Population-based comparative analysis of differentially expressed genes between Kashin-Beck disease grades I and II

Xi Wang, Yujie Ning, Wuhong Tan, Hanjie Yu, Zheng Li, Xiong Guo

1. Detailed Inclusion and Exclusion Criteria of Subjects

This study was approved by the Human Ethics Committee of Xi'an Jiaotong University. Written informed consent was obtained from all participants.

All subjects were randomly chosen from Yong-shou and Lin-you counties in Shaanxi province, the endemic areas of KBD, with prevalence of 20.4% and 10.5%, respectively. Patients were diagnosed according to the national diagnostic criteria of Kashin-Beck disease [WS/T207-2010] in China (Table S1). The KBD patients were classified as having a first degree (I) KBD manifested by multiple and symmetrical enlargements of phalanges or other joints, limitation of motion, arthralgia, mild muscle atrophy, and thickening of other limb joints. The patients with the symptoms of the firstdegree KBD with additional brachydactylia were diagnosed as the second degree (II) KBD. The KBD patients were compared with age and sex-matched healthy controls without symptomatic and clinical KBD according to criteria of[WS/T207-2010]. All the patients and the controls with other types of osteoarthropathyand other chronic diseases, such as cardiovascular disease, diabetes and hypertension were excluded.

2. Experimental Procedure of Microarray

(1) Blood collection and PBMCs isolation

Peripheral blood (4 ml) from each subject was collected into heparinized Vacutainer[®] tubes (Becton Dickenson, San Jose, CA, USA) for the gene expression analysis. Leukocyte cell numbers were determined using a Hemovet 950 (Drew Scientific, Oxford, CT, USA). The PBMCs were separated from plasma by centrifuging blood at 1500×g for 20 min. The cell pellets were resuspended in Hanks' balanced salt solution (Gibco BRL/Invitrogen, Carlsbad, CA, USA). The cell suspensions were layered over 5 ml of Lympholyte-H (Cedarlane Labs, Hornby, Canada) in a 15-ml Falcon tube, and centrifuged at 1500×g for 40 min. After rinsing twice with cold Hanks' balanced salt solution, the cells were stored in RNAlater[®] (Ambion Inc., Austin, TX, USA) until RNA isolation.

(2) RNA preparation

Total RNA was isolated from the PBMCs using Trizol[®] reagent (Life Technologies Inc., Carlsbad, CA, USA) following the manufacturer's recommended protocol. The quality and integrity of the extracted total RNA were determined using a high-resolution electrophoresis system (Agilent 2100

Bioanalyzer, Agilent Technologies, Palo Alto, CA, USA).

(3) Microarray hybridization

Isolated total RNA from each KBD/control pair was transcribed into complementary DNA (cDNA), and then reverse-transcribed into cRNA and labeled with CyDye using the Amino AllylMessageAmp[™] aRNA Amplification Kit (Ambion) according to the manufacturer's instructions. Before reverse transcription, RNase-free DNase I was used to remove residual genomic DNA from total RNA. For each pair, 0.5 µg of labeled cRNA was purified separately and then mixed together with hybridization buffer before applying to microarrays. The custom-made primer array, which contains 169 oligonucleotide probes representing the selected 169 human genes (manufactured by the National Engineering Research Center for Miniaturized Detection System in China), was used to perform microarray hybridization following the recommended protocol.

(4) Gene expression analysis

Microarray slides were scanned with Gene-Pix 4000B (Axon Instruments Inc., Foster City, CA, USA), and Gene PixPro 3.0 software (Axon Instruments Inc.) was used to analyze the 16-bit TIFF images produced by the scanner. Ratio images for all the spots were defined by accessing the gene list file describing the microarray location of each gene. After the average signal intensity was subtracted from the median back intensity, the gene expression data were imported into Excel spreadsheets (Microsoft Corp., Redmond, WA, USA) using Unigene and GeneBank descriptors. A global normalization was conducted to calculate the relative expression levels between two samples using all detected genes.

3. Further Statistical Analyses of Differentially Expressed Genes

Each gene was identified as “up-regulated” or “down-regulated” if the “fold change” was more than two-fold or less than 0.5-fold, respectively. The “fold change” value is the ratio of the signal intensity of the KBD sample compared to that of the control sample in each pair. Significant differences in gene expression between the KBD and the control samples were determined by Student's t-test. *P*-values were calculated by standard combinatorial approach and then adjusted for multiple testing by Bonferroni method. *P*-values less than 0.05 were considered statistically significant.

