Supplementary Table 1 - List of XLID cases analyzed in this study

Case#	Array	Gender	Family	Known Mutations	Sanger sequencing
ID1	AVG_Signal.1825371092_E	M			
ID2	AVG_Signal.1699538165_E	M			
ID3	AVG_Signal.1699538165_C	M			
ID4	AVG_Signal.1699538165_G	M			
ID5	AVG_Signal.1825371092_D	M		JARID1C	SS
ID6	AVG_Signal.1825371054_G	M			
ID7	AVG_Signal.1699538165_F	M			
ID8	AVG_Signal.1825371092_G	M			
ID9	AVG_Signal.1825371092_C	M			
ID10	AVG_Signal.1740115075_G	M			
ID11	AVG_Signal.1825371092_F	M		CCDC22 mutation identified in this study	SS
ID12	AVG_Signal.1699538152_B	M			
ID13	AVG_Signal.1825371054_D	M	PQBP1-F1(ID13,ID14,ID15). ID13-unaffected father		
ID14	AVG_Signal.1825371054_C	M	PQBP1-F1(ID13,ID14,ID15). ID13-unaffected father		
ID15	AVG_Signal.1825371054_H	M	PQBP1-F1(ID13,ID14,ID15). ID13-unaffected father		
ID16	AVG_Signal.1825371082_A	M	ID16,ID17		
ID17	AVG_Signal.1825371082_B	M	ID16,ID17		
ID18	AVG_Signal.1825371092_B	M		ND	SS
ID19	AVG_Signal.1699538152_D	M		HUWE1 duplication	
ID20	AVG_Signal.1825371093_E	M	ID20,ID21		
ID21	AVG_Signal.1825371093_F	M	ID20,ID21		
ID22	AVG_Signal.1699538165_D	M			
ID23	AVG_Signal.1825371093_A	M			
ID24	AVG_Signal.1740115075_H	M			
ID25	AVG_Signal.1825371082_E	M		CASK	SS
ID26	AVG_Signal.1825371092_H	M			
ID27	AVG_Signal.1740115074_H	M		ND	SS
ID28	AVG_Signal.1825371082_D	M			
ID29	AVG_Signal.1740115075_B	M			
ID30	AVG_Signal.1740115074_D	M			
ID31	AVG_Signal.1825371092_A	M		ND	SS

ID32	AVG_Signal.1740115074_B	M			
ID33	AVG_Signal.1740115074_C	M			
ID34	AVG_Signal.1699538152_C	M		ND, CCDC22 c.715G>A, p.E239K	SS
ID35	AVG_Signal.1740115074_E	M		ND	SS
ID36	AVG_Signal.1699538165_A	M		ND	SS
ID37	AVG_Signal.1740115075_C	M			
ID38	AVG_Signal.1740115075_D	M		ND	SS
ID39	AVG_Signal.1699538152_F	M			
ID40	AVG_Signal.1740115074_A	M			
ID41	AVG_Signal.1740115075_A	M		HUWE1 dup	
ID42	AVG_Signal.1825371093_B	M			
ID43	AVG_Signal.1825371093_C	M		ND	SS
ID44	AVG_Signal.1699538152_H	M			
ID45	AVG_Signal.1740115075_E	M		IL1RAPL1 insA	SS
ID46	AVG_Signal.1825371082_F	M			
ID47	AVG_Signal.1699538165_B	M			
ID48	AVG_Signal.1699538165_H	M		ND	SS
ID49	AVG_Signal.1699538152_G	M	ID49,ID50		
ID49	AVG_Signal.1825371082_H	M	ID49,ID50		TR
ID50	AVG_Signal.1740115074_G	M	ID49,ID50		
ID51	AVG_Signal.1825371082_G	M		PHF6	

