LRFN5 locus structure is influenced by the individual's sex and associated with autism

Helle Lybæk

Haukeland University Hospital: Haukeland Universitetssjukehus

Michael Robson

Max-Planck-Institute for Molecular Genetics: Max-Planck-Institut für molekulare Genetik

Nicole de Leeuw

Radboudumc Department of Human Genetics

Jayne Y. Hehir-Kwa

Prinses Máxima Centrum voor Kinderoncologie: Prinses Maxima Centrum voor Kinderoncologie

Aaron Jeffries

University of Exeter Medical School

Bjørn Ivar Haukanes

Haukeland University Hospital: Haukeland Universitetssjukehus

Siren Berland

Haukeland University Hospital: Haukeland Universitetssjukehus

Diederik de Bruijn

Radboudumc Department of Human Genetics

Stefan Mundlos

Max-Planck-Institute for Molecular Genetics: Max-Planck-Institut für molekulare Genetik

Malte Spielmann

UKSH Campus Kiel Klinik für Neuropadiatrie: Klinik für Kinder- und Jugendmedizin II

Gunnar Douzgos Houge (✉ gunnar.houge@helse-bergen.no)

Haukeland University Hospital: Haukeland Universitetssjukehus <https://orcid.org/0000-0002-6102-1513>

Research

Keywords: LRFN5, SALM5, developmental delay, chromatin, topologically associating domains, TAD, autism spectrum disorder, gene conversion

Posted Date: September 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-895724/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

***LRFN5* locus structure is influenced by the individual's sex and associated with autism**

Helle Lybæk,¹ Michael Robson,² Nicole de Leeuw,³ Jayne Y. Hehir-Kwa,⁴ Aaron Jeffries,⁵ Bjørn Ivar Haukanes,¹ Siren Berland,¹ Diederik de Bruijn,³ Stefan Mundlos,² Malte Spielmann⁶ and Gunnar Houge^{1,7,8*}

¹ Department of Medical Genetics, Haukeland University Hospital, 5021 Bergen, Norway

² Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, D-14195 Berlin

³ Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

⁴ Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands

⁵ University of Exeter, Stocker Road, Exeter, EX4 4QD, UK

⁶ Institute for Human Genetics, UKSH, Kiel

⁷ Institute of Clinical Medicine K2, Faculty of Medicine, University of Bergen,

⁸ Honorary Chair of Evolution and Genomic Sciences, University of Manchester, UK

*Corresponding author:

Gunnar Douzgos Houge

Department of Medical Genetics

Haukeland University Hospital

N-5021 Bergen, Norway

Tel: +47 55 97 54 75

E-mail: gunnar.douzgos.houge@helse-bergen.no

No of Figures: 4 + 6 supplementary

No of Tables: 2 + 1 supplementary

No of words abstract: 349

1 ABSTRACT

2 **Background:** *LRFN5* is a brain-specific gene needed for synaptic development and
3 plasticity. It is the only gene in a large 5.4 Mb topologically associating domain
4 (TAD) on chromosome 14, which we term the *LRFN5* locus. This locus is highly
5 conserved, but has extensive copy number variation.

6 **Methods:** Locus structure was studied by chromatin immunoprecipitation (chIP-on-
7 chip) in fibroblasts from individuals with autism and controls, supplemented with a
8 capture-HiC determination of TAD structures in a family trio. *LRFN5* expression was
9 studied in foetal brain cell cultures. In addition, locus interaction was studied in four
10 large and independent cohorts by measuring deviations from Hardy-Weinberg
11 equilibrium of a common deletion polymorphism.

12 **Results:** We found that locus structural changes are associated with developmental
13 delay (DD) and autism spectrum disorders (ASD). In a large family, ASD in males
14 segregated with a chromosome 14 haplotype carrying a 172 kb deletion upstream of
15 *LRFN5*. In a fibroblast capture-HiC study on an ASD-patient-parent trio, the ASD-
16 susceptible haplotype (in the mother and her autistic son) had a TAD pattern different
17 from both the father and a female control. When the trimethylated histone-3-lysine-9
18 chromatin (H3K9me3) profiles in fibroblasts from control males (n=6) and females
19 (n=7) were compared, a male-female difference was observed around the *LRFN5* gene
20 itself (p<0.01). Intriguingly, in three cohorts of individuals with DD (n=8757), the
21 number of heterozygotes of a common deletion polymorphism upstream of *LRFN5*
22 was 20-26% lower than expected from Hardy-Weinberg equilibrium. This indicates
23 early allelic interaction, and the genomic conversions from heterozygosity to wild-
24 type or deletion homozygosity were of equal magnitudes. In a control group of
25 medical students (n=1416), such conversions were three times more common than in

the DD-patient cohorts ($p=0.00001$). Hypothetically, such allelic interaction is needed to establish monoallelic expression, which we found in foetal brain cell cultures.

Limitations: The male-female difference in H3K9me3 profiles was based on fibroblast data from a small number of individuals, and the monoallelic expression data on a single experiment.

Conclusions: Taken together, allelic interaction, monoallelic expression and sex-dependent differences make the *LRFN5* locus attractive for exploring the genetic basis of synaptic memory and high-functioning male autism.

Keywords: *LRFN5*, *SALM5*, developmental delay, chromatin, topologically associating domains, TAD, autism spectrum disorder, gene conversion

BACKGROUND

Autism spectrum disorders (ASD) are 3-4 times more frequent in males than in females,^{1,2} and this sex difference is even more prominent in the high-functioning ASD group (previously called Asperger syndrome).³ The cause of this sex discrepancy is unknown and not X-linked.⁴ Despite high heritability,^{5,6} the genetic cause of mild ASD is mostly unknown in the absence of pathogenic genomic copy number gains or losses.^{7,8} Genome-wide association studies (GWAS) have not identified any major autism loci, but have linked ASD, intellectual disability (ID) and

1 schizophrenia to an overlapping set of low-risk variants. The *LRFN5* locus has only
2 been identified in a family-based GWAS study from 2009,⁹ a finding not replicated in
3 more recent population-based or meta-GWAS analyses.^{10,11} There are, however, other
4 results that link the *LRFN5* locus to autism. A copy number variation (CNV) study of
5 an ASD cohort found that nearly 1% of all ASD-linked rare CNVs were in the *LRFN5*
6 locus and none were reported *de novo*.¹² Even more relevant is a homozygous-
7 haplotype-sharing-in-autism-study that identified both *LRFN5* and the flanking gene
8 *FBXO33* as ASD candidate genes in at least two population clusters.¹³

9 *LRFN5*, called SALM5 (synaptic adhesion-like molecule-5) in mice, belongs to a
10 family of five small transmembrane protein genes involved in synaptic development,
11 organization and plasticity.¹⁴⁻¹⁹ *LRFN5* acts as a dimer, and induces presynaptic
12 differentiation through binding to the LAR family of receptor tyrosine phosphatases.²⁰
13 *LRFN5* is the only family member located in the middle of a gene desert that can also
14 form a large TAD of 5.4 Mb. The single-gene-in-a-TAD structure is not unique, but
15 this TAD is unusually large, the average TAD size being about 0.9 Mb.²¹ This
16 suggests complex gene regulation at the structural level. It is also the only *LRFN*-gene
17 with a large 5'UTR (of 1.9 kb), encoded by the first 2 (of 6) exons. This 5'UTR can
18 potentially form a complex hairpin with an estimated ΔG of -760 kcal/mol
19 (<http://www.unafold.org/>). This suggests gene regulation also at the translational
20 level.

21 Here we show that structural differences in this region are related to the sex of the
22 individual and associated with autism in males, that monoallelic expression do seem
23 to occur, and that locus homogenization by frequent allelic conversion could both be a
24 mechanism of structure maintenance and of *LRFN5* expression regulation.

