BUILDING AN AUTISM SEVERITY CLASSIFICATION MODEL FOR NF1 MOUSE

MARIA BEATRIZ MOREIRA PINTO E SOUSA

A dissertation submitted in partial fulfillment of the requirements for the Degree of Masters in Biomedical Research

Dissertação para obtenção do grau de Mestre em Investigação Biomédica

at Faculdade de Ciências Médicas | NOVA Medical School of NOVA University Lisbon

Fevereiro, 2021

BUILDING AN AUTISM SEVERITY CLASSIFICATION MODEL FOR NF1 MOUSE

Maria Beatriz Moreira Pinto e Sousa

Supervisors: Joana Gonçalves, PhD of ICNAS/CIBIT;

João Martins, PhD of ICNAS/CIBIT;

Rosalina Fonseca, MD, PhD of CEDOC/NMS

A dissertation submitted in partial fulfillment of the requirements for the Degree of Masters in Biomedical Research

Dissertação para obtenção do grau de Mestre em Investigação Biomédica

Fevereiro, 2021

Acknowledgments

I want to thank everyone who supported and believed in me during the whole process of writing this thesis.

My sincere thank you to my supervisor Doctor Joana Gonçalves and Doctor João Martins from ICNAS, who showed me how to be a competent researcher, provided all the data used in the construction of the classification model, and were always available to answer my questions and help me overcome any challenge. Thank you as well to my internal supervisor Doctor Rosalina Fonseca for being available to give advice on my project. And thank you to Professor Miguel Castelo-Branco and everyone at ICNAS for welcoming me into their team and all the tips that greatly improved my project.

Lastly, I must thank Shailendra Tomar for the unwavering support and all the help with the more technical parts of building a classification model.

Resumo

A perturbação do espectro do autismo é uma perturbação do neurodesenvolvimento caracterizada por défices na socialização e comunicação, comportamentos repetitivos/estereotipias e, em alguns casos, capacidade cognitiva reduzida. A apresentação e a gravidade dos sintomas são bastante heterogéneas. Em humanos, o autismo é classificado em três níveis, dependendo da severidade dos sintomas: nível 1 (leve), nível 2 (moderado) e nível 3 (severo).

Vários modelos animais têm sido usados em estudos na área do autismo os quais apresentam mutações genéticas para recapitular o fenótipo da perturbação. Contudo, mesmo em animais com o mesmo fundo genético, o nível de severidade dos sintomas varia, resultando em heterogeneidade dos dados obtidos, o que é, geralmente, visto como indesejável. Por outro lado, esta heterogeneidade é semelhante ao que é observado em humanos, e poderá ser usado para uma melhor translação dos estudos. Ao classificar os animais em três níveis de severidade diferentes, novas informações podem ser obtidas sobre a Perturbação do Espectro do Autismo e o impacto de diferentes fatores na severidade dos sintomas, bem como o impacto da severidade na eficácia de diferentes terapias.

Este estudo apresenta um modelo de classificação de severidade de autismo, o qual foi construído usando um modelo de perturbação do espectro do autismo, o modelo animal de neurofibromatose tipo 1 (*Nf1*$^{+/-}$). Para isso, pontuações de diferentes medidas comportamentais para marcos de desenvolvimento em período neonatal bem como para comportamentos social e repetitivo, e aprendizagem/memória foram utilizadas para obter uma pontuação composta. Este modelo foi bastante preciso na identificação dos animais *Nf1*$^{+/-}$ com fenótipo semelhante ao autismo, distinguindo-o dos seus irmãos de ninhada sem mutação (murganhos wild-type). Esta classificação foi seguida por um modelo de multi-classificação, que posteriormente atribuiu um nível de severidade aos animais *Nf1*$^{+/-}$. Este modelo mostra menos precisão e terá de ser melhorado com mais dados. A distribuição dos níveis de gravidade dentro da amostra obtida pelo modelo de multi-classificação está próxima da distribuição da severidade de autismo na população de neurofibromatose tipo 1 humana com perturbação do espetro do autismo.

Palavras Chave: Perturbação do Espetro do Autismo; Neurofibromatose tipo 1; Modelo de Classifição; Modelo Animal de Pertubação do Espectro do Autismo; Comportamento animal

Abstract

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social behaviour and communication, repetitive behaviours/stereotypies, and, in some cases, cognitive disability. Due to heterogeneity of symptomatology and severity of ASD, clinically autism is classified in three levels: level 1 (mild), level 2 (moderate), level 3 (severe).

To investigate autism features, several animal models with genetic mutations were used to recapitulate autism core symptoms. However, even in animals with the same genetic background, symptom severity varies, resulting in heterogeneity of data obtained. This is often seen as undesirable. But, if we look at it as a representation of what is observed in humans, it could be used to improve translatability of autism research. By classifying animals into three different severity levels, new insights can be obtained about the ASD, and the impact of different factors on severity, as well as the impact of severity in effectiveness of different therapies.

Here, using a model of ASD, the mouse model of neurofibromatosis type 1 (*Nf1*$^{+/-}$), was built a model of autism severity classification. For this, scores from different behavioural measures for developmental milestones as well as dimensions of social, repetitive behaviour, and learning/memory were used to obtain a composite score. This model was highly accurate and precise at identifying *Nf1*$^{+/-}$ and distinguishing it from their littermates wild-type. This was followed by a multi-classification model that further attributed a severity level to *Nf1*$^{+/-}$ animals. This model show less accuracy and should be improved with more data. The distribution of severity levels within the sample given by the multi-classification model is close to the distribution of autism severity in the human neurofibromatosis type 1 population with ASD comorbidity.

Keywords: Autism Spectrum Disorder; Neurofibromatosis type I; Classification Model; Autism spectrum disorder Mouse Model; Animal Behaviour

Contents

List of Figures and Tables

List of Abbreviations and Acronyms

ASD – Autism Spectrum Disorder
DNVS - Single de novo variants
DSM-5 - Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
Fmr1 - Fragile X mental retardation 1
fMRI – functional magnetic resonance imaging
GAP - GTPase-activating protein
IQ - intelligence quotient
mTOR - mammalian target of rapamycin
NF1 – Neurofibromatosis type 1
NLGN4 – Neuroligin 4
NRXN - Neurexin
PECS - Picture Exchange Communication Systems
PND – Postnatal day
QAT – Quantitative autistic trait
TSC2 - Tuberous Sclerosis Complex 2
WT – Wild-type

Introduction

Autism Spectrum Disorders

Autism Spectrum Disorder (ASD), a group of neurodevelopmental disorders, occur due to changes or deviations from normal development caused by genetic and/or environmental factors, before birth and/or in early childhood, leading to abnormal neuronal connectivity or structure. These disorders are characterized by deficits in social behaviour and communication, repetitive behaviours, restricted interests, and, in some cases, intellectual disability[1] (Figure 1).

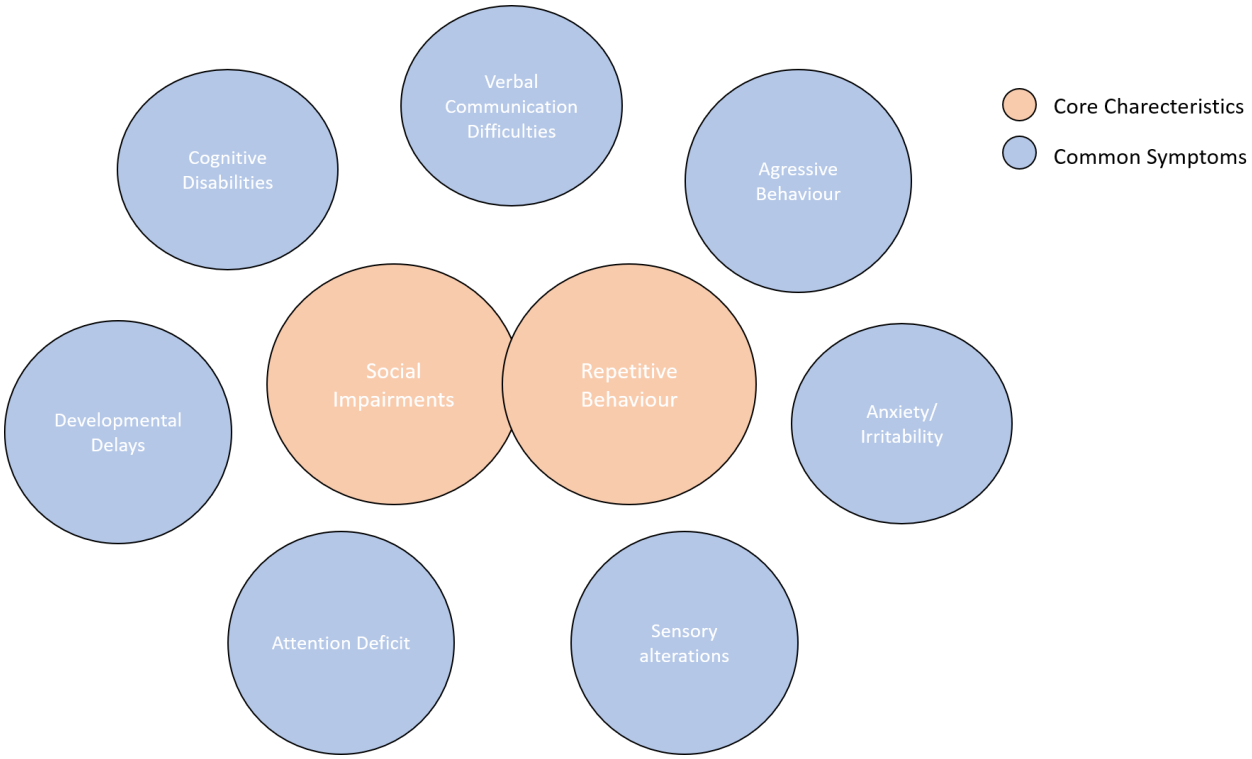

Figure 1 Core features of autism spectrum disorders and common associated symptoms

ASD is highly heterogenous and symptoms may present with different severity and frequency from individual to individual. The DSM-5 suggests that severity should be classified in three levels, depending on the level of assistance the patient needs:

- Level 1 (requiring support) – individuals have difficulties in social interaction and may be inflexible in some contexts (inflexibility of behaviour);
- Level 2 (requiring substantial support) – “marked deficits in verbal and nonverbal social communication” even when support is provided, inflexibility of behaviour, difficulty coping with change, and obvious repetitive behaviours;

- Level 3 (requiring very substantial support) – severe deficits in verbal and nonverbal communication, a person who speaks only a few words of intelligible speech, who rarely initiates social interaction.[1]

No single definitive aetiology has been established, as multiple genetic, epigenetic, and environmental factors seem to play a role in the development of ASD. Genetics play a major role in the development of these disorders. Early research in twins suggested ASD was heritable, while also under tight genetic control, postulating that multiple loci were involved in the development of autism.[2] In fact, multiple genomic studies have, since then, identified more than 100 ASD-risk genes and regions. In most instances, it is the interaction between single nucleotide changes that lead to ASD. The nucleotide changes are common in the population, present little individual risk and can be passed from parents to offspring. However rare, single de novo variants (DNVS), such as SHANK2, NRXN1, NLGN4, and NF1 have also been associated with ASD. These DNVS are sufficient to considerably increase the risk of autism.[3,4] Autism risk genes may affect several aspects of brain function at different times across the development process, making it harder to fully understand the mechanism behind the disorder. However, studies have shown that most of these genes encode proteins involved in synaptic structure and function, chromatin modification, regulation of gene expression, and development of excitatory and inhibitory neurons.[5–7]

Environmental factors also seem to be important contributors to ASD. Many of these are related to the mother, such as substance abuse, maternal nutrition, hormonal balance, stress, exposure to chemicals and air pollutants, and inflammatory events during pregnancy. [8,9] A recent study by Hegarty *et al. (2020)* performed on monozygotic and dizygotic twins with and without ASD identified some structures that are more influenced by genetic background and other which seem to be more influenced by environmental factors. In fact, cortical thickness, cerebellar white matter and ventricular volume seem to be considerably influenced by environmental factors.[10] Additionally, monozygotic twins exhibit different symptom severity scores, despite 96% concordance for ASD, highlighting the contribution of non-shared environmental factors in early infancy for the outcome of the disorder.[11]

Due to the high heterogeneity of factors and resulting phenotypes associated with ASD, research on possible mechanisms, treatments and therapies often faces challenges concerning translatability.

