Zebrafish models of Mucopolysaccharidosis types IIIA, B, & C show hyperactivity and changes in oligodendrocyte state

Ewan Gerken[1], Syahida Ahmad[2], Lakshay Rattan[3], Kim Hemsley[3], Karissa Barthelson[1,3*] and Michael Lardelli[1*]

[1] Alzheimer's Disease Genetics Laboratory, School of Biological Sciences, The University of Adelaide, Adelaide, SA, Australia

[2] Department of Biochemistry, Faculty of Biotechnology & Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[3] Childhood Dementia Research Group, College of Medicine & Public Health, Flinders Health and Medical Research Institute, Flinders University, Bedford Park, SA, Australia

*Equal senior authors

Key words: MPS III, zebrafish, RNA-seq, disease models, CRISPR editing, lysosomal storage disorder, iron, oligodendrocytes, hyperactivity, neuroinflammation

Abstract

Sanfilippo syndrome childhood dementia, also known as mucopolysaccharidosis type III (MPS III), is a rare inherited lysosomal storage disorder. Subtypes of MPS III are caused by deficiencies in one of four enzymes required for degradation of the glycosaminoglycan heparan sulfate (HS). An inability to degrade HS leads to progressive neurodegeneration and death in the second or third decades of life. Knowledge of MPS III pathogenesis is incomplete, and no effective therapies exist. We generated the hypomorphic mutations *sgsh*^{S387Lfs}, *naglu*^{A603Efs} and *hgsnat*^{G577Sfs} in the endogenous zebrafish genes orthologous to human *SGSH*, *NAGLU*, and *HGSNAT* that are loci for mutations causing MPS III subtypes MPS IIIA, B and C respectively. Our models display the primary MPS III disease signature of significant brain accumulation of HS, while behavioural analyses support anxiety and hyperactivity phenotypes. Brain transcriptome analysis revealed changes related to lysosomal, glycosaminoglycan, immune system and iron homeostasis biology in all three models but also distinct differences in brain transcriptome state between models. The transcriptome analysis also indicated marked disturbance of the oligodendrocyte cell state in the brains of MPS IIIA, B and C zebrafish, supporting that effects on this cell type are an early and consistent characteristic of MPS III. Overall, our zebrafish models recapture key characteristics of the human disease and phenotypes seen in mouse models. Our models will allow exploitation of the zebrafish's extreme fecundity and accessible anatomy to dissect the pathological mechanisms both common and divergent between the MPS IIIA, B, and C subtypes.

Introduction

Sanfilippo syndrome (mucopolysaccharidosis type III, MPS III) is a rare, recessive lysosomal storage disorder causing childhood dementia. MPS III is caused by accumulation in lysosomes of the glycosaminoglycan (GAG) heparan sulfate (HS), with four subtypes A, B, C, and D, distinguished by deficiencies in one of multiple enzymes acting in the HS degradation pathway. These four subtypes are due to hypomorphic mutations in the genes *SGSH*, *NAGLU*, *HGSNAT*, and *GNS* respectively. MPS III presents enormous challenges to the families of affected children due to the broad range of cognitive and behavioural changes accompanying the disease. The disease course begins with failure to achieve developmental milestones, followed by onset of behavioural symptoms such as hyperactivity, aggression and sleep disturbances [1-3]. Later stages of the disease involve progressive loss of cognition, speech, mobility and, eventually, deterioration of autonomic bodily functions such as swallowing and respiration [3, 4]. MPS IIIA is both the most common and, usually, most severe subtype, with mean age-at-death estimated at 15.2 ± 4.2 years, followed by 18.9 ± 7.3 years for IIIB and 23.4 ± 9.5 years for IIIC [5], although this is highly variable across geographic regions [6]. A lack of data exists for MPS IIID due to its scarcity. Significant heterogeneity also exists in the clinical presentation and rate of disease progression both between and within subtypes [7], as well as between affected siblings [8, 9].

MPS III is characterised not only by its clinical presentation but through a range of cellular and molecular pathologies. Analysis of post-mortem patient brain tissues has highlighted accumulation of gangliosides GM2 and GM3, misfolded protein aggregates, along with significant neuroinflammation in MPS disorders [10]. However, animal models are required to dissect the precise interplay of the earlier molecular events leading to neurodegeneration in MPS III, which remain poorly understood. Identifying molecular and behavioural phenotypes that animal models share with human MPS III disease is important for use of the models in research, particularly for screening and measuring the potential therapeutic benefit of pharmaceuticals. Currently, there are no established curative treatments for MPS III, and the treatment options available are limited in effect and mostly supportive [11]. Therefore, there is a pressing need for novel, versatile animal models of the disease to advance our understanding of MPS III pathophysiology and for development of new therapies. While *SGSH*, *NAGLU* and *HGSNAT* encode enzymes in the same HS catabolic pathway, MPS III-causative mutations in these genes have never been compared in the closely related individuals of an animal model organism to see how their phenotypes may differ. Neither have the differing sensitivities of the brain's various cell types to MPS III-causative mutations been studied in detail.

The zebrafish (*Danio rerio*) has emerged in recent years as a valuable model organism for studying lysosomal storage disorders. In total, over 60 zebrafish models of lysosomal storage disorders have been published (reviewed in [12]). Zebrafish share significant genetic [13], neurochemical and neuroanatomical [14] similarity with humans, and their small size, high fecundity, and rapid development make them an attractive system for high-throughput drug screening for discovery of therapeutics. In addition, the transparency of zebrafish embryos and larvae enables easy visualisation and imaging of internal organs, tissues, and subcellular structures including lysosomes. In particular, the reproductive characteristics of zebrafish are advantageous for reducing genetic and environmental noise through intra-family comparisons between mutant and non-mutant genotypes. We have previously had success in this approach with transcriptome analyses of zebrafish carrying Alzheimer's disease-like mutations [15-18]. Zebrafish are also a suitable model for behavioural analysis, with evidence suggesting that regulatory mechanisms behind learning, memory, aggression, anxiety and sleep are conserved between zebrafish and mammals (reviewed in [19]).

A zebrafish model of MPS IIIA has previously been described and recapitulates many features of the human disease including heparan sulfate (HS) accumulation and behavioural deficits [20]. Analysis of effects on brain proteome and transcriptome in this model revealed a variety of changes including to effectors of lysosomal catabolism and to immediate-early gene (IEG) expression important in regulating postsynaptic neuronal responses [21].

Here, we describe novel zebrafish models of MPS IIIA, MPS IIIB and MPS IIIC generated by CRISPR-Cas9 genome editing of zebrafish genes *sgsh*, *naglu* and *hgsnat*. We excluded *GNS* due to its duplication in teleosts [13] and the clinical scarcity of MPS IIID. We apply our previously successful intra-family strategy for comparison of different genotypes to investigation of MPS III-driven changes in behaviour and brain transcriptome states. In particular, we use families of zebrafish siblings including individuals of either of two MPS III subtypes to compare brain transcriptome effects between subtypes. Our results demonstrate that these zebrafish models recapitulate key features of MPS III, including GAG accumulation, metabolic defects, neuroinflammation, and behavioural deficits. Intriguingly, all three MPS III subtypes show marked disturbance of the oligodendrocyte cellular state suggesting that this neural cell type is particularly sensitive to loss of HS-degradative ability. These zebrafish MPS III models will serve as valuable tools for studying the pathophysiology of Sanfilippo syndrome and for the testing of novel therapeutic strategies.

Methods

Zebrafish husbandry and animal ethics statement

Experiments with zebrafish were performed under the auspices of the University of Adelaide Animal Ethics Committee (permit number S-2021-041) and Institutional Biosafety Committee (Dealing ID NLRD 15037). All experiments were performed with an inbred Tübingen (Tü) strain. Zebrafish embryos were collected and incubated in E3 embryo medium [22] at 28°C on a 14/10 hour light/dark cycle until 7 days post fertilisation (dpf). Hatched larvae were then moved to benchtop tanks (without water circulation) and fed live *Rotifera* from 7-21 dpf, followed by live *Artemia salina* from 21 dpf onwards. Juvenile fish were moved to recirculated water aquarium systems at 5 to 6 weeks post fertilisation. Adult zebrafish were maintained in shared recirculated water aquarium systems at 28°C at a density of 2-4 fish per litre (20-30 fish per 8L tank), on a 14/10 hour light/dark cycle, and fed NRD 5/8 dry flake food (Inve Aquaculture, Dendermonde, Belgium) in the morning and live *Artemia salina* (Inve Aquaculture) in the afternoon.

Generation of hypomorphic mutations in zebrafish *sgsh*, *naglu* and *hgsnat*

Mutagenesis

Targeted knock-in mutations were generated in zebrafish using CRISPR-Cas9 (for *sgsh* and *hgsnat*) and CRISPR-Cpf1 (for *naglu*) systems (Integrated DNA Technologies, Coralville, IA, USA) via a non-homologous end joining approach. The sgRNAs used to target each gene are summarised in **Table S1**, and visualised in **Fig.S1**.

The sgRNAs for *sgsh* and *hgsnat* sgRNAs were co-injected into the same G_0 embryos. For both genes, 1 µL of 100 µM crRNA (Integrated DNA Technologies) was mixed with 3 µL of tracrRNA (Integrated DNA Technologies) and 3 µL of nuclease-free duplex buffer (Integrated DNA Technologies). This solution was denatured at 95°C for 5 minutes, before being allowed to cool to room temperature to facilitate sgRNA heteroduplex formation. 3 µL of 64 µM Cas9 3NLS nuclease (Integrated DNA Technologies) was then added and the mixture incubated at 37°C for 10 minutes to allow ribonucleoprotein formation. These solutions for both genes were combined in a 1:1 ratio and ~2-5 nL was injected in zebrafish embryos at the one-cell stage. Injected embryos were incubated in 20 mL of E3 embryo medium at 28°C until 24 hours post fertilisation (hpf).

The mutagenesis of *naglu* was performed similarly to as described by Fernandez, Vejnar [23]. sgRNAs were diluted to 24 µM in 1X TE buffer and mixed in a 1:1 ratio (1 µL:1 µL) with Cpf1 (Cas12a) nuclease (Integrated DNA Technologies) which had been diluted to 20 µM in Cas9 Working Buffer (20 mM HEPES, 300 mM KCl). This mixture was incubated at 37°C for 10 minutes, then ~2-5 nL was

injected into zebrafish embryos at the one-cell stage. Injected embryos were incubated in 20 mL of E3 embryo medium at 34°C until 24 hpf, before being moved back to 28°C.

Screening

At 24 hpf, 10 embryos from each injected group were removed for detection of mutagenesis. The embryos were incubated in 50 µL of recombinant Proteinase K (Roche Holding AG, Basel, Switzerland) diluted to ~2 mg/mL in 1 x tris-EDTA (TE) at 55 °C for 3 hours, with each sample being agitated 3x during this period to disrupt the tissue. Proteinase K activity was inhibited by incubation at 95°C for 5 minutes before insoluble cellular debris was sedimented for 3 minutes at 16,000 g. The remaining genomic DNA (gDNA) was used for detection of mutation after PCR (see method below) using the primers described in **Table S2.**

PCR amplicons were tested for the presence of mutations using a T7 endonuclease I assay (New England Biolabs®, Ipswich, USA). 10 µL of each PCR was subjected to electrophoresis on a 1% agarose gel according to the protocol below. This was done to verify that non-specific fragments from PCR amplification smaller than the target amplicon (and which could interfere in the interpretation of T7 cleavage products) were not present. Following this check, samples were split into two separate reaction tubes containing 8 µL of PCR product + 1 µL 10X NEB® Buffer #2 (New England Biolabs®). These were incubated in a thermocycler with an initial denaturation of 95 °C for 5 minutes. Heteroduplexes were then formed via annealing by lowering the temperature from 95 down to 85 °C at a ramp rate of -2 °C/second, followed by a ramp rate of -0.1 °C/second from 85 °C down to 25 °C. 0.5 µL of T7 Endonuclease I (New England Biolabs®) was then mixed into one tube of each sample. All tubes were incubated at 37 °C for 20 minutes to allow cleavage to occur, before reactions were halted by cooling the tubes to 4 °C. Samples were mixed with 1/5th volume of 6X gel loading buffer (New England Biolabs®) and then electrophoresed in pairs alongside their undigested counterparts on an agarose gel as described below. An example gel image of the T7 assay for each mutant is shown in **Fig.S2**.

Sanger sequencing

Tail genomic DNA from fish determined to be homozygous (using the methods described in the **Genotyping** section below) for each of the three mutations was collected and subjected to PCR amplification as described below using primers flanking the mutation target site (**Table S2**). PCR reactions were resolved (see **Gel electrophoresis** section below) before the bands corresponding to the homozygous amplicon size were excised and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, Wisconsin, USA) according to the manufacturer's protocol. DNA

concentrations and the general quality of eluted samples were estimated on a NanoDrop® spectrophotometer (Thermo Fisher Scientific, Waltham, USA). Samples were prepared according to the recommended protocol of the Australian Genome Research Facility (AGRF, Adelaide, SA, Australia) before delivery to the AGRF. A single genotyping primer (**Table S2**) was used per sample according to which primer of a genotyping pair was more optimally located for sequencing relative to the mutation site; the Forward Primer for *sgsh*, the Forward Primer for *naglu* and the Reverse Primer for *hgsnat*. Samples were sequenced using Big Dye Terminator (BDT) chemistry version 3.1 (Applied Biosystems, Thermo Fisher Scientific).

Genotyping

Zebrafish were anaesthetised in 4 mg/mL tricaine methane sulfonate (Sigma-Aldrich®) before a 1 mm^2 piece of tail fin was sliced off with a sterile scalpel. Tissue was incubated with 100 µL of recombinant Proteinase K (Roche) diluted to ~2 mg/mL in 1 x tris-EDTA (TE) at 55 °C for 3 hours to release gDNA, with each sample being agitated 3x during this period to disrupt the tissue. Enzymatic activity was terminated at 95°C for 5 minutes before insoluble cellular debris was sedimented by centrifugation for 3 minutes at 16,000 g. Supernatants were stored at -20°C before use directly as template inputs for PCRs (see method below). PCR product(s) from each fish were resolved using gel electrophoresis (**Fig.S3**) to determine the genotype according to the method described below.

Polymerase chain reactions

PCRs were performed using primers flanking the mutation sites of *sgsh*, *naglu* and *hgsnat*. These reactions were assembled to a final volume of 20 µL using 5X Green GoTaq® Reaction Buffer and dNTPs (Promega) with laboratory-purified DNA Taq Polymerase. PCR conditions consisted of an initial 2 minute denaturation at 95°C, followed by 35 cycles of: 30 seconds denaturation at 95°C and then 30 seconds annealing and extension at 72°C. Reactions were finalised with a 5 minute extension period at 72°C before storage at -20°C for later gel electrophoresis. Primer sequences, annealing temperatures, and extension times can be found in **Table.S2**.

Gel electrophoresis

PCR products were loaded directly into ~2.5-3% w/v agarose gels in 1X Tris-acetate-EDTA (TAE) buffer (for Sanger sequencing and T7 assays) or 1X sodium borate (SB) buffer (for genotyping) with 1.5 mg/mL ethidium bromide (Sigma-Aldrich®). Samples were run alongside 5 µL of 1kb Plus DNA Ladder (New England Biolabs®) or pHAPE ladder [24] and separated at 90V (TAE) or 200V (SB) for 30-90 minutes before visualisation under UV light. All our MPS III zebrafish mutations are deletions, so mutation-site-spanning PCR amplicons from wild type fish will appear, after electrophoresis, as a

single higher molecular weight band, while amplicons from homozygous fish will appear as a single lower molecular weight band, and amplifications from heterozygous fish will display both bands.. An example of our PCR genotyping assay is found in **Fig.S3**.

Measurement of heparan sulfate (HS) in MPS III adult zebrafish brains

Adult zebrafish (3 months-of-age) were humanely euthanised in a loose ice slurry. Heads were removed by cutting at the level of the gills using sterile blades. These were then placed in petri dishes containing ice-cold 1 x PBS. Brains were carefully dissected from skulls and then placed, individually, into 500 µL aliquots of 10% methanol. The brains were then immediately homogenised on ice using a handheld pestle for two minutes. The brain homogenates were then sonicated using a Bioruptor® (Diagenode, Denville, New Jersey, USA) bath sonicator for 10 minutes on high setting with cycles of 30 seconds on, 30 seconds off. After sonication, the brain homogenates were cleared by centrifugation at 13,000 g for 5 minutes at 4°C. Total protein concentrations in the supernatants were measured using the EZQ® Protein Quantitation Kit (Molecular Probes, Inc. Eugene, Oregon, USA) following the manufacturer's protocol. 10 µg of total protein from each brain was delivered to the Mass Spectrometry Core Facility of the South Australian Health and Medical Research Institute (SAHMRI, Adelaide, SA, Australia) for measurement of disaccharide products of the butanolysis of heparan sulfate (which is directly related to the total amount of heparan sulfate) as described previously [25].

