Archaic introgression contributed to shape the adaptive modulation of angiogenesis and nitric oxide induction in human high-altitude populations from the Himalayas

Giulia Ferraretti[1*], Paolo Abondio[2*], Marta Alberti[1], Agnese Dezi[3], Phurba T. Sherpa[4], Paolo Cocco[5], Massimiliano Tiriticco[5], Marco di Marcello[5], Guido Alberto Gnecchi-Ruscone[1,6], Luca Natali[5,7], Angela Corcelli[8], Giorgio Marinelli[5], Davide Peluzzi[5], Stefania Sarno[1], Marco Sazzini[1,9#]

[1]Laboratory of Molecular Anthropology & Centre for Genome Biology, Dept. of Biological, Geological and Environmental Sciences, University of Bologna, Italy

[2]Dept. of Cultural Heritage, Ravenna Campus, University of Bologna, Italy

[3]Dept. of Precision and Regenerative Medicine and Ionic area, University of Bari Aldo Moro, Italy

[4]Mount Everest Summiters Club, Rolwaling, Dolakha, Nepal

[5]Explora Nunaat International, Montorio al Vomano, Teramo, Italy

[6]Current address: Dept. of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[7]Italian Institute of Human Paleontology, Rome, Italy

[8] Department of Translational Biomedicine and Neuroscience, University of Bari Aldo Moro, Bari, Italy

[9]Interdepartmental Centre Alma Mater Research Institute on Global Changes and Climate Change, University of Bologna, Italy

*These authors equally contributed

[#]For correspondence: marco.sazzini2@unibo.it (MS)

Abstract It is well established that several *Homo sapiens* populations experienced admixture with extinct human species during their evolutionary history. Sometimes, such a gene flow could have played a role in modulating their capability to cope with a variety of selective pressures, thus resulting in archaic adaptive introgression events. A paradigmatic example of this evolutionary mechanism is offered by the *EPAS1* gene, whose most frequent haplotype in Himalayan highlanders was proved to reduce their susceptibility to chronic mountain sickness and to be introduced in the gene pool of their ancestors by admixture with Denisovans. In this study, we aimed at further expanding the investigation of the impact of archaic

introgression on more complex adaptive responses to hypobaric hypoxia evolved by populations of Tibetan and Sherpa ancestry, which have been plausibly mediated by soft selective sweeps and/or polygenic adaptations rather than by hard selective sweeps. For this purpose, we used a combination of composite-likelihood and gene network-based methods to detect adaptive loci in introgressed chromosomal segments from Tibetan whole genome sequence data and to shortlist those enriched for Denisovan-like derived alleles that participate to the same functional pathways. According to this approach, we identified multiple genes putatively involved in archaic introgression events and that, especially as regards *EP300* and *NOS2*, have plausibly contributed to shape the adaptive modulation of angiogenesis and nitric oxide induction in high-altitude Himalayan peoples. These findings provided unprecedented evidence about the complexity of the adaptive phenotype evolved by these human groups to cope with challenges imposed by hypobaric hypoxia, offering new insights into the tangled interplay of genetic determinants that mediates the physiological adjustments crucial for human adaptation to the high-altitude environment.

Introduction

The scientific community currently agrees that the *Homo sapiens* species experienced admixture with extinct Hominins since traces of such inbreeding events are still detectable in the genomes of modern humans (Gouy et al., 2020). In fact, people belonging to non-African populations show 1–2% of Neanderthal ancestry (Green et al., 2010; Prüfer et al., 2014), while Melanesians and East Asians present 3% and 0.2% of Denisovan ancestry, respectively (Reich et al., 2010; Meyer et al., 2013; Prüfer et al., 2014; Racimo et al., 2017). Despite evidence supporting selection against introgressed alleles has been collected (Simonti et al., 2016; Racimo et al., 2017; McArthur et al., 2021), some of the genomic segments showing signatures ascribable to archaic introgression were also proved to have been targeted by positive selection in modern human populations, thus providing examples for the occurrence of Adaptive Introgression (AI) events (Racimo et al., 2017).

So far, several studies have indeed identified introgressed archaic alleles at high frequency in human genes involved in metabolism or in the response to environmental conditions, such as temperature, sunlight, and altitude (Prüfer et al., 2014; Benjamin et al., 2014; Sankararaman et al., 2014; Huerta-Sánchez et al., 2014; Gittelman et al., 2016; Racimo et al., 2017; Enard et al., 2018; Danneman et al., 2018). Moreover, some genes that play a role in immune responses to pathogens are found to be characterized by a similar pattern of variability (Laurent et al., 2011; Enard et al., 2018) and certain Neanderthal alleles have been shown to be associated with down-regulation of gene expression in brain and testes (McCoy et al., 2017; Racimo et al., 2017; Dannemann et al., 2018). These works collectively attest how genetic variants introduced in the human gene pool by admixture with archaic species can significantly impact our biology by possibly comporting modifications in the modulation of several functional pathways. In particular, the high frequency of some archaic alleles in protein-coding and/or regulatory genomic regions suggests a possible adaptive role for Neanderthal and/or Denisovan variants, pointing to a further evolutionary mechanism having potentially contributed to the processes of human biological adaptation to different environmental and cultural settings. By introducing new alleles in the gene pool of a given population, admixture in fact provides a very rapid opportunity for natural selection to act on it (Huerta-Sanchez et al., 2014; Jeong et al., 2014; Racimo et al., 2015; Hamid et al., 2021) and this is supposed to have likely occurred during the evolutionary history of *H. sapiens*, particularly after the last Out of Africa migration in the late Pleistocene (Israel et al., 2018; Vahdati et al., 2022). According to this view, gene flow from extinct Hominin species could have facilitated the adaptation of *H. sapiens* populations to peculiar Eurasian environments.

For instance, a Denisovan origin of the adaptive *EPAS1* haplotype, which confers reduced susceptibility to chronic mountain sickness to Tibetan and Sherpa highlanders (Beall et al., 2007; Bigham et al., 2010; Yi et al., 2010; Peng et al., 2011; Xu et al., 2011) is well established (Huerta-Sánchez et al., 2014; Zhang et al. 2021). However, the hard selective sweep experienced in high-altitude Himalayan populations by the *EPAS1* introgressed haplotype has been demonstrated to account only for an indirect aspect of their adaptive phenotype, which does not explain most of the physiological adjustments they evolved to cope with hypobaric hypoxia (Gnecchi-Ruscone et al., 2018). Therefore, how far gene flow between Denisovans

and the ancestors of Tibetan and Sherpa peoples facilitated the evolution of other key adaptive traits of these populations remains to be elucidated.

To fill this gap, and to overcome the main limitation of most approaches currently used to test for AI (i.e., inferring archaic introgression and the action of natural selection separately by means of different algorithms, which increases the risk of obtaining biased results due to confounding variables), we assembled a dataset of whole genome sequences (WGS) from 27 individuals of Tibetan ancestry living at high altitude (Cho et al., 2017; Jeong et al., 2018) and we analysed it using a composite-likelihood method specifically developed to detect AI events at once (Setter et al., 2020). Notably, this method was designed to recognize AI mediated by subtle selective events (as those involved in polygenic adaptation) and/or soft selective sweeps, which represent the evolutionary mechanisms that are supposed to have played a more relevant role than hard selective sweeps during the adaptive history of human groups characterized by particularly small effective population size, such as Tibetans and Sherpa (Gnecchi-Ruscone et al., 2018). Coupled with validation of the identified putative adaptive introgressed loci through the assessment of the composition of gene networks made up of DNA segments enriched for archaic derived alleles, as well as of the genetic distance between modern and archaic haplotypes, such an approach provided new evidence about the biological functions that have mediated high-altitude adaptation in Himalayan populations and that have been favourably shaped by admixture of their ancestors with Denisovans.

Results

Spatial distribution of genomic variation and ancestry components of Tibetan samples

After quality control (QC) filtering of the available WGS data, we obtained a dataset made up of 27 individuals of Tibetan ancestry characterized for 6,921,628 single nucleotide variants (SNVs). To assess whether this dataset represents a reliable proxy for the genomic variation observable in high-altitude Himalayan populations, we merged it with genome-wide genotyping data for 1,086 individuals of East-Asian ancestry belonging to both low- and high-altitude groups (Gnecchi-Ruscone et al., 2017; Landini et al.,

2021). We thus obtained an extended dataset including 231,947 SNVs, which was used to perform population structure analyses.

Results from ADMIXTURE and Principal Components Analysis (PCA) were found to be concordant with those described in previous studies (Jeong et al., 2014; Gnecchi-Ruscone et al., 2017; Gnecchi-Ruscone et al., 2018; Yang et al., 2021). According to the ADMIXTURE model showing the best predictive accuracy (K = 7) (Figure 1 – figure supplement 1), the examined WGS exhibited a predominant genetic component that was appreciably represented also in other populations speaking Tibeto-Burman languages, such as Tu, Yizu, Naxi, Lahu, and Sherpa (Figure 1A and Figure 1 – figure supplement 2). Such a component reached an average proportion of around 78% in individuals of Tibetan ancestry from Nepal included in the extended dataset, as well as of more than 80% in the subjects under investigation, who live in the Nepalese regions of Mustang and Ghorka (Figure 1A and 1B). This suggests that after their relatively recent migration in Nepalese high-altitude valleys, these communities might have experienced a higher degree of isolation and genetic drift with respect to populations that are still settled on the Tibetan Plateau, in which the same ancestry fraction did not exceed 64% (Figure 1A and 1B). Nevertheless, the overall ADMIXTURE profile of the considered WGS appeared to be quite comparable to those inferred according to genome-wide genotyping data for other Tibetan populations (Figure 1A and 1B, Figure 1 - figure supplement 2). Similarly, PCA pointed to the expected divergence of Tibetan and Sherpa high-altitude groups from the cline of genomic variation of East-Asian lowland populations (Abdulla et al., 2009; Jeong et al., 2014; Gnecchi-Ruscone et al., 2017; Zhang et al., 2017; Wang et al., 2022) (Figure 1C). Remarkably, the WGS under investigation clustered within the bulk of genome-wide data generated for other groups from the Tibetan Plateau, thus supporting their representativeness as concerns the overall genetic background of high-altitude Himalayan populations.

