The American Journal of Human Genetics, Volume *107*

Supplemental Data

De Novo Variants in *CNOT1*, a Central Component of the CCR4-NOT Complex Involved in Gene Expression and RNA and Protein Stability, Cause Neurodevelopmental Delay

Lisenka E.L.M. Vissers, Sreehari Kalvakuri, Elke de Boer, Sinje Geuer, Machteld Oud, Inge van Outersterp, Michael Kwint, Melde Witmond, Simone Kersten, Daniel L. Polla, Dilys Weijers, Amber Begtrup, Kirsty McWalter, Anna Ruiz, Elisabeth Gabau, Jenny E.V. Morton, Christopher Griffith, Karin Weiss, Candace Gamble, James Bartley, Hilary J. Vernon, Kendra Brunet, Claudia Ruivenkamp, Sarina G. Kant, Paul Kruszka, Austin Larson, Alexandra Afenjar, Thierry Billette de Villemeur, Kimberly Nugent, the DDD Study, F. Lucy Raymond, Hanka Venselaar, Florence Demurger, Claudia Soler-Alfonso, Dong Li, Elizabeth Bhoj, Ian Hayes, Nina Powell Hamilton, Ayesha Ahmad, Rachel Fisher, Myrthe van den Born, Marjolaine Willems, Arthur Sorlin, Julian Delanne, Sebastien Moutton, Philippe Christophe, Frederic Tran Mau-Them, Antonio Vitobello, Himanshu Goel, Lauren Massingham, Chanika Phornphutkul, Jennifer Schwab, Boris Keren, Perrine Charles, Maaike Vreeburg, Lenika De Simone, George Hoganson, Maria Iascone, Donatella Milani, Lucie Evenepoel, Nicole Revencu, D. Isum Ward, Kaitlyn Burns, Ian Krantz, Sarah E. Raible, Jill R. Murrell, Kathleen Wood, Megan T. Cho, Hans van Bokhoven, Maximilian Muenke, Tjitske Kleefstra, Rolf Bodmer, and Arjan P.M. de Brouwer

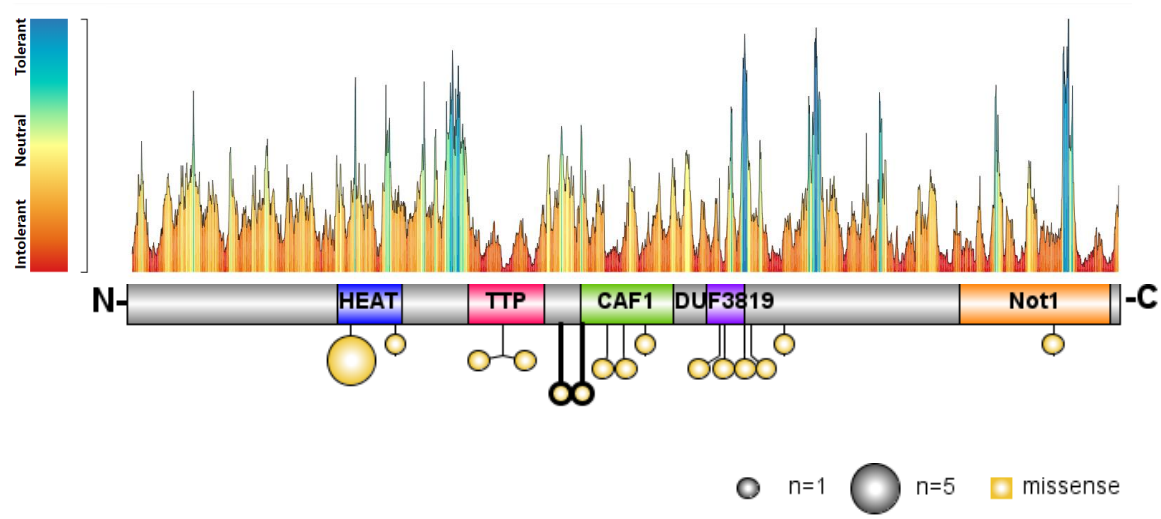
Supplemental Information

De novo variants in *CNOT1*, a central component of the CCR4-NOT complex involved in gene expression and RNA and protein stability, cause neurodevelopmental delay

Supplemental Figures

Supplemental Figure 1: Graphical representation of the CNOT1 missense variants in relation to the CNOT1 tolerance landscape

A

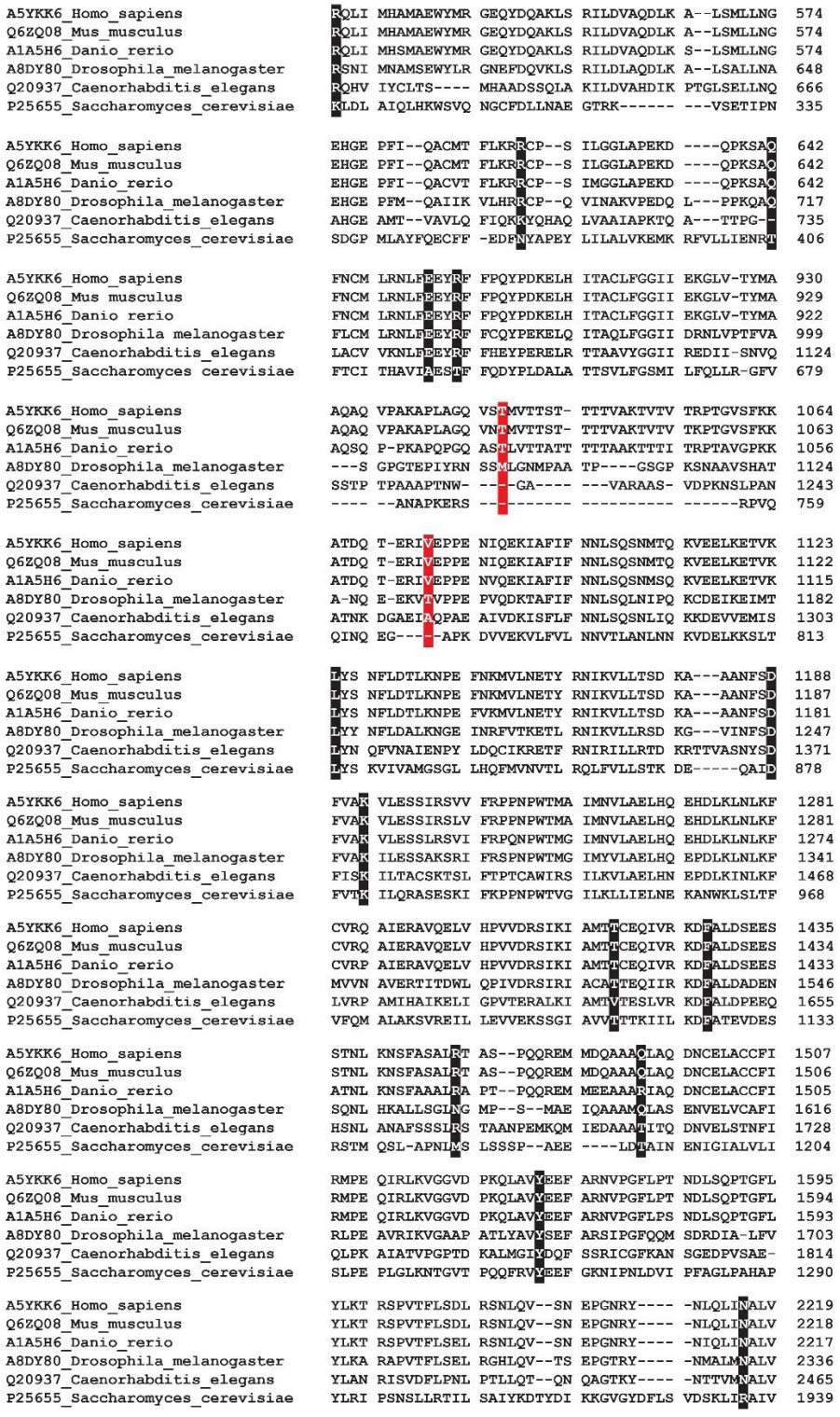


(**A**) Graphical representation of the linear protein structure of CNOT1 with the functional domains indicated by coloured boxes (blue: HEAT; pink: TTP; green: CAF; purple: DUF3819 and orange: Not1). Missense variants observed in this study are indicated by circles, in which the size of the circle corresponding to the number of recurrences of a given variant, and the colour of the circle the type of mutation. Note that Individual 28 was shown to have two *de novo* missense mutations, which are here shown by a bold outline. The tolerance landscape of CNOT1 was generated using Metadome (https://stuart.radboudumc.nl/metadome)[1], which is based on the non-synonymous over synonymous ratio in gnomAD. The colour in the plot is an indication for the tolerance (red: intolerant, blue: tolerant). The linear protein structure of CNOT1 with the missense variants is depicted underneath the tolerance profile. Together these indicate that the missense variants identified in CNOT1 are located in regions that are mostly intolerable for variation.

