

Novel patient-specific BBSOAS mouse models reveal genotype-phenotype correlations in brain structure and behavior

Johann G. Maass^{1,2}, Dominik Kamionek¹, Annabelle Mantilleri⁶, Susanne Theiss¹, Laura Dötsch¹, Felix Franke¹, Tim Schubert¹, Jonas G. Scheck^{3,4}, Claudia Pitzer⁵, Paolo Piovani⁶, Michele Bertacchi⁶, Olivier Deschaux⁶, Anubhav Singh¹, Chun-An Chen⁷, Henning Fröhlich¹, Michèle Studer^{6,*},[‡] and Christian P. Schaaf^{1,*},[‡]

¹Institute of Human Genetics, Heidelberg University Clinic, Heidelberg, Germany

²Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA

³Neuroradiology Department, University Hospital Heidelberg, Heidelberg, Germany

⁴Clinical Cooperation Unit Translational Radiation Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁵Interdisciplinary Neurobehavioral Core, Heidelberg University, Heidelberg, Germany

⁶Institute de Biologie Valrose (iBV), Université Côte d'Azur (UCA), CNRS 7277, Inserm 1091, Avenue Valrose 28, Nice 06108, France

⁷Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

*co-last authors

[‡]Corresponding emails:

Christian.Schaaf@med.uni-heidelberg.de

Michele.STUDER@univ-cotedazur.fr

Keywords: NR2F1, ASD, Autism, behavior, neurodevelopmental disorder, Bosch-Boonstra-Schaaf Optic Atrophy Syndrome, social behavior

Abstract

Bosch-Boonstra-Schaaf optic atrophy syndrome (BBSOAS) is a rare, autosomal dominant neurodevelopmental disorder caused by pathogenic variants in *NR2F1*, characterized by developmental delay, intellectual disability, optic nerve anomalies, and autism spectrum disorder. Most pathogenic variants cluster within the highly conserved DNA-binding domain (DBD) or ligand-binding domain (LBD) of NR2F1 and

are associated with variable clinical severity, suggesting a genotype-phenotype correlation. While previous mouse models have provided important insights, comprehensive behavioral characterization remains limited. Here, we present two novel BBSOAS mouse models harboring patient-specific variants in the DBD (*Nr2f1*^{+/*R139L*}) and LBD (*Nr2f1*^{+/*E397**}), alongside the established *Nr2f1*^{+/-} model. We analyzed both brain morphology and behavior to further expand the murine phenotype and investigate the genotype-phenotype correlation. We demonstrate that these models recapitulate key aspects of the BBSOAS phenotype, including deficits in cognition, social communication, and motor function, and that the presence and severity of behavioral abnormalities are dependent on variant type. Our findings provide new evidence for a genotype-phenotype correlation associated with domain-specific *NR2F1* variants and establish a robust platform for future mechanistic and therapeutic studies.

Summary statement

Novel mouse models of Bosch-Boonstra-Schaaf optic atrophy syndrome with patient-specific *Nr2f1* variants show strong face validity and offer insights into underlying genotype-phenotype correlation.

Introduction

Bosch-Boonstra-Schaaf optic atrophy syndrome (BBSOAS; OMIM 615722; ORPHA 401777) is a rare, autosomal dominant neurodevelopmental disorder caused by heterozygous pathogenic variants in *NR2F1* (Bosch et al. 2014). Clinically, BBSOAS is characterized by developmental delay, intellectual disability (ID), language impairments, optic nerve anomalies, and autism spectrum disorder (Bertacchi et al. 2022). *NR2F1* encodes an orphan transcription factor (Nuclear Receptor subfamily 2 group F member 1) that functions as a dimer and plays a critical role in neurogenesis. It is involved in cortical organization and regulates the balance between proliferation and differentiation (Tocco et al. 2021; Bertacchi et al. 2022; Cooney et al. 1992).

Most pathogenic variants in BBSOAS arise *de novo*. While no recurrent hotspot mutations have been identified, known pathogenic variants predominantly cluster within one of the two most highly conserved functional domains of the protein: the DNA-binding domain (DBD) and the ligand-binding domain (LBD) (Bertacchi et al.

2022). The syndrome is characterized by genetic heterogeneity and variable expressivity. For instance, cognitive impairment ranges from mild to severe ID, with some BBSOAS patients being able to attend regular school with support, while others are nonverbal and wheelchair-dependent (Bertacchi et al. 2022). A potential genotype-phenotype correlation was first proposed by Chen et al. (2016) (Chen et al. 2016) and later investigated in a study of 92 BBSOAS patients, which revealed that variants in the DBD are associated with a more severe clinical phenotype (Bertacchi et al. 2022). Developing and analyzing models that recapitulate this genotype-phenotype relationship is essential for advancing our understanding of BBSOAS.

Previous studies have evaluated individual mouse models carrying different *Nr2f1* variants, including a conditional knockout in neocortical and hippocampal neurons, a constitutive heterozygous knockout, and a heterozygous point mutation in the DNA-binding domain (DBD) (Tomassy et al. 2010; Contesse et al. 2019; Zhang et al. 2020; Chen et al. 2020; Armentano et al. 2007). These models have provided insights into the impact of *Nr2f1* deficiency on fundamental neurodevelopmental processes; however, several behavioral aspects of the BBSOAS phenotype in mice remain insufficiently characterized. Moreover, while only 16% of patients carry a deletion in *NR2F1*, most studies have relied on the *Nr2f1* knockout model, which exhibits mild behavioral abnormalities (2).

In this study, we generated two novel BBSOAS mouse models carrying patient-specific variants in the DNA-binding domain (*Nr2f1*^{+/*R139L*}) and the ligand-binding domain (*Nr2f1*^{+/*E397**}). Both missense variants in the DBD and nonsense variants are associated with a dominant negative effect (Fig. 1A) (Bertacchi et al. 2022). Using these two new models, alongside the established *Nr2f1*^{+/-} mouse model, we analyzed brain morphology throughout postnatal development and conducted a comprehensive battery of behavioral tests targeting key domains affected in individuals with BBSOAS, including cognition, social communication, and motor function (Fig. 1B). Notably, we identified anomalies in the ultrasonic vocalizations across all three mouse models, along with motor and cognitive impairments, which were most pronounced in mice carrying the DBD variant. Our results show that these models recapitulate the BBSOAS phenotype and further support the proposed genotype-phenotype correlation (2).

These mouse models represent a significant addition to BBSOAS research, providing critical insights into genotype-phenotype relationships of the disorder and contributing to the foundation for future therapeutic developments. Moreover, they offer the scientific community novel murine models for studying autistic-like phenotypes, anxiety, and other neurodevelopmental abnormalities.

Results

BBSOAS mouse models exhibit reduced body weight at several stages of development

To investigate the effects of patient-specific *Nr2f1* variants on brain development and behavior, we compared three BBSOAS mouse models (Fig. 1A). The first model harbors a deletion of the third exon and the polyA tail, resulting in a functional null allele (*Nr2f1*^{+/-}) (Armentano et al. 2006). The second model carries a missense variant in the DNA-binding domain (DBD) (*Nr2f1*^{+/*R139L*}) corresponding to a mutation identified in a severely affected patient with BBSOAS (Chen et al. 2016). The third model features a nonsense variant in the ligand-binding domain (LBD) (*Nr2f1*^{+/*E397**}), which has also been reported in a patient (Jurkute et al. 2021). Transcript analysis of *Nr2f1* with variant-specific primers confirmed either reduction of the WT allele (Fig. 1B-D) or increase of the mutant allele (Fig. 1E,F). All three mouse models were subjected to an identical behavioral testing timeline, assessing various aspects of murine behavior (Fig. 1G). In addition, brain morphology was analyzed at different developmental stages (Fig. 1G).

The body weight of *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} mice and their corresponding WT littermates was recorded at six developmental time points (P4, P8, P12, P20, P30, and P100). While the weight of *Nr2f1*^{+/-} and *Nr2f1*^{+/*E397**} animals did not differ from that of WT animals (Fig. 1H, I), *Nr2f1*^{+/*R139L*} mice showed significantly lower body weight at all measured time points (except P30) (Figs. 1H and S1B). At P100, both male and female *Nr2f1*^{+/*E397**} mice and *Nr2f1*^{+/-} males showed a sex-specific decrease in body weight (Fig. S1B,D). Together, these findings suggest that different *Nr2f1* variants differentially affect growth, with the DBD missense variant exerting a more substantial and persistent impact on body weight.

BBSOAS mouse models exhibit genotype-dependent alterations in brain morphology

To assess brain morphology, we performed histological analyses and magnetic resonance imaging (MRI) at multiple developmental stages (P7, P28, and adulthood) (Fig. 2). Although a total of 18 brain structures were examined, this section highlights the most prominent and consistent alterations observed across time points and *Nr2f1* mouse models. We observed that mice carrying pathogenic *Nr2f1* variants exhibited consistent changes in lateral ventricle and hippocampal size, replicated across multiple time points. At P7 *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice showed a tendency of enlarged lateral ventricles, while starting at P28, both lines exhibited a significant increase, with the effect being more pronounced in *Nr2f1*^{+/*R139L*} mice (804% increase) (Fig. 2A,B). The same trend could also be seen in adult *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice, with an increase in the lateral ventricle size of 163% and 155%, respectively. Importantly, this change occurred without an overall change in total brain volume (Table S2A) and in the absence of a clear obstruction in the cerebroventricular system.

This enlargement of the lateral ventricles is accompanied by a reduction in structures around these ventricles, such as the hippocampus or the striatum, see Fig. 2 and S2. In early adolescence (P28), the hippocampal size was reduced by 23% in the *Nr2f1*^{+/*R139L*} line and 26% in the *Nr2f1*^{+/*E397**}. This anomaly persisted into adulthood in *Nr2f1*^{+/*R139L*} mice, with a 22% reduction. Additionally, all three lines (*Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**}) exhibited corpus callosum thinning at P28 (17%, 22%, and 40% reductions, respectively (Fig. S2). The findings regarding other brain structures (e.g., striatum and globus pallidus) are summarized in Table S2.

Overall, the three *Nr2f1* variants lead to region-specific neuroanatomical abnormalities, including ventricular enlargement, hippocampal reduction, and corpus callosum thinning, highlighting the critical role of *Nr2f1* in embryonic and postnatal brain development.

BBSOAS mouse models show autism-like behavior and genotype-dependent impairments in social interaction

Clinical studies indicate that BBSOAS patients exhibit anomalies in social interactions, such as speech difficulties (80%), and autism spectrum disorder (39%) (Bertacchi et al. 2022). To investigate corresponding behavioral traits in mice, we

analyzed neonatal ultrasonic vocalizations (USV) and performed the social interaction test (SIT).

Analyzing the USVs emitted by the pups provides insights into early social communication deficits and autism-like behavior (Scattoni et al. 2008; Takumi et al. 2020; Ey et al. 2013). All three genetic mouse models showed a reduction in the number of USVs at P4 (Fig. 3A). Compared to their wildtype littermates, the number of calls was reduced by 30%, 80%, and 76% in *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} mice, respectively. However, by P8 and P12, no significant differences in call frequency were observed (Tab. S6).

In addition to quantitative comparisons, we analyzed 11 call characteristics identified as biologically relevant, see Table S3 (Coffey et al. 2019). To illustrate these findings, a subset of the results is highlighted below, focusing on selected call features altered in more than one mouse line. At P4, all three mouse lines exhibited reduced average call duration, with *Nr2f1*^{+/-} mice also showing reduced duration at P8. Call power and tonality were significantly diminished at P4 in both *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} models. By P8, call characteristics in these two models were comparable to controls; however, at P12, both *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice displayed a reduced slope of call frequency.

Call characteristics, such as duration, frequency changes, and complexity, can be integrated into a unified classification system. In this study, we applied a modified version of the categories originally described by Scattoni et al. (Scattoni et al. 2008). Using the deep learning algorithm and AI developed by Herdt et al., we classified calls into five distinct categories. Analysis of call types using this AI model, specifically built and trained for this task, revealed a significantly altered distribution of vocalization classes at P4, P8, and P12, with a pronounced shift toward simpler categories (classes 1, 2, and 5), as illustrated in Fig. 3B (Herdt et al. 2024). The proportion of complex calls (those featuring a frequency step or two simultaneous frequencies) was reduced at P4 across all mouse models. At P8, only *Nr2f1*^{+/-} mice displayed a significant reduction, whereas at P12, both *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice exhibited fewer complex calls, as assessed by the deep learning model developed by Herdt et al. (Fig. 3B) (Herdt et al. 2024).

In addition, we used the SIT approach to evaluate the social interest of early adolescent mice. In the third trial of the SIT, mice can choose to spend time in a compartment with a novel intruder animal or a littermate (for illustration of the setup, refer to Fig. S3A). Mice typically show strong social interest in novel conspecifics however, *Nr2f1*^{+/*R139L*} mice spent 30% less time in the intruder compartment during trial 3 of the 5 min test (Fig. 3C), and 41% more time in the littermate compartment (Fig. 3D), thus displaying reduced social interest in novel animals (Harro 2023). These mice also spent less time in the close-contact zone with the intruder (Fig. S3B). Consistently, *Nr2f1*^{+/*R139L*} mice showed a reduced Social Novelty Preference Index (Fig. 3E) (Baronio et al. 2015). All four effects were observed exclusively in the *Nr2f1*^{+/*R139L*} mouse line, suggesting a deficit in novelty preference and reduced social motivation. Notably, an increase in time spent in the close-contact zone with the littermate was observed only for the *Nr2f1*^{+/*E397**} mouse line compared to their WT littermates (Fig. 3F). This could imply heightened attachment to familiar conspecifics, potentially due to social anxiety or a reduced motivation to engage with novel individuals. Finally, *Nr2f1*^{+/-} mice did not display significant alterations during the SIT test.

Our data show that *Nr2f1*-deficient mouse models exhibit altered social behavior at different developmental stages, and that the nature of these alterations depends on the type of *Nr2f1* variant.

BBSOAS mouse models show genotype-dependent alterations in anxiety and exploratory behavior

Complementing our assessments of social communication, we further evaluated anxiety-related and exploratory behavior in BBSOAS mouse models using the elevated plus maze (EPM), open field (OF), and dark light box (DLB) tests. The hole board (HB) test was used to assess repetitive behaviors. Neither the DLB nor the HB assay revealed significant differences between mutant animals and their WT littermates for any of the parameters analyzed (Table S6). In the EPM, *Nr2f1*^{+/*R139L*} mice traveled a greater distance (30%) during the 10-minute test period and spent more time in the open arms (40%), as shown in Figs. 4A and S3C. In addition, both *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice entered the open arms more frequently, by 40%

and 27%, respectively (Fig. 4C), suggesting enhanced exploratory behavior or reduced anxiety-like traits in these models.

Nr2f1^{+/*R139L*} mice exhibited a similar trend in the OF test, where they traveled a greater distance (24%) and spent more time (218%) in the center of the arena (Fig. 4B,D during the 10-minute test period. Representative track maps of WT and *Nr2f1*^{+/*R139L*} animals are shown in Fig. 4E,F. In contrast, *Nr2f1*^{+/-} mice did not show significant changes in anxiety or exploratory behavior.

In the 24-hour home-cage monitoring test (LABORAS), which analyzes a wide range of behaviors, including anxiety and exploration, variant-carrying and WT mice exhibited largely similar behavior, with only minor differences (Table S6). Specifically, *Nr2f1*^{+/*E397*} mice showed a reduction in climbing distance and an increase in overall distance traveled. In contrast, *Nr2f1*^{+/*R139L*} mice displayed an increased climbing frequency, while all other parameters remained unchanged.

Taken together, the increased time spent in the open arms of the EPM and the center of the OF, along with the increased total distance traveled and the higher number of entries into the open arms, can be interpreted as indicators of hyperactivity or reduced anxiety, respectively.

BBSOAS mouse models show genotype-dependent motor function and coordination deficits

Given previous findings in *Nr2f1* conditional KO mouse models and the motor impairments observed in BBSOAS patients (Bertacchi et al. 2022; Tomassy et al. 2010), we further examined motor function, limb coordination, and gait stability across all genotypes using the CatWalk XT (CW) system. We observed an increase in front paw width in *Nr2f1* mutants: +0.075 cm (not significant) in *Nr2f1*^{+/-}, +0.094 cm ($P \leq 0.05$) in *Nr2f1*^{+/*R139L*}, and +0.099 cm ($P \leq 0.001$) in *Nr2f1*^{+/*E397*} mice (Fig. 5A). However, the width between the hind paws remained unchanged (Fig. S4B).

Beyond a wider front paw stance, stability can also be enhanced by reducing lateral paw support, which reflects the proportion of time the ipsilateral forelimb and hindlimb

remain in contact with the walkway (e.g., left-front and left-hind-paw). This parameter was significantly reduced in *Nr2f1*^{+/-} and *Nr2f1*^{+/*E397**} mice (Fig. S4D). In contrast, diagonal (e.g., left front and right hind or right front and left hind) and girdle (e.g., left front and right front or left hind and right hind) support were not altered (Fig. S4E,F).