Detecting adaptive introgression signatures across functionally related loci

To identify genomic regions showing signatures putatively ascribable to AI events, we scanned Tibetan WGS with the *VolcanoFinder* algorithm and we computed the composite Likelihood Ratio (LR) and the -logα statistics for each polymorphic site (Setter et al., 2020). After normalization of the obtained LR values, we

searched for chromosomal intervals enriched for significant LR scores by retaining and merging those overlapping genomic windows falling within top 0.1% of the related distribution (see Materials and methods). We thus identified 53 significant DNA segments distributed across nine chromosomes and encompassing a total of 486 genes. To shortlist loci potentially contributing to the same adaptive trait, we then explored functional relationships among the identified candidate AI genes by exploiting information about protein-protein interactions annotated in the STRING database. Accordingly, the *EP300* gene resulted one of the most interconnected loci, along with *MAPK1* and *MAPK11*, and occupied a pivotal position in the reconstructed framework of functionally related genes (Figure 2 - figure supplement 1). Interestingly, it showed established associations with genes that have been previously proved to be characterized by signatures left by natural selection in multiple high-altitude populations and that play a role in cellular responses to hypoxia, such as *EPAS1*, *EGLN1*, *PPARA* and *NOS2* (Figure 2- figure supplement 1) (Bigham et al., 2010; Simonson et al., 2010; Peng et al., 2011; Crawford et al., 2017; Zheng et al., 2017; Horscroft et al., 2017; Liu et al., 2022). In fact, when functional pathways enriched for putative AI genes were investigated, *HIF-1 signalling* resulted one of the few significant pathways including almost all the loci mentioned above (Supplement Table 2).

Collectively, this core set of genes (except for *EGLN1*) was characterized by a consistent number of variants showing non-normalized and/or normalized significant LR scores, as well as by elevated values of -logα (Figure 2 and Figure 2 - figure supplement 2), which suggest respectively their plausible archaic origin and an appreciable action of natural selection on them. In detail, high -logα and non-normalized significant LR values (i.e., 13 SNVs included in the upper tail of the distribution of non-normalized LR values calculated for chromosome 2) were obtained for the *EPAS1* genomic region, which is well-established as a positive control for AI, while high -logα values coupled with a remarkably low number of SNVs with significant LR scores characterized the *EGLN1* gene, which we instead envisaged as a negative control for AI (Figure 2 - figure supplement 2). These findings are concordant with previous evidence indicating these two loci as targets of natural selection in high-altitude Himalayan populations (Yang et al., 2017; Liu et al., 2022), although only *EPAS1* was proved to have been involved in an AI event (Huerta-Sanchez et al., 2014; Hu et al., 2017).

To validate results obtained by *VolcanoFinder* analysis, we used an approach based on the identification of networks of genes contributing to the same functional pathway and showing an overrepresentation of archaic (i.e., Denisovan)-like derived alleles (see Materials and methods). In line with what emerged from STRING enrichment analysis, many significant gene networks pointed out by the *Signet* algorithm were found to be involved in the same biological functions to which the core set of candidate genes identified by *VolcanoFinder* contributes (e.g., promotion of angiogenesis and induction of nitric oxide, NO) (Figure 3, Supplement Table 3). For instance, some significant gene networks were found to belong to the *JAK-STAT signalling*, *PI3K-Akt signalling*, *Apelin signalling* and *VEGF signalling* pathways that are known to be connected with each other and to be activated in hypoxic conditions to induce transcription of genes related with angiogenesis (Kanehisa et al., 2000, Fresno Vara et al., 2004; Ashley et al., 2005; Trân et al., 2018; Johnson et al., 2018; Mathari et al., 2021; Sahinturk et al., 2021) (Supplement Table 3). These signalling cascades determinate also the expression of *EP300* and *NOS2* genes and are capable to modulate tumour invasiveness, metastasis, and induction of NO (Kanehisa et al., 2000; Wang et al., 2005; Zhang et al., 2011; Johnson et al., 2018). In addition, other significant gene networks belonging to cancer-related pathways (e.g., *Prostate Cancer* and *Glioma*) include oncogenes that mediate functions related to angiogenesis, cell proliferation and inhibition of apoptosis (Kanehisa et al., 2000; Somanath et al., 2006 Yang et al., 2013; Lai et al., 2015; Chen et al., 2019; Xu et al., 2020) (Supplement Table 3). In detail, when results from *VolcanoFinder* and *Signet* analyses were compared at the single gene level, a total of twelve genes (i.e., *EP300, NOS2, TAB1*, *RAC2*, *MAPK1*, *MAPK11*, *MAP2K6*, *PPARA*, *FOXO1*, *ATF4*, *LEP*, *CSNK1E*) presented variation patterns supported by both methods as having been shaped by archaic introgression, and were found to be included in multiple significant gene networks involved in the modulation of biological functions such as angiogenesis, cell proliferation, and nitric oxide induction (Supplement Table 3). More specifically, the *EP300* gene was found to be involved, along with *ATF4* and *CREB cAMP* responsive elements, in the regulation of angiogenesis promotion, cell proliferation, and tumour invasiveness via cancer-related pathways (Halle et al., 2013; Voropaev et al., 2019; Huang et al., 2020). *MAPK1* was instead included together with *ATF4* and other members of the *MAPKinases family* in the significant gene network

ascribable to the *Relaxin signalling* pathway, which is known to exert both angiogenic and vasodilatory effects trough NO release (Kanehisa et al., 2000) (Figure 3A).

Estimating genetic distance between modern and archaic sequences

To explicitly test whether the Denisovan human species represented a plausible source of archaic alleles in the most convincing AI events inferred according to *VolcanoFinder* and *Signet* approaches, we assembled a dataset including *EP300* and *NOS2* candidate AI genomic sequences from high-altitude Tibetan, low-altitude Han Chinese (CHB), and Denisovan samples and we used the *Haplostrip* algorithm to estimate the genetic distance between modern and archaic haplotypes (see Materials and methods). For both genes, appreciable proportions of Tibetan haplotypes were found to cluster in the upper part of the heatmaps generated, thus showing among the lowest numbers of pairwise differences with respect to the Denisovan sequence observed in the entire dataset (Figure 4A and Figure 4 – figure supplement 2). Particularly, 33% and 16% of the total Tibetan *NOS2* and *EP300* haplotypes fell into the first quartile of haplotypes ranked according to their similarity with the Denisovan ones, in contrast to 50% and 9.5% of the *EPAS1* and *EGLN1* sequences used respectively as positive and negative controls for AI and non-AI loci. Moreover, as concerns an introgressed locus (e.g., *EPAS1*), most Tibetan haplotypes may be expected to present a smaller number of pairwise differences with Denisovan respect to non-Tibetan (i.e., CHB) haplotypes. Conversely, non-introgressed regions (e.g., *EGLN1*) should be characterized by a remarkably different pattern of haplotype variation. We thus calculated the cumulative proportion of Tibetan haplotypes for each haplotype class, defying each class according to the number of pairwise differences with respect to the Denisovan genome. Then, we counted the Tibetan and CHB haplotypes falling into the first quartile of the haplotype classes considered for each candidate gene to verify the assumption described above.

Results obtained for the *NOS2* gene seem to confirm the hypothesis of archaic introgression at such a locus. In fact, the cumulative proportion of Tibetan haplotypes turned out to be larger than that of CHB haplotypes in most of the defined classes and especially in those associated to the lowest numbers of pairwise differences with respect to the Denisovan sequence (Figure 4A and B). Such a trend is qualitatively

comparable to that observed for the *EPAS1* gene (Figure 4C), although the occurrence of a hard selective sweep at *EPAS1* (Huerta-Sánchez et al., 2014) has likely made the difference between Tibetan and CHB trends even more pronounced (Figure 4 – figure supplement 1). The pattern observed for the *EP300* gene was instead more nuanced, being characterized by nearly the same proportion of Tibetan and CHB haplotypes in the first quartile of haplotype classes and by an overall trend of cumulative proportions that was partially overlapped between the two populations (Figure 4 – figure supplement 2). However, the distribution of Tibetan *EP300* haplotypes in the different haplotype classes remained remarkably different with respect to that observed for the non-introgressed *EGLN1* locus, in which CHB haplotypes are systematically overrepresented in the haplotype classes presenting the lowest number of pairwise differences with respect to the Denisovan genome, thus determining higher cumulative proportions of CHB rather than Tibetan sequences even in classes enriched for Tibetan haplotypes (Figure 4D and Figure 4 – figure supplement 3).

Finally, we explored allele frequency trajectories of Denisovan derived alleles observed at these genes in Tibetan and CHB groups (see Materials and methods). In detail, when comparing frequencies for 85 archaic derived alleles in the *EP300* genomic region we pinpointed six SNVs showing significantly higher frequency (average value = 0.93) in Tibetans than in CHB (Fisher exact test, $p < 0.05$) (Figure 4 – figure supplement 4 and Supplement Table 4). However, no SNVs turned out to be significantly more represented in Tibetans after correction for multiple testing. As regards *NOS2*, no SNVs out of the 44 alleles considered showed a significantly higher frequency in the high-altitude group with respect to the CHB low-land population, although the average value of allele frequencies at Denisova-like derived alleles resulted slightly higher in Tibetans then in CHB (Figure 4 – figure supplement 4). Overall, these findings suggest that despite having exerted an adaptive role, introgressed variation at *EP300* and *NOS2* genes was plausibly not subjected to hard selective sweeps as instead previously attested for the *EPAS1* locus.

