Supplementary information

Two functional loci in the promoter of *EPAS1* gene involved in high-altitude adaptation of Tibetans

Xiang-Hong Xu[1 †], Xue-Wen Huang[2,3 †], Li Qun[4], Ya-Nan Li[1], Yi Wang[1], Chao Liu[5], Yanyun Ma[1], Qing-Mei Liu[1], Kang Sun[5], Feng Qian[1], Li Jin[1]*, Jiucun Wang[1]*

[1] State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R. China.

[2] Clinical laboratory of Huadong Sanatorium, Dajishan, Meiyuan Garden, Wuxi, Jiangsu 214065, P.R. China.

[3] Public Health Bureau for Shigatse District, 5 Keji Road, Shigatse District, Tibet 857000, P.R. China.

[4] Department of Gynecology and Obstetrics, The People's Hospital of Shigatse District, 28 Shanghai Middle Road, Shigatse District, Tibet 857000, P.R. China.

[5] Department of Physiology and Biophysics, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R. China.

[†] These two authors contributed equally to this work.

* Correspondence to: Li Jin or Jiucun Wang, Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 220 Handan Road, Shanghai 200433, P.R. China.

E-mail: lijin.fudan@gmail.com; jcwang@fudan.edu.cn

Supplementary methods

Stable cell line establishment

Flp-In-293 cells (Invitrogen) were used to stably express EPAS1 protein by integrating the pcDNA5/FRT/TO vector containing *EPAS1* coding sequence into the cell genomic locus via Flp recombinase-mediated DNA homologous recombination at the Flp recombination target (FRT) site, according to the manufacturer's instructions. Flp-In-293 cells were co-transfected with pcDNA5/FRT-EPAS1 plasmid together with pOG44 (which encodes the Flp-In recombinase) at the ratio of 1:9 (w/w) by lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. Forty-eight hours after transfection, cells were subject to positive selection with 15 μg/mL of blasticidin and 150 μg/mL hygromycin for two weeks. Selective media was replaced every three days to remove dead cells. Antibiotic-resistant colonies were isolated and re-cloned by serial dilution to establish individual clones, which were further confirmed by analyzing Zeocin-sensitivity and lack of β-galactosidase activity. To confirm the expression of *EPAS1*, cells were treated with 1 μg/mL tetracycline (Invitrogen) for 24 h.

Electrophoretic mobility shift assay (EMSA)

Non-radioactive EMSAs were performed with IKZF1 isoform 3 proteins and biotin-labeled double stranded probes, which were selected on the basis of predicted IKZF1 binding sites in the promoter of the *EPAS1* gene (Figure S1). Non-radioactive EMSAs were also performed with Sp1 proteins and biotin-labeled double stranded probes which were made by the 40-bp insertion fragment. Nuclear extract prepared from HEK293 cells transfected with or without pcDNA 3.1-IKZF1-myc plasmid using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific) was stored at -80℃ and thawed at room temperature just before use. Two kinds of double-stranded biotin-labeled probes were used for IKZF1 EMSAs, which were named G probe and C probe,