1 **METHODS**

2 **Aim of study**

3 Inspired by rare cytogenetic findings in individuals with translocations or copy
4 number aberrations affecting the *LRFN5* locus, in-depth investigation of locus
5 structure was done to find autism-associated changes, and to explain sex-biased
6 inheritance of autism susceptibility in a large family. Unexpectedly, violations of
7 rules of inheritance were also found, indicating frequent locus interactions with allelic
8 conversions.

9 **Patients and DNA samples**

10 DNA samples were obtained from the diagnostic genetic service in Bergen (Norway)
11 and in Nijmegen (the Netherlands). Patients were recruited from the clinical genetic
12 service in Bergen (Norway). Consent was obtained after genetic counselling of the
13 two families having boys with autism. They have also been counselled about the
14 results presented here. The senior author (GH) met these four males with autism 10-
15 20 years ago and again now recently. They are all unable to hold regular jobs because
16 they lack social/communication skills. Most have finished regular school and high
17 school, but not without difficulties (especially in mathematics). Language skills are
18 generally good (also in English). Typically, a digital clock display was easier to
19 understand than a classical clockface. Their range of daily-life problems is strikingly
20 similar. We have also counselled the families of the female with a *de novo* 14,21
21 translocation and the male with the *de novo* 6;14 translocation.

22 **ChIP-on-chip experiments**

23 Chromatin immunoprecipitation (chIP) was done on primary fibroblast cultures
24 obtained from skin biopsies from patients, family members and controls. Fibroblasts

1 were frozen in several batches after primary culture, and a single batch was thawed
2 for an experiment. This was done to avoid major differences in number of passages
3 (cell culture doubling times) between experiments. For each sample, a total of 4×10^7
4 fibroblast cells were cultured to ~90% confluence. Cells were cross-linked for 15 min
5 with 1% formaldehyde at 37°C before 125 mM glycine was added, followed by
6 incubation for 5 min at room temperature (RT). Next, cells were washed 3 times with
7 ice-cold phosphate-buffered saline (PBS) before harvesting by scraping, and cell
8 pellet were collected by centrifugation. Chromatin from cell pellet was isolated as
9 follows: Incubation for 10 min on ice in 4 ml cell lysis buffer [5 mM PIPES pH 8; 85
10 mM KCl; 0,5% NP-40; 1x protease inhibitors], centrifugation and incubation of cell
11 pellet for 30 min on ice in another lysis buffer [50 mM Tris pH 8; 10 mM EDTA; 1%
12 SDS; 1x protease inhibitors]. Chromatin was sheared on a S220 ultra-sonicator
13 (Covaris, MA, USA) according to the protocol “Chromatin shearing with SDS
14 detergent buffers” from the manufacturer. The level of fragmentation was examined
15 on a 1% agarose gel, and shearing was continued until the major part of the DNA
16 fragments had a size of ~600 bp. The cell solution was centrifuged, and the lysate was
17 frozen in batches of 100 µl at -80°C for later use, except for 50 µl lysate which was
18 used for determination of the chromatin concentration (=input DNA) by standard
19 DNA isolation procedures.

20 Chromatin immunoprecipitation (chIP) was done as follows: Per chIP, 50 µl
21 chromatin (on average 15 µg), 25 µl protein A/G PLUS beads (Santa Cruz
22 Biotechnology, TX; USA), and 425 µl incubation buffer [0,2% SDS, 1% Triton, 150
23 mM NaCl, 2 mM EDTA, 0,5 mM EGTA, 10 mM Tris pH 8,5, 1x protease inhibitors]
24 were incubated with rotation for 1 hour at 37°C. Protein-DNA complexes were
25 immunoprecipitated with a chIP-validated antibody at 4°C. The antibodies used were

1 against histone H3 (histone H3, #4620; Cell signalling, MA, USA), histone H3
2 trimethylated at lysine 4 (H3K4me3, ab8580; Abcam, Cambridge, UK), histone H3
3 trimethylated at lysine 27 (H3K27me3, ab6002; Abcam), histone H3 trimethylated at
4 lysine 9 (H3K9me3, ab8898; Abcam), histone H3 acetylated at lysine 9 (H3K9ac,
5 ab4441; Abcam), and rabbit IgG (#2729; Cell signalling). The amount of antibody
6 added was according to manufacturers' recommendations. Next day, 25 µl protein
7 A/G PLUS beads were added to each chIP and incubated for 2 hours at 4°C.
8 Subsequently, the protein-DNA complexes underwent a series of washes: 2x [0,1%
9 SDS, 0,1% DOC, 1% triton, 150 mM NaCl, 1 mM EDTA, 0,5 mM EGTA, 10 mM
10 Tris pH 8,5], 1x [0,1% SDS, 0,1% DOC, 1% triton, 500 mM NaCl, 1 mM EDTA, 0,5
11 mM EGTA, 10 mM Tris pH 8,5], 1x [0,25% LiCl, 0,5% DOC, 0,5% NP-40, 1 mM
12 EDTA, 0,5 mM EGTA, 10 mM Tris pH 8,5], and 2x [1 mM EDTA, 0,5 mM EGTA,
13 10 mM Tris pH 8,5]. The formaldehyde-induced crosslinks were reversed by
14 incubation for 30 min at RT with SDS-solution [0,1 M NaHCO₃ + 1% SDS], and
15 genomic DNA was recovered by standard DNA isolation procedures. Validation of
16 the chIPs were done by PCR on non-amplified chIP samples with primers recognizing
17 the *FBXO33* and *LRFN5* genes. In the PCR, rabbit IgG-chIP was used as a pre-
18 immune control and 2% of input DNA was used as a control of the
19 immunoprecipitation efficiency.

20 For the chIP-on-chip analysis, 60% of each of two replicate chIP samples were
21 concentrated using microcon YM-30 spin columns (Merck Millipore, MA, USA) and
22 amplified using the GenomePlex Complete Whole Genome Amplification kit (Sigma-
23 Aldrich, MO, USA). The chIP-on-chip hybridizations were done on a custom
24 designed high resolution NimbleGen 3x 720K array (Roche NimbleGen, WI, USA).
25 The array probes (50-mers, positions according to Chr37/hg19) covered unique

regions of chromosome 2 (168,500–178,500 Mb), chromosome 13 (94,000–113,000 Mb), chromosome 14 (16,475–70,975 Mb), and chromosome 17 (41,370–81,195 Mb) uniformly, with a median probe spacing of ~150 bp. DNA labelling, array hybridization, post-hybridization washes and scanning were performed according to the manufacturer's protocol: “NimbleGen Arrays User's Guide, ChIP-chip Array”, v6.2 (Roche NimbleGen). In short, the chIP and input DNA (DNA from non-precipitated chromatin) samples were labelled with Cy5- and Cy3-conjugated random nonamers, respectively. The labelled samples were purified, combined, denatured and hybridized to the array for 16 hours at 42°C. After stringent washing, the array was scanned using an Axon 4200AL Scanner (Molecular Devices, CA, USA) at 5-µm resolution. The acquired images were analysed by DEVA v1.2 software (Roche NimbleGen) creating pair reports, including raw intensities for each probe and per image. From these data, ratio files were generated. For data visualization, the average ratios of two replicate experiments were binned per kb, each adjusted for the number of probes per bin. These data were transferred to in Excel spreadsheets for further calculations and generation of plain text files in .bedgraph format, and then the data was plotted against chromosomal position using the UCSC browser's custom track option.

