UC Davis

UC Davis Previously Published Works

Title

No Evidence for Recent Selection at FOXP2 among Diverse Human Populations

Permalink

https://escholarship.org/uc/item/0mx934v1

Journal

Cell, 174(6)

ISSN

0092-8674

Authors

Atkinson, Elizabeth Grace
Audesse, Amanda Jane
Palacios, Julia Adela
et al.

Publication Date

2018-09-01

DOI

10.1016/j.cell.2018.06.048

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2019 September 06.

Published in final edited form as:

Cell. 2018 September 06; 174(6): 1424–1435.e15. doi:10.1016/j.cell.2018.06.048.

No evidence for recent selection at *FOXP2* among diverse human populations

Elizabeth Grace Atkinson[1,*,8,†], **Amanda Jane Audesse**[2,3], **Julia Adela Palacios**[4,5], **Dean Michael Bobo**[1], **Ashley Elizabeth Webb**[2,6], **Sohini Ramachandran**[4], and **Brenna Mariah Henn**[1,7,*]

[1]Dept. of Ecology and Evolution, Stony Brook University, Stony Brook, NY

[2]Dept. of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, 02912

[3]Neuroscience Graduate Program, Brown University, Providence, RI, 02912

[4]Dept. of Ecology and Evolutionary Biology and Center for Computational Molecular Biology, Brown University, Providence, RI, 02912

[5]Dept. of Statistics and Dept. of Biomedical Data Science, Stanford University, Stanford, CA 94305

[6]Carney Institute for Brain Science, Brown University, Providence, RI, 02912

[7]Dept. of Anthropology and the Genome Center, University of California – Davis, Davis, CA, 95616

SUMMARY

FOXP2 , initially identified for its role in human speech, contains two nonsynonymous substitutions derived in the human lineage. Evidence for a recent selective sweep in *Homo sapiens*, however, is at odds with the presence of these substitutions in archaic hominins. Here, we comprehensively reanalyze *FOXP2* in hundreds of globally distributed genomes to test for recent selection. We do not find evidence for recent positive or balancing selection at *FOXP2*. The original signal appears to have instead been due to sample composition. Our tests do identify an intronic region that is enriched for highly-conserved sites that are polymorphic among humans,

[*]Correspondence and requests for materials should be addressed to E.G.A. or B.M.H., bmhenn@ucdavis.edu or eatkinso@broadinstitute.org.

[8]Lead Contact

[†]First author current address: Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114; Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, 02142

DECLARATION OF INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization, B.M.H. and E.G.A; Methodology, B.M.H., E.G.A., A.W. and J.A.P.; Formal Analysis, E.G.A.; Investigation, E.G.A., A.A., J.A.P., D.M.B.; Writing – Original Draft, E.G.A; Writing – Review & Editing, E.G.A., B.M.H, A.W., S.R., J.A.P.; Funding Acquisition, E.G.A., B.M.H., S.R.; Resources, A.W. and S.R.; Supervision, B.M.H, S.R., A.W.

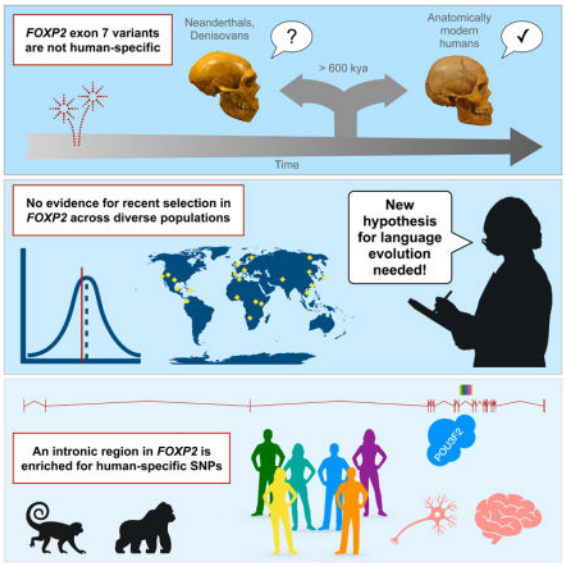
Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

compatible with a loss of function in humans. This region is lowly expressed in relevant tissue types, tested via RNA-seq on human prefrontal cortex and RT-PCR in immortalized human brain cells. Our results represent a substantial revision to the adaptive history of *FOXP2*, a gene regarded as vital to human evolution.

In brief

An in-depth examination of diverse sets of human genomes argues against a recent selective evolutionary sweep of FOXP2, a gene that was believed to be critical for speech evolution in early hominins.



INTRODUCTION

Complex spoken language is a uniquely human characteristic, ubiquitous among all global populations. Language development was vital to human evolution, likely enabling improved social organization, more efficient exchange of information, and potentially facilitating symbolic thought and abstraction (Klein, 2000). Isolating the genetic basis for speech, therefore, would allow reconstruction of the evolution of these behaviors, including in which hominin species they arose. The Forkhead box protein P2 (*FOXP2*) gene encodes a transcription factor that is expressed at high levels in the brain during fetal development, as well as in the lung and gut (Lai et al., 2001; Shu et al., 2001). Stop-gain mutations in *FOXP2* are associated with language dysfunction in individuals who are otherwise intellectually normal and follow an autosomal-dominant pattern of inheritance (Lai et al., 2003; MacDermot et al., 2005). *FOXP2* has been shown *in vitro* and *in vivo* to affect brain development and neural plasticity in humans and mice (Chiu et al., 2014; Enard et al., 2009; Reimers-Kipping et al., 2011; Španiel et al., 2011), and its altered expression affects brain function in language-related cortical areas (Fujita-Jimbo and Momoi, 2014; Konopka et al., 2009; Pinel et al., 2012; Spiteri et al., 2007). Intriguingly, it has also been shown to affect language-like behaviors in other animals, including in birdsong and juvenile mouse

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

vocalizations (Chabout et al., 2016; Haesler et al., 2004; Shu et al., 2005; Teramitsu et al., 2010) as well as other primates (Staes et al., 2017). To date, *FOXP2* is the only known autosomal-dominant language-related gene.

Initially investigated as the first gene implicated in speech and language (Fisher et al., 1998; Lai et al., 2000), Enard et al., (2002) found evidence for both accelerated evolution on the hominin lineage and a recent selective sweep in *FOXP2*. This led to speculation that *FOXP2* played a key role in the development of modern language, unique to *Homo sapiens*. The target of this sweep appeared to be two derived, protein-coding missense substitutions in humans that are absent in all other primates. Population genetic summary statistics (i.e. Tajima's *D*) calculated from intronic variation near the two derived amino acid substitutions indicated that the selective sweep occurred relatively recently – within the past 200,000 years. This model was later criticized, however, when ancient hominin genomes (Neanderthals and Denisovans) were found to also carry the substitutions, which is incompatible with a recent selective sweep specific to *Homo sapiens* (Krause et al., 2007 but see Coop et al., 2008); Neanderthals, Denisovans, and humans diverged at least ~600 *kya* (Kuhlwilm et al., 2016; Prüfer et al., 2014; Racimo et al., 2015).

Several alternate hypotheses have been proposed to reconcile these findings. The haplotype carrying the two derived amino acids could have been present as standing variation in the common ancestor of humans and other ancient hominins and then positively selected in humans after species divergence ~500 *kya*. However, for this to be the case, the haplotype would have had to be at high frequency in the ancestral population, which would dramatically reduce the ability to detect a recent selective signal in modern genomes (Przeworski et al., 2005). Krause et al. (2007) instead suggested a very old sweep targeted these two amino acids beginning before the split of humans and Neanderthals 300–400 *kya*, with the alleles reaching fixation ~260 *kya.* However, as Coop et al. (2008) point out, the observed pattern of an excess of high-frequency derived alleles in the intron preceding the exonic substitutions is indicative of a recent fixation of the linked allele, as low-frequency variation is quickly lost from populations. The ancestral alleles would have had to remain segregating at low frequency for over 300,000 years of human and Neanderthal history, which is unlikely. Gene conversion has been demonstrated to be insufficient for explaining how these two substitutions could be in Neanderthals (Ptak et al., 2009).

Despite extensive discussion over the past 15 years, the initial hypothesis of a recent selective sweep at exon 7 of *FOXP2* has not been systematically re-evaluated. This is especially concerning given that the model was based on a limited Sanger sequencing dataset of only three introns and a small sample of humans (n=20) (Enard et al., 2002). It has been suggested that alternative targets could be responsible for the sweep apart from these two substitutions (Maricic et al., 2013; Ptak et al., 2009). By comparing Neanderthal and Denisovan sequences to a panel of 50 modern humans, Maricic et al. (2013) identify a nearly-fixed, derived polymorphism (rs114972925) in intron 9 that affects the binding of an upstream transcription factor, POU3F2, in HeLa cells. The derived version of this variant was found to be less efficient at binding POU3F2 compared to the ancestral version, in addition to binding a different proportion of POU3F2 dimers versus monomers, which they argue is suggestive of an effect on subsequent FOXP2 expression (Maricic et al., 2013). We

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are not aware of any additional investigations of the selective history of this or other areas of *FOXP2* since the initial work or examination of this region in human brain cells, which is the tissue type most relevant to FOXP2's presumed primary function.

Major advances in genome sequencing technology and efforts to ascertain variation in global populations now provide the full sequence of *FOXP2* and its downstream target genes in thousands of individuals. Here, we re-analyze the evolution of *FOXP2* in diverse human genomes and reconcile contradictory previous *FOXP2* research using a combination of computational population genomic analysis and functional experiments. We test the *FOXP2* locus for signatures of recent selection, investigating the haplotype structure and site frequency spectrum of the region, and use Sanger sequencing to validate the key polymorphisms observed in error-prone next-generation sequencing data. We also follow-up an intronic area, identified in Maricic et al. (2013) and detected as an outlier in our genome-wide scans of sequence conservation, by analyzing RNA-seq data from fresh human prefrontal cortex samples and conducting RT-PCR in an immortalized human brain cell line, an appropriate cell type for such analysis. Our results substantially alter and advance the understanding of the processes that have shaped genetic variation in the famous *FOXP2* gene across diverse human groups.

RESULTS

No evidence for recent positive selection on FOXP2

To test for indications of a recent selective sweep at the *FOXP2* locus, we replicated the original analysis run by Enard et al. (2002) that led to this hypothesis using a large, modern dataset. Specifically, we calculated Tajima's D using two datasets: 53 high-coverage genomes from global human populations sequenced as part of the Human Genome Diversity Panel (HGDP) (Henn et al., 2016) and populations from the 1000 Genomes phase 1 dataset (1000G) (The 1000 Genomes Project Consortium, 2012) to test for signals of selection in *FOXP2*. These datasets have a substantially higher sample size compared to the earlier *FOXP2* studies, include a more diverse panel of individuals, and provide information across the entire *FOXP2* gene region and rest of the genome rather than only the three introns preceding exon 7. To detect deviations from expected values in each population, we constructed a null distribution for the D statistic in each data subset by calculating D on contiguous non-overlapping windows of the same genetic length as *FOXP2* throughout the autosomes. This provides the expected range of the D values in the genome as determined primarily by neutral population histories (STAR Methods).

