Loss-of-function of the Zinc Finger Homeobox 4 (ZFHX4) gene underlies a neurodevelopmental disorder

María del Rocío Pérez Baca1,2, Maria Palomares Bralo1,2, Michel Vanhooydonck1,2, Lisa Hamerlinck1,2, Eva D’haene1,2, Sebastian Leimbacher1,2, Eva Z. Jacobs1,2, Laurenz De Cock1,2, Erika D’haenens1,2, Annelies Dheedene1,2, Zoë Malfait1,2, Lies Vantomme1,2, Anamaria Silva6, Kathleen Rooney7, Fernando Santos-Simarro8, Roser Lleuger-Pujol9, Sixto García-Miñaur3,4, Itsaso Losantos-García10, Björn Menten1,2, Gaia Gestri11, Nicola Ragge12, ZFH X4 consortium, Bekim Sadikovic6,7, Elke Bogaert1,2, Delfien Syx1,2, Bert Callewaert1,2, & Sarah Vergult1,2*

*shared first authors
°shared last authors

1Center for Medical Genetics Ghent, Ghent University Hospital, Ghent, Belgium
2Department of Biomolecular Medicine, Ghent University, Ghent, Belgium
3CIBERER-ISCIII and INGEMM, Institute of Medical and Molecular Genetics, Hospital Universitario La Paz, Madrid, Spain
4ITHACA- European Reference Network, Spain
5Department of Diagnostic Sciences, Ghent University, Ghent, Belgium
6Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada
7Verspeeten Clinical Genome Centre, London Health Sciences Centre, London, ON, Canada
8Unit of Molecular Diagnostics and Clinical Genetics, Hospital Universitari Son Espases, Health Research Institute of the Balearic Islands (IDISBa), Palma, Spain
9Hereditary Cancer Program, Catalan Institute of Oncology, Doctor Josep Trueta University Hospital; Precision Oncology Group (OncoGIR-Pro), Institut d’Investigació Biomèdica de Girona (IDIBGI), Girona, Spain
10Department of Biostatistics. Hospital Universitario La Paz, Madrid, Spain
11University College London, London, England, Great Britain
12Birmingham Women’s and Children’s NHS Foundation Trust, Clinical Genetics Unit, Birmingham Women’s Hospital, Lavender House, Mindelsohn Way, Edgbaston, Birmingham B15 2TG

Abstract

8q21.11 microdeletions encompassing the gene encoding transcription factor ZFHX4, have previously been associated with us with a syndromic form of intellectual disability, hypotonia, decreased balance and hearing loss. Here, we report on 57 individuals, 52 probands and 5 affected family members, with protein truncating variants (n=36), (micro)deletions (n=20) or an inversion (n=1) affecting ZFHX4 with variable developmental delay and intellectual disability, distinctive facial characteristics, morphological abnormalities of the central nervous system, behavioral alterations, short stature, hypotonia, and occasionally cleft palate and anterior segment dysgenesis. The phenotypes associated with 8q21.11
microdeletions and ZFHX4 intragenic loss-of-function variants largely overlap, identifying ZFHX4 as the main driver for the microdeletion syndrome, although leukocyte-derived DNA shows a mild common methylation profile for (micro)deletions only. We identify ZFHX4 as a transcription factor that is increasingly expressed during human brain development and neuronal differentiation. Furthermore, ZFHX4 interacting factors identified via IP-MS in neural progenitor cells, suggest an important role for ZFHX4 in cellular and developmental pathways, especially during histone modifications, cytosolic transport and development. Additionally, using CUT&RUN, we observed that ZFHX4 binds with the promoter regions of genes with crucial roles in embryonic, neuron and axon development. Since loss-of-function variants in ZFHX4 are found with consistent dysmorphic facial features, we investigated whether the disruption of zfhx4 causes craniofacial abnormalities in zebrafish. First-generation (F0) zfhx4 crispant zebrafish, (mosaic) mutant for zfhx4 loss-of-function variants, have significantly shorter Meckel’s cartilages and smaller ethmoid plates compared to control zebrafish. Furthermore, behavioral assays show a decreased movement frequency in the zfhx4 crispant zebrafish in comparison with control zebrafish larvae. Although further research is needed, our in vivo work suggests a role for zfhx4 in facial skeleton patterning, palatal development and behavior.

Introduction

Neurodevelopmental disorders (NDDs) affect approximately 2 to 5% of the children worldwide. In up to 40% of the affected individuals, a monogenic cause can be identified, but many gene-disease relationships in NDDs remain to be discovered\(^1\). In 2011, Palomares et al. reported eight unrelated individuals with a recognizable intellectual disability (ID) syndrome, all sharing a (sub)microscopic deletion in the 8q21.11 region containing ZFHX4\(^2\). Since then, others reported individuals with 8q21.11 deletions ranging from 0.66 to 14.5 Mb in size, all with similar clinical characteristics\(^3\-7\). In 2019, a loss-of-function ZFHX4 splice variant was reported in a 7-year-old male from a cohort of patients with apraxia of speech\(^8\). In 2020, ZFHX4 was identified as a novel NDD candidate gene in a large-scale exome-sequencing study\(^9\). Hence, ZFHX4 haploinsufficiency might be one of the drivers of the 8q21.11 microdeletion phenotype. In addition, Bishop et al. identified ZFHX4 as a novel orofacial cleft candidate gene\(^10\). In both exome driven studies, de novo ZFHX4 loss-of-function (LoF) variants appeared to be enriched in the affected individuals. In alignment with these data, the latest gnomAD version (v4.1.0) predicts ZFHX4 to be intolerant to protein truncating variants (PTVs), with a pLI score (probability of being loss-of-function intolerant) of 1, and a constraint score of observed/expected of 0.24\(^11\).

ZFHX4 comprises 12 exons and encodes a transcription factor that carries 22 zinc fingers and four homeodomains\(^12\). Expression of ZFHX4 is high in the developing brain of different species\(^13\-5\) and shown to decrease in adult mice \(^13\). In addition, Zfhx4 is highly expressed in cartilage\(^17\). In mice, Zfhx4 deficiency is associated with cleft palate, and its expression in the palatal shelves supports a role for Zfhx4 in palatal development\(^17\). Homozygous Zfhx4 knockout mice die of neurogenic respiratory disorders within 24 hours after birth\(^18\). In humans, ZFHX4 expression has been detected in adult brain, muscle, and liver\(^12\). Altogether,
ZFHX4 has been identified as a transcription factor involved in neural cell maturation, region-specific brain differentiation\textsuperscript{14} and skeletal development\textsuperscript{17}.

Moreover, in accordance with a role in cell maturation, ZFHX4 appears to be involved in the biology of different types of cancer, such as glioblastoma\textsuperscript{19}, esophageal squamous cell carcinoma (ESCC)\textsuperscript{20,21} and ovarian cancer\textsuperscript{22,23}. Chudnovsky et al. showed that ZFHX4 regulates therapy-resistant tumour initiating (TIC) glioblastoma cell state by interacting with CHD4, a core member of the nucleosome remodeling and deacetylase (NuRD) complex\textsuperscript{19}. Additionally, several studies showed that in ESCC and ovarian cancer, ZFHX4 expression can be used as a prognostic biomarker for survival outcomes in patients with metastasis\textsuperscript{22}.

Here, we report a cohort of 57 individuals with ZFHX4 aberrations. We evaluate the common features and differences between ZFHX4 protein truncating variants (PTVs) and copy number variants (CNVs) including ZFHX4 to inform clinical management. Our observations confirm that ZFHX4 loss-of-function drives the microdeletion phenotype. Furthermore, we evaluated the protein and DNA interactome of ZFHX4, indicating an important role for ZFHX4 in normal cell and tissue maturation, and in neurodevelopmental pathways including axon development. Finally, zebrafish zfhx4 crisants show craniofacial malformations and behavioral abnormalities. Overall, our data indicate that ZFHX4 is essential for neural and craniofacial development.

**Material and methods**

**Patient ascertainment**

Individuals with ZFHX4 alterations were identified through Clinical Genetics Services worldwide using web-based databases such as GeneMatcher\textsuperscript{24} and DECIPHER. The clinical research activity relating to this report has been in accordance with the Declaration of Helsinki on the Ethical Principles for Medical Research Involving Human Subjects. Genetic tests were performed clinically or using research protocols approved by each center ethics committee. Informed consent for genetic analyses, clinical data and biological sample collection, and publication of relevant findings were obtained for all individuals. We collected the detailed phenotype data through a standard form specific for ZFHX4 (Table S1 and S2).

**Statistics**

Statistical studies were performed in SAS 9.4 (SAS Institute, Cary, NC, USA). Comparative studies between females and males and between CNV and LoF variants were calculated using two-tailed Fischer's exact test. P values <0.05 were considered statistically significant. Counts for clinical characteristics were given only for those where the respective information was available, therefore the total count can be lower than the total number of individuals.

**Facial feature analysis**

In order to make the description of facial features as homogeneous as possible, the facial features of all individuals were evaluated by two experienced dysmorphologists, Fernando Santos-Simarro and Sixto García-Miñana. The DeepGestalt technology by Face2 Gene
(Face2Gene, FDNA Inc., Boston, MA) was used to validate the presence of distinct facial patterns. The cohort was divided into two groups to analyze facial features. The first group included a total of 31 frontal photographs of 20 individuals with CNVs. The second group included 26 facial photographs of 18 individuals with predicted LoF variants in ZFHX4. In both cases, age-, sex- and ethnicity-matched healthy individuals were used as a comparison cohort to generate a composite image of the affected individuals. The composite image (not included as a figure) was generated in both cases using DeepGestalt facial analysis and an assessment of the binary comparison between controls and affected individuals was performed by measuring the receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) as described by Mak et al. 2021.

