

ORIGINAL ARTICLE

Open Access



# Association of *EGLN1* genetic polymorphisms with SpO<sub>2</sub> responses to acute hypobaric hypoxia in a Japanese cohort

Yoshiki Yasukochi<sup>1\*</sup> , Takayuki Nishimura<sup>2</sup>, Midori Motoi<sup>2</sup> and Shigeki Watanuki<sup>3</sup>

## Abstract

**Background:** Recent studies have explored various genetic and physiological factors related to high-altitude adaptation in highlander populations. However, the effects of single nucleotide polymorphisms (SNPs), influencing such adaptation, on physiological responses to hypobaric hypoxia have not been examined in lowlanders with lowlander ancestry. Thus, we aimed to investigate the association between SNPs around the *EGLN1* genomic region, possibly involved in high-altitude adaptation, and physiological changes to hypobaric hypoxia exposure in a cohort of Japanese lowlanders.

**Methods:** Physiological data were obtained from 46 healthy Japanese male students under different atmospheric pressure conditions (equivalent to sea level and altitudes of 2500 and 4000 m). Genotypes of seven SNPs around *EGLN1* were determined in all subjects by PCR-direct sequencing or TaqMan SNP genotyping assay.

**Results:** Results of the association study suggest that percutaneous arterial oxygen saturation (SpO<sub>2</sub>) responses of individuals with rs12097901 and rs2790859 alleles, whose frequencies are high in highlander populations (HL alleles), may be susceptible to acute hypobaric hypoxia. SpO<sub>2</sub> levels of individuals with HL alleles were lower than those of individuals with non-HL alleles. At the same time, the subjects with HL alleles did not appear to have any remarkable hematological or pulmonary features that may counteract the low levels of SpO<sub>2</sub>. One may hypothesize that the low SpO<sub>2</sub> levels in HL allele carriers could be a risk factor for acute mountain sickness in Japanese population.

**Conclusions:** Our findings suggest that rs12097901 and rs2790859 genotypes affect SpO<sub>2</sub> responses and may be associated with the susceptibility to acute hypobaric hypoxia in Japanese population.

**Keywords:** *EGLN1*, High-altitude adaptation, Hypobaric hypoxia, SNP, SpO<sub>2</sub>

## Background

Recent studies have identified single nucleotide polymorphisms (SNPs) related to physiological phenotypes that may be adaptive in high-altitude (hypobaric hypoxia) environments. In particular, physiological and genetic factors that contribute to high-altitude adaptations have been extensively studied in Tibetan, Ethiopian, and Andean highlander populations. Beall [1] suggested that

physiological responses to hypobaric hypoxia stress were different in these three populations.

In recent years, genome-wide scans in Tibetan highlanders detected signals of recent positive selection in several chromosomal regions containing genes involved in the hypoxia-inducible factor (HIF) pathway, especially *EPAS1* and *EGLN1* [2–6]. HIFs play a central role in oxygen homeostasis and regulate erythropoiesis, angiogenesis, anaerobic metabolism, and glycolytic pathway [7, 8]. Under normoxic condition, prolyl hydroxylase domain-containing proteins [PHD1 (encoded by *EGLN2*), PHD2 (*EGLN1*), and PHD3 (*EGLN3*)], which are cellular oxygen sensors, hydroxylate specific proline residues of HIF- $\alpha$

\* Correspondence: [hyasukou@proof.ocn.ne.jp](mailto:hyasukou@proof.ocn.ne.jp)

<sup>1</sup>Department of Human Functional Genomics, Advanced Science Research Promotion Center, Organization for the Promotion of Regional Innovation, Mie University, 1577 Kurima-machiya, Tsu, Mie 514-8507, Japan  
Full list of author information is available at the end of the article



(HIF-1 $\alpha$ , HIF-2 $\alpha$ , and HIF-3 $\alpha$ ), leading to the degradation of HIFs in the proteasome pathway involving the von Hippel-Lindau protein. Under hypoxia, HIF- $\alpha$  is not degraded and dimerizes with HIF- $\beta$ , leading to the upregulation of proangiogenic molecules, such as the vascular endothelial growth factor and erythropoietin. Tibetan-specific *EPAS1* and *EGLN1* alleles were associated with low hemoglobin concentration in the population, and these alleles can be protective against chronic hypoxia caused by excessive erythrocytosis [2, 9].

A signal of recent positive selection in Andean and Tibetan populations has been commonly detected in a chromosomal region containing *EGLN1* that encodes the PHD2 protein [10]. This suggests that *EGLN1* may be a key factor for the high-altitude adaptation of modern humans. However, there have been no studies examining the effect of *EGLN1* genetic polymorphism on physiological responses of lowlanders with lowlander ancestry under hypoxia. Therefore, in the present study, we explored the relationship between *EGLN1* SNPs and physiological responses of Japanese lowlanders under acute hypobaric hypoxia.

## Methods

### Study subjects

A total of 46 healthy Japanese male students of Kyushu University (Fukuoka, Japan; mean age  $22.5 \pm 1.17$  years) who had no clinical problems participated in the present study. None of them was a medium- or long-term highland resident within the last 2 months preceding the experiment. The subjects had been prohibited from taking exercise and drinking alcohol for 1 day, and from eating, smoking, and drinking caffeine for  $> 2$  h before entering the hypobaric chamber. Their physical characteristics (mean values  $\pm$  standard deviation) were as follows: height,  $1.71 \pm 0.05$  m; weight,  $61.1 \pm 12.3$  kg; body mass index (BMI),  $20.8 \pm 3.96$  kg/m<sup>2</sup>; and body surface area,  $1.66 \pm 0.15$  m<sup>2</sup>.

The study was approved by the Ethical Committee of Kyushu University Institutional Review Board for Human Genome/Gene Research and by the Committee on Human Research of Mie University Graduate School of Medicine.

### Physiological data collection

The experiment was carried out with the subjects wearing T-shirts and shorts. Measurement sensors were attached to the participants for approximately 30 min at 28 °C prior to the experiment, and then the subjects entered the hypobaric chamber. Physiological data, including oxygen consumption (VO<sub>2</sub>), carbon dioxide output (VCO<sub>2</sub>), minute ventilation (MV), percutaneous arterial oxygen saturation (SpO<sub>2</sub>), pulse rate (PR), and perfusion index (PI), were measured in the subjects that had rested

in the sitting position. Hematological and pulmonary phenotypic data were independently collected in the following order of conditions: 765 Torr (equivalent to sea level) before decreasing the ambient atmospheric pressure, 562 Torr (equivalent to the altitude of 2500 m), 465 Torr (equivalent to 4000 m), and 765 Torr after pressure recovery. The detailed methods of the experiment and data collection are described elsewhere [11, 12].

### DNA extraction and SNP genotyping

Genomic DNA was extracted from subjects' saliva using a Saliva DNA Isolation Kit (Norgen Biotek Corporation, Thorold, ON, Canada) according to the manufacturer's protocol. Because the amount of genomic DNA was scarce, an Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp, NJ, USA) was used to perform whole-genome amplification. We selected seven SNPs (rs12097901, rs186996510, rs480902, rs479200, rs2808611, rs2790859, and rs2275279) around *EGLN1* to examine the association with time series variation of physiological measurements [at 0, 30 (equivalent to 2500 m), 60 (equivalent to 4000 m), and 90 min]. These SNPs have been previously suggested to be associated with high-altitude adaptation in highlander populations [3, 13, 14]. Of the seven selected SNPs, two (rs12097901 and rs186996510) are located in exon 1 of *EGLN1*. A DNA fragment covering a partial exon 1 sequence, including the two SNPs, was amplified by PCR using the following set of primers: *EGLN1*ex1-F, 5'-CAGTAACGGCCCCCTATCTC TC-3' (forward), and *EGLN1*ex1-R, 5'-TACTCGAGCGC CAGCTTC-3' (reverse). PCR amplification was conducted in a 10- $\mu$ L reaction mixture containing 0.2  $\mu$ L (0.2  $\mu$ M) each of forward and reverse primers, 5  $\mu$ L of 2  $\times$  AmpliTaq Gold 360 Master Mix, 0.5  $\mu$ L of the extracted DNA, and 4.1  $\mu$ L of dH<sub>2</sub>O, using a TaKaRa PCR Thermal Cycler Dice (TaKaRa Bio). After incubation at 95 °C for 10 min, 30–35 cycles were performed as follows: 30 s at 95 °C, 30 s at 62 °C, 45 s at 72 °C, with a post-cycling extension at 72 °C for 7 min. Direct sequencing of the PCR products was performed by Eurofins Genomics (Tokyo, Japan). Five SNPs (rs480902, rs479200, rs2808611, rs2790859, and rs2275279) were genotyped by TaqMan SNP genotyping assay (Applied Biosystems, CA, USA).