Calculation of D in the entire pooled HGDP genomes dataset (composed of ~75% non-African individuals in a ratio approximately matching the African/Eurasian proportion in the original Enard et al. work) replicates a significant D value of −1.305 (Figure 1A, purple; Table 1). This places *FOXP2* in the 5% lower tail of the empirical distribution, which is generally interpreted as an indication of positive selection (Yu et al., 2009). However, testing D in the African individuals in this dataset separately from those whose ancestors underwent the Out-of-Africa (OoA) expansion erases such a signal (African D= −0.573, OoA D= −0.659) (Figure 1A, red, blue). The D values for African and OoA individuals against their own background calibrations remain non-significant (Figure 1A, Table 1). This pattern is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mirrored when conducting the same test in another dataset, 1000 Genomes Phase 1 (Figure 1B).

The results presented in Figure 1 were conducted with a similar analysis design to Enard et al. (2007) for the sake of a direct comparison to the analyses that generated the hypothesis of recent positive selection on *FOXP2*. These results remain robust when comparing *FOXP2* to genes with similar exonic content and thus that are arguably under similar levels of evolutionary constraint (STAR Methods, Figure S1A–F). Examining DNA sequence constraint directly using gene-level Genomic Evolutionary Rate Profiling (GERP) scores, *FOXP2* is not significantly different from the average autosomal GERP score for all canonical transcripts (Figure S1G). This eliminates concern that the genomic makeup or overall level of DNA sequence constraint on *FOXP2* is skewing our Tajima's *D* results.

We additionally tested for selection using Fay and Wu's *H*, as this statistic may better control for demographic processes and was also calculated in Enard et al. (2002). Rather than comparing to Watterson's theta, θ_w, *H* is the difference in the estimates of θ based on heterozygosity of ancestral variants relative to the weighted homozygosity of derived variants. This approach tests for an excess of derived variants at high frequency, which is then interpreted to result from hitchhiking with a linked adaptive variant. We calculated *H* on the entire *FOXP2* gene region in the HGDP genomes using the same population subsets as in Tajima's *D*. Values mirrored the *D* results qualitatively, though none reached α =0.05 significance (STAR Methods).

In sum, our results do not support recent positive selection at *FOXP2*. We demonstrate that the original finding suggesting evidence for a recent selective sweep instead appears to have been a result of sample ancestry composition, namely the inclusion of predominantly but not entirely individuals of Eurasian descent.

Inconclusive evidence for ancient selection at FOXP2

Ancient selection could also explain the presence of the two derived substitutions in exon 7 identified by Enard et al. (2002). To test for the possibility of ancient selection (>200 *kya*) on *FOXP2*, we conducted a McDonald-Kreitman (MK) test on *FOXP2*'s coding sequence by comparing the variation within the HGDP genomes to a population of 10 chimpanzees from the PanMap collection (Auton et al., 2012). We obtained a non-significant McDonald-Kreitman value of −1.0625 (p=0.68). A negative test result is generally interpreted to suggest positive selection, though the test did not reach significance in this case.

No evidence for balancing selection at FOXP2

We ran a coalescent-based test to assess if balancing selection was operating on the *FOXP2* region rather than a selective sweep. A chromosomal region under balancing selection has a deeper genealogy than expected under neutrality, with older time to the most recent common ancestor (TMRCA) because different haplotypes are maintained for long periods of time (Kaplan et al., 1988). We ran ARGweaver to estimate the TMRCA of *FOXP2* compared to the background TMRCA values in an effort to assess whether *FOXP2* had a significantly deeper coalescence time, which would be indicative of balancing selection. We did not find any significantly elevated TMRCA across *FOXP2* in these 7 samples compared to the 10 Mb

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of comparison sequence (Figure S2, STAR Methods). Therefore, we do not find evidence of balancing selection affecting the *FOXP2* locus, either.

Window size as well as population composition dramatically affect summary statistics

We report that the value of *D* shifts dramatically depending on the population composition. Next, we investigated the effect of window size on summary statistics. We conducted a "gene crawl" along the length of the *FOXP2* gene to examine the distribution of *D* values for windows of various sizes across *FOXP2* (STAR Methods). In both 1.5kb and 6kb windows, *D* varied substantially even within quite proximal regions (Figure 1C–F). This demonstrates that the scale of the area selected for calculation has a dramatic effect on *D*, rendering uncalibrated interpretation of the value obtained from any given region challenging.

An intronic area in FOXP2 is enriched for constrained, human-specific polymorphisms

To avoid these complications of window-base metrics, we next scanned genome-wide using a site-based metric of evolutionary conservation, the GERP score. We identified an intronic region of interest ("ROI") in *FOXP2* of 2251 bp that contains several common SNPs with GERP scores >3. Sites with GERP scores greater than 3 are generally considered to be evolutionarily conserved. The ROI lies between exons 8 and 9 of the primary FOXP2 transcript (isoform 1, NM_014491); we define the ROI by the bordering SNPs at hg19 positions chr7:114288164 and chr7:114290415. In all cases save for rs115978361, the SNPs in the ROI are uniquely derived in humans as compared to other primates and archaic hominins, and are present in more than one individual in the 1000G and HGDP data (Table 2). No SNPs are present in the ROI in a sample of 10 chimpanzees (Auton et al., 2012) and no fixed substitutions are present compared to three archaic hominins (Meyer et al., 2012; Prüfer et al., 2014, 2017) across the entire 2251 bp region. There is one fully derived site in our human samples that is heterozygous in the Altai Neanderthal (chr7:114288583) and 9 fixed differences compared to the PanMap chimps (Table S1). Interestingly, this region includes rs114972925**,** the SNP identified in Maricic et al. (2012).

As there are relatively high false negative and false positive error rates in next-generation sequencing data (Bobo et al., 2016), we validated SNPs in the ROI with Sanger sequencing. Specifically, we Sanger sequenced a 2,749 bp region including the ROI with 7 primers spanning both directions from DNA extracted from 40 HGPD-CEPH individuals (Cann et al., 2002) and 7 South African Khomani San individuals (Uren et al., 2016) (STAR Methods). Three additional SNPs were detected in the KhoeSan at high frequency that were not found in the HGDP NGS data, one of which (rs577428580) had an exceptional GERP score (4.89).

Of the 8 SNPs we identify with a GERP score >3, none are fully fixed among human populations. rs114972925 was highlighted in Maricic et al. (2012) as an altered TFBS and a potential target for a selective sweep. However, we find that the ancestral allele segregates at 43%, 17% and 7% frequency in the Khomani San, Namibian San and Kenyan Luhya, respectively. This pattern of variation is inconsistent with a selective sweep in the human ancestral population, and likely reflects the general pattern of declining south to north effective population size within Africans (Henn et al., 2011). We do not observe any

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

homozygotes for rs114972925 in the 1000G, although given the low frequency of the SNP, the lack of homozygotes does not significantly depart from Hardy-Weinberg expectations. The KhoeSan contain other common SNPs with remarkably high GERP scores, including: rs191654848, rs560859215, and rs577428580 which vary from GERPs of ~4 to 6 (Table 2). GERP SNPs >4 generally have a large demonstrable biological effect and are subject to purifying selection (Henn et al. 2016). To summarize, the constrained, human-specific sites are common and include a site previously suggested to affect FOXP2 function.

To assess whether detecting so many extremely high-scoring GERP SNPs in a region of this size was exceptional, we calculated the average GERP score for SNPs located in contiguous, partially overlapping windows with same base pair width as the ROI throughout the genic regions of the autosomes (STAR Methods). After removing singletons to minimize bias resulting from sequencing error, the average GERP score for SNPs in the ROI is 4.677 in the HGDP dataset and 4.828 in 1000G, both highly significant compared to the background distribution averages of −0.070 and −0.474 (p<0.001) (Figure 2A,B). The extreme values for the GERP and PhyloP conservation scores are visualized in Figure 2C. To assess how unusual it is to find a region of this size with multiple tightly constrained SNPs, we conducted a similar scan tallying the counts of high-GERP scoring SNPs (GERP > 3), again requiring allele count to be greater than 1 to reduce inflation by sequencing error. We find that the number of observed constrained SNPs in the ROI is statistically significantly larger than the genic background for windows of the same size. The ROI is thus also an outlier considering the frequency of conserved SNPs in the window (Figure S1H,I). We underscore that the ROI remains a statistical outlier when we partition the data in multiple ways (Figures 2, S1H–I). It is thus unlikely to randomly detect this many polymorphisms (7 in HGDP, 6 in 1000G) clustered together at so highly conserved a locus (Table 2).

Haplotype networks do not bear signatures of a recent selective sweep

To examine the distribution of genetic variation in the *FOXP2* area from a haplotype perspective, we created median-joining haplotype networks for the ROI for the 1000G dataset and Sanger sequencing data, as well as for the entire *FOXP2* region in the HGDP genomes (Figures 3A,B and S3A). In the 1000G network, most haplotypes from Europeans, Asians, and Americans cluster as a closely related set. The European and American haplotypes tend to cluster more closely relative to Asian and African haplotypes. There are also a number of divergent Yoruba, Luhya and African-American haplotypes distinct from the main global cluster. Given that 1000G is impoverished for African diversity, we also examined a global HGDP dataset supplemented with additional Sanger sequencing of southern African KhoeSan. Representatives from all populations are present in the central frequent haplotype, but we also observe numerous divergent haplotypes in the KhoeSan, Mbuti, Mozabites (i.e. Africans) and the Pathan from Pakistan. Generally, a complete selective sweep is expected to result in one primary haplotype and a “star” shape. Our haplotype networks do not bear signatures of a classical sweep in the ancestors of all humans.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The intronic ROI is expressed in human brain cells and resembles an enhancer element

To directly investigate a potential function of the ROI, we examined this region for evidence of expression in relevant cell types, beginning with human brain cells in culture. Using U87-MG immortalized brain cells, which are a human glioma cell line, we isolated total RNA and performed reverse-transcription-PCR (RT-PCR) using primers targeting the ROI. We detected expression of this region using a primer set specific to the ROI (Figure 4A,B). We confirmed the identity of the amplified cDNA as identical the ROI using Sanger sequencing and BLAT analysis (Figure 4C) (Kent, 2002; Kent et al., 2002).