Cell culture

Fibroblast culture, treatment and RNA extraction

Fibroblasts derived from skin biopsies were cultured in DMEM medium supplemented with Fetal Calf Serum (FCS), Penicillin/Streptomycin, Kanamycin, Non-Essential Amino Acids (NEAA) and Fungizone until we had at least 600,000 cells. Afterwards, 300,000 cells were seeded in a P60 plate on Day 1 for cycloheximide treatment. On Day 2 the culture medium was replenished with or without cycloheximide (5μg/mL, C7698, Sigma Aldrich) and on Day 3 RNA was extracted with the Maxwell. In short, the cells were loosened with a cell scraper and transferred to a 15mL tube, which was then centrifuged for 3 minutes at 300rcf. Afterwards the supernatant was removed and 200μL 1-thioglycerol/homogenization solution (Promega) was added. The cells were then vortexed and 200μL lysis buffer (Promega) was added. Finally, the sample (400μL) was transferred to a well from the Maxwell and RNA was extracted with the RSC Simply RNA Tissue kit (Promega) following the RSC Simply RNA Cells method.

Human embryonic stem cells (hESCs)

The naïve hESCs were cultured according to the protocol described in Pérez Baca et al, 2024.

Induced pluripotent cells (hiPSCs)

The iPSC line was commercially obtained from Axol BioScience and maintained in mTESR™ Plus (Stemcell Technologies, #100-0276) on coated cultured plates with Geltrex® LDEV-Free hESC-Qualified Reduced Growth Factor Basement Membrane Matrix (GibcoTM, ThermoFisher), according to the manufacturer’s instructions. The induced pluripotent stem cells were cultured at 37°C in a 5% CO₂ humid atmosphere. Double volume was replaced every two days as a fixed feeding interval.

Neural progenitor cells (NPCs)

The NPCs were differentiated starting with the above mentioned hiPSCs 0.5 x 10⁶ on coated cultured plates with Geltrex® LDEV-Free hESC-Qualified Reduced Growth Factor Basement Membrane Matrix (GibcoTM, ThermoFisher), proceeding with the monolayer protocol using the STEMdiff™ SMAD™ neural induction kit (Stemcell Technologies, #05835) according to the manufacturer’s instructions for 6-9 days and further cultured up to 18-21 days prior to freezing. The NPCs were cultured at 37°C in a 5% CO₂ humid atmosphere. Daily full-medium change as a fixed feeding interval until ready to be passaged.
**Hi-C library preparation and analysis**

A Hi-C library was generated according to the *in situ* Hi-C protocol adopted by the 4D Nucleome consortium with a few adaptations. Briefly, five million patient fibroblasts were crosslinked using 2% formaldehyde, lysed, snap frozen in liquid nitrogen and stored at -80°C until further processing. Pre-lysed, crosslinked nuclei were then digested overnight using 250 U *DpnII* restriction enzyme (New England Biolabs, R0543L). DNA ends were marked by incorporating biotin-14-dATP (Life Technologies, 19524-016) and ligated for 4 hours using 2000 U T4 DNA ligase (New England Biolabs, M0202L). Subsequently, crosslinks were reversed overnight using proteinase K (Qiagen, 19131) and Hi-C template DNA was purified using 1x AMPure XP beads (Becton Dickinson, A63881) and stored at 4°C until library preparation. Hi-C template DNA was sheared to a size of 300-500 bp using microTUBE snap-caps (Covaris, 520045) in a Covaris M220 sonicator and MyOne Streptavidin T1 beads (Life Technologies, 65601) were used to pull down biotinylated ligation junctions. Next, the sample was split into 5 µg aliquots for sequencing library preparation using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, E7645L) and NEBNext Multiplex Oligos (New England Biolabs, E7335L). The amplified library was purified, and size selected using 0.55x and 1.2x AMPure XP beads (Beckman Coulter, A63881). Pooled libraries were sequenced on an Illumina NovaSeq 6000 using 100 bp paired-end reads to a depth of ~500 million reads.

FASTQ files containing raw sequencing data were processed into a Hi-C contact matrix containing both raw and normalized counts using the Juicer pipeline (v1.6) with BWA-MEM mapping (v0.7.17) to the hg38 reference genome. A processed and normalized Hi-C matrix for control fibroblasts was obtained from Melo *et al.*. We used FAN-C to visualize Knight-Ruiz (KR) normalized and observed/expected patient vs. control Hi-C matrices for the region of interest.

**Expression profiling**

RNA of cultured cells was isolated using the Direct-zol™ RNA MiniPrep Kit (Zymo Research) according to manufacturer’s instructions, and the concentration determined by NanoDrop (Thermofischer). cDNA synthesis was performed using the iScript Advanced cDNA synthesis kit (BioRad) according to manufacturer’s instructions. Subsequently, RT-qPCR was performed with a PCR mix containing 5 ng cDNA, 2.5 µl of SsoAdvanced SYBR qPCR mix (BioRad) and 0.25 µl of forward and reverse primers (250 µM concentration, Integrated DNA Technologies). The RT-qPCR was performed on a LC-480 device (Roche) and gene expression levels were analyzed using the qBase+ software 3.2 (Biogazelle). The software qBase+ uses a generalized model of the delta-delta-Ct approach, thereby supporting the use of gene-specific amplification efficiencies and normalization with multiple reference genes, with up to 204 genes. All formulas of this model are detailed in Hellemans et al., 2007. The RT-qPCR primers used in this study can be found in Table S12. Three biological and two technical replicates were collected per timepoint.

**Cell harvesting & protein extraction**

The cells were scraped in PBS and collected in 15ml falcons. The collected cells were centrifuged at 4°C during 5 minutes at 2000rpm. The RIPA buffer (250 mg sodium deoxycholate, 40.5 ml water, 5M NaCl, 1M TrisHCl, 10% SDS solution, 10% NP-40) is used to lyse the cells for 45 minutes, following centrifugation (10min, 4°C, 8000g) the proteins were
extracted. The concentration was determined using the BCA protein assay kit (Thermo Fisher, catalog number 23225).

**Immunoprecipitation followed by mass-spectrometry (IP-MS)**

Protein from 15 to 20 million cells were extracted as mentioned above (Materials). Next, 2μg antibody (Anti-ZFHX4 antibody HPA023837, Rabbit (DA1E) mAb IgG XP® Isotype Control #3900S for WT/IP vs IGG1P sample groups, respectively) was added and after four hours, 20 μl extensively washed protein A Ultralink resin beads (Thermo Scientific, catalog number 53139) were added, further rotated overnight. The samples were centrifuged and washed five times with 50mM digestion buffer (50 mM Tris HCl pH 8, 2 mM CaCl$_2$) to remove unbound proteins. Beads containing antibody and proteins of interest were resuspended in 175 μl digestion buffer. For IP, 25 μl of samples were incubated at 95°C for 10 min with 6.25 μl of 5x laemmli buffer at 1000 rpm to elute the proteins. Supernatants containing the proteins of interest were used for western blot analysis. For IP-MS, the remaining 150 μl of the samples was incubated for 4 hours with 1 μg trypsin (Promega) at 37°C, while shaking. Beads were removed by centrifugation at maximum speed and proteins were digested again with 1 μg of trypsin, overnight at 37°C.

Peptides were re-dissolved in 20 μl loading solvent A (0.1% trifluoroacetic acid in water/acetonitrile (ACN) (98:0.5, v/v)) of which 2 μl was injected for LC-MS/MS analysis on an Ultimate 3000 ProFlow nanoLC system in-line connected to an Orbitrap Fusion Lumos mass spectrometer (ThermoFischer scientific). Trapping was performed at 20 μl/min for 2 min in loading solvent A on a 5 mm trapping column (ThermoFischer scientific, 300 μm internal diameter (I.D.), 5 μm beads). The peptides were separated on a 250 mm Aurora Ultimate, 1.7μm C18, 75 μm inner diameter (Ionopticks) kept at a constant temperature of 45°C in a butterfly heater (Phoenix S&T). Peptides were eluted by a non-linear gradient starting at 1% MS solvent B reaching 26.4% MS solvent B (0.1% FA in acetonitrile) in 30 min, 44% MS solvent B in 38 min, 56% MS solvent B in 40 minutes followed by a 5-minute wash at 56% MS solvent B and re-equilibration with MS solvent A (0.1% FA in water), all at a flow rate of 250 nL/min. The mass spectrometer was operated in data-independent mode, automatically switching between MS and MS/MS acquisition. Full-scan MS spectra ranging from 390-910 m/z with a target value of 4E5, a maximum fill time of 50 ms and a resolution at of 60,000 were followed by 30 quadrupole isolations with a precursor isolation width of 10 m/z for HCD fragmentation at an NCE of 34% after filling the trap at a target value of 4E5 for maximum injection time of 54 ms. MS2 spectra were acquired at a resolution of 30,000 at 200 m/z in the Orbitrap analyser without multiplexing. The isolation intervals were set from 400 – 900 m/z with a width of 10 m/z using window placement optimization. QCloud$^{32,33}$ has been used to control instrument longitudinal performance during the project.

