Cell, Volume *137*

Supplemental Data

A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice

Wolfgang Enard, Sabine Gehre, Kurt Hammerschmidt, Sabine M. Hölter, Torsten Blass, Mehmet Somel, Martina K. Brückner, Christiane Schreiweis, Christine Winter, Reinhard Sohr, Lore Becker, Victor Wiebe, Birgit Nickel, Thomas Giger, Uwe Müller, Matthias Groszer, Thure Adler, Antonio Aguilar, Ines Bolle, Julia Calzada-Wack, Claudia Dalke, Nicole Ehrhardt, Jack Favor, Helmut Fuchs, Valérie Gailus-Durner, Wolfgang Hans, Gabriele Hölzlwimmer, Anahita Javaheri, Svetoslav Kalaydjiev, Magdalena Kallnik, Eva Kling, Sandra Kunder, Ilona Moßbrugger, Beatrix Naton, Ildikó Racz, Birgit Rathkolb, Jan Rozman, Anja Schrewe, Dirk H. Busch, Jochen Graw, Boris Ivandic, Martin Klingenspor, Thomas Klopstock, Markus Ollert, Leticia Quintanilla-Martinez, Holger Schulz, Eckhard Wolf, Wolfgang Wurst, Andreas Zimmer, Simon E. Fisher, Rudolf Morgenstern, Thomas Arendt, Martin Hrabé de Angelis, Julia Fischer, Johannes Schwarz, and Svante Pääbo

Contents

S1 Generation of mice

Cloning the targeting vector. The 3' homology arm, extending 188 bp upstream from exon7 until 1993 bp downstream of exon 8 (chr6:15344452–15348836 in assembly mm8), was amplified by PCR from C57BL/6J genomic mouse DNA and subcloned. During subcloning the two human-specific non-synonymous changes were introduced through PCR mutagenesis (correspond to mutations c15344772a and a15344838g on mouse chromosome 6, resulting in T302N and N324S in mouse FOXP2 (NP_444472)). In addition, a loxP site (ATAACTTCGTATAATGTATGCTATACGAAGTTAT) was inserted in forward orientation between position chr6:15345511 and chr6:15345512, 658 bp downstream from exon7. This modified 3' homology arm was cloned into a vector containing a PGK driven Neomycin resistance cassette flanked by two FRT sites so that *Neo* is transcribed from the opposite strand as *Foxp2*. In this vector, the 5' homology arm (chr6:15341073 – 15344451) was cloned, after it was amplified by PCR from BAC RP23-103L11, subcloned and sequence verified. The primer for the 3' end of the 5' homology arm contained a loxP site in forward orientation to allow for a conditional removal of both exon7 and the neomycin cassette in targeted mice.

Generation of targeted mice. The targeting construct was linearized and electroporated into Bruce4 C57BL/6 embryonic stem (ES) cells (Kontgen et al., 1993) by Ozgene Inc. (Bentley, Australia). Genomic DNA isolated from about 1000 G418-resistant ES cell colonies was digested with EcoRV and analyzed by Southern blotting. One targeted ES cell clone (5C10) was identified and the correct integration was confirmed by Southern blotting using a 5' probe (chr6:15337837-15338336), a 3' probe (chr6:15349263-15349690) and a neomycin resistance cassette probe (Figure S1A). This analysis, as well as the sequencing of a fragment spanning from exon7 to the 3' loxP site, confirmed that the locus was correctly targeted, but also revealed an additional, stronger signal above the expected signal from the targeted chromosome, when using the 3' probe (Figure S1A). This suggested a duplication and rearrangement covering the 3' probe, which was also confirmed in Southern analyses using two other restriction enzymes (data not shown). Chimeras were generated from this clone by blastocyte injection, backcrossed to C57BL/6J and their offspring analyzed by Southern blotting. All 22 offspring carrying the targeted allele also showed the additional signal, suggesting linkage between the targeted allele and the rearrangement (data not shown). Using a quantitative PCR assay, strong linkage was confirmed in subsequent generations, because the targeted $Foxp2^{hum}$ allele and the 3' rearrangement cosegregated in all 93 mice (corresponding to 183 meioses) analyzed. However, mRNA and protein expression levels and mRNA structure of *Foxp2* seem to be unaffected by these changes (Figure S1). A second targeted ES cell clone (5H11) was identified among 260 additional G418-resistant ES cell colonies and did not show the duplication/rearrangement in Southern blot analyses (Figure S2A). Chimeric mice generated from this clone were crossed to Cre deleter mice (Ozgene) to generate founders carrying the $Foxp2^{ko}$ allele and founders carrying the $Foxp2^{hum}$ allele (Figure S2A). Offspring from the $Foxp2^{hum}$ founders were crossed to mice transgenic for the recombinase FLPe under the control of the human *ACTB* promoter ((Rodriguez et al., 2000); Jackson Laboratory, Stock Number 003800; C57BL/6J) to generate a $Foxp2^{hum}$ allele in which the FRT-flanked neomycin resistance cassette had been removed ($Foxp2^{hum\Delta neo}$). The recombinase transgenes were outcrossed in the next generation.

For all colony expansions, mice were bred to C57BL/6J (Harlan Winkelmann, Borchen, Germany) and the offspring of heterozygous parents were used for all analyses except for neuroanatomical comparisons at P1 and P20. Mice were housed at a 10h/14h dark/light cycle under standard conditions. All animal work was performed in accordance with governmental and institutional guidelines.

Genotyping assays. The *Foxp2*^{hum} allele was genotyped using a PCR assay containing one forward primer located in the neomycin cassette, one forward primer located in the 3' end of the 5' homology arm and one reverse primer located downstream of exon7. This results in an amplification of a 601 bp fragment from the *Foxp2*^{wt} allele and a 710 bp fragment from the *Foxp2*^{hum} allele. Details on this and the other PCR assays used are available upon request. Mice described phenotypically were generally genotyped in at least two independent assays.

Expression levels and structure of the *Foxp2*^{hum} mRNA allele To investigate the effects of the *Foxp2*^{hum} allele on mRNA levels and mRNA structure of *Foxp2*, we analyzed *Foxp2* expression in the cerebrum, cerebellum and lung of two heterozygous male and two heterozygous females (line 5H10) by RT-PCR and sequencing. Peak ratios of both human-specific substitutions gave no indication that the expression levels of the *Foxp2*^{wt} and *Foxp2*^{hum} allele differed in any of the three tissues (Figure S1B). Furthermore, neither qualitative analysis of RT-PCR products (with primers located in different exons) spanning the entire coding region from heterozygous and wildtype animals nor sequencing the entire *Foxp2* coding region from cerebral cDNA from a heterozygous and a wildtype animal, indicated any change in the *Foxp2* mRNA structure due to the genetic changes of the *Foxp2*^{hum} allele (data not shown). Hence, it is unlikely that the phenotypic effects of the *Foxp2*^{hum} allele are caused by changes in mRNA levels or mRNA structure of *Foxp2*, which makes it less likely that the loxP sites, the neomycin cassette and the 3' rearrangement cause any of the observed *Foxp2*^{hum} –dependent effects.

Immunoblotting. 20µg of proteins or an protein equivalent of 1mg of embryonic brain were separated using precast gels (Criterion Tris-HCl , 10.5 – 14 %, Bio-Rad, Munich, Germany) and blotted on an Immun-Blot® PVDF Membran (Bio-Rad) using a Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad). The antibodies HPA000382 (1:2500, Atlas Antibodies, Stockholm, Sweden) and sc21069 (1:750 diluted, Santa Cruz Biotechnology, Santa Cruz, USA) were used to detect Foxp2 and anti-actin C4 (1:3000, monoclonal, MP Biomedicals, Strasbourg, France) was used to detect beta-actin. These were detected using rabbit IgG HRP-Linked F(ab')2 (1:10000, GE Healthcare) for HPA000382, donkey anti goat IgG-HRP (1:10000, Santa Cruz Biotechnology) for sc21069 and mouse IgG, HRP-Linked Whole Ab(1:10000, GE Healthcare) for anti-actin C4 and the ECL-Plus detection system (GE Healthcare) on a Typhoon scanner (GE Healthcare). Magic Mark XP Western Standard (Invitrogen) was used as a molecular weight marker (M).

S2 Mouse Clinic screen *Foxp2*$^{hum/hum}$ mice

Housing conditions and workflow.

Mice from 17 litters, all born within one week from crossings of parents heterozygous for *Foxp2*^{hum} (5C10 line), were genotyped independently twice, and 15 mice of each genotype and sex were selected as evenly across litters as possible. Five mice were grouped per cage, independent of their genotype and shipped at six weeks of age to the German Mouse Clinic. There, they were housed on wood fibre (Altromin, Lage, Germany) in type II polycarbonate cages in individually ventilated caging systems operating under positive pressure (VentiRack Bioscreen TM, Biozone, Margate, UK). Mice were transferred in weekly intervals to new cages with forceps in Laminar Flow Class II changing stations. Mice were fed the irradiated standard rodent high energy breeding diet (Altromin 1314) and given semidemineralized, filtered (0.2 µm) water *ad libitum*. Light was adjusted to a 12h/12h light/dark cycle; temperature and relative humidity were regulated to 22 ± 1°C and 55 ± 5%, respectively. In specified modules, husbandry conditions were adjusted according to the experiment requirements (see corresponding sections). All people attending the facility completely changed their garments (jackets and trousers autoclaved) and shoes, and wore caps and masks before entering the GMC (Brielmeier et al., 2002). Outbred 8-week-old male SPF Swiss mice were used as sentinels and housed on a mixture of new bedding and aliquots of soiled bedding (50:50) from all cages on the rack. In addition, the sentinels were also exposed to soiled air from all cages on the rack. Health monitoring was performed by on-site examination of the sentinel mice by certified laboratories according to FELASA recommendations (www.felasa.org). Mice were kept in accordance with German legal requirements. Tests were performed with the approval of the Regierung von Oberbayern.
Mouse mutants entering the GMC were examined in a primary screen according to a standard workflow (Gailus-Durner et al., 2005). After the mice arrived at the German Mouse Clinic, they were acclimatized in the new environment for one week. The males then started the Behavior Screen. They remained there for three weeks. Immediately following the behavior tests, the anatomical inspection of the Dysmorphology Screen was performed. In the next week, the Neurology Screen was completed. One week later, the mice underwent the Eye Screen. When the mice were 12 weeks old, blood was taken, and samples were distributed to the blood-based screens for Clinical Chemistry, Immunology and Allergy. The following week, the animals were tested in the Nociceptive Screen. One week later, the mice were passed to the Cardiovascular Screen, where the mice remained for two weeks. In parallel, 10 mutant animals (five males / five females) and 10 controls (five males / five females) left the animal facility for the Lung Function Analysis, which for technical reasons is located elsewhere. Organs of male mice were frozen for future molecular phenotyping. All other animals underwent the bone and cartilage tests of the Dysmorphology Screen. After the Dysmorphology Screen the remaining mice were transferred for three weeks to the Metabolic Screen. After completion of the primary screen, all animals were analyzed macro- and microscopically in the Pathology. The screening of female animals began one week later and followed the same workflow with the exception that the females used for the Lung Function Screen went straight to pathology afterwards. If not otherwise stated (Table S1), data were analyzed using an ANOVA with sex and genotype as factors. Significant differences between genotypes are

indicated stepwise from 0.05, 0.01 to 0.001. Measurements showing significant differences between genotypes are summarized in Table S1 and S4. Summaries of all measurements and/or raw data are available upon request and on www.europhenome.org.

Morphology.

At six weeks of age, i.e. when the mice entered the facility, their general physical condition and health were checked. At the age of nine weeks, a morphological observation as a whole-body checkup was performed on all 60 mice using the protocol for morphological analysis from (Fuchs et al., 2000), as adapted for the German Mouse Clinic. Using a clickbox (supplied by the MRC Institute of Hearing Research, Nottingham, UK) we tested their ability to detect a sound at 20 kHz. The animals' reaction was classified into six categories (0=no reaction, 1=no Preyer reflex, 2= retarded reaction, 3= normal reaction, 4= strong reaction, 5= particularly strong reaction).
At 11 weeks, the eyes of all 60 mice were examined. The posterior parts of both eyes were examined by funduscopy. After pupil dilation with one drop of atropine (1%), the mouse was grasped firmly in one hand and clinically evaluated using a head-worn indirect ophthalmoscope (Sigma 150 K, Heine Optotechnik, Herrsching, Germany), in conjunction with a condensing lens (90D lens, Volk, Mentor, OH, USA) mounted between the ophthalmoscope and the eye. Mice were also examined biomicroscopically for eye abnormalities as previously described (Favor, 1983). Briefly, pupils were dilated with a 1% atropine solution applied to the eyes at least 10 min prior to examination. Both eyes of each mouse were examined by slit lamp biomicroscopy (Zeiss SLM30) at 48x magnification with a narrow beam slit lamp illuminated at a 25-30° angle from the direction of observation. Observed phenotypic variants in the eyes were carefully documented.
At the age of 16 to 17 weeks, X-ray analysis (10 males and 10 females of each genotype) and bone densitometry (10 males of each genotype, 10 *Foxp2*$^{wt/wt}$ and 9 *Foxp2*$^{hum/hum}$ females) were performed on anesthetized mice. X-ray images were taken using a Faxitron X-ray Model MX-20 (Specimen Radiography System, Illinois, USA) and a NTB Digital X-ray Scanner EZ 40 (NTB GmbH, Diepholz, Germany). Using iX-Pect software supplied by the manufacturer of the X-ray scanner, the image was taken and analyzed. Analysis was completed qualitatively by visually inspecting the images, as well as quantitatively using the ruler tool of iX-Pect software.
Bone density analysis was done using a pDEXA Sabre X-ray Bone Densitometer (Norland Medical Systems. Inc., Basingstoke, Hampshire, UK; distributed by Stratec Medizintechnik GmbH, Pforzheim, Germany) and analysis was completed on the entire body, including and excluding the skull.

Blood parameters.

Blood samples were collected from ether-anesthetized mice (isoflurane for *Foxp2*$^{wt/ko}$ screen) at 12 weeks of age by puncturing the retro-orbital sinus with a non-heparinized capillary (0.8 mm in diameter; Laborteam K&K; Munich, Germany; Art.No. 1.28.13.1.2). 500µl of blood was collected from each mouse in a heparinized tube (Li-heparin, KABE; Nümbrecht, Germany; Art.No. 078028). An additional, smaller sample (~ 50µl) was collected (using the same capillary) in an EDTA-coated tube (KABE, Art.No 078035). Each tube was immediately inverted five times to distribute the anticoagulant homogenously. The Li-heparin-coated tubes were stored in a rack at room temperature

for two hours. Afterwards, cells and plasma were separated in a centrifugation step (10 min, 9503 x g; Biofuge, Heraeus; Hanau, Germany). Plasma was distributed between the Immunology Screen (30 µl), the Allergy Screen (30 µl) and the Clinical Chemical Screen (130 µl), while the cell pellet was given to the Immunology Screen for FACS-analysis. Two female mice (*Foxp2*$^{hum/hum}$) did not recover from narcosis.

The plasma sample for the clinical chemical analysis was transferred into an Eppendorf tube and diluted 1:2 with aqua dest. The solution was mixed for a few seconds (Vortex genie, Scientific Industries; New York, USA) to prevent clotting and then centrifuged again for 10 min at 4656 x g. Additionally, the Clinical Chemical Screen received the EDTA-blood sample for hematological investigations.

The clinical chemistry screen was performed using an Olympus AU 400 autoanalyzer and adapted reagents from Olympus (Hamburg, Germany).

A volume of 50 µl EDTA-blood was used to measure basic hematological parameters with a blood analyzer, validated for the analysis of mouse blood (ABC-Blutbild-Analyzer, Scil Animal Care Company GmbH, Viernheim, Germany). The number and size of red blood cells, white blood cells and platelets were measured by electrical impedance, and hemoglobin by spectrophotometry. Mean corpuscular volume (MCV), mean platelet volume (MPV) and red blood cell distribution width (RDW) were calculated directly from the cell volume measurements, and the hematocrit (HCT) from MCV * red blood cell count. Mean corpuscular hemoglobin (MCH) and the mean concentration of corpuscular hemoglobin (MCHC) were calculated from the hemoglobin/red blood cell count (MCH) and hemoglobin/hematocrit (MCHC).

Plasma IgE concentrations were determined from 30µl of plasma, essentially as described (Alessandrini et al., 2001). The plasma levels of IgM, IgG1, IgG2a, IgG2b, IgG3, and IgA were determined from 30µl by standard sandwich ELISAs using goat anti-mouse immunoglobulin antibodies and alkaline phosphatase (AP) conjugates (SouthernBiotech, Birmingham, USA). The presence of rheumatoid factor and anti-DNA antibodies was evaluated by indirect ELISA with rabbit IgG (Sigma-Aldrich, Steinheim, Germany) and calf thymus DNA (Sigma-Aldrich), respectively, as antigens and with AP-conjugated goat anti-mouse secondary antibody (Sigma-Aldrich). Serum samples from MRL/MpJ-Tnfrsf6lpr mice (Jackson Laboratory, Bar Harbor, ME, USA) were used as positive controls in the autoantibody assays.