To further assess locomotor stability, we analyzed temporal and spatial coordination and interlimb synchronization, using coupling parameters, which measure the timing difference between initial paw contacts (e.g., right front paw and right hind paw) as a percentage of the step cycle duration of the reference paw (e.g., right front paw). Coupling was assessed for three specific paw pairings: diagonal, ipsilateral, and girdle across all 12 possible combinations. In mice, paw movements follow a consistent and precisely regulated timing pattern. For example, ipsilateral paw pairs (e.g., left front and left hind or right front and right hind) and girdle paw pairs typically exhibit coupling values around 50 CSTAT (circular phase lag variable), while diagonal paw pairs approach 0 or 100 CSTAT, respectively (Kloos et al. 2005). As shown in Figs. 5C and S5A,B, overall coupling parameters across combined paw pair types did not significantly differ between experimental lines and wild-type (WT) controls. However, deviations in individual paw pair couplings indicated specific coordination deficits. Across the CatWalk XT dataset, we observed a cumulative effect of these alterations, with 0, 1, and 7 parameters in the *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} line, respectively (Table S6). An example of impaired coupling in *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice for the right front (RF) and left front (LF) paw is shown in Fig. 5B.

Quadrupeds adapt various step sequences (AA, AB, CA, CB, and more, see Maricelli et al.) depending on context and motor ability (Mendes et al. 2015; Maricelli et al. 2016). While the total number of step sequences used remained unchanged (Fig. S4C), *Nr2f1*^{+/*R139L*} mice showed a 12.1% increase in AB usage and a 7.1% decrease in AA compared to WT littermates (Fig. 5D). *Nr2f1*^{+/*E397**} mice exhibited a similar trend with AB increasing by 15.1% and AA decreasing by 9.9% (Fig. 5D). No significant step sequence changes were observed in *Nr2f1*^{+/-} mice.

Analysis of the regularity and precision of step sequences, as quantified by the regularity index, which integrates multiple parameters to assess interlimb

coordination consistency, revealed no significant differences between groups (Fig. S4G). Similarly, the P_{LDA} score, computed by the machine-learning model (Timotius et al.), showed no significant variation in overall motor performance (Fig. S4H) (Timotius et al. 2021).

Consistent with our observations during animal handling, the data indicate that all three mouse models exhibit mild but evident motor dysfunction. CatWalk XT analysis showed that while overall performance, reflected in P_{LDA} scores or walking speed, is not severely compromised, specific parameters corresponding to coordination and balance are impaired in a genotype-dependent manner.

Mice carrying *Nr2f1* variants exhibit impaired spatial learning

To test the spatial learning and memory properties of our *Nr2f1* mutant mice, we used the APA test, with the primary readout being the total number of shocks received over eight training trials. Both *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice received significantly more shocks (484% and 188% increase in number, respectively) and spent more time (652% and 197% increase, respectively) in the shock zone compared to their WT littermates during training (Fig. 6A and Table S6). To evaluate memory retention, animals were reintroduced to the arena 24 hours later for a memory trial in which the shock zone was turned off. We assessed their ability to recall the location of the previously active shock zone. One parameter, the latency to the first entry into the previously shock-associated active zone, was on average 13% shorter for *Nr2f1*^{+/-} mice, 26% shorter for *Nr2f1*^{+/*R139L*}, and 15% longer for *Nr2f1*^{+/*E397**} mice compared to WT littermate controls; however, these differences were not statistically significant (Table S6). Notably, *Nr2f1*^{+/*R139L*} mice spent significantly more time in the previously active zone during the memory trial, suggesting a possible deficit in spatial memory recall (Fig. 6B).

To illustrate the learning process, Fig. 6C-E depicts the average latency (in seconds) for animals of each line to first enter the shock zone. During the eight learning trials, animals gradually associate the shock zone with pain and learn to avoid it. As a result, latency is expected to increase over the initial trials. The data reveal that *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice exhibit impairments in this learning process. Although these mice eventually learn how to identify and avoid the shock zone

effectively, they require more trials to do so. *Nr2f1*^{+/-} mice did not show alterations in the APA test.

These results correlate different patient-specific *Nr2f1* mutations with deficits in learning and spatial memory recall, with DBD- and LBD-mutated mouse models displaying the most significant anomalies.

***Nr2f1*^{+/*R139L*} mice exhibit the most severe phenotype among BBSOAS mouse models analyzed**

Among the three mouse models analyzed in this study, *Nr2f1*^{+/*R139L*} mice displayed the most severe phenotype. Due to differences in experimental cohorts, timing, and potential genetic drift, direct cross-line comparisons were not feasible. Therefore, we performed line-specific analyses, systematically comparing each mutant line to its respective WT littermates. To estimate the overall phenotypic impact of each variant, we quantified the total number of significantly altered parameters. *Nr2f1*^{+/*R139L*} mice exhibited the highest number of significantly altered parameters (30), followed by *Nr2f1*^{+/*E397**} mice (19) and *Nr2f1*^{+/-} mice (6) (Fig. 7A and Table S4).

For parameters that were significantly altered in more than one line, we compared the effect sizes using Cohen's *d*. According to Cohen's classifications, effect sizes above 0.2 are considered a mild effect, values above 0.5 are moderate, and values exceeding 0.8 represent a strong effect. The average *d* values were 0.6810, 1.1035, and 1.0054 for *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} mice, respectively (Fig. 7B and Table S5), further supporting the conclusion that the *R139L* variant exerts the strongest phenotypic effect

Discussion

To date, our pathophysiological understanding of the diverse *NR2F1* variants underlying BBSOAS remains incomplete, limiting the development of targeted therapies. Existing mouse models, such as the heterozygous *Nr2f1* KO, fail to fully capture the range of genetic alterations reported in human patients and do not sufficiently recapitulate the broad spectrum of behavioral alterations characteristic of the syndrome (Chen et al. 2020; Flore et al. 2017; Contesse et al. 2019; Parisot et al. 2017; Tomassy et al. 2010). This highlights a critical need for novel mouse models

with greater construct validity, that more accurately phenocopy both the clinical manifestations and the proposed genotype-phenotype correlations.

In this study, we established two new mouse models carrying patient-specific *NR2F1* variants, selected to represent the more severe end of the phenotypic spectrum observed in BBSOAS. All mouse models assessed in behavioral tasks were bred, maintained, and tested under standardized conditions in the same facility. To ensure consistency and comparability, behavioral testing was performed using identical protocols by the same investigator. Our findings demonstrate that all three models exhibit robust, multifaceted phenotypes, with clear genotype-dependent differences. Consistent with previous studies, heterozygous null mice displayed only mild behavioral changes and minor brain abnormalities. In contrast, and in line with BBSOAS clinical data, the patient-specific point mutation models (*Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice) showed more pronounced and complex phenotypes, supporting the hypothesis of a dominant-negative effect associated with missense or nonsense variants in patients.

One example of the dominant-negative effect associated with the DBD variant is reflected in the brain morphology across the models. Mice harboring a DBD variant exhibit several significant anomalies across different brain regions, whereas heterozygous KO mice show alterations limited to thinning of the corpus callosum. However, previous studies have additionally reported hippocampal volume loss in both tissue-specific and constitutive heterozygous *Nr2f1* KO mice (Flore et al. 2017; Chen et al. 2020). The hippocampus is of particular interest, given its relevance to intellectual disability and potential improvements in long-term memory reported in individuals with BBSOAS. Patients also show structural abnormalities such as medial temporal lobe dysgyria (Desai et al. 2023; Pinter et al. 2001; Bertacchi et al. 2020). However, direct comparison with human MRI data remains challenging due to interspecies differences in brain development, as well as methodological limitations. While we did not observe reduced hippocampal volume in the *Nr2f1*^{+/-} model, both the *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice exhibited marked reductions at various postnatal stages. A consistently observed feature across models, and one that aligns with neuroimaging findings in BBSOAS patients, is the thinning of the corpus callosum, a hallmark phenotype in the disorder (Chen et al. 2016; Desai et al. 2023;

Bertacchi et al. 2020). Notably, our analysis has revealed a previously undescribed phenotype in the point mutation models: significant enlargement of the lateral ventricles. This novel finding merits further investigation to determine whether it reflects underlying brain atrophy, developmental hypoplasia, or disrupted cerebrospinal fluid dynamics.

In parallel with the observed neuroanatomical alterations, we identified several behavioral abnormalities across the three *Nr2f1* mouse models. Our behavioral assessments focused particularly on autism-like features in rodents (USV and SIT). Notably, all models exhibited a robust reduction in USVs at P4, with the strongest impairments observed in *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice, and a milder deviation in *Nr2f1*^{+/-} mice. USVs are considered innate behaviors in rodents, and a reduced number of calls is a well-established hallmark in mouse models with autism-like phenotypes, making this parameter a robust and intuitive readout for modeling BBSOAS-related social impairments (Kikusui et al. 2011; Ey et al. 2013). Consistent with clinical data indicating that social deficits in BBSOAS are not confined to a particular developmental window, *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice exhibited persistent qualitative alterations in USV patterns at later developmental stages, alongside anomalies in social behavior as assessed by the SIT at P28/29. Although *Nr2f1*^{+/*E397**} mice displayed a comparatively weaker phenotype in the SIT, they still demonstrated reduced social interest in line with the findings from *Nr2f1*^{+/*R139L*} mice. The reasons for the mild SIT phenotype observed in *Nr2f1*^{+/*E397**} mice and the lack of detectable abnormalities in *Nr2f1*^{+/-} mice have not been thoroughly investigated, but could be influenced by several confounding and systemic factors, such as affective state of the animals, experimental setup, limited sample size, and the age of the mice, all of which can obscure smaller behavioral effects. Interestingly, our findings in the *Nr2f1*^{+/*R139L*} model are consistent with the results reported by Zhang et al., who observed similar social impairments in mice carrying a variant in the DBD (Zhang et al. 2020).

The hyperactive phenotype observed in our experiments, characterized by increased total distance traveled, is particularly notable in light of the subtle yet distinct motor dysfunctions identified through CatWalk XT analysis. Gait alterations, such as changes in stride length and paw placement, suggest impaired coordination that mirrors clinical findings in BBSOAS patients, where hyperactivity often coexists with

motor delays or hypotonia. These results imply that hyperactivity in the mouse models may not solely reflect behavioral changes but could also stem from underlying neuromotor dysfunction, highlighting the value of combining behavioral and gait assessments to capture the full spectrum of BBSOAS-related phenotypes.

To date, four studies have evaluated motor function in different *Nr2f1* models (Tomassy et al. 2010; Contesse et al. 2019; Chen et al. 2020; Zhang et al. 2020). Conditional knockout of *Nr2f1* in neocortical and hippocampal neurons led to deficits in fine motor coordination tasks (e.g., skilled-reaching and adhesive-removal task), while general motor performances (rotarod and wire hang test) remained unaffected (Tomassy et al. 2010; Contesse et al. 2019). Similarly, Chen et al. reported no major motor deficits in constitutive heterozygous *Nr2f1* KO (Chen et al. 2020). Furthermore, Zhang et al. found no abnormalities in rotarod in their DBD model, although only male mice were assessed (Zhang et al. 2020). Based on these findings and our own experience with the three mouse models, we did not anticipate a pronounced motor phenotype. However, given previous evidence of *Nr2f1*-related prefrontal/motor area defects, we employed the CatWalk XT system to detect any subtle motor coordination deficits (Tomassy et al. 2010; Armentano et al. 2007). This approach revealed a more nuanced motor phenotype, with significant gait alterations of *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice. Notably, all three models consistently showed increased front paw width, a hallmark often reported in other models of coordination deficits (Scarrott et al. 2023). In addition, a wider stance suggests impairments in balance and coordination. Since front paw width typically correlates with body weight, this effect is expected to be more pronounced in lighter mice, which would show a narrower stance (Pitzer et al. 2021). However, despite their lower body weight, variant-carrying mice show an increased front paw width. The disproportionate stance suggests a fundamental disruption in balance and motor control.

The mutation-dependent effects observed in paw coupling, particularly pronounced in *Nr2f1*^{+/*R139L*} mice, underscore deficits in a coordination task closely linked to fine motor functions. These findings support the hypothesis that *Nr2f1* plays a greater role in fine motor control than in gross motor performance (Tomassy et al. 2010). Overall, the lack of significant impairments in general motor performance aligns with both our

direct observations and previous reports, reinforcing the idea that *Nr2f1*-related dysfunction primarily affects coordination rather than basic motor ability.

BBSOAS is associated with a range of neurological and behavioral features, among which intellectual disability (ID) is one of the most impactful symptoms, significantly affecting the ability of patients to engage in daily life. Although nearly all individuals with BBSOAS present some degree of ID, its severity varies widely, ranging from mild to severe (Bertacchi et al. 2022). This phenotypic variability presents challenges for modeling in animals, yet it also underscores the value of cognitive testing in mouse models as a means to dissect genotype-phenotype correlations.

Previous studies have demonstrated spatial learning and memory deficits in *Nr2f1* neocortical- and hippocampal-specific knockout models using the Morris water maze (Flore et al. 2017). In line with the improved long-term memory observed in some patients, Chen et al. also observed altered fear memory (Chen et al. 2020) in constitutive heterozygous *Nr2f1* knockout mice using the fear conditioning test (Chen et al. 2020; Rech et al. 2020). To build on these findings, we conducted the APA test. In accordance with earlier reports, *Nr2f1*^{+/-} mice showed no significant changes in learning and memory. However, both *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice displayed marked learning impairments.

A potential explanation for the absence of learning and memory deficits in *Nr2f1*^{+/-} mice may be their reliance on non-spatial strategies. Rather than encoding specific spatial locations, these mice might compensate for hippocampal dysfunction, particularly deficits in allocentric spatial processing, by adopting alternative strategies such as memorizing distances or movement patterns (DiMattia & Kesner 1988; Flore et al. 2017). Although the heterozygous KO mouse is an established model for BBSOAS, its behavioral phenotype remains mild, potentially due to the greater sensitivity of human cognitive assessments in detecting subtle impairments not captured by standard rodent paradigms.

BBSOAS patients are frequently described as having remarkably good long-term memory (personal communication, Dr. Schaaf). While Chen et al. reported enhanced long-term memory in a *Nr2f1* variant model based on contextual fear conditioning,

our APA memory trials revealed an opposing trend (Chen et al. 2020). This discrepancy may be attributable to impairments during the learning phase, which could impact subsequent memory performance. However, as illustrated in Fig. 5C, the animals ultimately learn the location of the active zone, indicating that learning might be delayed rather than absent. Further investigations are required to delineate the specific memory-related phenotypes associated with *Nr2f1* variants.

Nevertheless, the novel mouse models presented in this study offer a good representation of the key features of BBSOAS symptoms, including autism-like behaviors, hyperactivity, fine motor dysfunction, and intellectual disability. Importantly, our results provide the first experimental evidence that pathogenic point mutations in the DBD (and to a lesser extent in the LBD) lead to more severe phenotypes than the heterozygous KO. When compared to the symptom severity index proposed by Bertacchi et al. for the different classes of *NR2F1* variants (deletions, DBD and truncation variants), our finding align closely with clinical observations, with heterozygous variants in the DBD causing the most pronounced phenotype followed by truncations (*Nr2f1*^{+/*E397**} variants in the LBD) and deletions. These results support the notion of a genotype-phenotype correlation and suggest a dominant-negative mechanism, which needs to be further understood. Bertacchi et al. proposed functional obligatory dimerization as a possible explanation, but neomorphic effects altering *Nr2f1* DNA-binding specificity may also contribute and need to be investigated. Furthermore, it is important to note that only 63% of the known patients fall into one of the three main variant categories (16% deletions, 35% DBD variants, and 12% truncations) (Bertacchi et al. 2022). Due to practical constraints, rarer genotypes, such as frameshift variants (8%), were not represented in this study. Moreover, while we analyzed variants that we consider broadly representative, most variants occur *de novo*. Although clinical data suggest that variants can be categorized by locus and type, the absence of recurrent hotspot mutations limits our ability to generalize findings. In addition, further investigation into the potential contribution of visual impairment observed in affected individuals to behavioral abnormalities such as hyperactivity, anxiety, or motor dysfunction will be crucial to further define and understand the phenotype.

In summary, the three mouse models presented here faithfully replicate key aspects of disease-relevant symptoms of BBSOAS and represent a valuable platform for

investigating disease mechanisms. In particular, the novel *Nr2f1*^{+/R139L} and *Nr2f1*^{+/E397*} models provide new opportunities for investigating the underlying pathophysiology of more severe cases of BBSOAS. The pronounced phenotype observed in these models also opens up new therapeutic avenues for symptom mitigation. Reliable models are essential for exploring therapeutic strategies, and based on the strong dominant-negative effect seen in the *Nr2f1*^{+/R139L} mice, we have initiated therapeutic studies using antisense oligonucleotides to selectively downregulate the pathogenic *R139L* allele. In addition, these models are relevant beyond the scope of BBSOAS, providing the scientific community with novel tools to study autism-like behaviors, coordination deficits, and broader neurodevelopmental abnormalities.

