Supplemental methods, figures and tables

Generation of Prostate Tumor-initiating Cells is Associated with Elevation of Reactive Oxygen Species and IL6/STAT3 Signaling

Yi Qu[1], Anne Margrete Oyan[1,2], Runhui Liu[3], Yaping Hua[1,3], Jigang Zhang[3], Randi Hovland[4], Mihaela Popa[5], Xiaojun Liu[1,3], Karl A. Brokstad[6], Ronald Simon[7], Anders Molven[1,8], Biaoyang Lin[9,10,11], Wei-dong Zhang[3], Emmet McCormack[12] Karl-Henning Kalland[1,2#], Xi-Song Ke[1#]

Supplemental methods

Mass spectrometric analysis. Tryptic peptide mixtures were separated by the Ettan MDLC nanoflow/capillary LC system (GE Healthcare, Pittsburgh, PA USA) equipped with a trapping column (Dionex/LC Packings µ-Precolumn Cartridge P/N 160454 C18 PepMap 100, 5 µm, 100 Å, 300 µm i. d. x 5 mm, Sunnyvale, CA USA) and a nanocolumn (Dionex/LC Packings P/N 160321 150 x 0.075 mm i. d., C18 PepMap, 3 µm, 100 Å, Sunnyvale, CA USA), and then analyzed using LTQ-Orbitrap (Thermo Finnigan, Bremen, Germany) with a nanospray configuration. The precursor ion scan MS spectra (m/z 300-1600) were acquired in the Orbitrap with the resolution R = 60,000 at m/z 400 with the number of accumulated ions being 1 $X10^6$. The five most intense ions were isolated and fragmented in linear ion trap (number of accumulated ions: 3 X 10^4). The resulting fragment ions were recorded with the resolution R = 15,000 at m/z 400.

The extract_msn of the BioWorks program V3.2 (Thermo Electron, Inc., Waltham, MA USA) was used to generate the MS peak list with the default parameters. The ICIS peak-detection algorithm was used. The SEQUEST algorithm (Thermo Fisher Inc.) For the SEQUEST database search, the spectra were searched against the IPI.HUMAN.v3.58.fasta protein database (with 79,794 entries) (http://www.ebi.ac.uk/IPI/IPIhuman.html) using the BioWorks program V3.2 (Thermo Electron, Inc., Waltham, MA USA). In the TurboSEQUEST search

parameter setting, the threshold for Dta generation was 10,000, and precursor mass tolerance for Dta generation was 1.4. For the SEQUEST search, peptide tolerance was set at 3 daltons and fragment ions tolerance was set at 0.01 dalton. PeptideProphet™ [2] was used to assess the MS/MS spectra quality and a threshold score for accepting individual MS/MS spectra is set at P value of 0.9, which corresponds to a 0.5% error rate in our dataset. One missed tryptic cleavage was permitted.

Carboxyamidomethyl cysteine (Cys_CAM) (+ 57) was included as a fixed modification for iodoacetamide reduction and alkylation. As the proteins were prepared by polyacrylamide gel electrophoresis, the cysteines might react with free acrylamide monomers to form propionamide cysteine (Cys_PAM). We included an optional 14 daltons in the search to account for potential propionamide cysteine (the mass difference between Cys-PAM and Cys-CAM is 14). Methionine oxidation (+15.999 Daltons) was chosen as another optional modification for the database search. Proteins with ProteinProphet P value greater than 0.9 and with more than two unique peptide hits were considered as true hits. A randomized database of the IPI.HUMAN.v3.58.fasta was used as a decoy database to calculate the false discovery rate of protein identification. The perl script used for randomization was from www.matrixscience.com/downloads/decoy.pl.gz. The false discovery rate (FDR, 0.568%) was calculated by the ratio of the number of matches to the randomized database to that to the IPI.HUMAN.v3.58.fasta database.

