The MER41 family of HERVs is uniquely involved in the immune-mediated regulation of cognition/behavior-related genes: pathophysiological implications for autism spectrum disorders

Serge Nataf*[1, 2, 3], Juan Uriagereka[4] and Antonio Benitez-Burraco [5]

[1]CarMeN Laboratory, INSERM U1060, INRA U1397, INSA de Lyon, Lyon-Sud Faculty of Medicine, University of Lyon, Pierre-Bénite, France.
[2] University of Lyon 1, Lyon, France.
[3]Banque de Tissus et de Cellules des Hospices Civils de Lyon, Hôpital Edouard Herriot, Lyon, France.
[4]Department of Linguistics and School of Languages, Literatures & Cultures, University of Maryland, College Park, USA.
[5]Department of Spanish, Linguistics, and Theory of Literature (Linguistics). Faculty of Philology. University of Seville, Seville, Spain

* Corresponding author: serge.nataf@inserm.fr

ABSTRACT

Interferon-gamma (IFNγ), a prototypical T lymphocyte-derived pro-inflammatory cytokine, was recently shown to shape social behavior and neuronal connectivity in rodents. STAT1 (Signal Transducer And Activator Of Transcription 1) is a transcription factor (TF) crucially involved in the IFNγ pathway. It binds consensus sequences that, in humans, are located with a high frequency in the LTRs (Long Terminal Repeats) of the MER41 family of HERVs (Human Endogenous Retrovirus). However, the role of an IFNγ/STAT1/MER41 pathway in human cognition is still poorly documented. Here, we identified a unique set of human cognition/behavior-related genes harboring MER41 LTR sequence(s) in their promoter regions. Interestingly, these MER41 LTRs exhibited TF binding sites (TFBSs) for not only STAT1 but also a significant number of immune TFs belonging to the NFKB (Nuclear Factor Kappa B) complex. These include notably STAT3 (Signal Transducer And Activator Of Transcription 3) and YY1 (YY1 transcription factor). Furthermore, a survey of proteomics and transcriptomics databases allowed to map a unique network linking TFs of the NFKB complex, with FOXP2 (Forkhead Box P2), currently known as a key TF involved in language evolution. Confirming the specificity of our findings, we demonstrated that a significant number of autism spectrum disorders (ASD) susceptibility genes exhibit a promoter-localized MER41 LTR sequence. Of note, this result was specific to MER41 HERVs since none of the more than 100 HERVs analyzed, most of which are common to all mammals, harbored such a feature. Finally, enrichment analyses of ASD susceptibility genes pointed again to the crucial involvement of STAT3 in human cognitive functions.

INTRODUCTION

Interferon gamma (IFNγ), the prototypical T Helper 1 (TH1) cytokine, is a T-cell derived pro-inflammatory molecule exerting several effects on innate immune cells and others on non-immune cells including neurons. Indeed, IFNγ was recently shown to be a social behavior regulator and to shape neuronal connectivity in rodents (Filiano et al., 2016). Irrespective of cell type, binding of IFNγ to its receptor induces the transcriptional regulation of target genes via the recognition of promoter consensus sequences, by the transcription factor STAT1 (Signal Transducer And Activator Of Transcription 1) (Green et al., 2017; Ramana et al., 2002). In humans, an important share of the IFNγ /STAT1 pro-inflammatory pathway is mediated by the binding of STAT1 to consensus sequences localized in the LTRs (Long Terminal Repeats) of the MER41 family of HERVs (Human Endogenous Retrovirus) (Chuong et al., 2016). MER41 integrated the genome of a primate ancestor 45-60 million years ago and a total of 7,190 LTR elements belonging to 6 subfamilies (MER41A-MER41G) are detectable in the modern human genome (Chuong et al., 2016). That being said, the hypothesis of an IFNγ/STAT1/MER41 pathway shaping social behavior in humans (Filiano et al., 2016) is not yet supported by experimental data. As a first step, human candidate gene(s) regulated by such a pathway need(s) to be identified. Additionally, it is important to determine if STAT1 is the sole immune TF that potentially regulate behavior- or cognition-related genes via MER41 LTRs in humans. Finally, it is worth considering whether the stepped integration of HERVs into the human genome and more generally any evolutionary change dictated by infectious events could be related to key cognitive specificities experienced by hominins, in particular the emergence of language.

To address these issues, we followed a bioinformatics workflow relying on the use of two recently-generated web tools allowing a survey of HERV sequences and their associated

transcription factor binding sites (TFBSs) in the entire human genome. Using this approach, we identified a specific set of genes harboring MER41 LTRs in their promoter region located 2 Kb upstream of the transcription starting site (TSS). Surprisingly, this list of genes was significantly enriched in cognition-related genes. Moreover, the identified MER41 LTRs were enriched in binding sites recognized by specific TFs of the NFKB complex, notably NFKB1, STAT3, CEBPB and YY1. We then performed phylogenetic comparisons between humans and chimpanzees regarding: i) MER41 LTR insertions sites in the promoter regions of candidate genes, and ii) the functional protein domains of immune TFs binding MER41 LTRs in such promoter regions. This was so as to infer putative consequences for the expression pattern of cognition-related genes, which might ultimately account for some of the cognitive differences between species. Our results indicate a possible evolutionary advantage to humans in the impact of MER41 LTRs on the transcriptional regulation of behavior/cognition-related genes. Finally, since FOXP2 is currently known as a key TF regulating aspects of brain development and function important for speech and language processing (Konopka et al., 2009; Oswald et al., 2017; Spiteri et al., 2007; Vernes et al., 2011, 2007), we searched genomics and proteomics databases to map a putative functional interactome linking TFs of the NFKB complex, MER41 LTRs and FOXP2. This way, we were able to infer from experimental data an evolutionary-determined connection between behavior/cognition-related genes, infectious diseases, and immunity. Finally, as a confirmatory investigation, such a connection was assessed in the pathological context of autism spectrum disorders (ASD).

MATERIALS AND METHODS

A scheme summarizing the workflow followed in the present work is shown in Figure 1.

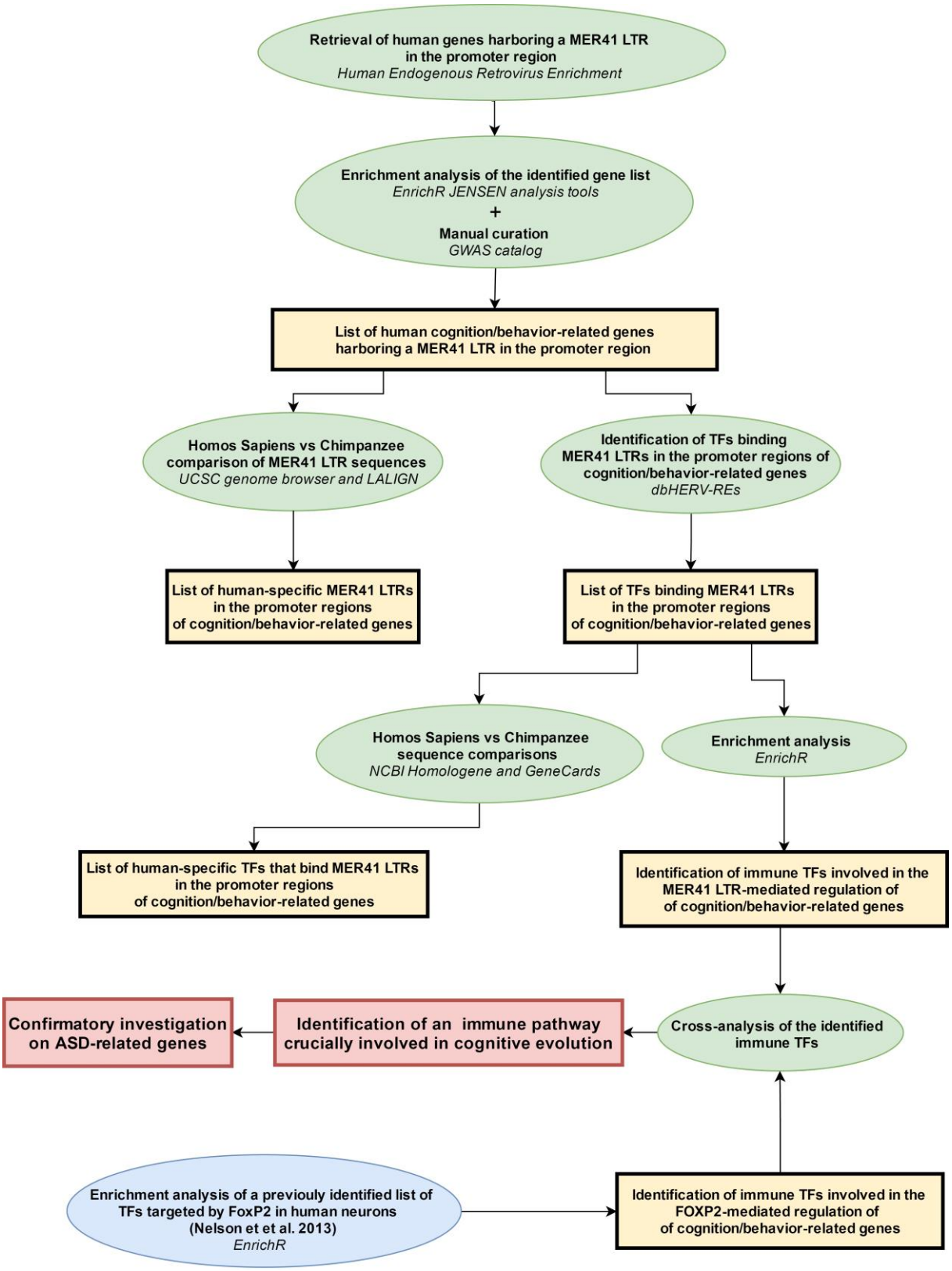

Figure 1: Wokflow of the sudy. Rectangles (yellow or red) frame the main results obtained following each of the analytical steps briefly described in ellipse shapes (green or blue). Terms in italics correspond to the name of the bioinformatics tools used for each analytical step. LTR: long terminal repeats; GWAS: genome-wide association studies; ASD: autism spectrum disorders .

Bioinformatic tools and corresponding tasks performed in this study are described below.

- The EnHERV database and web tool (Tongyoo et al., 2017): identifying human genes harboring MER41 LTR sequence(s) in the promoter region located 2 KB upstream the TSS ; only solo LTRs oriented in the sense direction relative to the gene orientation were taken into account.

