Genomics 116 (2024) 110854

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Polymorphic transposable elements provide new insights on high-altitude adaptation in the Tibetan Plateau

Giorgia Modenini [a,*], Paolo Abondio [b], Marco Sazzini [a,c], Alessio Boattini [a]

[a] Dept. of Biological, Geological and Environmental Sciences, University of Bologna, Bologna, Italy
[b] IRCCS Istituto Delle Scienze Neurologiche Di Bologna, Bologna, Italy
[c] Interdepartmental Centre - Alma Mater Research Institute on Global Changes and Climate Change, University of Bologna, Italy

ARTICLE INFO

Keywords:
Structural variants
Retrotransposons
Polymorphic transposable elements
High altitude

ABSTRACT

Several studies demonstrated that populations living in the Tibetan plateau are genetically and physiologically adapted to high-altitude conditions, showing genomic signatures ascribable to the action of natural selection. However, so far most of them relied solely on inferences drawn from the analysis of coding variants and point mutations. To fill this gap, we focused on the possible role of polymorphic transposable elements in influencing the adaptation of Tibetan and Sherpa highlanders. To do so, we compared high-altitude and middle/low-lander individuals of East Asian ancestry by performing *in silico* analyses and differentiation tests on 118 modern and ancient samples. We detected several transposable elements associated with high altitude, which map genes involved in cardiovascular, hematological, chem-dependent and respiratory conditions, suggesting that metabolic and signaling pathways taking part in these functions are disproportionately impacted by the effect of environmental stressors in high-altitude individuals. To our knowledge, our study is the first hinting to a possible role of transposable elements in the adaptation of Tibetan and Sherpa highlanders.

1. Introduction

The peopling of the Tibetan Plateau, with an average elevation of 4000 m, by the ancestors of Tibetan and Sherpa highlanders, is one of the most compelling examples of Anatomically Modern Humans (AMH) adapting to a new and extreme environment [1–3]. The establishment of stable high-altitude settlements by the ancestors of Tibetan and Sherpa populations seems to have occurred only after the Last Glacial Maximum. Moreover, different studies suggested that more recent instances of migrations, admixture and geographical/cultural isolation could have further influenced the genetic variation of present-day Tibetan and Sherpa groups [4,5].

In the last decades, several population genomics and genome-wide association studies (GWAS) [1–3,6–9] tried to disentangle the genetic basis of high altitude adaptation (HAA), but so far, most of the studies have relied on single nucleotide polymorphisms (SNPs) to search for evidence of natural selection in Tibetan and Sherpa populations. Two genes, related to the hypoxia-inducible transcription factor (HIF) pathway, have been identified as under positive selection in these populations: *EPAS1* (endothelial PAS domain protein 1) and *EGLN1* (egl-9 family hypoxia-inducible factor 1) [1,10]. Furthermore, it has been demonstrated that *EPAS1* carries signals of adaptive introgression from Denisovan archaic hominins [11,12], who admixed with the ancestral population of both modern high-altitude and low-altitude East Asians.

Structural variation (SV) is an essential mutational force shaping the evolution and function of the human genome [13]. However, few studies [13,14] have analyzed the link between SVs and HAA and, to our knowledge, no one has focused on retrotransposons to date.

Retrotransposons are mobile genetic elements with the ability to replicate themselves and increase the number of their copies: indeed, sequences from retrotransposons constitute at least 40% of the human genome [15]. These elements are divided into long terminal repeats (LTR) retrotransposons, to which the human endogenous retrovirus (HERV) family belongs, and non-LTR retrotransposons, represented by short interspersed nuclear elements (SINEs, such as Alu-like elements, ~300 bases long), long interspersed nuclear elements (LINEs, complete elements are ~6 kilobases long, but frequently they are shorter due to 5′ truncation during insertion) and the composite family of SINE-VNTR-Alu (SVAs of variable length because of the presence of a Variable Number Tandem Repeat [VNTR] region). Among non-LTR retrotransposons, only LINE-1 s are autonomously active [16], while Alus and SVAs rely on LINE-1's machinery to mobilize themselves [17].

* Corresponding author.
E-mail address: giorgia.modenini2@unibo.it (G. Modenini).

https://doi.org/10.1016/j.ygeno.2024.110854
Received 16 January 2024; Received in revised form 23 March 2024; Accepted 30 April 2024
Available online 1 May 2024

Transposable Elements (TEs) have been an important source of genetic variation throughout human evolution, many of them being polymorphic and showing population-specific stratification [18]. Accordingly, they are valuable genetic markers for the study of human populations variability, as shown by several recent works [19–21]. Therefore, we decided to focus on the study of polymorphic TEs to disentangle the genetic basis of high-altitude adaptation in the Tibetan Plateau.

Here we provide the first large-scale study on 114 high-coverage published genomes from the Tibetan Plateau and East Asia in order to assess the role of polymorphic TEs in the HAA of populations settled along the Himalayan Arc. Together with modern genomes, we analyzed four high-coverage ancient and archaic DNA samples (two ancient Tibetans and two archaic hominins, namely Altai Neanderthal and Denisova) to provide a temporal context for the most significant results emerging from modern genomes analyses.

2. Materials and methods

2.1. Samples and study design

114 published modern high-coverage genomes (30×-45×) from high-altitude (HA), middle-altitude (MA) and low-altitude (LA) populations of East Asia and one African population (Yoruba) were included in this study, along with four high-coverage ancient/archaic DNA samples: one from a Denisovan individual [22], one from a Neanderthal individual [23] and two "ancient Tibetans" [24] (C1 and S10), spanning 3150–1250 years before present (yBP) (Supplementary Table S1 and Fig. 1). The 114 modern samples are represented by: 10 Sherpa [4,7,25,26], 28 Tibetans [4]; 8 Tujia, 8 Yi, 6 Naxi, 29 Han Chinese from the Human Genome Diversity Project (HGDP) [27] and 25 Yoruba from the 1000 Genomes Project (1KGP) phase 3 [28]. Individuals were selected avoiding relatives.

These samples were chosen to represent the modern genome variability of the Tibetan Plateau and East Asia. The African population was selected as an outgroup. The ancient samples were included to analyze the evolution of the HA populations of the Tibetan Plateau by searching for common variants between modern and ancient DNA samples.

Original *bam* files were first converted to the *fastq* format with the *bedtools* command *bamtofastq* [29]. All *fastq* files were then treated with AdapterRemoval [30] and subsequently aligned to the human reference genome GRCh38dh (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/GRCh38_reference_genome/GRCh38_reference_genome/GRCh38_full_analysis_set_plus_decoy_hla.fa) with bwa-mem [31]. The resulting *sam* files were processed with samtools fixmate to clean up read pairing information and flags and then sorted with samtools sort [32]. The obtained *bam* files were then indexed and processed with MarkDuplicates (see Picard: http://broadinstitute.github.io/picard).

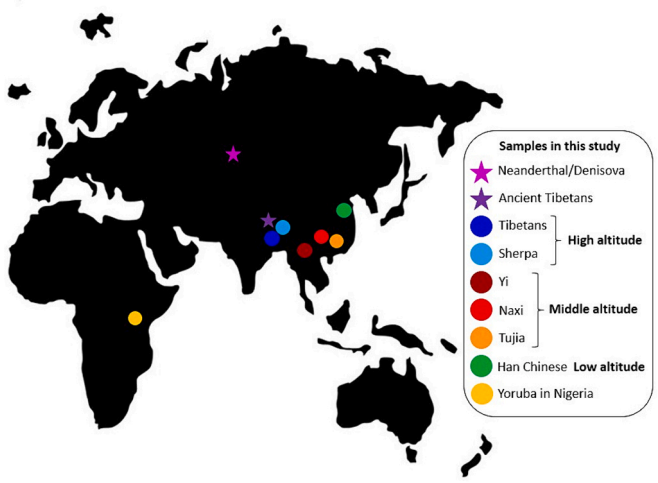

Fig. 1. Location of the modern and ancient samples analyzed in this study.

2.2. Identification of TEs in ancient and modern samples

The identification of polymorphic non-reference TEs (Alu, LINE-1, SVA and HERV-K) was performed on both ancient and modern samples using the software MELT v2.2.2 with the function 'Split' [20]. MELT-DEL was applied to retrieve information about polymorphic reference TEs (Alu, LINE-1 and SVA) and TypeTE [33] was then used to better genotype reference Alus. Only "PASS" sites were included in a single final VCF file. Fisher tests of independence with one and two degrees of freedom were performed to identify TEs with significantly different frequencies in HA compared to MA-LA populations. Tests were performed for both allelic and genotype frequencies. TEs that yielded significant results (nominal p-value <0.01) at least for allele frequencies ("differentiated TEs") were considered as putatively contributing to Tibetan and Sherpa differentiation.