Peripheral blood leukocytes (PBLs) were isolated from those cells previously isolated from the 500 µl Li-heparin blood by erythrocyte lysis with NH_4Cl (0.17M) - Tris buffer (pH 7.45), and placed directly onto 96-well microtiter plates. After subsequent washing with FACS staining buffer (PBS, 0.5% BSA, 0.02% sodium azide, pH 7.45), PBLs were incubated for 20 min with 1 µM ethidium monazide bromide (EMA, Molecular Probes, The Netherlands) and Fc block (clone 2.4G2, PharMingen, San Diego, CA, USA). EMA bound to the DNA of dead cells was photocrosslinked by brief light exposure. Cells were then stained with fluorescence-conjugated monoclonal antibodies (PharMingen). The following main cell populations were analyzed: B cells ($CD19^+$ clone 1D3), B1 B cells ($CD19^+CD5^+$, clone 53-7.3), B2 B cells ($CD19^+CD5^-$), T cells ($CD3^+$, clone 145-2C11), $CD4^+$ T cells (clone RM4-5), $CD8^+$ T cells (CD8α, clone 53-6.7; CD8β, clone H35-17.2), γ/δT cells (clone GL3), granulocytes ($Gr\text{-}1^+$, clone RB6-8C5), and NK cells ($CD49b^+$, clone DX5). We also analyzed additional subpopulations based on the following surface antigens: IgD (clone 11-26c.2a), B220 (clone RA3-6B2), CD11b (clone M1/70), CD103

(clone 2E7), CD25 (clone PC61), CD62L (clone MEL-14), CD45RB (clone 16a), Ly-6C (clone AL-21), and CD44 (clone IM7). Data were acquired on a FACS LSR II (BectonDickinson, San Diego, CA, USA) and were analyzed using FlowJo software (TreeStar Inc, USA). Samples were collected until a total number of 25,000 cells was reached.

Energy metabolism

26 animals (7 males of each genotype, 7 *Foxp2*$^{wt/wt}$ females and 5 *Foxp2*$^{hum/hum}$ females) were single caged on grid panels (0.5°cm grid hole diameter) and fed *ad libitum* for 14 days. Body weight, food consumption, rectal temperature, daily feces production and energy content in both feces and food were recorded. Feces were collected in three day intervals. Samples of lab food and feces (~1 g) were dried at 60°C for two days, homogenized in a coffee grinder and then squeezed into pill form for determining energy content in a bomb calorimeter (IKA Calorimeter C7000). Energy uptake was determined as the product of food consumed and the caloric value of the food. To calculate the amount of metabolized energy, the energy content of feces was subtracted from energy uptake, as well as 2% of the energy uptake to account for energy loss due to urination (Drozdz, 1975). To test the effect of the factors genotype and sex, an ANOVA with body weight as covariate was performed (with body weight as covariate for mass specific energy assimilation).

Lung function

Five males and five females of each genotype were studied at the age of 15 weeks using a commercially available system from Buxco® Electronics (Sharon, Connecticut), essentially as described (Drorbaugh and Fenn, 1955). It measures the pressure changes that arise from inspiratory and expiratory temperature and humidity fluctuations during breathing. Calibration of the system allows transforming these pressure swings into flow and volume signals so that tidal volumes, respiratory rates, minute ventilation, inspiratory and expiratory times, as well as peak inspiratory and peak expiratory flow rates can be measured. Measurements were taken between 8 a.m. and 11 a.m. for 40 minutes each. Respiratory parameters were determined separately for the active animal (mostly occurring at the beginning of the measurement when the chamber is explored by the animal) and the resting animal.

Cardiovascular function

The mice began the Cardiovascular Screen at the age of 14 weeks; 19 female mice (10 *Foxp2*$^{wt/wt}$, 9 *Foxp2*$^{hum/hum}$) and 20 male mice (10 each) underwent blood pressure and electrocardiogram (ECG) analysis. Blood pressure was measured in conscious mice with a non-invasive tail-cuff method using the MC4000 Blood Pressure Analysis Systems (Hatteras Instruments Inc., Cary, North Carolina, USA). Four animals were restrained on a pre-warmed metal platform in metal boxes. The tails were looped through a tail-cuff and fixed to a notch containing an optical path with an LED light and a photo sensor. The blood pulse wave in the tail artery was detected and transformed into an optical pulse signal by measuring light extinction. Pulse detection, cuff inflation and pressure evaluation were automated by the system software. After five initial inflation runs for habituation, 12 measurement runs were performed for each animal in one session. Runs with movement artifacts were excluded. After one day of training, in which the animals

were habituated to the apparatus and protocol, the measurements were taken on four consecutive days between 8:30 and 11:30 AM. 20 to 48 individual measurements were pooled to obtain a mean over the four measurement days for each animal. Data were analyzed by two-way ANOVA with sex and genotype as fixed factors. ECG was performed in anesthetized (isoflurane/pressured air inhalation) mice using three metal bracelets placed on the joints of the feet together with electrode gel. The complete setup was situated in a faraday cage. The electrodes were positioned on the front-paws and the left hind-paw, resulting in the bipolar standard limb leads I, II and III and the augmented unipolar leads AVF, AVR, AVL. ECG was recorded for about seven minutes. A shape analysis of the ECG traces was performed with the software ECG-auto (EMKA technologies, Paris, France). For each animal, intervals and amplitudes were evaluated from five different sets of averaged beats (lead II). The parameter Q-T interval was also corrected for the RR interval (QTc). In addition, the recordings were visually screened for arrhythmias, including supraventricular and ventricular extrasystoles and conduction blockages.

We found the R wave amplitude to be lower ($p<0.05$) in *Foxp2*$^{hum/hum}$ females and non-significantly so ($p>0.05$) in *Foxp2*$^{hum/hum}$ males. The R wave is part of the QRS complex that reflects the excitable ventricular mass. However, the values were within normal ranges and the QRS complex showed no other alterations. Hence, the origin of the difference is not clear and might reflect a spurious significant result rather than a biological effect of *Foxp2*^{hum} on heart function.

Neurology

Animals were assessed at 10 weeks of age, essentially as described (Schneider et al., 2006). First, undisturbed behavior was observed in an 11 cm diameter glass cylinder. The mice were then transferred to an arena consisting of a clear plexiglass box (42 x 26 x 18 cm) in which a plexiglass sheet on the floor was marked with 15 squares. Locomotor activity and motor behavior within this area was observed and followed by a sequence of manipulations testing reflexes and grip strength. The grip strength meter system (TSE, Bad Homburg, Germany) determined the fore limb grip strength by exploiting the tendency of a mouse to grasp a horizontal metal bar while being pulled by its tail. For each mouse five trials were undertaken within one minute. Measurements were completed with the recording of body weight. Throughout the entire procedure, abnormal behavior, biting, defecation and vocalization were also recorded.

The TSE-RotaRod 3375 apparatus (Accelerating Model, TSE, Bad Homburg) was used to measure motor coordination, balance and motor learning ability. The machine was configured in an environment with minimal stimuli, such as noise and movement. Mice were placed perpendicular to the axis of rotation, with head facing the direction of the rotation, and the rotarod accelerated from 4 to 40 rpm for 300 sec, with a 20 minute interval between each trial. In motor coordination testing, mice were given four trials at the accelerating speed on one day. The mean latency for falling off the rotarod during the trials was recorded. Before the start of the first trial, mice were weighed. For testing motor learning, any improvement in mouse performance occurring over three consecutive days was measured. All experimental equipment was thoroughly cleaned with Pursept-A and dried prior subsequent tests. Statistical analyses were completed as described (Schneider et al., 2006).

Nociception was tested in 59 mice (14 *Foxp2*$^{hum/hum}$ females, 15 *Foxp2*$^{wt/wt}$ females and 15 males of each genotype) using a hot-plate assay. A mouse was placed on a 28 cm diameter metal surface maintained at 52+0.2°C surrounded by a 20 cm high Plexiglass wall (TSE, Bad Homburg, Germany). Mice remained on the plate for 60 seconds, or until they performed one of three behaviors regarded as indicative of nociception: hind paw licking, hind paw shake/flutter or jumping. For each of these three behaviors, the latency was compared for sex and genotype using an ANOVA.

Behavior

The *modified Hole Board* test (Ohl et al., 2001) was performed according to the procedures described (Barrantes Idel et al., 2006). The test apparatus (Figure S3B) consisted of a box (150 x 50 x 50 cm) divided into an experimental compartment (100 x 50 cm) and a group compartment (50 x 50 cm) by a transparent PVC partition (50 x 50 x 0.5 cm) with 111 holes (1 cm diameter) staggered in 12 lines to allow group contact. A board (60 x 20 x 2 cm) with 23 holes (1.5 x 0.5 cm) staggered in 3 lines with all holes covered by motile lids was placed in the middle of the test arena. The area around the board was divided into 12 similarly sized quadrants by lines taped onto the floor of the box. Both box and board were made of dark grey PVC. All lids were closed before the start of a trial. For each trial, an unfamiliar object (a round blue plastic tube lid, diameter approx. 2 cm) and a familiar object of similar size (solid, brass hexagon, diameter and height approx. 2 cms), were placed into the test arena, in the corner quadrant diametrical to the starting point, with a distance of 2 cm between them. A copy of the familiar object had been placed in the animals' home cages 3 days earlier and removed one day before testing. The illumination levels were set at approximately 150 lux in the corners and 200 lux in the middle of the test arena.

At the beginning of the experiment, all animals from one cage were habituated together in the group compartment for 20 min. Then each animal was placed individually into the experimental compartment and allowed to explore it freely for 5 min, during which the cage mates remained in the group compartment. During the 5 min trial, the animal's behavior was recorded by a trained observer blind to the genotype with a hand-held computer. Behaviors scored included line crossings, rearings, board entries, hole exploration, group contact, object exploration, grooming, defecation and immobility. Data were analyzed with the Observer 4.1 Software (Noldus, Wageningen, Netherlands). Additionally, a camera was mounted 1.20 m above the center of the test arena, the animal's track was videotaped and its locomotor path analyzed with a video-tracking system (Ethovision 2.3, Noldus, Wageningen, Netherlands). After each trial, the test animal was returned to the group compartment until all animals from one cage had been tested. The test arena was then cleaned carefully with a disinfectant before starting with the next animal.

Acoustic startle reflex and prepulse inhibition were assessed during the same session for 57 mice (15 males of each genotype, 14 *Foxp2*$^{wt/wt}$ females and 13 *Foxp2*$^{hum/hum}$ females). Experiments were performed between 08:30AM and 5:00PM. Background noise was 65 dB, and startle pulses were bursts of white noise (40 msec). A session was initiated with a 5 min acclimation period, followed by five presentations of leader startle pulses (110 dB), which were excluded from statistical analysis. Trial types included seven different startle stimulus intensities (70, 80, 85, 90, 100, 110, 120 dB) and one NS (no stimulus) trial, in which only the background noise was present to determine the baseline movement of the

animal. Trial types for the PPI included four different prepulse intensities (67, 69, 73, 81 dB); each prepulse preceded the startle pulse (110 dB) by a 50 msec inter stimulus interval. Each trial type for ASR and PPI was presented 10 times in random order, organized in 10 blocks, with each trial type occurring once per block. ITIs varied from 20-30 sec. This protocol is based on the Eumorphia protocol (see www.eumorphia.org) adapted to the specifications of our startle equipment (Med Associates, St. Albans, VT, USA). Data were analyzed by two-way analysis of variance (ANOVA) with sex and genotype as fixed factors.

Pathology

Five 15 week old females of each genotype and another 38 animals, 22-23 weeks of age (11 *Foxp2*$^{wt/wt}$ females, 7 *Foxp2*$^{hum/hum}$ females and 10 males of each genotype), were weighed and analyzed macroscopically (http://eulep.anat.cam.ac.uk/Necropsy_of_the_Mouse/). For 23 of these (7 *Foxp2*$^{wt/wt}$ females, 5 *Foxp2*$^{hum/hum}$ females, 5 *Foxp2*$^{wt/wt}$ males and 6 *Foxp2*$^{hum/hum}$ males) a histological analysis was performed. To this end, organs (skin, bone, heart, lung, cerebrum, cerebellum, trachea, salivary gland, esophagus, thymus, spleen, cervical lymph nodes, thyroid, parathyroid, adrenal gland, stomach, intestine, liver, pancreas, kidney, reproductive organs and urinary bladder) were fixed in 4% buffered formalin, embedded in paraffin, cut into 2µm thick sections and stained with hematoxylin and eosin. Particular attention was paid to Purkinje cell abnormalities, since these are affected in *Foxp2* knock-out mice (Fujita et al., 2008; Shu et al., 2005), but no abnormalities were found. Non-genotype-specific results included a proliferation of subcapsular cells in the adrenal cortex (9/23), microgranulomas in the liver (8/23) and a dilated urinary bladder (7/23).

Second batch of *Foxp2*$^{hum/hum}$ mice

Mice from 15 litters, all born on the same day from crossings of parents heterozygous for a *Foxp2*^{hum} allele in which the FRT-flanked neomycin resistance cassette had been removed (derived from the 5H11 line as described in section S1) were genotyped independently twice and 15 mice of each genotype and sex were selected as evenly across litters as possible. These mice were shipped and housed as described in S2 and behavior on the modified hole board and electrocardiogram were measured as described in S2. Whereas the measurements for the electrocardiogram hardly differ between genotypes in this new batch of mice (Table S2), measurements on the modified hole board (Table S2), including the significant longer latency to enter the hole board indicate a more cautious exploratory behavior for *Foxp2*$^{hum/hum}$ mice as observed in the first batch. In none of the cases did we observe a significant genotype*batch interaction (Table S3), indicating that the effects of the *Foxp2*^{hum} allele are independent of either the neomycin cassette or the ES cell clone in which *Foxp2*^{hum} was generated. Means of all factors significant for the factor genotype are displayed in Figure S3.

In addition, these mice were also tested in an elevated plus-maze and a light-dark box after the test on the modified hole board had been finished. The *elevated plus-maze* apparatus was made out of light grey PVC and consisted of 4 arms (each arm: 5 cm wide, 30 cm long) arranged in the form of a plus sign around a centre platform (5 X 5 cm), and elevated 75 cm above the ground on a pole underneath the centre platform. Two opposite arms had 15 cm high walls (closed arms), the other two arms (open arms) had an 0.3 cm

rim. The illumination level was 100 Lux on the centre platform. For testing, a mouse was placed at the end of a closed arm (distal to the centre) facing the wall, and allowed to explore the apparatus for 5 min. A camera was mounted above the centre of the apparatus and arm and centre entries of the mice were scored via a video monitor by a trained observer blind to the genotype in an adjacent room. The entry criterion was the placement of all 4 paws in an arm or the centre.

The *light/dark* test box was made of PVC and divided into two compartments, connected by a small tunnel (4 × 6 x 9 cm high). The lit compartment (29 × 19 × 24 cm high) was made of white PVC and was illuminated by cold light with an intensity of 650 lux in the middle, the dark compartment (14 × 19 × 24 cm high) was made of black PVC and not directly illuminated (appr. 20 lux in the centre). The mouse was placed in the centre of the black compartment and allowed to freely explore the apparatus for 5 min. Behaviours were observed by a trained observer sitting next to the box using a hand-held computer. Data were analyzed with respect to (1) the number of entries, latency to first entry, and time spent in both compartments and the tunnel; (2) the number of rearings in both compartments and the tunnel. Additionally, grooming behavior was recorded. An entry into a compartment was defined as placement of all four paws into the compartment. Additionally, a camera was mounted above the center of the test arena to videotape the trial.

In each test, the apparatus was cleaned carefully with a disinfectant after every mouse before starting with the next animal. Observation data were analyzed with the Observer 4.1 Software, and video-tracking data were analyzed using the Ethovision 2.3 Software (both from Noldus, Wageningen, Netherlands).

We used SPSS v.13.0 (SPSS Inc, Chicago, IL USA) for all statistical analyses.

S3 Mouse Clinic screen *Foxp2*$^{wt/ko}$ mice

Mice from 12 litters, all born within 3 days from crossings of *Foxp2*$^{wt/ko}$ mice with C57BL/6J mice, were genotyped twice and 30 females (13 *Foxp2*$^{wt/wt}$, 17 *Foxp2*$^{wt/ko}$) and 30 males (19 *Foxp2*$^{wt/wt}$ and 11 *Foxp2*$^{wt/ko}$) were selected across litters as evenly as possible. These mice were essentially screened as described for the *Foxp2*$^{hum/hum}$ mice in Section S2, except that the mice were anesthetized with isoflurane for blood sample collection and a second blood sample was taken at 17 weeks of age to test the robustness of differences obtained for immunological and chemical parameters at 12 weeks. Further, the acoustic startle reflex and prepulse inhibition were not assessed. Results and the number of animals measured for each parameter are shown in Table S4.

In order to compare *Foxp2*$^{hum/hum}$ mice and *Foxp2*$^{wt/ko}$ mice with respect to their behavior on the modified hole board, we combined data from all 157 mice (27 *Foxp2*$^{hum/hum}$ (12 females, 15 males) and 27 wildtype littermates (12 females, 15 males) from the first batch (line 5C10), 28 *Foxp2*$^{hum/hum}$ (15 females, 13 males) and 27 wildtype littermates (15 females, 12 males) from the second batch (line 5H11 *Foxp2*$^{hum\Delta neo}$), 23 *Foxp2*$^{wt/ko}$ (13 females, 10 males) and 25 wildtype littermates (15 females, 10 males) from the second batch (line 5H11 *Foxp2*^{hum}deltaNeo). We analyzed the data using an ANOVA with the

factors strain (levels *Foxp2*^{ko} and *Foxp2*^{hum}), genotype (levels mutated or wildtype) and sex for all variables (partly log, square-root or rank transformed to meet equality of variance assumptions). Measurements for which a strain*genotype interaction is significant are thus influenced differently in *Foxp2*$^{wt/ko}$ mice and *Foxp2*$^{hum/hum}$ mice. Results of this analysis are summarized in Table S5 and all significant strain*genotype interactions are displayed in Figure 2.