Capture HiC-based *LRFN5* locus TAD-structure determination

***LRFN5* locus selection:** The capture Hi-C (CHiC) SureSelect library was designed over the genomic interval (chr14:539,000,000–47,000,000, GRCh37) using the SureDesign tool from Agilent (Agilent Technologies, Santa Clara, CA). The coverage was 70,5% by 159698 probes of total size 4,668 Mb and 5x tiling density.

Fixation of fibroblast nuclei: Capture HiC experiments were performed on dermal fibroblasts from a family trio (parents and child with ASD) and a control female of

1 the same age as the mother. Trypsinised fibroblasts were washed in PBS and then
2 transferred to a 50-ml Falcon tube and complemented with 10% FCS/PBS. 37%
3 formaldehyde was added to a final concentration of 2% and cells were fixed for
4 10 min at room temperature. Crosslinking was quenched by adding glycine (final
5 concentration; 125 mM). Fixed cells were washed twice with cold PBS and lysed
6 using fresh lysis buffer (10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM MgCl₂, 0.1 mM
7 EGTA with protease inhibitor) to isolate nuclei. Cell lysis was assessed
8 microscopically after 10-min incubation in ice. Nuclei were centrifuged for 5 min at
9 480g, washed once with PBS and snap frozen in liquid N₂.

10 **Chromosome conformation capture library preparation and sequencing:**

11 3C libraries were prepared from fixed nuclei as described previously.²² Briefly, lysis
12 buffer was removed by centrifugation at 400 g for 5 min at 4 °C, followed by
13 supernatant aspiration, snap-freezing, and pellet storage at -80 °C. Later, nuclei pellets
14 were thawed on ice, resuspended in 520 µl 1× DpnII buffer, and then incubated with
15 7.4 µl 20% SDS shaking at 900 rpm. at 37 °C for 1 hour. Next, 75 µl 20% Triton X-
16 100 was added and the pellet was left shaking at 900 rpm. at 37°C for 1 hour. A 15-µl
17 aliquot was taken as a control for undigested chromatin (stored at -20°C). The
18 chromatin was digested using 40 µl 10 U/µl DpnII buffer shaking at 900 rpm. at 37°C
19 for 6 h; 40 µl of DpnII was added and samples were incubated overnight, shaking at
20 900 rpm. at 37°C. On day three, 20 µl DpnII buffer was added to the samples
21 followed by shaking for 5 more hours at 900 rpm. at 37 °C. DpnII subsequently was
22 inactivated at 65°C for 25 min and a 50-µl aliquot was taken to test digestion
23 efficiency (stored at -20°C). Next, digested chromatin was diluted in 5.1 ml H₂O, 700
24 µl 10×ligation buffer (Thermo Fisher Scientific), 5 µl 30 U/µl T4 DNA ligase and
25 incubated at 16°C for 4 h while rotating. Ligated samples were incubated for a further

1 30 min at room temperature. Chimeric chromatin products and test aliquots were de-
2 crosslinked overnight by adding 30 μ l and 5 μ l proteinase K, respectively, and
3 incubated at 65°C overnight. On the fourth day, 30 μ l or 5 μ l of 10 mg/ml RNase was
4 added to the samples and aliquots, respectively, and incubated for 45 min at 37°C.
5 Next, chromatin was precipitated by adding 1 volume phenol-chloroform to the
6 samples and aliquots, vigorously shaking them, followed by centrifugation at 4,000
7 rpm. at room temperature for 15 min. To precipitate aliquoted chromatin, 1 volume
8 100% ethanol and 0.1 volume 3M NaAc, pH 5.6 was added and the aliquots placed at
9 -80°C for 30 min. DNA was then precipitated by centrifugation at 5,000 rpm. for 45
10 min at 4°C followed by washing with 70% ethanol, and resuspension in 20 μ l with 10
11 mM Tris-HCl, pH 7.5. To precipitate samples, extracted sample aqueous phases were
12 mixed with 7 ml H₂O, 1 ml 3M NaAc, pH 5.6, and 35 ml 100% ethanol. Following
13 incubation at -20°C for at least 3 h, precipitated chromatin was isolated by
14 centrifugation at 5,000 rpm. for 45 min at 4 °C. The chromatin pellet was washed
15 with 70% ethanol and further centrifuged at 5,000 rpm. for 15 min at 4°C. Finally, 3C
16 library chromatin pellets were dried at room temperature and resuspended in 10 mM
17 Tris-HCl, pH 7.5. To check the 3C library, 600 ng were loaded on a 1% gel together
18 with the undigested and digested aliquots. The 3C library was then sheared using a
19 Covaris sonicator (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: 6 cycles
20 of 60 s each; set mode: frequency sweeping; temperature: 4-7°C). Adaptors were
21 added to the sheared DNA and amplified according to the manufacturer's instructions
22 for Illumina sequencing (Agilent). The library was hybridized to the custom designed
23 SureSelect beads and indexed for sequencing (150 bp paired-end) following the
24 manufacturer's instructions (Agilent).

25 **Capture-HiC analysis:** Raw reads were pre-processed with cutadapt v1.15 to trim

1 potential low-quality bases (-q 20 -m 25) and any remaining sequencing adapters
2 (-a and -A option with Illumina TruSeq adapter sequences according to
3 the cutadapt documentation) at the 3' ends of the reads. Mapping, filtering and
4 deduplication of the short reads were performed with the HiCUP pipeline v0.7.04
5 (<https://doi.org/10.12688%2Ff1000research.7334.1>) (no size selection, Nofill: 1,
6 Format: Sanger). The pipeline employed Bowtie2 v2.2.6
7 (<https://doi.org/10.1038%2Fnmeth.1923>) for mapping short reads to the hg19 human
8 reference genomes. Juicer tools 0.7.5 (<https://doi.org/10.1016%2Fj.cels.2016.07.002>)
9 was used to generate binned-contact maps from valid and unique read pairs with
10 MAPQ ≥ 30 and to normalize contact maps by Knight and Ruiz (KR) matrix
11 balancing (<https://doi.org/10.1016%2Fj.cell.2014.11.021>
12 <https://doi.org/10.1093%2Fmanum%2Fdr019>). For the generation of capture-HiC
13 contact maps, only reads pairs mapping to the enriched genomic region were
14 considered and shifted by the offset of the enriched genomic region. For the import
15 with Juicer tools, we used a custom chrom.sizes file containing only the size of the
16 enriched part of the genome. Afterward, KR-normalized maps were exported at 10-kb
17 resolution, and coordinates were shifted back to their original values.

18 **DNA methylation studies**

19 CpG methylation of the *LRFN5* locus was initially interrogated by the Illumina 450K
20 DNA methylation array (Infinum HumanMethylation450 BeadChip), following
21 manufacturer's protocol. No methylation differences were found between male and
22 female controls. Specific positions in the *LRFN5* locus (namely the region
23 corresponding to the peak of the locus-TAD and the *LRFN5* promoter) were analysed
24 by both DNA sequencing after bisulphite treatment, and custom designed MLPA tests
25 (using the P-300 kit from MRC-Holland, and following the manufacturer's protocol).

1 Still, no sex differences were found. All these methylation studies were done on
2 leukocyte DNA from peripheral blood samples.

3 ***LRFN5* expression in foetal brain cell cultures**

4 Human neural stem cell lines were established from 12-week foetal brain cell cultures
5 using a conditionally immortalization methodology, as described.²³ STROC05,
6 STROC08 and STROC11 are striatal cell lines that are identical and from the same
7 dissection, just different isolated clones. An Illumina SNP array was used to identify
8 heterozygous SNPs in *LRFN5*, and the *LRFN5* cDNA sequence of polyclonal brain
9 cells was then compared to the cDNA sequence of these cell lines.

