Polyol pathway links glucose metabolism to the aggressiveness of cancer cells.

Annemarie Schwab[1], Aarif Siddiqui[1], Maria Eleni Vazakidou[1], Francesca Napoli[1], Martin Böttcher[2], Bianca Menchicchi[3], Umar Raza[4], Özge Saatci[4], Angela M. Krebs[5], Fulvia Ferrazzi[6], Ida Rapa[7], Katja Dettmer-Wilde[8], Maximilian J. Waldner[3], Arif B. Ekici[6], Suhail Ahmed Kabeer Rasheed[9], Dimitrios Mougiakakos[2], Peter J. Oefner[8], Ozgur Sahin[4], Marco Volante[7], Florian R. Greten[10], Thomas Brabletz[5] and Paolo Ceppi[1].*

Authors affiliations:

1. Junior Research Group 1, Interdisciplinary Center for Clinical Research, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Erlangen, Germany.
2. Department of Internal Medicine 5, Hematology and Oncology, University Hospital Erlangen, Erlangen, Germany;
3. Department of Medicine 1, University Hospital Erlangen, Erlangen, Germany;
4. Bilkent University, Department of Molecular Biology and Genetics, Ankara, Turkey.
5. Experimental Medicine I, FAU Erlangen-Nürnberg, Erlangen, Germany;
6. Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
7. Pathology Unit, San Luigi Hospital, University of Turin, Turin, Italy;
8. Institute of Functional Genomics University of Regensburg, Regensburg, Germany
9. Duke-NUS Medical School, Singapore;
10. Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, Frankfurt am Main, Germany

[*] **Address correspondence to:** Dr. Paolo Ceppi, Junior Research Group 1, Interdisciplinary Center for Clinical Research, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Nikolaus-Fiebiger-Zentrum, Glückstrasse 6, 91054 Erlangen, Germany. tel: +49 (0)91318539300, fax: +49 (0)91318536386

Conflict of interest: The authors declare no conflicts of interests.

Running title: Polyol pathway in aggressive cancers.

Word count: 5064 words.

Figures: 7 Main + 6 Supplementary Figures, and 1 Supplementary Table.

Keywords: Polyol pathway, Cancer, AKR1B1, SORD, Epithelial-to-mesenchymal transition; TGF-β;

Précis: *A glucose-transforming pathway in TGF-Beta-driven epithelial-to-mesenchymal transition provides novel mechanistic insights into the metabolic control of cancer differentiation.*

ABSTRACT

Cancer cells alter their metabolism to support their malignant properties. In this study, we report that the glucose-transforming polyol pathway (PP) gene aldo-keto-reductase-1-member-B1 (AKR1B1) strongly correlates with epithelial-to-mesenchymal transition (EMT). This association was confirmed in samples from lung cancer patients and from an EMT-driven colon cancer mouse model with p53 deletion. In vitro, mesenchymal-like cancer cells showed increased AKR1B1 levels, and AKR1B1 knockdown was sufficient to revert EMT. An equivalent level of EMT suppression was measured by targeting the downstream enzyme sorbitol-dehydrogenase (SORD), further pointing at the involvement of the PP. Comparative RNA sequencing confirmed a profound alteration of EMT in PP-deficient cells, revealing a strong repression of TGF-β signature genes. Excess glucose was found to promote EMT through autocrine TGF-β stimulation, while PP-deficient cells were refractory to glucose-induced EMT. These data show that PP represents a molecular link between glucose metabolism, cancer differentiation, and aggressiveness, and may serve as a novel therapeutic target.

INTRODUCTION

Differentiation and phenotypical heterogeneity have a fundamental impact on the aggressiveness of tumors (1,2). Epithelial-to-mesenchymal transition (EMT) is a de-differentiation process that cancer cells use to acquire an invasive, chemo-resistant phenotype as well as the properties of cancer stem cells (CSCs) (3,4). EMT/CSC are therefore therapeutically very attractive, but a further understanding of the underlying

biological processes is required in order to develop novel effective therapeutics (5). Some cancer types, and the corresponding CSC population in particular (6), are highly dependent on glucose metabolism and aerobic glycolysis, which they use as a major pathway for biosynthesis (7), and the molecular connection between glucose metabolism, EMT and CSCs has recently started to emerge (6,8,9). Importantly, epidemiological and experimental studies have connected sucrose/fructose consumption with increased cancer risk (10-12), but the molecular determinants are still undefined and the impact of other glucose-related pathways has been poorly investigated. By means of an unbiased transcriptomic analysis aimed at identifying genes highly correlated with aggressive EMT-driven tumors, we found the aldo-keto reductase family 1, member B1 (*AKR1B1*). This gene encodes for a specific member of the aldo-keto reductase superfamily catalyzing the reduction of glucose to sorbitol (13), the first step of the polyol pathway (PP). Using different models, we provide evidence that the PP can functionally regulate EMT, representing a novel link between glucose metabolism and tumor aggressiveness.

MATERIALS AND METHODS

Cell lines and treatments All cell lines were cultured in media supplemented with 5% FBS, 1% Pen/Strep and 1% L-Glutamine (Sigma) at 37 °C and 5% CO_2 in a humidified incubator. Cells were from the National Cancer Institute (those part of the NCI-60 panel, except MDA-MB-231 and HCT116) or from ATCC, were STR-profiling authenticated and used between passages 3 and 15, were examined for the presence of mycoplasma and maintained in Plasmocin to prevent contamination (detection kit and reagents from Invivogen). Breast (T47D, MCF-7, MDA-MB-231, BT549, Hs578T), ovarian (Ovcar -4,-5,-8, HeyA8), lung (A549, Calu-1, H1299), colorectal (KM12 and Colo205) as well as renal CAKI-1 and glioblastoma U251 cancer cell lines were cultured in RPMI (Sigma). Other colorectal cancer cell lines (HCT116, HT29) were cultured in McCoy's 5A (Gibco). 293T and KPC(Z) cell lines were cultured in DMEM (Sigma).

In vitro treatments For EMT induction, cells were treated with 10 ng/ml TGF-β (R&D Systems) for 2-3 days or with 50 or 100 mM glucose (Sigma) for 3-6 days. Inhibition of TGF-β signaling was performed with a neutralizing TGF-β 1,2,3 antibody (R&D systems, 1D11, 1 µg/ml) in the presence of either 11 mM (concentration of standard

media formulation) or 100 mM glucose for 6 days. The corresponding IgG1 isotype antibody (R&D systems, 11711, 1 µg/ml) served as a control. To test the sensitivity towards cisplatin, cells were treated with 350 or 500 µM cisplatin (Sigma) the day after plating, using DMSO as control.

Stable overexpression and knockdown of plasmid DNA and shRNA Overexpression experiments were performed as previously described (14) using the EX-C0237-Lv105 construct (Genecopoeia) for AKR1B1. Stable knockdown of human AKR1B1 and SORD expression was achieved using Sigma MISSION lentiviral transduction particles (TRCN0000288741 and TRCN0000288812 for AKR1B1-; TRCN0000028052, TRCN0000028069 and TRCN0000028106 for SORD-knockdown) and non-targeting controls. Cells were selected in media containing 3 µg/ml Puromycin (Sigma).

Transient siRNA transfection Transient transfections were performed using siPORT NeoFX transfection agent (Applied Biosystems) according to the manufacturer's instruction. 80.000 cells were transfected in 12-well plates with 50 nM siRNA (Ambion).

Real-time live-cell assays Cell proliferation and migration were monitored using live time-lapse images recorded with the IncuCyte ZOOM live cell analysis system (Essen BioScience). As readout for proliferation the occupied area (% of confluence) is plotted over time. For migration, homogeneous scratch wounds were created using the 96-pin WoundMaker (Essen BioScience) and relative wound density is plotted over time.

Western Blotting For protein isolation cells were lysed in RIPA buffer with Halt Protease&Phosphatase Inhibitor Cocktail (both Thermo-Fisher), followed by centrifugation at 12.000 rpm. Proteins were quantified with the Pierce BCA Protein Assay Kit (Thermo-Fisher), separated by SDS-PAGE (10%) and transferred onto a PVDF membrane (Thermo-Fisher). Proteins were detected using AKR1B1 (Thermo-Fisher, polyclonal), SORD (Thermo-Fisher, polyclonal), E-Cadherin (Cell Signaling, 4A2), Vimentin (Cell Signaling, D21H3), ZEB1 (Sigma, polyclonal), Fra-1 (Cell Signaling, D80B4), Smad3 (Cell Signaling, C67H9), pSmad3 (Cell Signaling, C25A9),

β-Actin HRP conjugated (Cell Signaling, 8H10D10) and species-specific HRP-conjugated secondary antibodies (Southern Biotech). Bands were detected using the Pierce ECL Western Blotting Substrate (Thermo-Fisher), X-ray CL-XPosure films (Thermo-Fisher) and the automatic film processor CP1000 (AGFA).

Immunofluorescence Immunofluorescence staining of cells grown on cover slips was performed as previously described (14). Signals were visualized using Leica DM5500B fluorescence microscope and Leica Application Suite-X software.

Colony formation assay To assess clonogenic ability, a colony formation assay was performed using cells plated at a low density (1000 cells/well), as previously described (14).

Extracellular Flux Assays Bioenergetics of control and knockdown cell lines were determined using the XFe96 Extracellular Flux Analyzer (Seahorse Bioscience/Agilent Technologies, North Billerica, MA). Cells were seeded in specialized cell culture microplates at a density of 2000 cells/well and cultured for 72 h. One hour before the measurement cells were incubated at 37°C in a CO_2-free atmosphere. For the determination of glycolytic parameters basal extracellular acidification rate (ECAR; indicative of glycolysis) was first determined under glucose-free conditions. Secondly, the rate of glycolysis was calculated using the ECAR after glucose supplementation (10 mM). For the determination of respiratory parameters basal oxygen consumption rate (OCR, indicator for mitochondrial respiration) was measured. All experiments were performed in pentaplicates. Raw values were normalized to the total protein content for each well.

Flow cytometry The proportion of cancer stem cells was determined using the Aldefluor Kit (StemCell Technologies) according to the manufacturer's instruction. Propidium Iodide (PI, Sigma) staining was performed as previously described (14). Cells were analyzed with FACSCalibur (BD Biosciences) and FlowJo (Treestar Inc, version 8.8.6) or CytoFLEX and CytExpert 2.0 (Beckman Coulter).

Sphere-forming assays Tumorsphere assays were performed by seeding 20.000 cells/well in ultra-low attachment plates (Sigma) in complete MammoCult Medium

(StemCell Technologies) supplemented according to the manufacturer's instructions for 8 days. Analysis was performed by counting the number of spheres (1/20 dilution in quadruplicates) or lysing proteins for western blot analysis.

RNA isolation and quantitative real time PCR RNA was isolated using miRNeasy Mini Kit (Qiagen) according to the manufacturer's instruction. 500 ng of RNA were used to perform reverse transcription with the Tetro cDNA synthesis Kit (Bioline) according to the manufacturer's instructions. DNA was amplified with gene-specific primers using TaqMan Universal Master Mix II (Thermo-Fisher). Expression values were measured using ABI Prism 7300. Data was analyzed using the ΔΔCt method and normalized to GAPDH. Results are shown as relative fold expression compared to Ctrl cells.