Measurement of enzyme activity in *sgsh* mutant brains

A fluorometric enzyme activity assay was performed for *sgsh*$^{+/+}$, *sgsh* $^{S387Lfs/+}$ and *sgsh* $^{S387Lfs/S387Lfs}$ sibling fish to determine the remaining levels of enzyme function. Adult zebrafish (16 months-of-age) were humanely euthanised in a loose ice slurry. Heads were removed by cutting at the level of the gills using sterile blades. These were then placed in petri dishes containing ice-cold 1 x PBS. Brains were carefully dissected from skulls and then placed, individually, into 100 µL of cold 0.2M pH 6.5 sodium acetate (NaAC). The brains were then immediately homogenised on ice using handheld pestles for two minutes (1 minutes on,30 seconds off, 1 minute on). Zebrafish brain homogenates were then sonicated using a Bioruptor® (Diagenode) bath sonicator for 10 minutes on high setting with repetitions of 30 seconds on, 30 seconds off. After sonication, the brain homogenates were cleared by centrifugation at 13,000 g for 4 minutes at 4°C. Total protein concentrations in the supernatants were measured using the EZQ® Protein Quantitation Kit (Molecular Probes, Inc.) following the manufacturer's protocol. On ice, an aliquot of each protein sample was diluted to a concentration of 10 µg in 10uL of 0.2M pH 6.5 NaAC in a separate 1.5mL screwcap tube.

In order to determine the activity level of the Sgsh enzyme, these protein samples were then incubated with 20 µL of 5 mM 4-Methylumbelliferyl N-sulpho-α-D-glucosaminide (4MU-αGlcNS) substrate (Carbosynth Limited, Berkshire, UK; cat. no. EM06602) for exactly 17 hours at 47 °C in the dark. Then, samples were removed onto ice and the reaction terminated by adding 6 µL of McIlvaine's phosphate/citrate buffer (Pi/Ci buffer). To liberate the fluorescent 4MU from the glucosamine residue, samples were incubated for exactly 24 hours at 37 °C in the dark with α-Glucosidase from *Bacillus stearothermophilus* (0.1 U in 10 µL; Sigma-Aldrich®, Castle Hill, Australia; cat. no. G3651). This second reaction was terminated by adding 100 µL of stop buffer (0.5 M Na2CO3/NaHCO3, pH 10.7). 4-methylumbelliferone (4-MU) standard (Sigma-Aldrich®; Cat# M1381) was included in 0, 355, 710, 1065 and 1420 picomole quantities to generate a standard curve for the assessment of sample activity. All samples were loaded into a flat-bottomed 96-well plate and their fluorescence measured on a SpectraMax ID Multi-Mode Microplate Reader (Molecular Devices LLC, CA, USA) at excitation wavelength 355 nm and emission wavelength 460 nm. This procedure follows that described by Whyte, Hopwood [26] and Karpova, Voznyi [27]. Statistical analysis of the data was performed using a Wilcoxon rank sum test with the *stat_compare_means* function in the package *ggplot2* [28].

Larval behaviour data collection

Two adult zebrafish heterozygous for MPS III mutations were in-crossed and the resulting progeny collected and housed together within petri dishes containing ~ 50 mL of E3 medium [22] at 28.5°C. Lighting conditions were set on a cycle of 16 hours in the light and 8 hours in the dark to simulate natural conditions. At 2 dpf, the larvae were transferred into a ~ 26 x 16 cm tank and resuspended in a fresh batch of E3 media at a depth of 3cm. At 4 dpf, larvae were placed individually into wells of clear Falcon 24-well plates (Corning, New York, United States) in 1 mL of E3 medium (**Fig.S4**).

The movement of larvae in each plate was tested within the DanioVision Observation Chamber (Noldus, Wageningen, Netherlands) with the temperature control unit set at 28.5°C. The larvae within each plate were allowed to habituate to the behaviour room 30 minutes prior to testing. The movement of these experimentally naïve larvae was recorded using the DanioVision camera at 15 frames per second and 1280×960 resolution for one hour. Behavioural tracking was performed with EthoVision XT Version 11.5.1026 (Noldus). A 11.5 mm diameter circle was added inside each individual well to represent the central exploratory zone. The time spent moving was set at a starting threshold at 2 mm/s and stopping threshold at 1 mm/s and was averaged over 13 frames for all tests. Maximum velocity was set at 16 mm/s and samples were filtered out if exceeded. Videos were manually inspected to ensure proper tracking. Then, the overall distance travelled, distance per

specified time bin, time and crossings into the central zone data was exported into a spreadsheet which contained these parameters summed over three 20 minute time bins. After behavioural data was collected, individual larvae were euthanised in ice water and genomic DNA extracted. Samples were then genotyped using PCRs and gel electrophoresis.

Analysis of larval behaviour data

All statistical analysis was performed in *RStudio* [29] using *R* [30]. To determine whether MPS III genotype had a significant effect on the total distance travelled (a marker of activity), the time spent in the centre and the frequency of crossings into the centre zone, we fitted these data to linear mixed effect models using *lme4* [31], logit-linked beta generalised linear models with *glmmTMB* [32], and log-linked negative binomial models respectively.

In all models, we specified genotype, time bin and an interaction of genotype and time bin as fixed effects, and behaviour trial and a fish identifier (to account for repeated measures) as random effects. In the total distance travelled dataset, model assumptions were violated for all MPS III subtypes. Subsequently, the distance travelled response variables were transformed by square root or log10 so that model assumptions were no longer violated. The time spent in the centre and time spent moving data were transformed into ratios of each time bin by dividing by 1200. All zero values were converted to an extremely small positive value before being fitted to logit linked beta models. The significance of the fixed effects in all models were determined with a Wald type II χ^2 test (implemented in the *car* [33] package) with a threshold of significance at p-value < 0.05. Estimated marginal means were calculated using *emmeans* [34] and plotted using *ggplot2* [28].

Zantiks Free Movement Pattern Y-maze

Data collection

We created families of simultaneous siblings with wild type, heterozygous, and homozygous genotypes by in-crossing pairs of fish heterozygous for each mutation (**Fig.3A**). Due to some difficulties with embryo collection, multiple mating events per parent pair were required to produce the ~100 individual fish required for testing. Families possessing the same mutation were raised side-by-side in tanks within the same recirculated water aquarium system.

We performed the Free-Movement-Pattern Y-maze using the Zantiks LT automated behavioural testing environment (Zantiks, Cambridge, UK) on our heterozygous in-cross families at 20 (*sgsh*^{S387Lfs}), 21 (*naglu*^{A603Efs}) and 17 (*hgsnat*^{G577Sfs}) months of age. To reduce the confounding effects of circadian rhythm differences, we only performed behavioural testing between 9am and 5pm. The testing of each family took place over the span of a single day. Following the protocol described previously

[35], fish were isolated eight at a time for 30 minutes in the behavioural testing room to acclimatise to the new room environment. Then, they were placed in the Y-mazes containing 6L of system water in the Y-maze-containing tank at an initial temperature of 27 °C, and allowed to swim freely in lit conditions for 1 hour in the absence of any external stimuli. Fish movement was tracked using the in-built Zantiks tracking system, which exports a video recording, and generates a spreadsheet containing the time points at which each fish enters and exits a zone of the Y maze (**Fig.3B**).

Behavioural data analysis

All behavioural analysis was performed using the *R* programming language [30], in the *RStudio* integrated environment [29]. Briefly, raw data files were processed using an *R* batch script developed by Fontana et al. [36] and available from GitHub (https://github.com/clay-j/ZANTIKS_YMaze_Analysis_Script). The output of this script gives the frequencies of 16 tetragrams (four consecutive arm entries, consisting of all possible left (L) or right (R) turn choice combinations from LLLL to RRRR) over six time intervals (bins) of 10 minutes each. Data visualisation was performed using the *ggplot2* package [28].

To assess any changes in the frequencies of the alternation tetragrams (LRLR and RLRL tetragrams) performed by our zebrafish in the Y-maze, we fitted the tetragram frequency data given by the batch script to a generalised linear mixed effects model using the *glmmTMB* package [32] with a logit link function. We employed the beta-binomial distribution to account for overdispersion in the dataset. Random effects were a fish identifier number (e.g. fish 1, fish 2 etc.), and the start time of the behavioural test. Fixed effects were specified to be genotype (wild type, heterozygous, or homozygous), whether a fish displayed a behavioural lateralisation (left bias, right bias or no bias), sex, time bin during the Y-maze test, and an interaction effect between genotype and time bin. To test whether these fixed effects significantly impacted the alternation probability of the fish, we used Type II Wald $\chi 2$ tests on the generalised linear model using the *Anova* function of the *car* package [33]. Fixed effects were considered to be significant if they had a p-value < 0.05. Estimated marginal means were calculated using *emmeans* [34].

To assess whether our zebrafish models showed any changes in general activity or locomotion, we also calculated the total number of arm entries (turns) performed by each fish in the Y-maze over the six time bins. This data was fitted to another generalised linear model with a negative binomial distribution with a logit link function using the *glm.nb* function of the *MASS* package [37]. Random effects were fish ID and start time, while included fixed effects were genotype, time bin, clutch, sex and an interaction effect between genotype and time bin. As above, we used Type II Wald $\chi 2$ tests

using the *Anova* function, with effects considered to be significant at a p-value < 0.05. Estimated marginal means were calculated as above.

Transcriptome analysis

Data generation

To compare brain transcriptomes in our three MPS III zebrafish lines, we generated families of 2nd Filial generation zebrafish originating from G_0 in-crosses of parental zebrafish heterozygous for two of the MPS III mutations (**Fig.4**, **Fig.S5-7**). We initially aimed to analyse n = 6 fish per genotype from single families of progeny from in-crosses of doubly heterozygous mutant fish. We chose n = 6 as, according to a previous power calculation, this would allow approximately 70% power to detect differentially expressed genes in zebrafish models of Alzheimer's disease [38]. However, due to smaller than expected clutch sizes, and questionable purified RNA quality, we were forced to use multiple families of siblings, or n<6, in some analyses. For the *A and B* (AB) arm of the experiment, we analysed n = 4 fish per genotype, which also included 4 *sgsh* $^{S387Lfs/+}$ fish to explore cellular processes affected in the brains of presumably unaffected carriers of MPS III mutations. For the *A and C* (AC) arm, we sequenced all homozygous *sgsh* ^{S387Lfs} zebrafish in the family (n = 7) to account for the lower number of fish with this genotype in the AB arm. For the *B and C* (BC) arm, we could not obtain a large enough family in a single breeding event within our timeframe to be able to compare with the other families. Therefore, we analysed fish spawned by two pairs of parents (P1 and P2, both doubly heterozygous for the *naglu*^{A603fs} and *hgsnat*^{G577Sfs} mutations) in a total of 3 spawning events. This information is summarised in **Fig.S5-7**.

Each family of fish was raised until 3 months of age, at which time the entire family was humanely euthanised in an ice slurry. Entire heads were excised at the level of the gills and were each incubated in 600 µl of RNA*later*™ Stabilization Solution (Invitrogen, Thermo Fisher Scientific, Waltham, USA) overnight at 4°C, before storage at -80°C until use. The tail of each fish was also removed for genomic DNA extraction and genotype determination by PCR as described above. Total RNAs were isolated from RNA*later*-preserved brains using the *mir*Vana™ miRNA Isolation Kit (Ambion, Life Technologies, Thermo Fisher Scientific, Waltham, USA) following the manufacturer's protocol. To remove any carried-over genomic DNA during RNA purification, we treated the total RNA samples with *DNaseI* using the DNA-*free*™ DNA Removal Kit (Invitrogen, Thermo Fisher Scientific, Waltham, USA) following the manufacturers protocol for routine DNase treatment. Then, 500 ng of high-quality total RNA (with RIN_e generally ≥ 8) was delivered on dry ice to the South Australian Genomics Centre (SAGC, Adelaide, Australia) for preparation of stranded, polyA+ libraries using Nugen Universal Plus mRNA-seq (NuGEN, Ltd., UK). Cluster generation was performed using the Illumina to MGI Library Conversion (MGIEasy Universal Library Conversion Kit, Part No.

MGI1000004155). Then, 2 x 98bp paired end sequencing was performed including 8 bp unique molecular identifiers (UMIs) using MGI DNBSEQ-G400 chemistry (MGI Tech, Shenzhen, China). Each of the libraries was sequenced over multiple lanes and the data from each library subsequently combined.

Pre-processing

Pre-processing was performed using a custom pipeline implemented in *snakemake* [39]. Briefly, the raw fastq files were subjected to initial quality checks using *fastQC* version 0.11.9 [40]. Then, adaptors were trimmed and low-quality reads removed using *fastp* version 0.23.1 [41]. The remaining reads were aligned to the zebrafish genomes e (GRCz11, Ensembl release 101 [42]) using the splice-aware alignment software *STAR* version 2.7.0d [43]. Aligned reads associated with the same unique molecular identifier (UMI) correspond to PCR duplicates generated using library preparation. These were deduplicated using the *dedup* function of *umi_tools* v1.0.1 [44]. A gene-level counts matrix was then generated using *featureCounts* version 2.0.1 [45].

Analysis

Statistical analysis of the RNA-seq data was performed using *R [30]*. Lowly expressed genes are considered uninformative for differential expression analysis. Therefore, for each experimental comparison, we established a counts per million threshold value that equated to 10 counts in the sample with smallest library size. Genes were removed if their expression was lower than this value in at least the number of samples equalling the smallest genotype group (as recommended in [46]). The remaining genes (i.e. above the detection threshold) were normalised using the trimmed mean of M-values (TMM) method [47]. Principal component analysis was performed to assess whether batch effects (such as “Experimental arm”, RIN_e or “Home Tank”) showed significant influence on brain transcriptomes. The factors of “Experimental Arm” and “Home tank” were found to show effects on the brain transcriptome (**Fig.S8-11**). Therefore, each experimental arm was analysed separately, and the “Home Tank” was included in the design matrices for differential expression analyses. Differentially expressed (DE) genes were identified using generalised linear models and likelihood ratio tests in *edgeR* [48]. We considered a gene to be differentially expressed (DE) if the FDR-adjusted p-value was less than 0.05.

In initial differential gene expression analyses, we noted some small biases for genes to be called DE due to their %GC content and length (**Fig.S12-14**). Therefore, we used the conditional quantile normalisation (CQN, [49]) method to account for this bias. For CQN, we used the average length (in base pairs) of transcripts per gene, and a weighted average (by transcript length) of the %GC content

of detected transcripts per gene to generate an offset term which was included in DE analyses in *edgeR*.

For gene set analyses, we used the KEGG [50] and gene ontology [51] gene sets obtained from the Molecular Signatures Database [52] using *msigdbr* [53]. The KEGG and GO term gene sets were filtered to retain only genes detected in each arm of the experiment. The GO terms themselves were also filtered so that only GO terms with more than 3 steps to the root node were retained (to reduce redundancy). The KEGG gene sets were additionally filtered to retain only gene sets containing more than 5 genes. To test for changes to iron homeostasis, we used the lists of genes identified to contain iron-responsive elements (IREs) in the untranslated regions of their encoded transcripts from [54]. To test for changes in cell type proportions, we used the marker genes of adult zebrafish brain cell types (23 different cell types) from a single-cell RNA-seq dataset [55]. We tested whether these gene sets were enriched with DE genes using *goseq* [56] specifying the co-variate as the average transcript length per gene. We also utilised the fast implementation of the *ROAST* algorithm: *fry* [57], and the *GSEA* algorithm: *fgsea* [58]. We calculated a directional ranking statistic for *fgsea* using the sign of the $\log_2$FC (logFC, -1 or 1) multiplied by the $-\log_{10}$ of the inverse of the p-value (sign(logFC) * -log10(1/p-value)). This ranked list of genes was then used as input for *fgseaMultiLevel*().