- The db-HERV-RE database and web tool (Ito et al., 2017): identifying experimentally-demonstrated TFBSs in HERV LTRs.

- The EMBL GWAS database "GWAS catalog" (Welter et al., 2014): identifying among a queried list of genes those for which a SNP associated with cognitive and/or behavioral traits or conditions had been previously reported

- The “Home genetics” NCBI database (Collins et al., 2016): identifying among a queried list of genes those for which a causative link has been established with an inherited disorder affecting cognition in humans.

- The enrichment web platform EnrichR (Kuleshov et al., 2016): performing enrichments analyses on queried lists of genes. The EnrichR website allows surveying simultaneously 132 libraries gathering 245,575 terms and their associated lists of genes or proteins. Enrichment analysis tools provided by the Enrichr bioinformatics platform provides adjusted P-values computed from the Fisher’s exact test, Z-scores assessing deviation from an expected randomly obtained rank, and combined scores computed from the Z-scores and adjusted P-values obtained with the Fisher exact test. We essentially focused our analysis on 3 ontology libraries based on text-mining: i) the “Jensen TISSUES” library (Santos et al., 2015), to determine whether a list of genes is significantly associated with a specific tissue or cell type, ii) the “Jensen COMPARTMENTS” library (Binder et al., 2014), to determine whether a list of

genes is significantly associated with a specific cellular compartment or macromolecular complex and iii) the "Jensen DISEASES" library (Pletscher-Frankild et al., 2015), to determine whether a list of genes is significantly associated with a specific disease. When enrichment analyses were performed on lists of genes coding exclusively for TFs, which might potentially bias results, we performed control analyses starting from random lists of TFs. To this aim, we used a recently updated list of human TFs for which experimentally-proven or inferred TFBSs were identified (Lambert et al., 2018). From this list, we generated 50 random lists of 34 TFs using the randomization web tool "Random.Org" (https://www.random.org/lists/).

- The UCSC genome browser (Rosenbloom et al., 2015): retrieving the sequences of MER41 LTRs and their precise localization in the promoter region of cognition/behavior-related genes in the human genome (Human genome assembly GRCh38/hg38) and in the Pan troglodytes genome (Chimpanzee genome assembly CSAC 2.1.4/panTro4).

- The Swiss Institute of Bioinformatics (SIB) sequence alignment web tool LALIGN (SIB Swiss Institute of Bioinformatics Members, 2016): performing sequence comparisons between human and chimpanzee MER41 LTRs located in the promoter region of cognition/behavior-related genes. For each of these genes we checked the presence, nature and precise localization of MER41 LTR sequences in the promoter region.

- The NCBI web tool "Homologene" (NCBI Resource Coordinators, 2014): performing a comparison between human and chimpanzee regarding the protein sequences and functional domains of TFs that bind MER41 LTRs in the promoter region of cognition/behavior-related genes.

- The GeneCards database and web tool (Rebhan et al., 1997; Stelzer et al., 2016): retrieving chimpanzee orthologs of human genes of interest

RESULTS

1- Identification of human cognition/behavior-related genes harboring a MER41 LTR in the promoter region

We queried the EnHERV database and web tool (Tongyoo et al., 2017) to identify human genes harboring MER41 LTR sequence(s) in the promoter region located 2 KB upstream the TSS. 79 coding genes were identified (data supplement 1). From those, enrichment analyses were performed using the Enrichr website as described in the Materials & Methods section. While no statistically significant enrichments were found with regard to tissue-specific expression or sub-cellular localization of gene products, our identified list of 79 coding genes harbors a significant enrichment in genes associated with the term "Intellectual disability" (Table 1).

Table 1

Term	Adjusted P-value	Z-score	Combined Score
Intellectual disability	0.041	-4.73	32.10
Bullous keratopathy	0.012	-2.16	18.95
Bardet-Biedl syndrome	0.055	-2.91	16.30
Acrodermatitis enteropathica	0.049	-2.57	16.28
Senior-Loken syndrome	0.012	-1.90	15.97

Legend: an enrichment analysis was performed with the web platform EnrichR on the list a 79 coding genes harboring MER41 LTR sequences in their promoter region. Based on the analysis of "Jensen DISEASES" library, the highest statistical scores were obtained with the term "Intellectual disability". The five most significant enrichments are shown.

Indeed, 6 out of 79 genes were annotated with the term "Intellectual disability". Moreover, when performing a survey of the Genetics Home Reference database (the NCBI repository of genetic diseases) (Collins et al., 2016), we identified 4 additional genes whose

genetic alterations are responsible for human CNS inherited disorders that induce cognitive alterations. So overall, out of 79 human genes harboring a MER41 LTR sequence in their promoter region, 10 had an established causative link with human inherited disorders translating into intellectual disability. The genes and associated genetic conditions are summarized in Table 2 and described below:

- *BBS10* (Bardet-Biedl syndrome 10) *:* Bardet-Biedl syndrome 10 (vision loss, obesity, polydactily, kidney abnormalities and intellectual disability)
- *DEAF1 (*DEAF1 transcription factor): Mental retardation, autosomal dominant 24 (intellectual disability and impairments in adaptive behavior)
- *AP1S1* (Adaptor Related Protein Complex 1 Subunit Sigma 1): MEDNIK syndrome (Mental retardation, enteropathy, deafness, peripheral neuropathy, ichthyosis and keratoderma)
- *ST3GAL5* (ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 5): Salt and pepper developmental regression syndrome (epilepsy, abnormal brain development and intellectual disability)
- *CDH15* (Cadherin 15): Mental retardation, autosomal dominant 3 (intellectual disability and impairments in adaptive behavior)
- *CEP290* (Centrosomal Protein 290): Bardet-Biedl syndrome 14 (vision loss, obesity, type 2 diabetes, hypercholesterolemia, polydactily, intellectual disability, impaired speech, delayed psychomotor development and behavioral alterations)
- *GAMT* (Guanidinoacetate N-Methyltransferase): Cerebral creatine deficiency syndrome 2 (epilepsy, intellectual disability and altered speech development)

- *CDCA7* (Cell Division Cycle Associated 7): Immunodeficiency-centromeric instability-facial anomalies syndrome 3 (recurrent infections, facial anomalies and psychomotor retardation)
- *DDHD2* (DDHD Domain Containing 2): Spastic paraplegia 54, autosomal recessive (delayed psychomotor development, intellectual disability and early-onset spasticity of the lower limbs)
- *GCSH* (Glycine Cleavage System Protein H): Glycine encephalopathy (hypotonia, delayed psychomotor development and epilepsy)

Table 2

Gene symbol	Disease name	OMIM Ref
BBS10	Bardet-Biedl syndrome 10	# 615987
DEAF1	Mental retardation, autosomal dominant 24	# 615828
AP1S1	MEDNIK syndrome	# 609313
ST3GAL5	Salt and pepper developmental regression syndrome	# 609056
CDH15	Mental retardation, autosomal dominant 3	# 612580
CEP290	Bardet-Biedl syndrome 14	# 615991
GAMT	Cerebral creatine deficiency syndrome 2	# 612736
CDCA7	Immunodeficiency-centromeric instability-facial anomalies syndrome 3	# 616910
DDHD2	Spastic paraplegia 54, autosomal recessive	# 615033
GCSH	Glycine encephalopathy	# 605899

Legend: genes responsible for inherited cognition/behavior-related diseases and harboring a MER41 LTR in their promoter region were identified. Gene symbols (left column), names of the inherited disorder (middle column), and the corresponding OMIM (Online Mendelian Inheritance in Man) (Amberger et al., 2015) entries (right column) are shown.

To complement these findings, we then queried the EMBL GWAS database "GWAS catalog" (Welter et al., 2014) to identify, among our initial list of 79 genes, those for which a SNP associated with cognitive and/or behavioral traits or conditions had been previously reported. We identified 14 genes with such cognition/behavior-related SNPs (Table 3). Of note, only one gene, DDHD2, was present in both lists (Tables 2 and 3):

Table 3

Gene symbol	Associated cognitive/behavioral trait or condition
DRAXIN	Obsessive-compulsive disorder or autism spectrum disorder, Schizophrenia
CCDC136	Reading disability or specific language impairment
DDHD2	Schizophrenia, autism spectrum disorder
DDX24	Oppositional defiant disorder dimensions in attention-deficit hyperactivity disorder
DNAJB7	Autism spectrum disorder or schizophrenia
ERAP1	Alcohol dependence
GFRA4	Night sleep phenotypes
IL19	Alzheimer's disease
NENF	Loneliness
PSMG3	Executive inhibition in attention deficit hyperactivity disorder
PTN	Schizophrenia
SYNE1	Bipolar disorder, autism spectrum disorder
TRIM27	Daytime sleep phenotypes, Social communication problems, ASD; Schizophrenia
WDR19	Cognitive function

Legend: genes associated with cognition/behavior-related diseases or traits in GWAS studies and harboring a MER41 LTR in their promoter region were identified. Gene symbols (left column) and the names of associated cognititive/behavioral traits or conditions (right column) are shown. *DRAXIN*: Dorsal Inhibitory Axon Guidance Protein; *CCDC136*: Coiled-Coil Domain Containing 136; *DDX24*: DEAD-Box Helicase 24; *DNAJB7*: DnaJ Heat Shock Protein Family (Hsp40) Member B7; *ERAP1*: Endoplasmic Reticulum Aminopeptidase 1; *GFRA4*: GDNF Family Receptor Alpha 4; *IL19*: Interleukin 19; *NENF*: Neudesin Neurotrophic Factor; *PSMG3*: Proteasome Assembly Chaperone 3; *PTN*: Pleiotrophin; SYNE1: Spectrin Repeat Containing Nuclear Envelope Protein 1; TRIM27: Tripartite Motif Containing 27; *WDR19*: WD Repeat Domain 19.

Overall, 23 cognition/behavior-related genes were found to be putatively regulated at the transcriptional level by the IFNγ/STAT1/MER41 pathway in the human species.

2- LTRs from specific members of the MER41 family of HERVs are inserted in the promoter regions of cognition/behavior-related genes

TFBSs in LTRs from the MER41 family (MER41 A-E and MER41G) were shown to vary depending of the MER41 member considered (Chuong et al., 2016). Using the EnHERV database and web tool, we thus sought to precisely identify the MER41 member(s) for which LTRs can be demonstrated in the promoter regions of cognition/behavior-related genes. Surprisingly, as shown in Table 4, we identified only 4 cognition/behavior-related genes

harboring a MER41B LTR in their promoter region, thus potentially regulated at the transcriptional level by the IFNγ/STAT1 pathway: *CEP290*, *DDHD2*, *PTN*, and *GCSH*.

