

# Fetal Gene Variants Associated With Birth Weight Protection in a Native High Altitude Ladakhi Population

**Sushil Bhandari**

Royal College of Surgeons in Ireland

**Padma Dolma**

Sonam Norboo Memorial Hospital

**Mitali Mukerji**

Institute of Genomics and Integrative Biology

**Bhavana Prasher**

Institute of Genomics and Integrative Biology

**Hugh Montgomery**

University College London Institute for Human Health

**Dalvir Kular**

King's College Hospital

**Vandana Jain**

All India Institute for Medical Sciences

**Vatsla Dadhwal**

All India Institute for Medical Sciences

**David J Williams**

University College London Institute for Women' Health

**Aniket Bhattacharyaa**

Institute of Genomics and Integrative Biology

**Edmund Gilbert**

Royal College of Surgeons in Ireland

**Gianpiero L Cavalleri**

Royal College of Surgeons in Ireland

**Sara L Hillman** (✉ [sara.hillman@ucl.ac.uk](mailto:sara.hillman@ucl.ac.uk))

University College London Institute for Women' Health



---

## Research Article

**Keywords:** fetal growth restriction (FGR), ZBTB38, ZFP36L2, HMGA2, CDKAL1, PLCG1, Ladakhi neonates

**Posted Date:** October 14th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-931753/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

Pathological low birth weight 'fetal growth restriction' (FGR) is an important predictor of adverse obstetric outcomes including stillbirth. It is more common amongst native lowlanders when gestating in the hypoxic environment of high altitude, whilst populations who have resided at high altitude for many generations are relatively protected. Genetic study of pregnant populations at high altitude allows for exploration of the hypoxic influence on FGR pathogenesis.

Pregnant women were recruited from Sonam Norboo Memorial Hospital, Ladakh between February 2017-January 2019 in this study. Principal component, admixture and genome wide association analysis (GWAS) were applied on umbilical cord blood DNA samples from 316 neonates, to explore ancestry and the genetic influence on low birth weight.

Our findings support Tibetan ancestry in the Ladakhi population, with subsequent admixture with neighboring Indo-Aryan populations. Fetal growth protection was evident in Ladakhi neonates. Seven loci from five different genomic regions (*ZBTB38*, *ZFP36L2*, *HMGA2*, *CDKAL1*, *PLCG1*) previously associated with birthweight, were likewise similarly associated here.

In summary, the Ladakhi population show evidence of enrichment of variants in genes that may help mitigate altitude-associated fetal growth restriction, supporting novel biological pathways and therapeutic targets for FGR, worthy of further investigation.

## Introduction

Birth weight is a complex trait driven by metabolic, vascular and immune interactions between mother and fetus. Successful placental and fetal growth leads to appropriate birth weight, reduced neonatal morbidity and ultimately success of a species [1–3]. Fetal growth restriction (FGR), defined as an inability for a fetus to reach its growth potential, leads to lower birth weight and compromised neonatal survival [1, 4–6].

Studies support a genetic role in influencing birth weight [7]. Genome wide association studies (GWAS) have identified approximately 70 single nucleotide polymorphisms (SNPs) of robust influence [8], although these studies have been primarily based around European ancestry cohorts of normal birth weight. Understanding the genetic contribution to *in-utero* fetal growth, particularly where significant genetic adaptation may have protected birth weight in populations at greater risk of FGR, might suggest new interventions to mitigate FGR in other situations.

Although FGR is more prevalent in those residing above 1500m (high altitude, where oxygen availability is reduced) [4, 9, 10], populations who have resided at high altitude for many generations appear to be protected [11, 12]. For example, at altitudes above 3000m, babies born to Tibetan women are more than 500g heavier than those born to native lowland Han Chinese [13]. Similarly, in La Paz, Bolivia (3,600m), native Andean babies are born heavier than their European counterparts [14].

To date, genetic studies of birth weight at high altitude have primarily focused on maternal genotypes. The maternal *PRKAA1* gene locus (coding for *AMPK*, a central regulator of cellular energy metabolism) has been associated with birth weight and maternal uterine artery diameter in high altitude Andean residents [15]. Less is known about fetal genotype and its association with birth weight at high altitude in Tibetan ancestry populations.

Ladakh, in the Jammu and Kashmir region of India, lies between the Karakoram and Himalayan mountain ranges. The Ladakhi population that have been resident at  $\geq 3400$  meters altitude for many generations. The term “La-dvags” in Tibetan means “land of high passes” and “Ladakh” is the Persian spelling [16]. The highland Ladakh region connects South Asia with the Tibetan plateau via the ancient trade corridor, “the Silk Road”. Genetic adaptation to the hypoxia of high altitude has now been well described [17]. The Ladakhi population, although sharing similar linguistic, cultural and religious practices with Tibetans [16, 18], are very poorly studied at the level of population structure and genetic selection. Sonam Norboo Memorial Hospital in Leh [3,540m], provides maternity care for the region of Ladakh. It has a very high institutional birth rate of more than 90%<sup>19</sup> making it a unique site to study a pregnant population at high altitude.

We hypothesized that the fetal genotype of native Ladakhis would be enriched for gene variants which protect *in-utero* growth despite the adverse effects of high altitude gestation. We first performed a genomic structural survey of the population. Then, in those subjects identified as distinctly Ladakhi from the genomic survey, we investigated fetal genetic elements that associated with birth weight.

## Methods

Pregnant women and their partners were recruited from Sonam Norboo Memorial (SNM) Hospital, Leh [altitude 3,540m] from antenatal clinics and the ultrasound department, over a 2-year period (February 2017-January 2019). Ethical permissions were granted for the study from the Indian Health Ministry’s Screening committee (HMSC) on the 7th September 2016, the Office of the Chief Medical Officer Leh (3rd August 2016), the All India Institute for Medical Sciences, and the University College London Research Ethics Committee (3634/002). All methods were performed in accordance with the relevant guidelines and regulations and informed consent was obtained from all participants and/or their legal guardians. The research was performed in accordance with the Declaration of Helsinki.