Neurofibromatosis Type 1

Neurofibromatosis type 1 (NF1) is a monogenic autosomal dominant disorder caused by mutations, inherited or spontaneous, in the *Nf1* gene resulting in the loss of function of neurofibromin. This protein is expressed in early melanocyte development and melanogenesis[12] , is necessary for the correct development of neurons and glia[13,14], and hematopoietic development[15].

Neurofibromin is a tumour suppressor protein that regulates the activity of RAS guanosine triphosphatase (GTPase-activating protein, GAP). Loss of neurofibromin function leads to excessive cell growth and survival due to the hyperactivation of RAS, which promotes cellular growth through the mTOR and MEK-ERK pathways (Figure 2).[16]

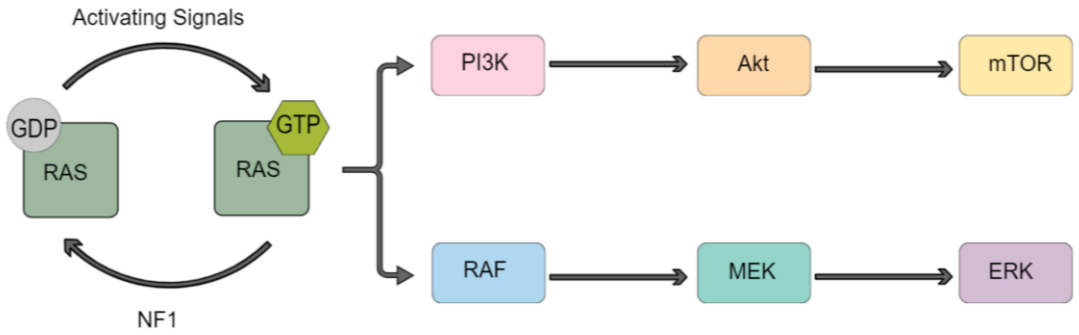

Figure 2 NF1 regulates cell growth and survival through mTOR and MEK-ERK pathways. Loss of neurofibromin leads to hyperactivation of RAS, which activates phosphoinositol 3'kinase (PI3K), Akt and mTOR. Loss of neurofibromin and RAS activation also leads to hyperactivation of the MEK-ERK pathway.

NF1 is characterized mainly by the occurrence of neurofibromas, pigmentary abnormalities ("café-au-lait" spots), and skeletal deformities. However, there is a great level of variability in symptom presentation, even between patients with identical germline mutations. Some individuals may also present symptoms of NF1 in only one area of the body. This is known as segmental or mosaic NF1.[17]

Deficits in cognition and behaviour are also present in some patients, such as learning disabilities, impaired spatial perception, attention deficits, social and communication difficulties, and motor delays.[18–20] Consequently, multiple studies have identified an increased risk of ASD in NF1 patients, with approximately 39.2% NF1 patients scoring above-threshold in quantitative autistic trait (QAT) scores.[21]

Several authors have reported that the high risk of ASD associated with NF1 could be caused by an excitation/inhibition (E/I) imbalance, characterized by a desequilibrium

between glutamatergic and GABAergic inputs. This dysfunction has been reported in NF1 patients in the parieto-occipital cortex, midbrain, and thalamus with a decrease in GABA(A) receptors binding.[22] Learning impairment in NF1 patients has also been associated with deficient modulation of GABA mediated intracortical inhibition[23]. Accordingly, studies performed in the *Nf1*$^{+/-}$ mouse model have confirmed the disruptions in GABA/glutamate ratio and GABA(A) receptor distribution and levels. This work also found that these alterations occurred in a region-dependent manner [24]. Moreover, it was showed a link between abnormal GABA and glutamate neurotransmission in the amygdala, and social deficits. These deficits were rescued by blocking of Pak1 in the amygdala.[25]

Due to these characteristics, the *Nf1*$^{+/-}$ mouse model is valuable to study ASD, especially its social and cognitive dimensions, and the impact of E/I imbalance in autistic-like behaviour.

Animal Models in Autism Research

Animal models are commonly used in research to study ASD. Several mouse models have been created using genetic alterations associated with ASD such as NLGN4, Fmr1, NF1, TSC2. Others are generated through the re-creation of environmental factors related with high risk of ASD, such as maternal immune activation and prenatal exposure to valproic acid. These models recapitulate not only the molecular and structural aspects of the disorder, but it's behavioural phenotype, allowing researchers to study all aspects of the disorder. Further, animals models allow to understand how different factors and possible therapies might impact on behaviour[26], using a battery of behavioural and neurodevelopmental tests. These tests evaluate different behaviours and cognitive function relevant to the different dimensions of autism.

Understanding the developmental process in pups is crucial to understand how ASD might impact development in infancy, as it is a neurodevelopmental disorder. For this purpose, multiple tests can be used at different stages of the pups' development to analyse milestones such as motor ability and coordination, sensorimotor development, and early communication. Methods for assessing these developmental milestones have been thoroughly describe in the literature.[27] Behavioural tests performed in juvenile and adult mice are used to describe how ASD affects social behaviour, memory and learning and compare findings in animal models with human ASD population features.

Social behaviour can be assessed in a variety of tests that evaluate different aspects of social interactions, from sociability to social memory, vocalizations and dominance. The three-chamber test, for instance, is routinely used to assess sociability and social

memory in ASD mouse models, allowing researchers to better understand which components of social behaviour differ from wild-type (WT) control animals.[28]

Another core feature of autism is the repetitive behaviour. In mouse models, this can be studied using tests such as marble burying test, where the number of marbles buried by the animal in a certain period is used as a measure of repetitiveness. Another behaviour associated with repetitiveness is grooming, which can be observed during the open field test, spontaneously.[29]

The Morris Water Mazer is frequently used to evaluate spatial learning and memory, while other motor tasks may also be used to assess other dimensions of memory or learning.[30]

Animal models are also useful for more invasive studies that could not be performed in patients, or even for some longitudinal studies, as they have a shorter life cycle than humans. They are indispensable in the search for the mechanisms behind ASD and possible therapies, as well as for drug development.

However, similar to humans, ASD mice may show interindividual variability in symptom severity, even in groups with the same genetic mutation.[31] This can make the process of obtaining statistically significant results strenuous and have a negative impact on translatability of the results from animal studies to the clinic. However, when properly addressed, this variability may offer deeper insights into the disorder, its mechanisms, and the best ways to help patients with different symptom severity.

Statistical models and their importance in biomedical research and in clinic

Mathematical and statistical models are used in most areas of research, from economy to physics, chemistry and biomedicine, due to their capability of describing complex events or predicting outcomes from large amounts of data.

With the rapid technological development, Data Science has become an almost ubiquitous discipline, providing tools to improve fundamental research as well as clinical practices, by automating and speeding up complex analysis and decision processes.

Machine learning - algorithms that can become more accurate through experience - is gaining popularity in the clinic and multiple algorithms have been developed with the goal to help identify risk factors for multiple diseases, diagnose, predict clinical outcomes.

For ASD, machine learning algorithms have shown to be able diagnose ASD based on characteristic autistic behaviours[32], fMRI data[33] that can even be combined with personal data such as full-scale intelligence quotient (IQ), followed by verbal IQ and performance IQ[34]. By allowing the analysis of home videos and short questionnaires that can be filled

out by the parents, these algorithms allow for easier screening and facilitate early diagnosis.[35]

Further, machine learning has also been used to improve the quality of life and therapies for ASD patients. For instance, machine learning improves robot-assisted therapy for emotion expression by improving robot perception of affect and engagement in children with ASD[36]. The possibility of using algorithms to help neurotypical individuals better understand ASD patients has also been explored by Tang et al. (2018), using sensors to predict emotion from facial expression and body movement.[37] Picture Exchange Communication Systems (PECS) – which help ASD patients communicate by using picture cards to express their needs - can also be refined by using data such as location and event sensing to offer context-sensitive picture cards.[38]

In research, computational approaches may help tackle the high heterogeneity of ASD by analysing large-scale datasets and identifying risk factors, guiding future research.[39]

Another approach involves searching for biomarkers that can help early detection of the disorder. Artoni et al. (2020) trained a neural network to identify abnormalities in cholinergic modulation in idiopathic and genetic mouse models of ASD that was then successfully applied to human data, showing that this type of approach can quickly provide reliable and translational biomarkers.[40]

ASD composite score

El-Kordi et al. (2013) developed an autism severity score for mice using the Nlgn4 null mutant model. This score was obtained by combining readouts from several behavioural tests for three dimensions of ASD: social interaction, communication, and repetitive behaviour/stereotypies. The severity score was able to predict the correct genotype between WT and *Nlgn4*$^{-/-}$ mice with an accuracy of almost 100%.[31] Different severities in animals with the same genetic background, housed in the same conditions were reported as well, confirming that even with similar conditions and identical genetic background, symptom severity may vary. This study used only behaviour tests performed in 11-12-week-old mice, without any neurodevelopmental read-outs, which could be important contributors in the severity of ASD, as it is a neurodevelopmental disorder.

In present work, a classification system was built to further characterize the *Nf1*$^{+/-}$, an ASD mouse model. A composite severity score was calculated using read-outs from behavioural tests performed in a battery of neurodevelopmental milestones tests to assess the contribution of early development to symptom severity, together with 8-week-

old mice for the social, repetitive behaviour and cognitive dimensions of ASD. This score distinguished between *Nf1*$^{+/-}$ and WT genotypes with high accuracy and precision. The multi-classification model was then built to classify subjects into level 1 (mild), level 2 (moderate), or level 3 (severe) ASD-like symptoms. Due to the small sample size and increased level of complexity, the accuracy of this model was considerably lower than the first. However, accuracy should improve as more data is included in the model. It is also worth noting that the distribution of severity obtained by the multi-classification model is similar to what is observed in the NF1 human population with ASD.

Hypothesis and Aims

Many studies have been dedicated to furthering the knowledge on Autism Spectrum Disorder, including its mechanisms and aetiology, molecular pathways, phenotype characteristics, risk factors and possible therapies. Mice models are commonly used in this field, as they allow for more complete and invasive studies to be developed. However, similar to what is observed in the human population, there is a high variability in symptom severity and presentation. This heterogeneity hinders translatability of results obtained through animal research to the clinic. Variability is generally compensated for by using more animals, rather than by a finer division of subjects into groups according to phenotype. This means that the effect of symptom severity and presentation are often masked, and valuable information on this core characteristic of ASD is not achieved from these studies.

The present study aimed to improve the quality of data and information obtained from animal models of ASD by building a severity classification model. For that, we will used *Nf1*$^{+/-}$ mouse as a ASD animal model. Here, we intended to provide a useful tool to classify subjects according to autism severity rather than just genotype.

In this work, three specific objectives are identified:

1 - Explore the impact of NF1 depletion in development milestones, behaviour and cognitive processes;

2 - Build a severity score using behavioural and neurodevelopmental measures to distinguish WT from *Nf1*$^{+/-}$ genotypes;

3 - Categorize symptom severity in three different levels similar to those used to classify ASD severity in humans.

Exploratory analysis was performed on data from developmental milestones and adult behavioural tests to fully characterize the *Nf1*$^{+/-}$mouse model (Aim 1). This data was also used to build a ASD severity composite score. For that, single read-outs from each test were z-standardized and combined (by averaging z-scores) to calculate a single composite score. This score was then used to identify *Nf1*$^{+/-}$ and WT genotypes with high accuracy and precision (Aim 2). Finally, a multi-classification model was developed to classify animals identified as *Nf1*$^{+/-}$ by the composite score into three different severity levels. As NF1 has been described to significantly affect social behaviour, high scores from social tests were used as a required condition for the highest level of severity (level 3). Parameters for each level were proposed based on the distribution of ASD severity in the human NF1 + ASD population (Aim 3).

