DLX5/6 GABAergic expression affects social vocalization: implications for human evolution

Giovanni Levi[1], Camille de Lombares[1], Cristina Giuliani[2], Mey El Soudany[1], Rym Aouci[1], Paolo Garagnani[3,4], Claudio Franceschi[5], Dominique Grimaud-Hervé[6], and Nicolas Narboux-Nême[1,+]

1. Physiologie Moléculaire et Adaptation, CNRS UMR7221, Département AVIV, Muséum National d'Histoire Naturelle, Paris, France.

2. Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Italy.

3. Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Bologna, Italy.

4. Clinical Chemistry, Department of Laboratory Medicine, Karolinska Institutet at Huddinge University Hospital, Stockholm, Sweden.

5. Laboratory of Systems Medicine and Healthy Aging, and Department of Applied Mathematics, Lobachevsky State University, Nizhny Novgorod, Russia

6. Histoire Naturelle de l'Homme Préhistorique, CNRS UMR 7194, Département H&E, Muséum National d'Histoire Naturelle, Paris, France.

+ Correspondence should be addressed to:

Nicolas Narboux-Nême
Évolution des Régulations Endocriniennes,
Centre National de la Recherche Scientifique, UMR-7221,
7, rue Cuvier
75231 Paris, Cedex 05, France
Tel. : Fax. : Email : nnarboux-nême@mnhn.fr

ABSTRACT

DLX5 and DLX6 are two closely related transcription factors involved in brain development and in GABAergic differentiation. The *DLX5/6* locus is regulated by FoxP2, a gene involved in language evolution and has been associated to neurodevelopmental disorders and mental retardation. Targeted inactivation of *Dlx5/6* in mouse GABAergic neurons (*Dlx5/6*^{VgatCre} mice) results in behavioural and metabolic phenotypes notably increasing lifespan by 33%.

Here, we show that *Dlx5/6*^{VgatCre} mice present an hyper-vocalization and hyper-socialization phenotype. While only 7% of control mice emitted more than 700 vocalizations/10min, 30% and 56% of heterozygous or homozygous *Dlx5/6*^{VgatCre} mice emitted more than 700 and up to 1400 calls/10min with longer, uninterrupted, call sequences and a higher proportion of complex and modulated calls. Hyper-vocalizing animals were more sociable: the time they spent in dynamic interactions with an unknown visitor was more than doubled compared to low-vocalizing individuals.

We then compared the *DLX5/6* genomic regions from Neanderthal and modern humans to examine if *DLX5/6* GABAergic expression could have been involved in human social evolution. We identify an introgressed Neanderthal haplotype (*DLX5/6*-N-Haplotype) present in 12.6% of European individuals that covers *DLX5/6* coding and regulatory sequences. The *DLX5/6*-N-Haplotype is not significantly associated to Autism Spectrum Disorders (ASDs) but includes the binding site for GTF2I, a gene associated to Williams-Beuren syndrome, a neurodevelopmental disorder characterized by hyper-sociability and hyper-vocalization. Interestingly, the *DLX5/6*-N-Haplotype is significantly underrepresented in semi-supercentenarians (>105y of age), a well-established human model of healthy ageing and longevity, suggesting its potential involvement in the co-evolution of longevity, sociability and speech.

Keywords: *DLX5/6* genes, ultrasonic vocalization, hyper-vocalization, social behaviour, Human evolution.

INTRODUCTION

Within groups of animals, vocal calls play an essential role in directing social behaviour, in transmitting information about food availability and predators and for reproductive and maternal activities. In humans, vocal communication and speech constitute major socialization determinants; their dysfunction represents a core trait of neurodevelopmental conditions such as Autism Spectrum Disorders (ASDs). Several mouse models characterized by genetic lesions associated to human neurodevelopmental disorders have been used to study the genetic and neurological mechanisms supporting the interplay between vocalization and socialization (Caruso, et al. 2020). Mice social behaviours are accompanied by the production of ultrasonic vocalizations (USVs) (Sales 1972), which, in the case of most laboratory strains, present a frequency included between 35 and 110 kHz (Holy and Guo 2005). Mice USVs profiles vary with the type of social activity performed (e.g. courtship, aggression, fear), with the animal's developmental stage, with their genetic background and can be altered by induced genetic modifications. Although the link between USVs profiles and social activity is still incompletely analysed, a recent study has implemented artificial intelligence algorithms to associate USVs of individual mice to specific social actions (Ey, et al. 2020; Sangiamo, et al. 2020) unveiling a functional correlation between specific USVs and social behaviours. The generation of voluntary structured vocalizations requires the coordination of different levels of cognitive and motor activities and involves a complex and diffuse neural network ranging from the prefrontal and motor cortical areas for voluntary initiation of vocalization to several forebrain and brainstem regions (Jurgens 2002; Simonyan 2014) for the control of laryngeal movements and respiration. The inputs/outputs to this network are integrated in the midbrain periaqueductal gray (PAG), the central neural region controlling vocalization in most mammalians (Jurgens 2009; Tschida, et al. 2019).

The study of developmental disorders affecting vocal communication such as stuttering, verbal dyspraxia and some types of ASDs has led to the identification of genes involved in the control of speech and/or socio/cognitive faculties (Konopka and Roberts 2016). The most intensively studied of them, the transcription factor FOXP2, is associated to inherited dyspraxia and poor control of the facial musculature. Other "language-associated" genes include for example: *SHANK1/2/3, FOXP1, GNPTAB, GNPTG, CNTNAP2,* and *NAGPA*; for most of these genes mouse models have been generated resulting almost invariably in alteration of USV and social behaviour (Wohr 2014; Caruso, et al. 2020). These animal models suggest that a certain level of conservation exists between the mechanisms controlling vocalization.

Homozygous disruption of *Foxp2* in the mouse cerebral cortex results in abnormal social approach and altered USVs (Medvedeva, et al. 2019). One of the targets of Foxp2 is the *Dlx5/6* locus: the mouse line Foxp2-S321X, which does not produce Foxp2 protein, displays a more than 200-fold increased expression of *Dlx6as1* (also known as *Evf1/2* or *Shhrs*) (Vernes, et al. 2006) a long noncoding RNA generated from the *Dlx5/6* locus known to play a central role in the control of *Dlx* gene expression (Faedo, et al. 2004; Feng, et al. 2006).

DLX5/6 are pleiotropic genes that control multiple aspects of embryonic development (Merlo, et al. 2000), including limb, craniofacial and laryngeal morphogenesis (Beverdam, et al. 2002; Depew, et al. 2002; Robledo, et al. 2002; Conte, et al. 2016) and GABAergic neuronal differentiation and function (Perera, et al. 2004; Wang, et al. 2010; Cho, et al. 2015; de Lombares, et al. 2019).

In humans, mutations in a 1.5 Mb genomic region including *DLX5/6* are associated with Split Hand/Foot Malformation type 1 (SHFM1) (Scherer, et al. 1994; Crackower, et al. 1996), a form of ectrodactyly which can be accompanied by sensorineural deafness, craniofacial malformations with or without intellectual disability (Birnbaum, et al. 2012; Rasmussen, et al. 2016). Recently a largescale transcriptomic study on post-mortem brains has suggested that the *DLX5/6* locus participates to genetic modules involved in ASDs and schizophrenia (Gandal, et al. 2018).

Mice with a *Dlx5/6* deletion present a SHFM1 phenotype associated with severe craniofacial defects (Beverdam, et al. 2002; Depew, et al. 2002), while targeted inactivation of these genes in GABAergic neurons modifies behaviour, metabolism and life expectancy (de Lombares, et al. 2019).

Modern humans, the sole survivor of the group of hominins, differ from other *Hominidae* by many anatomical and physiological characters and, most remarkably, for their cognitive and linguistic capacities which reflect the singularity of our brain. Comparison of modern and archaic *Homo* genomes has played a pivotal role in the quest of *Homo sapiens* evolutionary history and in defining what makes us modern humans (see for ex. (Prufer, et al. 2014 ; Racimo, et al. 2014; Racimo 2016; Peyregne, et al. 2017; Hajdinjak, et al. 2018)).

Comparison of *Neanderthal, Denisovan* and modern *Human* genomes has shown that interbreeding occurred between these different human groups until 45,000-65,000 years ago (Sankararaman, et al. 2014). During later evolution, there has been strong selection against archaic introgression of most Neanderthal or

Denisovan sequences in the *Homo sapiens* genome while some introgressed regions were conserved and were even under positive selection. There is increasing interest to decode the biological, phenotypic and medical consequences of Neanderthal legacy in the modern human genome (Simonti, et al. 2016; Dolgova and Lao 2018; Rotival and Quintana-Murci 2020).

To understand the coevolution of human traits, a particular attention has been given to pleiotropic genes, which can simultaneously control the development of numerous characters. For example, in prospective studies trying to identify genes capable of affecting simultaneously craniofacial shape, osteogenesis and language abilities, Boeckx and Benitez-Burraco have proposed a network of physically or functionally interconnected genes including the *RUNX*, *FOXP* and *DLX* families and their regulators (Boeckx and Benitez-Burraco 2014, 2015).

Here we show that inactivation of *Dlx5/6* in mouse GABAergic neurons results in a hyper-vocalization phenotype and in a sharp increase of dynamic social interactions. Furthermore, we identify an introgressed Neanderthal haplotype (N-Haplotype) present in 6.3% of Europeans alleles which covers *DLX6* and *DLX5* coding and regulatory sequences including SNPs associated to either hyper- (Williams-Beuren syndrome) or hypo- (ASDs) sociability and vocalization syndromes (Poitras, et al. 2010).

RESULTS

Dlx5/6 expression in GABAergic neurons affects mouse social ultrasonic vocalizations

To analyze the social ultrasonic vocalization phenotypes (UVs) resulting from the simultaneous deletion of *Dlx5/6* homeodomains in adult GABAergic neurons, we have applied the resident/intruder experimental paradigm to groups of control, $Dlx5/6^{VgatCre/+}$ and $Dlx5/6^{VgatCre}$ female mice (Moles, et al. 2007). In this experimental setting, the animals to be tested (resident) are isolated 3 days before being exposed to an unfamiliar group-housed control adult mouse (intruder). Although UVs are simultaneously recorded for the pair of mice (resident and intruder), previous studies have shown that the contribution of the new comer is minor in comparison to that of the occupant (Hammerschmidt, et al. 2012).

The number and types of vocalizations emitted during a 10min test period (v/10min) by 10 $Dlx5/6^{VgatCre/+}$, 9 $Dlx5/6^{VgatCre}$ and 15 control females confronted to a control female intruder were individually recorded. Most (14/15; 93%) control mice emitted less than 700 v/10min and only one emitted 849 v/10min. In contrast, 30% (3/10) of the $Dlx5/6^{VgatCre/+}$ and 56% (5/9) of the $Dlx5/6^{VgatCre}$ mice presented a unique vocalization profile emitting more than 700 v/10min, with two heterozygous and four homozygous emitting more than 1000 v/10min and up to 1400 v/10min (Figure 1 A, B).

