A mutation associated with stuttering alters mouse pup ultrasonic vocalizations

Terra D. Barnes[a], **David Wozniak**[b], **Joanne Gutierrez**[c], **Tae-Un Han**[c], **Dennis Drayna**[c], and **Timothy E. Holy**[a]

Terra D. Barnes: terra.barnes@gmail.com; Dennis Drayna: drayna@nidcd.nih.gov; Timothy E. Holy: holy@wustl.edu

[a]Department of Neuroscience, Washington University in St. Louis, School of Medicine, Campus Box 8108, 660 S. Euclid Avenue, St. Louis, MO 63110-1093

[b]Department of Psychiatry, Washington University in St. Louis, School of Medicine, Campus Box 8108, 660 S. Euclid Avenue, St. Louis, MO 63110-1093 and the Taylor Family Institute for Innovative Psychiatric Research

[c]National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, Maryland 20892, USA

Summary

A promising approach to understanding the mechanistic basis of speech is to study disorders that affect speech without compromising other cognitive or motor functions. Stuttering, also known as stammering, has been linked to mutations in the lysosomal enzyme-targeting pathway, but how this remarkably specific speech deficit arises from mutations in a family of general "cellular housekeeping" genes is unknown. To address this question, we asked whether a missense mutation associated with human stuttering causes vocal or other abnormalities in mice. We compared vocalizations from mice engineered to carry a mutation in the *Gnptab* (N-acetylglucosamine-1-phosphotransferase subunits alpha/beta) gene with wild type littermates. We found significant differences in the vocalizations of pups with the Gnptab stuttering mutation compared to littermate controls. Specifically, we found that mice with the mutation emitted fewer vocalizations per unit time, had longer pauses between vocalizations, and that the entropy of the temporal sequence was significantly reduced. Furthermore, Gnptab missense mice were similar to wild type mice on an

extensive battery of non-vocal behaviors. We then used the same language-agnostic metrics for auditory signal analysis of human speech. We analyzed speech from people who stutter with mutations in this pathway and compared it to control speech, and found abnormalities similar to those found in the mouse vocalizations. These data show that mutations in the lysosomal enzyme targeting pathway produce highly specific effects in mouse pup vocalizations, and establish the mouse as an attractive model for studying this disorder.

Abstract
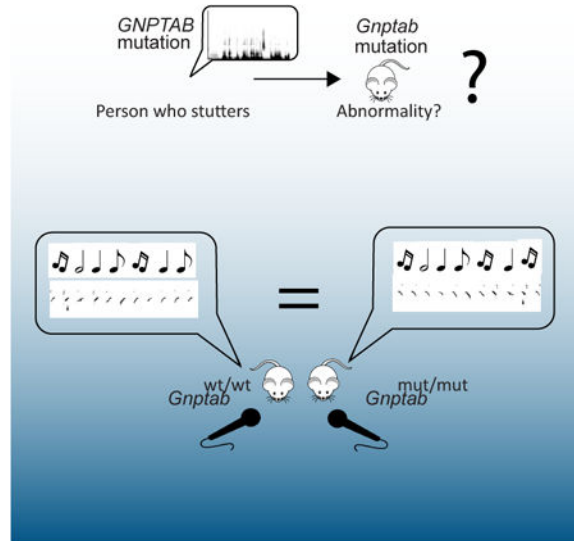
Introduction

Speech disorders affect millions of people worldwide and are usually treated with behavioral therapy[1, 2]. A better understanding of human speech disorders may lead to a wider array of treatment options, and can provide insights into the genetic and neural underpinnings of human speech. To date, there are remarkably few speech disorders with a clearly identified genetic component [3-5]. Developmental stuttering is one of the most common speech disorders, with the persistent form of the disorder affecting 3 million adults in the United States. The disorder is characterized by frequent repetitions or prolongations of syllables or words, or speech that has frequent hesitations or pauses – known as "blocks" – that disrupt the smooth flow of speech [1]. As an inroad into mechanisms of speech production, stuttering is remarkable because affected individuals are, on average, normal by all other known measures of language, cognitive, and motor function [1]. Recently, it was found that persistent developmental stuttering unaccompanied by other deficits or symptoms can be linked to mutations in the lysosomal enzyme-targeting pathway (LETP) [6]. Mutations in the LETP pathway account for 9-16% of all cases of persistent nonsyndromic stuttering[7].

Lysosomes are degradative cellular organelles that contain acid hydrolases. These enzymes are targeted to lysosomes by tagging them with mannose 6-phosphate[8]. The mannose 6-phosphate is added to these acid hydrolases in a two-step process carried out by the products of three genes. Mutations in all three of these genes have been found in humans who stutter

[6, 7]. However, the mechanisms by which mutations in an apparent "housekeeping gene" produce a deficit with the remarkable specificity of stuttering remain unclear.

While humans are irreplaceable for discovery of the genetic underpinnings of speech disorders, they have a number of disadvantages for analysis of the underlying mechanisms. An animal model would be an invaluable tool for investigations into the circuit and cellular mechanisms of a vocal disorder, despite the obvious difficulties in comparing human speech with any animal vocalization [3]. While no exact correspondence with human speech can be expected, it is possible that one or more "low-level" features—such as planning, initiation, timing, breath control, and/or temporal sequencing—may be shared both in phenotype and underlying neuronal circuitry across mammalian species.

Mouse vocalizations have been extensively studied and have been used successfully in the study of several disorders [9-13]. Mice produce ultrasonic vocalizations in a range of social situations. These vocalizations have repeated syllables and complex structure, and have been characterized as "songs" [14]. One type of innate mouse vocalization is the isolation call of pups which has also been extensively studied, especially in disease [3, 11, 15-20]. Mouse pups, when separated from their mother during the first two weeks of life, spontaneously vocalize [20]. The rate of vocalization peaks at around 5-7 days postnatal and decrease thereafter.

This approach is exemplified by studies of the speech disorder associated with mutations of the FOXP2 transcription factor. The *Foxp2* gene is conserved across species [21]. In mice, disruption of the *Foxp2* gene has been shown to cause changes in pup isolation calls [15-17, 22] (though not necessarily in adults [23]); in turn, such studies have been used to obtain numerous insights into the molecular and cellular functions of this gene and its role in human speech [3, 4, 11, 15-17, 24].

Here, we analyze pup isolation calls and other behaviors of mice with a knock-in mutation in the *Gnptab* gene, for which the equivalent mutation has been shown to cause stuttering in humans [6]. We show that such mice exhibit alterations in timing and sequencing of vocalization that are reminiscent of specific deficits observed in human stuttering.

Results

We engineered mice to carry a homozygous Glu1179Lys mutation in *Gnptab (Gnptab* [mut/mut)], homologous to the Glu1200Lys mutation in human *GNPTAB* well characterized for its role in stuttering [6, 25]. Mice were constructed on a pure BALB/c background (Figure 1A and S1), using a heterozygote × heterozygote breeding strategy to ensure the availability of littermate controls with matched pre- and post-natal environment. We recorded pup isolation calls on postnatal day 3, 5, and 8; data from all days are presented in figure supplements, with data from day 8 presented in the main figures below.

Mouse vocalizations: rate, spectral characteristics, and timing

We first asked whether *Gnptab* [mut/mut] mice vocalized. Both wild-type littermates (*Gnptab* [wt/wt]), and *Gnptab* [mut/mut] mice produced vocalizations over the 3.5 minute recording session

that appeared to have normal amplitude and spectral characteristics (Figure 1B-1D). We found that *Gnptab* mut/mut mice exhibited significantly fewer vocalizations compared to littermate controls (*Gnptab* mut/mut 64.2 ± 34.8 vocalizations; *Gnptab* wt/wt 184.5 ± 39.4 vocalizations; *t*-test p<0.034, Figure 2A) in the three and half minutes of a recording.