Discussion

In the present study, we aimed at investigating how far gene flow between the Denisovan archaic human species and the ancestors of modern populations settled in high-altitude regions of the Himalayas contributed to the evolution of key adaptive traits of these human groups, in addition to having conferred them reduced susceptibility to chronic mountain sickness (Huerta-Sánchez et al., 2014). For this purpose, we used WGS data from individuals of Tibetan ancestry to search for genomic signatures ascribable to AI mediated by weak selective events rather than by hard selective sweeps, under the assumption that soft sweeps and/or processes of polygenic adaptation are more likely to have occurred in such remarkably isolated and small-effective population size groups (Gnecchi-Ruscone et al., 2018).

By assembling a large genome-wide dataset including both low- and high-altitude populations, we first framed the available WGS data into the landscape of East Asian genomic variation. This confirmed that the genomes under investigation are well representative of the overall profiles of ancestry components observable in high-altitude Himalayan peoples (Figure 1A-B). In fact, the considered individuals were found to show close genetic similarity to other populations of Tibetan ancestry (Jeong et al., 2014) and to Sherpa people from Nepal (Gnecchi-Ruscone et al., 2017), as well as to appreciably divergence from the cline of variation of lowland East-Asians (Abdulla et al., 2009) (Figure 1C).

Based on this evidence, we submitted Tibetan WGS to a pipeline of analyses that implemented multiple independent approaches aimed at identifying genomic regions characterized simultaneously by signatures putatively ascribable to AI events and by tight functional correlations with each other. According to such a rationale, we shortlisted the candidate introgressed loci that most likely contribute to the same adaptive trait by searching for chromosomal intervals enriched for: i) significant LR scores computed by the *VolcanoFinder* algorithm, ii) Denisovan-like derived alleles belonging to significant gene networks reconstructed with the *Signet* method (Supplement table 3), iii) genes that have been also proved to participate to the same functional pathway according to information about established protein-protein interactions (Figure 2 – figure supplement 1 and Supplement table 2).

Among the genes encompassed within the putative introgressed chromosomal intervals identified by the *VolcanoFinder* method, *EP300, NOS2, TAB1, RAC2, MAPK1, MAPK11, MAP2K6, PPARA, FOXO1, ATF4, LEP and CSNK1E* were retained after the filtering procedure describe above. In fact, although 13 *EPAS1* SNVs

were observed among the top positive non-normalized LR values for chromosome 2, this locus did not present significant scores after normalization. We can interpret such a finding as suggestive of the tendency of the normalization procedure to enable the identification of relatively weak and widespread AI signatures plausibly ascribable to soft selective sweeps and/or polygenic adaptive mechanisms rather than to hard selective sweeps.

When we used STRING enrichment analysis to explore the functional role of the identified candidate AI loci, *mitogen-activated protein kinase* genes (e.g., *MAPK1* and *MAPK11*) were found to systematically occur in diverse significant pathways (Supplement Table 2). These proteins are key actors in the activation of the signalling cellular cascades in the *MAPK signalling* pathway, comporting the exploitation of multiple effects among which the modulation of the *VEGF signalling* pathway in different types of cancers is included (Nicolas et al., 2019). For instance, the activation of *MAPK/CREB* pathways upregulates the expression of anti-apoptotic genes in renal tubular cells during hypoxia-reoxygenation injury (Dong et al., 2019). Moreover, the *TAB1* gene directly binds and activates *MAP3K7*/*TAK1* kinase protein thus establishing strong functional connections within the *MAPK signalling* pathway (Kanehisa et al., 2000; Sayers et al., 2022). Finally, *ATF4* encodes for the *activating transcription factor 4* that belongs to a family of *CREB*-like proteins whose expression is detected in hypoxic and nutrient-deprived regions in tumours (Singleton et al., 2019; Sayers et al., 2022).

Despite such suggestive findings and according to the signatures ascribable to the action of natural selection on *EP300* and *NOS2* previously detected by other studies focused on high-altitude human populations, we considered these latter genes as the most convincing candidate AI loci to have played a relevant role in the adaptive history of Tibetan and Sherpa peoples. In detail, they were found to be recurrent in almost all the significant pathways pointed out by STRING enrichment analysis and, as members of the *HIF-1 signalling* pathway, are known to act in conjunction with *EPAS1* to regulate cellular responses to the hypoxic stress (Supplement table 2) (Narravula et al., 2001; Belanger et al., 2007; Chiu et al., 2020, Wu et al., 2022). Interestingly, *EP300* has been demonstrated to modulate directly *EPAS1* transcriptional activity and resulted significantly co-expressed whit *EPAS1* in the placenta of Tibetan women (Kanehisa et al., 2000; Sayers et al., 2022; Wu et al., 2022). The inducible nitric oxide synthase 2 enzyme

encoded by *NOS2* is instead expressed in a variety of cell types under hypoxic conditions where it controls a wide range of biological processes, among which activation of the *HIF-1 signalling* pathway and NO production (Kanehisa et al., 2000; Bigham et al., 2010; Fago et al., 2015; Ho et al., 2012; Sayers et al., 2022). The resulting signalling cascade thus impacts on target loci, such as *VEGF*, *Flt-1*, *Flk-1*, *Tie2*, *EPO*, *NOS*, which act as the main regulators of angiogenesis, erythropoiesis, and vascular tone (Kanehisa et al., 2000; Verheul et al., 2003; Takeda et al., 2004; Carroll et al., 2006; Vočanec et al., 2019). Especially angiogenesis represents a biological function that has been demonstrated to be improved in Tibetan and Sherpa high-altitude populations, enabling efficient O_2 delivery despite the hypoxic stress due to tissues' increased blood perfusion (Gnecchi-Ruscone et al., 2018). In fact, the potential adaptive role of *EP300* and *NOS2* was already proposed for Tibetan and Andean populations respectively (Bigham et al., 2010; Simonson et al., 2010; Peng et al., 2011; Crawford et al., 2017; Zheng et al., 2017Bigham et al., 2010; Simonson et al., 2010; Peng et al., 2011; Crawford et al., 2017; Zheng et al., 2017), along with that of multiple genes (e.g., *PPARA*, *STAT3*, *IL6*, *EPAS1*, *VEGF*, *EGLN1*, *EPO*, *ERK*, *PKC*) functionally related with them or with other members of the *HIF-1 signalling* pathway whose expression is modified under hypoxic conditions when tested *in vitro* and *in vivo* in different animal models (Foll et al., 2014; Yang et al., 2017; Tenzing et al., 2021; Liu et al., 2022; Mrakic-Sposta et al., 2022; Soliman et al., 2022; Yang et al., 2022; Yuan et al., 2022) (see Supplement information). In addition, both *NOS2* and *EP300* resulted implicated in mechanisms of NO production and regulation (Kanehisa et al., 2000; Zheng et al., 2017), which have been previously hypothesized as possibly subjected to selective pressures in Tibetans because of the high NO concentration in their blood and its physiological consequences, including increased O_2 delivery and reversibly inhibition of cellular respiration (Hoit et al., 2005; Erzurum et al., 2007; Fago et al., 2015; Poderoso et al., 2019). Results obtained by *Signet* analysis seem to further corroborate such a hypothesis by validating *EP300* and *NOS2* as genes plausibly involved in AI events. Some significant gene networks including *EP300* or functionally related to *NOS2* indeed showed higher-than-expected proportions of Denisovan-like derived alleles and again pointed to NO induction and promotion of angiogenesis as functional cascades having evolved peculiarly in populations of Tibetan ancestry (Figure 3, Supplement table 3). For instance, one of these networks participate to the *Jack-Stat signalling* pathway responsible for phosphorylation and activation of the *STAT3* transcriptional

factor (Supplement table 3) (Kanehisa et al., 2000). Interestingly, *STAT3* expression was found to be downregulated in the placenta of Tibetan women compared with that of individuals of European ancestry living at high-altitude for fewer than three generations, suggesting a potential adaptive role of such gene in populations that successfully reproduce in hypoxic conditions since thousands of years (Tenzing et al., 2021). Another significant network strictly linked from a functional perspective with both *NOS2* and *EP300* was ascribable to the *Apelin signalling* pathway (Supplement table 3). Activation of the Apelin-receptor system was shown to induce vessel relaxation and cardiac contractility in rodents, both *in vivo* and in isolated heart models, through the production and release of NO (Zhong et al., 2007; Trân et al., 2018; Mathari et al., 2021). The complex functional connections between *EP300*/*NOS2* and genes playing a role in such processes thus provide additional evidence supporting their potential adaptive role in high-altitude Tibetan and Sherpa populations. Moreover, *EP300* resulted included in one significant gene network belonging to the *Prostate cancer* pathway together whit *ATF4* and *CREB* genes, supporting once again its strong connections with cellular mechanisms deeply involved in the modulation of angiogenesis (Figure 3A). Particularly, the *CREB1* gene is defined as a proto-oncogene whose knockout in uveal melanoma cells dramatically decreases the tumour progression (Voropaev et al., 2019; Sapio et al., 2020). This gene, along with *PPARA,* was also proved to upregulate the expression of Highly Up-regulated in Liver Cancer (*HULC*) long non-coding RNA in hepatocellular carcinoma, affecting the aggressivity and proliferation of this type of cancer (Yu et al., 2017).

Overall, results obtained by means of the *Signet* and *VolcanoFinder* algorithms partially overlapped both from a functional perspective and at the single gene level, thus supporting the hypothesis that the evolution of especially *EP300* and *NOS2* genes has been shaped by AI events. More in general, these results indicate that archaic introgression remarkably impacted those biological functions, such as angiogenesis and NO induction, that have been then targeted by natural selection in high-altitude populations, suggesting that a portion of the introgressed variation was positively selected in Tibetan ancestors after their settlement at high altitudes.