Supplemental Figure 2: Photos of hands and feet of Individuals with (*de novo*) *CNOT1* variants

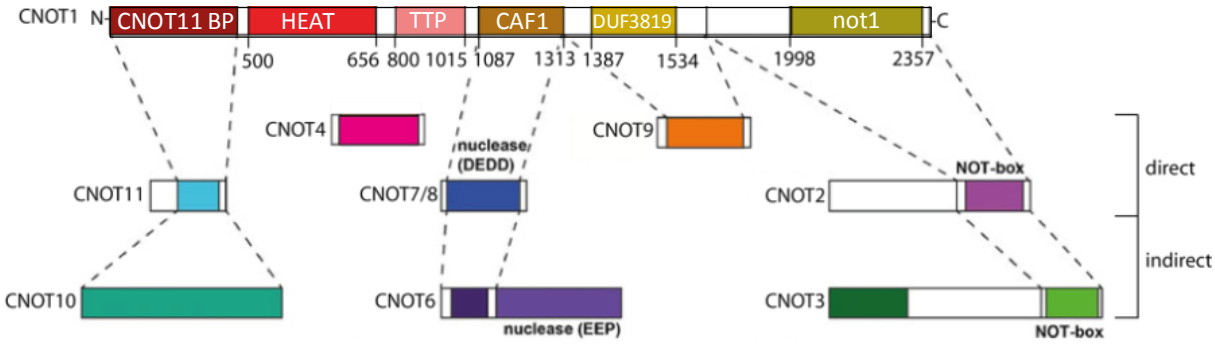
Hands and feet of Individuals with (likely) pathogenic *CNOT1* variants are shown, with individuals grouped based on the type of mutation identified. A variety of subtle abnormalities of hands and feet can be observed in the cohort, e.g. clinodactyly (Individual 7, 12, 18, 29, 39), tapered fingers (Individual 1, 19, 30), small hands (Individual 6, 7, 12 18, 19), broad forefeet with short broad toes (Individual 7, 12, 30). The subtle abnormalities were detected in both individuals with a presumed loss-of-function variant as well as in individuals with a missense variant. The green line below photos of individuals missense variants indicates that variants were located in the CAF domain.

Supplemental Figure 3: Evolutionary conservation of missense variants observed in CNOT1.



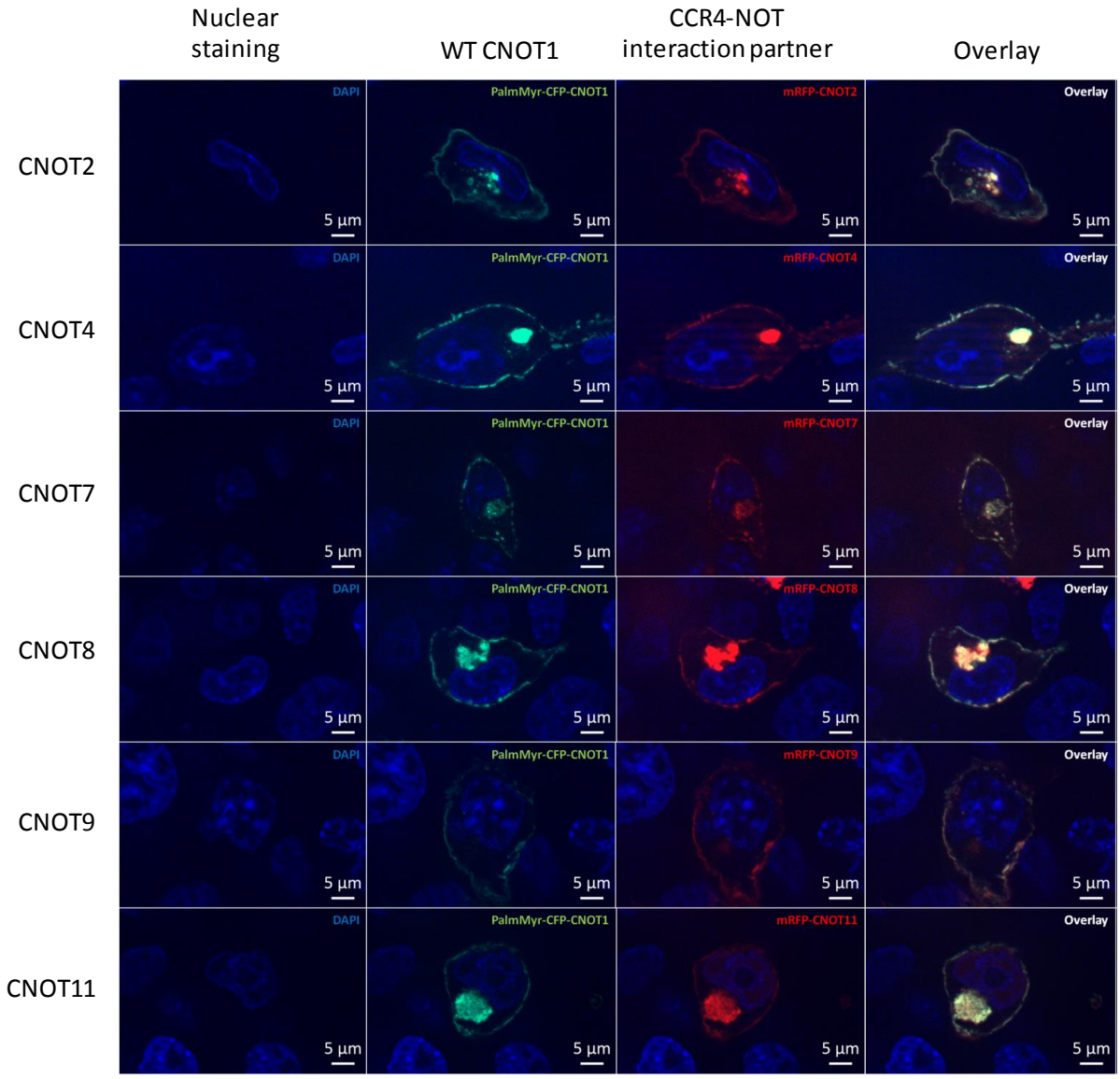
Positions of the *CNOT1* variants in CCR4-NOT transcription complex subunit 1. Cross-species alignment by Clustal Omega[2] of the protein sequences directly surrounding the missense variants. The unique changes are highlighted by the black boxes. The two *de novo* changes identified in the same individual are indicated by the red boxes. Protein accession numbers used for alignment are given before the sequences and include the specific species. The position of the last amino acid residue in each row is given right after the respective sequences.

Supplemental Figure 4: Direct and indirect CNOT1 CCR4-NOT interaction partners



Schematic representation of interaction partners of CNOT1 in its scaffolding function in the CCR4-NOT complex. Direct interaction partners include CNOT11, binding to the CNOT11-binding domain (CNOT11 BP), CNOT7/8 to the CAF1 domain, CNOT9 to the DUF3819 domain and CNOT2 to the Not1 domain. In addition, through indirect interactions, CNOT1 fulfils a scaffolding function for CNOT10, CNOT6 and CNOT3. Prior to this study, it was unknown whether CNOT4 directly or indirectly binds CNOT1.

Supplemental Figure 5: Co-localization of wildtype CNOT1 and wildtype CCR4-NOT1 interaction partners in COS-1 cells



Prior to assessing the (co-)localization of wildtype (WT) CNOT1 and its CCR4-NOT1 interaction partners, transfection rates were determined. Approximately 10% of all cells was transfected (CNOT1) and between 28-53% of the transfected cells was double transfected (CFP-CNOT1+ CCR4-NOT interaction partner). For these double transfected cells, colocalization was studied. The palmMyr-CFP-CNOT1 (green) was targeted to the cell membrane and co-localized with the mRFP-labeled CNOT2, 4, 7, 8, 9 and 11 (red), respectively. Of note, in addition to the membrane localization, partners CNOT7, 8, 9 and 11 also located to the nucleus and cytosol.

Supplemental Figure 6: CNOT1 missense variants do not seemingly disrupt binding of CCR4-NOT1 interaction partners