To determine if the ROI was a FOXP2 exon that is spliced to the adjacent exons 8 and 9, we designed primer pairs to detect this splicing event (Exon 8/ROI and ROI/Exon 9). Interestingly, although we could detect spliced FOXP2 using control primers, we did not observe amplification using either exon/ROI primer set (Figure 4D). Thus, it is unlikely that the ROI functions as a novel FOXP2 exon spliced to adjacent exons, at least in this cell type. Notably, amplification of the intronic region flanking the ROI, as well as intronexon junctions, could also be detected by RT-PCR (Figure S4). Thus, it possible that either unspliced FOXP2 is detected in this assay, or the ROI is expressed as a transcript independent of FOXP2, such as an enhancer RNA or other non-coding RNA.

To complement our *in vitro* work on immortalized human brain cells with *in vivo* data from relevant areas of human brains, we performed transcript-agnostic mapping of raw data from an RNA-seq dataset from fresh human dorsolateral prefrontal cerebral cortex tissue (Scheckel et al., 2016) (STAR Methods). We observed that some sequencing reads consistently mapped to the ROI (Figures 4A, S5A), indicating at least a low level of expression of this region in human prefrontal cortex. Importantly, while paired-end reads mapping to coding exons of *FOXP2* had mate pairs in other exons, none of the reads in the ROI had mates in *FOXP2* exons. Thus, consistent with our RT-PCR findings, it is unlikely that the ROI encodes a novel exon of *FOXP2*, though the depth of coverage was too shallow for reliable *de novo* isoform assembly or quantification.

To further investigate if there is any evidence for a role of the ROI as an alternative exon, we conducted two *in silico* tests. First, we computationally predicted the functional effect of SNPs in the ROI to see if any would be predicted to affect splicing (Cingolani et al., 2012). We did not find any SNPs predicted to affecting splicing in the ROI in either the full 1000G or the HGDP genomes datasets. Secondly, we scanned for open reading frames across the ROI to assess whether a potentially functional protein could be generated from this region. We found that translation of the ROI using any reading frame would only produce very short peptides due to the presence of multiple stop codons (Figure S5B). Therefore, both of these computational analyses do not support a hypothesis of the ROI being a novel exon.

We also explored the possibility that the ROI functions as an enhancer regulatory region. Enhancers are DNA elements that function at a distance, and can be located upstream or downstream of genes, or within introns. These regions are bound by specific transcription factors, are marked by covalent histone and DNA modifications, and can be expressed at low levels. We asked if the ROI met these three criteria using publicly available data (Figure S6). First, we found that this region has been previously identified as a chromatin regulatory

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

region in ORegAnno, a literature-curated database identifying potential regulatory regions (Griffith et al., 2007; Montgomery et al., 2006), displays DNAse-accessibility, and can be methylated. Second, we observed that the ROI contains consensus binding sites for a number of transcription factors, including three binding site for POU3F2 (also known as BRN2), which has been previously suggested to regulate *FOXP2* (Maricic et al., 2013). Finally, we interrogated the Sestan Lab Human Brain Atlas, which includes microarray expression data from 13 different brain regions in the human (Kang et al., 2011). Interestingly, we found that expression of the ROI could be detected in the Brain Atlas, with the strongest enrichment in striatum, thalamus and cerebellum. This expression pattern is consistent with previously published analysis of FOXP2 expression in humans (Lai et al., 2003). Altogether, these data are consistent with the ROI historically functioning as a putative enhancer to regulate the expression of FOXP2.

DISCUSSION

No evidence for a selective sweep at FOXP2 in the time frame relevant for language evolution

Our primary goal in this study was to critically assess evidence for a selective sweep targeting *FOXP2* exon 7 in current, larger datasets and to account for the effect of recent human demographic history on patterns of variation in this region. Considering two large next-generation genome datasets, we do not find evidence for a selective sweep in or near *FOXP2*'s exon 7. However, when we curate our datasets such that they consist of individuals with the same ancestry composition as originally tested in Enard et al. (2002) (predominantly individuals whose ancestors underwent the OoA expansion), we replicate a significant negative value for Tajima's D in both the HGDP and 1000 genomes datasets (*FOXP2*'s D is in the lower 5% tail of the empirical genic null distribution of each pooled dataset). Notably, testing for D in the African individuals separately from those who underwent the Out-of-Africa (OoA) expansion erases such a signal (Figure 1A,B, Table 1) when compared to background distributions of D representing the values expected for these datasets as influenced only by population history.

It has been demonstrated that unbalanced sample sizes within structured populations will result in a negative D (Wakeley, 2008). This is due to the inflation of ζ_1 and ζ_{n-1}, the singleton classes in the unfolded site frequency spectrum, in the Tajima's D equation: $D = \frac{\pi - \frac{S}{a_1}}{\sqrt{\widehat{Var}\left[\pi - \frac{S}{a_1}\right]}}$, the numerator of which determines the sign of D where π is the average number of pairwise differences between samples in the dataset, S is the number of segregating sites, $a_1 = \sum_{i=1}^{n-1} \frac{1}{i}$, and n is the sample size. The numerator can be expanded to $\sum_{i=1}^{n-1} \left(\frac{2i(n-i)}{n(n-1)} - \frac{1}{\sum_{j=1}^{n-1} \frac{1}{j}} \right) \xi_i$, where ζ_i the number of polymorphic sites that have i copies of the mutant base. Including mostly individuals whose ancestors underwent the OoA migration – a severe bottleneck (Henn et al., 2012, 2016) – led to individuals who are

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

genetically similar to one another (less pairwise difference between sites, decreased π). After the bottleneck, the populations expanded, and each subsequent new mutation resulted in a new segregating site. We therefore expect to find a negative value for *D* in most genes, simply due to the OoA demographic event. Indeed, the average values for the background distributions in all subsets of the data are below 0 (Figure 1A,B, Table 1). Including a small number of African individuals in the analysis introduces many new segregating sites (further inflating *S*) due to African individuals' high level of genetic diversity not found in OoA groups. However, the addition of African samples does not drastically affect π, as the average pairwise difference is less affected by including only a few such individuals. *D* will thus become more negative due solely to the pooled sample composition and the nature of these populations' demographic histories rather than supporting an argument relevant to natural selection acting on the DNA sequence in this area.

Table 1 compares the *FOXP2 D* values we obtained to their respective genomic background distributions. Firstly, as described above, there is a negative average *D* for all subsets of the data, highlighting the importance of controlling for the genome-wide expectation in each particular dataset rather than simply using the historical *D* threshold of zero. Secondly, accounting for gross population substructure (i.e. if an individual's ancestors underwent the OoA migration and expansion) dramatically changes the empirical expectation for *D*. And thirdly, when African and non-African populations are analyzed separately, the significantly negative *D* signal on the *FOXP2* gene area is lost. Therefore, our re-analysis does not support a recent sweep in *FOXP2* and instead suggest that the prior findings could have been an unintended consequence of the sampling strategy, specifically the lower number of individuals of African descent in the analysis, and the pooling together of individuals with different ancestral demographic histories.

Insufficiently calibrated selection scans can have misleading results

The complications in interpreting window-based screens of natural selection like Tajima's *D* are further highlighted by calculations run on the same genomic region at different scales (Figure 1C–F). The size of the window employed for analysis has a major effect on *D*, and the correct window size is often not obvious. *FOXP2* is a large gene (over 600 kb including the 5' UTR), which makes analysis of the gene as one unit challenging to interpret. Different portions of the gene could be undergoing different selective processes, so treating them as one functional unit would modify the site frequency spectrum in ways that would be impossible to predict without prior knowledge. This raises an additional concern for the hypothesis of a sweep at exon 7, namely the treatment in the earlier analysis of non-contiguous areas of the gene as one unit (combining the 3 introns preceding exon 7). These three introns might be selectively distinct from one another, resulting in unpredictable *D* results when combined, as we note in Table 1. Though it has been argued that there is strong LD across *FOXP2* (Ptak et al., 2009), there are several recombination peaks within the gene in the combined HapMap recombination map and African American recombination maps (Hinch et al., 2011; The International HapMap Consortium, 2005) (Figure S3B). Our work serves as a cautionary tale for careful data treatment in the context selection summary statistics.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Haplotype networks, too, are not reflective of a recent selective sweep at the locus

We additionally used other approaches to comprehensively check for putative selection at the *FOXP2* locus. First, we visualized the haplotypic diversity across *FOXP2* with haplotype networks from newly generated Sanger sequencing data (consisting of 40 global human HGDP individuals and 7 additional Khomani San individuals) and non-admixed 1000G African and OoA populations (Figures 3, S3A, STAR Methods). These networks indicate several major haplogroups for *FOXP2* segregating across populations, including at least one exclusively maintained within Africa. Generally, a complete selective sweep is expected to result in one primary haplotype representing linkage between the selected allele and hitchhiking mutations nearby, resulting in a star-like shape after the introduction of new mutations. Our haplotype networks therefore do not reflect a classical sweep in the ancestors of all humans. Instead, the presence of one primary node for OoA individuals likely reflects the major OoA event which carried a subset of the genetic variation within Africa to the rest of the world. A haplotype network for the entire FOXP2 gene region in the HGDP genomes (Figure S3A) possessed long internal branches, which are sometimes interpreted as indications of balancing selection, however explicitly testing for balancing selection using ancestral recombination graphs did not provide evidence for this type of selection, either (Figure S2).

An intronic area in FOXP2 is enriched for common, tightly-constrained, human-specific SNPs

To investigate *FOXP2* from another angle, we next looked at a site-based metric of DNA sequence evolution – the GERP score (Cooper, 2005). GERP scores estimate site-wise DNA sequence conservation across multiple species alignments and reflect the level of evolutionary constraint on a DNA element as determined by the number of rejected substitutions; that is, mutations removed by purifying selection. Regions with high GERP scores, i.e. conserved, are therefore interpreted to be functionally important. We found that a region in intron 9 does harbor a significant number of extreme GERP SNPs compared to the remainder of the genome (Figure 2). The high-GERP SNPs seen in the ROI are at common frequencies in at least one human population surveyed in our datasets and, in all cases but one, are uniquely derived in humans as compared to other primates and archaic hominins (Table 2). No SNPs are present across the entire ROI in the 10 PanMap chimps and only 1 heterozygous site is present in three other archaic hominin genomes (the Alai and Vindija Cave Neanderthals and the Denisovan) (Table S1). Sanger sequencing the ROI further elucidated 3 additional SNPs that had not previously been called in NGS data for the 1000G or HGDP datasets. These 3 sites (rs577428580, rs563023653, rs7799652) are at common frequency in KhoeSan individuals and are seen in Sanger data in some of the same individuals present in the NGS datasets. This highlights the need for careful evaluation of low-frequency variation in NGS data given the tendency of pipelines to generate a high false-negative rate in order to minimize false positives (Bobo et al., 2016). To summarize, we see an unexpectedly large number of common, derived SNPs in the human lineage as compared to chimps and the Neanderthal in the ROI. High evolutionary constraint amongst taxa but variability within *Homo sapiens* is compatible with a modified functional role for this locus in humans, such as a recent loss of function.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The intronic ROI is expressed in human brain cells and resembles an enhancer element

Finally, we followed up on the *FOXP2* ROI using *in vitro* and *in vivo* experiments to investigate its potential current or past functional role in the human brain. Using RNA-seq and RT-PCR, we were able to detect RNA transcribed from the ROI in human prefrontal cortex and in human brain cells in culture (Figures 4, S4, S5A). However, we did not find evidence that the ROI encodes a novel exon of *FOXP2* (Figure S5B). Our assays cannot rule out that detection of amplicons in RT-PCR experiments and reads in RNA-seq data result from an unspliced FOXP2 transcript. It is also possible that the element is expressed at higher levels at specific developmental stages.