The difference in rate of vocalizing in the homozygous mutants could be due to the length of vocalizations increasing, the length of pauses increasing, or a mixture of these. We found no significant difference in the duration of the vocalizations between the two groups of mice (*t*-test, p >0.5, Figure 2B). However, there was a significant difference in the mean pause length between vocalizations (*Gnptab* mut/mut 6.18 ± 1.23 s; *Gnptab* wt/wt 2.72 ± 0.69 s; *t*-test p<0.02, Figure 2C). This increase in the mean pause length in the *Gnptab* mut/mut could be due to increased duration of all pauses or an increase in a specific subset of pauses. Much like humans, mice exhibit "bouts" of vocalizations, separated by longer pauses (see Figure 1B and 1C). We found that the *Gnptab* mut/mut showed no difference in the length of intra-bout pauses, those short pauses between vocalizations inside a bout (Figure S2 and Experimental Procedures). Similarly, *Gnptab* mut/mut had a similar proportion of large pauses when the criterion for large pauses was not stringent (less than 0.8 seconds). However, using more stringent criteria for large pauses (setting the cutoff to 0.8 seconds or higher), the proportion of long pauses was increased in the *Gnptab* mut/mut mice (Figure S3). This result suggests that the key factor contributing to the reduced number of vocalizations in *Gnptab* mut/mut animals compared to their wild-type littermate controls was an increase in the duration of long pauses between bouts of vocalization.

We analyzed the same features of the vocalizations from heterozygous *Gnptab* mut/wt littermates, and found them to be more similar to the wildtype *Gnptab* wt/wt mouse vocalizations (Figure S3). We also found that the vocalization anomalies of the *Gnptab* mut/mut animals became more pronounced from day 3 to day 8 (Figure S3).

To obtain further insight into the changes in number of vocalizations, we also examined the number of bouts per recording and the number of vocalizations per bout in the *Gnptab* mut/mut mice. As expected from the increase in duration of long pauses, the number of bouts per recording was smaller in the *Gnptab* mut/mut mice (*Gnptab* mut/mut 14.74 ± 4.3 bouts; *Gnptab* wt/wt 45.35 ± 7.8 bouts; *t*-test p < .0023) (Figure 3A). There was no difference in the average number of vocalizations per bout (*Gnptab* mut/mut 3.0 ± .3 vocalizations; *Gnptab* wt/wt 3.6 ± .3 vocalizations; *t*-test p > .65, Figure 3B). In the *Gnptab* mut/mut mice, there was a small but significant increase in the percentage of bouts that contained only one vocalization compared to their wild type littermates (*Gnptab* mut/mut 45.5% ± .04%, *Gnptab* wt/wt 34.2% ± .03%, p <0.029). The percentage of bouts that contained two vocalizations was slightly smaller in the homozygous mutants mice (*Gnptab* mut/mut 13.0 % ± .02% *Gnptab* wt/wt 20.0% ± .03, p<0.05). There was no difference in the percentage of bouts that contained 3, 4, or 5 syllables (Figure 3C).

Mouse vocalizations: usage and temporal sequencing of syllables

In humans, some of the disfluencies in stuttering consist of silent blocks, but others—such as sound or syllable repetitions—are voiced. To ascertain whether *Gnptab* mut/mut mice showed such abnormalities in their vocalizations, we analyzed both the usage and temporal

sequencing of syllable types as categorized by an established classification scheme based on presence and size of abrupt pitch jumps [10, 14] (Figure 4A). We chose this scheme because it was initially introduced based on quantitative observations of clustering, which makes it easy to automate and therefore unbiased. Several other syllable classification schemes have been used [11, 12, 18, 23, 26]; these alternative schemes differ primarily in the classification of syllables lacking pitch jumps, but they also exhibit many points of commonality. Wild type *Gnptab* wt/wt and knock-in *Gnptab* mut/mut mice were indistinguishable in the syllable types and their usage (see Experimental Procedures, p>.39, Figure S4). This indicates that these mutations do not affect the ability to produce different syllable types.

We next looked at syllable repetition. Wild type mice normally exhibit a strong tendency towards syllable repetition, with doublets (defined as two of the same syllable type in a row, in bouts of two or more syllables) occurring 38% of the time. We found no significant increase in the presence of doublets (Figure 4C) in the vocalizations of the *Gnptab* mut/mut mice. However, within bouts, *Gnptab* mut/mut mice exhibited a higher percentage of syllables of a single type, suggesting that *Gnptab* mut/mut mice might be more stereotyped in their vocalizations (Figure 4D). We tested this possibility using a first order Markov process model to compare the entropy present in the temporal structure of these vocalizations, and found a tendency towards increased repetition in the vocalizations of the *Gnptab* mut/mut animals (*t*-test, p<0.022, Figure S5), indicating an overall reduction in temporal diversity. Consequently, *Gnptab* mut/mut mice use all syllable types in proportions similar to wild-type, but exhibit higher stereotypy in temporal sequencing.

Non-vocal phenotypes

We also asked whether phenotypic differences between wildtype *Gnptab* wt/wt and homozygous knock-in *Gnptab* mut/mut mice were specific to vocal behavior. *Gnptab* mut/mut pups exhibited normal weights (Figure S5).

We next examined adult wild type and homozygous knock-in mice using a wide variety of behavioral tests. Mice were evaluated first on a 1-h locomotor activity test. There was a significant decrease in ambulatory activity (total ambulations) on the part of the *Gnptab*^{mut/mut} mice across the 60-min test session (Figure 5A, p=0.0498). In contrast, no differences were found between groups with regard to vertical rearing frequency, or time spent, entries made, or distance traveled in the central zone of the test field.

The mice were then tested on a battery of sensorimotor measures to assess possible functional deficits. No performance differences were observed between groups on the ledge (Figure 5B), pole (Figure 5C), inverted screen tests (Figure 5D), or on the walking initiation, platform, or inclined screens tests. These results suggest that balance, strength, co-ordination, and initiation of movement were not grossly disturbed in the *Gnptab*^{mut/mut} mice. These animals and their wild type littermates also performed equivalently on tests of acoustic startle (Figure 5E), prepulse inhibition of startle (PPI) (Figure 5F), and rotarod tests (Figure 5G).

Spatial learning and memory capabilities of the *Gnptab*^{mut/mut} mice were assessed using the Morris water maze task with a computerized tracking system as previously described [27].

Cued (visible platform, variable location), and place (submerged, hidden platform, constant location) trials were conducted. No significant effects involving genotype were found (Figure 5H and 5I).

We next examined the sociability of the mice as assessed by measuring the time investigating a conspecific stimulus mouse that was loosely sequestered in a small withholding cage compared to the investigation times observed for an empty cage and found no significant differences between genotypes [28] (Figure 5J).

Mice were next evaluated for differences in exploratory behaviors and olfactory preference using a previously published procedure [29], where hole poking with the nose served as the main behavioral response. The apparatus contained 4 corner and 4 side holes in the floor. One corner hole contained a familiar odorant and the diagonally opposite corner hole contained a novel odorant. *Gnptab^mut/mut* and control mice performed similarly on several test variables including the frequency of nose pokes made into the side or corner holes, and with regard to pokes made into odorant-containing versus empty corner holes, and for pokes made into the corner holes containing a novel versus a familiar odorant. However, wild type mice showed a significant preference for the hole containing the familiar odorant (p=0.006) while the *Gnptab^mut/mut* mice did not show a significant preference for the familiar odorant, though there was a trend in that direction (p=0.11). In addition, the *Gnptab^mut/mut* mice had decreased poke durations on average across empty and odorant-containing holes as indicated by a significant Genotype effect (Figure 5K, p=0.028). The *Gnptab^mut/mut* and Wild type groups each had significantly longer poke durations for the odorant-containing holes versus the empty ones, (p=0.024 and 0.012, respectively).

People who stutter (PWS) have been shown to have a higher incidence of anxiety disorders [30]. Therefore we assessed anxiety-like behaviors using the elevated plus maze (EPM) and fear conditioning. Analysis of the classic open arm variables in the EPM including time spent, entries made, and distance traveled in the open arms did not yield any significant effects involving genotype (Figure 5L). We also evaluated the performance of *Gnptab^mut/mut* mice on a fear-associated Pavlovian conditioning test and found no genotype differences (Figure 5M).