respectively, with only one base-pair difference at the site of rs56721780:G>C. The double-stranded biotin-labeled probes were made by annealing the 5'-biotin end-labeled sense and antisense complementary oligonucleotides at a 1:1 molar ratio in the annealing buffer (100 mM Tris-HCl [pH 7.5], 500 mM NaCl, 10 mM EDTA) at 95℃ for 5 min and slowly cooled down to room temperature. The annealed biotin-labeled probes were diluted to 10 fmol/μl with Tris-EDTA solution and stored at -20℃ before use. EMSAs were performed using a LightShift chemiluminescent electrophoretic mobility shift assay kit (Thermo Scientific) according to the manufacturer's instructions. Briefly, an aliquot of nuclear extract containing 3 μg of protein was incubated with 2 μl of 10 × binding buffer, 2.5% glycerol, 0.05% NP-40, 5 mM $MgCl_2$, and 1 μg of poly (dI-dC) in a total volume of 20 μl for 30 min at room temperature. The reaction mixtures were loaded on a 6% or 5% native polyacrylamide gel and transferred to Immobilon-NY^+ membrane (Millipore). The membrane was then crosslinked for 15 min with the membrane face down on a transilluminator equipped with 312 nm bulbs. After the membrane incubated with LightShift stabilized streptavidin-horseradish peroxidase conjugate and the luminol/enhancer stable peroxide solution, the protein and probe complexes were visualized by ImageQuant LAS 4000 mini system. For competition assays to test binding specificity of predicted proteins, excess (100 × or 200 ×) unlabeled G or C competitor or excess (200 × or 400 ×) unlabeled insertion competitor was added to the EMSA reaction. The unlabeled competitors have the same sequence with the corresponding labeled probes and were annealed as previously described for the labeled probes. For supershift EMSAs, 1 μl of anti-myc mouse monoclonal antibody (Cell Signaling Technology, Danvers, MA, USA) or anti-Sp1 polyclonal antibody (Millipore) was pre-incubated with the nuclear protein for 1 h on ice before adding to the binding reaction. Supplementary Table S3 listed the nucleotide sequences of the labeled probes and unlabeled competitors.

Western blot assay

Amnion was homogenized in RIPA lysis buffer (Beyotime, Jiangsu, China) to get total proteins. Extracted protein samples were separated by SDS-PAGE and transferred to a PVDF membrane (Millipore, Billerica, MA, USA). After being blocked with 5% non-fat milk in TBS-T at room temperature for 1 h, the membrane was incubated with primary antibodies against human IKZF1 (R&D System, Minneapolis, MN, USA) or Sp1 (Millipore) or cyclophilin B (Abcam, Cambridge, UK) overnight at 4℃. The bound antibodies were detected with horseradish peroxidase-conjugated horse anti-goat IgG (Ding Guo Biotechnology Co., Shanghai, China) for 2 h at room temperature and visualized using enhanced chemiluminescence solution (Thermo Scientific) with ImageQuant LAS 4000 mini system (GE Healthcare, Piscataway, NJ, USA).

Supplementary tables

Supplementary Table S1: Microarray detection of differentially expressed genes in *EPAS1* stably expressed Flp-In-293 cells

Probe ID	Gene Name	Gene Description	Fold change	CI_low	CI_high	P value
200879_s_at	EPAS1	endothelial PAS domain protein 1	4.481	3.793	5.294	1.20×10^{-69}
215446_s_at	LOX	lysyl oxidase	1.583	1.467	1.708	2.78×10^{-32}
204298_s_at	LOX	lysyl oxidase	2.318	2.013	2.669	1.35×10^{-31}
36711_at	MAFF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog F	2.315	1.922	2.787	8.02×10^{-19}
218507_at	C7orf68	chromosome 7 open reading frame 68	1.262	1.190	1.339	7.19×10^{-15}
229700_at	ZNF738	zinc finger protein 738	0.466	0.382	0.570	7.60×10^{-14}
200621_at	CSRP1	cysteine and glycine-rich protein 1	0.548	0.467	0.644	2.32×10^{-13}
1554452_a_at	C7orf68	chromosome 7 open reading frame 68	1.217	1.152	1.284	1.25×10^{-12}
227209_at	CNTN1	contactin 1	1.461	1.310	1.629	8.85×10^{-12}
231798_at	NOG	noggin	1.764	1.487	2.092	7.16×10^{-11}
204595_s_at	STC1	stanniocalcin 1	1.649	1.412	1.924	2.34×10^{-10}