S4 Immunohistochemistry

Adults, P10 pups and embryos analyzed immunohistochemically came from the same litter, whereas pups at postnatal day one (P1) and P20 were derived from crossings of homozygous parents. P10 pups, P20 pups and adult animals were anesthetized with CO_2 and transcardially perfused first with saline (0.9% NaCl) for 1–2 min and then with a fixative containing 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4) for 30 min. Brains were removed and postfixed for 24h in the same solution. Mothers pregnant 16.5 day after overnight matings were anesthetized with CO_2 and the heads of the pups were fixed by immersion for 24h-48h in a fixative containing 4% paraformaldehyde in 0.1 M phosphate buffer (pH 7.4). For cryosections, samples were equilibrated with 30% sucrose for two to three days and sectioned with a freezing microtome coronally and/or sagittally in 12µm (P1, P10) or 30µm (P20, adults) slices. 12µm slices were immediately mounted on slides and all samples were stored at -20° until use. Samples embedded in paraffin were cut at 7µm (Jung Histoslide 2000, Leica, Nussloch, Germany*)* and mounted on coated slides (Superfrost-Plus, Menzelgläser, Braunschweig). Animals were treated following the guidelines set by the Panel on Euthanasia and the Guide for the Care and Use of Laboratory Animals. Protocols were approved by the Ethical Committee of Leipzig University. 30µm cryosections were incubated free-floating in Tris-buffered saline (TBS, 0.01M Tris, 0.145 M NaCl, pH adjusted to 7.4 with HCl) at room temperature. Sections were exposed for 10 min to 3% H_2O_2 in TBS, treated with blocking solution of 3% normal goat serum (dianova, Hamburg, Germany), rinsed once with TBS and then incubated overnight with an antibody. The FOXP2 antibodies HPA000382 (derived from an internal 150 aa recombinant peptide; Atlas Antibodies, Stockholm, Sweden) and ab16046 (Abcam plc, Cambridge, UK; epitope is synthetic peptide derived from the last 15 aa of FOXP2) produced identical results in one parallel staining. All were used at a dilution of 1:1000 (adults) and 1:2000 (E 16.5 – P20) in TBS. Sections were rinsed several times with TBS-T (TBS with 0.1% Triton X-100 (Sigma-Aldrich, Steinheim, Germany), incubated for 1h with a biotinylated secondary antibody (goat anti-rabbit IgG, dilution 1:1000, dianova, Hamburg, Germany), rinsed several times with TBS-T and incubated for 1h with Extravidin-POD (Sigma-Aldrich, 1:1000). After final rinsing with TBS, the reaction was visualized by incubating 5-10 min with 0.2mg/ml 3,3'-diaminobenzidine (Sigma-Aldrich)/ 0.015% H_2O_2 (Merck, Darmstadt, Germany) in TBS. Sections were mounted and coverslipped with Entellan in toluene (Merck). Some adult and P20 sections were used for double-immunofluorescence staining with antibodies to FOXP2 and Calbindin D-28k (monoclonal, Swant, Bellinzona, Switzerland, 1:2000). The secondary antibodies (dianova, Hamburg, Germany) used were the red fluorescent Cy3-goat-anti-rabbit IgG (1:1000) and green fluorescent Cy2-goat-anti-mouse

IgG (1:500). Subsequently, sections were washed with TBS, mounted on fluorescence-free slides, air-dried and coverslipped with Entellan (Merck). For staining mounted 12µm cryosections, slides were adapted to room temperature, fixed for 10 minutes in 4% paraformaldehyde, rinsed several times with TBS, treated with 3% H_2O_2 in 70% methanol for 20 min and then processed as described for cryosections above. Paraffin sections were dewaxed by immersion in Xylol and ethanol of decreasing concentration (absolute, 96%, 80%, 70%, 50%), washed with TBS, treated with 3% H_2O_2 in 70% methanol for 20 min and then also processed as described above. Control experiments omitted the primary antibody in the given protocol. Slides were examined using Zeiss Axiophot and/or Zeiss Axiovert 200M and Axiovision software (Zeiss, Jena, Germany). Regions were identified according to Paxinos and Franklin (1997).

S5 Neurotransmitter measurements

10 *Foxp2*$^{hum/hum}$ males and 10 male wildtype littermates (19-25 weeks of age) and 10 *Foxp2*$^{wt/ko}$ and 10 male littermates (22 weeks of age) were used to determine neurotransmitter levels. Mice were decapitated, their brains removed within seconds, frozen in -30° cold isopentane and stored at -80°. Frozen coronal sections (0.5-1 mm) were prepared at the following bregma coordinates (Paxinos and Franklin, 1997): anterior-posterior (+2.34)-(+1.34), (+1.34)-(+0.84), (-0.2)-(-0.7), and (-6.5)-(-7.0) and stored at –80°C. For neurochemical analyses, tissue samples of 5 different brain regions (frontal cortex, caudate-putamen, nucleus accumbens, lateral globus pallidus and the cerebellum) from both hemispheres were quickly isolated by a punch of 1 mm in diameter. Tissue samples from each area were homogenized by ultrasonication in 250 or 500 µl in 0.1 N perchloric acid at 4°C immediately after processing. 100 µl of the homogenates were added to equal volumes of 1 N sodium hydroxide for measuring protein content. The remaining homogenates were centrifuged at 17,000 g and 4 °C for 10 min. Aliquot of the supernatants were added to equal volumes (20 µl) of 0.5 M borate buffer and stored at -80° for subsequent amino acid analyses. The remaining supernatants were used for immediate measurement of monoamines (dopamine and serotonin [5-hydroxytryptamine; 5-HT]) by high performance liquid chromatography (HPLC) with electrochemical detection as previously described (Felice et al., 1978; Sperk, 1982; Sperk et al., 1981). Glutamate and GABA were determined as described previously (Piepponen and Skujins, 2001). Briefly, amino acids were precolumn derivatized with o-phthalaldehyde-2-mercaptoethanol using a refrigerated autoinjector and then separated on an HPLC column (ProntoSil C18 ace-EPS) at a flow rate of 0.6 ml/min and a column temperature of 40°C. The mobile phase was 50 mM sodium acetate pH = 5.7 in a linear gradient from 5% to 21% acetonitrile. Derivatized amino acids were detected by their fluorescence at 450 nm after excitation at 330 nm. To assess an overall significance between genotypes within a single strain we performed for each neurotransmitter separately an ANOVA with brain region as repeated measure and genotype as factor (using SPSS v.13.0). We combined the data from both strains and used an ANOVA with brain regions as repeated measure and the factors genotype (with levels wildtype and mutated (i.e. *Foxp2*$^{wt/ko}$ or *Foxp2*$^{hum/hum}$)) and strain (with levels *Foxp2*^{ko} and *Foxp2*^{hum}). A significant interaction between genotype and strain then indicates a significant

difference between *Foxp2*$^{hum/hum}$ and *Foxp2*$^{wt/ko}$ mice. Table S6 displays results of this analysis, as well as p-values of Student's t-tests performed separately for each neurotransmitter and brain region.

S6 Neuronal cell culture

Striatal neural precursor cells (NPCs) were dissected from ~ E14 old embryos. Pregnant females were sacrificed according to NIH guidelines and the approval of the local animal care committee. Tissue samples were incubated in 0.1 mg/ml papain (Roche, Mannheim, Germany)/ DNase solution (100 µg/ml, Roche) for 30 min at 37°C, rinsed in PBS, incubated in antipain (50 µg/ml; Roche) for 30 min at 37°C, and finally homogenized by gentle triturating using a fire-polished Pasteur pipette. Cells were expanded as a three-dimensional (neurosphere) culture in serum-free DMEM (high glucose)/F-12 mixture (1:1) medium, supplemented with both 20 ng/ml of human recombinant epidermal growth factor (EGF) and 20 ng/ml basic fibroblast growth factor (bFGF; both from PromoCell, Heidelberg, Germany). NPCs were expanded in normal atmospheric oxygen. During the five week proliferation period, cell survival rate was determined weekly for each culture using trypan blue exclusion (Invitrogen, Carlsbad, CA USA) and after 3 weeks of proliferation using propidium iodide FACS analysis (Nicoletti et al., 1991), but no differences between genotypes were found (data not shown). After four weeks of proliferation, differentiation of NPCs was induced for one week via replacement of expansion media by defined media without mitogens but with 1% FCS and 5 µM forskoline (Sigma-Aldrich Chemie GmbH Munich, Germany) on precoated poly L-lysine dishes. Cells were stained with DAPI (Roche Diagnistics GmbH, Mannheim, Germany) and antisera against the neural marker beta III tubulin (Tuj-1; PRB-435P, rabbit polyclonal, Covance research products, Inc, Denver, USA). For three cultures per genotype we quantified neurite length of five neurons each using the NeuronJ plugin (Meijering et al., 2004) of Image J (Abràmoff et al., 2004). All analyzed differentiated cultures expressed FOXP2 as shown by immunoblotting using the FOXP2 antibody ab16046 (Abcam plc, Cambridge, UK).
Total neurite length between genotypes was compared by a Mann-Whitney U test using SPSS v.13.0 software (SPSS Inc, Chicago, IL, USA).

S7 Electrophysiology

Coronal slices (300 µM thick) containing the striatum were prepared from the brains of 15- to 27-day-old *Foxp2*$^{hum/hum}$ mice and their wildtype littermates (all derived from the 5H11 line in which the Neomycin cassette had been removed as described above). Slices were superfused with an external Ringer solution containing (in mM): 125 NaCl, 2.5 KCl, 2 CaCl2, 1 MgCl2, 26 NaHCO3, 1.25 NaH2PO4, and 20 glucose, bubbled with 95% O2/5% CO2. The pH was adjusted to 7.3 – 7.4 and the osmolarity to 305 mOsM. We did not suppress GABAA receptor currents, since we were recording at a holding potential of – 70 mV close to the chloride reversal potential. All recordings were performed at room temperature. Medium spiny neurons were identified by their

morphology and characteristic electrophysiological properties including negative resting membrane potentials and slow capacitance transients. To confirm the phenotype of recorded neurons, we labeled some cells (n =5) using 200 µM Oregon-green-BAPTA1 (Invitrogen) and imaged these cells using a confocal microscope (Olympus Fluoview FV1000). All imaged cells showed the characteristic morphology of medium spiny neurons (Figure S7D). For analyses, cells were accepted if the leak current was ≤ 100 pA and the resting membrane potential below – 70 mV to ensure homogenous recordings. Glass electrodes (5-8 MΩ) were filled with a solution containing (in mM): 150 K-gluconate, 10 NaCl, 3 Mg-ATP, 0.3 GTP, 10 Hepes and 0.05 EGTA, adjusted to pH 7.3 and an osmolarity of 305 mOsM. The liquid junction potential of 15 mV was corrected post hoc. Excitatory medium spiny neuron afferents were stimulated with a glass electrode filled with external saline and placed between the recorded medium spiny neuron and cortex, typically ~50-100 µm from the cell body using current clamp mode. Stimulus intensity was adjusted to yield largest EPSC amplitudes without eliciting an action potential or inducing direct stimulation. Typically, we used 100 – 200 µA for a duration of 0.1 s. Continuous stimulation was performed at a frequency of 0.5 Hz. Following 20 min. of baseline stimulation, we applied 3 tetani of 3 s duration and a frequency of 100 Hz separated by 30 s. LTD was then measured for 30 min. Data points were calculated every 30 seconds by averaging the last 15 EPSCs (0.5 Hz). All recordings were performed using an Multiclamp amplifier (Axon Instruments), filtered at 2 kHz and digitized at 10 kHz. Acquisition and analysis were performed using custom Clampex 9 software.
LTD data was analyzed using a repeated measure ANOVA of all measured timepoints before and after high-frequency stimulation using SPSS 13.0. There was a significant interaction between time and genotype ($F(3.5,48.8)=7.51$; $p=0.04$; Greenhouse-Geisser corrected) and also a significant effect of the factor genotype ($F(1,14)=7.51$, $p=0.016$).

S8 Gene expression analysis

Sample preparation.

Embryos (E16.5) were dissected and briefly washed in ice-cold dissection buffer (15mM HEPES (hemisodium salt, Research Organics, Cleveland, Ohio, USA), 30mM glucose (Sigma-Aldrich, Steinheim, Germany), 1x Hanks balanced Salt solution (without Calcium, Magnesium, Phenol Red, Sodium Bicarbonate; Invitrogen, Carlsbad, California, USA) and 1.34 mM NaHCO3 (Sigma-Aldrich); sterile filtered). The brain of each embryo was then removed, dissected under a stereomicroscope in RNAlater (Ambion, Austin, Texas, USA) and stored in RNAlater until all brains of the litter had been prepared. For preparation of the striatum, the cerebellum was removed, the brain cut into the two hemispheres and the striatum (i.e. caudate nucleus, putamen and pallidum) was prepared from each hemisphere according to (Schambra et al., 1992). The striata of both hemispheres were combined and stored in RNAlater at -20° until further use. Tailbuds of the embryos were used for genotyping and sex determination (Lambert et al., 2000). We picked individuals as balanced across sex and litter as possible: The following samples were used (genotype_sex_litter): wt/wt_F_243, wt/wt_F_243, hum/hum_F_243, hum/hum_F_243, wt/wt_M_243, hum/hum_M_243, hum/hum_M_256,

hum/hum_M_256, wt/wt_M_256, hum/hum_M_256, wt/wt_M_256, wt/wt_M_261, hum/hum_M_261, wt/wt_F_261. A second batch of embryos was prepared in a similar fashion: hum/hum_F_309, hum/hum_F_309, hum/hum_F_309, hum/hum_F_304, hum/hum_F_308, hum/hum_F_308, wt/wt_M_284, wt/wt_M_284, wt/wt_M_304, wt/wt_F_304, wt/wt_F_208. Striatal tissue from adult females was prepared in a similar fashion (see also (Zapala et al., 2005) for a description). All embryonic and adult samples were derived from crosses of *Foxp2*$^{wt/hum}$ animals (5C10 line). Striatal samples from young animals P15, P18 and P21) were prepared slightly different in that brains were cut in ice-cold PBS by a custom-build razorblade holder that allowed making 2 mm thick coronal sections. 2-3 slices were used to obtain striatal biopsy punches each 2 mm in diameter, which were then frozen in liquid nitrogen. We picked individuals so that genotypes were balanced as much as possible for sex, litter and run in the two different "strains" ("Ki" for pups from crosses of *Foxp2*$^{wt/hum}$ animals (5H11 line with the Neomycin cassette removed(*Foxp2*$^{hum\Delta neo}$)) and "Ko" for pups from crosses of *Foxp2*$^{wt/ko}$ and C57BL/6J animals): The following 35 samples were used (strain_genotype_day_sex_run_litter): Ki_hum/hum_15_F_3_5, Ki_hum/hum_15_F_2_5, Ki_wt/wt_15_F_1_6, Ki_wt/wt_15_M_2_5, Ki_hum/hum_15_M_1_6, Ki_hum/hum_15_M_3_6, Ki_wt/wt_18_F_3_1, Ki_hum/hum_18_F_1_1, Ki_hum/hum_18_F_2_4, Ki_wt/wt_18_M_3_1, Ki_hum/hum_18_M_2_1, Ki_hum/hum_18_M_1_4, Ki_wt/wt_21_F_2_2, Ki_hum/hum_21_F_1_2, Ki_hum/hum_21_F_3_3, Ki_wt/wt_21_M_1_2, Ki_hum/hum_21_M_3_2, Ko_wt/ko_15_F_1_8, Ko_wt/wt_15_F_1_8, Ko_wt/ko_15_F_3_8, Ko_wt/ko_15_M_2_9, Ko_wt/ko_15_M_1_10, Ko_wt/wt_15_M_3_11, Ko_wt/ko_18_F_2_12, Ko_wt/wt_18_F_2_12, Ko_wt/ko_18_M_2_7, Ko_wt/ko_18_M_3_12, Ko_wt/wt_18_M_1_12, Ko_wt/ko_21_F_3_10, Ko_wt/ko_21_F_2_10, Ko_wt/wt_21_F_3_10, Ko_wt/ko_21_F_1_12, Ko_wt/ko_21_M_3_10, Ko_wt/ko_21_M_1_7/10 (pool of two inds), Ko_wt/wt_21_M_2_10.

For RNA isolation, the tissue was transferred to 1ml of TRIZol reagent (Invitrogen) and immediately homogenized using a teflon pestle (cylindrical) in a 2ml glass homogenizer on a Schütt homgenplus homogenizer (Schütt Labortechnik, Göttingen, Germany). RNA and genomic DNA were isolated according to the TRIZol protocol with addition of 250µg Glycogen (Roche Diagnostics, Mannheim, Germany) to 1ml of homogenate. DNA was used to confirm genotypes and RNA was further purified using the MiniElute kit (Qiagen, Valencia, California, USA). High and equal quality RNA was ensured on an Agilent 2100 Bioanalyzer system (Foster City, California, USA). Labeling of 3.5 µg (embryonic striatum), 3.3 µg (adult striatum) and 1µg (young animlas) of total RNA, hybridization to Affymetrix Mouse Genome 430 2.0 arrays, staining, washing and array scanning were carried out following Affymetrix protocols. All primary expression data are publicly available at the NCBI GEO database (accession number GSE13588).

Data analysis.

All data preparation and analysis was conducted in the R statistical environment (Gentleman et al., 2004). Gene expression levels were extracted and summarized from the 71 Affymetrix .CEL files using the MG-Mm430 Ensembl Custom CDF file based on mouse Ensembl genes (Dai et al., 2005). We used the "rma" algorithm (R-Bioconductor

"affy" package; (Gautier et al., 2004) for background correction and quantile normalization of log base 2 transformed of expression levels. Genes with expression levels above background were determined based on the Wilcoxon test for detection, conducted with the "mas5" function (R-Bioconductor "affy" package) using a cutoff of $p < 0.05$. Among the 15,768 Ensembl genes present on the array, 11,329 had detectable expression levels among 11 or more arrays and these 11,329 genes were used for all further analyses. We first normalized the data for the factors *Run, Sex* and *Age* using multiple regression. *Run* has 6 levels (two embryonic data sets, three runs of the P15-P21 set and one adult data set) and *Age* is the number of days post conception. We then compared the 30 *Foxp2*$^{hum/hum}$ mice and their 23 *Foxp2*$^{wt/wt}$ littermates using an F-test and found 34 genes significant at a p-value threshold of 0.001. In 1000 permutations of *Genotype* labels we found on average (median) 4 genes at that cut-off and 34 or more genes only in 36 of the 1000 permutations (i.e. permutation test p-value = 0.036). When analyzing embryos, young animals and adults separately, the number of genes significant at this threshold were 9 (in permutations: median = 6, p = 0.360), 11 (5, p=0.233) and 7 (3, p=0.302), respectively. Hence, there is a non-significant excess of genes in each age group and a significant excess of significant genes when all age groups are combined. To take into account the direction of change, we calculated the effect size of genotype differences using Cohen's *D*. We chose 0.8 as a cut-off for a high effect size (Cohen, 1992) and 74 genes that are lower expressed in *Foxp2*$^{hum/hum}$ mice (KiEYA_dn) and 133 genes that are higher expressed in *Foxp2*$^{hum/hum}$ mice (KiEYA_up) meet this criterion.

