The American Journal of Human Genetics, *91*

Supplemental Data

A Noncoding, Regulatory Mutation Implicates *HCFC1* in Nonsyndromic Intellectual Disability

Lingli Huang, Lachlan A. Jolly, Saffron Willis-Owen, Alison Gardner, Raman Kumar, Evelyn Douglas, Cheryl Shoubridge, Dagmar Wieczorek, Andreas Tzschach, Monika Cohen, Anna Hackett, Michael Field, Guy Froyen, Hao Hu, Stefan A. Haas, Hans-Hilger Ropers, Vera M. Kalscheuer, Mark A. Corbett, and Jozef Gecz

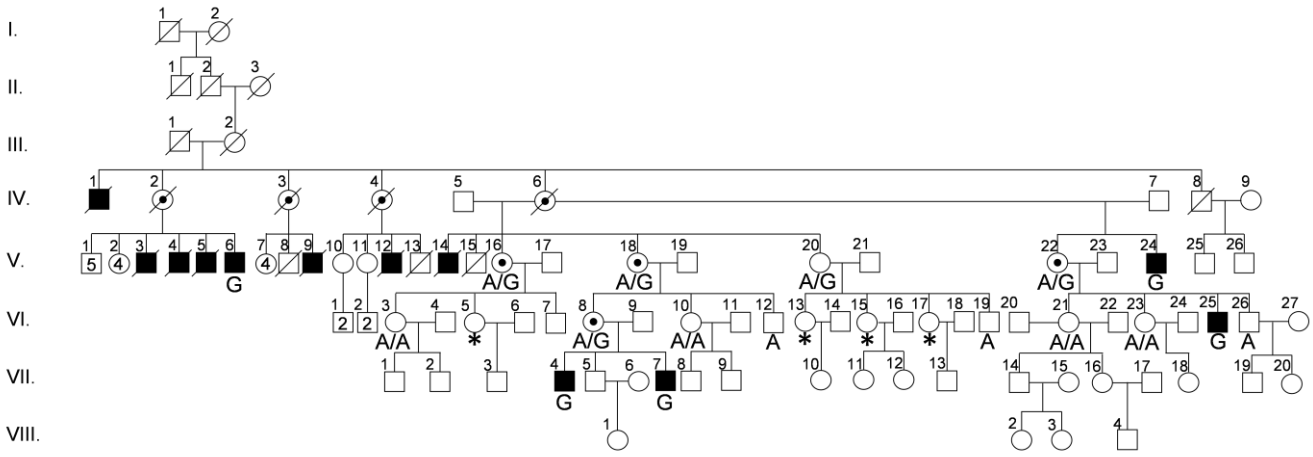

Figure S1. Pedigree of the MRX3 Family

Updated pedigree of the MRX3 family is shown. *HCFC1* chrX:152,890,455A>G (hg18) genotyped individuals are shown with either their genotype (A, A/A, G, or A/G) or an asterisk (*), for those individuals, where the genotype is not disclosed (but was typed and segregated with their clinical status). Open circles show females; circles with a dot in the middle show obligate carrier females; empty squares show males, solid squares show affected males, crossed symbols show deceased individuals.

