

# Researchers Construct a Genome-wide Map of Adaptive Genetic Variants of Tibetan Highlanders

A recent study led by Dr. XU Shuhua from CAS-MPG Partner Institute for Computational Biology (PICB), CAS Shanghai Institute of Nutrition and Health (SIH), was published online in *National Science Review*, entitled “Prioritizing natural selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders”. This study constructed a map of adaptive genetic variants of Tibetan highlanders on the whole-genome scale. It is the very first effort for such a purpose based on the whole-genome deep sequencing data, and provides the first comprehensive panel of functional variants closely related to adaptive evolution of the Tibetan people.

The last decade witnessed intensive studies of Tibetan high-altitude adaptation. *EPASI*, a gene encodes hypoxia-inducible transcription factor to activate downstream genes to cope with particularly hypoxia, is so far the most well-recognized candidate gene for high-altitude adaptation of Tibetans, as well as other human and non-human species on the highlands. However, no adaptive functional variant has been identified in *EPASI*, leaving an unsolved problem in this field, and meanwhile bringing about some more fundamental questions: (1) How many genes or genetic variants have driven the adaptive evolution of humans to the high altitude? (2) Is the causal variant located within *EPASI* gene or anywhere else underlying the association between *EPASI* and the high-altitude adaptation? (3) Are there any other genes with greater contribution than *EPASI* to human adaptation to high-altitude?

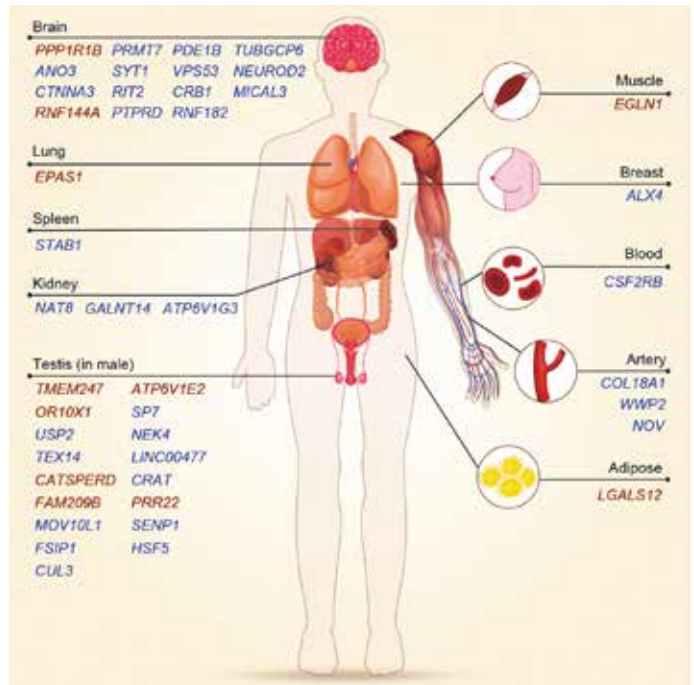
To address these questions, researchers conducted a systematic analysis of the Tibetan genomes. Taking advantage of the whole-genome deep sequencing, they constructed a map of adaptive genetic variants of high-altitude adaptation, including 1,877 key variants with

known functions: 63 missense, 7 loss-of-function, 1,298 evolutionarily conserved variants and 509 expression quantitative traits loci. These functional variants are very likely to contribute to the adaptive evolution of the Tibetans, but not necessarily in a direct way. High-altitude adaptation involves a wide range of phenotypic variations driven by enormously large numbers of variants and genes, and it could be even more complicated than some complex diseases. Moreover, the research team developed a statistic (*FIS*, functional importance score) to prioritize these identified adaptive genetic variants, and found that the top signal is not the well-known *EPASI*, but *TMEM247*, a transmembrane protein coding gene. Especially, a missense variant (rs116983452) located in *TMEM247* showed rather high frequency specifically in Tibetans. This key variant may largely increase the genetic differentiation between the lowlanders and Tibetan highlanders, by transforming Alanine, the common wild type in the lowlanders, to Valine, the Tibetan-specific mutant type. This mutant is carried by around 94% Tibetans, but is in low frequency or even missing in other world-wide modern human populations. It is so far the most highly differentiated missense variant between Tibetans and the lowlanders. Interestingly, this mutant also presents as a homozygote in a 50,000-year-old human genome that was found in the Denisova Cave in Siberia. Researchers estimated the age of the adaptive Tibetan sequences carrying *TMEM247*-rs116983452-T to be about 60,000 years ago, implying that this Tibetan-specific variant could be inherited from early inhabitants of Qinghai-Tibet Plateau with archaic ancestry.