Post-hoc power analysis was performed using the *ssizeRNA_vary*() function of the *ssizeRNA* package [59]. The calculations were performed at the controlled false discovery rate of 0.05, with a pseudo sample size of 40 and under the assumption that 90% of the genes were not differentially expressed. The mean gene counts and dispersions in the control group were set from those calculated in each individual arm of the experiment.

Data and code availability statement

The RNA-seq datasets generated in this study will be deposited in the GEO database upon publication. All code to reproduce the transcriptome analysis and the behaviour analysis will be uploaded to GitHub.com upon publication.

Results

Creation of zebrafish models of MPS IIIA, B and C

Severe disease mutations disrupting translational reading frames are known to exist in the most downstream exons of the human *SGSH* [60], *NAGLU* [60], and *HGSNAT* [61] genes. We sought to identify similarly disruptive frameshift mutations in the respective orthologous zebrafish genes, *sgsh*, *naglu*, and *hgsnat*, represented in **Fig.S1**. The advantage of identifying frameshift mutations leading to stop codons in the final exons of these genes is that such mutations are less likely to cause nonsense-mediated decay (NMD) of mutant mRNAs and so may avoid the transcriptional adaptation phenomenon [62] that can complicate the interpretation of mutation phenotypes. Mutagenesis was carried out using micro-injection of CRISPR-Cas and CRISPR-Cpf1 systems into wild type embryos at the single-cell stage. Ultimately, we isolated the mutations $sgsh^{S387Lfs}$, $naglu^{A603Efs}$, and $hgsnat^{G577Sfs}$ (**Fig.1A-C**), all of which cause frameshifts upstream of the equivalent positions of known severe human mutations, and so presumably are similarly detrimental to the enzyme activities of the genes' protein products.

Zebrafish models of MPS IIIA, B and C show biochemical characteristics consistent with the human disease

We next aimed to confirm that our MPS III mutations were reducing the functions of the enzymes encoded by their respective genes. Therefore, we measured heparan sulfate (HS) accumulation in the brains of 3-month-old zebrafish by mass spectroscopy. Each mutation was analysed by in-crossing two heterozygous fish and then examining the levels of butanolysis disaccharide product of HS (directly proportional to HS [63]) in the brains of their wild type, heterozygous, and homozygous progeny (n = 3 per genotype). Similar to recessive human MPS III, the wild type and heterozygous zebrafish brains showed only low basal levels of HS while homozygous mutants showed HS levels many fold higher (**Fig.1D**) indicative of decreased activity of enzymes degrading HS.

To investigate further the loss of enzyme activity of the mutant Sgsh protein in our MPS IIIA model, we performed a fluorometric enzyme activity assay to estimate the levels of sulfamidase enzyme activity in 16-month old brains from the progeny of a single mating of two fish heterozygous for the $sgsh^{S387Lfs}$ mutation. Considerable variation in measured enzyme activity was observed between individuals of the same genotype, but a linear decline was observed between the mean activities of the wild type, heterozygous, and homozygous brains (**Fig.1E**). The difference from the heterozygous group to both the wild type and homozygous group was marginally insignificant ($p = 0.052$), likely

due to insufficient statistical power. However we observed a significant difference in enzyme activity between the wild type and homozygous mutant brains ($p = 0.016$), consistent with an activity of the mutant enzyme (**Fig.1E**) at approximately 30% of wild type levels.

Disease-specific behavioural changes are not observed in 5 day old MPS III larvae

Children with MPS III show severe hyperactivity early in disease progression [64]. To determine whether our MPS III zebrafish larvae also show this disease characteristic, we assessed the total distance they travelled in one hour in an open field. Overall, zebrafish of each genotype travelled less with time (i.e., increasing bins), possibly reflecting habituative behaviour. We did not observe any statistically significant effect of genotype with time bin in any MPS III model at 5 dpf (**Fig.2A–C**, raw data visualised in **Fig.S15**).

To investigate the characteristics of movement, which may differ between mutant genotypes (e.g., whether mutant larvae travel in shorter bursts of movement compared to wild type, whilst travelling a similar distance overall) we inspected the ratio of time spent moving within each time bin. As expected, there was a general decrease in time spent moving over each successive time bin (**Fig.2D-F**), however, no statistically significant difference was observed by genotype. On this basis, MPS III larvae do not exhibit hyperactivity at 5 dpf.

The MPS IIIA zebrafish model of Douek et al. [20] was found to exhibit less thigmotactic behaviour, interpreted as reduced anxiety [65]. Our independent MPS IIIA zebrafish model showed increased thigmotactic behaviour (i.e. less time in the central zone of the well), particularly in the first 40 minutes of the test (**Fig.2G**). No mutant-specific significant differences were observed for MPS IIIB and MPS IIIC (**Fig.2H,I**). Another measurement of thigmotactic behaviour is the number of crossings into the central zone. Modelled data demonstrated no significant effect of genotype. Therefore, it is unlikely that differences in anxiety-related behaviour emerge by 5 dpf in any of the three MPS III models.

Zebrafish models of MPS IIIA and B show hyperactivity in adulthood

The changes in brain molecular state in children affected by MPS III are reflected in changes in cognition and behaviour, so we sought to see whether similar changes could be observed in our mutant fish. The free movement pattern (FMP) Y-maze test [35] can be used to monitor short term spatial memory by measurement of left and right turn choice alternation (represented in **Fig.3B**). It can also be used to compare overall motion activity between fish. Behavioural testing is typically subject to large variation in individual behaviour, but genetic and environmental influences on the behaviour of individual zebrafish can be reduced by examining families of zebrafish siblings raised in

similar environments. For tests of the behavioural effects of each of our three MPS III mutations, we raised multiple families of progeny from spawnings of single pairs of heterozygous mutant fish (**Fig.3A**) (one pair produced all the progeny families for any particular mutation). Thus, for each mutation examined, genetic variation was reduced by examining siblings, with minor temporal variation between siblings in terms of spawning date. Each zebrafish family was composed of wild type, heterozygous, and homozygous siblings (**Fig.3A, Fig.S16**), however their genotypes could only be distinguished by PCR tests on tail clip biopsies taken after behavioural testing was complete. An advantage of this experimental constraint is that behavioural data collection was blinded, thereby avoiding possible observer bias.

FMP Y-maze behavioural data from each individual fish was collected for one hour and was subsequently analysed in 10 minute time bins to account for changes in behaviour during testing as the fish became more familiar with their environment. A generalised linear mixed effects model with a beta-binomial distribution was used for the statistical analysis. We tested for fixed effects including genotype, left-right bias, sex, time bin, and an interaction effect between genotype and time bin. We found no significant effects of genotype observed on the estimated probability of MPS III fish performing left and right turn choice alternation during the y-maze test (**Fig.3C**). Therefore, there is no evidence supporting altered spatial working memory in these fish. However, we did observe a higher than expected rate of failure to alternate turn choice in our tests that may be indicative of higher than usual stress levels (**Fig.S17**). This observation was not affected by fish genotype.

To assess whether our three zebrafish MPS III models might show the hyperactivity characteristic of the human condition, we examined the number of turns made by fish of wild type, heterozygous, or homozygous genotype during one hour in the free movement Y-maze. The raw total turns data (**Fig.S16**) from 10 minute time bins across the hour (**Fig.S18**) was analysed using a generalised linear mixed effects model with a negative-binomial distribution and tested for fixed effects including genotype, time bin, clutch/tank, sex and an interaction effect between genotype and time bin. We found that the effect of genotype on the number of turns predicted by the model was significant for *sgsh* and *naglu* mutant families (**Fig.3D, Table S3**). For these MPS III model zebrafish families, we observed an increase in predicted average turns per bin from wild type (lowest) to heterozygous to homozygous fish (highest). This indicated that homozygous $sgsh^{S387Lfs}$ and homozygous $naglu^{A603Efs}$ mutant zebrafish showed a hyperactivity phenotype at 20 and 21 months of age respectively, and that zebrafish heterozygous for these mutations may display an attenuated effect. $hgsnat^{G577Sfs}$ mutant zebrafish did not show statistically significant changes in activity during this experiment, however a subtle trend towards hyperactivity was apparent most similar to that observed in the

naglu^{A603Efs} family. Next, to test whether this hyperactivity was consistent across the duration of the test, we visualised the model-predicted data over the course of the six 10-minute time bins (**Fig.S19**). We saw that the predicted turns decreased over the course of the test, with the effect of time bin on predicted turns found to be highly significant (**Table S3**). This decrease in activity over the course of the test is not surprising given we expect zebrafish to habituate to a changed environment over time. We saw from our genotype:bin interaction effects (**Table S3**) that this decrease in predicted turns over the time bins was not significantly altered by the genotype of the fish for the *naglu* or *hgsnat* fish. However, in homozygous *sgsh*^{S387Lfs} fish, the genotype:bin interaction effect was significant, and the decrease in turns during the test was reduced in these fish compared to their wild type and heterozygous siblings (**Fig.S19**). This suggests that homozygous *sgsh*^{S387Lfs} fish show impaired habituation to this Y-maze test. Interestingly, we also observed that the model-predicted turns for heterozygous *naglu*^{A603Efs} and heterozygous *hgsnat*^{G577Sfs} fish were closer to those of their homozygous siblings than to their wild type siblings wild type (**Fig.3D**) and this was reflected across the time bins (**Fig.S19**).

Brain transcriptome analysis of zebrafish models of MPS IIIA, B and C

One motivation behind construction of our zebrafish models of MPS IIIA, B, & C mutations was to make an unbiased survey of the effects of these mutations on the brain's molecular biology – and to identify possible differences between these types of MPS III. Transcriptome analysis is currently the single technique giving the greatest volume of molecular phenotypic data from analysis of cells and tissues. Therefore, we sought to apply transcriptome analysis to characterisation of the molecular state of brains from our models. We enhanced the sensitivity of this analysis to detect differences between MPS III types by applying our intra-family strategy of genetic and environmental noise reduction (**Fig.4**). The intra-family analysis strategy also facilitated an experimental structure incorporating internal replication; we bred families allowing comparisons of MPS IIIA & B to wild type, MPS IIIA & C to wild type, and MPS IIIB & C to wild type (the "AB", "AC", and "BC" experimental arms respectively). We analysed brains at 3 months of age (around the time at which fish are becoming sexually mature and so equivalent to early teenage in humans) as we wished to detect early effects of the mutations before possible disturbance of brain cell type number. We had already detected hyperactivity at later ages (above) and our previous transcriptome analysis of familial Alzheimer's disease-like mutations (expected to have much milder phenotypes than those causing MPS III) at 6 months of age had detected significant differences [66].

Principal component analysis (PCA) performed on RNA-seq data from each of the experimental arms revealed the MPS IIIA zebrafish brain transcriptomes as, apparently, the most perturbed among our

three MPS III models, as these generally formed distinct clusters from their wild type siblings (particularly in the family comparing MPS IIIA and IIIC mutations, **Fig.5A-C**). Consistent with this observation, the number of differentially expressed (DE) genes detected in each mutant in each family was highest for the MPS IIIA mutants (**Fig.5D-I, Supplemental Data File 2**). Note, however, that subsequent analysis revealed our analyses of MPS IIIB brain transcriptomes to be statistically underpowered (see below).

Gene set enrichment analysis reveals both common and distinct changes to KEGG pathways in MPS III zebrafish brains

To identify alterations of cellular function implied by the differential expression of genes between genotypes we tested for over-represented gene ontology terms using *goseq* (using the average transcript length per gene as a covariate). We detected significant over-representation of GO terms related to oligodendrocyte function such as myelination and fatty acid metabolism in MPS IIIA zebrafish (and not MPS IIIC) in the arm of the experiment comparing MPS IIIA and MPS IIIC (**Fig.S20**). This is explored further below. In all other DE gene lists, the significantly over-represented GO terms were only driven by one or two genes (**Supplemental Data File 3**). For example, the DE genes in MPS IIIC zebrafish brains in the BC arm were significantly over-represented with the GO term TIGHT_JUNCTION_ORGANIZATION (p_{FDR} = 0.035). However, this GO term has 54 genes which were detected in this experiment, and only two of them were identified to be differentially expressed. Therefore, the biological relevance of these significantly over-represented GO terms is questionable.

Over-representation analysis using goseq relies on a hard threshold for the classification of DE genes (a FDR-adjusted p-value < 0.05 in *edgeR*). To overcome this limitation, we also performed another type of enrichment analysis, *ROAST* [57] implemented in *fry* from the *limma* package. *Fry* does not assume a cut-off for a single gene to be called DE, and uses rotation as an alternative to permutation to generate levels of statistical significance [57]. Performing *fry* on the KEGG gene sets revealed both common and distinct changes to gene expression across the MPS III models in each of the experimental arms. The KEGG gene sets significantly altered in at least one of the comparisons of an MPS III genotype to wild type are summarised in **Fig.6**.

Significant changes to gene expression in the KEGG gene sets could not be detected in the BC arm of the experiment. This is likely due to the low number of biological replicates within each clutch of fish analysed in this arm. In the AB arm, we could only detect significant changes to gene expression in the LYSOSOME gene set. This is the gene set which is the most markedly changed in both genotypes of MPS III zebrafish brains in the AC arm. This suggests that, in the AB arm, the analysis of n = 4 has only allowed detection of the most extreme changes to gene expression. Consistent with this, post-

hoc power calculations performed using ssizeRNA [59] indicated that n = 4 achieved approximately 20% power to detect DE genes (**Fig.S21**). Since the AC arm had the most statistical power to detect DE genes, hereafter we focus our attention on this comparison.

Reassuringly, the KEGG gene sets altered in both MPS IIIA and MPS IIIC zebrafish in the AC arm include LYSOSOME, GLYCOSAMINOGLYCAN_DEGRADATION, and OTHER_GLYCAN_DEGRADATION (**Fig.6**). The significant changes to gene expression identified within these KEGG pathways are consistent with heparan sulfate accumulation in lysosomes as the primary disease mechanism in MPS III, and with the accumulation of monosialic gangliosides (GM_2 AND GM_3) due to the HS accumulation (reviewed in [67]). The LYSOSOME gene set was also significantly altered in MPS IIIA and MPS IIIB brains in the AB comparison. This supports that the changes to gene expression in the LYSOSOME gene set are extensive and can be detected with lower statistical power. The GLYCOSAMINOGLYCAN_DEGRADATION and OTHER_GLYCAN_DEGRADATION gene sets did not reach statistical significance in the AB and BC comparisons. These gene sets were, in general, upregulated in all MPS III comparisons (**Fig.7A-C**). To determine whether the LYSOSOME, GLYCOSAMINOGLYCAN_DEGRADATION, and OTHER_GLYCAN_DEGRADATION pathways were altered in a similar manner in MPS IIIA and MPS IIIC zebrafish brains, we calculated Pearson correlations between the $\log_2$FC (logFC) values within each gene set. We found that the correlations between the logFC values were generally higher within the experimental arms of the experiment (i.e. between sibling zebrafish), rather than within the same subtype of MPS III (not siblings). However, whether this is due to true biological differences between the families of zebrafish, or due to the differing sample cohort structures cannot be discerned. The Pearson correlations of logFC values for the genes within these gene sets were found to be higher between sibling MPS IIIA and MPS IIIC zebrafish within the AC arm, compared to correlations observed between the zebrafish of the same MPS III subtype between different arms (**Fig.7D-L**). Notably, the correlations were lower in the GLYCOSAMINOGLYCAN_DEGRADATION gene set, which is consistent with the loss of different enzymes in these subtypes, resulting in the accumulation of a different HS degradation products.

Other KEGG gene sets which were significantly altered in both MPS IIIA and MPS IIIC mutants in the AC arm included pathways involving the immune system (COMPLEMENT_AND_COAGULATION_CASCADES and ANTIGEN_PROCESSING_AND_PRESENTATION), and metabolism (AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM). Similar to the lysosomal gene sets described above, the genes in these gene sets generally showed positive correlations, and the correlations were stronger between siblings than within MPS III subtypes (**Fig.S22-24**).

There were more gene sets significantly altered only in MPS IIIA (15 gene sets) compared to only in MPS IIIC (1 gene set), consistent with MPS IIIA generally being a more severe subtype than MPS IIIC in humans (reviewed in [67]). These gene sets spanned metabolism, the immune system, cellular growth and adhesion, and vesicular transport.