Human Vocalizations

Given that the deficits of mice were primarily vocal in nature, we compared these vocal deficits with those of PWS. Stuttered and normal speech have been compared using a wide variety of language-based and acoustic analyses [1, 31-37]. To ensure optimal comparability, we applied our mouse-centric analyses to the vocalizations of PWS. Moreover, we were able to analyze vocalizations from normally fluent human controls (CT), PWS, and PWS and carry a well-characterized mutation in one of the three genes encoding the LETP (PWS-L) (Table S1). We thus tested whether the vocal abnormalities found in the *Gnptab* ^mut/mut^ mice could be seen in the more complex vocalizations of humans. We obtained recorded speech samples using a standard 500-word stuttering diagnostic reading passage containing balanced representation of all speech sound classes [38]. To ensure comparability with mouse data, no language-specific criteria were used to determine vocalization segments.

Computer-assigned vocalization segments could therefore include more than one syllable or word.

We compared the rate of vocalizations in PWS versus controls (CT). Consistent with previous studies [1, 39-41], stuttering case subjects had significantly fewer vocalizations per minute (PWS, 90.2 ± 3.0 vocalizations; CT, 124.9 ± 3.1; mean ± s.e.m.; t-test, p< 10^{-11}, Figure 6B). Like the *Gnptab* $^{mut/mut}$ mice, the PWS-L group also displayed significantly fewer vocalizations per minute than controls (PWS-L, 81.1 ± 4 vocalizations per minute; t-test, p< 10^{-11}, Figure 6B). These differences were comparable to the difference found between the *Gnptab* $^{mut/mut}$ mice and *Gnptab* $^{wt/wt}$ mice.

An analysis of the durations of vocalizations revealed no significant difference between the general stuttering cases and controls (PWS mean 0.335 ± 0.020 sec; CT mean 0.313 ± 0.012 sec; t-test, p>0.3). However, there was a significant increase in the duration of the vocalizations in stuttering cases carrying an identified mutation in the LETP (PWS-L mean 0.385 ± 0.033 sec, t-test, p < 0.02, Figure 6C).

A primary feature of stuttering is involuntary pauses known as blocks, and such pauses have been found to be important in assessing fluency of speech [11, 34]. Like the differences in the pause lengths in the *Gnptab* $^{mut/mut}$ mice and *Gnptab* $^{wt/wt}$ mouse vocalizations, the speech of PWS had significantly longer pauses between vocalizations compared to that of control subjects (PWS 0.385 ± 0.021 sec; CT, 0.170 ± 0.005 sec; t-test, p < 10^{-12}, Figure 6D). This was also true for case subjects carrying an LETP mutation (PWS-L, 0.375 ± .038 sec; t-test, p <10^{-9}). Moreover, we found a significantly greater proportion of long pauses in both case groups compared to the control subjects over a wide range of cutoff values for defining "long" pauses (Figure 6E). These data indicate that we can detect a statistically significant difference in specific features of vocalization in PWS using the same methods used to analyze the mouse data. Such measures are "language agnostic" because they do not depend on detailed features (formant structure, vowels and consonants, etc.) of human speech. Using just two such features, number of vocalizations per minute and length of pauses, a k-nearest neighbor classifier could distinguish a PWS from a control subject 79% of the time, significantly better than chance (shuffled data, p < 10^{-5}).

To determine whether these same phenomena apply to stuttering during extemporaneous speech, we also examined the length and number of speech intervals in 80 three-minute samples obtained from publicly-available podcasts, for both controls and PWS. Consistent with the results from people reading a script, we found that in extemporaneous speech, PWS had fewer vocalizations per unit time compared to controls. This difference was significant (PWS 274.7 ± 8.4 vocalizations; CT 434.8 ± 6.5 vocalizations; t-test p < 10^{-12}, Figure 7A). The average length of vocalization also differed significantly between groups (PWS 415 ± 17 ms; CT 258 ± 6; t-test p < 10^{-10}; Figure 7B). We also found that the mean length of pauses between speech intervals was significantly longer in PWS compared to controls (PWS 255 ± 12 ms; CT 160 ± 3; t-test p < 10^{-8}, Figure 7C).

Discussion

Vocal alterations: mice

Identification of some of the genetic underpinnings of human stuttering has created an opportunity to study whether homologous mutations could affect vocalization in a genetically-tractable animal model. We found that the vocalizations of *Gnptab* mut/mut mice were significantly different than those of their *Gnptab* wt/wt littermates. *Gnptab* mut/mut mice had fewer vocalizations per unit time and showed an increased incidence of long pauses. *Gnptab* mut/mut mice produced all the same syllables as their wildtype littermates, and used them in similar proportions. We note that these conclusions are specific to the syllable classification scheme adopted here, and it remains possible that other classification schemes would come to a different conclusion. From manual inspection of the sonograms, at present we have not seen evidence leading us to suspect any kind of difference in syllable repertoire.

However, by two measures mice with an LETP mutation showed more temporal stereotypy in their vocalizations. First, during bouts of vocalization, a higher percentage of syllables were of a single type. Second, the entropy of the first order Markov process was reduced, indicating an overall reduction in temporal diversity.

Vocal alterations: human

The techniques used to analyze mouse vocalization were validated by applying them to human speech from PWS, PWS with a mutation in the LETP, and controls. Stuttered speech has been extensively analyzed using a variety of measures including words per minute, voice onset time, formant frequencies of consecutive glottal periods, and other language-based acoustic techniques [31, 33, 34, 40]. Previous studies have demonstrated that stuttering is strongly correlated with fewer vocalizations per unit time and an increase in pause length [39, 40]. We confirmed and extended these results using language agnostic measures designed to be generalizable to mice. In addition, we found that using just three such language-agnostic features (number of vocalizations per minute, length of syllables and length of pauses), we could classify a recording as coming from a PWS or control.

Stuttering is heterogeneous disorder, and there have been many proposed subtypes of stuttering [42]. Here we have analyzed the speech of a subset of PWS with a genetically identified etiology and compared it to the speech of a general group of PWS. Our results show that using our language-agnostic analyses, PWS with a known mutation in the LETP showed similar abnormalities as those displayed by PWS of an unknown etiology.

Vocal alterations: comparison between mouse and human

Human speech is produced by air flow induced vocal fold vibrations [43]. In contrast, mouse vocalizations are thought to be a form of whistle [44]. Many types of comparisons between human speech and mouse vocalization are unlikely to be informative due to anatomical differences and the complex and communicative nature of human speech. In the case of stuttering, it is important to note that while stuttering may manifest as deficiencies in the control of vocal tract movements, there is no evidence that these are the proximal cause of the disorder. In fact, there is evidence that persons who are deaf or hearing impaired stutter

when using sign language suggesting the cause of stuttering resides elsewhere [45]. In addition, mouse vocalizations require control of airflow, subglottal pressure, and glottal adduction and these share some common neural pathways with human vocal production [3, 10, 44]. Moreover, there is evidence linking these more cognitive pathways to stuttering [46, 47].

GNPTAB mutations exhibit several common features in both organisms. PWS, and specifically subjects carrying an LETP mutation (PWS-L), showed fewer vocalizations per unit time and as fluency blocks, which acoustically correspond to large pauses. Also like the mouse, the tendency towards syllable repetition is a defining characteristic of human stuttering. That some of the prominent features of human stuttering are recapitulated by the mouse model suggests that such a model may be useful in discovering the circuit and neurochemical underpinnings of at least some aspects of stuttering. One interpretation of these results is that a primary deficit in stuttered speech is the inability to initiate vocalization sequences, and that this specific aspect of the behavior is shared between organisms and can be disrupted by similar genetic means.

It is worth noting that mice are deaf during the postnatal period until at least P10 depending on the frequency [48]. Auditory feedback in humans has been alternatively shown to reduce or increase stuttering [49]. However with humans being able to hear is not necessary for stuttering behavior. It has been found that people that are hard of hearing and people who are deaf, stutter, both in oral speech and sign language [45, 50].

