Supplementary Information: Structure and function of neonatal social communication in a genetic mouse model of autism  T. Takahashi, S. Okabe, P. Ó Broin, A. Nishi, K. Ye, M. V. Beckert, T. Izumi, A. Machida, G. Kang, S. Abe, J. L. Pena, A. Golden, T. Kikusui, N. Hiroi

Supplementary Figures. S1 to S8
Supplementary Table S1
Supplementary Experimental Procedures
Supplementary Figure S1. Duration, pitch and peak amplitude of calls. a) The mean (+SEM) duration of each of 10 vocal call types. As homogeneity of variance was violated (Cochran’s C=0.17, P<0.01), data were square-root transformed before analysis; for clarity, the averages of raw data are shown. Genotype was used as an independent factor and age and call types were used as repeated factors in a two-way, one repeated measure design ANOVA. Interactions between genotype and age (F(1,37)=4.91, P = 0.0329) and between genotype and call type (F(9,333)=11.00, P< 0.0001) were significant. ** indicates a statistically significant difference between wild-type and heterozygous pups at 1%, as determined by Newman-Keuls post-hoc comparisons. The mean (+SEM) of pitch (b) and peak amplitude (c) of vocalization at P8 (postnatal day 8). The two genotype groups were indistinguishable for pitch (B, t(24)=1.56, P = 0.13) or peak amplitude (c, t(24)=1.97, P = 0.0601), as determined by two-sided Student’s t-tests. In c, the values are indicated relative to the maximum amount of sound amplitude accurately detectable by the microphone used. Abbreviations: Cx, complex; Ts, two-syllable; Fs, frequency steps; Ha, harmonics; C, composite; and H, hump (a.k.a., chevron); Sh, short; D, downward; F, flat; and U, upward.
Supplementary Figure S2. Representative spectrogram. A spectrogram of a P8 wild-type pup is shown.
**Supplementary Figure S3. Entropy scores of vocal call numbers of raw data.** A two-way, one repeated measure ANOVA with genotype as an independent factor and H levels as a repeated measure determined that genotype effect (F(1,26)=13.11, *P* = 0.0012) and model order effect (F(4,104)=262.65, *p* < 0.0001), and interaction between genotype and model order were significant (F(4,104)=4.79, *P* = 0.0014). ++ indicates a statistically significant difference between wild-type and heterozygous pups at 1%; * and ** indicate statistically significant difference at an H[x] level from H[x-1] level at 5% and 1%, respectively, as determined by Newman-Keuls post-hoc comparisons. Abbreviations: WT, wild type; HT, heterozygous.
Supplementary Figure S4. Cross validation of two-call combinations. A leave-one out cross validation of two-call combination within strings was performed using models incorporating different numbers of components and calls. The lowest predictive error rate was seen with a seven component model with five two-call combinations.
Supplementary Figure S5. Proximity of pups to the group average. a) In raw data, pup IDs A340 and A338 most closely represented the averages of wild-type (WT) and heterozygous (HT) groups, respectively. b) These two pups were also among the three pups most resembling the call sequence proportion average of each genotype groups as determined by the Markov model.
Supplementary Figure S6. Ranking of random sequences.

The degree of difference from the original wild-type call sequence among the top ten random sequences out of 100 generated is shown. Abbreviations: Cx, complex; Ts, two-syllable; Fs, frequency steps; Ha, harmonics; C, composite; and H, hump (a.k.a., chevron); Sh, short; D, downward; F, flat; and U, upward.
Supplementary Figure S7. Ranking and percentages of pup call strings. a) Ranking of the average numbers of 10 call types within strings at P8 (postnatal day 8). b) Cumulative percentages indicate that the most frequently used 7-8 call types accounted for the majority of data. Abbreviations: Cx, complex; Ts, two-syllable; Fs, frequency steps; Ha, harmonics; C, composite; and H, hump (a.k.a., chevron); Sh, short; D, downward; F, flat; and U, upward. WT, wild-type and HT, heterozygous.
Supplementary Figure S8. String data loadings plots: Correlation circle produced by the 'plotVar' function of the R 'mixomics' package indicating the correlations between each of the variables (two-call sequences) and the selected components from the associated scores plot in Fig. 2e. Abbreviations: Cx, complex; Ts, two-syllable; Fs, frequency steps; Ha, harmonics; C, composite; and H, hump (a.k.a., chevron); Sh, short; D, downward; F, flat; and U, upward.
Supplementary Table S1.