2.3. Assessing population structure with PCA and Admixture

To assess the genetic relationships among the individuals included in our dataset, as well as their shared ancestry, a principal component analysis (PCA) and ADMIXTURE analysis [34] were performed on the TEs dataset. Quality control (QC) was performed with the PLINK software [35] on modern samples, including the removal of genetic elements belonging to sexual chromosomes, a check for the proportion of missing data (using the commands –geno 0.01 and –mind 0.01), the respect of Hardy-Weinberg equilibrium after Bonferroni correction for multiple testing (−-hwe 0.01/α, where α is equal to the number of variants remaining in the dataset at this stage of the QC procedure), the removal of rare variants (−-maf 0.01) and an assessment of linkage disequilibrium along the genome, using a sliding window of 500 bp, a moving step of 50 bp and a threshold value of 0.1 (−-indep-pairwise 500 50 0.1). After QC, PCA was performed on the TEs dataset by applying file format conversions as provided by the *convertf* and *smartpca* tools from the EIGENSOFT package v6.0.1 [36].

Similarly, the ADMIXTURE software [34] was employed to perform an estimation of shared genetic ancestry across populations. A number K of putative ancestral components between 2 and 12 was tested, and 50 iterations of each run were performed to minimize the error and maximize the log-likelihood of each ancestry estimate.

2.4. Testing the extent and direction of differentiation with Fst and PBS

To measure population differentiation due to genetic structure, the fixation index (Fst) was computed for all TEs used to perform PCA and Admixture analyses [37]. For the purpose of computing their genetic distance, three groups of modern individuals with shared genetic ancestry were extracted from our samples: high-altitude Tibetans and Sherpas (HA, 38 individuals), middle-altitude Tibeto-Burman speaking populations (MA, 22 individuals) and low-altitude Han Chinese (LA, 29 individuals). Fst was computed for all three population pairs (HA/MA; HA/LA; MA/LA) and each distribution was independently standardized by subtracting the average Fst from each score, then dividing the obtained value by the standard deviation of the distribution. Only normalized Fst scores falling in the tails of the distribution (*i.e.*, positive and negative regions exceeding 2 standard deviations) were considered significant for further inquiry. Significant Fst scores distinguishing HA from both MA and LA were detected and, to obtain signals potentially related to HA-induced differentiation, TEs with significant Fst scores also emerging in the MA/LA comparison were removed. This allowed us to focus specifically on putative HA-linked TEs by taking away contributions to MA/LA differentiation.

In addition, Population Branch Statistics (PBS) was also computed to

corroborate Fst observations and infer the directionality of differentiation, with the African Yoruba group (25 individuals) added to explore results coming from an outgroup of different ancestry [3]. PBS was computed for the following trios (where the third population is the outgroup): HA/MA/Yoruba; HA/LA/Yoruba; HA/MA/LA; MA/LA/Yoruba. The top 0.1% of PBS scores for each trio was deemed significant; moreover, TEs emerging from the last trio (middle altitude/Han Chinese/Yoruba) were used to filter out signals otherwise confirmed by the other three, so as to highlight genetic signatures characterizing unequivocally the high-altitude Tibetan and Sherpa groups.

2.5. In silico functional analyses

Information about the position of all detected non-reference TEs was retrieved from the MELT output, which includes the location of the TE ("null", when the TE is in an non-genic area; "intronic", "exon", "3_UTR", "5_UTR", "Promoter", "Terminator" when the TE is in a genic or regulatory region) and the gene name, if any, in RefSeq format. On the other hand, information about reference TEs location was retrieved using a self-customized python script after downloading genes annotation in *bed* format from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTables) and using as a reference the genomic locations identified by the following works: [38,39] for 3'UTRs and 5'UTRs (1000 bp from gene end and 210 bp from gene start, respectively); [40] for promoters (500 bp from 5' UTRs); [41] for terminators (between 250 and 1050 bp after 3' UTRs). All RefSeq IDs were then converted into the Official Gene Symbol with the software DAVID (Database for Annotation, Visualization and Integrated Discovery; https://david.ncifcrf.gov/conversion.jsp) [42,43] and used for further investigations. We also looked for the relevant diseases potentially related to the detected sets of genes using DAVID, which refers to the GAD database [44]. The statistical overrepresentation test was performed with Panther (http://pantherdb.org/), selecting the organism *Homo sapiens* and calculating false discovery rate (FDR) and Fisher test. The analysis of Gene Network and Pathways was done with the Kyoto Encyclopedia of Genes and Genomes, KEGG (https://www.genome.jp/kegg/mapper/search.html), selecting the organism *Homo sapiens* (*"hsa"*). Genes mapped by significant TEs were also compared with literature lists of genes deemed under positive selection in Tibetans [45] and candidate genes for polygenic adaptation in Tibetan and Sherpa populations [26].

After the identification of candidate TEs that possibly contributed to HAA, we checked whether these TEs were also present in the four considered archaic (Neanderthal and Denisova) and ancient (C1 and S10) individuals, by inspecting the MELT output.

Finally, we inferred a possible function for the candidate TEs by cross-checking our results with those provided by Cao and colleagues [46], who identified a list of TEs that act as expression/alternative splicing Quantitative Trait Loci (eQTL/sQTL).

2.6. Association test with GEMMA

An association test using "high altitude" (1) and "low altitude" (0) as binary pseudo-phenotypes was computed with the software GEMMA [47] on the whole variants dataset. Only individuals from HA (Tibetans and Sherpa) and MA + LA populations (Tibeto-Burman speaking populations and Han Chinese samples grouped together) were included in the analysis. Following instructions from the manual, we first calculated the relatedness matrix with the command -gk 2, meaning that the software calculated a standardized (rather than centered) relatedness matrix. Then, Wald's test was performed applying a linear model (−lm 1) on the previously estimated matrix. Only TEs with an adjusted *p*-value <0.001 (after Benjamini-Hochberg correction) were considered as significantly associated with high altitude ("associated TEs"). Results were plotted using the R package "CMplot" [48] (https://github.com/YinLiLin/CMplot).

3. Results

3.1. Identification of TEs in HA and LA populations

After using MELT-Split on 118 ancient and modern samples, we successfully identified 9144 polymorphic non-reference TEs, of which 7438 Alu, 1193 LINE-1, 492 SVA, 21 HERV-K, and 3754 polymorphic reference TEs, of which 3492 Alu, 169 LINE-1, 93 SVA. Based on the information provided by the MELT output, 49.4% of non-reference TEs are in non-genic regions ("null"), while 50.6% are in genic regions (42.1% are in introns, 3.8% in promoters, 3.9% in terminators, 0.6% in 3' UnTranslated Regions (UTRs), 0.1% in 5' UTRs and 0.1% in exons). Moreover, when looking at reference TEs, 54.7% are intergenic, 43.4% are intronic, 0.2% in exons, 0.6% at 3' UTRs, 0.05% at 5' UTRs, 0.15% in promoters and 0.9% in terminators. Finally, reference and non-reference TEs feature systematic differences in allele frequencies (Supplementary Fig. S1), in particular reference TEs are enriched for higher allele frequencies (> 0.5) while non-reference TEs are enriched in lower allele frequencies (< 0.5). In general, this is due to the fact that ref. TEs are identified based on their presence on a "single" genome (the reference) while non-ref TEs are detected based on many genomes (*i.e.* all the analyzed genomes).

We considered as putatively related with HAA only TEs that yielded significant Fisher tests (*p*-value <0.01) based on allele frequencies. Accordingly, when comparing HA (Sherpa and Tibetan) and MA-LA (Tibeto-Burman speaking and Han Chinese) populations we detected 271 significant TEs (154 non-reference TEs and 117 reference TEs) (Supplementary Table S2).