In conclusion, the presented models provide a robust platform for dissecting BBSOAS pathophysiology and improving our understanding of genotype-phenotype correlations. Further studies, including complementary behavioral phenotyping and therapeutic approaches, are ongoing. Addressing these challenges will ultimately require a coordinated effort among teams specializing in neurodevelopmental disorders, and we are open to collaborating with others. We anticipate that our work will significantly enhance the understanding of BBSOAS and serve as a valuable resource for the NR2F1 research community.

Materials and Methods

Animals

Mice were maintained in a specific-pathogen-free facility at the Interfaculty Biomedical Facility (IBF) of Heidelberg University and at the Centre d'Exploration Fonctionnelle Pré-clinique of the Institute of Biology Valrose, University Côte d'Azur, Nice, France, on a 12-hour light/dark cycle. They had ad libitum access to food and water. Environmental conditions were controlled at 50-60% humidity and 22°C (± 2°C).

The behavioral experiments were performed at the Interdisciplinary Neurobehavioral Core in Heidelberg by the same experimenter, who was blinded to the genotype. All animals started on the same behavioral timeline. This was done to ensure that all

animals had the same experimental experience at all stages. Testing was performed in two cohorts, consisting of 66 and 142 animals, respectively. Animals carrying a variant were tested alongside their wild-type littermates. Genotyping was performed after the USV test, and the cohorts were subsequently reduced to the target group size of ~14 animals per group by randomly excluding surplus litters or littermates (see Table S7). The tests were conducted during the light phase and in strict compliance with national and international guidelines for the Care and Use of Laboratory Animals. The experimenter ensured that background noises during testing were kept to a minimum. The study was approved by the Regional Council Regierungspräsidium Karlsruhe, Germany (G-172/21, G-177/21, and G-252/19), and by the local ethical committee (CIEPAL Azur 28 NCE/2024-976) and the Ministry for Higher Education and Research, France.

We tested three genetic mouse lines carrying patient-specific *NR2F1* variants: *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**}, all maintained on a C57BL/6J background (Armentano et al. 2006). Littermates from each respective mouse line served as wild-type controls. Each model represents a distinct patient subgroup, and both male and female mice were tested in balanced proportions and showed similar behavior (Fig. S1A and Table S7). For details and genotyping information, refer to Table S1.

Generation of the *Nr2f1*^{+/-} mouse model

Nr2f1^{+/-} mouse embryos were obtained via the European Mouse Mutant Archive (EMMA-international strain name: B6.129S2-Nr2f1^{tm1Mist}/Cnrm), and the colony at the IBF was established via embryo transfer. For details on the generation of this mutation, see Armentano et al. (Armentano et al. 2006).

Generation of the *Nr2f1*^{+/*R139L*} mouse model

The *Nr2f1*^{+/*R139L*} mouse line was generated at the Baylor College of Medicine in Houston, Texas, by the Genetically Engineered Rodent Models Core using CRISPR/Cas9-mediated genome editing. The mutation was introduced at the end of the first exon of *Nr2f1*. The mutation replaced the CGC codon (arginine) with a CTC codon, resulting in the substitution of arginine by leucine (R139L), see Fig. S6 for CRISPR/Cas9 design. To facilitate genotyping, donor DNA was designed to introduce the desired mutation along with silent mutations, and sequencing primers

were developed targeting the mutated region. The Cas9 protein and the single-guide RNA (sgRNA), together with the ssODN repair template, were microinjected into fertilized C57BL/6J oocytes. Following microinjection, chimeric founders were generated and subsequently mated with C57BL/6J females to achieve germline transmission of the mutation.

Generation of the *Nr2f1*^{+/E397*} mouse model

The E397* mouse line was generated at PHENOMIN-iCS (Institut Clinique de la Souris, Illkirch, France) as part of the PHENOMIN-2020 call for proposals, with funding provided by PHENOMIN. Three successive point mutations (CGA>ATG) were introduced at the beginning of exon 3 of the *Nr2f1* 1 gene, targeting nucleotide position (1188 to 1190). Two of the mutations converted the GAA codon (glutamic acid) into a TGA stop codon, resulting in a premature termination at amino acid residue 397 (E397*), see Fig. S7 for CRISPR/Cas9 design. The third mutation was silent at the protein level but introduced an NdeI restriction site to facilitate genotyping. Wild-type Cas9 protein and the crRNA/tracRNA (crRNA sequence 5'-tctcggatgagagtttcgat-3' score 88 according to <https://crispor.gi.ucsc.edu/>) and a single-stranded repair template containing the desired mutations, were electroporated into fertilized C57BL/6N oocytes. Two mosaic founders were obtained and mated with C57BL/6N females to achieve germline transmission of the mutation. Prior to experiments, the line was backcrossed (8-10 generations) into the C57BL/6J background. For Primer details see Table S1. The amplified products were digested with NdeI (30 min at 37°C) to discriminate the wild-type allele from the mutant.

Real time qRT-PCR

E14.5 mouse embryo brains were dissected, cut in half and frozen at -80° C. Total RNA was isolated with NucleoSpin RNA II columns (Macherey-Nagel; 740902.50). RNA quantity and quality were evaluated with Nanodrop and agarose gel electrophoresis. For each sample, 200 ng of total RNA were reverse transcribed using random nonamers (Reverse Transcriptase Core Kit, Eurogentec, qRT-RTCK-03). qRT-PCR was performed with the GoTaq SYBR Green qPCR Mix (Promega A6001) on a LightCycler® 480 (Roche). cDNA, stored at -20 °C, was diluted to ensure 2 ng per reaction. Amplification take-off values were determined using the built-in LightCycler® 480 relative quantification analysis function, and relative

expression was calculated with the $2^{-\Delta\Delta C_t}$ method, normalizing to the housekeeping gene β -Actin. At least two reactions were assembled per sample/gene analyzed during qRT-PCR amplification (technical replicates). Primer sequences used are listed in Table S1.

Cresyl Violet (Nissl) staining protocol

Cryostat sections were air-dried for 60 minutes at room temperature (RT). Mice were anesthetized using a mixture of Tiletamine-Zolazepam-Xylazine-Buprenorphine (TZXB) and perfused intracardially with phosphate-buffered saline (PBS), followed by 4% paraformaldehyde (PFA) in PBS. Brains were dissected, postfixed in 4% PFA for 24 hours at 4°C, and then cryoprotected by sequential immersion in sucrose-PBS solutions (10% sucrose for 12h, then 25% for 20h at 4°C) before being embedded in Cryomatrix (Epredia, ref: 6769006) and stored at -80°C. Coronal brain sections (40 μ m thick) were cut using a Leica cryostat and collected as free-floating sections in 12-well plates containing PBS with 0.05% sodium azide. Brain sections were then mounted and air-dried for 12h at RT. Slides were briefly post-fixed in 4% PFA for 10 minutes at RT, followed by three washes in PBS (5 minutes each). Sections were then defatted in a 1:1 acetone: ethanol solution for 10 min at RT and equilibrated in MilliQ water for 2 minutes. Staining was performed by incubating slides for 1 hour at 37°C in 0.1% Cresyl Violet Acetate (Sigma, C5042) prepared in 0.2% acetic acid in water. After staining, slides were rinsed in 96% ethanol, quickly dehydrated in 100% ethanol, and cleared in xylene for 5 minutes. Finally, sections were coverslipped using VectaMount mounting medium (ref: H-5000).

Brain Magnetic Resonance Imaging

Brain Magnetic Resonance Imaging (MRI) was performed in adult mice under isoflurane anesthesia (induction: 4%, maintenance: 1.5–2%) using a 3T scanner. T2-weighted images were acquired with a RARE sequence (TR: 3350 ms, TE: 48 ms) to assess brain morphology. Images were acquired at a resolution of 0.104 mm $\times$ 0.104 mm with an image matrix of 192 $\times$ 192 and a field of view (FOV) of 20 mm $\times$ 20 mm. Body temperature was maintained at 37°C, and respiration was monitored throughout the procedure. Image analysis was conducted using Horos, 3D Slicer, and IMARIS to assess volumetric and structural differences. In brief, brains and ventricles were manually segmented based on T2 hyperintense signal and exported

as binary labelmaps. A surface mesh was generated using ITKSnap and imported to IMARIS for 3D visualization and rendering.

Neonatal Ultrasonic Vocalization

At postnatal days (P) 4, 8, and 12, pups were randomly selected, separated from their mother and littermates, and placed in a sound-attenuating chamber (42 cm × 42 cm × 42 cm) with 350 lx illumination. All sounds emitted by the pups were recorded for 5 minutes. Following the session, pups were immediately returned to their home cages.

Ultrasonic Vocalizations (USVs) were recorded using an UltraSoundGate condenser microphone (CM16/COMPA, Avisoft Bioacoustics) positioned 30 cm above the testing arena. For analysis, a custom-developed software was utilized, employing both entropy-based thresholding and a convolutional neural network (Herdt et al. 2024). The software enables the detection and classification of individual USVs into predefined categories. To validate the accuracy of the analysis, randomly selected samples were manually verified using SASLabPro software (Avisoft Bioacoustics). Calls were detected with 94.9% recall and 99.3% precision (Herdt et al. 2024). The subsequent classification achieved an accuracy of 86.8% (Herdt et al. 2024).

Elevated Plus Maze (EPM)

At P21, mice were placed in the center of the arena. Voluntary movements were automatically tracked (Sygnis Tracker, custom-made) for 10 minutes. The brightness was controlled at 210 lx in the center, 200 lx in the open arms, and 170 lx in the closed arms (Ayala et al. 2010).

Open Field (OF)

Based on established literature, the central area brightness was set to 250 lx (Martin-Arenas & Pintado 2019). At P22, mice were placed in alternating corners of the arena, and their spontaneous movement was recorded (Sygnis Tracker) for 10 minutes.

Dark/Light Box (DLB)

At P24, mice were placed at the threshold between the light and the dark compartment, and their movements were automatically tracked (Sygnis Tracker) for

10 minutes. Brightness levels were set at 300 lx in the light compartment and 2 lx in the dark compartment (Bourin & Hascoët 2003).

Hole Board (HB)

At P26, mice were placed on the board for a 10-minute session, and head dips were manually recorded. Tests were conducted at 170 lx brightness.

Social Interaction Test (SIT)

At P28 or 29, mice were placed in the arena in three consecutive trials. The arena was illuminated at 160 lx (Saw n.d.). During the first trial, the test animal explored the three chambers. In the second trial, a same-sex littermate was placed inside a pencil cup in one of the side chambers. In the third trial, a novel intruder mouse, a same-sex adult CD-1 mouse, was introduced into the other chamber. Each trial lasted 5 minutes, with the test animal initially placed in the middle chamber. A 15-minute interval was maintained between trials. The time spent and distance traveled in each chamber were automatically recorded.

LABORAS

The LABORAS (Laboratory Animal Behavior Observation Registration and Analysis System) platform is a home-cage monitoring system. We measured behavioral parameters between P31 and P35, including locomotion, grooming, rearing, and climbing. The testing environment during the light phase was maintained at 160 lx. Mice were individually housed in LABORAS cages for 24 hours with ad libitum access to food and water. Behavior was tracked automatically via the LABORAS software.

CatWalk XT (CW)

The CatWalk XT gait analysis system (Noldus, Wageningen, The Netherlands) assessed locomotion at P36, including step sequences, speed, and paw coordination. Mice were placed in a 1 m long, red-light illuminated corridor (10 lx). Based on the literature, no training was done (Deumens et al. 2007). Testing continued until three recordings with a run variation below 30% were obtained. For analysis, both the standard parameters provided by the CatWalk XT system and a machine learning model were used. The computerized gait analysis examines a

comprehensive array of 317 parameters. To ensure a focused and interpretable analysis, we only examined specific parameters, e.g., gait speed or the width between the front paws. Additionally, to address the complexity inherent in the CW dataset, we employed a novel machine-learning algorithm developed by Timotius et al. (Timotius et al. 2021). The machine learning model integrates key CW parameters into a composite index representing overall locomotor performance.

For the analysis of individual parameters, we prioritized a subset of commonly used parameters, selected for their established relevance and intuitive interpretability (Heinzel et al. 2020) (Timotius et al. 2021) (Machado et al. 2015). Notably, 90% of the 317 parameters assessed by the CW system are speed-dependent (Heinzel et al. 2020). However, since none of the mouse models exhibited significant differences in speed (Fig. S4A), we performed a motor function deficit-focused comparison without confounding effects from variations in gait speed.

Active Place Avoidance (APA)

The Active Place Avoidance task was performed on two consecutive days between P40 and P50, based on previously established protocols (Zalucki et al. 2018). The round arena (77 cm in diameter) rotated at 1 revolution per minute. Lighting was set to 100 lx. The EthoVision XT software defined a 60° shock zone and delivered mild electric shocks upon entry to promote avoidance and motivate the animals to exit the zone. Learning was assessed over 8 trials with a 20 min break between each, followed by a memory test the next day.

Data Analysis

Outliers (values deviating by more than 3 standard deviations from the mean) were excluded from further analysis. Behavioral data were analyzed using a two-tailed homoscedastic (Student's) t-test. All datasets met our prerequisite of having a skewness less than 3. This approach follows recent mathematical recommendations regarding pre-testing, statistical test selection, and average sample sizes (Fagerland & Sandvik 2009; Rasch et al. 2011). All our findings are explorative. Statistical computations were performed using Python (libraries: pandas, numpy, and scipy.stats). Statistical significance was set at $P \leq 0.05$ for all analyses.

Acknowledgments

The authors thank Barbara Kurpiers for her support at the Interdisciplinary Neurobehavioral Core, Christine Fischer for her statistical advice, and Ilia Valentin for her help in writing this paper. We thank the Core Facility Small Animal Imaging at the DKFZ for the MRI scans.

Competing interests

The authors declare no financial or competing interests.

Funding

This study was supported by the German Research Foundation (Grant #532002382 to Christian Schaaf), by the Fondation pour la Recherche Médicale (Grant #EQU202003010222), Fondation de France (Grant #00123416), ERA-NET Neuron (Grant Number: ANR- 21-NEU2-0003-03), and by the PHENOMIN Program and the Institut Clinique de la Souris (ICS) in Strasbourg to M.S. Jonas G. Scheck was supported by a fellowship from the German Neurology Society. Paolo Piovani was supported by a fellowship from the “Région PACA Emplois Jeunes Doctorants 2021” and from the Fondation de France (#00154573).

Data and resource availability

All relevant data can be found within the article and its supplementary information. Additional original data can be shared by the corresponding author upon request.

Author contributions statement

Conceptualization: C.P.S., M.S., J.G.M., H.F.; Methodology: A.M., S.T., C.P., H.F., J.G.M., M.B., O.D., L.D., C.C., P.P.; Software: J.G.M., T.S., A.S.; Validation: J.G.M., A.M., S.T.; Formal analysis: J.G.M., C.P., A.S., D.K., A.M., M.S., J.S., A.M., P.P.; Investigation: J.G.M., D.K., A.M.; Resources: C.P.S., M.S., C.P.; Data curation: J.G.M., A.M., A.S.; Writing - original draft: J.G.M., C.P.S., F.F., T.S., D.K., O.D., M.S.; Writing - review & editing: J.G.M., M.S., C.P.S.; Visualization: J.G.M., C.P.S., A.M.; Supervision: C.P., H.F., M.S., M.B.; Project administration: J.G.M., C.P.S., M.S., C.P.; Funding acquisition: M.S., C.P.S.

References

- Armentano, M. et al., 2006. COUP-TFI is required for the formation of commissural projections in the forebrain by regulating axonal growth. *Development (Cambridge, England)*, 133(21), pp.4151–4162. Available at: <https://pubmed.ncbi.nlm.nih.gov/17021036/> [Accessed October 22, 2024].
- Armentano, M. et al., 2007. COUP-TFI regulates the balance of cortical patterning between frontal/motor and sensory areas. *Nature neuroscience*, 10(10), pp.1277–1286. Available at: <http://dx.doi.org/10.1038/nn1958>.
- Ayala, J. et al., 2010. Standard operating procedures for describing and performing metabolic tests of glucose homeostasis in mice. *Disease models & mechanisms*, 3, pp.525–534. Available at: <https://journals.biologists.com/dmm/article/3/9-10/525/53437/Standard-operating-procedures-for-describing-and>.
- Baronio, D. et al., 2015. Effects of an H3R antagonist on the animal model of autism induced by prenatal exposure to valproic acid. *PloS one*, 10(1), p.e0116363. Available at: <https://pubmed.ncbi.nlm.nih.gov/25560049/> [Accessed November 28, 2024].
- Bertacchi, M. et al., 2020. NR2F1 regulates regional progenitor dynamics in the mouse neocortex and cortical gyrification in BBSOAS patients. *The EMBO journal*, 39(13), p.e104163. Available at: <http://dx.doi.org/10.15252/emboj.2019104163>.
- Bertacchi, M. et al., 2022. Pathophysiological Heterogeneity of the BBSOA Neurodevelopmental Syndrome. *Cells*, 11(8). Available at: <http://dx.doi.org/10.3390/cells11081260>.
- Bosch, D.G.M. et al., 2014. NR2F1 mutations cause optic atrophy with intellectual disability. *American journal of human genetics*, 94(2), pp.303–309. Available at: <http://dx.doi.org/10.1016/j.ajhg.2014.01.002>.
- Bourin, M. & Hascoët, M., 2003. The mouse light/dark box test. *European journal of pharmacology*, 463(1-3), pp.55–65. Available at: [http://dx.doi.org/10.1016/s0014-2999\(03\)01274-3](http://dx.doi.org/10.1016/s0014-2999(03)01274-3).
- Chen, C.-A. et al., 2020. Nr2f1 heterozygous knockout mice recapitulate neurological phenotypes of Bosch-Boonstra-Schaaf optic atrophy syndrome and show impaired hippocampal synaptic plasticity. *Human molecular genetics*, 29(5), pp.705–715. Available at: <http://dx.doi.org/10.1093/hmg/ddz233>.