We also noted some differences between human stuttering and the mouse model. There have been several cases of PWS who are heterozygous for a mutation in the LETP; in contrast, we found that the vocalizations of heterozygous mice were more similar to those of wild type controls. In humans, repetition of a single syllable is unusual and strongly indicative of stuttered speech; in mice, syllable repetition is commonplace, with the consequence that only longer sequences of repetition exhibited significant differences. Finally, an important outstanding question is whether the *Gnptab* [mut/mut] mice were attempting to vocalize during the abnormal pauses or simply remaining quiet.

Non-vocal behaviors

We examined sensory-motor function and learning in a series of non-vocal behavioral paradigms. We ran an exhaustive behavioral screening on these animals including 50 statistical comparisons. We found that *Gnptab* [mut/mut] mice were normal over the large majority of these measurements. Two measures which did exhibit differences were general ambulatory activity ($p = 0.007$ (corrected for 6 bins, p=.042) during the final time bin, or $p = 0.049$ over the entire time interval) and exploratory behaviors related to hole pokes ($p = 0.01$-0.03). This is not significantly better than chance (one would expect 2.5 tests to pass this criterion by chance). Nor was any single p-value sufficiently strong to survive Bonferroni correction for multiple comparisons. With such an extensive battery of tests we therefore conclude that our finding appears to be a specific abnormality in vocalization. An alternative interpretation of these findings is that *Gnptab* [mut/mut] mice may show reduced rate of certain spontaneous behaviors.

Summary

Our data show that mutations in the lysosomal enzyme targeting pathway affect mouse pup isolation calls and human speech in similar ways, suggesting that there is a commonly disrupted vocalization function in both. We suggest that this animal model could allow the large array of genetic and neurobiological tools that exist in mice to be applied to the study of the cellular and molecular features of stuttering and perhaps normal speech.

Experimental Procedures

Data Acquisition

Studies were conducted under institutionally reviewed and approved animal study protocols. (NIDCD/NIH #1318-13, D.D. Principal Investigator and Washington University protocol 20130179, T.H. Principal Investigator, and 20120110, D.W., Principal Investigator). Knock-in mice were constructed on a BALB/c background using the strategy shown in Figure S1. To generate test subjects, heterozygous mice were mated. Females were separated from all other mice before they gave birth. Dams were checked for pups daily and the first day that a litter was discovered was considered as postnatal day zero (P0). Mice came from a total of 18 litters. We recorded from mice on postnatal three (P3), five (P5), and eight (P8). Mouse data were collected blind to genotype.

Analysis of Vocalizations—All analyses were done using in-house MATLAB software, available online at http://holylab.wustl.edu/. Analysis code implemented a fully-automated algorithm and were therefore blind to genotype. For mice, waveforms were pre-processed, band-pass filtered (25–110 kHz), and vocalizations identified using mean frequency, "spectral purity" (fraction of total power concentrated into a single frequency bin), and the "spectral discontinuity" (the change in the allocation of power across frequencies between two adjacent time bins) [14]. Stored acoustical waveforms were processed using MATLAB to compute the sonogram (512 samples/block, half-overlap, resulting in a time resolution of 1.02 ms and a frequency resolution of 0.98 kHz). We also examined broad band 'clicks' in the mouse vocalizations that were defined as non-contiguous milliseconds where less than 200 of the 512 sampled half frequencies were empty. No difference in the number of clicks was detected between groups. Stored acoustical waveforms were processed in MATLAB to compute the sonogram (512 samples/block, half-overlap, resulting in a time resolution of 1.02 ms and a frequency resolution of 0.98 kHz).

For humans, vocalizations were identified in terms of supra-threshold power in either of two frequency bands (0.6 -1.6 kHz and 5.6-9.6 kHz). Vocalization periods separated by gaps shorter than 20 ms were merged. Comparison with manual scoring demonstrated that breathing and other non-speech noises were excluded (Figure 5A). We examined the length and number of speech intervals and non-speech intervals (pauses between speech) in the speech samples. These intervals did not necessarily have only one word; many times words were run together without an appreciable non-speech interval.

To calculate the number of syllables and the duration of syllables and pauses, each vocalization or pause contributed to the mean for each animal or subject; each individual's

mean was then averaged to obtain the group average. A *t*-test was then performed to compare groups with each individual's mean. An alternative analysis, where each syllable contributes to the overall mean was also performed and produced comparable results, not shown.

For the syllable analysis, the definitions set forth in Arriaga et al. were used, with the exception of classes "i, j, k" these syllable types were grouped into the category 'other' [10, 14]. This classification was done using a fully-automated algorithm. To determine whether the usage frequency differed significantly between genotypes, we first picked a random pair of recordings from the *Gnptab* [mut/mut] mice. We performed a bootstrap analysis of the differences in syllable usage between two individual mice, measuring the difference by

$$D = \sum_{s=1}^{9} \left(\sqrt{\tilde{a}_s} - \sqrt{\tilde{b}_s} \right)^2$$

where s is the syllable type index, and $\tilde{a}_s$ and $\tilde{b}_s$ are the normalized numbers of observed or predicted syllables of type s. The normalization was performed in terms of $A = \Sigma a_s$ and $B = \Sigma b_s$ (where variables without ~ denote raw counts), scaling $\tilde{a}_s = \frac{B}{\max(A, B)} a_s$ and $\tilde{b}_s = \frac{A}{\max(A, B)} b_s$. As a consequence, the recording with more syllables was scaled to a number equal to that of the recording with fewer syllables, and the one with fewer was unmodified. The bootstrap was calculating the mean difference D among randomly-selected pairs of *Gnptab* [mut/mut] mice, and then comparing this mean against the distribution of mean D obtained from an identical number of wildtype/mut pairs, with wildtypes resampled to match the number of syllables produced by mut animals.

We used a *t*-test to compare the number of doublets and found no difference between experimental groups. Bout-level analyses defined bouts based on histograms of pause lengths for all groups of mice in each day we recorded. Histograms were constructed with a range of bin sizes (50 ms to 300 ms). The middle of minimum bin in the range of 0.15 to 0.33 seconds was averaged across all bin sizes to determine the criteria for an inter-bout pause. The resulting minimum intra-bout/inter-bout cutoff was determined to be 0.33 s in day 3 recordings, 0.28 s in day 5 and 0.22 s in day 8 recordings. To analyze bout-level repetition, the syllables in each bout were analyzed to determine how many of the syllables in each bout were the same type as the mode of that bout. In Figure 4D, the confidence intervals were calculated by bootstrapping bouts; bout-sizes with fewer than 3 exemplars are depicted only in terms of their mean. Entropy of syllable usage was calculated from the proportion of different syllable types; entropy for the temporal sequence (modeled as a first-order Markov process) was given by $H2 = -\Sigma p(X) \Sigma p(X|Y) \log 2 p(X|Y)$ with X and Y being each syllable type [19].

For the classification analysis, the number of vocalizations per minute, the average duration of each vocalization and the average duration of each pause were normalized to a range of 0-1 and plotted in 3 dimensional space. We then categorized each recording as being either

from a control or person who stutters based on nearest neighbors. Choices of 1, 3, and 5 closest neighbors all yielded very similar results. We analyzed three cases: 1) all recordings from controls and PWS, 2) controls and PWS with an unknown cause, and 3) controls and PWS with a known mutation in the LETP. Data in all cases were significant compared to empirically calculated p value of .05 (data labels (e.g. controls, person who stutters) shuffled 10,000 times, 95% confidence intervals).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Bloodstein, O.; Bernstein Ratner, N., editors. A Handbook on Stuttering. 6th. 2008.

2. Conture, EG. Stuttering: Its Nature, Dignosis, and Treatment. Allyn and Bacon; 2001.

3. Fisher SE, Lai CSL, Monaco AP. Deciphering the Genetic Basis of Speech and Language Disorders. Annual Review of Neuroscience. 2003; 26:57–80.

4. MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CSL, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, et al. Identification of FOXP2 Truncation as a Novel Cause of Developmental Speech and Language Deficits. The American Journal of Human Genetics. 2005; 76:1074–1080. [PubMed: 15877281]

5. Fedorenko E, Morgan A, Murray E, Cardinaux A, Mei C, Tager-Flusberg H, Fisher SE, Kanwisher N. A highly penetrant form of childhood apraxia of speech due to deletion of 16p11. 2. European Journal of Human Genetics. 2015

6. Kang C, Riazuddin S, Mundorff J, Krasnewich D, Friedman P, Mullikin JC, Drayna D. Mutations in the Lysosomal Enzyme–Targeting Pathway and Persistent Stuttering. New England Journal of Medicine. 2010; 362:677–685. [PubMed: 20147709]

7. Raza MH, Domingues CEF, Webster R, Sainz E, Paris E, Rahn R, Gutierrez J, Chow HM, Mundorff J, Kang Cs, et al. Mucolipidosis types II and III and non-syndromic stuttering are associated with different variants in the same genes. Eur J Hum Genet. 2015

8. Kornfeld S. Trafficking of lysosomal enzymes. The FASEB Journal. 1987; 1:462–468. [PubMed: 3315809]

9. Fischer J, Hammerschmidt K. Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. Genes, Brain and Behavior. 2010:no–no.

10. Arriaga G, Zhou EP, Jarvis ED. Of Mice, Birds, and Men: The Mouse Ultrasonic Song System Has Some Features Similar to Humans and Song-Learning Birds. PLoS ONE. 2012; 7:e46610. [PubMed: 23071596]

11. Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T, Somel M, Brückner MK, Schreiweis C, Winter C, Sohr R, et al. A Humanized Version of Foxp2 Affects Cortico-Basal Ganglia Circuits in Mice. Cell. 2009; 137:961–971. [PubMed: 19490899]

12. Scattoni ML, Ricceri L, Crawley JN. Unusual repertoire of vocalizations in adult BTBR T+ tf/J mice during three types of social encounters. Genes, Brain and Behavior. 2011; 10:44–56.

13. Anne-Marie EENT, Sourd L, Boeckers CSLTM, Thomas PF. The Autism ProSAP1/Shank2 Mouse Model Displays Quantitative and Structural Abnormalities in Ultrasonic Vocalisations. 2013

14. Holy TE, Guo Z. Ultrasonic Songs of Male Mice. PLoS Biol. 2005; 3:e386. [PubMed: 16248680]

15. Fujita E, Tanabe Y, Shiota A, Ueda M, Suwa K, Momoi MY, Momoi T. Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proceedings of the National Academy of Sciences. 2008; 105:3117–3122.

16. Gaub S, Groszer M, Fisher SE, Ehret G. The structure of innate vocalizations in Foxp2-deficient mouse pups. Genes, Brain and Behavior. 2010; 9:390–401.

17. Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, De Gasperi R, Sosa MAG, Rabidou D, et al. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:9643–9648. [PubMed: 15983371]

18. Grimsley JMS, Monaghan JJM, Wenstrup JJ. Development of Social Vocalizations in Mice. PLoS ONE. 2011; 6:e17460. [PubMed: 21408007]

19. Ey E, Torquet N, Le Sourd AM, Leblond CS, Boeckers TM, Faure P, Bourgeron T. The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations. Behavioural Brain Research. 2013; 256:677–689. [PubMed: 23994547]

20. Ehret G. Categorical perception of mouse-pup ultrasounds in the temporal domain. Animal Behaviour. 1992; 43:409–416.

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

22. Groszer M, Keays DA, Deacon RMJ, de Bono JP, Prasad-Mulcare S, Gaub S, Baum MG, French CA, Nicod J, Coventry JA, et al. Impaired Synaptic Plasticity and Motor Learning in Mice with a Point Mutation Implicated in Human Speech Deficits. Current biology : CB. 2008; 18:354–362. [PubMed: 18328704]

23. Hammerschmidt K, Schreiweis C, Minge C, Pääbo S, Fischer J, Enard W. A humanized version of Foxp2 does not affect ultrasonic vocalization in adult mice. Genes, Brain and Behavior. 2015; 14:583–590.

24. Groszer M, Keays DA, Deacon RMJ, de Bono JP, Prasad-Mulcare S, Gaub S, Baum MG, French CA, Nicod J, Coventry JA, et al. Impaired Synaptic Plasticity and Motor Learning in Mice with a Point Mutation Implicated in Human Speech Deficits. Current Biology. 2008; 18:354–362. [PubMed: 18328704]

25. Fedyna A, Drayna D, Kang C. Characterization of a mutation commonly associated with persistent stuttering: evidence for a founder mutation. J Hum Genet. 2011; 56:80–82. [PubMed: 20944643]

26. Grimsley JM, Gadziola MA, Wenstrup JJ. Automated classification of mouse pup isolation syllables: from cluster analysis to an Excel-based "mouse pup syllable classification calculator". Frontiers in Behavioral Neuroscience. 2013; 6:89. [PubMed: 23316149]

27. Wozniak DF, Hartman RE, Boyle MP, Vogt SK, Brooks AR, Tenkova T, Young C, Olney JW, Muglia LJ. Apoptotic neurodegeneration induced by ethanol in neonatal mice is associated with profound learning/memory deficits in juveniles followed by progressive functional recovery in adults. Neurobiology of disease. 2004; 17:403–414. [PubMed: 15571976]

28. Dougherty JD, Maloney SE, Wozniak DF, Rieger MA, Sonnenblick L, Coppola G, Mahieu NG, Zhang J, Cai J, Patti GJ. The disruption of Celf6, a gene identified by translational profiling of serotonergic neurons, results in autism-related behaviors. The Journal of Neuroscience. 2013; 33:2732–2753. [PubMed: 23407934]

29. Wozniak DF, Diggs-Andrews KA, Conyers S, Yuede CM, Dearborn JT, Brown JA, Tokuda K, Izumi Y, Zorumski CF, Gutmann DH. Motivational disturbances and effects of L-dopa administration in neurofibromatosis-1 model mice. PLoS ONE. 2013; 8:e66024. [PubMed: 23762458]

30. Menzies RG, Onslow M, Packman A. Anxiety and stuttering: Exploring a complex relationship. American Journal of Speech-Language Pathology. 1999; 8:3.

31. Czyzewski A, Kaczmarek A, Kostek B. Intelligent Processing of Stuttered Speech. Journal of Intelligent Information Systems. 2003; 21:143–171.

32. Healey EC, Ramig PR. Acoustic Measures of Stutterers' and Nonstutterers' Fluency in Two Speech Contexts. Journal of Speech, Language, and Hearing Research. 1986; 29:325–331.

33. Howell P, Sackin S, Glenn K. Development of a Two-Stage Procedure for the Automatic Recognition of Dysfluencies in the Speech of Children Who Stutter: II. ANN Recognition of Repetitions and Prolongations With Supplied Word Segment Markers. Journal of speech, language, and hearing research : JSLHR. 1997; 40:1085–1096.

34. Lustyk T, Bergl P, Cmejla R. Evaluation of disfluent speech by means of automatic acoustic measurements. The Journal of the Acoustical Society of America. 2014; 135:1457–1468. [PubMed: 24606282]

35. Riper, Cv. The nature of stuttering. Englewood Cliffs, nj: Prentice–Hall; 1982.

36. Robb M, Blomgren M, Chen Y. Formant frequency fluctuation in stuttering and nonstuttering adults. Journal of Fluency Disorders. 1998; 23:73–84.

37. Seery CH, Watkins RV, Mangelsdorf SC, Shigeto A. Subtyping stuttering II: Contributions from language and temperament. Journal of Fluency Disorders. 2007; 32:197–217. [PubMed: 17825669]

38. Webster RL. Empirical considerations regarding stuttering therapy. Controversies about stuttering therapy. 1979:209–240.

39. Andrade, CRFd; Cervone, LM.; Sassi, FC. Relationship between the stuttering severity index and speech rate. Sao Paulo Medical Journal. 2003; 121:81–84. [PubMed: 12870056]

40. Johnson W. Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. The Journal of speech and hearing disorders. 1961:1.