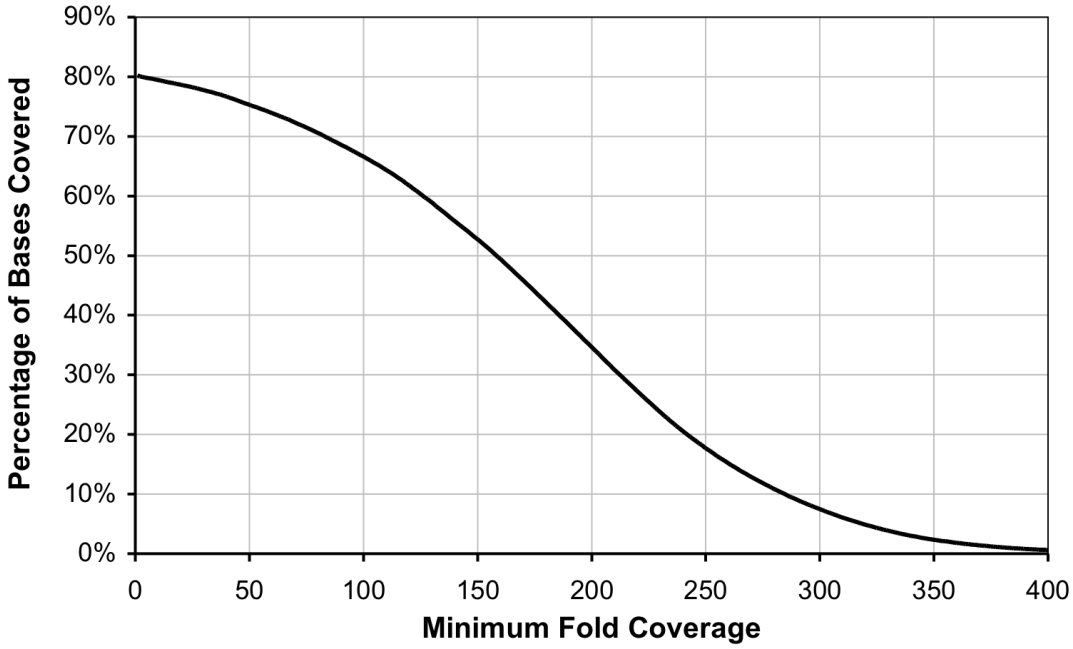

Figure S2. Coverage Plot for All Bases Tiled on the Targeted Capture Array

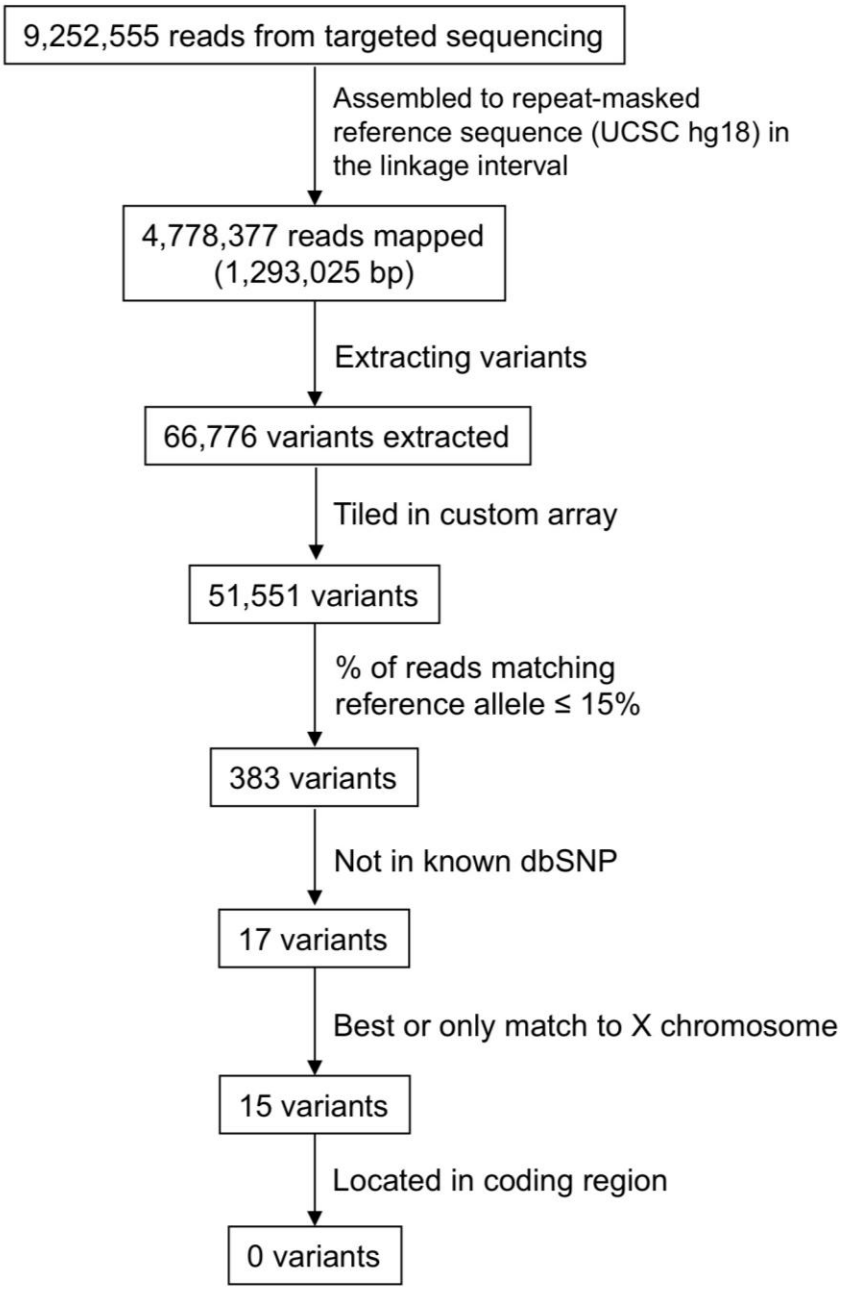

Figure S3. Work Flow of Detecting and Prioritizing Variants

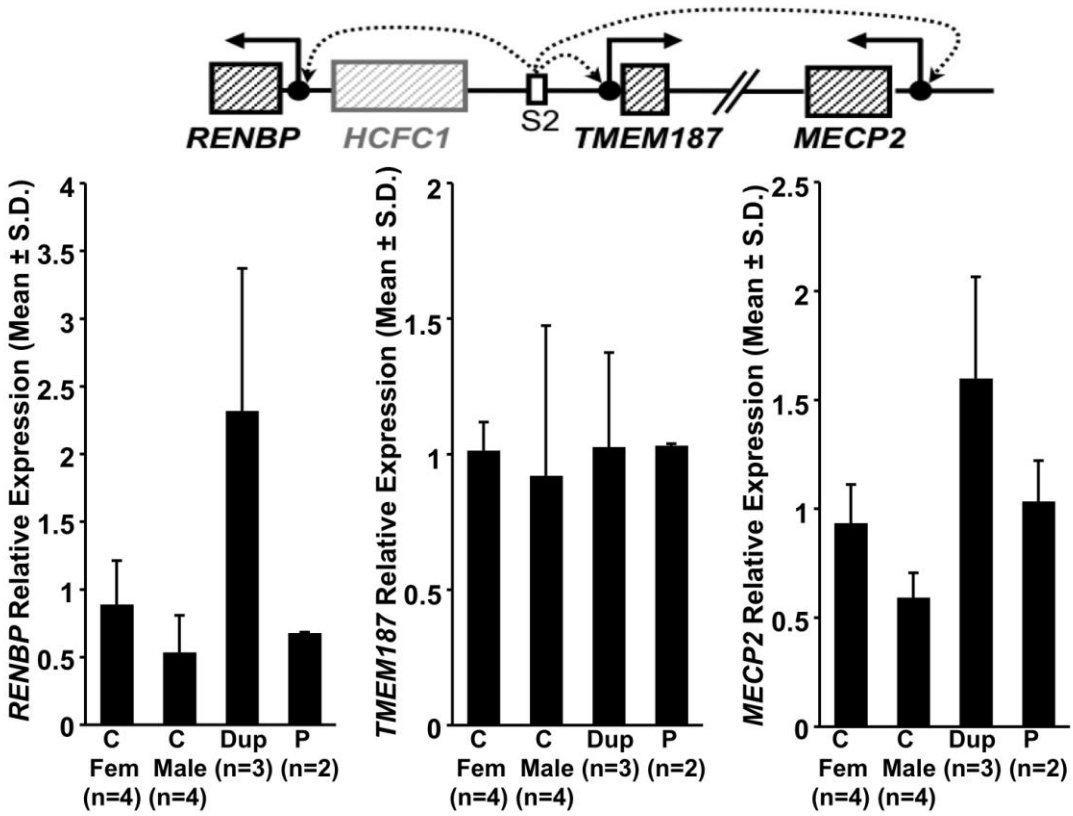

Figure S4. Possible Effects of YY1 Binding Site S2 Mutation on *RENBP*, *TMEM187*, and *MECP2*

Upper panel, schematic diagram of the genomic structure of the *HCFC1* gene and possible responsive genes to the mutation in the YY1 binding site S2. Arrows with dashed lines represent the possible pattern of effects on the transcriptional activity of genes in the region. Lower panel, qRT-PCR showing the relative expression of *RENBP*, *TMEM187* and *MECP2* to β-actin expression in controls (females n = 4, C Fem; and males n = 4 C Male), *MECP2* gene duplications involving these genes (n = 3; Dup), and the MRX3 patients (n = 2; P). qRT-PCR was carried out as previously described (Nat. Genet. 39, 1127–1133, 2007), using primers listed in Table S4. Data was derived from three independent experiments. Graphs represent mean expression ± SD.

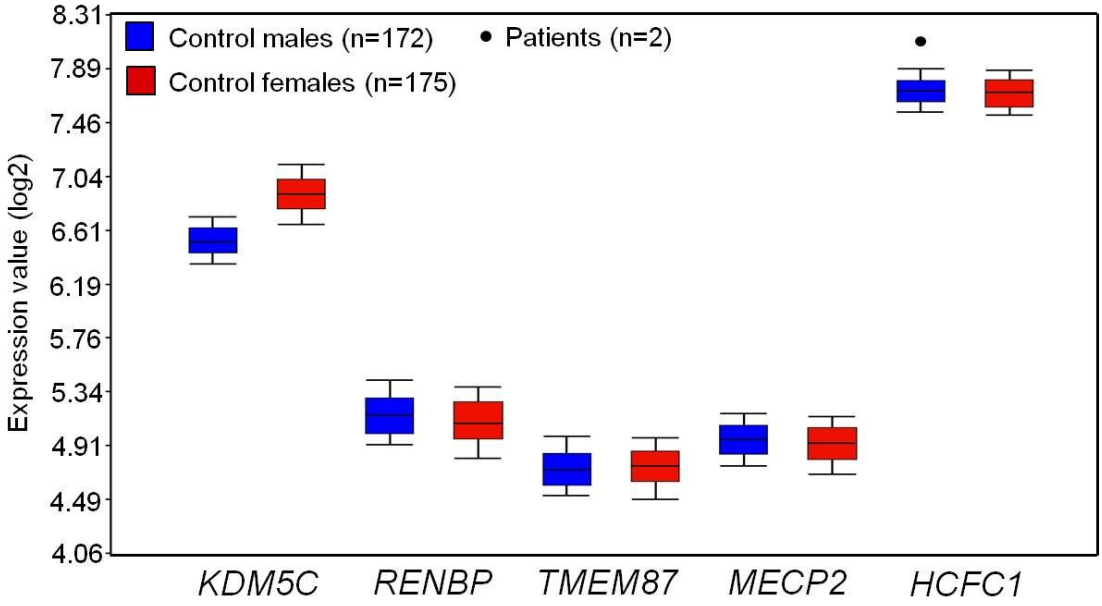

Figure S5. Relative Expression of *KDM5C, RENBP, TMEM87, MECP2*, and *HCFC1* of the Patients (n = 2) as Determined by qRT-PCR to Control Males (n = 192) and Females (n = 175) Measured by Microarray (GSE1485)

Box plots illustrate the expression distribution in LCLs of normal males and females of *KDM5C*, which is known to escape X inactivation, and *RENBP*, *TMEM87*, and *MECP2* that are known to be subjected to X inactivation. Expression of the patients (black dot near the *HCFC1* mRNA box plot) was determined as fold change by qRT-PCRs and converted into log2 expression relative to the mean of the control group. Microarray analysis was carried out using Partek Genomic Suite V6.6 as previously described (Nature 430, 743-747, 2004). qRT-PCR was carried out using primers listed in Table S4.

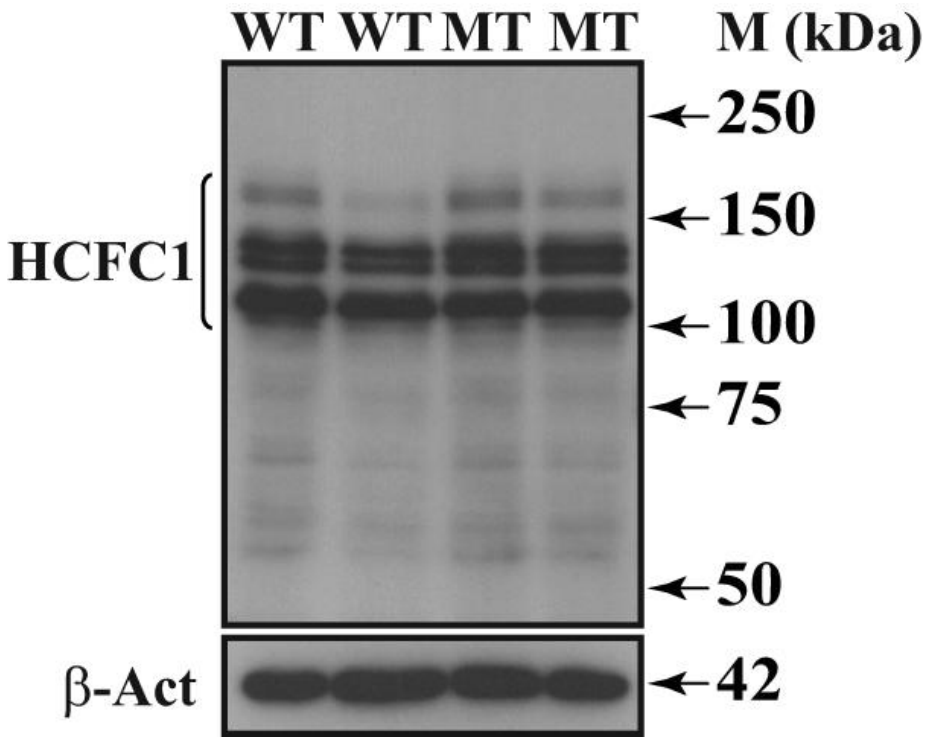

Figure S6. Total Protein Lysates (12 µg each) from Normal (WT; n = 2) and MRX3 (MT; n = 2) Male LCLs Were Immunoblotted According to Published Method (Mol Cancer Res. 4, 655-665, 2006) using a Rabbit Anti-HCFC1 Antibody (Cell 74, 115-125, 1993)

HCFC1 protein is post-translationally processed into multiple species, which are shown on the left hand side of the blot. Anti-β-actin antibody (Sigma) was used as a loading control (bottom panel).