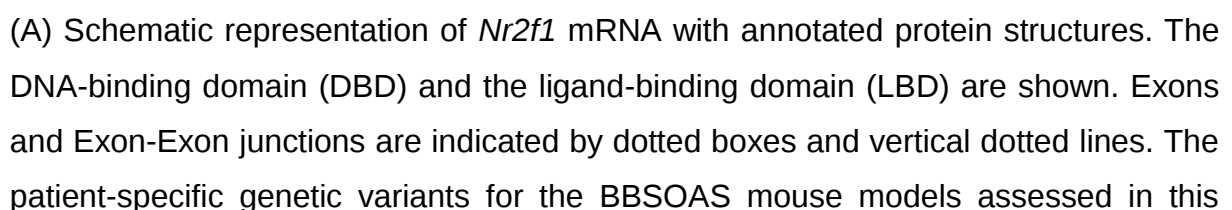
- Chen, C.-A. et al., 2016. The expanding clinical phenotype of Bosch-Boonstra-Schaaf optic atrophy syndrome: 20 new cases and possible genotype-phenotype correlations. *Genetics in medicine: official journal of the American College of Medical Genetics*, 18(11), pp.1143–1150. Available at: <http://dx.doi.org/10.1038/gim.2016.18>.
- Coffey, K.R., Marx, R.E. & Neumaier, J.F., 2019. DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, 44(5), pp.859–868. Available at: <http://dx.doi.org/10.1038/s41386-018-0303-6>.
- Contesse, T. et al., 2019. Hyperactive and anxiolytic-like behaviors result from loss of COUP-TF1/Nr2f1 in the mouse cortex. *Genes, brain, and behavior*, 18(7), p.e12556. Available at: <https://pubmed.ncbi.nlm.nih.gov/30653836/> [Accessed December 12, 2024].
- Cooney, A.J. et al., 1992. Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTCA response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors. *Molecular and cellular biology*, 12(9), pp.4153–4163. Available at: <http://dx.doi.org/10.1128/mcb.12.9.4153-4163.1992>.
- Desai, N.K. et al., 2023. Common neuroimaging findings in Bosch-Boonstra-Schaaf optic atrophy syndrome. *AJNR. American journal of neuroradiology*, 44(2), pp.212–217. Available at: <https://pubmed.ncbi.nlm.nih.gov/36702506/> [Accessed March 21, 2025].
- Deumens, R. et al., 2007. The CatWalk gait analysis in assessment of both dynamic and static gait changes after adult rat sciatic nerve resection. *Journal of neuroscience methods*, 164(1), pp.120–130. Available at: <https://pubmed.ncbi.nlm.nih.gov/17532474/> [Accessed November 8, 2024].
- DiMattia, B.D. & Kesner, R.P., 1988. Spatial cognitive maps: differential role of parietal cortex and hippocampal formation. *Behavioral neuroscience*, 102(4), pp.471–480. Available at: <https://pubmed.ncbi.nlm.nih.gov/3166721/> [Accessed December 13, 2024].
- Ey, E. et al., 2013. The Autism ProSAP1/Shank2 mouse model displays quantitative and structural abnormalities in ultrasonic vocalisations. *Behavioural brain research*, 256, pp.677–689. Available at: <http://dx.doi.org/10.1016/j.bbr.2013.08.031>.

- Fagerland, M.W. & Sandvik, L., 2009. Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemporary clinical trials*, 30(5), pp.490–496. Available at: <https://pubmed.ncbi.nlm.nih.gov/19577012/> [Accessed November 17, 2024].
- Flore, G. et al., 2017. Gradient COUP-TFI expression is required for functional organization of the hippocampal septo-temporal longitudinal axis. *Cerebral cortex (New York, N.Y.: 1991)*, 27(2), pp.1629–1643. Available at: <https://pubmed.ncbi.nlm.nih.gov/26813976/> [Accessed December 12, 2024].
- Harro, J. ed., 2023. *Psychiatric vulnerability, mood, and anxiety disorders* 1st ed., New York, NY: Springer.
- Heinzel, J. et al., 2020. Use of the CatWalk gait analysis system to assess functional recovery in rodent models of peripheral nerve injury - a systematic review. *Journal of neuroscience methods*, 345(108889), p.108889. Available at: <https://pubmed.ncbi.nlm.nih.gov/32755615/> [Accessed November 8, 2024].
- Herd, R. et al., 2024. Enhancing the analysis of murine neonatal ultrasonic vocalizations: Development, evaluation, and application of different mathematical models. *The journal of the Acoustical Society of America*, 156(4), pp.2448–2466. Available at: <https://pubmed.ncbi.nlm.nih.gov/39400270/>.
- Jurkute, N. et al., 2021. Pathogenic NR2F1 variants cause a developmental ocular phenotype recapitulated in a mutant mouse model. *Brain communications*, 3(3), p.fcab162. Available at: <http://dx.doi.org/10.1093/braincomms/fcab162>.
- Kikusui, T. et al., 2011. Cross fostering experiments suggest that mice songs are innate. *PloS one*, 6(3), p.e17721. Available at: <http://dx.doi.org/10.1371/journal.pone.0017721>.
- Kloos, A.D. et al., 2005. Stepwise motor and all-or-none sensory recovery is associated with nonlinear sparing after incremental spinal cord injury in rats. *Experimental neurology*, 191(2), pp.251–265. Available at: <https://pubmed.ncbi.nlm.nih.gov/15649480/> [Accessed November 29, 2024].
- Machado, A.S. et al., 2015. A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. *eLife*, 4. Available at: <https://pubmed.ncbi.nlm.nih.gov/26433022/> [Accessed December 6, 2024].

- Maricelli, J.W. et al., 2016. Trendelenburg-like gait, instability and altered step patterns in a mouse model for limb girdle muscular dystrophy 2i. *PloS one*, 11(9), p.e0161984. Available at: <https://pubmed.ncbi.nlm.nih.gov/27627455/> [Accessed December 2, 2024].
- Martin-Arenas, F.J. & Pintado, C.O., 2019. Results of the Open Field Test at different light intensities in C57 mice. Available at: https://www.researchgate.net/publication/332112484_Results_of_the_Open_Field_Test_at_different_light_intensities_in_C57_mice [Accessed March 8, 2024].
- Mendes, C.S. et al., 2015. Quantification of gait parameters in freely walking rodents. *BMC biology*, 13(1), p.50. Available at: <https://pubmed.ncbi.nlm.nih.gov/26197889/> [Accessed December 5, 2024].
- Parisot, J. et al., 2017. COUP-TFI mitotically regulates production and migration of dentate granule cells and modulates hippocampal Cxcr4 expression. *Development (Cambridge, England)*, 144(11), pp.2045–2058. Available at: <https://pubmed.ncbi.nlm.nih.gov/28506990/> [Accessed December 12, 2024].
- Pinter, J.D. et al., 2001. Amygdala and hippocampal volumes in children with Down syndrome: a high-resolution MRI study. *Neurology*, 56(7), pp.972–974. Available at: <https://pubmed.ncbi.nlm.nih.gov/11294940/> [Accessed March 24, 2025].
- Pitzer, C., Kurpiers, B. & Eltokhi, A., 2021. Gait performance of adolescent mice assessed by the CatWalk XT depends on age, strain and sex and correlates with speed and body weight. *Scientific reports*, 11(1), p.21372. Available at: <https://pubmed.ncbi.nlm.nih.gov/34725364/> [Accessed November 1, 2024].
- Rasch, D., Kubinger, K.D. & Moder, K., 2011. The two-sample t test: pre-testing its assumptions does not pay off. *Statistical papers (Berlin, Germany)*, 52(1), pp.219–231. Available at: <https://link.springer.com/article/10.1007/s00362-009-0224-x>.
- Rech, M.E. et al., 2020. Phenotypic expansion of Bosch-Boonstra-Schaaf optic atrophy syndrome and further evidence for genotype-phenotype correlations. *American journal of medical genetics. Part A*, 182(6), pp.1426–1437. Available at: <http://dx.doi.org/10.1002/ajmg.a.61580>.
- Saw, N.L., Stanford Behavioral and Functional Neuroscience Laboratory. Available at: <https://med.stanford.edu/sbfnl/services/bm/si/three-chamber.html>.

- Scarrott, J.M. et al., 2023. Ap4b1-knockout mouse model of hereditary spastic paraplegia type 47 displays motor dysfunction, aberrant brain morphology and ATG9A mislocalization. *Brain communications*, 5(1), p.fcac335. Available at: <https://pubmed.ncbi.nlm.nih.gov/36632189/> [Accessed December 12, 2024].
- Scattoni, M.L. et al., 2008. Unusual repertoire of vocalizations in the BTBR T+tf/J mouse model of autism. *PloS one*, 3(8), p.e3067. Available at: <http://dx.doi.org/10.1371/journal.pone.0003067>.
- Takumi, T. et al., 2020. Behavioral neuroscience of autism. *Neuroscience and biobehavioral reviews*, 110, pp.60–76. Available at: <http://dx.doi.org/10.1016/j.neubiorev.2019.04.012>.
- Timotius, I.K. et al., 2021. Combination of defined CatWalk gait parameters for predictive locomotion recovery in experimental spinal cord injury rat models. *eNeuro*, 8(2), pp.ENEURO.0497–20.2021. Available at: <https://pubmed.ncbi.nlm.nih.gov/33593735/> [Accessed November 1, 2024].
- Tocco, C., Bertacchi, M. & Studer, M., 2021. Structural and Functional Aspects of the Neurodevelopmental Gene NR2F1: From Animal Models to Human Pathology. *Frontiers in molecular neuroscience*, 14, p.767965. Available at: <http://dx.doi.org/10.3389/fnmol.2021.767965>.
- Tomassy, G.S. et al., 2010. Area-specific temporal control of corticospinal motor neuron differentiation by COUP-TFI. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), pp.3576–3581. Available at: <https://pubmed.ncbi.nlm.nih.gov/20133588/> [Accessed December 12, 2024].
- Zalucki, O. et al., 2018. Analysis of hippocampal-dependent learning and memory behaviour in mice lacking Nfix from adult neural stem cells. *BMC research notes*, 11(1), p.564. Available at: <http://dx.doi.org/10.1186/s13104-018-3652-7>.
- Zhang, K. et al., 2020. Imbalance of Excitatory/Inhibitory Neuron Differentiation in Neurodevelopmental Disorders with an NR2F1 Point Mutation. *Cell reports*, 31(3), p.107521. Available at: <http://dx.doi.org/10.1016/j.celrep.2020.03.085>.

Disease Models & Mechanisms • DMM • Accepted manuscript



paper are annotated. (B-F) Real-Time qRT-PCR analysis of *Nr2f1* expression (total expression and expression of specific variants) in whole brain extracts from E14.5 mouse embryos. Primer pairs used for *Nr2f1* transcript detection are depicted below and colour-coded with the boxes in B-F. Bars indicate the standard error of mean (SEM). Primers marked with asterisks specifically recognize the mutant allele. *Nr2f1* WT allele expression is reduced in (B) *Nr2f1*^{+/-}, (C) *Nr2f1*^{+/*R139L*} and (D) *Nr2f1*^{+/*E397**} genotypes. The mutant allele is detected exclusively in (E) *Nr2f1*^{+/*R139L*} and (F) *Nr2f1*^{+/*E397**} embryos. Datapoints are normalized to the respective HET mean. (G) The experimental timeline outlines the sequence of behavioral tests and brain morphology analysis conducted from P4 to P60. Behavioral assessments were performed in two cohorts. Cohort sizes were progressively reduced by randomly excluding litters as the tests became more time-intensive. Brain morphology analyses were carried out in three separate cohorts for each experiment. (H) *Nr2f1*^{+/*R139L*} neonates exhibit reduced body weight at P4, P8, and P12. Bars indicate the SEM. (I) *Nr2f1*^{+/*R139L*} mice show a reduced body weight at P100. Datapoints are normalized to the respective WT mean. Box plots represent the median and 25–75th percentiles; whiskers indicate the range of the data. *P ≤ 0.05, ****P ≤ 0.0001; Student's t-test.

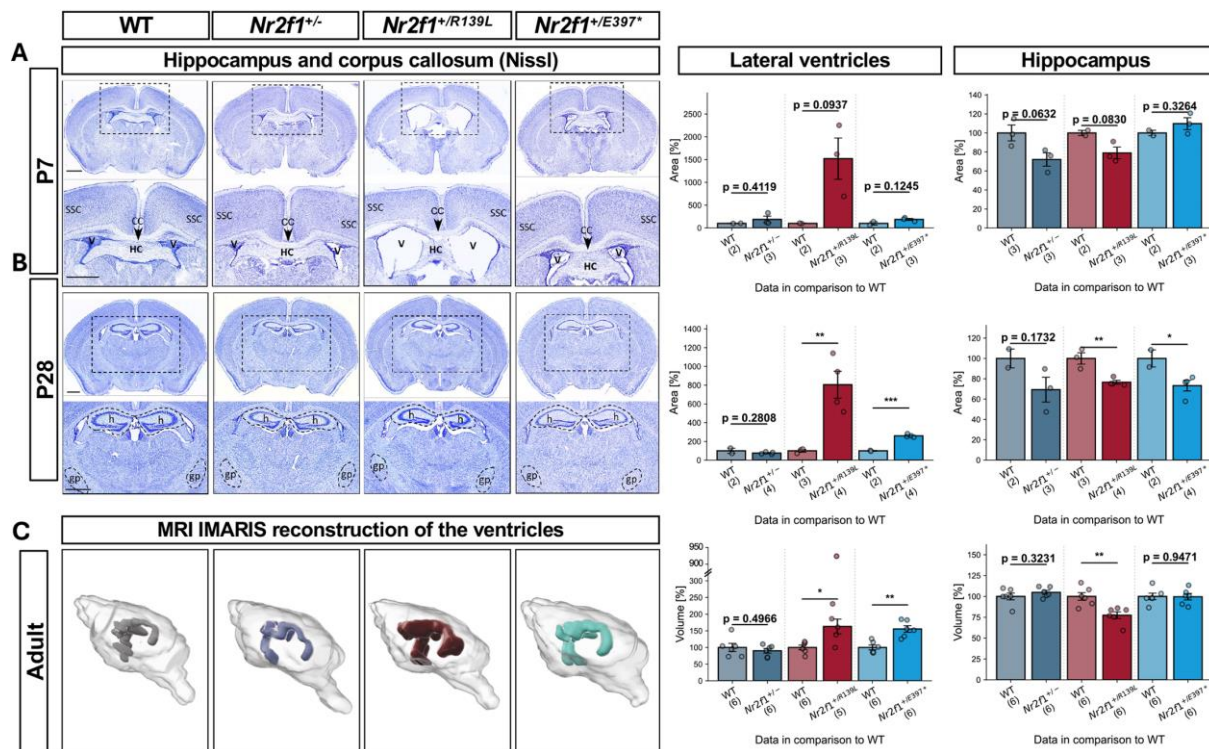


Fig. 2. Brain morphology in P7, P28, and adult mice.

(A) Nissl staining of *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} brains. Somatosensory cortex (SSC), corpus callosum (CC), hippocampal commissure (HC), and ventricles (V). *Nr2f1*^{+/*R139L*} mice tend to have enlarged lateral ventricles. (B) Nissl staining of P28 *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} brains. Hippocampus (h), globus pallidus (gp). *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice have statistically enlarged lateral ventricles, and a reduced hippocampal size. (C) MRI scan of 2-month-old WT, *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} brains. 3D reconstruction (IMARIS) of representative brains and the ventricular system. *Nr2f1*^{+/*R139L*} and *Nr2f1*^{+/*E397**} mice have statistically enlarged lateral ventricles. *Nr2f1*^{+/*R139L*} mice show a reduced hippocampus. Bars indicate the SEM. Datapoints are normalized to the respective WT mean. Box plots represent the median and 25–75th percentiles; whiskers indicate the range of the data. *P ≤ 0.05, ***P ≤ 0.001; Student's t-test. Scale bars = 1mm.