41. Ryan BP. Articulation, Language, Rate, and Fluency Characteristics of Stuttering and Nonstuttering Preschool Children. Journal of Speech, Language, and Hearing Research. 1992; 35:333–342.

42. Yairi E. Subtyping stuttering I: A review. Journal of Fluency Disorders. 2007; 32:165–196. [PubMed: 17825668]

43. Honda M. Human speech production mechanisms. NTT Technical Review. 2003; 1:24–29.

44. Johnson AM, Ciucci MR, Russell JA, Hammer MJ, Connor NP. Ultrasonic output from the excised rat larynx. The Journal of the Acoustical Society of America. 2010; 128:EL75–EL79. [PubMed: 20707418]

45. Montgomery BM, Fitch JL. The prevalence of stuttering in the hearing-impaired school age population. Journal of Speech and Hearing Disorders. 1988; 53:131–135. [PubMed: 3361855]

46. Fox PT, Ingham RJ, Ingham JC, Zamarripa F, Xiong JH, Lancaster JL. Brain correlates of stuttering and syllable production. Brain. 2000; 123:1985–2004. [PubMed: 11004117]

47. Rosenbek J, Messert B, Collins M, Wertz RT. Stuttering following brain damage. Brain and language. 1978; 6:82–96. [PubMed: 698787]

48. Ehret G. Development of absolute auditory thresholds in the house mouse (Mus musculus). Ear and Hearing. 1976; 1:179–184.

49. Kalinowski J, Armson J, Stuart A, Gracco VL. Effects of alterations in auditory feedback and speech rate on stuttering frequency. Language and Speech. 1993; 36:1–16. [PubMed: 8345771]

50. Snyder, G. Unpublished manuscript. University of Mississippi, University; 2006. The existence of stuttering in sign language and other forms of expressive communication: Sufficient cause for the emergence of a new stuttering paradigm.

Highlights

- Mutations in the lysosomal enzyme-targeting pathway cause stuttering in humans

- Mice with a mutation in this pathway also show abnormalities in vocalizations

- Other behaviors in the knock-in mice appear largely intact

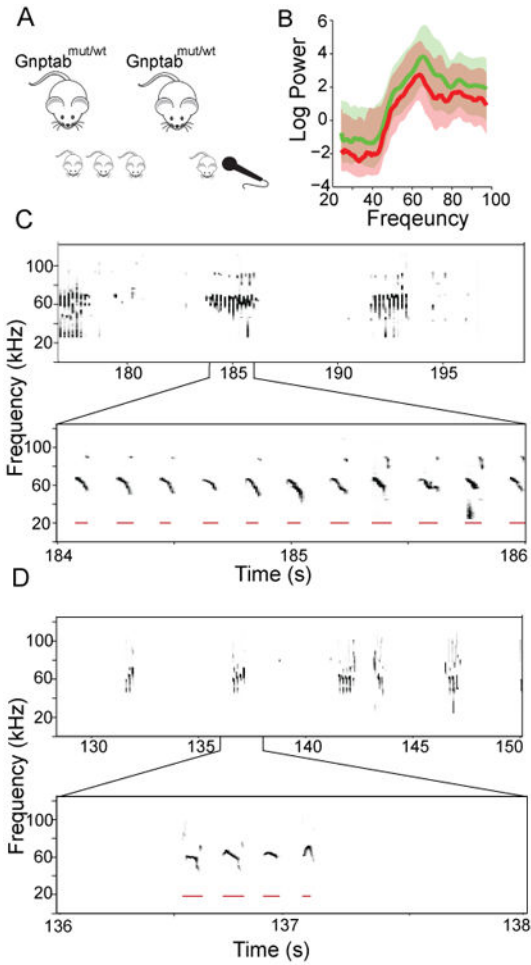- People who stutter with this type of mutation have similar features in their speech

Figure 1. Experimental scheme

(A) Knock-in mice with a Glu1179Lys mutation in exon 19 of the *Gnptab* gene were constructed on a BALB/c background. Recordings were performed on mice pups separated from dam for recording on postnatal day 3, 5 and 8. **(B)** Frequency distribution of mouse pup vocalizations on day 8 in wild type littermates from the *Gnptab* $^{wt/wt}$ mice (green, n=20) and *Gnptab* $^{mut/mut}$ mice (red, n=19) **(C)** Sonogram of a wild type mouse pup vocalizing in isolation. Lower window is an expanded version of one portion, showing the repeated structure in the mouse vocalizations. Red bars indicate a detected vocalization. **(D)** Sonogram of a *Gnptab* $^{mut/mut}$ as in c. See also Figure S1.
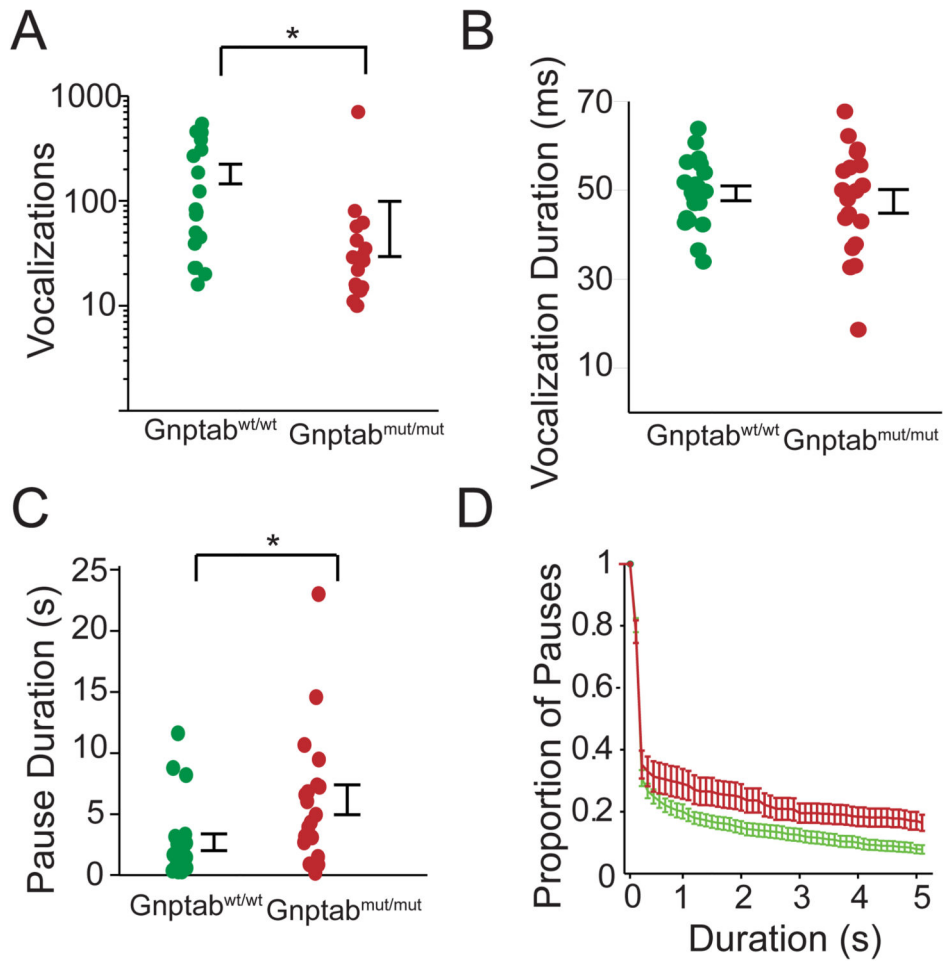
Figure 2. Differences in the vocalizations of mice with a mutation in the LETP compared to wild type littermates