| Mice group     | cross point (msec) |
|----------------|--------------------|
| WT P8 Female   | 235.39             |
| HT P8 Female   | 378.65             |
| WT P12 Female  | 285.75             |
| HT P12 Female  | 327.28             |

A cross point between the theoretical and observed distribution curves of inter-call intervals is indicated for each genotype/age. Calls separated by intervals longer than the cross point were not included in a sequence.
Supplementary Experimental Procedures

**Mouse:** We used female congenic Tbx1 heterozygous pups and their wild-type littermates with C57BL/6J background for vocal call recording. Heterozygous and wild-type littermates from 5 or more cohorts were tested. Female C57BL/6J mice (Japan Clea Co. Ltd, Yokohama, Japan) were used to test the effects of vocal calls on maternal approach. While there is a known high boy to girl ASD ratio, such a difference disappears particularly for social communication when samples are matched and controlled for IQ. Moreover, there is no male predominance of ASD among carriers of 22q11.2 hemizygosity. All mice were placed in a temperature-controlled room and given access to food and water ad libitum and kept under a 14 h:10 h light/dark cycle. Female mice were separated from male breeders when identified as pregnant and placed singly in a cage. All pups were born in heterozygous male-female breeders and litters were maintained with their own mothers till they were weaned. Female C57BL/6J mothers were used for analysis of maternal approach to pup vocal calls. Animal handling and use followed protocols approved by the Animal Care and Use Committees of the Albert Einstein College of Medicine and Azabu University, in accordance with NIH guidelines.

**Recording of pup vocal calls:** Congenic Tbx1 heterozygous pups and their wild-type littermates were individually recorded for vocalization within ±12 hrs of postnatal day 8 or day 12, as described elsewhere. Briefly pups were separated from dams and placed in the recording apparatus (26 °C) for 5 min. We used an UltraSoundGate condenser microphone capsule (CM16, Avisoft, Germany) connected to a computer equipped with Avisoft-Recorder software (Version 3.2) for recording, as described in our previous publication. The sampling rate was set at 300 kHz and format 16 bit. The recording and display range were set from 0 to 150 kHz. For analysis, a fast Fourier transformation (FFT) was applied to recorded signals using Avisoft SASLab Pro (Version 4.40). We analyzed the frequency range from 20kHz to 120 kHz, and used the lower cut-off frequency of 15 kHz to reduce background noise outside the relevant frequency band. Spectrograms were generated with a FFT length, 256 points, a time
window overlap of 50% with a 100% frame, a flat-top window, a frequency resolution of 488 Hz and a time resolution of 1 ms.

We previously analyzed the number and duration of vocal calls at P8 \(^1\), but for the present study we analyzed new P12 data and re-analyzed the raw P8 data set and applied a whole new set of statistical and functional analyses to both age data. While the percentages of each call type (i.e., number of each call type/total number of call types) were also used for analyses in other studies to reduce variability, we did not use this parameter here, as the actual numbers of each call type were determinants for how they are expected to be theoretically spaced and how the actual calls were spaced as sequences.

There are many classification schemes based on acoustic parameters, including frequencies at the start, end and peak of a call, frequency modulation and duration. Some schemes categorize calls into three \(^6\), five \(^7,8\), nine \(^9,10\) or ten or more distinct call types \(^11,12\). We used the call classification scheme developed by Scattoni and colleagues \(^11\) for two reasons. First, grouping different call types as one would mask otherwise detectable differences \(^13\); the more call types are categorized, the more sensitive the classification becomes in detecting subtle, type-specific differences. Second, this call classification and its variations reliably detect type-dependent atypicalities in pups of genetic mouse models of ASD \(^1,11,14-16\). Other classification schemes have not been used to detect atypical pup vocalization during maternal separation. Data on spectrograms were examined by an experimenter blinded to genotype and each call type was categorized using the same parameters of each of the ten distinct call types \(^11\). Our PLS-DA analysis validates this classification scheme in that the call types were well separated (see Figure 1c), thereby justifying the use of this classification scheme, insofar as they do not justify grouping some of these into a single call type.