Methods

Animals: Data from developmental milestones and behavioural tests performed in 29 mice [15 *Nf1*$^{+/-}$ and 14 WT littermates (controls)] was used in the construction of the classification model. Developmental milestones were obtained in pups from postnatal day (PND) 6 to PND 14. Behaviour tests were performed in 8-weeks-old male mice. To keep the genetic background of the *Nf1*$^{+/-}$ mice constant across experiments and studies, mice were backcrossed to Taconic C57Bl/6 mice at least 10 times and bred once with 129T2/SyEmsJ before experiments. Animals were group housed (2-4) on a 12h light/dark cycle at ICNAS animal facility. All experiments were carried out in accordance with the European Union Council Directive (2010/63/EU), the National Regulations, and the Internal Review Board of the University of Coimbra. All animals were healthy (discomfort score 0), and all efforts were made to minimize the number of animals used and their suffering.

Developmental milestones and behavioural tests: Data from NF1 mice developmental milestones and behavioural tests was analysed and scores were calculated for each measure. Importantly, experimenters were blinded to genotype during the experiments and analysis. The tests were used to characterize four different dimensions of ASD.

Developmental milestones: A battery of developmental milestones was analysed on pups on PND 6,8,10,12 and 14, following the same task sequence at every time point. Pups were returned to their home cage once the battery of tests was finalized. For the calculation of the composite score, all the read-outs from each day were averaged to create one score per test.

1) Surface righting reflex: This test was used to evaluate the motor ability of pup to return to a four-limb position from a supine position. The pup was placed on its back on a flat surface and held in position for 5 seconds (Figure 3). The time the pup took to flip onto its feet was recorded, until a maximum time of 30 seconds .[41]

Figure 3 Surface Righting Test. Pups are held on their backs for 5 seconds. Time to return to a supine position is record. Image by Ferreira, H. (2020)[42]

2) Negative geotaxis reflex: Motor coordination was assessed with this test. Here, each pup was placed facing down on a 35º incline, covered with fabric to facilitate traction (Figure 4). The time the animal took to rotate and face the top of the incline was recorded, with a cut-off of 30 seconds. Mice that fell off or rolled down the platform were re-tested to a maximum of 2 extra tries, after which a time of 30 seconds was attributed.[41]

Figure 4 Negative Geotaxis Test. Pups were placed facing down on a 35º incline. Time to rotate and face the top of the incline was recorded.Image by Ferreira, H. (2020)[42]

3) Locomotion: This test was used to assess extinction of rotary movement. A 13 cm diameter circle was drawn on a flat surface and the animal placed at the centre of it (Figure 5). Time to exit the circle with both forepaws was recorded to a maximum time of 30 seconds. Failure to exit the circle earned the score of 30 seconds.[41]

Figure 5 Locomotion test. Pups were place in the middle of a 13 cm diameter circle. Time to leave the circle with two forepaws was measured.Image by Ferreira, H. (2020)[42]

4) Nest seeking: Sensorimotor development was assessed with the nest seeking test. Here, a rectangular plastic arena (25cm x 10cm) divided into 3 compartments was used. Bedding from home cage was placed on the left compartment and similar amount of fresh bedding was placed on the right compartment (Figure 6). Lines were drawn in each compartment 6,5 cm from the centre to mark each zone. The pup was placed in the center and allowed to explore. Two trials were performed, with a 30 second intertrial interval, during which the operator held the animal. In each trial, time to cross home bedding line with both forepaws was measured, up to a cut-off time of 120 seconds. Final score was calculated by averaging the scores attributed in each trial. Pups were positioned facing opposite sides of the arena, in each trial, to rule out possible head turning preferences.[41]

Figure 6 Nest Seeking test. Bedding from home cage was placed on the left compartment of an arena divided in three. Fresh bedding was placed on the right compartment. Pups were placed in the middle of the arena and time to find the nest was measured. Image by Ferreira, H. (2020)[42]

Adult Behaviour: A battery of tests for repetitive and social behaviours and learning/memory skill was performed on 8-week-old male mice.

1) Marble burying test:

This test was used to assess repetitive and obsessive behaviour. Mice were placed, individually, in a standard mouse cage with a 5 cm layer of sawdust. Fifteen marbles were placed equidistantly on the surface of the sawdust, with a 3x5 disposition, as illustrated in Figure 7. To prevent animals from escaping the arena, the walls were extended with acrylic sheets. Mice were left to explore the arena for 30 min, and the number of buried marbles was recorded at 5, 10, 15, 20, 25 and 30 minutes.

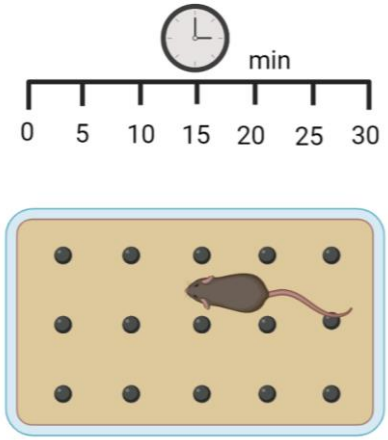

Figure 7 Representation of the Marble Burying Test. Number of buried marbles was recorded at 0,5,10,15,20,25 and 30 minute time points. Image created with BioRender.com

A score was calculated with the following formula:

$$\text{Index= Avg}\left(\frac{number\ marbles\ buried\ at\ each\ time\ point}{Total\ marbles}\right)$$

2) Morris Water Maze: To evaluate learning and memory, data from the Morris water maze was used. Mice were placed in a round pool filled with tap water at 23ºC. The escape platform was exposed 0.5 cm above the water during training. Water was made opaque with non-toxic white paint. Animals were tested for three trials per day for a period of 4 days. Before the beginning of the test, animals were allowed to swim freely for 30 seconds and placed on the escape platform for 30 seconds. During testing, the platform was placed in north-west quadrant and submerged (Figure 8). Mice were placed in the pool starting at the edge of one of the four quadrants and allowed to swim for 60 seconds. If the animal did not reach the platform within the 60 second period, it was guided toward it and held for 15 seconds on the platform. This procedure was followed for three trials,

starting at a different position each time. At the end of each trial block, the mouse was dried and placed back on its home cage (kept warm with a heating pad) for 30 to 40 minutes. Swim time and path length were recorded using a tracking software. After the final session, the platform as removed for a probe trial to test spatial strategy and retention. Mice were allowed to swim for 60 seconds.

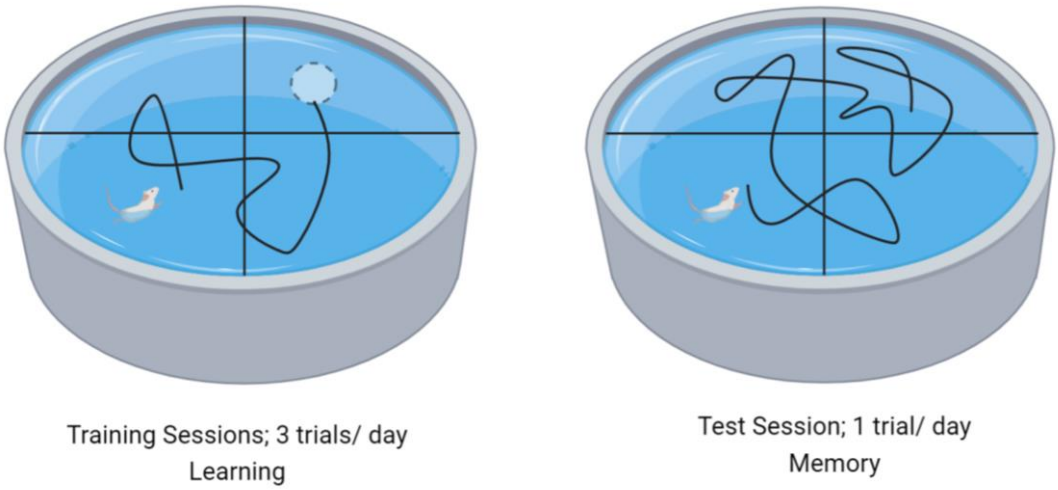

Figure 8 Representation of the Morris Water Maze Test for Learning and Memory. Trials have a maximum duration of 60 seconds. Image created with BioRender.com

The percentage of time spent in the quadrant where the platform was previously located was recorded.[43] The following scores were obtained:

Increase in Learning on day x+1 (dx+1): $Avg(\frac{L(dx+1)-L\,(dx)}{L(dx)}$x100,

where L corresponds to Learning: L= $\frac{Time\ in\ platform\ zone}{60}$

Increase in Memory on day x+1 (dx+1): $Avg(\frac{M(dx+1)-M\,(dx)}{M(dx)}$x100,

where M corresponds to Memory: M= $\left(\frac{Time\ Plat+Quadrant\ 4}{60}\right)$

3) Open Field: Grooming and rearing were obtained as measures of repetitive and obsessive behaviour, during open field task. Each animal was placed in an open polycarbonate arena (50 x 50 cm, 40 cm high) for ten minutes.[41] Time spent grooming and rearing was recorded. For analysis purposes, grooming was defined as a sequence that began with licking the paws, proceeding to the nose using elliptical brushing

movements, followed by unilateral strokes to clean vibrissae and eyes, bilateral strokes beginning behind the ears and over the face, and ending with licking the rest of the body and cleaning of the tail (Figure 9). In addition, rearing was considered as moments when the mouse lifted its forelimbs off the ground, putting the weight on its hind paws, lifting his head upwards. Average time spent on the activity was used as the scores for the ASD Severity composite score.

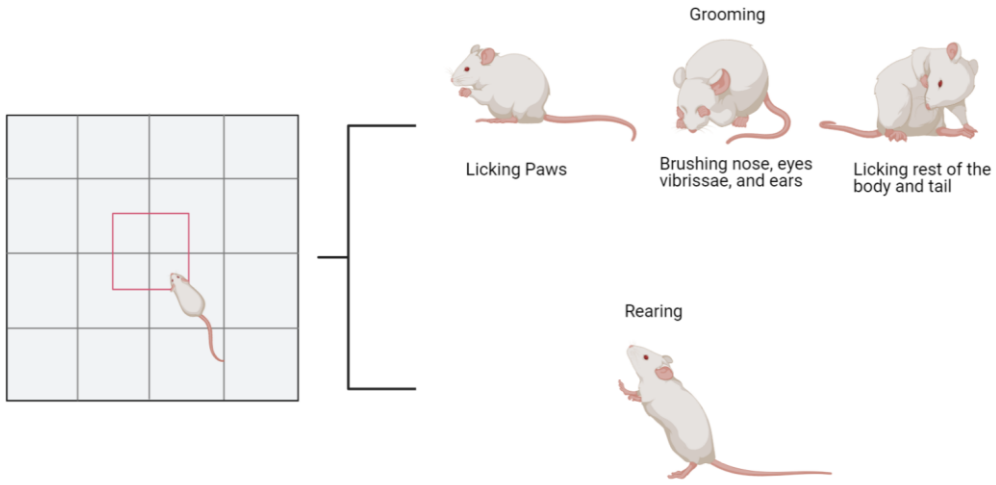

Figure 9 Representation of Grooming and Rearing behaviours observed during the Open Field test. Image created with BioRender.com

4) Three Chamber Social Test: To evaluate social behavior and the deficits in social interactions, the three-chamber test was used (Figure 10). This test measures the "sociability", the propensity to spend time with another mouse, as compared to time spent alone in an identical but empty chamber, and "social novelty", the propensity to spend time with a previously never-before-met mouse rather than with a familiar mouse. A three compartments' box was used (Panlab Light Grey Methacrylate floor, transparent walls, 60 x 42 x 22 (H) cm,). On day one, after adaption to the three-chamber arena, the subject mouse is placed into the middle chamber and allowed to explore freely for 10 minutes. In the left chamber of the test apparatus, a stimulus mouse is placed under a wire cage. In the right chamber, a similar wire cage is located without the stimulus mouse. To quantify the sociability of the experimental mouse, the time it spends in each chamber and the time spent sniffing at each wire cage are measured. On the following day, a novel mouse is placed under the wire cage in the empty right chamber, while the already known mouse stays in the left chamber. Time spent exploring each cage was recorded. The arena was cleaned with alcohol 70% between each animal. The position of the cups changed between each mouse. The ovariectomised female was habituated to the cup for 15 minutes per day, during 3 days before the tests.