The one gene set significantly altered in MPS IIIC mutants and not in MPS IIIA mutants was FATTY_ACID_METABOLISM. Four downregulated genes appear to drive the statistical significance of this gene set in MPS IIIC zebrafish brains (**Fig.S25**). These genes are acyl-CoA synthetase long chain family member 1b (*acsl1b*), hydroxyacyl-coenzyme A dehydrogenase (*hadh*), alcohol dehydrogenase 8b (*adh8b*) and enoyl-CoA, hydratase/3-hydroxyacyl CoA dehydrogenase (*ehhadh*). Measurements of fatty acid concentrations in MPS IIIC zebrafish brains will be necessary to understand why transcript levels from these genes might be downregulated.

Changes to gene expression within neural stem cells, oligodendrocytes and microglia

The brain is a complex organ made of up many different cell types. If there were changes to the proportions of these cell types within the samples generated for bulk brain total RNA, this could lead to artefactual observations of changes in gene expression unrelated to impaired cellular function. Rather, the transcriptome changes would be reflective of the lack or increase of that cell type within the bulk brain tissue samples. To assess this possibility in our dataset, we utilised a single-cell RNA-seq dataset [55] to obtain gene sets listing the genes expressed in 23 different cell types found in adult zebrafish brains. Using *fry* with a directional hypothesis, we found statistical evidence for overall reductions to gene expression in marker gene sets of oligodendrocytes, neural stem cells and in microglia with high *apoc1* expression (**Fig.8,8A-D**). These 3 gene sets contain a substantial number of shared genes (**Fig.S26**), and the possibility arises that the statistical significance of the gene set could be driven by the same genes. To assess this, we inspected the “leading-edge” genes from the fgsea [58] algorithm (**Fig.9E,F**). The leading-edge genes can be interpreted as the genes which drive the statistical significance of the gene set, and so can be used to determine whether the same genes are driving the enrichment of the gene sets. In the MPS IIIA zebrafish brain transcriptomes, the neural stem cell gene set contained the most unique leading-edge genes (32 out a total 39 leading edge genes). The oligodendrocytes and microglia gene sets shared almost all of their leading-edge genes (only 7 unique leading-edge genes were found to drive the statistical significance of the oligodendrocyte gene set). A similar trend is observed in the MPS IIIC zebrafish. From this RNA-seq dataset, we cannot say with confidence whether there are changes to cell type proportions in these bulk brain samples, or whether these cell types are dysfunctional. Further analysis will be required to clarify this issue.

Evidence supporting iron dyshomeostasis in young MPS III zebrafish brains

Changes to iron homeostasis are observed in many neurodegenerative diseases (reviewed in [68]). Iron accumulation has been observed in a human MPS IIIB case [69] and a mouse model of MPS IIIB [70]. However, the broader role of iron homeostasis in all MPS III subtypes is poorly understood. To explore whether iron dyshomeostasis might occur in our zebrafish models of MPS III, we performed an enrichment analysis on our "IRE" gene sets [54]. These gene sets consist of the genes in zebrafish encoding transcripts containing iron-responsive elements (IREs) in their 5' or 3' untranslated regions (UTRs). These gene sets also differentiate between those genes which encode consensus IRE sequences (those which have "high-quality" (ire_hq) IREs), and those with near-consensus sequences previously shown or predicted to function as IREs [71, 72], the "ire_all" gene set. In the AC arm, we identified significant changes to the expression of all IRE gene sets in MPS IIIA zebrafish but not MPS IIIC (**Fig.10A**). The changes to the IRE genes were not replicated in the other MPS IIIA zebrafish (in the AB arm), possibly due to the reduced statistical power in that experimental arm. Intriguingly, significant changes to the expression of genes containing 5' IREs were detected in the MPS IIIB zebrafish in the BC arm, despite the lower statistical power in this experimental arm. This was not replicated in the MPS IIIB zebrafish in the AB arm, suggesting that this observation may be specific to that experiment.

Hin et al.'s study of changes in IRE-containing transcript levels under iron deficient and iron excess conditions showed that the particular direction of change at any transcript level was not a general consequence of any IRE type (5' or 3' UTR). Therefore, the IRE changes observed cannot be used to impute iron deficiency or excess in the brains of the zebrafish MPS IIIA or IIIB models.

No evidence that heterozygosity for loss-of-function mutations in *sgsh* has effects on cellular functions in the brain

Single, heterozygous mutations in genes associated with other lysosomal storage disorders have been reported to be implicated in adult-onset neurodegenerative diseases [73-75]. Recently, Douglass et al. [76] identified subtle changes to motor coordination and brain pathology due to heterozygosity for a loss-of-function mutation in *Sgsh*. To investigate the molecular consequences of heterozygosity for a hypomorphic mutation in zebrafish *sgsh,* we analysed the brain transcriptomes of heterozygous *sgsh* ^{S387Lfs} fish in the AB arm. A PCA performed on the brain transcriptome data from the AB arm (and omitting the MPS IIIB zebrafish for clarity) revealed that the heterozygous *sgsh* ^{S387Lfs} fish appear to cluster with the wild type, rather than with MPS IIIA fish (**Fig.S28**), supporting that the heterozygous *sgsh* ^{S387Lfs} fish brain transcriptomes more closely represent a wild type state. Consistent with this observation, only 14 DE genes were detected in heterozygous *sgsh*^{S387Lfs} fish

relative to wild type (**Fig.5D,G, Supplemental Data File 2**) and these were not significantly over-represented with any GO terms (**Supplemental Data File 3**). We also did not detect any significant changes to gene expression in the KEGG, IRE or cell type marker gene sets (**Supplemental Data Files 4, 5 and 6**). However, a significant enrichment of genes on chromosome 22 (which contains *sgsh*) was detected (fry p_{FDR} = 2.71e^{-10}, **Fig.11**). These chromosomally co-located differentially expressed genes (CC-DEGs) may reflect artefactual differential expression due to enrichment for linked expression quantitative trait loci (eQTLs), a phenomenon that can occur due to the breeding structure used to generate the families analysed [77, 78]. Together, these observations do not support that zebrafish *sgsh* mutation carriers show significant disturbances of brain function at 3 months of age.

Discussion

We have described the creation and initial characterisation of zebrafish models of MPS IIIA, B and C. In all three models we see highly significant accumulation of HS in the brains of 3-month-old fish. Our models thereby display the primary metabolic signature of the human disease. The absence of any appreciable HS accumulation in zebrafish heterozygous for our mutations is also consistent with the recessive nature of MPS III mutations in humans.

The assay of *sgsh*-encoded enzyme activity indicated that *sgsh*^{S387Lfs} homozygous fish do not display complete loss of sulfamidase function. Our results suggest that approximately 30% of the enzyme function may be retained in these homozygous fish. However, this is insufficient to prevent gross accumulation of HS in the brain. This contrasts with the effects of the *sgsh*$^{\Delta ex5-6}$ mutation generated by Douek et al. that deletes much of two exons and their separating intron and causes complete loss of sulfamidase enzyme function. The *sgsh*^{S387Lfs} mutation deletes only 16 nucleotides in the most downstream exon 7 of *sgsh*, but is still expected to display a severe phenotype due to its similarity in position to numerous other known severe nonsense and frameshift mutations in *SGSH* [60].

While we have not yet measured the enzyme activities remaining for our mutations of *naglu* or *hgsnat*, we anticipate that these will represent a greater loss of function than seen for our mutation of *sgsh*. In the case of *naglu* this prediction is supported by a greater accumulation of HS seen in *naglu*^{A603Efs} homozygous brains than in *sgsh*^{S387Lfs} homozygous brains. The *hgsnat*^{G577Sfs} mutation caused the least HS accumulation of all the mutations, and showed milder changes to gene set expression and behaviour than for *sgsh*^{S387Lfs}. However, it is highly likely that this frameshift mutation represents a complete loss of function since the encoded protein is predicted to lack almost the entirety of the two most C-terminal transmembrane domains. Interestingly, in the HS degradative

pathway, the acetyl CoA:α-glucosaminide-acetyltransferase enzyme activity encoded by *hgsnat* acts prior to the α-N-acetyl-D-glucosaminidase enzyme activity encoded by *naglu*. Presumably, the loss of acetyl CoA:α-glucosaminide-acetyltransferase activity from *hgsnat* is partially compensated for by the presence of a related enzyme activity or an alternative degradative pathway (at least in zebrafish). The relatively mild phenotype of the *hgsnat*^{G577Sfs} mutation relative to *sgsh*^{S387Lfs} in our study is consistent with the observation that cases of MPS IIIA tend to show greater severity than those of the other MPS III subtypes [4, 79].

Zebrafish have great utility for behavioural studies due to their high rates of reproduction and short embryo development time. We employed intra-family behavioural analysis on our zebrafish models, since this is the best approach for reducing potentially confounding genetic and environmental effects. We also genotyped these fish after data collection. This has the benefits of preventing the results from being affected by genotyping procedure-imposed stress and of blinding the observer of the experiment. Our goal was to identify cognitive differences in our MPS III homozygous zebrafish. Cognitive functions, including memory deficits, can be difficult to assess accurately clinically in MPS III [3]. Children living with Sanfilippo syndrome are known to exhibit alterations in memory. However, to our knowledge, spatial working memory specifically has not been formally assessed in children with MPS III. Previously, spatial working memory differences have been observed in a male MPS IIIB mouse model [80]. We tested spatial working memory in our three zebrafish MPS III models as a representation of broader memory phenotypes but saw no significant differences between genotypes at the ages studied. Rather than reflecting any insensitivity of the Y-maze test to changes in spatial working memory in zebrafish compared to when these tests are conducted with mice and humans [35], the reason may lie in the remarkable neural regenerative abilities of zebrafish that contrast with those of mice and humans (reviewed by Zambusi and Ninkovic [81]). Although we did not observe genotype-based differences in alternation, we did observe higher than normal proportions of repetition behaviour and lower than normal proportions of alternation across each MPS III zebrafish family. Previous work in this field has demonstrated that zebrafish, like humans and mice, employ an alternation-dominant search strategy in the Y-maze [35, 36]. However, increased stress has been shown disrupt performance of zebrafish in similar spatial memory tests [82]. This suggests that our zebrafish may have been under additional stress due to unavoidable external stimuli occurring during data collection, emphasising the need for follow-up experiments.

Hyperactivity is a well-established behavioural phenotype commonly displayed by children living with MPS III [1, 83]. We observed a significant hyperactivity phenotype in our aged MPS III *sgsh* and *naglu* model zebrafish but not in 5 dpf larvae. This is consistent with hyperactivity observed in MPS IIIA [84, 85] and MPS IIIB [86] at widely varied ages, although other mouse studies have produced conflicting

results potentially related to environmental conditions and genetic background variation [85, 87, 88]. While hyperactivity has been observed in MPS IIIC mice [89], we did not observe statistically significant hyperactivity in our *hgsnat* zebrafish, again consistent with the lower severity of MPS IIIC in humans. In humans with MPS III, hyperactivity is an early behavioural change preceding severe cognitive deficits [1] and so may be due to changes in, for example, synaptic activity before neurodegenerative cell loss [90-92]. Therefore, in zebrafish models of MPS III with considerable regenerative ability, we might not expect to see changes in those behaviours due to cell loss, but could still observe them due to altered neuronal activity in largely intact brains. Therefore, additional testing of our MPS III models is warranted using behavioural tests for phenotypes not due to neurodegeneration. Differences in anxiety-related behaviours are common in human patients [1, 2], and have been observed in MPS IIIA [87], MPS IIIB [80, 93] and MPS IIIC [89] mice. These behaviours can be assessed robustly in larval [94] and adult [95] zebrafish and can be applied to our models in the future.

We implemented a three-arm construction for our transcriptome analysis which, optimally, would enable direct comparison of each subtype to each other, as well as internal replication in comparing each subtype to wild type siblings twice. Unfortunately, this was largely unsuccessful due to the lower than normal breeding capacity of our zebrafish during the generation of the families for analysis. Despite insufficient statistical power in multiple arms of our experiment, the PCA results demonstrated the importance of our intra-family approach, as zebrafish brain transcriptomes clustered very differently between parent pairs.

Despite these limitations, in the subtype AC comparison family we were able to detect differences in gene set enrichments in lysosomal and glycosaminoglycan degradation pathways, which are expected in the context of primary MPS III biology. The difference in fold and significance between the changes we found in MPS IIIA zebrafish compared to MPS IIIC zebrafish clearly demonstrate the greater severity of the mutation in *sgsh* compared to the mutation in *hgsnat*. This is consistent with the greater severity of MPS IIIA onset and progression in humans compared to MPS IIIC [2, 5].

The gene set analysis should be interpreted with a degree of caution given our observation of a differential signal in the oligodendrocyte cell gene set. This observation was consistent across all of our three MPS III models, suggesting that oligodendrocytes may be one of the earliest and/or most severely affected cell types in MPS III.

A recent preprint demonstrated a variety of oligodendrocyte-related effects in the brains of an MPS IIIC mouse model [96]. Specifically, these researchers found substantially reduced numbers of

oligodendrocytes and substantially reduced myelin thickness in their mouse and human patient MPS IIIC samples.

The apparent sensitivity of oligodendrocytes in our MPS III models is biologically intuitive as the myelin sheathes of these cells give them the greatest surface area of any neural cell type and this is likely accompanied by the greatest burden of extracellular matrix (ECM) maintenance and turnover. This implies that oligodendrocytes would likely be more sensitive to deficits in degradation of HS (that is found abundantly in the ECM) than other cell types and would be the first to show pathological effects from mutations causing MPS III. Notably, oligodendrocytes are also the cell type showing the greatest accumulation of iron in the brain [97, 98] and, if HS accumulation interferes with iron homeostasis (as our IRE analysis supported, see below), then oligodendrocytes would be expected to be particularly sensitive to this.

The acidic internal environment of the lysosome is required for the reduction of the biologically inactive ferric form of iron (Fe^{3+}) to the active ferrous form of iron (Fe^{2+}) [99, 100]. When HS accumulates in lysosomes, it may upset the correct acidification of these compartments and interfere with the proper conversion of ferric to ferrous iron. This would have significant implications for cells given the central role of ferrous iron in energy metabolism and other crucial molecular processes. With this in mind, we searched for changes in iron homeostasis in our MPS III zebrafish using our previously developed IRE analysis approach (Hin et al. 2021).Significant IRE transcriptome signals were predominantly associated with the MPS IIIA fish. This is not surprising given that MPS IIIA is the most severe subtype in humans. MPS IIIB is the next most severely affected subtype, however we lacked statistical power in the AB and BC transcriptome experiment arms which likely explains the relative absence of statistically significant perturbations in expression of transcripts with these IREs in MPS IIIB zebrafish brains.

Conclusions

MPS III is devastating for children living with the disease and their families. With limited therapeutic options available, animal models are a crucial part of efforts to understand and treat the underlying pathologies. In this study, we successfully generated novel zebrafish models of MPS IIIA, B and C by creating the hypomorphic mutations *sgsh*^{S387Lfs}, *naglu*^{A603Efs} and *hgsnat*^{G577Sfs}. To our knowledge, our MPS IIIB *naglu* and MPS IIIC *hgsnat* models are the first published in zebrafish. Importantly, these models feature significant heparan sulfate accumulation in the brain, displaying the primary signature of the disease. We found that, in contrast to the already published zebrafish mutation, *sgsh*$^{\Delta ex5-6}$, our *sgsh*^{S387Lfs} mutation appears to retain some enzyme function, a feature which could be useful in future investigations of this heterogeneous disease. While zebrafish intra-family behavioural analysis did not provide evidence of effects on adult memory or larval behaviour, two of our models displayed statistically significant hyperactivity in the homozygous state at later ages, a phenotype in-common with MPS III in humans. Intra-family RNA-seq transcriptome analysis revealed expected changes in gene expression related to lysosomal and glycan-degradative biology and the immune system in MPS III zebrafish brains. Interestingly, we also observed evidence of iron dyshomeostasis in MPS IIIA zebrafish and changes in oligodendrocyte cell state in all three mutants, implying that this cell type may be particularly affected in MPS III. Ultimately, our investigations of the behavioural and molecular features of our MPS III zebrafish show concordance both with existing mouse models and the human condition. Further investigation is required to deepen our understanding of the molecular pathways involved and for replication of our findings. We envisage that the *sgsh*^{S387Lfs}, *naglu*^{A603Efs} and *hgsnat*^{G577Sfs} mutant zebrafish will be valuable platforms for discovery and testing of novel MPS IIII therapeutic approaches.