**MS data analysis**

LC-MS/MS runs of all samples were searched together using the DiaNN algorithm (version 1.9), library free. Spectra were searched against the human protein sequences in the Swiss-Prot database (database release version of 2022_01), containing 20,588 sequences (www.uniprot.org) and the imbedded protein contaminant database. Enzyme specificity was set as C-terminal to arginine and lysine, also allowing cleavage at proline bonds with a
maximum of two missed cleavages. Variable modifications were set to oxidation of methionine residues and acetylation of protein N-terminals. Mainly default settings were used, except for the addition of a 400-900 m/z precursor mass range filter, the match between runs (MBR) option. Further data analysis of the shotgun results was performed with an in-house script in the R programming language, version 4.2.2. Protein expression matrices were prepared as follows: the DIA-NN main report output table was filtered at a precursor and protein library q-value cut-off of 1% and only proteins identified by at least one proteotypic peptide were retained. After pivoting into a wide format, iBAQ intensity columns were then added to the matrix using the DIAgui’s R package get_IBAQ function (https://rdrr.io/github/mgerault/DIAgui/man/DIAgui.html). PG.MaxLFQ intensities were log2 transformed and replicate samples were grouped. Proteins with less than three valid values in at least one group were removed and missing values were imputed from a normal distribution centered around the detection limit (package DEP)\(^34\) leading to a list of 2,795 quantified proteins in the experiment, used for further data analysis. To compare protein abundance between pairs of sample groups (WTIP vs IGGIP sample groups), statistical testing for differences between two group means was performed, using the package limma\(^35\). Statistical significance for differential regulation was set to a false discovery rate (FDR) of <0.05 and fold change of >4- or <0.25-fold (\(|\log2FC| = 2\)). Results are presented in Supplementary Tables S8 and S9. Z-scored PG.MaxLFQ intensities from significantly regulated proteins were plotted in a heatmap after non-supervised hierarchical clustering.

We analysed the significantly enriched genes from the MS data analysis with thousands of annotated gene set libraries retrieved from high-quality projects with the EnrichR tool\(^36-38\), and the results are visualized with a horizontal bar graph where the bars plotted horizontally show the gene sets found in different libraries and their corresponding pathways along the x-axis with the combined score with the measured p-values, q-values and z-scores which represents the measure of significance of enrichment level of the associated pathways.

**Western Blot**

Western blot was performed according to an in-house protocol. Either protein lysates or eluted proteins after IP were subjected to NuPage gel (ThermoFischer Scientific, catalog number: EA0375BOX) electrophoresis, subsequently transferred to a nitrocellulose membrane, and detected using the appropriate antibodies: anti-ZFHX4 antibody (Sigma-Aldrich, HPA023837), anti-ZFHX4 antibody (Abnova, H00079776-M02), anti-vinculin (Merck Life Sciences, V9131). Blots were stripped twice. Imaging was performed with an Amersham Imager 600 CCD camera (GE Healthcare, Chicago, Illinois, USA).

**Zebrafish experiments**

**Breeding**

Zebrafish lines were housed at the zebrafish facility Ghent in semi-closed recirculating housing systems (ZebTEC and WTU systems, Tecniplast) at a constant temperature (27-28°C), pH (~7.5), conductivity (~550µS) and light dark cycle (14/10). Zebrafish were fed every day with dry food (Gemma Micro) and with artemia (Ocean Nutrition). Zebrafish lines were maintained according to standard protocols\(^39,40\). Embryos/larvae were raised in E3 medium\(^41\) in a petri dish and placed in an incubator at 28°C, until 5 days-post-fertilization (dpf), the age at which experiments were performed.

*zfhx4 crispant RNP complex generation & injections*
We followed the protocol from Kroll et al., 2021 for the generation of F0 knockout larvae for a single gene. Following the proper guidelines and by using CRISPOR, we were able to find the best corresponding gRNAs for exons 2, 3 and 4 of zfhx4. The crRNAs were designed and produced by IDT (Integrated DNA Technologies) and checked in InDelphi. The possible off targets with the highest probability found by CRISPOR, were assessed via PCR and sequencing on a Miseq instrument (Illumina). Sequences of the zfhx4 crRNAs, scrambled crRNAs and the corresponding primers used for genotyping of the injected embryos, are mentioned in Table S13.

To assemble the RNP, 2 nmol crRNA and 5 nmol tracrRNA pellets were respectively resuspended in 10 or 25 μL of the corresponding Duplex buffer (IDT), to obtain a 200 μM stock of each. These were stored at -20°C and were aliquoted to protect the RNA from the continuous thaw-freeze cycles and longer storage time. The crRNA was annealed with the tracrRNA to form the gRNA by mixing 1 μL crRNA with 1 μL of tracrRNA and 1.51 μL Duplex buffer. To obtain the duplex crRNA:tracrRNA (61 μM), this mix was heated to 95°C for 5 min, then cooled on ice. Each crRNA:tracrRNA complex was individually mixed with an equal molar amount Cas9. Each RNP complex was incubated 5 min at 37°C and had a concentration of 30.5 μM. The three RNP solutions were pooled in equal amounts (10.1 μM, 2 μL) making a total mix of 6 μL. The RNPs were stored at −20°C before the injections.

Around 1 nL (first experiment) and 1 nL (second repeat experiment) of the three-RNP pool was injected into the single-cell yolk before cell inflation. This amounted to around 5029/800 pg of Cas9 and 1070/1000 pg of total gRNA. Each unique RNP was present in equal amounts in the pool. Therefore, in the case of three RNPs, 10.1/9.5 fmol of each RNP were injected in total with the Femtojet injector. The solution contained a dye to follow the successful injections.

Behavioral studies
Locomotor activity in response to light and dark stimuli was analyzed in zebrafish larvae (5 dpf) using the DanioVision Observation Chamber (Noldus, The Netherlands) equipped with a Basler GenICam camera capturing 30 frames/second for image acquisition. Data was recorded and analyzed using the EthoVision XT 15 software (Noldus, The Netherlands). At 5 dpf, zfhx4 crisprants and scrambled zebrafish were transferred individually to a 24-well plate containing 1ml of E3 medium. Plates were subsequently placed in the DanioVision Observation Chamber with a protocol consisting of 20 min of 5% light (corresponding to 139 lux), 20 min dark, followed by alternating periods of 5 min 5% light and 5 min dark, for a total of 4 cycles. During the experiment, the temperature was kept constant at 28 °C using a Temperature Control Unit (Noldus). The first 10 minutes of the experiment were considered an acclimatization period and were not used for analyses. For each zebrafish, center-point detection was used. Movement frequency is depicted in time intervals of 1 min. This behavior experiment was repeated twice with n=60/genotype.

Genotyping
From each experimental and control condition, 30 5 dpf larvae were collected and DNA was extracted. The PCR reactions were performed on the exons of the sequence of interest (Table S13), following paired-end 2x150 bp sequencing on a Miseq instrument (Illumina) to
determine the knockout efficiency following the protocol by Clement et al. The primers were used in the following PCR reaction: 2' 94°C, 12x (30° 94°C, 30° 60°*C, 1° 72°C), 25x (40° 94°C, 40° 48°C, 30° 72°C), 10° 72°C) PCR reaction. The symbol “*” indicates in each cycle, lowering of the temperature by 1°C. The primers, target sequences and knockout efficiencies can be found in Table S13.

**Alcian blue stainings and imaging**

The following protocol is based on the published work from Stephan C.F NeuHauss and Delbaere et al., 2020. Image analysis was performed to confirm the phenotype-genotype of crispsants in comparison with scrambled zebrafish based on measuring different structures on the head of the 5 dpf injected larvae with the Leica fluorescent microscope M165 FC and a microscope light source. The magnification of the images was 12x. These stainings and measurements were repeated twice. We evaluated the length of the Meckel’s, the ethmoid plate, the distance between the middle point of the Meckel’s cartilage to the frontal point on the basihyal cartilage and the angle, the length of each ceratohyal cartilage and the angle between them with Fiji (Image J). Each measurement was normalized to the length of the middle point of both Meckel’s structures to the parachordal structure.

**DNA methylation analysis**

DNA was obtained from the peripheral blood of individuals with ZFHX4-related syndrome and control subjects. Subsequently, a bisulfite-based technique was employed, and the levels of methylation were evaluated utilizing Illumina Infinium EPIC bead chip arrays (San Diego, CA, USA) following the prescribed procedures by the manufacturer. The assessment of sample quality, as well as the importation of normalized methylation and non-methylation signal intensity data, was conducted using the R minfi package (version 1.44.0).

The methylation data underwent analysis via a standard pipeline, outlined by Aref-Eshghi et al., within the R environment (version 4.2.2). To ensure that disparities noted between the case and control cohorts were solely due to alterations in methylation rather than other potentially influencing factors, probes located on the X and Y chromosomes, probes prone to cross-reactivity with other genomic regions, and those with a detection p-value > 0.01 were removed from the analysis.

Probes exhibiting beta values of 0 and the top 1% most variable probes in terms of variance within either the case or control samples were excluded. Methylation levels, represented as beta values, underwent conversion into M-values through logit transformation. These transformed values were then employed for linear regression modeling utilizing the limma package (version 3.54.2).

The control samples were chosen from the EpiSign™ Knowledge Database (EKD) via the R software package Matchit (version 4.5.2). Controls were matched by age, sex, and array type, resulting in the identification of 56 samples.