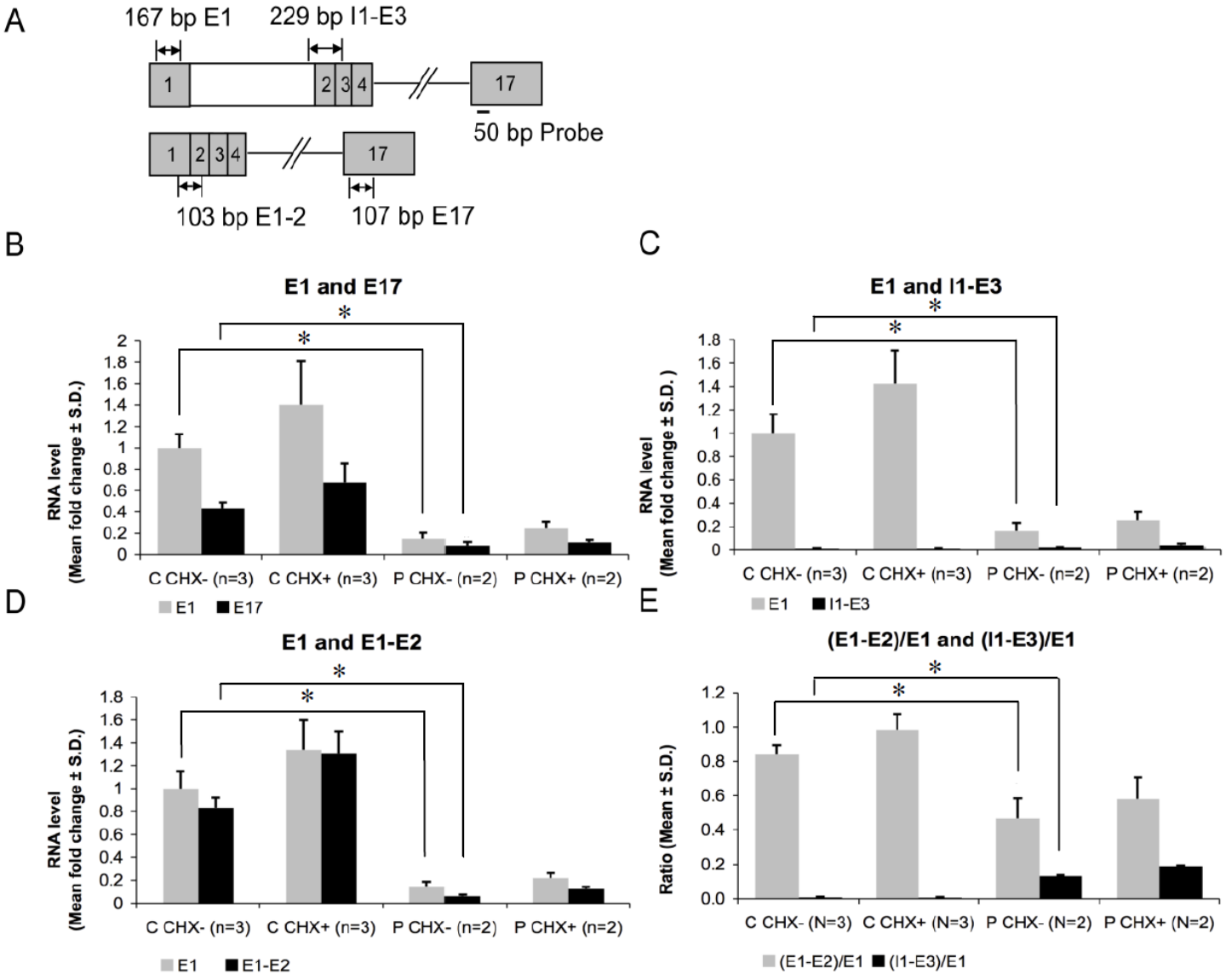
The table lists all ID cases for which the RNA and microarray data passed the quality control criteria. The array name identifies the corresponding expression data in the raw data file. For cases where more than one family member was included in the study, all family members are listed in the fourth column. SS- Proband Sanger sequenced by Tarpey et al (ref#5), ND- No obvious disease-causing variant identified yet by Sanger sequencing, TR- technical replicate.