Acknowledgements/Funding statement

The authors of this paper would like to express their sincere appreciation to several individuals who made valuable contributions to this project. We thank Dr. Morgan Newman and Rebel Bertram for their assistance with the initial screening for generation of the mutants, Barb King for her invaluable assistance with fluorometric enzyme assays, Dr Marten Snel and Dr Paul Trim (SAHMRI) for performing the heparan sulfate measurements, Dr Matt Parker for advice regarding behaviour analysis, the South Australian Genomics Centre (SAGC) for performing the library preparation and RNA-sequencing, Angel Allen for help with figure production, and Lachlan Baer for help with the *snakemake* pipeline.

This work was supported by funding from the Sanfilippo Children's Foundation. KB is supported by a Race Against Dementia – Dementia Australia Research Foundation Postdoctoral Fellowship and funds from Flinders University. EG is supported by an Australian Government Research Training Stipend and a Sanfilippo Children's Foundation PhD Top-Up Scholarship.

Author Statement

ML was responsible for project conception, funding and oversight. EG, KB and SA generated and characterised the mutants. EG and LR generated and analysed the adult and larval behaviour data respectively. LR was supervised by KH. KB and EG generated the RNA-seq data, which was then analysed by KB. Initial drafting of the manuscript was performed by EG and KB. Subsequent revisions and editing were performed by EG, LR, KB and ML.

Figures

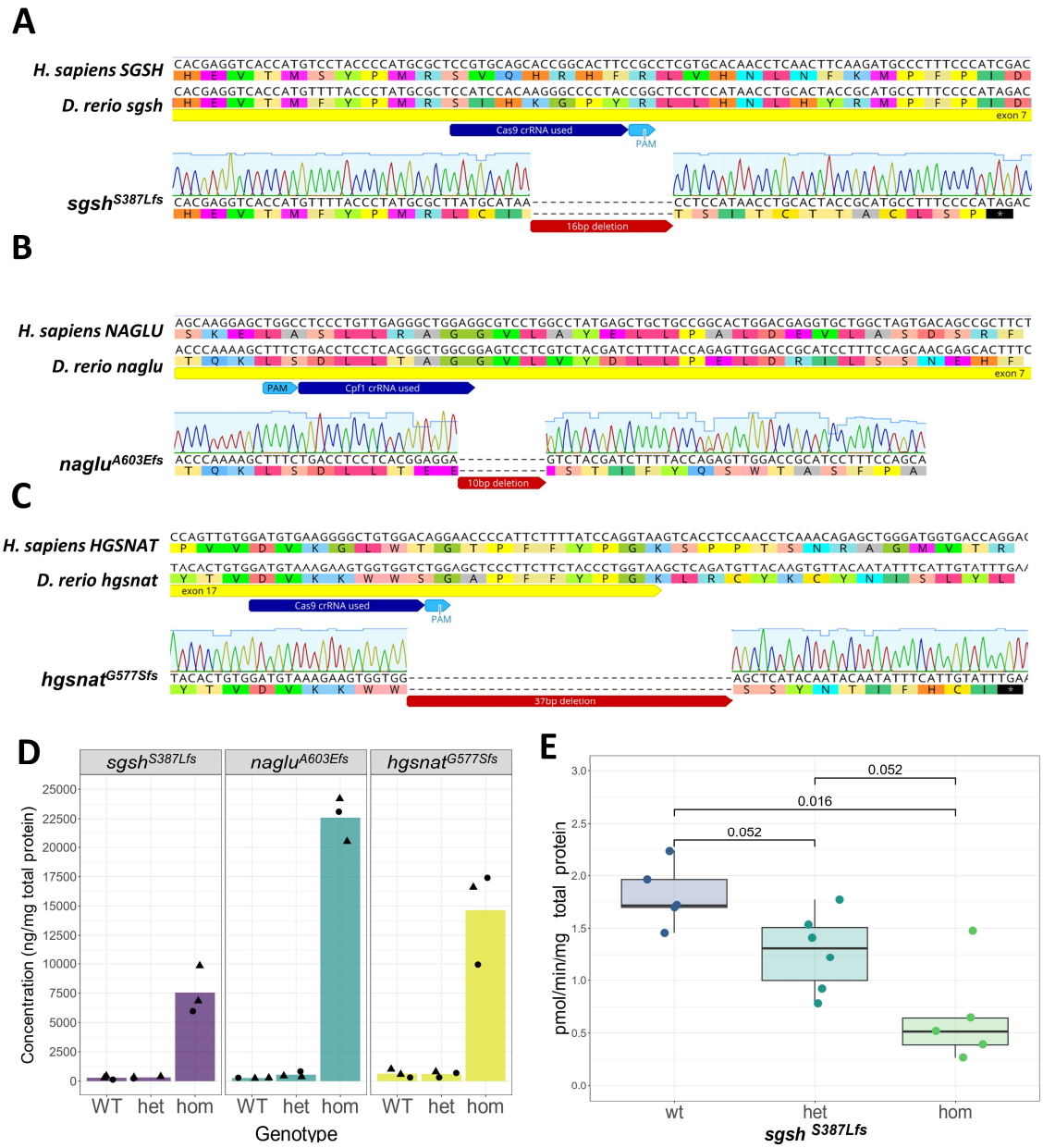


Fig.1: Summary of MPS III mutant characterisation. A-C, Sanger sequencing chromatograms of PCR amplicons from homozygous mutant zebrafish (*D. rerio*). **A, *sgsh*^{S387Lfs}**; **B, *naglu*^{A603Efs}**; and **C, h*gsnat*^{G577Sfs}** zebrafish, aligned to the corresponding genomic sequences of *Homo sapiens* and zebrafish. Light blue shading on each chromatogram indicates the sequencing base call quality. CRISPR guide RNAs (dark blue), PAMs (light blue), exonic regions (yellow), and deletion sites (red) are indicated. Amino acid sequences are displayed under each nucleotide sequence. **D,** Mass spectrometry measurement of heparan sulfate disaccharide concentration in wild type, heterozygous and homozygous 3-month old *sgsh*^{S387Lfs} (left panel), *naglu*^{A603Efs} (centre panel) and *hgsnat*^{G577Sfs} (right panel) zebrafish brains. Raw mass spectrometry data can be found in **Supplemental Data File 7**. **E,** Fluorometric assay of enzyme activity in 16-month old brains of *sgsh*^{S387Lfs} wild type (n = 5), heterozygous (n = 6) and homozygous (n = 5) fish. Enzyme activity is plotted as pmol of substrate converted per min per mg of total protein.

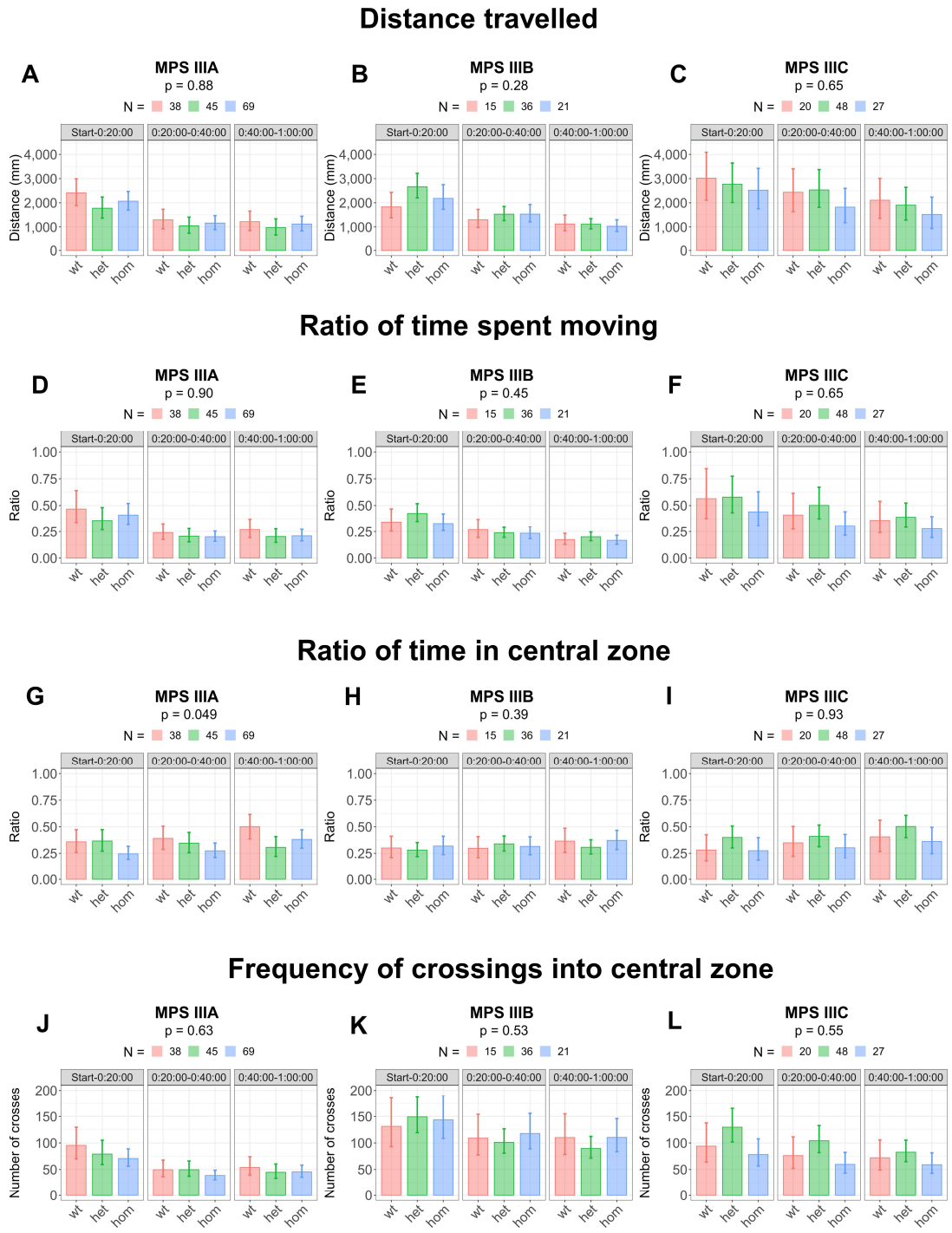


Fig.2: MPS III zebrafish larvae generally do not show behaviour phenotypes reminiscent of the human disease at 5 days post fertilisation. A-C, Linear mixed model-predicted total distance travelled during 1 hour (split across three 20 minute time bins) by MPS III larvae. Families consisted of larvae wild type (wt), heterozygous (het) and homozygous (hom) for each MPS III subtype indicated in the panel title. The number (N) of larvae per genotype are also indicated in the legend. **D-F**, Generalised linear mixed effect model-predicted ratio of time spent moving during 1 hour by MPS III larvae, where a greater ratio represents a greater proportion of time spent moving. **G-I**, Generalised linear mixed model-predicted ratio of time in the central zone, where a greater ratio represents a greater proportion of time in the centre. **J-L**, Generalised linear mixed model-predicted frequency of crossings into central zone. Raw data points for these estimated marginal means are shown in **Fig.S15**.

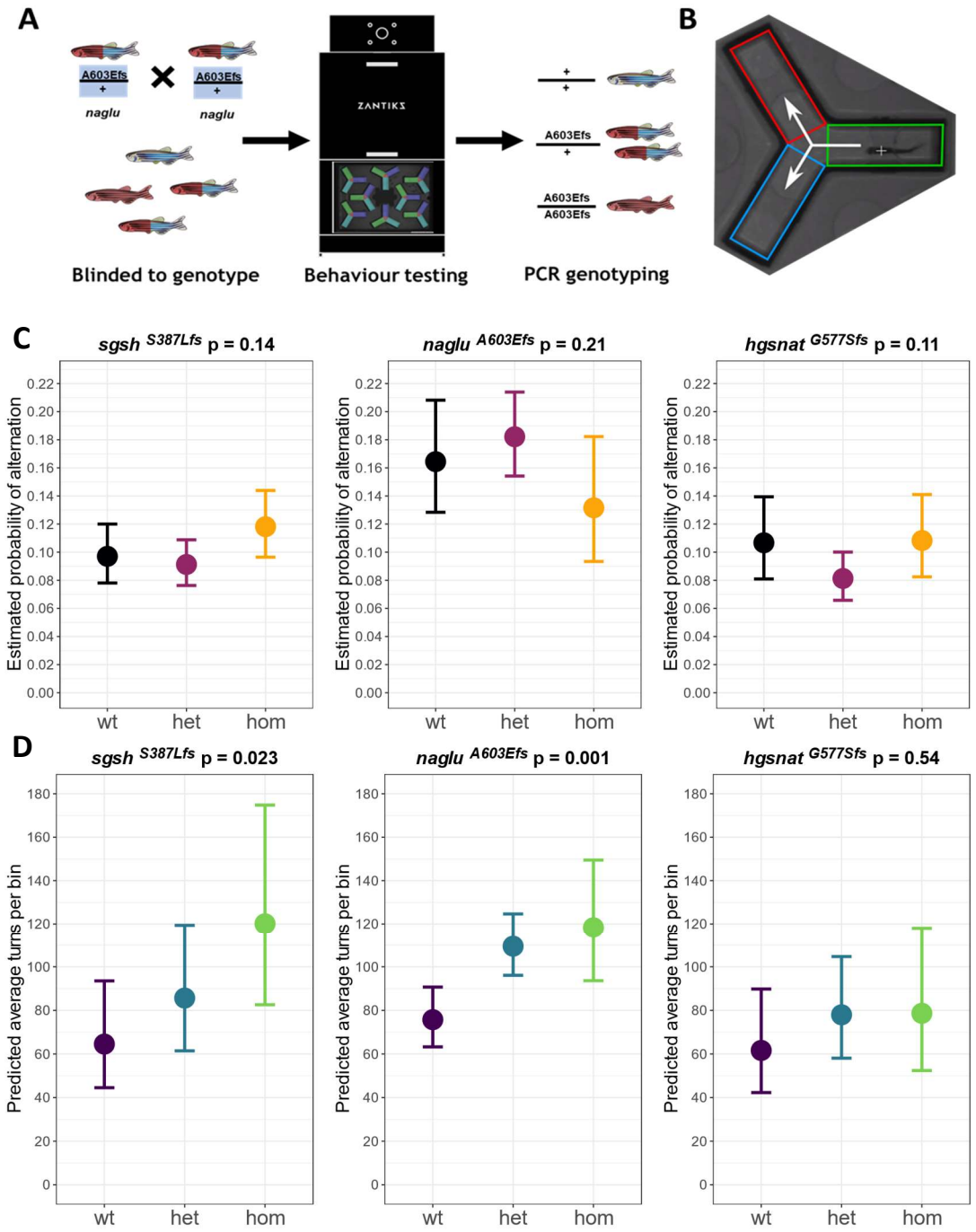
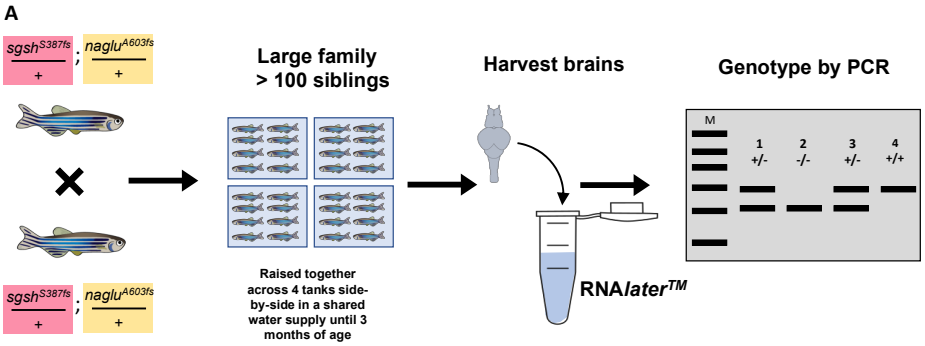


Fig.3: MPS III zebrafish adults show behavioural changes reminiscent of the human disease. A, The single heterozygous in-cross breeding strategy used in which experimental families are generated from a single parent pair. All genotypes of interest are present within each family, and the genotype of each individual fish is determined after behavioural testing is concluded. This figure displays this strategy as used to analyse the *naglu*^{A603Efs} mutation and was identical for analysis of the other mutations. **B,** Diagram of a single Y-maze arena divided into three zones. Left-right turn choices are recorded. **C,** Generalised linear mixed model predicted probabilities of zebrafish performing an alternation tetragram in the Y-maze at 20 months (*sgsh*^{S387Lfs}, left panel), 21 months (*naglu*^{A603Efs}, middle panel) and 17 months (*hgsnat*^{G577Sfs}, right panel). Effects of genotype are not significant for alternation probability in each case (**Table.S3**). **D,** Generalised linear mixed model-predicted average number of turns per 10 minute time bin in the same behavioural experiments as reported in C. Significance of fixed effects for predicted turn values can be found in **Table.S3**.