### Statistical analyses

Genotype data of all subjects were converted into numeric data according to the dominant, additive, and recessive inheritance models. The dominant and recessive models were defined as "0, AA; 1, AB + BB" and "0, AA + AB; 1, BB" (A, major allele; B, minor allele), respectively, whereas the additive model was defined as "0, AA; 1, AB; 2, BB." The significance of difference in physiological measurements between subjects with different genotypes in the dominant or recessive models was assessed using Welch's *t*-test. To detect

potential confounders, the generalized linear model (GLM) was applied to examine the correlation between anthropometric phenotypes and SpO<sub>2</sub> in the additive model. A Gaussian distribution was selected for the family with an identity link in the model for the phenotypic data, because the SpO<sub>2</sub> measurement data are continuous data. The GLM method showed no correlations between anthropometric phenotypes and SpO<sub>2</sub> ( $P > 0.05$ ), and thereby the GLM analysis without covariates was employed to test the relation among genotypes and the SpO<sub>2</sub> measurements in the additive model. Two-way analysis of variance (ANOVA) was conducted using time (the experimental conditions) and genotype as independent variables and the repeated measurements as dependent variables. The statistical tests described above were performed using R software version 3.4.2 [15] via RStudio version 1.1.383 [16]. Linear regression analysis was conducted to examine the relationships between SpO<sub>2</sub> and several hematological and pulmonary phenotypes. For regression analysis, we applied the analysis of covariance (ANCOVA) to test the differences in the regression coefficients (i.e., slope of regression line) between genotypes and in the adjusted mean values (i.e., the intercept of the regression line) [17].

### Population genetics

Haplotype phase and linkage disequilibrium (LD) between SNPs were estimated using IMPUTE2 version 2.3.2 [18] and Haploview version 4.2 [19] programs, respectively. We estimated  $D'$  values among SNPs using R package “genetics” [20]. The LDs between pairs of the *EGLN1* SNPs in four ethnic groups (East Asian, South Asian, European, and African populations) from the 1000 Genomes Project (<http://www.internationalgenome.org/>, [21]) were surveyed, employing LDlink web-based tools (<https://analysistools.nci.nih.gov/LDlink/>, [22]). The categories

of four ethnic groups are listed at <http://www.internationalgenome.org/data-portal/population>. A large integrated variant dataset of JPT (Japanese in Tokyo, Japan) from the 1000 Genomes Project was obtained through the 1000 Genomes Browser (<http://phase3-browser.1000genomes.org/>). Allele frequency data were obtained through the Ensembl genome browser (<http://www.ensembl.org>, [23]). The allele whose frequency was high in African populations was defined as the ancestral allele. Significant deviations from the Hardy-Weinberg equilibrium (HWE) were tested by Fisher's exact test, using PLINK 1.90 (<http://zzz.bwh.harvard.edu/plink/>, [24]). Network and neighbor-joining (NJ) tree [25] of the estimated haplotypes were constructed with the neighbor-net method [26] by using Splits-tree4 ver. 4.14.4 [27] and with p-distance by using MEGA version 7 [28], respectively.

## Results

### Genotypes of seven SNPs around *EGLN1* in 46 Japanese subjects

We genotyped the seven selected SNPs around *EGLN1* in 46 healthy Japanese males (Table 1). Frequencies of derived alleles (~0.6) at two SNPs (rs480902 and rs2808611) were higher than those of ancestral alleles (~0.4). In addition, frequency data of derived alleles in African populations from the Ensembl genome browser were approximately 0.4, suggesting that the frequencies of the derived alleles may have increased after Out-of-Africa. Derived alleles of three SNPs (rs479200, rs12097901, and rs186996510) or ancestral alleles of the remaining four SNPs were frequently observed in highlander populations [3, 13, 14]. We refer to the allele whose frequency is high in highlander populations as “HL allele” (Table 1).

**Table 1** Summary of SNPs examined in the present study

RefSNP ID	Position <sup>a</sup>	Region	N	Allele frequency <sup>b</sup>		HL allele <sup>d</sup>	Allele frequency in highlanders <sup>d</sup>	$H_{obs}$	$H_{exp}$	HWE (P value) <sup>e</sup>
				Ancestral <sup>c</sup>	Derived <sup>c</sup>					
rs480902	1: 231,395,881	<i>EGLN1</i> intron 1	46	T: 0.413 (38)	C: 0.587 (54)	T	0.71	0.565	0.485	0.37
rs479200	1: 231,408,034	<i>EGLN1</i> intron 1	46	T: 0.609 (56)	C: 0.391 (36)	C	0.71	0.478	0.476	1.00
rs2808611	1: 231,412,734	<i>EGLN1</i> intron 1	46	G: 0.424 (39)	A: 0.576 (53)	G	NA	0.500	0.488	1.00
rs12097901	1: 231,421,509	<i>EGLN1</i> exon 1 (C127S) <sup>f</sup>	46	G: 0.533 (49)	C: 0.467 (43)	C	0.90	0.457	0.498	0.56
rs186996510	1: 231,421,877	<i>EGLN1</i> exon 1 (D4E) <sup>f</sup>	46	C: 0.967 (89)	G: 0.033 (3)	G	0.81	0.022	0.063	0.03
rs2790859	1: 231,465,611	Intergenic region	46	T: 0.620 (57)	C: 0.380 (35)	T	0.76	0.543	0.471	0.37
rs2275279	1: 231,591,348	<i>TSNAX-DISC1</i>	46	A: 0.663 (61)	T: 0.337 (31)	A	0.69	0.283	0.447	0.02

Abbreviations: SNP single nucleotide polymorphism,  $H_{obs}$  observed heterozygosity,  $H_{exp}$  expected heterozygosity, HWE Hardy-Weinberg equilibrium, NA not available

<sup>a</sup>Position in NCBI build GRCh38

<sup>b</sup>Values indicate allele frequencies, and the observed numbers are indicated in parentheses

<sup>c</sup>The allele whose frequency was high in African populations was defined as the ancestral allele, according to allele frequency data obtained through the Ensembl genome browser (<http://www.ensembl.org>, [23])

<sup>d</sup>Allele frequencies are high in highlander populations [3, 13, 14]

<sup>e</sup>Statistically significant P values ( $P < 0.05$ , by Fisher's exact test) are shown in italics