(**A**) Vocalizations per recording in mouse pup vocalizations in wild type littermates (*Gnptab* wt/wt, green, n=20), and mice homozygous for the Glu1179Lys mutation (*Gnptab* mut/mut, red, n=19). Note the significant decrease in the *Gnptab* mut/mut mice compared *Gnptab* wt/wt littermates (*t*-test, p < 0.034). (**B**) Duration of vocalization in *Gnptab* wt/wt and *Gnptab* mut/mut mice. No significant difference was found. (**C**) Pause duration between vocalizations, with a significant increase in the duration of pauses in the *Gnptab* wt/wt compared to the *Gnptab* mut/mut mice (*t*-test p<0.020). (**D**) Proportion of long pauses with a steadily increasing criterion for criterion for 'long pauses'. *Gnptab* mut/mut mice had significantly more long pauses. Each dot represents the mean of one subject. Error bars represent standard error of the mean across individuals. Color scheme as in panels A-C. See also Figure S2 and S3.
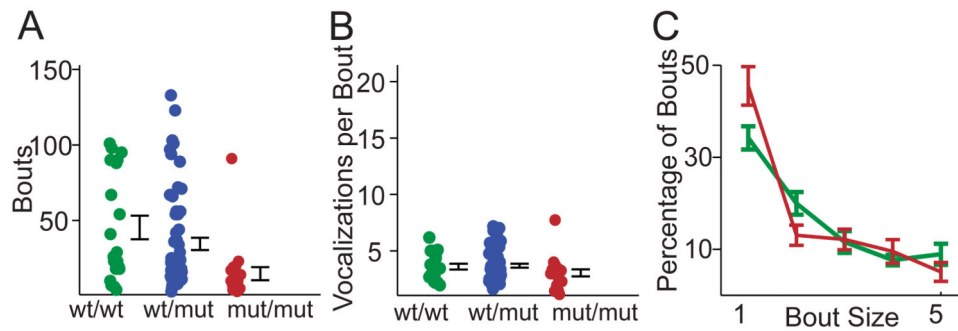
Figure 3. Number of bouts in *Gnptab* ^{wt/mut} mice vocalizations compared to *Gnptab* ^{mut/mut} and *Gnptab* ^{wt/wt} mice across days

(**A**) Number of bouts per recording in mouse pup vocalizations in wild type littermates (*Gnptab* ^{wt/wt}, green), (*Gnptab* ^{wt/mut}, blue) and mice homozygous for the Glu1179Lys mutation (*Gnptab* ^{mut/mut}, red). Note there was a significant difference between *Gnptab* ^{wt/wt} and *Gnptab* ^{mut/mut} mice (*t*-test $p < .0023$). Error bars represent standard error of the mean across individuals. (**B**) No significant difference was found in the number of vocalizations per bout. (**C**) Percentage of bouts containing 1 (*Gnptab* ^{mut/mut}, red, 45.5% ± .04%, *Gnptab* ^{wt/wt}, green, 34.2% ± .03%, $p < .029$), 2 (*Gnptab* ^{mut/mut} 13.0 % ± .02% *Gnptab* ^{wt/wt} 20.0% ± .03, $p<.05$), 3, 4 or 5 vocalizations.
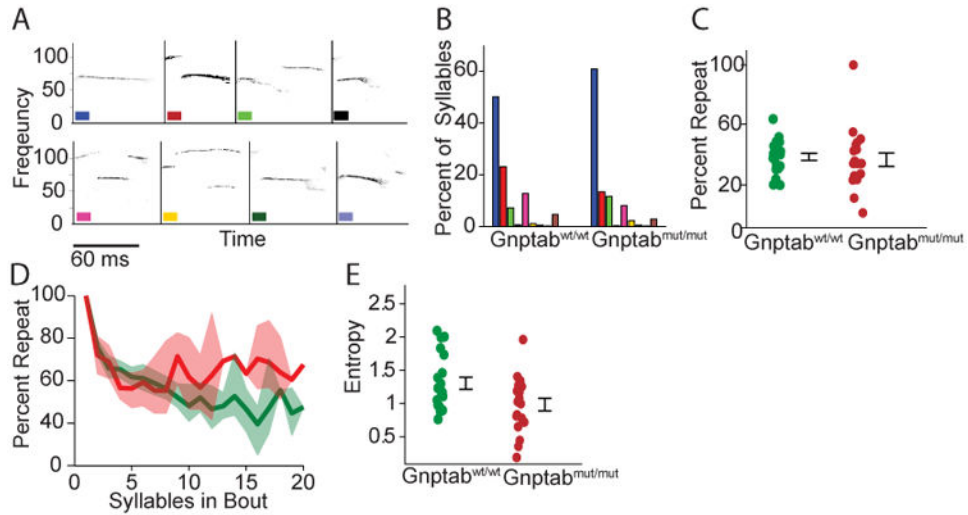
Figure 4. Mouse vocalizations from *Gnptab* wt/wt compared to vocalizations of *Gnptab* mut/mut
(A) Syllable identification scheme showing examples of each type of syllable. **(B)** Percentage of each type of syllable in wild type and knock-in mice. Differences were not statistically significant. Each color represents one syllable type. **(C)** Percentage of times that one syllable type was followed by the same syllable type. Differences were not statistically significant. **(D)** Percentage of syllables that were of the same type as the mode of each bout, for bout sizes ranging from 1 to 20. Data from *Gnptab* wt/wt and *Gnptab* mut/mut . Shaded areas represent 95% confidence intervals. **(E)** Diversity of vocalization sequences as quantified by the entropy of the corresponding first order Markov process (*t*-test, p<0.022). See also Figure S4 and S5.
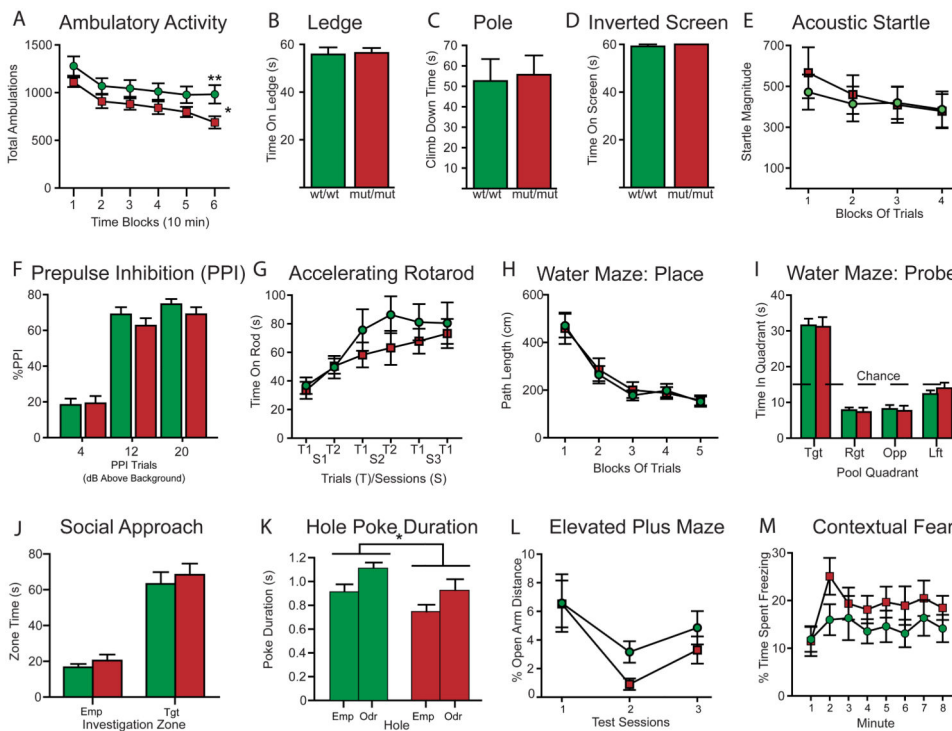
Figure 5. Non-vocal behavioral test results from *Gnptab* wt/wt and *Gnptab* mut/mut mice
(**A**) Ambulatory activity was significantly decreased (*p<0.05) in mice homozygous for the Glu1179Lys mutation (*Gnptab* mut/mut, red, n=15) compared to wild type (*Gnptab* wt/wt, green, n=15) mice with differences being greatest during block 6 (**p=0.017). (**B-D**) The two groups of mice did not differ on several sensorimotor tests such as the time spent on an elevated ledge (**B**), the time to climb down a pole (**C**), or the time spent hanging upside down on an inverted screen before falling (**D**). (**E-F**) The *Gnptab* wt/wt and *Gnptab* mut/mut mice also exhibited similar magnitudes of the acoustic startle response (arbitrary units) (**E**) and PPI (**F**). (**G**) Time spent on accelerating rotarod before falling. No significant differences were found. (**H**) Path length on Morris Water maze for place trials. (**I-J**) The performance of the *Gnptab* mut/mut mice was not impaired during the spatial learning (place) trials (**I**) nor in terms of retention performance during a probe trial where each group showed spatial bias for the target quadrant by spending significantly more time in it versus the times spent in each of the other quadrants (p<0.00005). (**J**) The *Gnptab* wt/wt and *Gnptab* mut/mut mice showed comparable levels of sociability in that each group spent significantly more time in an investigation zone surrounding a stimulus mouse that remained inside a withholding cage compared to levels exhibited toward an empty cage (p<0.00005), although there were no differences between the two groups for either of the two conditions. (**K**) For each group, the duration of hole pokes was significantly greater for the odorant-containing versus the empty holes (p<0.024), although the *Gnptab* wt/wt mice had significantly higher poke durations on average across the two hole types (*p=0.028). (**L**) The two groups did not differ in levels of anxiety-related behaviors in the elevated plus maze as exemplified by the lack of significant effects involving genotype on the percentage of distance traveled in the open arms out of the total distances traveled in all of the arms. (**M**) The absence of any