Table 4

MER41A	*BBS10*
MER41A	*DRAXIN*
MER41A	*CDCA7*
MER41A	*CDH15*
MER41A	*DEAF1*
MER41A	*DNAJB7*
MER41A	*ERAP1*
MER41A	*GFRA4*
MER41A	*PSMG3*
MER41A	*SLC25A37*
MER41A	*TRIM27*
MER41A	*GAMT*
MER41A	*CCDC136*
MER41B	*CEP290*
MER41B	*DDHD2*
MER41B	*PTN*
MER41B ; MER41E	*GCSH*
MER41C	*DDX24*
MER41C	*SYNE1*
MER41C	*WDR19*
MER41D	*AP1S1*
MER41D	*IL19*
MER41D	*NENF*

Legend: genes associated with cognition/behavior-related diseases or traits in GWAS studies and harboring a MER41 LTR in their promoter region were identified. Gene symbols (left column) and the names of associated cognititive/behavioral traits or conditions (right column) are shown. *DRAXIN*: Dorsal Inhibitory Axon Guidance Protein; *CCDC136*: Coiled-Coil Domain Containing 136; *DDX24*: DEAD-Box Helicase 24; *DNAJB7*: DnaJ Heat Shock Protein Family (Hsp40) Member B7; *ERAP1*: Endoplasmic Reticulum Aminopeptidase 1; *GFRA4*: GDNF Family Receptor Alpha 4; *IL19*: Interleukin 19; *NENF*: Neudesin Neurotrophic Factor; *PSMG3*: Proteasome Assembly Chaperone 3; *PTN*: Pleiotrophin; SYNE1: Spectrin Repeat Containing Nuclear Envelope Protein 1; TRIM27: Tripartite Motif Containing 27; *WDR19*: WD Repeat Domain 19.

In contrast, a majority of MER41 LTRs located in the promoter regions of cognition/behavior-related genes were MER41A LTRs, which lack STAT1 binding sites (Chuong et al., 2016). On

this basis, to determine if other immune pathways (non IFNγ/STAT1-mediated) may regulate the transcription of cognition/behavior-related genes via MER41 LTRs, we then used the HERV database and web tool db-HERV-RE (Ito et al., 2017), which allow the identification of experimentally-demonstrated TFBSs in HERV LTRs.

3- YY1 is the sole transcription factor harboring TFBSs in LTRs from* all *MER41A-E members

Using the approach described above, we identified 34 TFs that bind MER41A-E LTRs (Table 5).

Table 5

BATF	MER41E
CEBPB	MER41A
CTCF	MER41A
EBF1	MER41B
EGR1	MER41B
ELF1	MER41B
ELK4	MER41A, MER41B
ESR1	MER41A, MER41B
FOS	MER41B, MER41D, MER41E
FOSL1	MER41B
FOSL2	MER41B, MER41E
FOXA1	MER41C
FOXA2	MER41C
GATA1	MER41B, MER41E
GATA2	MER41A, MER41B, MER41E
GATA4	MER41A, MER41B
GATA6	MER41A, MER41B
JUN	MER41A
JUNB	MER41A, MER41B, MER41E
JUND	MER41D, MER41E
MEF2A	MER41B
NANOG	MER41A, MER41B
NFE2	MER41A, MER41B
NFKB1	MER41B
POU2F2	MER41B
POU5F1	MER41A, MER41B

SP1	MER41B
SPI1	MER41B
SRF	MER41A, MER41B, MER41E
STAT1	MER41B, MER41E
STAT3	MER41B, MER41E
TAL1	MER41E
USF1	MER41B
YY1	MER41A, MER41B, MER41C, MER41D, MER41E

Legend: TFs (left column) harboring TFBSs in LTR(s) of MER41 member(s) (right column) located in the promoter regions of cognition/behavior-related genes are listed along with MER41 member(s) in which corresponding TFBSs are observed (right column). BATF: CCAAT Enhancer Binding Protein Beta; CEBPB: CCAAT Enhancer Binding Protein Beta; CTCF: CCCTC-Binding Factor ; EBF1: Early B cell Factor 1; EGR1: Early Growth Response 1; ELF1: E74 Like ETS Transcription Factor 1; ELK4: ELK4, ETS Transcription Factor; ESR1: Estrogen Receptor 1; FOS: Fos Proto-Oncogene, AP-1 Transcription Factor Subunit; FOSL1: FOS Like 1, AP-1 Transcription Factor Subunit; FOSL2: FOS Like 2, AP-1 Transcription Factor Subunit; FOXA1: Forkhead Box A1; FOXA2: Forkhead Box A2; GATA1: GATA Binding Protein 1; GATA2: GATA Binding Protein 2; GATA4: GATA Binding Protein 4; GATA6: GATA Binding Protein 6; JUN: Jun Proto-Oncogene, AP-1 Transcription Factor Subunit; JUNB: JunB Proto-Oncogene, AP-1 Transcription Factor Subunit; JUND: JunD Proto-Oncogene, AP-1 Transcription Factor Subunit; MEF2A: Myocyte Enhancer Factor 2A; NANOG: Nanog Homeobox; NFE2: Nuclear Factor, Erythroid 2; NFKB1: Nuclear Factor Kappa B Subunit 1; POU2F2: POU Class 2 Homeobox 2; POU5F1: POU Class 5 Homeobox 1; SP1: Sp1 Transcription Factor; SPI1: Spi-1 Proto-Oncogene; SRF: Serum Response Factor; STAT1: Signal Transducer And Activator Of Transcription 1; STAT3: Signal Transducer And Activator Of Transcription 3; TAL1: TAL BHLH Transcription Factor 1, Erythroid Differentiation Factor; USF1: Upstream Transcription Factor 1; YY1: YY1 Transcription Factor.

As expected, MER41B LTR comprises a STAT1 consensus sequence while MER41A LTR does not. Interestingly, an YY1 consensus sequence is present in the LTRs of all the MER41A-E members. Moreover, it is worth noting that among the 34 TFs identified above, only *YY1* and *CTCF* are causative genes of inherited disorders translating into mental retardation. These comprise: 1) Gabriele-de Vries syndrome, an autosomal dominant neurodevelopmental disorder induced by mutations/deletions in *YY1* and characterized by delayed psychomotor development, intellectual disability and frequent autistic symptoms; 2) Mental retardation, autosomal dominant 21, a developmental disorder induced by mutations in *CTCF* and characterized by significantly below average general intellectual functioning associated with impairments in adaptive behavior) (Collins et al., 2016). Other inherited disorders associated to the above identified TF genes are summarized in Table 6

and include notably 3 groups of immune-related diseases induced respectively by genetic alterations of *STAT1*, *STAT3* and *NFKB1* (Table 6).

Table 6

Gene symbol	Disease name	OMIM Ref
STAT1	Immunodeficiency 31A	# 614892
	Immunodeficiency 31B	# 613796
	Immunodeficiency 31C	# 614162
STAT3	Autoimmune disease, multisystem, infantile-onset, 1	# 615952
	Hyper-IgE recurrent infection syndrome	# 147060
NFKB1	Immunodeficiency, common variable, 12	# 616576
CTCF	Mental retardation, autosomal dominant 21	# 615502
YY1	Gabriele-de Vries syndrome	# 617557
ESR1	Oestrogen resistance syndrome	# 615363
GATA4	Testicular anomalies with or without congenital heart disease	# 615542
	Atrial septal defect 2	# 607941
	Atrioventricular septal defect 4	# 614430
	Tetralogy of Fallot	# 187500
	Ventricular septal defect 1	# 614429

Legend: genes responsible for inherited disorders and coding for TFs (left column) which bind MER41 LTRs in the promoter regions of cognition/behavior-related genes are listed in left column. Corresponding names are shown of the inherited diseases (middle column) and OMIM (Online Mendelian Inheritance in Man) (Amberger et al., 2015) entries (right column).

While our approach allowed listing TFs that bind MER41 LTRs in the promoter region of specific cognition/behavior-related genes, one may argue that such TFs also bind TFBSs located outside MER41 LTRs. Their potential impact on cognition and behavior may thus be independent from MER41 LTRs. Interestingly, a recent study analyzed for a large set of TFs the proportions of TFBSs located within HERV LTRs relative to those outside HERV LTRs (Ito et al., 2017). A survey of this study shows that among the 34 TFs we identified as potential MER41 LTR-mediated transcriptional regulators of cognition/behavior-related genes, almost half (15 of the 34) belong to the top 25 TFs harboring the highest proportions of HERV TFBSs (Table 7).

Table 7

TF	TFBSs	HERV-TFBSs	Proportion
GATA6	32,23	7,073	21.9%
USF1	138,147	29,524	21.4%
GATA4	81,738	16,359	20.0%
TAL1	35,321	6,951	19.7%
STAT1	21,727	3,515	16.2%
YY1	193,183	30,506	15.8%
GATA1	49,133	7,714	15.7%
SPI1	131,487	19,731	15.0%
GATA2	112,602	16,309	14.5%
SRF	32,776	4,623	14.1%
NANOG	102,008	13,903	13.6%
NFE2	56,155	7,528	13.4%
JUNB	31,088	4,086	13.1%
JUND	220,44	28,583	13.0%
STAT3	114,24	14,509	12.7%

Legend: table adapted from Ito et al. (Ito et al., 2017). TFs of interest were retrieved from a list of 25 TFs which, from a large list of human TFs, harbor the highest proportions of TFBSs in HERV LTRs. Among these, 25 TFs are listed in the left column 10 TFs for which TFBSs are observed in MER41 LTRs of cognition/behavior-related genes. The column entitled "TFBSs" indicates the corresponding numbers of TFBSs in the whole genome. The column entitled "HERV-TFBSs" indicates the corresponding numbers of TBSSs located within HERV-LTRs. The right column indicates for each TF the proportion of TFBSs located within HERV-LTRs relative to all TFBSs.

4- Consensus sequences in MER41A-E LTRs are bound by a unique network of immune TFs belonging to the NFKB complex

We then performed an enrichment analysis of the 34 TFs for which TFBSs were found in the MER41 LTRs located in the promoter regions of cognition/behavior-related genes. When querying in EnrichR the Jensen TISSUES library of tissue-specific gene sets, we observed a highly significant enrichment in genes expressed in the immune system (Table 8).

