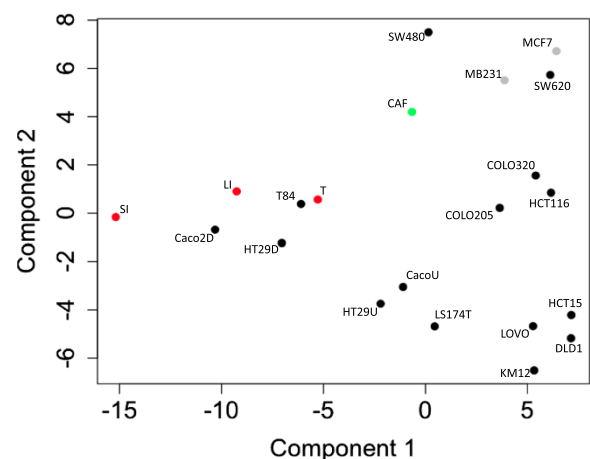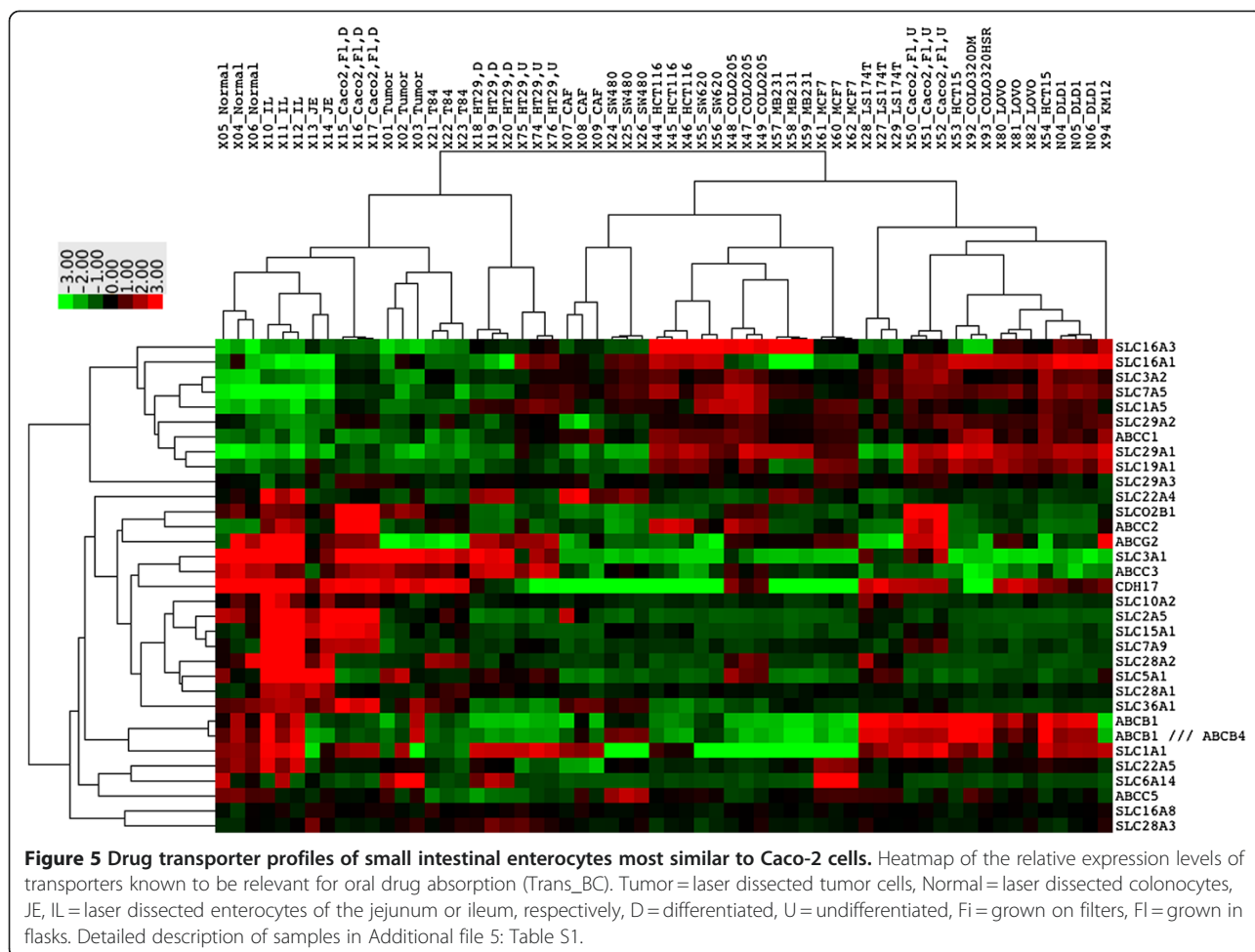

Figure 4 Caco-2 cells most similar to small intestinal enterocytes with respect to expression of solute carriers. Principal component analysis (PCA) as shown in Figure 1, but carried out on data set filtered using probe sets representing solute carriers.