This observation demonstrates the presence of an incompletely penetrant vocalization phenotype associated with the reduction of the level of expression of Dlx5/6 in GABAergic neurons characterized by an allel-dosage-dependent increase in the number of calls per minute.

Figure 2 (A-H) presents some illustrative examples of vocalization profiles recorded from normal and mutant resident mice. Control animals (Fig. 2 A-D) and low-vocalizing mutants (Fig. 2 G, H) emitted bursts of vocal signals separated by silent intervals or did not emit any call for long periods of time (Fig. 2H). In contrast, the high-vocalizing $Dlx5/6^{VgatCre}$ mice emitted uninterrupted series of calls extending for many seconds without any detectable interruption or pause (Fig. 2 E, F), this pattern of vocalization was never observed in control animals.

The total time spent vocalizing by mice emitting a high number of vocalizations was about 4 times longer than that of low-vocalizing littermates (Figure 3A) suggesting an increased frequency of long-lasting calls. We categorized the types of vocalization emitted in five major groups short, simple, complex, frequency

jumps and unstructured as previously described (Ey, et al. 2013). The analysis of the specific types of calls emitted by each animal showed that high-vocalizing animals presented a significant increase in the number of complex vocalizations ("modulated + complex" according to (Ey, et al. 2013)) which are the longer vocalization types emitted by mice and can last up to 0.15ms (Figure 3 B). Complex vocalizations have been recently associated to more social behaviours (Ey, et al. 2020). Therefore, in this experimental paradigm, designed to detect social vocalization, we observed a new partially-penetrant phenotype characterized by an increase in the number calls, and differences in vocal repertoire and "prosody" (Lahvis, et al. 2011).

Dlx5/6 expression in brain regions which control vocalization

Vocal communication depends on a complex neural network that includes regions of the forebrain, of the basal ganglia and of the brainstem (Jurgens 2002). These multiple inputs are integrated by the periaqueductal grey (PAG), a midbrain territory that is at the interface between the telencephalon and phonation/respiration motor centers. As all these brain territories include GABAergic neurons, we determined in which of them *Dlx5* is expressed. To this end we analyzed the distribution of ß-galactosidase staining in $Dlx5^{lacZ}$ adult mouse brains (Figure 4) (Acampora, et al. 1999). We find that *Dlx5* is not expressed in the PAG or other midbrain regions (Fig. 4A) nor in posterior motor areas. On the contrary, *Dlx5* is strongly expressed in the cerebral cortex (Fig. 4B), in the hypothalamus (Fig. 4C) and in the striatum (Fig. 4D). These observations suggest that Dlx5/6 affect central cognitive and emotional inputs to vocal communication, but do not affect the "mechanics" of phonation.

Social behaviour of hyper-vocalizing mice

The social behaviour of the mice on which we had quantified UVs was analyzed with the resident/intruder experimental setting described above. Behaviour and social interactions were filmed and analyzed as previously described (de Chaumont, et al. 2019). Social behaviour between individuals of the same genotype was very variable resulting in no significant difference in sociability between control, $Dlx5/6^{VgatCre/+}$ and $Dlx5/6^{VgatCre}$ mice (de Lombares, et al. 2019). However, when we confronted the groups of hyper-vocalizing (>700 Calls/10min) and low-vocalizing (<700 calls/10min) animals, we found a

highly-significant association between hyper-vocalization and hyper-sociability. Hyper-vocalizing animals spent more than twice time following or approaching the intruder than the low-vocalizing individuals (Figure 5 A,B).

Therefore, the same animals in which the mutation of *Dlx5/6* in GABAergic neurons induced hyper-vocalization presented also an hyper-sociability phenotype.

Introgression of a Neandertal-like haplotype in the *DLX5/6* locus

Patients carrying mutations in the *DLX5/6* locus or mouse models in which these genes are deregulated present abnormalities in at least five set of characters which differentiate *Homo sapiens* from *Homo Neanderthalensis*: 1) brain size and shape, 2) limb morphology, 3) craniofacial shape and 4) middle ear morphology and 5) bone density.

Comparing the *DLX5/6* locus of modern humans (Figure 6A) to those of Neanderthal (Altaï, Vindija), Denisovian, and other *Hominidae* we identified 298 SNPs with rare alleles that are shared between modern humans and Neanderthals, but absent from the African populations.

In the proximal region of *DLX5/6*, we identified a group of SNPs, all located in non-coding regions, that co-segregate at more than 80% in individuals from the 1000 Genomes database (The 1000 Genomes Project Consortium), forming a "core" 37,160bp haplotype (*DLX5/6*-N-Haplotype) defined by all consecutive sites with Neandertal-like alleles (Table 1, Figure 6 A). This haplotype includes *DLX6* coding and regulatory sequences, the intergenic region between *DLX6* and *DLX5* and *DLX5* intron 3 and exon 3. Allele sharing with archaic genomes at polymorphic sites in non-African populations can result either from incomplete lineage sorting or admixture. To distinguish between these two scenarios we looked at the length of the haplotypes. Introgressed haplotypes present fewer recombinations than those predicted from the time interval that separates modern humans from the common ancestral population. The *DLX5/6-N*-Haplotype is significantly longer than expected under the null hypothesis of incomplete lineage sorting (p value = $1.08e^{-5}$), supporting an archaic origin of the haplotype.

The proportion of individuals carrying the *DLX5/6*-N-Haplotype is variable depending on the population, from 1.5% in East Asia to 6.3% in Europe and absent in continental Africa (Figure 6A, B). Moreover, the presence of the different SNPs of Neanderthal origin composing the haplotype varies within the same

geographical region and is particularly visible in South Asia (Figure 6 SAS) suggesting the presence of haplotypes of different size.

Overall, the proportion of the *DLX5/6*-N-Haplotype is low in East Asia and America (less than 4%) and completely absent from Africans and Peruvians from Lima. In South Asia and Europe, the *DLX5/6*-N-Haplotype is much more frequent reaching up to 8.2% in the British population of England and Scotland and in Sri Lankan Tamil in the UK. In African-American and Caribbean the haplotype is poorly represented (Figure 6 B).

Association of the *DLX5/6* locus with human cognitive impairment and speech defects

To evaluate the possible implications of the *DLX5/6* locus and of the *DLX5/6*-N-Haplotype in particular for human development and evolution, we reexamined the phenotype of patients carrying known genetic alterations within this region. Deletions or mutations at the *DLX5/6* locus are associated to Split Hand Foot Malformation Type 1 (SHFM1), a form of ectrodactyly in which limb malformations can be associated to cognitive, craniofacial or hearing defects. The analysis of SHFM1 patients for which the genetic lesion had been localized, permits associate specific genomic sub-regions to other syndromic traits such as mental retardation, language delay and autism (Figure 7).

Most families carrying mutations limited to the genomic region encompassing only limb-specific enhancers (Birnbaum, et al. 2012) present SHFM1 with no other physical or mental problems, with the exception of one case of mild cognitive defects. Families with deletions or translocations including craniofacial-specific enhancers present SHFM1 with hearing loss with or without other craniofacial anomalies and a high incidence of cognitive defects (Rasmussen, et al. 2016). Large heterozygous deletions covering both regulatory and coding *DLX5/6* sequences present SHFM1 with craniofacial defects associated in 5 out of 7 cases to severe mental retardation. Patients carrying mutations restricted to *DLX5/6* intergenic region or their coding sequences have a higher probability to include members with ASDs or speech delay but do not present the limb deformities of SHFM1 (Hamilton, et al. 2005; Nakashima, et al. 2010; Poitras, et al. 2010; Chang, et al. 2011).

The *DLX5/6*-N-Haplotype includes enhancers (I56i, I56ii and MEF2) capable to direct *DLX5/6* expression in the brain (Fazel Darbandi, et al. 2016; Assali, et al. 2019). Interestingly a polymorphism in enhancer I56i

is associated to the binding of GTF2I, a transcription factor deleted in Williams-Beuren syndrome, a complex condition often characterized by hyper-vocalization and hyper-sociability (Figure 7).

The *DLX5/6*-N-Haplotype allele is under-represented in semi-supercentenerians

We then evaluated the possible association between the *DLX5/6*-N-Haplotype and specific traits found in modern human populations. As the *DLX5/6* locus has been associated to ASDs and mental retardation ((Gandal, et al. 2018) and (Figure 7)), we analysed the *DLX5/6*-N-Haplotype frequency in the MSSNG database of ASDs families (RK, et al. 2017) and we did not find a significant association (p=0.16) with the disease. This finding, however, cannot exclude an association of the haplotype to specific mental traits not necessarily represented in the ASDs database. We then focused on other traits that can be associated to social vocalization. Notably, mortality risk has been correlated to social environment in humans and other social mammals (Snyder-Mackler, et al. 2020). Since, inactivation of *Dlx5/6* in GABAergic neurons results in a 33% life span extension (de Lombares, et al. 2019) we then focused on the population of semi-supercentenarians (CEN), i.e. people over 105 years of age (105+) characterized by exceptional longevity. The frequency of the *DLX5/6*-N-Haplotype or of truncated haplotypes containing at least 2 consecutive Neandertal SNPs were calculated for all the European populations present in the 1000Genomes database and for a cohort of 81 Italian semi-supercentenarians. Fixation index (F_{st}) analysis showed a significant difference in *DLX5/6*-N-Haplotype frequency (p-value < 0.05). The 105+ (CEN) presented a 50% under-representation of the haplotype when compared to any other European population with the only exception of IBS and constituted the most divergent group. The comparisons with the Italian control group termed "ITA" (the group closest to 105+ CEN in term of genetic ancestry) also showed a significant difference (p-value = 0.026). Chi-square analysis based on the allelic distribution between centenarians and the total European population also showed significant underrepresentation of the haplotype (p<0.05) (Figure 8).

DISCUSSION

Here we have shown that reduction of *DLX5/6* expression in mouse GABAergic neurons results in an hyper-vocalization and hyper-socialization phenotype. Furthermore, we identify a *DLX5/6*-Neanderthal-derived haplotype present in up to 8% of non-African alleles that includes enhancers controlling gene expression in the central nervous system (Fazel Darbandi, et al. 2016) and sequences associated to either hyper- or hypo-socialization/vocalization pathologies such as Williams-Buren syndrome (Poitras, et al. 2010) or ASDs (Gandal, et al. 2018). Interestingly the *DLX5/6*-N-Haplotype is underrepresented in the genomes of healthy semi-supercentenarians suggesting an association between *DLX5/6* brain expression and healthy aging reminiscent of what has been shown in mouse models (de Lombares, et al. 2019).