Pregnancies were included if both parents were aged over 18 years and unrelated (first cousin or closer), were having a singleton pregnancy and planned to have their baby at the hospital. Women were recruited if the estimated due date could be calculated from last menstrual period or dating ultrasound. Those pregnancies where the fetus had a clinically obvious fetal structural or chromosomal abnormality were excluded. Women completed a questionnaire documenting their social (including nutritional), family, obstetric and medical histories (including smoking, alcohol, chronic medical problems, and medications). These data have already been reported by our group [28]. Geographical ancestry was recorded for more than three generations for both parents.

Following birth and delivery of the placenta and umbilical cord, at least three 1ml aliquots of residual whole umbilical cord blood with a separate serum sample were frozen immediately at  $-80^{\circ}\text{C}$ . Information concerning the birth process (mode of delivery) and the neonatal characteristics routinely collected by the hospital (sex, weight, head circumference, crown-heel length and APGAR score) were recorded.

Cord blood from studied offspring (and whole blood from their parents for future testing) was then shipped to Delhi and genomic DNA extracted at the Institute for Genomics and Integrative Biology, Delhi using Qiagen kits and validated, standardized protocols. Common bi-allelic SNPs were genotyped in samples using Illumina Global Screening Array SNP-microarray technology. Genotype assignment from the microarray fluorescence data was performed using Illumina's Genome Studio software.

We combined the genotypes of 316 Ladakhi individuals with reference individuals from surrounding populations living at high and low altitude. Human Genome Diversity project<sup>35</sup> datasets were included for the following different populations; Burusho, Han, Hazara, Indo Aryan, Japanese, Yakut Siberian. Similarly other population datasets included in this study were Munda [36], Sherpa [21], Tibetan [37] and Tibeto Burman [38]. These individual cohorts were merged using autosomal SNPs found in common between all separate datasets. This merged dataset was then processed through several quality control steps using the software PLINK 1.9 [39, 40]. We included only individuals or SNPs that had  $< 5\%$  missing genotypes, SNPs with a minor allele frequency (MAF)  $>1\%$ , and SNPs with a HWE at significance of  $>1e^{-6}$ . Furthermore, we calculated closely related individuals using King<sup>41</sup>, pruning one individual from a pair related by 2nd degree or closer. This left a final dataset of 1,413 individuals. Due to the variety of genotyping platforms between the different population references, the final set of common SNPs contained 60,280 markers.

We calculated principal components using PLINK [39], first pruning SNPs in linkage disequilibrium using PLINK's `-indep-pairwise` command using a window size of 1000, moving 50 SNPs, and using an  $r^2$  threshold of 0.2 – leaving 41,718 SNPs in approximate linkage.

We conducted a genetic clustering analysis of Ladakhi populations with its neighboring populations using ADMIXTURE v1.2 [42], estimating individual ancestry proportions in an unsupervised analysis. The same cohort assembled in PCA was used for ADMIXTURE analysis, i.e., non-missing autosomal SNPs, individuals filtered for relatedness, and SNPs pruned for linkage disequilibrium. Unsupervised ADMIXTURE analysis was carried out over  $k=2-4$  populations.

We further investigated evidence of admixture within the Ladakh population's history by applying  $f$ -statistics<sup>20</sup> testing the strength of evidence that the Ladakh population is admixed between pairs of neighboring populations using the  $f_3$  implementation within *ADMIXTOOLS 2*

(<https://uqrmaie1.github.io/admixtools/index.html> - a manuscript describing *ADMIXTOOL 2* is currently under preparation by its authors). The same cohort of individuals and SNPs that were used for PCA and ADMIXTURE analysis was used for  $f_3$ -statistic work.

We detected Runs of Homozygosity (ROH) [43] in Ladakhi individuals and compared population-averages to individuals from neighboring lowland and highland populations using PLINK's `-homozyg` command, with

the following specific parameters; `-homozyg -homozyg-window-snp 50 -homozyg-snp 50 -homozyg-kb 1500 -homozyg-gap 1000 -homozyg-density 50 -homozyg-window-missing 5 -homozyg-window-het 1`. This was carried out on a subset of the individuals described in principal component, ADMIXTURE, and  $f_3$  analyses, selecting individuals who were genotyped on SNP-microarray chips with an overlap of common SNPs >100K in number to faithfully detect ROH. This subset included individuals from the Indian Indo-Aryan, Ladakhi, Sherpa, Tibeto-Burman Bhutanese, Tibeto-Burman Nepalese, or Tibetan population labels. Common SNPs between these individuals were filtered according to the same parameters as in the PCA, leaving 117,044 common SNPs.

Genome-wide association tests were conducted using 601,887 genotyped SNPs. Linear regression was applied, implemented in PLINK v1.9 with the `-linear` command. We performed an additive genetic model adjusting sex and the first four principal components obtained from genome-wide SNP data. Gene variants associated with birthweight in the Ladakhi were compared to variants in the Global Biobank Engine [23] to confirm their association with birthweight in other populations.

Sample size calculations using algorithms incorporated in GWApower [44] were based on mean birth weights obtained from audit data collected in Leh prior to the study commencing. Calculations identified the study had an 80% power to detect a variant that would explain 11.4% (or greater) of the variability observed in birth weight given a discovery sample size of 300 neonates and alpha level of  $5 \times 10^{-8}$ .

## Results

Over a two year period (February 2017-January 2019), all women who presented for antenatal care at Sonam Norboo Memorial Hospital in Leh were approached to take part in the study. In total, 316 families were recruited. Maternal and fetal baseline characteristics are summarized in Table 1. Mean birth weight was 3.18kg (mean birth weight centile 44.5th) in term individuals (>37 weeks' gestation). According to intergrowth charts (<https://intergrowth21.tghn.org>), 14% of infants were born at less than the 10th centile (which defines a small for gestational age [SGA] newborn) and only 7.8% of all infants were born with a low birth weight (LBW, defined as less than 2.5kg). The average age for pregnant women was 28.9 years and 37.9% (120/316) were primigravida. The majority of subjects (306/316; and 96.8%) reported >3 generations of Ladakhi ancestry. The ten women who reported a Tibetan ancestry all gave birth at term and with a mean birth weight of 3.6kg. Preterm birth was infrequent (15/316 [4.7%] infants were born before 37 weeks). Amongst these, birth weight was, as expected, lower (mean 2.4kg) than those born at term (mean 3.2kg) [ $P=0.0001$ ] although birth weight centile was similar (52nd vs 45th  $P=0.324$ ) suggesting appropriate growth for gestational age. Recognizing that being preterm would confound birth weight if used in isolation, for further analysis we instead used birth weight centiles as a more useful measure of growth potential, given that it adjusts for gestational age.