We tested whether these genes are enriched in (i) functional categories annotated in the Gene Ontology (The_Gene_Ontology_Consortium, 2000), (ii) genes identified as Foxp2 targets in a human cell line system (Vernes et al., 2007), (iii) genes containing putative Foxp2 binding sites in their promoters and (iv) genes enriched in either D1 or D2 positive medium spiny neurons (Heiman et al., 2008). For (i) we used gene annotations (August 2008) obtained from Ensembl Biomart (http://www.ensembl.org/biomart/martview/) as an input for the hypergeometric test implemented in FUNC (Prufer et al., 2007). This allows testing each functional category for an overrepresentation of genes compared to all other annotated genes with detectable expression and computes an overall enrichment p-value (Prufer et al., 2007). The most enriched ($p<0.01$) biological function categories are for KiEYA_up genes: GO:0051028 (mRNA transport; 4 out of 109 genes in the group are KiEYA_up genes; 0.6 expected; hypergeometric test p=0.003); GO:0045449 (regulation of transcription; 31 out of 109; 19.2 expected; p=0.003); GO:0048193 (Golgi vesicle transport; 4 out of 109; 0.7 expected; p=0.005); GO:0051252 (regulation of RNA metabolic process; 28 out of 109; 18 expected; p=0.009). For KiEYA_dn genes: GO:0051056 (regulation of small GTPase mediated signal transduction; 6 out of 55; 1.1 expected; p=0.005); GO:0007268 (synaptic transmission; 5 out of 55; 1.2 expected; p=0.007). We then analyzed 300 data sets in which the genotype label had been permuted, but found the same or lower overall enrichment p-value for KiEYA_up genes (p=0.5) and Ki_EYA_dn genes (p=0.7), indicating no significant enrichment of KiEYA_up and KiEYA_dn genes in known functional categories. (ii) We used data from (Vernes et al., 2007), who used promoter arrays with 5,790 represented genes to identify 303 potential FOXP2 targets in a human neuroblastoma cell line. We used Ensembl Biomart to map 3,656 of the 5,790 human genes to mouse orthologs of which 2,453 had detectable expression levels in our data set. Among the 112 FOXP2 targets one and two

genes overlap with KiEYA_up genes and KiEYA_dn genes, respectively (permutations: 1 expected on average, p=0.8 and 0.2, respectively). (iii) To determine whether there is an association between putative Foxp2 binding motifs and the observed gene expression patterns, we collected the promoter sequence (5kbp upstream and 2kbp downstream of the transcription start site) of all genes with an annotated 5'UTR from Ensembl Biomart. The number of each motif (and its reverse complement) matching the consensus sequence TATTTRT shown to repress FOXP1 and FOXP2-dependent transcription in vitro (Wang et al., 2003) and its subsequence ATTT (and its reverse complement) used for crystallizing FOXP2 (Stroud et al., 2006) was counted. For each motif we compared its mean number in the promoters of KiEYA_up and KiEYA_dn genes and compared these means to the 1000 permutations. We found no enrichment in KiEYA_dn genes (p>0.8 for all six motifs), but found a significant enrichment in KiEYA_up genes (tatttat: average occurrence=2.4, average occurrence in permutations=1.8, permutations p-value= 0.02; ataaata: 1.6, 1.5, 0.36; tatttgt: 1.1, 0.8, 0.05; acaaata: 1, 0.7, 0.03; aaat: 53.3, 40.7, 0.02; attt: 58.4, 42.7, 0.01). (iv) In Table S2 of Heiman et al. are Affymetrix probesets (Mouse Genome 430 2.0 arrays) listed as differently expressed in striatonigral (D1 positive) medium spiny neurons and striatopallidal (D2 positive) medium spiny neurons. We used Ensembl Biomart to map these probesets to Ensembl gene identifiers and classified among our 11,329 genes with detectable expression levels 137 as D1 and 52 as D2 enriched. None of the KiEYA_up and none of the KiEYA_dn genes overlapped with D2-enriched genes (median in permutations = 0, p-value =1) and 2 of the KiEYA_up and 4 of the KiEYA_dn genes overlapped with D1-enriched genes (median in permutations = 1 and 0, respectively, p-value =0.4 and 0.06, respectively). This indicates that there is no major shift in the proportion of striatonigral and striatopallidal medium spiny neurons in *Foxp2*$^{hum/hum}$ mice, but indicates a tendency for a higher proportion of effects in striatonigral medium spiny neurons (see also below).

When we compared gene expression in the 12 *Foxp2*$^{wt/ko}$ and their 6 *Foxp2*$^{wt/wt}$ littermates (P15-P21) using an F-test as above, we found 6 genes significant at a p-value threshold of 0.001 (permutations: median = 5, p-value = 0.5). Hence, the effect on striatal gene expression is similar in both *Foxp2*$^{wt/ko}$ and *Foxp2*$^{hum/hum}$ mice (see above). *Foxp2* is the gene that is most strongly affected in *Foxp2*$^{wt/ko}$ mice (F-test p-value=0.00002), showing a reduction of expression in *Foxp2*$^{wt/ko}$ mice to 68% of that in their *Foxp2*$^{wt/wt}$ littermates that is in agreement with the reduction to 75% seen by quantitative PCR in embryonic brains (Figure S2). This shows that relevant signals can be detected in the data and that the effects on striatal gene expression in *Foxp2*$^{wt/ko}$ and *Foxp2*$^{hum/hum}$ mice are rather subtle.

To test whether gene expression is affected in a similar or dissimilar way in *Foxp2*$^{wt/ko}$ and *Foxp2*$^{hum/hum}$ mice, we correlated the effect size across genes within the young (P15-P21) animals. The negative rank correlation of -0.18 (Spearman rank correlation) indicates that gene expression patterns tend to be affected in opposite directions, although not significantly so since more negative correlations are found in 20% of the permutations. Since we found opposite tendencies also for other phenotypes, we nevertheless explored the properties of the two gene groups that are upregulated (absolute effect size >0.8) in *Foxp2*$^{hum/hum}$ mice and downregulated in *Foxp2*$^{wt/ko}$ mice (64 KiYup_KoYdn genes) and vice versa (106 KiYup_KoYdn genes). Among the tests (i) – (iv) described above we find a significant effect (p< 0.05) compared to permutations only

for KiYup_KoYdn genes that are enriched in striatonigral (D1 positive) medium spiny neurons (24 of the 106 overlap, 1 expected from permutations, p=0.002).

S9 Vocal Behavior

To generate pups to be recorded for isolation calls, heterozygous parents (2 females, one male per cage) were mated and cages with pregnant females were checked daily for litters (these were the same matings set up for the Mouse Clinic screen described in Note S2). The first day that a litter was detected was considered as postnatal day zero (P0). For each recording at P4, P7, P10 and P13, the cage with the litter to be measured was taken under a clean bench and a pup was selected randomly and placed on a cotton pad in a plastic beaker (6 cm high, 6.2 cm in diameter for P4-P7and 6.7 cm high and 8.5 cm in diameter for P10-P13). The pup was weighed and then placed in a soundproof plexiglas box (custom made) with an ultrasound microphone situated 6 cm above the bottom of the plastic beaker. The microphone (UltraSoundGate CM16) was connected to a preamplifier (UltraSoundGate 116), which was connected to a computer using the recording software Avisoft SASLab Pro v4.33c set to a sampling frequency of 300 kHz (hardware and software from Avisoft Bioacoustics, Berlin, Germany). The clean bench was closed, switched off and the pup was recorded for 2 minutes (P13: 3 minutes). For identification, pups were marked (Microtattoo System, Fine Science Tools, Heidelberg, Germany) after the measurement at P4, and a tail clip was taken for genotyping after the measurement at P13. Altogether, recordings for all four days were made from 137 pups originating from 13 litters (6 to 15 pups per litter), comprised of 16, 34, 15 females and 24, 31 and 17 males of the genotype *Foxp2*$^{wt/wt}$, *Foxp2*$^{wt/hum}$ and *Foxp2*$^{hum/hum}$, respectively.

We counted the number of calls per recording session with an AVISOFT Recorder 2.97 (Avisoft Bioacoustics, Berlin). We used the following recorder settings: energy level = 0, cut off < 35 kHz and entropy < 35 %. First, we visually inspected all recordings to ensure that the automated sampling routine selected only calls of mouse pups and no other sounds, such as toe clicking. The Avisoft recorder stores the selected sounds in separate wave files, and, in addition, logs the time of call onset. We used these data to calculate the distribution of call intervals. For the detailed acoustic analysis, we increased the recorder's energy level setting to 2. Second, we visually selected the calls chosen from the automated selection routine to ensure that all calls had sufficient amplitude throughout the call. From the selected calls, we calculated spectrograms (frequency range: 150 kHz, frequency resolution: 293 Hz, time resolution: 0.21 ms). We submitted the resulting spectrograms to the custom software program LMA 2005 to extract a set of acoustic parameters (see Table S7). As mice typically concentrate the energy of their calls in one small frequency band, so-called 'pure tone-like sounds' (see Figure S9B), we focused on the peak frequency, i.e. the loudest frequency. For each call we then determined the mean, start, end, maximum and minimum peak frequency, the location of the maximum frequency within the call, as well as the greatest difference [Hz] in peak frequency between two consecutive 0.21 ms bins. We also calculated the slope of the call from the start to the end peak frequency, as well as the slope of a linear trend through the peak frequencies of a call. To capture information describing the modulation of calls, we calculated the global modulation by averaging the absolute difference

between the peak frequency of a 0.21 ms bin and the linear trend. We also calculated the local modulation by averaging the absolute difference between the peak frequency of a 0.21ms bin and a curve over the entire call, which is smoothed using a three point floating average.

For statistical comparison, we classified the mouse calls into four different categories. The first two call categories (call type 1 and 2) included all calls that contained no or only minor frequency jumps (< 50% of the frequency value before the jump with a mean jump size around 3 kHz). A change point analysis was used to distinguish between short and long calls: call type 1 was comprised of all calls < 50 ms (N=5234; 57 % of all calls), call type 2 all calls ≥ 50 ms (N=612; 7 % of all calls), with 50 ms being the change point of the distribution of call duration of all calls. The third category (call type 3) was comprised of all calls with frequency jumps that exceeded 50 % of the frequency value before the jump (jump size > 25 kHz; N=2242; 25 % of all calls). Call category 4 consisted of the various remaining sounds, which were not analyzed (11% of all sounds). Either these could not be unambiguously identified as sound from the mouse pups, or consisted of some rare sounds that were uttered occasionally from a few subjects. None of these sounds were observed more frequently in one of the two genotypes. To eliminate an overrepresentation of subjects with high vocal activity, we randomly selected a maximum of 10 calls per subject and recording day. We then calculated for each parameter the mean value per subject and recording day. To test for differences between genotypes, we used a general linear mixed model (SPSS 13.0), with days (P4, P7, P10 and P13) as within-subject factor and genotype, sex, and litter as fixed between-subject factors. To control for weight differences, weight was included as a covariant. To correct for multiple testing of the various acoustic measurements, we first used Fisher's omnibus test as a global test (test statistic: -2 * (sum ln P_i), df = 2*N P_i). Since this global test indicated a significant effect, we report uncorrected P-values (Table S7). We also performed an ANOVA on call parameters with sex, genotype and day as fixed factor, i.e. treated calls independently and obtained similar results (Table S7). Since only very few subjects uttered call types 2 and 3 at all days, we analyzed genotype effects for parameters of these call types by calculating the mean value per subject and day and using an ANOVA with day, sex and genotype as between subject factors and weight as a covariate (Table S9). The complete dataset is available upon request.

A second batch of animals, derived from parents heterozygous for a *Foxp2*^{hum} allele derived from the 5H11 line (Supplemental Data S1), was recorded and analyzed as described above for call type 1. Except for local modulation, all parameters showed the same tendency between genotypes as before (Table S8), but were not significantly different, probably due to the smaller number of individuals analyzed (17 *Foxp2*$^{hum/hum}$ and 11 *Foxp2*$^{wt/wt}$). However, when calls were analyzed in an ANOVA with sex, genotype and day as fixed factor, i.e. were treated independently, the same parameters (except for local modulation) were clearly significantly different between genotypes (Table S8). Hence, the second batch confirms the findings of the first batch that *Foxp2*^{hum} influences the structure of isolation calls.

Supplemental Tables

Supplemental Table S1: German Mouse Clinic screen of *Foxp2*$^{hum/hum}$ mice. All significant differences between genotypes. The full table is available upon request. Data available at www.europhenome.org.

Screen	Parameter	Male wt/wt	Fem. wt/wt	Male h/h	Fem. h/h	Sex[a]	Gen.[b]	Test	N[c]
m.holeboard_observer	Line crossing [frequency]	130.87 ± 4.92	127.92 ± 5.65	119.6 ± 5.11	106.5 ± 5.63	n.s.	<0.01	ANOVA	15/12/15/12
m.holeboard_observer	Hole exploration [frequency]	63.8 ± 4.52	73.5 ± 4.48	48.67 ± 4.89	60.17 ± 5.77	<0.05	<0.01	ANOVA	15/12/15/12
m.holeboard_observer	Board entry [frequency]	9.87 ± 1.1	14.67 ± 1.46	7.53 ± 1.19	9.75 ± 1.16	<0.01	<0.01	ANOVA	15/12/15/12
m.holeboard_observer	Unfamiliar object exploration [frequency]	8.13 ± 0.74	7.58 ± 0.92	6.27 ± 0.61	6.17 ± 0.64	n.s.	<0.05	ANOVA	15/12/15/12
m.holeboard_ethovision	Total distance moved [cm]	3361.94 ± 98.59	3540.27 ± 131.53	3038.91 ± 116.04	2933.49 ± 130.58	n.s.	<0.001	ANOVA	15/12/15/12
m.holeboard_ethovision	Mean velocity [cm/sec.]	19.93 ± 0.52	19.68 ± 0.47	18.37 ± 0.43	17.64 ± 0.64	n.s.	<0.01	ANOVA	15/12/15/12
m.holeboard_ethovision	Maximum velocity [cm/sec.]	60.38 ± 1.65	60.97 ± 2.85	55.61 ± 2.33	55.06 ± 3.51	n.s.	<0.05	ANOVA	15/12/15/12
m.holeboard_ethovision	Turns [frequency]	1780.87 ± 36.25	1887.58 ± 46.95	1697.07 ± 45.37	1675.67 ± 36.64	n.s.	<0.001	ANOVA	15/12/15/12
m.holeboard_ethovision	Mean distance to wall [cm]	7.56 ± 0.26	8.33 ± 0.24	6.66 ± 0.34	7.81 ± 0.32	<0.01	<0.05	ANOVA	15/12/15/12
m.holeboard_ethovision	Mean distance to board [cm]	8.32 ± 0.17	8.03 ± 0.16	9.13 ± 0.26	8.34 ± 0.17	<0.05	<0.01	ANOVA	15/12/15/12
Neurology_arena	Locomotor activity: lines crossed	23.3± 1.2	28.3± 1.9	19.1± 1.1	23.9± 2.0		<0.05	ANOVA	15/15/15/15
Heart; ECG	R amplitude [mV]	3.09± 0.15	3.34± 0.13	2.69± 0.11	2.82± 0.25	n.s.	<0.01	ANOVA	10/10/10/9

[a] significance for factor sex; when two values are given, tests were performed separately for *Foxp2*$^{wt/wt}$ and *Foxp2*$^{hum/hum}$ mice; n.s. represents a p-value >0.05; empty cells were not explicitly tested; [b] significance for factor genotype [c] Number of individuals analyzed: males *Foxp2*$^{wt/wt}$/ females *Foxp2*$^{wt/wt}$ / males *Foxp2*$^{hum/hum}$ / females *Foxp2*$^{hum/hum}$.

Supplemental Table S2: *Foxp2*$^{hum/hum}$ results, second batch (line 5H11). All significant differences between genotypes and all parameters explored in Table S1. The full table is available upon request. Data available at www.europhenome.org.

Screen	Parameter	Male wt/wt	Fem. wt/wt	Male h/h	Fem. h/h	Sex[a]	Gen.[b]	N
m.holeboard_ observer	Line crossing (frequency)	102±3 .78	106.4 ±4.01	91.77 ±4.75	102±8 .13	n.s.	n.s.	12/15/1 3/15
m.holeboard_ observer	Hole exploration (frequency)	38.17 ±4.76	39.87 ±3.09	34±5. 34	29.87 ±3.64	n.s.	n.s.	12/15/1 3/15
m.holeboard_ observer	Board entry (frequency)	5.42± 0.96	6.73± 0.98	4.92± 1	4.33± 0.77	n.s.	n.s.	12/15/1 3/15
m.holeboard_ observer	Board entry (latency)	91.7± 22.3	54.29 ±8.21	114.5 7±25. 82	143.4 3±23. 73	n.s.	<0.01	12/15/1 3/15
m.holeboard_ observer	Unfamiliar object exploration (frequency)	9.08± 0.78	6.27± 0.99	7.77± 0.64	6.67± 1.27	n.s.	n.s.	12/15/1 3/15
m.holeboard_ ethovision	Total distance moved (cm)	2681. 51±88 .62	2705. 78±10 1.64	2398. 65±12 8.88	2566. 49±18 4.89	n.s.	n.s.	12/15/1 3/15
m.holeboard_ ethovision	Mean velocity (cm/sec.)	17.64 ±0.36	17.82 ±0.35	16.25 ±0.68	17.13 ±0.89	n.s.	n.s.	12/15/1 3/15
m.holeboard_ ethovision	Maximum velocity (cm/sec.)	56.25 ±1.72	56.39 ±1.96	56.33 ±3.64	53.47 ±2.28	n.s.	n.s.	12/15/1 3/15
m.holeboard_ ethovision	Turns (frequency)	1518. 25±38 .06	1523. 87±50 .88	1410. 46±46 .52	1442. 07±80 .24	n.s.	n.s.	12/15/1 3/15
m.holeboard_ ethovision	Mean distance to wall (cm)	6.8±0. 28	7.19± 0.33	6.91± 0.39	6.71± 0.27	n.s.	n.s.	12/15/1 3/15
m.holeboard_ ethovision	Mean distance to board (cm)	8.9±0. 19	8.55± 0.26	8.71± 0.3	8.71± 0.26	n.s.	n.s.	12/15/1 3/15
ECG	R amplitude [mV]	2.45± 0.11	2.74± 0.12	2.37± 0.14	2.7±0. 17	<0.05	n.s.	14/15/1 5/15

[a] significance for factor sex in a two-way ANOVA with sex and genotype as factors;
[b] significance for factor genotype; note that there was no significant interaction between genotype and sex for any parameter;

Supplemental Table S3: ANOVA results of the two batches of *Foxp2*$^{hum/hum}$ mice assessed on the modified hole board. All significant differences between genotypes and all parameters explored in Table S1 and S2. The full table is available upon request.