DNA microsatellite fingerprinting and DNA copy number analysis. DNA microsatellite fingerprinting was performed following the protocols described previously (10). For DNA copy number analysis Affymetrix Genome-Wide human SNP arrays 6.0 were used to examine acquired genomic copy number changes and loss of heterozygosity (LOH) of all the cell lines. Genomic DNA was purified using the Tissue DNA kit (Cat.# D3396-02, EZNA, OMEGA Biotek). DNA pre-handling and array hybridization were performed according to the

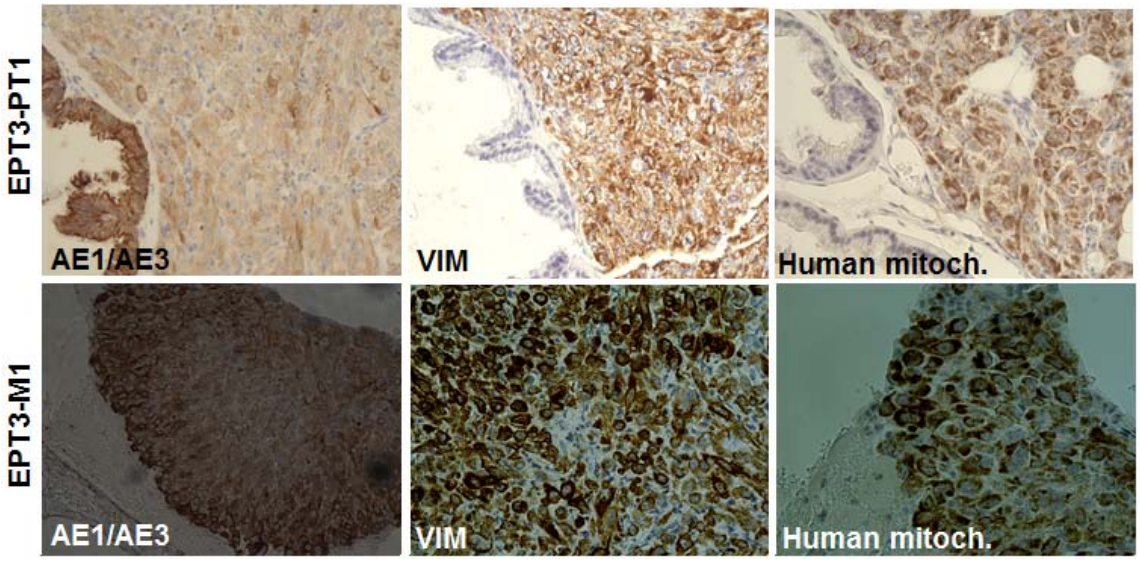
manufacturer's instructions (P/N 702504, Rev3, Affymetrix, Santa Clara, CA, USA) and arrays were scanned in an Affymetrix GeneChip Scanner 3000. Quality control, genotype calling, probe level normalization and copy number normalization to produce log2 ratios were done in Affymetrix GeneChip® Genotyping Console v3.0.1. An in-house reference file generated from 59 healthy blood donors was used. Data analysis and visualization was performed in Chromosome analysis suite v 1.2.2 with a threshold of minimum 50kb and 25 markers, 1Mb LOH. Aberrations are reported according to ICSN nomenclature and build GRCh37 (Hg19). Variants were classified as germline benign variants if present in all cell lines and in the database of normal variants (http://dgvbeta.tcag.ca/dgv/app/home) and these are removed from the Supplementary tables 1and 2.