B

Genotype		% of progeny
sgsh	*naglu*	
+/+	+/+	6.25% (1 in 16)
mut/+	+/+	12.5% (1 in 8)
mut/mut	+/+	6.25% (1 in 16)
+/+	mut/+	12.5% (1 in 8)
mut/+	mut/+	25% (1 in 4)
mut/mut	mut/+	12.5% (1 in 8)
+/+	mut/mut	6.25% (1 in 16)
mut/+	mut/mut	12.5% (1 in 8)
mut/mut	mut/mut	6.25% (1 in 16)

Fig.4: Intra-family brain transcriptome analysis. **A,** Zebrafish doubly heterozygous for two MPS III mutations (*sgsh* ^{S387Lfs}]MPS IIIA] and *naglu* ^{A603fs} [MPS IIIB] shown here) are in-crossed to generate a large family of at least 100 siblings. The family is raised across, at least, 4 tanks, at a density of 25 fish per tank until 3 months of age. Then, all fish are euthanised and their brains harvested and preserved in RNA*later* solution. Each fish is genotyped using PCR tests. **B,** Expected proportions of mutant genotypes assuming normal Mendelian inheritance (illustrated using *sgsh* and *naglu*). The wild type and single homozygous mutation genotypes of interest each represent 6.25% of the entire progeny (approx. 6 fish per genotype). The brains of these fish were then selected for RNA-seq, see **Fig.S5-7**.

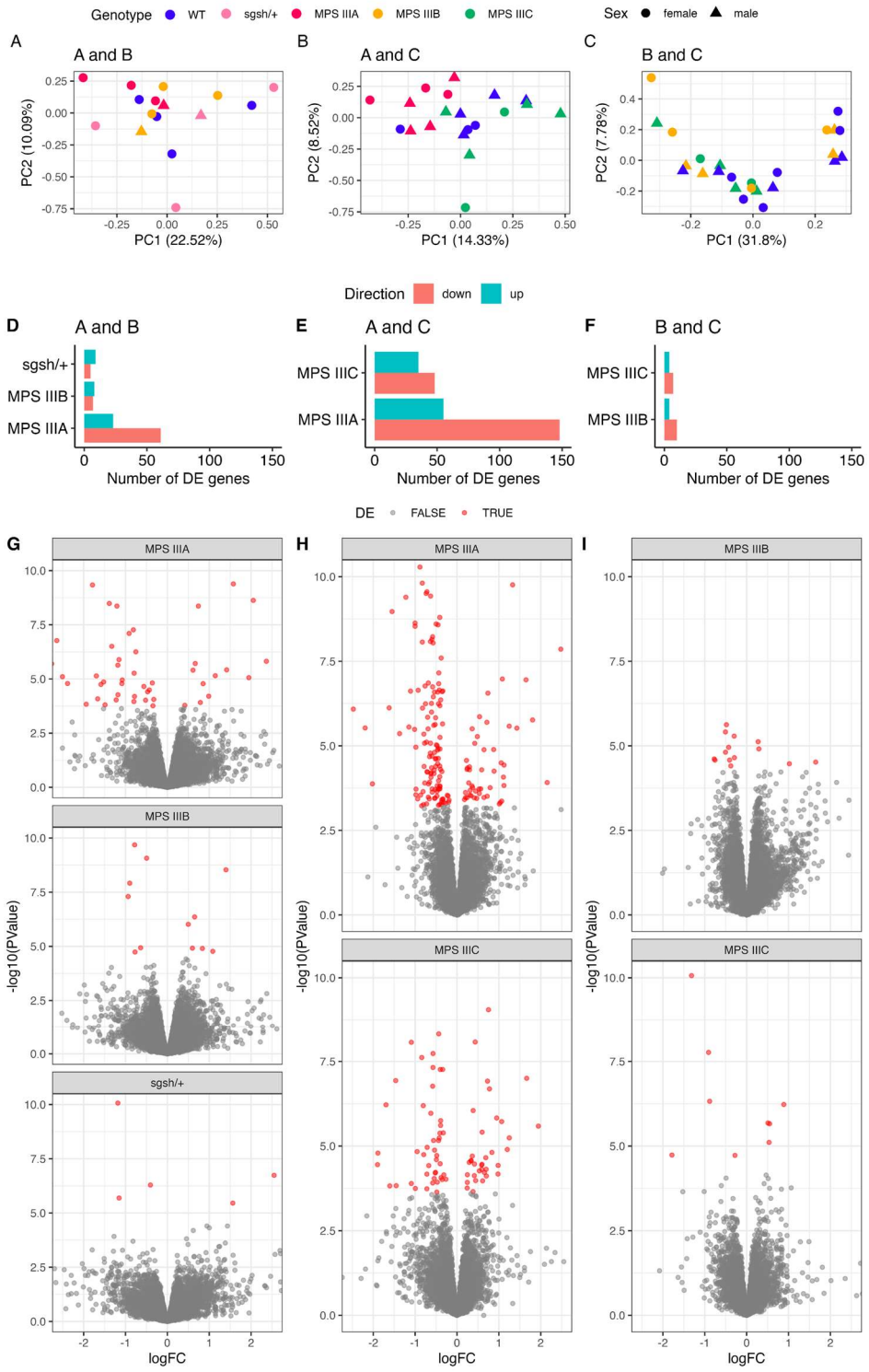


Fig.5: Principal component analysis of zebrafish brain transcriptome datasets. **A,** A and B; **B,** A and C; and **C,** B and C. Each point represents a brain transcriptome, which is coloured according to genotype and shaped according to sex. The percentage of variance explained by principal components (PC) 1 and 2 are indicated within the x and y axis labels. **D-F,** Bar plots indicating the numbers of differentially expressed (DE) genes (up and downregulated) detected due to each mutant genotype in the **D,** A and B; **E,** A and C; and **F,** B and C families. **G-I,** Volcano plots of differential gene expression analyses. Plots are constrained between -2.5 and 2.5 on the x axis and between 0 and 10 on the y axis for visualisation purposes.

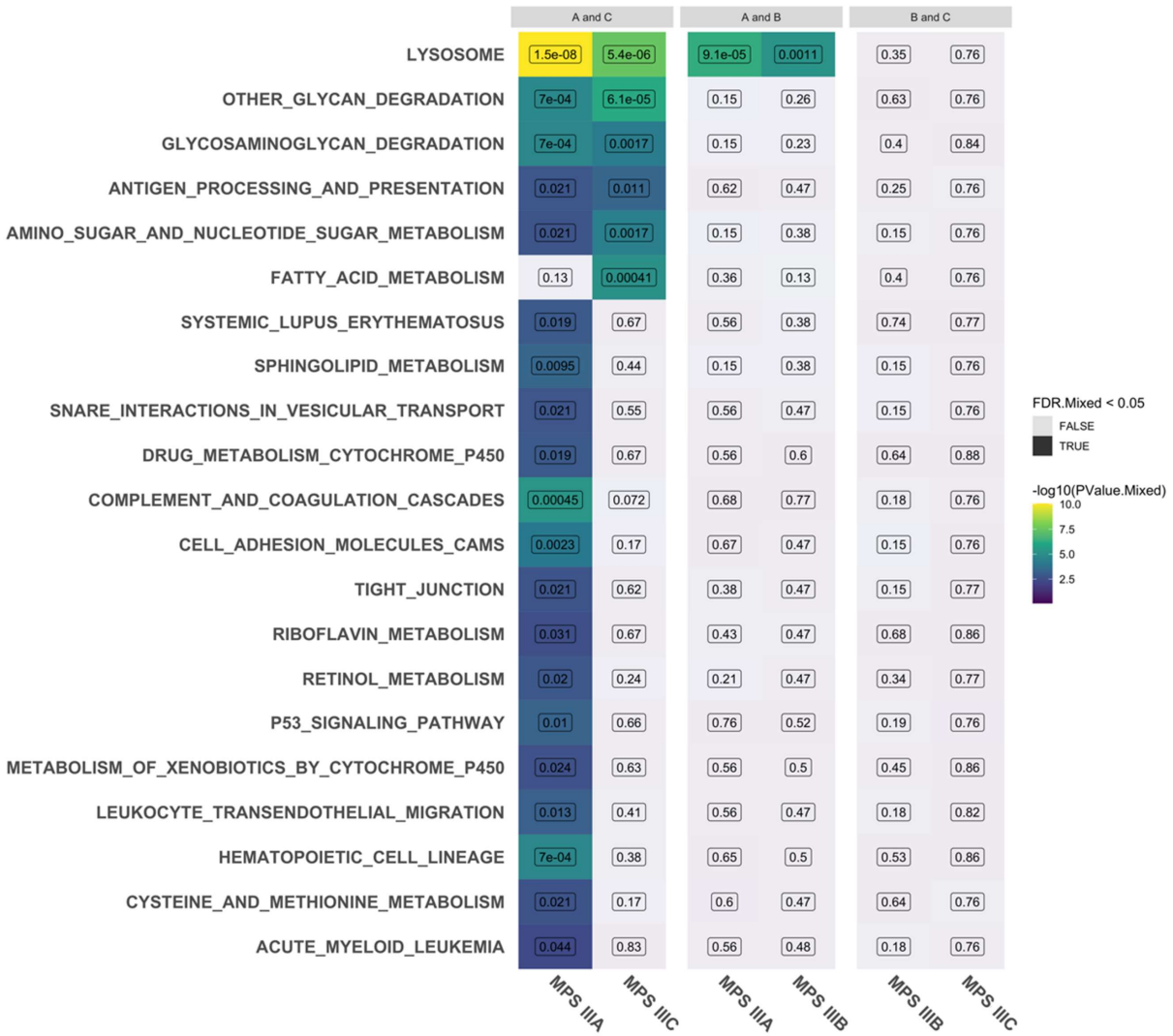


Fig.6: KEGG gene sets significantly altered in MPS III mutant zebrafish brain transcriptomes. The cells in the heatmap are coloured according to statistical significance (brighter, more yellow colours indicate greater statistical significance). Cells appear grey if the FDR-adjusted p-value (indicated within each cell) did not reach the threshold of 0.05.

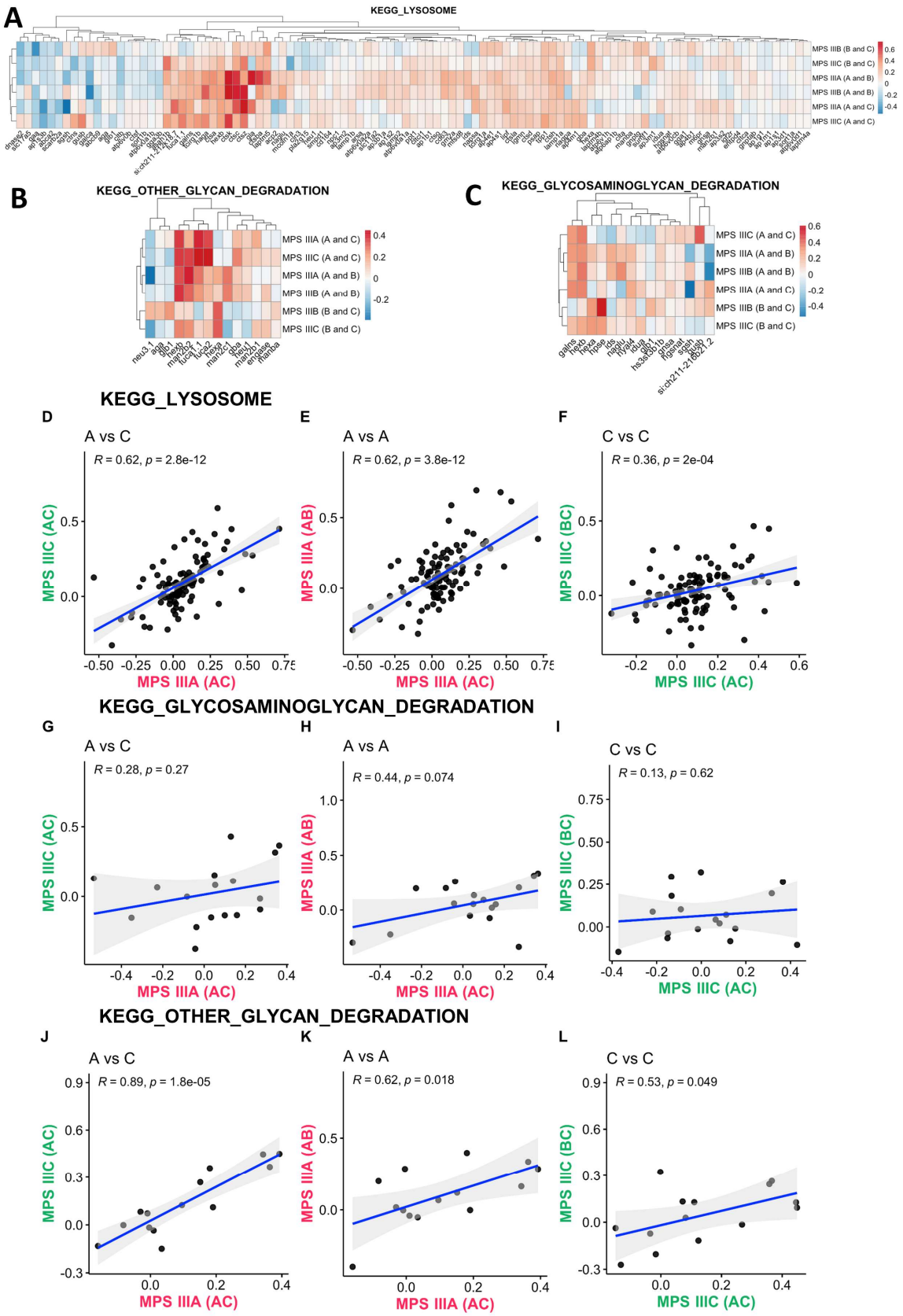


Fig.7: Heatmap indicating the logFC values of genes in the **A,** LYSOSOME; **B,** OTHER_GLYCAN_DEGRADATION; and **C,** GLYCOSAMINOGLYCAN_DEGRADATION gene sets in all MPS III mutants. **D-L,** Pearson correlation of logFC values between individual zebrafish replicates of MPS III subtype groups.

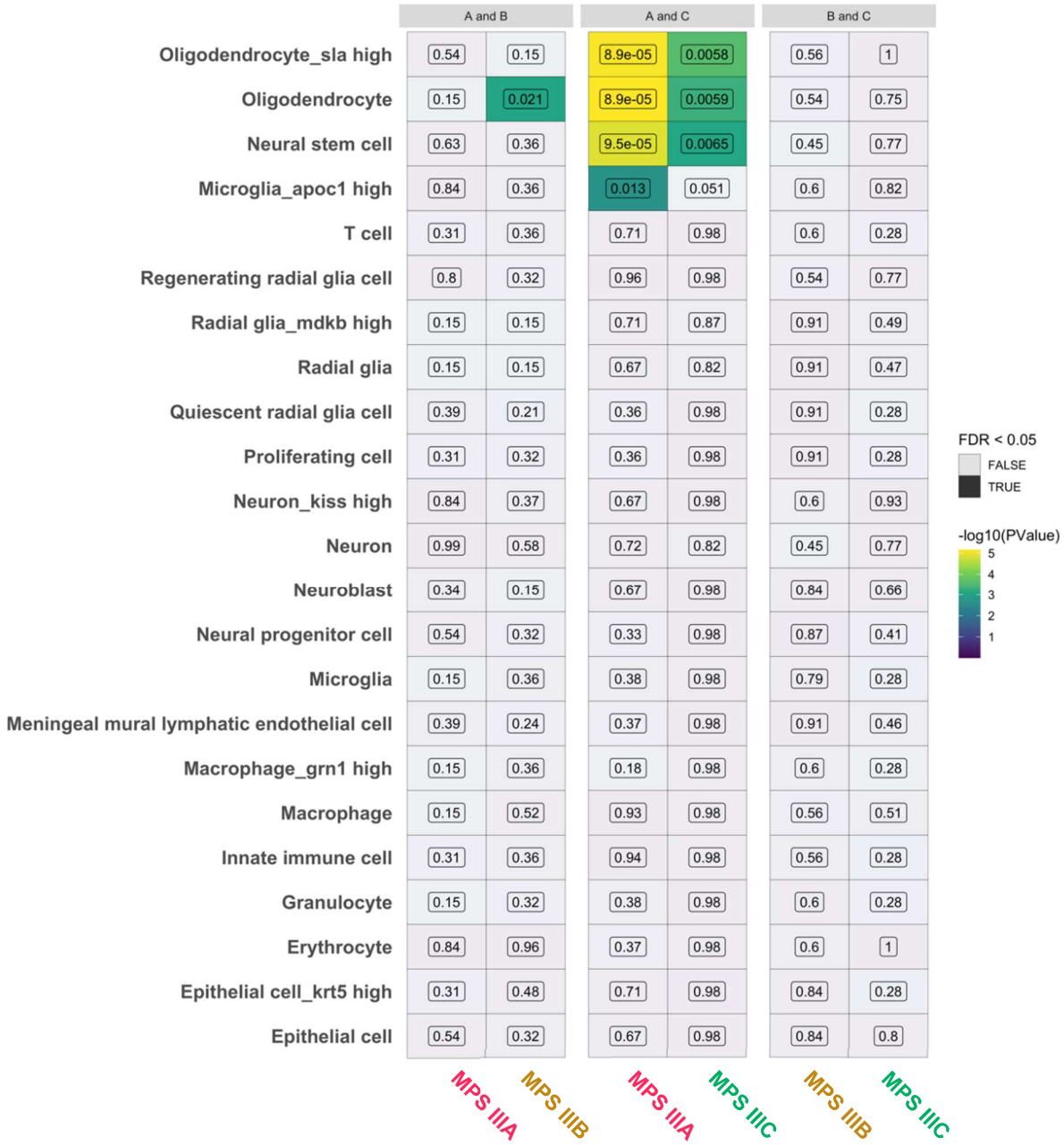


Fig.8: Potential changes to cell type marker gene expression in young-adult MPS III zebrafish brain transcriptomes. The cells in the heatmap are coloured by statistical significance (Brighter, more yellow colours indicate greater statistical significance). The cells appear grey if the FDR-adjusted p-value (the numbers within the cells) did not reach the threshold of 0.05.