The model matrix was refined by incorporating estimated proportions of diverse blood cell types as covariates to address potential confounding influences. Following this adjustment, p-values were refined for statistical significance using the eBayes function within the limma package.
The parameters for probe selection underwent adjustment to enhance the discrimination between the case and control samples. This evaluation utilized hierarchical clustering, employing Ward's method based on Euclidean distance, and multidimensional scaling (MDS), which entailed scaling the pairwise Euclidean distances between samples for visualization purposes.

We employed the R package e1071 (version 1.7-13) to train a support vector machine (SVM) and construct a binary prediction model, following the methodology outlined in the referenced article. The training process involved case samples, matched control samples, 75% of the remaining control samples and other rare genetic disorders from the EKD, while the remaining 25% of samples were reserved for testing purposes.

For each sample, the Methylation Variant Pathogenicity (MVP) score, ranging from 0 to 1, was computed to designate the similarity between its methylation pattern and the identified methylation profile for ZFHX4 case samples. The conversion of SVM decision values into MVP scores was accomplished using the Platt scaling method.

**CUT&RUN sequencing and data analysis**

CUT&RUN was performed using Cutana pA/G-MNase (Epicypher, 15-1016) according to the manufacturer’s protocol. Briefly, cells (0.5M cells/sample) were washed, NE buffer was added to extract the nuclear protein and incubated with activated Concanavalin A beads for 10 min at room temperature. Cells were then resuspended in antibody buffer containing EDTA 0.5M, 1:100 dilution of each antibody (anti-ZFHX4 antibody, HPA023837; Rabbit (DA1E) mAb IgG XP® Isotype Control #3900S; anti-CTCF, Merck 07-729; H3K27ac antibody) was added to individual cell aliquots and tubes were rotated at 4 °C overnight. The following day, targeted chromatin digestion and release was performed with 2.5 mL Cutana pA/G-MNase and 100mM CaCl₂. The genomic DNA was retrieved with the CUTANA™ DNA Purification Kit. Sequencing libraries were prepared with the NEBNext End Prep, following with the adaptor ligation and after size selection with Ampure Beads, the indexes were added, and the fragments are amplified. After the PCR amplification, the clean-up with the Ampure beads predominantly elutes mostly the fragments between 150-500 bp. The sequencing libraries were then sequenced with Illumina Nextseq 2000 using 2x150 bp paired-end reads to obtain 10 million read pairs per sample.

Read processing and peak calling was performed using a dedicated nextflow CUT&RUN pipeline available through nf-core [https://doi.org/10.5281/zenodo.5653535]. Briefly, raw reads were trimmed, QC filtered and aligned to the hg38 reference genome using Bowtie2 (v2.4.4). Upon Q-score filtering, duplicate removal and CPM normalization, peaks in target samples were called with SEACR (v1.3) using corresponding IgG samples as controls. A consensus peak list per sample was obtained by requiring the presence of peaks in both replicates. Downstream annotation of consensus peaks was performed using the ChiPseeker R package.

The top 20 of enriched GO terms (q value < 0.05) is visualized as a dot plot in Figure 6B. The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis provides the annotation of biological pathways, which is represented in Figure 6C. The top 10 of enriched pathways (q value < 0.05) is visualized as a dot plot. Homer37 (v4.11) was used to perform motif enrichment analysis, with 200 bp around the peak summit as input. The top 20 de novo and known motifs which were found from the analyses are summed up in Table S10.
Human expression datasets

We examined the expression of ZFHX4 and its different isoforms in 54 non-diseased adult tissues (see Figure S1). To this end, the ZFHX4 expression levels in terms of TPM (transcripts per million) were retrieved from GTEx (https://www.gtexportal.org/home/). Furthermore, the ZFHX4 expression levels in terms of RPKM (reads per kilobase million) during neurodevelopment and adulthood were retrieved from Brainspan (http://www.brainspan.org, Figure 4). The brain-tissue-specific bulk RNA-seq data was downloaded from their website ("RNA-Seq Gencode v10 summarized to exons"). The brain regions were matched between development stage 2A and the later ones based on Table 2 in their technical white paper59. Read counts of the human embryonic craniofacial tissue bulk RNA-seq by Yankee et al. were accessed through their website (http://cotneyweb.cam.uchc.edu/craniofacial_bulkrna/)60. The ZFHX4 expression levels in human embryonic craniofacial tissue60 are represented in terms of RPKM (reads per kilobase million) from neurocranial crest cells to craniofacial tissue in embryonic periods.

The scRNA-seq neural organoid datasets61 (23 to 180 day of age) were accessed as RDS objects accessed through PsychENCODE on synapse (syn36139275), harmonized and integrated on a high performance computing unit (R version 4.1.0) using the "harmony" package (version 1.2.0)62. Plots were created using the ggplot2 package (version 3.5.2) in R (version 4.3.2)63. Lines in the plots were fitted using the ggplot2 function geom_smooth (method = "loess") on the normalized counts (Figure 4C) or RPKM (Figure 4C). Plot from Figure 4D was created using the ggplot2 function "geom_bar()" with error bars made with stat_summary (fun.data = "mean_sdl").

Results

Individuals harboring a ZFHX4 deletion or protein truncating variant present with a clinically distinctive NDD

Demographic features

Data on 57 individuals (52 probands and 5 affected parents) with an aberration affecting ZFHX4 (i.e. inversion, copy number variants (CNVs) and single nucleotide variants (SNVs) are summarized in Table S3 and more comprehensively documented in Table S1 and S2. There are 33 males and 24 females (Table S4). Ages ranged between five months and around their 50s. The median age at the last examination was around 8 years old.

Genotype

Twenty probands present with a structural variant affecting ZFHX4, including one chromosomal inversion and 19 CNVs. The inversion between genomic positions chr8:69893659 and chr8:76806725 (hg38) in proband 1 results in a breakpoint at chromosome 8q21.11 disrupting the ZFHX4 intron between exons 4 and 5 (Figure 1). The CNVs are between 1.80 kb and 13.03 Mb in size, and comprise 13 multigenic deletions, six intragenic ZFHX4 deletions (less than 90 kb in size), and one intragenic duplication that potentially causes a frameshift. The smallest region of overlap between the multigenic deletions is 361.9 kb (chr8:76444811-76806725 (hg38)) containing ZFHX4 and PEX2 (Figure 1A). Detailed information can be found in Table S1.
Heterozygous frameshift, nonsense, and canonical splice site variants were identified in 23, seven, and two probands, respectively (Table S2). These PTVs were found in exon 2 (probands 21-30), exon 3 (probands 31 and 32), exon 5 (proband 34) and in the penultimate exon 10 (probands 36-51) and are all predicted to result in nonsense-mediated decay (i.e. > 55 bp away from last exon-exon boundary) (Figure 1B). All nonsense variants for which a CADD (combined annotation dependent depletion) score was available, have a score of 35 or higher, indicating that they are within the 0.1% most deleterious possible substitutions in the human genome. Probands 33 and 35 harbor a ZFH4 splice donor variant (ZFH4 (NM_024721.5):c.3093+1G>T) and a splice acceptor variant (ZFH4 (NM_024721.5):c.3965-1G>A), respectively. According to SpliceAI, the c.3093+1G>T variant results in loss of a canonical donor sequence (Figure S2) possibly leading to skipping of exon 3, while the c.3965-1G>A variant results in loss of the canonical acceptor sequence at exon 10 and gain of a novel acceptor causing a frameshift in exon 10 (Figure S2).

Additionally, we identified one individual (proband 52) who is compound heterozygous for a nonsense (ZFH4 (NM_024721.5):c.6528G>A/p.(Trp2176*)) and a frameshift variant (ZFH4 (NM_024721.5):c.6564_6567del/p.(Asn2188Lysfs*12)). Because the phenotypic consequences of the variants in probands 1 and 52 might be different, we excluded both individuals from the clinical characteristics analysis.

Twelve probands (4, 5, 8, 10, 15, 23, 29, 45, 47, 48, 49, and 52) inherited the variant from a similarly affected parent. The inheritance patterns of the ZFH4 variant in probands 21 & 50, mother of proband 23, mother of proband 29, mother of proband 47 could not be determined. In the remaining 38 probands, the aberration occurred de novo (i.e. 15 de novo CNVs and 23 de novo SNVs) (Table S1 & S2).

Only three variants (identified in probands 23, 25, 26 (same variant as proband 25) and 44) were found in the public variant database gnomAD (release v4.1.0). Twenty missense variants present in gnomAD affect the same codon as the truncating variant in probands 31, 32, 34, 36, 38-41, 44, 46-52 (Table S5). While ZFH4 LoF variants appear to be under high constraint (as indicated by a pLI-score of 1 and a LOEUF score of 0.24), missense variants are not (as indicated by a Z score of 1.38), thereby suggesting that these missense variants do not result in LoF. Nevertheless, the presence of pathogenic variants in gnomAD does indicate that phenotypic variability may be wide.