Supplementary Table 2. Phenotype characterization of PQBP1 mutations described in this study	
Family/Case	**Phenotype**
PQBP1-F1[a] (ID14,ID15)	Moderate mental retardation (3/3), normal head circumference (3/3), short stature (2/3), submucous cleft palate (1/3), velopharyngeal incompetence (1/3), hypospadias and unilateral renal agenesis (1/3). Facial features (3/3): high nasal bridge, arched eyebrows, upslanting palpebral fissures, hypertelorism, long palpebral fissures, short philtrum, small mouth and narrow chin.
ID30	Mild to moderate mental retardation, short stature, microcephaly and brachycephaly with a sloping forehead, slightly upslanting palpebral fissures, high nasal bridge and bulbous nasal tip, prominent ears, long fingers and short bulbous great toes.
[a-] The PQBP1-F1 family consists of three affected brothers, two of which were included in the expression profiling data set (ID14, ID15). All three brothers carry the (AG)2 deletion. The frequency of specific clinical features is given between brackets	

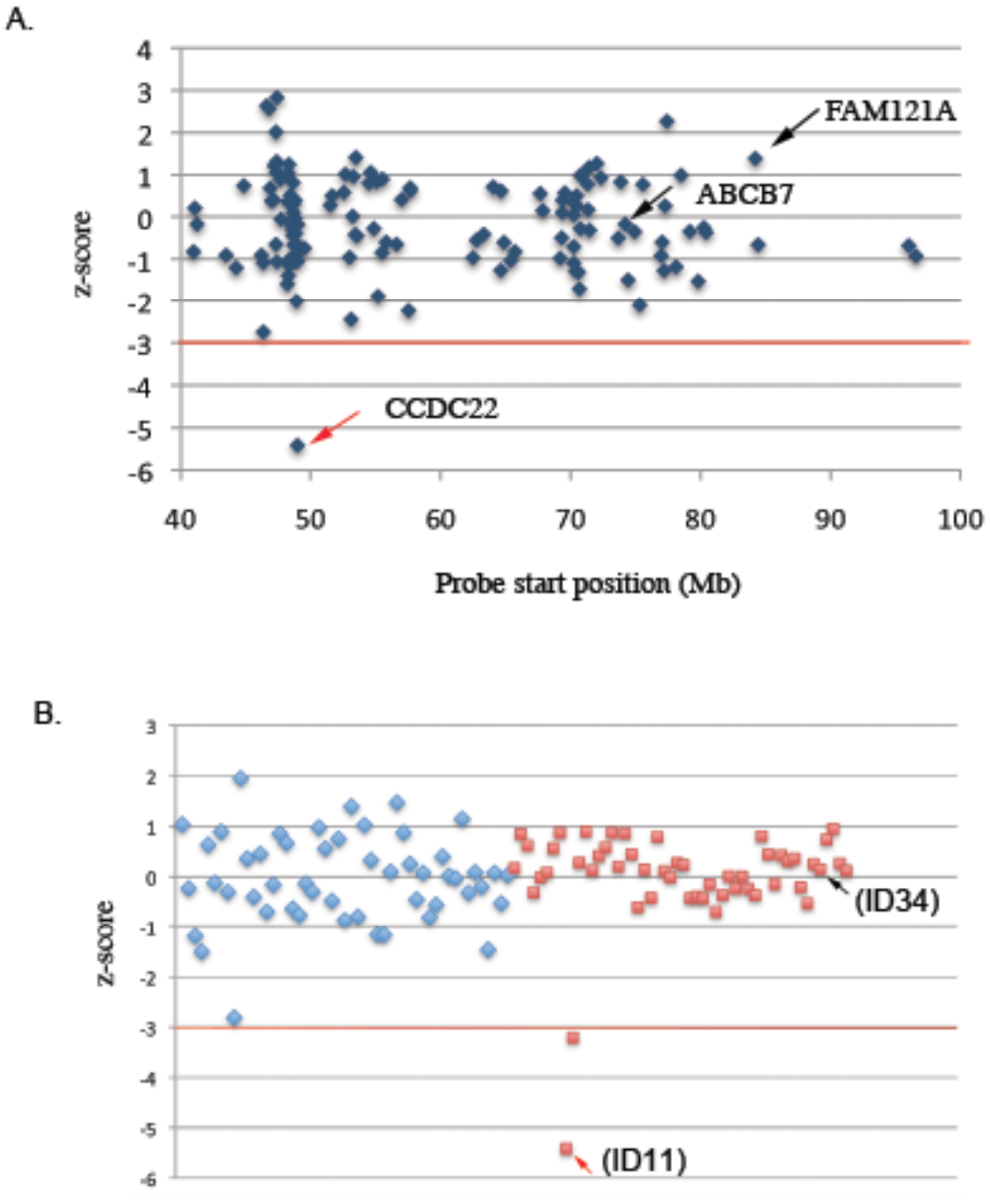