The potential former function of the ROI remains unclear, but several possibilities exist. One possibility is that the ROI codes for a long non-coding RNA (lncRNA). lncRNAs have been implicated in a variety of biological processes, including developmental patterning, behavior, and synaptic plasticity. Mechanistically, lncRNAs have been shown to impact chromatin states through cis- and trans-acting mechanisms, thereby functioning to fine-tune gene expression or modulate transcription factor activity (Kung et al., 2013). Interestingly, it has been estimated that at least 50% of lncRNAs are expressed in the nervous system, and many exhibit cell-type specific patterns of neuronal expression (Quan et al., 2017; Qureshi and Mehler, 2013).

It is also possible that the ROI is an enhancer region that is transcribed at low levels, termed an Enhancer RNA (eRNA). eRNAs were initially discovered in neurons and are transcribed by RNA polymerase II (Kim et al., 2010, 2015). Functionally, these regions promote activation of their target genes. Consistent with this possibility, the ROI houses many transcription factor binding sites and hallmarks of regulatory regions (Figure S6), which may suggest a regulatory function. A particularly noteworthy binding site in the ROI is for POU3F2 (at rs114972925), which has been suggested to regulate the expression of FOXP2 (Maricic et al., 2013), where the derived allele in humans has reduced binding affinity compared to the ancestral allele. Notably, since enhancers can act at a distance, we cannot rule out that this putative enhancer region is controlling expression of a different gene. However, given that the nearest gene (*MDFIC*) is over 250kb away and there is no evidence supporting this mode of regulation, this possibility is unlikely. Here, we find that rs114972925 is polymorphic in southern African human populations. Therefore, this SNP cannot be not necessary for language function as both alleles persist at high frequency in modern human populations. Though perhaps obvious, it is important to note that there is no evidence of differences in language ability across human populations.

Conclusions

In conclusion, we do not find evidence that the *FOXP2* locus, or any previously implicated site within *FOXP2*, is associated with recent positive selection in humans. Specifically, we demonstrate that there is no evidence that the original two amino-acid substitutions were targeted by a recent sweep limited to modern humans <200 *kya* as suggested by Enard et al. (2002). It is possible that these two substitutions were the targets of an ancient selective sweep, though an examination of ancient selection on this region did not reach significance here. We also do not find consistent evidence to argue that the intronic SNP rs114972925,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

previously discussed in Maricic et al. (2013), is responsible for the selective sweep, as we do not see evidence for recent selection targeting this region, either. Rather, we find that the ancestral allele persists at high frequency in some modern African populations. This intronic ROI containing rs114972925 is unusual in having so many tightly constrained sites that are variable within humans compared to other species, which is suggestive of a loss of function. Any modified function of the ROI does not appear to be related to language, however, as modern southern African populations tolerate high minor allele frequencies with no apparent consequences to language faculty. We do not dispute the extensive functional evidence supporting *FOXP2*'s important role in the neurological processes related to language production (Lai et al., 2001; MacDermot et al., 2005; Torres-Ruiz et al., 2016). However, we show here that recent natural selection in the ancestral *Homo sapiens* population cannot be attributed to the *FOXP2* locus and thus *Homo sapiens'* development of spoken language. We hypothesize that the ROI contains an enhancer element, which while strongly conserved in other species, has recently experienced a loss of function in humans, consistent with low expression of the ROI in brain-related tissues.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Elizabeth Atkinson (eatkinso@broadinstitute.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Datasets—The human sequences used for population genetic analyses are 53 adult human genomes from global human populations sequenced as part of the Human Genome Diversity Panel (HGDP) (Henn et al., 2016) and a subset of representative individuals from the 1000 Genomes Project phase 1 (The 1000 Genomes Project Consortium, 2012). The populations represented by HGDP include two from Africa (KhoeSan, Mbuti) and four populations whose ancestors underwent the Out-of-Africa (OoA) expansion (Maya, Yakut, Cambodian, Mozabite). The subset of individuals from the 1000 Genomes Project used in comparisons includes individuals from the following populations: GBR, JPT, CHB, TSI, PEL, MXL (OoA) and the YRI and LWK (African). Subject details, including age and sex where available, for all HGDP individuals can be found at http://www.hagsc.org/hgdp.html and 1000G individuals can be found at http://www.internationalgenome.org/. We did not treat males and females or different aged individuals differently across computational analyses in this article as we are only interested in autosomal DNA sequence variation unrelated to individual phenotypes.

The chimpanzees used as an outgroup were a population of 10 individuals from the PanMap dataset (Auton et al., 2012). Three ancient hominin genomes were included for comparison with modern genomes: the Altai Neanderthal (Prüfer et al., 2013), which is a composite genome of several individuals with an average of 52x coverage, the Vindija Cave Neanderthal (Prüfer et al., 2017) and the Denisovan (Meyer et al., 2012).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cell cultue for RT-PCR

U87-MG immortalized cells were purchased from ATCC and cultured in DMEM with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin/glutamine. U87-MG cells are derived from brain tissue of an adult male.

Sanger sequencing of HGDP-CEPH individuals

Sanger sequencing of the ROI was conducted on DNA from 40 HGDP-CEPH immortalized lymphoblastoid cell lines derived from global adult human samples (Cann et al., 2002), including both African and non-African individuals and all those in whom we had detected the alternate allele in the ROI. This was done in order to verify the presence of the SNPs we observed using next-generation sequencing data. We did not treat males and females differently, as sex should not affect the sequence of this autosomal region. Extracted DNA from the HGDP-CEPH cell lines was purchased from Foundation Jean Dausset-CEPH (Paris, France). DNA samples are distributed dissolved in TE (10:1) at a concentration of ~60ng/μL, with ~5μg of DNA per well. The specific cell lines/individuals that we Sanger sequenced are as follows:

Individual ID	Population	Sex
HGDP00712	Cambodian	F
HGDP00713	Cambodian	F
HGDP00716	Cambodian	M
HGDP00719	Cambodian	F
HGDP00720	Cambodian	M
HGDP00721	Cambodian	F
HGDP00987	HGDP San	M
HGDP00991	HGDP San	M
HGDP00992	HGDP San	M
HGDP01029	HGDP San	M
HGDP01032	HGDP San	M
HGDP01036	HGDP San	M
SA007	Khomani San	F
SA017	Khomani San	F
SA036	Khomani San	F
SA043	Khomani San	F
SA053	Khomani San	M
SA055	Khomani San	M
SA057	Khomani San	M
HGDP00449	Mbuti	M
HGDP00456	Mbuti	M
HGDP00462	Mbuti	M
HGDP00471	Mbuti	F

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Individual ID	Population	Sex
HGDP00474	Mbuti	M
HGDP00476	Mbuti	F
HGDP01081	Mbuti	M
HGDP01258	Mozabite	M
HGDP01259	Mozabite	M
HGDP01262	Mozabite	M
HGDP01264	Mozabite	M
HGDP01267	Mozabite	F
HGDP01274	Mozabite	F
HGDP01275	Mozabite	F
HGDP01277	Mozabite	F
HGDP00213	Pathan	M
HGDP00222	Pathan	M
HGDP00232	Pathan	F
HGDP00237	Pathan	F
HGDP00239	Pathan	F
HGDP00243	Pathan	M
HGDP00247	Pathan	F
HGDP00258	Pathan	M
HGDP00950	Yakut	M
HGDP00955	Yakut	F
HGDP00960	Yakut	M
HGDP00963	Yakut	F
HGDP00964	Yakut	M

METHOD DETAILS

Reverse Transcription and PCR

Total RNA was isolated using the RNeasy RNA Purification (QIAGEN) with an on-column DNase digestion (QIAGEN). Total RNA (1 ug) was reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. After reverse transcription, cDNA was diluted 1:5 dH2O and 1.9ul was used for PCR on the BioRad C1000 Touch Thermal Cycler using PowerUp SYBR Green Master Mix (BioRad) according to the manufacturer's protocol. Primers were designed using Primer3 software (Koressaar and Remm, 2007). ROI primers were designed in the intronic region of chr7: 114,288,164–114,290,415 (hg19 coordinates). To determine the length of the expressed product in the ROI, the ROI was separated into approximate 500 bp sections. Primers were designed within those sections, RT-PCR was performed, and then analyzed on an agarose gel. Other FOXP2 isoform primers and primers for the exons surrounding the ROI were designed to span one intron. We highlight that the RT reactions were primed with random primers and that we used total RNA, eliminating any concern that

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

we would only be isolating polyA RNA and thus missing enhancer RNAs or long non-coding RNAs that are not polyadenylated in these tests.

Primer Sets Used for RT-PCR (all primers listed 5’ to 3’):

FOXP2-Ctl-F	CATCATTCCATAGTGAATGGACAG
FOXP2-Ctl-R	CCATGGCCATAGAGAGTGTG
FOXP2-ROI1-F	AGGCAGCATGGCTACAAAAT
FOXP2-ROI1-R	GATGCCTTAGTTTTCAGAGATGG
FOXP2-ROI2-F	CTTTCTGAAGGCACCCTTTG
FOXP2-ROI2-R	AGAGCAAAGCAAGCTCTGGA
FOXP2-ROI3-F	CCACTTGGTCCTTTTGAAGC
FOXP2-ROI3-R	ATGTCCCTTGCAGGAAGTTG
FOXP2-ROI4-F	AAGCAGTAAACAAGTGTAGAAAATCA
FOXP2-ROI4-R	AGGGTGTAAATGCAGGAAGC
FOXP2-Int1-F	TTTTTCACATTACATCTCAAACAAAC
FOXP2-Int1-R	ATGTGTTATACATAAGCAACTGTCCT
FOXP2-Int2-F	TTGAGCCCACTTGGGTAAAT
FOXP2-Int2-R	ATGGGCAAAGTAAGGCAACA
FOXP2-Exon8-Intron F	GGCTGTGAAAGCATTTGTGA
FOXP2-Exon8-Intron R	AATGTGCCTAAAATGCCCATA
FOXP2-Exon9-Intron F	GCCTATGCCACTAAGATCGAC
FOXP2-Exon9-Intron R	ATTTGCACTCGACACTGAGC

Purification and Sequencing of cDNA

Amplified cDNA was run on a 2% agarose and the 130 bp product corresponding to the ROI was isolated and then purified with the QIAquick Gel Extraction Kit (QIAGEN). The purified product was then prepped for sequencing according to Genewiz specifications. Sequencing results were confirmed by comparison to GRCh37 with BLAT (Kent et al., 2002, 2002).