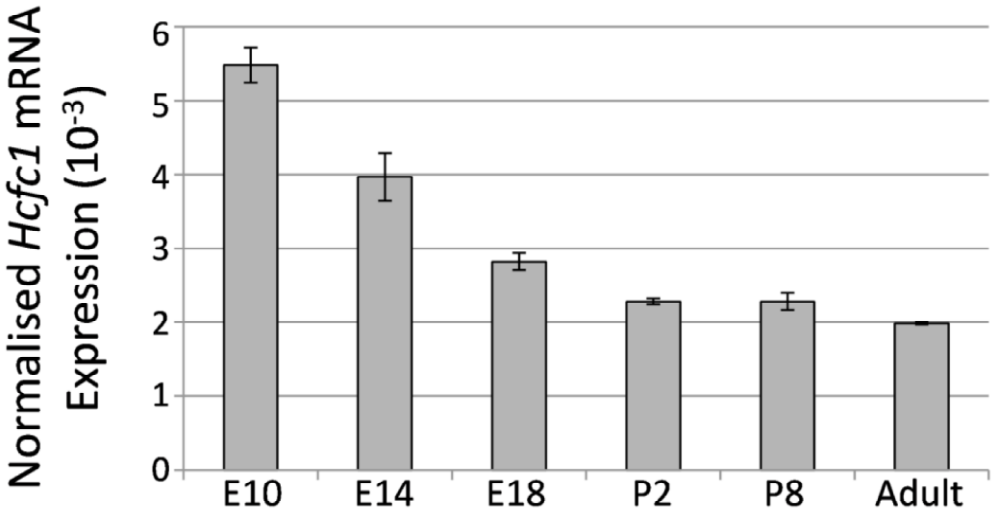

Figure S7. *Hcfc1* Is Highly Expressed during Embryonic Brain Development

qRT-PCR used to quantify *Hcfc1* mRNA at different stages of embryonic (E) and post-natal (P) stages of murine brain development. n ≥ 4 brains at each time point. qRT-PCR was carried out as previously described (Nat. Genet. 39, 1127–1133, 2007), using primers listed in Table S4 below. The embryonic expression profile is consistent with a role in proliferative cells – as these cells deplete through the generation of terminally differentiated cells so does *Hcfc1* expression. Expression during postnatal development, however suggests likely role in post-mitotic cells as well.

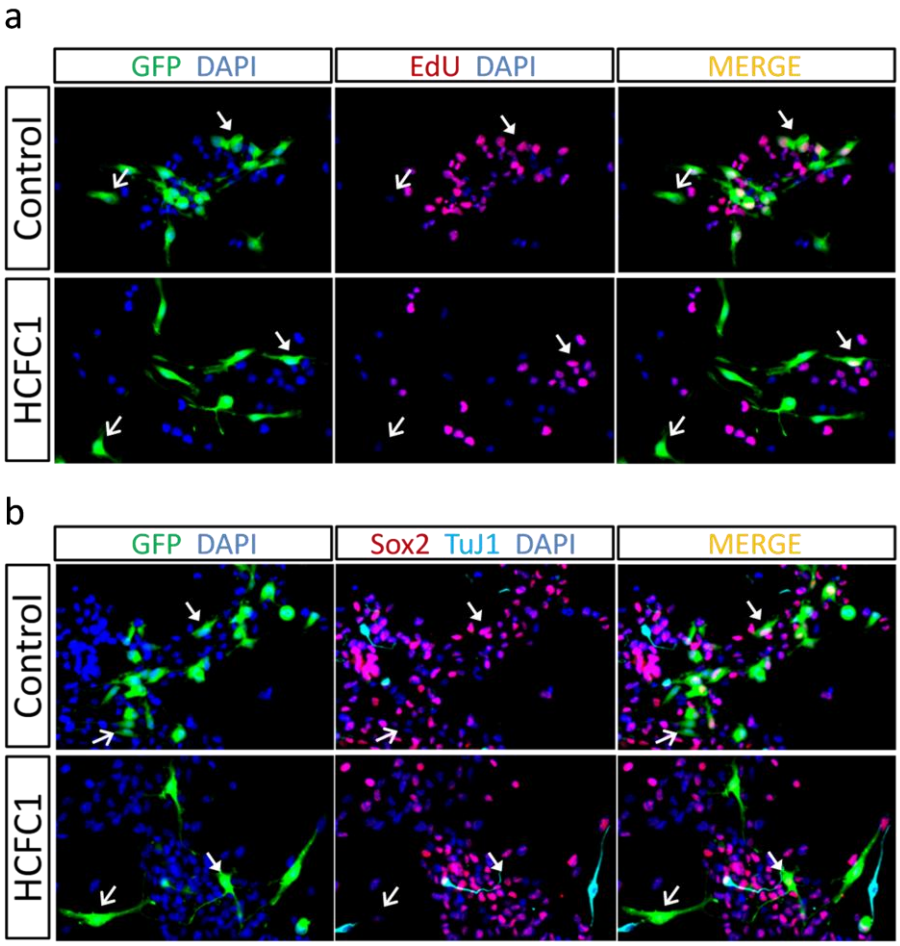

Figure S8. *HCFC1* Overexpression Reduces Number of NSCs

NSCs were isolated, grown and differentiated using described methods (Methods Mol Biol. 482, 143-158, 2009), and transfected via nucleofection as per manufactures instructions (Lonza). NSCs were transfected with plasmids expressing *GFP* (pMax-EGFP; Lonza) together with either an empty vector control (pcDNA3.1; Invitrogen) or *HCFC1* expression plasmid (pCGN-*HCFC1*). Transfected NSCs were grown for 3 days in adherent culture before analysis.

(A) Cells labelled for 12 hours with EdU. EdU labelling and detection was achieved using the Click-it EdU AlexaFluor647-Azide kit (Invitrogen). Representative immunofluorescent images identifying transfected cells via EGFP expression (green) and the presence of EdU incorporation (red). Nuclei are stained with DAPI (blue). Closed arrowheads indicate transfected cells labelled with EdU, open arrowheads indicate transfected cells not labelled with EdU.

(B) Representative immunofluorescent images of transfected cells (green) stained using antibodies against cell type specific marker proteins: NSCs (Sox2; red) and Neurons (TuJ1; cyan); Nuclei are stained with DAPI (blue). Closed arrowheads indicate transfected cells expressing Sox2, open arrowheads indicate transfected cells that do not. Note that cells over-expressing HCFC1 are less likely to be labelled with EdU, and also less likely to co-express Sox2 compared to controls. All staining and microscopy carried out as previously described (Mol. Biol. Cell. 20, 2015-2029, 2009).

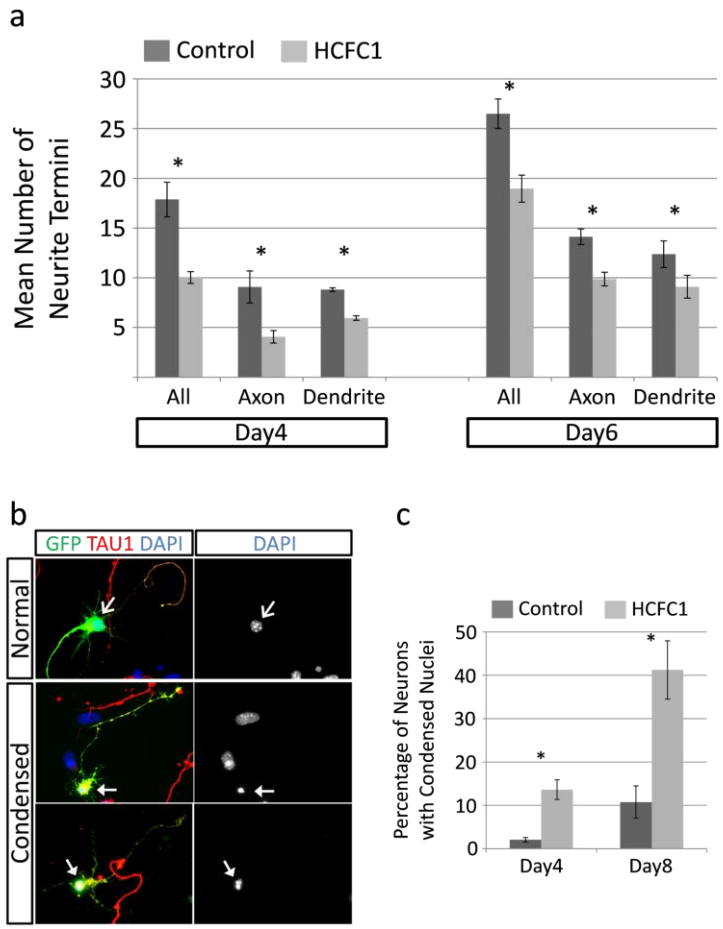

Figure S9. *HCFC1* Overexpression Reduces Hippocampal Neuronal Arborisation and Increases Neurotoxicity

Primary hippocampal neurons were isolated, transfected and grown as previously described (Am. J. Hum. Genet. 87, 371-375, 2005). Neurons were transfected with plasmids expressing *GFP* (pMax-EGFP; Lonza) together with either an empty vector control (pcDNA3.1; Invitrogen) or a *HCFC1* expression plasmid (pCGN-*HCFC1*).

(A) *HCFC1* over-expression reduces hippocampal outgrowth. Transfected cells were identified using GFP expression, and axons and dendrites were identified using immunoreactivity to TAU1 and MAP2 antibodies respectively, as previously described (Am. J. Hum. Genet. 87, 371-375, 2010). Average number of neurite termini identified at days 4 and 8 of differentiation determined as previously described (Am. J. Hum. Genet. 87, 371-375, 2010). At least 25 neurons scored per experiment done in triplicate.