10 **SNP array analysis of a deletion polymorphism upstream of *LRFN5***

11 The allele frequencies of a common 60 kb deletion, chr14(GRCh37):g.41609383-
12 41669664, detected by at least 12 oligonucleotides on the Affymetrix 6.0 SNP array
13 and by at least 18 oligonucleotides on the Affymetrix CytoScan SNP array
14 (Affymetrix, ThermoFischer Scientific, USA), were determined in four cohorts of
15 individuals: three ascertained due to developmental disorders, and one group of
16 mostly medical students (Table 2). All cohorts were anonymized. Three of the cohorts
17 (patient cohort I and III and the student cohort) had their genomic copy number status
18 determined by the Affymetrix 6.0 array, while patient cohort II was tested by the
19 Affymetrix CytoScan array.

20 **Statistical analysis**

21 Basic and on-line statistical tools were used to investigate the statistical significance
22 of the following findings: the segregation of the A-haplotype in the autism family:
23 Fisher Exact Test ($p = 0.0476$); the female-male comparison in the table of Figure 3:
24 Student two-sided t-test ($p = 0.005$); the difference between Hardy-Weinberg

1 equilibrium and observed allelic distribution in the patient cohorts in Table 2: Chi-
2 square test ($p < 0.00001$); and the difference between the allelic distribution in the
3 patient cohorts and the student cohort in Table 2: Chi-square test ($p < 0.00001$).

4 **RESULTS**

5 ***LRFN5* locus structural changes can be associated with developmental** 6 **delay and autism**

7 Our attention was drawn to the *LRFN5* locus in 2007 when a girl with intellectual
8 disability (ID) that gave no social contact was found to have two *de novo*
9 chromosome changes: a small 2q31.1 deletion and a balanced 14;21-translocation.²⁴
10 One translocation breakpoint was in the *LRFN5* locus at 14q21.1, and the other on the
11 acrocentric 21p arm. To explore the consequences of these chromosome aberrations
12 for chromatin structure, we performed ChIP-on-chip experiments on patient skin
13 fibroblasts. The results were compared to a non-autistic boy with another *LRFN5*
14 locus translocation: a *de novo* t(6;14)(q26;q21.1) causing Coffin-Sirin syndrome
15 because *ARID1B* was disrupted on chromosome 6 (Supplementary Figure 1). He
16 speaks well and has good social function. A clear difference in the *LRFN5* locus
17 chromatin profiles was only seen in the autistic girl. Possibly this is because her
18 translocation fused the *LRFN5* locus to an acrocentric p-arm, and a trimethylated
19 histone-3-lysine-9 (H3K9me3) spreading effect from *FBXO33* and centromeric could
20 be seen on derivative chromosome 14, while more trimethylated histon-3-lysin-4
21 (H3K4me3) and less trimethylated histon-3-lysin-27 (H3K27me3) was seen in the
22 *LRFN5* locus on derivative chromosome 21 (Supplementary Figure 1). In addition, an
23 apparent “compensatory” H3K4me3-effect corresponding to the 2.6 Mb 2q31.1
24 deletion could also be seen. However, this unexpected and unprecedented effect could

1 be 2q31.1-locus specific as it was not found in an individual with a 2.2 Mb
2 chromosome 17q23.1 deletion (Supplementary Figure 2). These two unique 14;21-
3 translocation patients initiated our *LRFN5* research, and we noted that autistic
4 behaviour was only seen in the individual with chromatin changes to this locus.

5 Of more general relevance for *LRFN5* function is the number of individuals with
6 DD/ASD and copy number variants (CNVs) in the *LRFN5* locus in the DECIPHER
7 database (n=30) and our own records (n=3) (Supplementary Figures 3 and 4, plus
8 reference ²⁵). Only one small deletion containing *LRFN5* itself was registered as *de*
9 *novo*. Because many CNVs are inherited from a seemingly unaffected parent, most
10 have been regarded as non-pathogenic. Nevertheless, if this truly is an ASD
11 susceptibility locus, even reduced penetrance should not have prevented that from
12 being discovered. However, if sex-of-origin also matters for autism susceptibility, an
13 ASD link could easily be missed.

14 **Autism in males segregated with a specific *LRFN5* locus haplotype**
15 **inherited from their mothers**

16 High-functioning ASD was diagnosed in two pairs of brothers from two families from
17 the same geographical region (Table 1; fam #1 and fam #2). These four males have
18 been seen by the senior author both as children and adults, and their autism phenotype
19 is strikingly similar. They all had a 172 kb deletion just upstream of the *LRFN5*
20 promoter and the same locus haplotype (called the A-haplotype in Table 1), inherited
21 from their normal mothers. The families were too distant to know about any
22 relatedness. The A-haplotype was also found in a normal maternal uncle, inherited
23 from his mother. Of the nine individuals sharing the A-haplotype in these families,
24 4/5 males had ASD, and 4/4 *SRY*-negatives (three females and one XX-male) were

1 normal. This may suggest that the A-haplotype increased ASD susceptibility in males
2 in this family (Fisher Exact Test $p=0.0476$). We also found overlapping deletions on
3 different haplotypes in four other individuals or families, ascertained by copy-number
4 high-resolution SNP-array testing because of developmental delay, ID, ASD or
5 schizophrenia (Table 1). This could indicate that the A-haplotype is linked to male
6 ASD-susceptibility, not the 172 kb deletion *per se*. An effect of the deletion is not
7 excluded, but then it must be haplotype dependent. The 172 kb deletion does not
8 contain enhancer-like chromatin profiles or lncRNAs.

9 **The structure of the *LRFN5* locus is influenced by the sex of the** 10 **individual**

11 To explore if the A-haplotype influenced TAD-structure of the *LRFN5* locus, we
12 performed a capture HiC-experiment on skin fibroblasts from a family trio (boy with
13 ASD and his unaffected parents), with a normal unrelated female as control (Figure
14 1). Three patterns emerged: In the autistic boy and his mother, both sharing the A-
15 haplotype, the whole-locus 5.4 Mb mega-TAD had three sub-TADs with boundaries
16 at the *LRFN5* promoter (arrow A in Figure 1) and the middle of the *LRFN5*-
17 downstream gene desert (arrow B in Figure 1). The small red diamond just
18 centromeric to the *LRFN5* promoter marks the 172 kb deletion. In the father, only two
19 TADs could be discerned inside the mega-TAD. No distinct “B-junction” could be
20 seen. In the control female, only the mega-TAD was distinct; an “A-junction” was
21 diffuse if at all present (Figure 1).

22 These subtle capture-HiC differences point to variation in TAD structure and that the
23 autistic boy inherited his mother’s structure. Hypothetically, this specific structure
24 (with a “B-junction”) could be A-haplotype dependent and the one linked to autism in

1 this family. We also noted that the father and the control female had TAD differences
2 (Figure 1).

3 To explore if this could reflect differences in their chromatin profiles, chIP-on-chip-
4 generated chromatin data (H3K4me3, H3K27me3 and H3K9me3) on fibroblasts from
5 four family members, five control males and seven control females, were compared.
6 A sex difference was found for the H3K9me3 profiles, i.e. heterochromatin protein-1
7 (HP1)-associated constitutive heterochromatin, but only corresponding to the *LRFN5*
8 gene itself and the region downstream to the H3K4me3/H3K27me3 signals marking
9 the peak of the mega-TAD (Figure 2). In this area, males had more heterochromatin
10 than females. This difference was significant, but individual differences in the degree
11 of this effect should be noted (Figure 2). In fact, the groups are overlapping (Figure
12 3). In contrast, the sex difference was zero in a region 1 Mb upstream (Figure 3).