In social mammals, such as mice and humans, vocal communication (VC) is essential for the survival of the group. VC is needed to convey diverse information such as location, social status, reproductive availability, emotional context (fear, aggression), food availability and presence of predators. Such a vast set of inputs and outputs is coordinated by a complex neural network that combines regions of the forebrain, of the basal ganglia and of the brainstem (Jurgens 2002). A key brain region involved in the integration of these neuronal activities is the periaqueductal grey (PAG), a midbrain territory that receives multiple telencephalic inputs and transmit them to phonation and respiration motor centers. As *Dlx5/6* are not expressed in the PAG, but only in telencephalic interneurons, their effect on vocalization must derive from cognitive initiation and modulation and not from impairment of laryngeal or respiratory motor centers.

In comparison to their normal controls, many $Dlx5/6^{VgatCre}$ and $Dlx5/6^{VgaCre/+}$ mice emit a higher number of longer "complex" calls often not separated by any interruption. These long-modulated USVs are emitted prevalently when mice are engaged in socially-intense behaviours and more rarely during isolated behaviour (Ey, et al. 2020). Therefore, the high-vocalization profile of $Dlx5/6^{VgatCre}$ mice is predictive of an enhanced social behaviour and is in line with our finding of an increase in dynamic socialization of high-vocalizing individuals.

Many mouse models have been generated to study the effect of genetic lesions associated to human neurodevelopmental disorders. In particular, the simultaneous study of ultrasonic vocalizations and

behaviour has permitted to evaluate the impact of specific mutations on social communication and interaction in mouse models of ASDs (Caruso, et al. 2020). For example, female-female social interactions and USV have been analysed in *Shank2*$^{-/-}$ and *Shank3*$^{-/-}$ mutant mice showing reduction in the number and duration of USVs associated to diminished sociability (Schmeisser, et al. 2012; Ey, et al. 2013; Wohr 2014). Extrapolation of our results to the human condition would suggest that reduced *DLX5/6* expression in GABAergic neurons could be associated to increased social vocalization, a phenotype opposite to ASDs.

Our understanding of the origin of psychiatric pathologies such as schizophrenia and ASDs has recently progressed substantially thanks to large Genome Wide Association studies (GWAs) which have shown that, beyond some specific cases where monogenic mutations cause highly penetrant phenotypes, psychiatric conditions are complex polygenic disorders associated with hundreds of common genetic loci involving genes associated to synaptic function and neural development (Huguet, et al. 2016; Dennison, et al. 2019). Module analysis of gene-networks implicated in interneuron-dependent ASD and schizophrenia has identified gene expression variations and splicing of *DLX5*, *DLX6* and *asDLX6nc*. Indeed, the modules including these *DLX* genes and other key genetic components of GABAergic neurons constitute the major hub for both ASD and schizophrenia (Gandal, et al. 2018). Remarkably, a point mutation within the *I56i* enhancer, is associated with binding of the transcription factor GTF2I which is deleted in Williams-Beuren syndrome characterized by hypervocalization and hypersociability (Poitras, et al. 2010).

The origin of vocal language in the genus *Homo* has been approached through anatomical analyses of the brain and larynx and, more recently, through the comparison of genes involved in language acquisition such as *FOXP2* (Dediu and Levinson 2013; Dediu and Christiansen 2016). Remarkably *FOXP2* is located in 7q31, close to the *DLX5/6* locus (7q21), and regulates the expression of a large antisense noncoding RNA of *DLX6* (*asDLX6nc* or *Evf1/2*) (Vernes, et al. 2006; Vernes, et al. 2011) which is directly involved in the control of *Dlx5* and *Dlx6* expression (Feng et al 2006). Furthermore, it has been proposed that *DLX5/6* could have been implicated in brain and skull globularization and acquisition of human-specific language capacities (Boeckx and Benitez-Burraco 2014).

The *DLX5/6*-N-Haplotype covers the complete sequence of *DLX6*, the intragenic region and the 2^{d} intron and 3rd exon of *DLX5* and three known enhancers for brain expression of *DLX5/6*: *I56i, I56ii* and *Mef2-BA* (Birnbaum, et al. 2012), and the three first exons of the long non-coding RNA *asDLX6nc*, which

has been implicated in GABAergic neuronal differentiation (Cajigas, et al. 2018). The major SNPs composing the *DLX5/6*-N-Haplotype are located in non-coding sequences in proximity of the known enhancers and splice sites of *DLX5* and *asDLX6nc* and could well alter their expression profile.

The comparison of human genomes with those of other vertebrate species and, in particular, with other *Hominidae*, has revealed the presence of Human Accelerated Regions (HARs), short sequences presenting an important number of Single Nucleotide Changes (SNCs) specific to *Homo sapiens* (Lindblad-Toh, et al. 2011; Capra, et al. 2013). Interestingly, the *DLX5/6* locus includes at the same time 5 HARs and, as shown in this report, an introgressed Neanderthalian region; this paradoxical situation might suggest that enhancers implicated in non-neural functions of *DLX5/6* such as bone remodeling and reproduction could have been strongly counter-selected, whereas brain-expression driving enhancers could have been conserved in the human population.

Vocal communication and socialization are pivotal traits of the evolution of modern human society. The conservation of the *DLX5/6*-N-Haplotype in 6.3% of EUR alleles (which represent about 12,6% of the population as the *DLX5/6*-N-Haplotype is almost invariably heterozygous), suggests that it presence in a proportion of individuals is beneficial for the sustainability of modern human society. As *DLX5/6* are pleiotropic genes with multiple interconnected functions including craniofacial morphogenesis, neuronal differentiation, bone development and reproduction (Merlo, et al. 2000) one could speculate that the presence of the *DLX5/6*-N-Haplotype might also affect other traits characteristic of the evolution of *Homo sapiens*, such as metabolism, skull shape and bone density. In particular one of the main differences between modern humans and Neanderthals has been the composition and size of social groups: while Neanderthal groups were small, composed by 10 to 30 individuals, modern humans developed complex societies (Duveau, et al. 2019). It is possible, therefore, that the *DLX5/6*-N-Haplotype could influence characters affecting the structure of modern human society.

Interestingly the *DLX5/6*-N-Haplotype is underrepresented in semi-supercentenarians, a population considered as a paradigm of healthy ageing as it has the capacity to slow the progression of pathological conditions determining reduced mental and physical functions (Vacante, et al. 2012). This observation could be related to the 33% lifespan extension described for $Dlx5/6^{VgatCre}$ mice, suggesting that the *DLX5/6*-N-Haplotype could regulate *DLX5/6* neuronal expression (de Lombares, et al. 2019). Our finding implies

that the presence of the *DLX5/6*-N-Haplotype, although positive for certain aspects, could be associated to traits reducing human healthy aging and life expectancy. This observation could well be integrated within current ideas on the biology of ageing which suggest that ageing rates, and consequently lifespans, evolve as a function of trade-offs with other characters including social, cognitive and reproductive capacities (Yang, et al. 2016; Snyder-Mackler, et al. 2020).

FIGURE LEGENDS

Figure 1: Number of calls/10 min emitted by *Dlx5/6*$^{VgatCre/+}$ and *Dlx5/6*^{VgatCre} mice.

Number of calls emitted by of *Dlx5/6*$^{VgatCre/+}$ (n=10), *Dlx5/6*^{VgatCre} (n=9) and control (n=15) six months old female mice confronted to an intruder genetically normal female mouse of the same age over a period of 10 minutes. The number and types of vocalization were quantified.

A) 93% of control mice emitted less than 700 vocalizations during the 10 minute tests, only one outlier emitted 849 calls. In contrast, 30% of *Vgat*$^{\Delta Dlx5\text{-}6/+}$ (3/10) and 56% of *Vgat*$^{\Delta Dlx5\text{-}6}$ (5/9) mice emitted between 850 and 1400 calls/10 min. B) Percentage of mice emitting less (pink) or more (red) than 700 calls/ 10 min. The difference between each genotype is highly significant (***= $p<0.001$).

Figure 2: Spectrograms of control and *Dlx5/6*^{VgatCre} mice.

Representative spectrograms from 4 controls (A-D) and 4 *Dlx5/6*^{VgatCre} (E-H) mice. In each group of animals some individuals (e.g. D, H) emitted very few vocalizations while others emitted sequences of calls separated by silent intervals (A-C, G). Hyper-vocalizing *Dlx5/6*^{VgatCre} mice emitted long, uninterrupted, sequences of calls (E, F) which were never observed in controls or low-vocalizing animals.

Figure 3: Total vocalization time and call types emitted by hyper-vocalizing and low-vocalizing mice.

A) Total time spent vocalizing by mice emitting less or more than 700 calls/10 min. B) Percentage of call types produced by mice emitting less or more than 700 calls/10 min. Five types of emissions have been considered as in (Ey, et al. 2013): Short, Simple (Flat+Upward+Downward), Complex (Modulated + Complex), Unstructured, and Frequency jumps. Hyper-vocalizing animals vocalize, in average, four times more than low vocalizing animals and emit significantly more complex and modulated calls.

(***= $p<0.001$; *=$p<0.05$).

Figure 4: Expression of *Dlx5* in different adult brain regions involved in vocalization.

X-gal staining on serial sections of *Dlx5*$^{lacZ/+}$ brains (Acampora, et al. 1999). *β*-D-galactosidase activity was not detected in the brainstem (A), in the cerebellum and in the midbrain including the ventro-lateral

periacqueductal grey (vlPAG) a region which integrates sensory and cognitive inputs and controls vocalization motor centers. *Dlx5* was expressed in the prefrontal cortex (B) and in other cortical regions (cx), in the striatum (C) and in hypothalamic nuclei (D). AHP, Anterior hypothalamic area; cc, corpus callosum; cx, cortex; DM, Dorsomedial hypothalamus; PVN, para-ventricular nucleus; SCh, suprachiasmatic nucleus; VMH, ventro medial hypothalamus).

Figure 5: Social behaviour of hyper-vocalizing and low-vocalizing mice.

The social interaction behaviour of $Dlx5/6^{VgatCre/+}$ (n=10), $Dlx5/6^{VgatCre}$ (n=9) and control (n=15) mice placed in the same cage with a control visitor was monitored using a protocol (de Chaumont, et al. 2019; de Lombares, et al. 2019) which associates computer vision and artificial intelligence to identify and categorize individual social interactions.