Table 1

Maternal and fetal characteristics summarized as mean [min/max] with birth weight of infant (kg) divided into those born at term gestations of >37 weeks and those born preterm at <37 weeks, birth weight centile (%) derived from Intergrowth and those born at <10th centile recorded (defined as small for gestational age) and ancestry by history (Ladakhi and Tibetan).

Characteristic	Term>37 <42+1 weeks n=301	Preterm<37 weeks n=15	Mean n=316
Maternal Age (years)	29.1 [18/40]	26 [20/35]	29.0[18/40]
Weight (kg)	57.0 [37/92]	56.9 [40/82]	57.0[37/92]
Hb (g/dl)	12.7 [10.4/14.8]	12.3 [7.5/17]	12.3 (7.5/17)
Primiparous n (%)	111 (36.9)	6 (40)	117 (37)
Birth weight (kg)	3.18[1.98-4.40]	2.45[1.5-3.2]	3.15[1.5-4.4]
Birth weight centile %	44.5 [0.13/98.7]	49 [1.5/99]	44.9 [0.13/99.9]
<10th centile (%)	42 (14)	2 (13.3)	44 (13.9)
Tibetan by history kg (centile)			3.6 (70) [n=10]
Ladakhi by history kg (centile)			3.17 (43.9) [n=306]

### Reconstruction Of Ladakhi Population History

To optimize our GWAS, we first sought to understand the population history of the Ladakhi population. We compared Ladakhi with other surrounding populations based on major language groups including Tibeto-Burman (Tibetans, Sherpa, Ladakhi), Indo-European (Indo-Aryan, Hazara), Austro-Asiatic (Munda) and isolate languages (Bursushaski) (Fig. 1) using Principle Component Analysis (PCA). To understand the nature and extent of Himalayan ancestry in the Ladakhi population, we compared them with high altitude populations (Sherpa, Tibetans) and other Tibeto-Burman populations adjacent to the foothills of the Himalayas in Nepal and Bhutan.

In the PCA, the Ladakhi individuals project together, forming a genetic cluster that is distinct from the neighboring Indo-Aryan Indian population. This Ladakhi cluster forms a genetic cline between the two termini of principal component 1, which separates Indo-Aryan Indian individuals from individuals with Himalayan or East Asian ancestry. The second principal component appears to separate out Han (East Asian) individuals from Sherpa (high altitude), where Ladakhi individuals are placed towards the high-altitude-associated Sherpa population alongside individuals from Himalayan regions such as Tibet, Nepal, and Bhutan. The intermediate position of Ladakhi individuals on principal component 1 is suggestive of admixture between Indo-Aryan and Himalayan populations. The position of Ladakhi individuals on principal

component 2 suggests the Himalayan genetic group that the Ladakhi share affinity to is Tibetan or Sherpa with no evidence in PCA of admixture with Han-related East Asian ancestries.

We further contextualized our population structure analysis using the maximum-likelihood estimation of individual ancestries using ADMIXTURE (Fig. 2). At  $k=2$ , the two ancestral components, are maximized in either Himalayan or the East Asian Han populations (in red), and Indo-Aryan Indians (green). We observe that Ladakhi individuals are modelled with a slightly higher proportion of the East Asian ancestry than South Asian ancestral component, in agreement with their average location on principal component 1. At  $k=3$ , one ancestry component (represented by blue in Fig. 2) is maximized in the East Asian Han, the second (green) maximized in Sherpa, and the third (red) in Indo-Aryan Indians. The majority average component in Ladakhi individuals is the Sherpa-maximised ancestry (52.4%), then the Indo-Aryan-maximised component (32.3%), suggesting the majority of Ladakhi are closer to Tibeto-Burmans than Indo-Aryans. At  $k=4$  an ancestry component separates ancestry which is maximized in Tibetan-Bhutanese individuals from other East Asian sources. This component (in purple) is also found in appreciable proportions in other high-altitude populations. Ladakhi individuals also show appreciable proportions of this ancestry component common in both Nepalese Tibeto-Burman and Tibetans. However, the blue ancestry component at  $k=4$  in Ladakhi individuals continues to support admixture with other South Asian population such as Hazara, Burusho or Indo-Aryan Indians. Thus, Ladakhi appear genetically closer to Tibeto-Burman speakers of Himalayan region than Indo-Aryan populations of South Asia. Long-term isolation in remote highland, practices of endogamy culture and a founder effect might have added greatly to the formation of distinct genetic cluster in the Ladakhi agreeing with PCA result (Fig. 1). In summary, the data supported Ladakhi as a genetically distinct Himalayan population closer to Tibeto-Burman population and Tibetan populations than to Indo-Aryan.

We next performed two additional analyses to further confirm that the Ladakhi are admixed between lowland Indo-Aryan and highland Tibeto-Burman sources. Firstly, we leveraged  $f_3$ -statistics [20] in the form of  $f_3(X; A, B)$  where  $X$  is tested for evidence of admixture between sources  $A$  and  $B$ , where a negative  $f_3$ -statistic is indicative of admixture. We placed Ladakhi as  $X$ , and tested combinations of lowland/Indo-Aryan populations (Burusho, Indo\_Aryan, Hazara, Munda) as  $X$  and highland/East or North Asian populations (Han, Japanese, Sherpa, Tibetan, TB\_Bhutanese, TB\_Nepalese, Yakut\_Siberian) as  $Y$ , reporting those  $f_3$  results with an absolute  $Z$  score  $>3$  (Fig. 3). The strongest evidence of admixture (most negative  $f_3$ -statistic) is between Tibetan or Sherpa sources and Indo-Aryan or Burusho sources – supportive of results from PCA. Additionally, we performed Runs of Homozygosity (ROH) detection using PLINK, comparing Ladakhi to a subset of neighbouring lowland and highland references. We detect elevated ROH (See Supplementary Fig. S1) in the high altitude Tibeto-Burman and Sherpa populations, agreeing with previous estimates [21], but only modest levels of ROH in the Ladakhi, more comparable to the general Tibetan or Indo-Aryan labelled individuals. These modest levels of ROH would be consistent with admixture between different ancestries.