Figure 5 Drug transporter profiles of small intestinal enterocytes most similar to Caco-2 cells. Heatmap of the relative expression levels of transporters known to be relevant for oral drug absorption (Trans_BC). Tumor = laser dissected tumor cells, Normal = laser dissected colonocytes, JE, IL = laser dissected enterocytes of the jejunum or ileum, respectively, D = differentiated, U = undifferentiated, Fi = grown on filters, Fl = grown in flasks. Detailed description of samples in Additional file 5: Table S1.

cancer cells with EMT properties. Thus, we examined their data set (GEO, GSE17215) and found that our EMT signature was also significantly enriched when performing GSEA (p = 0.000, FDR = 0.000, normalized enrichment score = 2.01). Using the same approach, however, we could not observe any significant enrichment (i.e. p < 0.05) of this EMT signature in either oxaliplatin-treated HT29, DLD1 and LOVO cells (GEO, GSE10405), or methotrexate-treated HT29 and Caco-2 cells (GEO, GSE16648) as compared to untreated cells. Of note, however, that the stem cell signature HuISC was enriched in methotrexate treated Caco-2 cells (p = 0.000, FDR = 0.000, normalized enrichment score = 2.08).
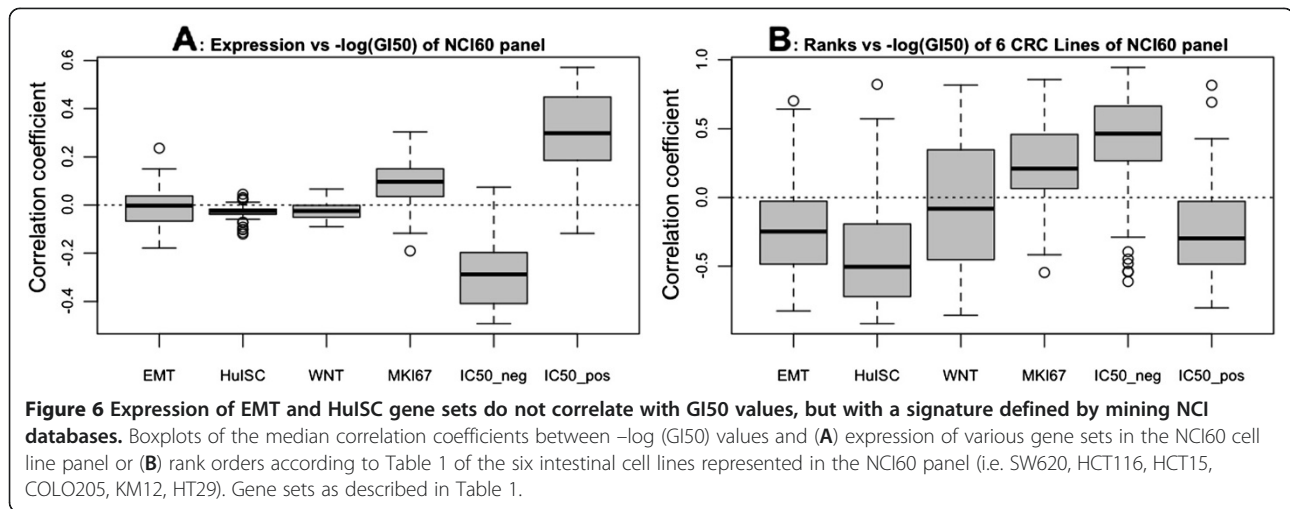
To extend this analysis we addressed more comprehensively to what extent EMT, HuISC and WNT signatures correlated with GI50 values of 50 commonly used chemotherapeutics in the NCI60 cell line panel (http://dtp.nci.nih.gov/docs/cancer/cancer_data.html).

Unexpectedly, no correlation between expression of EMT, HuISC and Wnt signatures and GI50 values was observed (Figure 6A), and furthermore, rank orders in terms of EMT, HuISC and Wnt signatures of 6 CRC cell

lines comprised within the NCI60 panel (Table 1) did not reveal any positive correlation with GI50 values, yet even a tendency to a negative link (Figure 6B).

Most chemotherapeutics interfere with the cell cycle activities and accordingly cell proliferation has been associated with susceptibility to chemotherapy [42]. As expected, a positive correlation between expression of proliferation associated genes and GI50 values was observed across the NCI60 cell lines (Figure 6A). The CRC cell lines tested herein (Table 1) displayed no significant difference in proliferation (data not shown) and no link was observed between their rank orders (Table 1) and GI50 values (Figure 6B).

Finally, we scrutinized the NCI60 data sets for genes whose expression was either positively or negatively correlated with GI50 values of chemotherapeutics (i.e. for at least three compounds correlation coefficient > |0.5|). Clustering of the correlation coefficients revealed two distinct gene sets (Additional file 4: Figure S3) for which the median positive and negative correlation coefficients were determined (Figure 6A). Interestingly, a link could be observed between rank orders of the CRC cell lines

Figure 6 Expression of EMT and HuISC gene sets do not correlate with GI50 values, but with a signature defined by mining NCI databases. Boxplots of the median correlation coefficients between –log (GI50) values and (**A**) expression of various gene sets in the NCI60 cell line panel or (**B**) rank orders according to Table 1 of the six intestinal cell lines represented in the NCI60 panel (i.e. SW620, HCT116, HCT15, COLO205, KM12, HT29). Gene sets as described in Table 1.