3.2. PCA and Admixture

In order to contextualize the variability of the human groups considered in this work, we performed PCA and ADMIXTURE analyses

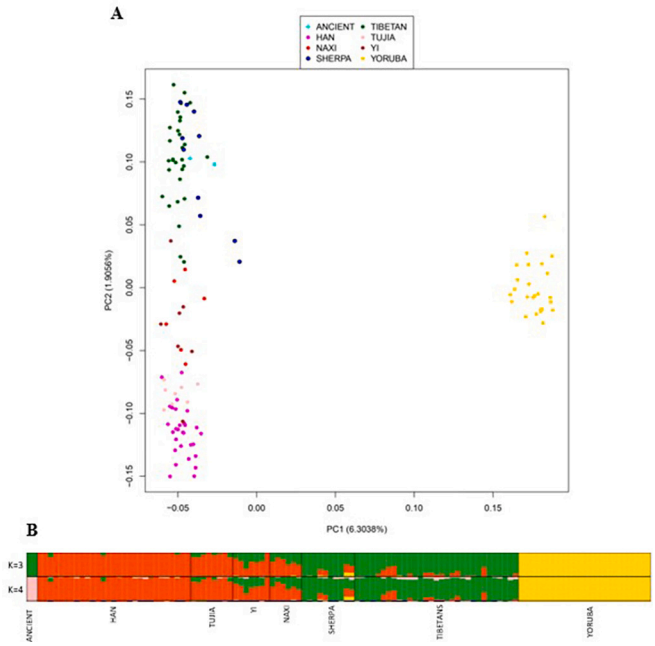

Fig. 2. PCA and Admixture plots. A) In the PCA, the first PC discriminates between African and non-African groups, while PC2 highlights a high-to-low altitude gradient with Tibetans and Sherpas at the top and Han Chinese at the bottom. B) Admixture plots showing results for K = 3 (lowest CV error = 0.44385) and K = 4, where the two Ancient Tibetans (C1 and S10) carry their own ancestry component ("pink"). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

based on TEs. The obtained results (Fig. 2B) confirm a clear distinction between African and non-African populations, as shown by the first principal component (PC1) in the PCA and by a complete separation of ancestry components between the Yoruba group ("yellow") and the individuals of Asian ancestry ("red" and "green") at K = 3, which is the number of components associated with the lowest CV error (0.44385). The second principal component (PC2) highlights some geographic structure between HA and LA individuals, who are distributed along a high-to-low altitude gradient including Tibetans and Sherpa at the top, then the Tibeto-Burman speaking groups (Naxi, Yi and Tujia), and finally the Han Chinese representatives at the bottom (Fig. 2A). The same gradient can be observed in the ADMIXTURE graph at K = 3, where the "green" component is predominant in HA Tibetans and Sherpa, while the "red" component is characteristic of the LA Han Chinese individuals and the Tibeto-Burman speaking groups show variable combinations of "red" and "green" components, with an increasingly higher proportion of Tibetan-like ancestry for the Tujia, Yi and Naxi groups, respectively. The Ancient Tibetan samples clearly overlap with Tibetans and Sherpas in the PCA, while they carry the "green" HA ancestry component in the ADMIXTURE plot at K = 3, confirming their tight relationship with the modern HA groups, although at K = 4 they acquire their own ancestry component ("pink"). The two individuals from the Kusunda population separate themselves from the other HA individuals both in the PCA and ADMIXTURE plots (Fig. 2B), where they show a mix of three ancestry components, whereas the other Sherpa and Tibetan samples have at most two components ("green" and/or "red").

3.3. Fst and PBS (scans testing differentiation)

Population differentiation has been evaluated by computing the fixation index, Fst, for all TEs in our panel and by comparing the high-altitude (HA, Tibetan and Sherpa), Tibeto-Burman speaking (MA) and low-altitude (LA) groups in pairs.

Although all scores are relatively low, the HA populations (Tibetan+Sherpa) exhibit the highest differentiation (Fst = 0.013) from the LA group (Han Chinese) and an intermediate level of differentiation with respect to the MA Tibeto-Burman speaking representatives (Fst = 0.006). Similarly, these last two groups show the lowest mean Fst score (Fst = 0.003).

After normalizing each Fst distribution, a total of 103 non-reference TEs significantly discriminate between the HA and MA groups, while 131 distinguish the HA and LA cohorts, and 115 characterize the comparison between MA and LA populations. By cross-referencing the significant scores for the three distributions, a total of 32 non-reference TEs emerge as able to discriminate between HA and both MA and LA populations (see Supplementary Table S3). To further corroborate this finding, the resulting non-reference TEs were cross-checked with those having a significant Fisher test score for allele counts between the HA and non-HA groups: all non-reference TEs discriminating HA from both MA and LA show Fisher p-values <0.01, confirming the significantly different presence of the transposable elements under scrutiny in the Tibetan and Sherpa groups with respect to the other two (see Supplementary Table S3).

To further define which TEs are preferentially differentiated in the HA group, rather than in the other two, three-way PBS distance analysis was carried out as described in the Materials and Methods section. Keeping into consideration the top 0.1% scores as significant among the performed tests and cross-referencing the results, a total of 62 non-reference TEs (58 Alus, 2 LINE-1 s, 2 SVAs) appear to be characteristic of differentiation in the direction of the HA group (see Supplementary Table S4), with 22 falling into promoters or introns of known genes. When intersecting the PBS and the afore-mentioned 32 non-reference TEs with significant Fst discriminating HA from MA and LA groups, eight non-reference TEs are shared (highlighted in bold in the Supplementary Tables S3 and S4), two of them being located in the genes *ASAH1-AS1* (acid ceramidase antisense RNA 1) and *PHF21 A* (PHD finger protein 21 A) on chromosomes 8 and 11, respectively.

3.4. In silico functional analyses

After the identification of 271 "differentiated" TEs, we retrieved information about their location and type and found that 126 of them are located in genic regions (71 non-reference TEs + 55 reference TEs).

By analyzing the corresponding genes with the software DAVID, we observed that the most represented disease classes, according to the GAD database [44], are "cardiovascular" (p-value = $1.5*e^{-6}$), "hematological" (p-value = $1.4*e^{-6}$) and "chem-dependency" (p-value = $1.9*e^{-9}$). Indeed, the most significant disorders/affected traits (p-value <0.001 after Bonferroni correction) are: Tobacco use disorder, Lipoproteins VLDL and Cholesterol LDL (Supplementary Table S5).

3.5. Association test with GEMMA

The association test with GEMMA on HA and MA-LA populations retrieved a total of 266 significant results ("associated" TEs, Supplementary Table S6: Adjusted p-value <0.001, as highlighted by the red-dotted line in Fig. 3). Since 123 TEs are located in genic regions, we used DAVID to perform functional annotation clustering, and retrieved information about the relationship between those genes and diseases from the GAD database [44]. Interestingly, the most significant conditions reflect the previously mentioned patterns, such as "Tobacco use disorder", represented by 51 genes (p-value = $2.7*e^{-11}$). Other significant conditions were related to cardiovascular traits (*i.e.*, "Blood Pressure", "Erythrocyte Count", "Heart Rate", "Hemoglobin A Glycosylated", "Glomerular Filtration Rate"), body measurements ("Body Mass Index"), "cholesterol LDL" and "Insulin".

3.6. Identification of TEs that possibly contributed to HAA

After the identification of four sets of TEs potentially contributing to HAA in Tibetan and Sherpa populations (271 "differentiated" TEs, 266 "associated" TEs, 32 significant TEs for Fst and 62 for PBS), we compared these four lists: 11 TEs are shared among three tests out of four (Table 1). Then, we verified their presence or absence in the two archaic hominins (Neanderthal and Denisova) and two ancient Tibetan individuals (C1 and S10). We also performed an *in silico* analysis to infer a possible function for the candidate TEs by cross-checking our results with those provided by Cao and colleagues [46]. We also cross-checked the list of genes mapped by the four sets of significant TEs with those under positive selection in [45,49] and found 10 genes in common (Supplementary Table S7). For instance, the gene *SUPT3H* is mapped by an AluYe that is significant for PBS and differential allele frequencies analyses: the gene is under positive selection in Tibetans according to Deng and colleagues [49], who performed a composite of multiple signals (CMS) analysis. Moreover, by looking at the lists of candidate genes for polygenic adaptation in Tibetan and Sherpa populations [26], it emerged that the same gene *SUPT3H* - and other five genes - are shared.

The most interesting TE, based on all our *in silico* analyses, is an AluYe on chromosome 12:44670594 in the gene *NELL2* (Neural EGFL-like 2), which is involved in tobacco use disorder. The Alu has significantly different allele frequencies between HA and MA-LA populations (Fisher p-value = 0.00131) and is a significant result also for the association test performed with GEMMA (adjusted p-value = $1.81*e^{-04}$). Moreover, it is one of the significant results for PBS, with a score = 0.168776. By looking at the putative function of this Alu, it acts as eQTL in skin and as sQTL in brain cortex and putamen basal ganglia. Finally, it is present in Denisova and one ancient Tibetan (C1).