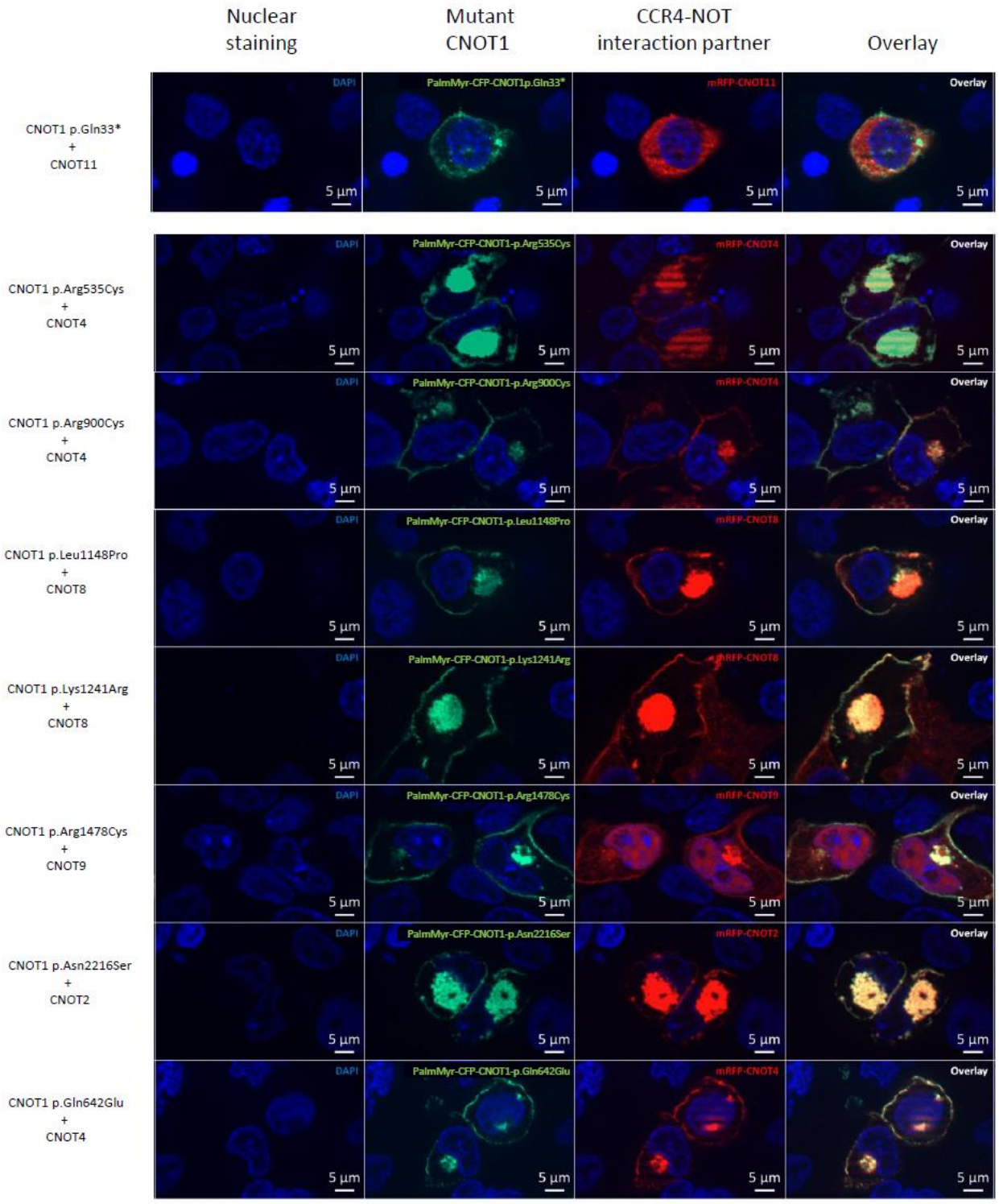
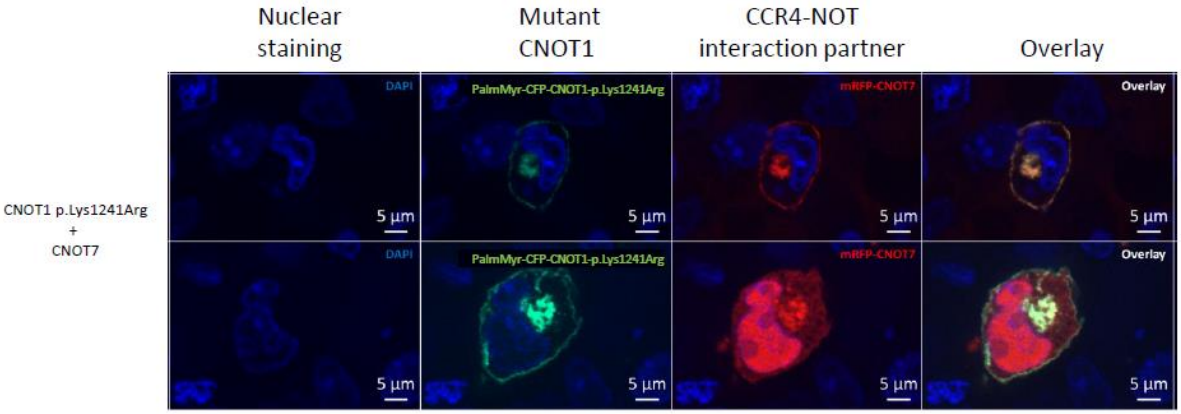


Figure S6 continued



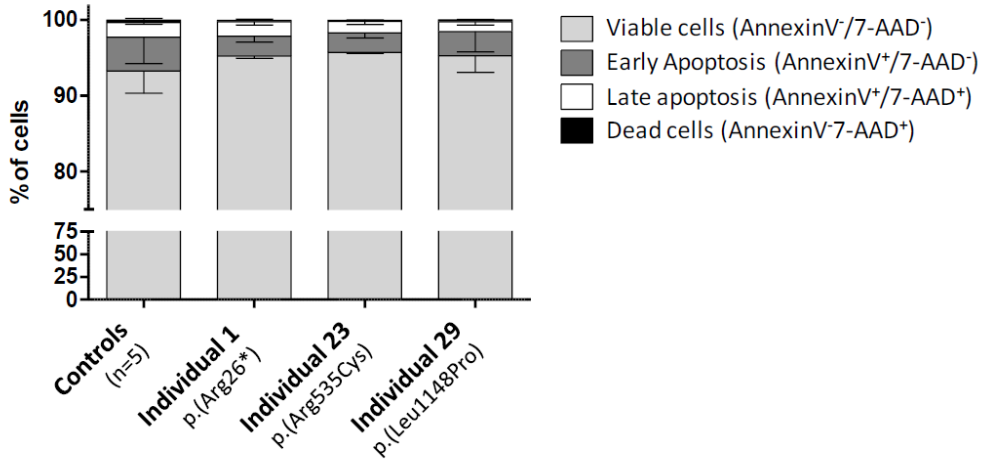
Co-transfection of mutant CNOT1 and CNOT2, 4, 8, 9 and 11 in COS-1 cells. PalmMyr-CFP-CNOT1-mut (green) were targeted to the cell membrane and transfected with RFP-tagged (red) subunits. Top panel shows the CNOT1-p.Gln33* with its interaction partner CNOT11. The variant clearly disrupts co-localization. Whereas this nonsense variant is expected to be degraded by nonsense mediated RNA decay in patient-derived cells, the loss of interaction shows the validity of the assay. For all (but one) missense variants, co-localization is not affected. For the co-transfection of CNOT1-Lys1241Arg with CNOT7, we noted that the interaction between was disturbed in approximately half of the cells, while in the other part the interaction remained intact.

Supplemental Figure 7: Apoptosis assay in CNOT1-mutant patient-derived EBV cell lines

A

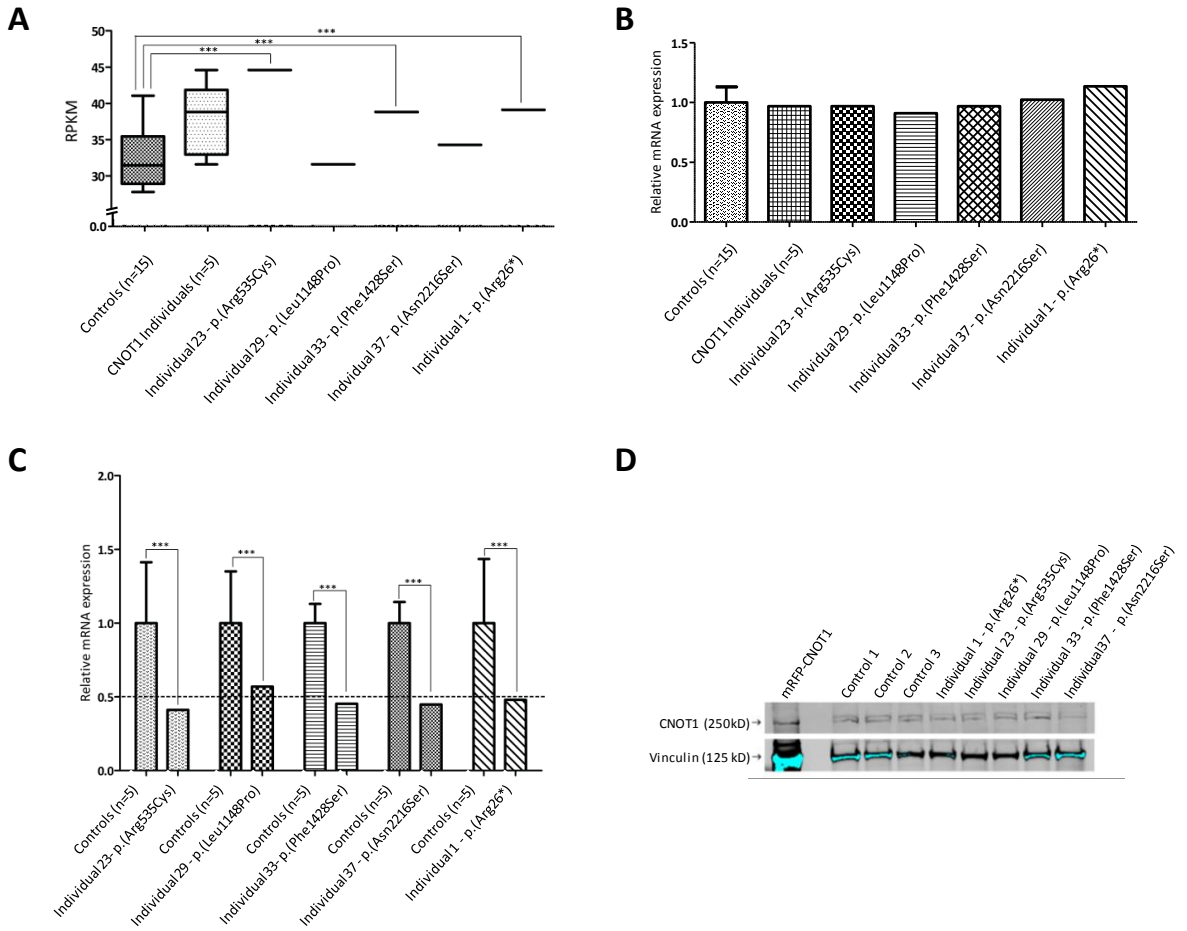
	AnnexinV $^-$ / 7-AAD $^-$ Alive cells (%)		AnnexinV $^+$ / 7-AAD $^-$ Early apoptosis (%)		AnnexinV $^+$ / 7-AAD $^+$ Late apoptosis (%)		AnnexinV $^-$ / 7-AAD $^+$ Dead cells (%)	
	exp 1	exp 2	exp 1	exp 2	exp 1	exp 2	exp 1	exp 2
Controls (n=5)	95.4	91.2	2.0	6.9	2.1	1.8	0.5	0.1
Individual 1 - p.(Arg26*)	95.4	95.0	2.1	3.2	2.2	1.6	0.3	0.1
Individual 23 - p.(Arg535Cys)	95.8	95.6	2.1	3.1	1.8	1.2	0.2	0.1
Individual 29 - p.(Leu1148Pro)	96.8	93.7	1.3	5.1	1.6	1.0	0.3	0.2