(Table 1) and GI50 values (Figure 6B), suggesting that these new signatures might be worth pursuing for assessing chemo-sensitivity.

Discussion

Selection criteria for cell-based models are often ill defined and arbitrary. Here we propose to make use of publicly available data generated by high throughput technologies to guide the choice of cell lines according to more objective criteria, such as gene expression signatures for selected traits. Focusing on CRC cell lines, we validated this approach for three different fields of application, i.e. the identification of cell lines most relevant for modeling and investigating tumor cell invasion, drug absorption and/or transport, and response to chemotherapy. Though we have focused on specific selection criteria, our strategy may be applied to any type of selection criteria (i.e. gene sets of different pathways and/or properties).

Models to study tumor biology

In view of the heterogeneity of solid tumors [9,43] and the challenge to identify adequate cell-based models, we assessed to what extent comparative data mining for identifying tumor traits that overlap in tumor samples and cell lines might contribute to improving the relevance of cell-based models. We therefore ranked a panel of CRC cell lines using EMT and HuISC signatures to assess to what extent such signatures reflect invasive tumor behavior. Reports on in vivo studies with these cell lines indeed validate our ranking criteria. Thus, SW480, SW620, COLO320 and HCT116 cells, which all ranked high in our assessment for EMT and HuISC signatures, showed the highest propensity for local invasion upon orthotopic grafting [6,8], whereas HT29, COLO205 and DLD1, which ranked lower for these

traits, showed less aggressive behavior. De Vries and co-workers observed that SW620 cells were indeed strongly invasive in vitro, in vivo, however, they proved less invasive than HT29 and LS174T cells [44]. Collectively, unbiased data mining allowed identifying cell lines with prominent mesenchymal traits (e.g. SW480), which display migratory properties and represent an invasive phenotype, whereas HT29, the most frequently used CRC model line, and T84 cells, which both ranked low for these traits came closest to gene expression profiles of bulk tumor tissue cells (Figure 1 and Table 1).

Caco-2 cells, a model of choice for drug absorption

Screening the literature with key words such as "absorption" and "transporters" revealed that, within the CRC cell line panel assessed herein, Caco-2 is the most widely used intestinal cell line for absorption-related studies [45-47]. Comparing transporter gene expression profiles revealed that differentiated Caco-2 cells indeed most accurately recapitulate those of normal enterocytes. Similar to Hayeshi et al. [37], we observed that culture conditions influence the expression pattern of selected transporters. Based on our ranking method we show that Caco-2 cells grown for three weeks, independently of their origin and culture conditions, best match the genomic profiles of small intestinal enterocytes.

Chemo-sensitivity

Selection of tumor cells with EMT and stem cell properties and higher chemo-resistance has recently emerged as a novel mechanism underlying therapy-resistance [10-12,48,49]. Vast databases with GI50 values, a standard to assess drug sensitivity [15,16], are available for the intestinal CRC lines HT29, COLO205, KM12, SW620, HCT116, HCC_2998 and HCT15 through the NCI database (http://dtp.nci.nih.gov/docs/cancer/cancer_

data.html). We therefore probed these data sets for a cor-relation between EMT and CSC signatures and chemore-sistance. Using the EMT signature defined herein, we could indeed confirm such a correlation in the breast cancer data set of Gupta et al. [11]. However, we could not observe any significant association of EMT-related genes with the GI50 values of CRC cell lines (Figure 6). Interestingly, novel chemoresistance-associated signa-tures could be identified; indicating that in CRC cell lines expression of EMT and CSC signatures are not dominant factors in the response to therapy. Cell line- or treatment-specific effects may account for this discrepancy. It is of note that Gupta and co-authors compared paclitaxel vs. salinomycin-treated cells, while the other data were based on treatment versus non-treatment comparisons.

A correlation between drug resistance and EMT fea-tures has, however, been observed using CRC biopsies [13,14]. As these may contain variable amounts of stroma and as stroma content may influence response to therapy [50], these data are not directly comparable to GI50 values of CRC cell lines.

Conclusions

We have presented a straightforward strategy to exploit publicly available gene expression data to guide the choice of cell-based models. While this approach does not overcome the major limitations of such models, introducing a rank order of selected features may allow selecting model cell lines that are more adapted and per-tinent to the addressed biological question. In summary, we conclude that SW480 represent an EMT-high, migra-tory phenotype and scored highest in terms of signatures associated to worse overall survival and higher risk of re-currence based on patient derived databases, therefore a good model to study invasive tumor cells. On the other hand, differentiated HT29 and T84 cells showed gene expression patterns closest to tumor bulk derived cells. Regarding drug absorption, we confirmed that differen-tiated Caco-2 cells are the model of choice for active up-take studies in the small intestine.