### **Birth Weight Analysis On Ladakhi Subjects**

In order to determine if there were any genome-wide significant predictors of birth weight in the Ladakhi population, we performed a GWAS of birth weight in the 176 individuals, who were genetically identified as



Ladakhi from the population study (see Supplementary Fig S2). We did not find any signals that were genome-wide significant after correction for multiple testing (i.e.  $p < 5 \times 10^{-8}$ ). However, when we looked in the tail of the association statistics from our birth weight GWAS (i.e.  $p = 1 \times 10^{-4}$  to  $1 \times 10^{-7}$  uncorrected), we noted the presence of multiple variants previously associated with traits relevant to high altitude adaptation, including body mass index (rs10968576 in *LINGO2*) and blood related traits (rs16893892 in *RP5-874C20.3*, rs2298839 in *AFP*, rs9261425 in *TRIM31* and rs362043 in *TUBA8*) [23].

Next, in order to determine whether the genetic architecture of birth weight in the Ladakhi population overlaps with that observed in lowland populations, we sought the association with birth weight in Ladakhis of the 70 genetic signals previously associated with birth weight in lowland populations [8]. Overall, 32 of these 70 signals were either directly genotyped or captured through linkage disequilibrium ( $r^2 > 0.8$ ) in our Ladakhi dataset. Seven of these 32 signals were significantly associated ( $P < 0.05$ , uncorrected) with birth weight in the Ladakhi GWAS study (see Table 2). These variants mapped to five different birth weight associated genes (*ZBTB38*, *ZFP36L2*, *HMGA2*, *CDKAL1* and *PLCG1*) with the direction of effect being consistent with the original discovery reporting [8] all seven cases.

Table 2

Single nucleotide polymorphism (SNP) and gene location along with P and Beta values of 176 babies born with a >3 generation ancestry in Leh Ladakh compared with UK biobank and GWAS\_Catolog

SNP	Mapped Gene	SNP_ Annotation	P-values			Beta values	
			Ladakhi	UK biobank	GWAS_Catolog	GWAS_Catolog	Ladakhi
rs6440006	<i>ZBTB38</i>	Upstream	0.03515	$1.099e^{-5}$	$4 \times 10^{-12}$	-0.021224	-0.0818
rs8756	<i>HMGA2</i>	3_prime_UTR	0.000078	$1.97e^{-138}$	$1 \times 10^{-19}$	-0.027582	-0.176
rs1351394	<i>HMGA2</i>	intronic	0.000078	$3.28e^{-137}$	$2 \times 10^{-33}$	-0.043	-0.176
rs7968682	<i>HMGA2</i>	intergenic	0.000078	$4.97e^{-130}$	$4 \times 10^{-60}$	-0.041831	-0.176
rs4952673	<i>ZFP36L2</i>	intergenic	0.04	NA-not available	$2 \times 10^{-11}$	-0.020476	-0.09315
rs35261542	<i>CDKAL1</i>	intronic	0.01864	NA	$3 \times 10^{-45}$	-0.040621	-0.08755
rs753381	<i>PLCG1</i>	missense variant	0.02988	NA	$3 \times 10^{-9}$	-0.01512	-0.0792

## Discussion

This study is the first to undertake a comprehensive fetal genotypic exploration of a high altitude population in relation to a birth weight phenotype. We showed that the Ladakhi are more closely related to Tibeto-Burman speaking populations than to the Indo-Aryan groups of South Asia, and provide clear evidence that Tibeto-Burman expansion occurred in North East India crossing the Himalayan range. [24]. The majority of Ladakhi form a common clustering PCA, indicating the presence of a Ladakhi specific genetic ancestry component that is intermediate between Tibeto-Burman and Indo-Aryan ancestries. Further analysis utilising *f*-statistics and ROH data further supports a demographic history of admixture between these sources. It suggests a distinct demographic history in the Ladakh region, plausible due to its close proximity with the Tibetan plateau. Lack of large overlap between Ladakhi genotype data and existing references limited haplotype analyses of Identity-by-Descent segments or “ChromoPainter”-based methods [25], which limited the power to analyse the demographic history of the population in more detail. Future work overcoming these technical limitations may provide further insights.

Individuals in Leh are born at a significantly higher birth weight than expected, based on existing literature for average birth weights in India at 2.8-3kg [26, 27] and seen from our previously published work [28]. One such mechanism of protection of birth weight is through genetic selection, as evidenced in other phenotypic traits of high altitude populations [17].

Birthweight is a complex trait with potentially conflicting parental inheritance patterns. As such, it is perhaps unsurprising that no one clear signal was identified at GWAS level, especially given the low numbers in the final study. Study of parental DNA and genetic signals in genes responsive to the hypoxia-responsive transcription factors *hypoxia inducible factors (HIFs)* in relation to offspring birth weight is an obvious follow up study to further explore HIF function and high altitude adaptation.

In relation to SNPs previously associated with birth weight by GWAS, we identified that rs1351394 of the *HMGA2* gene (encoding the high mobility group-A2 protein) was replicated in Ladakhi offspring. High-mobility group (HMG) proteins are ubiquitous nuclear proteins that binds to DNA and induce structural change in chromatin, thereby regulating gene expression [15]. *HMGA2* has been associated with human height and birth weight in lowland populations [29, 30] and with adipose mass variation in pigs [31], making it a biologically plausible functional candidate.

*ZBTB38* encodes a zinc finger transcription factor that binds methylated DNA to enhance or repressing transcription in a way which is complex and likely cell-dependent [32]. Its expression appears to play a role in skeletal development [30]. Recent GWAS have reported SNPs in this gene to be associated with human height- perhaps through upregulation of insulin growth factor 2, a potent fetal growth factor [33]. The derived allele frequency of the associated *ZBTB38* variant (rs6440006) is higher (> 15%) and more highly differentiated in Ladakhis compared with low altitude resident Han Chinese ( $F_{st} > 0.01$ ), supporting significant enrichment in Ladakhis residing at high altitude.