(B and C) *HCFC1* overexpression is neurotoxic.

(A) Representative immunofluorescent images used to score neurons undergoing cell death. Transfected neurons identified using GFP expression, and TAU1 immunoreactivity. Neurons, which displayed nuclei with overt signs of condensation and fragmentation (closed arrowheads), were scored as dying, whilst neurons with normal nuclei (open arrowheads) were not. Note that dying neurons commonly displayed other morphological differences including blebbing membranes and vesicular inclusions. At least 300 neurons scored per experiment conducted in triplicate. All graphs represent the mean of triplicate experiments +/- SD. *$p<0.05$ by 2-tailed Student's t-test.

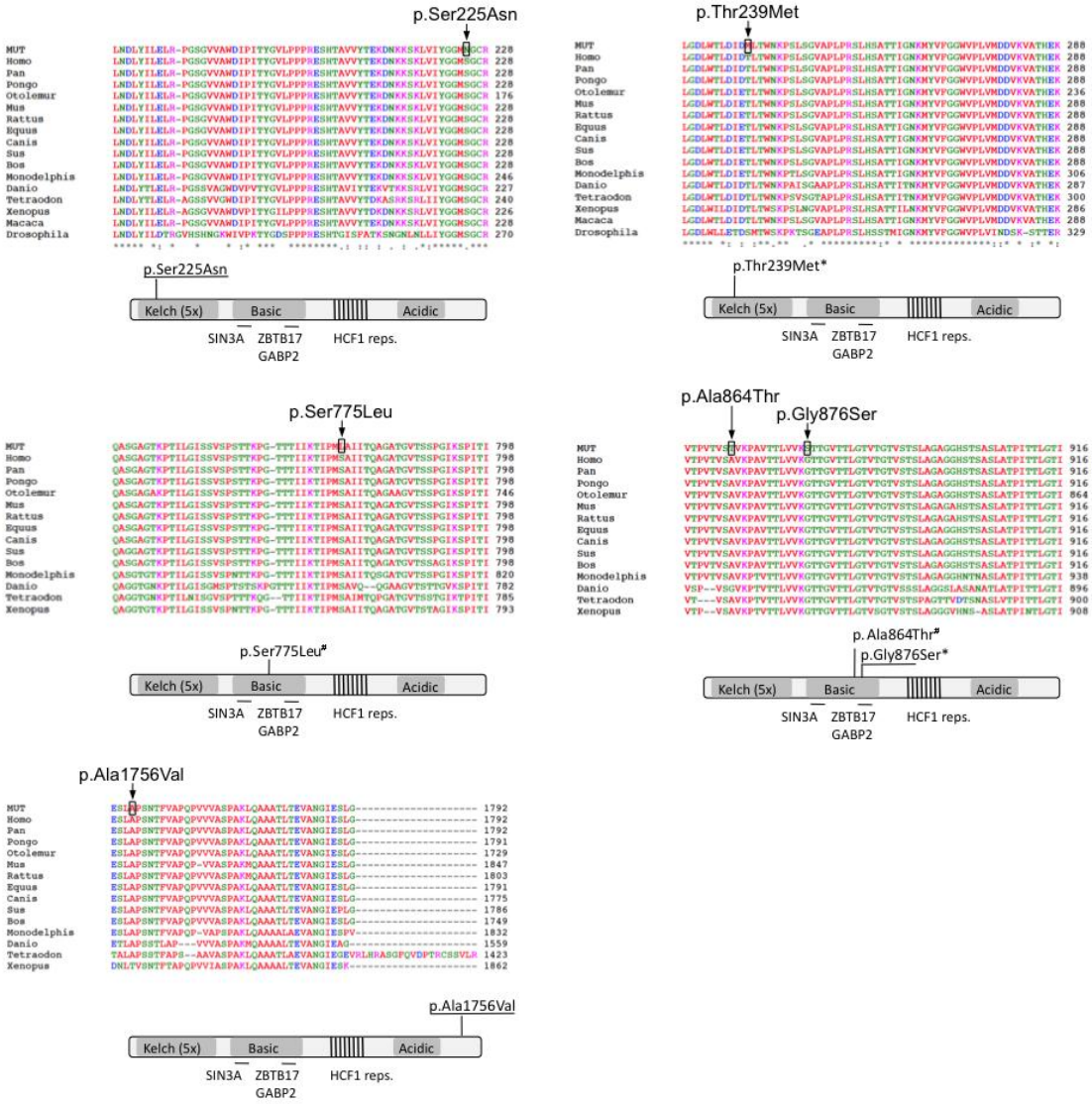

Figure S10. ClustalW Multiple Protein Alignment of Selected *HCFC1* Orthologs

ClustaW alignment has been performed at http://www.ebi.ac.uk/. The "MUT" sequence on top of each alignment indicates the mutant HCFC1 protein with that specific amino acid highlighted by a box and variant annotation. At the bottom of each panel there is a schematic of HCFC1 protein (similar to that shown in Figure 2E) showing the position of that specific amino acid within the HCFC1 protein. The underlined amino acid changes indicated those found as part of this study. The asterisk (*) indicate those also found by Piton et al. (Mol. Psychiatry 16, 867-880, 2011) and the hash (#) indicates those found by Tarpey et al. (Nat. Genet. 41, 535-543, 2009).

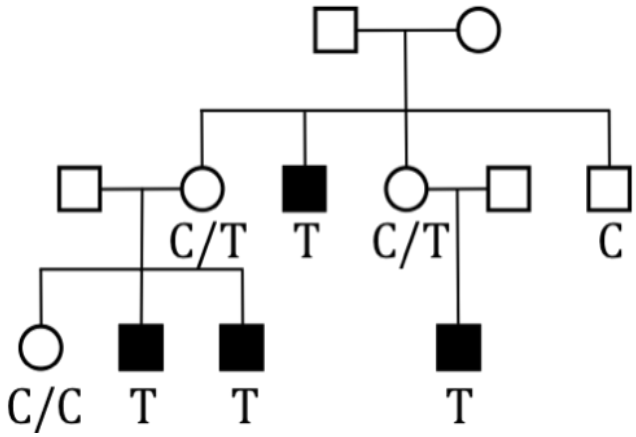

Figure S11. Pedigree of Family D144

The segregation of the *HCFC1* change (hg18) chrX:152881908C>T; p.Ser225Asn is shown. "C" refers to the wild type allele and "T" to the mutant allele.

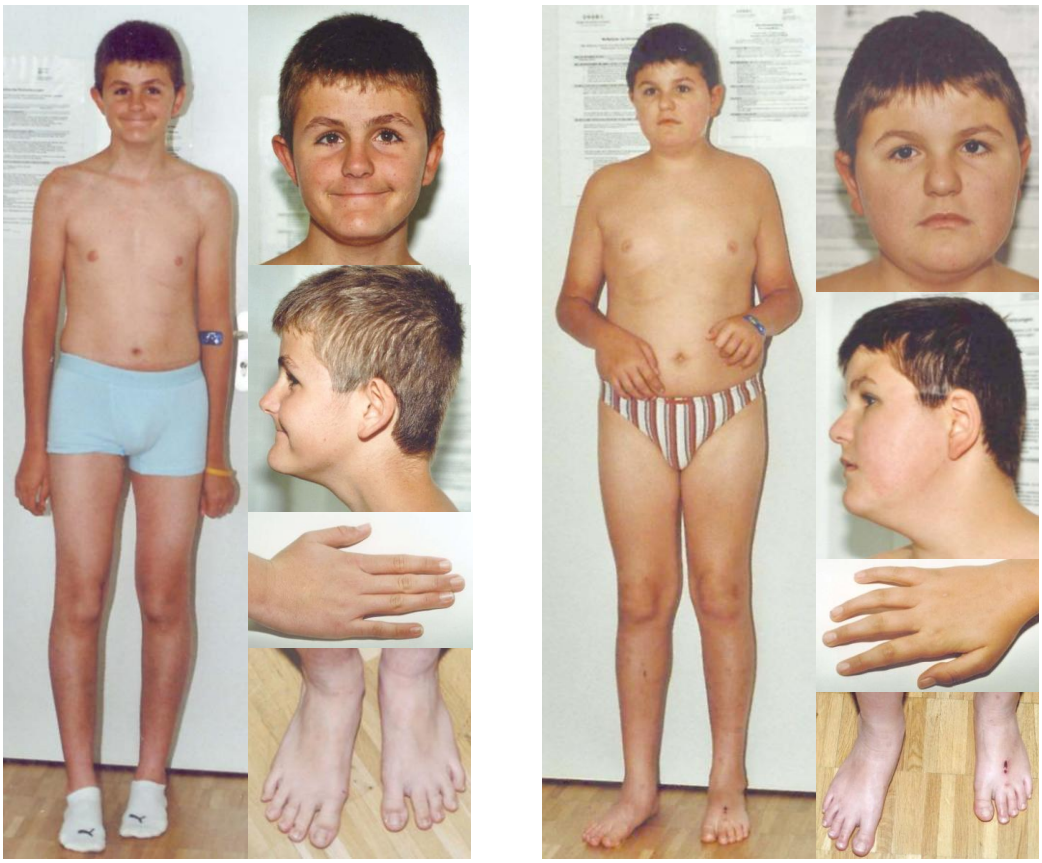

Figure S12. Patients from the D144 Family

Table S1. Statistics of the Data from Targeted Sequencing

	Total Tiled	Tiled and Covered	Tiled not Covered	Amplified[2]	Not Amplified[2]
Bases (nt)					
Total	941,355	755,814	185,541		
Repeat-Masked	179,194	978	178,216		
Not Repeat-Masked	762,161	754,836	7,325		
Percentage					
Total	100%	**80.29%**[1]	19.17%[1]		
Repeat-Masked	19.04%	0.13%	**96.05%**		
Not Repeat-Masked	80.96%	**99.87%**	3.95%		
GC Content					
Total	51.69%	53.20%	45.58%		
Repeat-Masked	44.80%	46.88%	44.79%		
Not Repeat-Masked	53.31%	53.20%	**64.28%**		
CCDS (nt)					
Total	114,762	113,418	1,344	988	356
Repeat-Masked	965	161	804	669	135
Not Repeat-Masked	113,797	113,257	540	319	221
CCDS Percentage					
Total	100%	**98.83%**[1]	1.17%[1]	100%	100%
Repeat-Masked	0.84%	0.14%	59.82%	67.71%	37.92%
Not Repeat-Masked	**99.16%**	**99.86%**	40.18%	32.29%	62.08%

[1] The total percentage would be 100% when calculating the percentage of Repeat-Masked and Not Repeat-Masked items in total items.
[2] Amplified by Sanger exon-sequencing (Nat. Genet. 41, 535-543, 2009) or not.