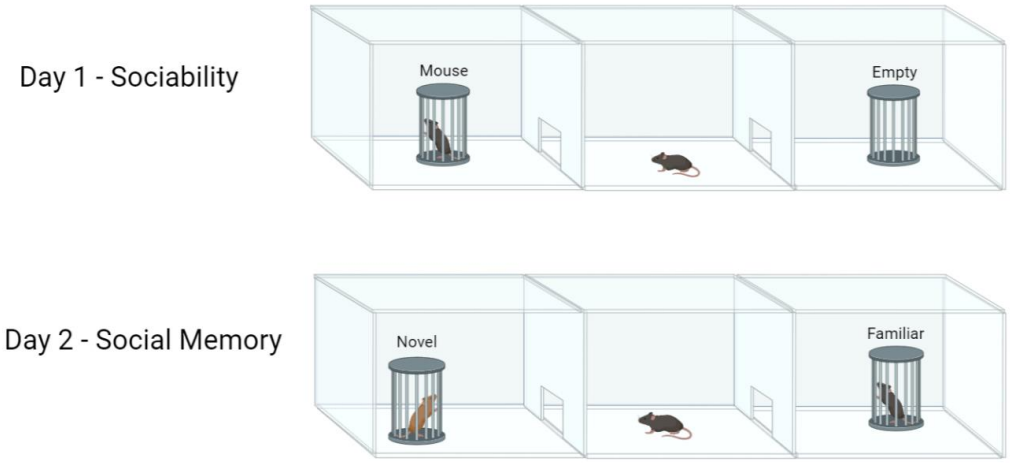

Figure 10 Representation of the Three-chamber social test. Mice were allowed to explore for 10 minutes. The position of the cup was changed between each mouse and the arena was cleaned with alchool 70% between each trial. Image created with BioRender.com

Sociability Index and Social Memory were calculated as:

$$\text{Sociability Index} = \frac{Time\ with\ stranger}{time\ stranger + time\ empty} \times 100$$

$$\text{Social Memory} = \frac{Time\ with\ new}{Time\ new + time\ familiar} \times 100$$

Statistical Analysis: Statistical tests and graphs were obtained using GraphPad Prism version 9.0.0 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com. Data was compared by 2-way ANOVA with post hoc planned comparisons and Mann-Whitney U tests. A p value below 0.05 was considered significant. Figures are expressed as mean ± SEM, unless stated otherwise**.**

Composite Score: The autism composite score was calculated by combining the scores of all behaviour measures, or only the ones from statistically significant measures (for comparison purposes). These scores were z-standardized and the composite score was the mean of all z-standardized scores. Scores from readouts where WT group were originally higher were inverted, to ensure that NF1 group would correspond to higher score. Missing values were replaced by the average of the group (WT or NF1). One *Nf1*$^{+/-}$ and one WT mouse had to be excluded from the composite score due to multiple missing neurodevelopmental and adult behaviour readouts. Therefore, 27 animals were used to build the composite score. Redundant readouts were identified through

correlation analysis and removed. Internal consistency of each composite score scale was assessed using Cronbach's α and McDonald's ω. Scores over >0.6 were considered acceptable (Kontig et al, 2009). Correlation and internal consistency analysis were performed using JASP Software (JASP Team (2020). JASP (Version 0.14.1). Scores were compared to genotype in order to assess whether the composite score was a good predictor of genotype and could be used to distinguish between *Nf1*$^{+/-}$ and WT.

Classification Model: To build autism severity classification model, the composite score scale that better predicted genotype was used. Thresholds for levels 1, 2 and 3 (mild, moderate, and severe, respectively) were proposed based on composite score and social behaviour readouts. An independent observer rated each animal on a scale of level 1 (healthy of mild) to level 3 (severe impairment) based on the videos from the social behaviour readouts. These classifications were then used as a comparison to evaluate the model's performance.

Results

Development Milestones

1) Locomotion: Although there were no significant differences between NF1 and WT groups on each of the testing days (Figure 11a), overall, NF1 mice spent more time to exit the circle in the locomotion test (p=0.005, NF1>WT) (Figure 11b).

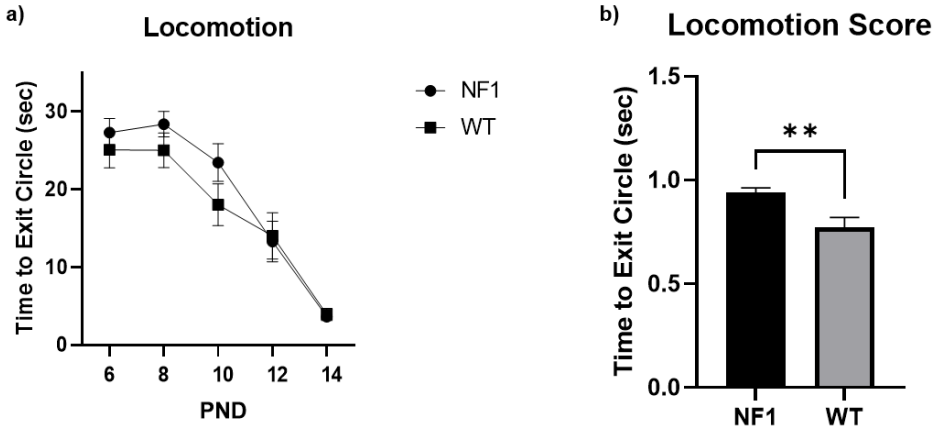

Figure 11 Neurodevelopmental milestones locomotion test showed a) no daily differences in time to exit the circle between NF1 and WT groups. However b) NF1 mice spent more time to exit the circle that WT mice. Two-way ANOVA with post hoc planned multiple comparisons (a) and one-tailed Mann-Whitney U test (b); n=28

2) Nest Seeking: In agreement with locomotion test, no significant differences were detected on each separate testing day between NF1 and WT mice in the nest seeking test (Figure 12a). However, overall, the NF1+/− group was faster than the WT group to reach the nest (p=0.007, NF1<WT) (Figure 12b).

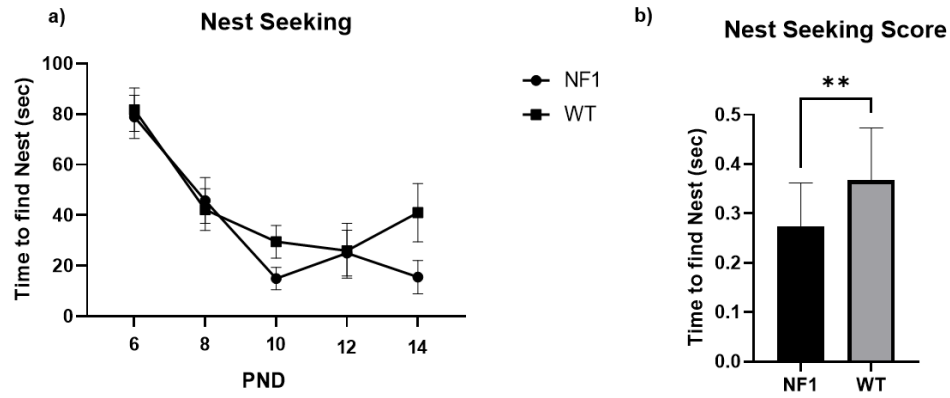

Figure 12 The nest seeking test revealed a) no significant daily differences between NF1 and WT mice in time to find the nest. However, b) on average, NF1 mice took less time to reach the nest. Two-way ANOVA with post hoc planned multiple comparisons (a) and one-tailed Mann-Whitney U test (b); n=28 3)

3) Negative Geotaxis: No difference between WT and NF1 groups were observed in negative geotaxis test. Both groups spent similar time to turn to face the top of the incline in all testing days and overall (p=0.325, NF1>WT) (Figure 13a and b).

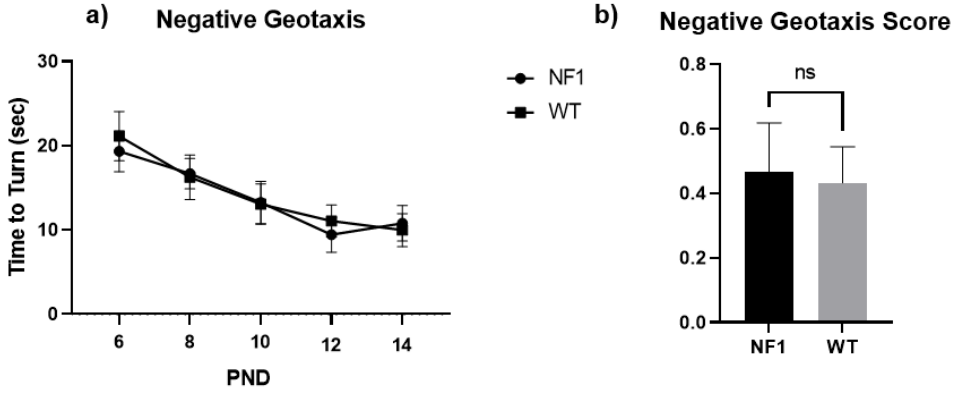

Figure 13 In the negative geotaxis test showed no difference a) in time to turn daily or b) overall between NF1 and WT mice. Two-way ANOVA with post hoc planned multiple comparisons (a) and one-tailed Mann-Whitney U test (b); n=28

4) Surface Righting: Both groups had similar performance in the surface righting test and no significant differences between NF1 and WT mice were detected in time to flip to a supine position during the testing period (p=0.160, NF1+/−<WT) (Figure 14a and b).

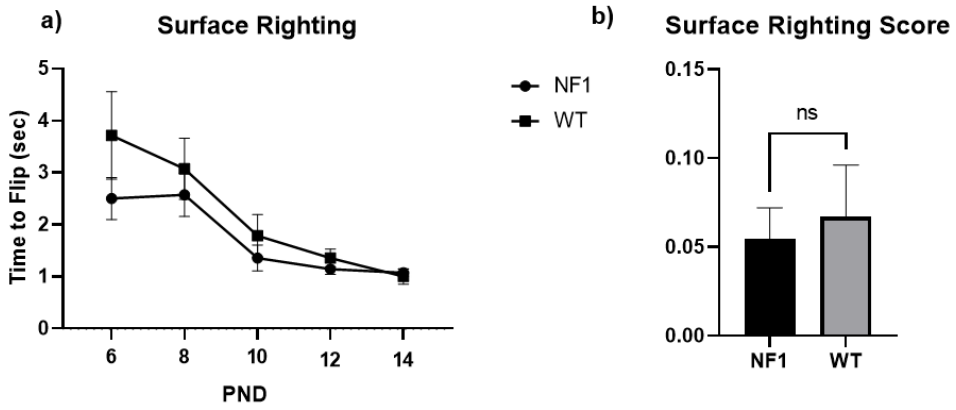

Figure 14 *Nf1*$^{+/-}$ depletion does not seem to affect surface righting reflex. a) no differences between NF1 and WT mice in surface righting reflex on any of the testing days nor b) overall. Two-way ANOVA with post hoc planned multiple comparisons (a) and one-tailed Mann-Whitney U test (b); n=28

Adult Behaviour

1) Social: In the three-chamber test, NF1 mice did not show a clear preference between the chamber with the mouse over the empty chamber (p=0.674), in contrast to the WT group, who showed a significant preference for the chamber with the mouse (p=0.046) (Figure 15a). The sociability scores (preference for mouse vs empty) for each group,

however, were not statistically different (p=0.161, NF1<WT) (Figure 15b). Neither NF1 nor WT showed a marked preference for novel or familiar animal on the second day (p=0.4, and p=0.686 respectively) (Figure 15c). However, social memory (preference for new vs familiar) scores were lower in the NF1 group than in the WT group (p=0.047, NF1<WT) (Figure 15d).