Table 8

Name	Adjusted P-value	Z-score	Combined score
Immune system	3.73e-11	-6.92	204.50
Adult	3.19e-10	-7.11	183.60
Finger	7.89e-8	-5.21	104.88
Mesoderm	1.19e-10	-3.52	96.24

Embryoid_body	2.62e-10	-3.15	82.65

Legend: an enrichment analysis was performed with the web platform EnrichR on the list of 34 TF-coding genes that bind MER41 LTR sequences in the promoter regions of cognition/behavior-related genes. Based on the analysis of "Jensen TISSUES" library, the highest statistical score was obtained with the term "Immune system". The five most significant enrichments are shown.

Correspondingly, a significant enrichment in genes involved in immune disorders such as Crohn's disease or arthritis was also found (Table 9).

Table 9

Name	Adjusted P-value	Z-score	Combined score
Crohn's disease	2.99e-4	-4.53	56.99
Cancer	0.003	-4.53	40.13
Arthritis	0.003	-4.39	36.10
Atrial heart septal defect	0.002	-3.35	32.60
Hematologic_cancer	0.003	-3.15	28.32

Legend: an enrichment analysis was performed with the web platform EnrichR on the list of 34 TF-coding genes that bind MER41 LTR sequences in the promoter regions of cognition/behavior-related genes. Based on the analysis of "Jensen DISEASES" library, the highest statistical score was obtained with the term "Crohn's disease". The five most significant enrichments are shown.

Finally, enrichment analysis with regard to sub-cellular localization of gene products (Table 10) showed a highly significant enrichment in proteins associated to the term "nucleoplasm" (as expected, since the analyzed genes are all TFs) and a similarly significant enrichment in proteins belonging to the NFKB complex (20 out of 34 TFs). This macromolecular protein complex is formed by first degree (direct) or higher degrees (indirect) protein partners of NFKB1, a TF involved in the positive transcriptional regulation of key immune genes (Refs).

Table 10

Name	Adjusted P-value	Z-score	Combined score
Nucleoplasm	6.91e-14	-8.61	297.25
NF-kappaB complex	5.53e-17	-6.18	263.23

Bcl-2 family protein complex	1.12e-7	-7.98	153.20
Chromatin accessibility complex	1.02e-18	-2.99	141.16
BCL-2 complex	2.09e-7	-7.56	139.39

Legend: an enrichment analysis was performed with the web platform EnrichR on the list a 34 TF-coding genes harboring MER41 LTR sequences in their promoter region. Based on the analysis of "Jensen COMPARTMENTS" library, the highest statistical scores were obtained with the term "Nucleoplasm" (an expected finding considering that analyzed list comprise exclusively TF-coding genes) and the term "NF-kappaB complex". The five most significant enrichments are shown.

To ascertain that our findings were not biased by analyzing a list exclusively composed of TFs, we generated 50 random lists of TFs as described in Materials & Methods and retrieved for each the adjusted P-values obtained for the term "Immune system" with the "Jensen TISSUES" tool. For all 50 TF random lists analyzed, adjusted P-values were < 0.05 (data supplement 2).

These results show that MER41A-E LTRs located in the promoter regions of cognition/behavior-related genes provide regulatory consensus sequences that are recognized by a specific network immune TFs, in particular TFs associated to the NFKB complex. On this basis, we explored the BioGRID database of human protein interactions (Chatr-aryamontri et al., 2015) to map the protein network formed by the 20 TFs binding MER41 LTRs and belonging to the NFKB complex. The retrieved network is shown in **Figure 2.**

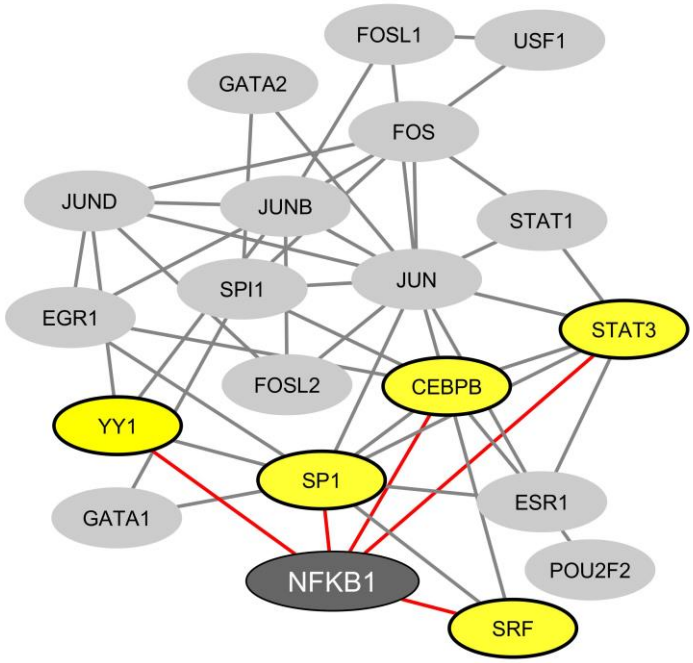

Figure 2: Mapping of the protein network formed by members of the NFKB complex that bind MER41 LTRs in the promoter regions of cognition/behavior-related genes. A survey of the human proteome was performed by querying the protein-protein interaction database BioGRID (Chatr-aryamontri et al., 2015). Nodes in yellow indicate direct (fist degree) NFKB1 protein interactor; edges in red indicate direct (first degree) interactions with NFKB1; other protein interactions are depicted by grey lines.

5- Chimpanzees vs humans comparative analysis of MER41A-E LTRs sequences and locations in the promoter regions of candidate cognition/behavior-related genes

As mentioned, MER41 HERVs integrated the genome of a primate ancestor 45-60 million years ago. The process of so-called "ERV domestication" (Dewannieux and Heidmann, 2013) relies on mechanisms that are not only species-specific, but may have partly shaped speciation (Johnson, 2015). Accordingly, in primates the species-specific domestication of MER41 HERVs translates into the existence of species-specific differences regarding the insertion sites and/or sequences of integrated (fixed) LTRs. On this basis we investigated whether the promoters of cognition/behavior-related genes harbored the same MER41A-E LTRs in human and chimpanzees. Out of 23 candidate genes examined we found that 14 exhibited, in both species, MER41 LTR sequences belonging to the same family, displaying 95

to 100% homology (Data supplement 3). That being said, for 9 of 22 candidate genes (*DRAXIN*, *CDH15*, *CEP290*, *CDCA7*, *DDHD2*, *IL19*, *GCSH*, *SYNE1* and *PSMG3*) MER41 LTR sequences were not observed in the 2 Kb promoter regions of the chimp orthologs. In four cases (*DRAXIN*, *CEP290*, *CDCA7*, and *DDHD2*), they were found at distances larger than 2 Kb from the TSS. In other 5 cases (*CDH15*, *IL19*, *GCSH*, *SYNE1* and *PSMG3*), they were absent, at least up to 10 Kb from the TSS. These results are indicative of differences that may prove functionally relevant with regard to the MER41 LTR-mediated transcriptional regulation of selected behavior/cognition-related genes. This remains to be experimentally explored. However, overall, the observed differences are suggestive of some evolutionary advantage to humans in the integration of MER41 LTRs in the promoter of specific behavior/cognition-related genes. It is of note that, according to the classification provided by the Gene ontology (GO) consortium (The Gene Ontology Consortium, 2017), several of these genes are annotated with "Biological process" GO terms that may possibly render account for distinctive features between chimpanzees and human CNS (data supplement 4). These include the terms "visual learning" and "locomotor behavior" for *DDHD2*, "hindbrain development" for *CEP290* and "forebrain development", "spinal cord development", "axon guidance" and "negative regulation of neuron apoptotic process" for *DRAXIN*.

6- TFs that are binding MER41 LTRs in the promoter regions of cognition/behavior-related genes show species-specific features that discriminate chimpanzees from humans

To further assess the potential role of MER41 LTRs in the species-specific transcriptional regulation of cognition/behavior-related genes, the amino acid sequences of the 34 TFs that were found to bind MER41 LTRs in humans were then compared to those of their orthologs in chimpanzees. Using the NCBI-provided web tool "Homologene" (NCBI Resource

Coordinators, 2014), we found that for 3 TFs (SPI1, FOXA2, and JUND), domain comparisons with "Homologene" could not be performed because of the sequence homology being too low (71% for SPI1) or a lack of chimp orthologs (FOXA2 and JUND) (Rebhan et al., 1997; Stelzer et al., 2016). While 27 of the remaining 31 TFs displayed a 100% homology, 4 TFs, namely JUNB, STAT1, EGR1, and YY1, harbored differences with regard to functional domains (data supplement 5). Of note, these 4 TFs belong to the network of NFKB complex TFs we identified (Figure 2). Among these genes, the most promising candidate is YY1. This is the only direct protein interactor of NFKB1 in the NFKB network. Moreover, YY1 is the only TF for which a TFBS is present in all subtypes of MER41 (i.e., MER41 A to E) LTRs in the promoter regions of candidate cognition/behavior-related genes. Interestingly, the difference between the human and chimp YY1 sequences resides in functional DNA binding domains (data supplement 5). Also, as compared to humans, the STAT1 protein in chimps is missing a TAZ2 binding domain which is required for the binding of STAT1 to CREB-binding protein, a TF coactivator whose mutations are responsible for the neurodevelopmental disorder called Rubinstein-Taybi Syndrome (characterized by short stature, moderate to severe intellectual disability, distinctive facial features, broad thumbs and first toes).

7- Immune TFs binding MER41 LTRs in the promoter regions of cognition/behavior-related genes are linking FOXP2 to the NFKB pathway

FOXP2, a TF abundantly expressed in cortical neurons, is uniquely involved in cognitive and behavioral evolution and developmental emergence of human language (Fisher and Scharff, 2009; Scharff and Petri, 2011; Vernes et al., 2011; Xu et al., 2018). Until recently, the transcriptional activity FOXP2 was thought to rely on the recognition of specific TFBs by FOXP2 homodimers or by FOXP1/FOXP2 or FOXP4/FOXP2 heterodimers (Sin et al., 2015;

Wang et al., 2003). However, a recent work established a short list of TFs that bind FOXP2 and are likely to regulate FOXP2 availability and/or DNA binding properties in neurons (Estruch et al., 2018). Interestingly, YY1 was identified as one of the 7 newly identified FOXP2-interacting TFs. We thus sought to determine whether YY1 could potentially represent a molecular link between FOXP2 and the NFKB/MER41 pathway we identified. To this aim, we retrieved TF-coding genes from the entire list of 863 human FOXP2 neuronal target genes that can be currently obtained from the ChEA 2016 and ENCODE 2015 libraries (available via the EnrichR enrichment web tool)(Kuleshov et al., 2016), two manually-curated databases of Chip-seq or Chip-Chip experiments. The 863 human FOXP2 neuronal targets listed in these libraries were retrieved from a study by Nelson et al. (Nelson et al., 2013), in which results obtained from two distinct human neuronal cell lines (PFSK-1 1 and SK-N-MC cell lines) were crossed and then checked by microfluidic affinity assays. Interestingly, from this initial set of genes, the recovered list of 60 TFs targeted by FOXP2 in human neurons (data supplement 6) was found to be significantly enriched in genes annotated with the terms “Immune System”, “NF-kappaB complex” or “Crohn’s disease” (Table 11). Other enrichments are shown in data supplement 6.