Partial Least Square Discriminant Analysis (PLS-DA): To extract statistical components that best
explain the covariance between the 10 different call types and genotype/age, we applied this analysis using the publically available software package DiscriMiner (http://gastonsanchez.com/software/). We excluded from analysis pups that emitted less than 10 calls. The proportions of 10 call types of each pup were used as the predictors for genotype and age. We evaluated the PLS-DA models with different number of components. The first 4 components explained 22%, 11%, 5% and 5% of variance, respectively; the 5th-9th components collectively added no more than 1%. The prediction error with only one component is 0.30 (11 out of 37 observations). It is 0.27 with 2 components, and 0.24 with 3 or 4 components, 0.22 with each of 5 to 9 components, indicating that the prediction error reached a plateau at 5 components. We also applied leave-one-out cross-validation. The error rate was 0.43 with 2 components, 0.41 with 3 components, and 0.38 with 4 components. The first two components most effectively explain the covariance in call proportions between the 10 different call types and genotype/age.

Definition of a string: To eliminate widely separated calls in a sequence, we first computed the distribution of the theoretical and observed distribution curve of inter-call intervals. The theoretical distribution was computed as a function of the number of intervals with a given number of calls emitted within a fixed 5-min recording time. The observed distribution yielded one peak between 100 and 1000 msec and a much smaller peak between 10 and 100 msec. We statistically defined a string as a series of calls that occur with an inter-call interval smaller than the cross point of the theoretical and observed distribution curves. By definition, the string differs from other measures of a cluster of calls (e.g., bout, phrase and burst) that do not eliminate a series of calls that can be explained by a random distribution of inter-call intervals.

Shannon entropy analysis: The entropy rate was used to determine the non-random nature of call sequences. The entropy of a signal is a measure of its associated uncertainty. If there is no inherent
structure, i.e. random order among vocal calls, the entropy rate remains flat when we consider the signal at different entropy orders. If there is some level of signal embedded in call sequences, a deviation from this constant is expected. By assessing the entropy rate using models that incorporate increasing amounts of contextual information about the relative frequencies of call type configurations, we determined the entropy level at which this structure appears.

In the zero-order model, the entropy (in bits) was simply calculated as the log of the number of call types used (denoted $m$):  
$$H[0] = \log_2 m$$

In the first-order model ($H[1]$), each of the call types was considered statistically independent and the entropy is based on the single call numbers:

$$H[1] = -\sum_{i=1}^{m} (p_i) \log_2 (p_i)$$

In the second-order model, conditional probabilities (where $p_{j|i}$ indicates the probability of observing call type $j$ given that call type $i$ has just been emitted) were incorporated to expand the entropy calculation to include two-call numbers:

$$H[2] = -\sum_{i=1}^{m} (p_i) \sum_{j=1}^{m} (p_{j|i}) \log_2 (p_{j|i})$$

Similarly, the third-order entropy model included a component $p_{k|j,i}$, the conditional probability of observing call type $k$, given that calls $j$ and $i$ preceded it:

$$H[3] = -\sum_{i=1}^{m} (p_i) \sum_{j=1}^{m} (p_{j|i}) \sum_{k=1}^{m} (p_{k|j,i}) \log_2 (p_{k|j,i})$$

The fourth-order entropy model was an expansion of the above to calculate the sum of the probabilities of each call type based on the probabilities of the three preceding calls:

$$H[4] = -\sum_{i=1}^{m} (p_i) \sum_{j=1}^{m} (p_{j|i}) \sum_{k=1}^{m} (p_{k|j,i}) \sum_{l=1}^{m} (p_{l|k,j,i}) \log_2 (p_{l|k,j,i})$$
Markov model: To describe the connections and their directions between call types, we used Markov chains. We assumed that when a pup emits a given call-type, it is in an explicit state, and then calculated the probability that a mouse would transition to another state (S₂) from this first state (S₁). We additionally assumed that the probability of S₂ being chosen is dependent solely upon the call-type of S₁. The probability for each transition (Sₜ) is calculated using custom-written MatLab routines (Mathworks). The number of occurrences of each Sₜ is tallied and then divided by the total number of times the call-type of S₁ was made.