We also attempted to explicitly test whether the Denisovan human species represented a plausible source of archaic alleles in the framework of the inferred AI events, as previously demonstrated for *EPAS1* (Huerta-

Sánchez et al., 2014). When direct comparison between modern and archaic haplotypes at the putative introgressed loci was performed, an overall higher degree of similarity between Tibetan and Denisovan rather than between Han Chinese and Denisovan sequences was observed for *NOS2* in those haplotype classes presenting the smallest numbers of pairwise differences with the archaic sequence, in line with what reported for *EPAS1* (Figure 4A). Instead, as concerns *EP300* the distribution of haplotypes among individuals with Tibetan and Han Chinese ancestry was found to be more homogeneous, suggesting that after Denisovan admixture with their common ancestors natural selection might have targeted this locus less intensively in Tibetan and Sherpa populations with respect to what happened at *NOS2*. Nevertheless, for both genes the observed patterns of haplotype variation substantially diverged from that described for *EGLN1*, which was previously proved to have been not involved in an AI event. At the same time, it is noteworthy that *EP300* and *NOS2* presented considerably less striking divergence between Tibetan and Han Chinese haplotype frequencies and similarity with the Denisovan sequence than what observed for *EPAS1* and thus than what expected after the occurrence of a hard selective sweep (Figure 4 – figure supplement 1). This suggests that the adaptive evolution of *NOS2* and *EP300* after the introgression event has been mediated by different mechanisms, which did not exceptionally increase the frequency of few variants/haplotypes and consequently did not drastically reduce genetic diversity at the selected locus.

This is in accordance with the hypothesis of a relevant role played by soft selective sweeps and/or polygenic adaptation during the evolutionary history of the ancestors of high-altitude Himalayan populations (Gnecchi-Ruscone et al., 2018). These mechanisms are generally characterized by weak selection strength on single alleles due to their limited individual contribution to the adaptive trait, which enables recombination to erode haplotypes carrying advantageous variants and to produce signatures that are hardly distinguishable from those observed at neutrally evolving genomic regions (Pritchard at al., 2010).

Overall, we collected multiple evidence supporting both the archaic origin and the adaptive role of variation at *EP300* and *NOS2* genes in populations of Tibetan ancestry, as well as at other loci presenting tight functional relationships with them, thus expanding the knowledge about AI events having involved the ancestors of modern high-altitude Himalayan populations and Denisovans. The results obtained

emphasized once more the complexity of the adaptive phenotype evolved by these human groups to cope with challenges imposed by hypobaric hypoxia, offering new insights for future studies aimed at elucidating the molecular mechanisms by which several genes along with *EP300* and *NOS2* interact with each other and contribute to mediate physiological adjustments that are crucial for human adaptation to the high-altitude environment.

Materials and methods

Dataset composition and curation

The dataset used in the present study included WGS data for 27 individuals of Tibetan ancestry recruited from the high-altitude Nepalese districts of Mustang and Ghorka (Jeong et al., 2018). Although these subjects reside in Nepal, they have been previously proved to speak Tibetan dialects and to live in communities showing religious and social organizations proper of populations settled on the Tibetan Plateau, being also biologically representative of high-altitude Tibetan people (Cho et al., 2017). To filter for high-quality genotypes, the dataset was subjected to QC procedures using the software PLINK v1.9 (Purcell et al., 2007). In detail, we retained autosomal SNVs characterized by no significant deviations from the Hardy-Weinberg Equilibrium ($P > 5.3 \times 10^{-9}$ after Bonferroni correction for multiple testing), as well as loci/samples showing less than 5% of missing data. Moreover, we removed SNVs with ambiguous A/T or G/C alleles and multiallelic variants, obtaining a dataset composed by 27 individuals and 6,921,628 SNVs. WGS data were finally phased with SHAPEIT2 v2.r904 (Delaneau et al., 2013) by applying default parameters and using the 1000 Genomes Project dataset as a reference panel (1000 Genomes Project Consortium et al., 2015) and HapMap phase 3 recombination maps.

Population structure analyses

To assess representativeness, genetic homogeneity, and ancestry composition of Tibetan WGS included in the dataset, we performed genotype-based population structure analyses. For this purpose, we merged the unphased WGS dataset with genome-wide genotyping data for 34 East-Asian populations (Gnecchi-Ruscone

et al., 2017; Landini et al., 2021) and we applied the same QC procedures described above. The obtained extended dataset included 231,947 SNVs and was used to assess the degree of recent shared ancestry (i.e., identity-by-descent, IBD) for each pair of individuals. Identity-by-state (IBS) estimates were thus used to calculate an IBD kinship coefficient and a threshold of PI_HAT = 0.270 was considered to remove closely related subjects to the second degree (Ojeda-Granados et al., 2022). To discard variants in linkage disequilibrium (LD) with each other we then removed one SNV for each pair showing $r^2 > 0.2$ within sliding windows of 50 SNVs and advancing by five SNVs at the time. The obtained LD-pruned dataset was finally filtered for variants with minor allele frequency (MAF) < 0.01 and used to compute PCA utilizing the *smartpca* method implemented in the EIGENSOFT package (Patterson et al., 2006), as well as to run the ADMIXTURE algorithm version 1.3.0 (Alexander et al., 2009) by testing K = 2 to K = 12 population clusters. In detail, 25 replicates with different random seeds were run for each K and we retained only those presenting the highest log-likelihood values. In addition, cross-validation errors were calculated for each K to identify the value that best fit the data. Both PCA and ADMIXTURE results were visualized with the R software version 4.0.5.

Detecting signatures of adaptive introgression

To identify chromosomal regions showing signatures putatively ascribable to AI events, we submitted the phased WGS dataset to the *VolcanoFinder* pipeline, which relies on the analysis of the allele frequency spectrum of the population that is supposed to have experienced archaic introgression (Setter et al., 2020). This method enabled us to scan the genome to calculate two statistics for each polymorphic site: α (subsequently converted in -logα) and LR which are informative respectively of i) the strength of natural selection and ii) the conformity to the evolutionary model of adaptive introgression (see Supplement Information for further details about the evolutionary model tested). Since composite likelihood statistics are not associated with p-values, we implemented multiple procedures to filter SNVs according to the significance of their LR values. First, for each chromosome, variants were grouped into frequency bins and a statistical distribution of LR values was constructed for each of them. LR distributions were normalized, obtaining bell curves in which we considered as significant those SNVs whose LR values fell into the

extreme positive tail (i.e., LR ≥ 2). Then, the genome was divided into 200 kb windows with an overlap of 50 kb and for each of them we calculated the ratio between the number of significant SNVs and the total number of variants. We finally ranked windows according to such a ratio and to shortlist those enriched in significant LR values we retained chromosomal intervals within the top 0.1% of the obtained distribution, subsequently merging overlapping significant windows into larger genomic segments.

Investigating functional relationships among putative introgressed loci

Functional relationships among genes included in the chromosomal segments pointed out by *VolcanoFinder* analysis as putatively enriched for AI alleles were investigated by exploring protein-protein interactions and protein activity in specific cellular pathways, as well as by performing formal enrichment analyses, with functions implemented in the STRING tool (available at: string-db.org) (Jensen et al., 2009). For this purpose, we considered only protein-protein interactions showing confidence scores ≥ 0.7 and the obtained protein frameworks were integrated using information available in literature regarding the functional role of the related genes and their possible involvement in high-altitude adaptation.

Identifying gene networks enriched for Denisovan-like derived alleles

To validate *VolcanoFinder* results by using an independent approach, we set up a pipeline of analyses aimed at identifying the biological functions whose underlying genomic regions might have been significantly shaped by Denisovan introgression. For this purpose, we used the *Signet* algorithm (Gouy et al., 2020) to reconstruct network of genes that participate to the same biological pathway and that are also characterized by a higher-than-expected proportion of Denisovan-like derived alleles. To do so, we first compared the Tibetan and Denisovan genomes to assess which SNVs were present in both modern and archaic sequences. These loci were further compared with the ancestral reconstructed refence human genome sequence (1000 Genomes Project Consortium et al., 2015) to discard those presenting an ancestral state (i.e., that we have in common with several primate species). Accordingly, although we cannot rule out the possibility that *H. sapiens* inherited these variants from the most recent common ancestor we share

with Denisovans, we were able to shortlist SNVs that have at least the potential to have been introgressed due to gene flow from such an archaic species. We then calculated the proportion of Denisovan-like derived alleles for each Tibetan gene and we used the related genome-wide distributions to inform *Signet* analysis (see Supplement Information for details). We performed five independent runs of the *Signet* algorithm to check for consistency of the significant gene networks and functional pathways identified and we finally depicted the confirmed results using Cytoscape v3.9.1 (Shannon et al., 2003).

Haplostrip **analysis**

To explicitly test whether the putative introgressed loci pointed out by *VolcanoFinder* and *Signet* analyses present variation patterns compatible with a scenario of introgression from the Denisovan archaic human species, genetic distance between modern and archaic haplotypes were estimated using the *Haplostrip* pipeline, as described in previous studies (Huerta-Sánchez et al., 2014; Marnetto et al., 2017). For this purpose, we extracted the sequences of genes showing putative introgression signatures ± 50 kb upstream and downstream of these loci from the 27 Tibetan genomes under study (Jeong et al., 2018), from 27 CHB WGSs (Auton et al., 2015) and from the Denisovan genome (Meyer et al., 2012). The CHB population, which is known to share a relatively recent common ancestry with Tibetans, was used as a "negative low-altitude control" (i.e., as a group whose ancestors experienced Denisovan introgression, but did not evolve high-altitude adaptation). In addition to the identified candidate loci, we included in this dataset also the *EPAS1* and *EGLN1* genes as positive and negative control loci, which enabled us to perform a direct comparison of the observed haplotype patterns to those of genomic regions that have been previously proved to be involved or not in AI events. We then phased the assembled dataset with SHAPEIT2 v2.r904 (Delaneau et al., 2013), as described in the *Dataset composition and curation* section and we run the *Haplostrip* pipeline.