Parameter[a]	gen.	sex	batch	gen.* batch	gen.* sex	sex* batch	gen.*sex* batch
Total distance moved (cm)	<0.001	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Mean velocity (cm/sec.)	<0.001	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Maximum velocity (cm/sec.)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Turns (frequency)	<0.01	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Absolute meander (Sqrt degrees/sec.)	<0.05	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Board entry (rank latency)	<0.05	n.s.	<0.05	n.s.	n.s.	n.s.	n.s.
Mean distance to wall (cm)	<0.05	<0.05	<0.01	n.s.	n.s.	n.s.	n.s.
Mean distance to board (cm)	n.s.	<0.05	n.s.	n.s.	n.s.	n.s.	n.s.
Line crossing (frequency)	<0.01	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Hole exploration (frequency)	<0.01	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Board entry (frequency)	<0.01	<0.05	<0.001	n.s.	n.s.	<0.05	n.s.
Unfamiliar object exploration (sqrt total duration %)	n.s.	n.s.	<0.01	n.s.	n.s.	n.s.	n.s.

[a] some parameters were transformed (log(ln), square root (sqrt) or rank) in order to meet the equality of error variance assumption;

Supplemental Table S4: German Mouse Clinic screen of *Foxp2*$^{wt/ko}$ mice. All significant differences between genotypes and parameters explored in Table S1 and S2. The full table is available upon request. Data available at www.europhenome.org.

Screen	Parameter	Male wt/wt	Fem. wt/wt	Male wt/ko	Fem. wt/ko	Sex[a]	Gen.[b]	Test	N[c]
modified hole board_observer	Line crossing (frequency)	115.5 ±6.64	121.31±8.08	94.3± 5.29	106.73±4.46	n.s.	p<0.01	ANOVA	10/13/10/15
modified hole board_observer	Hole exploration (frequency)	48.1± 5.16	54.54 ±4.8	58.4± 5.13	55.47 ±4.72	n.s.	n.s.	ANOVA	10/13/10/15
modified hole board_observer	Board entry (frequency)	9.2±1.26	10.62 ±1.31	10.1± 1.34	10.27 ±1.26	n.s.	n.s.	ANOVA	10/13/10/15
modified hole board_observer	Board entry (latency)	46.97 ±7.9	58.17 ±15.18	53.33 ±15.5	44.13 ±3.74	n.s.	n.s.	ANOVA	10/13/10/15
modified hole board_observer	Group contact (frequency)	24.4± 2.32	25.08 ±3.54	17.5± 1.88	18.87 ±1.7	n.s.	p<0.05	ANOVA	10/13/10/15
modified hole board_observer	Unfamiliar object exploration (frequency)	8.2±1.32	3.54± 0.37	6.8±0.9	5.4±0.64	p<0.001	n.s.	ANOVA	10/13/10/15
modified hole board_observer	Unfamiliar object exploration (total duration %)	1.78± 0.3	0.62± 0.09	1.94± 0.26	1.33± 0.17	p<0.001	p<0.05	ANOVA	10/13/10/15
modified hole board_ethovision	Total distance moved (cm)	2912.87±132.7	3135.65±150.27	2604.53±114.61	2879.14±84.44	p<0.05	p<0.05	ANOVA	10/13/10/15
modified hole board_ethovision	Mean velocity (cm/sec.)	18.78 ±0.62	19.22 ±0.63	17.27 ±0.54	18.69 ±0.47	n.s.	n.s.	ANOVA	10/13/10/15
modified hole board_ethovision	Maximum velocity (cm/sec.)	56.75 ±1.54	53.63 ±1.84	58.01 ±2.36	61.49 ±2.83	n.s.	p=0.06	ANOVA	10/13/10/15
modified hole board_ethovision	Turns (frequency)	1559.4±42.54	1668.08±48.46	1462.2±48.96	1543.6±25.86	p<0.05	p<0.05	ANOVA	10/13/10/15
modified hole board_ethovision	Mean turn angle (degrees)	24.52 ±0.8	23.4± 0.53	26.16 ±0.59	24.82 ±0.79	n.s.	p<0.05	ANOVA	10/13/10/15
modified hole board_ethovision	Absolute meander (degrees/sec.)	17.06 ±0.69	16.21 ±0.44	18.55 ±0.49	17.49 ±0.65	n.s.	p<0.05	ANOVA	10/13/10/15
modified hole board_ethovision	Mean distance to board (cm)	8.48± 0.18	7.82± 0.29	7.66± 0.24	7.86± 0.18	n.s.	n.s.	ANOVA	10/13/10/15
modified hole board_ethovision	Mean distance to wall (cm)	7.2±0.27	8.09± 0.4	8.5±0.35	8.04± 0.26	n.s.	p=0.07	ANOVA	10/13/10/15
Morphology_Hearing	hearing mean score (N)	3	3.08	2.3	2.07	n.s.	<0.001	Kruskal-Wallis	10/13/10/15
Morphology DEXA	Fat mass[units]	5.71± 0.79	4.27± 0.66	7.23± 0.85	6.49± 0.99	n.s.	<0.05	ANOVA	10/7/10/6
Morphology DEXA	Fat Content[units x 100/g]	18.99 ±2.40	17.5± 2.71	24.14 ±2.55	25.19 ±3.37	n.s.	<0.05	ANOVA	10/7/10/6
Morphology DEXA	Lean mass[units]	20.91 ±0.66	16.99 ±0.73	18.91 ±0.72	15.83 ±0.49	<0.0001	<0.05	ANOVA	10/7/10/6
Morphology DEXA	Lean Content[units x 100/g]	70.51 ±2.57	69.51 ±2.78	64.68 ±2.62	62.89 ±3.24	n.s.	<0.05	ANOVA	10/7/10/6
Neurology_arena	Locomotor activity: lines crossed	19.2± 1.79	23.2± 1.4	21.4± 1.67	21.5± 1.92	n.s.	n.s.	ANOVA	10/13/10/15
Neurology rotarod	day 1; Trial 1; Latency to fall [s]	58.1± 10.39	42.9± 15.8	57.1± 15.98	71.7± 14.2	Geno x trial-no. <0.01		linear mixed-	10/13/10/1

Neurology rotarod	day 1; Trial 2; Latency to fall [s]	135.2 ±13.3 8	81.1± 18	132.2 ±12.2 3	69.9± 15.94	Geno n.s. Sex n.s. Body weight n.s. Trial-no. <0.0001		effects model	5
Neurology rotarod	day 1; Trial 3; Latency to fall [s]	153.2 ±7.5	130.8 ±22.5 8	115.8 ±20.9 5	79.4± 11.48				
Neurology rotarod	day 1; Trial 4; Latency to fall [s]	143±1 2.28	142.5 ±20.6 2	109.5 ±14.0 7	117.2 ±16.8 2				
Clinical Chemistry	Inorganic Phosphate [mmol/l]	1.26± 0.09	1.34± 0.09	1±0.0 8	1.16± 0.07	n.s.	<0.05	ANOVA	10/13/ 10/15
Clinical Chemistry	Ferritin [ng/ml]	24.3± 1.7	27.2± 2.07	29.9± 4.39	31.9± 1.63	n.s.	<0.05	ANOVA	10/13/ 10/15
Clinical Chemistry; Hematology	Red blood cell count [10^3/µl]	10±0. 08	9.8±0. 13	10.2± 0.08	10±0. 08	n.s.	<0.05	ANOVA	10/13/ 10/15
Clinical Chemistry; Hematology	Hemoglobin[g/dl]	14.9± 0.08	14.7± 0.13	15.1± 0.08	14.9± 0.1	n.s.	<0.05	ANOVA	10/13/ 10/15
Clinical Chemistry; Hematology	Hematocrit[%]	46.9± 0.28	45.8± 0.44	47.7± 0.26	46.8± 0.38	<0.05	<0.05	ANOVA	10/13/ 10/15
Immunology	CD4+[%]	9.05± 0.59	8.93± 0.35	9.7±0. 61	10.47 ±0.5	n.s.;n .s.	n.s.;p <0.05	t-test	10/13/ 10/15
Immunology	CD8a+	8.05± 0.36	8.23± 0.29	8.07± 0.39	9.74± 0.38	n.s.;p <0.01	n.s.;p <0.01	t-test	10/13/ 10/15
Immunology	IgM[mg/l]	1701. 1±164 .13	1959. 8±85. 36	2015. 9±154 .28	1680. 9±95. 93	n.s.;n .s.	n.s.;p <0.05	t-test	10/13/ 10/15
Immunology; 2nd bleeding	CD4+[%]	9.83± 0.27	10.62 ±0.3	9.54± 0.5	11.72 ±0.33	n.s.; p<0.0 1	n.s.; p<0.0 5	t-test	10/7/1 0/6
Immunology; 2nd bleeding	l/45+[#]		30376 .1±30. 63		30538 ±57.6 7		p<0.0 5	t-test	0/7/0/ 5
Immunology; 2nd bleeding	nonBcells/CD4+/CD3+ Freq.45+[%]		9.9±0. 32		11.23 ±0.41		p<0.0 5	t-test	0/7/0/ 5
Nociception_h ot plate test	Latency to jumping [s]	50.6± 2.9	55.2± 2.8	48.9± 2.9	43.7± 2.8	n.s.	<0.05	ANOVA	10/13/ 10/15
Heart; blood pressure	Pulse [bpm]	622.2 ±14.4	584.4 ±13.3	550.5 ±20.2	569.5 ±21.9	n.s.	<0.05	ANOVA	10/7/1 0/6
Heart; electrocardiog ram	R amplitude [mV]	2.77± 0.14	3.10± 0.42	2.58± 0.15	3.32± 0.33	p<0.0 5	n.s.	ANOVA	10/7/1 0/6
Lung; active phase	respiratory rate [1/min]	n.d.	480.6 ±7.1	n.d.	508.5 ±5.7		<0.05	t-test	0/5/0/ 4
Lung; active phase	inspiratory time [ms]	n.d.	43.1± 0.4	n.d.	41.1± 0.6		<0.05	t-test	0/5/0/ 4
Lung; active phase	peak inspiratory flow rate [ml/s]	n.d.	8.8±0. 2	n.d.	9.9±0. 3		<0.05	t-test	0/5/0/ 4
Lung; active phase	mean inspiratory flow rate [ml/s]	n.d.	5.4±0. 1	n.d.	5.9±0. 2		<0.05	t-test	0/5/0/ 4
Lung; active phase	mean expiratory flow rate [ml/s].	n.d.	2.8±0. 1	n.d.	3.2±0. 1		<0.05	t-test	0/5/0/ 4
Lung; resting phase	respiratory rate [1/min]	n.d.	333±5 .5	n.d.	362.8 ±12.1		<0.05	t-test	0/5/0/ 3
Lung; resting phase	minute ventilation [ml/min]	n.d.	76±2. 2	n.d.	88±2. 8		<0.02	t-test	0/5/0/ 3
Lung; resting phase	expiratory time [ms]	n.d.	125.2 ±2.9	n.d.	111.4 ±4.5		<0.05	t-test	0/5/0/ 3
Lung; resting phase	relative duration of inspiration	n.d.	0.31± 0.01	n.d.	0.33± 0		<0.05	t-test	0/5/0/ 3
Lung; resting phase	peak inspiratory flow rate [ml/s]	n.d.	7.2±0. 1	n.d.	7.7±0. 1		<0.05	t-test	0/5/0/ 3
Lung; resting phase	mean inspiratory flow rate [ml/s]	n.d.	4.3±0. 1	n.d.	4.6±0. 1		<0.02	t-test	0/5/0/ 3
Lung; resting phase	mean expiratory flow rate [ml/s].	n.d.	1.9±0. 1	n.d.	2.3±0. 1		<0.01	t-test	0/5/0/ 3
Metabolism	Food consumption[g day-1]	2.6±0. 1	2.7±0. 1	3.0±0. 1	2.8±0. 2	n.s.	<0.05	ANOVA	7/5/7/ 4

Metabolism	Metabolized energy [kJ day-1]	33.81 ±0.92	34.72 ±1.79	40.11 ±0.9	36.12 ±2.55	n.s.	<0.05	ANOVA	7/5/7/4
Metabolism	Metabolized energy[kJ g-1 day-1]	1.16± 0.03	1.41± 0.06	1.40± 0.05	1.45± 0.10	<0.05	<0.05	ANCOVA	7/5/7/4
Metabolism	Food assimilation coefficient [%]	75.8± 0.3	75.5± 0.7	77.2± 0.4	76.2± 0.3	n.s.	<0.05	ANOVA	7/5/7/4

[a] significance for factor sex; when two values are given, tests were performed separately for *Foxp2*$^{wt/wt}$ and *Foxp2*$^{wt/ko}$ mice; n.s. represents a p-value >0.05; empty cells were not explicitly tested; [b] significance for factor genotype; when two values are given, tests were performed separately for males and females; a third value indicates significance when males and females are combined. [c] Number of individuals analysed: males *Foxp2*$^{wt/wt}$/ females *Foxp2*$^{wt/wt}$ / males *Foxp2*$^{wt/ko}$ / females *Foxp2*$^{wt/ko}$.

Supplemental Table S5: ANOVA results comparing *Foxp2*$^{hum/hum}$ mice with *Foxp2*$^{wt/ko}$ mice on the modified hole board

Parameter[a]	gen. [b]	sex	strain[c]	gen.* strain	gen.* sex	sex* strain	gen.*sex* strain
Absolute meander (Sqrt degrees/sec.)	<0.05	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Angular velocity (degrees/sec.)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Board entry (frequency)	n.s.	n.s.	<0.01	n.s.	n.s.	n.s.	n.s.
Board entry (rank latency)	n.s.	n.s.	<0.001	n.s.	n.s.	n.s.	n.s.
Board entry (total duration %)	n.s.	n.s.	<0.01	<0.05	n.s.	n.s.	n.s.
Familiar object exploration (sqrt total duration %)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Group contact (latency)	n.s.	<0.05	n.s.	n.s.	n.s.	n.s.	n.s.
Group contact (sqrt_frequency)	<0.01	n.s.	<0.05	n.s.	n.s.	n.s.	n.s.
Group contact (total duration %)	n.s.	n.s.	n.s.	<0.05	n.s.	n.s.	n.s.
Hole exploration (frequency)	n.s.	n.s.	n.s.	<0.05	n.s.	n.s.	n.s.
Hole exploration (sqrt_latency)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Line crossing (frequency)	<0.001	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Maximum velocity (cm/sec.)	n.s.	n.s.	n.s.	<0.05	n.s.	n.s.	n.s.
Mean distance to board (cm)	n.s.	n.s.	<0.001	<0.05	n.s.	n.s.	n.s.
Mean distance to wall (cm)	n.s.	n.s.	<0.001	<0.01	n.s.	n.s.	n.s.
Mean turn angle (ln degrees)	<0.05	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Mean velocity (cm/sec.)	<0.01	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Object Index (ln)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Rearings in box (frequency)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Rearings in box (latency)	n.s.	n.s.	<0.05	n.s.	n.s.	n.s.	n.s.
Total distance moved (cm)	<0.01	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Turns (frequency)	<0.01	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
Unfamiliar object exploration (sqrt total duration %)	n.s.	<0.01	n.s.	<0.05	n.s.	<0.01	n.s.