200632_s_at	NDRG1	N-myc downstream regulated 1	1.173	1.115	1.234	5.03×10^{-10}
242517_at	KISS1R	KISS1 receptor	1.446	1.282	1.630	1.85×10^{-09}
208763_s_at	TSC22D3	TSC22 domain family, member 3	1.103	1.068	1.139	2.33×10^{-09}
227671_at	XIST	X (inactive)-specific transcript (non-protein coding)	0.942	0.923	0.960	2.62×10^{-09}
212063_at	CD44	CD44 molecule (Indian blood group)	0.635	0.543	0.743	1.21×10^{-08}
219908_at	DKK2	dickkopf homolog 2 (Xenopus laevis)	1.874	1.503	2.335	2.28×10^{-08}
241404_at	NA	NA	1.799	1.456	2.222	5.27×10^{-08}
218177_at	CHMP1B	chromatin modifying protein 1B	1.575	1.333	1.861	9.84×10^{-08}
1554876_a_at	S100Z	S100 calcium binding protein Z	0.540	0.429	0.680	1.68×10^{-07}
200665_s_at	SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	0.582	0.474	0.715	2.58×10^{-07}
238066_at	RBP7	retinol binding protein 7, cellular	0.720	0.635	0.816	2.58×10^{-07}
1553470_at	DNAH17	dynein, axonemal, heavy chain 17	1.694	1.381	2.078	4.35×10^{-07}
201170_s_at	BHLHE40	basic helix-loop-helix family, member e40	1.235	1.137	1.342	5.99×10^{-07}
201100_s_at	USP9X	ubiquitin specific peptidase 9, X-linked	0.930	0.904	0.957	6.19×10^{-07}

200878_at	EPAS1	endothelial PAS domain protein 1	0.669	0.571	0.784	6.78×10^{-07}
211175_at	GPR45	G protein-coupled receptor 45	1.723	1.389	2.136	7.30×10^{-07}

Note: NA, not available

Supplementary Table S2: Sequences of primers used in this study

Primer Name	Target region	Sequence 5' to 3'	Roles
pcDNA3.1-IKZF1-F	IKZF1 CDS region	CCGGAATTCGCCACCATGGATGCTGATGAGGGTCAAG	Cloning PCR
pcDNA3.1-IKZF1-R	IKZF1 CDS region	CCCAAGCTTGCTCATGTGGAAGCGGTGCTC	Cloning PCR
pcDNA3.1-Sp1-F	Sp1 CDS region	CCGGAATTCGCCACCATGGGAAGCGACCAAGATCACTCC	Cloning PCR
pcDNA3.1-Sp1-R	Sp1 CDS region	CGGGGTACCGAAGCCATTGCCACTGATATTAATG	Cloning PCR
pEGFP-EPAS1-F	EPAS1 CDS region	CCCAAGCTTATGACAGCTGACAAGGAGAAG	Cloning PCR
pEGFP-EPAS1-R	EPAS1 CDS region	CGCGGATCCCTACTAGGTGGCCTGGTCCAGGGCTCTG	Cloning PCR
pcDNA5.1-EPAS1-F	EPAS1 CDS region	CGCGGATCCGCCACCATGGACTACAAGGATGACG	Cloning PCR
pcDNA5.1-EPAS1-R	EPAS1 CDS region	GCGGATATCCTAGGTGGCCTGGTCCAGGGCTCTGAGG	Cloning PCR
EPAS1-pro-F	EPAS1 promoter	CGGGGTACCGAGGCTGTTGTGGTGAACTACT	Cloning PCR
EPAS1-pro-R	EPAS1 promoter	CCGCTCGAGTCAGGACACTGCCGAGGAT	Cloning PCR
LOX-pro-F	LOX promoter	TCCCCCGGGTTCGCCTGTCTGAGTT	Cloning PCR
LOX-pro-R	LOX exon 1	CCCAAGCTTCACTCCTTTTGCCAGA	Cloning PCR