Table S2. Overview of Coding Regions Not Covered by Sequence Reads

Chromosome	chr_Start[1]	chr_End[1]	Span (bp)	Repeats (bp)	Unique Bases	Gene	Location	GC_Content	Sequence Characters
chrX	149612662	149612807	146	98	48	MTMR1	Exon1	84.93%	Repeats
chrX	152607224	152607266	43	0	43	SLC6A8 (known XLID genes)	Exon1	58.14%	Flanked with repeats
chrX	152853933	152853988	56	37	19	RENBP	Last exon	83.93%	Repeats
chrX	153325777	153325887	111	0	111	FAM50A	Exon1	64.86%	Flanked with repeats
		Total	356	135	221				

1 bp positions based on UCSC hg18 March 2006 genome build

Table S3. Overview of All Sequenced Variants in Custom Array but Not in dbSNP

Position on chrX (hg18)	Coverage	% Variant Allele	Sequence Variant[1]	Locations	Gene or Closest Gene	Alternative Allele in Other Species[2]
149676594	281	98.9%	A/G	Intronic	*MTMR1*	**A (13)**[4], G (1), - [4](30)
149681465	170	95.9%	C/T	Intronic	*MTMR1*	**C (10)**, G (1), - (33)
149688488	242	96.3%	G/A	Intronic	*CD99L2*	**G (12)**, A (8), C (1), - (23)
149714813	237	99.6%	T/C	Intronic	*CD99L2*	**T (20)**, C (3), - (21)
149726254	147	98.0%	G/A	Intronic	*CD99L2*	**G (7)**, A (1), - (36)
149818442	67	98.5%	C/G	Intergenic (between *CD99L2* and *HMGB3*)	*CD99L2*	**C (4)**, G (14), T (1), - (26)
149819749	331	96.7%	T/-	Intergenic (between *CD99L2* and *HMGB3*)	*CD99L2*	**T (14)**, A (2), C (1), - (27)
152421658	90	96.7%	C/T	Intronic	*BGN*	**C (11)**, T (2), G (1), A (1), - (29)
152426100	159	94.3%	A/G	Intronic	*BGN*	**A (8)**, G (4), C (1), - (31)
152453830	187	93.6%	G/T	Intergenic region (between *ATP2B3* and *BGN*)	*ATP2B3*	**G (13)**, A (3), C (2), -
152569224	67	94.0%	G/A	3'UTR	*DUSP9*	**G (15)**, - (29)
152683343	122	90.2%	C/-	Intronic	*KIAA1206*	**C (6)**, A (5), G (6), T (1), - (26)
152863377[4]	22[4]	86.4%[4]	-/G[4]	5'UTR	*RENBP*	
152890455	102	99.0%	A/G	Intergenic (between *HCFC1* and *TMEM187*)	*HCFC1*	**A (31)**, - (13)
153960377	14	92.9%	A/G	Intronic	*BRCC3*	**A (17)**, G (1), - (26)

[1] All variants are based on positive strands.
[2] Alternative alleles in all available vertebrates were based on the data from UCSC conversation track
[3] Numbers in bracket are counted species with the allele. '-' represents noninformatative allele in the species.
[4] The position chrX: 152863377 '-/G' is a complex variant.

Table S4. List of Primers Used in This Study for Sequencing and qPCR

Primer Name	F/R	Sequence (5' to 3')	Size (bp)	Location	Product Sizes
HCFC1 Total	F	GGCAACGAGGGAATAGTGGAC	21	Exon 1	73 bp
	R	TCACGGCTGGGATGAACC	18	Exon 2	
HCFC1	F	GCTTCTTCTGCCGTACAGTTGC	22	Intergenic	91 bp
S2 Specific	R	CTCACTTCCCGCCTTATGACTC	22	Intergenic	
RENBP	F	TGCGCCTCTTCTACCAAGTG	20	Exon 9	82 bp
	R	CTCAGGTAGCCAAACCATTCC	21	Exon 10	
TMEM187	F	AGGCGAGTGAGGCGAAATAC	20	Exon 1	232 bp
	R	GAGAGCAGCCTTCAATACGC	20	Exon 2	
MECP2	F	GCTTTTCCCTGGGGATTGA	19	Exon 3	120 bp
	R	GGGACCCATGTATGATGACC	20	Exon 3	
Hcfc1	F	GCAACGAGGGGATAGTGGAC	20	Exon 1	74 bp
	R	TCTCACAGCTGGGATGAACC	20	Exon 2	
HCFC1	F	TACGGCAGAAGAAGCGGTA	20	intergenic	199 bp
Sequencing	R	GGGTCCTTTAAGAGGCTTGG	21	Intergenic	

RNA isolation, reverse transcription reactions and qPCR reactions were performed as previously described (Nat. Genet. 39, 1127–1133, 2007).

Table S5. List of Oligonucleotides Used in Electrophoretic Mobility Shift Assays

Binding Site	Genomic Location	Sequence (5'-3')
YY1 Consensus	N/A	CGCTCCCCGGCCATCTTGGCGGCTGGT
S2	chrX:152,890,443-152,890,469	CTGGGAGCCGCCATCTTGTGTGAAGAA
mS2	chrX:152,890,443-152,890,469	CTGGGAGCCGCC**g**TCTTGTGTGAAGAA

Oligonucleotides matching the YY1 consensus binding site (sc-2533, Santa Cruz Biotech), the S2 binding site in *HCFC1* (152,890,443-152,890,469), or the variant mS2 binding site of affected individuals in the MRX3 family. Variant nucleotide is in lowercase bold font in mS2.

Table S6. **Primers for ChIP-PCR Amplifications**

Target Region	**F/R**	**Sequence (5' to 3')**	**Size (bp)**
S2	F	TACGGCAGAAGAAGCGGTAAC	199
	R	GGGTCCTTTAAGAGGCTTGG	
S6	F	TCGTCTAAGGCAGCTCTCACGGAGAAG	169
	R	GCTTCTTGAGCCTAGGCGCTCGACAGT	
-C	F	AACTTTCCCCAGTTCAGTGC	103
	R	TGGGAACGACTTAGCATGAG	

PCR amplifications were performed using the KAPA HiFi PCR Kit (Kapa Biosystems) under the following conditions: S2 and S6: 98 °C for 3 min; 31 cycles of 98 °C for 15 s, 66 °C for 15 s, and 72 °C for 30 s and *HUWE1* negative control: 98 °C for 3 min; 31 cycles of 98 °C for 15 s, 60 °C for 15 s, and 72 °C for 30 s.

Table S7. List of Primers for the PCR Amplification of Genomic DNA Fragments Used in Luciferase Reporter Constructs

Genomic fragments	F/R	Sequence (5' to 3')	Size (bp)	Retriction enzyme
1 205 bp	F	GG**gctagc**AGTCTCATCCGGGCATTTC[a]	27	*Nhe*I
	R	CGGAGA**aagctt**TCTCACACAGCGGTAGACG	31	*Hind*III
705 bp	F	GG**gctagc**AGTCTCATCCGGGCATTTC	27	*Nhe*I
	R	GGC**aagctt**CGTACGGCTTGTGAAGTCTCG	30	*Hind*III
R-705 bp	F	CTGAGC**aagctt**GCAGTCTCATCCGGGCATTTC	33	*Hind*III
	R	AGGGGGC**gctagc**CGTACGGCTTGTGAAGTC	31	*Nhe*I

[a]Bold lowercase indicates the restriction enzyme site.