Supplementary Table 3. Primers for *CCDC22 and PQBP1*					
Primer Name	**F/R**	**Sequence**	**Size (bp)**	**Location**	**Product Sizes**
***CCDC22* gDNA sequencing**					
P-I	F	CAGCAGCTATTGCAAGCTCAAC	22	Promoter	373 bp
	R	GAGCCCTGAATGTCAGGATC	20	Intron 1	
***CCDC22* RT-qPCR**					
E1-E4	F	ACTTTCCAACTCTCCCCACAC	21	5'UTR	515 bp
	R	CCGGAGGAGAATAGCTGAGT	20	Exon 4	
Long I1-E3	F	CACATCCGGGACTCTAAAGC	20	Intron 1[a]	1 646 bp[b]
	R	AGACGCTCAGCCAAGAAGAG	20	Exon 3	
E1-E2	F	GGACCGAATCCTCATCCATTC	21	Exon 1	103 bp
	R	ACAGCCTCTACAACCAGCTCA	21	Exon 2	
I1-E3	F	ATCAAGCTGGTCCCCTTCTT	20	Intron 1[c]	229 bp[d]
	R	AGCCAAGCTCCAAGGGATAG	20	Exon 3	
E1	F	ACTTTCCAACTCTCCCCACAC	21	5'UTR	167 bp
	R	CAGCGAATGGATGAGGATTC	20	Exon 1	
E17	F	CTCGAGGAGCAGATCGAGAC	20	Exon 17	105 bp
	R	TAGGAGGCCAGCGTTCTCC	19	Exon 17	
***PQBP1* RT-qPCR**					
PQBP1-F	F	GAGAGAGCGAGACAGGGAAC			
PQBP1-R	R	TTTCGGCTTACTGCCTTCTT			

[a] 33 bp apart from exon 1; [b] 6 397bp if amplified from gDNA; [c] 3 bp apart from exon 2; [d] 4 980 bp if amplified from gDNA. F, forward primer; R, reverse primer