<sup>f</sup>Amino acid substitution

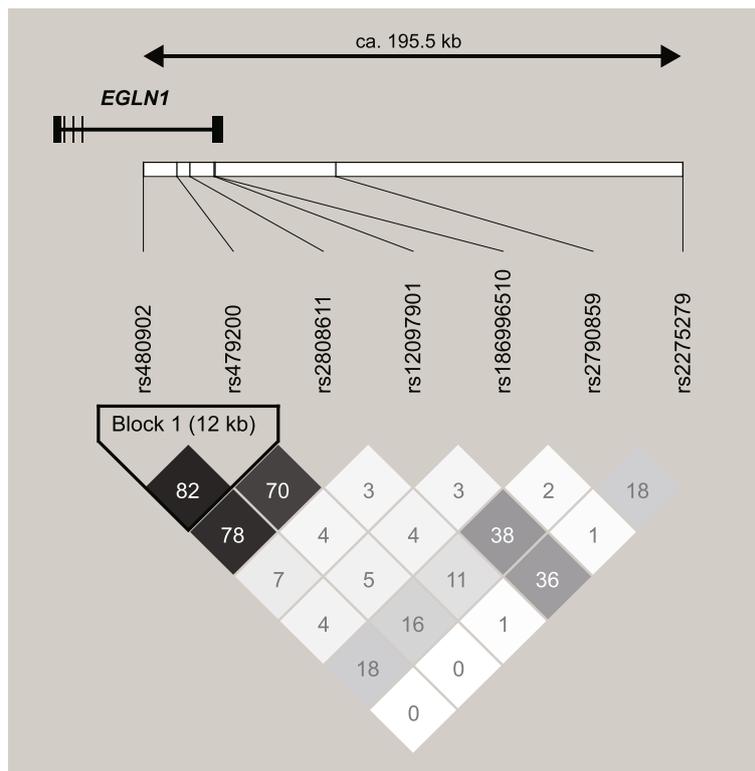
The target seven SNPs are closely located in the chromosomal region 1q42.2. We thus examined LDs between pairs of these SNPs (Fig. 1). The LDs between three SNPs (rs480902, rs479200, and rs2808611) were relatively strong ( $D' = 0.90\text{--}0.95$ ,  $r^2 = 0.70\text{--}0.82$ ). In African populations, from the 1000 Genomes Project database,  $r^2$  values between pairs of the three SNPs were low ( $r^2 = 0.16\text{--}0.34$ ), although  $D'$  values were relatively high ( $D' = 0.43\text{--}0.99$ ), according to LDmatrix, an LDlink application. In three ethnic groups (East Asians, South Asians, and Europeans), the  $r^2$  and  $D'$  values were higher than those in the African populations ( $D' = 0.94\text{--}0.95$  and  $r^2 = 0.82\text{--}0.98$  for East Asians;  $D' = 0.98\text{--}1.00$  and  $r^2 = 0.73\text{--}0.84$  for South Asians;  $D' = 0.96\text{--}1.00$  and  $r^2 = 0.51\text{--}0.87$  for Europeans). These results suggest that the strong LDs may have been established since the Out-of-Africa event. We additionally examined LDs between pairs of 1504 biallelic sites across a ~399-kb genomic region at 1q42.2, based on datasets from JPT in the 1000 Genomes Project database. The LD plot indicated that the three SNPs (rs480902, rs479200, and rs2808611) were in a large LD block while the other SNPs examined were located outside the LD block in the chromosomal region (Additional file 1: Figure S1). The values of  $r^2$  between pairs of the remaining four

SNPs examined in the present study were less than 0.4 ( $r^2 = 0.00\text{--}0.38$ ), whereas the  $D'$  values were relatively high ( $D' = 0.50\text{--}1.00$ ).

Among the seven SNPs examined, two (rs2275279 and rs186996510) showed significant deviations from HWE ( $P = 0.02\text{--}0.03$ ) (Table 1). Thus, we removed these SNPs from further analyses. The mean observed and expected heterozygosities for the remaining five SNPs among the 46 subjects were 0.509 and 0.484, respectively. Based on Bonferroni's correction,  $P$  values of  $< 0.01$  ( $0.05/5$  SNPs examined) were considered statistically significant for further analyses.

**Association of physiological responses to acute hypobaric hypoxia with the genotypes of five SNPs around EGLN1**

We examined the relationships between genotypes of the five SNPs around *EGLN1* and physiological responses to acute hypobaric hypoxia exposure (at 0, 30, 60, and 90 min) in the three inheritance models. Time series variations of pulmonary and hematological traits ( $VO_2$ ,  $VCO_2$ , MV,  $SpO_2$ , PR, and PI) and  $SpO_2$  latency (latency in the first reading and time to stabilization of  $SpO_2$ ) in individuals with different genotypes of each SNP are shown in Table 2 and Additional file 2: Table S1. Through the GLM analysis, the rs12097901 genotype



**Fig. 1** The physical position and linkage disequilibrium of seven SNPs across a ~195.5-kb genomic region at 1q42.2 in 46 Japanese subjects. The diagram was created by Haploview version 4.2. The number in a diamond represents  $r^2$  value ( $\times 100$ ). The haplotype block was defined by the method of Gabriel et al. [36]

**Table 2** Time series variations of SpO<sub>2</sub> in individuals with different genotypes

RefSNP ID	Position <sup>a</sup>	Genotype <sup>b</sup>	2N <sup>c</sup>	SpO <sub>2</sub> <sup>d</sup>				SpO <sub>2</sub> latency <sup>e</sup>	P value <sup>f</sup>
				0 min	30 min	60 min	90 min		
rs480902	1: 231,395,881	CC	14	98.3 ± 0.24	93.6 ± 0.35	86.0 ± 0.84	97.0 ± 0.30	58.1 ± 1.61	0.367
		TC	26	98.5 ± 0.19	94.1 ± 0.31	86.2 ± 0.57	97.6 ± 0.29	57.2 ± 1.46	
		TT	6	98.2 ± 0.67	93.7 ± 0.41	85.7 ± 0.56	97.2 ± 0.63	55.0 ± 2.27	
rs479200	1: 231,408,034	TT	17	98.3 ± 0.22	93.6 ± 0.30	86.2 ± 0.73	97.1 ± 0.27	57.5 ± 1.52	0.527
		TC	22	98.5 ± 0.22	94.2 ± 0.36	86.1 ± 0.66	97.5 ± 0.34	57.5 ± 1.63	
		CC	7	98.3 ± 0.57	93.6 ± 0.38	85.8 ± 0.48	97.4 ± 0.55	55.1 ± 1.92	
rs2808611	1: 231,412,734	AA	15	98.3 ± 0.25	93.7 ± 0.32	86.1 ± 0.80	97.0 ± 0.30	58.1 ± 1.52	0.766
		GA	23	98.5 ± 0.21	94.0 ± 0.36	86.0 ± 0.63	97.5 ± 0.31	57.4 ± 1.55	
		GG	8	98.4 ± 0.50	93.9 ± 0.34	86.2 ± 0.58	97.5 ± 0.57	54.8 ± 2.25	
rs12097901	1: 231,421,509	GG	14	99.0 ± 0.10	94.5 ± 0.31	87.2 ± 0.60	97.2 ± 0.33	61.2 ± 1.35	0.006
		CG	21	98.2 ± 0.20	93.4 ± 0.32	85.5 ± 0.68	97.3 ± 0.30	56.1 ± 1.56	
		CC	11	98.1 ± 0.46	93.9 ± 0.47	85.8 ± 0.79	97.6 ± 0.51	54.1 ± 1.76	
rs2790859	1: 231,465,611	TT	16	98.0 ± 0.33	93.5 ± 0.35	85.4 ± 0.71	97.3 ± 0.35	55.2 ± 1.44	0.014
		TC	25	98.5 ± 0.17	94.0 ± 0.30	86.1 ± 0.56	97.5 ± 0.28	56.9 ± 1.38	
		CC	5	99.1 ± 0.19	94.6 ± 0.55	88.2 ± 0.77	96.6 ± 0.61	64.8 ± 1.71	