Then, to assess whether Tibetan haplotypes at candidate AI genes overall present a lower number of differences from the archaic sequence with respect to those of the CHB low-land control population, we defined different haplotype classes according to the number of pairwise differences from the Denisovan genome and we calculated the percentage of Tibetan and non-Tibetan haplotypes belonging to each of

them. The rationale for such an approach was that in the case of introgressed loci (e.g., *EPAS1*), most haplotype classes may be expected to be characterized by a greater cumulative proportion of Tibetan haplotypes rather than non-Tibetan ones. Conversely, for non-introgressed loci (e.g., *EGLN1*), we might expect a remarkably different pattern of haplotypes distribution, with almost all haplotype classes presenting a larger proportion of non-Tibetan haplotypes rather than Tibetan ones. In line with this expectation, we also counted the number of Tibetan haplotypes falling into the first quartile of haplotype classes defined for a given genomic region, considering as evidence supporting archaic introgression those cases in which the top 25% of haplotype classes included a greater number of Tibetan haplotypes than CHB ones.

Comparing frequencies of Denisovan-like derived alleles between Tibetans and Han Chinese

We investigated the differences in allele frequency of archaic derived alleles (i.e., SNVs presenting the same derived state in Tibetans, CHB and the Denisovan genome) at candidate AI loci between Tibetan and CHB populations by performing Fisher exact test using functions implemented in the PLINK software v1.9 (Purcell et al., 2007). For this purpose, we considered a dataset made up of 27 individuals with Tibetan ancestry and 27 CHB individuals. Particularly, we merged Tibetan WGS analysed in the present study with 27 CHB WGS from the 1000 Genomes Project panel (Auton et al., 2015). We applied to the obtained dataset the same QC filter procedures described in the *Dataset composition and curation* section. Then, we extracted the SNVs showing archaic derived alleles at candidate AI genes, as well as in 50 kb upstream and downstream of these loci. Finally, we performed the Fisher exact test on this subset of SNPs considering as significant those variants with an associated p-value < 0.05 (Figure 5 and Supplement Table 4-5).

Acknowledgements

We acknowledge support from the Fondazione Cassa di Risparmio in Bologna through the project "*Genetic adaptation and acclimatization to high altitude as experimental models to investigate the biological mechanisms that regulate physiological responses to hypoxia*", which was granted to MS (n. 2019.0552).

Additional information

Authors contributions

Giulia Ferraretti and Paolo Abondio, Data curation, Software, Formal analysis, Investigation, Writing - original draft; Marta Alberti, Software, Formal analysis; Agnese Dezi, Phurba T. Sherpa, Paolo Cocco, Massimiliano Tiriticco, Marco di Marcello, Guido Alberto Gnecchi-Ruscone, Luca Natali, Angela Corcelli, Giorgio Marinelli, Davide Peluzzi, Data curation, Writing - review and editing; Stefania Sarno, Formal analysis, Writing - original draft; Marco Sazzini, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Writing - review and editing.

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., … Schloss, J. A. (2015). A global reference for human genetic variation. In *Nature* (Vol. 526, Issue 7571, pp. 68–74). Nature Publishing Group. https://doi.org/10.1038/nature15393

Benjamin, V., & M, A. J. (2014). Resurrecting Surviving Neandertal Lineages from Modern Human Genomes. *Science*, *343*(6174), 1017–1021. https://doi.org/10.1126/science.1245938

Bergström, A., McCarthy, S. A., Hui, R., Almarri, M. A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J. F., Cann, H., Mallick, S., Reich, D., Sandhu, M. S., Skoglund, P., Scally, A., Xue, Y., … Tyler-Smith, C. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science*, *367*(6484). https://doi.org/10.1126/science.aay5012

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., Scherer, S. W., Julian, C. G., Wilson, M. J., López Herráez, D., Brutsaert, T., Parra, E. J., Moore, L. G., & Shriver, M. D. (2010). Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *PLOS Genetics*, *6*(9), e1001116. https://doi.org/10.1371/journal.pgen.1001116

Cho, J. I., Basnyat, B., Jeong, C., di Rienzo, A., Childs, G., Craig, S. R., Sun, J., & Beall, C. M. (2017). Ethnically Tibetan women in Nepal with low hemoglobin concentration have better reproductive outcomes. *Evolution, Medicine, and Public Health*, *2017*(1), 82–96. https://doi.org/10.1093/emph/eox008

Crawford, J. E., Amaru, R., Song, J., Julian, C. G., Racimo, F., Cheng, J. Y., Guo, X., Yao, J., Ambale-Venkatesh, B., Lima, J. A., Rotter, J. I., Stehlik, J., Moore, L. G., Prchal, J. T., & Nielsen, R. (2017). Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. *American Journal of Human Genetics*, *101*(5), 752–767. https://doi.org/10.1016/j.ajhg.2017.09.023

Dannemann, M., & Racimo, F. (2018). Something old, something borrowed: admixture and adaptation in human evolution. In *Current Opinion in Genetics and Development* (Vol. 53, pp. 1–8). Elsevier Ltd. https://doi.org/10.1016/j.gde.2018.05.009

Dong, Q., Jie, Y., Ma, J., Li, C., Xin, T., & Yang, D. (2019). Renal tubular cell death and inflammation response are regulated by the MAPK-ERK-CREB signaling pathway under hypoxia-reoxygenation injury. *Journal of Receptors and Signal Transduction*, *39*(5–6), 383–391. https://doi.org/10.1080/10799893.2019.1698050

el Mathari, B., Briand, P., Corbier, A., Poirier, B., Briand, V., Raffenne-Devillers, A., Harnist, M. P., Guillot, E., Guilbert, F., & Janiak, P. (2021). Apelin improves cardiac function mainly through peripheral vasodilation in a mouse model of dilated cardiomyopathy. *Peptides*, *142*. https://doi.org/10.1016/j.peptides.2021.170568

Enard, D., & Petrov, D. A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell*, *175*(2), 360-371.e13. https://doi.org/10.1016/j.cell.2018.08.034

Foll, M., Gaggiotti, O. E., Daub, J. T., Vatsiou, A., & Excoffier, L. (2014). Widespread signals of convergent adaptation to high altitude in Asia and America. *American Journal of Human Genetics*, *95*(4), 394–407. https://doi.org/10.1016/j.ajhg.2014.09.002

Gittelman, R. M., Schraiber, J. G., Vernot, B., Mikacenic, C., Wurfel, M. M., & Akey, J. M. (2016). Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. *Current Biology*, *26*(24), 3375–3382. https://doi.org/https://doi.org/10.1016/j.cub.2016.10.041

Gnecchi-Ruscone, G. A., Abondio, P., de Fanti, S., Sarno, S., Sherpa, M. G., Sherpa, P. T., Marinelli, G., Natali, L., di Marcello, M., Peluzzi, D., Luiselli, D., Pettener, D., & Sazzini, M. (2018a). Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes. *Genome Biology and Evolution*, *10*(11), 2919–2930. https://doi.org/10.1093/gbe/evy233

Gnecchi-Ruscone, G. A., Abondio, P., de Fanti, S., Sarno, S., Sherpa, M. G., Sherpa, P. T., Marinelli, G., Natali, L., di Marcello, M., Peluzzi, D., Luiselli, D., Pettener, D., & Sazzini, M. (2018b). Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes. *Genome Biology and Evolution*, *10*(11), 2919–2930. https://doi.org/10.1093/gbe/evy233

Gnecchi-ruscone, G. A., Jeong, C., Fanti, S. de, Sarno, S., Gentilini, D., Blasio, A. M. di, Sherpa, M. G., Sherpa, P. T., Marinelli, G., Marcello, M. di, Natali, L., Peluzzi, D., Pettener, D., Rienzo, A. di, Luiselli, D., & Sazzini, M. (2017). The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the recent peopling of Southern Himalayas. *Scientific Reports*, *January*, 1–12. https://doi.org/10.1038/s41598-017-15862-z

Gouy, A., Excoffier, L., & Nielsen, R. (2020). Polygenic Patterns of Adaptive Introgression in Modern Humans Are Mainly Shaped by Response to Pathogens. *Molecular Biology and Evolution*, *37*(5), 1420–1433. https://doi.org/10.1093/molbev/msz306

Green, R., Krause, J., Briggs, A., Rasilla Vives, M. de la, & Fortea Pérez, F. J. (2010). A draft sequence of the neandertal genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021PubMedID:20448178

Hamid, I., Korunes, K. L., Beleza, S., & Goldberg, A. (2021). Rapid adaptation to malaria facilitated by admixture in the human population of cabo verde. *ELife*, *10*, 1–24. https://doi.org/10.7554/eLife.63177

Hu, H., Petousi, N., Glusman, G., Yu, Y., Bohlender, R., Tashi, T., Downie, J. M., Roach, J. C., Cole, A. M., Lorenzo, F. R., Rogers, A. R., Brunkow, M. E., Cavalleri, G., Hood, L., Alpatty, S. M., Prchal, J. T., Jorde, L. B., Robbins, P. A., Simonson, T. S., & Huff, C. D. (2017). Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genetics*, *13*(4). https://doi.org/10.1371/journal.pgen.1006675

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194–197. https://doi.org/10.1038/nature13408

Israel, H., W, W. G., Rolf, Q., Mathieu, D., Rainer, G., Leslie, K., Avner, A., Miryam, B.-M., Helene, V., Norbert, M., Luis, A. J., María, M.-T., María, B. de C. J., Cinzia, F., Laura, M.-F., Rachel, S., Hila, M., A, K. V., Viviane, S., … Mina, W.-E. (2018). The earliest modern humans outside Africa. *Science*, *359*(6374), 456–459. https://doi.org/10.1126/science.aap8369

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., Beall, C. M., & di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications*, *5*, 1–7. https://doi.org/10.1038/ncomms4281

Jeong, C., Witonsky, D. B., Basnyat, B., Neupane, M., Beall, C. M., Childs, G., Craig, S. R., Novembre, J., & di Rienzo, A. (2018). Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. *PLOS Genetics*, *14*(9), e1007650. https://doi.org/10.1371/journal.pgen.1007650

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. In *Nucleic Acids Research* (Vol. 28, Issue 1). http://www.genome.ad.jp/kegg/