Table S8. Functional Annotation of 218 Differentially Expressed Genes

Category	Term	Count	%	Fold Enrichment	P-Value	Benjamini
SP_PIR_KEYWORDS	acetylation	48	23.3	1.8	8.10E-05	2.10E-02
SP_PIR_KEYWORDS	mitochondrion	21	10.2	2.5	3.60E-04	4.70E-02
UCSC_TFBS	NCX	100	48.5	1.3	4.00E-04	6.70E-02
GOTERM_BP_FAT	regulation of synaptic transmission, GABAergic	4	1.9	21.5	7.60E-04	6.60E-01
GOTERM_BP_FAT	intracellular transport	18	8.7	2.5	7.80E-04	4.20E-01
INTERPRO	Pyridoxal phosphate-dependent transferase, major region, subdomain 1	5	2.4	11.2	1.00E-03	3.60E-01
GOTERM_CC_FAT	mitochondrion	24	11.7	2	1.20E-03	2.60E-01
GOTERM_MF_FAT	poly(U) RNA binding	3	1.5	43	1.90E-03	5.20E-01
UCSC_TFBS	GATA2	77	37.4	1.3	2.40E-03	1.90E-01
INTERPRO	Aminotransferase, class V/Cysteine desulfurase	3	1.5	39.2	2.40E-03	4.10E-01
UCSC_TFBS	SPZ1	72	35	1.4	3.10E-03	1.60E-01
GOTERM_BP_FAT	translation	11	5.3	3	3.30E-03	7.90E-01
UP_SEQ_FEATURE	transit peptide:Mitochondrion	13	6.3	2.7	3.40E-03	9.00E-01
GOTERM_MF_FAT	vitamin B6 binding	5	2.4	8	3.50E-03	4.80E-01
GOTERM_MF_FAT	pyridoxal phosphate binding	5	2.4	8	3.50E-03	4.80E-01
GOTERM_MF_FAT	poly-pyrimidine tract binding	3	1.5	32.2	3.50E-03	3.60E-01
BLOCKS	IPB004839	3	1.5	32.4	3.50E-03	4.00E-01
OMIM_DISEASE	Multiple Genetic Loci for Bone Mineral Density and Fractures	3	1.5	31.4	3.60E-03	2.00E-01
SP_PIR_KEYWORDS	pyridoxal phosphate	5	2.4	7.8	3.70E-03	2.90E-01
SP_PIR_KEYWORDS	transit peptide	13	6.3	2.7	3.80E-03	2.20E-01
GOTERM_MF_FAT	vitamin binding	7	3.4	4.6	4.00E-03	3.10E-01
SP_PIR_KEYWORDS	transferase	26	12.6	1.8	4.40E-03	2.10E-01
GOTERM_MF_FAT	aminopeptidase activity	4	1.9	11.5	4.90E-03	3.10E-01
GOTERM_MF_FAT	transferase activity, transferring nitrogenous groups	4	1.9	11.5	4.90E-03	3.10E-01
GOTERM_BP_FAT	glycoprotein metabolic process	8	3.9	3.6	6.60E-03	9.00E-01
GOTERM_BP_FAT	nucleocytoplasmic transport	7	3.4	4.1	7.10E-03	8.70E-01
SP_PIR_KEYWORDS	alternative splicing	95	46.1	1.2	7.20E-03	2.80E-01
SP_PIR_KEYWORDS	cytoplasm	49	23.8	1.4	7.40E-03	2.50E-01
GOTERM_BP_FAT	nuclear transport	7	3.4	4	7.50E-03	8.30E-01
UP_SEQ_FEATURE	splice variant	95	46.1	1.2	7.70E-03	9.30E-01
GOTERM_BP_FAT	intracellular protein transport	11	5.3	2.7	7.70E-03	7.90E-01
UCSC_TFBS	FOXO4	110	53.4	1.2	8.90E-03	3.30E-01
UCSC_TFBS	AHRARNT	107	51.9	1.2	9.30E-03	2.80E-01
SP_PIR_KEYWORDS	rna-binding	13	6.3	2.3	9.90E-03	2.90E-01
GOTERM_BP_FAT	macromolecular complex subunit organization	16	7.8	2.1	1.00E-02	8.40E-01
GOTERM_MF_FAT	exopeptidase activity	5	2.4	5.6	1.20E-02	5.40E-01
GOTERM_BP_FAT	polysaccharide biosynthetic process	4	1.9	8.1	1.30E-02	8.70E-01
REACTOME_INTERACTION	vesicle-associated membrane protein 8 (endobrevin)	3	1.5	16.1	1.30E-02	1.00E+00
GOTERM_CC_FAT	cytosol	24	11.7	1.7	1.40E-02	7.00E-01
GOTERM_BP_FAT	cellular protein localization	11	5.3	2.4	1.40E-02	8.40E-01
GOTERM_CC_FAT	Golgi apparatus part	9	4.4	2.8	1.40E-02	8.30E-01
GOTERM_BP_FAT	Golgi vesicle transport	6	2.9	4.2	1.40E-02	8.70E-01
INTERPRO	Aminotransferase, class I and II	3	1.5	16.2	1.40E-02	8.80E-01
GOTERM_CC_FAT	plasma membrane part	35	17	1.5	1.50E-02	4.70E-01
GOTERM_BP_FAT	cellular macromolecule localization	11	5.3	2.4	1.50E-02	8.30E-01
GOTERM_CC_FAT	Golgi membrane	7	3.4	3.5	1.50E-02	5.30E-01
GOTERM_MF_FAT	nucleotidyltransferase activity	6	2.9	4.1	1.50E-02	5.60E-01
GOTERM_CC_FAT	mitochondrial ribosome	4	1.9	7.7	1.50E-02	6.10E-01
GOTERM_CC_FAT	organellar ribosome	4	1.9	7.7	1.50E-02	6.10E-01
UCSC_TFBS	GATA1	156	75.7	1.1	1.70E-02	4.00E-01
UCSC_TFBS	TST1	90	43.7	1.2	1.70E-02	3.50E-01
GOTERM_BP_FAT	immune response	15	7.3	2	1.80E-02	8.60E-01
REACTOME_INTERACTION	vesicle-associated membrane protein 7	3	1.5	13.7	1.80E-02	1.00E+00
UCSC_TFBS	SRF	137	66.5	1.1	2.10E-02	3.70E-01
SP_PIR_KEYWORDS	disease mutation	26	12.6	1.6	2.10E-02	4.40E-01
SP_PIR_KEYWORDS	Acyltransferase	6	2.9	3.8	2.10E-02	4.70E-01
UP_SEQ_FEATURE	domain:t-SNARE coiled-coil homology	3	1.5	13.2	2.10E-02	9.90E-01
UCSC_TFBS	CDPCR1	80	38.8	1.2	2.20E-02	3.60E-01
GOTERM_BP_FAT	protein complex biogenesis	12	5.8	2.2	2.20E-02	8.80E-01
GOTERM_BP_FAT	protein complex assembly	12	5.8	2.2	2.20E-02	8.80E-01
GOTERM_BP_FAT	peroxisome localization	2	1	91.4	2.20E-02	8.90E-01
GOTERM_BP_FAT	microtubule-based peroxisome localization	2	1	91.4	2.20E-02	8.90E-01
BIND	5'-3' exoribonuclease 2	3	1.5	12.3	2.30E-02	1.00E+00
UCSC_TFBS	HTF	106	51.5	1.2	2.50E-02	3.60E-01
SP_PIR_KEYWORDS	ribonucleoprotein	8	3.9	2.8	2.50E-02	4.60E-01
MINT	804156:tumor necrosis factor receptor superfamily, member 1A	5	2.4	4.4	2.50E-02	1.00E+00
GOTERM_BP_FAT	post-Golgi vesicle-mediated transport	4	1.9	6.3	2.50E-02	8.90E-01
GOTERM_CC_FAT	mitochondrial part	13	6.3	2	2.60E-02	6.10E-01
GOTERM_BP_FAT	aminoglycan biosynthetic process	3	1.5	11.9	2.60E-02	8.80E-01
SP_PIR_KEYWORDS	phosphoprotein	89	43.2	1.2	2.70E-02	4.60E-01
GOTERM_MF_FAT	translation factor activity, nucleic acid binding	5	2.4	4.4	2.70E-02	7.20E-01
GOTERM_BP_FAT	nuclear export	4	1.9	6.1	2.70E-02	8.90E-01
MINT	805245:FYN oncogene related to SRC, FGR, YES	3	1.5	11.5	2.70E-02	1.00E+00
GOTERM_CC_FAT	trans-Golgi network transport vesicle	3	1.5	11.6	2.70E-02	5.80E-01
GOTERM_BP_FAT	macromolecular complex assembly	14	6.8	1.9	2.90E-02	8.80E-01
GOTERM_BP_FAT	glycoprotein biosynthetic process	6	2.9	3.5	2.90E-02	8.90E-01
GOTERM_BP_FAT	carbohydrate biosynthetic process	5	2.4	4.3	2.90E-02	8.60E-01
SMART	t_SNARE	3	1.5	11.1	2.90E-02	9.50E-01
SP_PIR_KEYWORDS	Aminopeptidase	3	1.5	11.2	2.90E-02	4.50E-01
SP_PIR_KEYWORDS	Signal-anchor	10	4.9	2.3	3.00E-02	4.40E-01
REACTOME_INTERACTION	nerve growth factor (beta polypeptide)	4	1.9	5.6	3.00E-02	1.00E+00
GOTERM_BP_FAT	superoxide metabolic process	3	1.5	11	3.00E-02	8.60E-01
UCSC_TFBS	CDP	117	56.8	1.1	3.10E-02	3.70E-01
UCSC_TFBS	NKX61	87	42.2	1.2	3.10E-02	3.90E-01
UCSC_TFBS	PBX1	112	54.4	1.1	3.20E-02	3.60E-01
UCSC_TFBS	IRF1	54	26.2	1.3	3.30E-02	3.40E-01
GOTERM_MF_FAT	identical protein binding	14	6.8	1.9	3.40E-02	7.70E-01
INTERPRO	Target SNARE coiled-coil region	3	1.5	10.2	3.40E-02	9.80E-01
UCSC_TFBS	CEBPB	108	52.4	1.1	3.50E-02	3.30E-01
UCSC_TFBS	MYB	90	43.7	1.2	3.50E-02	3.40E-01
GOTERM_CC_FAT	perinuclear region of cytoplasm	8	3.9	2.6	3.50E-02	6.00E-01
GOTERM_CC_FAT	mitochondrial matrix	7	3.4	2.9	3.50E-02	6.40E-01
GOTERM_CC_FAT	mitochondrial lumen	7	3.4	2.9	3.50E-02	6.40E-01
BIND	heterogeneous nuclear ribonucleoprotein U (scaffold attachment factor A)	4	1.9	5.5	3.50E-02	1.00E+00
UCSC_TFBS	PAX6	106	51.5	1.1	3.60E-02	3.10E-01
UCSC_TFBS	MEIS1BHOXA9	99	48.1	1.2	3.60E-02	3.00E-01
GOTERM_BP_FAT	activation of protein kinase activity	5	2.4	4	3.60E-02	8.90E-01
GOTERM_CC_FAT	intracellular organelle lumen	28	13.6	1.5	3.70E-02	5.80E-01
GOTERM_BP_FAT	actin cytoskeleton organization	7	3.4	2.8	3.70E-02	8.90E-01
REACTOME_INTERACTION	vesicle-associated membrane protein 2 (synaptobrevin 2)	3	1.5	9.4	3.70E-02	1.00E+00
GOTERM_MF_FAT	RNA binding	15	7.3	1.8	3.80E-02	7.70E-01
GOTERM_BP_FAT	lipid modification	4	1.9	5.3	3.90E-02	9.00E-01
GOTERM_MF_FAT	single-stranded RNA binding	3	1.5	9.6	3.90E-02	7.40E-01
BIND	growth factor receptor-bound protein 2	7	3.4	2.7	4.00E-02	1.00E+00
MINT	802764:HLA-B associated transcript 2	3	1.5	9.3	4.00E-02	1.00E+00
UCSC_TFBS	HFH1	89	43.2	1.2	4.10E-02	3.20E-01
GOTERM_BP_FAT	cellular macromolecular complex subunit organization	9	4.4	2.3	4.10E-02	9.00E-01
PANTHER_FAMILY	PTHR15706:SH3 MULTIPLE DOMAIN	2	1	47.9	4.10E-02	1.00E+00
PANTHER_FAMILY	PTHR11258:(2-5)OLIGOADENYLATE SYNTHETASE	2	1	47.9	4.10E-02	1.00E+00
UCSC_TFBS	YY1	134	65	1.1	4.20E-02	3.10E-01
INTERPRO	Heavy-metal-associated, conserved site	2	1	45.8	4.30E-02	9.80E-01
INTERPRO	2'-5'-oligoadenylate synthetase 1, domain 2/C-terminal	2	1	45.8	4.30E-02	9.80E-01
INTERPRO	2-5-oligoadenylate synthetase, conserved site	2	1	45.8	4.30E-02	9.80E-01
INTERPRO	Heavy metal transport/detoxification protein	2	1	45.8	4.30E-02	9.80E-01
SP_PIR_KEYWORDS	ribosomal protein	6	2.9	3.1	4.40E-02	5.60E-01
GOTERM_BP_FAT	actin filament organization	4	1.9	5.1	4.40E-02	9.00E-01
GOTERM_MF_FAT	structural constituent of ribosome	6	2.9	3.1	4.50E-02	7.70E-01
GOTERM_BP_FAT	positive regulation of defense response	4	1.9	5	4.50E-02	9.00E-01
BLOCKS	IPB006117	2	1	43.1	4.50E-02	9.70E-01
UCSC_TFBS	CART1	93	45.1	1.2	4.60E-02	3.30E-01
UCSC_TFBS	SOX5	94	45.6	1.2	4.70E-02	3.20E-01
GOTERM_CC_FAT	organelle lumen	28	13.6	1.4	4.70E-02	6.10E-01
GOTERM_CC_FAT	Golgi apparatus	16	7.8	1.7	4.70E-02	6.30E-01
MINT	818071:serpin peptidase inhibitor, clade B (ovalbumin), member 9	3	1.5	8.5	4.70E-02	1.00E+00
SP_PIR_KEYWORDS	hydrolase	24	11.7	1.5	4.80E-02	5.60E-01
GOTERM_BP_FAT	actin filament-based process	7	3.4	2.7	4.80E-02	9.10E-01
REACTOME_PATHWAY	REACT_11061:Signalling by NGF	6	2.9	2.9	4.90E-02	8.40E-01