13 The four family members, including the mother and the *SRY*-negative brother of the
14 index boy that shared the A-haplotype (Table 1), all had a “male-type” of chromatin
15 pattern (Figure 2). Possibly, this pattern is *SRY*-dependent, but not if one inherits the
16 A-haplotype. The control male and female with the largest H3K9me3-associated
17 chromatin differences are also shown at the bottom of the locus overview figure
18 (Figure 4), where their H3K9me3-chromatin profiles are aligned to the ASD-
19 associated A-haplotype TAD structure, CTCF-binding sites, and open-chromatin
20 neuronal single-cell ATAC sites.

21 To explore if we could test large numbers of individuals for sex-related or autism-
22 related differences in this region, we performed a pilot MLPA-based study on
23 anonymized blood leukocyte DNA to see if these sex difference in chromatin profiles
24 was also reflected in methylation differences of a CpG located at the conserved CTCF

1 binding site in the *LRFN5* promoter (chr14:42,069,895-42,069,949, hg19). An
2 average methylation degree of 0.19 (CI 0.13-0.25) in control males (n=16) and control
3 females (n=6) were found with no sex difference, and the same average level (0.20)
4 was found in patients investigated because of non-ID associated autism (n=14). In a
5 separate experiment, using the Illumina 450K BeadChip methylation test, we found
6 an average degree of CpG methylation of 0.14 at the *LRFN5* promoter (19 CpGs were
7 interrogated) and 0.52 in the *LRFN5* gene itself (8 CpGs were interrogated) in 10
8 other control individuals, still without sex difference. This shows that the difference in
9 fibroblast H3K9me3 profiles at the *LRFN5* gene is not reflected in differences in CpG
10 methylation of leukocyte DNA in the same region.

11 **Allelic interaction must be frequent in the *LRFN5* locus**

12 If the *LRFN5* locus structure is critical for brain function, how is it maintained when
13 copy number changes are so common? A possible answer to this question was
14 unexpectedly found upon examining the allelic distribution of a common 60 kb
15 deletion in the middle of the *LRFN5*-upstream gene desert (chr14:41,609,383-
16 41,669,664 (hg19); Table 2 (position of deletion indicated in Figure 4). We examined
17 three large and independent cohorts of individuals (two from Norway and one from
18 the Netherlands) who had been investigated with a high-density SNP-array because of
19 a developmental disorder, usually including variable degrees of developmental delay
20 (DD). In addition, one Dutch cohort (n=1,416) of mainly medical students served as a
21 control. The deletion's minor allele frequency (MAF) was around 8.5% in the Dutch
22 population, 15% in the Norwegian population, and 14% in gnomAD
23 (gnomad.broadinstitute.org).²⁶ When comparing the observed allelic distribution in
24 these cohorts with the expected distribution as per Hardy-Weinberg equilibrium, there

were too few heterozygotes in all cohorts. We found that the loss of heterozygote wild-type (wt) / deletion (del) was of equal magnitude in both directions, i.e. to wt/wt and to del/del. In the three patient cohorts, this loss was 20%, 24% and 26%, respectively, but in the student cohort, it was 80% (Table 2). This surprising difference was not a technical artefact. 12-18 SNP-array oligonucleotides covered the 60 kb deletion, and >5 in a row is usually sufficient for correct calling. Furthermore, if this was due to missed wt/del calling, the high number of del/del homozygotes would still be incompatible with Hardy-Weinberg equilibrium. Also, when examining SNP array results from our diagnostic routine, we found that there was an excess of homozygosity regions >1 Mb in the *LRFN5* locus compared to many other regions in the genome. We are not aware of any other potential explanations for these observations than an early mitotic allelic conversion event with a frequency of at least 1 in 5 individuals in the DD group and as high as 4 in 5 individuals in the student group. This is far beyond expectation. Maybe this allelic interaction is needed to establish monoallelic expression. In human fetal brain cell cultures from the striatum, we did indeed find evidence indicating monoallelic *LRFN5* expression (Supplementary Figure 5).

DISCUSSION

Despite extensive research, it has remained elusive why autism, and especially the higher-functioning variants, is more common in males than in females.² Here we show that the synaptic regulation and maintenance gene *LRFN5*, situated in the middle of a conserved gene desert capable of forming a 5.4 Mb mega-TAD, may be part of the answer to this question. The three main reasons for this are the large family with remotely related pairs of brothers with a very similar form of higher-functioning

1 autism sharing the same maternally inherited *LRFN5* locus haplotype (Table 1), the
2 sex-influenced differences in locus chromatin 2D structure in fibroblasts from normal
3 males and females (Figure 2 and 3), and the allelic interaction that must take place in
4 early embryonic development, as evidenced by the striking deviation from Hardy-
5 Weinberg equilibrium of a 60 kb deletion polymorphism (Table 2).

6 There are eight putative SOX9 and SRY binding sites flanking the *LRFN5* locus; 3+2
7 SOX9 sites and 1+2 SRY sites (Figure 2 and Supplementary Figure 6; TFBS_conc
8 track in the UCSC browser). This could have importance for generation of the
9 observed male/female chromatin difference (Figures 2 and 3). The mechanism behind
10 the early locus interaction is more difficult to explain. We are unaware of such
11 frequent allelic interaction and locus conversion in any other part of the genome, e.g.
12 the frequency of inter-chromosomal gene conversion in gene families with more than
13 two alleles has been estimated to be around 0.2%.²⁷ Gene conversion events are more
14 common towards the 3'UTR end of protein-coding genes, and this is believed to be
15 due to RNA transcription aiding the process.²⁷ Maybe *LINC02315*, an RNA gene
16 upstream of *LRFN5* ending in the common 60 kb deletion polymorphism, has a
17 similar role (Figure 4).

18 The most interesting question is why frequent locus homozygotisation by gene
19 conversions occurs. We hypothesize that this process is needed to establish stochastic
20 monoallelic expression, advantageous for fine-tuning *LRFN5* expression. Given the
21 high conversion frequencies, allelic interaction is probably the rule, and allelic
22 conversion a consequence. Exploratory foetal brain expression data suggests that
23 monoallelic expression does occur in humans (Supplementary Figure 5), but we do
24 not know if this is always the case or if it is stochastic. The next question is why fine-

1 tuning of *LRFN5* expression is so critical that the gene, on top of having a 1.9 kb
2 5'UTR encoded by exons 1 and 2, needs to be framed by 2-3 Mb of conserved,
3 presumed regulatory, gene deserts. Maybe the difference in locus conversion
4 frequencies between a student cohort (4 in 5 were converted to homozygosity) and
5 three DD-cohorts (1 in 4 were converted to homozygosity) provides a clue (Table 2).
6 Hypothetically, if the *LRFN5* locus is a genetic basis for synaptic memory, then both
7 allelic interaction and conversion could be advantageous for this function. This fits
8 well with *LRFN5*'s role as a protein involved in synapse strength and dynamics,^{15,20,28}
9 and it could have relevance for the photographic memory of details that some autistic
10 individuals may have.

11 The concept of autism-related risk haplotypes of the *LRFN5* locus, variable locus
12 structure influenced by the individual's sex, and early allelic homozygotisation, fits
13 well with the complex pattern of ASD inheritance.⁴ In the families with the ASD-
14 susceptible A-haplotype described here, all ASD males were born to carrier mothers,
15 and one non-penetrant male was also recorded (Table 1). Of note, this male was hemi-
16 methylated at rs144497930, i.e. at the major TAD-peak around the CTCF binding site
17 in Figure 4 (Supplementary Table 1). Other family members and control individuals
18 were fully methylated at this CpG. Maybe the epigenetic pattern predisposing to
19 autism was not established during his early embryogenesis. In a "locus-interaction-
20 determines-ASD-susceptibility" model, this makes sense and would also explain why
21 rare deletions recorded in DECIPHER may be pathogenic despite inheritance from a
22 presumed normal male or female parent (Supplementary Figure 3). Such deletions
23 could interfere with allelic interaction and epigenetic regulation. It should also be
24 noted that while many different deletions may be seen in the *LRFN5* locus
25 (Supplementary Figure 6), duplications are rare, also in DECIPHER (Supplementary

1 Figure 4). Finally, the whole *LRFN5* locus or the *LRFN5* gene itself can be disrupted
2 or deleted, apparently without (additional) phenotypic consequences, as indicated
3 from the DECIPHER database, a gnomAD pLI of 0.56,²⁶ and our own diagnostic
4 genomic copy-number records. This is as expected if stochastic monoallelic
5 expression is the rule.