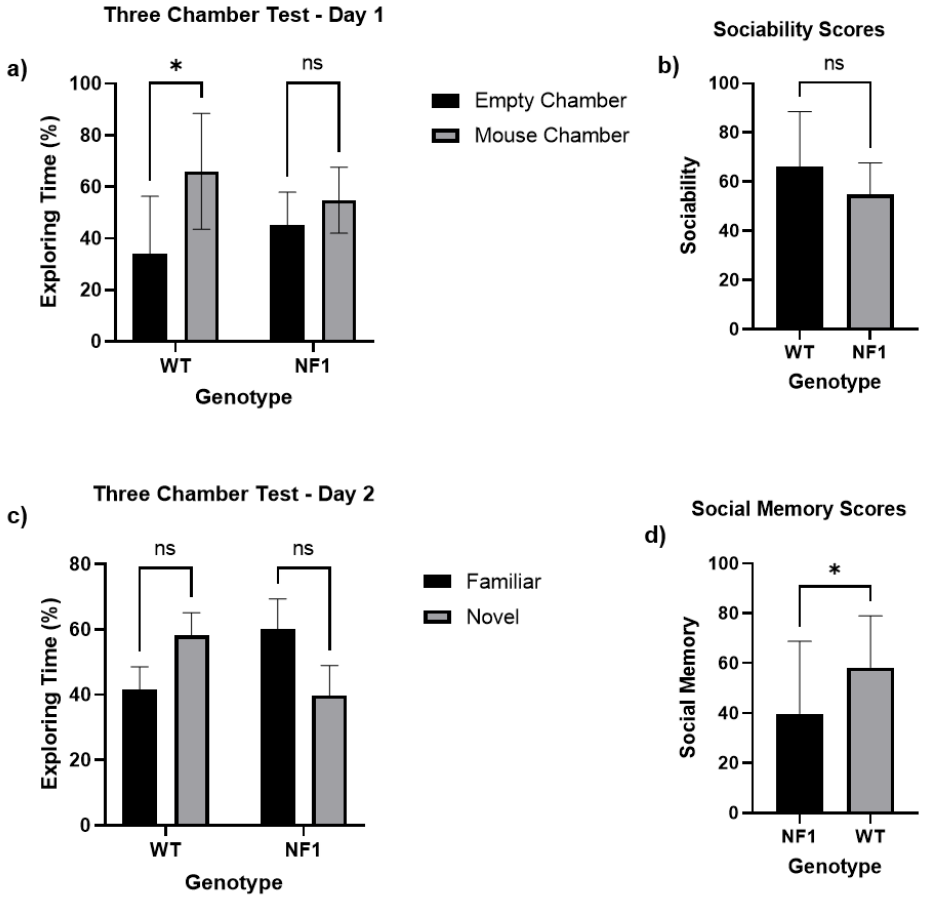

Figure 15 *Nf1*$^{+/-}$depletion affects social behavior. a) NF1 mice show no preference between a conspecific and an empty chamber.Two-way ANOVA b) No clear difference was observed in sociability (time spent with conspecific) between WT and NF1 mice. One-tailed Mann-Whitney U test c) Neither WT nor NF1 show marked preference for either familiar or novel mice. Two-way ANOVA However, d) NF1 mice spent less time with novel mouse.One-tailed Mann-Whitney U test; n=19

2) Repetitive Behaviour: At the 15-minute of marble burying test, NF1 mice had buried significantly more marbles than the WT group (p=0.031). Differences in number of marbles buried at the other time points (5, 10, 20, 25 and 30 minutes) were not statistically significant (Figure 16a). However, NF1 group buried, on average, more marbles than the WT group (p=0.007, NF1>WT) (Figure 16b).

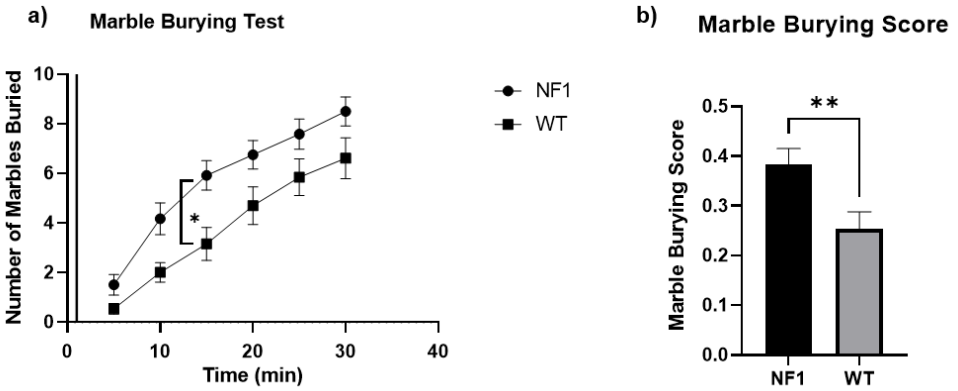

Figure 16 NF1 exhibit repetitive behavior in the marble burying test. a) NF1 mice bury significantly more marbles at the 15 min time point than their WT littermates. Two-Way ANOVA with post hoc planned comparisons b) The total number of marbles buried by the NF1 group was significantly higher than WT. One-tailed Mann-Whitney Test; n=25

No significant differences were found in time spent grooming between the two groups for each time interval (0-5 minutes: p=0.947; 5-10 minutes: p=0.541) (Figure 17a) nor in rearing time (0-5 minutes: p=0.965; 5-10 minutes: p=0.994) (Figure 17b).

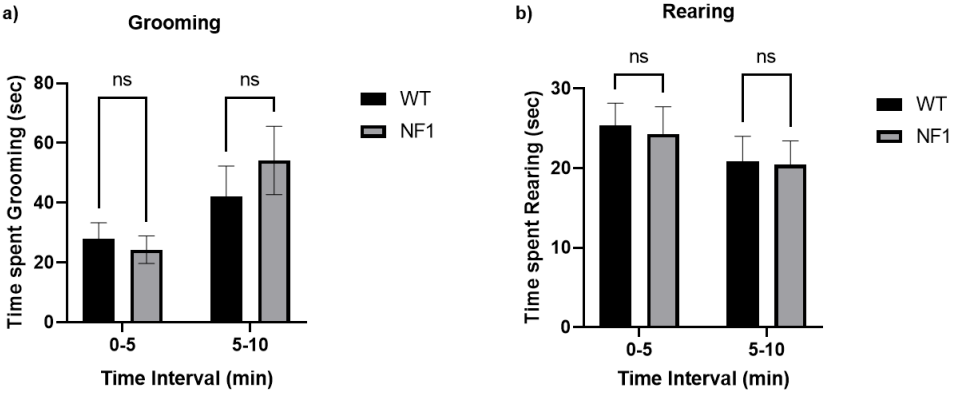

Figure 17 Read-outs from the open field test did not show any difference between NF1 and WT group in a) grooming or in b) rearing. Two-way ANOVA; n=29.

3) Learning and Memory: No differences were observed between NF1 and WT groups in time spent in zone in each of the 4 days of the Morris Water Maze, nor in daily increase in learning (p=0.77, NF1<WT) (Figure 18a and b)**.** Likewise, no significant differences were found between NF1 and WT mice regarding time spent in quadrant 4 and target zone during the last trial of each day. However, when day 4 is dismissed, the increase in memory from day 2 and day 3 is lower for the NF1 group (p=0.041, NF1<WT) (Figure 18c and d).

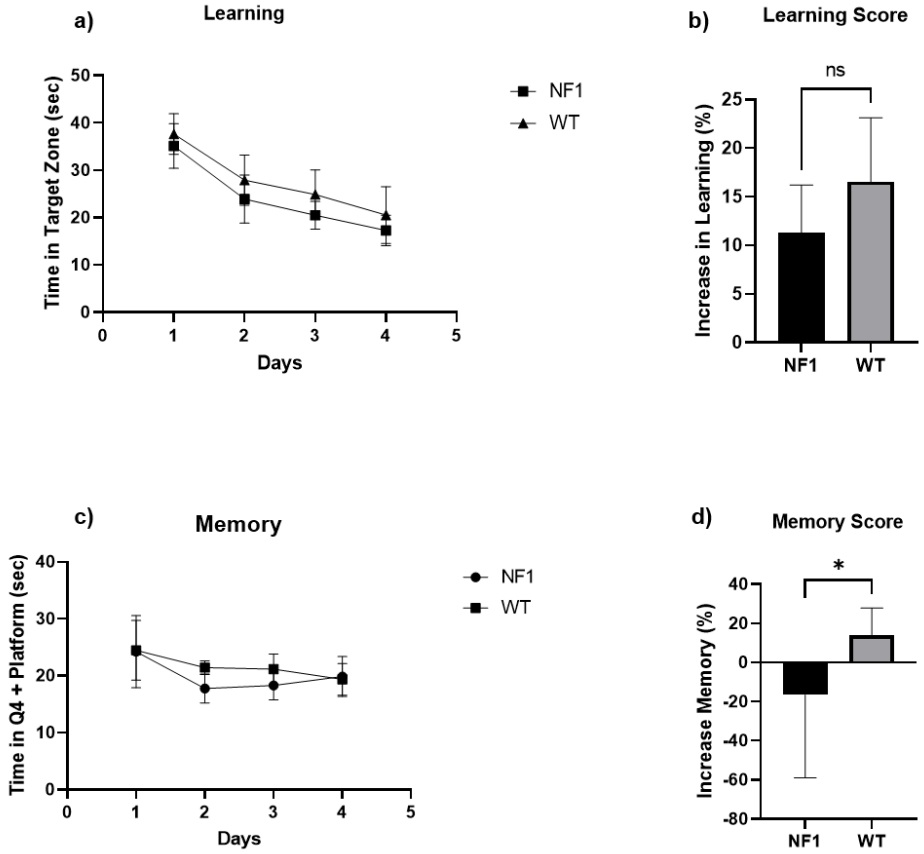

Figure 18 Read-outs from the Morris Water Maze show a) no significant differences on each day between NF1 mice and WT littermates. Two-Way ANOVA with posthoc planned comparisons b) no differences in learning throughout the testing period between NF1 and WT groups. One-tailed Mann-Whitney U Test c) No differences were observed in memory on each single day between NF1 and WT mice. Two-way ANOVA with posthoc planned comparisons. However, d) NF1 show a decrease in memory during days 2 and 3 of the test, while WT mice seem to increase their memory on the same days. One-tailed Mann-Whitney U test; n=21

ASD Severity Composite Score

To build the severity classification model, one score for each measure was used, and missing values replaced by mean score in group (WT or *Nf1*$^{+/-}$). Scores were z-standardized. Measures where NF1group had lower scores than WT were inversed, so that higher scores would correspond to the *Nf1*$^{+/-}$genotype. A composite score was calculated as the average of ranks from all indexes and another one using only the measures that were statistically significant according to the Mann-Whitney U test, for comparison purposes.

Correlation analysis identified redundant variables and have poor relationship with the composite score. Negative geotaxis was excluded from the “ALL” model (Figure 19).