RNA Sequencing Before and during the library preparation the quality of RNA was analyzed using a 2100 Bioanalyzer system (Agilent Technologies). Barcoded RNA sequencing libraries were prepared from 100 ng total RNA using Illumina's TruSeq stranded mRNA kit according to the manufacturer's instructions. Libraries were subjected to single-end sequencing (101 bp) on a HighSeq-2500 platform (Illumina, San Diego, CA). Quality filtering was performed using cutadapt v. 1.9.1; then reads were mapped against the human reference genome using the STAR aligner v. 2.5.2b, and a STAR genome directory created by supplying the Ensembl gtf annotation file (release 87) for GRCh37. Read counts per gene were obtained using featureCounts program v. 1.5.1 and the Ensembl gtf annotation file. Following analyses were performed using R version 3.3.1. In particular, differential expression analysis was performed with the DESeq2 package v.1.12.3. The total number of expressed genes was 19.165. Data are deposited in the GEO database with the accession number GSE106169. Genes were considered differentially expressed when they were at least two fold up-/down-regulated and if their Benjamini-Hochberg adjusted p-value was less than 0.01. The statistical significance of the overlap between differentially expressed genes was assessed by means of the hypergeometric distribution test. Gene set enrichment analysis was performed with the online application available at http://software.broadinstitute.org/gsea/index.jsp).

Generation and prognostic evaluation of gene signatures Lists of differentially expressed genes between shAKR1B1 and shSORD versus control cells were compared to calculate the degree of overlap. A previously generated signature comprising 494 genes differentially expressed during EMT in different cancers (15) was compared with protein-coding genes differentially expressed upon AKR1B1 and SORD knockdown considering the directionality, i.e. genes down-regulated upon shAKR1B1 and shSORD were compared with genes up-regulated during EMT and vice versa. Overlaps are shown as Venn diagrams. The significance of the overlap between the gene lists was tested by a hypergeometric distribution test as implemented in http://nemates.org/MA/progs/overlap_stats.html. The knockdown (KD) score in patients was defined as the sum of z scores of the commonly up-regulated genes minus the sum of z scores of the commonly down-regulated genes upon AKR1B1 and SORD knockdown for each patient. z scores were calculated in SPSS software. The separation of patients based on their KD score was done from lower and upper quartiles for GSE14333 and GSE7390, and from median for GSE17710. Kaplan Meier plots were generated in GraphPad software. Statistical significance of the separation between the two curves was assessed by log-rank test. The lists of genes used in all the analyses are provided in the **Supplementary Table**.

Chromatin Immunoprecipitation (ChIP). ChIP was performed as previously described (16), except additional crosslinking with 1.5 mM ethylene glycol bis(succinimidyl succinate) (EGS) for 30 min before fixing with 1% formaldehyde. 100 µg chromatin per experimental condition was incubated with 5 µg anti-ZEB1 (Santa Cruz Biotechnology, H102) or 5 µg normal rabbit IgG control antibody (Santa Cruz). After precipitation and de-crosslinking, DNA was purified with a QIAquick PCR purification kit (Qiagen). *AKR1B1* promoter regions in ZEB1 bound chromatin were quantified using real Time PCR. Primers used: *EPCAM*, forward (5′-GCC AGG TAA AAG CTC AAA GG-3′) and reverse (5′- GCG GGA ACT GGA TAG AGG A -3′); *AKR1B1-1* forward (5′- AGC CGT CTC CTG CTC AAC-3′) and reverse (5′-CGG AGA GTG TGA GGC GAG-3); *AKR1B1-2* forward (5′-CGC TTT CCC ACC AGA TAC AG-3′) and reverse (5′- TAG TGG CAG CGGATT CTT TC -3); *AKR1B1-3* forward (5′-CCT AGA GTG GGG TGC AAA GA -3′) and reverse (5′- TCC TTT TGC AAA GCA CCT TC -3); *HPRT* forward (5′- TGA GAG TTC AAG TTG AGT TTG GA -3′) and reverse (5′- TGA TAA TTT TAC TGG CGA TGT CA -3).

ELISA For measurement of extracellular TGF-β, culture supernatants were collected 48h after plating in serum-free media and centrifuged at 1,000 rpm and 4 °C for 10 min. ELISA was performed using the human TGF-β 1 ELISA kit (Sigma) according to the manufacturer's instruction. Absorption rates were measured with the SpectraMax190 (Molecular Devices) and normalized to the cell number at the time of collection.

Human specimens Formalin-fixed paraffin embedded (FFPE) surgical specimens of tumor tissues of 59 NSCLC patients completely resected between 2005 and 2006 at the San Luigi Hospital were consecutively collected. The main patients' characteristics are shown in **S3A**. None of the patients received pre-surgical chemo/radiation therapy. All cases were reviewed and classified (WHO classification) using anonymized samples. The studies were conducted in accordance with the Declaration of Helsinki. The Research Ethics Committee of the San Luigi Hospital, University of Turin, had approved the retrospective use of solid tumor tissues for immunohistochemistry (IHC)-staining (approvals n.167/2015 and 204/2016).

Mouse specimens Mouse specimens were obtained from a murine model of AOM-induced colon tumorigenesis with an intestinal epithelial cell-specific p53 deletion (p53ΔIEC) and corresponding wild type animals (17). Two investigators (IR and MV) independently reviewed the H&E stained FFPE tissue sections to identify tumors. Of the n=30 p53ΔIEC tumors n=24 were non-invasive and n=6 were locally invasive. Samples were stained and scored as described below. All mouse experiments were reviewed and approved by the Regierungspräsidium Darmstadt, Darmstadt, Germany.

Immunohistochemistry Tissue sections were deparaffinized and rehydrated with a series of decreasing ethanol concentrations. Antigen retrieval was performed using a pressure cooker in citrate buffer (pH 6) for 30 min. For AKR1B1 detection, endogenous peroxidase activity was blocked using 3% H2O2 for 15 min, followed by blocking of non-specific binding sites with 1% BSA in TBS-T for 30 min. The primary antibody incubation (AKR1B1 from Thermo-Fisher at 1:25, or the E-Cadherin from Neomarkers at 1:30) was performed for 1h. After washing, all sections were

incubated with rabbit Dako REAL Envision-HRP for 40 min. DAB substrate (Dako) was used for visualization of antibody binding. Sections were counterstained with haematoxylin, rehydrated and mounted. Immunoreactivity scorings were evaluated independently by two pathologists with an H-score, which was generated from the following equation: H-score= ΣPi (i + 1), where 'I' represents the intensity of staining (0–3+), and 'Pi' stands for the percentage of stained tumor cells (0% to 100%).

Gas chromatography–mass spectrometry Cells were harvested for metabolite analysis by direct scraping in 80% methanol as recently described (18). For sample preparation, the cell suspension was thawed, vortexed and spiked with 10 µL internal standard containing $^{13}C_6$-glucose at a concentration of 1mM. The sample was then centrifuged (9,560 g, 6 min, 4 °C) and the supernatant was removed. The cell pellet was washed twice with 200 µL 80% methanol, the supernatants were combined and dried in a vacuum evaporator (CombiDancer, Hettich AG, Bäch, Switzerland). Sugars were analyzed by GC-MS after methoximation and silylation using the derivatization protocol and instrumental setup recently described (18). An injection volume of 1 µL and splitless injection were employed. Quantification was performed using calibration curves with $^{13}C_6$-glucose as internal standard. Intracellular metabolite levels were normalized to protein amount. The protein pellet obtained after extraction was resuspended in a NaH_2PO_4 buffer (20 mM with 1.2 % SDS) and the protein amount was determined using the FluoroProfile Protein Quantification Kit (Sigma).

Statistical analysis In the transcriptomic analysis (performed on Novartis and U133 gene array datasets) correlations were assessed by Pearson's statistics, statistical difference between groups was assessed with Mann–Whitney U tests, and post-hoc analysis was performed with the Bonferroni method for multiple comparisons. Significant associations with clinical-pathological variables were tested with the Mann–Whitney U and the Kruskal–Wallis tests. *In vitro* experiments were performed in replicates, independently repeated and representative results are shown. T-tests in the *in vitro* experiments were performed comparing groups of different conditions with replicates. In western blot experiments, integrated densitometry data were obtained using the ImageJ software. Statistical significance in proliferation and migration assay was analyzed with two-way ANOVA by GraphPad Prism 7.0. Results were

considered statistically significant if $p<0.05$ (* indicates $p<0.05$, ** $p<0.01$ and *** $p<0.001$).

RESULTS

AKR1B1 correlates with the EMT phenotype of cancer cells. A mesenchymal-like phenotype is a common feature of de-differentiated aggressive cancers (5,19). To identify genes associated with EMT in an unbiased fashion, we performed a transcriptomic analysis using two gene-array datasets from the NCI-60 panel of cancer cell lines (**Figure 1A**). For each dataset, we generated a Vimentin/E-Cadherin gene ratio (*VIM/CDH1*) and the two ratios were found to be strongly correlated (**S1A**). The ability of this gene ratio to differentiate between epithelial-like and mesenchymal-like cancer cells was validated dividing cells by EMT status according to a previously reported western blot quantification (20) (**Figure 1B**). Pearson's statistics was then used to identify genes most significantly positively correlating with *VIM/CDH1* in both datasets (mesenchymal genes, **1A**). The genes present in both lists were re-ranked for the fold level of increase (mesenchymal- versus epithelial-like) and the final list included a total of 35 genes (**Figure 1C** and **S1B**). Several known mesenchymal markers were present, like fibronectin (*FN1*), TGF-β1 (*TGFB1*) and N-Cadherin (*CDH2*), as well as recently identified EMT determinants, like *LOXL2* (21) and *VCAN* (22) (**S1B**). Among the top-ranked genes, we focused our attention on the aldo-keto reductase family 1, member B1 *(AKR1B1),* which exhibited in average 7-fold higher mRNA expression in mesenchymal-like cells (**Figure 1D**). This was also further confirmed analyzing the dataset from the Cancer Cell Line Encyclopedia (CCLE) on more than 900 cancer cells, dividing the cells in epithelial- and mesenchymal-like based on the *VIM/CDH1* gene ratio (**Figure 1E**). This gene encodes an enzyme catalyzing the reduction for numerous aldehydes, most importantly glucose as part of the polyol pathway. Other members of the same superfamily without any connection to the PP, such as AKR1B10 (71% sequence identity) (23) did not show a correlation with EMT (**S1C**), suggesting that the role in EMT could be a specific property of the B1 member.

In order to validate these findings at the protein level, we quantified AKR1B1 in lysates from a panel of cancer cells (mainly belonging to the NCI-60). The results of this analysis (**Figure 1F**) clearly indicated a significant increase in the expression of

AKR1B1 in mesenchymal-like cells (indicated by no E-Cadherin, high Vimentin and ZEB1 expression), with the appearance of a specific band at 37-38 kDa. Few exceptions were observed, but overall the analysis of the protein level confirmed the data from mRNA levels and emphasizes the possible role of this enzyme in EMT.