Table 11

Name	Adjusted P-value	Z-score	Combined score
NF-kappaB complex	1.16e-4	-6.00	80.15
Immune system	0.002	-6.75	65.28
Crohn's disease	0.02	-4.39	30.55

Legend: an enrichment analysis was performed with the web platform EnrichR on the list 60 TF-coding genes previously demonstrated to be transcriptionally regulated by FOXP2 in human neurons as assessed by Chip-seq experiments. Based on the analysis of the “Jensen COMPARTMENTS”, “Jensen DISEASES” and “Jensen TISSUES” libraries, statistically significant scores were obtained with the terms “NF-kappaB complex”, “Immune system”, and “Crohn’s disease”.

Importantly, out of 60 TFs targeted by FOXP2 in human neurons, 13 TFs belong to the NFKB complex including NFKB1, CEBPB, STAT3, and YY1 (Figure 3). Thus, a transcriptional regulatory pathway may link FOXP2 to MER41 LTRs located in the promoter regions of specific cognition/behavior-related genes, as depicted in Figure 3.

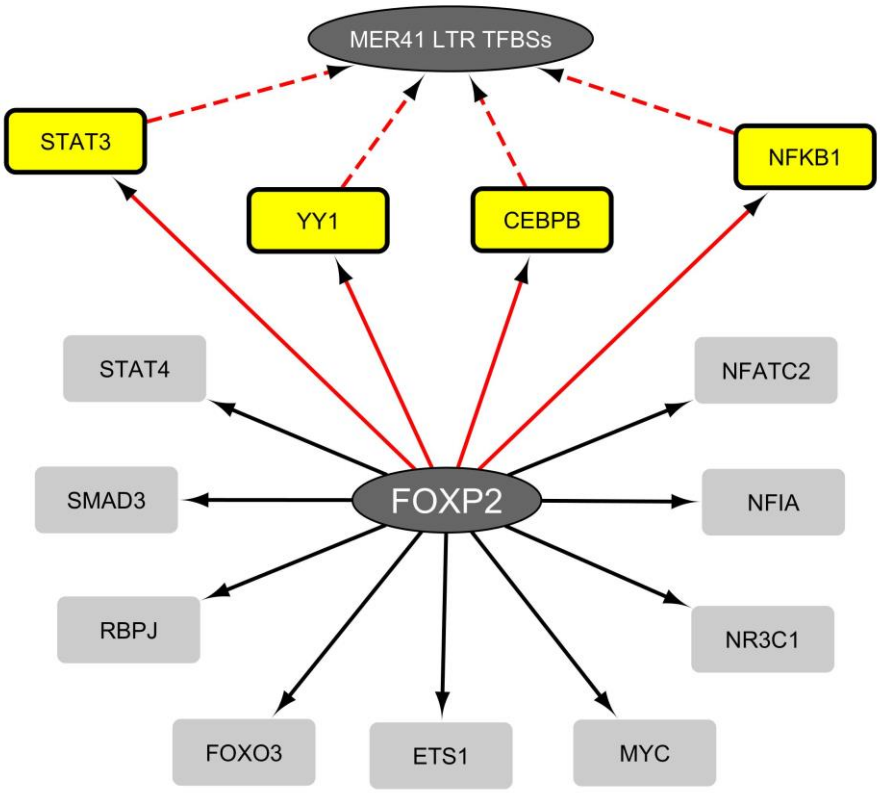

Figure 3: FOXP2 in human neurons is a transcriptional regulator of key genes coding for TFs of the NFKB complex. The list of TFs targeted by FOXP2 in the human neuronal cell lines PFSK-1 and SK-N-MC was retrieved from Ref X. Among these TFs, only those coding for molecules of the NKB complex are represented in this figure. Plain arrows indicate transcriptional regulatory links identified by Chip-seq experiments. Dashed arrows indicate binding of TFs to TFBSs as inferred from experimental and/or *in silico* analyses. Nodes in yellow and arrows in red indicate a putative pathway linking FOXP2 to cognition/behavior-related genes harboring a MER41 LTR in their promoter region.

To complement and confirm these findings, we explored data obtained from a recent work (Oswald et al., 2017) attempting to identify in the human neuronal cell line SH-SY5Y a set of FOXP2 neuronal targets that are specific to human FOXP2. More precisely, data from this

study were obtained by: i) a meta-analysis of previous works reporting on FOXP2 neuronal targets (based notably on Chip-Seq analyses of the neuronal cell line SH-SY5Y) (Enard et al., 2009; Hilliard et al., 2012; Konopka et al., 2009; Spiteri et al., 2007; Vernes et al., 2011, 2007) and ii) a comparison of neuronal genes that are targeted by human FOXP2 vs non-human primates orthologs of FOXP2 in the neuronal cell line SH-SY5Y (Oswald et al., 2017). A set of 40 candidate proteins coded by FOXP2-targeted genes was identified. Protein interactors of these candidate molecules were added in order to establish a final list of 80 neuronal proteins that are putatively regulated by FOXP2 in a human-specific manner. It is of note that only 3 molecules are shared between the 2 lists of FOXP2 neuronal targets established by Oswald et al. (Oswald et al., 2017) and Nelson et al. (Nelson et al., 2013) respectively. These are PABPC1 (Poly(A) Binding Protein Cytoplasmic 1), EIF2C2 (Protein argonaute-2), and STAT3. Such a discrepancy may relate to the genomic differences under basal conditions between the neuronal cell lines analyzed (SH-SY5Y cell line in Oswald et al. vs PFSK-1 1 and SK-N-MC cell lines in Nelson et al.). Interestingly, however, enrichment analysis of the list established by Oswald et al. (i.e., the 80 neuronal proteins that are putatively regulated by FOXP2 in a human-specific manner) showed again a highly significant enrichment in genes annotated with the terms “NF-kappaB complex” or “Immune System” (data supplement 7). Also, such a list comprised 11 TFs that, similarly, were significantly enriched in TFs annotated with the term “NF-kappaB complex”, according to the “Jensen COMPARTMENTS” classification (adjusted p-value: 1.23e-5), or with the term “Immune system”, according to the “Jensen TISSUES” classification (adjusted p-value: 2.47e-3) (data supplement 7).

We then sought to determine whether the identified FOXP2-targeted TFs belonging to the NFKB complex may form a protein network with FOXP2 via YY1, the recently identified FOXP2 partner. We focused our analysis on the set of FOXP2 targets identified by Nelson et

al., since it includes YY1. To achieve our goal, we surveyed the BioGRID database of human protein interactions (Chatr-aryamontri et al., 2015) and retrieved the direct (first degree) protein partners that are currently identified for each of the NFKB complex TFs whose coding genes are targeted by FOXP2 in the study by Nelson et al. (Nelson et al., 2013). Using this approach, we could map a protein network connecting NFKB1, FOXP2, and YY1 along with 8 TFs belonging to the NFKB complex (Figure 4).

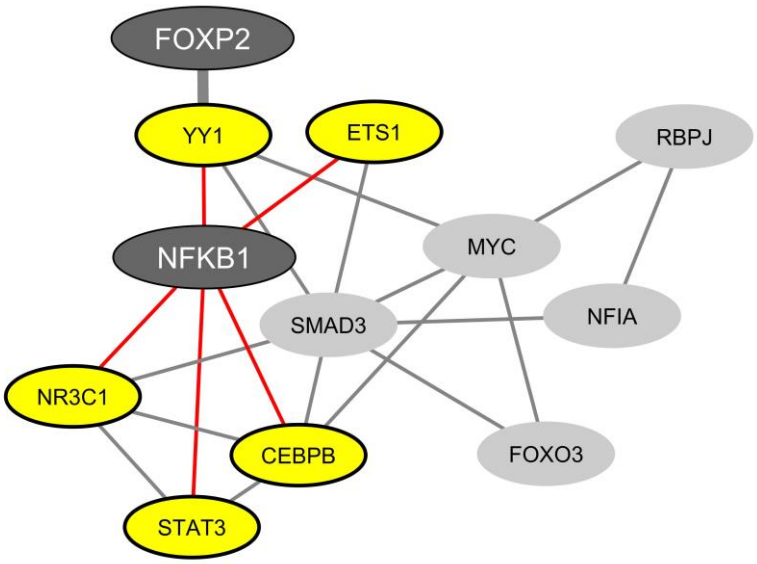

Figure 4: Mapping of the protein network linking FOXP2 to members of the NFKB complex that are themselves transcriptionally regulated by FOXP2 in human neurons. Protein interactions linking NFKB-related TFs that are transcriptionally regulated by FOXP2 were retrieved from the human proteome database BioGRID (Chatr-aryamontri et al., 2015). Edges in red indicate direct (first degree) interactions with NFKB1; nodes in yellow indicate direct (fist degree) NFKB1 protein interactors that bind MER41 LTRs in the promoter regions of cognition/behavior-related genes; other protein interactions are depicted by grey lines. Homodimeric interactions were not depicted.

8- Confirmatory investigation on ASD susceptibility genes

Overall, our analyses identified a specific set of cognition/behavior-related genes that are potentially regulated via a NFKB/MER41 LTR pathway. However, since we tested only 1 out of over 100 currently known families of HERV, the specificity of our findings remained to be

tested. In addition, the potential pathophysiological implications of such a newly identified connection between immunity and cognition had to be explored.