Pearson's Correlations

Variable		Social Memory	Marble Burying	Sociability	Grooming	Rearing	Learning	Memory	Locomotion	Negative Geotaxis	Surface Righting	Nest Seeking	composite_all
1. Social Memory	Pearson's r	—											
	p-value	—											
2. Marble Burying	Pearson's r	0.040	—										
	p-value	0.844	—										
3. Sociability	Pearson's r	0.451	0.146	—									
	p-value	0.018	0.468	—									
4. Grooming	Pearson's r	−0.018	0.099	0.126	—								
	p-value	0.928	0.622	0.533	—								
5. Rearing	Pearson's r	0.159	0.121	0.237	0.222	—							
	p-value	0.428	0.549	0.234	0.266	—							
6. Learning	Pearson's r	0.291	−0.087	0.210	0.420	0.321	—						
	p-value	0.141	0.666	0.293	0.029	0.102	—						
7. Memory	Pearson's r	0.082	0.428	0.051	0.353	0.041	0.181	—					
	p-value	0.684	0.026	0.800	0.071	0.840	0.367	—					
8. Locomotion	Pearson's r	−0.055	0.057	−0.031	0.090	−0.331	−0.072	0.274	—				
	p-value	0.786	0.777	0.878	0.656	0.092	0.720	0.167	—				
9. Negative Geotaxis	Pearson's r	−0.148	0.055	−0.001	−0.278	0.050	−0.076	−0.067	0.558	—			
	p-value	0.462	0.787	0.995	0.161	0.805	0.705	0.741	0.002	—			
10. Surface Righting	Pearson's r	−0.123	−0.077	−0.110	0.315	−0.155	0.233	0.354	0.232	−0.186	—		
	p-value	0.542	0.702	0.585	0.109	0.440	0.242	0.070	0.245	0.354	—		
11. Nest Seeking	Pearson's r	0.326	0.108	0.337	0.268	0.140	0.345	0.339	0.071	−0.215	0.242	—	
	p-value	0.097	0.592	0.086	0.177	0.487	0.078	0.083	0.723	0.280	0.223	—	
12. composite_all	Pearson's r	0.447	0.371	0.489	0.605	0.357	0.584	0.643	0.263	−0.073	0.402	0.658	—
	p-value	0.019	0.057	0.010	< .001	0.067	0.001	< .001	0.185	0.717	0.037	< .001	—

Figure 19 Correlation analysis for the composite score calculation with all behavioral and neurodevelopmental scores. Negative geotaxis is not significantly correlated with the composite score, and has a negative effect on it. Locomotion and rearing are also not significantly correlated with the final composite score.

Three models were built and their performance on distinguishing between WT and NF1 was tested, summarised in Table 1:

a) **Model 1** included all the statistically significant developmental milestones and behavioural indexes. This model could identify *Nf1*$^{+/-}$genotype with a precision of 87,5% and distinguish between *Nf1*$^{+/-}$ and WT with an accuracy of 92,59%. The error rate of the model was 7,41%. Internal consistency was poor (cronbach's α= 0,458; McDonald's ω= 0,486). This model was created for comparison purposes, but is not a good option, since by using only the statistically significant measures in the same data set, redundancy is introduced, and the model tends to overfit the data, explaining the recall of 100%. It might not be as accurate and precise with different sets of data.

b) **Model 2** was built using all measures except negative geotaxis. It was 85,7% precise at identifying the *Nf1*$^{+/-}$ genotype, with 85,7% recall and 85,2% accuracy at distinguishing between *Nf1*$^{+/-}$ and WT. The error rate of this model is 14,8%. Internal consistency is acceptable (cronbach's α= 0,634; McDonald's ω= 0,654).

c) **Model 3** used all the measures except ones identified through correlation analysis to have a weak correlation with the final composite score. These measures were negative geotaxis, locomotion, rearing. The model distinguished between *Nf1*$^{+/-}$and WT with an accuracy of 81,48%. It identified *Nf1*$^{+/-}$genotype with a precision of 80% and recall of 85,71%. Error rate was 18,52%. Internal consistency was acceptable (cronbach's α= 0,653; McDonald's ω= 0,667).

Table 1 Summary of the 3 different models created to calculate the ASD severity composite score

Model	**Precision**	**Recall**	**F Score**	**Accuracy**	**Error Rate**	**Cronbach's α**	**McDonald's ω**
1 -Only Significant	87,50%	100,00%	93,33%	92,59%	7,41%	0,458	0,486
2- ALL - Neg Geotaxis	85,7%	85,7%	85,7%	85,2%	14,8%	0,634	0,654
3- All - Negative Geotaxis, Loc, Rearing	80,00%	85,71%	82,76%	81,48%	18,52%	0,653	0,667

ASD Severity Multi-classification Model

Model 2 was selected as the best model due to its ability to distinguish between genotypes with high accuracy (Figure 20). It was then used to determine threshold for the three different severity levels of autism spectrum disorder. The three levels of severity are only given to animals classified as NF1 by the model.

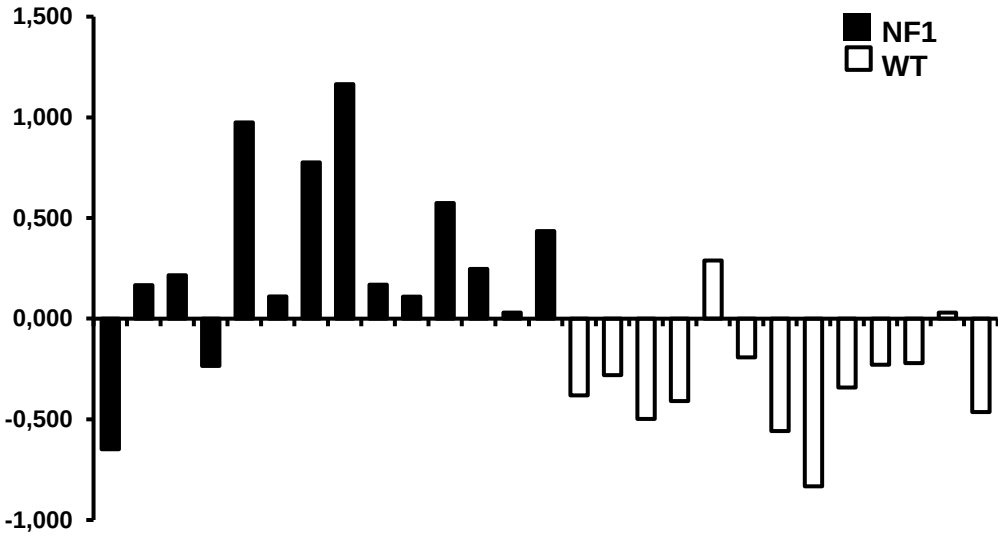

Figure 20 NF1 vs WT composite scores obtained with model 2. NF1 mice consistently scored higher (>0) than WT mice (<0); n=27

Considering NF1 as a model for autistic-like social behaviour, conditions were added to ensure that animals classified as level 3 presented deficits in this dimension. Hence, for a level 3 classification, NF1 mice had to have a composite score higher than 0.5 and either social memory or sociability score higher than 1. For levels 1 and 2, only the composite score was considered (level 1: composite score<0.25; level 2: 0.25< composite score >0.5).

Table 2 Multi-classification model performance

Classification	Precision	Recall	F Score
WT	85%	100%	92%
Level 1	12%	25%	17%
Level 2	33%	25%	29%
Level 3	67%	25%	36%
Accuracy	56%		

Classification was compared against the severity classification performed by an independent observer. The model showed a 56% accuracy (
Table **2**). The model classified 57,14% of the NF1 mice as level 1 (mild), 21,4% as level 2 (moderate), and 21,4% as level 3 (severe) (Figure 21).

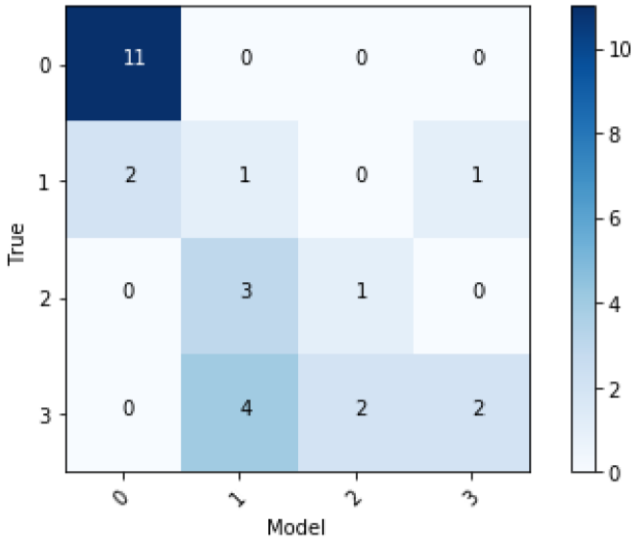

Figure 21 Confusion matrix for the multi-classification model. Number of actual animals per classification (True) vs number of animals classified by the model (Model). n=27

Discussion

During neonatal period, NF1 pups show impairments in locomotion, since they spent more time than their WT littermates to exit the circle during the locomotion test. This delay means that NF1 animals loose rotatory movements later than WT. Accordingly, NF1 patients often show motor deficits, including gait impairment, speed, and agility, as well as muscle weakness.[44,45] It has been proposed that motor deficits may be linked with abnormal dopamine levels or structural and functional irregularities in brain areas related to motor function, such as corpus callosum, thalamus, and caudate nucleus, given the importance of neurofibromin to the correct development of the brain.[46] Motor impairments have also been reported in ASD patients.[47]

On the other hand, the NF1 group was faster in reaching the nest, during the nest seeking test. These observations were unexpected, as NF1 has been associated with deficits in olfactory learning in fly models [48] and abnormal olfactory preference in mice[49]. However, it could be that WT pups are more exploratory and explore the arena before actively looking for the nest, thus taking longer to reach it. In fact, NF1 mice tend to show reduced exploratory behaviour than their WT counterparts.[50]

No significant differences between groups were found in negative geotaxis and surface righting reflex. Previous studies in NF1 mice have reported similar results in surface righting reflex. Maloney et al (2017) tested surface righting reflex of NF1 mice on PND14 by placing the mice in a 50-mL conical containing a lid with a hole, which was turned 180º when the belly of the mouse was facing down, placing the mouse on its back. Time the animal took to right itself with all four paws underneath its belly was recorded. The authors reported no differences in righting reflex between NF1 and WT mice.[51] No studies were found in NF1 mouse model for negative geotaxis. However, studies done in maternal immune activation model of ASD also show no significant alterations in negative geotaxis.[52] On the other hand, the C58/J mouse model of ASD has been reported to show enhanced negative geotaxis on PNDs 8 and 10.[53]

Regarding adult behaviour, NF1 male mice showed social impairments associated with social recognition/memory. NF1 mice did not show any marked interest in the mouse when presented with the option to choose between an empty chamber and a chamber with a conspecific. On the second day of the test, NF1 group spent more time with the familiar mouse than the WT group, when given the choice between familiar and novel mice, indicating less social memory. Previous studies in the NF1 mouse model have also reported impairments in long-term (24h) social memory.[25] Multiple authors have suggested that social impairments in ASD may be caused by impairments in facial

recognition or social memory. In fact, NF1 patients have been reported to show less attention to faces[54].