Abbreviation: SpO<sub>2</sub> percutaneous arterial oxygen saturation

<sup>a</sup>Position in NCBI build GRCh38

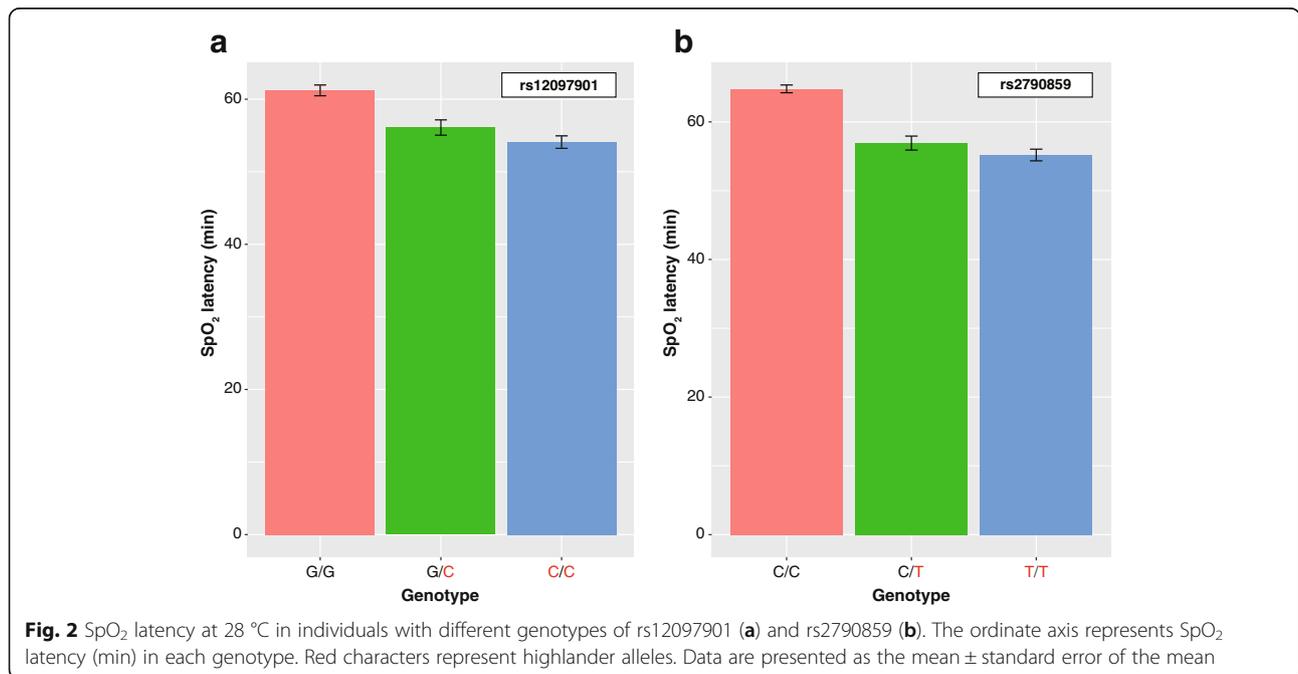
<sup>b</sup>Highlander alleles are shown in italics (see Table 1)

<sup>c</sup>The number of chromosomes

<sup>d</sup>Quantitative data are presented as the mean ± standard error of the mean

<sup>e</sup>Latency in the first reading and time to stabilization of SpO<sub>2</sub>

<sup>f</sup>P values estimated by the GLM analysis for SpO<sub>2</sub> latency



showed a significant association with SpO<sub>2</sub> latency ( $P = 0.006$ ) in the additive model, whereas the rs2790859 genotype was related to the latency with borderline significance ( $P = 0.014$ ) (Table 2 and Fig. 2). In addition, the significant relations of SpO<sub>2</sub> latency to rs12097901 and rs2790859 genotypes were shown in the dominant and recessive models, respectively ( $P = 0.003$ ). SpO<sub>2</sub> latencies of individuals with HL alleles of these SNPs were shorter than those of individuals with non-HL alleles. The differences of SpO<sub>2</sub> latencies may be attributable to effects on physiological reaction time to acute hypobaric hypoxia exposure. In the present study, three of the 46 subjects examined showed high BMI values ( $> 30 \text{ kg/m}^2$ ). It has been reported that BMI may influence SpO<sub>2</sub> levels [29]. However, the corresponding SpO<sub>2</sub> values were not extremely high (Additional file 3: Figure S2). Thus, these subjects were included for further analyses.

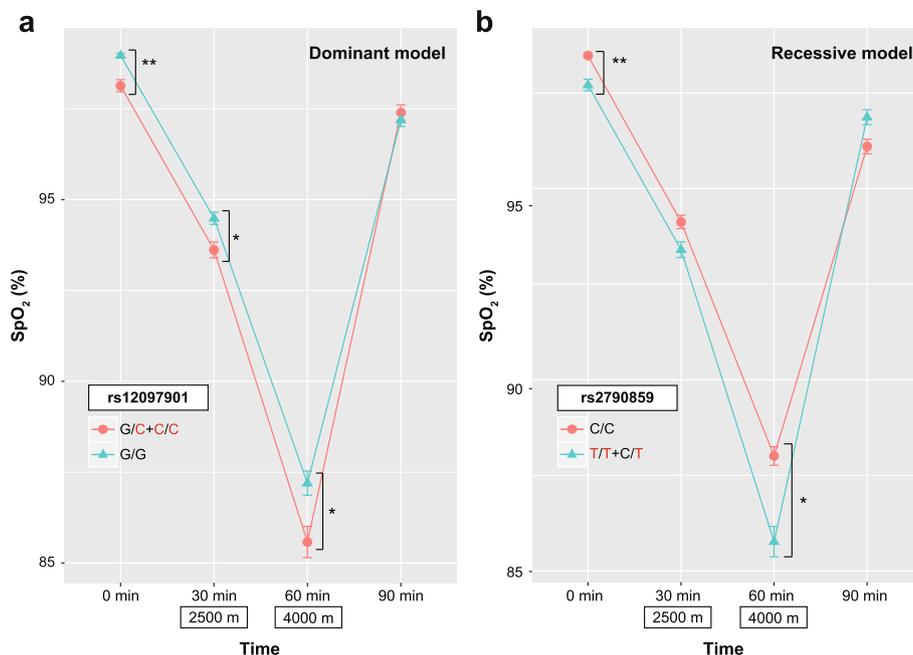
In the dominant model, the mean SpO<sub>2</sub> values from 0 to 60 min were affected by rs12097901 genotype: the mean values in the combined groups of homozygotes with derived allele (HL allele) and heterozygotes with ancestral and derived alleles were lower than those in homozygotes with ancestral allele (Table 2 and Fig. 3a). The difference in SpO<sub>2</sub> values at 0 min between the groups was significant ( $P = 0.003$ , by Welch's  $t$ -test). In contrast, the mean SpO<sub>2</sub> value at 90 min in the combined group was slightly higher than that in homozygotes with ancestral allele. Although the interaction between time and genotype

was not statistically significant, the  $P$  value was  $\sim 0.05$  ( $F_{(3, 132)} = 2.6$ ,  $P = 0.052$ , by two-way ANOVA).

In the recessive model, rs2790859 showed a similar effect to that of rs12097901 (Table 2 and Fig. 3b). The mean SpO<sub>2</sub> values from 0 to 60 min in the combined group of homozygotes with ancestral allele (HL allele) and heterozygotes with ancestral and derived alleles were lower in homozygotes with derived allele. The difference in SpO<sub>2</sub> values at 0 min was statistically significant ( $P = 0.008$ ), and the interaction between time and genotype was borderline significant ( $F_{(3, 132)} = 3.6$ ,  $P = 0.015$ ): the mean SpO<sub>2</sub> value at 90 min in individuals with HL alleles was higher than that in individuals with non-HL allele.

### Effects of estimated haplotypes or diplotypes on SpO<sub>2</sub> values

Using IMPUTE2 program, we estimated phased haplotypes comprising the five SNPs examined in all subjects. Haplotype and diplotype frequencies are shown in Additional file 4: Table S2. Here, we focus on haplotypes of rs12097901 and rs2790859 related to SpO<sub>2</sub> (Table 3). Haplotype "CT" (from left, rs12097901 and rs2790859) comprised HL alleles (HL haplotype), whereas haplotype "GC" included all non-HL alleles (non-HL haplotype). In the 46 subjects, the proportion of haplotype "GT" was the highest (32%), followed by "CT" (30%), "GC" (22%), and "CC" (16%). The proportions of diplotypes "CT/CT" (HL diplotype) and "GC/GC" (non-HL diplotype) were



**Fig. 3** Time series variations of SpO<sub>2</sub> at 28 °C in two groups of subjects differentiated by rs12097901 (a) and rs2790859 (b) genotypes. The ordinate axis represents SpO<sub>2</sub> and the abscissa axis represents time (min): 30 and 60 min are equivalent to altitudes of 2500 and 4000 m, respectively. Red characters represent highlander alleles. Line colors represent each group of genotypes. Data are presented as the mean  $\pm$  standard error of the mean. \* $P < 0.05$ ; \*\* $P < 0.01$  (by Welch's  $t$ -test)

**Table 3** Haplotype and diplotype frequencies of rs12097901 and rs2790859 in 46 Japanese subjects

Haplotype <sup>a</sup>	Observed number	Frequency	Diplotype <sup>a</sup>	Observed number	Frequency
<i>GT</i>	29	0.32	<i>CC/GT</i>	13	0.28
<i>CT</i>	28	0.30	<i>CT/CT</i>	9	0.20
<i>GC</i>	20	0.22	<i>GT/GC</i>	6	0.13
<i>CC</i>	15	0.16	<i>GC/GC</i>	5	0.11
			<i>CT/GC</i>	4	0.09
			<i>CT/GT</i>	4	0.09
			<i>GT/GT</i>	3	0.07
			<i>CC/CT</i>	2	0.04