Landini, A., Yu, S., Gnecchi-Ruscone, G. A., Abondio, P., Ojeda-Granados, C., Sarno, S., de Fanti, S., Gentilini, D., di Blasio, A. M., Jin, H., Nguyen, T. T., Romeo, G., Prata, C., Bortolini, E., Luiselli, D., Pettener, D., & Sazzini, M. (2021). Genomic adaptations to cereal-based diets contribute to mitigate metabolic risk in some human populations of East Asian ancestry. *Evolutionary Applications*, *14*(2), 297–313. https://doi.org/10.1111/eva.13090

Laurent, A.-R., J., J. M., Subhash, K., Alasdair, M., Klara, D., Loren, G., Farbod, B., Baback, G., Ma, L., A., P. F., Joshua, K., Mary, C., Derek, M., Raja, R., Meral, B., E., M. S. G., Martin, M., A., G. L., Sofia, T., … Peter, P. (2011). The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. *Science*, *334*(6052), 89–94. https://doi.org/10.1126/science.1209202

Liu, C. C., Witonsky, D., Gosling, A., Lee, J. H., Ringbauer, H., Hagan, R., Patel, N., Stahl, R., Novembre, J., Aldenderfer, M., Warinner, C., di Rienzo, A., & Jeong, C. (2022). Ancient genomes from the Himalayas illuminate the genetic history of Tibetans and their Tibeto-Burman speaking neighbors. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-28827-2

Marnetto, D., & Huerta-Sánchez, E. (2017). Haplostrips: revealing population structure through haplotype visualization. *Methods in Ecology and Evolution*, *8*(10), 1389–1392. https://doi.org/10.1111/2041-210X.12747

McArthur, E., Rinker, D. C., & Capra, J. A. (2021). Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. *Nature Communications*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-24582-y

McCoy, R. C., Wakefield, J., & Akey, J. M. (2017). Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression. *Cell*, *168*(5), 916-927.e12. https://doi.org/https://doi.org/10.1016/j.cell.2017.01.038

Meyer, M., Kircher, M., Gansauge, M., Li, H., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., De, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Siebauer, M., Green, R. E., Bryc, K., Briggs, A. W., Dabney, J., … Eichler, E. E. (2013). *A High Coverage Genome Sequence From an Archaic*. *338*(6104), 222–226. https://doi.org/10.1126/science.1224344.A

Mrakic-Sposta, S., Montorsi, M., Porcelli, S., Marzorati, M., Healey, B., Dellanoce, C., & Vezzoli, A. (2022). Effects of Prolonged Exposure to Hypobaric Hypoxia on Oxidative Stress: Overwintering in Antarctic Concordia Station. *Oxidative Medicine and Cellular Longevity*, *2022*, 1–14. https://doi.org/10.1155/2022/4430032

Nicolas, S., Abdellatef, S., Al Haddad, M., Fakhoury, I., & El-Sibai, M. (2019). Hypoxia and EGF stimulation regulate VEGF expression in human Glioblastoma multiforme (GBM) cells by differential regulation of the PI3K/Rho-GTPase and MAPK pathways. *Cells*, *8*(11). https://doi.org/10.3390/cells8111397

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., de Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., … Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–49. https://doi.org/10.1038/nature12886

Racimo, F., Marnetto, D., & Huerta-Sánchez, E. (2017). Signatures of archaic adaptive introgression in present-day human populations. *Molecular Biology and Evolution*, *34*(2), 296–317. https://doi.org/10.1093/molbev/msw216

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354–357. https://doi.org/10.1038/nature12961

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Setter, D., Mousset, S., Cheng, X., Nielsen, R., DeGiorgio, M., & Hermisson, J. (2020). VolcanoFinder: Genomic scans for adaptive introgression. In *PLoS Genetics* (Vol. 16, Issue 6). https://doi.org/10.1371/journal.pgen.1008867

Simonti, C. N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D. S., Chisholm, R. L., Crosslin, D. R., Hebbring, S. J., Jarvik, G. P., Kullo, I. J., Li, R., Pathak, J., Ritchie, M. D., Roden, D. M., Verma, S. S., Tromp, G., Prato, J. D., Bush, W. S., Akey, J. M., … Capra, J. A. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science*, *351*(6274), 737–741. https://doi.org/10.1126/science.aad2149

Soliman, M. M., Aldhahrani, A., Althobaiti, F., Ahmed, M. M., Sayed, S., Alotaibi, S., Shukry, M., El-Shehawi, A. M., & Sa, A. M. (2022). Characterization of the Impacts of Living at High Altitude in Taif: Oxidative Stress Biomarker Alterations and Immunohistochemical Changes. *Issues Mol. Biol*, *2022*, 1610–1625. https://doi.org/10.3390/cimb

Tenzing, N., van Patot, M. T., Liu, H., Xu, Q., Liu, J., Wang, Z., Wang, Y., Wuren, T., & Ge, R. L. (2021). Identification of a miRNA–mRNA Regulatory Networks in Placental Tissue Associated With Tibetan High Altitude Adaptation. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.671119

Trân, K., Murza, A., Sainsily, X., Coquerel, D., Côté, J., Belleville, K., Haroune, L., Longpré, J. M., Dumaine, R., Salvail, D., Lesur, O., Auger-Messier, M., Sarret, P., & Marsault, É. (2018). A Systematic Exploration of Macrocyclization in Apelin-13: Impact on Binding, Signaling, Stability, and Cardiovascular Effects. *Journal of Medicinal Chemistry*, *61*(6), 2266–2277. https://doi.org/10.1021/acs.jmedchem.7b01353

Voropaev, H., Gimmelshein Vatkin, M., Shneor, D., Luski, S., Honigman, A., & Frenkel, S. (2019). Infectious Knockdown of CREB and HIF-1 for the Treatment of Metastatic Uveal Melanoma. *Cancers*, *11*(8). https://doi.org/10.3390/cancers11081056

Wang, Y., Zou, X., Wang, M., Yuan, D., Yang, L., Zeng, Y., Cheng, F., Tang, R., & He, G. (2022). The genomic history of southwestern Chinese populations demonstrated massive population migration and admixture among proto-Hmong–Mien speakers and incoming migrants. *Molecular Genetics and Genomics*, *297*(1), 241–262. https://doi.org/10.1007/s00438-021-01837-3

Wu, D., Liu, Y., Chen, W., Shao, J., Zhuoma, P., Zhao, D., Yu, Y., Liu, T., Yu, R., Gan, Y., Yuzheng, B., Huang, Y., Zhang, H., Bi, X., Tao, C., Lai, S., Luo, Q., Zhang, D., Wang, H., … Zeng, C. (2022). How Placenta Promotes the Successful Reproduction in High-Altitude Populations: A Transcriptome Comparison between Adaptation and Acclimatization. *Molecular Biology and Evolution*, *39*(6). https://doi.org/10.1093/molbev/msac120

Yang, J., Jin, Z. B., Chen, J., Huang, X. F., Li, X. M., Liang, Y. B., Mao, J. Y., Chen, X., Zheng, Z., Bakshi, A., Zheng, D. D., Zheng, M. Q., Wray, N. R., Visscher, P. M., Lu, F., & Qu, J. (2017). Genetic signatures of high-altitude adaptation in Tibetans. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), 4189–4194. https://doi.org/10.1073/pnas.1617042114

Yang, X. Y., Rakha, A., Chen, W., Hou, J., Qi, X. bin, Shen, Q. K., Dai, S. S., Sulaiman, X., Abdulloevich, N. T., Afanasevna, M. E., Ibrohimovich, K. B., Chen, X., Yang, W. K., Adnan, A., Zhao, R. H., Yao, Y. G., Su, B., Peng, M. S., & Zhang, Y. P. (2021). Tracing the Genetic Legacy of the Tibetan Empire in the Balti. *Molecular Biology and Evolution*, *38*(4), 1529–1536. https://doi.org/10.1093/molbev/msaa313

Yang, Y., Gao, C., Yang, T., Sha, Y., Cai, Y., Wang, X., Yang, Q., Liu, C., Wang, B., & Zhao, S. (2022). Vascular characteristics and expression of hypoxia genes in Tibetan pigs' hearts. *Veterinary Medicine and Science*, *8*(1), 177–186. https://doi.org/10.1002/vms3.639

Yang, Y., Yuan, H., Yang, Q., Cai, Y., Ren, Y., Li, Y., Gao, C., & Zhao, S. (2022). Post-transcriptional regulation through alternative splicing in the lungs of Tibetan pigs under hypoxia. *Gene*, *819*. https://doi.org/10.1016/j.gene.2022.146268

Yu, X., Zheng, H., Chan, M. T. V., & Wu, W. K. K. (2017). HULC: an oncogenic long non-coding RNA in human cancer. In *Journal of Cellular and Molecular Medicine* (Vol. 21, Issue 2, pp. 410–417). Blackwell Publishing Inc. https://doi.org/10.1111/jcmm.12956

Zhang, C., Lu, Y., Feng, Q., Wang, X., Lou, H., Liu, J., Ning, Z., Yuan, K., Wang, Y., Zhou, Y., Deng, L., Liu, L., Yang, Y., Li, S., Ma, L., Zhang, Z., Jin, L., Su, B., Kang, L., & Xu, S. (2017). Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1242-y

Zheng, W. S., He, Y. X., Cui, C. Y., Ouzhu, L., Deji, Q., Peng, Y., Bai, C. J., Duoji, Z., Gongga, L., Bian, B., Baima, K., Pan, Y. Y., Qu, la, Kang, M., Ciren, Y., Baima, Y., Guo, W., Yang, la, Zhang, H., … Su, B. (2017). EP300 contributes to high-altitude adaptation in Tibetans by regulating nitric oxide production. *Zoological Research*, *38*(3), 163–170. https://doi.org/10.24272/j.issn.2095-8137.2017.036