B



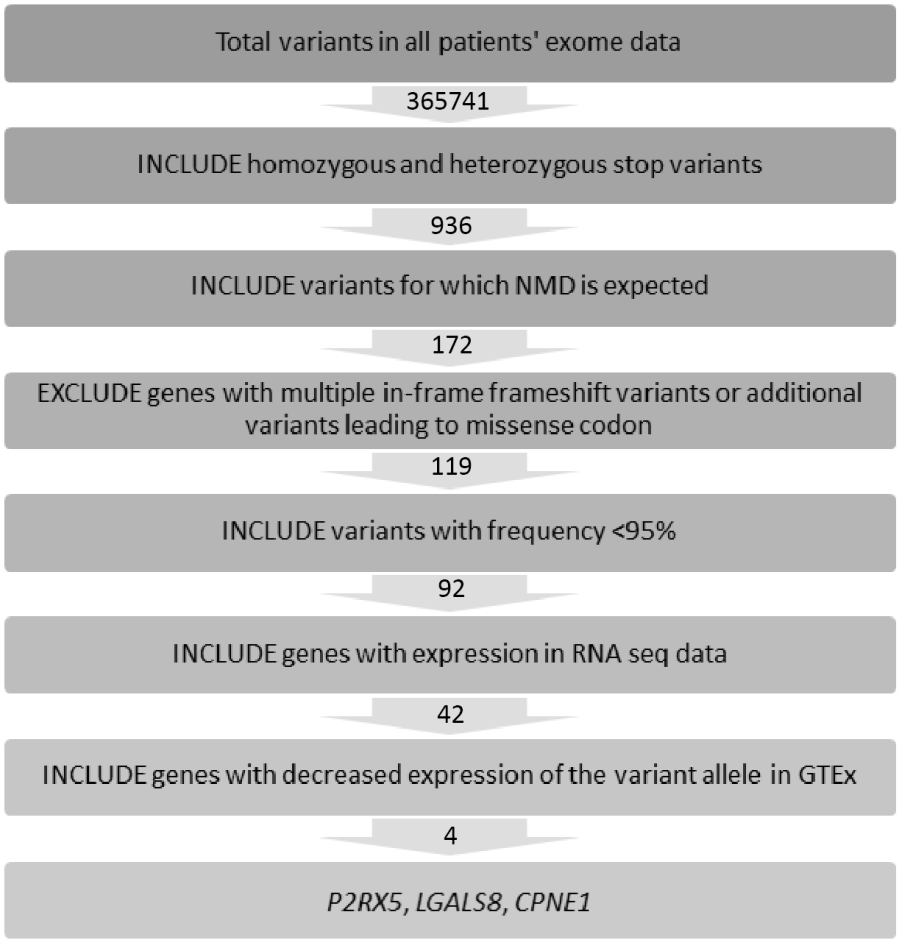
(**A**) FACS results denoting the percentage of cells in various apoptotic stages using Annexin V staining for apoptotic cells, and 7-AAD for viability. The experiment was repeated twice (exp 1 and exp 2). (**B**) Graphical representation of FACS results presented in A. From these data, it was concluded that there is no significant increase in apoptosis in patient-derived EBV-LCLs with *de novo CNOT1* variants when compared to controls (n=5).

Supplemental Figure 8: CNOT1 RNA and protein levels in Individuals with a CNOT1 de novo variant do not differ from controls



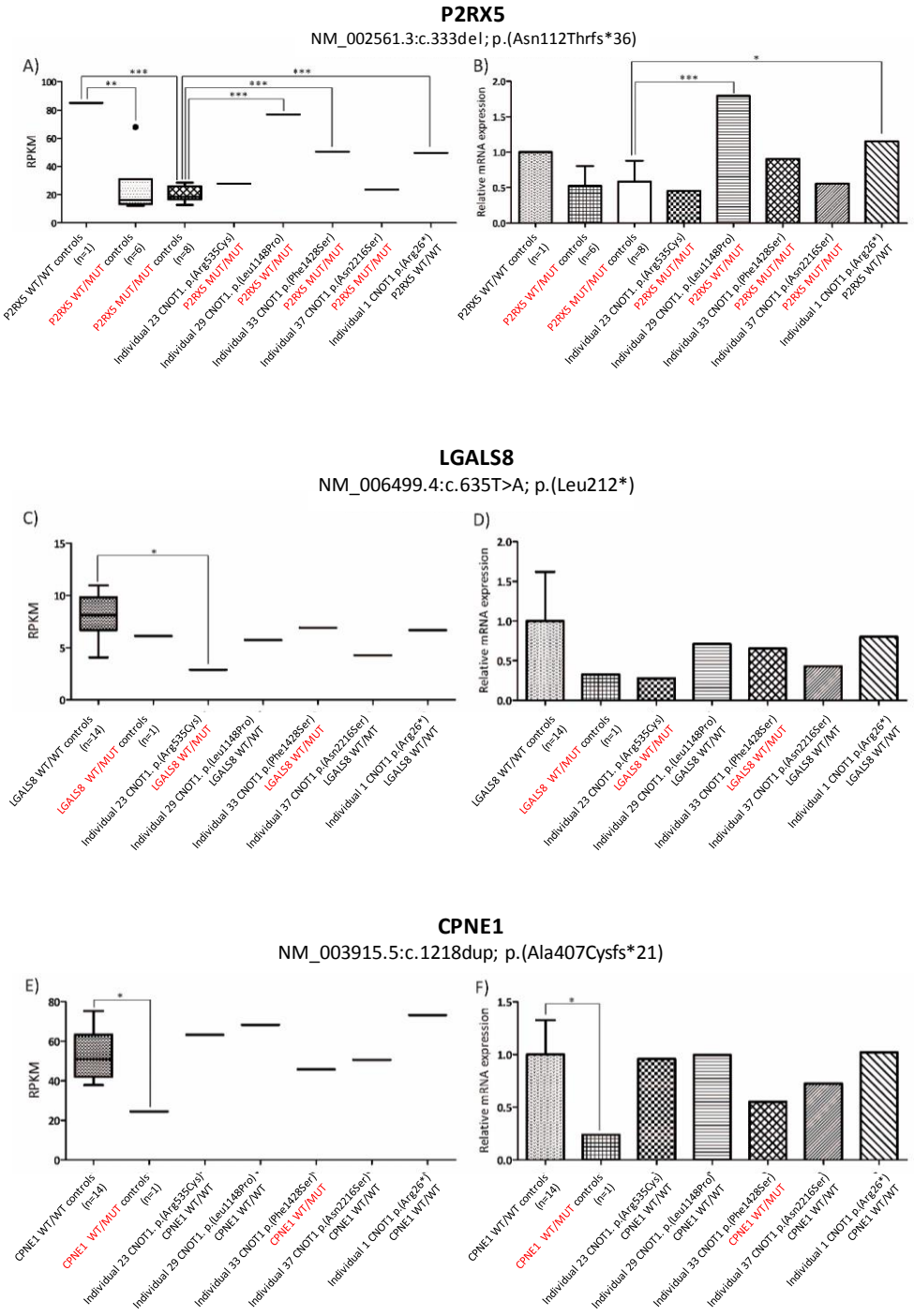
Evaluation of CNOT1 RNA (**A,B**, **C**) and protein (**D**) levels in patient-derived EBV cell lines of individuals with CNOT1 de novo variants and controls. In (**A**) normalized QuantSeq 3'mRNA counts are visualized for controls (n=15) and Individuals with *de novo* CNOT1 variants (n=5). Comparison between Individuals with *CNOT1* variants and controls does not show a statistical difference in CNOT1 expression. When individual controls are compared such difference can be observed (***= p<0.001). (**B**) Validation of QuantSeq results by conventional qPCR, indicating that there is no difference in expression between individuals with a *de novo* variant in CNOT1 and controls. (**C**) More detailed allele-specific qPCR shows that when only the WT allele is targeted by qPCR (tailor-made per individual; Table S6), only half the expression is observed in the patient-derived cell lines. These results indicating that the total amount of expression as observed in (**B**) is obtained by equal expression of both the WT and mutant allele. (**D**) Analysis of CNOT1 protein levels in these individuals again do not show significance difference when compared to controls. Vinculin is used as loading control.