4. Discussion

Different studies used TEs as genetic markers to study modern human variability and differentiation [19–21]. Therefore, we decided to

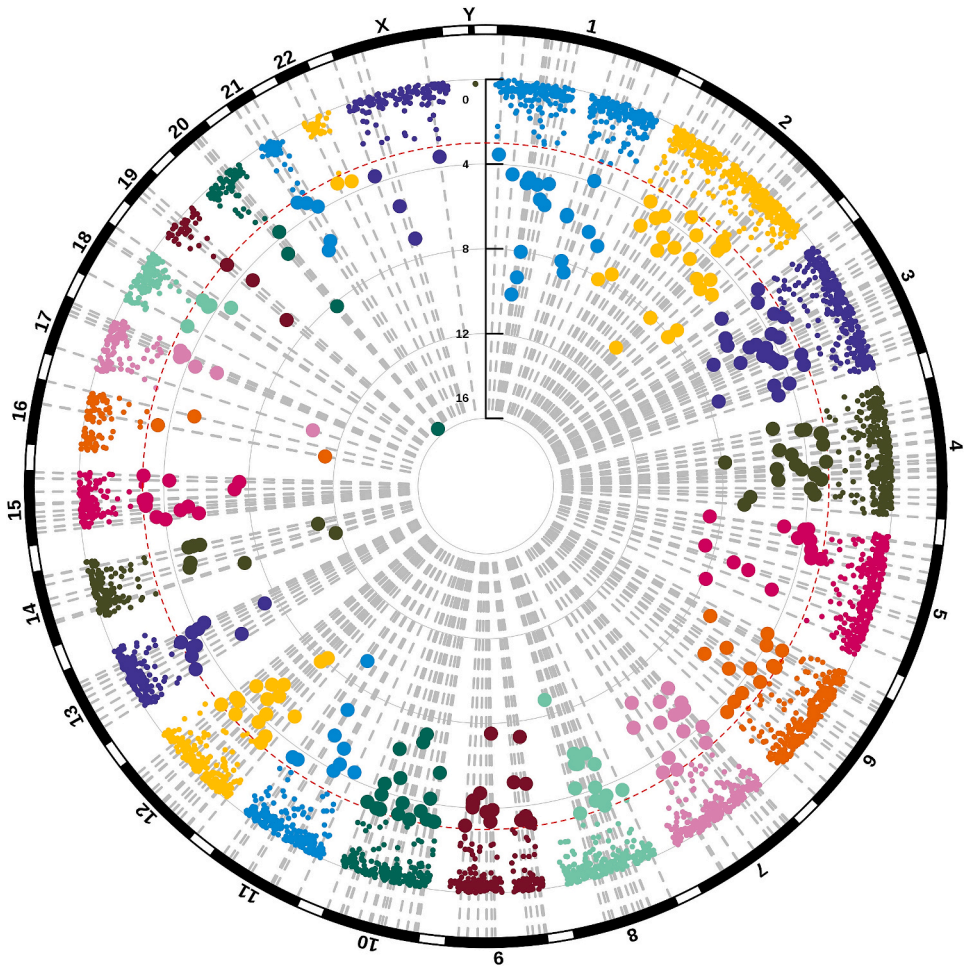

Fig. 3. Circular manhattan plot of the significant TEs associated with high altitude. TEs that yielded an adjusted p-value <0.001 (red-dotted line) have a bigger size than the others. *P*-values are shown as $Log_{10}P$. Plotting of results was performed with the "CMplot" R package (https://github.com/YinLiLin/CMplot). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
The 11 TEs shared among the four different tests (differential allele frequencies; association test with GEMMA; Fst; PBS). Chr_pos = chromosome + position; TE type = TE subfamily (INS = non-reference TE); Location = location of the TE ("null" = intergenic); Gene = gene in which is located the TE; Tests = tests for which the TE is significant (1 = differential allele frequencies; 2 = association test with GEMMA; 3 = Fst; 4 = PBS); eQTL/sQTL = the polymorphic TE acts as eQTL and/or sQTL; Altai/Den/C1/S10 = the TE was found also in ancient or archaic individuals (Altai Neanderthal, Denisova, C1 and S10; 1 = presence of the TE; 0 = absence; NA = missing data).

Chr_pos	TE type	Location	Gene	Tests	eQTL	sQTL	Altai	Den	C1	S10
12_44670594	INS-AluYe	intronic	NELL2	1, 2, 4	X	X	0	1	1	NA
14_64710808	INS-AluY	intronic	PLEKHG3	1, 2, 4			0	0	NA	0
3_120524105	INS-AluY	null	null	1, 3, 4			0	0	NA	1
11_45940658	INS-AluY	intronic	PHF21A	1, 3, 4			0	0	1	0
13_28811284	INS-AluY	null	null	1, 3, 4			0	0	0	0
2_137868315	INS-AluYg	null	null	1, 3, 4			0	0	NA	0
5_144951450	INS-L1Ta	null	null	1, 3, 4			0	0	0	1
6_81458432	INS-AluY	null	null	1, 3, 4			0	0	1	1
8_18085627	INS-AluYb8	intronic	ASAH1-AS1	1, 3, 4	X	X	NA	0	0	0
9_100940618	INS-AluYb3a1	null	null	1, 3, 4			NA	0	0	0
6_16950725	INS-L1Ambig	null	null	1, 2, 3	X		0	0	NA	0

investigate the possible influence of polymorphic TEs on the differentiation of East Asian populations exposed and adapted to an extreme environment, as represented by the high-altitude villages of the Tibetan Plateau inhabited by Tibetan and Sherpa ethnic groups.

As a first step, we performed two population genetics analyses, PCA and Admixture, on 118 ancient and modern individuals to contextualize their variability. Both of them confirm that retrotransposons are valuable genetic markers to study population differentiation: indeed, as shown in Fig. 2A, there is a clear distinction between the African group (represented by Yorubas) and the East Asian populations in terms of principal components and ancestry components. Moreover, the PCA plot shows an altitudinal gradient between low/middle-landers (Han Chinese, Yi, Tujia and Naxi) and high-landers (Tibetan and Sherpa). The Admixture graph (Fig. 2B) identifies two main ancestry components: "green" for Tibetans and Sherpa, and "red" for Han Chinese, with the Tibeto-Burman speaking groups (Yi, Tujia and Naxi) showing a mixture

of the two. Previous works based on SNPs [4,7,24,26] detected a similar pattern of variability in the same populations.

Fixation index (Fst) and PBS statistics computed using TEs confirm a subtle but tangible differentiation among the three analyzed groups, with an increase in average Fst following an altitudinal gradient: the HA cohort shows a higher Fst value when compared to the LA group (0.013) than the MA subjects (0.006). This is once again in line with what has emerged in previous works based on SNPs [4,7,24,26]. Only non-reference TEs were deemed significant by Fst, while for PBS there is a mixture of both reference and non-reference significant TEs. Such discrepancies are likely to emerge as a consequence of the systematic differences in allele frequencies between the two TEs groups (Supplementary Fig. S1), with reference TEs enriched for higher allelic frequencies (> 0.5) and non-reference TEs enriched for lower allelic frequencies (< 0.5), combined with the analytical approach of each method.

PBS analysis has been useful to detect two polymorphic TEs falling in genetic elements, *ASAH1-AS1* (acid ceramidase antisense RNA 1) and *PHF21 A* (PHD finger protein 21 A). Limited literature exists on *ASAH1-AS1*, which codes for the antisense sequence of the proximally located *ASAH1* gene. A recently published analysis of long noncoding RNAs (lncRNAs) in head and neck squamous cell carcinoma does associate its methylation status with predictive prognostic power [50]. The gene itself has never been studied in individuals of Asian ancestry, but several genome-wide association studies (GWAS) in cohorts of European and African descent in America and the Caribbean detected several SNPs associated with traits indicative of obesity [51], bone mineral density and serum urate levels in females [52], and coronary artery calcification in type 2 diabetes [53] among others, although few signals reach the *p*-value threshold of significance for GWAS (10^{-8}). Indeed, phenotypes similar to those above listed have been described by their relationship with high altitude in cohorts of Tibetan and Chinese ancestry, highlighting significant differences with low altitude groups [54–56]. Moreover, the *ASAH1* gene regulated by its antisense RNA encodes a ceramidase enzyme, which drives the degradation of the waxy lipid ceramide in its two components, sphingosine and a fatty acid. *ASAH1* has been variably described in the context of senescent cell permanence, neuron survival and neurite structure maintenance [57,58], possibly influencing the composition of lipid rafts in the cellular membrane and, therefore, its stability and the functionality of the attached proteins. Interestingly, the *ASAH1*-coded acid ceramidase is a lysosomal protein that also protects the cell from oxidative stress, as shown both in retinal cells overexpressing *ASAH1* [59] and, indirectly, in a cellular model of Parkinson's disease through inhibition of ceramide synthesis [60]. Furthermore, recent studies have highlighted a role of *ASAH1* genetic variants in regulating the outcome of physical activity and exercise interventions [61,62]. In this context, *ASAH1* may serve a crucial role, as it is known that hypobaric hypoxia due to high altitude exposure induces inflammation by increasing the circulating levels of reactive oxygen species, which are pervasive, unstable molecules with high oxidizing power, and this is exacerbated by physical activity [62], as well as being responsible for premature cell senescence and apoptosis [62–65].

The *PHF21A* gene has been identified in the context of RNA switches induced by hypoxia that regulate oncogenic genes, as already verified for example with the vascular endothelial growth factor alpha (*VEGFA*) [66]. Indeed, it appears as though *PHF21A* mRNA is regulated at a translational level by hypoxia-induced switches and that, together with other genes, it composes a vast translational regulon modulating hypoxia resistance and cell survival [66]. The expression of *PHF21A* is also found to be significantly reduced, and the gene possibly downregulated by miRNAs, in human fibroblast-like synoviocytes, which produce pro-inflammatory mediators in osteoarthritis (OA) and are the cause of synovial pathology associated with OA [67]. Again, this may be in line with epidemiological observations pointing to an increased incidence of inflammatory-mediated musculoskeletal pathologies at high altitude and in colder environments [68,69], as well as generalized hypoxia-induced osteopenia [70], although direct dissection of the involved molecular mechanisms are still unknown in this case.