AKR1B1 associates with ZEB1 expression. The transcription factor ZEB1 is a master regulator of EMT (24). Gene array analysis from the NCI-60 panel indicated a significant positive correlation between *AKR1B1* and *ZEB1* expression (**Figure 1G**). To functionally explore this association, we measured mRNA and protein levels in MDA-MB-231 cells stably knocking down ZEB1, a condition that suppresses EMT morphology and markers expression and found that AKR1B1 was significantly down-regulated (**Figure 1H-J**), strengthening its association with EMT. Conversely, analysis of GEO profiling data from breast cancer cells in which miR-200c (ZEB1 inhibitor and EMT-suppressing miRNA) was overexpressed (25) indicated a significant *Akr1b3* (*AKR1B1* mouse orthologue) reduction, along with EMT suppression (**S2A**). A significant AKR1B1 downregulation was also obtained knocking down ZEB1 with siRNA and shRNA in lung cancer A549 cells (**S2B-C**). Similarly, protein analysis of cell lines derived from pancreatic tumors of *Kras*^{G12D};*Trp53*^{R172H};*Pdx1*-Cre (KPC) mice and their *Zeb1* knockout counterpart (KPCZ) (24), indicated higher AKR1B1 levels in the mesenchymal-like KPC and a trend towards a lower expression in the KPCZ, see **S2D-E**. To understand if ZEB1 was directly involved in AKR1B1 regulation we performed chromatin immunoprecipitation (ChIP) and transfection experiments, but the results failed to indicate a direct regulation, as no significant binding to *AKR1B1* promoter was detected (**S2F**), and no AKR1B1 up-regulation followed ZEB1 overexpression (**S2G-H**).

AKR1B1 correlates with more aggressive and invasive cancers. In order to test the association between AKR1B1 and EMT in patient tissues, samples from a cohort of consecutively resected non-small cell lung cancer (NSCLC) patients (n=59, **S3A**) were stained by immunohistochemistry (IHC) and a significant negative correlation between AKR1B1 and E-Cadherin was found (**Figure 2A-B**). Also, a trend for a higher AKR1B1 expression in tumors with a higher N status was observed (extent of lymph nodal involvement, **S3B**). Interestingly, analysis of comparative profiling (GEO profiles) on airway epithelial cells isolated from non-smoker vs. smoker subjects,

revealed a striking increase in AKR1B1 levels, suggesting a possible role on lung tumorigenesis (**Figure 2C**). Survival analysis from large datasets of NSCLC, ovarian and gastric cancer patients confirmed a strong negative prognostic role for AKR1B1 (**Figure 2D and S3C**).

To test the *in vivo* relevance of AKR1B1 in tumorigenesis and invasiveness, an IHC analysis was performed on samples from a mouse model of AOM-induced colon tumorigenesis with an intestinal epithelial cell-specific *Trp53* deletion ($p53^{\Delta IEC}$) (17), **Figure 2E**. This model has been previously characterized as shifting from a non-invasive to an invasive EMT-like phenotype upon p53 inactivation, (17), in line with p53's EMT-suppressing properties (26). Samples were scored for AKR1B1 expression and the results indicated significantly higher expression in invasive tumors from $p53^{\Delta IEC}$ mice as compared to both p53-deficient non-invasive or wildtype counterparts (**Figure 2F-H**). All together, these results obtained in tissues confirmed the association from cultured cells and strengthened the correlation of AKR1B1 with aggressive EMT-driven tumors.

AKR1B1 suppression inhibits EMT and cancer cell growth. To understand if AKR1B1 plays a role in cancer EMT, we subjected A549 cells to shRNA-mediated knockdown using two independent sequences targeting AKR1B1, and western blot analysis indicated that the EMT phenotype was suppressed (**Figure 3A**). Knockdown cells acquired a cobblestone shape (**Figure 3B**), and immunofluorescence staining confirmed increased E-Cadherin and reduced Vimentin expression (**Figure 3C**). To check if EMT reversal affected migratory ability, a hallmark of EMT, a wound-healing assay was performed and indicated a significant suppression of motility in shAKR1B1 cells (**Figure 3D**). Migration experiments were also analyzed with a matrix (relative wound density) that normalizes for proliferation. This is because, along with a reduced motility, a significant reduction in growth rate (**Figure 3E**) and colony formation (**Figure 3F**) was observed, in line with previous reports (27). The reduction in proliferation was comparable to that measured in shZEB1 cells (**S3D**), possibly linking the loss of growth potential to the EMT suppression. No significant difference was observed in the relative proportion of cells in each phase of the cell cycle (**Figure 3G** and **S3E**). Similar results in terms of E-Cadherin expression and growth suppression upon AKR1B1 knockdown were obtained in the ovarian cancer cell line Ovcar-5 (**Figure 3H-I**). Conversely, AKR1B1 overexpression in MCF-7 cells reduced

E-Cadherin and slightly increased ZEB1 (**Figure 3J**), while stimulating proliferation (**S3F**). Altogether, these results indicated that AKR1B1 has a direct impact on EMT-associated phenotypes and proliferation.

AKR1B1 regulates the CSC phenotype. EMT and the CSC phenotype are strictly functionally associated (28). We analyzed GEO profiling data of ovarian cancer cells grown as monolayers compared to 3D spheroids, a condition used to enrich for CSCs (29), and *AKR1B1* was found to be the top-ranked up-regulated gene (**Figure 4A**). We therefore compared the AKR1B1 protein levels in HeyA8 (ovarian) and A549 cells maintained in adherent or spheroid-forming conditions, and an increased AKR1B1 expression in the tumorspheres was detected (**Figure 4B**). Consistently, shAKR1B1 cells displayed reduced ability to form spheres (**Figure 4C** and **S3G**). FACS analysis of activated aldehyde dehydrogenase H1 (ALDH1), an established CSC marker (30), indicated a significant reduction of $ALDH1^+$ cells upon knockdown (**S3H**). Comparable results were also obtained treating cells with the AKR1B1 inhibitor Sorbinil (**Figure 4D-F**), confirming the involvement of AKR1B1 in cancer stemness.

SORD knockdown phenocopies AKR1B1 inhibition and suppresses EMT. AKR1B1 belongs to the polyol pathway (PP), a two-step enzymatic reaction converting excess intracellular glucose into fructose, and sorbitol dehydrogenase (SORD) is the second-in-line enzyme (13) (**Figure 5A**). We therefore investigated the impact of SORD loss on EMT. As a result, shRNA-mediated knockdown of SORD was found to significantly block EMT and CSC marker expression (**Figure 5B-C**), and to inhibit colony formation (**Figure 5D**), migratory ability and growth (**Figure 5E-F**) to a similar extent as the AKR1B1 knockdown. Comparable results were also obtained from other cancer cells (**Figure 5G** and **S4A-C**) and *SORD* gene expression revealed a prognostic role in NSCLC (**Figure 5H**). Importantly, both PP-deficient cells still displayed sensitivity towards cisplatin and the combination of knockdown and drug treatment induced an almost complete proliferation arrest (**S4D**). Assessment of cells' metabolic phenotype with the Seahorse analyzer indicated a comparable significant loss of activity in both knockdown cells, affecting basal respiration (**S4E**) as well as glycolysis, a condition that could not be rescued by the addition of external glucose (**Figure 5I**). Interestingly, SORD knockdown was found to induce a significant reduction in AKR1B1 levels in some cells (e.g. Ovcar-5 in **Figure 5G**), indicating the

possible existence of a cell- or tumor-specific co-regulation of these genes. However, a comprehensive analysis of *AKR1B1/SORD* co-expression levels in tumors from the TCGA dataset indicated tissue-specific patterns and a higher prevalence of a negative correlation (**S5**).

RNA profiling of PP-deficient cells reveals alterations of EMT signature genes. To better characterize changes associated with the loss of PP genes at an unbiased and high-throughput level, we performed RNA sequencing profiling of shAKR1B1 and shSORD cells compared to scrambled controls (**Figure 6A**). The results indicated a very significant overlap ($p<0.0000001$) between the genes up- and down-regulated in both knockdowns (**Figure 6B**) and were successfully validated by qPCR (**Figure 6C**). Gene set enrichment analysis (GSEA) revealed that the pathway most significantly altered in the PP-deficient cells (overlapping genes) was EMT (**Figure 6D**). The comparison with a previously reported EMT signature comprising 494 coding genes (15) indicated a very significant overlap (**Figure 6E** and **Supplementary Table**), further confirming the functional association with EMT. Importantly, a signature generated with the genes most differentially down-regulated in shAKR1B1/shSORD cells (see Supplementary methods) proofed to be strongly prognostic in different types of cancer (**Figure 6F**).

Glucose-induced activation of PP controls EMT via TGF-β autocrine stimulation. Among the list of EMT genes suppressed by the knockdown of PP enzymes, we noticed the presence of several TGF-β-related genes, including *TGFB1* (**Figure 7A**). This was verified by a further comparison with a previously reported TGF-β-specific gene signature (31) ($p<0.0001$). TGF-β is a cytokine with an established major role in promoting EMT (32). Reduced levels of the soluble active form of TGF-β were measured in the supernatant of knockdown cells (**Figure 7B**), and exogenous TGF-β addition rescued EMT/CSC markers in PP-deficient cells (although with delayed kinetics, **Figure 7C-E**). Based on this evidence and on the previously reported connection between glucose uptake (33)**,** metabolism (8) and EMT, we decided to further explore the interplay between glucose, PP and TGF-β-driven EMT. First, we tested if excess glucose had polyol-generating and EMT-promoting effects. Cells grown in the presence of high glucose levels displayed a marked increase in PP metabolites, as evaluated by gas chromatography–mass spectrometry (GC-MS) analysis (**Figure 7F** and **S6A**), along with a significant

increase in EMT markers (see the Vimentin/E-Cadherin ratios), and AKR1B1 expression levels (**Figure 7G** and **S6B**), and an augmented migratory ability (**Figure 7H)**. Then, we hypothesized that glucose could promote EMT via the PP by an autocrine TGF-β stimulation. High glucose treatments were therefore repeated in the presence of a TGF-β neutralizing antibody, and this was sufficient to abrogate the glucose-induced increase of EMT markers (**Figure 7I**). Augmented TGF-β signaling following stimulation with excess glucose was further confirmed by an upregulation of phosphorylated Smad3 (**S6C**) and *RUNX2* transcript levels (**S6D**) as well as a significant increase in the Fra-1 protein (*FOSL1* gene) (**S6E**), previously shown to control *TGFB1* gene and EMT (34). Conversely, exogenous overexpression of AKR1B1 in epithelial-like cells was capable of increasing *TGBF1* and *RUNX2* (**S6F**). Finally, PP-deficient cells exposed to high glucose were found more refractory to EMT induction compared to control cells (**Figure 7J**). The knockdown of PP enzymes also consistently reduced cells' ability to increase the proportion of $ALDH1^+$ cells upon glucose treatment (**S6G**). We could therefore conclude a model in which excess glucose is partly metabolized in the PP supporting EMT via TGF-β autocrine stimulation (see **Figure 7K**).