Table S9. List of Genes with Greater Than Two-Fold Upregulation or Downregulation

GeneSymbol	entrez GeneID	nuID	logFC [a]	Aproximate FC [a,b]	t [c]	P.Value	adj.P.Val	B [d]	Control		MRX3	Male Control		Female Control	
									Average	SD	Average	Average	SD	Average	SD
UCHL1	7345	ipJcOd7qKh.nj9V31c	-2.1320	0.2281	-2.5152	0.0273	0.8693	-3.3937	9.0452	1.3196	6.9132	9.7296	1.3235	8.4976	1.1515
LOC644936	644936	KTqOT7xKC1fldcAlF0	-1.7837	0.2904	-2.2008	0.0482	0.8951	-3.7335	11.8459	1.0280	10.0622	12.4354	0.2374	11.3743	1.2024
TUBB2B	347733	ircfyu5nq3eud9Hvl4	-1.3955	0.3801	-1.7390	0.1078	0.9140	-4.2115	11.0573	1.2474	9.6618	10.7713	1.5692	11.2861	1.0574
PSAT1	29968	Ek9TyQi_xVdVLfZfXc	-1.1607	0.4473	-3.0914	0.0094	0.8117	-2.7712	9.6895	0.4692	8.5288	9.4124	0.5517	9.9112	0.2716
SERPINB1	1992	QZcruSjd.BLJCnje3o	-1.1277	0.4576	-2.3850	0.0346	0.8708	-3.5350	8.4800	0.7328	7.3522	8.0230	0.8031	8.8455	0.4630
HLA-K	649853	KGhlVCGTOHRRnt3hTg	2.8822	7.3729	1.8053	0.0964	0.9140	-4.1454	9.4482	2.4895	12.3304	7.9637	2.2864	10.6357	2.1233
RPS4Y1	6192	QEpD_7OuProlBdrbvo	2.8364	7.1423	1.6134	0.1329	0.9196	-4.3334	9.2278	2.7409	12.0642	12.1141	0.1818	6.9188	0.0743
MMP7	4316	rkHTgfpT3_rOoT13jU	1.9585	3.8865	3.1580	0.0083	0.8014	-2.7007	8.0686	0.8387	10.0271	7.5380	0.9769	8.4931	0.4293
EIF1AY	9086	WeKOvcS_o.6U00FCCI	1.9434	3.8461	1.6133	0.1329	0.9196	-4.3335	8.2765	1.8767	10.2199	10.2499	0.2018	6.6978	0.0610
MMP7	4316	3nR9jijgZ6GjrJLcqM	1.8793	3.6791	3.2856	0.0066	0.7949	-2.5669	7.8442	0.7728	9.7236	7.4106	0.9561	8.1911	0.4127
CLLU1OS	574016	rXotrr84O65TeqB4QU	1.6035	3.0387	1.9241	0.0786	0.9140	-4.0245	8.2585	1.2942	9.8619	8.8066	1.4123	7.8200	1.1458
MAL	4118	6dFQSN.UitTrolYwV4	1.4210	2.6777	2.7744	0.0169	0.8425	-3.1119	8.9930	0.7711	10.4140	9.3313	0.9388	8.7223	0.5677
FXYD2	486	T6KSVPReqA4er7Y6Gs	1.4124	2.6617	1.8912	0.0832	0.9140	-4.0584	7.7453	0.9319	9.1576	7.1788	0.6934	8.1984	0.8938
PRRX1	5396	rjl96N8JS_0uitdKnk	1.4004	2.6397	4.6636	0.0006	0.6814	-1.2744	6.9560	0.2485	8.3563	6.8878	0.2091	7.0105	0.2870
IL17RB	55540	QnnX7JUVOIJCIXwKfU	1.3697	2.5841	2.2202	0.0466	0.8951	-3.7127	8.5868	0.8203	9.9564	7.9616	0.7110	9.0869	0.5127
GSTT1	2952	EuJ6gF0V_RbXSS1ECQ	1.3655	2.5766	2.7588	0.0174	0.8425	-3.1288	7.8014	0.7668	9.1669	7.3379	0.9794	8.1722	0.2642
FAM38B	63895	ZgI5e7.Xe6pW5H16Ig	1.2859	2.4384	2.9769	0.0116	0.8359	-2.8935	7.6843	0.6675	8.9702	7.7363	0.6483	7.6427	0.7556
KDM5D	8284	WN_KCVC5sX8._VUUV0	1.2377	2.3583	2.1898	0.0492	0.8951	-3.7452	7.3999	0.8764	8.6377	8.2827	0.4192	6.6937	0.0406
MSX1	4487	QrVSSD4S_BInIpTt14	1.2326	2.3499	2.2357	0.0453	0.8951	-3.6962	8.1052	0.8035	9.3378	7.4586	0.3813	8.6225	0.6554
UGT2B7	7364	WwEX2r5JUYXRevUsUc	1.2223	2.3331	1.4989	0.1600	0.9331	-4.4404	8.4412	1.2559	9.6634	8.7499	1.1540	8.1941	1.4087
TMEM51	55092	QcZ6LXge6sgBGF6f7I	1.2210	2.3310	6.0922	0.0001	0.4734	-0.2708	7.6048	0.2895	8.8257	7.6798	0.2208	7.5447	0.3478
UGT2B17	7367	uSUBX5cKA6CL_5Ineg	1.2192	2.3282	2.0890	0.0589	0.9061	-3.8525	7.8879	0.9027	9.1071	8.1341	0.9415	7.6909	0.9250
UGT2B7	7364	WicBF9q_SVGF0Xr1LE	1.2038	2.3034	1.5072	0.1578	0.9331	-4.4328	8.3383	1.2311	9.5421	8.5968	1.0993	8.1315	1.4159
ASCL1	429	BddMriNCKUUvsfSRQ4	1.2005	2.2981	2.0356	0.0647	0.9101	-3.9086	7.7095	0.6469	8.9100	7.8551	0.7657	7.5931	0.5991
HOXB2	3212	ZpbcegA2Jd1RXS2TIc	1.1934	2.2869	4.8795	0.0004	0.6814	-1.1011	7.1833	0.3724	8.3767	7.0380	0.4555	7.2996	0.2894
CXCR4	7852	ryQLhlipeLnUslFk3o	1.1709	2.2515	2.1309	0.0547	0.9061	-3.8080	8.1929	0.7911	9.3637	8.1554	0.9579	8.2228	0.7490
OAS1	4938	HdefXV5T01eVd1OCfg	1.1060	2.1525	2.0429	0.0639	0.9070	-3.9009	8.8313	0.8309	9.9373	8.2417	0.7590	9.3030	0.5681
OXTR	5021	iKl6Ke4iOKThxwq.qs	1.0942	2.1349	3.3755	0.0056	0.7804	-2.4737	8.5197	0.4855	9.6139	8.2223	0.5586	8.7576	0.2797
CR2	1380	BQof3gl0fOI5eCUpTo	1.0772	2.1099	3.3615	0.0057	0.7804	-2.4882	9.0340	0.4932	10.1112	9.3255	0.4454	8.8008	0.4299
TGM5	9333	Ed2iSvih7ueHugoLpc	1.0748	2.1064	3.8937	0.0022	0.7414	-1.9587	7.5378	0.4227	8.6126	7.6512	0.4169	7.4472	0.4515
B3GNT2	10678	9pXinsUzR4UrU3cVT4	1.0621	2.0879	3.5673	0.0039	0.7620	-2.2784	8.4860	0.4337	9.5480	8.6191	0.3629	8.3795	0.4955
MAL	4118	xH4ngBUjrvB6WVR.Wk	1.0587	2.0830	2.6838	0.0200	0.8425	-3.2102	7.7242	0.5890	8.7828	7.9627	0.7146	7.5334	0.4565
HAVCR2	84868	fhrMp_N7fnz75VeZck	1.0419	2.0589	3.9573	0.0019	0.7414	-1.8983	8.6321	0.3865	9.6739	8.8645	0.3823	8.4461	0.3032
CKLF	51192	rSWAMATZV959LuApRs	1.0187	2.0261	3.5496	0.0040	0.7620	-2.2963	8.5509	0.4126	9.5696	8.6313	0.5485	8.4865	0.3212
S100A4	6275	cNeHntOm5XeiClejuM	1.0112	2.0156	2.0324	0.0650	0.9101	-3.9120	9.2362	0.6767	10.2475	9.0621	0.9724	9.3755	0.3899