EPAS1-SNP-Mut-F	EPAS1 promoter	TTTCCAACACCTGTAGCCTTTGCGTTTCCCAGGAC	Site mutation PCR
EPAS1-SNP-Mut-R	EPAS1 promoter	GCAAAGGCTACAGGTGTTGGAAATTTGTAGATTTAA	Site mutation PCR
EPAS1-seq-F1	EPAS1 promoter	GGTCCTTCCACAGCCTTCA	Sequencing
EPAS1-seq-R1	EPAS1 promoter	TAAGGGCTCACACAGGCAAC	Sequencing
EPAS1-seq-F2	EPAS1 promoter	GCCGAGAAATTATCCCCACCT	Sequencing
EPAS1-seq-R2	EPAS1 promoter	CACAGCGTCGGTGTCCTCGTA	Sequencing
EPAS1-indel-F	EPAS1 promoter	GATGCAGTGACTTGAGGGCACAT	Indel-PCR
EPAS1-indel-R	EPAS1 promoter	TTTTGCCAGATTGACCCCG	Indel-PCR
EPAS1-F	EPAS1 coding region	ATAAGTTCACCCAAAACCCCAT	qRT-PCR
EPAS1-R	EPAS1 coding region	GGCAGCAGGTAGGACTCAAAT	qRT-PCR
GAPDH-F	GAPDH coding region	GAGTCAACGGATTTGGTCGT	qRT-PCR
GAPDH-R	GAPDH coding region	CATGGGTGGAATCATATTGGA	qRT-PCR
LOX-F	LOX coding region	AGCATACAGGGCAGATGTCAGAG	qRT-PCR
LOX-R	LOX coding region	CTTGGTCGGCTGGGTAAGAAAT	qRT-PCR

ChIP-LOX-F	LOX promoter	TGGCATTGCTTGGTGGAGA	ChIP-qPCR
ChIP-LOX-R	LOX promoter	TTTTGCCAGATTGACCCCG	ChIP-qPCR

Note: Underlined bases indicated the restriction sites

Supplementary Table S3: Sequences of probes used for EMSA

Probe name	Sequence (5'to3')	Base number
5’-biotin end-labeled G oligonucleotides (sense)	GTAGCCTTTGGGTTTCCCAGGACTTCCATC	30
5’-biotin end-labeled G oligonucleotides (antisense)	GATGGAAGTCCTGGGAAACCCAAAGGCTAC	30
Unlabeled G oligonucleotides (sense)	GTAGCCTTTGGGTTTCCCAGGACTTCCATC	30
Unlabeled G oligonucleotides (antisense)	GATGGAAGTCCTGGGAAACCCAAAGGCTAC	30
5’-biotin end-labeled C oligonucleotides (sense)	GTAGCCTTTGCGTTTCCCAGGACTTCCATC	30
5’-biotin end-labeled C oligonucleotides (antisense)	GATGGAAGTCCTGGGAAACGCAAAGGCTAC	30
Unlabeled C oligonucleotides (sense)	GTAGCCTTTGCGTTTCCCAGGACTTCCATC	30
Unlabeled C oligonucleotides (antisense)	GATGGAAGTCCTGGGAAACGCAAAGGCTAC	30
5’-biotin end-labeled insertion oligonucleotides (sense)	CGCGCAGGAGCGGCGGGGTGCCGAGAGTAGGGACAGAAGC	40
5’-biotin end-labeled insertion oligonucleotides (antisense)	GCTTCTGTCCCTACTCTCGGCACCCCGCCGCTCCTGCGCG	40
Unlabeled insertion oligonucleotides (sense)	CGCGCAGGAGCGGCGGGGTGCCGAGAGTAGGGACAGAAGC	40

Unlabeled insertion oligonucleotides (antisense)	GCTTCTGTCCCTACTCTCGGCACCCCGCCGCTCCTGCGCG	40