By calculating differential allele frequencies on the 11,192 identified TEs, we detected 271 TEs that discriminate between high-altitude and low/middle-altitude groups (Fisher *p*-value <0.01): 126 are in genic regions and, interestingly, the most represented disease classes (according to DAVID) [42,43] are "cardiovascular", "hematological" and "chem-dependency".

To strengthen our results, we also applied an association test using GEMMA [47] that identified 266 TEs associated with the "altitude" context (Adjusted p-value <0.001): of these, four are shared between the two analyses. The condition "tobacco use disorder" (represented by 51 genes mapped by as many TEs, p-value $= 2.7*e^{-11}$) was deemed significant by both differential allele frequencies analysis and association test. As for tobacco use, some studies investigated the influence of smoking at high-altitude, but the results have been so far controversial: some groups identified a correlation between smoking and a lower incidence of Acute Mountain Sickness (AMS) [71–73], while others point towards an absence of such correlation [74–76], with Wu and colleagues [71] suggesting that smoking "slightly decreases the risk of AMS but impairs long-term altitude acclimatization and lung function during a prolonged stay at high altitude". A previous work by Ramirez and colleagues [77], who studied the populations of the Tuquerres Plateau in southern Colombian Andes (3000 m above sea level), reported that in smokers there is "an increase in hemoglobin and hematocrit and a higher mean corpuscular volume and mean corpuscular hemoglobin concentration than in non-smoking high altitude subjects".

However, this hypothesis seems inconsistent for the Tibetan populations case. In fact, tobacco, whose origins can be traced back to the American continent [78], has been introduced in Europe and Asia in the last few centuries, making it an unlikely source of selective pressures in the Tibetan Plateau; on the contrary, it may have had a more relevant influence in Andean populations, which have been exposed to this substance for millennia. Therefore, we speculate that the detected condition acts as a proxy for other physiological traits or that the involved genes, such as *NELL2*, fulfill other pleiotropic roles in metabolisms influenced by the high altitude condition.

Indeed, it is well known that hypoxia (*i.e.*, reduction of oxygen intake with increasing altitude) induces physiological and morphological variations in the human brain [79,80], even though the genetic underpinnings of these changes remain largely unknown. Accordingly, we hypothesize that *NELL2*, which is specifically expressed in neural tissues [81,82] and stimulates neuronal polarization as well as axon growth [83], could contribute to altitude-driven functional changes in the brain, even if at the moment there is no experimental evidence of it.

On the whole, according to all our *in silico* analyses, the AluYe on chromosome 12:44670594, located in the gene *NELL2*, arose as the most significant result (Table 1). The Alu discriminates between HA and MA-LA groups (based on allele frequencies, Fisher *p*-value $= 0.00131$); more precisely, we observe that the Alu is characterized by a steady reduction of both presence and homozygosity following a decreasing altitudinal cline. In addition, it is also associated with the "altitude" condition (according to GEMMA: adjusted p-value $= 1.81*e^{-04}$), suggesting a meaningful relationship between the presence of the polymorphic TE in this gene and physiological responses to high altitude. Moreover, it is one of the significant results according to PBS (score $= 0.168776$), pinpointing a relevant role for this variant in the process of local differentiation between high- and low- altitude populations. This Alu also acts as eQTL in skin and as sQTL in brain cortex and putamen basal ganglia [46], indicating a possible regulatory role for this element on *NELL2*. The AluYe is present in Denisova and in the ancient Tibetan C1 (3150 yBP), suggesting that this element could have anciently emerged.

Finally, we cross-checked genes mapped by significant TEs for the four applied tests (differential allele frequencies analysis, Fst, PBS and association test with GEMMA) with the lists of genes under positive selection in Tibetan populations [45,49], and candidate gene for

polygenic adaptation in Tibetan and Sherpa groups [26]. This way, we found 15 genes mapped by significant TEs which experienced instances of positive selection/polygenic adaptation in high-altitude populations of the Tibetan Plateau. For instance, the gene *SUPT3H* is mapped by an AluYe significant for PBS and differential allele frequencies analysis and the gene has been found under positive selection/polygenic adaptation in Tibetan groups [26,45,49]. More precisely, *SUPT3H* is a probable transcriptional activator [84]. Interestingly, the locus containing this protein-coding gene was reported by a previous study to be associated with bone and cartilage phenotypes [85] and another work found variants in the *SUPT3H-RUNX2* locus to be involved in craniofacial phenotypes [86].

The present study highlights the putative contribution of the non-coding genome in high-altitude adaptation in a cohort of Tibetan and Sherpa individuals, when compared to mid-altitude and low-altitude populations of similar ancestry and geographic origins. Indeed, several transposable elements show a significant differentiation between the inhabitants of the Tibetan plateau and the other analyzed groups (as detected by Fst and PBS statistics), revealing a possible role in the control of peculiar genes involved in oxidative stress and hypoxia-induced inflammation, which themselves modulate the expression or translation of other genes. Furthermore, the prevalence of TEs with significantly different frequencies between HA and MA-LA groups and associated to the "altitude" context (according to GEMMA) similarly highlights genes involved in cardiovascular, hematological, chem-dependent and respiratory conditions, indicating that metabolic and signaling pathways taking part in these functions are disproportionately impacted by the effect of environmental stressors in HA individuals through both coding and regulatory elements. This extensive nested modulation, also pointed out by the fact that some of the detected TEs are quantitative trait loci influencing expression and/or alternative splicing, may be suggestive of a wider network of relationships between coding and non-coding elements, which intervenes in fine-tuning the physiological responses to high altitude environments.

Funding

The Authors declare that no funding was associated with this work.

CRediT authorship contribution statement

Giorgia Modenini: Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Paolo Abondio:** Writing – review & editing, Writing – original draft, Formal analysis, Investigation. **Marco Sazzini:** Writing – review & editing, Supervision, Conceptualization. **Alessio Boattini:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The Authors declare that they have no competing interests.

Data availability

vcf files containing information about transposable elements identified in this work are available from the Authors upon reasonable request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2024.110854.

References

[1] C.M. Beall, G.L. Cavalleri, L. Deng, R.C. Elston, Y. Gao, J. Knight, C. Li, J.C. Li, Y. Liang, M. McCormack, H.E. Montgomery, H. Pan, P.A. Robbins, K.V. Shianna, S. C. Tam, N. Tsering, K.R. Veeramah, W. Wang, P. Wangdui, M.E. Weale, Y. Xu, Z. Xu, L. Yang, M.J. Zaman, C. Zeng, L. Zhang, X. Zhang, P. Zhaxi, Y.T. Zheng, Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 11459–11464, https://doi.org/10.1073/pnas.1002443107.

[2] T.S. Simonson, Y. Yang, C.D. Huff, H. Yun, G. Qin, D.J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, L.B. Jorde, J.T. Prchal, R. Ge, Genetic evidence for high-altitude adaptation in Tibet, Science 329 (2010) 72–75, https://doi.org/10.1126/science.1189406.

[3] X. Yi, Y. Liang, E. Huerta-Sanchez, X. Jin, Z.X.P. Cuo, J.E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T.S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, P. Asan, G. Ni, J. Tian, X. Xu, T. Liu, R. Jiang, G. Wu, M. Zhou, J. Tang, T. Qin, S. Wang, G. Li Feng, J. Huasang, W. Luosang, F. Wang, Y. Chen, X. Wang, Z. Zheng, Z. Li, G. Bianba, X. Yang, S. Wang, G. Tang, Y. Gao, Z. Chen, L. Luo, Z. Gusang, Q. Cao, W. Zhang, X. Ouyang, H. Ren, H. Liang, Y. Zheng, J. Huang, L. Li, K. Bolund, Y. Kristiansen, Y. Li, X. Zhang, R. Zhang, S. Li, H. Li, R. Yang, J. Nielsen, J. Wang, Sequencing of 50 human exomes reveals adaptation to high altitude, Science 329 (2010) 75–78, https://doi.org/10.1126/science.1190371.

[4] D. Lu, H. Lou, K. Yuan, X. Wang, Y. Wang, C. Zhang, Y. Lu, X. Yang, L. Deng, Y. Zhou, Q. Feng, Y. Hu, Q. Ding, Y. Yang, S. Li, L. Jin, Y. Guan, B. Su, L. Kang, S. Xu, Ancestral origins and genetic history of Tibetan highlanders, The American Journal of Human Genetics 99 (2016) 580–594, https://doi.org/10.1016/j.ajhg.2016.07.002.