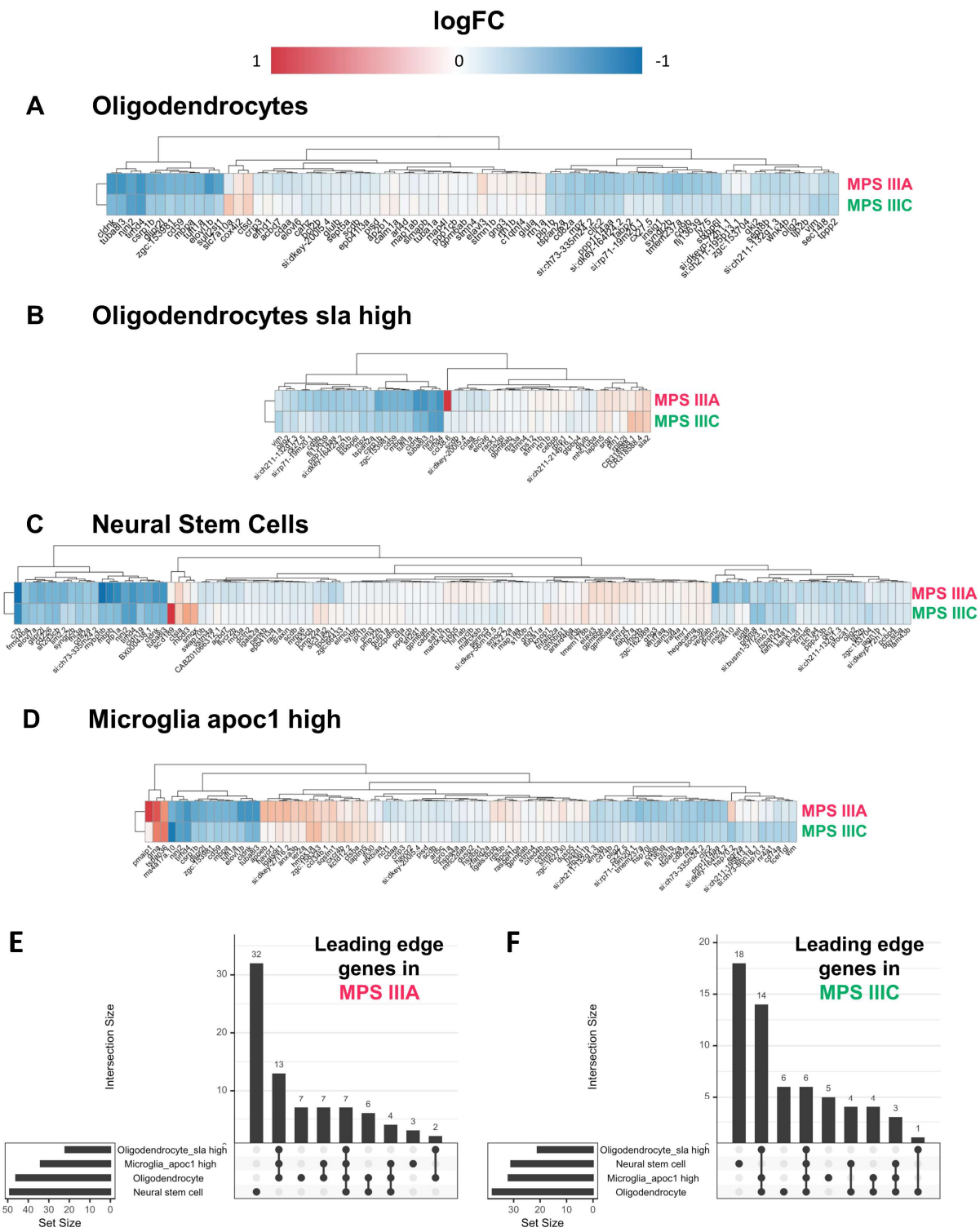


Fig.9: Direction of change in oligodendrocyte, neural stem cell and microglial gene sets in MPS IIIA and MPS IIIC zebrafish brains. The $\log_2$FC (logFC) of detectable genes in the **A,** oligodendrocyte; **B,** oligodendrocyte with high sla expression; **C,** neural stem cells; and **D,** microglia with high *apoc1* expression gene sets. The colour bar represents the value of the logFC for **A-D**. **E,** Overlap of leading edge genes in the significantly altered marker gene sets in MPS IIIA mutants in the AC arm. **F,** Overlap of leading edge genes in the significantly altered marker gene sets in MPS IIIC mutants in the AC arm.

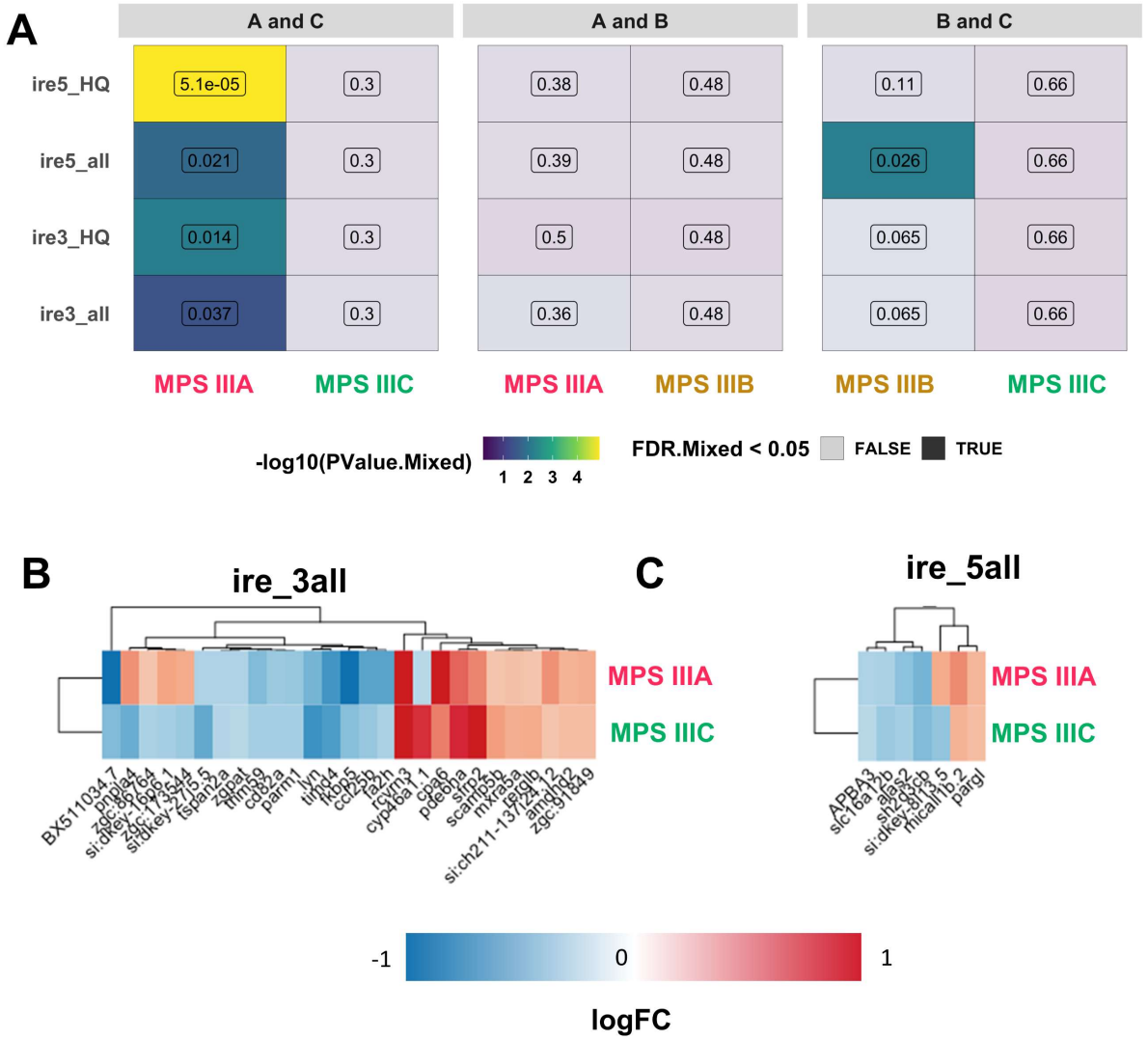


Fig.10: Iron homeostasis changes in MPS III zebrafish. **A,** Heatmap showing the statistical significance of the iron-responsive element (IRE) gene sets in MPS III zebrafish brains. The colour of the cells indicates the level of significance (brighter colour indicates greater statistical significance, while the cells appear darker if the FDR-adjusted p-value is > 0.05). Numbers within cells are the FDR-adjusted p-values**. B,** Heatmap indicating the $\log_2$FC (logFC) values for genes encoding transcripts with iron-responsive elements (IREs) in their 3' or **C,** 5' untranslated regions. Only genes showing absolute $\log_2$ fold change values greater than 0.3 are displayed for visualisation purposes. (The complete heatmap without this filtering is shown in **Fig.S27**). Clustering was performed using Euclidean distances.

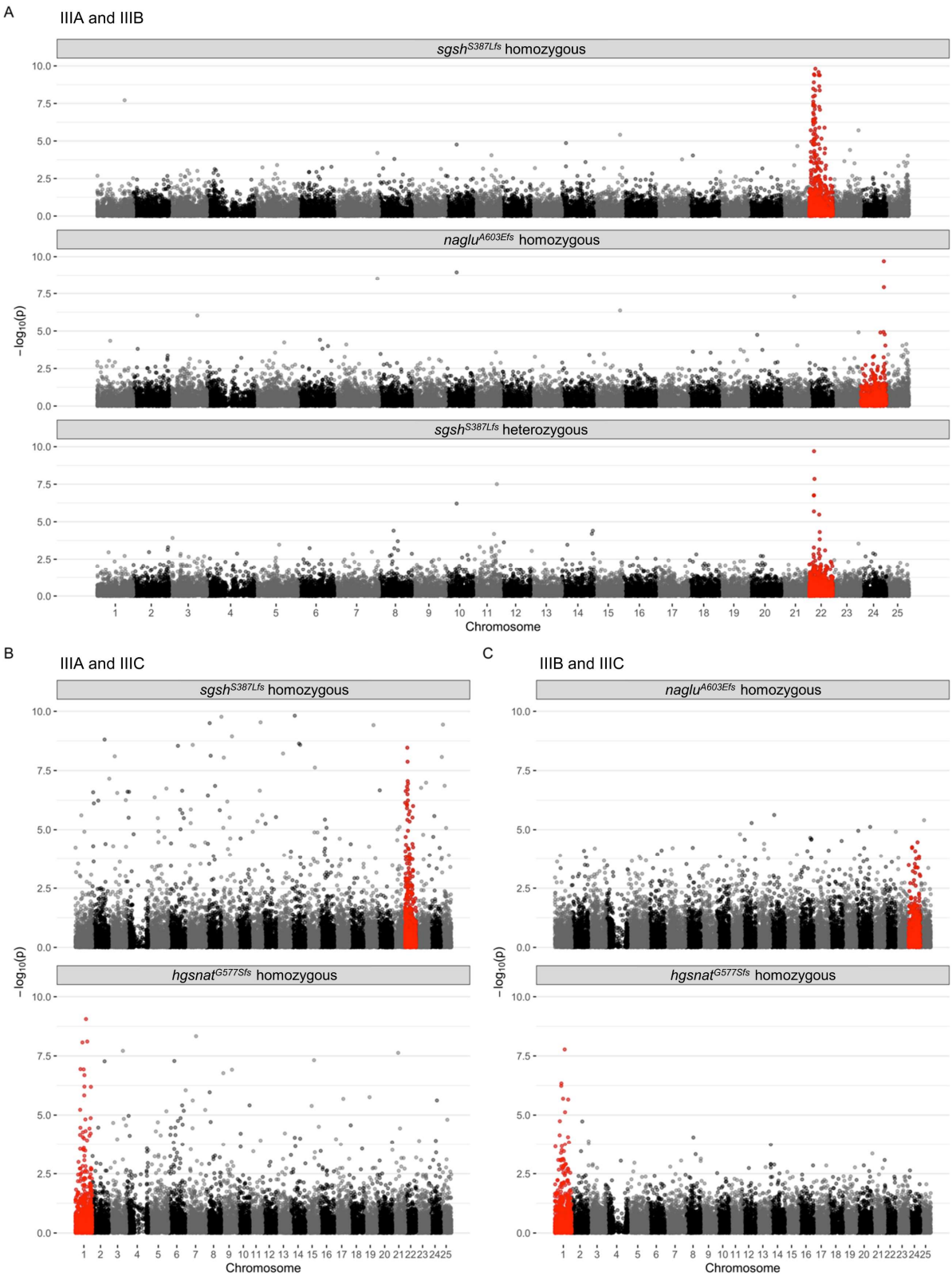


Fig.11. Manhattan plots indicating p-values from the differential gene expression analyses in the **A,** AB arm; **B,** AC arm; and **C,** BC arm. Points appear red if they are located on the chromosome bearing the mutation.