Methods
Cell culture

Cell culture media were as follows: SW480, RPMI 1640, 10% fetal calf serum (FCS); Caco-2, Dulbecco's modified Eagle essential medium (DMEM) Glutamax, 10% FCS, 1% non-essential amino acids; HT29, LS174T, DMEM Glutamax, 10% FCS; T84, RPMI 1640, 10% FCS, 2 mM glutamine. Primary fibroblasts, DMEM Glutamax, 10% FCS, 100 U/ml penicillin, 100 μg/ml streptomycin, pt? >0.5 mL gentamycin (50 mg/mL). All cell lines were cul-tured at 37°C and 5% CO_2 (Samples 15–29 see Additional file 5: Table S1). CDX2-shCaco-2 cells were maintained under puromycin selection (2 ug/mL) (Samples 86–88 see

Additional file 5: Table S1). RNA was extracted using Nucleo-Spin RNA-extraction kit from Machery-Nagel (Oensingen, Switzerland). CacoReady^{TM} kit cells (Ready-cell, Barcelona, Spain) were cultured on 24-transwells filters for either 14 days in DMEM low glucose contain-ing 10% FCS, 1% L-glutamine and 100 U/ml penicillin, 100 μg/ml streptomycin, then 4 days of culture in semi-solid shipping media and again 3 days in culture media (Readycell's patented technology) (Samples 54–56 see Additional file 5: Table S1), or for 3 weeks in the absence of the semi-solid shipping media (Samples 57–59 see Additional file 5: Table S1). Differentiated Caco-2LD cells were grown at low density and then differentiated on 0.4 μm PET filters as described earlier [51] (Samples 49–51 see Additional file 5: Table S1).

Laser dissection microscopy

Immediately after resection of human tissue, samples were put on ice and sections of one cm length of small intestine and colon (tumor regions and adjacent normal epithelium), respectively, were cut, washed in chilled PBS, embedded in OCT and frozen immediately. One section of the ileum und two of the jejunum were first stored in RNAlater at 4°C for shipping, then rinsed with chilled PBS and embedded in OCT and frozen immedi-ately. 12 μm frozen sections were cut and mounted on Leica or Zeiss membranes for dissecting microscopy (Leica Microsystems, Germany and PALM Microbeam, Germany), fixed in 96% ethanol for 30 s and colored for an equivalent time with hematoxylin/eosin solution, re-spectively. Membranes were then rinsed in water for 30 s, transferred for 10 s to 70% ethanol, followed by 96% ethanol and air-dried. Frozen samples were pro-cessed using a laser dissecting microscope coupled to a CCD camera (Leica Microsystems, Germany and PALM Microbeam, Germany). For each patient sample dissected cells were pooled in a tube cap containing 20 μL RNA lysis buffer. Collection of human intestinal samples was approved by the corresponding local ethic commission (Samples 1–6, 10–14 see Additional file 5: Table S1).

Gene expression using GeneChip® human genome U133 plus 2.0

Total RNA was extracted using the total RNA extraction Nucleospin II kit by Machery-Nagel (Oensingen, Switz-erland). The quality and quantity of all RNA samples was examined by the Agilent 2100 Bioanalyzer (Agilent Biotechnologies, Germany) and by a NanoDrop (Witec AG, Switzerland), respectively. RINs for cell lines were between 9 and 10, for LDM material between 5 and 7 except for two small intestinal samples. 100 ng of total RNA were used as the starting material for all individual samples. Labeling and fragmentation of cRNA, array hybridization and scanning was performed according to

the protocol by Affymetrix. Details see Additional file 6. Cell lines were measured in triplicates with different passage numbers. Normal tumor intestinal human tissue was obtained from different individuals. The complete data set is publicly available at http://www.ncbi.nlm.nih.gov/geo/ through the accession number GSE30292.

Migration and proliferation assay

Chemotaxis was assayed in 48-well Boyden microchambers (Neuro Probe, Cabin John, MD, USA). Cells were cultured in serum-free media overnight. For the migration studies, chemotaxis buffer containing 30% serum was placed in the lower wells, and 10^4 cells, suspended in chemotaxis buffer (culture media with 1% BSA), in the upper wells. A PVPF membrane (Poretics 25 × 80 mm) with 8 μm pores was coated with rat collagen type I. After 15 h of incubation the membrane was removed, washed on the upper side with PBS, fixed and stained with Diff-Quick staining kit. Migrated cells were counted at 1000-fold magnification in five randomly selected fields. Values are given as average cell count of five high-powered fields (5 HPF). Proliferation was assessed using an MTT assay (Sigma-Aldrich, Switzerland) over five days. Briefly, 3000 cells were seeded in five replicates per day, per well in a 96-well plate. At each time point 10 μL MTT working solution was added to each well. Then the plate was incubated for 3 h at 37°C and the cells lysed with lysis buffer overnight. Optical density was measured at 540 nm. ANOVA was done for migration data (p value ≤ 0.05).