To that aim, we retrieved the whole list of 1019 human genes that are considered ASD susceptibility genes, as assessed and regularly updated via manual curation of the literature on the SFARI knowledgebase for ASD (Abrahams et al., 2013) . We used the EnHERV database and web tool to evaluate if this list of ASD susceptibility genes is significantly enriched in genes harboring a HERV LTR in their promoter regions. In a manner similar to the strategy followed for the genome-wide identification of genes harboring a promoter-localized MER41 LTR-sequence, we chose the following analytical parameters: solo LTRs located in the 2 KB region upstream the TSS and oriented in the sense direction relative to gene orientation. The 133 families of HERV that can be mined on EnHERV were successively queried and, strikingly, the only significant enrichment with a p-value < 0.05 (Fisher exact test) was observed for the MER41 family of HERV. Further supporting the specificity of our findings, when analyzing the 22 lists of non-CNS related genes provided as training lists by the EnHERV server, we did not find any significant enrichment in genes with promoter-localized MER41 LTR sequences. Interestingly, among the 8 ASD-related genes harboring a MER41 LTR sequence in their promoter region, only 3 belonged to the list we initially identified from our genome-wide exploratory strategy (*CEP290*, *DEAF1* and *SYNE1*). Indeed, the SFARI gene list is based on a manual curation of data deriving from unique clinical case reports or studies performed in animal models. Thus, of 79 human genes harboring a MER41 LTR in their promoter region, 23 exert cognition/behavior-related functions according to the GWAS catalog and a survey of the Home Genetics Reference website, while 5 additional genes relate to ASD pathophysiology, according to the SFARI website. These are: *ZNF8* (zinc

finger protein 8), *NR1H2* (nuclear receptor subfamily 1 group H member 2), *EXOC6* (exocyst complex component 6), *ST7* (suppression of tumorigenicity 7) and *SLC30A5* (solute carrier family 30 member 5). Regarding the MER41 LTRs associated to these genes, we did not observe major differences in terms of sequences or distances from the TSS when comparing chimpanzee vs human genomes (data not shown). Of note, the 1019 ASD-related genes listed in the SFARI website comprise 79 TFs that include YY1, STAT1, ESR1 and a significant number of TFs belonging to the NFKB complex (adjusted p-value: 0.008 according to the Jensen COMPARTMENT analysis web tool).

To get further insights into the potential links between ASD pathophysiology and the NFKB pathway, we then attempted to determine whether ASD susceptibility genes had been previously reported to be targeted byTFs belonging to the NFKB complex. To this aim, we mined, via the Enrichr platform, the ChEA 2016 library of Chip-Seq and Chip-Chip experiments in which, for a given TF and a given cell type, putative target genes are listed on the basis of identified TFBSs. Surprisingly, STAT3 was the human TF harboring the most statistically significant number of binding sites putatively regulating ASD-related genes (adjusted p-value: 1.31E-31) (data supplement 8). Thus, out of 1019 currently identified ASD susceptibility genes, 312 (30.6%) are putative STAT3 targets. To a lesser extent, a significant enrichment was also found for JUN and SPI1 (two TFs binding MER41 LTRs and belonging to the NFKB complex) and for EGR1 and FOXA2 (two TFs binding MER41 LTRs but not belonging to the NFKB complex) (data supplement 8). Overall, STAT3, JUN and SPI1 are 3 members of the NFKB complex that, collectively, bind putative regulatory regions for as much as 418 (41%) ASD susceptibility genes. Finally, since these results pointed to a specific role of STAT3 in ASD pathophysiology, we attempted to determine whether, conversely, STAT3

transcriptional targets are enriched in CNS-related genes. An enrichment analysis of the currently known 6014 STAT3 gene targets, as identified by Chip-Seq or Chip-Chip experiments and listed on the "Harmonizome" (Rouillard et al., 2016) ChEA library, showed that "Hypothalamus" and "Brain" are the terms that most significantly associate with these gene targets according the "Jensen TISSUES" enrichment tool (adjusted p-value: 1.70 e-75 and 3.25e-55 respectively) (data supplement 8).

DISCUSSION

In our species, we found a significant enrichment in cognition/behavior-related functions among genes harboring MER41 LTRs in their promoter regions. Indeed, of 79 human genes harboring a MER41 LTR in their promoter region, 28 (i.e 35%) exert cognition/behavior-related functions. Interestingly, in these specific regulatory regions, we also observed substantial differences between human and chimpanzees regarding the localization and sequences of MER41 LTRs. These results suggest that the MER41 family of ERV could have been involved in cognitive changes after our split from chimps . In this scheme, infection and horizontal transmission of MER41 ERVs in a community of primate ancestors would have led to germline infection followed by vertical transmission. Then, genomic evolution from these primate ancestors would have been, at least in part, dictated by the processes of ERV endogenisation and domestication, which is itself mainly dictated by the host's immune system (Dewannieux and Heidmann, 2013). Accordingly, differences regarding the insertion sites of MER41 LTRs in the promoter region of a large range of genes, including cognition/behavior-related genes, might have played roles in cognitive speciation.

Our work also shows that, in humans, the immune-mediated MER41 LTR-mediated transcriptional regulation of cognition/behavior-related genes is not limited to the

IFNγ/STAT1 pathway. MER41 LTRs in the promoter regions of cognition/behavior-related genes are enriched in TFBSs recognized by a unique set of immune TFs belonging to the NFKB complex and forming an interconnected protein network. At the protein level, several of these immune TFs (namely STAT1, YY1, JUNB and EGR1) exhibited major differences in terms of functional domains when comparing humans to chimpanzees. Thus, cognitive evolution after our split from chimps might have been influenced by the process of endogenisation and domestication of MER41 ERVs and by the parallel genomic evolution of key immune TFs. In this view, it is worth noting that immune genes harbor the highest levels of purifying selection in the human genome, which reflects the key functions of immunity in the defense against life-threatening infectious agents (Deschamps et al., 2016). This is notably the case for STAT1, a functionally important member of the NFKB complex (Deschamps et al., 2016).

Of note, YY1 was identified as a functional hub in this network of immune TFs. Indeed, YY1 binding sites are observed in the LTRs from all MER41 subtypes (MER41 A to E) and YY1 is a direct protein partner of NFKB1. Moreover, in humans, YY1 harbors Zinc finger type DNA binding domains that are distinct from those observed in chimpanzees. Finally, YY1 is not only recognized as being crucially involved in CNS development (as shown notably in the inherited brain disorder “Gabriel-de Vries syndrome”) but as exerting major functions in the adaptive immune system. In particular, YY1 was demonstrated to inhibit differentiation and function of regulatory T cells by blocking Foxp3 expression (Hwang et al., 2016) and to regulate effector cytokine gene expression and T(H)2 immune responses (Guo et al., 2008).

While assigning a putative role to the NFKB/MER41 pathway in cognitive evolution, our results also demonstrate that the NFKB complex is connected to FOXP2 at the mRNA and

protein levels. Such unexpected connection unravels a new molecular pathway between immunity and cognition and, even more specifically, between immunity and language. We previously proposed that the nervous and immune systems have somehow co-evolved to the benefits of both systems (Benítez-Burraco and Uriagereka, 2016a; Serge Nataf, 2017). The present paper suggests that, independently from the existence of MER41 LTRs in the promoter of specific cognition/behavior-related genes, important molecular pathways supporting cognition and language co-opted innate immune molecules of the NFKB complex. Supporting this view, previous works provided evidence that, in both invertebrates and mammals, neuronal NFKB is essential to behavior and cognitive functions (Dresselhaus et al., 2018; Kaltschmidt and Kaltschmidt, 2009; Mattson and Meffert, 2006; Meffert and Baltimore, 2005). In the central nervous system of rodents, components of the NFKB complex are detectable in neuronal processes and synapses under physiological conditions (Dresselhaus et al., 2018; Salles et al., 2014). Moreover, synaptic transmission as well as exposure to neurotrophins activates the NFKB pathway in neurons (Kaltschmidt and Kaltschmidt, 2009; Mattson and Meffert, 2006; Meffert and Baltimore, 2005). In turn, NFKB activation in neurons triggers the transcription of multiple neuronal genes that may favor cognition and shape behavior. This is notably the case for neuropeptide Y and BDNF (Snow and Albensi, 2016). Finally, a previous paper reported that, in the FOXP2 gene, regulatory changes that distinguish modern humans from Neanderthals affect 6 TFBSs among which 4 are recognized by NFKB1, STAT1, or CEBPB (Snow and Albensi, 2016). Thus, although NFKB-mediated regulation of FOXP2 has not been demonstrated yet, our work calls for further investigation on interplay between FOXP2 and the NFKB pathway in human neurons. If formally demonstrated, the molecular connections between FOXP2, the NFKB pathway, and MER41 LTRs would shed new light on the impact of immunity on cognitive evolution. Such a

molecular connection would also provide new cues on the pathophysiology of neurodevelopmental disorders like ASD and schizophrenia, which are thought to involve immunity. In particular, our data are in line with the growing body of evidence pointing to maternal infections as an etiology of ASD (Estes and McAllister, 2015; Li et al., 2018; Murray et al., 2018; Varghese et al., 2017). Similarly, one may speculate on the fact that, in human neurons, FOXP2 targets immune TFs that are involved in the pathophysiology of Crohn's disease, a dysbiosis-associated inflammatory condition of the intestinal tract (de Souza and Fiocchi, 2016; Neurath, 2014; Zhang et al., 2017). Since gut microbiota-driven inflammatory signals influence the development and maintenance of brain neuronal networks (Codagnone et al., 2018; Hoban et al., 2016; Kelly et al., 2017; Osokine and Erlebacher, 2017; Sarkar et al., 2018), it could be hypothesized that neuronally-expressed TFs of the NFKB complex are involved in a physiological inflammatory pathway linking gut microbiota to cognition and, more specifically, language (Benítez-Burraco and Uriagereka, 2016b; Kelly et al., 2017).

More than 100 families of HERV are integrated in our genome. It was thus important to determine whether and to which extent our findings are specific to MER41. We found that MER41 was the only HERV family whose LTRs were found with statistically high frequency in the promoter regions of ASD-related genes. It must be emphasized that, while many HERV families were inherited from ancestors common to all mammals, the MER41 family is detected exclusively in the genome of primates. We also observed that STAT3, a immune TF involved in functionally crucial NFKB-mediated pathways, is the human TF harboring the most highly significant number of targets among ASD-related genes (312 out of 1019 ie 30.6%). Conversely, text mining analysis showed that "Hypothalamus" and "Brain" are the terms being the most highly significantly associated with the list of currently known target genes of STAT3. Overall, these data indicate that the connection between immunity

in one hand (specifically, TFs of the NFKB complex) and cognition/behavior on the other (specifically, ASD-related genes) largely overcomes the sole presence of MER41 LTRS in the promoter region of a restricted set of cognition/behavior-related genes. These results warrant further studies to investigate the potential involvement of STAT3-mediated hypothalamic events in ASD pathophysiology. Interestingly, however, recent works showed that STAT3 inhibition exerted therapeutic effects in animal models of ASD (Ahmad et al., 2018; Parker-Athill et al., 2009). Furthermore, the hypothalamus hosts specific neuronal populations, notably oxytocin- or vasopressin-secreting neurons, that have been involved in ASD pathophysiology (Insel, 2010; Meyer-Lindenberg et al., 2011; Romano et al., 2016).

