Electronic Supplementary Information

Identification and localization of Xylose-binding proteins and as potential biomarkers for liver fibrosis/cirrhosis

Yaogang Zhong, Xiu-Xuan Sun, Peixin Zhang, Xinmin Qin, Wentian Chen, Yonghong Guo, Zhansheng Jia, Huijie Bian*, Zheng Li*

CONTENTS:

Materials and Methods

Protein extraction.

The total protein of HSCs was extracted with T-PER Reagent (Pierce Biotechnology Inc., USA) according to the manufacturer's instructions. The procedures were briefly described as before.[S1] The supernatant was immediately transferred for use or stored at -80°C. Protein concentration in cell lysates was determined by Bradford assay kit (Pierce) (Fig. S1).

Isolation of XBPs.

The XMPCs were prepared according to the protocol[S2-S4] with some modifications. The XMPCs were blocked with 400 μL blocking buffer (including 2% ethanolamine and 0.1% BSA) at 37°C for 1 h and then washed three times with the binding buffer (20 mmol·L^{-1} Tris-HCl, 0.5 mol·L^{-1} NaCl, 10 mmol·L^{-1} $CaCl_2$, 10 mmol·L^{-1} $MnCl_2$, 6 mmol·L^{-1} $MnCl_2$, pH7.2), 2 mg protein, and 20 μL 100 mmol·L^{-1} PMSF were added, then incubated with the binding buffer for 2 h under gentle shaking. Unbound proteins were removed from the conjugates by washing five or six times with the washing buffer (20 mmol·L^{-1} Tris-HCl, 0.05% Tween-20, pH7.2) until proteins did not exist in supernatant. The XBPs were eluted with 400 μL eluting fluid (8 mol·L^{-1} Urea, 0.1 mol·L^{-1} NH_4HCO_3) for 1 h, then added 10 μL PMSF and stored at -80°C. The purified XBPs were quantitated by the Bradford Assay (Fig. S1).

Identification by LC-OrbitrapMS/MS.

The purified XBPs were treated with reduction, alkylation and trypsin digestion according to the protocol.[S2] The peptide sequencing was identified by LC-Orbitrap MS/MS,[S5] LC-MS system consist of LTQ-Orbitrap XL ETD mass spectrometers (ThermoFisher) equipped with Easy-nLC II System. Five microliters of trypsin-digested solutions were loaded by an auto-sampler on a Easy-nLC II system with an pre-column (2 cm×100 μm, 5 μm C18-A1) and analytical column (10 cm×75 μm, 3 μm C18-A2). Peptides were eluted from the capillary column at a flow rate of 300 nL·min^{-1} to the MS through a nano-spray emitter tip. Spray voltage was set to 1.95 kV and capillary temperate was 200°C. The mobile phases consisted of A (water with 0.1% formic acid) and B (acetonitrile with 0.1% formic acid). The three-step linear gradient of 3-9% B in 10 min, 9-41% in 70 min, 41-81% in 15 min, and 81% B for 25 min was used throughout this study. MS scans were performed to select 10 intense peaks for subsequently MS/MS scans.

Proteome Discoverer 1.2 software (Thermo Fisher) with SEQUEST search managed the LC pump and the automatic spectral recording. MS/MS ion search was performed in International Protein Index (IPI.HUMAN. v3.73. fasta) databases. A mass tolerance of ±10ppm for the precursor ions and a tolerance of ± 1 Da for fragment ions were used. One missed cleavage was allowed. Oxidation (M) was set as a variable modification and carboxymethyl (C) was set as a fixed modification. The identified proteins were filtered using high confidence (FDR=0.01) on the peptide level. The obtained data were also used to calculate the peak area. The peak area values of proteins in one sample were compared to those in another sample, and expression levels of proteins were evaluated according to the peak area correlation between the two samples.

Distribution and ontology analysis of XBPs.

The accession number list of the proteins identified also underwent functional analysis by

Database for Annotation, Visualization, and Integrated Discovery (DAVID)[S6] online tool (http://david.abcc.ncifcrf.gov/). Functional annotation clustering of our accession numbers was performed on the basis of GO, Genomes pathway database (KEGG), selecting uniprot_accession as identifier and peptide list as list type.

Network and functional analysis.