Sparse Partial Least Squares Discriminant Analysis (sPLS-DA): To objectively extract statistical components that best explain variance in two-call sequences within strings and separate genotypes, we used the R 'mixomics' software package to apply sPLS-DA to a matrix of bi-gram frequencies, with each row in the frequency matrix, X, being associated with an entry in the vector of class labels, Y. This method incorporates an L1 (or Lasso) regularization when computing the Singular Value Decomposition (SVD) of the covariance matrix simultaneously performing classification and variable selection. This allowed us to identify vocal call sequences that provided the best class separation between wild-type and heterozygous calls and to measure the relative importance of these features on class prediction.

We first filtered out call sequences that show near-zero variance. These calls are not likely to be of value as predictors due to having either only one unique value, or relatively few unique values. In order to identify both a) the optimal number of components needed to explain the dependent variable, and b) the number of variables to keep in the regression model, (the regularization parameter λ is automatically chosen by the algorithm based on these selections), we performed a leave-one-out cross-validation to assess the overall model error rate for each combination.
After cross validation, we examined the scores and loadings plots for the sPLS-DA analysis using the selected number of components and call types. The scores plot provides a summary of the relationship between the samples and shows their projection into the new space defined by the indicated pair of latent variables, while the loadings plot shows the relationships between the predictor variables and their contribution to this projection (Supplementary Figures S4, S8). We also attempted to classify observations based on both age and genotype. The cross validation process indicated that the best predictive model was obtained by selecting 25 two-call sequences on a single component. This model achieved an error rate of 0.375, a rate much higher than that of the model built on genotype alone (see Supplementary Figure S4 for comparison). This is primarily due to the fact that heterozygous pups did not differ from P8 to P12, making it more difficult to distinguish vocalization sequences by age.

**Code availability:** Codes for PLS-DA, Shannon entropy analysis, Markov modeling and sPLS-DA were custom-written. The codes are available upon request.

**Maternal approach:** The test apparatus was a modified Plexiglas mouse home cage (see Figure 3b). One ns-Si sound emitter was placed facing the wire mesh net-covered end of one of the two tubes. The odor of bedding, in the form of bedding itself and cotton smeared with bedding, was placed between the emitter and the mesh. Ultrasonic vocal calls were played back using nc-Si emitter. The emitter was composed of a surface-heating thin film electrode, a nanocrystalline silicon (ns-Si) layer, and a single-crystalline silicon wafer. The electrical signals generated were supplied as voltage inputs through a high frequency amplifier. This device is capable of efficiently generating ultrasound waves using heat transfer at the surface into air and does not use any mechanical vibration. It maintains a constant sound amplitude up to 160kHz. This is technically critical because components of complicated vocal waves are emitted at high pitches above 100kHz. It more faithfully reproduces pup ultrasonic vocal calls than commercially available speakers, which do not maintain a constant amplitude.
at pitches higher than 100kHz and therefore artificially over- and under-represent low and high
frequency components, respectively, for each call. To evaluate the accuracy of sound emitted from our
device, ultrasonic sound was monitored by a condenser microphone (CM16/CMPA, Avisoft
Bioacoustics, Germany). An amplifier (UltraSoundGate116H, Avisoft Bioacoustics, Germany)
analyzed data using an analog-to-digital converter, frequency filters, digital first-Fourier-transform
analysis, and signal input-output terminals. Input signals were monitored with spectrum software
(Avisoft-SASLab Pro 3.0, Avisoft Bioacoustics, Germany). The calls played back were presented as the
original calls in terms of call types, duration, intervals and amplitude.

Calls from many pups cannot be presented to a single female mouse, because females tend to show a
rapid habituation to repeated presentations of calls 20. Use of calls of all pups is also impractical, as our
pilot study showed that each call sample requires 10 or more mothers to achieve statistical significance
due to inherently variable responses among different mothers. Thus, we chose calls of a single
representative wild-type pup and a single representative heterozygous pup. To do this, we computed the
difference scores (D) between each pup and the group average in the proportion of each of the ten call
types. This difference is squared (D^2) to reflect a deviation from the average regardless of the direction
(i.e., higher or lower than the average). These scores were summed for all call types, yielding a
deviation score (Σ^2). The pup with the smallest Σ^2 score was chosen from each genotype
(Supplementary Figure S5a). We further validated that these mice are among pups with the Markov
values closest to the averages (Supplementary Figure S5b).