Supplemental Figure 9: Filter strategy to identify genes harbouring PTCs expected to be targeted by NMD in controls and individuals with *de novo CNOT1* variants



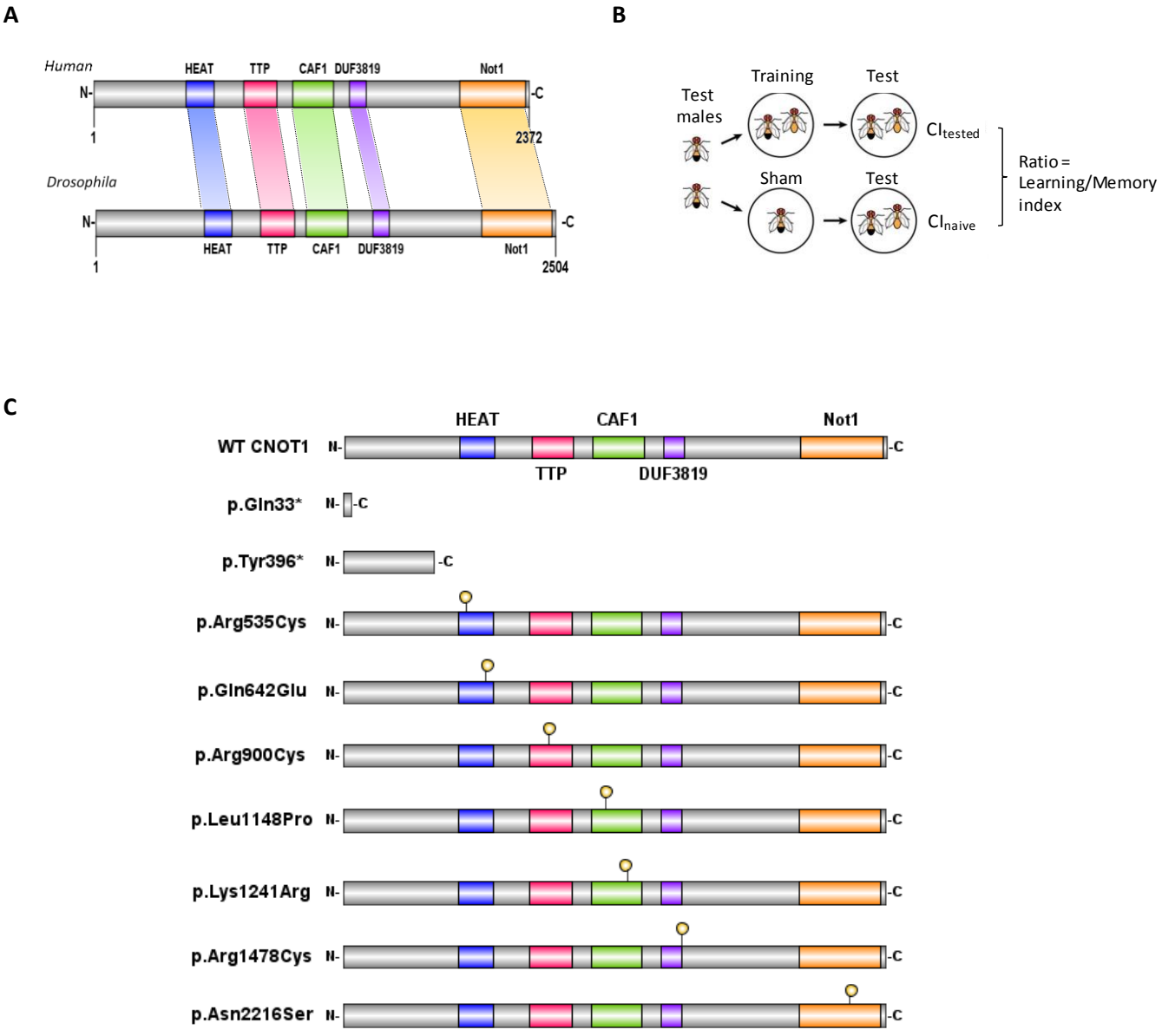
Selection criteria for filtering of exome sequencing data of individuals with *de novo CNOT1* variants (Individuals 1, 22, 28, 32 and 36) to evaluate NMD in QuantSeq 3'mRNA seq data. Overall QC thresholds used for filtering WES data: read counts for heterozygous and homozygous were set at 6x and 4x coverage respectively, with homozygous variants showing >90% of variant reads and heterozygous variants between 20-80%. Only "stop gain", "canonical splice sites" and "frameshift" were selected if they had a frequency higher than 10% of the alleles in a population (gnomAD frequency).

Supplemental Figure 10: Expression of genes of interest for which NMD is expected



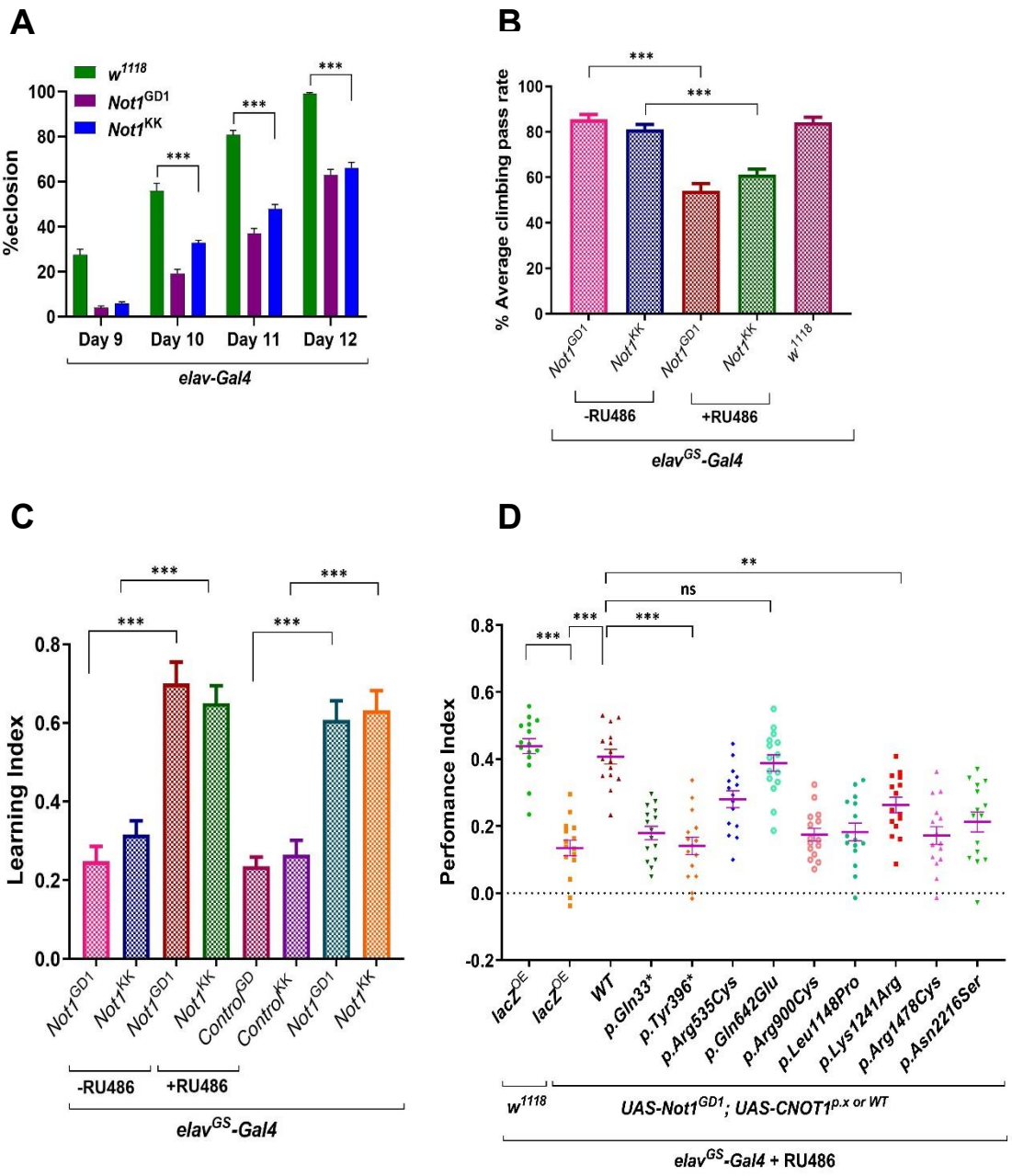
The genotype for the nonsense of frameshift variant in the gene of interest is indicated below each plot (in red, individuals with heterozygous or homozygous premature termination codons). For each gene, left panel shows the RPKM values by QuantSeq, with on the right panel, the validations thereof as relative mRNA expression obtained by qPCR. Significant results of two-tailed unpaired t-tests comparing control groups and two-tailed z-test comparing single samples with the control groups are indicated with stars ($p \leq 0.05$=*, $p \leq 0.01$=**, $p \leq 0.001$=***). Whereas subtle differences in NMD between controls and individuals with CNOT1 variants are observed from RNA seq data, none of them withstand validation by qPCR and statistical analysis. Hence, we concluded that individuals with de novo CNOT1 variants degrade transcripts with protein truncating variants in similar ways as controls, and that the endonucleolytic NMD pathway is not affected.