Haplotype consists of two candidate SNPs related to SpO<sub>2</sub> (from left, rs12097901 and rs2790859)

<sup>a</sup>Highlander alleles are shown in italics (see Table 1)

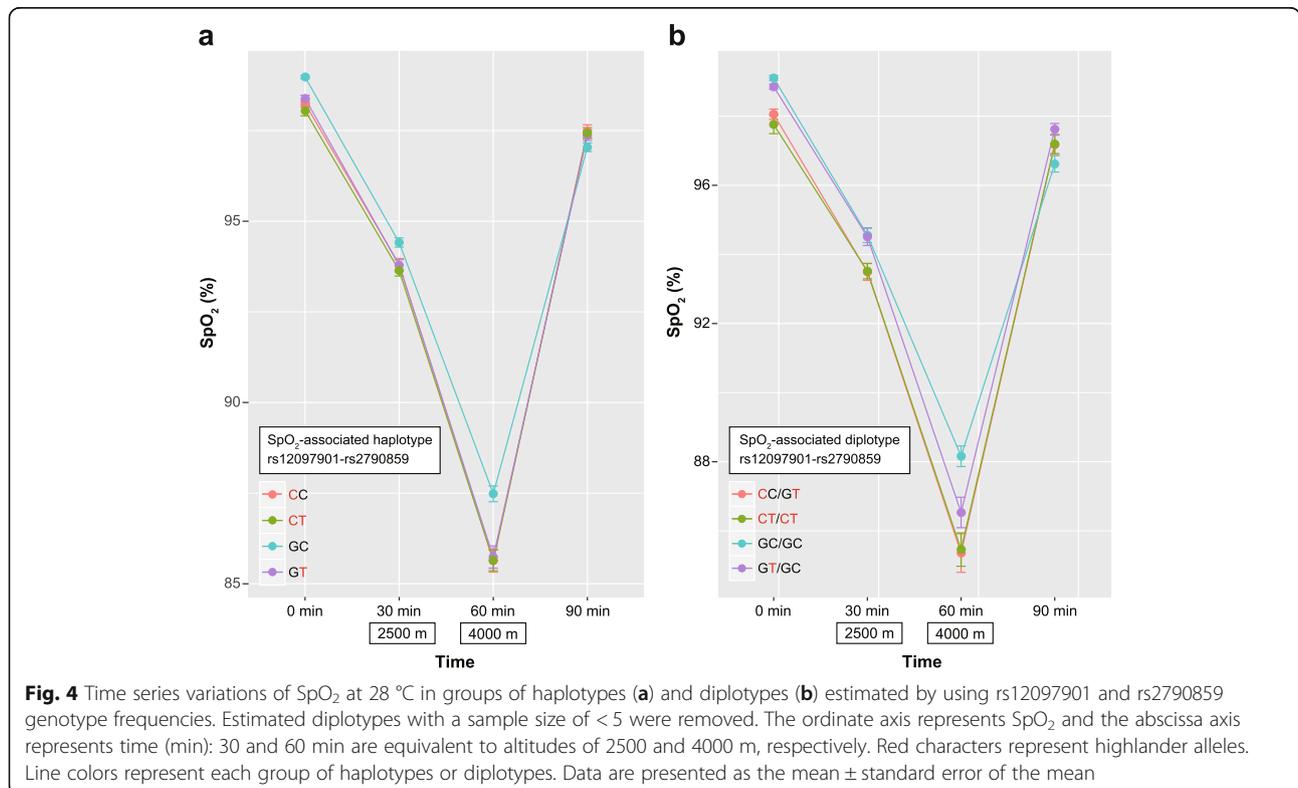
20 and 11%, respectively. We compared time series variations of SpO<sub>2</sub> levels among individuals with different haplotypes (Fig. 4a) and diplotypes (Fig. 4b). The mean SpO<sub>2</sub> values of non-HL haplotype “GC” from 0 to 60 min were higher than those of other haplotypes, whereas SpO<sub>2</sub> levels of non-HL haplotype at 90 min were lower than those in individuals with other haplotypes. Similarly, SpO<sub>2</sub> values of diplotypes in subjects with non-HL haplotypes (“GC/GC” and “GT/GC”) from 0 to 60 min were higher than those in subjects with other diplotypes without non-HL haplotypes.

These results were consistent with the results of the genotype-based association studies.

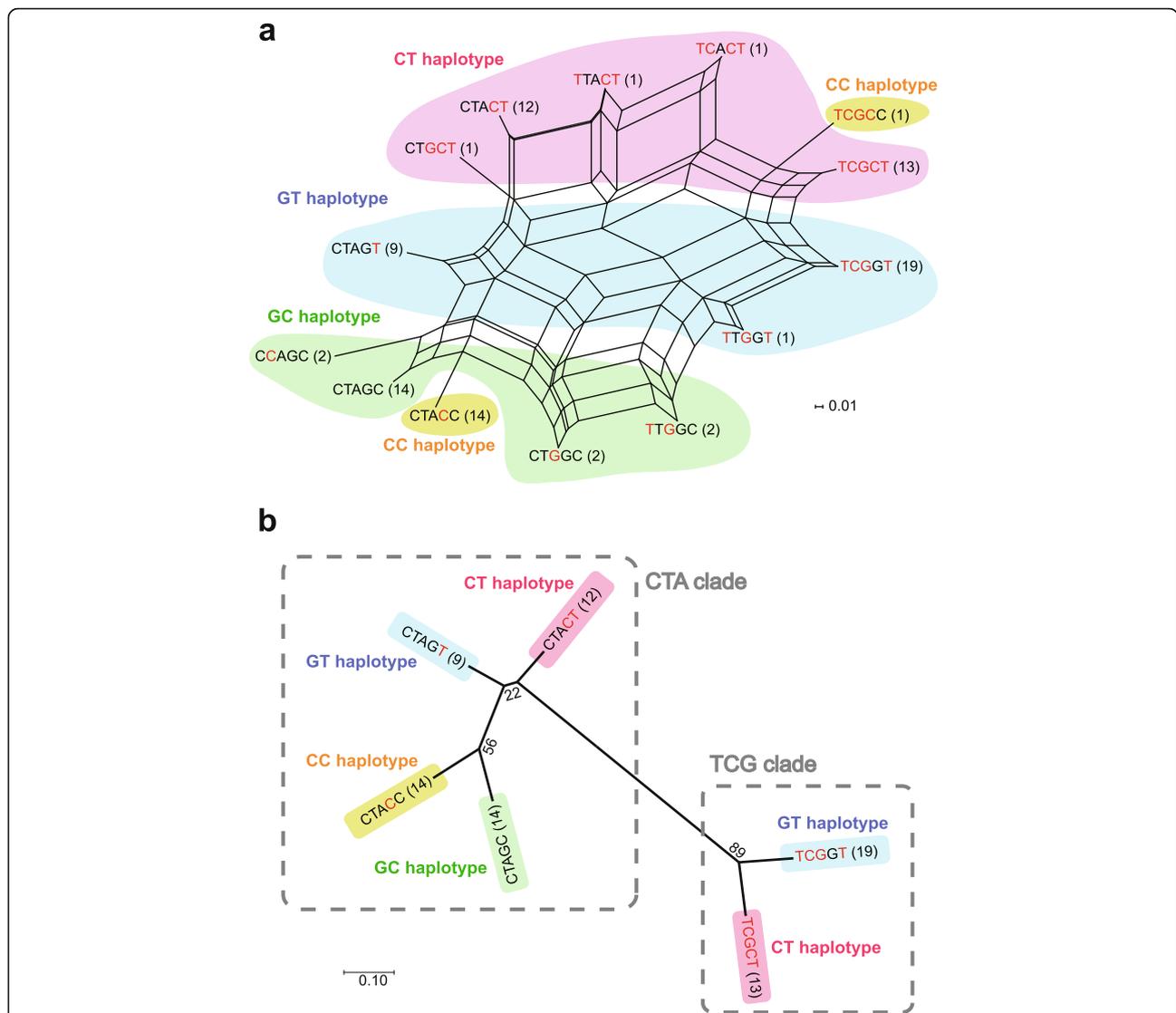
Using estimated haplotype sequences comprising all SNPs examined, we generated a haplotype network (Fig. 5a). This network indicated that HL or non-HL haplotypes of two SpO<sub>2</sub>-associated SNPs (i.e., rs12097901 and rs2790859) and the other three SNPs were mixed with each other, suggesting that the chromosomal region around *EGLN1* has undergone recombination events. To simplify the relationships between the haplotypes, an unrooted NJ tree was constructed after removing haplotypes with a sample size of < 5 (Fig. 5b). The topology of this tree showed that all haplotype groups of SpO<sub>2</sub>-associated SNPs (GT, CT, GC, and CC) were included in CTA clade with non-HL alleles of the other three SNPs (from left, rs480902, rs479200, and rs2808611), whereas TCG clade with HL alleles had two SpO<sub>2</sub>-associated haplotypes only (CT and GT). Thus, the diversity of CTA lineage appears to be higher than that of TCG lineage.