[a] some parameters were transformed (log(ln), square root (sqrt) or rank) in order to meet the equality of error variance assumption; [b] levels mutated (*Foxp2*$^{wt/ko}$ or *Foxp2*$^{hum/hum}$) and wildtype (*Foxp2*$^{wt/wt}$); [c] levels *Foxp2*^{hum} (both batches) and *Foxp2*^{ko};

Supplemental Table S6: Neurotransmitter measurements

Strain	Statistic	Region[a]	Dopamine[b]	Serotonin[b]	Glutamate[b]	GABA[b]
Foxp2^{hum}	Mean±SEM for *Foxp2*$^{hum/hum}$	CPu	718±32.9	20±2.2	79±2.5	13±0.7
		CRB	16±1.3	8±1.2	71±1.3	11±0.4
		GP	315±23.1	30±3.6	66±3.3	32±2.6
		FC	8±1	22±2.4	106±5.3	15±0.9
		NAc	362±34.1	49±5.2	67±2.5	46±3.6
	Mean±SEM for *Foxp2*$^{wt/wt}$	CPu	826±22.6	19±1.8	75±2.1	13±0.4
		CRB	22±3.3	8±1.4	66±2.1	11±0.6
		GP	406±18.7	32±3.6	64±3.4	28±2.6
		FC	14±2.2	29±3	97±5.7	14±0.6
		NAc	460±32.5	46±5.2	72±3.4	39±3.1
Foxp2^{ko}	Mean±SEM for *Foxp2*$^{ko/wt}$	CPu	1119±157.6	44±6.1	66±1.8	13±0.6
		CRB	15±3.3	18±3.7	61±1.3	10±0.4
		GP	483±83.6	80±5.4	52±1.9	32±2.3
		FC	12±2.5	73±4.6	80±2.3	12±0.3
		NAc	661±108.8	159±13.6	59±1.4	34±2.5
	Mean±SEM for *Foxp2*$^{wt/wt}$	CPu	788±164.9	52±5.5	65±3.7	13±0.7
		CRB	15±3.7	11±1.4	63±1.9	11±0.5
		GP	291±32.9	78±3.6	54±2.5	35±2.5
		FC	12±2.3	71±4.7	88±3.3	13±0.4
		NAc	356±55.3	112±6.5	56±3.6	33±1.8
Foxp2^{hum}	p-value t-test	CPu	0.014	0.886	0.286	0.721
		CRB	0.162	0.727	0.054	0.933
		GP	0.007	0.753	0.798	0.285
		FC	0.021	0.082	0.24	0.401
		NAc	0.052	0.706	0.205	0.137
	p (geno) rep. meas. ANOVA[c]	all	<0.001	0.634	0.22	0.089
Foxp2^{ko}	p-value t-test	CPu	0.163	0.321	0.764	0.434
		CRB	0.973	0.071[f]	0.379	0.064
		GP	0.047[f]	0.775	0.474	0.482
		FC	0.903	0.725	0.072	0.026
		NAc	0.02[f]	0.005	0.552	0.815
	p (geno) rep. ANOVA[c]	all	0.039	0.044	0.523	0.437
Combined	*p* (geno*strain) ANOVA[d]	CPu	0.066[f]	0.318[f]	0.654	0.733
		CRB	0.417	0.066[f]	0.041	0.19
		GP	0.005[f]	0.673	0.539	0.204
		FC	0.123	0.224	0.058	0.078[f]
		NAc	0.003[f]	0.014	0.219	0.242
	p (geno*strain) rep. ANOVA[e]	all	0.005	0.039	0.785	0.323

[a] CPu = Caudate Putamen, CRB = Cerebellum, GP = Globus pallidus, FC = frontal cortex, NAc = Nucleus accumbens; [b] Concentrations are given as nmol/mg protein [c] Regions were treated as repeated measure and the p-value for the between-subjects factor genotype is reported; [d] the p-value for the Greenhouse-Geisser corrected interaction between genotype and strain, which indicates differences between *Foxp2*$^{wt/ko}$ and *Foxp2*$^{hum/hum}$ mice; [e] as for [d] with regions treated as repeated measure; [f] assumption of equality of error variances is violated (Levene's Test, $p<0.05$);

Supplemental Table S7: Analysis of call structure (calls <50 ms) comparing *Foxp2*$^{hum/hum}$ with *Foxp2*$^{wt/wt}$ pups (5C10 line)

Parameter	gen. [o]	sex [p]	wt/wt [q]	h/h [q]	gen. [r]
Amplitude gap [ms] [a]	n.s.	n.s.	0.8 ± 0.12	0.6 ± 0.1	n.s.
Duration [ms] [b]	n.s.	n.s.	20.8 ± 0.71	20.9 ± 0.76	n.s.
General slope [Hz/ms] [c]	<0.001	n.s.	-363 ± 32.2	-173 ± 43.3	<0.001
Global modulation [kHz] [d]	n.s.	n.s.	1.2 ± 0.06	1 ± 0.07	<0.001
Local modulation [Hz] [e]	<0.01	n.s.	150 ± 7.2	115 ± 8.6	<0.001
Peak frequency end [kHz] [f]	n.s.	n.s.	72.2 ± 0.65	70.1 ± 0.81	<0.001
Peak frequency jump [kHz] [g]	n.s.	n.s.	3 ± 0.23	2.5 ± 0.36	<0.05
Peak frequency max [kHz] [h]	<0.01	n.s.	80.8 ± 0.75	77.1 ± 0.85	<0.001
Peak frequency maxloc [(1/durat.)*location] [i]	n.s.	n.s.	0.3 ± 0.01	0.3 ± 0.02	n.s.
Peak frequency mean [kHz] [j]	<0.01	<0.05	77 ± 0.67	73.7 ± 0.71	<0.001
Peak frequency min [kHz] [k]	<0.01	<0.05	71 ± 0.63	68.3 ± 0.68	<0.001
Peak frequency start [kHz] [l]	<0.001	<0.05	78 ± 0.76	74 ± 0.75	<0.001
Slope of trend [m]	<0.001	n.s.	-0.3 ± 0.03	-0.2 ± 0.03	<0.001
Start/ max slope [Hz/ms] [n]	n.s.	n.s.	539 ± 63.5	512 ± 60	n.s.

[a] Duration of breaks in amplitude within call; [b] Time in milliseconds (ms) between onset and offset of call; [c] General slope from start to end peak frequency; [d] Mean deviation between peak frequency and linear trend; [e] Mean frequency difference between original and floating average curve; [f] End of peak frequency; [g] Maximum difference of peak frequency between successive bins; [h] Maximum peak frequency; [i] Location of maximum peak frequency ; [j] Mean of peak frequency; [k] Minimum peak frequency; [l] Start frequency of peak frequency; [m] Factor of linear trend of peak frequency; [n] Slope from start to maximum peak frequency; [o] p-value for factor genotype in a general linear mixed model with days as within-subject factor, genotype, sex and litter as fixed between-subject factors and weight as covariate on values averaged from a maximum of 10 randomly chosen calls per individual; [p] p-value for sex to put the genotype effects into perspective; [q] means ± S.E.M. averaged per individual across all days (in total 119 measures for *Foxp2*$^{wt/wt}$ and 88 for *Foxp2*$^{hum/hum}$); [r] p-value for factor genotype in an ANOVA with day, genotype and sex as fixed between-subject factors on a maximum of 10 randomly chosen calls per individual.

Supplemental Table S8: Analysis of call structure (calls <50 ms) comparing *Foxp2*$^{hum/hum}$ mice with *Foxp2*$^{wt/wt}$ pups (5H11 line)

Parameter [a]	gen. [b]	sex [c]	wt/wt [d]	h/h [d]
Amplitude gap [ms]	n.s.	n.s.	0.2 ± 0.03	0.3 ± 0.07
Duration [ms]	n.s.	n.s.	19.3 ± 0.68	19.1 ± 0.75
General slope [Hz/ms]	<0.05	n.s.	-242 ± 24	-190 ± 33.1
Global modulation [kHz]	n.s.	<0.01	1 ± 0.05	0.9 ± 0.07
Local modulation [Hz]	n.s.	n.s.	108 ± 4.9	107 ± 6.8
Peak frequency end [kHz]	n.s.	n.s.	69.7 ± 0.7	68.9 ± 0.72
Peak frequency jump [kHZ]	n.s.	n.s.	2 ± 0.15	2.1 ± 0.27
Peak frequency max [kHz]	<0.01	n.s.	75.8 ± 0.75	74.6 ± 0.75
Peak frequency maxloc [(1/duration)*location]	<0.05	n.s.	0.3 ± 0.02	0.3 ± 0.03
Peak frequency mean [kHz]	<0.01	n.s.	73.3 ± 0.7	72.2 ± 0.71
Peak frequency min [kHz]	<0.05	n.s.	68.7 ± 0.68	67.7 ± 0.74
Peak frequency start [kHz]	<0.01	n.s.	74 ± 0.76	72.7 ± 0.81
Slope of trend	<0.05	n.s.	-0.2 ± 0.02	-0.1 ± 0.02
Start/ max slope [Hz/ms]	n.s.	n.s.	270 ± 20.8	264 ± 28.2

[a] for a description of parameters see Table S7; [b] p-value for factor genotype in an ANOVA with day, genotype and sex as fixed between-subject factors on a maximum of 10 randomly chosen calls per individual; [c] p-value for sex to put the genotype effects into perspective; [d] means ± S.E.M. averaged per individual across all days (in total 57 measures for *Foxp2*$^{wt/wt}$ and 36 for *Foxp2*$^{hum/hum}$);

Supplemental Table S9: Analysis of call structure (calls with frequency jumps) comparing *Foxp2*$^{hum/hum}$ mice with *Foxp2*$^{wt/wt}$ pups (5C10 line)

Parameter [a]	gen. [b]	sex [c]	wt/wt [d]	h/h [d]
Amplitude gap [ms]	<0.05	n.s.	4.3 ± 0.36	5.6 ± 0.62
Duration [ms]	<0.05	n.s.	31.6 ± 1.32	36 ± 1.55
Peak frequency end [Hz]	n.s.	n.s.	84.2 ± 0.9	86.8 ± 0.91
Peak frequency jump [kHz]	n.s.	n.s.	37.1 ± 0.49	36.5 ± 0.56
Peak frequency jumploc [(1/duration)*location]	n.s.	n.s.	0.7 ± 0.02	0.7 ± 0.02
Peak frequency max [kHz]	n.s.	n.s.	94.1 ± 0.67	95.7 ± 0.86
Peak frequency maxloc [(1/duration)*location]	n.s.	n.s.	0.7 ± 0.02	0.7 ± 0.02
Peak frequency min [kHz]	<0.05	n.s.	56.6 ± 0.54	58.7 ± 0.65
Peak frequency start [kHz]	<0.01	n.s.	63.9 ± 0.62	66.9 ± 1.13

[a] for a description of parameters see Table S7; [b] p-value for factor genotype in an ANOVA with day, genotype and sex as fixed between-subject factors and weight as covariate on averages per individual and day; [c] p-value for sex to put the genotype effects into perspective; [d] means ± S.E.M. averaged per individual across all days (in total 89 measures for *Foxp2*$^{wt/wt}$ and 60 for *Foxp2*$^{hum/hum}$);

Supplemental References

Abràmoff, M. D., Magalhães, P. J., and Ram, S. J. (2004). Image Processing with ImageJ. Biophotonics International, vol. 11, no. 7, July 2004, pp. 36-42. Biophotonics Int *11*, 36-42.

Alessandrini, F., Jakob, T., Wolf, A., Wolf, E., Balling, R., Hrabé de Angelis, M., Ring, J., and Behrendt, H. (2001). Enu mouse mutagenesis: generation of mouse mutants with aberrant plasma IgE levels. Int Arch Allergy Immunol *124*, 25-28.

Barrantes Idel, B., Montero-Pedrazuela, A., Guadano-Ferraz, A., Obregon, M. J., Martinez de Mena, R., Gailus-Durner, V., Fuchs, H., Franz, T. J., Kalaydjiev, S., Klempt, M., *et al.* (2006). Generation and characterization of dickkopf3 mutant mice. Mol Cell Biol *26*, 2317-2326.

Brielmeier, M., Fuchs, H., Przemeck, G., Gailus-Durner, V., Hrabé de Angelis, M., and Schmidt, J. (2002). The GSF-Phenotype Analysis Center (German Mouse Clinic, GMC): A sentinel-based health-monitoring concept in a multi-user unit for standardized characterization of mouse mutants. In Proceedings of the 8th FELASA Symposium, J.-L. Guenet, and C. Herweg, eds. (Aachen, Laboratory Animals Ltd), pp. 19-22.

Cohen, J. (1992). A power primer. Psychological Bulletin *122*, 155-159.

Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., Bunney, W. E., Myers, R. M., Speed, T. P., Akil, H., *et al.* (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res *33*, e175.

Drorbaugh, J. E., and Fenn, W. O. (1955). A barometric method for measuring ventilation in newborn infants. Pediatrics *16*, 81-87.

Drozdz, A. (1975). Food habits and food assimilation in mammals. In Methods for Ecological Bioenergetics, W. Grodzinski, R. Z. Klekowski, and A. Duncan, eds. (Oxford, Blackwell), pp. 23-47.

Favor, J. (1983). A comparison of the dominant cataract and recessive specific-locus mutation rates induced by treatment of male mice with ethylnitrosourea. Mutat Res *110*, 367-382.

Felice, L. J., Felice, J. D., and Kissinger, P. T. (1978). Determination of catecholamines in rat brain parts by reverse-phase ion-pair liquid chromatography. J Neurochem *31*, 1461-1465.

Fuchs, H., Schughart, K., Wolf, E., Balling, R., and Hrabé de Angelis, M. (2000). Screening for dysmorphological abnormalities--a powerful tool to isolate new mouse mutants. Mamm Genome *11*, 528-530.

Fujita, E., Tanabe, Y., Shiota, A., Ueda, M., Suwa, K., Momoi, M. Y., and Momoi, T. (2008). Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proc Natl Acad Sci U S A *105*, 3117-3122.

Gailus-Durner, V., Fuchs, H., Becker, L., Bolle, I., Brielmeier, M., Calzada-Wack, J., Elvert, R., Ehrhardt, N., Dalke, C., Franz, T. J., *et al.* (2005). Introducing the German

Mouse Clinic: open access platform for standardized phenotyping. Nat Methods *2*, 403-404.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307-315.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5*, R80.

Heiman, M., Schaefer, A., Gong, S., Peterson, J. D., Day, M., Ramsey, K. E., Suarez-Farinas, M., Schwarz, C., Stephan, D. A., Surmeier, D. J., *et al.* (2008). A translational profiling approach for the molecular characterization of CNS cell types. Cell *135*, 738-748.

Kontgen, F., Suss, G., Stewart, C., Steinmetz, M., and Bluethmann, H. (1993). Targeted disruption of the MHC class II Aa gene in C57BL/6 mice. Int Immunol *5*, 957-964.

Lambert, J. F., Benoit, B. O., Colvin, G. A., Carlson, J., Delville, Y., and Quesenberry, P. J. (2000). Quick sex determination of mouse fetuses. J Neurosci Methods *95*, 127-132.

Meijering, E., Jacob, M., Sarria, J. C., Steiner, P., Hirling, H., and Unser, M. (2004). Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytometry A *58*, 167-176.

Nicoletti, I., Migliorati, G., Pagliacci, M. C., Grignani, F., and Riccardi, C. (1991). A rapid and simple method for measuring thymocyte apoptosis by propidium iodide staining and flow cytometry. J Immunol Methods *139*, 271-279.

Ohl, F., Holsboer, F., and Landgraf, R. (2001). The modified hole board as a differential screen for behavior in rodents. Behav Res Methods Instrum Comput *33*, 392-397.

Paxinos, G., and Franklin, K. B. J. (1997). The Mouse Brain in Stereotaxic Coordinates (San Diego, Academic Press).

Piepponen, T. P., and Skujins, A. (2001). Rapid and sensitive step gradient assays of glutamate, glycine, taurine and gamma-aminobutyric acid by high-performance liquid chromatography-fluorescence detection with o-phthalaldehyde-mercaptoethanol derivatization with an emphasis on microdialysis samples. J Chromatogr B Biomed Sci Appl *757*, 277-283.

Prufer, K., Muetzel, B., Do, H. H., Weiss, G., Khaitovich, P., Rahm, E., Paabo, S., Lachmann, M., and Enard, W. (2007). FUNC: a package for detecting significant associations between gene sets and ontological annotations. BMC Bioinformatics *8*, 41.

Rodriguez, C. I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., Stewart, A. F., and Dymecki, S. M. (2000). High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. Nat Genet *25*, 139-140.

Schambra, U., Lauder, J. M., and Silver, J. (1992). Atlas of the prenatal mouse brain (San Diego, Academic Press).

Schneider, I., Tirsch, W. S., Faus-Kessler, T., Becker, L., Kling, E., Busse, R. L., Bender, A., Feddersen, B., Tritschler, J., Fuchs, H., *et al.* (2006). Systematic, standardized and

comprehensive neurological phenotyping of inbred mice strains in the German Mouse Clinic. J Neurosci Methods *157*, 82-90.

Shu, W., Cho, J. Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G. A., Schmeidler, J., De Gasperi, R., Sosa, M. A., Rabidou, D., *et al.* (2005). Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A *102*, 9643-9648.

Sperk, G. (1982). Simultaneous determination of serotonin, 5-hydroxindoleacetic acid, 3,4-dihydroxyphenylacetic acid and homovanillic acid by high performance liquid chromatography with electrochemical detection. J Neurochem *38*, 840-843.

Sperk, G., Berger, M., Hortnagl, H., and Hornykiewicz, O. (1981). Kainic acid-induced changes of serotonin and dopamine metabolism in the striatum and substantia nigra of the rat. Eur J Pharmacol *74*, 279-286.

Stroud, J. C., Wu, Y., Bates, D. L., Han, A., Nowick, K., Paabo, S., Tong, H., and Chen, L. (2006). Structure of the forkhead domain of FOXP2 bound to DNA. Structure *14*, 159-166.

The_Gene_Ontology_Consortium (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., Geschwind, D. H., and 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.

Wang, B., Lin, D., Li, C., and Tucker, P. (2003). Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J Biol Chem *278*, 24259-24268.

Zapala, M. A., Hovatta, I., Ellison, J. A., Wodicka, L., Del Rio, J. A., Tennant, R., Tynan, W., Broide, R. S., Helton, R., Stoveken, B. S., *et al.* (2005). Adult mouse brain gene expression patterns bear an embryologic imprint. Proc Natl Acad Sci U S A *102*, 10357-10362.