Public data

Publicly available data was obtained at http://www.ncbi.nlm.nih.gov/geo/ through the accession numbers GSE2361, GSE2361, GSE7303 (Pool normal; colon GSM175905, small intestine GSM175908, lung GSM176012, breast GSM175792, prostate GSM175923, uterus GSM175945, kidney GSM175911), GSE2109 (Pool tumor; colon GSM38055, small intestine GSM38068, lung GSM38051, breast GSM38051, prostate GSM38053, uterus GSM38052, kidney GSM38073), GSE10843, GSE13059, GSE16648, GSE22572. MB231 and MCF7 cell lines (GSE10890) were included in the analysis as controls for cells with strong EMT properties and with weak, respectively. Expression data for NCI60 cell line panel was obtained through http://discover.nci.nih.gov/cellminer/ [52] and GI50 data at http://dtp.nci.nih.gov/docs/cancer/cancer_data.html.

Expression data analysis

Robust multi-array averaging (RMA) and quantile normalization were used to quantify gene expression. Significant differences were identified applying a Bayesian approach using the limma package (R 2.12.0, Bioconductor 2.7). A threshold of an adjusted p value ≤ 0.05 was used to identify significant changes if not indicated otherwise. Gene set enrichment analysis (GSEA) was carried out according to Subramanian et al. [53] and p values were computed using a bootstrap distribution created by resampling gene sets of the same cardinality. See details in supplementary data section "expression data analysis" and Additional file 1: Table S2 regarding data sets. Principle component analysis was performed using the affy package (R, Bioconductor). Correlation coefficients between −log(GI50) of selected drugs (see Additional file 6) and normalized expression values were calculated using R. Heatmaps were generated using Cluster 3.0 and TreeView (http://rana.lbl.gov/EisenSoftware.htm) [54]. Centered correlation or Spearman rank correlation and average linkage clustering, respectively, were used for similarity measurement and clustering. Effects of signatures on overall and recurrence free survival were studied using the Kaplan-Meier survival curves and log rank tests in three different publicly available data sets (GSE12945, GSE14333, GSE17537) using the survival package in the statistical software R version 2.13.0. Patient samples were divided into two groups splitting at the average score.

Ranking according to signature strength

Gene expression data was filtered by selecting only one probe set, the one with the highest standard deviation across all samples, per gene. For each gene of a given gene set the relative expression (i.e. gene-wise zero-centering of expression values) across all samples was determined. The score of a gene set was calculated as the average of the expression values of the genes in the set. The cell lines were then ranked according to this score. See details in supplementary data section "expression data analysis" and Additional file 1: Table S2 regarding data sets.

Drug sensitivity analysis

Correlation coefficients between −log(GI50) and log2 gene expression values were calculated for a panel of 50 chemotherapeutics (CTX, list of drugs see Additional file 7: Figure S4) and a panel of drugs (NSC) for which for all 59 cell lines GI50 values were available and SD across 59 lines of −log(GI50) was > 0.

Additional files

> **Additional file 1: Table S2.** Gene sets used for data analysis.
>
> **Additional file 2: Figure S1.** Solute carriers of plasma membrane profiles of small intestinal enterocytes most similar to Caco-2 cells. Heatmap of relative expression levels of solute carriers of the plasma membrane.

Additional file 3: Figure S2. Differentiated Caco-2, HT29 and T84 cell lines most similar to enterocytes with respect to expression ABC transporters. Heatmap of the relative expression levels of ABC transporters.

Additional file 4: Figure S3. Two distinct groups of genes with respect to expected chemosensitivity. Heatmap of correlation coefficients –log(GI50) of 50 chemotherapeutics and expression values across the NCI60 cell line panel.

Additional file 5: Table 1S. List of expression files used for microarray analysis and corresponding accession numbers at Gene Expression Omnibus.

Additional file 6: Supplementary information for methods and additional results [29,30,41,55-58].

Additional file 7: Figure S4. Enrichment of small intestinal signature in Caco-2 cells upon differentiation. Heatmap representing enrichment scores obtained by gene set enrichment analysis of the selected panel of gene sets. Genes in the various cell lines were ranked according to their relative average expression (n = 2-3) in the given cell line compared to a panel of healthy epithelial tissues. Details see Additional file 6, material and methods.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute for Macromolecular Chemistry and Center for Biological Signaling Studies (BIOSS), University of Freiburg, Freiburg, Germany. [2]Institute of Biochemistry and Molecular Medicine, University of Berne, Berne, Switzerland. [3]Istituto di Biologia Cellulare e Neurobiologia, CNR, Rome, Italy. [4]OSC - Omics Science Center, RIKEN Yokohama Institute, Yokohama, Japan. [5]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [6]Swiss National Centre of Competence in Research Molecular Oncology, Swiss Institute for Experimental Cancer Research, Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences, Lausanne, Switzerland. [7]NCCR TransCure, University of Berne, Berne, Switzerland. [8]Institute of pathology of Lausanne, Centre hospitalier et universitaire vaudois, Lausanne, Switzerland. [9]Laboratorio de Investigaciones del Sistema Inmune (LISIN), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina. [10]Faculty of Pharmacy, University of Helsinki, Helsinki, Finland. [11]LEITAT Technological Center, Barcelona, Spain. [12]Medical Oncology Department, IOSI, Ospedale Regionale di Lugano, Lugano, Switzerland. [13]Swiss Institute for Experimental Cancer Research, Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences, Lausanne, Switzerland. [14]Institute of Biochemistry and Molecular Medicine, University of Bern, Buehlstrasse 28, 3000, Bern 9, Switzerland.