Supplementary figure legends

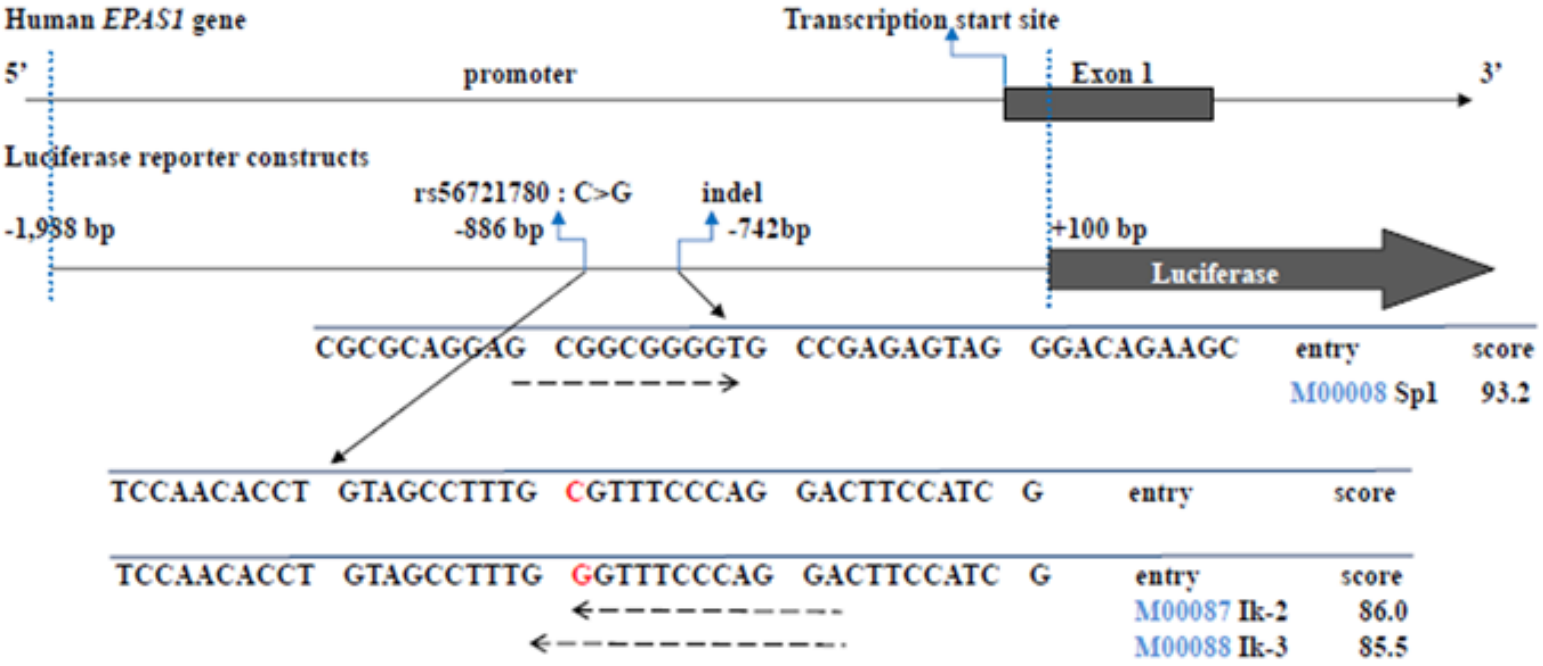
Fig. S1: Schematic representation of *EPAS1* gene is given. *EPAS1* promoter sequence from -1,988 bp to +100 bp relative to the transcription start site was cloned to pGL3-basic vector to generate luciferase reporter constructs, pGL3-G/deletion/C and pGL3-C/insertion/T. Luciferase reporter construct pGL3-C/deletion/C was generated by PCR-based site-directed mutagenesis at the position of rs56721780:G>C (-886 bp) using pGL3-G/deletion/C as the template. The sites of rs56721780:G>C and -742indel as well as the transcription start site are indicated. Mouse Ik-2 and Ik-3 (also known as IKZF1 isoform 3 and 5 in human, respectively), and Sp1 were computationally predicted as the possible target transcription factors at the sites of rs56721780:G>C and -742indel, respectively, by the website software TFSEARCH (version 1.3, threshold score 85.0, classification: vertebrate).