[5] J.Z. Li, D.M. Absher, H. Tang, A.M. Southwick, A.M. Casto, S. Ramachandran, H. M. Cann, G.S. Barsh, M. Feldman, L.L. Cavalli-Sforza, R.M. Myers, Worldwide human relationships inferred from genome-wide patterns of variation, Science 319 (2008) 1100–1104, https://doi.org/10.1126/science.1153717.

[6] S. Xu, S. Li, Y. Yang, J. Tan, H. Lou, W. Jin, L. Yang, X. Pan, J. Wang, Y. Shen, B. Wu, H. Wang, L. Jin, A genome-wide search for signals of high-altitude adaptation in Tibetans, Mol. Biol. Evol. 28 (2011) 1003–1011, https://doi.org/10.1093/molbev/msq277.

[7] C. Jeong, G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D.B. Witonsky, J. K. Pritchard, C.M. Beall, A. Di Rienzo, Admixture facilitates genetic adaptations to high altitude in Tibet, Nat. Commun. 5 (2014) 3281, https://doi.org/10.1038/ncomms4281.

[8] H. Hu, N. Petousi, G. Glusman, Y. Yu, R. Bohlender, T. Tashi, J.M. Downie, J. C. Roach, A.M. Cole, F.R. Lorenzo, A.R. Rogers, M.E. Brunkow, G. Cavalleri, L. Hood, S.M. Alpatty, J.T. Prchal, L.B. Jorde, P.A. Robbins, T.S. Simonson, C. D. Huff, Evolutionary history of Tibetans inferred from whole-genome sequencing, PLoS Genet. 13 (2017) e1006675, https://doi.org/10.1371/journal.pgen.1006675.

[9] J. Yang, Z.-B. Jin, J. Chen, X.-F. Huang, X.-M. Li, Y.-B. Liang, J.-Y. Mao, X. Chen, Z. Zheng, A. Bakshi, D.-D. Zheng, M.-Q. Zheng, N.R. Wray, P.M. Visscher, F. Lu, J. Qu, Genetic signatures of high-altitude adaptation in Tibetans, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 4189–4194, https://doi.org/10.1073/pnas.1617042114.

[10] K. Xiang, Y. Ouzhuluobu, Z. Peng, X. Yang, C. Zhang, H. Cui, M. Zhang, Y. Zhang Li, Gonggalanzi Bianba, Ciwangsangbu Basang, T. Wu, H. Chen, H. Shi, X. Qi, B. Su, Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation, Mol. Biol. Evol. 30 (2013) 1889–1898, https://doi.org/10.1093/molbev/mst090.

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

[12] X. Zhang, K.E. Witt, M.M. Bañuelos, A. Ko, K. Yuan, S. Xu, R. Nielsen, E. Huerta-Sanchez, The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans, Proc. Natl. Acad. Sci. U. S. A. 118 (2021) e2020803118, https://doi.org/10.1073/pnas.2020803118.

[13] C. Quan, Y. Li, X. Liu, Y. Wang, J. Ping, Y. Lu, G. Zhou, Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression, Genome Biol. 22 (2021) 159, https://doi.org/10.1186/s13059-021-02382-3.

[14] H. Lou, Y. Lu, D. Lu, R. Fu, X. Wang, Q. Feng, S. Wu, Y. Yang, S. Li, L. Kang, Y. Guan, B.-P. Hoh, Y.-J. Chung, L. Jin, B. Su, S. Xu, A 3.4-kb copy-number deletion near EPAS1 is significantly enriched in high-altitude tibetans but absent from the denisovan sequence, The American Journal of Human Genetics 97 (2015) 54–66, https://doi.org/10.1016/j.ajhg.2015.05.005.

[15] A.P.J. de Koning, W. Gu, T.A. Castoe, M.A. Batzer, D.D. Pollock, Repetitive elements may comprise over two-thirds of the human genome, PLoS Genet. 7 (2011) e1002384, https://doi.org/10.1371/journal.pgen.1002384.

[16] J.L. Goodier, Restricting retrotransposons: a review, Mob. DNA 7 (2016) 16, https://doi.org/10.1186/s13100-016-0070-z.

[17] L. Guio, J. González, New insights on the evolution of genome content: Population dynamics of transposable elements in flies and humans, in: M. Anisimova (Ed.), Evolutionary Genomics, Springer, New York, New York, NY, 2019, pp. 505–530, https://doi.org/10.1007/978-1-4939-9074-0_16.

[18] Y. Wang, B. Zhao, J. Choi, E.A. Lee, Genomic approaches to trace the history of human brain evolution with an emerging opportunity for transposon profiling of ancient humans, Mob. DNA 12 (2021) 22, https://doi.org/10.1186/s13100-021-00250-2.