Sanger sequencing of HGDP-CEPH individuals

Seven overlapping forward (F) and reverse (R) sequencing primers were designed to fully capture the ROI in the 40 HGDP-CEPH and ≠Khomani San individuals with at least a several hundred bp buffer on either end. Raw output files were processed and manually inspected in Sequencher (Sequencher version 5.4.6 DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI. www.genecodes.com). All polymorphic sites observed in the HGDP NGS genomes were manually validated across the individuals passing Sanger QC. The allele frequencies reported in Table 2 reflect the corrected frequencies of all SNPs post-Sanger validation of these sites. Primer sequences are as follows:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Primer Name	Probe Sequence
PCR F primer	AGAAAGGCAGGATGGCTAGT
PCR R primer	TGTGAACCTGTGAGGAGGAT
Seq Primer 1 F	TGAAGAGCAATAGAAAACAGTGGA
Seq Primer 2 R	GCTGCTTCAAAAGGACCAAG
Seq Primer 3 F	CCAGAGCTTGCTTTGCTCTT
Seq Primer 4 R	TGAACTCTTGGGGAACACAA
Seq Primer 5 R	AATGATGTGGTCGACTGACG
Seq Primer 6 R	CCAGAGCTGCCACTTTCTTT
Seq Primer 7 F	AGGGAGATGTGGCAGAGGTA

QUANTIFICATION AND STATISTICAL ANALYSIS

Tajima's D analyses

As every dataset will have variation in the range of Tajima's *D* values that can be expected based on the component individuals' population history, we designed a test to control for background *D* values in each dataset/subpopulation separately and provide a means to visualize an exceptional value. To build a distribution of expectations throughout the autosomes in each individual dataset, we calculated *D* for contiguous non-overlapping windows of the autosomes of the same bp length as the *FOXP2* gene (773,205 bp). These estimates provide the background range of *D* values in a given dataset, and thus represent the neutral expectations for *D* and a dataset/population-specific way to assess deviations from "normal" values. To ensure accurate calculation of *D*, we excluded windows that contained fewer than 5 SNPs, as these could produce unreliable estimates. For population subset-specific calculations (i.e., African, OoA), new background distributions were created for the relevant subset of individuals for direct comparison of *D* in *FOXP2* to those individuals' expected background values. This provides an indirect means to control for the population history of a particular group of interest. All calculations were conducted using VCFtools (Danecek et al., 2011). A significantly exceptional *D* value for *FOXP2* is represented by the value lying in a 5% tail of the empirical distribution of genome-wide Tajima's *D* values for each dataset.

We additionally carried out a "gene crawl" analysis across the length of *FOXP2* to visualize variation in the value of *D* across the gene as a factor of window size. For this, we calculated *D* in windows of 1500 bp and 6kb in both the HGDP genomes and 1000G datasets, again filtering for only regions that contained at least 5 SNPs so as not to bias results. In each dataset, we again calculated *D* in the full population, and subsequently in African and OoA individuals seperately.

Tajima's D value stratified by exon content and gene-level constraint

There is evidence that different genomic elements will produce different values of *D* (Rech et al., 2014). In order to compare *D* in *FOXP2* against genic regions under similar levels of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

selective constraint, we conducted two sets of null tests. First, we ran an additional batch of population-stratified selection tests in both HGDP and 1000G to create a background distribution of *D* values based on genes with similar percent exon sequence to *FOXP2*. Our goal with this test was to assess whether *FOXP2* shows an exceptional signal of selection when compared to other genes that are similar in exonic content (and thereby potentially subject to similar selection pressures), while controlling for population structure as in our previous analyses. The *FOXP2* canonical transcript is 2.311% exonic. We thus considered comparable genes to be those in which exons represented 1–4% of the total sequence length in the canonical transcript. Of all UCSC knownGene canonical transcripts, 3166 other genes exist in this comparable class, and we refer to these genes as "exon content-matched genes". We do not find evidence for selection at *FOXP2* in any data subset when comparing to exon content-matched genes (Figure S1A–F).

We additionally checked to see if *FOXP2* experienced an aberrant amount of gene-level constraint compared to other autosomal genes, which could also skew *D* values. To check for the direct possibility of gene-level constraint affecting results, we compared the average gene-level GERP value of *FOXP2* to the average genic GERP scores for all autosomal genes. We obtained gene-level scores by averaging the site-wise GERP value of all variants present within the canonical gene boundaries in the HGDP dataset. *FOXP2* does not appear exceptional compared to the background distribution of gene-level GERP scores for autosomal genes (considering only genes that contain at least 5 SNPs) (Figure S1G). Therefore, we do not believe that overall gene-level constraint is exerting a biasing influence over our estimate of Tajima's *D* in *FOXP2*.

Fay and Wu's H

Tests for Fay and Wu's *H* were conducted using the DnaSP software program (Librado and Rozas, 2009). *H* was calculated on the entire *FOXP2* gene region in the HGDP genomes (containing 2512 SNPs) using comparable population subsets to tests of *D*. Specifically, *H* was calculated for the entire pooled HGDP population, African and OoA subsets. When calculating *H* on the pooled sample, a strongly negative value was obtained ($H = -2.993$), traditionally interpreted as evidence for a a selective sweep. When calculated on only the OoA individuals, *H* remained negative but still non-significant ($H = -0.571$). When calculated on the African individuals only, *H* was positive ($H = 5.983$). We used the 1000 genomes human ancestor sequence generated by 6 primate species EPO alignments to orient HGDP variants to derived versus ancestral state.

Selection at Different Time Scales

The ability for statistical methods to detect the signatures of natural selection in modern human genomes depends on several factors, including the depth in time of the selective event, the strength of the event, and the starting allele frequency of the variant(s) in question. Tajima's *D* is most powerful for strong, complete or near-complete sweeps, with power to detect selection decreasing with increased elapsed since the sweep. Under very strong selection (s=0.1), *D* has been demonstrated to perform best at a time-frame of ~700–2000 generations ago (Ronen et al., 2013) and for weaker selective scenarios (s=0.01) to be best powered at ~1700 generations ago (~50 *kya*). Fay and Wu's *H* can detect events slightly

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

deeper back in time; it performs well under medium strength selective regimes at ~2000 generations ago, ~60 *kya* in humans, as it is not affected by the inflation of low-frequency alleles that accumulate after fixation. The power of both statistics tapers off by ~3000 generations, ~90 *kya*. As can be noted from these time frames, both of these analyses are equipped to detect recent selection, meaning a selective sweep occurring after the *Homo sapiens* populations diverge. We employed these statistics to replicate the original *FOXP2* work and assess whether there is evidence of recent selection at *FOXP2* in the modern human lineage, which is how the canonical hypothesis for *FOXP2* stood.

The McDonald-Kreitman test is designed to detect selection between species, and therefore can be used to examine whether there is evidence instead for ancient selection (>200 *kya*) on *FOXP2*. The MK test compares the number of variable sites that are polymorphic within a species vs. fixed between species for the classes of nonsynonymous vs. synonymous in order to estimate the proportion of substitutions that experienced positive selection between species. We ran the McDonald-Kreitman test on the *FOXP2* coding sequence comparing the HGDP genomes to a population of 10 chimpanzees (Auton et al., 2012). We found 11 fixed synonymous substitutions, 16 polymorphic synonymous sites, 2 fixed non-synonymous substitutions, and 6 polymorphic non-synonymous sites. This results in a non-significant (p=0.68) McDonald-Kreitman value of −1.0625. A negative test result is often interpreted to suggest positive selection, though it did not reach significance in this case.

Haplotype Networks

For genomic data, the program Shapeit2 (O'Connell et al., 2014) was used to phase chromosome 7 of the HGDP and 1000G datasets, informed by the HapMap combined b37 recombination map. Phased files were trimmed to include only the *FOXP2* region of interest and converted to FASTA format for manual inspection in the program DNA Alignment (Fluxus technologies). We generated median-joining haplotype networks (which allows for multi-allelic sites) for the HGDP and 1000G datasets and for varying segments of the gene, retaining branch lengths proportionate to mutational differences and node size proportionate to the number of haplotype samples, in the program Network 4.6.1.1 (Fluxus Technologies).

For Sanger sequencing data, we aligned individual fasta files to each other with Clustal Omega (Li et al., 2015) and phased them using the PHASE program (Stephens et al., 2001) with formatting scripts implemented with seqPhase (Flot, 2010). Haplotype visualization then followed the same pipeline as for genomic data.

GERP score analysis

GERP scores provide a measure of functional constraint based on DNA sequence conservation across a multi-species alignment. Higher GERP scores imply greater DNA sequence conservation. The average GERP score for the human autosomes is −0.04 (Davydov et al., 2010), and values greater than 3 are generally considered to be evolutionarily exceptional. Constraint intensity at each individual alignment position in our datasets was quantified and annotated in the VCF file using SNPeff (Cingolani et al., 2012) in terms of a "rejected substitutions" (RS) score, defined as the number of substitutions expected under neutrality minus the number of substitutions observed at the position. These

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

scores are equivalent in their interpretation to GERP scores. We used a standard threshold of RS>3 as indicative of a significantly high GERP score; in other words, a strongly conserved stretch of DNA sequence.

Examining *FOXP2* with GERP scores in mind, we discovered a 2251 bp area in intron 9 that contains a cluster of extremely highly conserved SNPs (6 SNPs with GERP > 3 in the 1000G and 7 SNPs in HGDP). We next assessed both whether it is unusual for an area of this size to have SNPs with so high an average GERP score and if it is unusual to find an area of this size containing this many high-GERP scoring SNPs. To do this, we tabulated the total number of SNPs, the number of high-GERP SNPs (RS > 3), and the average GERP score of SNPs in windows of the same size as the ROI (2251 bp), sliding 100bp along the genic-masked autosomes for each new window start site. Since functional constraint is expected to differ between genic and intergenic regions, we placed a "genic" mask on the datasets using the hg19 start and stop positions of annotated genes. We retained introns, as our region of interest in *FOXP2* is itself in an intron. We excluded SNPs that were defined by singletons to remove the potential influence of sequencing error. The results of these analyses, similar to the Tajima's *D* tests, provide a means to visualize the expected distribution of GERP scores across the genic regions in each individual dataset as compared to the values obtained from the ROI. The high-GERP SNP count and average GERP score in each relevant region can be compared to the background distribution and are deemed significant if they reside in a 5% tail (Figures 2, S4).