Legends for Supplemental figures

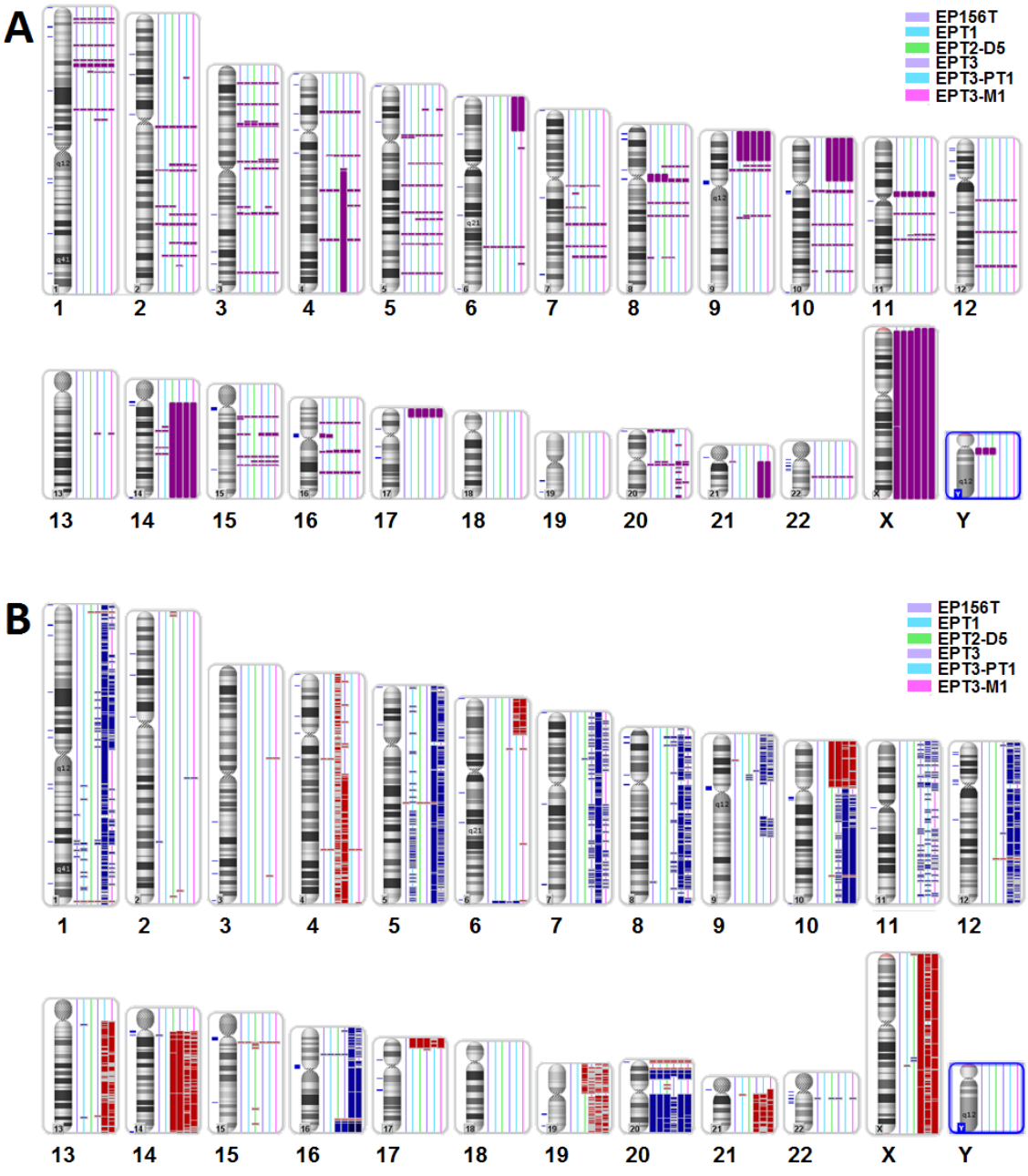
Supplemental figure 1 Immunohistochemical markers of EPT3 xenograft tumors and metastasis. The panels show histological sections with immunohistochemical staining of human prostate cells grown in mice. EPT3-PT1 and EPT3-M1 are primary tumor grown in the mouse prostate and diaphragmatic metastasis, respectively. AE1/AE3 pre-diluted antibody (ab961) detected both high and low molecular weight epithelial cytokeratins. Use of pre-diluted antibody against vimentin (VIM, ab8545) revealed strong staining throughout the tumor. Dual expression of epithelial and mesenchymal markers was detected in all tumors. The human cell specific anti-mitochondrion antibody (ab92824, diluted 1:1000) shows human EPT3-PT1 tumor cells with adjacent non-staining mouse ducts (middle row, rightmost panel) and metastatic EPT3-M1 tumor cells with adjacent non-staining mouse tissue (lower row, rightmost panel). The x40 microscopic objective was used to capture all images.

Supplemental figure 2 Genetic analysis of the cell lines in the prostate carcinogenesis model**.** The ideogram of the human genome labelled with the chromosomal regions with long continuous stretches of homozygosity larger than 1 Mb shown in (**A**), and copy number alterations (gains indicated blue and losses indicated red) larger than 50 kb shown in (**B**) of all the cell lines in the stepwise carcinogenesis model: From left to right: EPT156T, EPT1, EPT2-D5, EPT3, EPT3-PT1 and EPT3-M1. The exact aberrant genomic positions are shown in Supplemental data 4 and 5.

Supplemental figure 1



Supplemental figure 2



Supplemental tables

Supplemental table 1 DNA microsatellite profiling of EP156T, EPT3, EPT3-PT1 and EPT3-M1 cell lines. DNA microsatellite analyses using the markers displayed in the leftmost column showed near identical DNA patterns and validated common genetic origin of EP156T (passage 81), subcutaneous tumor derived EPT3 cells (passage 11), primary tumor derived EPT3-PT1 cells (passage 7) and metastasis derived EPT3-M1 cell lines (passage 5). Using the same method, the common genetic origin of EP156T, EPT1 and EPT2-D5 cells were published previously (1).