The types of behaviour recorded were classified as follows: 1) Isolated behaviour (moving alone, rearing alone, stopped alone, jump alone, huddling alone), 2) Static interactions (rearing in contact, oral-oral contact, oro genital contact, side-by- in same or opposite orientation, 3) Dynamic interactions (moving in contact, train of 2 mice, following the same path), 4) Social approach (making contact, approaching in rearing, approaching), 5) Social escape (breaking contact, escaping), a detailed description of each behaviour is found in (de Chaumont, et al. 2019). The time spent by each mouse (sec) adopting a specific behaviour was measured over a 10 min recording.

A) Normalized results: the average time spent in a specific behaviour by low-vocalizing animals was set as one (pink) and the ratio with the time spent in that same behaviour by hyper-vocalizing mice (red) is shown. For example “Dynamic Interaction” represents a trait that was adopted 2.3 times more frequently by hyper-vocalizing mice than by low-vocalizing animals.

B) Actual values in seconds recorded for each social behavioural trait. (***= $p<0.001$; *=$p<0.05$).

Figure 6: Linkage disequilibrium analysis of the distribution of the *DLX5/6-N*-Haplotype in different populations.

A) For each population, the width of the color bar represents the allele frequency for each SNP. SNPs linked by a solid bar cosegregate; if cosegregation is less than 100%, the width of the bar diminishes accordingly.

The allele frequency of the most represented Neanderthal SNP of the *DLX5/6*-N-Haplotype is indicated for each population. The major SNPs constituting the haplotype are indicated. The scale bar represents a 5% allelic frequency within a population.

EAS: East Asia populations; SAS: South Asia populations; EUR: European populations; AMR: American populations; AMR/AFR; Afro-americans; AFR: Africans.

B) Geographical distribution of the *DLX5/6*-N-Haplotype. The allele frequency of the most represented SNP constituting the *DLX5/6*-N-Haplotype is represented on the earth map.

The higher frequency of the *DLX5/6*-N-Haplotype is seen in Europe and South Asia, the haplotype is not present in African populations and in Peruvians from Lima.

CHB: Han Chinese in Beijing, China; JPT: Japanese in Tokyo, Japan; CHS: Southern Han Chinese; CDX: Chinese Dai in Xishuangbanna, China; KHV: Kinh in Ho Chi Minh City, Vietnam; CEU: Utah Residents (CEPH) with Northern and Western European Ancestry; TSI: Toscani in Italia; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian Population in Spain; YRI: Yoruba in Ibadan, Nigeria; LWK: Luhya in Webuye, Kenya; GWD: Gambian in Western Divisions in the Gambia; MSL: Mende in Sierra Leone; ESN: Esan in Nigeria; ASW: Americans of African Ancestry in South-West USA; ACB: African Caribbeans in Barbados; MXL: Mexican Ancestry from Los Angeles USA; PUR: Puerto Ricans from Puerto Rico; CLM: Colombians from Medellin, Colombia; PEL: Peruvians from Lima, Peru; GIH: Gujarati Indian from Houston, Texas; PJL: Punjabi from Lahore, Pakistan; BEB: Bengali from Bangladesh; STU: Sri Lankan Tamil from the UK; ITU: Indian Telugu from the UK

Figure 7: Association of different pathological conditions with the *DLX5/6* genomic region

The central black line represents the 7q21.3 locus, which includes at least 5 genes *DLX5*, *DLX6*, *SEM1*, *SLC25A13* and *DYNC1I1*. A long non-coding anti-sense RNA, *DLX6-AS*, is also transcribed from this

region. Purple boxes in the upper part of the figure represent the phenotypic consequences observed in Split Hand Foot Malformation syndrome (SHFM) patients carrying deletion/mutations in this locus. While all SHFM patients present ectrodactyly, in certain cases this malformation is accompanied by craniofacial defects, hearing loss and/or cognitive defects. The closer the deletion are to *DLX5/6* coding sequences the higher is the frequency of patients presenting also cognitive impairment. Point mutations within the *DLX5/6* region and proximal or intragenic enhancers have been associated to autism spectrum disorders (ASD) without ectrodactyly or craniofacial defects and are included in the *DLX5/6*-N-Haplotype. Three conserved enhancers (*I56i, I56ii* and *Mef2-BA*) that direct *DLX5/6* expression in the brain (indicated in yellow) are also part of the haplotype. A point mutation within the *I56i* enhancer, prevents binding of GTF2I, a transcription factor deleted in Williams-Beuren syndrome, a condition characterized by hypervocalization and hypersociability (Poitras, et al. 2010).

Figure 8: Allele frequency of the *DLX5/6*-N-Haplotype in the semi-supercentenarian population

The allele frequency of the archaic form of the *DLX5/6*-N-Haplotype of the European populations present in the 1000 Genomes project is confronted to the "Toscany" population and to a population of 81 semi-supercentenarians (105+) from central Italy. While the population from Toscany does not show any significant difference from the Europeans a significant reduction of the *DLX5/6*-N-Haplotype allele frequency is observed in the semi-supercentenarian population (*: $p<0.05$).

Table 1: SNPs included in the *DLX5/6*-N-Haplotype

rs_ID	introgressed	Altai	Vindija	Denisova	Chimp	Bononbo	Gorilla	Orang
rs12704907	C	CC	CC	TT	T	T	T	T
rs4729325	C	TT	CC	TT	T	T	T	T
rs1005169	A	CC	CA	CC	C	C	C	C
rs1724297	A	GG	AG	GG	G	G	G	G
rs1207733	A	AA	AA	GG	G	G	G	G

MATERIALS AND METHODS

Animals

Procedures involving animals were conducted in accordance with European Community (Council Directive 86/609) and French Agriculture Ministry directives (Council Directive 87–848, permissions 00782 to GL). The project was reviewed and approved by the "Cuvier" ethical committee of the Muséum National d'Histoire Naturelle (approval n° 68-028r1) validated by the French Ministry of Agriculture.

Mice were housed in light, temperature and humidity controlled conditions. Food and water were available ad libitum. Mice were individually identified by microchip implanted 3 weeks postnatal. Litter sizes and genotypes were kept in record. *Dlx5/6*$^{flox/flox}$ mice (Bellessort, et al. 2016) were bred with *Slc32a1*$^{tm2(cre)Lowl}$ knock-in mice (here referred as *Vgat*$^{cre/+}$ mice) on a mixed C57BL6/N X DBA/2 genetic background to obtain the double conditional mutant *Dlx5/6*$^{VgatCre/+}$ and *Dlx5/6*^{VgatCre} in which the DNA-binding regions of both *Dlx5* and *Dlx6* are deleted by GABAergic-specific cre-recombinase.

Vocalization and social behaviour analysis

Test females were isolated 3 days before the trial to increase their motivation for social interactions. On trial day, the tested female mouse was placed in a test cage (Plexiglas, 50 cm × 25 cm × 30 cm, 100 lux, with fresh bedding) for 30 min habituation. After this time, an unfamiliar group-housed control adult female was introduced. Ultrasonic vocalizations were recorded with a condenser ultrasound microphone Polaroid/CMPA, the interface UltraSoundGate 416-200 and the software Avisoft-SASLab Pro Recorded from Avisoft Bioacoustics (sampling frequency: 300 kHz; FFT-length: 1024 points; 16-bit format). Ultrasonic vocalizations were recorded for the pair of mice tested (one control newcomer and an habituated control, *Dlx5/6*$^{VgatCre/+}$ or *Dlx5/6*^{VgatCre} mouse). All analyzed animals were 6 months old. Previous studies suggest that the contribution of the newcomer to the vocalization is minor in comparison with that of the occupant (Hammerschmidt, et al. 2012). We measured the latency for the first ultrasonic vocalizations and the total number of ultrasonic vocalizations emitted. We manually established the distribution of ultrasonic vocalizations among the following call types (see (Ey, et al. 2013)): 1) "*short*", call duration less than 5 ms;

2) "*simple*", call duration longer than 5 ms and a frequency range either smaller than 6.25 kHz (*simple flat*) or with a frequency modulation in only one direction (*simple upward* or *downward*); 3) "*complex*", frequency modulations in more than one direction and frequency range larger than 6.25 kHz (*modulated*), or inclusion of one or more additional frequency components (harmonic or non-linear phenomena, but no saturation); 4) "*unstructured*", no pure tone component and 5) "*frequency jumps*", presence of one or more jump(s) in frequency, without time gap between the consecutive elements.

We used unpaired non-parametric Wilcoxon tests to compare the latency for the first call and the call rate between genotypes. We used Chi-squared (with a Bonferroni correction for multiple testing) tests to compare the proportions of calls between genotypes.

During vocalization analysis, animals were filmed for 10 min. Social behaviour was measured using a real-time approach that couples computer vision, machine learning and Triggered-RFID identification to track and monitor animals (de Chaumont, et al. 2019).

Histological analysis

Mice were fixed by transcardiac perfusion with 4% paraformaldehyde and brains were post-fixed by overnight immersion in 4% paraformaldehyde at 4°C. Samples were cryoprotected in 30% sucrose and frozen. Cryoprotected brains were embedded in OCT (Leica, France) and 60-µm-thick free-floating cryostat sections were prepared.

For *lacZ* expression analysis, X-gal staining was performed as described (Acampora, et al. 1999) on sections of $Dlx5^{lacZ/+}$ brains. Pictures were acquired using a Leica SP5 confocal microscope.

Identification of *DLX5/6*-Neanderthal haplotype introgression

To investigate whether polymorphisms identified in the *DLX5/6* locus belong to a Neanderthal introgressed haplotype, we analyzed individuals from the 1000 Genomes dataset (The 1000 Genomes Project Consortium, 2015) and looked for Neandertal-like alleles in a window of 200kb up and downstream from *DLX5*. We identified a haplotype of 37,160bp (Grch 38 – 7: 96,991,129 – 97,022,094) defined by five consecutive SNPs (rs1207733, rs4729325, rs1005169, rs1724297, rs1207733) with Neandertal-like alleles

shared by at least 80% of these individuals. We computed the probability of generating such a long haplotype under a scenario of incomplete lineage sorting by assuming the haplotype length follows a gamma distribution (Huerta-Sanchez, et al. 2014) using the parameters from (Dannemann, et al. 2016) and the local recombination rate of 2.046 cM/Mb (Hinch, et al. 2011). The haplotype is significantly longer than expected under the null hypothesis of incomplete lineage sorting (p value = $1.08628e^{-05}$), supporting an Archaic origin. The *DLX5/6*-N-Haplotype covers part of *DLX5*, all *DLX6*, the intergenic region and regulatory region upstream of *DLX6*.

***DLX5/6*-N-Haplotype allelic frequency**

The allelic frequency of each SNP composing the *DLX5/6*-N-Haplotype and their linkage disequilibrium was determined in populations of different geographical and ethnical origins present in the 1000 Genomes dataset (The 1000 Genomes Project Consortium, 2015).