6 **LIMITATIONS**

7 The capture-HiC and chromatin chIP-on-chip experiments were all done on human
8 skin fibroblasts, not brain cells, which is the relevant tissue for *LRFN5* expression. As
9 the TAD (3D) structure and probably also the chromatin (2D) structure can be activity
10 dependent, the relevance of these finding for *LRFN5* locus structure in brain cells is
11 yet unknown, but TAD structure tends to be conserved between cell types. The size of
12 the control cohort exploring sex-related differences in H3K9me3 profiles of the
13 *LRFN5* locus is small (6 males +7 females), and these data need to be confirmed in an
14 independent and preferably larger cohort to be certain of a sex difference. Hopefully,
15 an easier method to investigate such changes will be developed. Finally, the
16 monoallelic expression data (Supplementary Figure 5) is from a single experiment.

17 **CONCLUSIONS**

18 Our work suggests that *LRFN5* regulation could be Y-chromosome-dependent and
19 complex with a strikingly high frequency of early embryonic allelic interaction with
20 locus conversions. Structural variants in this locus are linked to autism, and we
21 speculate that this locus is a significant part of the genetic basis of synaptic memory.

1 **DECLARATIONS**

2 **Ethical board approvals**

3 The methylation study on the anonymized ASD cohort was approved by the Ethical
4 Review Board of Northern Norway, REK-Nord# 2013/965. The study on sex-
5 dependent differences in the epigenetic profile of the *LRFN5* locus was approved by
6 the Ethical Review Board of Western Norway, REK-Vest# 2016/25.

7 **Consent for publication**

8 All involved members of the two families with two autistic brothers, including the
9 brothers themselves, have been counselled about these results and have given consent
10 to publication, recorded in the hospital's journal. Other individuals participating in
11 this study were anonymised.

12 **Availability of data and materials**

13 Data generated during this study are available from the corresponding author on
14 reasonable request. The chIP-on-chip data can be shared through custom tracks in the
15 UCSC browser.

16 **Competing interest**

17 The authors have no conflicts of interest to declare.

18 **Funding**

19 This work was supported by HelseVest grant #911459.

20 **Author contributions**

21
22 HL did all chromatin profiling experiments, MR did the capture-HiC experiments
23 with data-analysis and figure preparation, NdL and JYH-K delivered the Dutch

1 deletion frequency data, AJ did the human foetal brain cell *LRFN5* expression test,
2 BIH, SB, DdB, SM and MS gave valuable input during the experimental and
3 scientific process and participated in analysing the data. GH counselled the patients
4 and families, designed the research project, wrote research applications, plotted and
5 analysed chromatin and allele frequency data, made most of the figures and all tables,
6 and wrote the manuscript. All authors have reviewed and approved the final version.

7 **Acknowledgements**

8 We are most grateful to the family members having contributed blood samples and
9 skin biopsies, and for their interest in contributing to research. They have all been
10 informed about the results by the senior author. This work could not have been done
11 without the expert technical assistance of Hilde E Rusaas, Lene Hjertnes and Atle
12 Brendehaug. We are also grateful to Sofia Douzgou Houge for critical feedback on
13 the manuscript.

14

15

16

1 **References**

- 2 1 Newschaffer, C. J. *et al.* The epidemiology of autism spectrum disorders. *Annu Rev*
3 *Public Health* **28**, 235-258, doi:10.1146/annurev.publhealth.28.021406.144007
4 (2007).
- 5 2 Amendola, L. M. *et al.* Performance of ACMG-AMP Variant-Interpretation
6 Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory
7 Research Consortium. *Am J Hum Genet* **98**, 1067-1076,
8 doi:10.1016/j.ajhg.2016.03.024 (2016).
- 9 3 Gillberg, C., Cederlund, M., Lamberg, K. & Zeijlson, L. Brief report: "the autism
10 epidemic". The registered prevalence of autism in a Swedish urban area. *J Autism*
11 *Dev Disord* **36**, 429-435, doi:10.1007/s10803-006-0081-6 (2006).
- 12 4 Zhao, X. *et al.* A unified genetic theory for sporadic and inherited autism. *Proc Natl*
13 *Acad Sci U S A* **104**, 12831-12836, doi:10.1073/pnas.0705803104 (2007).
- 14 5 Lauritsen, M. B., Pedersen, C. B. & Mortensen, P. B. Effects of familial risk factors
15 and place of birth on the risk of autism: a nationwide register-based study. *J Child*
16 *Psychol Psychiatry* **46**, 963-971, doi:10.1111/j.1469-7610.2004.00391.x (2005).
- 17 6 Devlin, B. & Scherer, S. W. Genetic architecture in autism spectrum disorder. *Curr*
18 *Opin Genet Dev* **22**, 229-237, doi:10.1016/j.gde.2012.03.002 (2012).
- 19 7 Weiss, L. A. *et al.* Association between microdeletion and microduplication at
20 16p11.2 and autism. *N Engl J Med* **358**, 667-675, doi:10.1056/NEJMoa075974
21 (2008).
- 22 8 Sebat, J. *et al.* Strong association of de novo copy number mutations with autism.
23 *Science* **316**, 445-449, doi:10.1126/science.1138659 (2007).
- 24 9 Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum
25 disorders. *Nature* **459**, 528-533, doi:10.1038/nature07999 (2009).
- 26 10 Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C. Meta-
27 analysis of GWAS of over 16,000 individuals with autism spectrum disorder

1 highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia.
2 *Mol Autism* **8**, 21, doi:10.1186/s13229-017-0137-9 (2017).

3 11 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum
4 disorder. *Nat Genet* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).

5 12 Pinto, D. *et al.* Functional impact of global rare copy number variation in autism
6 spectrum disorders. *Nature* **466**, 368-372, doi:10.1038/nature09146 (2010).

7 13 Casey, J. P. *et al.* A novel approach of homozygous haplotype sharing identifies
8 candidate genes in autism spectrum disorder. *Hum Genet* **131**, 565-579,
9 doi:10.1007/s00439-011-1094-6 (2012).

10 14 Ko, J. *et al.* SALM synaptic cell adhesion-like molecules regulate the differentiation
11 of excitatory synapses. *Neuron* **50**, 233-245, doi:10.1016/j.neuron.2006.04.005
12 (2006).

13 15 Morimura, N., Inoue, T., Katayama, K. & Aruga, J. Comparative analysis of
14 structure, expression and PSD95-binding capacity of Lrfr, a novel family of neuronal
15 transmembrane proteins. *Gene* **380**, 72-83, doi:10.1016/j.gene.2006.05.014 (2006).

16 16 Seabold, G. K. *et al.* The SALM family of adhesion-like molecules forms
17 heteromeric and homomeric complexes. *J Biol Chem* **283**, 8395-8405,
18 doi:10.1074/jbc.M709456200 (2008).

19 17 Ko, J. & Kim, E. Leucine-rich repeat proteins of synapses. *J Neurosci Res* **85**, 2824-
20 2832, doi:10.1002/jnr.21306 (2007).