The list of accession number of the proteins identified was processed by pathway analysis using KEGG pathway (http://www.kegg.jp/) and String (http://string-db.org/) tool.

Characteristics of XBPs.

The characteristics of the isolated XBPs were analyzed according to their consensus sequence with Motif-x software (v1.2 10.05.06). All complete and unbiased motif searches were performed in a set of the isolated XBPs using each of the 20 amino acids as a central residue with a 13-amino acid spread; the occurrence is set at 5, the significance is set at 10^{-5}, and the IPI human proteome is set as background. The motif scores were calculated by taking the sum of the negative log probabilities used to fix each position of the motif. The extracted non-redundant consensus sequences were applied to WebLogo to create relative frequency plots.[S7, S8]

Reference

S1. Y. Qin, Y. Zhong, G. Yang, T. Ma, L. Jia, C. Huang and Z. Li, *Molecules*, 2014, **19**, 19845-19867.

S2. G. Yang, W. Chu, H. Zhang, X. Sun, T. Cai, L. Dang, Q. Wang, H. Yu, Y. Zhong, Z. Chen, F. Yang and Z. Li, *Proteomics*. 2013, **13**, 878-892.

S3. X. Sun, G. Yang, S. Sun, R. Quan, W. Dai, B. Li, C. Chen and Z. Li, *Curr Pharm Biotechnol*. 2009, **10**, 753-760.

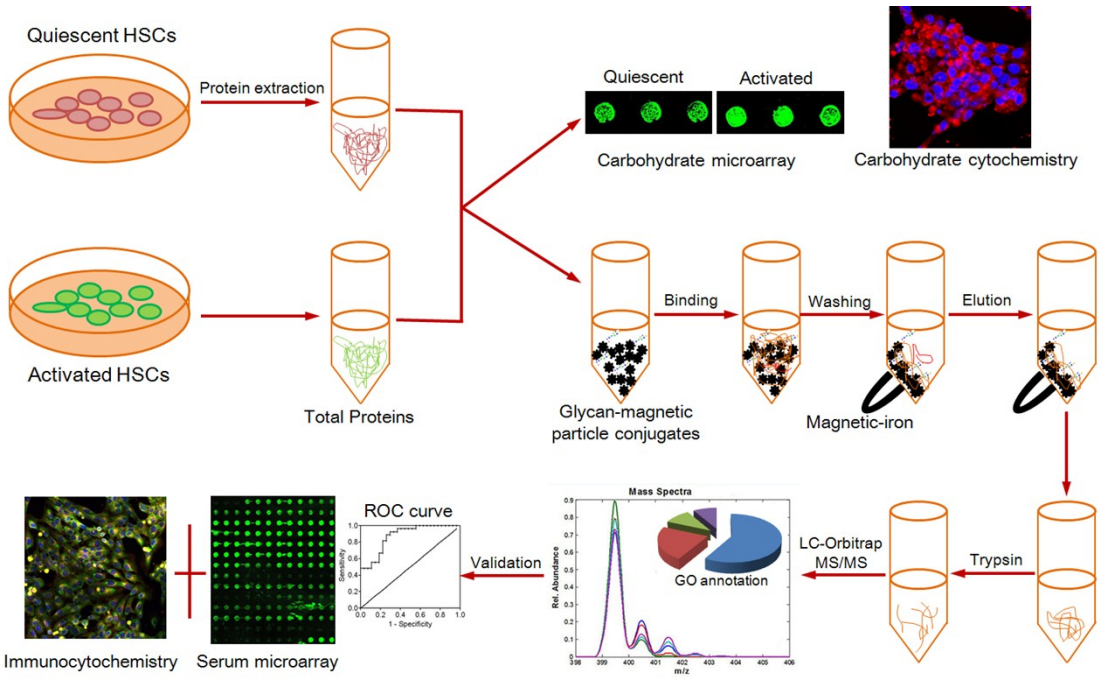
S4. G. Nan, H. Yan, G. Yang, Q. Jian, C. Chen and Z. Li, *Curr Pharm Biotechnol*. 2009, **10**, 138-146.

S5. J.C. Silva, M.V. Gorenstein, G. Li, J. Vissers, S. Geromanos, *Mol Cell Proteomics*. 2006, **5**, 144-156.

S6. D. Huang, B. Sherman and R. Lempicki, Nature Protoc, 2009, **4**, 44-57.