Supplemental Figure 11: Conservation of CNOT1 in Drosophila and schematic representation of courtship assay



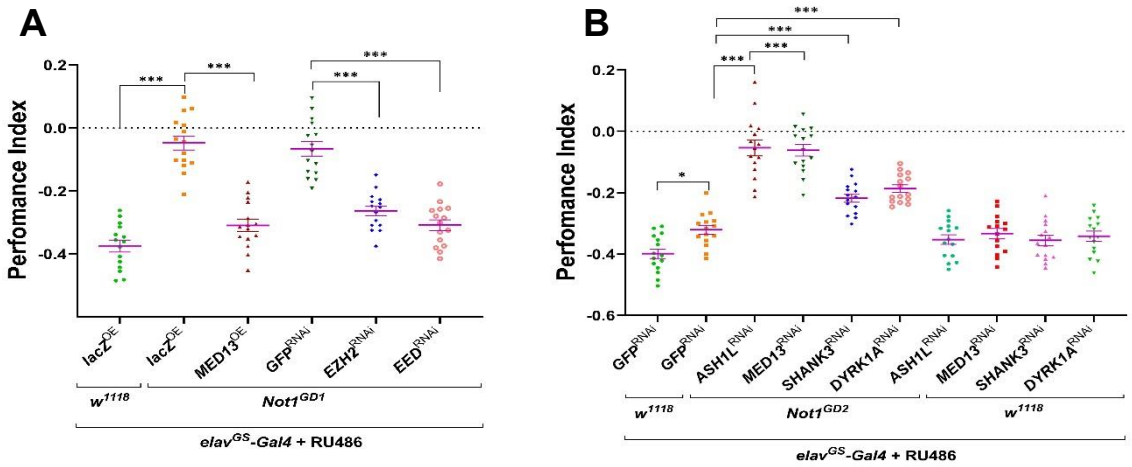
Schematic representation (**A**) of homology between human CNOT1 and its drosophila homolog Not1, and the principle of the courtship assay (**B**). In essence, if male flies have normal learning and memory, the learning and memory index, calculated by CI_{tested}/CI_{naive}, is low. Impaired learning and memory is shown by an index closer to 1.0. (**C**) Overview of CNOT1 constructs generated for the Drosophila functional follow-up, each harbouring a variant observed in an Individual with neurodevelopment disorders as observed in this study.

Supplemental Figure 12: Loss-of-Not1 function in Drosophila neurons results in multiple deficits.



(**A**) Persistent knockdown of *Not1* in all neurons under the control of *elav*-Gal4 induces a significant developmental delay. (**B**) Adult-only knockdown of CNOT1 for 3 weeks lead to motor dysfunction as assessed with the climbing assay described in Methods. (**C,D**) elav^{GS}-Gal4-induced knockdown of *Not1* in adults (**C**) or larval (**D**) stages results in learning and memory deficits. One–way ANOVA, **p<0.01, ***p<0.0001

Supplemental Figure 13: Genetic interactions between Not1 and known ID/ASD genes and rescue of Not1-induced neurodevelopmental defects at larval stages



(**A**) Overexpression of *MED13* or knockdown of PRC2 components *E(z)* and *esc* ameliorates *Not1* knockdown-induced memory deficits in larvae. (**B**) Neuronal RNAi knockdown of known ID/ASD genes (*ash1, MED13, shank3* and *Dyrk1a*) exacerbates *Not1* knockdown-induced memory defects in larvae. One–way ANOVA, *p<0.05, **p<0.01, ***p<0.0001

Supplemental Tables

Supplemental Table 1

Individual	Gender	Genomic annotation on Chr16(GRCh37):	c.DNA (NM_016284.4)	Protein	Type¥	Inheritance
Nonsense, Frameshift and Splice site variants						
Individual 1	F	g.58633166G>A	c.76C>T	p.(Arg26*)	Nonsense	*De novo*
Individual 2	M	g.58633166G>A	c.76C>T	p.(Arg26*)	Nonsense	Maternal mosaicism
Individual 3	F	g.58633166G>A	c.76C>T	p.(Arg26*)	Nonsense	*De novo*
Individual 4	F	g.58633166G>A	c.76C>T	p.(Arg26*)	Nonsense	*De novo*
Individual 5	M	g.58633145G>A	c.97C>T	p.(Gln33*)	Nonsense	*De novo*
Individual 6 *(daughter of 7)*	F	g.58633138A>G	c.102+2T>C	p.(?)	Splice site	Maternal
Individual 7 *(mother of 6)*	F	g.58633138A>G	c.102+2T>C	p.(?)	Splice site	Unknown
Individual 8	F	g.58622702C>A	c.210+1G>T	p.(?)	Splice site	*De novo*
Individual 9	M	g.58620475_58620478delTCTA	c.608_611delTAGA	p.(Ile203Thrfs*32)	Frameshift	*De novo*
Individual 10	F	g.58615276A>T	c.1188T>A	p.(Tyr396*)	Nonsense	Unknown: adopted
Individual 11	F	g.58583781_58583782insTTTAGCTTACCTTA	c.3363_3364insTAAGGTAAGCTAAA	p.(Val1122*)	Frameshift	*De novo*
Individual 12	F	g.58581153_58581159delTTGTCCT	c.3681_3687del	p.(Lys1227Asnfs*7)	Frameshift	Unknown: mother NA, not paternal
Individual 13	M	g.58581089C>A	c.3750+1G>T	p.(?)	Splice site	*De novo*
Individual 14	F	g.58581085C>T	c.3750+5G>A	p.(?)	Splice site	*De novo*
Individual 15	F	g.58577805T>G	c.4138-2A>C	p.(?)	Splice site	*De novo*
Individual 16	M	g.58575405C>G	c.4800G>C	(p.Lys1600Asn) / p.(?)	Missense / Splice site	*De novo*
Individual 17	M	g.58562529dup	c.6303dup	p.(Leu2102Serfs*4)	Frameshift	*De novo*

Individual 18 *(son of 19)*	M	g.58559939_58559967dup	c.6529_6557dup	p.(Arg2187Lysfs*63)	Frameshift	Maternal
Individual 19 *(mother of 18)*	F	g.58559939_58559967dup	c.6529_6557dup	p.(Arg2187Lysfs*63)	Frameshift	Unknown
Missense variants						
Individual 20*	M	g.58610468G>A	c.1603C>T	p.(Arg535Cys)	Missense	*De novo*
Individual 21*	F	g.58610468G>A	c.1603C>T	p.(Arg535Cys)	Missense	*De novo*
Individual 22#	M	g.58610468G>A	c.1603C>T	p.(Arg535Cys)	Missense	*De novo*
Individual 23#	F	g.58610468G>A	c.1603C>T	p.(Arg535Cys)	Missense	Unknown: mother NA, not paternal
Individual 24#	F	g.58610468G>A	c.1603C>T	p.(Arg535Cys)	Missense	*De novo*
Individual 25	M	g.58608568G>C	c.1924C>G	p.(Gln642Glu)	Missense	*De novo*
Individual 26	M	g.58589357C>T	c.2689G>A	p.(Glu897Lys)	Missense	*De novo*
Individual 27	M	g.58589348G>A	c.2698C>T	p.(Arg900Cys)	Missense	*De novo*
Individual 28	M	g.58585581G>A g.58585113C>G	c.3113C>T c.3265G>C	p.(Thr1038Ile) p.(Val1089Leu)	Missense Missense	*De novo* *De novo*
Individual 29	F	g.58583702A>G	c.3443T>C	p.(Leu1148Pro)	Missense	*De novo*
Individual 30	M	g.58581546T>C	c.3563A>G	p.(Asp1188Gly)	Missense	*De novo*
Individual 31	F	g.58581118T>C	c.3722A>G	p.(Lys1241Arg)	Missense	*De novo*
Individual 32	F	g.58577690T>C	c.4255A>G	p.(Thr1419Ala)	Missense	*De novo*
Individual 33	M	g.58577662A>G	c.4283T>C	p.(Phe1428Ser)	Missense	*De novo*
Individual 34	M	g.58577513G>A	c.4432C>T	p.(Arg1478Cys)	Missense	*De novo*
Individual 35	F	g.58576425T>A	c.4482A>T	p.(Gln1494His)	Missense	*De novo*
Individual 36	M	g.58575491A>C	c.4714T>G	p.(Tyr1572Asp)	Missense	*De novo*
Individual 37	M	g.58559220T>C	c.6647A>G	p.(Asn2216Ser)	Missense	*De novo*
(Partial) gene deletions						
Individual 38	M	46,XY,del(16)(q21) arr 16q21(58589134_58594329)x1	deletions of exon 17-21	p.(?)	Intragenic deletion	*De novo*
Individual 39	M	46,XY,del(16)(q12.2q21) arr 16q12.2q21(55,741,501-62,145,478)x1	-	-	Whole gene deletion	*De novo*

**Individual previously described by Kruszka et al. 2019; #: Individual previously described by De Franko et al. 2019.*

Supplemental Table 2: Clinical details for each individual with a (*de novo*) *CNOT1* variant

Excel file describing the clinical details for each individual included in this study.