Supplemental Figures

Enard et al. Figure S1

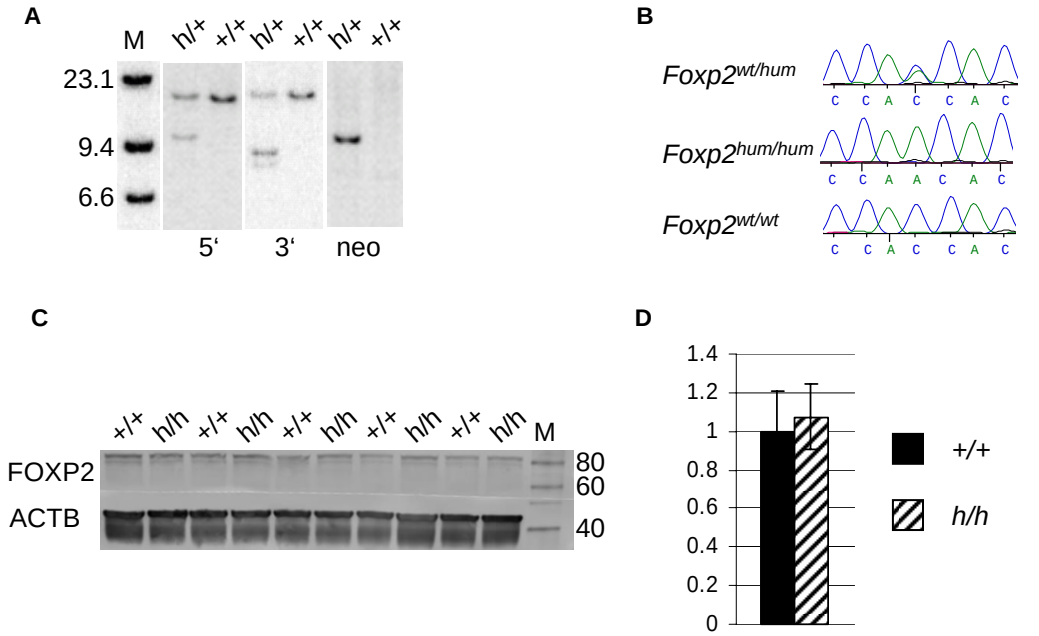


Figure S1 Characterization of the *Foxp2*^{hum} allele. (A) Southern blot analysis of genomic DNA from the targeted ES cell clone 5H10 (het) and from wildtype control DNA (wt). EcoRV digested genomic DNA together with a radioactively labelled marker (M) was probed with the 3‘ probe, the 5‘ probe and a probe for the neomycin resistance cassette (neo). The expected size of the *Foxp2*^{wt} allele is 16 kbp with the 3‘ and 5‘ probe. The expected size of the *Foxp2*^{hum} allele is 9.6 kbp with the 5‘or neo probe and 8.5 kbp with the 3‘probe. Note that the cell line shows signals at the expected size for all probes and an additional stronger signal slightly above the expected 8.5 kbp signal, indicating a duplication/rearrangement affecting a region containing the 3‘ probe (see Supplementary Data S1). This duplication/rearrangement is not present in the second targeted ES cell line (5H11, see Figure S2). **(B)** Trace files at the first introduced mutation (resulting in T302N) from cerebellar cDNA indicate that mRNA expression levels do not differ between the *Foxp2*^{hum} and *Foxp2*^{wt} allele. The same is seen for both mutations in cerebrum, cerebellum and lung of two heterozygous male and two heterozygous females (not shown). **(C)** Western blot analysis of Foxp2^{hum} protein expression. 20µg of protein extracted from E16.5 whole brains (line 5C10) were immunoblotted with antibodies against FOXP2 (HPA000382, Atlas Antibodies) and Actin (C4, MP Biomedicals). **(D)** Band intensities were calculated as volumes under automatically detected peaks using ImageQuant TL software (GE Healthcare) and are displayed as relative average intensities (± S.E.M) of the upper Foxp2 band, normalized to actin levels in and show no difference in the amount of Foxp2 detected between mice homozygous for Foxp2hum (*h/h*) and wildtype littermates (+/+).

Enard et al. Figure S2

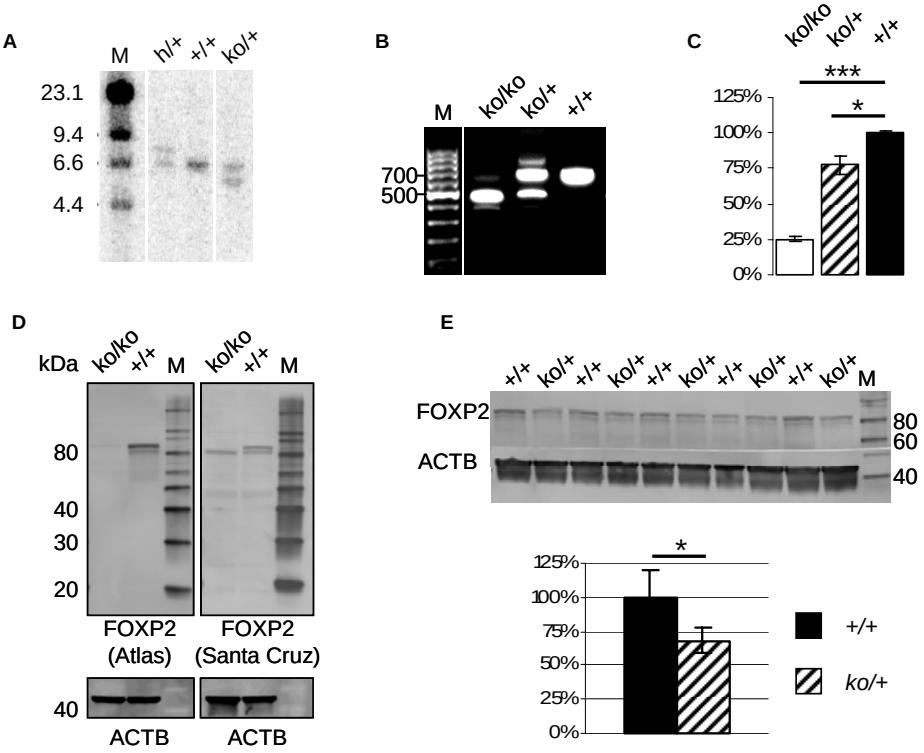
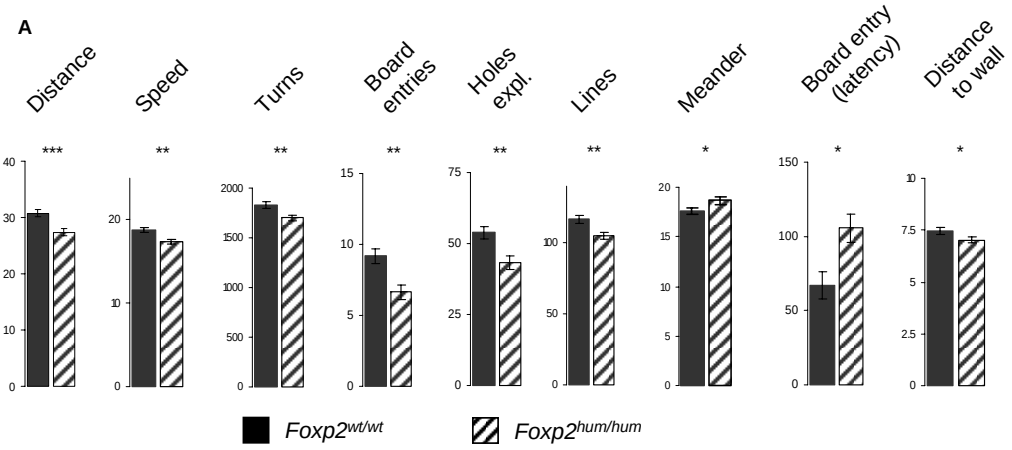


Figure S2 Characterization of the *Foxp2*^{ko} allele. (A) Southern analyses using the 3‘ probe of genomic DNA digested with NsiI from offspring of a cross between a Cre deleter strain (Ozgene) and a chimeric mouse generated from the targeted ES cell clone 5H11. Sizes of hybridizing fragments are as expected for a mouse heterozygous for the targeted *Foxp2*^{hum} allele *(h/+,* 8.7 kbp), heterozygous for the *Foxp2*^{ko} allele (*ko/+*, 5.7 kbp) and homozygous for the *Foxp2*^{wt} allele *(+/+,* 6.8 kbp). **(B)** Total RNA was isolated from whole brains of E16.5 embryos, reverse transcribed and used in a PCR with primers located in *Foxp2* exons 6 and 11. The deletion of exon 7 in the *Foxp2*^{ko} allele leads to a predicted size reduction of 214 bp, compatible with the observed results. The lower amount of the shorter product in *Foxp2*$^{wt/ko}$ mice (ko/+) suggests that non-sense-mediated RNA decay might reduce the amount of *Foxp2*^{ko} mRNA. **(C)** Quantitative real-time RT-PCR using *Foxp2* primers on RNA isolated from E16.5 whole brains. Shown are relative expression levels (± S.E.M.) averaged from six technical replicates derived from two individuals per genotype normalised to beta-2-microglobulin mRNA. **(D)** Western blot analysis of Foxp2 expression in *Foxp2*$^{ko/ko}$ mice. Proteins were extracted from E16.5 whole brains and an amount corresponding to 1mg of brain was loaded. Western analysis was performed using the antibody HPA000382 (1:2500, Atlas Antibodies, Stockholm, Sweden) and sc21069 (1:750, Santa Cruz Biotechnology, Santa Cruz, USA) that both were raised against epitopes N-terminal to exon seven. Note that the deletion of exon 7 leads to a predicted protein product of 291 amino acids (33.3 kDa) and that such a product is not detected. Blots were stripped and reprobed with Actin (1:5000, C4, MP Biomedicals) to control for equal loading. **(E)** Western blot analysis of Foxp2 expression in *Foxp2*$^{wt/ko}$ mice were done as described for Figure S1C. Band intensities (± S.E.M.) averaged from two independent Western blot analyses of the same brain samples show that Foxp2 is significantly reduced (Student’s t-test, $p<0.05$) to approximately 70% in *Foxp2*$^{wt/ko}$ embryonic brains.

Enard et al. Figure S3



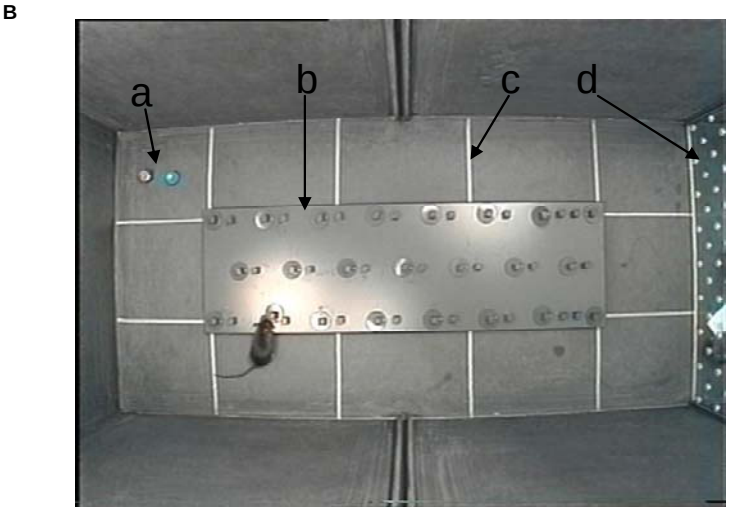


Figure S3: Reduced exploration of *Foxp2*$^{hum/hum}$ mice on the modified hole board. **(A)** Shown are means ± S.E.M of all parameters significant for the factor genotype in an ANOVA with the factors genotype, sex and batch (Table S3): the average distance travelled (in meters), the average speed (in centimetres/second), the number of turns, the number of board entries, the numbers of holes explored, the number of lines crossed, the absolute meander (in degrees/second), the latency to enter the board (in seconds) and the mean distance to the wall (in centimeter). Stars represent significance levels of 0.05 (*), 0.01 (**) and 0.001 (***). **(B)** Picture of the modifed hole board showing the familiar and the unfamiliar object (a), the hole board (b), the lines (c) and the group compartment with cage mates (d).

Enard et al. Figure S4

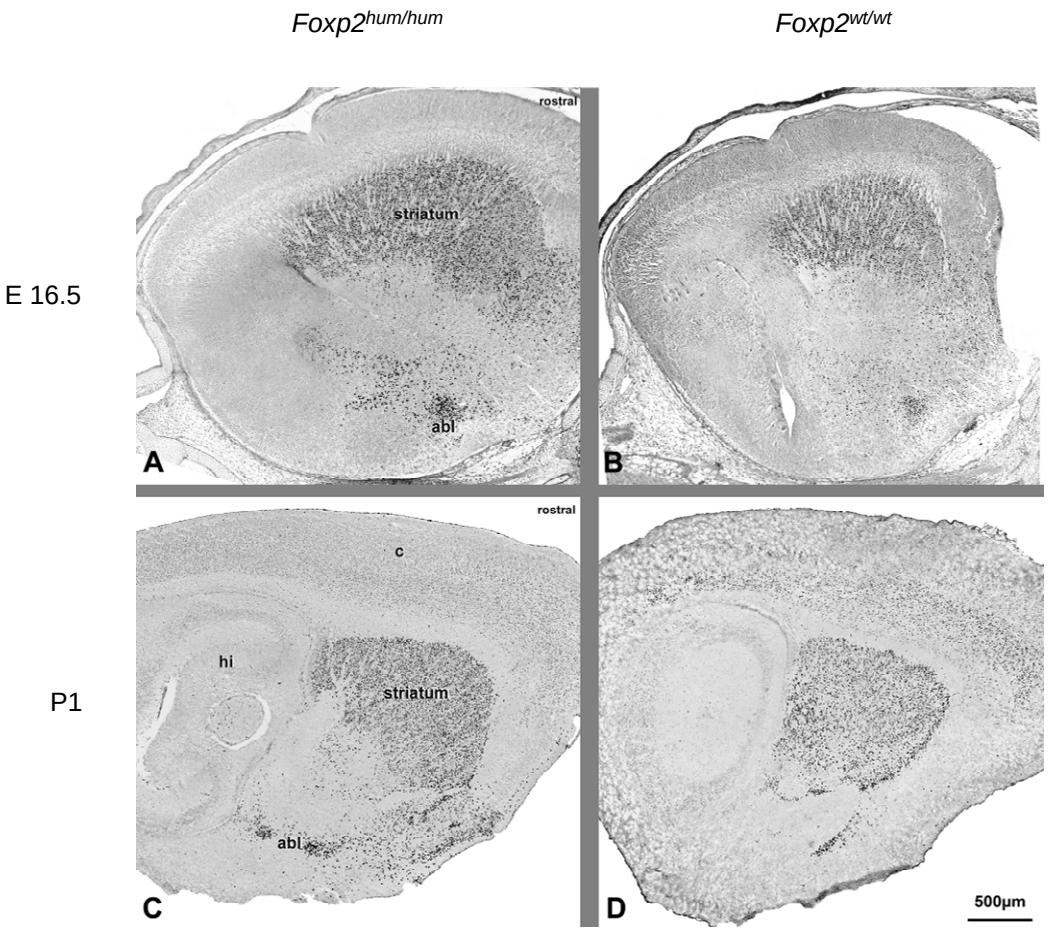


Figure S4: Representative Foxp2 stainings at E16.5 and P1. Shown are sagittal sections (7µm paraffin (**A**,**B**) and 12µm cryosections (**C**,**D**)) stained with Foxp2 antibody. Strong nuclear staining of striatal cells (centromedian striatal nucleus), and the basal lateral amygdaloid area (abl) is evident at both developmental stages.

Enard et al. Figure S5

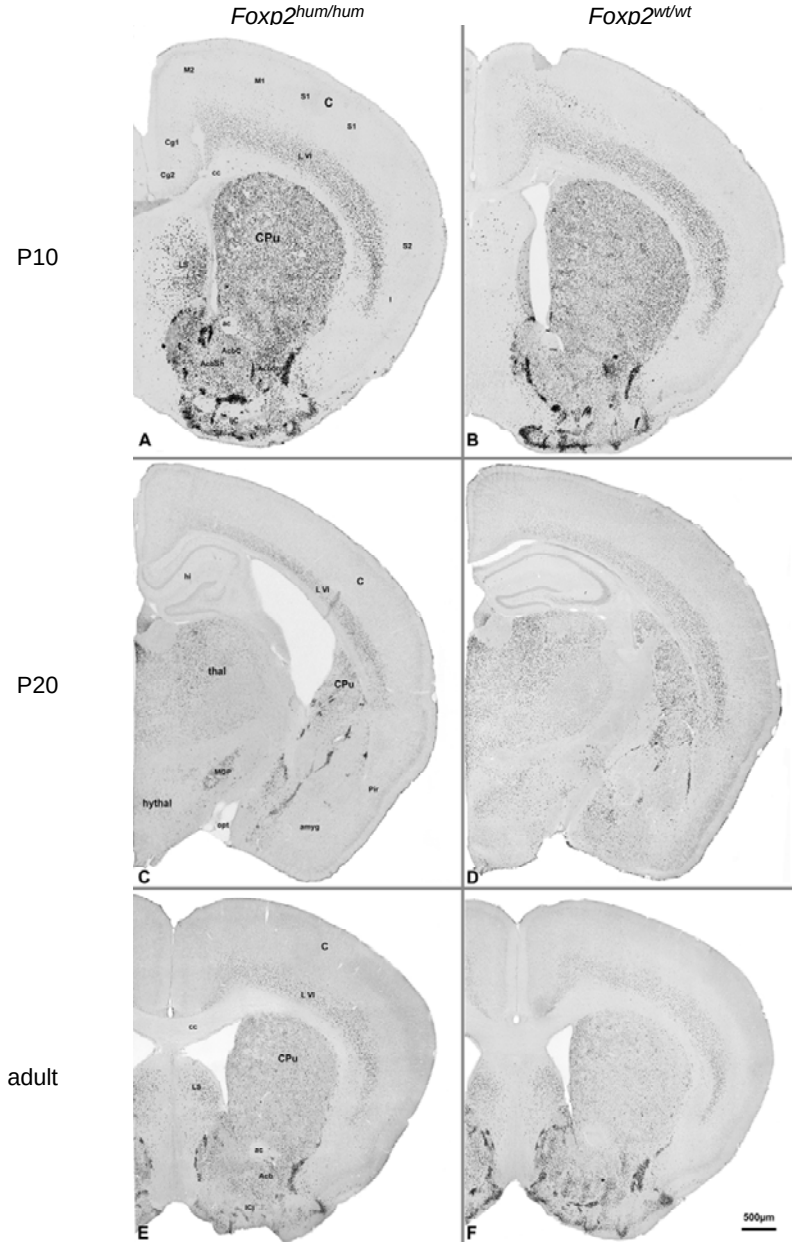


Figure S5: Representative Foxp2 stainings at P10, P20 and 3 month. Shown are 12µm coronal cryosections (**A**,**B**) approximately corresponding to Bregma 0.98 of the adult mouse brain (Paxinos & Franklin 1997) and 30µm cryosections (**C**-**F**) corrsponding to Bregma -1,34 (**C**), -1.46 (**D**) and Bregma 0.38 (**E**,**F**) stained with FOXP2 antibody (Abcam). Foxp2 staining is observed at all three stages for the striatum (Cpu), the cortical layer VI (LVI) from the middle line of the primary (M1) and secondary (M2) motor cortex, the primary (S1) and secondary somatosensory cortex (S2) and the insular cortex (I). Laterally, the nucleus accumbens (Acb) shows labeling that is different between its shell (AcbSh) and its core (AcbC)). Labelling is also evident in the islands of Calleja (IjC), the medial globus pallidus (MGP), the lateral septal nucleus (LS), the thalamic (thal) and hypothalamic (hythal) nuclei. Abbreviations: anterior commissure (ac), amygdala (amyg), cortex (C), corpus callosum (cc), cingulate cortex area 1 (Cg1) and area 2 (Cg2), hippocampus (hi), optic tract (opt), piriform cortex (Pir).