21 18 Choi, Y. *et al.* SALM5 trans-synaptically interacts with LAR-RPTPs in a splicing-
22 dependent manner to regulate synapse development. *Sci Rep* **6**, 26676,
23 doi:10.1038/srep26676 (2016).

24 19 Goto-Ito, S. *et al.* Structural basis of trans-synaptic interactions between PTPdelta
25 and SALMs for inducing synapse formation. *Nat Commun* **9**, 269,
26 doi:10.1038/s41467-017-02417-z (2018).

1 20 Lin, Z., Liu, J., Ding, H., Xu, F. & Liu, H. Structural basis of SALM5-induced
2 PTPdelta dimerization for synaptic differentiation. *Nat Commun* **9**, 268,
3 doi:10.1038/s41467-017-02414-2 (2018).

4 21 Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes.
5 *Annu Rev Cell Dev Biol* **33**, 265-289, doi:10.1146/annurev-cellbio-100616-060531
6 (2017).

7 22 Kragesteen, B. K. *et al.* Dynamic 3D chromatin architecture contributes to enhancer
8 specificity and limb morphogenesis. *Nat Genet* **50**, 1463-1473, doi:10.1038/s41588-
9 018-0221-x (2018).

10 23 El-Akabawy, G. *et al.* Implantation of undifferentiated and pre-differentiated human
11 neural stem cells in the R6/2 transgenic mouse model of Huntington's disease. *BMC*
12 *Neurosci* **13**, 97, doi:10.1186/1471-2202-13-97 (2012).

13 24 de Bruijn, D. R. *et al.* Severe Progressive Autism Associated with Two de novo
14 Changes: A 2.6-Mb 2q31.1 Deletion and a Balanced t(14;21)(q21.1;p11.2)
15 Translocation with Long-Range Epigenetic Silencing of LRFN5 Expression. *Mol*
16 *Syndromol* **1**, 46-57, doi:10.1159/000280290 (2010).

17 25 Cappuccio, G. *et al.* Microdeletion of pseudogene chr14.232.a affects LRFN5
18 expression in cells of a patient with autism spectrum disorder. *Eur J Hum Genet* **27**,
19 1475-1480, doi:10.1038/s41431-019-0430-5 (2019).

20 26 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation
21 in 141,456 humans. *Nature* **581**, 434-443, doi:10.1038/s41586-020-2308-7 (2020).

22 27 Benovoy, D. & Drouin, G. Ectopic gene conversions in the human genome.
23 *Genomics* **93**, 27-32, doi:10.1016/j.ygeno.2008.09.007 (2009).

24 28 Mah, W. *et al.* Selected SALM (synaptic adhesion-like molecule) family proteins
25 regulate synapse formation. *J Neurosci* **30**, 5559-5568,
26 doi:10.1523/JNEUROSCI.4839-09.2010 (2010).

27

1 Tables

2 **Table 1**

3 LRFN5 locus haplotypes in affected individuals and their families, all having
4 overlapping deletions on different haplotypes (A-E) just upstream of the LRFN5
5 promoter.

6

Fam #	Individual	DD/ASD	Deletion haplotype ¹
1	Father	-	-
1	Mother	-	A
1	Son ²	-	A
1	Son	ASD	A
1	Son	ASD	A
1	Son	-	-
1	Daughter	-	-
1	Daughter	-	-
2	Mat grandfather	-	-
2	Mat grandmother	-	A
2	Mat uncle	-	A
2	Father	-	-
2	Mother	-	A
2	Son	ASD	A
2	Son	ASD	A
2	Son	-	-
2	Daughter	-	-
3	Father	-	B
3	Mother	-	-
3	Son	DD	B
4	Father	-	-
4	Mother	-	C
4	Son	DD/ASD	C
5	Father	-	-
5	Mother	-	D
5	Daughter	Rett-like	D
6	Adult male	Mild ID/ schizophrenia	E

7

8 ¹ Deletion sizes (kb, hg19): A: 41846-42020, B: 41846-42023,
9 C: 41867-41979, D: 41822-41933, E: 41836-42004. The haplo-
10 types were also determined based on *LRFN5* locus SNP pattern.

11 ² This individual was a SRY-negative 46,XX-DSD male.

12

1 Table 2

2 Allelic distribution of a common 60 kb deletion* (from 41608 to 41657 kb in
3 GRCh37 based on SNP-array data) in three patient cohorts and one student cohort.
4 The D-column shows the fold difference of allele frequencies between the observed
5 del/del numbers and expected del/del numbers from Hardy-Weinberg equilibrium.

	n	wt/wt	wt/del	del/del	Deletion MAF	wt/wt	wt/del	del/del	Δ	% loss of wt/del to wt/wt and del/del
Patient cohort I (NO) <i>expected if HWE</i>	850	649	168	33	0.138	0.760 0.743	0.200 0.238	0.040 0.019	2.1	10% + 10%
Patient cohort II (NO) <i>expected if HWE</i>	4843	3582	1025	236	0.155	0.740 0.714	0.212 0.262	0.049 0.024	2.0	12% + 12%
Patient cohort III (NL) <i>expected if HWE</i>	3064	2600	389	75	0.088	0.850 0.832	0.130 0.161	0.020 0.008	2.5	13% + 13%
Student cohort (NL) <i>expected if HWE</i>	1416	1247	114	55	0.079	0.881 0.848	0.081 0.146	0.039 0.006	6.5	40% + 40%

*size based on gnomAD v3.1.1 whole-genome sequencing data is 41609383-41669664 (60.3 kb) with MAF 0.139 (2990/21518 alleles).

7

8

9 Figure legends

10 Figure 1

11 Capture-HiC results of a family trio (index male with ASD, his mother and father)
12 and a control female (of the same age as index' mother). Arrows indicate TAD
13 junctions (A and B). The small diamond just centromeric to TAD junction A
14 (corresponds to the *LRFN5* promoter) marks the 172 kb familial deletion.

15 Figure 2

16 Top: *LRFN5* locus H3K9me3 chIP-on-chip profiles from index with ASD, his 46,XX-
17 DSD non-autistic brother, his normal mother and his normal father (top four lanes).

1 Middle: Profiles from 5 control males. Bottom: Profiles from 7 control females. Note
2 the profile variability around the *LRFN5* gene in contrast to the flanking H3K9me3-
3 enriched domains.

4 **Figure 3**

5 Quantification of data from Figure 2: *LRFN5* locus H3K9me3 chromatin levels
6 (relative to genomic average) in fibroblasts from control males (M1-M6) and females
7 (F1-F7). The investigated positions are indicated in the table below, and the bar
8 diagram above illustrates the individual male/female distribution. Note that the groups
9 are overlapping.

10 **Figure 4**

11 Top: the capture-HiC result of the index male with ASD, showing three subTADs
12 inside the 5.4 Mb mega-TAD. The small diamond marked with an asterisk indicates
13 the 172 kb deletion, and the position of the common 60 kb deletion polymorphism is
14 also shown. Middle part shows CTCF sites (marking e.g. TAD boundaries and the
15 peak of the mega-TAD) and single-cell ATAC sites in selected cell types (indicating
16 areas of open chromatin). On the bottom is the H3K9me3 chromatin profiles of the
17 two individuals with the most pronounced male and female pattern from the control
18 profiles displayed in Figure 2.