References

1. Cleary, M.A. and J.E. Wraith, *Management of mucopolysaccharidosis type III.* Arch Dis Child, 1993. **69**(3): p. 403-6.
2. Muschol, N., et al., *Sanfilippo syndrome: consensus guidelines for clinical care.* Orphanet Journal of Rare Diseases, 2022. **17**(1): p. 391.
3. Valstar, M.J., et al., *Cognitive development in patients with Mucopolysaccharidosis type III (Sanfilippo syndrome).* Orphanet J Rare Dis, 2011. **6**: p. 43.
4. Valstar, M.J., et al., *Mucopolysaccharidosis type IIIA: clinical spectrum and genotype-phenotype correlations.* Ann Neurol, 2010. **68**(6): p. 876-87.
5. Lavery, C., C.J. Hendriksz, and S.A. Jones, *Mortality in patients with Sanfilippo syndrome.* Orphanet Journal of Rare Diseases, 2017. **12**(1): p. 168.
6. Zelei, T., et al., *Epidemiology of Sanfilippo syndrome: results of a systematic literature review.* Orphanet Journal of Rare Diseases, 2018. **13**(1): p. 53.
7. Kamp, J.J.P.V.D., et al., *Genetic heterogeneity and clinical variability in the Sanfilippo syndrome (types A, B, and C).* Clinical Genetics, 1981. **20**(2): p. 152-160.
8. Andria, G., et al., *Sanfilippo B syndrome (MPS III B): mild and severe forms within the same sibship.* Clin Genet, 1979. **15**(6): p. 500-4.
9. Lindor, N.M., et al., *Sanfilippo syndrome type A in two adult sibs.* American Journal of Medical Genetics, 1994. **53**(3): p. 241-244.
10. Viana, G.M., et al., *Brain Pathology in Mucopolysaccharidoses (MPS) Patients with Neurological Forms.* J Clin Med, 2020. **9**(2).
11. Pearse, Y. and M. Iacovino, *A Cure for Sanfilippo Syndrome? A Summary of Current Therapeutic Approaches and their Promise.* Med Res Arch, 2020. **8**(2).
12. Zhang, T. and R.T. Peterson, *Modeling Lysosomal Storage Diseases in the Zebrafish.* Frontiers in Molecular Biosciences, 2020. **7**.
13. Howe, K., et al., *The zebrafish reference genome sequence and its relationship to the human genome.* Nature, 2013. **496**(7446): p. 498-503.
14. Panula, P., et al., *The comparative neuroanatomy and neurochemistry of zebrafish CNS systems of relevance to human neuropsychiatric diseases.* Neurobiology of Disease, 2010. **40**(1): p. 46-57.
15. Barthelson, K., et al., *PRESENILIN 1 mutations causing early-onset familial Alzheimer's disease or familial acne inversa differ in their effects on genes facilitating energy metabolism and signal transduction.* 2021: p. 2021.01.26.428321.
16. Barthelson, K., et al., *Brain transcriptome analysis reveals subtle effects on mitochondrial function and iron homeostasis of mutations in the SORL1 gene implicated in early onset familial Alzheimer's disease.* Mol Brain, 2020. **13**(1): p. 142.
17. Hin, N., et al., *Accelerated brain aging towards transcriptional inversion in a zebrafish model of the K115fs mutation of human PSEN2.* PloS one, 2020. **15**(1): p. e0227258-e0227258.
18. Jiang, H., et al., *Transcriptome analysis indicates dominant effects on ribosome and mitochondrial function of a premature termination codon mutation in the zebrafish gene psen2.* PloS one, 2020. **15**(7): p. e0232559.
19. Norton, W. and L. Bally-Cuif, *Adult zebrafish as a model organism for behavioural genetics.* BMC Neuroscience, 2010. **11**(1): p. 90.
20. Douek, A.M., et al., *An Engineered sgsh Mutant Zebrafish Recapitulates Molecular and Behavioural Pathobiology of Sanfilippo Syndrome A/MPS IIIA.* International journal of molecular sciences, 2021. **22**(11): p. 5948.
21. Douek, A.M., et al., *Systems-level investigation of mucopolysaccharidosis IIIA identifies deficient synaptic activity as a key driver of disease progression.* bioRxiv, 2022: p. 2022.10.03.510585.

22. Westerfield, M., *The zebrafish book: a guide for the laboratory use of zebrafish.* University of Oregon Press, 2000.
23. Fernandez, J.P., et al., *Optimized CRISPR-Cpf1 system for genome editing in zebrafish.* Methods, 2018. **150**: p. 11-18.
24. Allen, A.G., K. Barthelson, and M. Lardelli, *pHAPE: a plasmid for production of DNA size marker ladders for gel electrophoresis.* bioRxiv, 2022: p. 2022.11.04.515137.
25. Trim, P.J., J.J. Hopwood, and M.F. Snel, *Butanolysis Derivatization: Improved Sensitivity in LC-MS/MS Quantitation of Heparan Sulfate in Urine from Mucopolysaccharidosis Patients.* Analytical Chemistry, 2015. **87**(18): p. 9243-9250.
26. Whyte, L.S., et al., *Variables influencing fluorimetric N-sulfoglucosamine sulfohydrolase (SGSH) activity measurement in brain homogenates.* Mol Genet Metab Rep, 2015. **5**: p. 60-62.
27. Karpova, E.A., et al., *A fluorimetric enzyme assay for the diagnosis of sanfilippo disease type A (MPS IIIA).* Journal of Inherited Metabolic Disease, 1996. **19**(3): p. 278-285.
28. Wickham, H., *ggplot2: elegant graphics for data analysis*. 2nd 2016 ed. 2016, New York: Springer-Verlag.
29. RStudio Team, *RStudio: Integrated Development Environment for R*. 2022, RStudio, PBC: Boston, MA.
30. R Core Team, *R: A language and environment for statistical computing*. 2021, R Foundation for Statistical Computing: Vienna, Austria.
31. Bates, D., et al., *Fitting Linear Mixed-Effects Models Using lme4.* Journal of Statistical Software, 2015. **67**(1): p. 1 - 48.
32. Brooks, M.E., et al., *glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling.* The R journal, 2017. **9**(2): p. 378.
33. Fox, J. and S. Weisberg, *An R Companion to Applied Regression*. 2019, Thousand Oaks, California: SAGE.
34. Russell V. Lenth, B.B., Paul Buerkner, Iago Giné-Vázquez, Maxime Herve, Maarten Jung, Jonathon Love, Fernando Miguez, Hannes Riebl, Henrik Singmann, *emmeans: Estimated Marginal Means, aka Least-Squares Means*. 2023.
35. Cleal, M., et al., *The Free-movement pattern Y-maze: A cross-species measure of working memory and executive function.* Behavior Research Methods, 2021. **53**(2): p. 536-557.
36. Fontana, B.D., et al., *Zebrafish (Danio rerio) behavioral laterality predicts increased short-term avoidance memory but not stress-reactivity responses.* Anim Cogn, 2019. **22**(6): p. 1051-1061.
37. Venables, W.N. and B.D. Ripley, *Modern Applied Statistics with S*. Fourth Edition ed. Statistics and Computing. 2002, NY: Springer New York.
38. Barthelson, K., et al., *Brain transcriptome analysis reveals subtle effects on mitochondrial function and iron homeostasis of mutations in the SORL1 gene implicated in early onset familial Alzheimer's disease.* Molecular Brain, 2020. **13**(1): p. 142.
39. Mölder, F., K.P. Jablonski, and B. Letcher, *Sustainable data analysis with Snakemake.[version 2; peer review: 2 approved]. F1000Research 10: 33.* Crossref, PubMed, 2021.
40. Andrews, S., *FastQC: a quality control tool for high throughput sequence data*. 2010, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
41. Chen, S., et al., *fastp: an ultra-fast all-in-one FASTQ preprocessor.* Bioinformatics, 2018. **34**(17): p. i884-i890.
42. Cunningham, F., et al., *Ensembl 2022.* Nucleic Acids Research, 2022. **50**(D1): p. D988-D995.
43. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
44. Smith, T.S., A. Heger, and I. Sudbery, *UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy.* Genome Research, 2017.
45. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-930.

46. Chen, Y., A.T. Lun, and G.K. Smyth, *From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline.* F1000Res, 2016. **5**: p. 1438.
47. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data.* Genome Biology, 2010. **11**(3): p. R25.
48. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
49. Hansen, K.D., R.A. Irizarry, and Z. Wu, *Removing technical variability in RNA-seq data using conditional quantile normalization.* Biostatistics, 2012. **13**(2): p. 204-16.
50. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27-30.
51. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
52. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.
53. Dolgalev, I., *msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format*. 2021.
54. Hin, N., et al., *Iron responsive element-mediated responses to iron dyshomeostasis in Alzheimer's disease.* Journal of Alzheimer's Disease, 2021. **84**(4): p. 1597-1630.
55. Jiang, M., et al., *Characterization of the Zebrafish Cell Landscape at Single-Cell Resolution.* Frontiers in Cell and Developmental Biology, 2021. **9**.
56. Young, M.D., et al., *Gene ontology analysis for RNA-seq: accounting for selection bias.* Genome Biology, 2010. **11**(2): p. R14.
57. Wu, D., et al., *ROAST: rotation gene set tests for complex microarray experiments.* Bioinformatics, 2010. **26**(17): p. 2176-2182.
58. Korotkevich, G., et al., *Fast gene set enrichment analysis.* bioRxiv, 2021: p. 060012.
59. Bi, R. and P. Liu, *Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments.* BMC Bioinformatics, 2016. **17**: p. 146.
60. Yogalingam, G. and J.J. Hopwood, *Molecular genetics of mucopolysaccharidosis type IIIA and IIIB: Diagnostic, clinical, and biological implications.* Human mutation, 2001. **18**(4): p. 264-281.
61. Feldhammer, M., et al., *Sanfilippo syndrome type C: mutation spectrum in the heparan sulfate acetyl-CoA: α-glucosaminide N-acetyltransferase (HGSNAT) gene.* Human mutation, 2009. **30**(6): p. 918-925.
62. El-Brolosy, M.A., et al., *Genetic compensation triggered by mutant mRNA degradation.* Nature (London), 2019. **568**(7751): p. 193-197.
63. Trim, P.J., J.J. Hopwood, and M.F. Snel, *Butanolysis derivatization: improved sensitivity in LC-MS/MS quantitation of heparan sulfate in urine from mucopolysaccharidosis patients.* Anal Chem, 2015. **87**(18): p. 9243-50.
64. Meyer, A., et al., *Scoring Evaluation of the Natural Course of Mucopolysaccharidosis Type IIIA (Sanfilippo Syndrome Type A).* Pediatrics, 2007. **120**(5): p. e1255-e1261.
65. Valle, F.P., *Effects of Strain, Sex, and Illumination on Open-Field Behavior of Rats.* The American Journal of Psychology, 1970. **83**(1): p. 103-111.
66. Barthelson, K., M. Newman, and M. Lardelli, *Brain transcriptomes of zebrafish and mouse Alzheimer's disease knock-in models imply early disrupted energy metabolism.* Disease Models & Mechanisms, 2022. **15**(1).
67. Andrade, F., et al., *Sanfilippo syndrome: Overall review.* Pediatrics International, 2015. **57**(3): p. 331-338.
68. Li, K. and H. Reichmann, *Role of iron in neurodegenerative diseases.* Journal of Neural Transmission, 2016. **123**(4): p. 389-399.

69. Brady, J., et al., *Mucopolysaccharidosis type IIIB (MPS IIIB) masquerading as a behavioural disorder.* BMJ Case Rep, 2013. **2013**.
70. Puy, V., et al., *Predominant role of microglia in brain iron retention in Sanfilippo syndrome, a pediatric neurodegenerative disease.* Glia, 2018. **66**(8): p. 1709-1723.
71. Butt, J., et al., *Differences in the RNA binding sites of iron regulatory proteins and potential target diversity.* Proceedings of the National Academy of Sciences, 1996. **93**(9): p. 4345-4349.
72. Henderson, B.R., et al., *Optimal sequence and structure of iron-responsive elements. Selection of RNA stem-loops with high affinity for iron regulatory factor.* Journal of Biological Chemistry, 1994. **269**(26): p. 17481-17489.
73. Kluenemann, H.H., et al., *Parkinsonism syndrome in heterozygotes for Niemann–Pick C1.* Journal of the Neurological Sciences, 2013. **335**(1): p. 219-220.
74. Clark, L.N., et al., *Gene-Wise Association of Variants in Four Lysosomal Storage Disorder Genes in Neuropathologically Confirmed Lewy Body Disease.* PLOS ONE, 2015. **10**(5): p. e0125204.
75. Gan-Or, Z., et al., *The p. L302P mutation in the lysosomal enzyme gene SMPD1 is a risk factor for Parkinson disease.* Neurology, 2013. **80**(17): p. 1606-1610.
76. Douglass, M.L., et al., *Is SGSH heterozygosity a risk factor for early-onset neurodegenerative disease?* Journal of Inherited Metabolic Disease, 2021. **n/a**(n/a).
77. White, R.J., et al., *Allele-specific gene expression can underlie altered transcript abundance in zebrafish mutants.* Elife, 2022. **11**.
78. Baer, L., et al., *Differential allelic representation (DAR) identifies candidate eQTLs and improves transcriptome analysis.* bioRxiv, 2023: p. 2023.03.02.530865.
79. Wagner, V.F. and H. Northrup, *Mucopolysaccharidosis type III*. 2019, Gene Reviews.
80. Kan, S.H., et al., *Behavioral deficits and cholinergic pathway abnormalities in male Sanfilippo B mice.* Behav Brain Res, 2016. **312**: p. 265-71.
81. Zambusi, A. and J. Ninkovic, *Regeneration of the central nervous system-principles from brain regeneration in adult zebrafish.* World J Stem Cells, 2020. **12**(1): p. 8-24.
82. Gaikwad, S., et al., *Acute stress disrupts performance of zebrafish in the cued and spatial memory tests: the utility of fish models to study stress-memory interplay.* Behav Processes, 2011. **87**(2): p. 224-30.
83. Porter, K.A., et al., *Parent Experiences of Sanfilippo Syndrome Impact and Unmet Treatment Needs: A Qualitative Assessment.* Neurol Ther, 2021. **10**(1): p. 197-212.
84. Langford-Smith, A., et al., *Female Mucopolysaccharidosis IIIA Mice Exhibit Hyperactivity and a Reduced Sense of Danger in the Open Field Test.* PLOS ONE, 2011. **6**(10): p. e25717.
85. Hemsley, K.M. and J.J. Hopwood, *Development of motor deficits in a murine model of mucopolysaccharidosis type IIIA (MPS-IIIA).* Behav Brain Res, 2005. **158**(2): p. 191-9.
86. Langford-Smith, A., et al., *Hyperactive behaviour in the mouse model of mucopolysaccharidosis IIIB in the open field and home cage environments.* Genes Brain Behav, 2011. **10**(6): p. 673-82.
87. Lau, A.A., et al., *Open field locomotor activity and anxiety-related behaviors in mucopolysaccharidosis type IIIA mice.* Behavioural Brain Research, 2008. **191**(1): p. 130-136.
88. Hemsley, K.M., B. King, and J.J. Hopwood, *Injection of recombinant human sulfamidase into the CSF via the cerebellomedullary cistern in MPS IIIA mice.* Molecular Genetics and Metabolism, 2007. **90**(3): p. 313-328.
89. Martins, C., et al., *Neuroinflammation, mitochondrial defects and neurodegeneration in mucopolysaccharidosis III type C mouse model.* Brain, 2015. **138**(Pt 2): p. 336-55.
90. Dwyer, C.A., et al., *Neurodevelopmental Changes in Excitatory Synaptic Structure and Function in the Cerebral Cortex of Sanfilippo Syndrome IIIA Mice.* Scientific Reports, 2017. **7**(1): p. 46576.
91. Wilkinson, F.L., et al., *Neuropathology in Mouse Models of Mucopolysaccharidosis Type I, IIIA and IIIB.* PLOS ONE, 2012. **7**(4): p. e35787.

92. Vitry, S., et al., *Enhanced degradation of synaptophysin by the proteasome in mucopolysaccharidosis type IIIB.* Molecular and Cellular Neuroscience, 2009. **41**(1): p. 8-18.
93. Cressant, A., et al., *Improved behavior and neuropathology in the mouse model of Sanfilippo type IIIB disease after adeno-associated virus-mediated gene transfer in the striatum.* J Neurosci, 2004. **24**(45): p. 10229-39.
94. Fontana, B. and M. Parker, *The larval diving response (LDR): Validation of an automated, high-throughput, ecologically relevant measure of anxiety-related behavior in larval zebrafish (Danio rerio).* Journal of Neuroscience Methods, 2022. **381**: p. 109706.
95. Fontana, B.D., N. Alnassar, and M.O. Parker, *The zebrafish (Danio rerio) anxiety test battery: comparison of behavioral responses in the novel tank diving and light–dark tasks following exposure to anxiogenic and anxiolytic compounds.* Psychopharmacology, 2022. **239**(1): p. 287-296.
96. Taherzadeh, M., et al., *Severe Central Nervous System Demyelination in Sanfilippo Disease.* bioRxiv, 2023: p. 2023.04.12.536631.
97. Connor, J.R. and S.L. Menzies, *Relationship of iron to oligondendrocytes and myelination.* Glia, 1996. **17**(2): p. 83-93.
98. Reinert, A., et al., *Iron concentrations in neurons and glial cells with estimates on ferritin concentrations.* BMC Neuroscience, 2019. **20**(1): p. 25.
99. Lambe, T., et al., *Identification of a Steap3 endosomal targeting motif essential for normal iron metabolism.* Blood, 2009. **113**(8): p. 1805-1808.
100. Yambire, K.F., et al., *Impaired lysosomal acidification triggers iron deficiency and inflammation in vivo.* eLife, 2019. **8**.