4. Details of the qRT-PCR Method

Quantitative real-time PCR was used to confirm the expression levels of 6 genes, which were the most significantly 4 up-regulated and 2 down-regulated ones. Total RNA was isolated from an additional 10 subjects (Table S2), prepared as for oligonucleotide microarray analysis, and then qRT-PCR was performed. The RNA was then converted into complementary DNA (cDNA) using Superscript II reverse transcriptase (Invitrogen, Karlsruhe, Germany) and random primers. QRT-PCR was performed using the ABI7500 Real-Time PCR system (Applied

Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. Primer sequences are listed in Table 3. All primer and probe sets (ABI3BP, NM_015429.3; BIRC3, NM_001165.4; SSBP1, NM_001256510.1; SIGLEC8, NM_014442.2; TTC25, NM_031421.2; and CSGALNACT1, NM_001130518.1) were supplied by TaqMan[®] Gene Expression Assays (Applied Biosystems). The relative fold-change for each individual gene was calculated by the comparative threshold cycle (Ct) equation as follows: $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct$ = mean ΔCt (KBD sample) – mean ΔCt (control sample); ΔCt = Ct (target gene) - Ct (housekeeping gene), where Ct values of target genes were normalized to Ct values of housekeeping gene (β -ACTIN).

Table S1: General information of the participants included in the analysis

	MicroarrayA		Microarray B	
	Control	Case	Control	Case
N	50	50	50	50
Women/men (% women)	28/22 (56.0)	28/22 (56.0)	30/20 (60.0)	30/20 (60.0)
Mean age in years (range)	57.91 (43-79)	55.43 (40-77)	55.62 (42-78)	56.94 (49-75)
Thickening of joint (%)	2.0	100.0	3.0	100.0
Brachydactyly (%)	0.0	0.0	0.0	100.0
RKBD (%)	0.0	100.0	0.0	100.0

Microarray A: Characteristics of 50 pairs of patients with KBD grade I and controls.

Microarray B: Characteristics of 50 pairs of KBD with grade patients II and controls.

RKBD: radiographic Kashin-Beck disease.

Table S2 Characteristics of the patients with KBD and control used for quantitative RT-PCR analysis

Sample pair	KBD				Normal			
	<i>n</i>	Age (years)	Male	Female	<i>n</i>	Age(years)	Male	Female
1	1	69	1	0	1	65	1	0
2	1	62	1	0	1	56	1	0
3	1	57	1	0	1	56	1	0
4	1	58	0	1	1	55	0	1
5	1	52	0	1	1	48	0	1
Total	5	59.6	3	2	5	56.0	3	2

Table S3 List of 69 differentially expressed genes in patients with KBD grade II ^{a)}

Gene name	Symbol	Public ID	Fold change ^{b)}
<i>Down-regulated genes</i>			
ATP-binding cassette, sub-family C, member 13, pseudogene	ABCC13	NR_003087	0.22±0.02
ABI family, member 3 (NESH) binding protein	ABI3BP	NM_015429	0.13±0.01
Acyl-CoA synthetase long-chain family member 6	ACSL6	NM_001009185	0.16±0.01
Anaphase promoting complex subunit 1	ANAPC1	XM_006712690	0.18±0.02
Aldehyde dehydrogenase 5 family, member A1	ALDH5A1	NM_001080	0.25±0.04
ATR serine/threonine kinase	AQP1	NM_001185060	0.34±0.08
Ataxin 3	ATXN3	NM_001127696	0.46±0.07
Branched chain amino-acid transaminase 1, cytosolic	BCAT1	NM_001178091	0.23±0.01
Chromosome 1 open reading frame 64	C1orf64	NM_178840	0.24±0.03
Complement component 4B (Chido blood group)	C4B	NM_001002029	0.45±0.11
Calcium channel, voltage-dependent, gamma subunit 6	CACNG6	NM_145814	0.24±0.02
Chondroitin sulfate N-acetylgalactosaminyltransferase 1	CSGALNACT1	NM_001130518	0.23±0.04
Cathepsin C	CTSC	NM_001114173	0.23±0.02
Cytochrome b5 type A (microsomal)	CYB5A	NM_001190807	0.49±0.04
Cytochrome b5 reductase 3	CYB5R3	NM_000398	0.21±0.03
Dystrophin, muscular dystrophy	DMD	NM_007868	0.26±0.03
Deleted in primary ciliary dyskinesia homolog	DPCD	NM_015448	0.23±0.03
Endosulfine alpha	ENSA	NM_004436	0.43±0.06
Enhancer of rudimentary homolog (Drosophila)	ERH	NM_004450	0.21±0.01
Family with sequence similarity 104, member A	FAM104A	NM_00109883	0.23±0.01
Family with sequence similarity 46, member C	FAM46C	NM_017709	0.17±0.01
F11 receptor	F11R	NM_016946	0.37±0.03
Fibrillin 1	FBN1	NM_000138	0.45±0.02
FK506 binding protein 9, 63kDa	FKBP9	NM_007270	0.43±0.12
Frizzled family receptor 1	FZD1	NM_003505	0.24±0.04
Growth differentiation factor 5	GDF5	NM_000557	0.21±0.02
Glutaredxin 5	GLRX5	NM_016417	0.33±0.04
G protein-coupled receptor 20	GPR20	NM_005293	0.41±0.03
G protein-coupled receptor, class C, group 5, member C	GPRC5C	NM_018653	0.44±0.04
Glutathione peroxidase 1	GPX1	NM_000581	0.30±0.03
Glutathione peroxidase 2	GPX2	NM_002083	0.32±0.03
Glutathione peroxidase 6	GPX6	NM_182701	0.22±0.03
Glutathione peroxidase 7	GPX7	NM_015696	0.35±0.04
Glutathione S-transferase theta 2	GSTT2	NM_000854	0.45±0.13
Hemoglobin, alpha2	HBA2	NM_000517	0.40±0.07
Ly1 antibody reactive	LYAR	NM_001145725	0.37±0.03
Mitochondrial carrier triple repeat 3 pseudogene	MCART3P	NR_026540	0.16±0.02
Major facilitator superfamily domain containing 3	MFSD3	NM_138431	0.49±0.07
Mannosyl glycoprotein-N-acetylglucosaminyltransferase	MGAT3	NM_001098270	0.48±0.14
Makorin ring finger protein 1	MKRN1	NM_001145125	0.47±0.06
Myeloid/lymphoid or mixed-lineage leukemia 5	MLL5	NM_018682	0.44±0.02
Phosphodiesterase 4B, cAMP-specific	PDE4B	NM_001037339	0.46±0.05
Progastricsin (pepsinogen C)	PGC	NM_001166424	0.42±0.05
Pleiotrophin	PTN	XM_005250516	0.40±0.19
RNA exonuclease 2	REXO2	NM_105523	0.38±0.05
Serine/threonine kinase 11	STK11	NM_000455	0.41±0.06
Tumor necrosis factor, alpha-induced protein 6	TNFAIP6	NM_007115	0.43±0.04
STE20-related kinase adaptor beta	STRADB	NM_001206864	0.17±0.02
Solute carrier family 52	SLC52A3	XM_005260655	0.31±0.01
Zic family member 5	ZIC5	NM_033132	0.23±0.02
<i>Up-regulated genes</i>			
Alpha-2-macroglobulin-like 1	A2ML1	NM_001282424	2.20±0.25