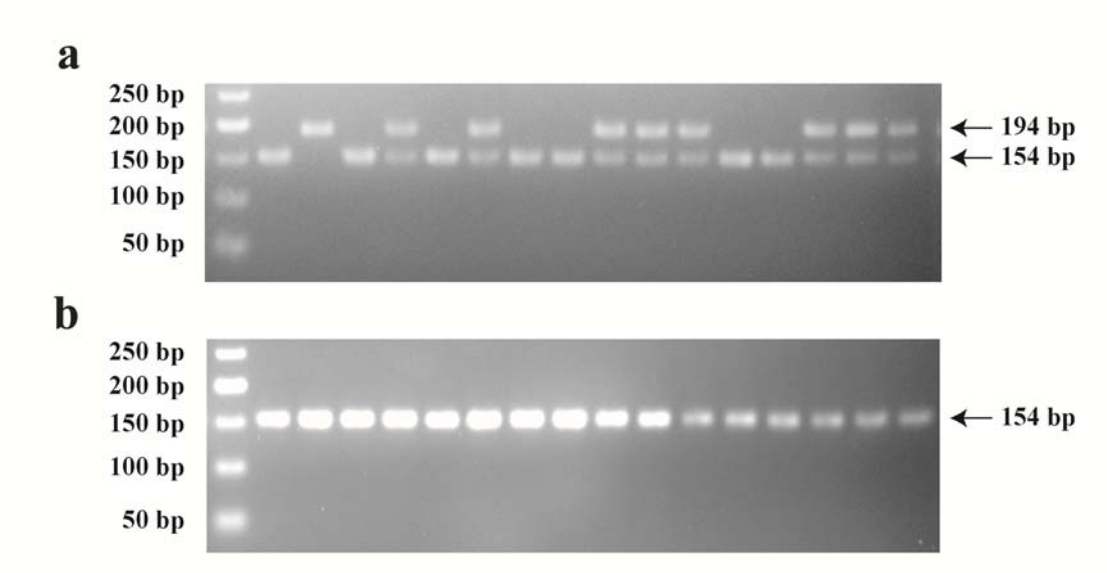
Fig. S2: *EPAS1*genotypes at the -742indel site were verified by PCR method. Samples with wild-type allele (deletion) yielded a PCR product of 154 bp, whereas the mutant allele (insertion) generated a PCR product of 194 bp. The first lane contains a 50-bp DNA size marker. (**a)** PCR results of 15 Tibetan subjects were shown. Totally, 43 Tibetan samples were studied. (**b**) PCR results of 15 Chinese Han subjects were shown. Totally, 25 Chinese Han samples were studied.

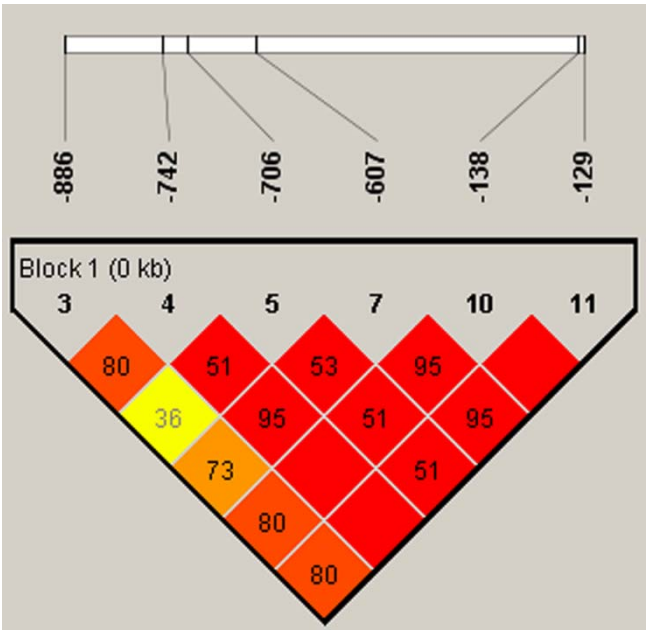
Fig. S3: Linkage disequilibrium (LD) plot of *EPAS1* gene promoter region in Tibetans was shown. The LD was measured using R^2, only variants with MAF larger than 0.25 were shown. The LD plots were generated using Haploview V3.32 software. LD color scheme is GOLD heat map and LD block was defined using confidence intervals rule.