Authors' contributions
Data acquisition: JC, SE, MN, PA, SB, MV, AF, SS, JH; data analysis: PA, JC, SE, TS, MD, MA; patient samples: MR, HB, PS, JC, PA; study design, funding, drafting of ms: PA, MA; supervision: PA. All authors read and approved the final manuscript.

References
1. Markowitz SD, Bertagnolli MM: Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med* 2009, **361**(25):2449–2460.
2. Gayet J, Zhou XP, Duval A, Rolland S, Hoang JM, Cottu P, Hamelin R: Extensive characterization of genetic alterations in a series of human colorectal cancer cell lines. *Oncogene* 2001, **20**(36):5025–5032.
3. Oliveira C, Pinto M, Duval A, Brennetot C, Domingo E, Espin E, Armengol M, Yamamoto H, Hamelin R, Seruca R, *et al*: BRAF mutations characterize colon but not gastric cancer with mismatch repair deficiency. *Oncogene* 2003, **22**(57):9192–9196.
4. Rosin-Arbesfeld R, Cliffe A, Brabletz T, Bienz M: Nuclear export of the APC tumour suppressor controls beta-catenin function in transcription. *EMBO J* 2003, **22**(5):1101–1113.
5. van Erk MJ, Krul CA, Caldenhoven E, Stierum RH, Peters WH, Woutersen RA, van Ommen B: Expression profiling of colon cancer cell lines and colon biopsies: towards a screening system for potential cancer-preventive compounds. *Eur J Cancer Prev* 2005, **14**(5):439–457.
6. Cespedes MV, Espina C, Garcia-Cabezas MA, Trias M, Boluda A, Gomez del Pulgar MT, Sancho FJ, Nistal M, Lacal JC, Mangues R: Orthotopic microinjection of human colon cancer cells in nude mice induces tumor foci in all clinically relevant metastatic sites. *Am J Pathol* 2007, **170**(3):1077–1085.
7. Chantret I, Barbat A, Dussaulx E, Brattain MG, Zweibaum A: Epithelial polarity, villin expression, and enterocytic differentiation of cultured human colon carcinoma cells: a survey of twenty cell lines. *Cancer Res* 1988, **48**(7):1936–1942.
8. Flatmark K, Maelandsmo GM, Martinsen M, Rasmussen H, Fodstad O: Twelve colorectal cancer cell lines exhibit highly variable growth and metastatic capacities in an orthotopic model in nude mice. *Eur J Cancer* 2004, **40**(10):1593–1598.
9. Thiery JP, Sleeman JP: Complex networks orchestrate epithelial-mesenchymal transitions. *Nat Rev Mol Cell Biol* 2006, **7**(2):131–142.
10. 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.
11. Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA, Lander ES: Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* 2009, **138**(4):645–659.
12. Singh A, Settleman J: EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* 2010, **29**(34):4741–4751.
13. Loboda A, Nebozhyn MV, Watters JW, Buser CA, Shaw PM, Huang PS, Van't Veer L, Tollenaar RA, Jackson DB, Agrawal D, *et al*: EMT is the dominant program in human colon cancer. *BMC Med Genet* 2011, **4**:9.
14. Oh SC, Park YY, Park ES, Lim JY, Kim SM, Kim SB, Kim J, Kim SC, Chu IS, Smith JJ, *et al*: Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. *Gut* 2011.
15. Sharma SV, Haber DA, Settleman J: Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. *Nat Rev Cancer* 2010, **10**(4):241–253.
16. Shoemaker RH: The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006, **6**(10):813–823.
17. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, *et al*: A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000, **24**(3):236–244.
18. Arango D, Mariadason JM, Wilson AJ, Yang W, Corner GA, Nicholas C, Aranes MJ, Augenlicht LH: c-Myc overexpression sensitises colon cancer cells to camptothecin-induced apoptosis. *Br J Cancer* 2003, **89**(9):1757–1765.
19. Augenlicht LH, Velcich A, Klampfer L, Huang J, Corner G, Aranes M, Laboisse C, Rigas B, Lipkin M, Yang K, *et al*: Application of gene expression profiling to colon cell maturation, transformation and chemoprevention. *J Nutr* 2003, **133**(7 Suppl):2410S–2416S.
20. Boyer J, Allen WL, McLean EG, Wilson PM, McCulla A, Moore S, Longley DB, Caldas C, Johnston PG: Pharmacogenomic identification of novel determinants of response to chemotherapy in colon cancer. *Cancer Res* 2006, **66**(5):2765–2777.
21. Hoshino H, Miyoshi N, Nagai K, Tomimaru Y, Nagano H, Sekimoto M, Doki Y, Mori M, Ishii H: Epithelial-mesenchymal transition with expression of

SNAI1-induced chemoresistance in colorectal cancer. *Biochem Biophys Res Commun* 2009, **390**(3):1061–1065.