[19] L. Rishishwar, C.E. Tellez Villa, I.K. Jordan, Transposable element polymorphisms recapitulate human evolution, Mob. DNA 6 (2015) 21, https://doi.org/10.1186/s13100-015-0052-6.
[20] E.J. Gardner, V.K. Lam, D.N. Harris, N.T. Chuang, E.C. Scott, W.S. Pittard, R. E. Mills, 1000 Genomes Project Consortium, S.E. Devine, The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology, Genome Res. 27 (2017) (1916–1929), https://doi.org/10.1101/gr.218032.116.
[21] W.S. Watkins, J.E. Feusier, J. Thomas, C. Goubert, S. Mallick, L.B. Jorde, The simons genome diversity project: a global analysis of mobile element diversity, Genome Biol. Evol. 12 (2020) 779–794, https://doi.org/10.1093/gbe/evaa086.
[22] M. Meyer, M. Kircher, M.-T. Gansauge, H. Li, F. Racimo, S. Mallick, J.G. Schraiber, F. Jay, K. Prüfer, C. de Filippo, P.H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R.E. Green, K. Bryc, A.W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J. Kitzman, M.F. Hammer, M.V. Shunkov, A.P. Derevianko, N. Patterson, A.M. Andrés, E.E. Eichler, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, A high-coverage genome sequence from an archaic denisovan individual, Science 338 (2012) 222–226, https://doi.org/10.1126/science.1224344.
[23] K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P.H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J.C. Mullikin, S.H. Vohr, R. E. Green, I. Hellmann, P.L.F. Johnson, H. Blanche, H. Cann, J.O. Kitzman, J. Shendure, E.E. Eichler, E.S. Lein, T.E. Bakken, L.V. Golovanova, V. B. Doronichev, M.V. Shunkov, A.P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains, Nature 505 (2014) 43–49, https://doi.org/10.1038/nature12886.
[24] C. Jeong, A.T. Ozga, D.B. Witonsky, H. Malmström, H. Edlund, C.A. Hofman, R. W. Hagan, M. Jakobsson, C.M. Lewis, M.S. Aldenderfer, A. Di Rienzo, C. Warinner, Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 7485–7490, https://doi.org/10.1073/pnas.1520844113.
[25] S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y.S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I.G. Romero, A.R. Jha, D.M. Behar, C.M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M.S. Abdullah, A. Ruiz-Linares, C.M. Beall, A. Di Rienzo, C. Jeong, E.B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B.M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J.T.S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M.F. Hammer, T. Kivisild, W. Klitz, C.A. Winkler, D. Labuda, M. Bamshad, L.B. Jorde, S.A. Tishkoff, W.S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons genome diversity project: 300 genomes from 142 diverse populations, Nature 538 (2016) 201–206, https://doi.org/10.1038/nature18964.
[26] G.A. Gnecchi-Ruscone, P. Abondio, S. De Fanti, S. Sarno, M.G. Sherpa, P.T. Sherpa, G. Marinelli, L. Natali, M. Di Marcello, D. Peluzzi, D. Luiselli, D. Pettener, M. Sazzini, Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes, Genome Biol. Evol. (2018), https://doi.org/10.1093/gbe/evy233.
[27] A. Bergström, S.A. McCarthy, R. Hui, M.A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S. Felkel, P. Hallast, J. Kamm, H. Blanché, J.-F. Deleuze, H. Cann, S. Mallick, D. Reich, M.S. Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic variation and population history from 929 diverse genomes, Science 367 (2020) eaay5012, https://doi.org/10.1126/science.aay5012.
[28] 1000 Genomes Project Consortium, A. Auton, L.D. Brooks, R.M. Durbin, E. P. Garrison, H.M. Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G. R. Abecasis, A global reference for human genetic variation, Nature 526 (2015) 68–74, https://doi.org/10.1038/nature15393.
[29] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.
[30] S. Lindgreen, AdapterRemoval: easy cleaning of next-generation sequencing reads, BMC. Res. Notes 5 (2012) 337, https://doi.org/10.1186/1756-0500-5-337.
[31] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.
[32] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.
[33] C. Goubert, N.A. Zevallos, C. Feschotte, Contribution of unfixed transposable element insertions to human regulatory variation, Philos. Trans. R. Soc. B 375 (2020) 20190331, https://doi.org/10.1098/rstb.2019.0331.
[34] D.H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals, Genome Res. 19 (2009) 1655–1664, https://doi.org/10.1101/gr.094052.109.
[35] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575, https://doi.org/10.1086/519795.
[36] A.L. Price, N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet. 38 (2006) 904–909, https://doi.org/10.1038/ng1847.
[37] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (1984) 1358, https://doi.org/10.2307/2408641.
[38] F. Mignone, C. Gissi, S. Liuni, G. Pesole, Untranslated regions of mRNAs, Genome Biol. 3 (2002), https://doi.org/10.1186/gb-2002-3-3-reviews0004. REVIEWS0004.
[39] P. Dvorak, V. Hlavac, P. Soucek, 5' Untranslated region elements show high abundance and great variability in homologous ABCA subfamily genes, Int. J. Mol. Sci. 21 (2020) 8878, https://doi.org/10.3390/ijms21228878.
[40] T.H. Kim, L.O. Barrera, M. Zheng, C. Qu, M.A. Singer, T.A. Richmond, Y. Wu, R. D. Green, B. Ren, A high-resolution map of active promoters in the human genome, Nature 436 (2005) 876–880, https://doi.org/10.1038/nature03877.
[41] S. West, N.J. Proudfoot, Transcriptional termination enhances protein expression in human cells, Mol. Cell 33 (2009) 354–364, https://doi.org/10.1016/j.molcel.2009.01.008.
[42] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57, https://doi.org/10.1038/nprot.2008.211.
[43] B.T. Sherman, M. Hao, J. Qiu, X. Jiao, M.W. Baseler, H.C. Lane, T. Imamichi, W. Chang, DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update), Nucleic Acids Res. 50 (2022) W216–W221, https://doi.org/10.1093/nar/gkac194.
[44] K.G. Becker, K.C. Barnes, T.J. Bright, S.A. Wang, The genetic association database, Nat. Genet. 36 (2004) 431–432, https://doi.org/10.1038/ng0504-431.
[45] W. Zheng, Y. He, Y. Guo, T. Yue, H. Zhang, J. Li, B. Zhou, X. Zeng, L. Li, B. Wang, J. Cao, L. Chen, C. Li, H. Li, C. Cui, C. Bai, Baimakangzhuo, X. Qi, Ouzhuluobu, B. Su, Large-scale genome sequencing redefines the genetic footprints of high-altitude adaptation in Tibetans, Genome Biol. 24 (2023) 73, https://doi.org/10.1186/s13059-023-02912-1.
[46] X. Cao, Y. Zhang, L.M. Payer, H. Lords, J.P. Steranka, K.H. Burns, J. Xing, Polymorphic mobile element insertions contribute to gene expression and alternative splicing in human tissues, Genome Biol. 21 (2020) 185, https://doi.org/10.1186/s13059-020-02101-4.
[47] X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies, Nat. Genet. 44 (2012) 821–824, https://doi.org/10.1038/ng.2310.
[48] L. Yin, H. Zhang, Z. Tang, J. Xu, D. Yin, Z. Zhang, X. Yuan, M. Zhu, S. Zhao, X. Li, X. Liu, rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study, Genomics Proteomics Bioinformatics 19 (2021) 619–628, https://doi.org/10.1016/j.gpb.2020.10.007.
[49] L. Deng, C. Zhang, K. Yuan, Y. Gao, Y. Pan, X. Ge, Y. He, Y. Yuan, Y. Lu, X. Zhang, H. Chen, H. Lou, X. Wang, D. Lu, J. Liu, L. Tian, Q. Feng, A. Khan, Y. Yang, Z.-B. Jin, J. Yang, F. Lu, J. Qu, L. Kang, B. Su, S. Xu, Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders, Natl. Sci. Rev. 6 (2019) 1201–1222, https://doi.org/10.1093/nsr/nwz108.
[50] X. Zheng, D. Zheng, C. Zhang, H. Guo, Y. Zhang, X. Xue, Z. Shi, X. Zhang, X. Zeng, Y. Wu, W. Gao, A cuproptosis-related lncRNA signature predicts the prognosis and immune cell status in head and neck squamous cell carcinoma, Front. Oncol. 13 (2023) 1055717, https://doi.org/10.3389/fonc.2023.1055717.
[51] K. Wang, W.-D. Li, C.K. Zhang, Z. Wang, J.T. Glessner, S.F.A. Grant, H. Zhao, H. Hakonarson, R.A. Price, A genome-wide association study on obesity and obesity-related traits, PloS One 6 (2011) e18939, https://doi.org/10.1371/journal.pone.0018939.
[52] Y. Yao, X. Chu, M. Ma, J. Ye, Y. Wen, P. Li, B. Cheng, S. Cheng, L. Zhang, L. Liu, X. Qi, C. Liang, O.P. Kafle, C. Wu, S. Wang, X. Wang, Y. Ning, F. Zhang, Evaluate the effects of serum urate level on bone mineral density: a genome-wide gene-environment interaction analysis in UK biobank cohort, Endocrine 73 (2021) 702–711, https://doi.org/10.1007/s12020-021-02760-8.
[53] J. Divers, N.D. Palmer, C.D. Langefeld, W.M. Brown, L. Lu, P.J. Hicks, S.C. Smith, J. Xu, J.G. Terry, T.C. Register, L.E. Wagenknecht, J.S. Parks, L. Ma, G.C. Chan, S. G. Buxbaum, A. Correa, S. Musani, J.G. Wilson, H.A. Taylor, D.W. Bowden, J. J. Carr, B.I. Freedman, Genome-wide association study of coronary artery calcified atherosclerotic plaque in African Americans with type 2 diabetes, BMC Genet. 18 (2017) 105, https://doi.org/10.1186/s12863-017-0572-9.
[54] B.Y. Lin, W.-D. Lin, C.-K. Huang, M.-C. Hsin, W.-Y. Lin, A.D. Pryor, Changes of gut microbiota between different weight reduction programs, Surg. Obes. Relat. Dis. 15 (2019) 749–758, https://doi.org/10.1016/j.soard.2019.01.026.
[55] H. Zuo, T. Zheng, K. Wu, T. Yang, L. Wang, Q. Nima, H. Bai, K. Dong, Z. Fan, S. Huang, R. Luo, J. Wu, J. Zhou, H. Xu, Y. Zhang, S. Feng, P. Zeng, X. Xiao, B. Guo, Y. Wei, X. Pei, X. Zhao, China Multi-Ethnic Cohort (CMEC), high-altitude exposure decreases bone mineral density and its relationship with gut microbiota: results from the China multi-ethnic cohort (CMEC) study, Environ. Res. 215 (2022) 114206, https://doi.org/10.1016/j.envres.2022.114206.
[56] X. Wang, J. Liu, Q. Wang, Q. Chen, The transcriptomic and epigenetic alterations in type 2 diabetes mellitus patients of Chinese Tibetan and Han populations, Front Endocrinol (Lausanne) 14 (2023) 1122047, https://doi.org/10.3389/fendo.2023.1122047.
[57] K. Kyriakou, C.W. Lederer, M. Kleanthous, A. Drousiotou, A. Malekkou, Acid ceramidase depletion impairs neuronal survival and induces morphological defects in neurites associated with altered gene transcription and sphingolipid content, Int. J. Mol. Sci. 21 (2020) 1607, https://doi.org/10.3390/ijms21051607.
[58] R. Munk, C. Anerillas, M. Rossi, D. Tsitsipatis, J.L. Martindale, A.B. Herman, J.-H. Yang, J.A. Roberts, V.R. Varma, P.R. Pandey, M. Thambisetty, M. Gorospe, K. Abdelmohsen, Acid ceramidase promotes senescent cell survival, Aging (Albany NY) 13 (2021) 15750–15769, https://doi.org/10.18632/aging.203170.