One of the strengths of this study relates to the accurate dating of gestation from last menstrual periods and early dating ultrasounds. This allowed for calculation of birth weight centiles which adjusts neonatal weight for gestational age and sex at birth, allowing for better correlation with pathological poor growth.

This study focused on fetal genotype, but it is recognized that maternal [34] and paternal genotype also have an effect. We have parental DNA, and it would be beneficial in the future to analyze inheritance patterns of these fetal gene variants of interest.

A limitation of the study was the overall number of subjects in the birth weight analysis. Despite most respondents describing themselves as ‘many generation’ Ladakhi, the cohort was more genetically diverse than anticipated. This limitation meant that our birth weight analysis focused only on 176 infants of the original cohort of 316 (55.7%). Our interpretations would be strengthened by the addition of more subjects, stratified according to their genomic ancestry.

Although birth weight is a complex multifactorial trait to correlate with high altitude ancestry, this study provides evidence that variants are enriched in genes associated with height and weight along with skeletal growth and development in Ladakhi infants who attain an optimal birth weight. Further studies are warranted to confirm this influence and better understand the functional implications of the genetic variants identified in relation to the hypoxic environment. This detailed knowledge may help inform new treatments in growth restriction, acting on development pathways intrinsic in reaching optimized *in-utero* growth.

## Conclusion

We found Tibetan ancestry in the Ladakhi population, and evidence of recent limited admixture with neighboring Indo-Aryan populations. We replicated in the Ladakhi population, the effects of a subset of variants known to predict birthweight in European-descent populations. Enriched variants were focused on body anthropomorphic traits and hematological parameters. Further genomic study of high altitude infants is indicated to validate these findings in conjunction with mechanistic study to gain insight in potential mechanisms of action that may be amenable to therapeutic intervention.

## Declarations

### Acknowledgments

This work would not have been possible without the tireless commitment of the staff and patients at Sonam Norboo Memorial Hospital, Leh Ladakh. The work has been supported by several hospital Medical Superintendents, in particular Dr Phunchok Wangchuk Kokluk and by the Chief Medical Officer, Dr Motup Dorje. This work was largely funded by a Wellcome Trust SEED trust award (to SH). A small UCL Global Health Challenges award was received by HM and SH and used for travel expenses. The Royal College of Obstetricians and Gynaecologists also awarded the Eden Travelling Fellowship to SH to develop the relationship with the maternity team at Sonam Norboo Memorial Hospital. HM and DJW are supported by the National Institute for Health Research’s Comprehensive Biomedical Centre at University College London Hospitals (UCLH BRC). SB was supported by a Government of Ireland Postdoctoral Fellowship from the Irish Research Council (GOIPD/2018/408).

### Author Contributions

HM, DK, SH and DJW conceived the study. PD led the study in Ladakh with assistance from SH and DK. VJ and VD completed the study in Delhi and liaised with IGIB for sample processing. MM, BP and AB processed blood samples and run analysis. SB led the combined analysis and population reconstruction with EG and GLC supervising. All authors have approved the submitted version and have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

### Additional Information

The author (s) declare no competing interests

### Data Availability

The datasets supporting the current study have not been deposited in a public repository yet but are available on request.