To choose one randomized sequence of the representative wild-type calls, we generated 100 randomized
sequences of the representative wild-type calls used in our experiment. The percentage of each two-call
combination was computed and compared to that of the wild-type calls. We then compared 100
randomized sequences with the original wild-type sequence structure in terms of 1) the most frequent
combination type (i.e., Cx-Cx), 2) the four most frequent combinations (i.e., Cx-Cx, H-H, Ts-Ts, H-Cx) and the seven most frequent call combinations (i.e., Cx-Cx, H-H, Ts-Ts, H-Cx, Cx-F, Cx-H, Fs-Cx) (see **Supplementary Figure S6**). A random call sequence #93 had the least degree of resemblance to the original wild-type two-call call sequences and was chosen as the exemplar random call sequence.

Virgin male and virgin female C57BL/6JCl mice (Japan Clea Co. Ltd, Yokohama, Japan) were paired at ≥8 weeks of age and pregnant female mice were separated singly when a vaginal plug was noted. The lactating mother was used at 5-7 days postpartum. The basic procedure is detailed in our published work 17,18. Day 1 consisted of habituation in the home cage placed in the soundproof room. Days 2 and 3 were test days. On day 2, after the mother was placed in the test apparatus for 30 min in the soundproof test box, we exposed the mother to vocal calls for 5 min (**Figure 3a**). On day 3, the mother was similarly tested with the call type not used on day 2. Each mother was tested with the vocal sound of a wild-type pup and that of a heterozygous pup (or the vocal sound of a wild-type pup and that of randomized wild-type calls) in the two sessions. Half of the mothers were presented with wild-type calls on day 2 and heterozygous calls (or randomized wild-type calls) on day 3; the other half was exposed in the opposite order. Different sets of lactating mothers were used for two sets of experiments: 1) consecutive presentation of wild-type calls and heterozygous calls and 2) consecutive presentation of wild-type calls and randomized wild-type calls. Wild-type, heterozygous or randomized wild-type calls were played back at the end of one of the two tubes; no sound was presented at the other tube. The position of the sound tube was counterbalanced. Simultaneous playback of a pair of two sound types from the two tubes is not ideal. First, such a presentation induces far fewer maternal responses from mothers 21. Second, as the response to one call type invariably affects the time simultaneously spent responding to the other call type, the impact of each sound cannot be accurately evaluated. Third, the degree of merging of two different sounds in the center of the apparatus has never been established, and the accidental position of a mother at the onset of the sounds in relation to the relative vicinity of the mother to the two sound
sources cannot be controlled. Fourth, we monitor the accuracy of reproduced sounds. If two sounds are simultaneously played back, the accuracy of each reproduced sound cannot be evaluated. Smell of the mothers’ own pups was placed between the emitter and the wire mesh at the end of both tubes. Although the odor stimulus does not serve as a discriminative stimulus between the two tubes, it is necessary to ensure that call sound reliably induces maternal approach\textsuperscript{17,18}. The mother mouse was allowed to freely choose, move and stay anywhere in the apparatus during the 5-min test. After testing, the mother was returned to the home cage and placed in the sound proof room. The mother’s behavior was recorded by a video camera. Recorded behavior was rated by an observer blinded to the genotypes of the pup calls. Data were analyzed for the time the mother peeked at the entrance of each tube, and the time the mother spent at the closest vicinity to the emitter in the sound tube, and the latency to enter the tubes.

**Other statistical analyses**

For detection of statistically significant difference between groups, we used ANOVAs, followed by Newman-Keuls post-hoc comparisons. When two groups were compared, two-sided Student t-test was used. When homogeneity of variance was violated, square-root transformation was applied for statistical testing. The non-parametric Kolmogorov-Smirnov test was used to compare the frequency distribution of two groups. When normality was violated, data were analyzed using the Wilcoxon signed rank order test. A probability of 0.05 or less was considered to be significant.
References

1. Hiramoto T, Kang G, Suzuki G, Satoh Y, Kucherlapati R, Watanabe Y et al. Tbx1: identification of a 22q11.2 gene as a risk factor for autism spectrum disorder in a mouse model. *Hum Mol Genet* 2011; 20: 4775-4785.