[59] E. Sugano, G. Edwards, S. Saha, L.A. Wilmott, R.C. Grambergs, K. Mondal, H. Qi, M. Stiles, H. Tomita, N. Mandal, Overexpression of acid ceramidase (ASAH1) protects retinal cells (ARPE19) from oxidative stress, J. Lipid Res. 60 (2019) 30–43, https://doi.org/10.1194/jlr.M082198.
[60] A. Mingione, F. Pivari, N. Plotegher, M. Dei Cas, A. Zulueta, T. Bocci, M. Trinchera, E. Albi, V. Maglione, A. Caretti, L. Bubacco, R. Paroni, D. Bottai, R. Ghidoni, P. Signorelli, Inhibition of ceramide synthesis reduces α-synuclein proteinopathy in a cellular model of Parkinson's disease, Int. J. Mol. Sci. 22 (2021) 6469, https://doi.org/10.3390/ijms22126469.
[61] L.S. Lewis, K.M. Huffman, I.J. Smith, M.P. Donahue, C.A. Slentz, J.A. Houmard, M. J. Hubal, E.P. Hoffman, E.R. Hauser, I.C. Siegler, W.E. Kraus, Genetic variation in acid ceramidase predicts non-completion of an exercise intervention, Front. Physiol. 9 (2018) 781, https://doi.org/10.3389/fphys.2018.00781.
[62] H. Sies, D.P. Jones, Reactive oxygen species (ROS) as pleiotropic physiological signalling agents, Nat. Rev. Mol. Cell Biol. 21 (2020) 363–383, https://doi.org/10.1038/s41580-020-0230-3.
[63] I. Liguori, G. Russo, F. Curcio, G. Bulli, L. Aran, D. Della-Morte, G. Gargiulo, G. Testa, F. Cacciatore, D. Bonaduce, P. Abete, Oxidative stress, aging, and diseases, Clin. Interv. Aging 13 (2018) 757–772, https://doi.org/10.2147/CIA.S158513.
[64] E. Pena, S. El Alam, P. Siques, J. Brito, Oxidative stress and diseases associated with high-altitude exposure, Antioxidants (Basel) 11 (2022) 267, https://doi.org/10.3390/antiox11020267.
[65] R. Faraonio, Oxidative stress and cell senescence process, Antioxidants (Basel) 11 (2022) 1718, https://doi.org/10.3390/antiox11091718.
[66] V. Subbiah, E.I. Dumbrava, Y. Jiang, K.Z. Thein, A. Naing, D.S. Hong, S. Fu, S. A. Piha-Paul, A.M. Tsimberidou, F. Janku, F. Meric-Bernstam, R. Kurzrock, G. Falchook, Dual EGFR blockade with cetuximab and erlotinib combined with anti-VEGF antibody bevacizumab in advanced solid tumors: a phase 1 dose escalation triplet combination trial, Exp. Hematol. Oncol. 9 (2020) 7, https://doi.org/10.1186/s40164-020-00159-1.
[67] Y.-J. Chen, W.-A. Chang, L.-Y. Wu, C.-F. Huang, C.-H. Chen, P.-L. Kuo, Identification of novel genes in osteoarthritic fibroblast-like Synoviocytes using next-generation sequencing and bioinformatics approaches, Int. J. Med. Sci. 16 (2019) 1057–1071, https://doi.org/10.7150/ijms.35611.
[68] O. Vega-Hinojosa, M.H. Cardiel, P. Ochoa-Miranda, Prevalence of musculoskeletal manifestations and related disabilities in a Peruvian urban population living at high altitude. COPCORD Study. Stage I, Reumatol Clin (Engl Ed) 14 (2018) 278–284, https://doi.org/10.1016/j.reuma.2017.01.011.
[69] E.H. Farbu, A.C. Höper, E. Reierth, T. Nilsson, M. Skandfer, Cold exposure and musculoskeletal conditions; a scoping review, Front. Physiol. 13 (2022) 934163, https://doi.org/10.3389/fphys.2022.934163.
[70] M.B. Brent, U. Simonsen, J.S. Thomsen, A. Brüel, Effect of acetazolamide and Zoledronate on simulated high altitude-induced bone loss, Front. Endocrinol. 13 (2022) 831369, https://doi.org/10.3389/fendo.2022.831369.
[71] T.-Y. Wu, S.-Q. Ding, J.-L. Liu, J.-H. Jia, Z.-C. Chai, R.-C. Dai, J.-Z. Zhao, Q.D. Tang, B. Kayser, Smoking, acute mountain sickness and altitude acclimatisation: a cohort study, Thorax 67 (2012) 914–919, https://doi.org/10.1136/thoraxjnl-2011-200623.
[72] H. You, X. Li, T. Pei, Q. Huang, F. Liu, Y. Gao, Predictive value of basal exhaled nitric oxide and carbon monoxide for acute mountain sickness, Wilderness Environ. Med. 23 (2012) 316–324, https://doi.org/10.1016/j.wem.2012.04.001.
[73] P. Song, J.-H. Zhang, J. Qin, X.-B. Gao, J. Yu, X.-G. Tang, C.-F. Tang, L. Huang, Smoking is associated with the incidence of AMS: a large-sample cohort study, Mil. Med. Res. 1 (2014) 16, https://doi.org/10.1186/2054-9369-1-16.
[74] S. Gaillard, P. Dellasanta, L. Loutan, B. Kayser, Awareness, prevalence, medication use, and risk factors of acute mountain sickness in tourists trekking around the Annapurnas in Nepal: a 12-year follow-up, High Alt. Med. Biol. 5 (2004) 410–419, https://doi.org/10.1089/ham.2004.5.410.
[75] D. Vinnikov, N. Brimkulov, P.D. Blanc, Smoking increases the risk of acute mountain sickness, Wilderness Environ. Med. 26 (2015) 164–172, https://doi.org/10.1016/j.wem.2014.10.006.
[76] D. Vinnikov, P.D. Blanc, C. Steinmaus, Is smoking a predictor for acute mountain sickness? Findings from a meta-analysis, Nicotine Tob Res 18 (2016) 1509–1516, https://doi.org/10.1093/ntr/ntv218.
[77] G. Ramirez, P.A. Bittle, G.L. Colice, R. Herrera, S.J. Agosti, P.R. Foulis, The effect of cigarette smoking upon hematological adaptations to moderately high altitude living, Journal of Wilderness Medicine 2 (1991) 274–286, https://doi.org/10.1580/0953-9859-2.4.274.
[78] D. Duke, E. Wohlgemuth, K.R. Adams, A. Armstrong-Ingram, S.K. Rice, D.C. Young, Earliest evidence for human use of tobacco in the Pleistocene Americas, Nat. Hum. Behav. 6 (2021) 183–192, https://doi.org/10.1038/s41562-021-01202-9.
[79] L. Zhang, J. Meng, H. Li, M. Tang, Z. Zhou, X. Zhou, L. Feng, X. Li, Y. Guo, Y. He, W. He, X. Huang, Hippocampal adaptation to high altitude: a neuroanatomic profile of hippocampal subfields in Tibetans and acclimatized Han Chinese residents, Front. Neuroanat. 16 (2022) 999033, https://doi.org/10.3389/fnana.2022.999033.
[80] X. Zhang, W. Xie, W. Du, Y. Liu, J. Lin, W. Yin, L. Yang, F. Yuan, R. Zhang, H. Liu, H. Ma, J. Zhang, Consistent differences in brain structure and functional connectivity in high-altitude native Tibetans and immigrants, Brain Imaging Behav. 17 (2023) 271–281, https://doi.org/10.1007/s11682-023-00759-5.
[81] S. Matsuhashi, S. Noji, E. Koyama, F. Myokai, H. Ohuchi, S. Taniguchi, K. Hori, New gene, nel, encoding a M(r) 93 K protein with EGF-like repeats is strongly expressed in neural tissues of early stage chick embryos, Dev. Dyn. 203 (1995) 212–222, https://doi.org/10.1002/aja.1002030209.
[82] M. Oyasu, S. Kuroda, M. Nakashita, M. Fujimiya, U. Kikkawa, N. Saito, Immunocytochemical localization of a neuron-specific thrombospondin-1-like protein, NELL2: light and electron microscopic studies in the rat brain, Brain Res. Mol. Brain Res. 76 (2000) 151–160, https://doi.org/10.1016/s0169-328x(99)00342-3.
[83] H.R. Kim, D.H. Kim, J.Y. An, D. Kang, J.W. Park, E.M. Hwang, E.J. Seo, I.H. Jang, C.M. Ha, B.J. Lee, NELL2 function in axon development of hippocampal neurons, Mol. Cells 43 (2020) 581–589, https://doi.org/10.14348/molcells.2020.0032.
[84] E. Martinez, V.B. Palhan, A. Tjernberg, E.S. Lymar, A.M. Gamper, T.K. Kundu, B. T. Chait, R.G. Roeder, Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo, Mol. Cell. Biol. 21 (2001) 6782–6795, https://doi.org/10.1128/MCB.21.20.6782-6795.2001.
[85] C.G. Boer, R. Narcisi, Y.F. Ramos, W.D. Hollander, N. Bomer, M.C.C. Betancourt, A. G. Uitterlinden, G. Van Osch, I. Meulenbelt, J.J. Van Meurs, Genetic variants in the SUPT3H-RUNX2 locus confer susceptibility for bone and cartilage related disorders via long-range regulation of RUNX2, Osteoarthr. Cartil. 23 (2015) A71, https://doi.org/10.1016/j.joca.2015.02.145.
[86] Z. Feng, Z. Duren, Z. Xiong, S. Wang, F. Liu, W.H. Wong, Y. Wang, hReg-CNCC reconstructs a regulatory network in human cranial neural crest cells and annotates variants in a developmental context, Commun Biol 4 (2021) 442, https://doi.org/10.1038/s42003-021-01970-0.