significant effects involving genotype suggested that there were no differences in contextual fear conditioning between the groups.
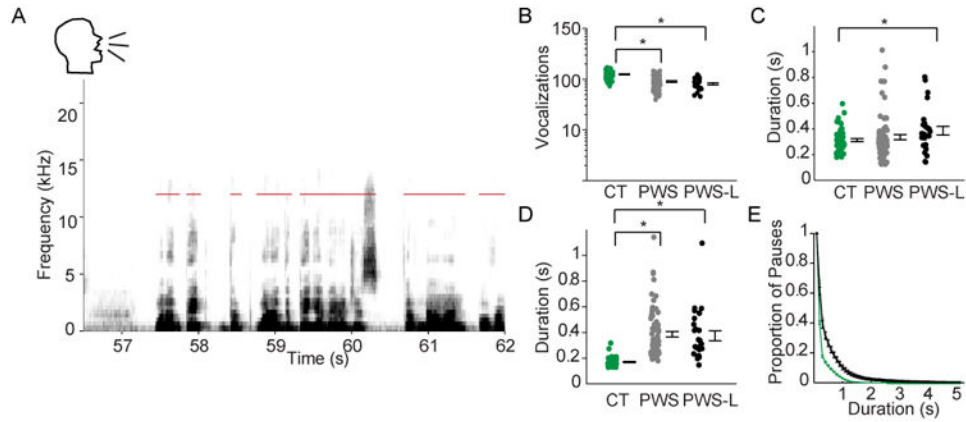
Figure 6. Vocalizations of human controls, PWS, and PWS with a known mutation in the LETP, as measured by a language agnostic analysis

(A) Sonogram of a person saying, "(Breath) One of big…the…one of the big advantages of nylon over" while reading from a standard diagnostic script. Red bars indicate detected vocalization intervals. **(B)** Vocalizations per minute in the control subjects reading the script (CT; green, n=51), PWS (PWS; gray, n=74), and PWS with a known mutation in the LETP reading the same script (PWS-L; black, n=26). Each dot represents the mean of one subject. Error bars represent standard error of the mean across individuals. Brackets and stars indicate significance. There was a significant decrease in the number of vocalizations in PWS group (*t*-test, $p < 10^{-11}$), and in the PWS-L (*t*-test, $p < 10^{-11}$) group compared to the CT group. Data on log scale. **(C)** Duration of vocalizations. Note there was no significant difference in the duration of vocalizations between the CT and PWS groups while there was a significant difference between the CT and PWS-L groups (*t*-test, $p < 0.02$). **(D)** Mean pause duration between vocalizations, with a significant increase in the PWS group (*t*-test, $p < 10^{-12}$) as well as in the PWS-L group (*t*-test, $p < 10^{-9}$) compared to the CT group. **(E)** Proportion of long pauses, with a steadily increasing criterion for 'long pauses', in the recordings of the subjects reading the script using the same color scheme as in panels B-D showing that over a range of criteria, PWS in both categories had more 'long' pauses. Bars indicate standard error of mean. See also Table S1.
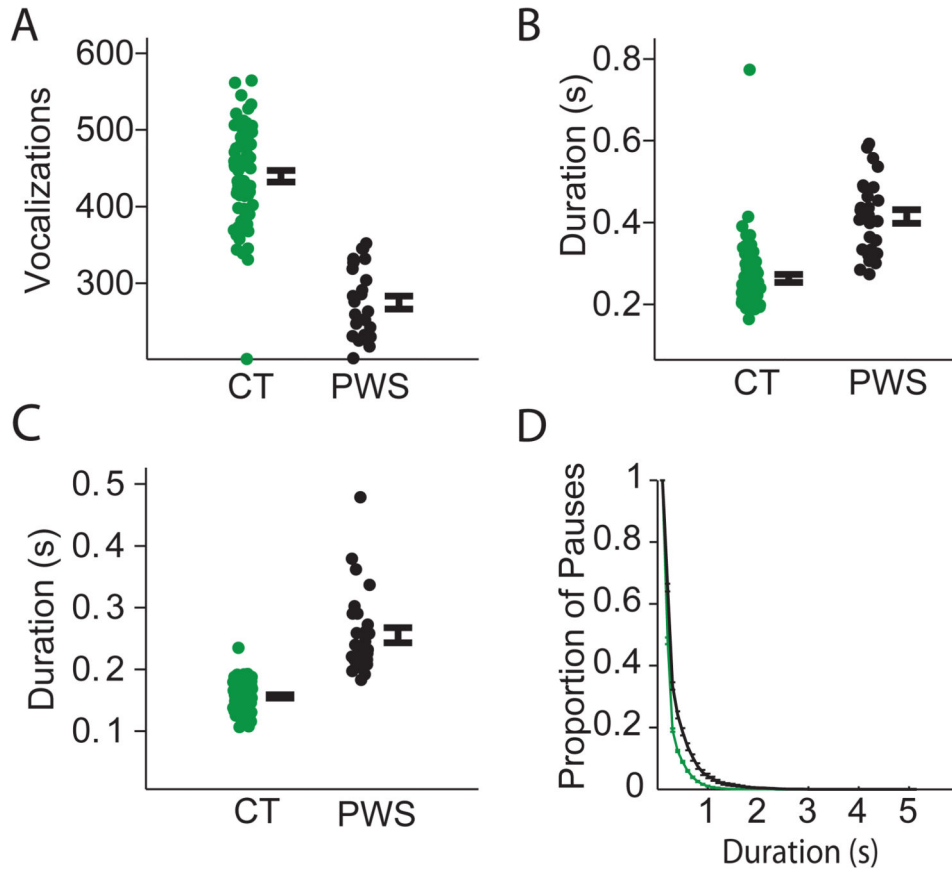
Figure 7. Extemporaneous vocalizations of human controls and PWS
(A) Vocalizations per minute in podcast segments from controls (CT; green, n=68), and PWS (PWS; black, n=28, t-test; $p<10^{-21}$). Each dot represents the mean of one subject. Error bars represent standard error of the mean across individuals. **(B)** Duration of vocalization periods. Note the significant increase in PWS group (t-test, $p< p<10^{-12}$) compared to the CT group. **(C)** Pause duration between vocalizations, with a significant increase in the duration of pauses in the PWS group (t-test, $p<10^{-17}$) compared to the CT group. **(D)** Proportion of long pauses with a steadily increasing criterion for 'long pauses', in the recordings of the controls subjects reading the script (CT; green), and PWS (ST; black). Bars indicate standard error of the mean. These results are similar to that of reading a script Figure 6.