**Facial features**

Dysmorphic facial features were reported in all individuals except for proband 21 for whom we have no data on the facial characteristics and probands 1 and 52 who were excluded from the clinical analysis (n=54/55; 98%). Frequently recurring facial characteristics are a broad lower half of the face (69%), a prominent forehead (71%), biparietal narrowing (47%), laterally sparse eyebrows (68%), eyelid ptosis (64%), periorbital fullness (53%), low-set ears (66%), prominent ear helices (66%), a wide nasal bridge (67%), flared nostrils (54%), underdeveloped nasal alae (64%) and an abnormal upper lip vermillion (51%). Facial characteristics observed in less than half of the individuals were hypertelorism (30%), upslanted palpebral fissures (44%), short palpebral fissures (37%), epicanthal folds (41%), posteriorly rotated ears (34%), a narrow nasal tip (41%), enlarged nares (29%), a smooth philtrum (34%), a thin upper lip (42%) and micrognathia (41%). Less frequently observed
characteristics were microphthalmia (6%), a short philtrum (23%), an everted upper lip (21%), downturned corners of the mouth (20%), and cleft lip and/or palate (24%) (Table S3). A computationally based facial analysis using DeepGestalt facial analysis on 20 individuals with a ZFHX4 deletion and 18 individuals with a LoF SNV in ZFHX4 versus age-, and sex-matched controls, indicates a highly specific and sensitive gestalt with an average area under the curve (AUC) of 0.988 and 0.986 for individuals with a CNV or SNV, respectively, \(p<0.005\). Comparison of the facial features of both affected groups (i.e. ZFHX4 deletion vs ZFHX4 LoF variants) showed no significant difference (AUC=0.656 \(p\) value=0.131, figure not included).

**Clinical findings**

Global developmental delay and/or ID was present in most individuals (81%, \(n=39/48\)). However, in 7/31 individuals (23%), intellectual functioning was reportedly normal. For those with a documented IQ score (n=31), most had mild ID (n=16), two had borderline ID, five had moderate ID and one had severe ID. No statistical differences were found in the levels of ID between individuals with a CNV affecting ZFHX4 and those with ZFHX4 PTVs.

In Table S6, we included other features that were observed in our patients. More than half of individuals (n=30/47) presented with muscular hypotonia. Two third had behavioural difficulties (n=32/48), with stereotypies (n=8) and autism (n=10) being the most frequently reported. Morphological abnormalities of the central nervous system were found in 42% of the individuals, mostly ventriculomegaly (n=5), aplasia/hypoplasia of the corpus callosum (n=3) and widened subarachnoid space (n=3). Ophthalmological abnormalities included strabismus (48%, n=22/46), refraction abnormalities (49%, n=19/39) either myopia (n=8) or hypermetropia (n=5), nystagmus (n=1) and amblyopia (n=2). Notably, 8 individuals had an abnormal anterior eye segment morphology, with astigmatism (n=4), sclerocornea (n=2), microcornea (n=2), anterior segment dysgenesis (n=2), corneal opacity (n=1), and/or persistent pupillary membrane (n=1). Musculoskeletal manifestations included digital abnormalities in 45% (n=22/49) with clinodactyly (n=7) and camptodactyly (n=6) being most frequently reported. Growth abnormalities were described in 45% (n=20/45) of the probands. Prenatally, three individuals presented with intrauterine growth restriction and two with fetal macrosomia. Five individuals showed a postnatal growth retardation and short stature was documented in 27% (n=14/51). Genital abnormalities were found in 28% (n=11/40), including cryptorchidism (n=6), hypogenitalism (n=4) and abnormality of the scrotum (n=3). Congenital heart defects were reported in 17% (n=7/42) of the probands, mainly pulmonic stenosis (n=1), ventricular septal defect (n=1), atrial septal defect (n=2), aortic arch hypoplasia (n=1) and right aortic arch (n=1). About one third of the children (n=15/46) experienced feeding difficulties in infancy. Recurrent infections were noted in 17% (n=7/41).

**Genotype-phenotype correlation**

We compared the frequencies of 27 facial features and 15 clinical features between individuals with CNVs (group 1, n=20) and individuals with PTV variants in ZFHX4 (group 2, n=34). Underdeveloped alae nasi (85% in CNVs, 54.5% in PTVs), abnormality of refraction (82% in CNVs, 31% in PTVs), behavioral problems (87.5% in CNVs, 58% in PTVs) and morphological abnormalities of the central nervous system (87.5% in CNVs, 29% in PTVs)
were more frequent in group 1 (p=0.023, p=0.004, p= 0.040 and p= 0.003 respectively),
while cleft lip and or palate and micrognathia was frequent in group 2 (p=0.013 and p=0.030 respectively). ID is generally more severe in individuals with a multigenic deletion.
Furthermore, cleft lip and/or palate and prominent ear helix are more frequently observed in males versus females (i.e. 34% vs 9% for cleft lip/palate and 83% vs 50% for ear helix),
while the penetrance of other clinical features did not differ between both sexes (p<0.05, Table S4).

A 6.9 Mb de novo inversion involving chromosome bands 8q13-8q21.13 results in changes in 3D chromatin structure and disruption of ZFHX4
Individual 1 presented with a compatible but more severe phenotype (severe intellectual disability, significant dysmorphic facial features, abnormal muscle tone and abnormal heart morphology). We delineated the underlying 6.9 Mb de novo inversion spanning chromosome bands 8q13-8q21.13 at basepair resolution via genome sequencing (GS). This revealed an intergenic breakpoint on the centromeric end of the inversion and a breakpoint within intron 4 of ZFHX4 on the telomeric right end (Figure 2). We next employed Hi-C to evaluate topologically associating domain (TAD) shuffling, revealing the typical bow tie configuration when mapped onto the reference genome (Figure 2). On the centromeric end, the inversion breakpoint resides within a TAD boundary (chr8:69875002-69900000 (hg38)), thereby possibly affecting two small (sub)TADs respectively containing the SLCO5A1 and PRDM14 gene. On the telomeric end of the inversion, the affected TAD contains ZFHX4 and PEX2 (Figure 2).

Patients with a deletion encompassing a ‘critical region’ seem to show a mild methylation profile
Since we previously identified a methylation pattern associated with loss-of-function of ZFHX3, a parologue of ZFHX4, we also evaluated this for ZFHX4 loss-of-function. To this end, we performed methylome analysis on blood derived DNA samples of 26 individuals, including nine individuals with a CNV of ZFHX4 (probands 4, 5 and their mothers, proband 10 and her father, probands 12-14), eleven individuals with a ZFHX4 PTV variant (proband 23 and his mother, probands 24-25, 29, 30, 33, 35, 38, proband 45 and his mother), proband 15 with an intragenic ZFHX4 duplication, proband 20 with an intragenic in-frame ZFHX4 deletion and proband 1 with an inversion affecting ZFHX4, as well as three individuals from the cohort published in 2011 (all listed in Table S7). No specific methylation difference was observed when comparing ZFHX4 truncating variants with matched controls. However, DNA methylation analysis identified 227 CpG probes displaying notable methylation disparities between cases with deletions that included a minimal region (of 53 kb; deletion observed in proband 14) containing the ZFHX4 promoter and matched controls, as evidenced by hierarchical clustering and MDS analyses (Figure 3A).
The support vector machine (SVM) classifier underwent training utilizing signature probes to precisely categorize individuals with ZFHX4-related syndrome, as observed in Figure 3B. It comprised 75% of control samples and other neurodevelopmental disorders (blue dots). The remaining 25% of control samples and other disorders were allocated for testing (grey dots).
The ZFHX4 samples with truncating variants (ZFHX4_truncating), the intragenic ZFHX4 duplication (ZFHX4_duplication), the inversion (ZFHX4_inversion) and the in-frame deletion (ZFHX4_in_frame_del) exhibited low methylation variant pathogenicity (MVP) scores, which differed from the methylation pattern observed in the ZFHX4 deletion samples (Figure 3B). Overall, we observed a distinct methylation profile for the deletion cases overlapping the deletion of proband 14, but not in individuals with intragenic ZFHX4 defects nor LoF variants.

**ZFHX4 expression increases during human neuronal differentiation, brain and craniofacial development**

GTEx v8 ([https://www.gtexportal.org/home/](https://www.gtexportal.org/home/)) indicates abundant ubiquitous ZFHX4 expression throughout the adult human body (Figure S1), except for several brain regions. However, expression profiles across seven organs (including brain and heart) through 23 developmental stages, show highest ZFHX4 expression during the embryonic and fetal period, especially in the human brain and cerebellum (Figure S1). In line with these observations, we confirmed increased expression of ZFHX4 in neural progenitor cells (NPCs) and early developed neurons upon *in vitro* neural differentiation starting from induced pluripotent stem cells (iPSCs) (Figure 4A). This increased expression was also confirmed at the protein level for NPCs in comparison to iPSCs (Figure 4B).

Furthermore, reanalysis of transcriptomic data within the BrainSpan atlas of the developing human brain, also confirmed the highest ZFHX4 expression in the brain between 8-16 weeks post conception, more specifically in the mediodorsal nucleus of the thalamus, striatum and the amygdaloid complex (left, Figure 4C). Through PsychENCODE, we retrieved single cell RNA-seq (scRNA-seq) datasets from developing neural organoids. In the early differentiation stages between day 23 and 45, the highest ZFHX4 expression is observed in the subcortical neurons (left, Figure 4C), while in day 120-150 neural organoids, this shifts to the outer radial glia and intermediate progenitors (right, Figure 4C). In addition, a clear increase in ZFHX4 expression is observed in later embryonic stages of the developing facial bones (Figure 4D), in line with observed facial skeletal changes in patients with ZFHX4 deficiency.