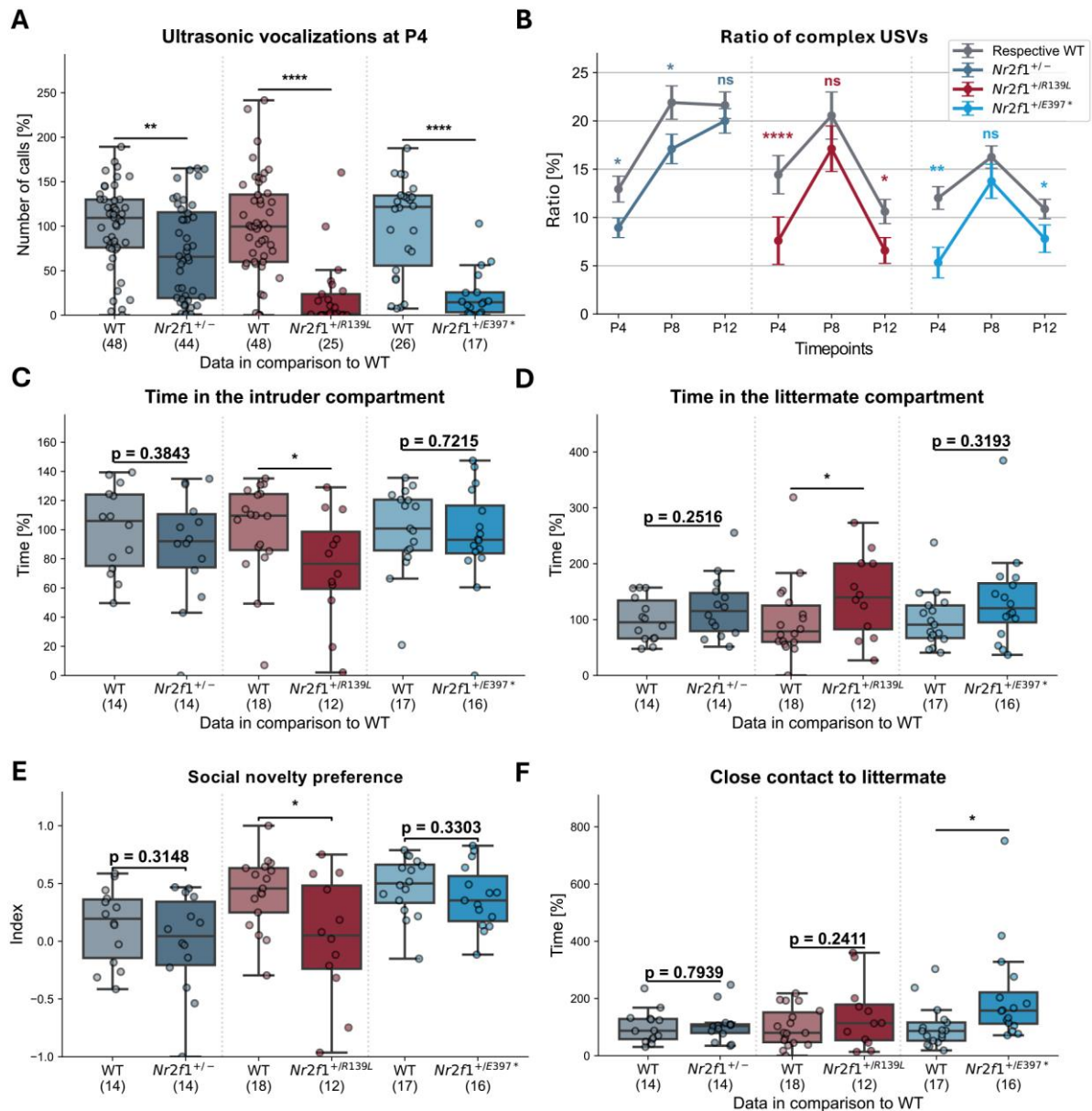


Fig. 3. Social and autism-like phenotype of mice with *Nr2f1* variants.

(A) Mice with *Nr2f1* variants emit fewer calls compared to WT littermates at P4. (B) Mice with *Nr2f1* variants show an altered call class profile, with a reduction in the prevalence of complex calls. Sample sizes for the *Nr2f1*^{+/-} line were n = 44, 41, and 40, and for the corresponding WT animals, n = 48, 46, and 39 at P4, P8, and P12, respectively. For the *Nr2f1*^{+/*R139L*} line, sample sizes were n = 25, 23, and 23, with corresponding WT animals n = 48, 45, and 42 at P4, P8, and P12, respectively. For the *Nr2f1*^{+/*E397**} line, sample sizes were n = 17, 17, and 16, and corresponding WT animals n = 26, 26, and 26 at P4, P8, and P12, respectively. Error bars represent the SEM. (C, D) *Nr2f1*^{+/*R139L*} mice spend less time in the compartment of the intruder and more time in the compartment of their littermate during the third trial of the social

interaction test (SIT). (E) *Nr2f1*^{+/*R139L*} mice have a lower social novelty preference index (“TimeWithIntruder” subtracted by “TimeWithLittermate” divided by “TimeWithIntruder” and “TimeWithLittermate”). (F) *Nr2f1*^{+/*E397**} mice spend more time in the close contact zone with their littermate during the third trial of the SIT. **P* ≤ 0.05, *****P* ≤ 0.0001; Student’s *t*-test. Datapoints in A, C, D, and F of each line are normalized to the respective WT mean. Box plots represent the median and 25–75th percentiles; whiskers indicate the range of the data.

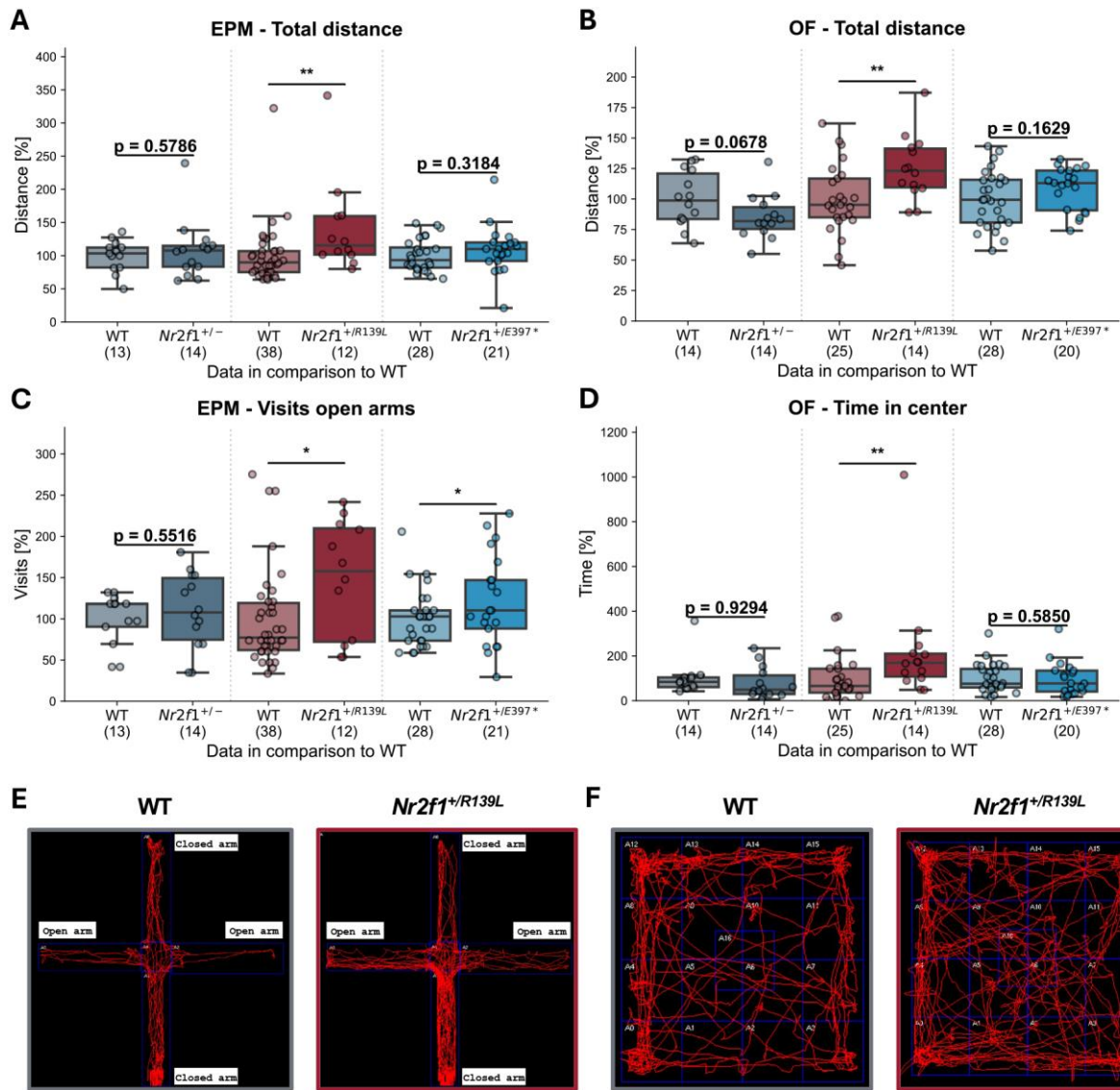


Fig. 4. Anxiety and exploratory behavior evaluated by the open field (OF) and elevated plus maze (EPM) tests in mouse models for BBSOAS.

(A, B) *Nr2f1*^{+R139L} show increased movement over the 10-minute test period compared to their respective WT littermates in both EPM and OF test. (C) *Nr2f1*^{+R139L} and *Nr2f1*^{+E397*} mice visit the open arms of the EPM arena more often. (D) *Nr2f1*^{+R139L} mice spend more time in the center of the OF arena and show increased movement over the 10-minute test period. *P ≤ 0.05, **P ≤ 0.01; Student's t-test.

Datapoints for each line are shown, normalized to the respective WT mean. Box plots represent the median and 25–75th percentiles; whiskers indicate the range of the data. (E, F) Representative track maps for WT and *Nr2f1*^{+R139L} mice in the EPM and OF. The selected animals have individual performance parameters close to the mean of their respective group.

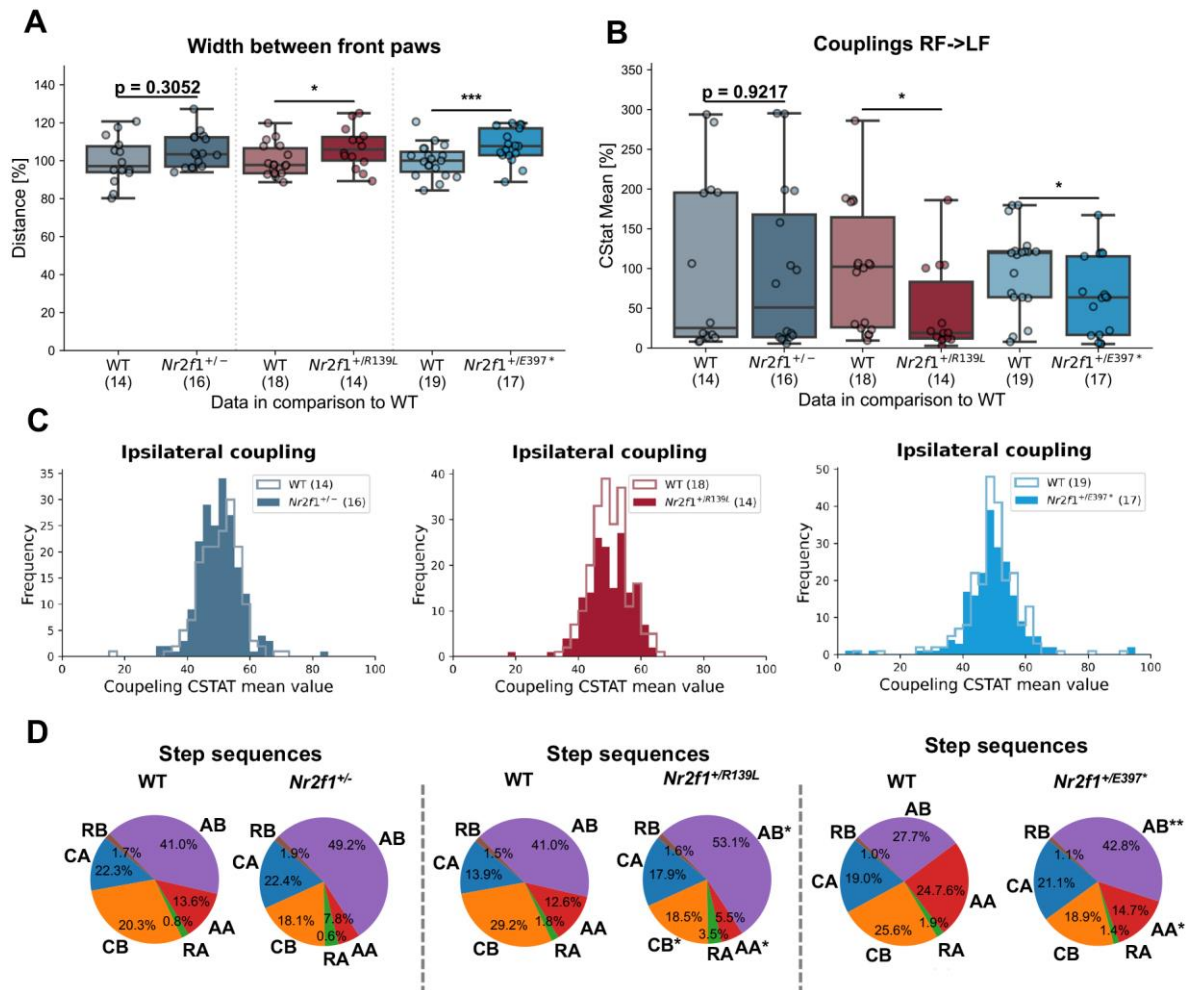


Fig. 5. Gait analysis using the CatWalk XT system.

(A) $Nr2f1^{+/-}$, $Nr2f1^{+/R139L}$, and $Nr2f1^{+/E397*}$ mice show an increase in the width between the front paws. (B) All mouse models display altered girdle coupling of the right front (RF) paw to the left hind (LF) paw. Datapoints for each line are shown, normalized to the respective WT mean. Box plots represent the median and 25–75th percentiles; whiskers indicate the range of the data. (C) Descriptive histograms showing the distribution of different coupling values for all 4 ipsilateral paw pairs for each mouse model. (D) Illustration of the step sequence type distribution for each line and genotype (WT and variant-carrying mice). Because of the data structure, no outlier analysis was done for the step sequence analysis. * $P \leq 0.05$, *** $P \leq 0.001$; Student's t-test.

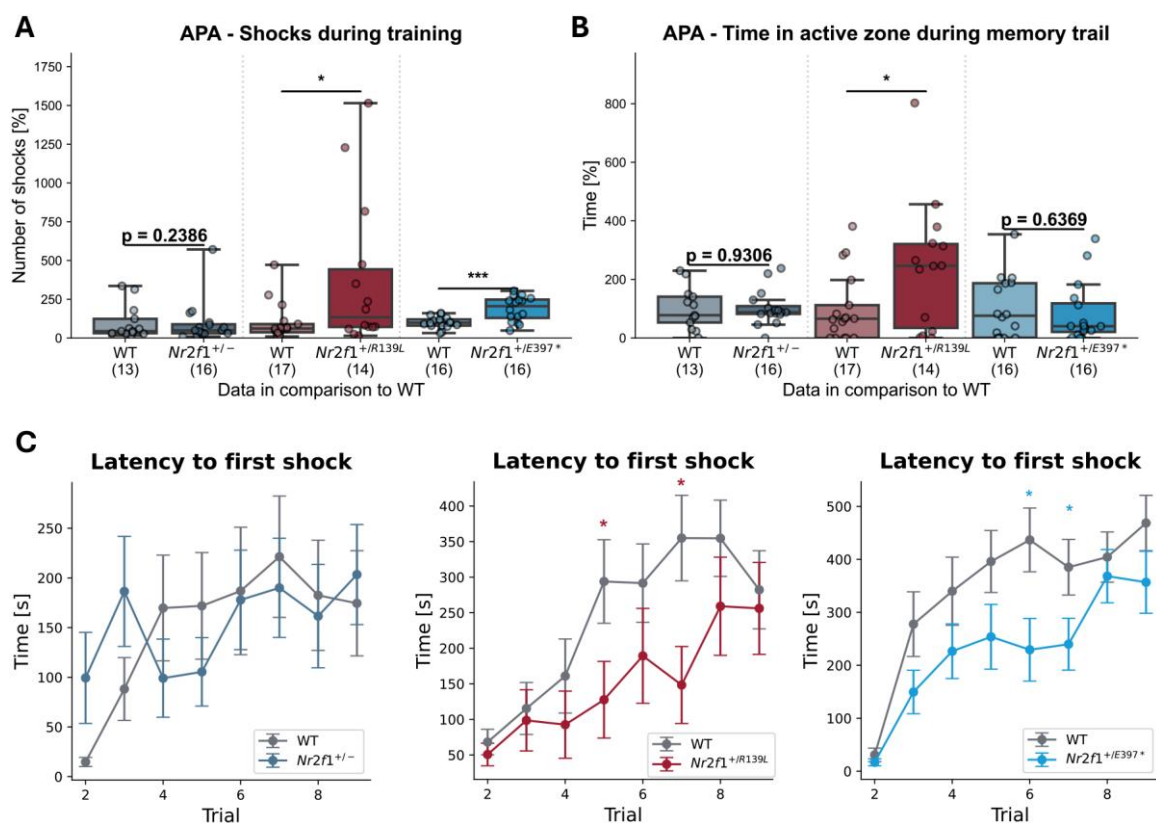


Fig. 6. Learning and spatial memory in the Active Place Avoidance test.