In fact, it was a long journey for human beings to conquer the Qinghai-Tibet Plateau. A previous study by XU's team estimated that the genetic origin of the Tibetan highlanders could be traced back to around 40,000 to

60,000 years ago, in the middle-late Paleolithic. The early migrants to the plateau had extensive genetic admixture with each other, and had further gene flows with the latecomers, leading to the admixed descendants with very complex genetic makeup -- inherited from ancestry lineages of modern humans and archaic hominins (e.g. Altai Neanderthal, Denisovan and other unknown archaic species). During this process, some archaic genomic segments that were advantageous to the high-altitude adaptation were retained, and were accumulated to high frequencies through natural selection. *TMEM247*-rs116983452-T is a typical case of such a scenario.

The team found that the frequency of *TMEM247*-rs116983452-T is strongly and positively correlated with altitude. Further analyses revealed that this variant is significantly associated with the expression levels of *EPAS1* and *TMEM247*, and could possible regulate the concentrations of hemoglobin and red blood cell of Tibetans under hypoxia. For most lowland people, a long period of exposure to hypoxia may induce the production of red blood cell to increase the blood oxygen, but may finally lead to polycythemia. In contrast, the hemoglobin and red blood cell are kept in relatively lower levels in the Tibetan highlanders, and *TMEM247*-rs116983452-T could be one of the key genetic factors of this protective mechanism. By statistical modeling, the team found that *TMEM247*-rs116983452 shows greater effect size and better predicts the phenotypic outcome than any *EPAS1* variants in the association with adaptive traits in Tibetans, but interactions were also observed between *TMEM247*-rs116983452 and *EPAS1* variants, indicating that multiple variants may jointly deliver the fitness of the Tibetans on the Plateau, where a complex model is needed to elucidate the adaptive evolution mechanism. The adaptive genetic variants map provided by this study help to narrow down the targets for further investigations on the genetic basis and molecular mechanism of high-altitude adaptation of Tibetans, and provides new perspective for unrevealing the mystery of human conquest of the extreme environment at high altitude. Notably, this study proposed



Tissue- and organ-specific expression of the candidate genes of high-altitude adaptation (highlighted in red, such as *EPAS1* and *TMEM247*, is more likely to be responsive for high altitude adaptation in Tibetans). (Credit: Dr. XU Shuhua's group, PICB).

that multiple variants may jointly deliver the fitness of the Tibetans on the Plateau, where a complex model is needed to elucidate the adaptive evolution mechanism.

This work was conducted by Dr. DENG Lian, Dr. ZHANG Chao, Dr. YUAN Kai, and PhD students GAO Yang (from ShanghaiTech University) and PAN Yuwen from Dr. XU Shuhua's team, in collaboration with researchers from Kunming Institute of Zoology of CAS, Wenzhou Medical University, Fudan University, Xizang Minzu University, etc. It was funded by the grants from CAS, the National Natural Science Foundation of China, the National Key Research and Development Program and Science and Technology Commission of Shanghai Municipality.

(PICB)

## Reference

Lian Deng *et al.*, Prioritizing Natural Selection Signals from the Deep-Sequencing Genomic Data Suggests Multi-Variant Adaptation in Tibetan Highlanders. *National Science Review*, (2019). doi: 10.1093/nsr/nwz108.