[a]FC: fold change
[b]relative to expression in controls
[c]moderated t-statistic
[d]B statistic

Table S10. Functional Annotation of 35 Two-Fold Upregulated or Downregulated Genes

Category	Term	Count	%	Fold Enrichment	P-Value	Benjamini
GOTERM_BP_FAT	neuron differentiation	5	15.6	6.4	5.90E-03	9.50E-01
GOTERM_BP_FAT	cell morphogenesis involved in differentiation	4	12.5	9.2	7.90E-03	8.70E-01
KEGG_PATHWAY	Metabolism of xenobiotics by cytochrome P450	3	9.4	18.2	9.80E-03	2.70E-01
KEGG_PATHWAY	Drug metabolism	3	9.4	17.6	1.00E-02	1.50E-01
SP_PIR_KEYWORDS	microsome	3	9.4	18.1	1.10E-02	7.10E-01
SP_PIR_KEYWORDS	transferase	7	21.9	3.2	1.60E-02	5.90E-01
PANTHER_FAMILY	PTHR11926~GLUCOSYL/GLUCURONOSYL TRANSFERASES	2	6.2	103.1	1.80E-02	3.50E-01
GOTERM_BP_FAT	neuron development	4	12.5	6.7	1.90E-02	9.60E-01
INTERPRO	UDP-glucuronosyl/UDP-glucosyltransferase	2	6.2	94.9	2.00E-02	8.10E-01
GOTERM_BP_FAT	cell morphogenesis	4	12.5	6.3	2.20E-02	9.40E-01
PANTHER_FAMILY	PTHR19418~HOMEOBOX PROTEIN	3	9.4	12	2.40E-02	2.40E-01
GOTERM_BP_FAT	forebrain development	3	9.4	11.1	2.70E-02	9.40E-01
BLOCKS	IPB002213	2	6.2	69.4	2.70E-02	5.00E-01
GOTERM_BP_FAT	cellular component morphogenesis	4	12.5	5.7	2.90E-02	9.20E-01
PIR_SUPERFAMILY	PIRSF005678:glucuronosyltransferase	2	6.2	63.2	2.90E-02	5.00E-01
SMART	HOX	3	9.4	9.7	3.10E-02	3.60E-01
UP_SEQ_FEATURE	DNA-binding region:Homeobox	3	9.4	10.1	3.30E-02	9.70E-01
GOTERM_BP_FAT	cell proliferation	4	12.5	5.2	3.70E-02	9.40E-01
GOTERM_BP_FAT	eating behavior	2	6.2	49	3.80E-02	9.20E-01
UP_SEQ_FEATURE	domain:MARVEL	2	6.2	49	3.90E-02	8.70E-01
SP_PIR_KEYWORDS	glycosyltransferase	3	9.4	9	4.10E-02	7.90E-01
GOTERM_MF_FAT	glucuronosyltransferase activity	2	6.2	45.4	4.20E-02	9.90E-01
GOTERM_BP_FAT	axonogenesis	3	9.4	8.8	4.20E-02	9.10E-01
KEGG_PATHWAY	Ascorbate and aldarate metabolism	2	6.2	42.7	4.30E-02	3.70E-01
GOTERM_BP_FAT	behavior	4	12.5	4.8	4.40E-02	9.00E-01
KEGG_PATHWAY	Pentose and glucuronate interconversions	2	6.2	40.4	4.50E-02	3.10E-01
GOTERM_BP_FAT	cell motion	4	12.5	4.7	4.50E-02	8.90E-01
GOTERM_CC_FAT	integral to membrane	13	40.6	1.6	4.70E-02	9.60E-01
UP_SEQ_FEATURE	transmembrane region	13	40.6	1.7	4.70E-02	8.10E-01
GOTERM_CC_FAT	microsome	3	9.4	8.1	4.80E-02	8.00E-01
GOTERM_BP_FAT	cell morphogenesis involved in neuron differentiation	3	9.4	8.1	4.90E-02	8.80E-01
SP_PIR_KEYWORDS	transmembrane	13	40.6	1.7	4.90E-02	7.50E-01

Table S11. Clinical Description of Individuals with Missense Variation in *HCFC1*

FAMILY	VARIANT	CLINICAL INFORMATION
D144	X:1528881908C>T; p.S225N	The index individual was born after an uneventful pregnancy at term to healthy and unrelated parents. He had mild intellectual disability and attended a school for special needs. On examination at the age of 17 years, he had behavioral problems including anxiety disorder and a tendency to compulsive behavior, but was otherwise healthy. His younger brother also had mild intellectual disability. At the age of 6 years, high-grade hyperopia was detected, and he attended a school for children with visual problems. On examination at the age of 9 years, deficits in fine motor skills were more pronounced than his verbal deficiencies. Body measurements were normal, and he had no dysmorphic signs or other additional health problems (see **Supplementary Fig. 11**)
D82	X:152870346G>A; p.A1756V	The index individual was born after an uneventful pregnancy to healthy and non-consanguineous parents. Psychomotor development was delayed. On examination at the age of 11 years, he had mild intellectual disability. Body measurements were normal, and, apart from asthma, he had no other health problems. Clinical information for other affected of the family was unavailable.
D147	X:152875279C>T; p.G876S	The individual was born to non-consanguineous healthy Turkish parents after an uneventful pregnancy at 37 weeks of gestation. His birth measurements were normal (weight: 3200 g, length: 49 cm, OFC: 35 cm). Breast feeding was not possible due to muscular hypotonia. He learned to sit at the age of 10 months and was able to walk without support at the age of 2 years. He spoke his first words at the same age. His childhood was complicated by frequent infections. An inguinal hernia and a cryptorchidism were surgically corrected. Seizures were successfully treated with carbamazepine. He visited a school for the mentally disabled. At the age of 8 9/12 years he had normal measurements for height and weight [height: 130 cm (10-25th centile; weight: 36 kg (90-97th centile)], but he presented with macrocephaly [OFC: 56 cm (>97th centile)]. He presented with macrostomia with a full and everted lower lip, a low-set posterior hairline, large and simple modelled ears, brachytelephalangy and hypertrichosis.