22. Jung JJ, Jeung HC, Lee JO, Kim TS, Chung HC, Rha SY: **Putative chemosensitivity predictive genes in colorectal cancer cell lines for anticancer agents.** *Oncol Rep* 2007, **18**(3):593–599.

23. Mariadason JM, Arango D, Augenlicht LH: **Customizing chemotherapy for colon cancer: the potential of gene expression profiling.** *Drug Resist Updat* 2004, **7**(3):209–218.

24. Sabbah M, Emami S, Redeuilh G, Julien S, Prevost G, Zimber A, Ouelaa R, Bracke M, De Wever O, Gespach C: **Molecular signature and therapeutic perspective of the epithelial-to-mesenchymal transitions in epithelial cancers.** *Drug Resist Updat* 2008, **11**(4–5):123–151.

25. Spitzner M, Emons G, Kramer J, Gaedcke J, Rave-Frank M, Scharf JG, Burfeind P, Becker H, Beissbarth T, Ghadimi BM, *et al*: **A gene expression signature for chemoradiosensitivity of colorectal cancer cells.** *Int J Radiat Oncol Biol Phys* 2010, **78**(4):1184–1192.

26. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J, Scherf U, Lee JK, Reinhold WO, Weinstein JN, *et al*: **Chemosensitivity prediction by transcriptional profiling.** *Proc Natl Acad Sci U S A* 2001, **98**(19):10787–10792.

27. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe JP, Tong F, *et al*: **A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.** *Cancer Cell* 2006, **10**(6):515–527.

28. Vermeulen L, De Sousa EMF, van der Heijden M, Cameron K, de Jong JH, Borovski T, Tuynman JB, Todaro M, Merz C, Rodermond H, *et al*: **Wnt activity defines colon cancer stem cells and is regulated by the microenvironment.** *Nat Cell Biol* 2010, **12**(5):468–476.

29. Deka J, Wiedemann N, Anderle P, Murphy-Seiler F, Bultinck J, Eyckerman S, Stehle JC, Andre S, Vilain N, Zilian O, *et al*: **Bcl9/Bcl9l are critical for Wnt-mediated regulation of stem cell traits in colon epithelium and adenocarcinomas.** *Cancer Res* 2010, **70**(16):6619–6628.

30. Merlos-Suarez A, Barriga FM, Jung P, Iglesias M, Cespedes MV, Rossell D, Sevillano M, Hernando-Momblona X, da Silva-Diz V, Munoz P: **The Intestinal Stem Cell Signature Identifies Colorectal Cancer Stem Cells and Predicts Disease Relapse.** *Cell Stem Cell* 2011, **8**(5):511–524.

31. Guo RJ, Suh ER, Lynch JP: **The role of Cdx proteins in intestinal development and cancer.** *Cancer Biol Ther* 2004, **3**(7):593–601.

32. Artursson P, Palm K, Luthman K: **Caco-2 monolayers in experimental and theoretical predictions of drug transport.** *Adv Drug Deliv Rev* 2001, **46**(1–3):27–43.

33. Fleet JC, Wang L, Vitek O, Craig BA, Edenberg HJ: **Gene expression profiling of Caco-2 BBe cells suggests a role for specific signaling pathways during intestinal differentiation.** *Physiol Genomics* 2003, **13**(1):57–68.

34. Mariadason JM, Bordonaro M, Aslam F, Shi L, Kuraguchi M, Velcich A, Augenlicht LH: **Down-regulation of beta-catenin TCF signaling is linked to colonic epithelial cell differentiation.** *Cancer Res* 2001, **61**(8):3465–3471.

35. Pshezhetsky AV, Fedjaev M, Ashmarina L, Mazur A, Budman L, Sinnett D, Labuda D, Beaulieu JF, Menard D, Nifant'ev I, *et al*: **Subcellular proteomics of cell differentiation: quantitative analysis of the plasma membrane proteome of Caco-2 cells.** *Proteomics* 2007, **7**(13):2201–2215.

36. Saaf AM, Halbleib JM, Chen X, Yuen ST, Leung SY, Nelson WJ, Brown PO: **Parallels between global transcriptional programs of polarizing Caco-2 intestinal epithelial cells in vitro and gene expression programs in normal colon and colon cancer.** *Mol Biol Cell* 2007, **18**(11):4245–4260.

37. Hayeshi R, Hilgendorf C, Artursson P, Augustijns P, Brodin B, Dehertogh P, Fisher K, Fossati L, Hovenkamp E, Korjamo T, *et al*: **Comparison of drug transporter gene expression and functionality in Caco-2 cells from 10 different laboratories.** *Eur J Pharm Sci* 2008, **35**(5):383–396.