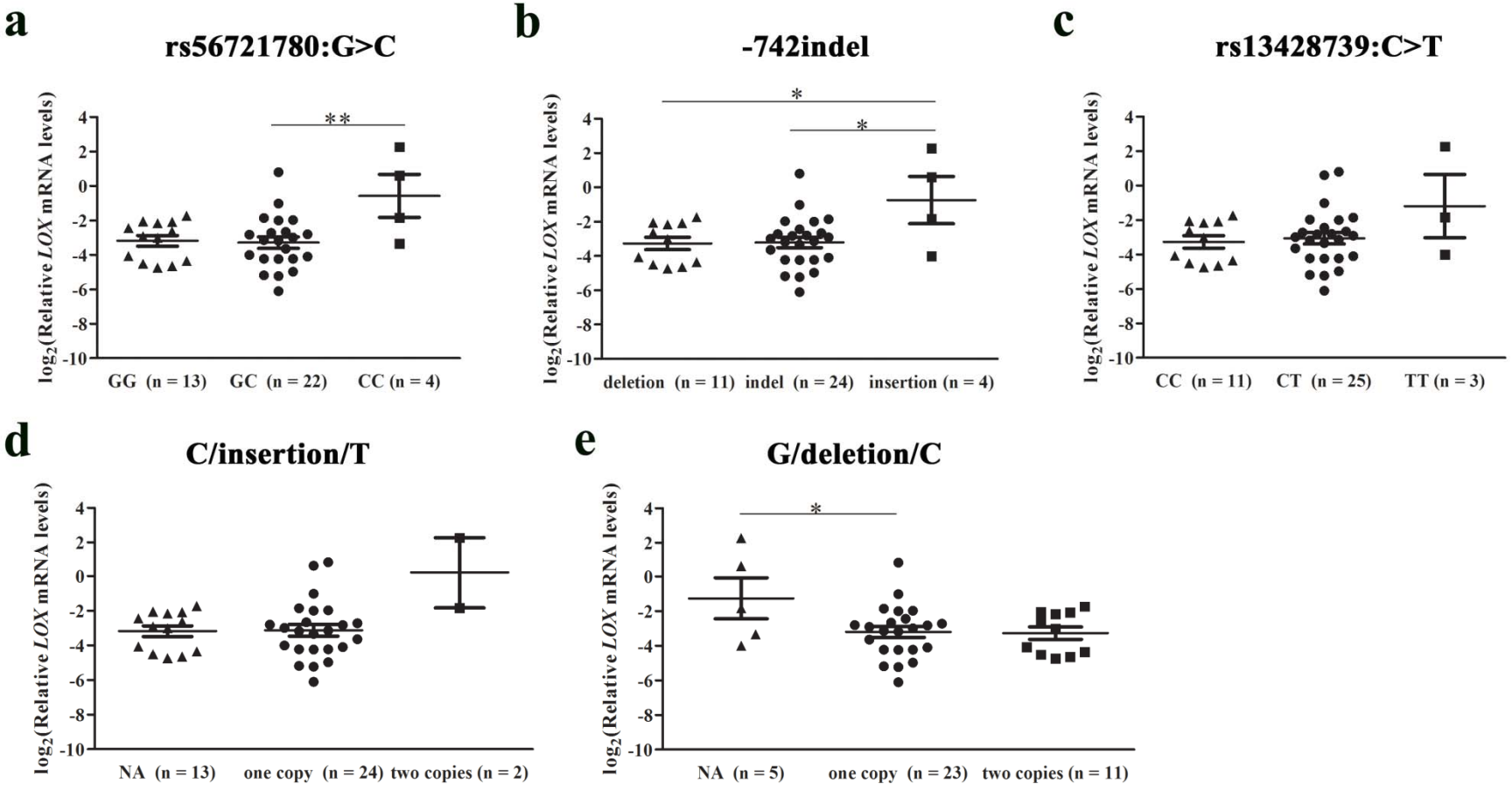
Fig. S4: Three variants in the promoter region of *EPAS1* gene were correlated with *LOX* gene expression levels in amnion. Individuals possessing wild-type alleles at the positions of (**a)** rs56721780:G>C, (**b**) -742indel and (**c**) rs13428739 had lower *LOX* expression levels in amnion. Individuals with two copies of (**d**) C/insertion/T haplotype had much higher *LOX* expression levels in amnion, whereas the (**e**) G/deletion/C haplotype decreased *LOX* expression levels. Relative expression levels were log2 transformed for plotting, with *GAPDH* as the endogenous control. Data were presented as mean ± S.E. Statistical analysis was performed using the *t* test or *t* test with Welch's correction implemented in the software of Graphpad Prism. *P < 0.05, **P < 0.01. The numbers in parentheses indicated the counts of individuals with the corresponding genotypes. NA, not available or without the haplotype indicated.