DLX5/6-N-Haplotype representation in autistic populations was determined using the MSSNG database (RK, et al. 2017) and the probability was calculated by transmission disequilibrium test (TDT).

The genomes from a cohort of 81 healthy semi-supercentenarians and supercentenarians [105+/110+] older than 105 years (mean age 106.6 ± 1.6) recruited in North, Centre and South of the Italian peninsula (CEN) were compared to a population (ITA) obtained combining the Tuscany 1000Genomes population (TSI) with 36 controls genomes matching the different geographical origin of CEN. European population correspond to TSI, CEU, FIN, GBR and IBS people present in 1000Genomes. Haplotypes were used to compute the F_{st} statistic (Wright 1951) between all the combinations of population pairs. To evaluate the significance of the F_{st} observed values we performed 500 simulations, where individuals were randomly assigned to one of the seven studied populations, and for each simulation F_{st} statistic was computed. At the end of this procedure we were able to fit a probability density function for the Fst statistic and to compute an empirical p-value for the F_{st} using the *GHap* (Utsunomiya, et al. 2016) R package. The significance of allelic differences between European population and semi-supercentenarians was also determined with a Chi-square test.

ACKNOWLEDGMENTS

This research was partially supported the ANR grants TARGETBONE (ANR-17-CE14-0024) and METACOGNITION (ANR-17-CE37-0007) to GL, grants of the "Fondation-NRJ – Institut de France" and the ATM "Cognitio" to NNN, CdL is supported by a grant of the French Ministry of Research.

We thank the following researchers for their valuable input:

Dr. Stéphane Peyregne, Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany for analysis of the DLX5/6 N-Haplotype.

Dr.s Thomas Bourgeron, Elodie Ey and Fabrice de Chaumont from the Pasteur Institute, Paris, for initial analysis of the vocalization and behavioural data.

Dr Robbie Davies from The Centre for Applied Genomics, Toronto, Canada for analysis of the association between the *DLX5/6*-N-Haplotype and autism in the MSSNG database.

A particular thank goes to the team in charge of mouse animal care and in particular M Stéphane Sosinsky and M Fabien Uridat and to Pr. Amaury de Luze in charge of the "Cuvier" ethical committee. We thank Mss. Aicha Bennana and Lanto Courcelaud for administrative assistance.

REFERENCES

Acampora D, Merlo GR, Paleari L, Zerega B, Postiglione MP, Mantero S, Bober E, Barbieri O, Simeone A, Levi G. 1999. Craniofacial, vestibular and bone defects in mice lacking the Distal-less-related gene Dlx5. Development 126:3795-3809.

Assali A, Harrington AJ, Cowan CW. 2019. Emerging roles for MEF2 in brain development and mental disorders. Curr Opin Neurobiol 59:49-58.

Bellessort B, Le Cardinal M, Bachelot A, Narboux-Neme N, Garagnani P, Pirazzini C, Barbieri O, Mastracci L, Jonchere V, Duvernois-Berthet E, et al. 2016. Dlx5 and Dlx6 control uterine adenogenesis during post-natal maturation: possible consequences for endometriosis. Hum Mol Genet 25:97-108.

Beverdam A, Merlo GR, Paleari L, Mantero S, Genova F, Barbieri O, Janvier P, Levi G. 2002. Jaw transformation with gain of symmetry after Dlx5/Dlx6 inactivation: mirror of the past? Genesis 34:221-227.

Birnbaum RY, Everman DB, Murphy KK, Gurrieri F, Schwartz CE, Ahituv N. 2012. Functional characterization of tissue-specific enhancers in the DLX5/6 locus. Hum Mol Genet 21:4930-4938.

Boeckx C, Benitez-Burraco A. 2015. Osteogenesis and neurogenesis: a robust link also for language evolution. Front Cell Neurosci 9:291.

Boeckx C, Benitez-Burraco A. 2014. The shape of the human language-ready brain. Front Psychol 5:282.

Cajigas I, Chakraborty A, Swyter KR, Luo H, Bastidas M, Nigro M, Morris ER, Chen S, VanGompel MJW, Leib D, et al. 2018. The Evf2 Ultraconserved Enhancer lncRNA Functionally and Spatially Organizes Megabase Distant Genes in the Developing Forebrain. Mol Cell 71:956-972 e959.

Capra JA, Erwin GD, McKinsey G, Rubenstein JL, Pollard KS. 2013. Many human accelerated regions are developmental enhancers. Philos Trans R Soc Lond B Biol Sci 368:20130025.

Caruso A, Ricceri L, Scattoni ML. 2020. Ultrasonic vocalizations as a fundamental tool for early and adult behavioural phenotyping of Autism Spectrum Disorder rodent models. Neurosci Biobehav Rev 116:31-43.

Chang SC, Pauls DL, Lange C, Sasanfar R, Santangelo SL. 2011. Common genetic variation in the GAD1 gene and the entire family of DLX homeobox genes and autism spectrum disorders. American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 156:233-239.

Cho KK, Hoch R, Lee AT, Patel T, Rubenstein JL, Sohal VS. 2015. Gamma rhythms link prefrontal interneuron dysfunction with cognitive inflexibility in Dlx5/6(+/-) mice. Neuron 85:1332-1343.

Conte D, Garaffo G, Lo Iacono N, Mantero S, Piccolo S, Cordenonsi M, Perez-Morga D, Orecchia V, Poli V, Merlo GR. 2016. The apical ectodermal ridge of the mouse model of ectrodactyly Dlx5;Dlx6-/- shows altered stratification and cell polarity, which are restored by exogenous Wnt5a ligand. Hum Mol Genet 25:740-754.

Crackower MA, Scherer SW, Rommens JM, Hui CC, Poorkaj P, Soder S, Cobben JM, Hudgins L, Evans JP, Tsui LC. 1996. Characterization of the split hand/split foot malformation locus SHFM1 at 7q21.3-q22.1 and analysis of a candidate gene for its expression during limb development. Hum Mol Genet 5:571-579.

Dannemann M, Andres AM, Kelso J. 2016. Introgression of Neandertal- and Denisovan-like Haplotypes Contributes to Adaptive Variation in Human Toll-like Receptors. Am J Hum Genet 98:22-33.

de Chaumont F, Ey E, Torquet N, Lagache T, Dallongeville S, Imbert A, Legou T, Le Sourd AM, Faure P, Bourgeron T, et al. 2019. Real-time analysis of the behaviour of groups of mice via a depth-sensing camera and machine learning. Nat Biomed Eng 3:930-942.

de Lombares C, Heude E, Alfama G, Fontaine A, Hassouna R, Vernochet C, de Chaumont F, Olivo-Marin C, Ey E, Parnaudeau S, et al. 2019. Dlx5 and Dlx6 expression in GABAergic neurons controls behaviour, metabolism, healthy aging and lifespan. Aging (Albany NY) 11:6638-6656.

Dediu D, Christiansen MH. 2016. Language Evolution: Constraints and Opportunities From Modern Genetics. Top Cogn Sci 8:361-370.

Dediu D, Levinson SC. 2013. On the antiquity of language: the reinterpretation of Neandertal linguistic capacities and its consequences. Front Psychol 4:397.

Dennison CA, Legge SE, Pardinas AF, Walters JTR. 2019. Genome-wide association studies in schizophrenia: Recent advances, challenges and future perspective. Schizophr Res:4-12.

Depew MJ, Lufkin T, Rubenstein JL. 2002. Specification of jaw subdivisions by Dlx genes. Science 298:381-385.

Dolgova O, Lao O. 2018. Evolutionary and Medical Consequences of Archaic Introgression into Modern Human Genomes. Genes (Basel) 9.
Duveau J, Berillon G, Verna C, Laisne G, Cliquet D. 2019. The composition of a Neandertal social group revealed by the hominin footprints at Le Rozel (Normandy, France). Proc Natl Acad Sci U S A 116:19409-19414.
Ey E, de Chaumont F, Bourgeron T. 2020. Spontaneous social communication in laboratory mice - placing ultrasonic vocalizations in their behavioural context. bioRxiv 2020.07.09.195362.
Ey E, Torquet N, Le Sourd AM, Leblond CS, Boeckers TM, Faure P, Bourgeron T. 2013. The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations. Behav Brain Res 256:677-689.
Faedo A, Quinn JC, Stoney P, Long JE, Dye C, Zollo M, Rubenstein JL, Price DJ, Bulfone A. 2004. Identification and characterization of a novel transcript down-regulated in Dlx1/Dlx2 and up-regulated in Pax6 mutant telencephalon. Dev Dyn 231:614-620.
Fazel Darbandi S, Poitras L, Monis S, Lindtner S, Yu M, Hatch G, Rubenstein JL, Ekker M. 2016. Functional consequences of I56ii Dlx enhancer deletion in the developing mouse forebrain. Dev Biol.
Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. 2006. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev 20:1470-1484.
Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, Won H, van Bakel H, Varghese M, Wang Y, et al. 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362.
Hajdinjak M, Fu Q, Hubner A, Petr M, Mafessoni F, Grote S, Skoglund P, Narasimham V, Rougier H, Crevecoeur I, et al. 2018. Reconstructing the genetic history of late Neanderthals. Nature 555:652-656.
Hamilton SP, Woo JM, Carlson EJ, Ghanem N, Ekker M, Rubenstein JL. 2005. Analysis of four DLX homeobox genes in autistic probands. BMC genetics 6:52.
Hammerschmidt K, Reisinger E, Westekemper K, Ehrenreich L, Strenzke N, Fischer J. 2012. Mice do not require auditory input for the normal development of their ultrasonic vocalizations. BMC Neurosci 13:40.
Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, et al. 2011. The landscape of recombination in African Americans. Nature 476:170-175.
Holy TE, Guo Z. 2005. Ultrasonic songs of male mice. PLoS Biol 3:e386.
Huerta-Sanchez E, Jin X, Asan, Bianba Z, Peter BM, Vinckenbosch N, Liang Y, Yi X, He M, Somel M, et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194-197.
Huguet G, Benabou M, Bourgeron T. 2016. The Genetics of Autism Spectrum Disorders. In: Sassone-Corsi P, Christen Y, editors. A Time for Metabolism and Hormones. Cham (CH). p. 101-129.
Jurgens U. 2009. The neural control of vocalization in mammals: a review. J Voice 23:1-10.
Jurgens U. 2002. Neural pathways underlying vocal control. Neurosci Biobehav Rev 26:235-258.
Konopka G, Roberts TF. 2016. Insights into the Neural and Genetic Basis of Vocal Communication. Cell 164:1269-1276.
Lahvis GP, Alleva E, Scattoni ML. 2011. Translating mouse vocalizations: prosody and frequency modulation. Genes Brain Behav 10:4-16.
Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, et al. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478:476-482.
Medvedeva VP, Rieger MA, Vieth B, Mombereau C, Ziegenhain C, Ghosh T, Cressant A, Enard W, Granon S, Dougherty JD, et al. 2019. Altered social behaviour in mice carrying a cortical Foxp2 deletion. Hum Mol Genet 28:701-717.
Merlo GR, Zerega B, Paleari L, Trombino S, Mantero S, Levi G. 2000. Multiple functions of Dlx genes. Int J Dev Biol 44:619-626.
Moles A, Costantini F, Garbugino L, Zanettini C, D'Amato FR. 2007. Ultrasonic vocalizations emitted during dyadic interactions in female mice: a possible index of sociability? Behav Brain Res 182:223-230.
Nakashima N, Yamagata T, Mori M, Kuwajima M, Suwa K, Momoi MY. 2010. Expression analysis and mutation detection of DLX5 and DLX6 in autism. Brain & development 32:98-104.
Perera M, Merlo GR, Verardo S, Paleari L, Corte G, Levi G. 2004. Defective neuronogenesis in the absence of Dlx5. Molecular and cellular neurosciences 25:153-161.