Zhong, J. C., Yu, X. Y., Huang, Y., Yung, L. M., Lau, C. W., & Lin, S. G. (2007). Apelin modulates aortic vascular tone via endothelial nitric oxide synthase phosphorylation pathway in diabetic mice. *Cardiovascular Research*, *74*(3), 388–395. https://doi.org/10.1016/j.cardiores.2007.02.002

Figures

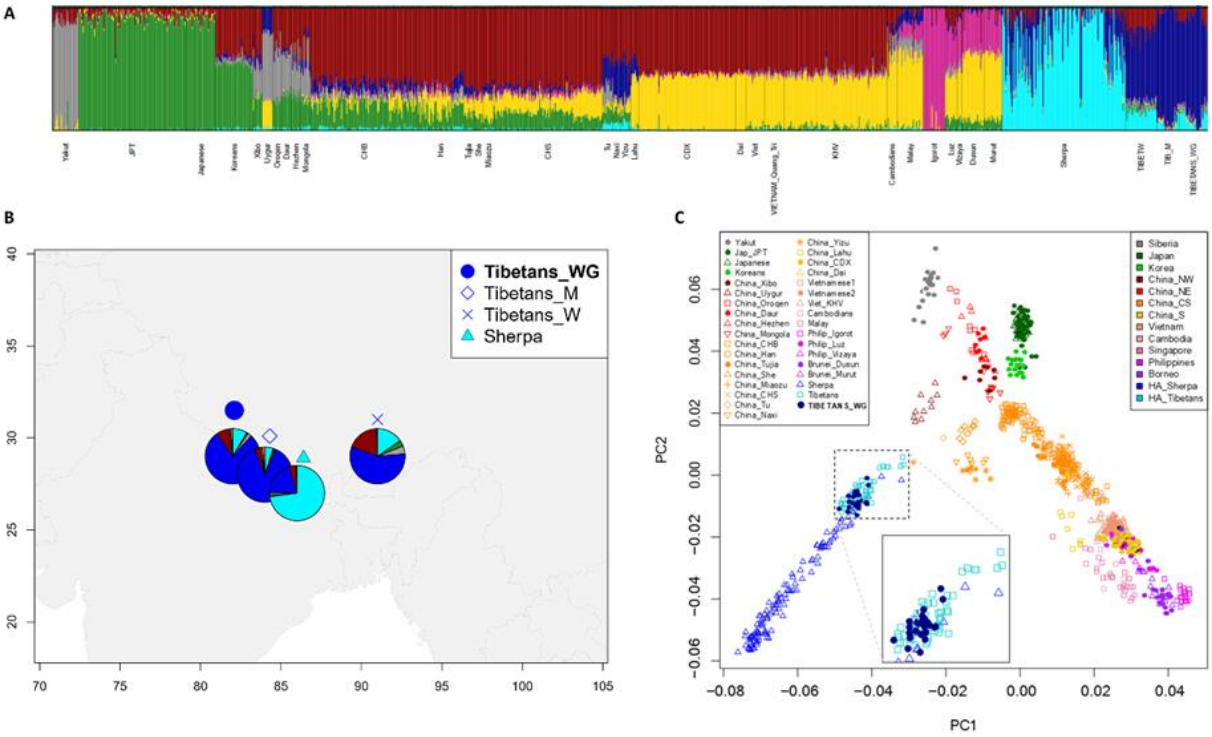

Figure 1. Population structure analyses performed on the extended dataset including Tibetan, Sherpa, and lowland East-Asian individuals. (A) Admixture analysis showed the best predictive accuracy when seven (K=7) population clusters were tested. Populations included in the dataset are labelled according to population names and acronyms reported in Supplement table 1. **(B)** Map showing geographic location and admixture proportions at K=7 of the high-altitude groups included in the extended dataset. The label *Tibetans_WG* indicates whole genome sequence data for individuals of Tibetan ancestry analysed in the present study. Additional information about the considered samples (e.g., number of individuals per group, reference study, and used abbreviations) are reported in Supplement table 1. **(C)** PCA plot considering PC1 vs PC2 and summarizing genomic divergence between high-altitude Tibetan/Sherpa people and the cline of variation observable for lowland East-Asian populations. The enlarged square displays clustering between Tibetan samples sequenced for the whole genome (i.e., blue dots) and Tibetan samples characterized by genome-wide data (i.e., light-blue squares).

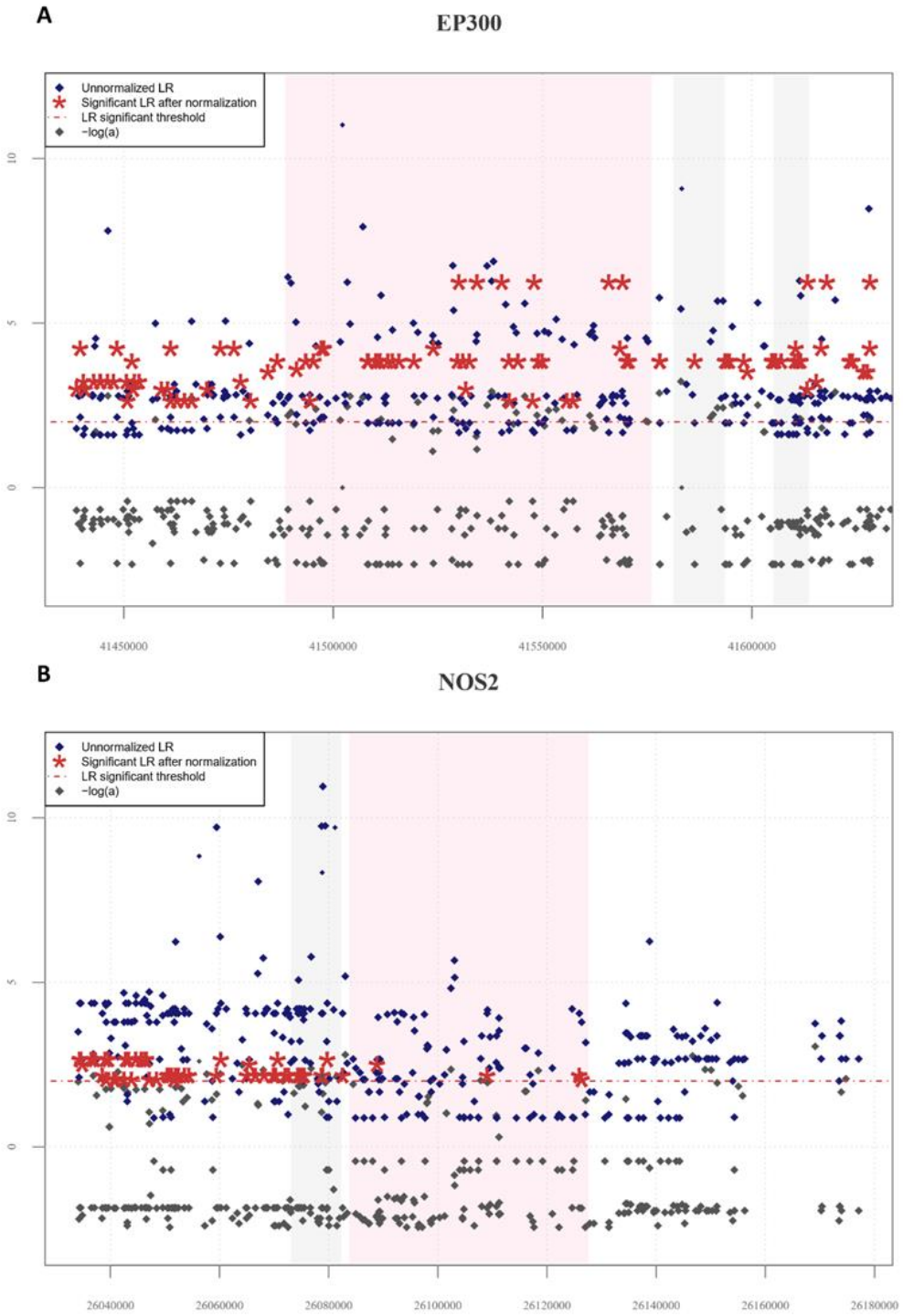

Figure 2. Distribution of *VolcanoFinder* statistics suggestive of putative adaptive introgressed loci across the *EP300* and *NOS2* genomic regions. On the x-axis are reported genomic positions of each SNV, while on the y axis are displayed the related statistics obtained. Pink background indicates the chromosomal interval occupied by the considered genes, while the grey background identifies those genes/pseudogenes (i.e., *EP300AS1* and *L3MBTL2* in the *EP300* downstream genomic region and the *LGALS9DP* pseudogene in the upstream *NOS2* region) possibly involved in regulatory transcription mechanisms. The dashed red line identifies the threshold set to filter for significant LR values. **(A)** Non-normalized LR (blue diamonds) and -logα (grey diamonds) values resulted collectively elevated in the entire *EP300* genomic region, which was characterized also by a consistent number of SNVs showing significant normalized LR values (red stars). **(B)** Both non-normalized LR and -logα scores appeared elevated in the upstream *NOS2* region, which also presented a remarkable number of SNVs showing significant normalized LR values.