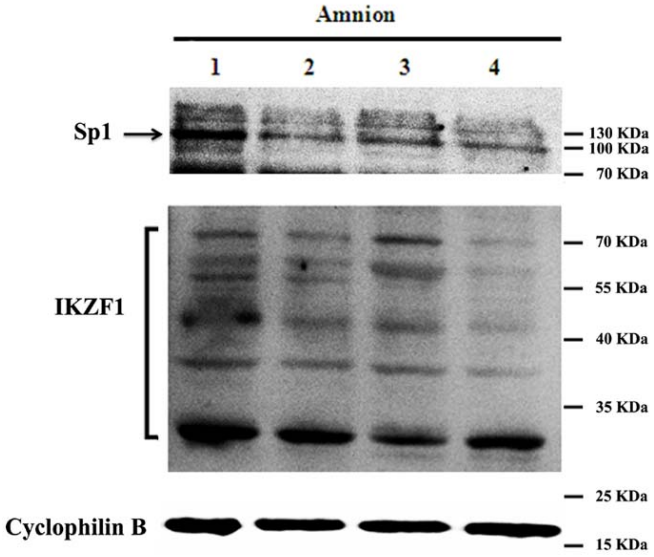
Fig. S5: Western blot analysis showed the expression of IKZF1and Sp1 in the amnion from Tibetan newborns.


Human EPAS1 gene
Transcription start site
5'
promoter
Exon 1
3'
Luciferase reporter constructs
rs56721780 : C>G
indel
-1,988 bp
-886 bp
-742bp
+100 bp
Luciferase
CGCGCAGGAG CGGCGGGGTG CCGAGAGTAG GGACAGAAGC
entry
score
M00008 Sp1 93.2
TCCAACACCT GTAGCCTTTG CGTTTCCCAG GACTTCCATC G
entry
score
TCCAACACCT GTAGCCTTTG GGTTTCCCAG GACTTCCATC G
entry
score
M00087 Ik-2 86.0
M00088 Ik-3 85.5


a
250 bp
200 bp
150 bp
100 bp
50 bp
194 bp
154 bp
b
250 bp
200 bp
150 bp
100 bp
50 bp
154 bp


-886
-742
-706
-607
-138
-129
Block 1 (0 kb)
3
4
5
7
10
11
80
51
53
95
36
95
51
95
73
51
80
80


a
rs56721780:G>C
log2(Relative LOX mRNA levels)
**
GG (n = 13)
GC (n = 22)
CC (n = 4)
b
-742indel
log2(Relative LOX mRNA levels)
*
*
deletion (n = 11)
indel (n = 24)
insertion (n = 4)
c
rs13428739:C>T
log2(Relative LOX mRNA levels)
CC (n = 11)
CT (n = 25)
TT (n = 3)
d
C/insertion/T
log2(Relative LOX mRNA levels)
NA (n = 13)
one copy (n = 24)
two copies (n = 2)
e
G/deletion/C
log2(Relative LOX mRNA levels)
*
NA (n = 5)
one copy (n = 23)
two copies (n = 11)


Amnion
1
2
3
4
Sp1
130 KDa
100 KDa
70 KDa
IKZF1
70 KDa
55 KDa
40 KDa
35 KDa
25 KDa
Cyclophilin B
15 KDa