19

SUPPLEMENTARY DATA

Supplementary Figure 1

14q21 chIP-on-chip chromatin profiles from a male control (Ctr), a male with Coffin-Siris syndrome and a *de novo* 6;14 translocation disrupting *ARID1B* (6;14), and a female with a *de novo* 14;21 translocation and severe ID/ASD (14;21). The arrows indicate approximate (FISH verified) translocation breakpoints. A: H3K4me3 profiles, B: H3K27me3 profiles, C: H3K9me3 profiles. Only the 14;21-chromatin pattern is markedly different from the Ctr and 6;14 cases: The H3K4me3 and H3K27me3 profiles for the *LRFN5* locus on derivative chromosome 21 show a more “open” chromatin pattern, and the H3K9me3 profiles upstream of the *FBXO33* gene on derivative chromosome 14 (upstream of 14;21 arrow marking the translocation breakpoint) show a more “closed” chromatin pattern.

Supplementary Figure 2

ChIP-on-chip chromatin profiles from the female with a *de novo* 14;21 translocation (same individual as in Supplementary Figure 1) as well as a *de novo* 2q31.1 deletion (panel A), compared to a male with a *de novo* 17q23.1 deletion (panel B). Black lines indicate deletions. The “compensatory” increase in H3K4me3 signal corresponding to the 2q31.1 deletion (panel A, top lane) was not reproduced by the 17q23.1 deletion.

Supplementary Figure 3

Overview of *LRFN5* locus deletions registered in DECIPHER. The registered phenotype is shown above the deletion: ID = intellectual disability, DD = developmental delay, ASD = autism spectrum disorder, incid? = likely incidental finding (no phenotype recorded), mat = maternally inherited, pat = paternally inherited. On the bottom, three deletions found in own patient records are shown (red

1 lines framed in black). The vertical green shades indicate *LRFN5* locus regions were
2 CNVs appear to be of no clinical consequence. The vertical red shades indicate
3 *LRFN5* locus regions where CNVs could increase susceptibility for DD/ASD.

4 **Supplementary Figure 4**

5 Overview of *LRFN5* locus duplications registered in DECIPHER. The registered
6 phenotype is shown above the deletion: ID = intellectual disability, DD =
7 developmental delay, epil = seizures, mat = maternally inherited, pat = paternally
8 inherited. Note that duplications are rarer than deletions in this region, which is also
9 the case in the DGV database (database of genomic variants; dgv.tcag.ca, see
10 Supplementary Figure 6).

11 **Supplementary Figure 5**

12 Cells from 12-week human foetal brain were examined for allelic balance of *LRFN5*
13 expression. Three SNP found to be informative (heterozygous) by an Illumina SNP
14 array were interrogated by cDNA Sanger sequencing: rs1137567 in exon 1 (5'UTR of
15 mRNA), rs6572117 in exon 4 (protein-coding sequence) and rs7148658 in intron 4. In
16 four (polyclonal) foetal brain samples (FB 11189/11242/11196/11192), both alleles
17 were expressed. In three striatal neuron cell cultures (STRO 05/08/11), only one allele
18 was expressed, suggesting monoallelic expression. Since the same allele was
19 expressed in all three cell lines, one cannot tell if the monoallelic expression was
20 stochastic.

21 **Supplementary Figure 6**

22 Overview of *LRFN5* locus, taken from the UCSC browser (version GRCh37), with a
23 transparent overlay of the capture-HiC result of the ASD index boy. Lane on top
24 shows computer predicted SOX9 and SRY binding sites, in the middle CTCF binding

1 sites in three different cell types can be found (note that junction B also corresponds
2 to CTCF signals), and on the bottom DECIPHER and DGV copy number variants are
3 found. In both, deletions are more common than duplications. Red vertical bars
4 indicate areas where copy number variants have been found in patients with ID/ASD,
5 and green areas where this so far is not the case.

Supplementary Table 1

Targeted MLPA-based DNA methylation results in the genomic area underlying the mega-TAD-peak in a small group of selected individuals. Please note that the normal male with the A-haplotype was hemi-methylated in a position fully methylated in other individuals.

Individual	Haplo-type	TAD-peak lncRNA CpG 42529925-6*	TAD-peak lncRNA CpG 42529958-9*	TAD-peak conserved CpG 42498027-8*	TAD-peak conserved CpG 42498046-7*
ASD male	A	~100%	~100%	~10%	~10%
ASD male	A	~100%	~100%	~5%	~5%
Normal male	A	~100%	~50%**	~5%	~5%
Normal female	A	~100%	tech. failure	~20%	~20%
ASD male	-	~100%	~100%	~10%	~10%
ASD male	-	~100%	~100%	~10%	~10%
Normal male	-	~100%	~100%	~10%	~10%
Normal female	-	~100%	~100%	~10%	~10%
Normal female	D	~100%	~100%	~50%	~50%
Normal female	C	~100%	~100%	~20%	~20%

*CpG positions corresponding to the mega-TAD peak on chromosome 14 are according to GRCh37. Haplotypes are according to Extended Data Table 1, i.e. all from individuals with a small locus deletion. ** Sanger sequencing verified that this finding was not due to a SNP, i.e. a C to T transition.

Figures

Figure 1

Capture-HiC results of a family trio (index male with ASD, his mother and father) and a control female (of the same age as index' mother). Arrows indicate TAD junctions (A and B). The small diamond just centromeric to TAD junction A (corresponds to the LRFN5 promoter) marks the 172 kb familial deletion.

Figure 2

Top: LRFN5 locus H3K9me3 chIP-on-chip profiles from index with ASD, his 46,XX,17 DSD non-autistic brother, his normal mother and his normal father (top four lanes). Middle: Profiles from 5 control males. Bottom: 1 Profiles from 7 control females. Note the profile variability around the LRFN5 gene in contrast to the flanking H3K9me3- enriched domains.

Display of the individual *LRFN5* chromatin data from males (M1-M6) and females (F1-F7) based on the table below

H3K9me3 chromatin scores (relative to genomic average) in fibroblasts from control males (M's) and females (F's)

Genomic position	M1	M2	M3	M4	M5	M6	F1	F2	F3	F4	F5	F6	F7	Average males	Average females	Diff
<i>LRFN5</i> (42050-42400 kb)	3	11	4	9	26	-5	-5	-19	-5	5	-24	-26	-19	8 (0 - 16)	-13 (-22 - -4)	21*
Reference pos. (40550-40900 kb)	54	57	51	38	68	60	46	58	60	61	43	60	55	55 (47 - 63)	55 (49 - 61)	0

The chromatin scores are based on 673-1907 measurement points (= one oligo on the array), on average around 700 measurements per position. The reference position was chosen to be in the middle of the heterochromatin stretch upstream of *LRFN5*. 95% confidence intervals are shown in parentheses. *p = 0.005 (Student two-sided t-test).

Figure 3

Quantification of data from Figure 2: *LRFN5* locus H3K9me3 chromatin levels (relative to genomic average) in fibroblasts from control males (M1-M6) and females (F1-F7). The investigated positions are indicated in the table below, and the bar diagram above illustrates the individual male/female distribution. Note that the groups are overlapping.

Figure 4

Top: the capture-HiC result of the index male with ASD, showing three subTADs inside the 5.4 Mb mega-TAD. The small diamond marked with an asterisk indicates the 172 kb deletion, and the position of the common 60 kb deletion polymorphism is also shown. Middle part shows CTCF sites (marking e.g. TAD boundaries and the peak of the mega-TAD) and single-cell ATAC sites in selected cell types (indicating areas of open chromatin). On the bottom is the H3K9me3 chromatin profiles of the two individuals with the most pronounced male and female pattern from the control profiles displayed in Figure 2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1teffect.pdf](#)
- [FigureS2del2qand17qeffect.pdf](#)
- [FigureS3DELLRFN5.pdf](#)
- [FigureS4DUPLRFN5.pdf](#)
- [FigureS5monoallelicbrainexpression.pdf](#)
- [FigureS6captureHiCinUCSC.jpg](#)
- [TableS1MLPAmetresults.docx](#)