Supplemental Table 3: Effect predictions of de novo missense variants in *CNOT1*

Unique missense variants	Domain	PDB structure	Location of missense	Predicted effect
p.(Arg535Cys)	HEAT	-		Loss of positive charge, and more hydrophobic than wildtype residue.
p.(Gln642Glu)	HEAT	-		Introduction of a negative charge.
p.(Glu897Lys)	TTP	4J8S	Buried in the core of the protein.	Introduction of a positive charge instead of a negative one. Loss of hydrogen bond, and sidechain might clash with Arg in vicinity. Possible destabilization of ZFP36 binding groove which affects the interaction with this protein.
p.(Arg900Cys)	TTP	4J8S	Buried in the core of the protein.	Loss of positive charge, and more hydrophobic than wildtype residue. Loss of salt-bridge formation with Glutamic Acid at position 940, which might cause loss of stability close to the ZFP63 binding groove. Interaction with ZFP36 might be disturbed.
p.(Thr1038Ile)	-	-	-	*No 3D structure available for effect prediction*
p.(Val1089Leu)	CAF1	4CT4	Surface of the protein	Change to hydrophobic amino acid. May disturb interactions with DDX6 and CNOT6/CNOT6L/CNOT7/CNOT8.
p.(Leu1148Pro)	CAF1	4CT4	Buried in the core of the protein.	The proline is expected to disturb the alpha-helix which may have severe effects on the structure of the protein. Variant will affect the interaction with DDX6 and CNOT6/CNOT6L/CNOT7/CNOT8.
p.(Asp1188Gly)	CAF1	4CT4	Semi-buried in core	Change of a relatively large to smaller residue ca affect the general stability of the conformation. No specific interactions of function found for this residue.
p.(Lys1241Arg)	CAF1	4CT4	Surface of the protein.	Loss of hydrogen bonds with Glutamic Acid at position 1244 as well as the salt bridges with Glutamic Acids 1244 and 1282. Interaction with CNOT6/CNOT6L/CNOT7/CNOT8 might be disturbed.
p.(Thr1419Ala)	DUF3819	4CRU	Surface of the protein.	Alanine is a bit smaller and hydrophobic. Maybe interaction with CNOT9 might be disturbed.
p.(Phe1428Ser)	DUF3819	4CRU	Buried in the core of the protein.	Less hydrophobic than wild type residue. Phe1428 reported by PISA-assembly be involved in protein interaction. Interaction with CNOT9 might be disturbed.
p.(Arg1478Cys)	DUF3819	4CRU	Surface of the protein.	Loss of positive charge, and more hydrophobic than wildtype residue. Structural information is unclear about function of this residue.
p.(Gln1494His)	DUF3819	4CRU	Surface of the protein	Size difference, distant from CNOT9 interaction site, no severe effect on structure expected
p.(Tyr1572Asp)	-	4CRU	Buried in the core of the protein.	Important for hydrophobic interactions in core. Loss of stability, close to CNOT9 interaction surface.
p.(Asn2216Ser)	Not1	4C0D	Buried in the core of the protein.	The variant will cause loss of hydrogen bonds (with Aspartic Acid at position 2219) in the core of the protein and as a result disturb correct folding.

Supplemental Table 4: *CNOT1* variants created by site directed mutagenesis

Protein annotation based on NM_001265612.1	Primer Orientation	Primer sequence (5'->3')
p.(Gln33*)	FW	GCAGGAAATATAGCATATTGTGAATC
	RV	TGGCTGGCTCGGTAATTT
p.(Tyr396*)	FW	ACCTCATATAAAGACCTTGGAAAC
	RV	CTACTGGGAACACTTCCATAC
p.(Arg535Cys)	FW	TCCCTCAATTTGCCAACTTATCATGC
	RV	GACTGTCCCTGCCCATGC
p.(Gln642Glu)	FW	CAAAAGTGCTGAACTTCCTCC
	RV	GGCTGGTCTTTTTCTGGG
p.(Arg900Cys)	FW	TGAAGAATATTGTTTTTTTCCCC
	RV	AACAAGTTCCTTAGCATAC
p.(Leu1148Pro)	FW	TTTCATAGCCCGTATTCAAACTTC
	RV	GTTTGGCTCAATACTGAC
p.(Lys1241Arg)	FW	TTTGTTGCCAGAGTCTTAGAATC
	RV	GGGCACTACATAGAGCAA
p.(Phe1428Ser)	FW	AGGAAGGATTCTGCCCTGGAT
	RV	GACTATTTGCTCACAAGTAGTCATG
p.(Arg1478Cys)	FW	CTCAGCCCTTTGTACTGCTTC
	RV	GCAAAACTGTTTTTTAAGTTGG
p.(Asn2216Ser)	FW	CAGCTCATCAGTGCACTGGTG
	RV	GAGGTTGTAGCGATTCCC
p.(Leu2102Serfs*4)	FW	TTTCTTTTGCATGATTTCCCAG
	RV	ACCAGCAGCACTCTTAAAG

Supplemental Table 5: Experimental set-up co-localization studies

CNOT1 mut	CNOT2	CNOT4	CNOT7	CNOT8	CNOT9	CNOT11
p.(Gln33*)						X
p.(Arg535Cys)		X				
p.(Gln642Glu)		X				
p.(Arg900Cys)		X				
p.(Leu1148Pro)			X	X		
p.(Lys1241Arg)			X	X		
p.(Arg1478Cys)					X	
p.(Asn2216Ser)	X					

Supplemental Table 6: Allele-specific primers to assess contribution of WT allele to total *CNOT1* expression

WT allele of	Sequence (5'-3')	(5'-3')
CNOT1 p.Arg26*	CAATTTAACCAAGAAAAATGACC	GAAATCCACATGCGAAAATAGGC
CNOT1 p.Arg535Cys	GGACAGTCTCCCTCACTTC	AAGTCCTGGGCCACATCAAG
CNOT1 p.Leu1148Pro	TGAGCCAAACTTTCATATCCT	CAGAGGTCAGGAGCACTTTAATG
CNOT1 p.Phe1428Ser	GCAAATAGTCAGGAAGGCTTT	AGCTGTCAAGTTACGCATCATG
CNOT1 p.Asn2216Ser	TACAACCTCCAGCTCAGCAA	TCCATGTGTGCTGAGTGAGT

For CNOT1, a general primer pair was used (Table S7), as well as primer pairs that are designed to be specific for the WT allele of each individual. For WT-allele-specific primers, the mutated nucleotide is marked red and the introduced mismatch is marked blue. The genes GUSB, PPIB and CLK2 were used for normalization (primer sequences in Table S7).

Supplemental Table 7: qPCR primers used for validation of RNAseq data and allele-specific expression

Gene	Sequence (5'-3')	Sequence (5'-3')
P2RX5	GTCTGTGCTGAGAATGAAGGC	GTCTTCACTCCGTTTCCAGC
LGALS8	GCATGTTCCTAGTGACGCAG	ACGAGGATTGAAATGAAAGGCC
CPNE1	ACTGACTCTCCCCTTGATGC	GGTTTCTGGCCTCTACCTCC
CNOT1 total expression	GCCCAAAAGTGCTCAACTTC	TGGCTACCATGGTGAGGATAG
RUVBL2	ATTGATCGACCAGCAACAGG	GACTCAATCATCTTGGTGCCC
PABPC1	CGCGTATGTGAACTTCCAGC	ACTGGCTTGCCCTTTATAACATC
PPP2CA	TGGTGGATGGGCAGATCTTC	TGGGGAACTTCTTGTAGGCG
GUSB*	AGAGTGGTGCTGAGGATTGG	CCCTCATGCTCTAGCGTGTC
PPIB*	CGGAAAGACTGTTCCAAAAAC	GATTACACGATGGAATTTGCTG
CLK2*	CGGCGAGAGGACAGCTAC	AGTATCGCCGGTCATACACC

** used as reference/control*

Supplemental Table 8: Targets identified for NMD assessment by integrating exome sequencing and QuantSeq data of Individuals with a *de novo CNOT1* variants

Gene	gDNA (GRCh37)	cDNA	Protein	Freq. (%)	Individual (allelic composition for variant)
P2RX5	chr17:g.3594277delG	NM_002561.4: c.333delC	p.(Asn112Thrfs*36)	58.51	1(**MUT**/**MUT**) 23 (**MUT**/**MUT**) 29 (**WT**/**MUT**) 33 (**MUT**/**MUT**)
LGALS8	chr1:g.236706300T>A	NM_006499.4: c.635T>A	p.(Leu212*)	8.041	23 (**WT**/**MUT**) 37 (**WT**/**MUT**)
CPNE1	chr20:g.34215234insA	NM_003915.5: c.1218dupT	p.(Ala407Cysfs*21)	9.653	33 (**WT**/**MUT**)

The positions of the variants in the genomic DNA (GRCh37), effect at transcript level (cDNA), and protein level are provided. The last column indicates the individuals who carry the nonsense or frameshift variant. In parenthesis is indicated whether the variant is identified as heterozygous (WT/MUT) or homozygous (MUT/MUT) variant. Freq. = frequency that refers to the overall frequency in GnoMAD.

Supplemental Material and Methods

Generation of expression vectors of the CCR4-NOT complex. To study the functional consequences of *de novo* variants in *CNOT1*, the coding region of wild-type (WT) cDNA (NM_001265612) was cloned into template vector pT7-EGFP-C1-HsNot1 (a kind gift from Elisa Izaurralde, Addgene plasmid # 37370)[3], after which the cDNA was subsequently cloned into destination vector pDONR201 using the gateway system according to the manufacturer's instructions (ThermoFisher Scientific, Landsmeer, the Netherlands), and transformed into DH5α cells using standard protocols. Plasmid DNA was isolated using a NucleoSpin® Plasmid QuickPure kit (Macherey-Nagel, Dueren, Germany), and, after digestion using HINDIII-HF (20 U/µl, New England Biolabs, Bioké, Leiden, the Netherlands), sequenced to exclude the presence of variants. Then, pDONR201-CNOT1 was cloned into two expression vectors: pPalmMyr/DEST(CFP), encoding PalmMyr-CFP-CNOT1, which contains a green fluorescently-labeled (CFP) PalmMyr cell membrane linker; and pDEST-733 (mRFP, ThermoFisher Scientific, Landsmeer, the Netherlands), encoding red fluorescent-labeled (RFP)-CNOT1, in which CNOT1 contains a RFP tag allowing visualization in the cytosol and/or nucleus. Using similar protocols, also the CCR4-NOT complex subunits CNOT2, CNOT4, CNOT7, CNOT8, CNOT9 and CNOT11 were cloned into expression vector pDEST-733 (mRFP; ThermoFisher Scientific, Landsmeer, the Netherlands).