**Relationships between SpO<sub>2</sub> and other hematological or pulmonary phenotypes**

We applied linear regression model to examine correlations between SpO<sub>2</sub> and the following hematological or pulmonary traits: MV, PR, PI, and respiratory exchange ratio ( $R = VCO_2/VO_2$ ). This analysis indicated that SpO<sub>2</sub> levels negatively correlated with PR ( $r^2 = 0.16, P = 0.003$ ;



**Fig. 4** Time series variations of SpO<sub>2</sub> at 28 °C in groups of haplotypes (a) and diplotypes (b) estimated by using rs12097901 and rs2790859 genotype frequencies. Estimated diplotypes with a sample size of < 5 were removed. The ordinate axis represents SpO<sub>2</sub> and the abscissa axis represents time (min): 30 and 60 min are equivalent to altitudes of 2500 and 4000 m, respectively. Red characters represent highlander alleles. Line colors represent each group of haplotypes or diplotypes. Data are presented as the mean ± standard error of the mean

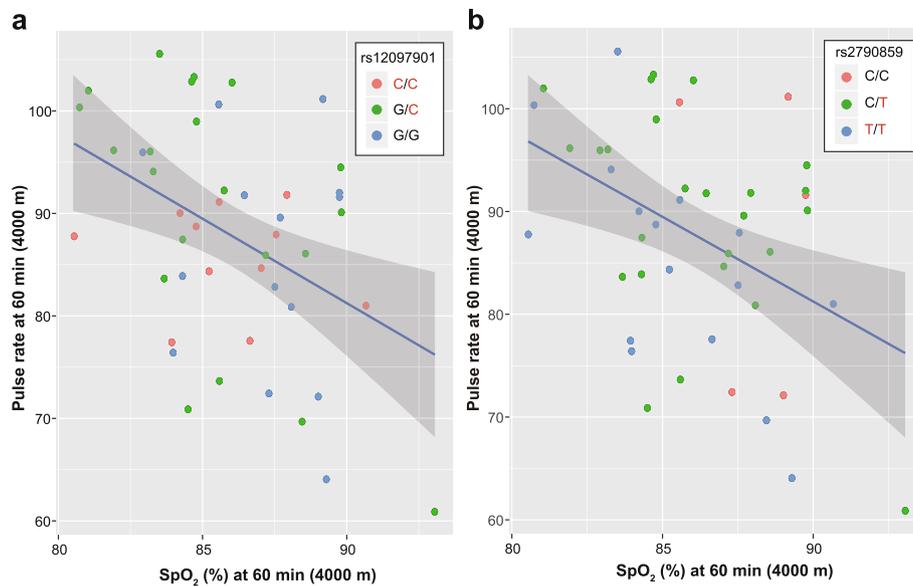


**Fig. 5** Neighbor-net network (a) and neighbor-joining tree (b) of the estimated haplotypes of five *EGLN1* SNPs. The observed number of haplotypes is indicated in parentheses. The neighbor-joining tree was constructed after removing haplotypes with a sample size of < 5, and the number of internal nodes represents bootstrap values (1000 replications). Red characters represent highlander alleles of each SNP. CT, GT, GC, and CC haplotypes were defined by rs12097901 and rs2790859 alleles. CTA and TCG clades were defined by rs480902, rs479200, and rs2808611 alleles

Fig. 6), whereas there was no significant correlation between SpO<sub>2</sub> and other phenotypes examined (Additional file 5: Figure S3, Additional file 6: Figure S4 and Additional file 7: Figure S5). For regression analysis, we performed ANCOVA to assess the differences in the regression coefficients between rs12097901 or rs2790859 genotypes and in the adjusted mean values. However, there were no significant differences in either slope ( $F_{(2, 40)} = 0.19-2.20, P = 0.124-0.825$  for rs12097901;  $F_{(2, 40)} = 0.04-0.98, P = 0.384-0.961$  for rs2790859) or intercept ( $F_{(2, 42)} = 0.59-2.97, P = 0.062-0.557$  for rs12097901;  $F_{(2, 42)} = 0.11-2.17, P = 0.127-0.894$  for rs2790859) of the regression lines (Fig. 6 and Additional file 5: Figure S3, Additional file 6: Figure S4 and Additional file 7: Figure S5).

### Discussion

In the present study, the HL alleles of rs12097901 and rs2790859 were associated with a shorter SpO<sub>2</sub> latency, compared to that in subjects harboring non-HL alleles. In addition, the recovery of SpO<sub>2</sub> from 60 to 90 min of the hypobaric hypoxia experiment was greater in HL allele carriers than in subjects with non-HL alleles. These results suggest that HL allele carriers are more susceptible to acute hypobaric hypoxia. SpO<sub>2</sub> levels in individuals with HL alleles from 0 to 60 min were lower than those in subjects with non-HL alleles. It is possible that the effect of HL alleles on SpO<sub>2</sub> levels under hypobaric hypoxia conditions may lead to the development of acute mountain sickness (AMS). AMS is defined by the



**Fig. 6** Relationship between SpO<sub>2</sub> and pulse rate at 60 min (equivalent to the altitude of 4000 m) for rs12097901 (a) and rs2790859 (b). Red characters represent highlander alleles. Blue line and gray band represent a regression line and its 95% confidence interval, respectively. Circle colors indicate genotypes of each SNP. The mean slope of the regression line was  $-1.65$ . Linear regression analysis showed significant correlation ( $r^2 = 0.16$ ,  $P = 0.003$ ). ANCOVA showed no significant differences in the regression coefficient between genotypes ( $F_{(2, 40)} = 1.31$ ,  $P = 0.282$  for rs12097901;  $F_{(2, 40)} = 0.19$ ,  $P = 0.824$  for rs2790859) and in the adjusted mean values ( $F_{(2, 42)} = 0.61$ ,  $P = 0.546$  for rs12097901;  $F_{(2, 42)} = 2.17$ ,  $P = 0.127$  for rs2790859)

presence of headache and at least one of the following symptoms: gastrointestinal upset (anorexia, nausea, or vomiting), fatigue or weakness, dizziness or lightheadedness, and difficulty falling asleep [30]. Although there is no reliable diagnostic modality or physical finding, arterial oxygen saturation (SaO<sub>2</sub>) could be a simple indicator of AMS development [31]. In the present study, the  $r^2$  value between rs12097901 and rs2790859 was low ( $r^2 = 0.38$ ), whereas the  $D'$  value was high ( $D' = 0.85$ ).  $D'$  tends to easily show high values in the absence of any of the four possible haplotypes, whereas  $r^2$  tends to show low values when SNPs with rare alleles are included in haplotypes being examined [32]. In addition, a small sample size could lead to biased LD estimates. Therefore, the effective independence of each SNP from SpO<sub>2</sub> variation remains unclear. Irrespective of the independence, genotyping of the two SNPs related to SpO<sub>2</sub> might be useful for the prevention and early diagnosis of AMS in Japanese population.