Marker	Chromosome	Position[#]	EP156T	EPT3	EPT3-PT1	EPT3-M1
Amelogenin	Xp22.3	9.1 Mb	104	104	104	104
	Yp11.2	4.3 Mb	110	110	-	-
D8S1179	8q24.1	121.2 Mb	142	142	142	142
			146	146	146	146
D21S11	21q21	20.5 Mb	206	206	206	206
			210	210	-	-
D18S51	18q21.3	57.6 Mb	305*	305*	305*	305*
			305	305	305	305
D3S1358	3p21	45.6 Mb	125	125	125	125
			129	129	129	129
vWA	12p13	5.9 Mb	184	184	184	184
			184	184	184	184
FGA	4q31	151.2 Mb	226	226	226	226
			230	230	230	230
D5S818	5q23	118.3 Mb	154	154	154	154
			158	158	158	158
D13S317	13q22-31	63.4 Mb	211	211	211	211
			223	223	223	223
D7S820	7q21	78.4 Mb	270	270	270	270
			274	274	274	274

[#] Distance from p-telomer, HuREF Map
*In addition a very small peak of 301 bp

Supplemental table 2 Secreted proteins detected in the culture supernatant of D5HS cells by mass spectrometry analysis. The mass spectrometry data including the probability of protein identification calculated by the ProteinProphet, and the number of unique peptides (num unique peps) and the total independent spectra counts (tol indep spectra) are shown. We only included proteins that were identified with at least two or more unique peptides by the mass spectrometry analysis. * means that the mRNA levels of the protein coding genes showed significant up-regulation in D5HS cells comparing to EPT2-D5 cells based supplemental data 1. # means growth factor or cytokine.

Gene symbol	Description	Protein probability	Num unique peps	Tot indep spectra
C3*	Complement C3 (Fragment)	1	65	172
COL12A1	Isoform 4 of Collagen alpha-1(XII) chain	1	49	79
QSOX1	Isoform 1 of Sulfhydryl oxidase 1	1	44	133
MMP2*	72 kDa type IV collagenase	1	41	104
COL6A1	Collagen alpha-1(VI) chain	1	39	220
FBN1	Fibrillin-1	1	36	55
ALB	Isoform 1 of Serum albumin	1	31	127
TGFBI	Transforming growth factor-beta-induced protein ig-h3	1	27	84
PXDN	Isoform 1 of Peroxidasin homolog	1	22	60
TIMP1	Metalloproteinase inhibitor 1	1	19	59
SERPINE1	Plasminogen activator inhibitor 1	1	17	26
COL7A1	Isoform 1 of Collagen alpha-1(VII) chain	1	17	24
LGALS3BP*	Galectin-3-binding protein	1	16	59
LAMC1	Laminin subunit gamma-1	1	16	37
SPARC	SPARC	1	15	43
ENPP2*	Isoform 1 of Ectonucleotide pyrophosphatase/phosphodiesterase family member 2	1	15	31
COL6A2*	Isoform 2C2 of Collagen alpha-2(VI) chain	1	14	54
SERPINE2*	Glia-derived nexin	1	14	35
PLAU	Urokinase-type plasminogen activator	0,9999	13	36
CLU	Clusterin	1	13	33
IGFBP7*	Insulin-like growth factor-binding protein 7	1	13	32
ECM1*	Isoform 1 of Extracellular matrix protein 1	1	13	29
GAS6	Isoform 2 of Growth arrest-specific protein 6	1	12	22
CALR	Calreticulin	1	11	26
SRGN	Serglycin	1	9	19
ADAMTS1*	A disintegrin and metalloproteinase with	1	9	18