DISCUSSION

The study of metabolic processes represents an emerging field in cancer biology (35) with a great translational potential (14,36-38). Here, we describe by several approaches that a glucose-metabolizing pathway is strongly correlated and functionally associated with cancer aggressiveness and the cancer differentiation program. Our data suggest that interfering with targetable cancer-relevant metabolic pathways may represent a valuable strategy to block or attenuate the EMT/CSC process and cancer malignancy.

PP has been implicated in the development of diabetic complications via the generation of reactive oxygen species (13). The activity of AKR1B1 has been previously linked to EMT in mouse normal lens epithelium (39,40) or in other normal cells (41). In cancer, AKR1B1 has been found to have a role in the growth and the aggressiveness of cancer cells (27), but was never associated with cancer cells' EMT, CSCs or related pathways, until recently. Wu and colleagues, in fact, reported that AKR1B1 can induce EMT via PGF2α synthesis and NF-κB activation in triple negative breast cancer cells (42). Based on the data here presented, which point at

the PP as strongly connected to EMT as well as the plasticity of other types of cancer cells, it is possible to speculate that the same enzyme has multiple ways to regulate EMT (in a tissue-specific manner). The expression of SORD has been previously detected in cancer cells (43,44), but was never functionally investigated, and the present data indicate an important role in cancer growth and differentiation worth of further investigations.

AKR1B1 deficiency in mice has been found to protect from AOM-induced pre-neoplastic aberrant crypt foci formation (45). The presence of an intestinal epithelial cell-specific p53 deletion was previously demonstrated to initiate EMT and promote invasive disease in this model (17), and we could here show a significant increase in AKR1B1 protein levels in samples from p53 deleted mice, particularly evident in invasive tumors. The *in vitro* and *in vivo* data (including from patients samples) of the present work strongly encourage further *in vivo* experimentations on AKR1B1 deficient mice in order to conclusively determine the role of AKR1B1 and associated pathways in tumorigenesis, in the EMT phenotype, and in cancer stemness. Importantly, an unbiased analysis of global RNA profiles in PP-deficient cells revealed a dramatic loss of EMT and TGF-β signature genes, most of which were previously found altered in the p53-deleted invasive mouse model (17), and this observation prompted us to investigate the interaction between glucose, PP and TGF-β in more detail. Several epidemiological and experimental studies have connected sucrose/fructose consumption with increased cancer risk (10-12). A few previous reports suggested that glucose uptake (33) and metabolism (8), as wells as fructose (46) could be connected to the occurrence of EMT, metastasis formation and with a more advanced stage in tumor progression (47), but the functional details were not defined. The present data highlight the role of alternative glucose metabolism pathways in cancer de-differentiation and aggressiveness clearly pointing at a direct role of glucose as an EMT promoter, as previously reported in some non-cancerous physiological contexts (48-50). A number of open questions remain to be answered. For instance, the detailed molecular mechanisms downstream of PP need to be identified for a deeper comprehension of the link between glucose/fructose and tumorigenesis. In addition, the exact causes for the observed loss of proliferation and metabolic activity in PP-deficient cells remain to be determined. Furthermore, the observed tissue-specific changes in the AKR1B1/SORD co-expression pattern (**S5**) suggest the importance of investigating the transcriptional regulation of these EMT-

promoting genes, as previously done for breast cancer (42), and the existence of a possible balance between them that controls cancer differentiation.

Noteworthy, PP genes and PP-related signatures were here found to be highly prognostic in various human cancer types, further confirming their important role in cancer progression. Targeting the expression or the activity of PP enzymes (alone or in combination with chemotherapy) may therefore represent a novel effective therapeutic strategy for the treatment of aggressive tumors.

ACKNOWLEDGMENTS. Work supported by the Interdisciplinary Center for Clinical Research (IZKF) of the University of Erlangen-Nuremberg, the Deutsche Krebshilfe grant number 70112536, the IALSC Lung Cancer Young Investigator Award to PC, and by the Clinical Research Group KFO262 funded by the German Research Foundation. Work partially presented at the American Association for Cancer Research 2017 annual meeting. Special thanks to Dr. H. Wurdak (University of Leeds) for critical discussion.

REFERENCES

1. Jogi A, Vaapil M, Johansson M, Pahlman S. Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors. Ups J Med Sci 2012;117(2):217-24.
2. Medema JP. Cancer stem cells: the challenges ahead. Nat Cell Biol 2013;15(4):338-44.
3. Visvader JE, Lindeman GJ. Cancer stem cells: current status and evolving complexities. Cell Stem Cell 2012;10(6):717-28.
4. Ceppi P, Peter ME. MicroRNAs regulate both epithelial-to-mesenchymal transition and cancer stem cells. Oncogene 2014;33(3):269-78.
5. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol 2017.
6. Liu PP, Liao J, Tang ZJ, Wu WJ, Yang J, Zeng ZL, et al. Metabolic regulation of cancer cell side population by glucose through activation of the Akt pathway. Cell Death Differ 2014;21(1):124-35.
7. Menendez JA, Joven J, Cufi S, Corominas-Faja B, Oliveras-Ferraros C, Cuyas E, et al. The Warburg effect version 2.0: metabolic reprogramming of cancer stem cells. Cell Cycle 2013;12(8):1166-79.
8. Lin CC, Cheng TL, Tsai WH, Tsai HJ, Hu KH, Chang HC, et al. Loss of the respiratory enzyme citrate synthase directly links the Warburg effect to tumor malignancy. Sci Rep 2012;2:785.
9. Dong CF, Yuan TT, Wu YD, Wang YF, Fan TWM, Miriyala S, et al. Loss of FBP1 by Snail-Mediated Repression Provides Metabolic Advantages in Basal-like Breast Cancer. Cancer Cell 2013;23(3):316-31.

10. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 2003;348(17):1625-38.
11. Dowman JK, Hopkins LJ, Reynolds GM, Nikolaou N, Armstrong MJ, Shaw JC, et al. Development of hepatocellular carcinoma in a murine model of nonalcoholic steatohepatitis induced by use of a high-fat/fructose diet and sedentary lifestyle. Am J Pathol 2014;184(5):1550-61.
12. Liu H, Heaney AP. Refined fructose and cancer. Expert Opin Ther Targets 2011;15(9):1049-59.
13. Tang WH, Martin KA, Hwa J. Aldose reductase, oxidative stress, and diabetic mellitus. Front Pharmacol 2012;3:87.
14. Siddiqui A, Vazakidou ME, Schwab A, Napoli F, Fernandez-Molina C, Rapa I, et al. Thymidylate synthase is functionally associated with ZEB1 and contributes to the epithelial-to-mesenchymal transition of cancer cells. J Pathol 2017;242(2):221-233.
15. Huang S, Holzel M, Knijnenburg T, Schlicker A, Roepman P, McDermott U, et al. MED12 controls the response to multiple cancer drugs through regulation of TGF-beta receptor signaling. Cell 2012;151(5):937-50.
16. Vannier C, Mock K, Brabletz T, Driever W. Zeb1 regulates E-cadherin and Epcam (epithelial cell adhesion molecule) expression to control cell behavior in early zebrafish development. J Biol Chem 2013;288(26):18643-59.
17. Schwitalla S, Ziegler PK, Horst D, Becker V, Kerle I, Begus-Nahrmann Y, et al. Loss of p53 in enterocytes generates an inflammatory microenvironment enabling invasion and lymph node metastasis of carcinogen-induced colorectal tumors. Cancer Cell 2013;23(1):93-106.
18. Dettmer K, Nurnberger N, Kaspar H, Gruber MA, Almstetter MF, Oefner PJ. Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. Anal Bioanal Chem 2011;399(3):1127-39.
19. Nieto MA, Cano A. The epithelial-mesenchymal transition under control: global programs to regulate epithelial plasticity. Semin Cancer Biol 2012;22(5-6):361-8.
20. Park SM, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors, ZEB1 and ZEB2. Genes Dev 2008;22:894-907.
21. Canesin G, Cuevas EP, Santos V, Lopez-Menendez C, Moreno-Bueno G, Huang Y, et al. Lysyl oxidase-like 2 (LOXL2) and E47 EMT factor: novel partners in E-cadherin repression and early metastasis colonization. Oncogene 2015;34(8):951-64.
22. Xia L, Huang W, Tian D, Zhang L, Qi X, Chen Z, et al. Forkhead box Q1 promotes hepatocellular carcinoma metastasis by transactivating ZEB2 and VersicanV1 expression. Hepatology 2014;59(3):958-73.

23. Zhang L, Zhang H, Zhao Y, Li Z, Chen S, Zhai J, et al. Inhibitor selectivity between aldo-keto reductase superfamily members AKR1B10 and AKR1B1: role of Trp112 (Trp111). FEBS Lett 2013;587(22):3681-6.
24. Krebs AM, Mitschke J, Lasierra Losada M, Schmalhofer O, Boerries M, Busch H, et al. The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. Nat Cell Biol 2017;19(5):518-29.
25. Knezevic J, Pfefferle AD, Petrovic I, Greene SB, Perou CM, Rosen JM. Expression of miR-200c in claudin-low breast cancer alters stem cell functionality, enhances chemosensitivity and reduces metastatic potential. Oncogene 2015;34(49):5997-6006.
26. Powell E, Piwnica-Worms D, Piwnica-Worms H. Contribution of p53 to metastasis. Cancer Discov 2014;4(4):405-14.
27. Saxena A, Tammali R, Ramana KV, Srivastava SK. Aldose reductase inhibition prevents colon cancer growth by restoring phosphatase and tensin homolog through modulation of miR-21 and FOXO3a. Antioxid Redox Signal 2013;18(11):1249-62.
28. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008;133(4):704-15.
29. Ceppi P, Hadji A, Kohlhapp FJ, Pattanayak A, Hau A, Liu X, et al. CD95 and CD95L promote and protect cancer stem cells. Nat Commun 2014;5:5238.
30. Kim IG, Kim SY, Choi SI, Lee JH, Kim KC, Cho EW. Fibulin-3-mediated inhibition of epithelial-to-mesenchymal transition and self-renewal of ALDH+ lung cancer stem cells through IGF1R signaling. Oncogene 2014;33(30):3908-17.
31. Bluthgen N, Kielbasa SM, Herzel H. Inferring combinatorial regulation of transcription in silico. Nucleic Acids Res 2005;33(1):272-9.
32. Xu J, Lamouille S, Derynck R. TGF-beta-induced epithelial to mesenchymal transition. Cell Res 2009;19(2):156-72.
33. Liu M, Quek LE, Sultani G, Turner N. Epithelial-mesenchymal transition induction is associated with augmented glucose uptake and lactate production in pancreatic ductal adenocarcinoma. Cancer Metab 2016;4:19.
34. Bakiri L, Macho-Maschler S, Custic I, Niemiec J, Guio-Carrion A, Hasenfuss SC, et al. Fra-1/AP-1 induces EMT in mammary epithelial cells by modulating Zeb1/2 and TGFbeta expression. Cell Death Differ 2015;22(2):336-50.
35. Pavlova NN, Thompson CB. The Emerging Hallmarks of Cancer Metabolism. Cell Metab 2016;23(1):27-47.
36. Colvin H, Nishida N, Koseki J, Konno M, Kawamoto K, Doki Y, et al. The Role of Glutaminase C and Glutamine Metabolism in Epithelial-Mesenchymal Transition in Colorectal Cancer. Ann Surg Oncol 2016;23:S84-S84.
37. Tsai YC, Chen SL, Chang HSV, Ch'ang HJ. Loss of Krupple-like factor 10 enhanced epithelial-mesenchymal transition of pancreatic cancer by modulating glucose metabolism via Sirt 6 and PKM2. Cancer Res 2016;76.