Balancing Selection Testing

The Sequential Markov Coalescent (SMC) model is an approximation of the ancestral recombination graph that traces the ancestry of samples along the genome. SMC models recombination explicitly by allowing genealogies to change along the length of the sequences. ARGweaver (Rasmussen et al., 2014) implements the SMC model to sample a sequence of local genealogies along a chromosomal segment from the posterior distribution. To test the TMRCA obtained from the *FOXP2* region to the background values, we ran ARGweaver on *FOXP2* and two comparison 5Mb chromosomal areas in chromosomes 6 and 7 using a set of genomes from seven African individuals (NA21302, NA19240, HG03428, HG02799, HG03108, HGDP01029 and HGDP00456). These samples were fosmid pool sequenced and phase resolved (n=14 haplotypes; Song et al., 2016). Fosmid-phasing obtains phase directly from DNA sequences without the need for extensive statistical phasing, which can have high error rates that affect the calculation of the TMRCA in coalescent simulations.

For our analysis, we assumed a recombination rate of 1.125×10^{-8}, a mutation rate of 1.5×10^{-8} bp per generation and effective population size of 15,000. We ran ARGweaver for 1000 iterations and thinned every 100 for a total of 100 samples. An ARGweaver sample of 2 Mb included 1956 local genealogies on average. To generate Figure S2, we estimated the posterior mean by the average TMRCA across the 100 ARGweaver posterior samples and the 95% Bayesian credible intervals (BCI) at every 200 bp. The posterior mean TMRCA for these samples across the chr7 5 Mb background region is ~46,400 *ya* with a 95% credible interval of (15,400; 124,400) and the chr6 posterior mean is ~66,000 *ya* with a 95% credible

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

interval of (27,400; 156,500). The posterior mean TMRCA across *FOXP2* is ~38,200 years ago with a 95% credible interval of (7,900; 97,400).

Human Brain RNA-seq

RNA-seq analysis was conducted on the 8 control individuals from the publicly-available GEO dataset PRJNA232669 (Scheckel et al., 2016). We processed raw paired FASTQ files with FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) using standard trimming and quality filters (minimum length=20, minimum quality score=20, minimum percent bases meeting quality score=50). We mapped reads to the human reference genome build GRCh37 using HISAT2 (Kim et al., 2015) with all parameters left at their default as of version 2.1.0. Visualization of low-depth read pile-up in the ROI was made using the Integrated Genomics Viewer (Robinson et al., 2011) for a representative example individual in Figures 4A and for all 8 individuals in Figure S5A.

Open Reading Frame Analysis

The ExPASy Translate tool was utilized to translate the ROI sequence to possible protein products, checking all three possible reading frames (Gasteiger et al., 2003). We included the *FOXP2* intron 9 ROI with a small buffer for open reading frame analysis, using hg19 coordinates chr7:114288114-114290465.

DATA AND SOFTWARE AVAILABILITY

Sanger individual FASTA files and primer sequences for the *FOXP2* ROI can be downloaded from Genbank as accession numbers MH234124 - MH234217. FOXP2 ROI Genotype calls from our Sanger sequencing are freely available for download at https://github.com/eatkinson/FOXP2_SangerROI.vcf. Other related scripts can be found on GitHub at https://github.com/eatkinson or acquired by emailing the Lead Contact, Elizabeth Atkinson (eatkinso@broadinstitute.org).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Krishna Veeramah, Gregory T. Smith, and the many scientists whose constructive discussions and suggestions at conferences shaped the development of this project. We thank Jeffrey Kidd for providing the fosmid-phased genomes. This work was supported by the National Institutes of Health (IRACDA postdoctoral grant K12 GM102778 to E.G.A., Interdisciplinary Predoctoral Neuroscience Training Grant 2T32MH020068 to A.A., COBRE award P20GM109035 to S.R., R01 GM118652 to B.M.H. and S.R.), the National Science Foundation (CAREER Award DBI-1452622 to S.R.), and a Terman Fellowship to J.A.P.

References

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Segurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, et al. A Fine-Scale Chimpanzee Genetic Map from Population Sequencing. Science. 2012; 336:193–198. [PubMed: 22422862]

Bobo D, Lipatov M, Rodriguez-Flores JL, Auton A, Henn BM. False Negatives Are a Significant Feature of Next Generation Sequencing Callsets. 2016 BioRxiv 066043.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, et al. A human genome diversity cell line panel. Science. 2002; 296:261–262. [PubMed: 11954565]

Chabout J, Sarkar A, Patel SR, Radden T, Dunson DB, Fisher SE, Jarvis ED. A Foxp2 Mutation Implicated in Human Speech Deficits Alters Sequencing of Ultrasonic Vocalizations in Adult Male Mice. Front Behav Neurosci. 2016; 10

Chiu YC, Li MY, Liu YH, Ding JY, Yu JY, Wang TW. Foxp2 regulates neuronal differentiation and neuronal subtype specification. Dev Neurobiol. 2014; 74:723–738. [PubMed: 24453072]

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Ruden DM, Lu X. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w11183; iso-2; iso-3. Fly (Austin). 2012; 6:80–91. [PubMed: 22728672]

Coop G, Bullaughey K, Luca F, Przeworski M. The Timing of Selection at the Human FOXP2 Gene. Mol Biol Evol. 2008; 25:1257–1259. [PubMed: 18413354]

Cooper GM. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005; 15:901–913. [PubMed: 15965027]

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011; 27:2156–2158. [PubMed: 21653522]

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++ PLoS Comput Biol. 2010; 6

Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, Kitano T, Pa APM, Pääbo S. Molecular evolution of FOXP2, a gene involved in speech and language. Nature. 2002

Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T, Somel M, Brückner MK, Schreiweis C, Winter C, Sohr R, et al. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell. 2009; 137:961–971. [PubMed: 19490899]

Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. Localisation of a gene implicated in a severe speech and language disorder. Nat Genet. 1998; 18:168–170. [PubMed: 9462748]

Flot JF. seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. Mol Ecol Resour. 2010; 10:162–166. [PubMed: 21565002]

Fujita-Jimbo E, Momoi T. Specific expression of FOXP2 in cerebellum improves ultrasonic vocalization in heterozygous but not in homozygous Foxp2 (R552H) knock-in pups. Neurosci Lett. 2014; 566:162–166. [PubMed: 24607928]

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003; 31:3784–3788. [PubMed: 12824418]

Griffith OL, Montgomery SB, Bernier B, Chu B, Kasaian K, Aerts S, Mahony S, Sleumer MC, Bilenky M, Haeussler M, et al. ORegAnno: an open-access community-driven resource for regulatory annotation. Nucleic Acids Res. 2007; 36:D107–D113. [PubMed: 18006570]

Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C. FoxP2 Expression in Avian Vocal Learners and Non-Learners. J Neurosci. 2004; 24:3164–3175. [PubMed: 15056696]

Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodriguez-Botigue L, Ramachandran S, Hon L, Brisbin A, et al. Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci. 2011; 108:5154–5162. [PubMed: 21383195]

Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. Proc Natl Acad Sci. 2012; 109:17758–17764. [PubMed: 23077256]

Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, Martin AR, Musharoff S, Cann H, Snyder MP, et al. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc Natl Acad Sci. 2016; 113:E440–E449. [PubMed: 26712023]

Hinch AG, Tandon A, Patterson N, Song Y, Rohland N, Palmer CD, Chen GK, Wang K, Buxbaum SG, Akylbekova EL, et al. The landscape of recombination in African Americans. Nature. 2011; 476:170–175. [PubMed: 21775986]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AMM, Pletikos M, Meyer KA, Sedmak G, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011; 478:483–489. [PubMed: 22031440]

Kaplan NL, Darden T, Hudson RR. The coalescent process in models with selection. Genetics. 1988; 120:819–829. [PubMed: 3066685]

Kent WJ. BLAT—The BLAST-Like Alignment Tool. Genome Res. 2002; 12:656–664. [PubMed: 11932250]

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The Human Genome Browser at UCSC. Genome Res. 2002; 12:996–1006. [PubMed: 12045153]

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. Widespread transcription at neuronal activity-regulated enhancers. Nature. 2010; 465:182–187. [PubMed: 20393465]

Kim TK, Hemberg M, Gray JM. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb Perspect Biol. 2015; 7:a018622. [PubMed: 25561718]

Klein RG. Archeology and the evolution of human behavior. Evol Anthropol Issues News Rev. 2000; 9:17–36.

Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, Geschwind DH. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature. 2009; 462:213–217. [PubMed: 19907493]

Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics. 2007; 23:1289–1291. [PubMed: 17379693]

Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, Hublin JJ, Hänni C, Fortea J, de Rasilla Mla, et al. The Derived FOXP2 Variant of Modern Humans Was Shared with Neandertals. Curr Biol. 2007; 17:1908–1912. [PubMed: 17949978]

Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, Fu Q, Burbano HA, Lalueza-Fox C, de la Rasilla M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. Nature. 2016; 530:429–433. [PubMed: 26886800]

Kung JTY, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. Genetics. 2013; 193:651–669. [PubMed: 23463798]

Lai CS, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Jamison DC, Green ED, et al. The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am J Hum Genet. 2000; 67:357–368. [PubMed: 10880297]

Lai CS, Fisher SE, Hurst Ja, Vargha-Khadem F, Monaco aP. A forkhead- domain gene is mutated in a severe speech and language disorder. Nature. 2001; 413:519–523. [PubMed: 11586359]

Lai CSL, Gerrelli D, Monaco AP, Fisher SE, Copp AJ. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain J Neurol. 2003; 126:2455–2462.

Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. 2015; 43:W580–4. [PubMed: 25845596]

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–1452. [PubMed: 19346325]

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CSL, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, et al. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet. 2005; 76:1074–1080. [PubMed: 15877281]

Maricic T, Günther V, Georgiev O, Gehre S, Curlin M, Schreiweis C, Naumann R, Burbano HA, Meyer M, Lalueza-Fox C, et al. A recent evolutionary change affects a regulatory element in the human FOXP2 gene. Mol Biol Evol. 2013; 30:844–852. [PubMed: 23197593]

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, et al. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science. 2012; 338:222–226. [PubMed: 22936568]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Montgomery SB, Griffith OL, Sleumer MC, Bergman CM, Bilenky M, Pleasance ED, Prychyna Y, Zhang X, Jones SJM. ORegAnno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation. Bioinformatics. 2006; 22:637–640. [PubMed: 16397004]

O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I, et al. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet. 2014; 10:e1004234. [PubMed: 24743097]

Pinel P, Fauchereau F, Moreno A, Barbot A, Lathrop M, Zelenika D, Le Bihan D, Poline JB, Bourgeron T, Dehaene S. Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions. J Neurosci Off J Soc Neurosci. 2012; 32:817–825.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014; 505:43–49. [PubMed: 24352235]

Prüfer K, de Filippo C, Grote S, Mafessoni F, Korlević P, Hajdinjak M, Vernot B, Skov L, Hsieh P, Peyrégne S, et al. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science. 2017; 358:655–658. [PubMed: 28982794]

Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. Evol Int J Org Evol. 2005; 59:2312–2323.