S7. G.E. Crooks, G. Hon, J.M. Chandonia and S.E. Brenner, *Genome Research*, 2004, **14**, 1188-1190.

S8. T.D. Schneider and R.M. Stephens, *Nucleic Acids Res*. 1990, **18**, 6097-6100.



Scheme 1. Schematic flow diagram of the integrated strategy.

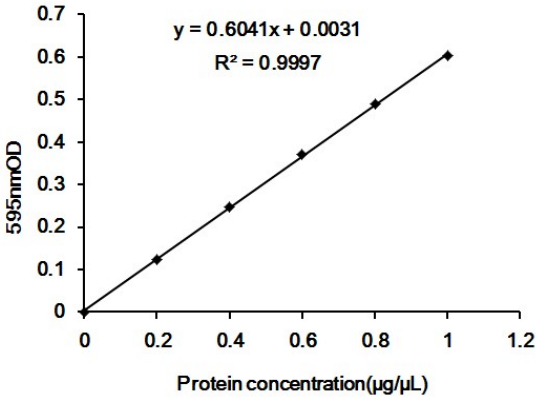


Fig. S1 Cellular protein and sera concentration. The protein concentrations of HSCs and sera samples were measured using Bradford assay.

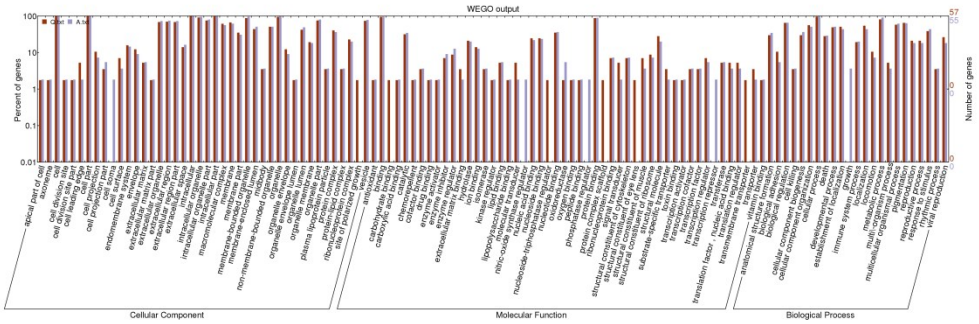


Fig. S2 GO classification of the identified XBPs using WEGO.

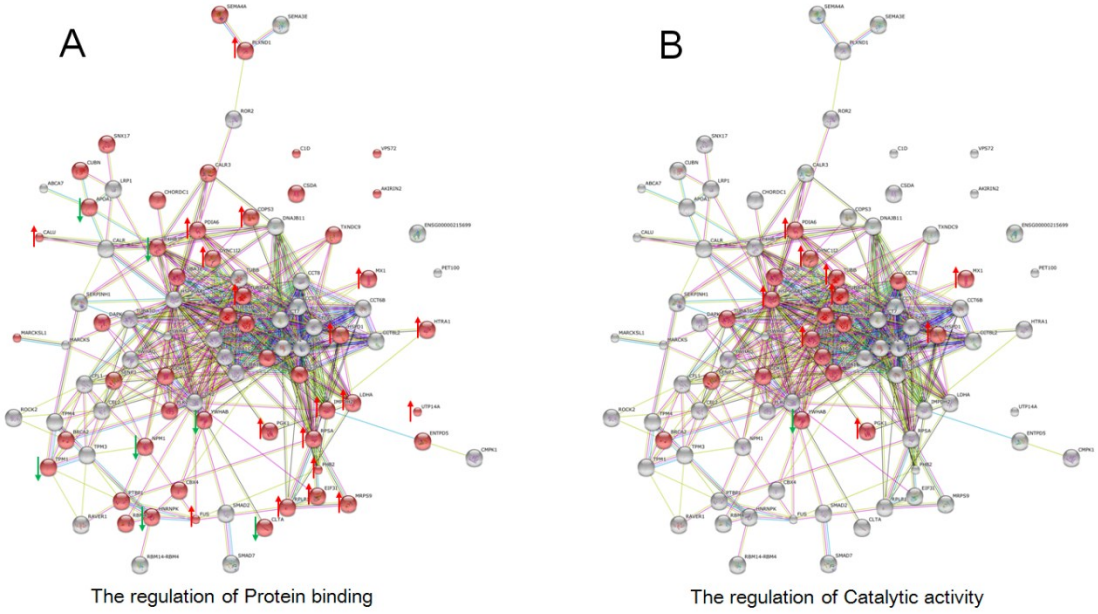


Fig. S3 Protein interaction network analysis revealed XBPs (red sphere) responsible for protein binding (A) and catalytic activity (B) that were up-regulated (red arrow) or down-regulated (green arrow) in the activated LX-2 through enrichment analysis of molecular function.