A previous study on comparisons of SaO<sub>2</sub> levels among three highlander populations (Tibetan, Ethiopian, and Andean populations) showed the lowest SaO<sub>2</sub> levels in Tibetans [1]. However, even though SaO<sub>2</sub> levels of Tibetans are low, they maintain high resting ventilation [33] and may counteract low levels of arterial oxygen by high levels of circulating nitric oxide that increases blood flow [9]. In the present study, however, we did not reveal any remarkable hematological or pulmonary features

in Japanese individuals with HL alleles of the two candidate SNPs related to SpO<sub>2</sub>. One may hypothesize that HL alleles in Japanese population are not adaptive to acute and short hypobaric hypoxia. At the same time, HL allele might be a risk factor for acute hypobaric hypoxia in Japanese population, because Japanese individuals with the HL allele do not appear to have remarkable hematological or pulmonary features that can possibly counteract low levels of SpO<sub>2</sub>. To validate this hypothesis, further analyses are required to examine effects of SNPs on long-term acclimatization and hemoglobin concentrations in lowlanders with lowlander ancestry under hypoxia.

In Indian lowlander populations, rs480902 and rs479200 were associated with the expression level of *EGLN1* and SaO<sub>2</sub> levels [34]. In addition, rs480902 also showed a significant relation to SaO<sub>2</sub> in Han Chinese AMS patients [35]. However, such association for rs480902 and rs479200 was not observed in the present study. The discrepancy of the relations between SNPs and the phenotypes might be due to differences in the genetic or physiological background among ethnic groups. Alternatively, given that the association of rs480902 with SaO<sub>2</sub> was not observed in the Han AMS patients, but not in the control group, the significance of the association might be detectable in Japanese AMS patients. The T allele could be a risk factor for the AMS, according to the results of Buroker et al. [35]. The frequencies of this allele tend to be high in highlanders and are consistent with low SpO<sub>2</sub> levels of HL alleles. Further

investigation is required to determine the association between rs480902 and SpO<sub>2</sub> in Japanese populations.

The haplotype network analysis suggested that recombination event had occurred around the *EGLN1* genomic region. The haplotype phase of the three SNPs with strong LDs (rs480902, rs479200, and rs2808611) was relatively conserved, and in the network, all SpO<sub>2</sub>-associated haplotype lineages were included in only the CTA lineage of the three SNPs. Furthermore,  $r^2$  values between rs2808611 and rs12097901 were extremely low ( $r^2 = 0.03$ ) for the physical distance. These results suggest that recombination breakpoints may be located around the boundary between these two SNPs. In fact, according to datasets of JPT in the 1000 Genomes Project database, the three SNPs with strong LDs were in an LD block outside rs12097901. In the haplotype network, the CTA clade included CTA, CTG, and CCA haplotypes, whereas TCG haplotype was only in the TCG clade. The difference in the number of nucleotides between CTA and the other haplotypes in the clade was only one, and the mutations (CTA → CTG and CTA → CCA) were of the transition type (In general, the frequency of transition mutation is higher than that of transversion mutation). Furthermore, the observed number of CTG and CCA haplotypes was small ( $\leq 2$ ). Consequently, CTG and CCA haplotypes may have arisen from the CTA haplotype by a single nucleotide mutation and have been maintained in the gene pool of the Japanese cohort. The paraphyletic clades of CTA haplotypes might be formed after recombination has occurred in the region including the other two SNPs (rs12097901 and rs2790859). Although we have not revealed the association between haplotypes consisting of the three SNPs with strong LDs and pulmonary or hematological phenotypes, the divergence of CTA and TCG haplotypes might influence the differences in the function of phenotypes that depend on the HIF pathway.

In Tibetan populations, rs12097901 and rs186996510 are in a strong LD because of strong positive selection (rapid expansion of the HL haplotype), and these SNPs may have epistatic effects on enzymatic activity [14]. In the Japanese cohort examined, the two SNPs were not in LD ( $r^2 = 0.03$ ), and there were two individuals with the HL haplotype (heterozygote and homozygote of the HL haplotype). However, these individuals did not show remarkably high or low quantitative values of pulmonary and hematological phenotypes, although the homozygote showed low SpO<sub>2</sub> values. Given that the nucleotide substitution at rs12097901 alters an amino acid of the PHD2 protein (encoded by *EGLN1*), and that pulmonary and hematological phenotypes are determined by polygenic inheritance, rs12097901 genotypes possibly have some effects on the enzymatic activity even if the effects are small.

There were several limitations in the present study. First, the effects of long-term acclimatization of the lowlander population on SpO<sub>2</sub> changes remain unclear. Second, the sample size is not sufficient to detect relations between SNPs and phenotypes with a high statistical power, although it is difficult to obtain data from a large number of subjects due to experimental limitations. Third, the functional relevance of the two candidate SNPs to SpO<sub>2</sub> variations remains unclear. A number of SNPs in various genes can contribute to high-altitude adaptation in highlanders. In future, genome-wide association study and functional analysis are required to evaluate the overall effect of genetic polymorphisms on inter-individual differences in physiological responses to hypobaric hypoxia.

## Conclusions

We showed that rs12097901 and rs2790859 were associated with SpO<sub>2</sub> variations and SpO<sub>2</sub> latency in an acute hypobaric hypoxia experiment in a Japanese cohort. SpO<sub>2</sub> values of individuals with HL alleles of these two SNPs were lower than those of individuals with non-HL alleles. It is possible that HL alleles may confer susceptibility to acute hypobaric hypoxia in Japanese populations.

## Additional files

**Additional file 1: Figure S1.** Linkage disequilibrium plot of 1504 biallelic sites across a ~399-kb genomic region at 1q42.2, based on datasets from JPT (Japanese in Tokyo, Japan) in the 1000 Genomes Project database. The position of SNPs examined is shown by the green arrow. The haplotype block was defined by the method of Gabriel et al. [36]. The genetic variants with the minor allele frequency of < 0.01 were removed from the analysis. (PDF 219 kb)

**Additional file 2: Table S1.** Time series variations of pulmonary and hematological phenotypes in individuals with different genotypes. (XLSX 15 kb)

**Additional file 3: Figure S2.** Relationship between BMI and SpO<sub>2</sub> at sea level in 46 Japanese male students. (PDF 101 kb)

**Additional file 4: Table S2.** Haplotype and diplotype frequencies of five SNPs around *EGLN1* in 46 Japanese subjects. (PDF 71 kb)

**Additional file 5: Figure S3.** Relationship between SpO<sub>2</sub> and minute ventilation at 60 min (equivalent to 4000 m) for rs12097901 (a) and rs2790859 (b). Red characters represent highlander alleles. Blue line and gray band represent a regression line and its 95% confidence interval, respectively. Circle colors indicate genotypes of each SNP. The mean slope of the regression line was -0.018. Linear regression analysis showed no significant correlation ( $r^2 = 0.001$ ,  $P = 0.818$ ). ANCOVA also showed no significant differences in the regression coefficient between genotypes ( $F_{(2, 40)} = 1.21$ ,  $P = 0.309$  for rs12097901;  $F_{(2, 40)} = 0.04$ ,  $P = 0.961$  for rs2790859) and in the adjusted mean values ( $F_{(2, 42)} = 2.97$ ,  $P = 0.062$  for rs12097901;  $F_{(2, 42)} = 0.11$ ,  $P = 0.894$  for rs2790859). (PDF 230 kb)

**Additional file 6: Figure S4.** Relationship between SpO<sub>2</sub> and perfusion index at 60 min (equivalent to 4000 m) for rs12097901 (a) and rs2790859 (b). Red characters represent highlander alleles. Blue line and gray band represent a regression line and its 95% confidence interval, respectively. Circle colors indicate genotypes of each SNP. The mean slope of the regression line was 0.011. Linear regression analysis showed no

significant correlation ( $r^2 = 0.011$ ,  $P = 0.481$ ). ANCOVA also showed no significant differences in the regression coefficient between genotypes ( $F_{(2, 40)} = 2.20$ ,  $P = 0.124$  for rs12097901;  $F_{(2, 40)} = 0.69$ ,  $P = 0.509$  for rs2790859) and in the adjusted mean values ( $F_{(2, 42)} = 1.29$ ,  $P = 0.285$  for rs12097901;  $F_{(2, 42)} = 1.08$ ,  $P = 0.348$  for rs2790859). (PDF 227 kb)