38. Anderle P, Rakhmanova V, Woodford K, Zerangue N, Sadee W: **Messenger RNA expression of transporter and ion channel genes in undifferentiated and differentiated Caco-2 cells compared to human intestines.** *Pharm Res* 2003, **20**(1):3–15.

39. Calcagno AM, Ludwig JA, Fostel JM, Gottesman MM, Ambudkar SV: **Comparison of drug transporter levels in normal colon, colon cancer, and Caco-2 cells: impact on drug disposition and discovery.** *Mol Pharm* 2006, **3**(1):87–93.

40. Tremblay E, Auclair J, Delvin E, Levy E, Menard D, Pshezhetsky AV, Rivard N, Seidman EG, Sinnett D, Vachon PH, *et al*: **Gene expression profiles of normal proliferating and differentiating human intestinal epithelial cells: a comparison with the Caco-2 cell model.** *J Cell Biochem* 2006, **99**(4):1175–1186.

41. Anderle P, Nielsen CU: **Transporters in the gastrointestinal tract**. In *Drug Bioavailability: The Future of Integrative Metabolism.* Edited by van de Waterbeemd H, Testa B.: Wiley VCH; 2008.

42. Mellor HR, Callaghan R: **Resistance to chemotherapy in cancer: a complex and integrated cellular response.** *Pharmacology* 2008, **81**(4):275–300.

43. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**(5):646–674.

44. de Vries JE, Dinjens WN, De Bruyne GK, Verspaget HW, van der Linden EP, de Bruine AP, Mareel MM, Bosman FT, ten Kate J: **In vivo and in vitro invasion in relation to phenotypic characteristics of human colorectal carcinoma cells.** *Br J Cancer* 1995, **71**(2):271–277.

45. Press B, Di Grandi D: **Permeability for intestinal absorption: Caco-2 assay and related issues.** *Curr Drug Metab* 2008, **9**(9):893–900.

46. Sambuy Y, De Angelis I, Ranaldi G, Scarino ML, Stammati A, Zucco F: **The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics.** *Cell Biol Toxicol* 2005, **21**(1):1–26.

47. Sun H, Chow EC, Liu S, Du Y, Pang KS: **The Caco-2 cell monolayer: usefulness and limitations.** *Expert Opin Drug Metab Toxicol* 2008, **4**(4):395–411.

48. Alison MR, Lim SM, Nicholson LJ: **Cancer stem cells: problems for therapy?** *J Pathol* 2011, **223**(2):147–161.

49. Zhou BB, Zhang H, Damelin M, Geles KG, Grindley JC, Dirks PB: **Tumour-initiating cells: challenges and opportunities for anticancer drug discovery.** *Nat Rev Drug Discov* 2009, **8**(10):806–823.

50. Farmer P, Bonnefoi H, Anderle P, Cameron D, Wirapati P, Becette V, Andre S, Piccart M, Campone M, Brain E, *et al*: **A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer.** *Nat Med* 2009, **15**(1):68–74.

51. Natoli M, Leoni BD, D'Agnano I, D'Onofrio M, Brandi R, Arisi I, Zucco F, Felsani A: **Cell growing density affects the structural and functional properties of Caco-2 differentiated monolayer.** *J Cell Physiol* 2011, **226**(6):1531–1543.

52. Pfister TD, Reinhold WC, Agama K, Gupta S, Khin SA, Kinders RJ, Parchment RE, Tomaszewski JE, Doroshow JH, Pommier Y: **Topoisomerase I levels in the NCI-60 cancer cell line panel determined by validated ELISA and microarray analysis and correlation with indenoisoquinoline sensitivity.** *Mol Cancer Ther* 2009, **8**(7):1878–1884.

53. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, *et al*: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15545–15550.

54. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci U S A* 1998, **95**(25):14863–14868.

55. Verzi MP, Shin H, He HH, Sulahian R, Meyer CA, Montgomery RK, Fleet JC, Brown M, Liu XS, Shivdasani RA: **Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2.** *Dev Cell* 2010, **19**(5):713–726.

56. Gassler N, Newrzella D, Bohm C, Lyer S, Li L, Sorgenfrei O, van Laer L, Sido B, Mollenhauer J, Poustka A, *et al*: **Molecular characterisation of non-absorptive and absorptive enterocytes in human small intestine.** *Gut* 2006, **55**(8):1084–1089.

57. van der Flier LG, van Gijn ME, Hatzis P, Kujala P, Haegebarth A, Stange DE, Begthel H, van den Born M, Guryev V, Oving I, *et al*: **Transcription factor achaete scute-like 2 controls intestinal stem cell fate.** *Cell* 2009, **136**(5):903–912.

58. Fritzmann J, Morkel M, Besser D, Budczies J, Kosel F, Brembeck FH, Stein U, Fichtner I, Schlag PM, Birchmeier W: **A colorectal cancer expression profile that includes transforming growth factor beta inhibitor BAMBI predicts metastatic potential.** *Gastroenterology* 2009, **137**(1):165–175.