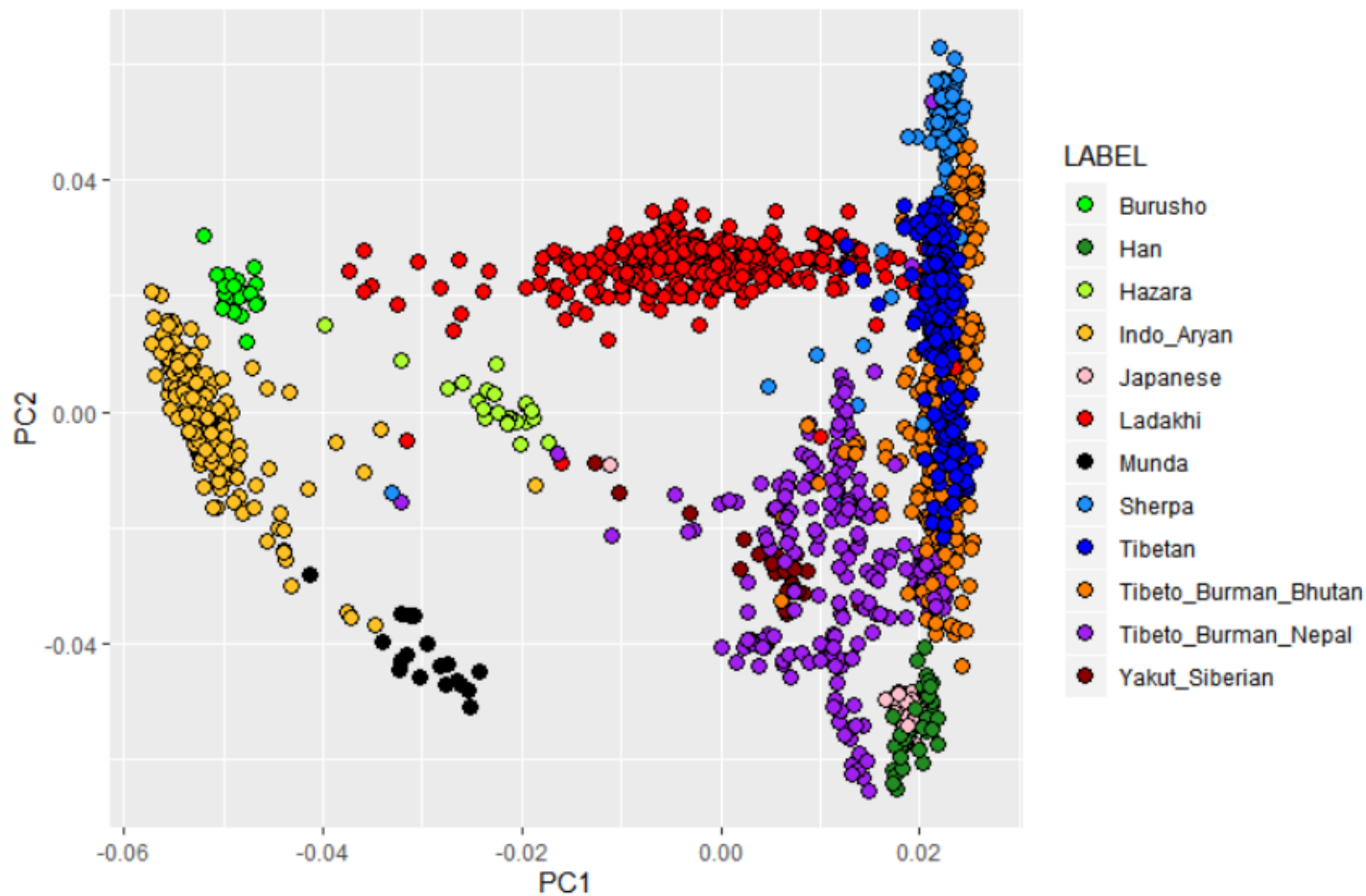
## References

1. Ludvigsson, J. F., Lu, D., Hammarstrom, L., Cnattingius, S. & Fang, F. Small for gestational age and risk of childhood mortality: A Swedish population study. *PLoS Med*, **15**, e1002717 <https://doi.org/10.1371/journal.pmed.1002717> (2018).
2. Hales, C. N. *et al.* Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ***303**, 1019-1022, 10.1136/bmj.303.6809.1019 (1991).
3. Gilbert, W. M. & Danielsen, B. Pregnancy outcomes associated with intrauterine growth restriction. *Am J Obstet Gynecol*, **188**, 1596–1599 <https://doi.org/10.1067/mob.2003.384> (2003).
4. Moore, L. G. Fetal growth restriction and maternal oxygen transport during high altitude pregnancy. *HAMB*, **4**, 141–156 (2003).
5. Barker, D. J., Osmond, C., Winter, P., Margetts, B. & Simmonds, S. J. Weight in infancy and death from ischaemic heart disease. *The Lancet*, **334**, 577–580 (1989).
6. McCowan, L. M., Figueras, F. & Anderson, N. H. Evidence-based national guidelines for the management of suspected fetal growth restriction: comparison, consensus, and controversy. *Am J Obstet Gynecol*, **218**, S855–S868 <https://doi.org/10.1016/j.ajog.2017.12.004> (2018).
7. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet*, **51**, 804–814 <https://doi.org/10.1038/s41588-019-0403-1> (2019).
8. Horikoshi, M. *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature*, **538**, 248–252 <https://doi.org/10.1038/nature19806> (2016).
9. Julian, C. G., Wilson, M. J. & Moore, L. G. Evolutionary adaptation to high altitude: a view from in utero. *Am J Hum Biol*, **21**, 614–622 <https://doi.org/10.1002/ajhb.20900> (2009).
10. Moore, L. G., Charles, S. M. & Julian, C. G. Humans at high altitude: hypoxia and fetal growth. *RespPhysiolNeurobiol*, **178**, 181–190 (2011).

11. Smith, C. The effect of maternal nutritional variables on birthweight outcomes of infants born to Sherpa women at low and high altitudes in Nepal. *Am J Hum Biol*, **9**, 751–763 (1997).
12. Moore, L. G., Zamudio, S., Zhuang, J., Sun, S. & Droma, T. Oxygen transport in Tibetan women during pregnancy at 3,658 m. *Am J Phys Anthropol*, **114**, 42–53 (2001).
13. Moore, L. G., Young, D., McCullough, R. E., Droma, T. & Zamudio, S. Tibetan protection from intrauterine growth restriction (IUGR) and reproductive loss at high altitude. *Am J Hum Biol*, **13**, 635–644 <https://doi.org/10.1002/ajhb.1102> (2001).
14. Haas, J. D., Stepcik, F. E., Beard, C. & Hurtado, J. L. Altitude, ethnic and sex differences in birth weight and length in Bolivia. *Hum Biol*, **52**, 459–477 (1980).
15. Hock, R., Furusawa, T., Ueda, T. & Bustin, M. HMG chromosomal proteins in development and disease. *Trends Cell Biol*, **17**, 72–79 <https://doi.org/10.1016/j.tcb.2006.12.001> (2007).
16. Jolden, T. & Rinchen Tundup, R. Cultural Relationship between the People of Ladakh and Tibet. *The Tibet Journal*, **43**, 65–71 (2018).
17. Bhandari, S. & Cavalleri, G. L. Population History and Altitude-Related Adaptation in the Sherpa. *Frontiers in physiology*, **10**, <https://doi.org/10.3389/fphys.2019.01116> (2019).
18. Rowold, D. J. *et al.* India: the land of high passes and genetic heterogeneity reveals a confluence of migrations. *Eur J Hum Genet*, **24**, 442–449 <https://doi.org/10.1038/ejhg.2015.80> (2016). Ladakh
19. Gutschow, K. D. P. Global policies and local implementation: Maternal mortality in Rural India Kappas M, Gross U, Kelleher D (eds) *Global Health: A Challenge for Interdisciplinary Research* 215-238(2012).
20. Patterson, N. *et al.* Ancient admixture in human history., **192**, 1065–1093 (2012).
21. Cole, A. M. *et al.* Genetic structure in the Sherpa and neighboring Nepalese populations. *BMC Genomics*, **18**, 102 (2017).
22. Horikoshi, M. *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet*, **45**, 76–82 <https://doi.org/10.1038/ng.2477> (2013).
23. McInnes, G. *et al.* Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics., **35**, 2495–2497 <https://doi.org/10.1093/bioinformatics/bty999> (2018).
24. Gayden, T. *et al.* The Himalayas as a directional barrier to gene flow. *Am J Hum Genet*, **80**, 884–894 <https://doi.org/10.1086/516757> (2007).
25. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS genetics*, **8**, e1002453 (2012).
26. Kumar, V. S. *et al.* New birth weight reference standards customised to birth order and sex of babies from South India. *BMC Pregnancy and Childbirth*, **13**, <https://doi.org/10.1186/1471-2393-13-38> (2013).
27. Marete, I. *et al.* Regional trends in birth weight in low- and middle-income countries 2013-2018. *Reproductive health*, **17**, <https://doi.org/10.1186/s12978-020-01026-2> (2020).
28. Dolma, P. *et al.* High-altitude population neonatal and maternal phenotypes associated with birthweight protection. *Ped Res*, 1-6 (2021).