	thrombospondin motifs 1			
NID1	Isoform 1 of Nidogen-1	1	9	16
IL6*#	Interleukin-6	1	8	27
GGH	Gamma-glutamyl hydrolase	1	8	15
LTF*	Growth-inhibiting protein 12	1	8	14
CLEC11A#	C-type lectin domain family 11 member A	1	8	13
AGRN	Agrin	1	8	10
MASP1	Isoform 2 of Mannan-binding lectin serine protease 1	1	7	23
LUM*	Lumican	1	7	20
DKK3*	cDNA FLJ52545, highly similar to Dickkopf-related protein 3	1	7	18
RNPEP	Aminopeptidase B	1	7	15
LAMB1	Laminin subunit beta-1	1	7	13
CHI3L1*	Chitinase-3-like protein 1	1	7	8
TNFRSF11B	Tumor necrosis factor receptor superfamily member 11B	1	6	25
GPI#	Glucose-6-phosphate isomerase	0,9987	6	18
STC2*	Stanniocalcin-2	1	6	16
DCN*	Isoform A of Decorin	1	6	13
TFRC	Transferrin receptor protein 1	1	6	8
B2M	Beta-2-microglobulin	1	6	7
MMP1	Interstitial collagenase	1	6	7
APOA1BP	cDNA FLJ56357, highly similar to Homo sapiens apolipoprotein A-I binding protein (APOA1BP), mRNA	1	5	15
COL5A1	Collagen type V alpha 1	1	5	13
ABP1*	Isoform 1 of Amiloride-sensitive amine oxidase [copper-containing]	1	5	11
CALU	cDNA FLJ31776 fis, clone NT2RI2008141, highly similar to CALUMENIN	1	5	10
LOX	Protein-lysine 6-oxidase	1	5	10
CPA4	Carboxypeptidase A4	1	5	7
LAMA2*	laminin alpha 2 subunit isoform b precursor	1	5	6
FN1*	fibronectin 1 isoform 2 preproprotein	0,9999	4	13
GPC1	Glypican-1	1	4	10
FSTL1*	Follistatin-related protein 1	1	4	8
FBLN1	Isoform B of Fibulin-1	0,9999	4	6
PCSK9*	Isoform 1 of Proprotein convertase subtilisin/kexin type 9	1	4	5
TNC*	Isoform 1 of Tenascin	1	4	5
HSPG2	Basement membrane-specific heparan sulfate proteoglycan core protein	1	4	4
IGFBP5	Insulin-like growth factor-binding protein 5	1	4	4
PRDX4	Peroxiredoxin-4	1	4	4
S100A9	Protein S100-A9	1	3	12
LTBP1	Latent-transforming growth factor beta-binding protein, isoform 1L	1	3	8
GRN*#	Isoform 1 of Granulins	1	3	6
SEMA3C	cDNA FLJ55486, highly similar to	1	3	6

	Semaphorin-3C			
COL3A1*	Isoform 1 of Collagen alpha-1(III) chain	1	3	5
DKK1	Dickkopf-related protein 1	1	3	5
MIF#	Macrophage migration inhibitory factor	1	3	5
SERPINB2*	Plasminogen activator inhibitor 2	1	3	5
ADAM9	Isoform 1 of ADAM 9	1	3	4
CST3	Cystatin-C	1	3	4
SPON2	Spondin-2	1	3	4
TF	Serotransferrin	1	3	4
GSN*	Isoform 1 of Gelsolin	1	3	3
TIMP2	Metalloproteinase inhibitor 2	1	3	3
S100A8#	Protein S100-A8	1	2	10
IGHG1	Putative uncharacterized protein DKFZp686N02209	0,9999	2	4
LAMC2*	Isoform Long of Laminin subunit gamma-2	1	2	4
PDGFD*#	Isoform 2 of Platelet-derived growth factor D	1	2	4
VEGFC#	Vascular endothelial growth factor C	1	2	4
CSF1#	Isoform 1 of Macrophage colony-stimulating factor 1	1	2	3
FST*	Isoform 1 of Follistatin	1	2	3
LAMB2*	Laminin subunit beta-2	1	2	3
LGALS1	Galectin-1	1	2	3
VCAN	Isoform V0 of Versican core protein	0,9999	2	3
ATRN*	Isoform 1 of Attractin	1	2	2
CTGF	Isoform 1 of Connective tissue growth factor	1	2	2
CXCL1*#	Growth-regulated alpha protein	0,9994	2	2
IL6ST*	Isoform 1 of Interleukin-6 receptor subunit beta	1	2	2
IL8*#	Isoform 2 of Interleukin-8	1	2	2
ISG15	Interferon-induced 17 kDa protein	1	2	2
PLA2G15	1-O-acylceramide synthase	0,9852	2	2
PRSS1	Putative trypsin-6	0,9998	2	2
PTX3	Pentraxin-related protein PTX3	1	2	2
SFN	Isoform 1 of 14-3-3 protein sigma	0,9975	2	2
STC1	Stanniocalcin-1	1	2	2
VASN*	Vasorin	1	2	2

References

1. Ke XS, Li WC, Hovland R, Qu Y, Liu RH, McCormack E, et al. Reprogramming of cell junction modules during stepwise epithelial to mesenchymal transition and accumulation of malignant features in vitro in a prostate cell model. Exp Cell Res. 2011;317(2):234-47.