2. Andersson GW, Gillberg C, Miniscalco C. Pre-school children with suspected autism spectrum disorders: do girls and boys have the same profiles? *Res Dev Disabil* 2013; 34: 413-422.

3. Harrop C, Shire S, Gulsrud A, Chang YC, Ishijima E, Lawton K et al. Does gender influence core deficits in ASD? An investigation into social-communication and play of girls and boys with ASD. *J Autism Dev Disord* 2015; 45: 766-777.

4. Volkmar FR, Szatmari P, Sparrow SS. Sex differences in pervasive developmental disorders. *J Autism Dev Disord* 1993; 23: 579-591.

5. Schneider M, Debbane M, Bassett AS, Chow EW, Fung WL, van den Bree MB et al. Psychiatric Disorders From Childhood to Adulthood in 22q11.2 Deletion Syndrome: Results From the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *Am J Psychiatry* 2014; 171: 627-639.

6. Young DM, Schenk AK, Yang SB, Jan YN, Jan LY. Altered ultrasonic vocalizations in a tuberous sclerosis mouse model of autism. *Proc Natl Acad Sci U S A* 2010; 107: 11074-11079.

7. Branchi I, Santucci D, Vitale A, Alleva E. Ultrasonic vocalizations by infant laboratory mice: a preliminary spectrographic characterization under different conditions. *Dev Psychobiol* 1998; 33: 249-256.

8. Panksepp JB, Jochman KA, Kim JU, Koy JJ, Wilson ED, Chen Q et al. Affiliative behavior, ultrasonic communication and social reward are influenced by genetic variation in adolescent mice. *PLoS One* 2007; 2: e351-

9. Portfors CV. Types and functions of ultrasonic vocalizations in laboratory rats and mice. *J Am Assoc Lab Anim Sci* 2007; 46: 28-34.

10. Sugimoto H, Okabe S, Kato M, Koshida N, Shiroishi T, Mogi K et al. A role for strain differences in waveforms of ultrasonic vocalizations during male-female interaction. *PLoS One* 2011; 6: e22093-

11. Scattoni ML, Gandhy SU, Ricceri L, Crawley JN. Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PLoS One* 2008; 3: e3067-

12. Grimsley JM, Monaghan JJ, Wenstrup JJ. Development of social vocalizations in mice. *PLoS One* 2011; 6: e17460-

13. Arriaga G, Jarvis ED. Mouse vocal communication system: are ultrasounds learned or innate? *Brain Lang* 2013; 124: 96-116.

14. Yang M, Mahrt EJ, Lewis F, Foley G, Portmann T, Dolmetsch RE et al. 16p11.2 Deletion Syndrome Mice Display Sensory and Ultrasonic Vocalization Deficits During Social Interactions. *Autism Res* 2015;

15. Ey E, Torquet N, Le Sourd AM, Leblond CS, Boeckers TM, Faure P et al. The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations. *Behav Brain Res* 2013; 256: 677-689.
16. Burkett ZD, Day NF, Penagarikano O, Geschwind DH, White SA. VoICE: A semi-automated pipeline for standardizing vocal analysis across models. Sci Rep 2015; 5: 10237-

17. Uematsu A, Kikusui T, Kihara T, Harada T, Kato M, Nakano K et al. Maternal approaches to pup ultrasonic vocalizations produced by a nanocrystalline silicon thermo-acoustic emitter. Brain Res 2007; 1163: 91-99.

18. Okabe S, Nagasawa M, Kihara T, Kato M, Harada T, Koshida N et al. Pup odor and ultrasonic vocalizations synergistically stimulate maternal attention in mice. Behav Neurosci 2013; 127: 432-438.

19. Kihara T, Harada T, Kato M, Nakano K, Murakami O, Kikusui T et al. Reproduction of mouse-pup ultrasonic vocalizations by nanocrystalline silicon thermoacoustic emitter. Applied Physics Letters 2006; 88: 1-3.

20. Hammerschmidt K, Radyushkin K, Ehrenreich H, Fischer J. Female mice respond to male ultrasonic 'songs' with approach behaviour. Biol Lett 2009; 5: 589-592.

21. Sewell GD. Ultrasonic communication in rodents. Nature 1970; 227: 410-