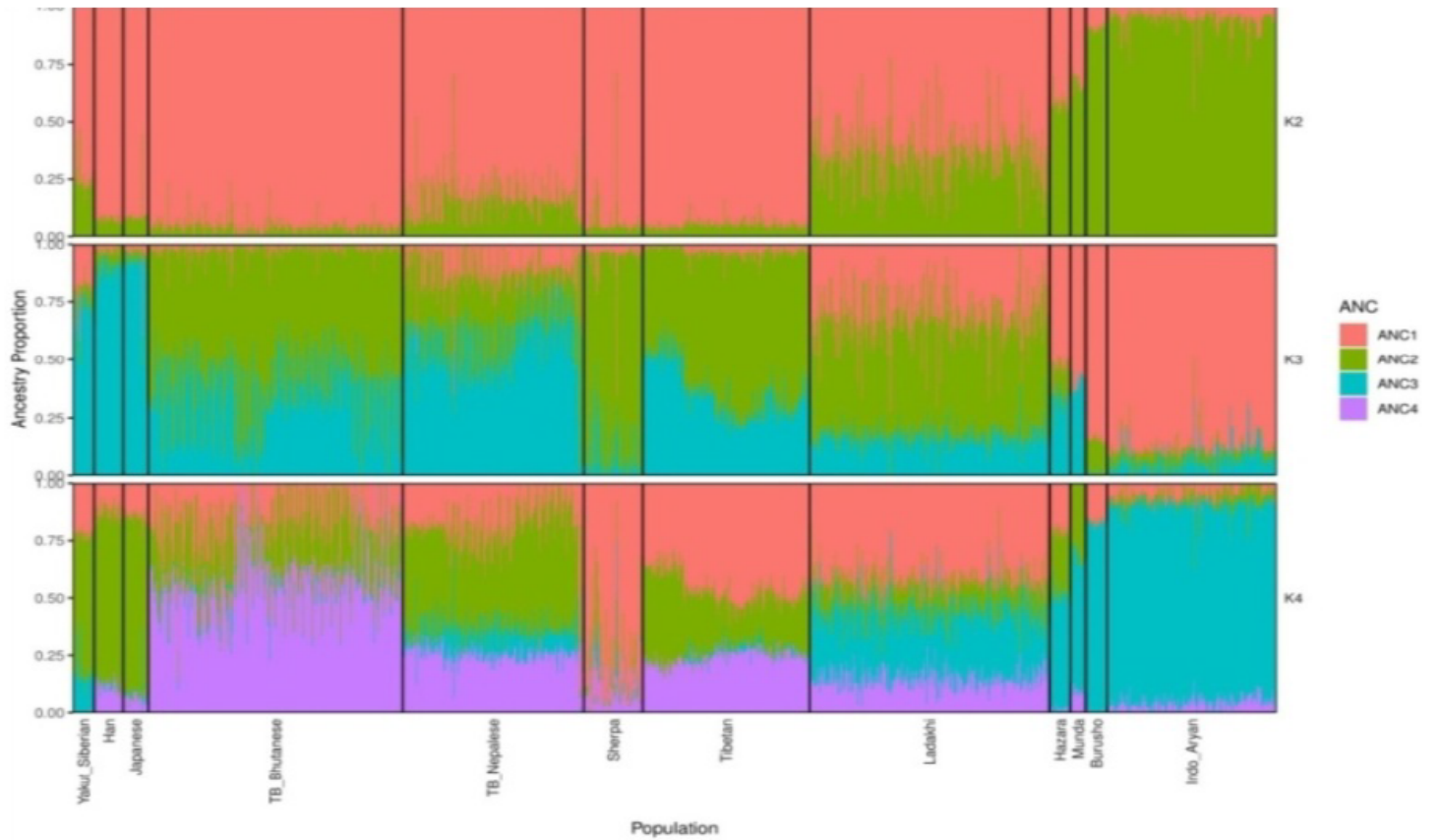
29. Weedon, M. N. *et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet*, **39**, 1245–1250 <https://doi.org/10.1038/ng2121> (2007).
30. Gudbjartsson, D. F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat Genet*, **40**, 609–615 <https://doi.org/10.1038/ng.122> (2008).
31. Kim, K. S. *et al.* Investigation of obesity candidate genes on porcine fat deposition quantitative trait loci regions. *ObesRes* **12**, 1981- 1994, [10.1038/oby.2004.249](https://doi.org/10.1038/oby.2004.249) (2004).
32. de Dieuleveult, M. & Miotto, B. D. N. A. Methylation and Chromatin: Role(s) of Methyl-CpG-Binding Protein ZBTB38. *Epigenet Insights*, **11**, <https://doi.org/10.1177/2516865718811117> (2018).
33. Lin, Y. J. *et al.* Association of human height-related genetic variants with familial short stature in Han Chinese in Taiwan. *Sci Rep*, **7**, <https://doi.org/10.1038/s41598-017-06766-z> (2017).
34. Hughes, A. E. *et al.* Fetal Genotype and Maternal Glucose Have Independent and Additive Effects on Birth Weight., **67**, 1024–1029 <https://doi.org/10.2337/db17-1188> (2018).
35. Cann, H. M. *et al.* A human genome diversity cell line panel., **296**, 261–262 <https://doi.org/10.1126/science.296.5566.261b> (2002).
36. Chaubey, G. *et al.* Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol Evol* **28**, 1013-1024, [10.1093/molbev/msq288](https://doi.org/10.1093/molbev/msq288) (2011).
37. Beall, C. M. *et al.* Natural selection on EPAS1 (HIF2 $\alpha$ ) associated with low hemoglobin concentration in Tibetan highlanders. *PNAS*, **107**, 11459–11464 (2010).
38. Arciero, E. *et al.* Demographic History and Genetic Adaptation in the Himalayan Region Inferred from Genome-Wide SNP Genotypes of 49 Populations. *Mol Biol Evol* **35**, 1916- 1933, [10.1093/molbev/msy094](https://doi.org/10.1093/molbev/msy094) (2018).
39. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, **81**, 559–575 (2007).
40. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, <https://doi.org/10.1186/s13742-015-0047-8> (2015).
41. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies., **26**, 2867–2873 <https://doi.org/10.1093/bioinformatics/btq559> (2010).
42. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, **19**, 1655–1664 (2009).
43. McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am J Hum Genet*, **83**, 359–372 <https://doi.org/10.1016/j.ajhg.2008.08.007> (2008).
44. Feng, S., Wang, S., Chen, C. C. & Lan, L. GWAPower: a statistical power calculation software for genome-wide association studies with quantitative traits. *BMC Genet*, **12**, <https://doi.org/10.1186/1471-2156-12-12> (2011).