**ZFHX4 interacts with proteins involved in neural tube morphology and cellular organization in neural progenitor cells**

Since NPCs showed increased ZFHX4 expression at the RNA and protein level, we performed immunoprecipitation followed by mass spectrometry (MS) in NPCs to discover potential interaction factors of ZFHX4. Forty-seven significantly enriched proteins were identified compared to an IgG isotype control (t-test FDR<0.05) (Figure 5A and B, Table S8 and S9). Biological process and pathway enrichment with EnrichR ([https://maayanlab.cloud/enrichr-kg](https://maayanlab.cloud/enrichr-kg)) revealed that ZFHX4 interacting proteins play a role in biological processes associated with histone modifications, development and cytosolic transport (Figure 5C). Furthermore, ZFHX4 interacts significantly with proteins encoded by genes associated with neurodevelopmental disorders: ZMIZ1 (MIM#: 618659), ZEB1 (MIM#:613270 & 609141), WDR81 (MIM#: 610185 & 617967), KCNQ2 (MIM#: 613720 & 121200), KMT2D (MIM#:
ZFHX4 binds to the promoter of genes involved in axonogenesis, regulation of the nervous system development, neurogenesis and neural migration

To assess the downstream target genes of the transcription factor ZFHX4, we performed CUT&RUN for ZFHX4 on two NPC replicates, retrieving around 1640 consensus peaks mainly consisting of promoter regions (55.62%) (Figure 6A & Table S10). Furthermore, Gene Ontology analysis of these peaks showed that ZFHX4 binds to the promoter of genes involved in (the regulation of) axonogenesis, forebrain development, cell differentiation nervous system development and neuron differentiation (Figure 6B). KEGG pathway analysis indicated that the proteins encoded by these genes are involved in axon guidance, the Wnt, Rap1 and other signaling networks regulating pluripotency of stem cells. All pathways are strongly involved in cell organization, embryonic and nervous system development (Figure 6C). Motif enrichment analysis of the consensus peaks revealed an enrichment of homeobox, transcription factors, SOX-related genes and zinc finger motifs (Table S11).

zfhx4 F0 zebrafish crispant larvae show craniofacial and behavioral phenotypes

The zebrafish zf hx4 gene (ENSDARG00000075542) resides on chromosome 24, contains 10 exons and has 71.37% sequence identity with the human ZFH X4 (Figure 7A). In 2021, Kroll et al. showed that first-generation (F0) mosaic crispant zebrafish recapitulate the phenotype of germline knock-outs. We generated zf hx4 crispants by injection of a mixture of three crRNAs obtaining a zf hx4 indel efficiency of approximately 79% (see Table S13).

To determine craniofacial cartilage defects, we stained neural crest-derived chondrocytes in 5dpf zf hx4 crispant larvae and scrambled crRNA control larvae with Alcian Blue (Figure 7B). zf hx4 F0 crispant larvae had a significantly smaller ethmoid plate and shorter Meckel’s cartilages in comparison with scrambled larvae ((p-value ≤ 0.001 (**), p-value ≤ 0.0001 (****), Figure 7C), thereby confirming that ZFHX4 LoF results in craniofacial anomalies.

Next, locomotor behavior was assessed in response to light/dark transitions using the Daniovision Observation Chamber. Analysis of the swimming pattern of scrambled zebrafish showed a higher movement frequency during the dark cycles. Zfhx4 crispant zebrafish showed a movement pattern comparable to scrambled zebrafish during the light and dark cycles, but with a consistently lower movement frequency (Figure 8D and Figure S6, p-value ≤ 0.05 (*), p-value ≤ 0.0001 (****)).

Discussion

In this study, we delineated a clinically distinctive neurodevelopmental phenotype caused by pathogenic variants in ZFHX4. Most individuals presented with mild to moderate intellectual disability and behavioral problems, often associated with anatomical brain anomalies. All assessed individuals showed a recurrent facial gestalt. Short stature, digital anomalies,
Genital defects, cleft lip/palate, ophthalmological anomalies and feeding difficulties were present in 25-50% of the affected individuals, while congenital heart defects were occasionally observed. The high phenotypic similarity, including the facial characteristics, between individuals with CNVs including ZFHX4 and individuals with intragenic ZFHX4 variants indicates that ZFHX4 loss-of-function is the mechanism underlying the common phenotype. However, individuals with CNVs appear to have an increased risk for behavioral abnormalities (Fisher’s exact test p=0.0402), while the prevalence of cleft lip/palate was lower compared to individuals with PTV variants in ZFHX4. Furthermore, the methylation profile identified in individuals with 8q21.11 (micro)deletions encompassing ZFHX4 could not be replicated in individuals with PTVs. Although other genes contained in the 8q21.11 (micro)deletions (Table S14) might contribute to or modify the methylation pattern, the smallest region of overlap among all deletion samples tested for a methylome signature, is the size of the deletion in individual 14 (53 kb). This deletion does not only partially delete ZFHX4, but also removes a non-coding region containing the ZFHX4 promoter. In the other tested individuals with a ZFHX4 PTV, an intragenic duplication, an intragenic deletion or an inversion, the aberration does not result in the removal of this critical region. This suggests that deletion of this critical region might be the main contributor to its pattern. To confirm a possible role for this ‘critical’ region as one of the drivers for the methylation profile, analysis of additional individuals is necessary.

Of note, twelve variants (i.e. six CNVs and six SNVs) were inherited from healthy parents, of whom three also had learning difficulties in childhood, although without a formal diagnosis. Although ZFHX4 is very intolerant for LoF variants (pLI=1; LOEF=0.24), 172 LoF variants (excluding low coverage ones) are present in gnomAD (v4.1.0), including three variants present in our cohort (albeit all with a very low frequency in the general population). As we observed a wide phenotypic spectrum in individuals with ZFHX4 CNVs and PTVs, analysis of ZFHX4 should not be limited to the identification of de novo aberrations. Moreover, careful recording of the phenotypic data of the parents remains important.

Individual 1 harbors a de novo inversion disrupting ZFHX4 and presented with a typical but more severe phenotype. Via Hi-C, we observed that the inversion affects two small (sub)TADs on the centromeric end, containing SLC05A1 and PRMD14 gene. At the telomeric end, the affected TAD contains ZFHX4 and PEX2 (Figure 2). SLC05A1 and PEX2 have been previously linked to disease. PEX2 is associated with the peroxisome biogenesis disorder 5A and 5B, both autosomal recessive multiple congenital anomaly disorders (MIM 614866 & 614867). SLC05A1 on the other hand, has been linked to the autosomal dominant Mesomelia synostoses syndrome (MSS, MIM: 600383)\(^{79,80}\). MSS is characterized by mesomelic shortening of the limb, acral synostoses and multiple congenital malformations. In 2010, Isidor et al. reported 8q13 deletions associated with MSS\(^{81}\). Although these deletions varied in size, they always encompassed SULF1 and SLC05A1. Furthermore, Roshandel et al. generated an siRNA mediated knockdown of Oatp30B (i.e., the Drosophila orthologue of SLC05A1) and observed a startling and seizure-like phenotype\(^{82}\). Although SLC05A1 is not directly affected in individual 1, the inversion might exert an effect on the expression of SLC05A1, thereby combining both the ZFHX4 and MSS phenotypes. However, further functional studies are necessary to unravel this.
Although previous reports suggested ZFHX4 as a candidate gene for non-syndromic orofacial clefting, we observed orofacial clefting in only about 30% of the individuals with ZFHX4 LoF variants. We investigated the craniofacial cartilage development in mosaic knockout zfhx4 zebrafish and observed significant differences between the Meckel’s cartilage and the ethmoid plate structures that correspond to the human jaw and palate, respectively. A homozygous Zfhx4 knock-out mice model further supports this hypothesis, presenting cleft palate at 100% penetrance. In addition to the craniofacial differences in our zebrafish model, we also observed behavioral changes. The decreased movement frequency could be attributed to (a combination of) neurological, skeletal and/or muscle defects in the zfhx4 crispant zebrafish larvae, given the role of ZFH X4 in different biological processes of these tissues. Interestingly, the International Mouse Phenotyping Consortium (IMPC) applied a panel of phenotyping screens on Zfhx4 homozygous and heterozygous knock-out mice. Similar to what was observed by Nakamura et al., the homozygous Zfhx4 deficient mice showed full penetrant preweaning lethality. However, the heterozygous Zfhx4 deficient mice were viable (11 females and 13 males), but did not present with cleft palate, and showed decreased exploration in a new environment. Our in vivo zebrafish results as well as the previous studies in mice, clearly indicate a relationship between ZFHX4, clefting and behavioral abnormalities.