Peyregne S, Boyle MJ, Dannemann M, Prufer K. 2017. Detecting ancient positive selection in humans using extended lineage sorting. Genome Res 27:1563-1572.
Poitras L, Yu M, Lesage-Pelletier C, Macdonald RB, Gagne JP, Hatch G, Kelly I, Hamilton SP, Rubenstein JL, Poirier GG, et al. 2010. An SNP in an ultraconserved regulatory element affects Dlx5/Dlx6 regulation in the forebrain. Development 137:3089-3097.
Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43-49.
Racimo F. 2016. Testing for Ancient Selection Using Cross-population Allele Frequency Differentiation. Genetics 202:733-750.
Racimo F, Kuhlwilm M, Slatkin M. 2014. A test for ancient selective sweeps and an application to candidate sites in modern humans. Mol Biol Evol 31:3344-3358.
Rasmussen MB, Kreiborg S, Jensen P, Bak M, Mang Y, Lodahl M, Budtz-Jorgensen E, Tommerup N, Tranebjaerg L, Rendtorff ND. 2016. Phenotypic subregions within the split-hand/foot malformation 1 locus. Hum Genet 135:345-357.
RK CY, Merico D, Bookman M, J LH, Thiruvahindrapuram B, Patel RV, Whitney J, Deflaux N, Bingham J, Wang Z, et al. 2017. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci 20:602-611.
Robledo RF, Rajan L, Li X, Lufkin T. 2002. The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. Genes Dev 16:1089-1101.
Rotival M, Quintana-Murci L. 2020. Functional consequences of archaic introgression and their impact on fitness. Genome Biol 21:3.
Sales GD. 1972. Ultrasound and aggressive behaviour in rats and other small mammals. Anim Behav 20:88-100.
Sangiamo DT, Warren MR, Neunuebel JP. 2020. Ultrasonic signals associated with different types of social behaviour of mice. Nat Neurosci 23:411-422.
Sankararaman S, Mallick S, Dannemann M, Prufer K, Kelso J, Paabo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507:354-357.
Scherer SW, Poorkaj P, Massa H, Soder S, Allen T, Nunes M, Geshuri D, Wong E, Belloni E, Little S, et al. 1994. Physical mapping of the split hand/split foot locus on chromosome 7 and implication in syndromic ectrodactyly. Hum Mol Genet 3:1345-1354.
Schmeisser MJ, Ey E, Wegener S, Bockmann J, Stempel AV, Kuebler A, Janssen AL, Udvardi PT, Shiban E, Spilker C, et al. 2012. Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. Nature 486:256-260.
Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebbring SJ, Jarvik GP, Kullo IJ, et al. 2016. The phenotypic legacy of admixture between modern humans and Neandertals. Science 351:737-741.
Simonyan K. 2014. The laryngeal motor cortex: its organization and connectivity. Curr Opin Neurobiol 28:15-21.
Snyder-Mackler N, Burger JR, Gaydosh L, Belsky DW, Noppert GA, Campos FA, Bartolomucci A, Yang YC, Aiello AE, O'Rand A, et al. 2020. Social determinants of health and survival in humans and other animals. Science 368:eaax9553.
Tschida K, Michael V, Takatoh J, Han BX, Zhao S, Sakurai K, Mooney R, Wang F. 2019. A Specialized Neural Circuit Gates Social Vocalizations in the Mouse. Neuron 103:459-472 e454.
Utsunomiya YT, Milanesi M, Utsunomiya AT, Ajmone-Marsan P, Garcia JF. 2016. GHap: an R package for genome-wide haplotyping. Bioinformatics 32:2861-2862.
Vacante M, D'Agata V, Motta M, Malaguarnera G, Biondi A, Basile F, Malaguarnera M, Gagliano C, Drago F, Salamone S. 2012. Centenarians and supercentenarians: a black swan. Emerging social, medical and surgical problems. BMC Surg 12 Suppl 1:S36.
Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE. 2006. Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum Mol Genet 15:3154-3167.
Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, et al. 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.

Wang Y, Dye CA, Sohal V, Long JE, Estrada RC, Roztocil T, Lufkin T, Deisseroth K, Baraban SC, Rubenstein JL. 2010. Dlx5 and Dlx6 regulate the development of parvalbumin-expressing cortical interneurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 30:5334-5345.
Wohr M. 2014. Ultrasonic vocalizations in Shank mouse models for autism spectrum disorders: detailed spectrographic analyses and developmental profiles. Neurosci Biobehav Rev 43:199-212.
Wright S. 1951. The genetical structure of populations. Ann Eugen 15:323-354.
Yang YC, Boen C, Gerken K, Li T, Schorpp K, Harris KM. 2016. Social relationships and physiological determinants of longevity across the human life span. Proc Natl Acad Sci U S A 113:578-583.

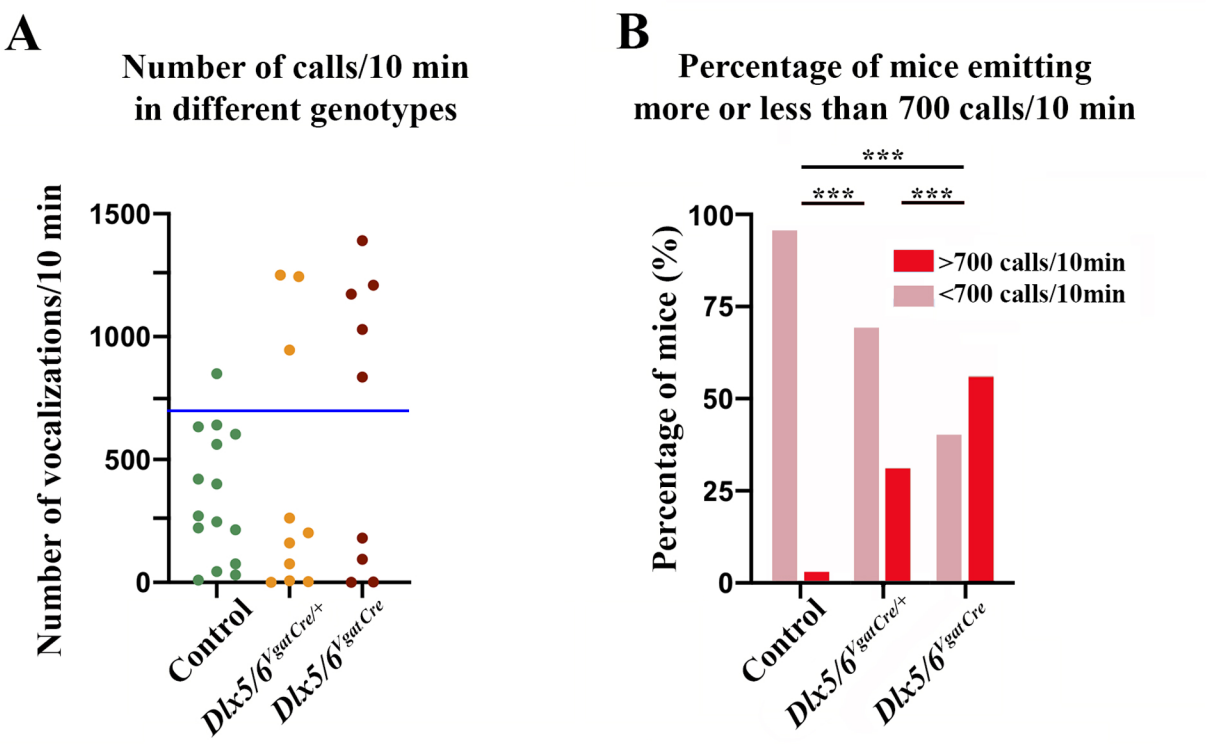
A
Number of calls/10 min
in different genotypes
Number of vocalizations/10 min
1500
1000
500
0
Control
Dlx5/6Vgat Cre/+
Dlx5/6Vgat Cre
B
Percentage of mice emitting
more or less than 700 calls/10 min
***
***
***
Percentage of mice (%)
100
75
50
25
0
>700 calls/10min
<700 calls/10min
Control
Dlx5/6Vgat Cre/+
Dlx5/6Vgat Cre