(A, B) $Nr2f1^{+/R139L}$ and $Nr2f1^{+/E397*}$ mice receive more cumulative shocks than their respective WT littermates during the 8 training trials of APA. $Nr2f1^{+/R139L}$ mice spend more time in the inactive shock zone during the recall trial, 24 hours after the training. Data points for each line are shown, normalized to the respective WT mean. Box plots represent the median and 25–75th percentiles; whiskers indicate the range of the data. (C) Descriptive figures show the learning process in comparison to the respective WT littermates. Bars represent the SEM. * $P \leq 0.05$, *** $P \leq 0.001$; Student's t-test.

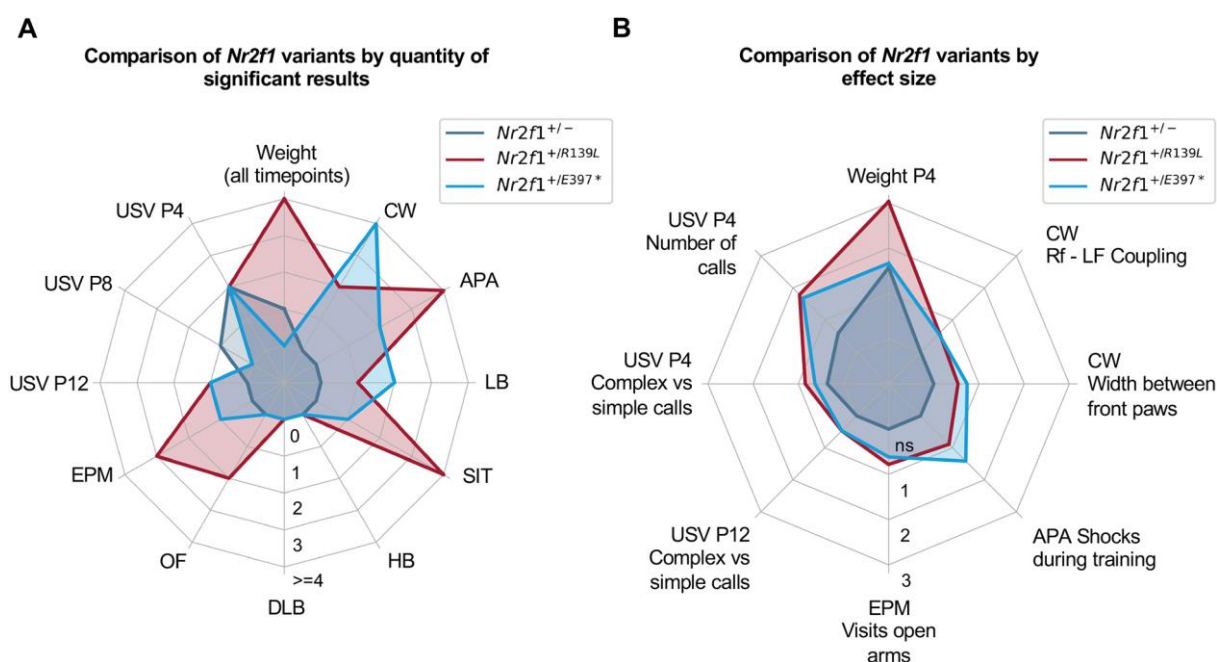


Fig. 7. Quantitative and qualitative analysis of the phenotypic differences between the mouse lines.

(A) Illustration of the number of significant results (as evaluated by the Student's *t*-test) for each experiment and mouse model. Since the tests vary in the number of measured parameters, and to maintain clarity and comparability, the radial scale was clipped at 4. For a detailed breakdown, see Table S1. (B) The effect of each mouse model for parameters significant across more than one mouse model was compared using Cohen's *d*. If the parameter was only significant in two of the three models, the non-significant value for the third line was set to 0 (ns). Cohen's *d* increases with the effect size, with values above 0.8 indicating a strong effect.

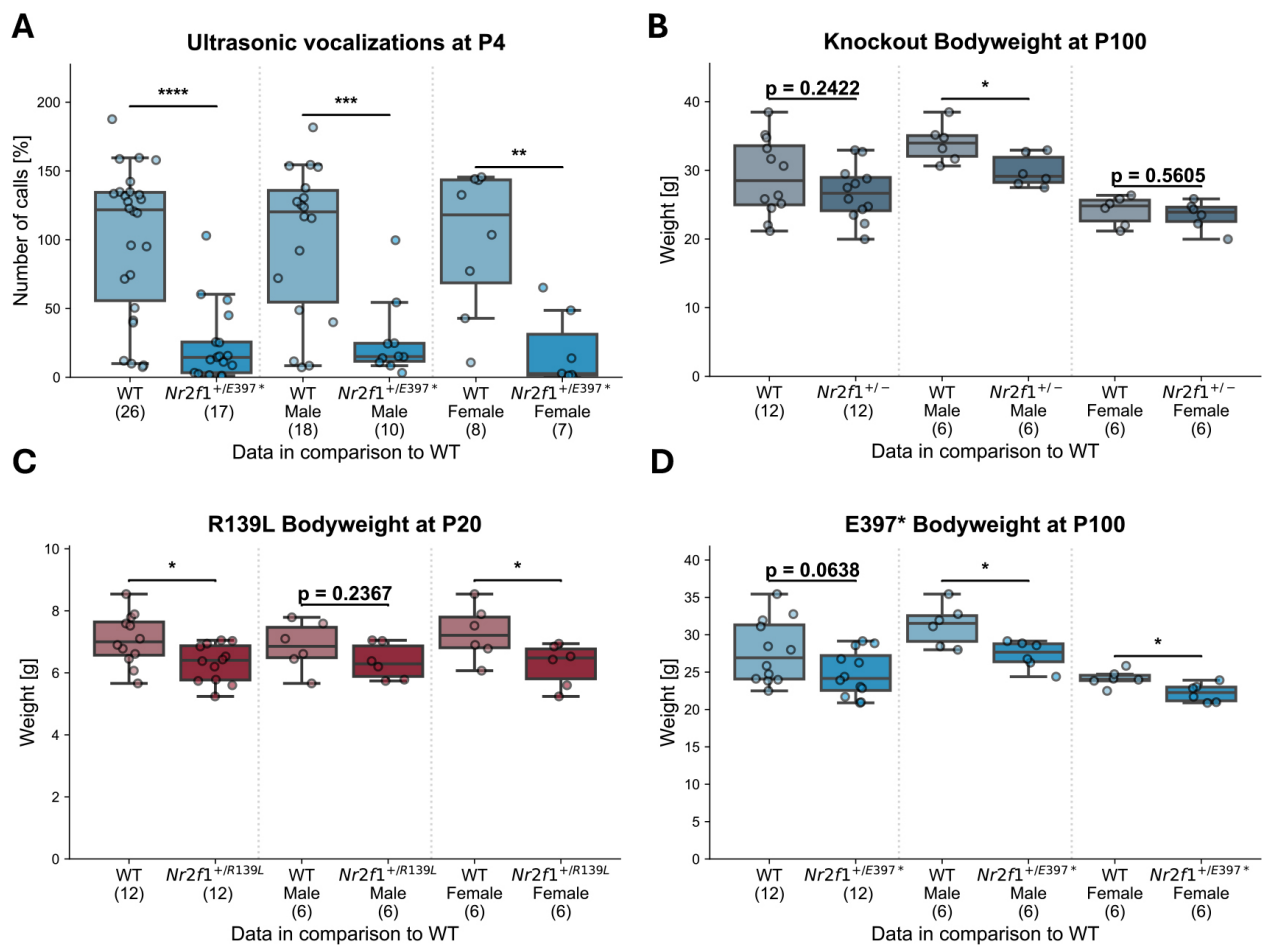


Fig. S1. Sex-specific analysis of selected parameters for distinct time points.

(A) An example of a sex-specific analysis, showing that both sexes show the same behavior. (B) At P100, $Nr2f1^{+/-}$ mice show a reduction in body weight for both male mice ($P=0.0215$). (C) At P20, $Nr2f1^{+/R139L}$ mice show a reduction in body weight for both sexes ($P=0.0176$), driven by the females ($P=0.0410$). (D) At P100, $Nr2f1^{+/E397*}$ mice show a reduction in body weight for both males ($P=0.0162$) and females. Data points for each line are shown, normalized to the respective WT mean. * $P \leq 0.05$. Student's t-test.

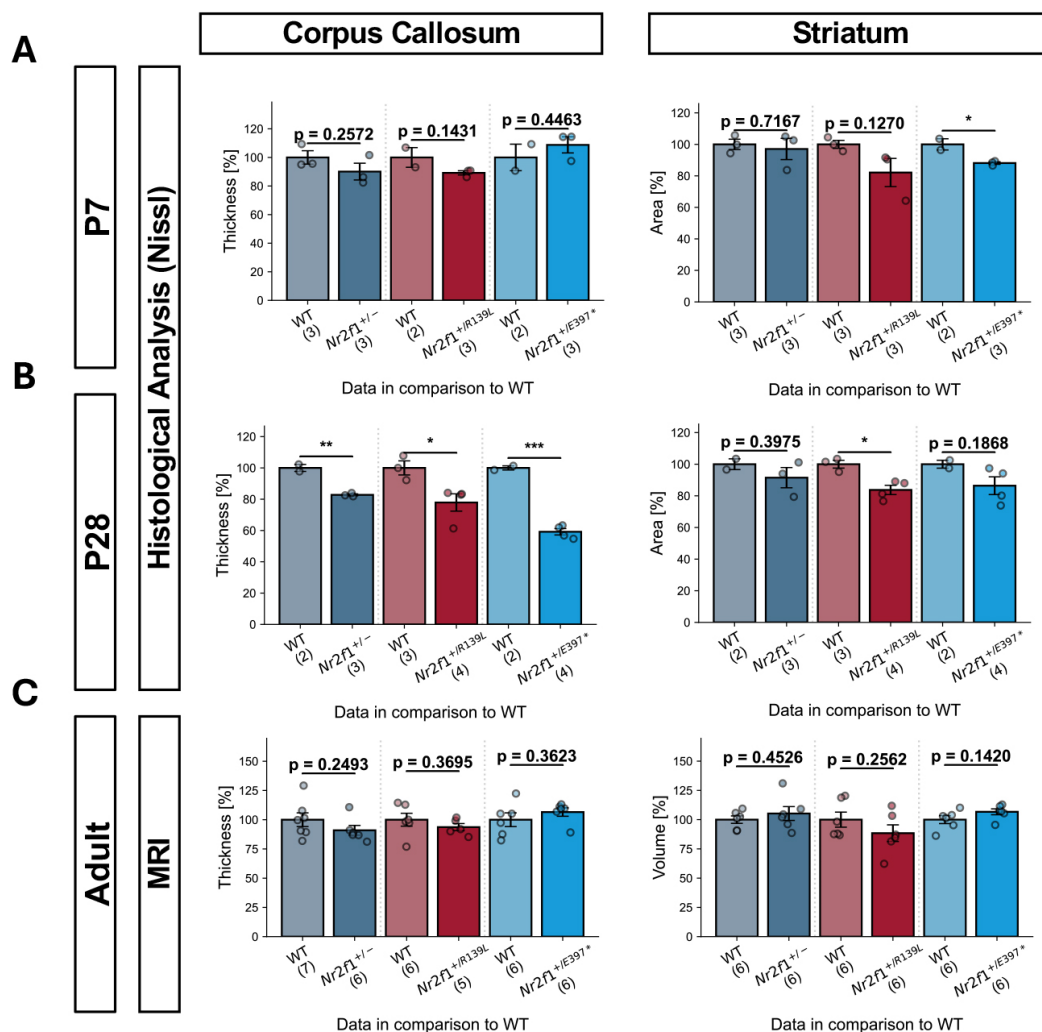


Fig. S2. Corpus Callosum and Striatum in P7, P28, and adult mice.

(A) Quantification of corpus callosum thickness and striatal area in P7 WT, $Nr2f1^{+/-}$, $Nr2f1^{+/R139L}$, and $Nr2f1^{+/E397*}$ brains (Nissl staining). No significant changes in corpus callosum thickness. Slightly reduced striatal area in $Nr2f1^{+/E397*}$ mice. (B) Quantification of corpus callosum thickness and striatal area in P28 WT, $Nr2f1^{+/-}$, $Nr2f1^{+/R139L}$, and $Nr2f1^{+/E397*}$ brains (Nissl staining). Significant thinning of the corpus callosum in all mutant lines and reduced striatal area in $Nr2f1^{+/R139L}$ mice. (C) Quantification of corpus callosum thickness and striatal volume in adult WT, $Nr2f1^{+/-}$, $Nr2f1^{+/R139L}$, and $Nr2f1^{+/E397*}$ mice (MRI analysis). No significant differences between genotypes. Data normalized to the respective WT mean. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$; Student's t-test.