Table S1 Characteristics of liver cirrhosis patient group

Hepatitis/ Cirrhosis	n	Males, (n) %	Age, years	HBsAg (IU /mL)	HbsAb (PEI U/mL)	HbeAg (PEI U/mL)	AST (IU/L)	ALT (IU/L)	GGT (IU/L)	Albumin (g/L)	Total bilirubin (μmol/L)	AFP (ng/mL)
	60	40 (66.0)	60.0 (55.5-63.5)	298.5 ±132. 8	17.15 ±8.15	42.88 ±15.66	87.21 ±69.06	69.82 ±51.91	142	36.80 ±5.91	34.88 ±43.45	58.9 ±193.4

AST, aspartate aminotransferase; ALT, alanine aminotransferase; GGT, glutamyltranspeptidase; AFP, α-fetoprotein.

Table S2. Detailed information of XBPs significantly differential expressed in the activated LX-2

IPI	Protein Acc. No.	Protein name	No. of unique peptides	M.W. (kDa)	Gene symbol	Mass	Score	HSCs		(A/Q)	Subcellular Location
								Q	A	Fold change (*p*-value)	
Up-regulated proteins											
IPI00953530	H9KVD3_HUMAN	Interferon-induced GTP-binding protein Mx1	2	15.1	MX1	1126.6	4.39	Q	A	1.95(0.032)	Cytoplasm; Endoplasmic reticulum membrane; Peripheral membrane protein; perinuclear region;
IPI00941900	CALU_HUMAN	Calumenin	3	37.1	CALU	825.91	3.64	Q	A	1.53(0.050)	Endoplasmic reticulum lumen; Secreted; Melanosome; Sarcoplasmic reticulum lumen[s]
IPI00940505	UT14A_HUMAN	U3 small nucleolar RNA-associated protein 14 homolog A	2	69.2	UTP14A	954.46	11.75	Q	A	1.63(0.009)	Nucleus & nucleolus;
IPI00917575	CH60_HUMAN	60 kDa heat shock protein	2	55.0	HSPD1	752.88	6.66	Q	A	1.56(0.013)	Mitochondrion matrix
IPI00916278	DC1I2_HUMAN	Cytoplasmic dynein 1 intermediate chain 2	2	12.3	DYNC1I2	975.93	3.67	Q	A	3.47(<0.001)	Cytoplasm & cytoskeleton[s]
IPI00910487	SERPH_HUMAN	Serpin H1	2	44.2	SERPINH1	830.40	6.64	Q	A	4.21(<0.001)	Endoplasmic reticulum lumen;
IPI00909158	PGK1_HUMAN	Phosphoglycerate kinase 1	2	30.0	PGK1	1012.0	6.24	Q	A	2.51(0.007)	Cytoplasm;
IPI00908770	TBB4A_HUMAN	Tubulin beta chain	5	35.9	TUBB	1354.7	40.63	Q	A	1.52(0.050)	Cytoplasm & cytoskeleton;
IPI00873656	SEPT7_HUMAN	Septin-7	2	44.8	7-Sep	762.35	3.44	Q	A	4.15(<0.001)	Cytoplasm; Chromosome & centromere & kinetochore; cytoskeleton & spindle; Midbody;
IPI00798361	RSSA_HUMAN	40S ribosomal protein SA	2	16.9	RPSA	999.50	3.87	Q	A	2.10(0.031)	Cell membrane; Cytoplasm; Nucleus[s]
IPI00789484	Q6P657_HUMAN	Plexin-D1	3	6.7	PLXND1	737.71	14.43	ND	A	3.57(<0.001)	Cell membrane;
IPI00644989	PDIA6_HUMAN	Protein disulfide-isomerase A6	4	48.1	PDIA6	764.43	6.09	Q	A	3.74(<0.001)	Endoplasmic reticulum lumen[s]; Cell membrane; Melanosome;