Figure 1

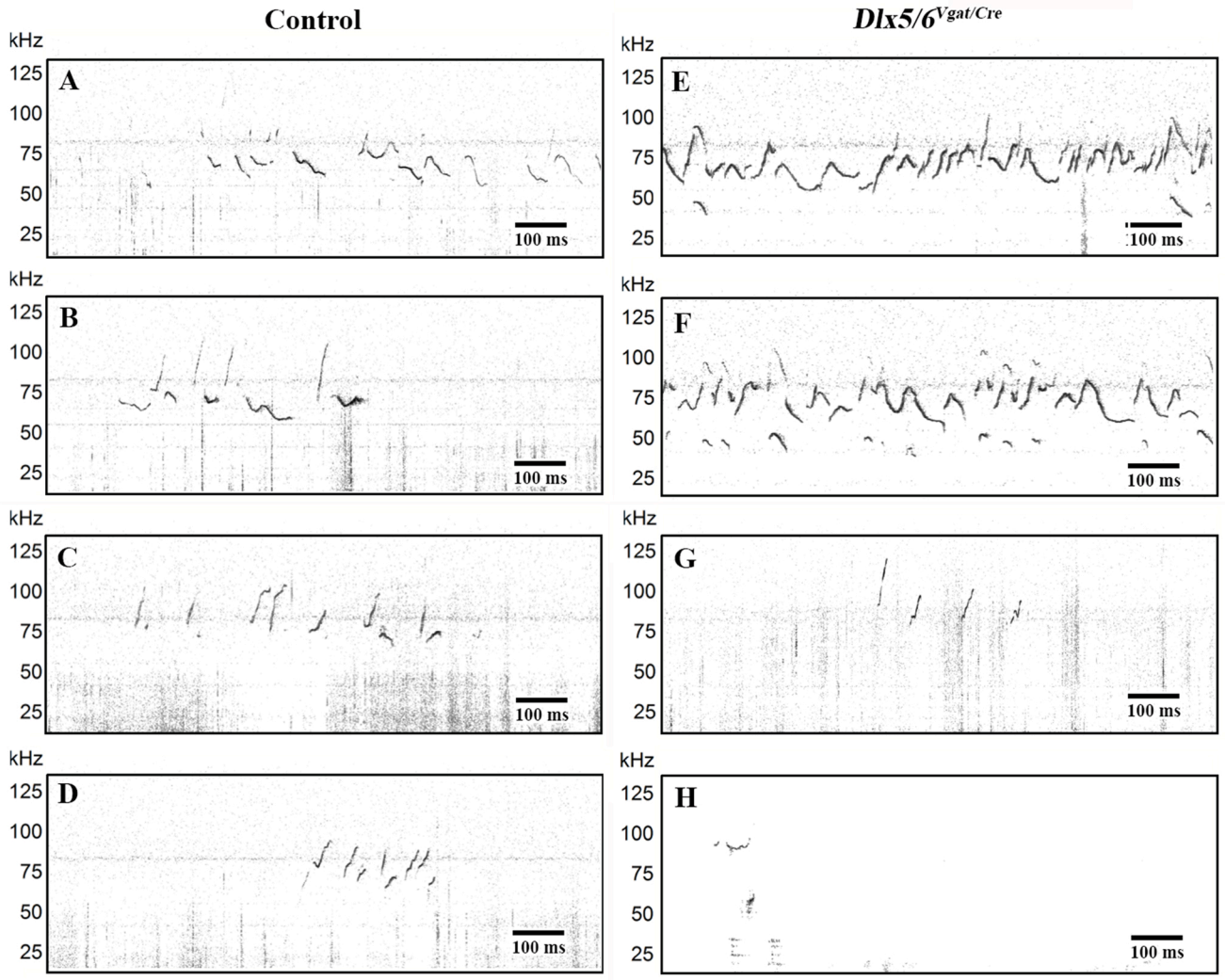

Figure 2

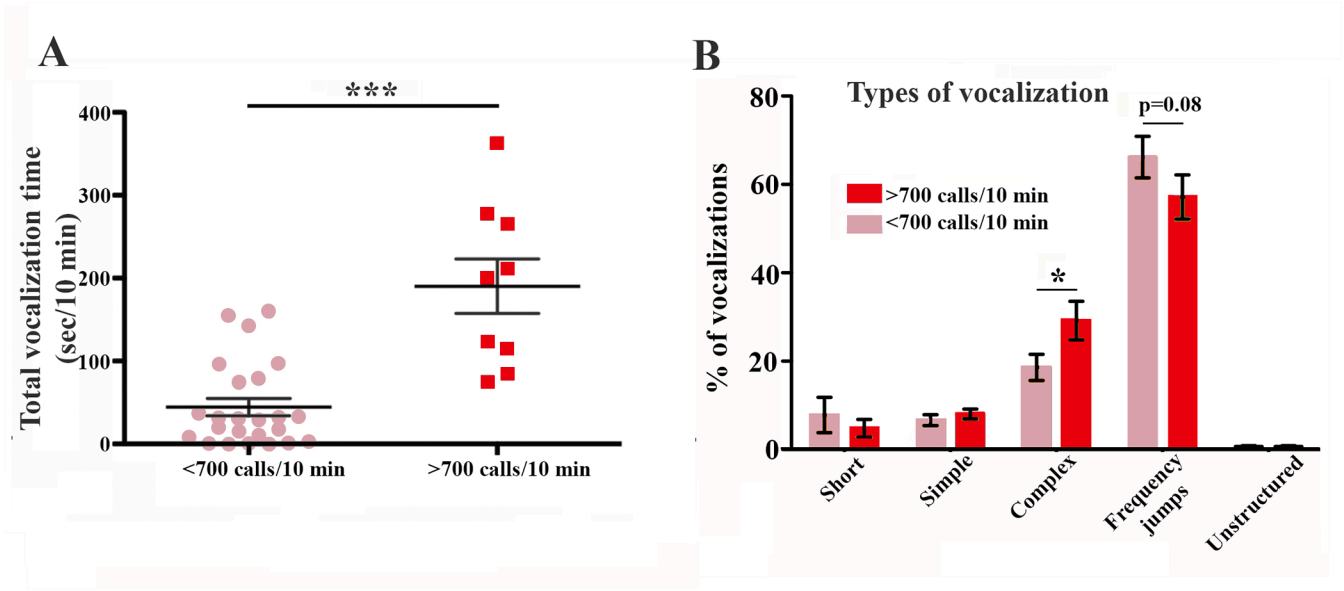
A
***
Total vocalization time (sec/10 min)
400
300
200
100
0
<700 calls/10 min
>700 calls/10 min
B
Types of vocalization
p=0.08
>700 calls/10 min
<700 calls/10 min
*
% of vocalizations
80
60
40
20
0
Short
Simple
Complex
Frequency jumps
Unstructured

Figure 3

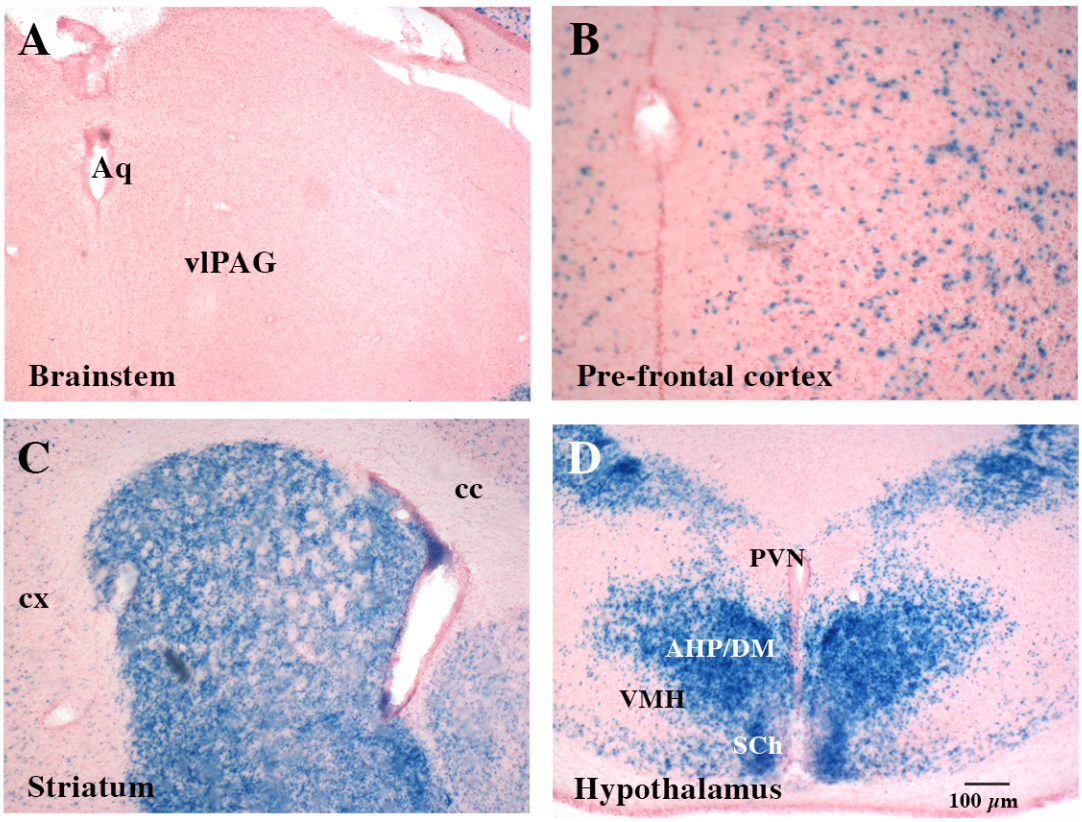
A
Aq
vlPAG
Brainstem
B
Pre-frontal cortex
C
cc
cx
Striatum
D
PVN
AHP/DM
VMH
SCh
Hypothalamus
100 μm

Figure 4

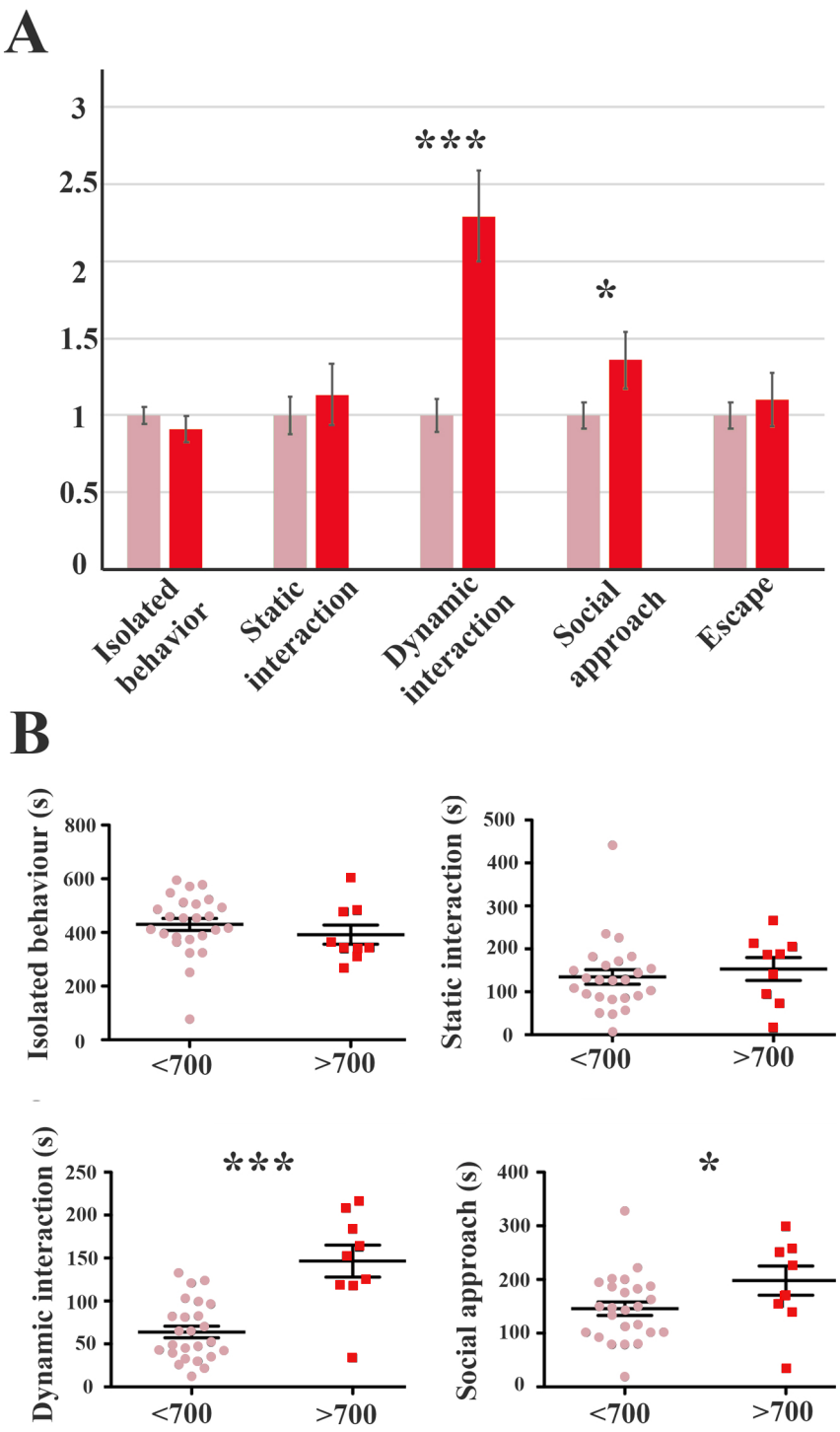
A
3
2.5
2
1.5
1
0.5
0
***
*
Isolated behavior
Static interaction
Dynamic interaction
Social approach
Escape
B
Isolated behaviour (s)
800
600
400
200
0
<700
>700
Static interaction (s)
500
400
300
200
100
0
<700
>700
Dynamic interaction (s)
250
200
150
100
50
0
***
<700
>700
Social approach (s)
400
300
200
100
0
*
<700
>700