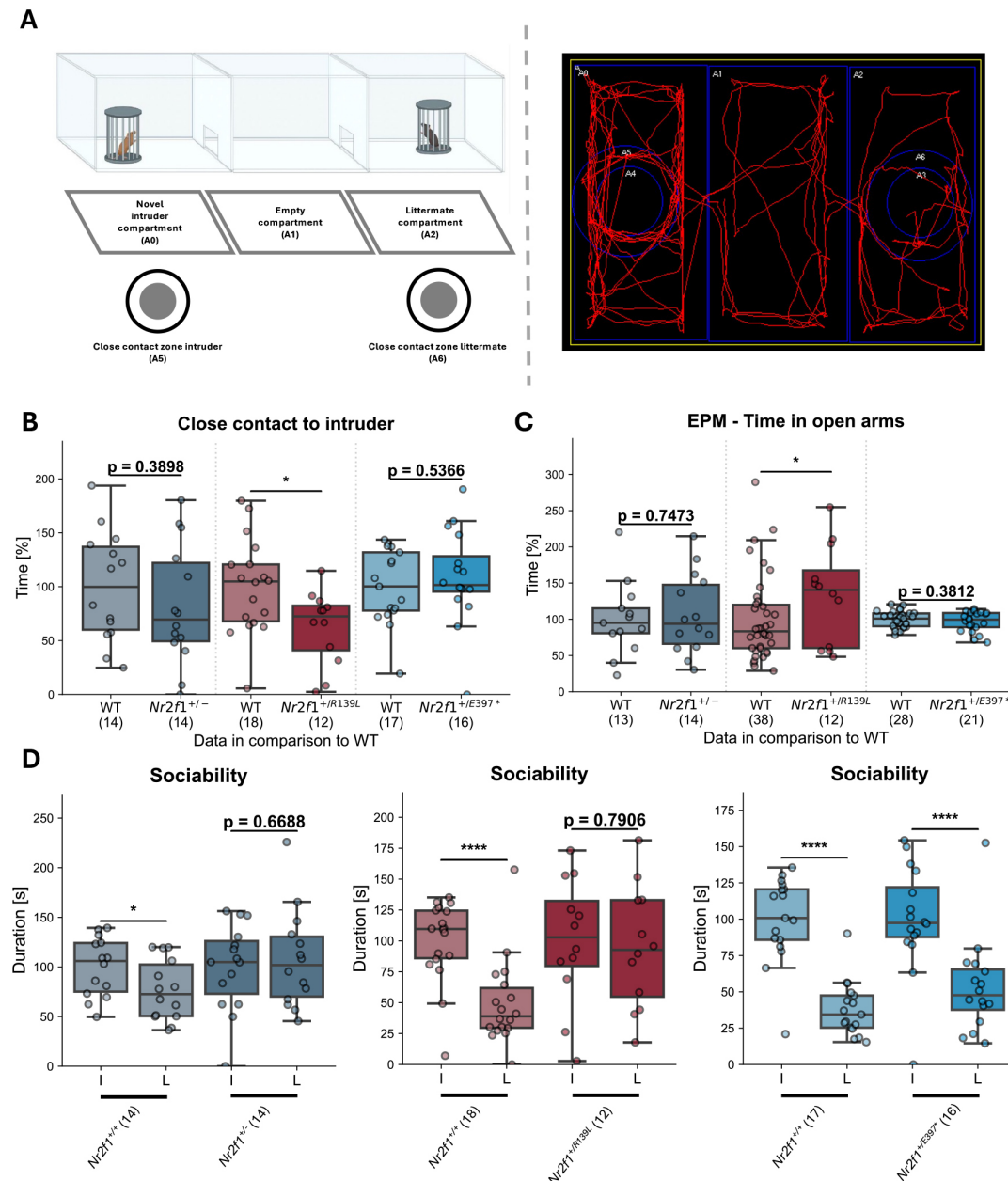


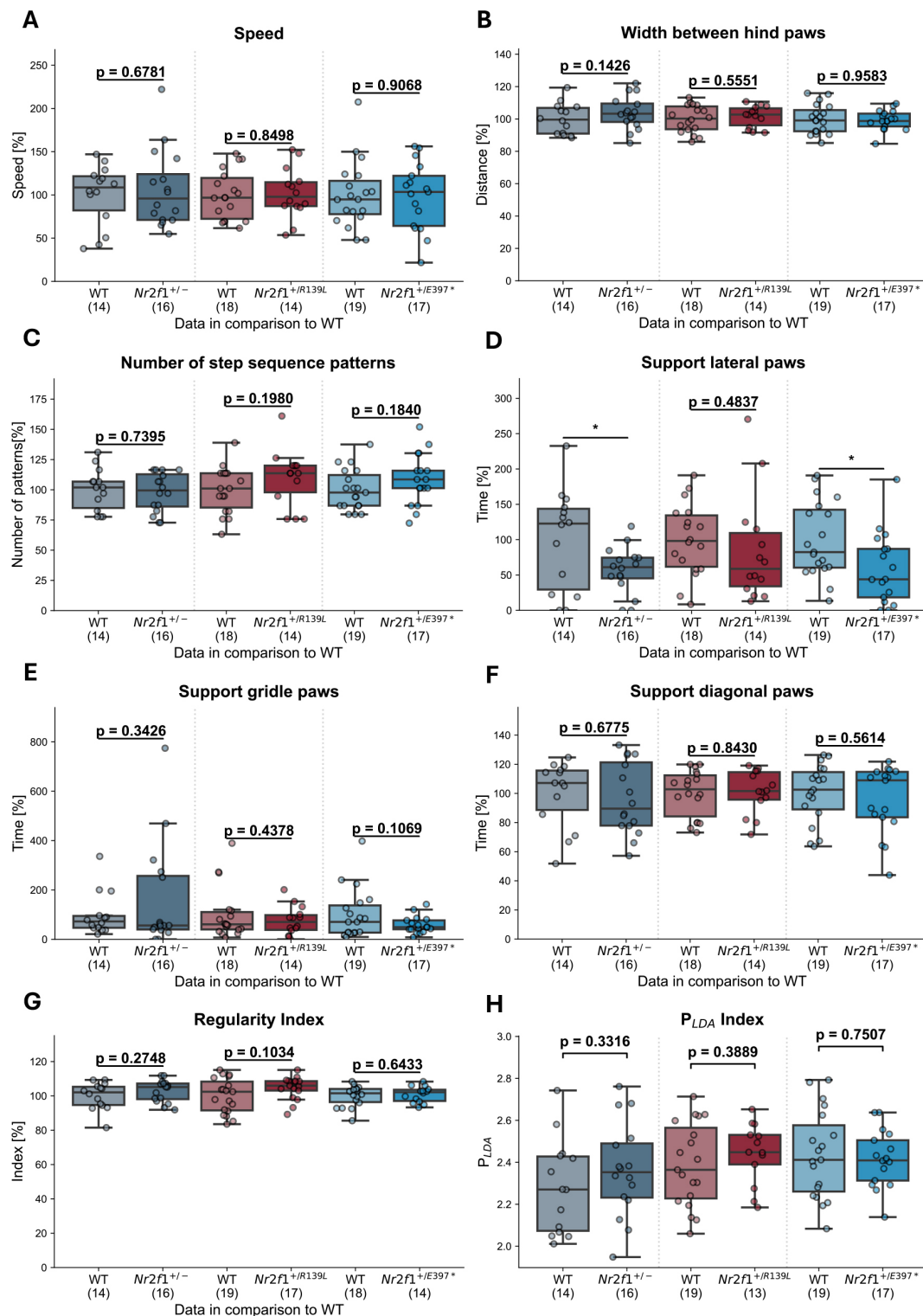
Fig. S3. Alterations in social interaction and elevated plus maze for mice with *Nr2f1* variants.

(A) An illustration of the SIT setup during trial 3 shows the different areas that were evaluated, as well as an example of a track map from a representative WT animal.

(B) *Nr2f1*^{+/-}/*R139L* mice spend less time in the close contact zone of the intruder as compared to their respective WT littermates. (C) The *Nr2f1*^{+/-}/*R139L* mice spend more time in the open arms of the EPM arena, as compared to their WT littermates. Data points for each line are shown, normalized to the respective WT mean. (D) The graphs compare the sociability of WT and mutant animals for each genetic line.

Typically, mice spend more time interacting with the intruder (social novelty) than with their littermates. This behavior is observed in all WT lines. However, animals carrying the R139L variant or a deletion in *Nr2f1* do not exhibit a preference for social novelty, suggesting impaired social interaction abilities. Labels: I - Intruder; L - Littermate.

Data points for each line are shown, normalized to the respective WT mean. *P ≤ 0.05, ****P ≤ 0.0001; Student's t-test.

**Fig. S4. Motorfunction analysis CatWalk XT.**

(A) The data shows no difference in body speed for all three mouse models. (B) The width between the hind paws is unchanged for the three mouse lines. (C) Mice do not show changes in the number of step sequence patterns in the CW. (D) $Nr2f1^{+/-}$ mice and $Nr2f1^{+/E397*}$ mice show a decrease in the lateral (e.g. left front, left hind) support. (E) The mice do not show a change in the girdle (e.g. left front, right front) support. (F) The mice do not show a change in the diagonal (e.g. left front, left hind) support. (G) The mice do not show a change in the regularity index, for overall limb coordination. (H) P_{LDA} Index shows no difference between the lines in overall motor performance. * $P \leq 0.05$; Student's t-test.

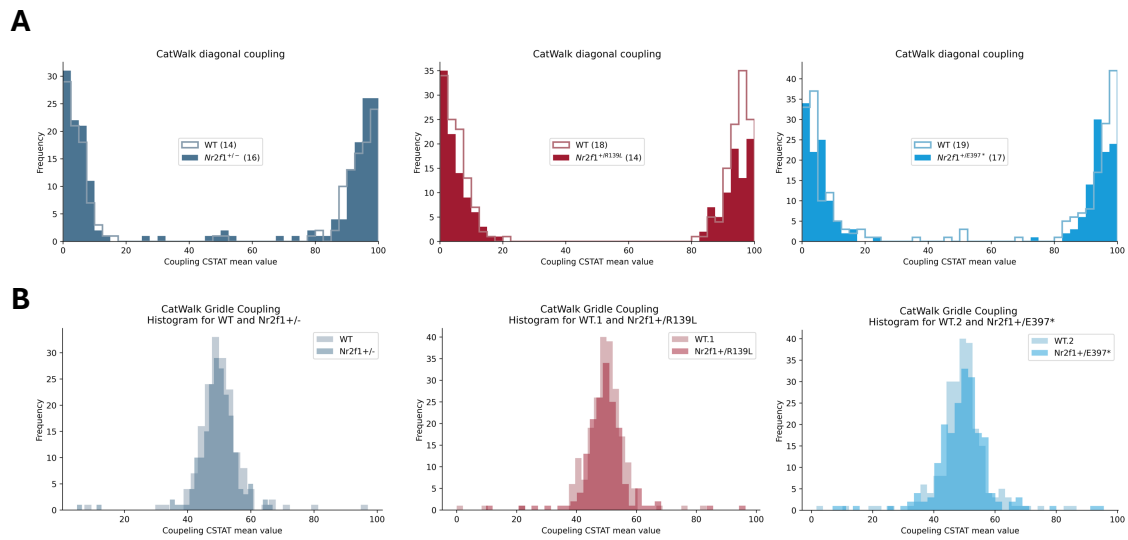


Fig. S5. Interpaw coordination analysis, CatWalk XT.

(A + B) Descriptive illustration of the overall diagonal, and gridle coupling, as measured in CSTAT mean. There are no apparent changes observable.

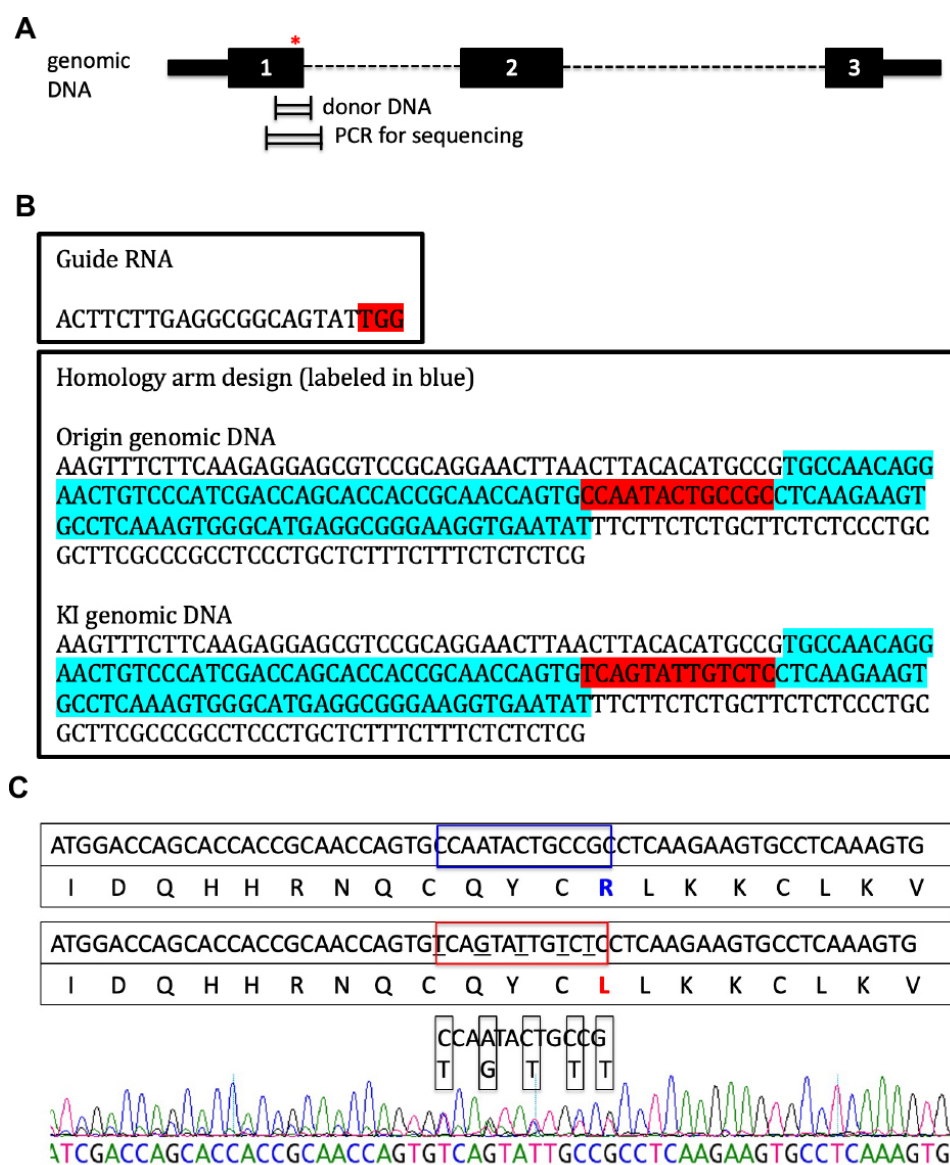


Fig. S6. Illustration of the creation of the Nr2f1^{+/-R139L} knockin mouse CRISPR/Cas9 design.

(A) Schematic representation of the Nr2f1 genomic locus showing exon 1–3 and the targeting site (red asterisk) within exon 1. The donor DNA and primers used for PCR and sequencing are indicated.

(B) Details of CRISPR/Cas9 targeting. The guide RNA sequence (black) and PAM site (red) are shown. The wild-type (origin) genomic sequence and knock-in (KI) genomic sequence are aligned, highlighting the homology arms (blue), the target codon (red), and the introduced mutation.

(C) Confirmation of the R139L mutation at the DNA level. The amino acid change from arginine (R) to leucine (L) is shown in the aligned translated sequences. Sanger sequencing chromatogram confirms the successful introduction of the c.416G>T mutation in the KI allele.

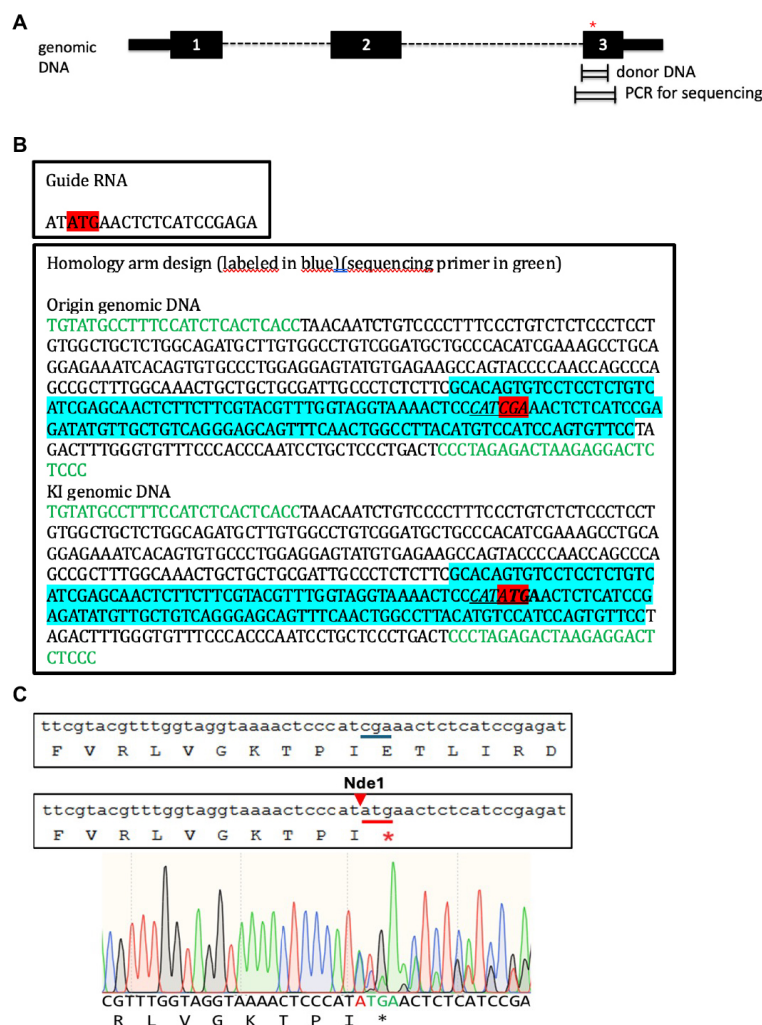


Fig. S7. Illustration of the creation of the Nr2f1^{+/E397*} knockin mouse CRISPR/Cas9 design.

(A) Schematic representation of the Nr2f1 genomic locus showing exon 1–3 and the targeting site (red asterisk) within exon 3. The donor DNA and primers used for PCR and sequencing are indicated.

(B) Details of CRISPR/Cas9 targeting. The guide RNA sequence (black) and PAM site (red) are shown. The wild-type (origin) genomic sequence and knock-in (KI) genomic sequence are aligned, highlighting the homology arms (blue), the target codon (red), and the introduced mutation.

(C) Confirmation of the E397* mutation at the DNA level. The amino acid change from glutamic acid (E) to a premature stopp codon (*) is shown in the aligned translated sequences. Sanger sequencing chromatogram confirms the successful introduction of the c.1189G>T mutation in the KI allele.

Table S1. Primer List.

For the KO line, the amplicon of the null allele is 36bp longer as it contains a loxP site. For the E397* genotyping, subsequently use NdeI to cut the amplicon into a 419 bp long WT band and the 280 bp and 139 bp long fragments of the E397* allele. For the R139L genotyping, there is only amplification in mutated animals as the primers are specific to the mutated sequence. For detailed genotyping protocols, reach out to the corresponding author of this paper

<i>Number</i>	<i>Name</i>	<i>Sequence</i>
1	genotyping-KO1	5'-CTGCTGTAGGAATCCTGTCTC-3'
2	genotyping-KO2	5'-AATCCTCCTCGGTGAGAGTGG-3'
3	genotyping-KO3	5'-ACATACACAGCCTGGCCTTGC-3'
4	Genotyping E397*-1	5'-TGTATGCCTTTCCATCTCACTCACC-3'
5	Genotyping E397*-2	5'-GGGAGAGTCCTCTTAGTCTCTAGGG-3'
6	Genotyping R139L-1	5'-AGCATCCTCACACACAAGCG-3'
7	Genotyping R139L-2	5'-CCGCAACCAGTGTCAGTATTG-3'
8	Nr2f1 exon3 F	5'-TCCCATCGAAACTCTCATCC-2
9	Nr2f1 exon3 F	5'-AGTGGGCTGCTCTTGTTCC-3'
10	Nr2f1 R139L WT F	5'-CAACCAGTGCCAATACTGCCGCC-3'
11	Nr2f1 R139L WT R	5'-CAGACAGGTAGCAGTGGCCA-3'
12	Nr2f1 R139L KI F	5'-CAACCAGTGTCAGTATTGTCT-3'
13	Nr2f1 R139L KI R	5'-TAGCCAGACAGGTAGCAGTG-3'
14	Nr2f1 E397* KI F	5'-AAGCCAGTACCCCAACCAG-3'
15	Nr2f1 E397* KI R F	5'-AACATATCTCGGATGAGAGTTCAT-2

Table S2. List of all p-values by Student's t-test for the histological and MRI analyses.