To explore learning and memory, we have used the Morris water maze test. Here, NF1 mice showed no significant difference in time spent in the platform area and in time to reach the platform. However, overall, the WT group showed more increase in memory, in days two and three of the test. Previous studies have suggested that despite learning and memory deficit, NF1 mice can reach the platform within the same time as WT by using different strategies and extended training can further improve their performance in this test.[55]. Deficits in spatial memory and learning have also been documented in children with NF1.[56] It has been suggested that these deficits may be related to visual impairments. A study performed in NF1 patients used fMRI to study early cortical visual pathways found alterations in activation of the visual cortex in response to low-level visual stimulation, which could explain the visuospatial impairments associated with NF1. The same study also reported abnormal default-mode network (DMN) activation during stimulation periods in the NF1 group, suggesting that attention deficits may also impair memory and learning in NF1 patients. [57]

Repetitive behaviour is one of the core characteristics of ASD. Therefore, to account for this dimension, NF1 mice and their WT littermates were tested using marble burying test. The NF1 group buried significantly more marbles than the WT group, showing that NF1 mice do engage in repetitive behaviours. The biggest difference between number of marbles buried by the NF1 group and the WT group was at the 15-minute time point. Also, grooming and rearing were used as measure for repetitive behaviour. However, NF1 mice did not seem to differ considerably from WT in either grooming time or rearing time. Repetitive behaviour has not yet been thoroughly characterized in this mouse model. High scores in the marble burying test have been reported in some ASD mouse models, such as BTBR T + tf/J mice[30], VPA[58], and maternal immune activation[59] models. However reduced burying activity has also been reported in several other ASD mouse models, such as Shank1$^{+/-}$, Shank1$^{-/-}$, Shank2$^{-/-}$, and mice with loss of maternal Ube3a.[30] Also, increased self-grooming has been reported in some mouse models of ASD, such as the VPA mouse model[58], BTBR T+tf/J inbred strain[60] , and NL1 KO mice[61] while some models, such as *Nlgn4* male mice[31], *Nlgn2*, *Shank1*$^{-/-}$ do not seem to show significant increase in grooming activity.[62] Several ASD mouse models have been reported to show decreased rearing, such as the *Shank1* $^{-/-}$, *Ephrin-A* $^{-/-}$ [63], and BTBR T+Itpr3tf/J[64]. Both grooming and rearing behaviour are also associated with anxiety.[65,66] Therefore, these activities may be influenced by levels of anxiety, either caused by external environment or by the genetic background, explaining the variability of these measures in different studies and models.

Repetitive behaviour and stereotypies are core characteristics of ASD in humans.[1] However, it has been reported that children with NF1 and ASD seem to exhibit less repetitive behaviours than the general ASD population.[67,68]

When the scores for each behavioural and neurodevelopmental measure were combined, composite scores were obtained for all subjects. The chosen model included all scores except negative geotaxis which was found to not only be negatively correlated with the final score but have a weak correlation with it. This model was able to distinguish between *Nf1*$^{+/-}$and WT genotypes with an accuracy of 85.2% and identify *Nf1*$^{+/-}$genotype with a precision of 85.7%, confirming that the chosen scores were appropriate. It is also worth noting that even within the NF1 group, different composite scores were obtained, corroborating that different symptom severity and presentation exist within the group.

To add severity levels to the classification model, thresholds for each level were proposed based on the composite scores of the NF1 group, as well as the average distribution of ASD severity in the NF1 population, and taking into account the main characteristics of the disorder – i.e., social impairments. When the severity levels were added to the model, the accuracy decreased considerably to 58.62%. This decrease was, however, not surprising, as extra levels of information require more data, which was not available. The model was, then, tested in a very small sample (16 animals classified as Nf1 through the composite score). It is expected that, in the future, as the sample size increases, the accuracy will also increase.

This multi-classification model classified 57,14% of the NF1 mice as level 1 (mild), 21,4% as level 2 (moderate), and 21,4% as level 3 (severe).This distribution is not far from what is reported in the NF1 population, where 42%- 70% of NF1 patients with ASD will have a ‘mild to moderate’ score, and 30-58% will have a severe score.[69,70]

Strengths, weaknesses, and recommendations

The classification model built in this project is accurate and precise enough to classify autism severity in the NF1 mouse model,

It is important to take into consideration that this project was done using a relatively small sample with high variability within the data, which may impact on statistical analysis. However, all tests performed were chosen with this in mind and are all possible to use in small samples. The model will continue to improve as more data is provided.

Another possible weakness are the missing values from the dataset that was used. These missing values had to be replaced with the average score of the measure for the group (NF1 or WT), which can introduce some bias to the model. This is expected to improve as more data is included in the model.

In this project, thresholds for the three different levels were proposed but need to be further validated. To ensure that was no redundancy in the true classification, an independent observer watched the social behaviour videos for each animal and attributed their own classification, which was then compared to the classification obtained by the model.

It is also worth noticing that adding the classification by level decreased the accuracy of the model. Still, this classification adds an extra level of information that should prove useful in future research in the area and should be improved as more data is fed to the model.

Lastly, the classification model should be validated in different animal models to ensure its replicability.

Conclusion

Data obtained from developmental milestones and adult behaviour of NF1 and WT littermates were consistent with previous published studies for the NF1 mouse model and/or for NF1 patients. Accordingly, these behaviours are also observed in ASD population, suggesting that the NF1 mouse model recapitulates behavioural features of both NF1 and ASD.

When combined, all behavioural and neurodevelopmental read-outs (except negative geotaxis) produced a composite score that clearly distinguished between WT and NF1 mice (with an accuracy of 85,2%) and identified *Nf1*$^{+/-}$genotype with 85,7% precision, which shows that the chosen read-outs were good distinguishing factors for this mouse model.

As expected, different severity scores were found within the NF1 group, confirming that animals with the same genetic background, and kept in similar housing conditions can present different severity of symptoms.

Differentiation into 3 levels of severity was possible with the multi-classification model. However, due to a small amount of data, adding the extra levels of information reduced the accuracy of the model. In the future, as more data is obtained and included in the model, accuracy should increase. Additionally, the distribution of ASD severity scores attributed by the model are similar to the distribution of severity in NF1 patients with ASD, which could indicate that the thresholds proposed are reasonable.

Since the read-outs used are characteristic of ASD, this model should be applicable to other mouse models of ASD, even if minor some adjustments need to be made and this will also be tested in the future.

The multi-classification model should provide researchers the opportunity to further the knowledge of ASD, allowing, for instance, studies on the impact of severity on therapeutic strategies, as well as the influence of other, non-genetic, factors on symptom severity, or possible correlations between brain structure and severity level, among others. Further, classification by severity level will allow experimental designs that reduce variability, allowing for less animals to be used in research.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. 5th, revised ed. American Psychiatric Publishing; 2013.

2. Bailey A, Le Couteur A, Gottesman I, et al. Autism as a strongly genetic disorder evidence from a british twin Study. *Psychol Med*. 1995;25(1):63-77. doi:10.1017/S0033291700028099

3. Searles Quick VB, Wang B, State MW. Leveraging large genomic datasets to illuminate the pathobiology of autism spectrum disorders. *Neuropsychopharmacology*. 2021;46(1):55-69. doi:10.1038/s41386-020-0768-y

4. An JY, Claudianos C. Genetic heterogeneity in autism: From single gene to a pathway perspective. *Neurosci Biobehav Rev*. 2016;68:442-453. doi:10.1016/j.neubiorev.2016.06.013

5. Satterstrom FK, Kosmicki JA, Wang J, et al. Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *bioRxiv*. 2018.

6. Gilbert J, Man HY. Fundamental elements in autism: From neurogenesis and neurite growth to synaptic plasticity. *Front Cell Neurosci*. 2017. doi:10.3389/fncel.2017.00359

7. Lord C, Brugha TS, Charman T, et al. Autism spectrum disorder. *Nat Rev Dis Prim*. 2020;6(1):5. doi:10.1038/s41572-019-0138-4

8. Cheroni C, Caporale N, Testa G. Autism spectrum disorder at the crossroad between genes and environment: contributions, convergences, and interactions in ASD developmental pathophysiology. *Mol Autism*. 2020;11(1):69. doi:10.1186/s13229-020-00370-1

9. Bilbo SD, Block CL, Bolton JL, Hanamsagar R, Tran PK. Beyond infection - Maternal immune activation by environmental factors, microglial development, and relevance for autism spectrum disorders. *Exp Neurol*. 2018. doi:10.1016/j.expneurol.2017.07.002

10. Hegarty JP, Pegoraro LFL, Lazzeroni LC, et al. Genetic and environmental influences on structural brain measures in twins with autism spectrum disorder. *Mol Psychiatry*. 2020;25(10):2556-2566. doi:10.1038/s41380-018-0330-z

11. Castelbaum L, Sylvester CM, Zhang Y, Yu Q, Constantino JN. On the Nature of

Monozygotic Twin Concordance and Discordance for Autistic Trait Severity: A Quantitative Analysis. *Behav Genet*. 2020;50(4):263-272. doi:10.1007/s10519-019-09987-2

12. Diwakar G, Zhang D, Jiang S, Hornyak TJ. Neurofibromin as a regulator of melanocyte development and differentiation. *J Cell Sci*. 2008;121(2):167 LP - 177. doi:10.1242/jcs.013912

13. Daston MM, Scrable H, Nordlund M, Sturbaum AK, Nissen LM, Ratner N. The protein product of the neurofibromatosis type 1 gene is expressed at highest abundance in neurons, Schwann cells, and oligodendrocytes. *Neuron*. 1992;8(3):415-428. doi:https://doi.org/10.1016/0896-6273(92)90270-N

14. Hegedus B, Dasgupta B, Shin JE, et al. Neurofibromatosis-1 Regulates Neuronal and Glial Cell Differentiation from Neuroglial Progenitors In Vivo by Both cAMP- and Ras-Dependent Mechanisms. *Cell Stem Cell*. 2007;1(4):443-457. doi:https://doi.org/10.1016/j.stem.2007.07.008

15. Bollag G, Clapp DW, Shih S, et al. Loss of NF1 results in activation of the Ras signaling pathway and leads to aberrant growth in haematopoietic cells. *Nat Genet*. 1996;12(2):144-148. doi:10.1038/ng0296-144

16. Ratner N, Miller SJ. A RASopathy gene commonly mutated in cancer: the neurofibromatosis type 1 tumour suppressor. *Nat Rev Cancer*. 2015;15(5):290-301. doi:10.1038/nrc3911

17. Gutmann DH, Ferner RE, Listernick RH, Korf BR, Wolters PL, Johnson KJ. Neurofibromatosis type 1. *Nat Rev Dis Prim*. 2017;3(1):17004. doi:10.1038/nrdp.2017.4

18. Lehtonen A, Howie E, Trump D, Huson SM. Behaviour in children with neurofibromatosis type 1: cognition, executive function, attention, emotion, and social competence. *Dev Med Child Neurol*. 2013;55(2):111-125. doi:10.1111/j.1469-8749.2012.04399.x

19. HUIJBREGTS S, JAHJA R, DE SONNEVILLE LEO, DE BREIJ S, SWAAB-BARNEVELD H. Social information processing in children and adolescents with neurofibromatosis type 1. *Dev Med Child Neurol*. 2010;52(7):620-625. doi:https://doi.org/10.1111/j.1469-8749.2010.03639.x

20. Molosh AI, Shekhar A. Neurofibromatosis type 1 as a model system to study molecular mechanisms of autism spectrum disorder symptoms. In: *Progress in*

Brain Research. Vol 241. 1st ed. Elsevier B.V.; 2018:37-62. doi:10.1016/bs.pbr.2018.09.014

21. Morris SM, Acosta MT, Garg S, et al. Disease Burden and Symptom Structure of Autism in Neurofibromatosis Type 1: A Study of the International NF1-ASD Consortium Team (INFACT). *JAMA Psychiatry*. 2016;73(12):1276-1284. doi:10.1001/jamapsychiatry.2016.2600

22. Violante IR, Patricio M, Bernardino I, et al. GABA deficiency in NF1: A multimodal [11C]-flumazenil and spectroscopy study. *Neurology*. 2016;87(9):897-904. doi:10.1212/WNL.0000000000003044

23. Zimerman M, Wessel MJ, Timmermann JE, et al. Impairment of Procedural Learning and Motor Intracortical Inhibition in Neurofibromatosis Type 1 Patients. *EBioMedicine*. 2015;2(10):1430-1437. doi:https://doi.org/10.1016/j.ebiom.2015.08.036

24. Gonçalves J, Violante IR, Sereno J, et al. Testing the excitation/inhibition imbalance hypothesis in a mouse model of the autism spectrum disorder: In vivo neurospectroscopy and molecular evidence for regional phenotypes. *Mol Autism*. 2017;8(47). doi:10.1186/s13229-017-0166-4

25. Molosh AI, Johnson PL, Spence JP, et al. Social learning and amygdala disruptions in Nf1 mice are rescued by blocking p21-activated kinase. *Nat Neurosci*. 2014;17(11):1583-1590. doi:10.1038/nn.3822