## Figures



**Figure 1**

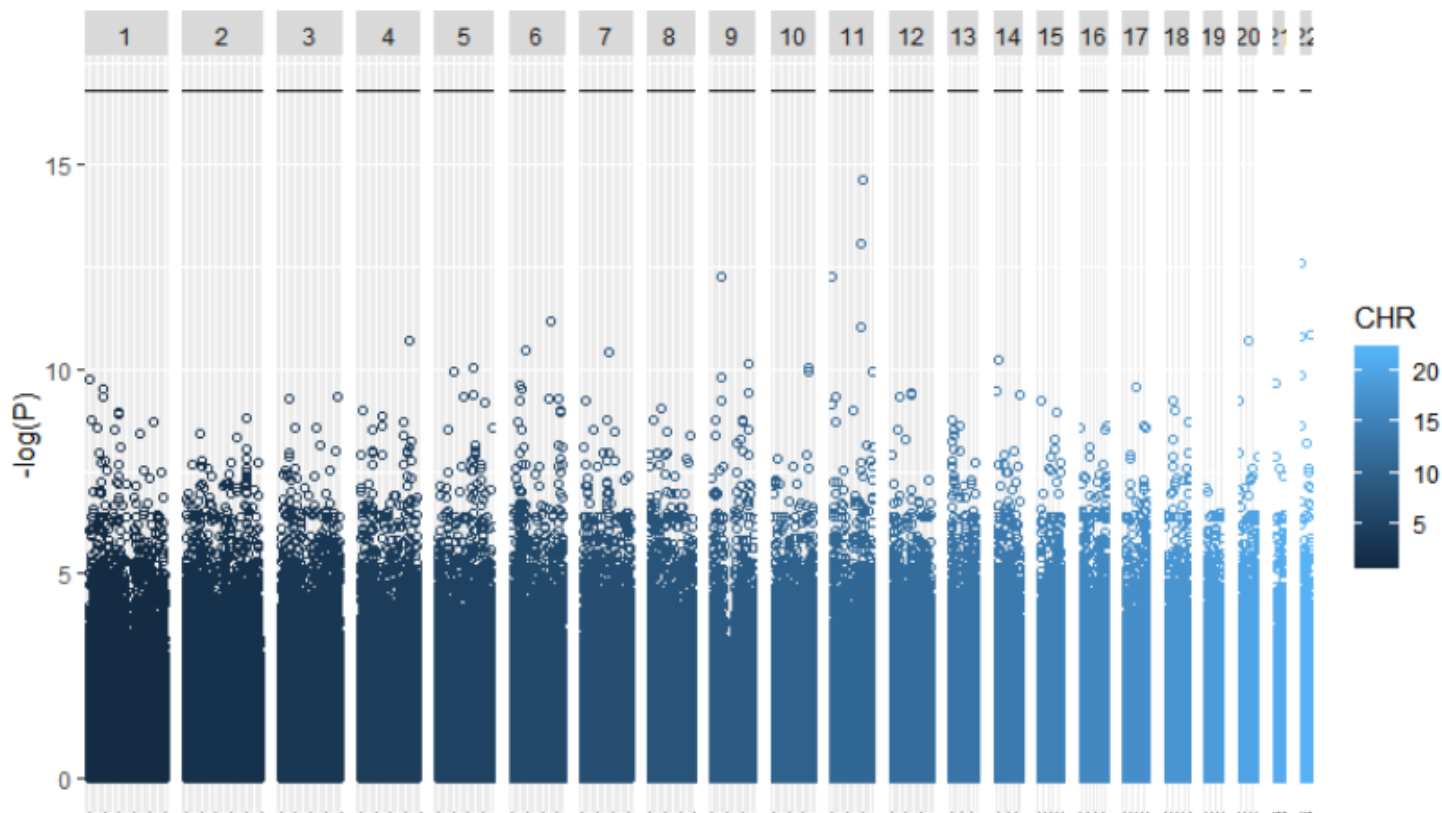
Principle Component Analysis (PCA) of Ladakhi population The first and second components of principal component analysis (PCA) with individual clusters labeled by different colours. PCA of Ladakhi and its neighboring populations are indicated in the Legend.



**Figure 2**

ADMIXTURE analysis of the Ladakhi population ADMIXTURE analysis for K=2 (top), K=3 (middle) and K=4 (bottom). ADMIXTURE analysis shows Ladakhi contain higher number of Tibeto-Burman ancestries and are relatively closer to Tibetans.





**Figure 3**

Evidence of admixture in Ladakhi between lowland or Indo-Aryan populations and highland or north/east Asian populations measured using  $f_3$  outgroup statistics. Along the x axis is the  $f_3$  statistic score, the more negative that statistic is the greater evidence of admixture. The error bars correspond to three standard errors of the estimated  $f_3$  statistic. Along the y axis are different pairs of putative source populations for an admixture event which creates the modern Ladakhi population.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementalinformation.docx](#)
- [SupplementaryFigures3.pdf](#)