Finally, in a more provocative view, the present paper may lead to reconsider the notion of cognitive evolution itself. A potential implication of our findings is, indeed, that cognitive evolution among primates might be a bystander effect of the immune system evolutionary adaptation to infections in general and, in particular, retroviral infections. Excluding this hypothesis would lead to consider the related and barely less provocative view that the last steps of our cognitive evolution are actively involved in immune responses against human-specific infectious agents. In any case, our work reinforces the notion of neuroimmune co-evolution that we previously put forward (Benítez-Burraco and Uriagereka, 2016a; S. Nataf, 2017b, 2017a). In this general frame, we would like to propose that, besides the potential role of endogenous immune cues (S. Nataf, 2017b, 2017a), immune signals triggered by infectious agents, including gut microbiota, might have been essential to cognitive evolution. In particular, depending on their pathogenicity, such infectious agents would have exerted a neuroimmune selection pressure over millions of years (for instance, via the self-domestication of HERVs) or during short periods of time (for instance, via the occurrence of life-threatening epidemics of viral or bacterial infections). In this view, our

findings allow a re-actualization of the previously enunciated hypothesis of a recent emergence of linguistic skills that would have been triggered by a fast propagating virus (Piattelli-Palmarini and Uriagereka, 2004).

REFERENCES

Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., Packer, A., 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol. Autism 4, 36. doi:10.1186/2040-2392-4-36

Ahmad, S.F., Ansari, M.A., Nadeem, A., Bakheet, S.A., Alshammari, M.A., Khan, M.R., Alsaad, A.M.S., Attia, S.M., 2018. S3I-201, a selective Stat3 inhibitor, restores neuroimmune function through upregulation of Treg signaling in autistic BTBR T+ Itpr3tf/J mice. Cell. Signal. 52, 127–136. doi:10.1016/j.cellsig.2018.09.006

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., Hamosh, A., 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 43, D789–D798. doi:10.1093/nar/gku1205

Benítez-Burraco, A., Uriagereka, J., 2016a. The Immune Syntax Revisited: Opening New Windows on Language Evolution. Front. Mol. Neurosci. 8, 84. doi:10.3389/fnmol.2015.00084

Benítez-Burraco, A., Uriagereka, J., 2016b. The Immune Syntax Revisited: Opening New Windows on Language Evolution. Front. Mol. Neurosci. 8, 84. doi:10.3389/fnmol.2015.00084

Binder, J.X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S.I., Schneider, R., Jensen, L.J., 2014. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database (Oxford). 2014, bau012. doi:10.1093/database/bau012

Chatr-aryamontri, A., Breitkreutz, B.-J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., Reguly, T., Nixon, J., Ramage, L., Winter, A., Sellam, A., Chang, C., Hirschman, J., Theesfeld, C., Rust, J., Livstone, M.S., Dolinski, K., Tyers, M., 2015. The BioGRID interaction database: 2015 update. Nucleic Acids Res. 43, D470–D478. doi:10.1093/nar/gku1204

Chuong, E.B., Elde, N.C., Feschotte, C., 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 351, 1083–7. doi:10.1126/science.aad5497

Codagnone, M.G., Spichak, S., O'Mahony, S.M., O'Leary, O.F., Clarke, G., Stanton, C., Dinan, T.G., Cryan, J.F., 2018. Programming Bugs: Microbiota and the Developmental Origins of Brain Health and Disease. Biol. Psychiatry. doi:10.1016/j.biopsych.2018.06.014

Collins, H., Calvo, S., Greenberg, K., Forman Neall, L., Morrison, S., 2016. Information Needs in the Precision Medicine Era: How Genetics Home Reference Can Help. Interact. J. Med. Res. 5, e13. doi:10.2196/ijmr.5199

de Souza, H.S.P., Fiocchi, C., 2016. Immunopathogenesis of IBD: current state of the art. Nat. Rev. Gastroenterol. Hepatol. 13, 13–27. doi:10.1038/nrgastro.2015.186

Deschamps, M., Laval, G., Fagny, M., Itan, Y., Abel, L., Casanova, J.-L., Patin, E., Quintana-Murci, L., 2016. Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. Am. J. Hum. Genet. 98, 5–21. doi:10.1016/j.ajhg.2015.11.014

Dewannieux, M., Heidmann, T., 2013. Endogenous retroviruses: acquisition, amplification and taming of genome invaders. Curr. Opin. Virol. 3, 646–56.

doi:10.1016/j.coviro.2013.08.005

Dresselhaus, E.C., Boersma, M.C.H., Meffert, M.K., 2018. Targeting of NF-κB to Dendritic Spines Is Required for Synaptic Signaling and Spine Development. J. Neurosci. 38, 4093–4103. doi:10.1523/JNEUROSCI.2663-16.2018

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S.M., Blass, T., Somel, M., Brückner, M.K., Schreiweis, C., Winter, C., Sohr, R., Becker, L., Wiebe, V., Nickel, B., Giger, T., Müller, U., Groszer, M., Adler, T., Aguilar, A., Bolle, I., Calzada-Wack, J., Dalke, C., Ehrhardt, N., Favor, J., Fuchs, H., Gailus-Durner, V., Hans, W., Hölzlwimmer, G., Javaheri, A., Kalaydjiev, S., Kallnik, M., Kling, E., Kunder, S., Mossbrugger, I., Naton, B., Racz, I., Rathkolb, B., Rozman, J., Schrewe, A., Busch, D.H., Graw, J., Ivandic, B., Klingenspor, M., Klopstock, T., Ollert, M., Quintanilla-Martinez, L., Schulz, H., Wolf, E., Wurst, W., Zimmer, A., Fisher, S.E., Morgenstern, R., Arendt, T., de Angelis, M.H., Fischer, J., Schwarz, J., Pääbo, S., 2009. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137, 961–71. doi:10.1016/j.cell.2009.03.041

Estes, M.L., McAllister, A.K., 2015. Immune mediators in the brain and peripheral tissues in autism spectrum disorder. Nat. Rev. Neurosci. 16, 469–486. doi:10.1038/nrn3978

Estruch, S.B., Graham, S.A., Quevedo, M., Vino, A., Dekkers, D.H.W., Deriziotis, P., Sollis, E., Demmers, J., Poot, R.A., Fisher, S.E., 2018. Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. Hum. Mol. Genet. 27, 1212–1227. doi:10.1093/hmg/ddy035

Filiano, A.J., Xu, Y., Tustison, N.J., Marsh, R.L., Baker, W., Smirnov, I., Overall, C.C., Gadani, S.P., Turner, S.D., Weng, Z., Peerzade, S.N., Chen, H., Lee, K.S., Scott, M.M., Beenhakker,

M.P., Litvak, V., Kipnis, J., 2016. Unexpected role of interferon-γ in regulating neuronal connectivity and social behaviour. Nature 535, 425–9. doi:10.1038/nature18626

Fisher, S.E., Scharff, C., 2009. FOXP2 as a molecular window into speech and language. Trends Genet. 25, 166–177. doi:10.1016/j.tig.2009.03.002

Green, D.S., Young, H.A., Valencia, J.C., 2017. Current prospects of type II interferon γ signaling and autoimmunity. J. Biol. Chem. 292, 13925–13933. doi:10.1074/jbc.R116.774745

Guo, J., Lin, X., Williams, M.A., Hamid, Q., Georas, S.N., 2008. Yin-Yang 1 regulates effector cytokine gene expression and T(H)2 immune responses. J. Allergy Clin. Immunol. 122, 195–201, 201.e1–5. doi:10.1016/j.jaci.2008.03.012

Hilliard, A.T., Miller, J.E., Fraley, E.R., Horvath, S., White, S.A., 2012. Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. Neuron 73, 537–52. doi:10.1016/j.neuron.2012.01.005

Hoban, A.E., Stilling, R.M., Ryan, F.J., Shanahan, F., Dinan, T.G., Claesson, M.J., Clarke, G., Cryan, J.F., 2016. Regulation of prefrontal cortex myelination by the microbiota. Transl. Psychiatry 6, e774–e774. doi:10.1038/tp.2016.42

Hwang, S.S., Jang, S.W., Kim, M.K., Kim, L.K., Kim, B.-S., Kim, H.S., Kim, K., Lee, W., Flavell, R.A., Lee, G.R., 2016. YY1 inhibits differentiation and function of regulatory T cells by blocking Foxp3 expression and activity. Nat. Commun. 7, 10789. doi:10.1038/ncomms10789

Insel, T.R., 2010. The Challenge of Translation in Social Neuroscience: A Review of Oxytocin, Vasopressin, and Affiliative Behavior. Neuron 65, 768–779.

doi:10.1016/j.neuron.2010.03.005

Ito, J., Sugimoto, R., Nakaoka, H., Yamada, S., Kimura, T., Hayano, T., Inoue, I., 2017. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. PLoS Genet. 13, e1006883. doi:10.1371/journal.pgen.1006883

Johnson, W.E., 2015. Endogenous Retroviruses in the Genomics Era. Annu. Rev. Virol. 2, 135–59. doi:10.1146/annurev-virology-100114-054945

Kaltschmidt, B., Kaltschmidt, C., 2009. NF-kappaB in the nervous system. Cold Spring Harb. Perspect. Biol. 1, a001271. doi:10.1101/cshperspect.a001271

Kelly, J.R., Minuto, C., Cryan, J.F., Clarke, G., Dinan, T.G., 2017. Cross Talk: The Microbiota and Neurodevelopmental Disorders. Front. Neurosci. 11, 490. doi:10.3389/fnins.2017.00490

Konopka, G., Bomar, J.M., Winden, K., Coppola, G., Jonsson, Z.O., Gao, F., Peng, S., Preuss, T.M., Wohlschlegel, J.A., Geschwind, D.H., 2009. Human-specific transcriptional regulation of CNS development genes by FOXP2. Nature 462, 213–7. doi:10.1038/nature08549

Kuleshov, M. V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W., Ma'ayan, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90–W97. doi:10.1093/nar/gkw377