The table shows p-values for histological measures at postnatal days 7 and 28 and for MRI-based volumetric analyses in adult mice. Statistical comparisons were made between wild-type (WT) and mutant (R139L and E397*) groups using Student's t-test. Changes (increase or decrease) compared to WT are indicated where relevant. Outliers (values >3 standard deviations from the group mean) were excluded prior to analysis. For details, see Methods; Histological Analysis, MRI Acquisition and Data Analysis.

Measure	WT vs Null	WT vs R139L	WT vs E397*
Histology Nissl P7			
CC thickness before Hipp (μm)	0.8044	0.4360	0.5682
CC thickness after Hipp (μm)	0.1282	0.0585 (decrease)	0.2542
Thickness CC (μm)	0.2572	0.1431	0.4463
Moyenne Anterior hippocampus area (μm^2)	0.0829	0.0044 (decrease)	0.5639
Moyenne posterior hippocampus area (μm^2)	0.1420	0.1981	0.1920
Moyenne Hippocampus total area (mm^2)	0.0632	0.0829	0.3260
Motor cortex thickness (μm)	0.4118	0.7904	0.6028
Somatosensory Cortex Thickness (μm)	0.3073	0.2438	0.2757
Visual Cortex Thickness (μm)	0.3217	0.7385	0.0418 (decrease)
Right ventricle (area μm^2)	0.5393	0.1709	0.1133
Left ventricle (area μm^2)	0.2597	0.0854	0.1397
Total	0.4119	0.0937	0.1245
Right ventricle (perimeter μm)	0.8400	0.1110	0.4706
Left ventricle (perimeter μm)	0.5213	0.0639	0.5167
Ant hippocampale commissure (Area μm^2)	0.6681	0.0254 (decrease)	0.2705
Ant hippocampale commissure (Perimeter μm)	0.8004	0.8456	0.6760
Striatum	0.7167	0.1270	0.0252
Globus Pallidus	0.7495	0.2443	0.0202
Histology Nissl P28			
CC thickness before Hipp (μm)	0.0119 (decrease)	0.0301 (decrease)	0.0129 (decrease)
CC thickness after Hipp (μm)	0.5432	0.1465	0.0001 (decrease)
Total Thickness CC (μm)	0.0026 (decrease)	0.0328 (decrease)	0.0002 (decrease)
Moyenne Anterior hippocampus area (μm^2)	0.0811	0.0001 (decrease)	0.3328
Moyenne posterior hippocampus area (μm^2)	0.2270	0.0639	0.0069 (decrease)
Moyenne Hippocampus total area (mm^2)	0.1733	0.0064 (decrease)	0.0466 (decrease)
Motor cortex thickness (μm)	0.3694	0.9382	0.3605
Somatosensory Cortex Thickness (μm)	0.0838	0.4023	0.5209
Visual Cortex Thickness (μm)	0.6170	0.0009 (decrease)	0.7984
Right ventricle (area μm^2)	0.3074	0.0187 (increase)	0.0081 (increase)
Left ventricle (area μm^2)	0.3063	0.0172 (increase)	0.0235 (increase)
Total	0.2808	0.0092 (increase)	0.0002 (increase)
Right ventricle (perimeter μm)	–	0.4318	0.1681
Left ventricle (perimeter μm)	–	0.0972	0.2045
Ant hippocampale commissure (Area μm^2)	0.7853	0.0003 (decrease)	0.0824
Ant hippocampale commissure (Perimeter μm)	0.5418	0.6106	0.1326
Striatum	0.3975	0.0100	0.1868
Globus Pallidus	0.4633	0.8861	0.7359
MRI Adults			
Striatum	0.4564	0.2562	0.1420
Hypothalamus	0.3818	0.8412	0.8699
Thalamus	0.3006	0.0479 (decrease)	0.7566
Hippocampus	0.3737	0.0043 (decrease)	0.9471
Cerebellum	0.2320	0.3113	0.8942
Corpus Callosum	0.2493	0.3695	0.3623
Lateral Ventricles	0.9368	0.1780	0.0012 (increase)
Total brain volume	0.5475	0.0502	0.0524

Table S3. Analysis and comparison of detailed USV characteristics.

Different USV parameters by significance and impairment. “-” indicates no significant change as assessed by double-sided homoskedastic t-test. Arrows indicate a significant increase or a decrease of the respective parameter.

Nr2f1^{+/-}: Call Length (P4) P=0.0401, Slope (P4) P<0.0001 P8: Call Length (P8) P=0.0005, Frequency Delta (P8) P=0.0038, Sinuosity (P8) P=0.0033

Nr2f1^{+/*R139L*}: Latency to first (P4) P=0.0010, Time to last (P4) P=0.0010, Average call length (P4) P<0.0001, Highest frequency (P4) P=0.0166, Power (P4) P<0.0001, Tonality (P4) P<0.0001, Slope (P12) P=0.0137.

Nr2f1^{+/*E397**}: Average call length (P4) P<0.0001, Frequency delta (P4) P=0.0036, Slope (P4) P=0.0213, Sinuosity (P4) P=0.0087, Power (P4) P=0.0006, Tonality (P4) P=0.0011, Slope (P12) P=0.0184.

Metric	P4			P8			P12		
	<i>Nr2f1</i> ^{+/-}	<i>Nr2f1</i> ^{+/<i>R139L</i>}	<i>Nr2f1</i> ^{+/<i>E397*</i>}	<i>Nr2f1</i> ^{+/-}	<i>Nr2f1</i> ^{+/<i>R139L</i>}	<i>Nr2f1</i> ^{+/<i>E397*</i>}	<i>Nr2f1</i> ^{+/-}	<i>Nr2f1</i> ^{+/<i>R139L</i>}	<i>Nr2f1</i> ^{+/<i>E397*</i>}
Latency to first call	-	↑	-	-	-	-	-	-	-
Time of last call	-	↑	-	-	-	-	-	-	-
Average call duration	↓	↓	↓	↓	-	-	-	-	-
Principal Frequency (kHz)	-	-	-	-	-	-	-	-	-
Low Freq (kHz)	-	-	-	-	-	-	-	-	-
High Freq (kHz)	-	↓	-	-	-	-	-	-	-
Delta Freq (kHz)	-	-	↓	↓	-	-	-	-	-
Slope (kHz/s)	↓	-	↓	-	-	-	-	↓	↓
Sinuosity	-	-	↓	↓	-	-	-	-	-
Power (dB/Hz)	-	↓	↓	-	-	-	-	-	-
Tonality	-	↓	↓	-	-	-	-	-	-

Table S4. Overview of the number of significant results for each experiment for each mouse model.

USV (Ultrasonic Vocalization), EPM (Elevated Plus Maze), OF (Open Field), DLB (Dark Light Box), HB (Hole Board), SIT (Social Interaction Test), LB (Laboras), APA (Active Place Avoidance) and CW (CatWalk XT). Detailed USV characteristics (Table S3) not included.

<i>Experiment</i>	<i>Nr2f1</i> ^{+/-}	<i>Nr2f1</i> ^{+/<i>R139L</i>}	<i>Nr2f1</i> ^{+/<i>E397*</i>}
Weight (all timepoints)	0	5	0
USV P4	2	2	2
USV P8	1	0	0
USV P12	0	1	1
EPM	0	3	1
OF	0	2	0
DLB	0	0	0
HB	0	0	0
SIT	0	4	1
LB	0	1	2
APA	0	4	2
CW	0	2	8
Total Significant Results	3	24	17

Table S5. Comparison of the mouse models by effect size (Cohen's d). Several parameters were significant in more than one line. Adding to the quantitative comparison we evaluated the effect size using Cohen's d. For each parameter and mouse model (if the result was significant as tested by Student's t-test), we calculated d. On average, d was 0.6810, 1.1035, and 1.0054 for *Nr2f1*^{+/-}, *Nr2f1*^{+/*R139L*}, and *Nr2f1*^{+/*E397**} mice, respectively.

<i>Parameter</i>	<i>Nr2f1</i> ^{+/-}	<i>Nr2f1</i> ^{+/<i>R139L</i>}	<i>Nr2f1</i> ^{+/<i>E397*</i>}
Weight P4 Males	1.5715	3.0305	1.6667
USV P4 Number of Calls	0.5808	1.7893	1.6818
USV P4 Complex vs Simple Calls	0.3578	0.8456	0.6307
USV P12 Complex vs Simple Calls	ns	0.4685	0.4664
EPM Visits Open Arms	ns	0.7818	0.6155
APA Shocks During Training	ns	0.8890	1.4136
CW BOS Width Between Front Paws	0.4238	0.5844	0.7054
CW RF - LF Coupling	0.4712	0.4386	0.5632
Average effect size (d)	0.6810	1.1035	1.0054

Table S6. List of all p-values by Student's t-test for the behavioral tests.

USV (Ultrasonic Vocalization), EPM (Elevated Plus Maze), OF (Open Field), DLB (Dark Light Box), HB (Hole Board), SIT (Social Interaction Test), LB (Laboras), APA (Active Place Avoidance) and CW (CatWalk XT). The table shows the p values for the different tests and parameters as calculated by the Student's T-test. Outliers (> 3 standard deviations from the mean) were not included in the T-test. For details see Methods; Behavioral Data Analysis and Statistics.

<i>Parameter</i>	<i>Nr2f1^{+/-}</i>	<i>Nr2f1^{+/R139L}</i>	<i>Nr2f1^{+/E397*}</i>
Weight P4			
Males and Females	0.8104	0.0111	0.3407
Males	0.0215	0.0004	0.0162
Weight P8	0.9495	0.0224	0.4089
Weight P12	0.6128	0.0365	0.4776
Weight P20	0.5658	0.0176	0.3443
Weight P30	0.9922	0.6604	0.4286
Weight P100	0.2422	0.0336	0.0638
USV P4			
Number of Calls	0.0066	<0.0001	<0.0001
Complex vs Simple Calls	0.0166	<0.0001	0.0040
USV P8			
Number of Calls	0.4739	0.6568	0.8930
Complex vs Simple Calls	0.0395	0.2269	0.3239
USV P12			
Number of Calls	0.1316	0.9095	0.8475
Complex vs Simple Calls	0.3882	0.0124	0.0322
EPM			
Time in Closed	0.6753	0.2455	0.3812
Time in Middle	0.2075	0.3099	0.6628
Time in Open Arms	0.7473	0.0370	0.3812
Total Distance	0.5786	0.0060	0.3184
Visits Open Arms	0.5516	0.0223	0.0398
OF			
Time in Middle	0.4474	0.0374	0.8326
Total Distance	0.0678	0.0074	0.1629
DLB			
Delay to First Entry	0.5691	0.5371	0.1710
Total Distance	0.6516	0.1328	0.4154
Visits	0.3424	0.9984	0.1699
Time in Light Area	0.6647	0.5087	0.4077
Distance in Light Area	0.8302	0.6736	0.4971
HB			
Nose Pokes	0.9917	0.1316	0.0764
SIT Trial 2			
Time with Intruder	0.3257	0.7458	0.3356
Time with Littermate	0.6366	0.9562	0.1261

Continued on next page

Table 1 – fortgesetzt

<i>Parameter</i>	<i>Nr2f1</i> ^{+/-}	<i>Nr2f1</i> ^{+/<i>R139L</i>}	<i>Nr2f1</i> ^{+/<i>E397*</i>}
Time in Close Contact with Intruder	0.4382	0.2599	0.7604
Time in Close Contact with Littermate	0.2794	0.7362	0.4893
SIT Trial 3			
Time with Intruder	0.5051	0.0119	0.7215
Time with Littermate	0.3587	0.0196	0.3193
Time in Close Contact with Intruder	0.3515	0.0186	0.5331
Time in Close Contact with Littermate	0.9566	0.2411	0.0230
Sociability Index	0.5865	0.0178	0.7215
LB			
Climbing Frequency	0.5428	0.0468	0.5406
Climbing Duration	0.5529	0.1079	0.0373
Locomotion Duration	0.1138	0.8262	0.0358
APA			
Recall Trial; Time in Zone	0.9306	0.0315	0.6369
Recall Trial; Number of Shocks	0.7526	0.0371	0.3954
Training Trial; Time in Zone	0.3166	0.0470	0.0166
Training Trial; Number of Shocks	0.2386	0.0218	0.0004
Latency to First Shock During Recall	0.8263	0.1668	0.6369
CW			
Width Between Front Paws	0.0468	0.0055	0.0004
Couplings_RF→LH_CStat_Mean	0.9217	0.0392	0.0430
Couplings_LF→RH_CStat_Mean	0.2868	0.3978	0.0234
Couplings_RH→LF_CStat_Mean	0.2859	0.6792	0.0026
Couplings_LF→RF_CStat_Mean	0.4148	0.1293	0.0161
Couplings_RF→LF_CStat_Mean	0.1021	0.0596	0.0375
Couplings_LF→LH_CStat_Mean	0.4784	0.1404	0.0001
Couplings_LH→LF_CStat_Mean	0.3037	0.0962	0.0001

Table S7. Overview of cohort sizes and sex distribution across behavioral tests.

This table summarizes the number of male and female mice, as well as the number of litters, tested at each stage of the behavioral pipeline for each mouse line. The tests include: USV (Ultrasonic Vocalizations), EPM (Elevated Plus Maze), OF (Open Field), DLB (Dark-Light Box), HB (Hole Board), SIT (Social Interaction Test), LB (Laboras), APA (Active Place Avoidance), and CW (CatWalk XT). All animals had identical experimental histories at each timepoint of testing. No new animals were added to any cohort, and wild-type and mutant mice were always tested in parallel. Testing was performed in two cohorts, consisting of 66 and 142 animals, respectively. For details see Methods;Animals.

Test	Line	Genotype	Males	Females	Total	Litters
USV P4	KO	WT	25	23	48	13
		Nr2f1 ^{+/-}	24	20	44	
	R139L	WT	22	26	48	12
		Nr2f1 ^{+/R139L}	14	11	25	
	E397*	WT	18	8	26	9
		Nr2f1 ^{+/E397*}	10	7	17	
USV P8	KO	WT	24	22	46	13
		Nr2f1 ^{+/-}	19	22	41	
	R139L	WT	21	24	45	10
		Nr2f1 ^{+/R139L}	13	10	23	
	E397*	WT	18	8	26	9
		Nr2f1 ^{+/E397*}	10	7	17	
USV P12	KO	WT	19	20	39	13
		Nr2f1 ^{+/-}	21	19	40	
	R139L	WT	20	22	42	10
		Nr2f1 ^{+/R139L}	13	10	23	
	E397*	WT	18	8	26	9
		Nr2f1 ^{+/E397*}	11	5	16	
EPM	KO	WT	7	6	13	5
		Nr2f1 ^{+/-}	7	7	14	
	R139L	WT	18	20	38	8
		Nr2f1 ^{+/R139L}	8	4	12	
	E397*	WT	19	9	28	8
		Nr2f1 ^{+/E397*}	11	10	21	
OF	KO	WT	7	7	14	5
		Nr2f1 ^{+/-}	8	6	14	
	R139L	WT	13	12	25	8
		Nr2f1 ^{+/R139L}	8	6	14	
	E397*	WT	20	8	28	8
		Nr2f1 ^{+/E397*}	10	10	20	
DLB	KO	WT	7	7	14	5
		Nr2f1 ^{+/-}	8	6	14	
	R139L	WT	13	12	25	8
		Nr2f1 ^{+/R139L}	8	6	14	
	E397*	WT	18	8	26	8
		Nr2f1 ^{+/E397*}	10	6	16	

Test	Line	Genotype	Males	Females	Total	Litters
HB	KO	WT	7	7	14	5
		Nr2f1 ^{+/-}	8	7	15	
	R139L	WT	13	12	25	8
		Nr2f1 ^{+/R139L}	8	6	14	
	E397*	WT	15	8	26	8
		Nr2f1 ^{+/E397*}	10	5	16	
SIT	KO	WT	7	7	14	5
		Nr2f1 ^{+/-}	8	7	14	
	R139L	WT	9	9	18	5
		Nr2f1 ^{+/R139L}	6	6	12	
	E397*	WT	10	7	17	5
		Nr2f1 ^{+/E397*}	11	5	16	
CW	KO	WT	7	7	13	5
		Nr2f1 ^{+/-}	8	8	16	
	R139L	WT	9	9	18	5
		Nr2f1 ^{+/R139L}	8	6	14	
	E397*	WT	10	9	19	5
		Nr2f1 ^{+/E397*}	10	7	17	
APA	KO	WT	7	7	13	5
		Nr2f1 ^{+/-}	8	8	16	
	R139L	WT	9	8	17	5
		Nr2f1 ^{+/R139L}	8	6	14	
	E397*	WT	10	6	16	5
		Nr2f1 ^{+/E397*}	10	6	16	