Ptak SE, Enard W, Wiebe V, Hellmann I, Krause J, Lachmann M, Pääbo S. Linkage disequilibrium extends across putative selected sites in FOXP2. Mol Biol Evol. 2009; 26:2181–2184. [PubMed: 19608635]

Quan Z, Zheng D, Qing H. Regulatory Roles of Long Non-Coding RNAs in the Central Nervous System and Associated Neurodegenerative Diseases. Regulatory Roles of Long Non-Coding RNAs in the Central Nervous System and Associated Neurodegenerative Diseases. Front Cell Neurosci Front Cell Neurosci. 2017; 11:175–175. [PubMed: 28713244]

Qureshi IA, Mehler MF. Long non-coding RNAs: novel targets for nervous system disease diagnosis and therapy. Neurother J Am Soc Exp Neurother. 2013; 10:632–646.

Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015; 16:359–371. [PubMed: 25963373]

Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. Genome-wide inference of ancestral recombination graphs. PLoS Genet. 2014; 10:e1004342. [PubMed: 24831947]

Rech GE, Sanz-Martín JM, Anisimova M, Sukno SA, Thon MR. Natural Selection on Coding and Noncoding DNA Sequences Is Associated with Virulence Genes in a Plant Pathogenic Fungus. Genome Biol Evol. 2014; 6:2368–2379. [PubMed: 25193312]

Reimers-Kipping S, Hevers W, Pääbo S, Enard W. Humanized Foxp2 specifically affects cortico-basal ganglia circuits. Neuroscience. 2011; 175:75–84. [PubMed: 21111790]

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011; 29:24–26. [PubMed: 21221095]

Ronen R, Udpa N, Halperin E, Bafna V. Learning Natural Selection from the Site Frequency Spectrum. Genetics. 2013; 195:181–193. [PubMed: 23770700]

Scheckel C, Drapeau E, Frias MA, Park CY, Fak J, Zucker-Scharff I, Kou Y, Haroutunian V, Ma'ayan A, Buxbaum JD, et al. Regulatory consequences of neuronal ELAV-like protein binding to coding and non-coding RNAs in human brain. ELife. 2016; 5

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem. 2001; 276:27488–27497. [PubMed: 11358962]

Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, Gasperi RD, Sosa MAG, Rabidou D, et al. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A. 2005; 102:9643–9648. [PubMed: 15983371]

Song S, Sliwerska E, Emery S, Kidd JM. Modeling Human Population Separation History Using Physically Phased Genomes. Genetics genetics. 2016; 116:192963.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Španiel F, Horáček J, Tintěra J, Ibrahim I, Novák T, Čermák J, Klírová M, Höschl C. Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. Neurosci Lett. 2011; 493:131–135. [PubMed: 21334420]

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH. Identification of the Transcriptional Targets of FOXP2, a Gene Linked to Speech and Language, in Developing Human Brain. Am J Hum Genet. 2007; 81:1144–1157. [PubMed: 17999357]

Staes N, Sherwood CC, Wright K, de Manuel M, Guevara EE, Marques-Bonet T, Krützen M, Massiah M, Hopkins WD, Ely JJ, et al. FOXP2 variation in great ape populations offers insight into the evolution of communication skills. Sci Rep. 2017; 7

Stephens M, Smith NJ, Donnelly P. A New Statistical Method for Haplotype Reconstruction from Population Data. Am J Hum Genet. 2001; 68:978–989. [PubMed: 11254454]

Teramitsu I, Poopatanapong A, Torrisi S, White SA. Striatal FoxP2 Is Actively Regulated during Songbird Sensorimotor Learning. PLoS ONE. 2010; 5

Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, et al. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]

The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005; 437:1299–1320. [PubMed: 16255080]

Torres-Ruiz R, Benitez-Burraco A, Martinez-Lage M, Rodriguez-Perales S, Garcia-Bellido P. Functional genetic characterization by CRISPR-Cas9 of two enhancers of FOXP2 in a child with speech and language impairment. 2016 BioRxiv 064196.

Uren C, Kim M, Martin AR, Bobo D, Gignoux CR, van Helden PD, Möller M, Hoal EG, Henn BM. Fine-Scale Human Population Structure in Southern Africa Reflects Ecogeographic Boundaries. Genetics. 2016; 204:303–314. [PubMed: 27474727]

Wakeley J. Coalescent Theory: An Introduction. W. H. Freeman; 2008.

Yu F, Keinan A, Chen H, Ferland RJ, Hill RS, Mignault Aa, Walsh Ca, Reich D. Detecting natural selection by empirical comparison to random regions of the genome. Hum Mol Genet. 2009; 18:4853–4867. [PubMed: 19783549]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HIGHLIGHTS

- No support for positive selection at *FOXP2* in large genomic datasets
- Sample composition and genomic scale significantly affect selection scans
- An intronic ROI within *FOXP2* is expressed in human brain cells and cortical tissue
- This ROI contains a large amount of constrained, human-specific polymorphism

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

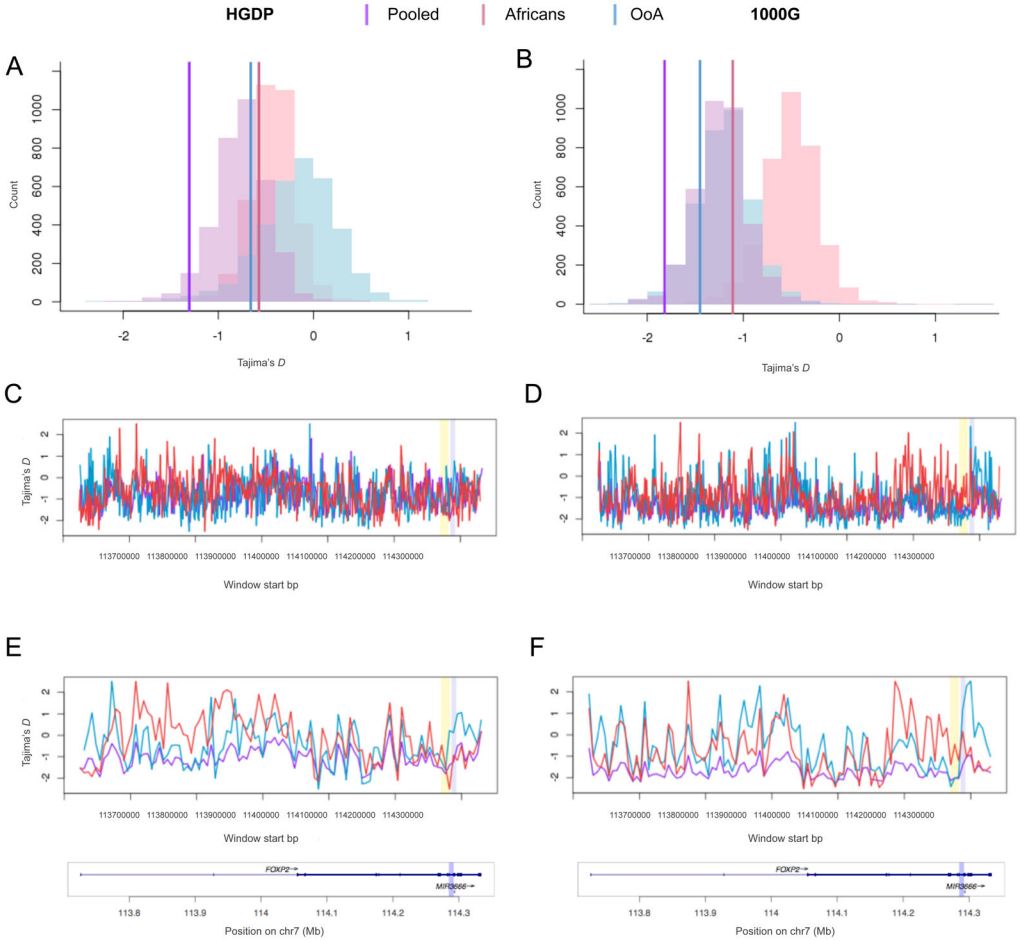


Figure 1. *FOXP2* gene region estimates of Tajima's *D* vary by population subset and window size

In all panels, purple represents the entire pooled dataset, red represents only African individuals, and blue represents only OoA individuals. HGDP results are in the left column, 1000G on the right. (A–B) Colored histograms show the background *D* values calculated from *FOXP2*-sized genic regions of the autosomes in representative population subsets of the data. Vertical lines are the *D* value for *FOXP2* in that sub-population. (C–F) "Gene crawl" of *D* values across the *FOXP2* region calculated in contiguous non-overlapping windows in the HGDP and 1000G datasets in window size of (C,D) 1500bp, and (E,F) 6kb. Below the plots is a schematic of the genes spanning this chromosomal area. The ROI is shown in lavender in the gene crawl plots and gene schematic. The 3 intron regions examined in Enard et al., (2002) are highlighted on gene crawl plots in yellow. See also Figure S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

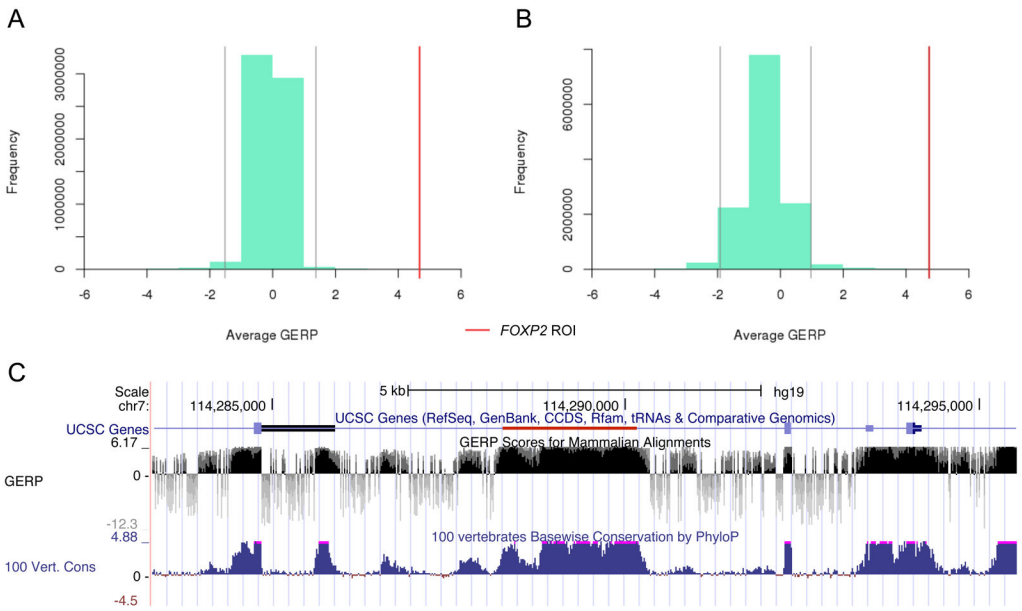


Figure 2. The ROI is a GERP outlier

The average GERP score for SNPs in the ROI in *FOXP2* (red line) is a significant outlier compared to other autosomal genic areas in both the HGDP (A) and 1000G (B) genomes datasets. Average GERP scores were calculated for all non-singleton SNPs located in windows of the same size as the ROI (2251 bp) sliding every 100 bp across the genic regions of the autosomes. Singletons were excluded to avoid bias due to sequencing error. The gray lines indicate 2 standard deviations. (C) UCSC genome browser image showing the primary transcript of FOXP2 (top track) and sequence conservation. The ROI is indicated by the red bar. The GERP conservation score for mammalian alignments is shown in black, while the PhyloP conservation score for vertebrates is plotted in navy. See also Figure S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