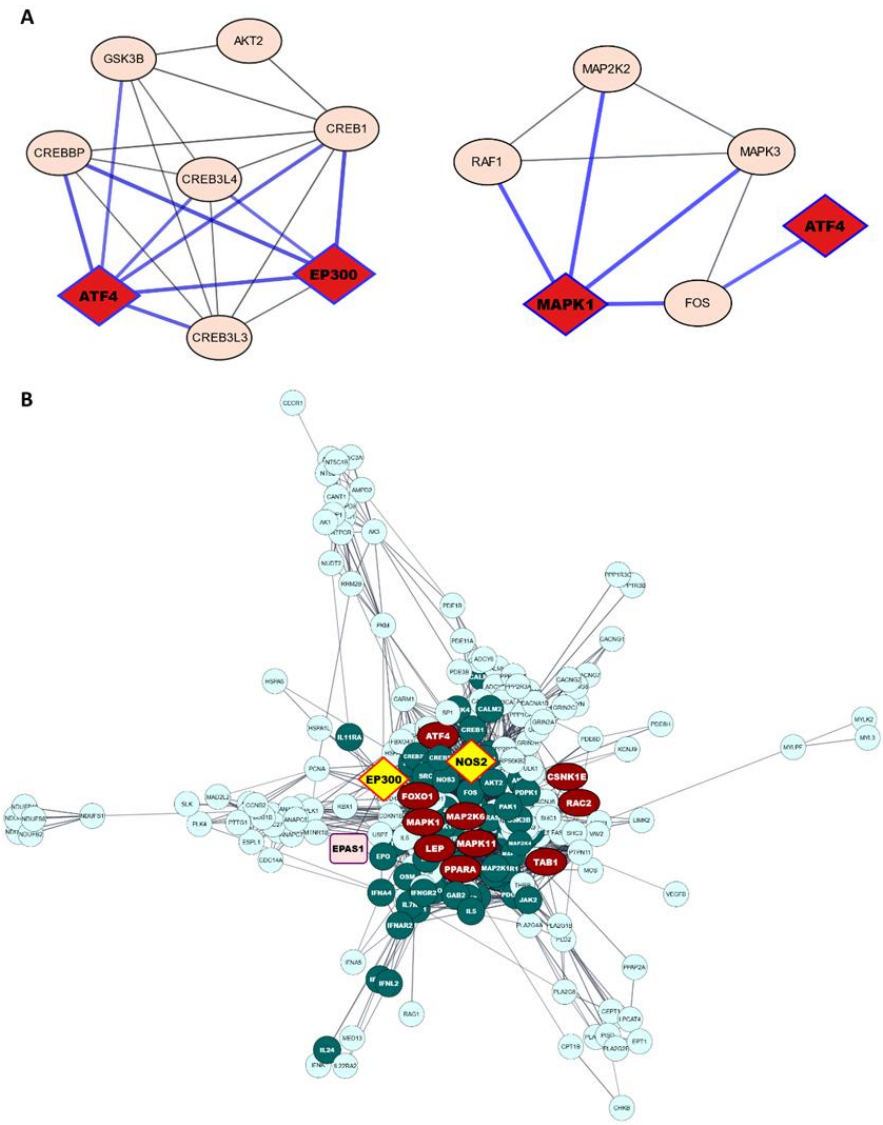

Figure 3. Significant gene networks enriched for Denisovan-like derived alleles according to the *Signet* analysis. (A) Partially overlapping significant networks belonging to *Cancer-related* and *Relaxin signalling* pathways including genes that mediate key functions in the modulation of cell proliferation and differentiation, in the promotion of angiogenesis, as well as in the regulation of vascular tone thought NO induction. Among them, *EP300* and genes functionally related to both *EP300* and *NOS2* candidate AI loci are highlighted in red. Genes supported by both *Signet* and *VolcanoFinder* analyses as potentially introgressed loci, as well as the associations that they establish in the networks are highlighted in blue. **(B)** Gene network built by setting co-expression as force function and by displaying the entire set of genes identified by the *Signet* algorithm as significantly enriched for Denisovan-like derived alleles. Genes whose variation pattern was supported by both *VolcanoFinder* and *Signet* analyses (e.g., *MAPK1*) as shaped by archaic introgression are displayed as dark red ellipses. *EP300* and *NOS2* genes, which we shortlisted as the most convincing candidate AI loci according to the *Signet* approach, *VolcanoFinder* analyses, and evidence advanced by previous studies supporting their adaptive role in high-altitude populations, are displayed as yellow diamonds. *EPAS1*, which was included manually in the network as a positive control locus that has been previously proved to have mediated adaptive introgression in Tibetan and Sherpa populations, was represented as pale pink rectangular. Genes included in pathways involved in angiogenesis and/or in the modulation of NO induction are reported as dark green circles, while the remining fraction of significant genes are represented as light-blue circles. The closeness or the distance between all nodes reflects the tendency to be co-expressed with each other and all the connections inferred are characterized by a confident score ≥ 0.7.

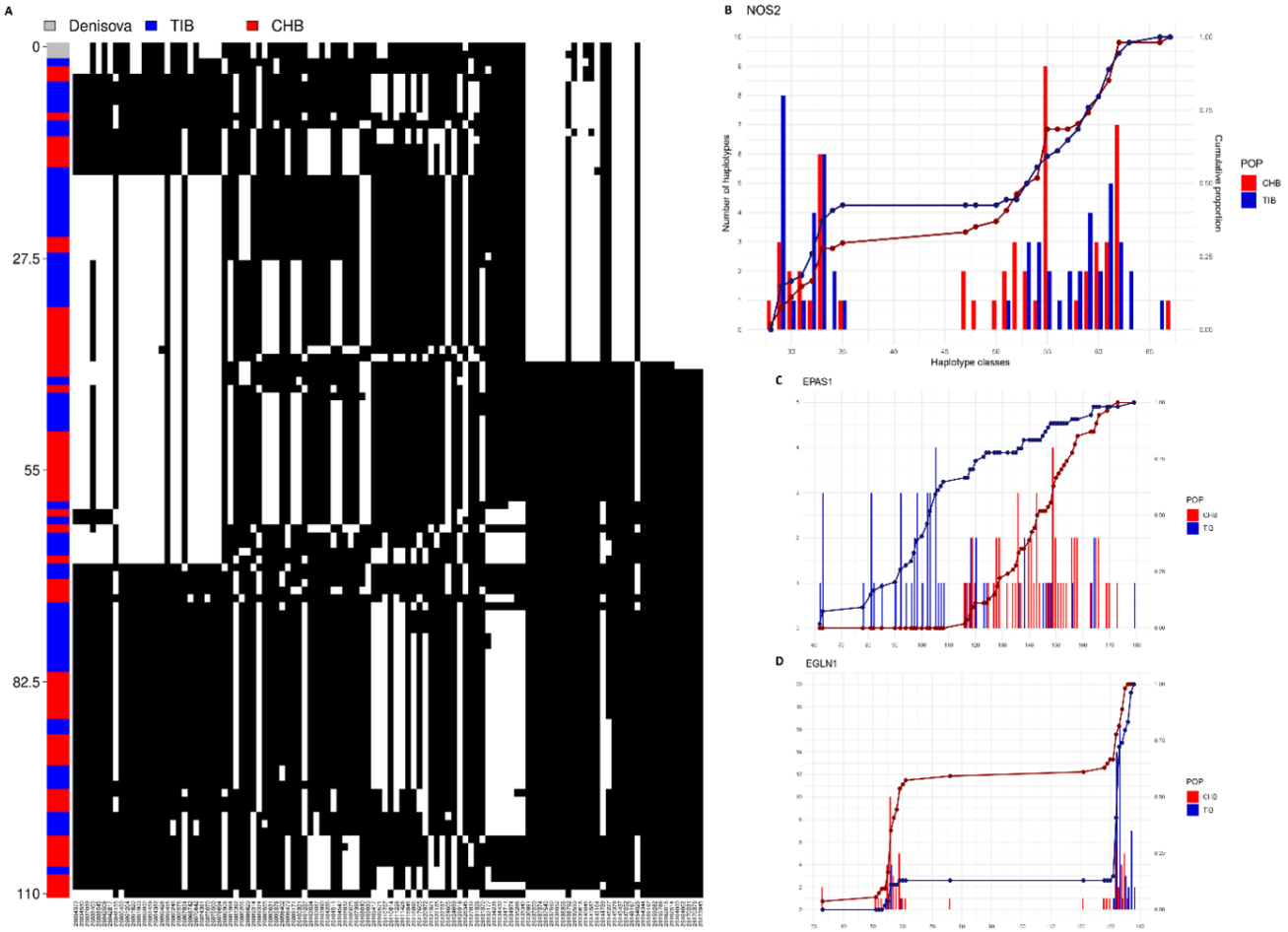

Figure 4 – Representation of genetic distances between modern and archaic haplotypes. (A) Heatmap displaying the divergence between Tibetan and Han Chinese *NOS2* haplotypes with respect to the Denisovan sequence. Haplotypes are reported in rows, while derived (i.e., black square) and ancestral (i.e., white square) alleles are displayed in columns. Haplotypes are ranked from top to bottom according to their number of pairwise differences with respect to the Denisovan sequence. A total of 14 haplotypes belonging to individuals with Tibetan ancestry are plotted in the upper part of the heatmap (i.e., the first quartile of haplotype distribution). These haplotypes account for 33% of the haplotypes inferred for Tibetans and are characterized by 32 or less pairwise differences with respect to the Denisovan sequence. **(B-D) Barplots showing the number and cumulative proportions of *NOS2, EPAS1* and *EGLN1* Tibetan and Han Chinese haplotypes in each haplotype class defined according to the number of pairwise differences between modern and archaic sequences.** In all plots are reported on the x-axis the haplotype classes, while on the first and on the second y-axes are indicated the number of haplotypes belonging to each haplotype class (i.e., blue/red bars) and the cumulative proportion of haplotypes (i.e., blue/red lines), respectively. **(B)** *NOS2* presents a pattern qualitatively comparable to the one displayed for the *EPAS1* gene, with 46% of haplotype classes presenting a greater value for the cumulative proportion of TIB haplotypes (i.e., blue line) rather than CHB ones (i.e., red line). **(C)** The *EPAS1* plot represents the trend expected for genes involved in AI events mediated by hard selective sweeps, in which TIB haplotypes (i.e., blue bars) are over-represented in all those haplotype classes presenting the smaller number of pairwise differences with Denisovan. In line with this, values obtained for the cumulative proportion of Tibetan haplotypes are higher with respect to those calculated for CHB haplotypes in all the haplotype classes considered. **(D)** The entire set of haplotype classes at *EGLN1* showed higher values for the cumulative proportion of CHB haplotypes with respect to the same proportion calculated for Tibetan haplotypes, with an over-representation of Han Chinese samples in the haplotype classes presenting few numbers of differences with Denisovan, as it may be expected for genes not involved in AI events.