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J.,

Hughes, T.R., Weirauch, M.T., 2018. The Human Transcription Factors. Cell 172, 650–665. doi:10.1016/j.cell.2018.01.029

Li, Y., Missig, G., Finger, B.C., Landino, S.M., Alexander, A.J., Mokler, E.L., Robbins, J.O., Manasian, Y., Kim, W., Kim, K.-S., McDougle, C.J., Carlezon, W.A., Bolshakov, V.Y., 2018. Maternal and Early Postnatal Immune Activation Produce Dissociable Effects on Neurotransmission in mPFC–Amygdala Circuits. J. Neurosci. 38, 3358–3372. doi:10.1523/JNEUROSCI.3642-17.2018

Mattson, M.P., Meffert, M.K., 2006. Roles for NF-κB in nerve cell survival, plasticity, and disease. Cell Death Differ. 13, 852–860. doi:10.1038/sj.cdd.4401837

Meffert, M.K., Baltimore, D., 2005. Physiological functions for brain NF-kappaB. Trends Neurosci. 28, 37–43. doi:10.1016/j.tins.2004.11.002

Meyer-Lindenberg, A., Domes, G., Kirsch, P., Heinrichs, M., 2011. Oxytocin and vasopressin in the human brain: social neuropeptides for translational medicine. Nat. Rev. Neurosci. 12, 524–538. doi:10.1038/nrn3044

Murray, K.N., Edye, M.E., Manca, M., Vernon, A.C., Oladipo, J.M., Fasolino, V., Harte, M.K., Mason, V., Grayson, B., McHugh, P.C., Knuesel, I., Prinssen, E.P., Hager, R., Neill, J.C., 2018. Evolution of a maternal immune activation (mIA) model in rats: Early developmental effects. Brain. Behav. Immun. doi:10.1016/j.bbi.2018.09.005

Nataf, S., 2017. Autoimmunity as a Driving Force of Cognitive Evolution. Front. Neurosci. 11, 582. doi:10.3389/fnins.2017.00582

Nataf, S., 2017b. Autoimmunity as a driving force of cognitive evolution. Front. Neurosci. 11. doi:10.3389/fnins.2017.00582

Nataf, S., 2017a. Evolution, immunity and the emergence of brain superautoantigens. F1000Research 6. doi:10.12688/f1000research.10950.1

NCBI Resource Coordinators, 2014. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 42, D7–D17. doi:10.1093/nar/gkt1146

Nelson, C.S., Fuller, C.K., Fordyce, P.M., Greninger, A.L., Li, H., DeRisi, J.L., 2013. Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. Nucleic Acids Res. 41, 5991–6004. doi:10.1093/nar/gkt259

Neurath, M.F., 2014. Cytokines in inflammatory bowel disease. Nat. Rev. Immunol. 14, 329–342. doi:10.1038/nri3661

Osokine, I., Erlebacher, A., 2017. Inflammation and Autism: From Maternal Gut to Fetal Brain. Trends Mol. Med. 23, 1070–1071. doi:10.1016/j.molmed.2017.10.008

Oswald, F., Klöble, P., Ruland, A., Rosenkranz, D., Hinz, B., Butter, F., Ramljak, S., Zechner, U., Herlyn, H., 2017. The FOXP2-Driven Network in Developmental Disorders and Neurodegeneration. Front. Cell. Neurosci. 11, 212. doi:10.3389/fncel.2017.00212

Parker-Athill, E., Luo, D., Bailey, A., Giunta, B., Tian, J., Shytle, R.D., Murphy, T., Legradi, G., Tan, J., 2009. Flavonoids, a prenatal prophylaxis via targeting JAK2/STAT3 signaling to oppose IL-6/MIA associated autism. J. Neuroimmunol. 217, 20–27. doi:10.1016/j.jneuroim.2009.08.012

Piattelli-Palmarini, M., Uriagereka, J., 2004. The immune syntax: the evolution of the language virus, in: Jenkins, L. (Ed.), Variation and Universals in Bioliguistics. pp. 341–377.

Pletscher-Frankild, S., Pallejà, A., Tsafou, K., Binder, J.X., Jensen, L.J., 2015. DISEASES: Text mining and data integration of disease–gene associations. Methods 74, 83–89. doi:10.1016/j.ymeth.2014.11.020

Ramana, C. V, Gil, M.P., Schreiber, R.D., Stark, G.R., 2002. Stat1-dependent and -independent pathways in IFN-gamma-dependent signaling. Trends Immunol. 23, 96–101.

Rebhan, M., Chalifa-Caspi, V., Prilusky, J., Lancet, D., 1997. GeneCards: integrating information about genes, proteins and diseases. Trends Genet. 13, 163.

Romano, A., Tempesta, B., Micioni Di Bonaventura, M.V., Gaetani, S., 2016. From Autism to Eating Disorders and More: The Role of Oxytocin in Neuropsychiatric Disorders. Front. Neurosci. 9, 497. doi:10.3389/fnins.2015.00497

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S., Hickey, G., Hinrichs, A.S., Hubley, R., Karolchik, D., Learned, K., Lee, B.T., Li, C.H., Miga, K.H., Nguyen, N., Paten, B., Raney, B.J., Smit, A.F.A., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn, R.M., Kent, W.J., 2015. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 43, D670–D681. doi:10.1093/nar/gku1177

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., Ma'ayan, A., 2016. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database 2016, baw100. doi:10.1093/database/baw100

Salles, A., Romano, A., Freudenthal, R., 2014. Synaptic NF-kappa B pathway in neuronal

plasticity and memory. J. Physiol. Paris 108, 256–62. doi:10.1016/j.jphysparis.2014.05.002

Santos, A., Tsafou, K., Stolte, C., Pletscher-Frankild, S., O'Donoghue, S.I., Jensen, L.J., 2015. Comprehensive comparison of large-scale tissue expression datasets. PeerJ 3, e1054. doi:10.7717/peerj.1054

Sarkar, A., Harty, S., Lehto, S.M., Moeller, A.H., Dinan, T.G., Dunbar, R.I.M., Cryan, J.F., Burnet, P.W.J., 2018. The Microbiome in Psychology and Cognitive Neuroscience. Trends Cogn. Sci. 22, 611–636. doi:10.1016/j.tics.2018.04.006

Scharff, C., Petri, J., 2011. Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 366, 2124–40. doi:10.1098/rstb.2011.0001

SIB Swiss Institute of Bioinformatics Members, 2016. The SIB Swiss Institute of Bioinformatics' resources: focus on curated databases. Nucleic Acids Res. 44, D27–D37. doi:10.1093/nar/gkv1310

Sin, C., Li, H., Crawford, D.A., 2015. Transcriptional Regulation by FOXP1, FOXP2, and FOXP4 Dimerization. J. Mol. Neurosci. 55, 437–448. doi:10.1007/s12031-014-0359-7

Snow, W.M., Albensi, B.C., 2016. Neuronal Gene Targets of NF-κB and Their Dysregulation in Alzheimer's Disease. Front. Mol. Neurosci. 9, 118. doi:10.3389/fnmol.2016.00118

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S.C., Fisher, S.E., Ren, B., Geschwind, D.H., 2007. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am. J. Hum. Genet. 81, 1144–57. doi:10.1086/522237

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., Lancet, D., 2016. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses, in: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. 1.30.1-1.30.33. doi:10.1002/cpbi.5

The Gene Ontology Consortium, 2017. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 45, D331–D338. doi:10.1093/nar/gkw1108

Tongyoo, P., Avihingsanon, Y., Prom-On, S., Mutirangura, A., Mhuantong, W., Hirankarn, N., 2017. EnHERV: Enrichment analysis of specific human endogenous retrovirus patterns and their neighboring genes. PLoS One 12, e0177119. doi:10.1371/journal.pone.0177119

Varghese, M., Keshav, N., Jacot-Descombes, S., Warda, T., Wicinski, B., Dickstein, D.L., Harony-Nicolas, H., De Rubeis, S., Drapeau, E., Buxbaum, J.D., Hof, P.R., 2017. Autism spectrum disorder: neuropathology and animal models. Acta Neuropathol. 134, 537–566. doi:10.1007/s00401-017-1736-4

Vernes, S.C., Oliver, P.L., Spiteri, E., Lockstone, H.E., Puliyadi, R., Taylor, J.M., Ho, J., Mombereau, C., Brewer, A., Lowy, E., Nicod, J., Groszer, M., Baban, D., Sahgal, N., Cazier, J.-B., Ragoussis, J., Davies, K.E., Geschwind, D.H., Fisher, S.E., 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet. 7, e1002145. doi:10.1371/journal.pgen.1002145

Vernes, S.C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J.M., Davies, K.E., Geschwind, D.H., Fisher, S.E., 2007. High-Throughput Analysis of Promoter Occupancy Reveals Direct

Neural Targets of FOXP2, a Gene Mutated in Speech and Language Disorders. Am. J. Hum. Genet. 81, 1232–1250. doi:10.1086/522238

Wang, B., Lin, D., Li, C., Tucker, P., 2003. Multiple Domains Define the Expression and Regulatory Properties of Foxp1 Forkhead Transcriptional Repressors. J. Biol. Chem. 278, 24259–24268. doi:10.1074/jbc.M207174200

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., Parkinson, H., 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42, D1001–D1006. doi:10.1093/nar/gkt1229

Xu, S., Liu, P., Chen, Y., Chen, Y., Zhang, W., Zhao, H., Cao, Y., Wang, F., Jiang, N., Lin, S., Li, B., Zhang, Z., Wei, Z., Fan, Y., Jin, Y., He, L., Zhou, R., Dekker, J.D., Tucker, H.O., Fisher, S.E., Yao, Z., Liu, Q., Xia, X., Guo, X., 2018. Foxp2 regulates anatomical features that may be relevant for vocal behaviors and bipedal locomotion. Proc. Natl. Acad. Sci. 115, 8799–8804. doi:10.1073/pnas.1721820115

Zhang, M., Sun, K., Wu, Y., Yang, Y., Tso, P., Wu, Z., 2017. Interactions between Intestinal Microbiota and Host Immune Response in Inflammatory Bowel Disease. Front. Immunol. 8, 942. doi:10.3389/fimmu.2017.00942

Acknowledgments: We thank Marine Guillen from Lyon-1 University (Histology Laboratory, UFR Lyon-Est) for the technical help she provided on internal quality control of bioinformatics analyses.

Conflict of interest declaration: the authors have no conflict of interest to declare regarding the present article

Authors contribution: SN performed the bioinformatics analyses en wrote the paper; ABB and JU wrote the paper.