Enard et al. Figure S6

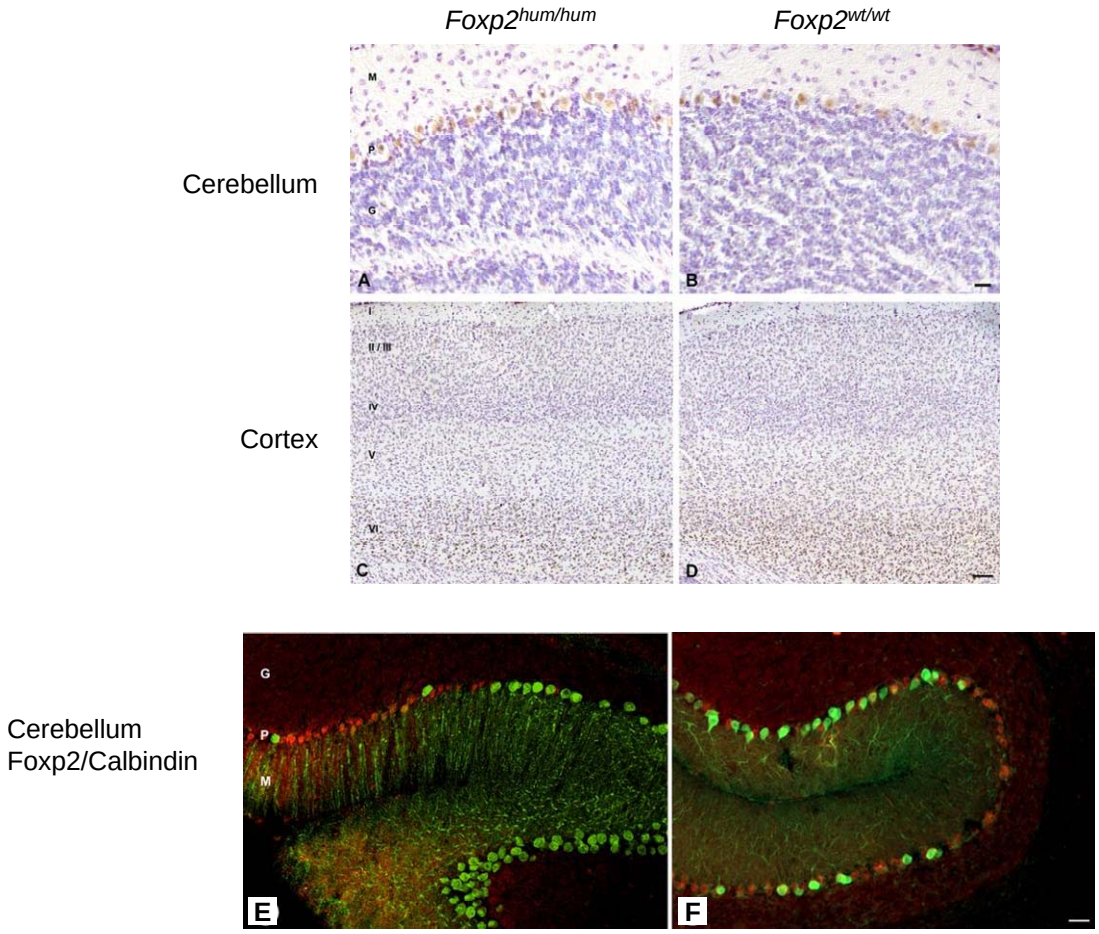


Figure S6: Details of Foxp2 staining. (A-D) 30µm cryosections of P20 pups were stained with Foxp2 (brown) and Nissl. Strong staining can be seen in the Purkinje cell layer (P) of the cerebellum (**A**&**B**; scale 20 µm) and in layer VI neurons of the somatosensory cortex (**C**&**D**, scale 100µm). Foxp2 staining is visualized with DAB (brown) in Nissl-stained slices. (**E-F**) 30µm cerebellar cryosections (scale 20µm) of P20 pups were double-stained with antibodies against Foxp2 (red) and Calbindin (green). The dendrites of calbindin-positive cells are visible in the molecular layer. Abbreviations: molecular layer (M), Purkinje cell layer (P), granule cell layer (G) and I to VI tag the cortical layers.

Enard et al. Figure S7

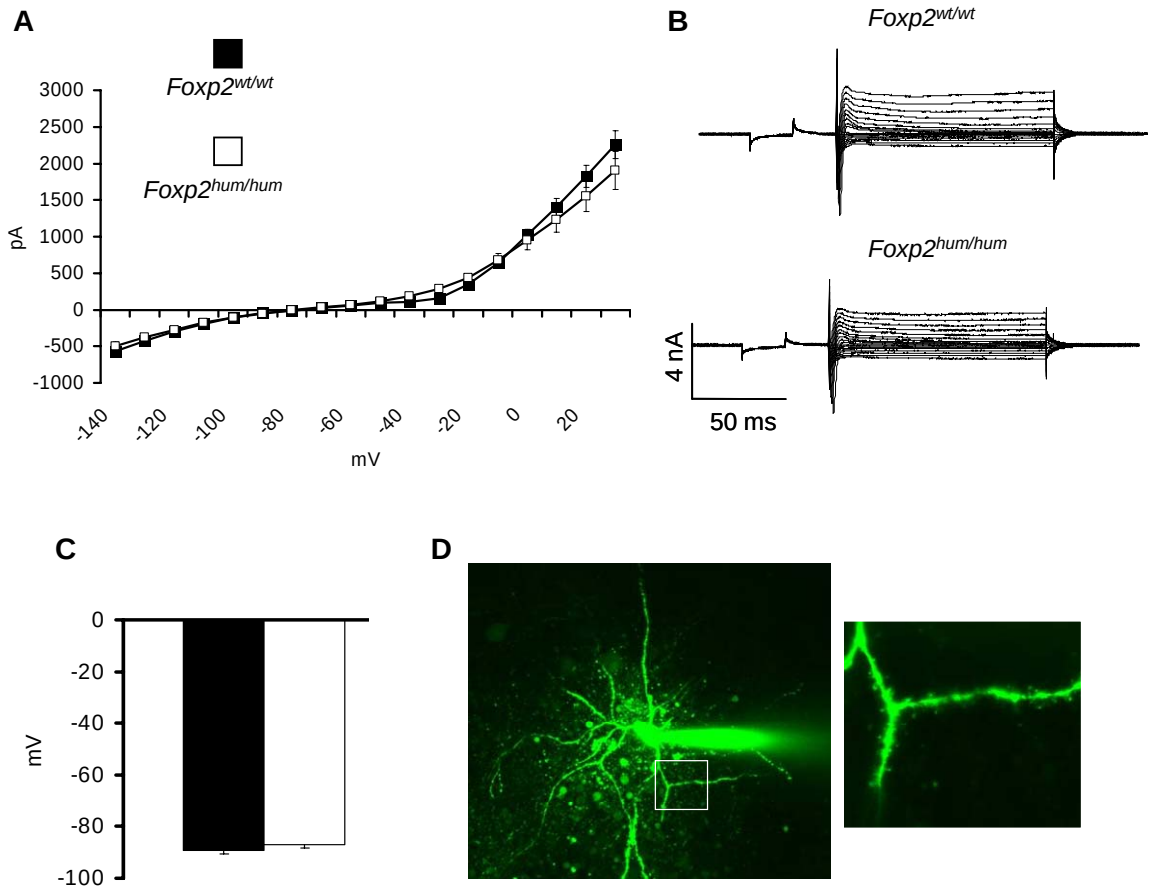


Figure S7: Standard current-voltage relations as well as resting membrane potential are not different between genotypes. Whole cell recordings of striatal neurons in acute slices showed normal magnitudes of sodium and potassium currents (panel **A** and **B**). The current-voltage relationship represents the terminal 20 ms of the recording. The resting membrane potential (panel **C**, current clamp) did also show no differences between genotypes. Panel **D** shows a representative example of one of the recorded neurons with a typical morphology of a medium sized neuron bearing multiple spines on its dendrites.

Enard et al. Figure S8

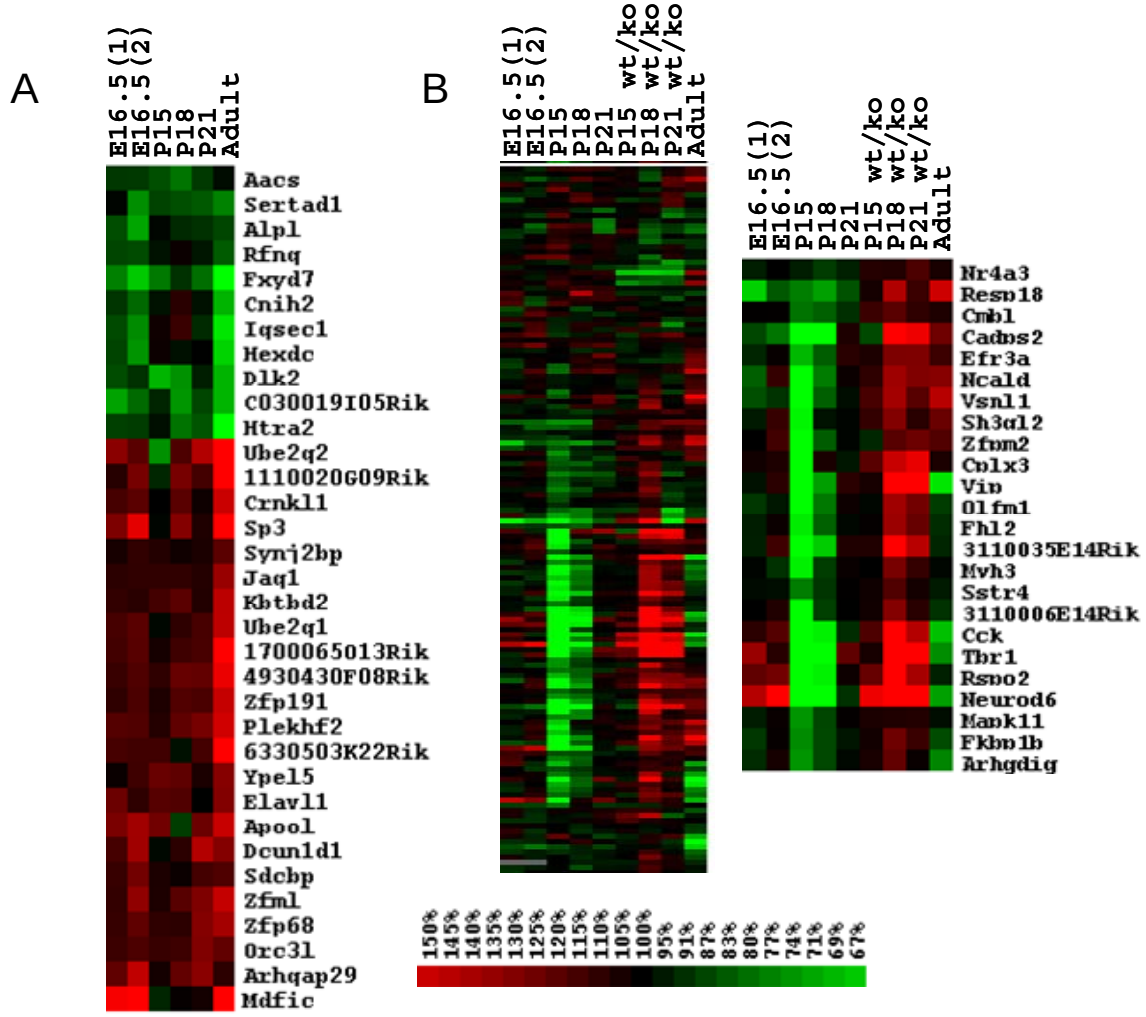


Figure S8: Differences in striatal gene expression patterns. (A) All genes significant ($p < 0.001$) for the factor genotype in the comparison of all *Foxp2*$^{hum/hum}$ mice and their wildtype littermates are shown. Expression ratios are shown relative to wildtype littermates and are normalized for age, sex and batch effects. The two separate batches of embryos are displayed seperately. Red indicates higher expression than in wildtype littermates and green lower expression as indicated in the scale. **(B)** The left panel shows all 137 genes that are preferentially expressed in Drd1a-positive medium spiny neurons (Heiman et al., 2008) and have detectable expression levels in our data. The right panel shows the subset of 24 genes whose expression is increased in *Foxp2*$^{wt/ko}$ mice (15-21 days old) and decreased in *Foxp2*$^{hum/hum}$ mice (15-21 days old) .

Enard et al. Figure S9

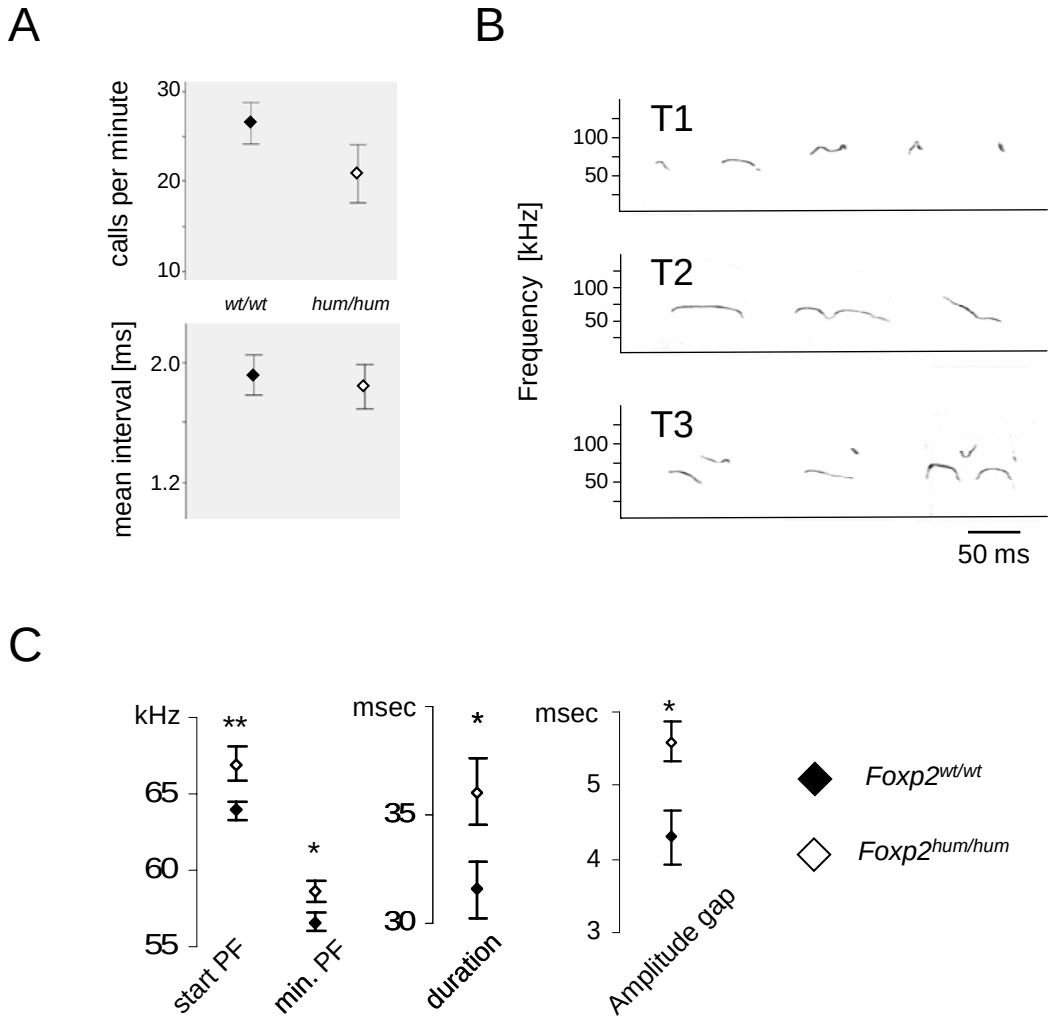


Figure S9: Ultrasonic vocalization (A) Effect of *Foxp2*^{hum} on number of calls and call intervals. Measurements show mean and standard error of the mean. **(B)** Examples of the three vocal types: Vocal type 1 (T1) are calls with no frequency jumps shorter than 50ms. Vocal type2 (T2) are calls with no frequency jumps longer than 50ms.Vocal type 3 (T3) are calls with frequency jumps. **(C)** Analysis of call types with frequency jumps. Plots of acoustic parameters differing between genotypes. Measurements are averaged across days and individuals (± S.E.M.). The peak frequency (PF) refers to the frequency with the maximum amplitude in each analyzed time window of 0.21 milliseconds. Stars indicate significant differences between genotypes (* <0.05; ** <0.01).