**Additional file 7: Figure S5.** Relationship between SpO<sub>2</sub> and respiratory exchange ratio at 60 min (equivalent to 4000 m) for rs12097901 (a) and rs2790859 (b). Red characters represent highlander alleles. Blue line and gray band represent a regression line and its 95% confidence interval, respectively. Circle colors indicate genotypes of each SNP. The mean slope of the regression line was 0.002. Linear regression analysis showed no significant correlation ( $r^2 = 0.012$ ,  $P = 0.474$ ). ANCOVA also showed no significant differences in the regression coefficient between genotypes ( $F_{(2, 40)} = 0.19$ ,  $P = 0.825$  for rs12097901;  $F_{(2, 40)} = 0.98$ ,  $P = 0.384$  for rs2790859) and in the adjusted mean values ( $F_{(2, 42)} = 0.59$ ,  $P = 0.557$  for rs12097901;  $F_{(2, 42)} = 0.17$ ,  $P = 0.840$  for rs2790859). (PDF 215 kb)

### Abbreviations

AMS: Acute mountain sickness; ANCOVA: Analysis of covariance; ANOVA: Analysis of variance; BMI: Body mass index; GLM: Generalized linear model; HIF: Hypoxia-inducible factor; HL allele: Highlander allele; HWE: Hardy-Weinberg equilibrium; JPT: Japanese in Tokyo, Japan; LD: Linkage disequilibrium; MV: Minute ventilation; NJ: Neighbor-joining; PHD: Prolyl hydroxylase domain-containing; PI: Perfusion index; PR: Pulse rate; SaO<sub>2</sub>: Arterial oxygen saturation; SNP: Single nucleotide polymorphism; SpO<sub>2</sub>: Percutaneous arterial oxygen saturation; VCO<sub>2</sub>: Carbon dioxide output; VO<sub>2</sub>: Oxygen consumption

### Acknowledgements

The authors deeply thank all the subjects for their participation in the study. We owe special thanks to Dr. Keita Ishibashi, Dr. Takafumi Katsumura, and Dr. Yuka Egashira for providing valuable comments or sample management.

### Funding

This work was supported by JSPS KAKENHI Grant-in-Aid for Young Scientists (B) Grant Numbers 15K18621 (to Y.Y.) and 15K18623 (to T.N.) and Grant-in-Aid for Scientific Research (A) Grant Numbers 17H01453 and 26251050 (to S.W. and T.N.) from the Ministry of Education, Culture, Sports, Science and Technology.

### Availability of data and materials

Scientifically motivated request for data sharing will be considered with reviewing ethical committee of our institute.

### Authors' contributions

YY contributed to the design of the study, analysis and interpretation of the data, and drafted the manuscript. TN contributed to the acquisition and interpretation of the data as well as the revision of the manuscript. MM contributed to the acquisition of the data and the revision of the manuscript. SW contributed to the conception of the study and acquisition of the data as well as the revision of the manuscript. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

This study was approved by the Ethical Committee of Kyushu University Institutional Review Board for Human Genome/Gene Research and by the Committee on Human Research of Mie University Graduate School of Medicine. All subjects were informed of the purpose and procedures of the study and possible risks involved. Written informed consent was obtained from all subjects.

### Consent for publication

All participants gave written informed consent for publication after a complete explanation of this study.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Department of Human Functional Genomics, Advanced Science Research Promotion Center, Organization for the Promotion of Regional Innovation, Mie University, 1577 Kurima-machiya, Tsu, Mie 514-8507, Japan. <sup>2</sup>Department of Public Health, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan. <sup>3</sup>Department of Human Science, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan.

Received: 28 December 2017 Accepted: 22 March 2018

Published online: 06 April 2018

### References

- Beall CM. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integr Comp Biol.* 2006;46:18–24.
- Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, et al. Natural selection on *EPAS1* (*HIF2alpha*) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci U S A.* 2010; 107:11459–64.
- Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, et al. Genetic evidence for high-altitude adaptation in Tibet. *Science.* 2010;329:72–5.
- Huerta-Sánchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature.* 2014;512:194–7.
- Hackinger S, Kraaijenbrink T, Xue Y, Mezzavilla M, Asan, van Driem G, et al. Wide distribution and altitude correlation of an archaic high-altitude-adaptive *EPAS1* haplotype in the Himalayas. *Hum Genet.* 2016; 135:393–402.
- Yang J, Jin Z-B, Chen J, Huang X-F, Li X-M, Liang Y-B, et al. Genetic signatures of high-altitude adaptation in Tibetans. *Proc Natl Acad Sci U S A.* 2017;114:4189–94.
- Maes C, Carmeliet G, Schipani E. Hypoxia-driven pathways in bone development, regeneration and disease. *Nat Rev Rheumatol.* 2012;8:358–66.
- Haase VH. Regulation of erythropoiesis by hypoxia-inducible factors. *Blood Rev.* 2013;27:41–53.
- Bigham AW, Lee FS. Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes Dev.* 2014;28:2189–204.
- Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, et al. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* 2010;6:e1001116.
- Motoi M, Nishimura T, Egashira Y, Kishida F, Watanuki S, Quintana-Murci L, et al. Relationship between mitochondrial haplogroup and physiological responses to hypobaric hypoxia. *J Physiol Anthropol.* 2015;35:12.
- Nishimura T, Katsumura T, Motoi M, Oota H, Watanuki S. Experimental evidence reveals the *UCP1* genotype changes the oxygen consumption attributed to non-shivering thermogenesis in humans. *Sci Rep.* 2017;7:5570.
- Aggarwal S, Negi S, Jha P, Singh PK, Stobdan T, Pasha MAQ, et al. *EGLN1* involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. *Proc Natl Acad Sci U S A.* 2010;107:18961–6.
- Lorenzo FR, Huff C, Myllymäki M, Olenchock B, Swierczek S, Tashi T, et al. A genetic mechanism for Tibetan high-altitude adaptation. *Nat Genet.* 2014; 46:951–6.
- R Core Team. R: a language and environment for statistical computing. 2016. <https://www.r-project.org/>.
- RStudio Team. RStudio: integrated development environment for R: RStudio, Inc; 2015. <http://www.rstudio.com/>.
- Snedecor GW, Cochran WG. *Statistical methods.* 6th ed. Ames: Iowa State University Press; 1967.
- Howie BN, Donnelly P, Marchini J, Rioux J, Xavier R, Taylor K, et al. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5:e1000529.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21:263–5.
- Warnes G, with contributions from Gorjanc G, Leisch F, Man M. *Genetics: Population Genetics.* 2013. R package version 1.3.8.1. <https://cran.r-project.org/package=genetics>.
- The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467:1061–73.

22. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555–7.
23. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res*. 2018;46:D754–61.
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
25. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987;4:406–25.
26. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol*. 2004;21:255–65.
27. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2006;23:254–67.
28. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
29. Korndewal MJ, Geurts Van Kessel WMH, Jak LG, Uiterwaal CSPM, Rövekamp MH, Van Der Ent CK. Influence of obesity on nocturnal oxygen saturation in young children. *Eur J Pediatr*. 2012;171:1687–93.
30. Roach RC, Bartsch P, Hackett PH, Olez O, Lake Louise AMS scoring consensus committee. The Lake Louise acute mountain sickness scoring system. In: Sutton JR, Houston CS, Coates G, eds. *Hypoxia and molecular medicine*. Burlington, Vt. Burlington: Queen City Press; 1993. p. 272–4.
31. Gallagher SA, Hackett PH. High-altitude illness. *Emerg Med Clin North Am*. 2004;22:329–55.
32. Marroni F, Pinosio S, Zaina G, Fogolari F, Felice N, Cattonaro F, et al. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (*CAD4*) gene. *Tree Genet Genomes*. 2011;7:1011–23.
33. Beall CM, Strohl KP, Blangero J, Williams-Blangero S, Almasy LA, Decker MJ, et al. Ventilation and hypoxic ventilatory response of Tibetan and Aymara high altitude natives. *Am J Phys Anthropol*. 1997;104:427–47.
34. Mishra A, Mohammad G, Thinlas T, Pasha MAQ. *EGLN1* variants influence expression and SaO<sub>2</sub> levels to associate with high-altitude pulmonary oedema and adaptation. *Clin Sci*. 2013;124:479–89.
35. Buroker NE, Ning X-H, Zhou Z-N, Li K, Cen W-J, Wu X-F, et al. *EPAS1* and *EGLN1* associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibetan plateau. *Blood Cells Mol Dis*. 2012;49:67–73.
36. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296:2225–9.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