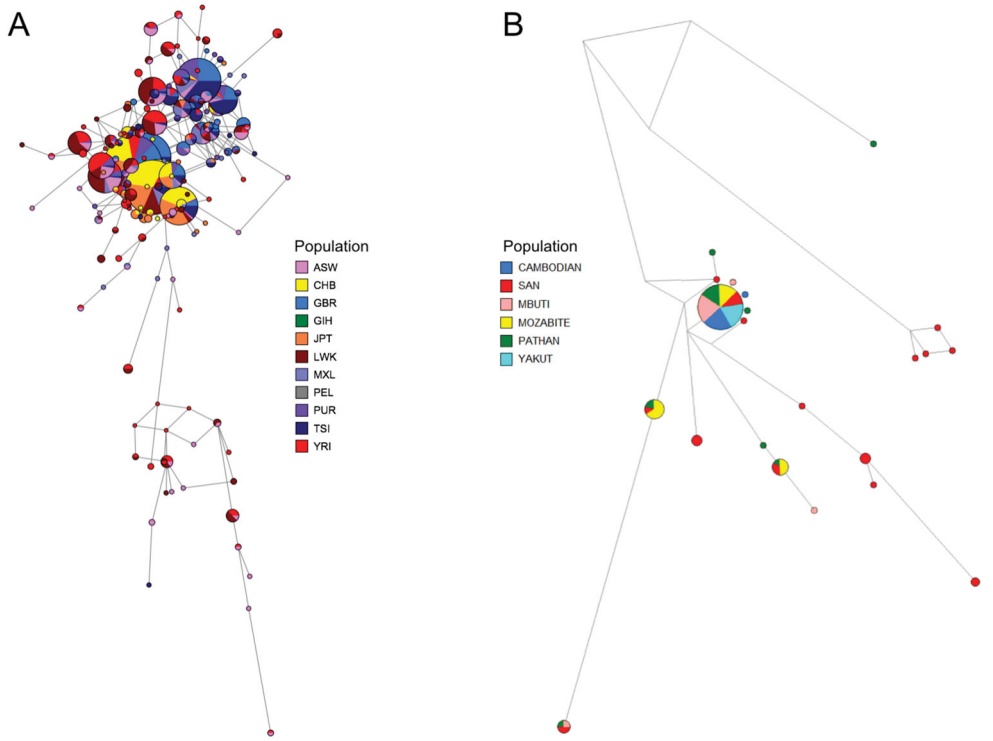


Figure 3. Haplotype networks are not reflective of a recent selective sweep at the locus

Median-joining haplotype networks for the intron 9 buffered ROI in (A) the 1000G dataset and (B) our Sanger sequencing dataset including 40 HGDP individuals and 7 additional Khomani San. Edge lengths are proportionate to mutations, node sizes are proportionate to the number of haplotypes. Nodes are colored by the population of the individual, with African individuals represented in shades of red (red, pink). Other colors represent OoA and admixed populations. See also Figures S2, S3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

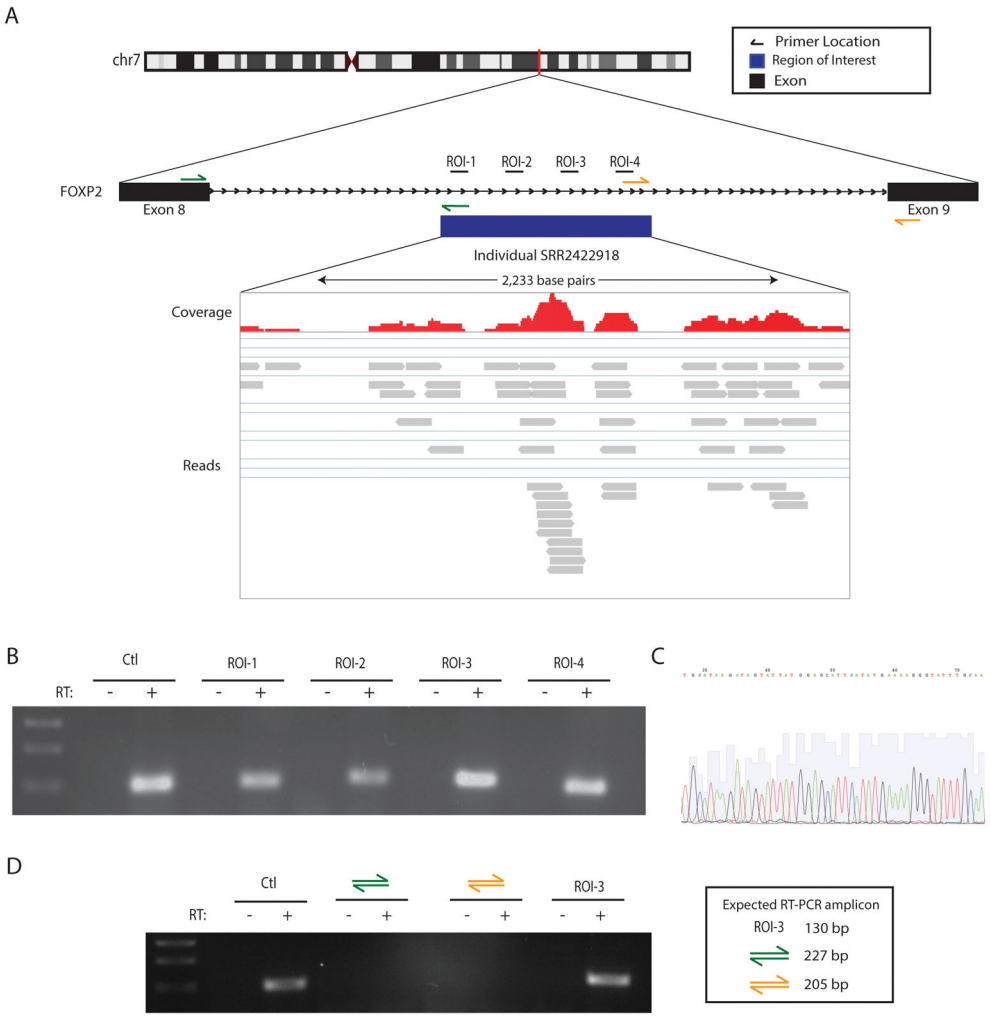


Figure 4. The intronic ROI is expressed in human brain cells

(A) Schematic representing the location of the ROI, regions amplified by RT-PCR, and RNA-seq read pileup in the ROI. RNA-seq data is from fresh human prefrontal cortex (Scheckel et al., 2016). Sequencing coverage at each position for the first individual (SRR2422918) is indicated with the red bar graph. Individual reads are represented as gray segments below the coverage plot. (B) RT-PCR analysis of the FOXP2 ROI in U87-MG immortalized human brain cells indicates that the ROI is expressed but not spliced to flanking exons in this cell type. Expression of the ROI was detected using primers within four section of the ROI (ROI-1 to 4). PCR product was amplified using combinations of primers from the ROI and surrounding exons (Exon 8/ROI or ROI/Exon 9). Positive control (Ctl) is a 101 base pair amplicon spanning exons 7–8 of FOXP2, upstream of the ROI. + and – indicate reverse transcription conditions with and without reverse transcriptase, respectively (+/− RT). Expected PCR amplicon sizes if splicing occurs from either exon 8 or exon 9 to the ROI are shown on the right. (C) Sanger sequencing analysis of the ROI amplified in (B) confirms amplification of the expected ROI sequence. (D) RT-PCR analysis

showing amplicons produced using four independent primer sets within the ROI. See also Figures S4–S6.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Values for Tajima's *D* in the partitioned HGDP and 1000G genomes datasets. The autosomal average was calculated from genic windows of the same bp width as *FOXP2* that contain at least 5 SNPs. 'Enard region' values are from tests on *FOXP2* introns 4, 5, and 6, the area examined in Enard et al. (2002). Statistics in the 5% tails of the genome-wide empirical distributions are indicated with an asterisk.

D value	HGDP genomes	1000G genomes
Autosomal Average: Pooled	−0.760	−1.223
Autosomal Average: Africans	−0.425	−0.502
Autosomal Average: OoA	−0.164	−1.193
FOXP2: Pooled	−1.305*	−1.819*
FOXP2: Africans	−0.573	−1.110
FOXP2: OoA	−0.657	−1.450
Enard region: Pooled	−1.858	−2.152
Enard region: Africans	−1.229	−1.180
Enard region: OoA	−1.279	−2.215*

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

High GERP SNPs and their derived allele frequencies in the 1000 genomes and HGDP genomes datasets. The ancestral and derived states of all alleles is presented alongside the versions found in 3 archaic hominins: the Altai and Vindija Cave Neanderthals (‘Nean A’ and ‘Nean V’) and the Denisovan (‘Deni’). The Namibian KhoeSan (‘N San’) and Mbuti are African HGDP populations, ‘OoA’ are HGDP non-African populations, CEU (European) and LWK (African) are from the 1000G, ‘K San’ are ≠Khomani San individuals representing an additional population from South Africa. HGDP and ≠Khomani San allele frequencies have been corrected based on site validation with Sanger sequencing. See also Table S1.

POS	ID	DER	ANC	Nean A	Nean V	Deni	GERP	N San	K San	Mbuti	LWK	OoA	CEU
114288164	rs115978361	C	T	C	C	C	4.48	92%	NA	100%	93%	92%	100%
114289621	rs191654848	A	G	G	G	G	3.98	8%	42%	0%	2%	0%	0%
114289641	rs114972925	T	A	A	A	A	5.87	83%	57%	100%	92%	100%	100%
114289880	rs560859215	T	G	G	G	G	6.08	17%	0%	0%	0%	0%	0%
114289918	rs116557180	G	A	A	A	A	5.78	0%	0%	0%	3%	0%	0%
114289934	rs577428580	T	C	C	C	C	4.89	8%	42%	0%	0%	0%	0%
114290213	rs7795372	A	T	T	T	T	3.32	50%	64%	86%	74%	95%	96%
114290415	rs7782412	C	T	T	T	T	4.33	25%	14%	7%	30%	9%	44%