The complex structure of ZFHX4, containing 4 homeodomains and 22 zinc fingers, suggests dynamic functions in diverse biological processes. Our IP-MS screen identified ZFHX4 protein interactors involved in the regulation of histone modifications, development, transport, gene expression and cell division. Most of the interacting factors are associated with histone modifications and development. YEATS2, Z33, KMT2D and ZMYND8 play a crucial role in chromatin remodeling, histone modifications, such as acetylation. Although CHD4, a crucial member of NuRD complex, was previously identified as ZFHX4 interacting protein in glioblastoma cells, it was not identified as an interaction partner in our NPC screen. However, we identified the CHD4-interacting multivalent histone reader ZMYND8 as an interacting factor of ZFHX4. Furthermore, in 2022, Dias et al. reported disruption of ZMYND8 as a cause for syndromic intellectual disability. Moreover, several of the other ZFHX4 interacting proteins have also been associated with NDDs: YEATS2 plays an important role in histone acetylation and has been involved with epilepsy (MIM: 615127; epilepsy, myoclonic, familial adult). ZEB1 has been associated in corneal dystrophy (MIM: 613270 and 609141). Another interacting factor is KCNQ2 (MIM: 613720; developmental and epileptic encephalopathy), encoding a potassium voltage-gated channel, which plays a key role in the regulation of neuronal excitability. Mutations in KCNQ2 have been associated with self-limited familial neonatal epilepsy (SeifNE) and other forms of epilepsy (early-onset developmental and epileptic encephalopathy). Additionally, mutations in KMT2D, encoding a histone methyltransferase give rise to Kabuki syndrome (MIM:147920, Kabuki syndrome 1), associated to neural crest cell differentiation and facial morphology. IMPA2 plays an important role in phosphatidylinositol signaling pathway and has been associated with susceptibility to bipolar disorder. Furthermore, GAMT is crucial for creatine biosynthesis and associated with cerebral creatine deficiency syndrome 2 (MIM:612736). Other interacting factors, such as NPTX2 and ZNF592, are involved in synaptic plasticity and neuronal development.
Furthermore, in NPCs, ZFHX4 mainly binds to the promoter regions of genes that are significantly enriched for homeobox, SOX-related genes and zinc fingers motifs according to de novo and known HOMER motif analysis (Table S11). Via SEA motif enrichment, the same consensus peaks seem to show ERF-, KLF (Krüppel-like factors) and SP (specificity protein) motif enrichment (Table S11). These transcription factors are crucial for transcriptional regulation and cell proliferation\textsuperscript{83,84}. Mutations in ERF genes have been associated with craniosynostosis with facial dysmorphism\textsuperscript{62}, while KLF and SP genes are involved in embryonic development and tissue differentiation\textsuperscript{83}.

In conclusion, we have delineated the ZFHX4-related NDD and identified ZFHX4 haploinsufficiency as a cause of autosomal dominant syndromic intellectual disability. Our initial findings indicate an important role for ZFHX4 in craniofacial and brain development, through interaction with proteins involved in chromatin organization.

**Figures**

**Figure 1: Overview of ZFHX4 aberrations in 52 probands and 5 affected family members.** (A) (Micro)deletions identified in probands 1–20 (#) are represented by the red bars and the duplication in proband 16 by the purple bar, all affecting ZFHX4 (light blue box). ReSeq coding genes are indicated respectively in blue\textsuperscript{131}. The numbers colored in black corresponding to their CNV, represent an individual with a training DNA sample for methylome analysis. The numbers colored in grey corresponding to each CNV variant, are individuals with their DNA samples used as testing samples for the methylome analysis. Genomic positions are according to hg38; ZFHX4 is located on the reverse strand. (B) **Nonsense, frameshift and splice variants in ZFHX4 are identified in probands 21-52 and visualized with IBS 2.0**\textsuperscript{132}. Protein and gene structure of the canonical transcript (GenBank: NM_024721.5) are shown in the top and bottom part, respectively. The majority of PTVs are located in exon 2 and 10. Top: LoF variants are present in the ZFHX4 cohort and located in different domains of ZFHX4. Blue dots represent the frameshift variants, the yellow dots the nonsense variant and the brown-grey dot represents the variants of proband 52 which contains a nonsense (Trp2176*) and a frameshift (Asn2188*fs) mutation. The dot has been placed on the position of the nonsense variant (Trp2176*). Bottom: coding exons are indicated as light blue boxes and the UTRs are represented with white boxes.

**Figure 2: Hi-C on proband 1 fibroblasts reveals disrupted chromatin structure at inversion breakpoints.** Hi-C contact frequency matrix for control and proband 1 fibroblast (inversion indicated by red arrows). Close-up of left and right breakpoint region, including observed/expected (O/E) contact frequencies for control and proband 1.

**Figure 3: (A) The Euclidean clustering plot depicts ZFHX4 cases in red and the matched controls are shown in blue.** The rows represent selected probes for the identified methylation pattern, while columns represent training cases and controls. A clear separation between cases and controls was observed. Methylation levels are color-coded: blue for 0 beta values and red for 1 beta values, with intensity indicating methylation level intensity. (B) **The MDS plot demonstrates the differentiation between case and control samples based on selected probes.** Cases are represented by red circles and controls by blue circles. C. Utilizing SVM classification trained with ZFHX4 samples against matched controls, as well
as 75% of other control samples and samples from other disorders (blue), in addition to the remaining 25% of the database samples used for testing (gray), produced significant findings. Except for the ZFHX4 deletion samples (training), all other conditions and ZFHX4 testing samples (ZFHX4_truncating) exhibited significantly low MVP scores, indicating an enhancement in the specificity of the model. The ZFHX4 testing samples, which encompass truncating variants (ZFHX4_truncating), a large duplication variant (ZFHX4_duplication), an inversion variant (ZFHX4_inversion) as well as an intragenic deletion (ZFHX4_in_frame_del) were included, enabling the model to predict class probability for ZFHX4 samples. Abbreviations: ADCADN, cerebellar ataxia, deafness, and narcolepsy, autosomal dominant; ARTHS, Arboleda-Thomas syndrome; ATRX, alpha-thalassemia/intellectual development syndrome, X-linked; AUTS18, autism, susceptibility to 18; BEFAHRS, Beck-Fahrner syndrome; BFLS, Börjeson-Forssman-Lehmann syndrome; BIS, blepharophimosis intellectual disability SMARCA2 syndrome; CdLS, Cornelia de Lange syndrome1–4; CHARGE, CHARGE syndrome; 16p11.2del, 16p11.2 deletion syndrome; CSS_c.6200, Coffin-Siris syndrome (c.6232G>A [GenBank: NM_006015.4 (ARID1A)]; p.Glu2078Lys) (c.6254T>G [GenBank: NM_006015.4 (ARID1A)]; p.Leu2085Arg) (c.6133T>C [GenBank: NM_017519.2 (ARID1B)]; p.Cys2045Arg); CSS4_c.2650, Coffin-Siris syndrome (c.2656A>G [GenBank: NM_001128849.1 (SMARCA4)]; p.Met886Val); CSS9, Coffin-Siris syndrome-9; Down, Down syndrome; Dup7, Williams-Beuren duplication syndrome (7q11.23 duplication syndrome); DYT28, Dystonia-28, childhood onset; EEC, epileptic encephalopathy, childhood onset; FLHS, Floating-Harbour syndrome; GADEVs, Gabriele de Vries syndrome; GTPTS, Genitopatellar syndrome; HMA, HVDAS_C, Helsmoortel-Van der Aa syndrome (ADNP syndrome [central]); HVDAS_T, Helsmoortel-Van der Aa syndrome (ADNP syndrome [Terminal]); ICF1 Immunodeficiency, centromeric instability, facial anomalies syndrome 1, ICF2_3-4, immunodeficiency, centromeric instability, facial anomalies syndrome 2, 3, and 4; IDDSELD, intellectual developmental disorder with seizures and language delay; Kabuki, Kabuki syndromes 1 and 2; KDM2B, KDM2B-related syndrome; KDM4B, KDM4B-related syndrome; KDVS, Koolen de Vries syndrome; Kleefstra, Kleefstra syndrome 1; LLS, Luscan-Lumish syndrome; MKHK_ID4, Menke-Hennekam syndrome-1, 2; MLASA2, myopathy, lactic acidosis, and sideroblastic anemia-2; MRD23, intellectual developmental disorder, autosomal dominant 23; MRD51, intellectual developmental disorder, autosomal dominant 51; MRX93, intellectual developmental disorder, X-linked, XLID93; MRX97, intellectual developmental disorder, X-linked 97, XLID97; MRXSA, Armfield syndrome; MRXSCJ, syndromic X-linked intellectual disability, Claes-Jensen type; MRXSN, syndromic X-linked intellectual disability, Nascimento type; MRXXSR, syndromic X-linked intellectual disability, Snyder-Robinson type; PHMDS, Phelan-McDermid syndrome; PRC2, RENS1, Renpenning syndrome; RMNS, Rahman syndrome; RSTS, Rubinstein-Taybi syndrome-1, 2; RSTS1, Rubinstein-Taybi syndrome-1; RSTS2, Rubinstein-Taybi syndrome-2; SBBYS, Say-Barber-Biesecker-Young-Simpson syndrome; Sotos, Sotos syndrome; TBSR, Tatton-Brown-Rahman syndrome; VCF5_comp, velocardiofacial syndrome; VCF5_core, velocardiofacial syndrome; WDSTS, Wiedemann-Steiner syndrome; WHS, Wolf-Hirschhorn syndrome; and Williams, Williams-Beuren deletion syndrome (7q11.23 deletion syndrome).

Figure 4: ZFHX4 expression increases during human neuronal differentiation, brain and craniofacial development (A) ZFHX4 expression levels during in vitro neural differentiation. RNA was sampled at the following stages: iPSCs, NSC (neural stem cells), neural progenitor cells (NPC) and direct differentiated neurons (Ns). An increasing expression is observed...
upon differentiation. The median expression value is indicated by the middle horizontal line. The whiskers indicate the minimum and maximum values. CNRQ, calibrated normalized relative quantity. (B) Endogenous ZFHX4 (399 kDa) levels in HEK293T, iPSCs, and NPCs as detected with western blot using the ZFHX4 antibody (top) and a vinculin antibody as a loading control (bottom). (C) ZFHX4 levels in developing human brain. ZFHX4 expression in the developing human brain (normalized RPKM data) showing higher expression during early prenatal development followed by decreased expression upon brain maturation. Data obtained from BrainSpan: http://www.brainspan.org. (D) Human embryonic craniofacial tissue bulk RNA-seq is visualized in a plot, showing an increase in ZFHX4 in craniofacial samples from five carnegie stages (CS13, CS14, CS15, CS17 and CS22).133

Figure 5: ZFHX4 interacts with 46 enriched proteins involved in axonogenesis, cell differentiation and neuron differentiation. (A) The enriched ZFHX4 interacting proteins in NPCs shown in a volcano plot. For each detected protein, the log2 fold change (FC) between the ZFHX4 IP-enriched samples and the IgG controls is shown on the x axis, while the -log10 of the p value is shown on the y axis. The horizontal dashed line indicates the -log10 value of a p value equal to 0.05. Proteins with an FDR < 0.05, log2 FC > 2 and log2 FC < -2 are, respectively, highlighted in blue and red. (B) Hierarchical clustering of potentially interesting proteins. The proteins shown to be differentially abundant between groups are visualized in a heatmap after non-supervised hierarchical clustering of z-scored log 2 PG.MaxLFQ intensities. The heatmap color scale from blue to red represents the level of abundance. The interacting factors which are absent in the IgG control, significant but not absent at the IgG control and significantly enriched protein and absent in the IgG control, are listed in Tables S7 and S8. (C) ZFHX4 interacting factors associated to their biological processes with EnrichR.