IPI00642205	EIF3I_HUMAN	Eukaryotic translation initiation factor 3 subunit I	2	14.1	EIF3I	658.80	3.16	Q	A	3.57(<0.001)	Cytoplasm[s]
IPI00514049	KCY_HUMAN	UMP-CMP kinase	4	19	CMPK1	760.88	3.16	Q	A	1.60(0.028)	Nucleus; Cytoplasm;
IPI00027252	PHB2_HUMAN	Prohibitin-2	2	33.3	PHB2	608.31	3.48	Q	A	2.31(0.043)	Mitochondrion inner membrane[s]; Cytoplasm[s]; Nucleus[s];
IPI00020599	CALR3_HUMAN	Calreticulin	2	48.1	CALR	1196.1	12.34	Q	A	1.62(0.011)	Endoplasmic reticulum lumen; Cytoplasm & cytosol; Secreted & extracellular space & extracellular matrix; Cell surface; Sarcoplasmic reticulum lumen[s];
IPI00012011	COF1_HUMAN	Cofilin-1	4	18.5	CFL1	597.61	22.67	Q	A	4.83(<0.001)	Nucleus matrix; Cytoplasm & cytoskeleton; Cell projection & ruffle membrane; Peripheral membrane protein; lamellipodium membrane;
IPI00008527	RLA1_HUMAN	60S acidic ribosomal protein P1	2	11.5	RPLP1	851.95	9.76	ND	A	—(<0.001)	Cytoplasm;
IPI00413344	COF2_HUMAN	Cofilin-2	2	18.7	CFL2	1098.6	18.23	ND	A	—(<0.001)	Nucleus matrix[s]; Cytoplasm & cytoskeleton[s];
IPI00025721	CSN3_HUMAN	COP9 signalosome complex subunit 3	3	47.8	COPS3	888.35	3.80	ND	A	—(<0.001)	Cytoplasm; Nucleus;
IPI00008454	DJB11_HUMAN	DnaJ homolog subfamily B member 11	2	40.5	DNAJB11	992.43	6.42	ND	A	—(<0.001)	Endoplasmic reticulum lumen;
IPI00902529	HTRA1_HUMAN	Serine protease HTRA1	3	48.4	HTRA1	744.83	3.50	ND	A	—(<0.001)	Secreted; Cytoplasm & cytosol;
IPI00641924	RT09_HUMAN	28S ribosomal protein S9	2	45.8	MRPS9	623.29	2.99	ND	A	—(<0.001)	Mitochondrion;
IPI00909890	FUS_HUMAN	RNA-binding protein FUS	4	44.8	FUS	751.66	9.70	ND	A	—(<0.001)	Nucleus;
IPI00219301	MARCS_HUMAN	Myristoylated alanine-rich C-kinase substrate	2	31.5	MARCKS	892.94	7.64	ND	A	—(<0.001)	Cytoplasm & cytoskeleton[p]; Membrane; Lipid-anchor[s];
IPI00795143	IMDH2_HUMAN	Inosine-5'-monophosphate dehydrogenase 2	3	10.4	IMPDH2	579.81	2.95	ND	A	—(<0.001)	Cytoplasm; Nucleus;