Aquaporin 1	ATR	NM_001184	4.60±0.39
Baculoviral IAP repeat containing 3	BIRC3	NM_001165	6.50±0.53
CD3g molecule, gamma (CD3-TCR complex)	CD3G	NM_000073	4.79±0.36
Deiodinase, iodothyronine, type 2	DIO2	NM_000793	2.07±0.11
EYA transcriptional coactivator and phosphatase 4	EYA4	NM_001301012	6.34±0.94
Fem-1 homolog a (C, elegans)	FEM1A	NM_018708	3.40±0.96
F-box protein 15	FBXO15	NM_001142958	2.94±0.29
FGFR1 oncogene partner 2	FGFR1OP2	NM_001171887	4.97±0.31
HIRA interacting protein 5	HIRIP5	NM_001122708	2.09±0.11
Killer cell lectin-like receptor subfamily C, member 1	KLRC1	NM_002259	2.48±0.12
Optineurin	OPTN	NM_001008211	3.04±0.17
Phosphodiesterase 8B	PDE8B	XM_005248623	2.70±0.21
Sialic acid binding Ig-like lectin 8	SIGLEC8	NM_014442	3.23±0.55
Solute carrier family 14 (urea transporter), member 1	SLC14A1	NM_001128588	4.04±0.31
Signal recognition particle 19kDa	SRP19	NM_001204193	2.03±0.37
Single-stranded DNA binding protein 1, mitochondrial	SSBP1	NM_001256510	3.12±0.25
Tumor necrosis factor (ligand) superfamily, member 11	TNFSF11	NM_003701	5.32±0.49
Tetratricopeptide repeat domain 25	TTC25	NM_031421	2.63±0.20

a) Differential expression genes between the patients with KBD grade II vs. Controls were assessed using the selection criteria described in materials and methods. According to the fold change value, only those genes showing significant differences (P -value<0.05) in expression, screening the differential expression genes in patients with KBD grade II are listed. b) Fold change, the mean and standard error of the mean (SEM) of the foldchange in expression of each gene.

Table S4 Results of linear discriminant analysis (LDA) with the expression ratio of common differentially expressed genes^{b,c}

		KBD Grade	Predicted	Group	Membership	Total
			KBD grade I		KBD grade II	
Original	Count	Grade I	47		3	50
		Grade II	16		34	50
	%	Grade I	94.0		6.0	100.0
		Grade II	32.0		68.0	100.0
Cross-validated ^a	Count	Grade I	39		11	50
		Grade II	17		33	50
	%	Grade I	78.0		22.0	100.0
		Grade II	34.0		66.0	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 81.0% of original grouped cases correctly classified.

c. 72.0% of cross-validated grouped cases correctly classified.