Figure 6: CUT&RUN sequencing data shows ZFHX4 binding to genes involved in axonogenesis, the regulation of the nervous system development, neurogenesis and neural migration. (A) Annotation of CUT&RUN consensus peaks reveals ZFHX4 binding to promoter regions. Left: Distribution of the ZFHX4 consensus peaks relative to transcription start sites (TSS). Half of the peaks are found between 0-1 kb of the TSS. Right: Pie chart visualizing the genomic annotation and the percentage of the peaks that overlap a TSS, 5’ UTR, 3’ UTR, exonic, intronic, downstream or distal intergenic regions. This shows that 55.62% of ZFHX4 consensus peaks are found near gene promoters. (B) ZFHX4 binds to genes involved in axonogenesis, forebrain and nervous system development, including the regulation of the neurogenesis and synapse organization; cell and neuron differentiation. The gene ontology (GO) dot plot displays the top 20 enriched biological processes (BPs) ranked by gene ratio (the number of genes related to GO term/total number of significant genes) and the p-adjusted values for these terms (colour, p-value 1e10^-7-1e10^-5). The size of the dot represents the gene counts per BP. (C) ZFHX4 plays a key role in embryonic, neuron and axon developmental pathways. The Kyoto Encyclopedia of Genes and Genomes (KEGG) dot plot provides all enriched biological pathways ranked by gene ratio (# of genes related to GO term/total number of significant genes) and the p-adjusted values for these terms (color). The size of the dot represents the gene counts per pathway.

Figure 7: Zebrafish zf/hx4 crisprants showed both craniofacial and behavioral changes. (A) Gene structure of the zebrafish zf/hx4 and ZFHX4. Top: The coding exons of ZFHX4 are
shown as light grey boxes and light blue boxes for the CDS (coding sequence), and untranslated regions in red (5'-UTR) and green (3'-UTR). Bottom: The coding exons of zfhx4 are shown as light grey boxes and light blue boxes for the CDS, and untranslated regions in red (5'-UTR) and green (3'-UTR). **(B) Representative craniofacial images for zf hx4 and scrambled zebrafish crispants are shown.** Left: microscopic ventral view scrambled (top panel) and zf hx4 crispants (bottom panel) of Alcian blue stained craniofacial structures at 5 dpf reveal smaller ethmoid plates and shorter Meckel’s cartilage in the F0 zf hx4 crispant larvae. Right: Adapted image\(^4^6\) from the representation of the cartilage structures of the zebrafish neurocranium (dark blue) at 5 dpf by lines. The red lines and circle show the measurements taken and normalized with the length of the head (purple line). **(C) Left: the surface area of the ethmoid plate was measured and normalized to the head length of the larvae.** Right: The length of the individual Meckel's cartilage structures was measured and normalized to the head length of the larvae. The ethmoid plate and the Meckel’s cartilage showed significant differences between crispant larvae compared to scrambled larvae (p-value ≤ 0.0001 (***) p-value ≤ 0.00001 (***)). **(D) zf hx4 crispants show a consistently lower movement frequency compared to scrambled control larvae during the complete behavior experiment but the trends during light and dark cycles remain similar.** Dark periods are indicated with grey boxes. Data is represented as mean ± SEM (n=60/genotype).

**Web Resources**

UniProt website: [www.uniprot.org](http://www.uniprot.org)

Evo-devo mammalian organs app: [https://apps.kaessmannlab.org/evodevoapp/](https://apps.kaessmannlab.org/evodevoapp/)

GTEx portal browser: [https://www.gtexportal.org/home/](https://www.gtexportal.org/home/)


Human embryonic craniofacial tissue bulk RNA-seq website: [http://cotneyweb.cam.uchc.edu/craniofacial_bulkrna/](http://cotneyweb.cam.uchc.edu/craniofacial_bulkrna/)


GnomAD v4.1.0: [https://gnomad.broadinstitute.org/about](https://gnomad.broadinstitute.org/about)

GTEx v8: [https://www.gtexportal.org/home/](https://www.gtexportal.org/home/)

**Acknowledgements**

We extend our gratitude to the participating families for their contribution in this study. We also appreciate the collaboration with the VIB Proteomics Core, led by Prof. Francis Impens. We are particularly grateful for the data analysis conducted by the bioinformatics team, specifically Simon Devos and Sara Dufour. Additionally, the zebrafish project was made possible with the support of the Zebrafish Facility Ghent Core Facility at the Center for
Medical Genetics Ghent (Ghent University), with special thanks to Andy Willaert and Hanne De Saffel.

Financial support has been provided by grants 1520518N and G055422N from the Research Foundation – Flanders (FWO) and BOF/STA/201909/009 from the Special Research Fund (BOF) from Ghent University. M.d.R.P.B is supported by a doctoral grant (STI.DIV.2023.0005.01) of the Marguerite-Marie Delacroix foundation. B.C. is a senior clinical investigator of the Research Foundation Flanders.

Research on proband 20 was possible by access to the data generated by the France Genomic Medicine Plan 2025. The work generated for proband 27 was performed within the European Reference Network for Intellectual disability, TeleHealth, Autism and Congenital Anomalies (ERN ITHACA). Sequencing and analysis of probands 35 and 45 were provided by the Broad Institute of MIT and Harvard Center for Mendelian Genomics (Broad CMG) and were funded by the National Human Genome Research Institute (NHGRI) grants UM1HG008900 (with additional support from the National Eye Institute, and the National Heart, Lung and Blood Institute), U01HG0011755, R21HG012397, and R01HG009141, and in part by the Chan Zuckerberg Initiative Donor-Advised Fund at the Silicon Valley Community Foundation (funder DOI 10.13039/100014989) grants 2019-199278, 2020-224274, 2022-316726 (https://doi.org/10.37921/236582yuakxy).

Angela Peron would like to thank ASST Santi Paolo e Carlo, San Paolo Hospital for the support during this study.

**Declaration of interests**

No declaration of interests is reported in this manuscript.

**Data and code availability**

The analyzed IP-MS data and the ZFHX4 consensus peaks of the CUT&RUN sequencing, as well as the results of the de novo and known motif analyses are available in the supplemental information.

**Supplementary material**

Supplemental information can be found online.

**References**


44. inDelphi https://indelphi.giffordlab.mit.edu/.
54. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders - ScienceDirect


63. ggplot2.


121. Nacc1 Mutation in Mice Models Rare Neurodevelopmental Disorder with Underlying Synaptic Dysfunction | Journal of Neuroscience https://www.jneurosci.org/content/44/14/e1610232024.long.


124. A Potassium Channel Mutation in Neonatal Human Epilepsy https://doi.org/10.1126/science.279.5349.403.


130. Anchoring and synaptic stability of PSD-95 is driven by ephrin-B3 | Nature Neuroscience https://www.nature.com/articles/nn.4140.


Hi-C
control fibroblasts
Melo et al.

Hi-C
patient fibroblasts

chr8 64 Mb 70 Mb 75 Mb 80 Mb 84 Mb

RefSeq genes
patient inversion

left breakpoint region

right breakpoint region

control

patient

chr8 68 Mb 69 Mb 70 Mb 71 Mb 72 Mb

control O/E

patient O/E

chr8 75 Mb 76 Mb 77 Mb 78 Mb 79 Mb

ZFHX4
A

Distribution of transcription factor binding loci relative to TSS

- Promoter (<=1kb) (49.39%)
- Promoter (1−2kb) (3.36%)
- Promoter (2−3kb) (2.87%)
- 3' UTR (1.28%)
- 1st Intron (0.92%)
- Other Intron (16.67%)
- Downstream (<=300) (0.06%)
- Distal Intergenic (16.3%)

B

GeneRatio

- 5e−07
- 4e−07
- 3e−07
- 2e−07
- 1e−07

p.adjust

- 2e−07
- 3e−07
- 4e−07
- 5e−07

Count

- 20
- 30
- 40
- 50
- 60
- 70

C

GeneRatio

- 0.02
- 0.04
- 0.06
- 0.08

p.adjust

- 0.005
- 0.010
- 0.015
- 0.020

Count

- 15
- 20
- 25
- 30
- 35
**A**

- **Meckel's cartilage - average length**
- **Ethmoid plate - surface area**

**B**  
- **Crispant larvae (5 dpf)**  
- **Control larvae (5 dpf)**

**C**

- **Ethmoid plate - surface area**
- **Meckel's cartilage - average length**

**D**

- **Movement frequency**