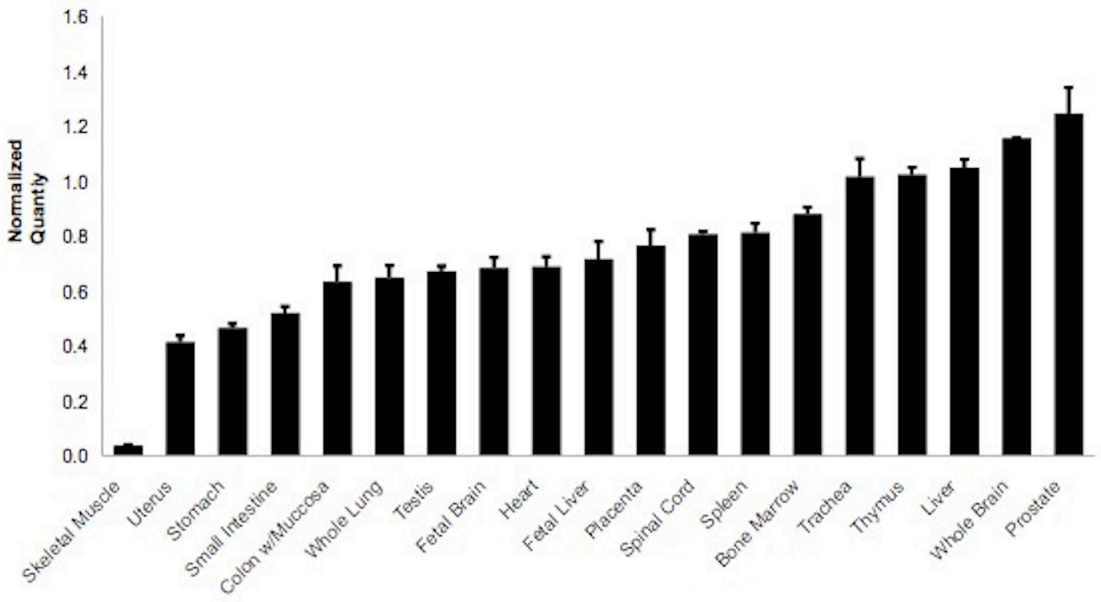
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3.



SUPPLEMENTARY METHODS

Patients

All of the cases used in this study (Supplementary Table 1) have been diagnosed with intellectual disability and had a pedigree consistent with X-linked inheritance. The majority of cases were probands from unrelated families. In addition, four families with more than one affected member were included in the study. A few cases had known causal mutations (Supplementary Table 1) none of which caused mRNA downregulation of the mutated gene.

The controls used in this study were unaffected siblings of autism patients from the AGRE cohort (http://research.agre.org/). Genome-wide expression profiles from controls LCLs were obtained as part of an independent study.

Microarrays

Total RNA was extracted from LCLs using the Qiagen miRNeasy kit, and RNA quality was assessed by Agilent Bioanalyzer. Genome-wide expression profiles were obtained on Illumina Ref8 v2 microarrays (XLID cases) and Illumina Ref8 v3 microarrays (controls). cDNA labeling and array hybridizations were performed following the manufacturer's protocol at the UCLA microarray core facility (http://microarray.genetics.ucla.edu/xowiki/).

Microarray data analysis was performed using the R software and the Bioconductor "Lumi" package(1). Raw data was log2 transformed and normalized by quantile normalization. Batch effects were adjusted using the ComBat package(2). Microarray data quality control criteria included high inter-array Pearson correlation

coefficients (PCC>0.9), low variance of mean inter-array correlation and probe detection P values <0.05 in at least 50% on the samples. 52 of the 64 RNA samples analyzed passed all the RNA and data quality control criteria and were retained for further analysis (Supplementary Table 1).

Genes with expression levels outside the 99.7% confidence interval (3 standard deviations below the mean) in a single case, or outside the 95% confidence interval (2 standard deviations below the mean) in two cases from the same family were considered significantly downregulated and followed up by RT-qPCR.