38. Schieber MS, Chandel NS. ROS links glucose metabolism to breast cancer stem cell and EMT phenotype. Cancer Cell 2013;23(3):265-7.
39. Zablocki GJ, Ruzycki PA, Overturf MA, Palla S, Reddy GB, Petrash JM. Aldose reductase-mediated induction of epithelium-to-mesenchymal transition (EMT) in lens. Chem Biol Interact 2011;191(1-3):351-6.
40. Chang KC, Petrash JM. Aldose Reductase Mediates Transforming Growth Factor beta2 (TGF-beta2)-Induced Migration and Epithelial-To-Mesenchymal Transition of Lens-Derived Epithelial Cells. Invest Ophthalmol Vis Sci 2015;56(8):4198-210.
41. Wei J, Zhang Y, Luo Y, Wang Z, Bi S, Song D, et al. Aldose reductase regulates miR-200a-3p/141-3p to coordinate Keap1-Nrf2, Tgfbeta1/2, and Zeb1/2 signaling in renal mesangial cells and the renal cortex of diabetic mice. Free Radic Biol Med 2014;67:91-102.
42. Wu X, Li X, Fu Q, Cao Q, Chen X, Wang M, et al. AKR1B1 promotes basal-like breast cancer progression by a positive feedback loop that activates the EMT program. J Exp Med 2017;214(4):1065-79.
43. Szabo Z, Hamalainen J, Loikkanen I, Moilanen AM, Hirvikoski P, Vaisanen T, et al. Sorbitol dehydrogenase expression is regulated by androgens in the human prostate. Oncol Rep 2010;23(5):1233-9.
44. Uzozie A, Nanni P, Staiano T, Grossmann J, Barkow-Oesterreicher S, Shay JW, et al. Sorbitol dehydrogenase overexpression and other aspects of dysregulated protein expression in human precancerous colorectal neoplasms: a quantitative proteomics study. Mol Cell Proteomics 2014;13(5):1198-218.
45. Tammali R, Reddy AB, Ramana KV, Petrash JM, Srivastava SK. Aldose reductase deficiency in mice prevents azoxymethane-induced colonic preneoplastic aberrant crypt foci formation. Carcinogenesis 2009;30(5):799-807.
46. Liu H, Huang D, McArthur DL, Boros LG, Nissen N, Heaney AP. Fructose induces transketolase flux to promote pancreatic cancer growth. Cancer Res 2010;70(15):6368-76.
47. Du S, Guan Z, Hao L, Song Y, Wang L, Gong L, et al. Fructose-bisphosphate aldolase a is a potential metastasis-associated marker of lung squamous cell carcinoma and promotes lung cell tumorigenesis and migration. PloS One 2014;9(1):e85804.
48. Lee YJ, Han HJ. Troglitazone ameliorates high glucose-induced EMT and dysfunction of SGLTs through PI3K/Akt, GSK-3beta, Snail1, and beta-catenin in renal proximal tubule cells. Am J Physiol Renal Physiol 2010;298(5):F1263-75.
49. Liu L, Wang Y, Yan R, Li S, Shi M, Xiao Y, et al. Oxymatrine Inhibits Renal Tubular EMT Induced by High Glucose via Upregulation of SnoN and Inhibition of TGF-beta1/Smad Signaling Pathway. PloS One 2016;11(3):e0151986.

50. Tang WB, Ling GH, Sun L, Zhang K, Zhu X, Zhou X, et al. Smad Anchor for Receptor Activation Regulates High Glucose-Induced EMT via Modulation of Smad2 and Smad3 Activities in Renal Tubular Epithelial Cells. Nephron 2015;130(3):213-20.

FIGURE LEGENDS

Figure 1. *An EMT gene ratio identifies AKR1B1 as strongly associated with the mesenchymal-like phenotype*. **A,** Scheme of the bioinformatic comparison of Vimentin/E-Cadherin gene expression ratios (*VIM/CDH1*) independently generated in two gene array datasets (U133 and Novartis) from the NCI-60 panel of cancer cells. Genes positively correlating with the ratios at $R^{Pearson}>0.4$ were extracted from each dataset and the overlap between the two datasets identified 61 genes (p-value for the overlap is $p<0.0001$), which were re-ranked based on the level of fold-increase between mesenchymal- vs epithelial-like cancer cells and tested for statistical significance (Mann-Whitney U test). Thirty-five genes were significant. **B,** Association between the *VIM/CDH1* gene ratio (Novartis) and a previously published determination of EMT-status of the NCI-60 cells, which defined three groups (epithelial-like (Ep), undefined (Undef) or mesenchymal-like (Mes)). P-value is a Kruskal–Wallis test. **C,** Gene expression analysis of the NCI-60 panel of cancer cells showing the top 35 genes most correlating with the *VIM/CDH1* ratio in both databases, ranked based on the level of fold-increase as described in A. **D,** *AKR1B1* mRNA expression levels (log2) in epithelial- and mesenchymal-like cancer cell lines from the NCI-60 (Novartis). P-value is a Mann-Whitney U test. **E,** *AKR1B1* mRNA expression data in cells from the CCLE collection, divided in two groups based on the 25th and 75th percentile of the *VIM/CDH1* ratios (originated from the same source and calculated for each cell line), indicating cells with epithelial and mesenchymal-like phenotype, respectively. P-value was calculated by a Mann-Whitney U test. **F,** Western blot analysis of EMT markers and AKR1B1 in a panel of various cancer cells. **G,** Correlation between *AKR1B1* and *ZEB1* mRNA levels in the NCI-60 dataset (Novartis). R is correlation coefficient; p-value was calculated with Pearson's statistics. **H,** Morphological appearance of MDA-MB-231 cells stably expressing shControl (shCtrl) and shZEB1. Scale bar is 50 µm. **I,** *AKR1B1* mRNA and **J,** protein levels in MDA-MB-231 cells stably expressing shRNA targeting ZEB1 compared to control cells. Bars are avg±SD. * indicates > than a 2-fold regulation. Experimental

data are representative from at least two independent experiments with similar results.

Figure 2. *AKR1B1 correlates with more aggressive and invasive cancers.* **A,** Immunohistochemical scores of AKR1B1 in cancer tissues from 59 NSCLC patients grouped by E-Cadherin positivity (membrane pattern vs negative or diffuse cytoplasmic staining). P-value is from a Mann-Whitney U test. **B,** Representative pictures of one AKR1B1^{neg}E-Cad^{pos} and one AKR1B1^{pos}E-Cad^{neg} case. Scale bars are 100 µm. **C,** Analysis of gene array data (GSE5060) showing *AKR1B1* levels from airway epithelial cells isolated from non-smoker vs. smoker subjects. P-value is from a Mann-Whitney U test. **D,** Survival analysis from a public database (KMplotter) showing poorer outcome of lung cancer (NSCLC) patients with higher *AKR1B1* levels (red lines) compared to lower expression levels (black lines). P value is log-rank test. **E,** Scheme of the tumor induction with azoxymethane (AOM) in wild-type (WT, or *TP53*$^{F/F}$) and p53 knockout mice (generated crossing the *TP53*$^{F/F}$ with the Villin-Cre mouse, to obtain the *TP53*$^{\Delta IEC}$, or p53$^{-/-}$). All mice were injected with 10 mg/kg AOM for 6 weeks and sacrificed at week 18. Pictures show hematoxylin and eosin staining of one non-invasive (right) and one invasive (left) tumor from the indicated genotypes. **F,** Representative immunohistochemical staining of AKR1B1 (mouse orthologue) in tumor tissues with the indicated genotypes. Scale bars are 100 µm. **G,** Immunohistochemical scores of AKR1B1 in tumor tissues from WT vs p53$^{-/-}$ mice. P-value is from a Mann-Whitney U test. **H,** Immunohistochemical scores of AKR1B1 in tissues from p53$^{-/-}$ mice, comparing non-invasive vs invasive tumors. P-value is from a Mann-Whitney U test. Number of observations (individual tumors) is indicated in brackets.

Figure 3. *AKR1B1 suppression inhibits EMT and the growth of cancer cells.* **A,** Western blot analysis of AKR1B1 and EMT markers in A549 cells infected with two independent shRNAs targeting the mRNA of AKR1B1, compared to scrambled-infected cells (shCtrl). **B,** Morphological appearance and **C,** immunofluorescence of E-Cadherin and Vimentin expression in A549 cells with control or AKR1B1 knockdown. DAPI was used as a nuclear counterstain. Scale bar are 50 and 20 µm, respectively. **D,** Representative pictures and quantification of wound-healing assay of A549 cells infected with shAKR1B1 or control cells. Lines in the pictures indicate the un-invaded area. Scale bar is 300 µm. Plotted are relative wound-width and relative

wound density over time. Points (n=6) are avg±SD. P-values are two-tailed t test and two-way ANOVA, respectively. **E,** Real-time proliferation assay of A549 cells infected with non-targeting shCtrl or shAKR1B1 expressing viruses. Plotted is cells' confluency over time. Points (n=4) are avg±SD. P-value is two-way ANOVA. **F,** Colony formation of indicated cells, stained 8 days after plating with crystal violet and quantified in triplicate dishes. Bars are avg±SD. P-values are two-tailed t-tests, ***, p< 0.001. **G,** FACS histograms of cell cycle analysis of indicated cells, and stained with PI (Nicoletti assay). **H,** Western blot analysis of AKR1B1 and E-Cadherin protein levels in Ovcar-5 cells infected with shRNA targeting AKR1B1, compared to shCtrl cells. **I,** Real-time proliferation assay of Ovcar-5 cells with shAKR1B1 or control cells. Points (n=4) are avg±SD. P-value is two-way ANOVA. **J,** Western blot analysis of AKR1B1, E-Cadherin and ZEB1 protein levels in MCF-7 cells overexpressing AKR1B1 or a control vector. Experimental data are representative from at least two independent experiments with similar results.

Figure 4. *AKR1B1 suppression inhibits CSC markers.* **A,** Analysis of gene array data (GSE80373) from ovarian cancer cells grown as 2D monolayers or 3D spheroids. Data are log2-transformed, corresponding to approximately 25-fold increase in spheroids. **B,** AKR1B1 protein levels in HeyA8 and A549 cells grown under adherent ('A') and in sphere-forming ('S') conditions. Indicated are the AKR1B1/β-Actin ratios. **C,** Representative pictures and quantification of spheres number in cells with AKR1B1 knockdown compared to control cells. Scale bar is 200 µm. Bars are avg±SD. P-values are two-tailed t-tests, ***, p< 0.001. **D,** FACS plots of cells treated with either 100 µM Sorbinil or DMSO and stained with Aldefluor reagent. Gates were set using DEAB reagent; percentages indicate the proportion of $ALDH1^+$ cells. **E,** Quantification and **F,** representative pictures of spheres formed following Sorbinil treatment and plating in low-adherence plates. Bars are avg±SD. P-values are two-tailed t-tests, **, p< 0.01. Scale bar is 100 µm. Experimental data are representative from at least two independent experiments with similar results.