26. Moy SS, Nadler JJ, Magnuson TR, Crawley JN. Mouse models of autism spectrum disorders: The challenge for behavioral genetics. *Am J Med Genet - Semin Med Genet*. 2006;142C(1):40-51. doi:10.1002/ajmg.c.30081

27. Heyser CJ. Assessment of developmental milestones in rodents. *Curr Protoc Neurosci*. 2004;Chapter 8:Unit 8.18. doi:10.1002/0471142301.ns0818s25

28. Crawley JN. Mouse behavioral assays relevant to the symptoms of autism. *Brain Pathol*. 2007;17(4):448-459. doi:10.1111/j.1750-3639.2007.00096.x

29. Silverman JL, Yang M, Lord C, Crawley JN. Behavioural phenotyping assays for mouse models of autism. *Nat Rev Neurosci*. 2010. doi:10.1038/nrn2851

30. Pasciuto E, Borrie SC, Kanellopoulos AK, et al. Autism Spectrum Disorders: Translating human deficits into mouse behavior. *Neurobiol Learn Mem*. 2015;124:71-87. doi:10.1016/j.nlm.2015.07.013

31. El-Kordi A, Winkler D, Hammerschmidt K, et al. Development of an autism severity score for mice using Nlgn4 null mutants as a construct-valid model of heritable monogenic autism. *Behav Brain Res*. 2013;251:41-49. doi:10.1016/j.bbr.2012.11.016

32. Duda M, Kosmicki JA, Wall DP. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Transl Psychiatry*. 2014;4(8):e424-e424. doi:10.1038/tp.2014.65

33. Kazeminejad A, Sotero RC. Topological Properties of Resting-State fMRI Functional Networks Improve Machine Learning-Based Autism Classification. *Front Neurosci*. 2019;12:1018. doi:10.3389/fnins.2018.01018

34. Parikh MN, Li H, He L. Enhancing Diagnosis of Autism With Optimized Machine Learning Models and Personal Characteristic Data. *Front Comput Neurosci*. 2019;13:9. doi:10.3389/fncom.2019.00009

35. Abbas H, Garberson F, Glover E, Wall DP. Machine learning approach for early detection of autism by combining questionnaire and home video screening. *J Am Med Informatics Assoc*. 2018;25(8):1000-1007. doi:10.1093/jamia/ocy039

36. Rudovic O, Lee J, Dai M, Schuller B, Picard RW. Personalized machine learning for robot perception of affect and engagement in autism therapy. *Sci Robot*. 2018;3(19). doi:10.1126/scirobotics.aao6760

37. Tang TY. Helping Neuro-Typical Individuals to “Read” the Emotion of Children with Autism Spectrum Disorder: An Internet-of-Things Approach. In: *Proceedings of the The 15th International Conference on Interaction Design and Children*. IDC ’16. New York, NY, USA: Association for Computing Machinery; 2016:666–671. doi:10.1145/2930674.2936009

38. Tang TY, Winoto P. An Interactive Picture Exchange Communication System (PECS) Embedded with Augmented Aids Enabled by IoT and Sensing Technologies for Chinese Individuals with Autism. In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. UbiComp ’18. New York, NY, USA: Association for Computing Machinery; 2018:299–302. doi:10.1145/3267305.3267629

39. Hassan MM, Mokhtar HMO. Investigating autism etiology and heterogeneity by decision tree algorithm. *Informatics Med Unlocked*. 2019;16:100215.

doi:https://doi.org/10.1016/j.imu.2019.100215

40. Artoni P, Piffer A, Vinci V, et al. Deep learning of spontaneous arousal fluctuations detects early cholinergic defects across neurodevelopmental mouse models and patients. *Proc Natl Acad Sci*. 2020;117(38):23298 LP - 23303. doi:10.1073/pnas.1820847116

41. VanRyzin JW, Yu SJ, Perez-Pouchoulen M, McCarthy MM. Temporary Depletion of Microglia during the Early Postnatal Period Induces Lasting Sex-Dependent and Sex-Independent Effects on Behavior in Rats. *eNeuro*. 2016;3(6):ENEURO.0297-16.2016. doi:10.1523/ENEURO.0297-16.2016

42. Ferreira H. Longitudinal Study of Ultrasonic Vocalizations in Pup and Juvenile Tsc2+/- Mice Reveals Communication Abnormalities. [Undergraduate thesis]. Coimbra, Portugal: Universidade de Coimbra; 2020.

43. Nguyen HT, Hinman MN, Guo X, et al. Neurofibromatosis type 1 alternative splicing is a key regulator of Ras/ERK signaling and learning behaviors in mice. *Hum Mol Genet*. 2017;26(19):3797-3807. doi:10.1093/hmg/ddx264

44. Champion JA, Rose KJ, Payne JM, Burns J, North KN. Relationship between cognitive dysfunction, gait, and motor impairment in children and adolescents with neurofibromatosis type 1. *Dev Med Child Neurol*. 2014;56(5):468-474. doi:10.1111/dmcn.12361

45. Souza JF, Passos RLF, Guedes ACM, Rezende NA, Rodrigues LOC. Muscular force is reduced in neurofibromatosis type 1. *J Musculoskelet Neuronal Interact*. 2009;9(1):15-17.

46. Summers MA, Quinlan KG, Payne JM, Little DG, North KN, Schindeler A. Skeletal muscle and motor deficits in Neurofibromatosis Type 1. *J Musculoskelet Neuronal Interact*. 2015;15(2):161-170.

47. Licari MK, Alvares GA, Varcin K, et al. Prevalence of Motor Difficulties in Autism Spectrum Disorder: Analysis of a Population-Based Cohort. *Autism Res*. 2020;13(2):298-306. doi:https://doi.org/10.1002/aur.2230

48. Kaas GA, Hawkins KE, Sweatt JD. 4.19 - Genetic Mechanisms of Memory Disorders (Excluding Alzheimer's Disease). In: Byrne JHBT-L and MACR (Second E, ed. Oxford: Academic Press; 2017:371-401. doi:https://doi.org/10.1016/B978-0-12-809324-5.21125-5

49. Wozniak DF, Diggs-Andrews KA, Conyers S, et al. Motivational Disturbances and Effects of L-dopa Administration in Neurofibromatosis-1 Model Mice. *PLoS One*. 2013;8(6):e66024. https://doi.org/10.1371/journal.pone.0066024.

50. Brown JA, Emnett RJ, White CR, et al. Reduced striatal dopamine underlies the attention system dysfunction in neurofibromatosis-1 mutant mice. *Hum Mol Genet*. 2010;19(22):4515-4528. doi:10.1093/hmg/ddq382

51. Maloney SE, Chandler KC, Anastasaki C, Rieger MA, Gutmann DH, Dougherty JD. Characterization of early communicative behavior in mouse models of neurofibromatosis type 1. *Autism Res*. 2018;11(1):44-58. doi:https://doi.org/10.1002/aur.1853

52. Singer HS, Morris C, Gause C, Pollard M, Zimmerman AW, Pletnikov M. Prenatal exposure to antibodies from mothers of children with autism produces neurobehavioral alterations: A pregnant dam mouse model. *J Neuroimmunol*. 2009;211(1):39-48. doi:https://doi.org/10.1016/j.jneuroim.2009.03.011

53. Ryan BC, Young NB, Crawley JN, Bodfish JW, Moy SS. Social deficits, stereotypy and early emergence of repetitive behavior in the C58/J inbred mouse strain. *Behav Brain Res*. 2010;208(1):178-188. doi:https://doi.org/10.1016/j.bbr.2009.11.031

54. Lewis AK, Porter MA, Williams TA, Bzishvili S, North KN, Payne JM. Attention to faces in social context in children with neurofibromatosis type 1. *Dev Med Child Neurol*. 2019;61(2):174-180. doi:https://doi.org/10.1111/dmcn.13928

55. Silva AJ, Frankland PW, Marowitz Z, et al. A mouse model for the learning and memory deficits associated with neurofibromatosis type I. *Nat Genet*. 1997;15(3):281-284. doi:10.1038/ng0397-281

56. Ullrich NJ, Ayr L, Leaffer E, Irons MB, Rey-Casserly C. Pilot Study of a Novel Computerized Task to Assess Spatial Learning in Children and Adolescents With Neurofibromatosis Type 1. *J Child Neurol*. 2010;25(10):1195-1202. doi:10.1177/0883073809358454

57. Violante IR, Ribeiro MJ, Cunha G, et al. Abnormal brain activation in neurofibromatosis type 1: a link between visual processing and the default mode network. *PLoS One*. 2012;7(6):e38785. doi:10.1371/journal.pone.0038785

58. Mehta M V, Gandal MJ, Siegel SJ. mGluR5-Antagonist Mediated Reversal of Elevated Stereotyped, Repetitive Behaviors in the VPA Model of Autism. *PLoS*

One. 2011;6(10):1-6. doi:10.1371/journal.pone.0026077

59. Malkova N V, Yu CZ, Hsiao EY, Moore MJ, Patterson PH. Maternal immune activation yields offspring displaying mouse versions of the three core symptoms of autism. *Brain Behav Immun*. 2012;26(4):607-616. doi:https://doi.org/10.1016/j.bbi.2012.01.011

60. Pearson BL, Pobbe RLH, Defensor EB, et al. Motor and cognitive stereotypies in the BTBR T+tf/J mouse model of autism. *Genes, Brain Behav*. 2011;10(2):228-235. doi:https://doi.org/10.1111/j.1601-183X.2010.00659.x

61. Blundell J, Blaiss CA, Etherton MR, et al. Neuroligin-1 Deletion Results in Impaired Spatial Memory and Increased Repetitive Behavior. *J Neurosci*. 2010;30(6):2115-2129. doi:10.1523/JNEUROSCI.4517-09.2010

62. Bey AL, Jiang Y. Overview of Mouse Models of Autism Spectrum Disorders. *Curr Protoc Pharmacol*. 2014;66(1):5.66.1-5.66.26. doi:https://doi.org/10.1002/0471141755.ph0566s66

63. Kim H, Lim C-S, Kaang B-K. Neuronal mechanisms and circuits underlying repetitive behaviors in mouse models of autism spectrum disorder. *Behav Brain Funct*. 2016;12(1):3. doi:10.1186/s12993-016-0087-y

64. Chao OY, Yunger R, Yang Y-M. Behavioral assessments of BTBR T+Itpr3tf/J mice by tests of object attention and elevated open platform: Implications for an animal model of psychiatric comorbidity in autism. *Behav Brain Res*. 2018;347:140-147. doi:https://doi.org/10.1016/j.bbr.2018.03.014

65. Smolinsky AN, Bergner CL, LaPorte JL, Kalueff AV. Analysis of Grooming Behavior and Its Utility in Studying Animal Stress, Anxiety, and Depression. In: Gould TD, ed. *Mood and Anxiety Related Phenotypes in Mice*. Neurometho. Humana Press; 2009:pp 21-36. doi:10.1007/978-1-60761-303-9

66. Ennaceur A. Tests of unconditioned anxiety — Pitfalls and disappointments. *Physiol Behav*. 2014;135:55-71. doi:https://doi.org/10.1016/j.physbeh.2014.05.032

67. Walsh KS, Vélez JI, Kardel PG, et al. Symptomatology of autism spectrum disorder in a population with neurofibromatosis type 1. *Dev Med Child Neurol*. 2013;55(2):131-138. doi:10.1111/dmcn.12038

68. Garg S, Plasschaert E, Descheemaeker M-J, et al. Autism Spectrum Disorder

Profile in Neurofibromatosis Type I. *J Autism Dev Disord*. 2015;45(6):1649-1657. doi:10.1007/s10803-014-2321-5

69. Hirabaru K, Matsuo M. Neurological comorbidity in children with neurofibromatosis type 1. *Pediatr Int*. 2018;60(1):70-75. doi:https://doi.org/10.1111/ped.13388

70. Adviento B, Corbin IL, Widjaja F, et al. Autism traits in the RASopathies. *J Med Genet*. 2014;51(1):10 LP - 20. doi:10.1136/jmedgenet-2013-101951