RT-PCR and RT-qPCR

Total RNA for RT-PCR and RT-qPCR was extracted from LCLs with Trizol Reagent (Invitrogen) combined with RNeasy mini kit (Qiagen) and treated with DNase I (Qiagen). 2 μg of RNA with 1 μg of random hexanucleotide primers (Invitrogen) were annealed at 65ºC and then incubated on ice for at least 2 minutes. Reverse transcriptional reagents (Invitrogen, SuperScript III Reverse Transcriptase Kit) are incubated at 25°C for 5mins and then 50° C for 60 mins Superscript III (Invitrogen). The efficiency of the reaction was tested by PCR using primers specific to the ubiquitously expressed ACTB gene. cDNAs were amplified with Taq DNA polymerase (Roche) and specific single-stranded DNA primers (35 cycles of denaturation, 94°C for 30 s; annealing for 30 s (specific Tm for each pair of primers, see Supplementary Table 2); extension, 72 °C for 30 s,). PCR products were analyzed by 1% agarose gel stained with Ethidium Bromide. Real-time PCR was performed using StepOnePlus Real-time PCR system (Applied Biosystems, Foster City, CA.). RT-qPCR reactions were performed in 20 microliters

volume containing iTaq Sybrgreen (Biorad) and primers at a concentration of 0.5μM each. Primers used in RT-PCR and RT-qPCR were designed by Primer 3 and specificity was checked by NCBI BLAST (Supplementary Table 2).

CHX Treatment

Approximately 3×10^6 cells were incubated in RPMI with 10% FCS and 100 μg ml^{-1} cycloheximide (Sigma) for 6 hrs. All samples were treated in triplicate. Before RNA extraction the cells were harvested by centrifugation and washed in PBS(3).

Statistical Analysis

P value was calculated using Student's paired, two-tailed t-test when comparing the difference between CHX treatment groups. When comparing the difference between controls and patients, P value was calculated using Student's unpaired, two-tailed t-test.

Western Blot

LCL cell pellets were lysed in ice-cold radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% Nonidet P-40, 0.5% sodium deoxycholate, 0.1% SDS and protease inhibitors) and incubated on ice for 30 min. Debris was removed by high-speed centrifugation for 10 min at 4°C. Cleared lysates were quantitated by Bradford Assay (Bio-Rad). Approximately equal quantity of proteins were separated on a NuPAGE Novex 4-12% Bis-Tris Gel (Invitrogen, 1.0 mm, 12 wells. Cat.No.NP0322BOX) and then transferred onto nitrocellulose membranes by semidry electroblotting using a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad). Membranes were blocked with 5% skim milk and 1% goat serum and then incubated with appropriate primary antibody (Sigma-Aldrich Anti-CCDC22 antibody, HPA000888) and subsequently with secondary HRP-conjugated antibody (Goat anti-Rabbit HRP-

conjugated antibody). The blots were developed using the enhanced chemiluminescence (ECL) method (Amersham). Densitometry was performed on low-exposured film to determine the protein level using GeneTools analysis software (Syngene software).

mTOM The mTOM analysis was performed using the available software (http://www.genetics.ucla.edu/labs/horvath/MTOM/) with the cortex expression data from Johnson et al.(4) as input. The default soft-threshold power of 6 was used for calculating the adjacency matrix, and the "signed" option was used in order to take into account the directionality of expression changes. The multi-node topological overlap module was built in a non-recursive manner, to limit the detection of co-expressed genes to those most specific to CCDC22.

Gene ontology analysis was performed using the online DAVID tool (http://david.abcc.ncifcrf.gov/) The p values reported are corrected for multiple comparisons (Benjamini and Hochberg).

1. Du P, Kibbe WA, Lin SM. *Bioinformatics* 2008 Jul 1; **24**(13)**:** 1547-1548.
2. Johnson WE, Li C, Rabinovic A. *Biostatistics* 2007 Jan; **8**(1)**:** 118-127.
3. Tarpey PS, Raymond FL, Nguyen LS, Rodriguez J, Hackett A, Vandeleur L *et al. Nat Genet* 2007 Sep; **39**(9)**:** 1127-1133.
4. Johnson MB, Kawasawa YI, Mason CE, Krsnik Z, Coppola G, Bogdanovic D *et al. Neuron* 2009 May 28; **62**(4)**:** 494-509.