IPI00604607	Q2VPJ6_HUMAN	Heat shock protein HSP 90-alpha	4	63.2	HSP90A A1	576.28	6.23	ND	A	—(<0.001)	Cytoplasm; Melanosome; Cell membrane;
IPI00007752	TBB4A_HUMAN	Tubulin beta-4A chain	7	35.9	TUBB4A	911.96	32.17	Q	A	4.61(<0.001)	Cytoplasm & cytoskeleton;
IPI00788938	LDHA_HUMAN	L-lactate dehydrogenase A chain	2	25.2	LDHA	624.80	2.76	ND	A	—(<0.001)	Cytoplasm;
Down-regulated proteins											
IPI00853525	APOA1_HUMAN	Apolipoprotein A-I	7	27.9	APOA1	966.97	24.57	Q	A	0.25(<0.001)	Secreted
IPI00910322	TPM4_HUMAN	Tropomyosin alpha-4 chain	3	17.6	TPM4	1170.6	10.78	Q	A	0.59(0.015)	Cytoplasm & cytoskeleton;
IPI00909308	B7Z596_HUMAN	Tropomyosin alpha-1 chain	2	17.8	TPM1	730.87	6.85	Q	A	0.55(0.023)	Cytoplasm & cytoskeleton;
IPI00878551	PDIA1_HUMAN	Protein disulfide-isomerase	5	51.2	P4HB	890.92	14.23	Q	A	0.26(0.004)	Endoplasmic reticulum lumen; Melanosome; Cell membrane; Peripheral membrane protein[p]
IPI00759832	1433B_HUMAN	14-3-3 protein beta/alpha	4	27.8	YWHAB	1080.0	11.08	Q	A	0.43(0.048)	Cytoplasm; Melanosome;
IPI00658013	NPM_HUMAN	Nucleophosmin	6	28.4	NPM1	1114.1	9.17	Q	A	0.63(0.050)	Nucleus & nucleolus; Cytoplasm & cytoskeleton & microtubule organizing center & centrosome；
IPI00514561	HNRPK_HUMAN	Heterogeneous nuclear ribonucleoprotein K	5	47.5	HNRNPK	863.80	10.48	Q	ND	—(<0.001)	Cytoplasm; Nucleus & nucleoplasm; Cell projection & podosome;
IPI00297779	TCPB_HUMAN	T-complex protein 1 subunit beta	3	57.5	CCT2	1145.1	11.87	Q	A	0.14(<0.001)	Cytoplasm;
IPI00216393	CLCA_HUMAN	Clathrin light chain A	2	23.6	CLTA	1176.5	3.88	Q	A	0.23(0.006)	Cytoplasmic vesicle membrane; Peripheral membrane protein; Cytoplasmic side; Membrane & coated pit;
IPI00020557	LRP1_HUMAN	Prolow-density lipoprotein receptor-related protein 1	3	504.2	LRP1	989.67	6.01	Q	ND	—(<0.001)	Cell membrane; Membrane & coated pit;
IPI00018146	1433T_HUMAN	14-3-3 protein theta	2	27.7	YWHAQ	1073.0	10.93	Q	A	0.25(<0.001)	Cytoplasm;

IPI00000816	Q5CZQ1_DANRE	14-3-3 protein epsilon	2	29.2	YWHAE	1044.5	11.69	Q	A	0.45(0.011)	Cytoplasm[s]; Melanosome;
IPI00290566	TCPA_HUMAN	T-complex protein 1 subunit alpha	2	60.3	TCP1	1177.2	6.57	Q	A	0.21(<0.001)	Cytoplasm; cytoskeleton & microtubule organizing center & centrosome;
IPI00555698	YBOX3_HUMAN	Y-box-binding protein 3	7	35.9	YBX3	891.93	12.26	Q	A	0.46(0.023)	Cytoplasm; Nucleus;

XBPs with a fold change of >1.5 or <0.67 are shown. Q: The quiescent LX-2, A: The activated LX-2; "A/Q" represents the fold change of a protein in the activated LX-2 vs. that in the quiescent LX-2. ND, not detected in the samples , "—"significantly differential expressed in the sample.

Table S3 The detail information of the main KEGG pathway

NO.	Name	Protein NO.(%)
ko01100	Metabolic pathways	9 (20.5%)
ko05169	Epstein-Barr virus infection	7 (15.9%)
ko04141	Protein processing in endoplasmic reticulum	7 (15.9%)
ko01110	Biosynthesis of secondary metabolites	6 (13.6%)
ko03010	Ribosome	6 (13.6%)
ko00010	Glycolysis / Gluconeogenesis	5 (11.4%)
ko01120	Microbial metabolism in diverse environments	5 (11.4%)
ko01130	Biosynthesis of antibiotics	5 (11.4%)
ko05203	Viral carcinogenesis	4 (9.09%)
ko05016	Huntington's disease	4 (9.09%)
ko01200	Carbon metabolism	4 (9.09%)
ko04612	Antigen processing and presentation	4 (9.09%)
ko05206	MicroRNAs in cancer	4 (9.09%)
ko04145	Phagosome	4 (9.09%)
ko01230	Biosynthesis of amino acids	4 (9.09%)