Figure 5. *SORD knockdown phenocopies AKR1B1 inhibition and blocks EMT.* **A,** Schematic presentation of the polyol pathway (PP). **B,** Western blot analysis of SORD and EMT markers in A549 cells infected with three independent shRNAs targeting SORD, compared to scrambled-infected cells. **C,** FACS plots of A549 cells infected with shSORD or shCtrl stained with Aldefluor reagent. Gates were set using

DEAB reagent; percentages indicate the proportion of $ALDH1^+$ cells. **D,** Colony formation of A549 cells infected with shAKR1B1, shSORD or shCtrl and stained with crystal violet and quantified 8 days after plating in triplicate dishes. Bars are avg±SD. P-values are two-tailed t-tests, **, p< 0.01, ***, p< 0.001. **E,** Representative pictures and quantification of the wound-healing assay of A549 cells transfected with non-targeting (shCtrl) or with shSORD. Lines in the pictures indicate the un-invaded area. Scale bar is 300 µm. Points (n=6) are avg±SD. P-value is two-way ANOVA. **F,** Real-time proliferation assay of A549 cells infected with non-targeting (shCtrl) or with shSORD constructs. Plotted is confluency over time. Points (n=4) are avg±SD. P-value is two-way ANOVA. **G,** Western blot analysis of AKR1B1, SORD and EMT markers in Ovcar-5 cells stably expressing a scrambled control or a shRNA targeting SORD. **H,** Survival analysis of data derived from a public database (KMplotter) showing the poorer outcome of NSCLC cancer patients with higher SORD levels (red line) compared to lower SORD levels (black line). P value is log-rank test. **I,** Left, quantification of ECAR (extracellular acidification rate, glycolysis) in A549 cells with AKR1B1, SORD or control knockdown. Right, real time proton flux analysis in PP-deficient or control cells under basal glucose conditions (11 mM) or after the addition of 100 mM glucose. Dotted line represents time point of glucose injection, '+' indicates cell lines injected with glucose. Bars (n=5) are avg±SD. P-values are two-tailed t-tests, *, p< 0.05. Experimental data are representative from at least two independent experiments with similar results.

Figure 6. *RNA profiling of PP-deficient cells reveals alterations of EMT signature genes.* **A,** Western blot analysis of AKR1B1, SORD and EMT markers in A549 cells subjected to RNA sequencing. **B,** Venn diagrams showing the overlap between genes up- and down-regulated in knockdown cells, compared to controls. P-values are hypergeometric distribution tests. **C,** qPCR validation of selected up- and down-regulated genes in knockdown cells. Bars are avg±SD. * indicates > than a 2-fold regulation. **D,** GSEA analysis of commonly down-regulated genes in knockdown cells indicating the most significantly enriched pathways. **E,** Overlap between differentially expressed genes upon AKR1B1 and SORD knockdown and genes whose expression is regulated during EMT. 57 (p-value=3.276×10^{-22}) and 53 (p-value=4.545×10^{-22}) genes from shAKR1B1-up and shSORD-up lists, respectively, were common to genes down-regulated during EMT (left panel). 33 (p-value=1.020×10-^{16}) and 21 (p-value=1.467×10^{-9}) genes from shAKR1B1-down and

shSORD-down lists, respectively, were common to the genes up-regulated during EMT (right panel). P values are hypergeometric distribution tests**. F,** Kaplan Meier plots representing the percentage of overall survival (OS) in lung cancer (GSE17710), distant metastasis free survival (DMFS) in breast cancer patients (GSE7390), and disease free survival (DFS) in colorectal cancer (GSE14333), separated based on the expression of knockdown (KD) score. P values are log-rank tests.

Figure 7. *Glucose-induced activation of PP controls EMT via TGF-β autocrine stimulation.* **A,** qPCR analysis of TGF-β-related genes in A549 cells stably expressing shAKR1B1, shSORD or non-targeting controls. Bars are avg±SD. * indicates > than a 2-fold regulation. **B,** Relative TGF-β levels measured by ELISA in supernatants from shCtrl, shAKR1B1 and shSORD-infected A549 cells, conditioned for 4 days and normalized to the cell number at the time of collection. Values are normalized to shCtrl. Bars are avg±SD. P-values are two-tailed t-tests, **, p< 0.01. **C,** Western blot analysis of EMT markers in shCtrl, shAKR1B1 and shSORD-infected A549 cells treated with TGF-β (10 ng/ml) for 48 hours. **D,** Representative images of cells treated as in C. Scale bar is 50 µm. **E,** FACS plots of cells infected with shAKR1B1 or shCtrl, either untreated or treated with TGF-β (10 ng/ml) for 72 hours and stained with Aldefluor reagent. Gates were set using DEAB reagent; percentages indicate the proportion of $ALDH1^+$ cells. **F,** GC-MS analysis of cellular extracts from cells grown under normal (blue) and high glucose (red) conditions. Figure shows part of the overlaid monosaccharide section of the chromatograms. Due to the methoximation step during derivatization, two signals are obtained for fructose and glucose. **G,** Western blot analysis of EMT markers and **H,** wound-healing assay in parental A549 cells treated with indicated concentrations of glucose for 6 days. Points (n=4) are avg±SD. P-value is two-way ANOVA. VE ratio is the Vimentin/E-Cadherin ratio as evaluated by band densitometry. **I,** Western blot analysis of EMT markers in cells treated as in G in the presence of either a TGF-β neutralizing antibody or IgG control. TGF-β 1 ng/ml was used as a positive control. **J,** Western blot analysis of AKR1B1, SORD and EMT markers in A549 cells infected with shRNA targeting AKR1B1, SORD or shCtrl cells treated with glucose for 3 days. **K,** Scheme of the proposed role of polyol pathway genes in EMT and CSCs. Experimental data are representative from at least two independent experiments with similar results.