Figure 5

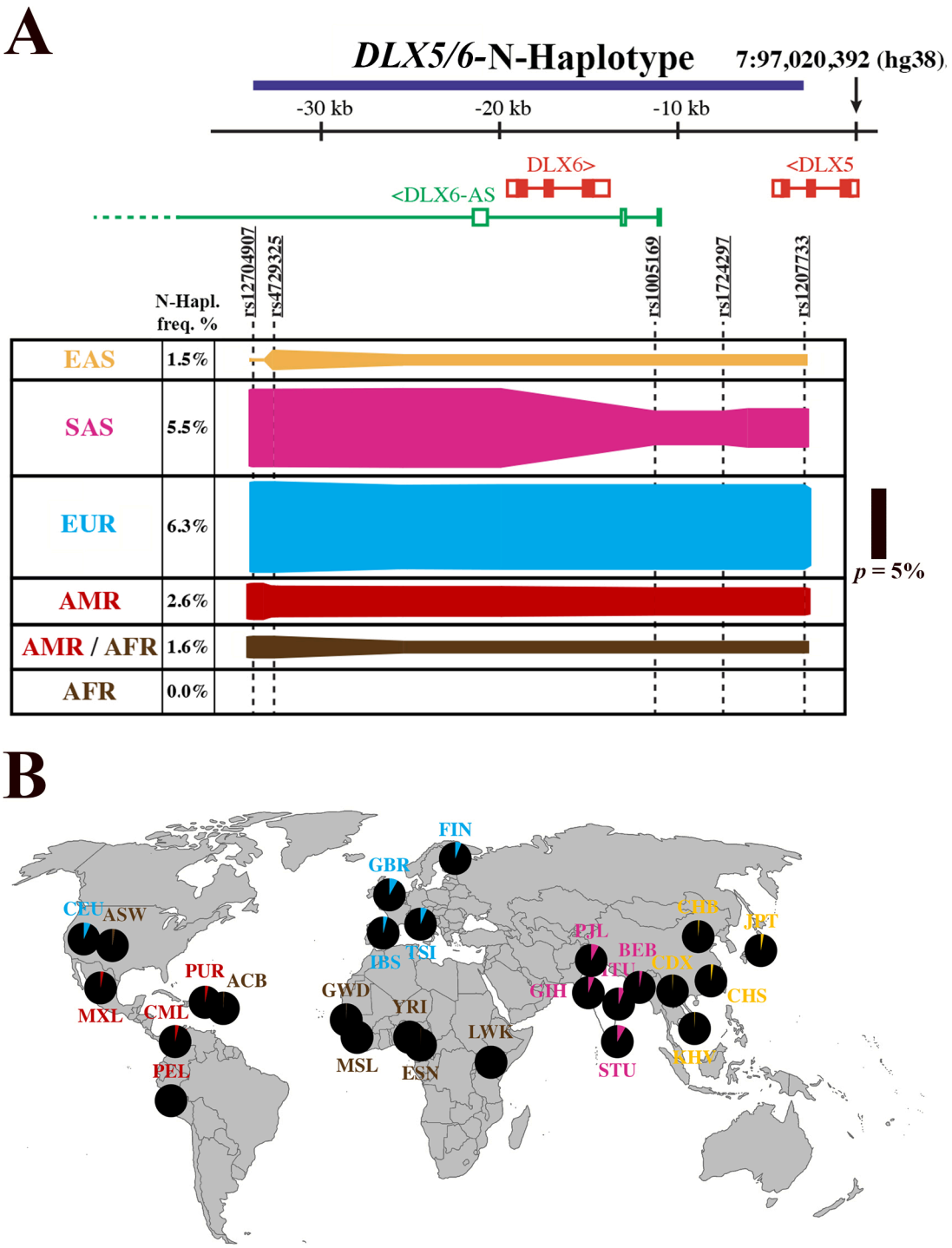
A
DLX5/6-N-Haplotype
7:97,020,392 (hg38)
-30 kb
-20 kb
-10 kb
DLX6>
<DLX5
<DLX6-AS
rs12704907
rs4729325
rs1005169
rs1724297
rs1207733
N-Hapl. freq. %
EAS 1.5%
SAS 5.5%
EUR 6.3%
AMR 2.6%
AMR / AFR 1.6%
AFR 0.0%
p = 5%
B
FIN
GBR
CEU
ASW
IBS
TSI
PUR
ACB
MXL
CML
PEL
GWD
YRI
MSL
ESN
LWK
PJL
BEB
ITU
GIH
STU
CHB
JPT
CDX
CHS
KHV

Figure 6

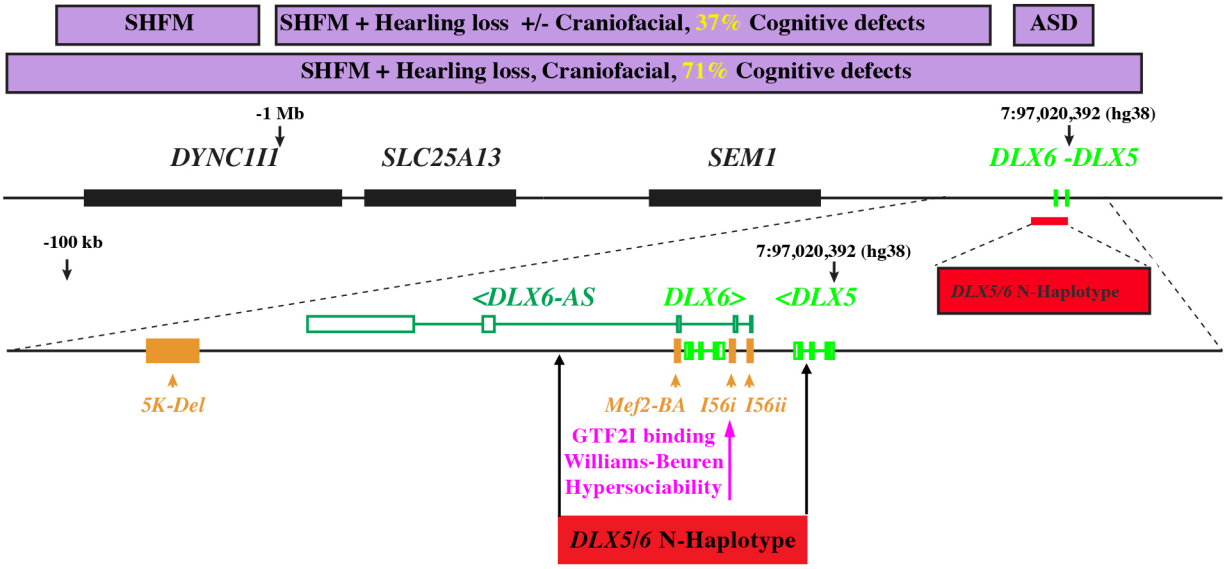
SHFM
SHFM + Hearling loss +/- Craniofacial, 37% Cognitive defects
ASD
SHFM + Hearling loss, Craniofacial, 71% Cognitive defects
-1 Mb
7:97,020,392 (hg38)
DYNC1I1
SLC25A13
SEM1
DLX6 -DLX5
-100 kb
7:97,020,392 (hg38)
<DLX6-AS
DLX6>
<DLX5
DLX5/6 N-Haplotype
5K-Del
Mef2-BA
I56i
I56ii
GTF2I binding
Williams-Beuren
Hypersociability
DLX5/6 N-Haplotype

Figure 7

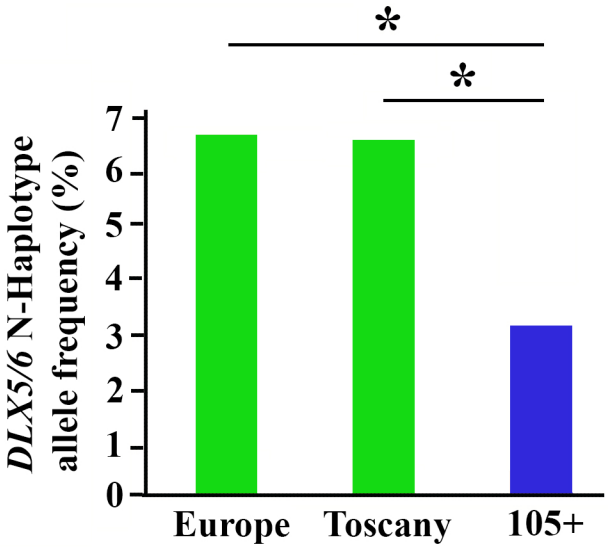
*
*
DLX5/6 N-Haplotype allele frequency (%)
7
6
5
4
3
2
1
0
Europe
Toscany
105+

Figure 8