Figure 1

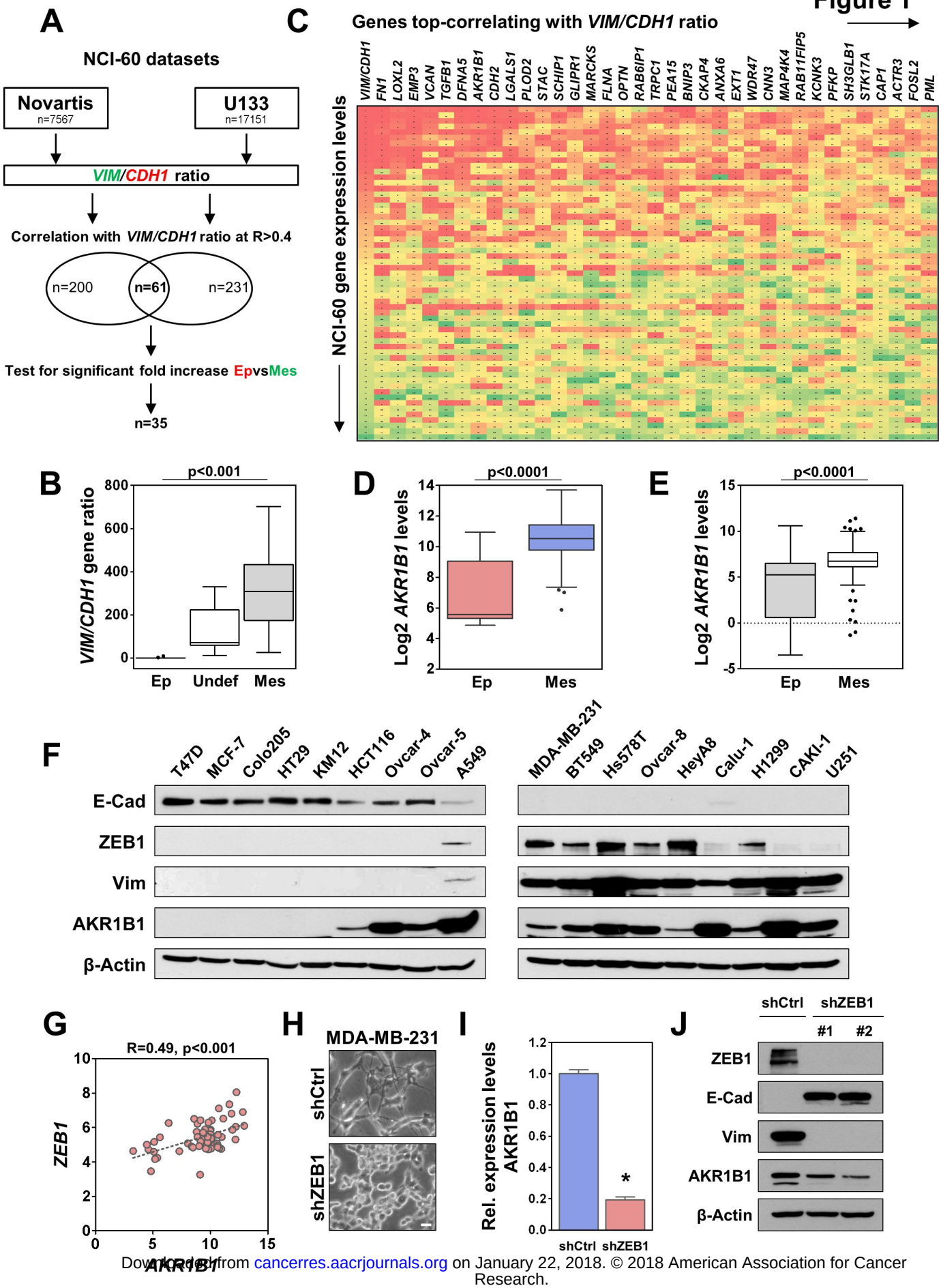


Figure 2

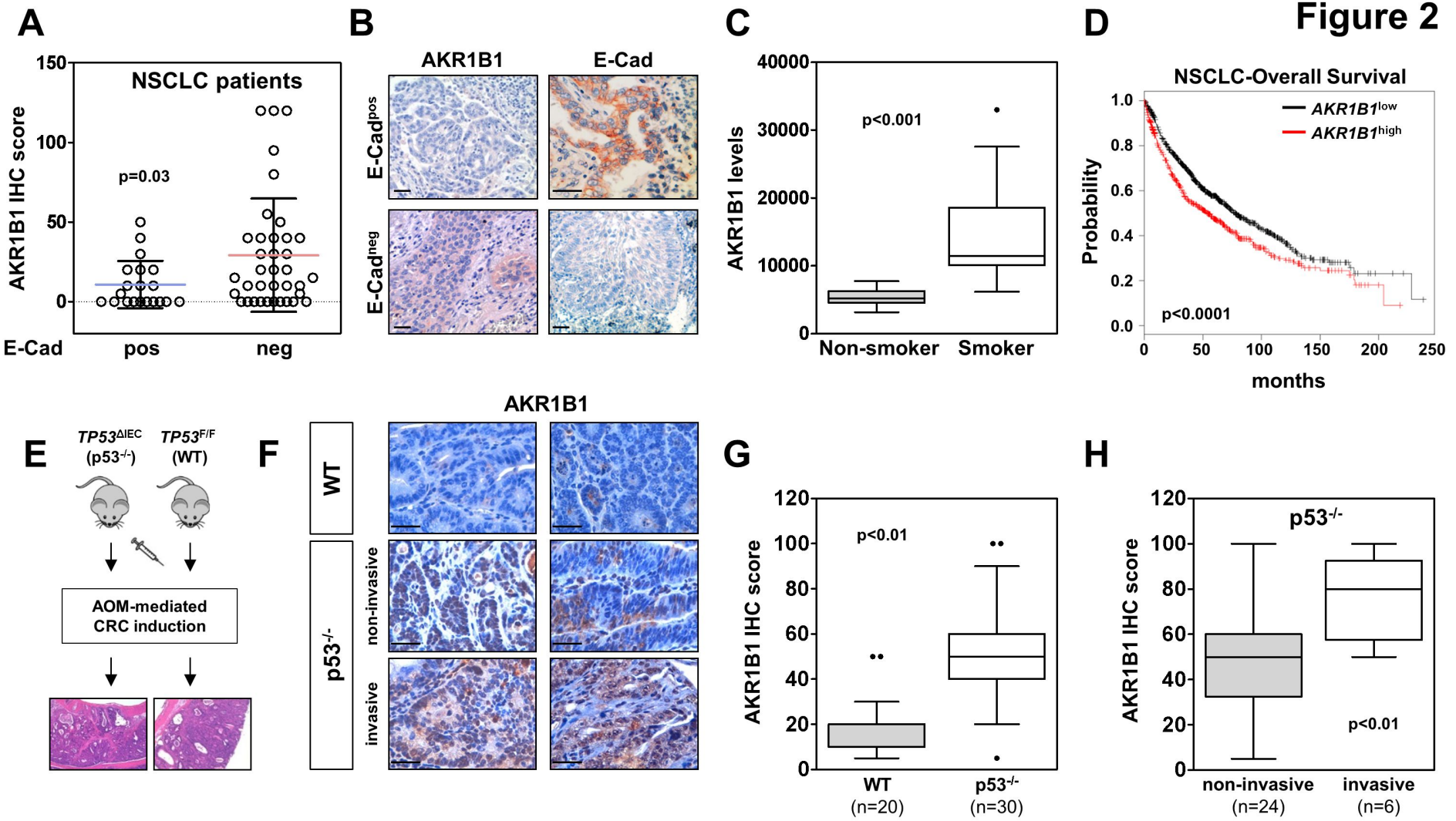


Figure 3

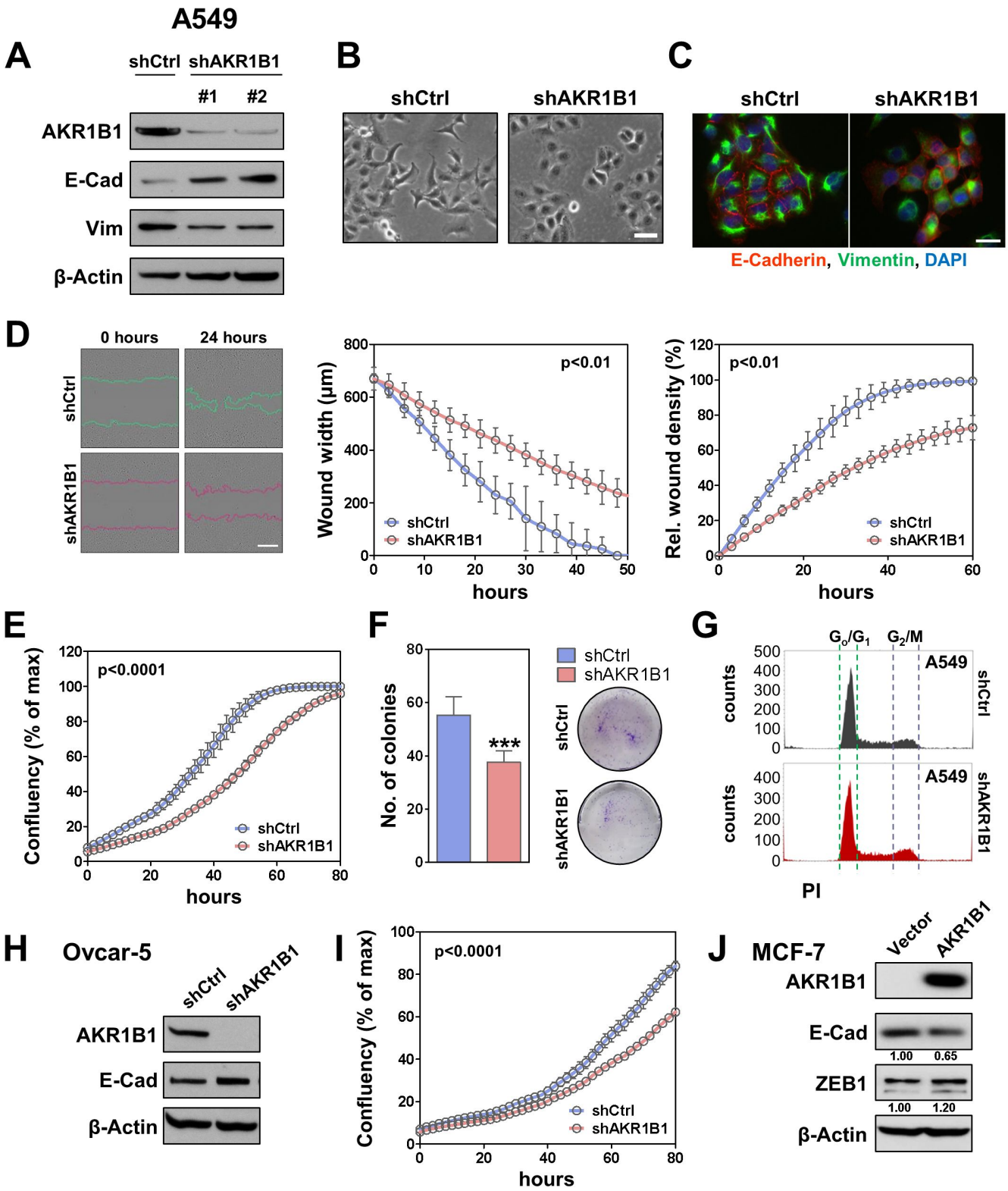


Figure 4

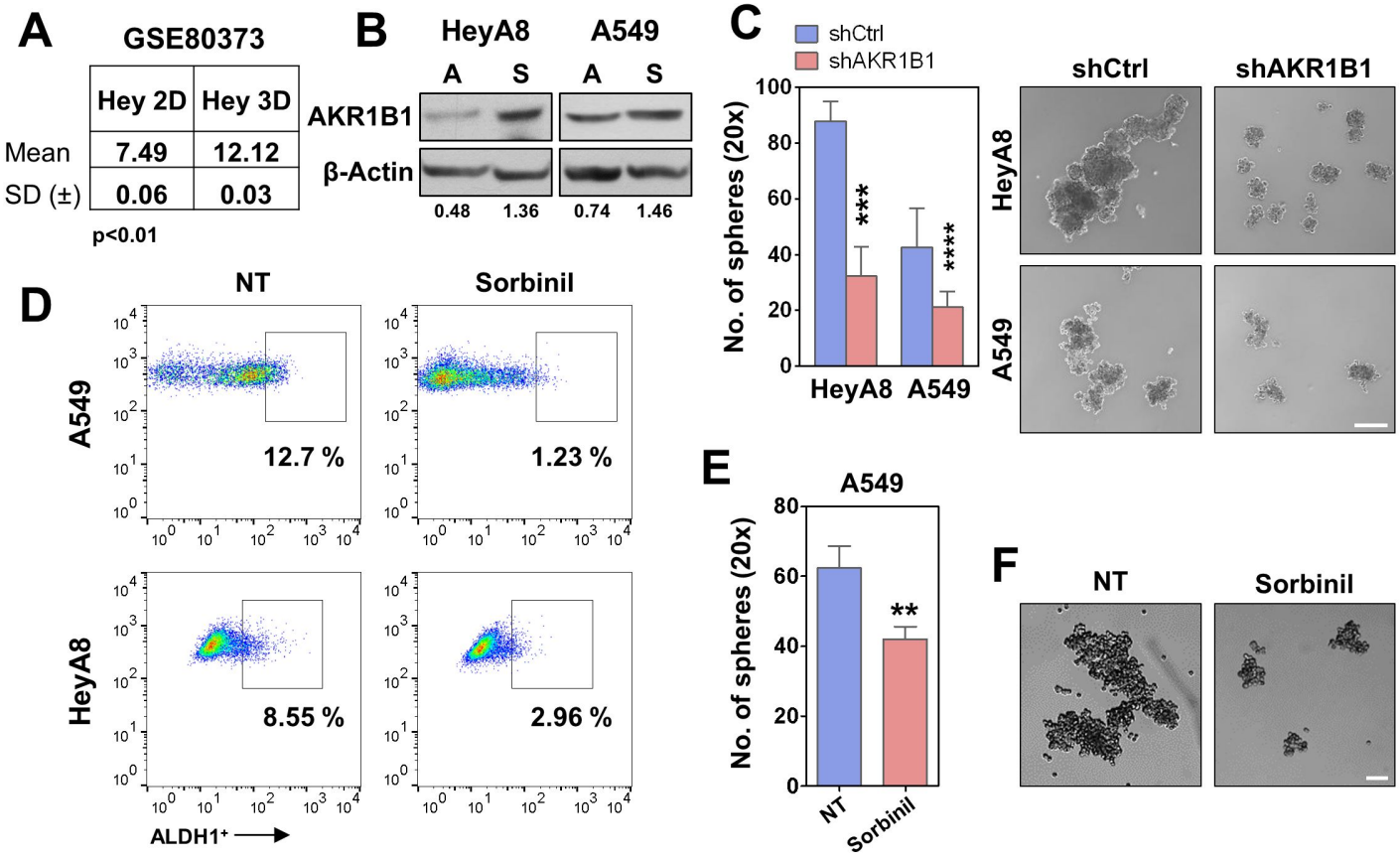


Figure 5

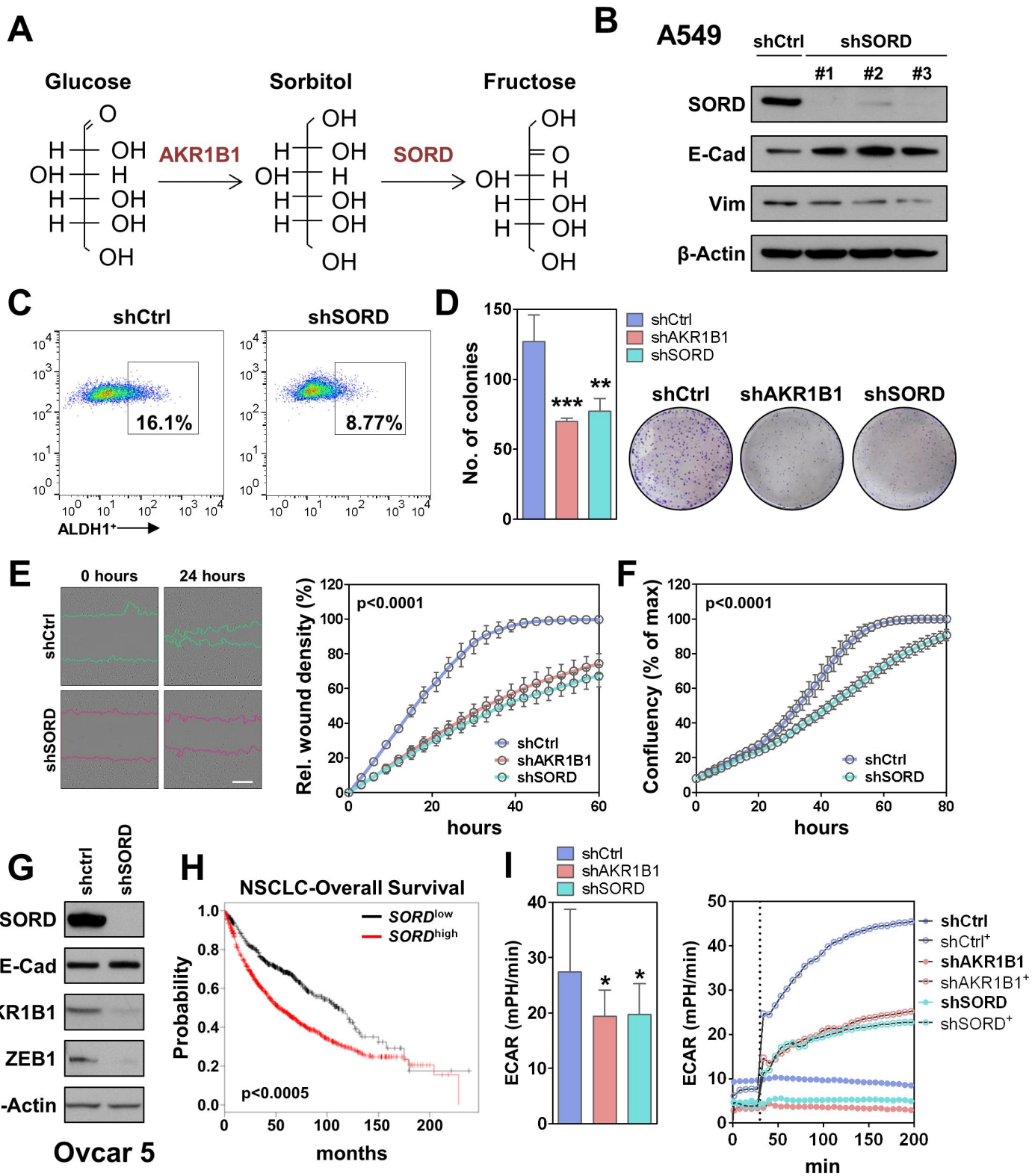
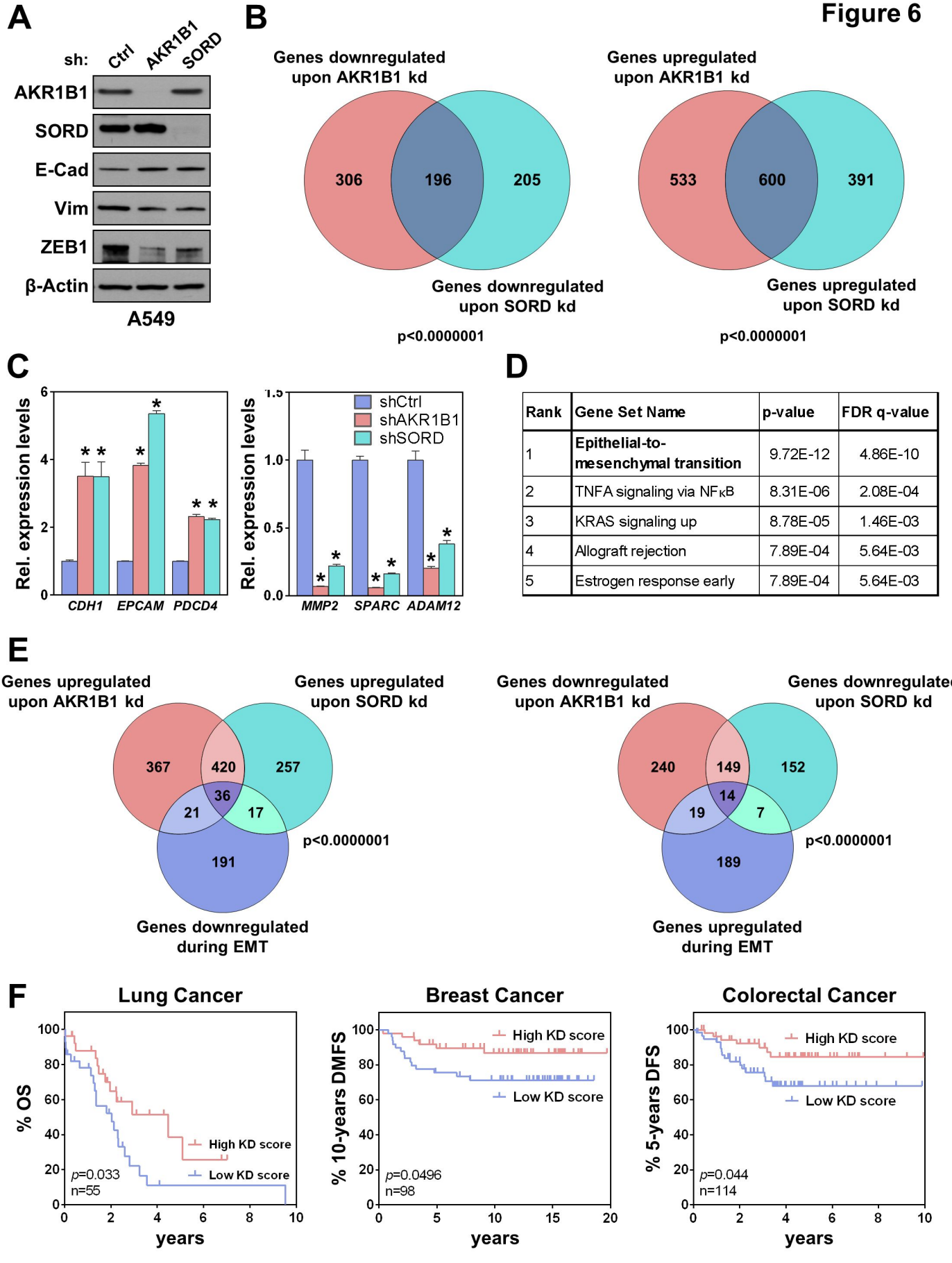
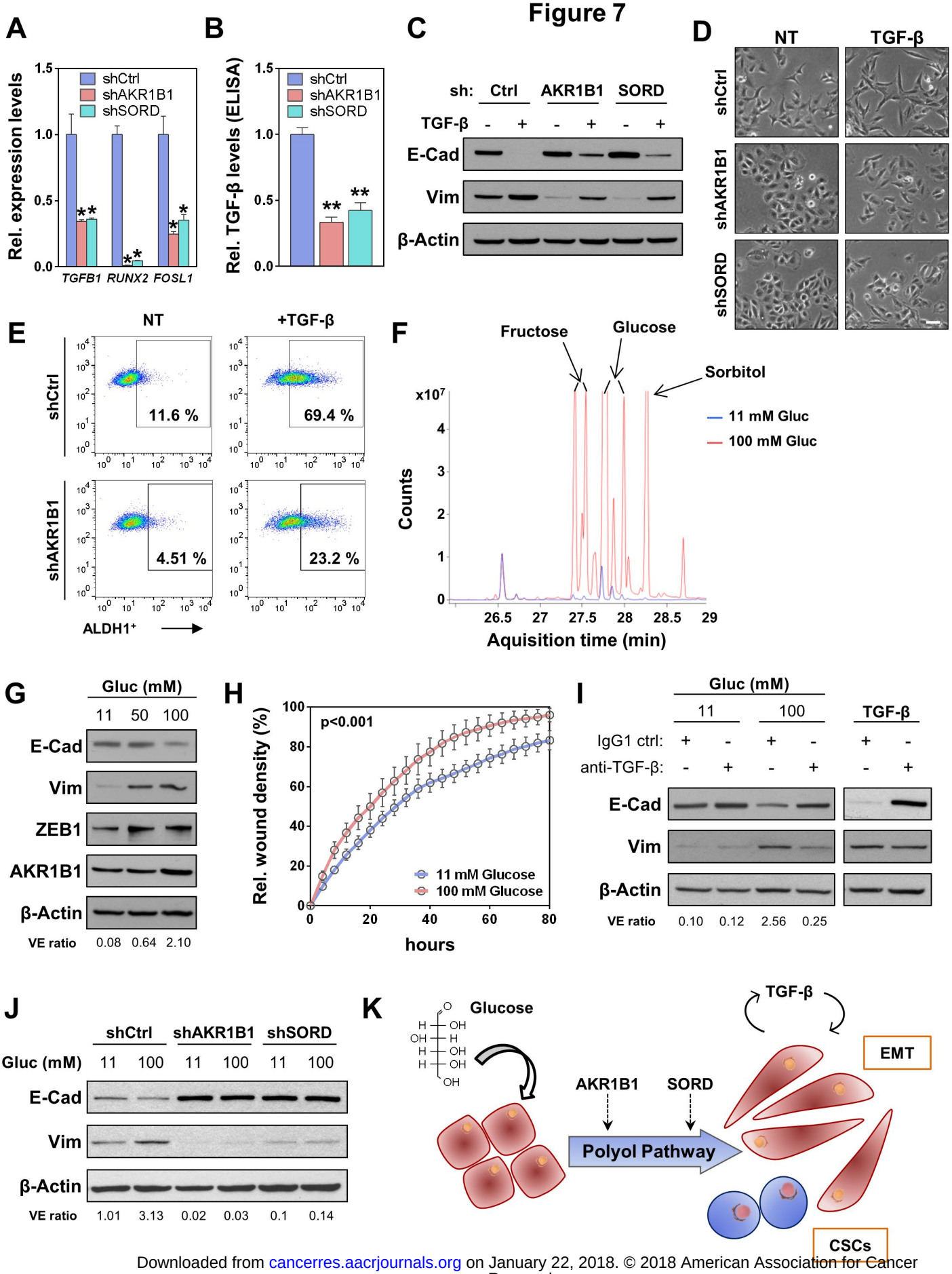


Figure 6



Rank	Gene Set Name	p-value	FDR q-value
1	**Epithelial-to-mesenchymal transition**	9.72E-12	4.86E-10
2	TNFA signaling via NFκB	8.31E-06	2.08E-04
3	KRAS signaling up	8.78E-05	1.46E-03
4	Allograft rejection	7.89E-04	5.64E-03
5	Estrogen response early	7.89E-04	5.64E-03

Figure 7



AACR American Association for Cancer Research

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Polyol pathway links glucose metabolism to the aggressiveness of cancer cells

Annemarie Schwab, Aarif Siddiqui, Maria Eleni Vazakidou, et al.

Cancer Res Published OnlineFirst January 17, 2018.