Detailed information of CCR4-NOT complex members used in this study:

Subunit	Catalog number	Accession number (EST/RefSeq)
CNOT2	GC-Z1136	ENST00000418359/NM_001199302
CNOT4	GC-A1384	ENST00000423368/NM_001190847
CNOT7	GC-T8806	ENST00000361272/NM_001322090
CNOT8	GC-A6058	ENST00000519404/NM_001301074
CNOT9	GC-U0151	ENST00000273064/NM_005444
CNOT11	GC-U0969	ENST00000289382/NM_017546

Site directed mutagenesis. In total, 11 CNOT1 variants were generated using site directed mutagenesis in pDONR201-CNOT1 (Table S3). In short, primers were designed and ordered (IDT) for each variant using NEBaseChanger, after which variants were introduced using the Q5 mutagenesis protocol of the

manufacturer (New England Biolabs, via Bioké Leiden the Netherlands). PCR products were sequenced to ensure that the variants were introduced correctly, and subsequently transformed into DH5α cells. Plasmids were isolated routine Miniprep kit (NucleoSpin® Plasmid QuickPure (Macherey-Nagel, Dueren, Germany). After digestion using PstI-HF (20 U/µl, New England Biolabs, via Bioké Leiden the Netherlands), the mutated CNOT1 constructs were cloned into expression vector pPalmMyr-CFP, thus encoding a variant-specific PalmMyr-CFP-CNOT1 expression.

Co-localization studies of CNOT1 and CCR4-NOT complex partners in COS-1 cells. COS-1 cells were grown to ~90% confluency on a glass coverslip in a 12-well plate, after which they were transfected with plasmid isolates according to the manufacturer's instructions (Lipofectamine® 3000; Cat# L3000001; Life Technologies, Landsmeer, the Netherlands) with minor modifications. In brief, 3µl Lipofectamine® 3000 was used for co-transfection of 500ng of the CFP-PalmMyr-CNOT1/mut plasmid DNA and 300ng of the RFP-CNOT subunits CNOT2, CNOT4, CNOT7, CNOT8, CNOT9 or CNOT11 (**Table S4**), respectively, while for single transfection 500ng of PalmMyr-CFP-CNOT1/mut or 300ng of the RFP-CNOT subunits CNOT2, CNOT4, CNOT7, CNOT8, CNOT9 or CNOT11, and 1 µl lipofectamine® 3000 was used. After ca. 6 hours the media was replaced by the normal DMEM media. The cells were washed two times in 1X PBS 36-48 hours after transfection and fixed in 2% paraformaldehyde (PFA, stock 4% PFA diluted in PBS) at room temperature for 15 min. Cells were stained with 4',6-diamidino-2-phenylindole (DAPI; Vector laboratories, via Brunschwig Chemie, Amsterdam, the Netherlands) while placed on a microscope slide, and localization of CNOT1 WT/mut and its subunits was analyzed using a Zeiss Axio Imager Z1/Z2 fluorescence microscope equipped with a 63x objective lens. Image processing was performed using Zen software (blue edition, Zeiss).

Apoptosis assay using patient-derived cell lines. For three Individuals (p.(Arg26*), p.(Arg535Cys), p.(Leu1148Pro)) and five control lines, an apoptosis assay was performed using cultured Epstein-Barr-

virus transformed lymphoblastoid cell lines (EBV-LCLs). Hereto, cells were incubated for 20-22 hours in medium (RPMI + 15%FSC + HEPES (1:50) and 1%Pen/Strep) containing 1µM Bortezomib to induce apoptosis. Next, cells from 2 ml culture were harvested and counted, to be re-suspended at a concentration of $2.5x10^6$ cells/ml using ice-cold PBS. Of these, 500,000 cells (200 µl) were spun down and re-suspended in 1xFACS buffer (PBS with 0.5% BSA) containing Annexin V-FITC (apoptosis marker) and 7-AAD (viable cell marker). FACS analysis, discriminating viable cells from those in apoptosis, was performed in duplicate using routine procedures.

qPCR Experiments for validation and allele-specific expression of individuals with a *de novo CNOT1* variant. For the quantitative real-time PCR (qPCR) experiments, RNA samples were isolated from EBV cell lines and converted to cDNA using routine procedures. qPCR primers at cDNA level were designed with Primer3web v4.1.0 (Table S7)[4]. Allele-specific primers were designed for the wildtype (WT) alleles of the patients' CNOT1 variants, according to Gaudet et al. (2009)[5]. *GUSB*, *PPIB*, and *CLK2* were taken along in the qPCR experiments as reference genes[6]. Standard curves were made for all primers with a cDNA control sample in a dilution series (20x, 80x, 320x, 1280x, 5120x dilutions in MilliQ). For all qPCR experiments, samples were tested in duplicate and a blank was taken along for each primer pair.

QuantSeq 3'mRNA sequencing. A selection of five individuals carrying *de novo CNOT1* mutations (Table S6) and 15 healthy controls were subjected to quantitative RNA sequencing (QuantSeq 3'mRNA-Seq Library Prep kit-FWD, Lexogen) in accordance with protocol. Prior to library generation, RNA from individuals with *CNOT1* variants and controls was isolated from EBV lymphoblastoid cell lines with either the NucleoSpin RNA kit (Macherey-Nagel) or the RNeasy mini kit (Qiagen) including on-column DNAse digestion for all samples (RNAse-Free DNAse kit, Qiagen), according to the manufacturer's protocol. Respective quantification of purified RNA was determined using Qubit RNA HS assay (Thermo Fisher Scientific). RNA input was normalized to 1000 ng for all samples. A mock qPCR on a 1:10 aliquot

of ds cDNA libraries indicated 17 cycles to be optimal for endpoint PCR. Quantification and quality control, including identification of the average fragment size, of the generated libraries was performed using the dsDNA HS assay (Qubit, Thermo Fisher Scientific) and HS-D1000 ScreenTape (2200 Tapestation, Agilent), respectively. Libraries were pooled equimolar to 100 fmol and diluted to a final concentration of 4 nM before sequencing on a NextSeq 500 instrument (75-cycle high output kit, Illumina). For RNAseq analysis, first quality was assessed using FastQC (v0.11.5, Babraham Bioinformatics), after which the raw data was subjected to TrimGalore! (v.0.4.4_dev, Babraham Bioinformatics) and Cutadapt[7] (v1.18) for removal of adapter sequences and poly(A) tails. Subsequent mapping of filtered and trimmed reads to the human reference genome (hg38) was performed using STAR aligner[8] (v2.6.0a). Lastly, gene level count data was generated using the HTSeq-count tool[9] (v0.11.0).

Evaluation of NMD efficiency using transcriptomics in patient-derived cell lines. Genes of interest for studying NMD of aberrant RNA transcripts contain a loss-of-function (LoF) in at least one individual with a de *novo CNOT1* variant. Selection was based on the following criteria (**Figure S8**). Homozygous and heterozygous loss-of-function variants were selected by filtering the exome sequencing (ES) data of five individuals with *de novo* CNOT1 variants. Next, variants were evaluated to fulfil the canonical rules for NMD: variants located >50-55 nucleotides upstream of the last exon-exon boundary or that were located in the last exon were excluded. Genes with multiple frameshift variants that were in-frame with each other and genes with additional variants that in combination with the selected stop variant led to a missense codon were excluded. In addition, only variants with a global SNP frequency <95% in the Genome Aggregation Database (gnomAD) were included to overcome problems with MAF wrongly represented as the common allele.[10] All genes without expression in the RNA sequencing data set were excluded from the analysis. The threshold for expression (QuantSeq) was set to a total read count of ≥10 in all twenty samples together. Lastly, the Genotype-Tissue Expression (GTEx) database

was consulted regarding the expression of the genes of interest with and without the LoF variants in EBV-LCLs[11]. Those genes for which the expression quantitative trait locus (eQTL) of the LoF variant showed a decreased expression of the variant allele in EBV-LCLs in the GTEx database were included.

***Drosophila* Genetics.** Fruit fly husbandry was done at 25℃ unless otherwise mentioned. Fly RNAi lines that include *Not1* (GD1#v12571, GD2#v41680, KK#v106587), *ash1* (GD#v28982, KK#v108832), *MED13* (SH#v330726), *shank3* (GD#44830, KK#v103592), *Dyrk1a* (GD#v28628, KK#v107066) were obtained from Vienna stock center (Vienna Biocenter Core Facilities GmbH, Vienna, Austria). The pan-neuronal Gal4 driver lines, *elav-Gal4* (BL-458), *elav^{GS}-Gal4* (BL-43642), *OK107-Gal4* (BL-854) were obtained from Bloomington stock center (Dept Biology, Indiana University, Indianapolis, United States). Additional RNAi lines for *MED13* (BL-34630), *Dyrk1a* (BL-35222) and UAS-*MED13* (BL-63800) were obtained from Bloomington stock center. Human *CNOT1* cDNA variants were cloned into a *pUAST*-attB vector and transgenes were produced (BestGene, Chino Hills, California, Unites States) using a site-specific integration method to avoid position effects on gene expression. Flies were raised on a standard molasses-corn meal food. For the courtship memory assays, adult flies of desired genotypes were treated with 100µM RU486 (Mifepristone) mixed in fly food for 1-week after eclosion in order to conditionally induce *elav^{GS}-Gal4*-mediated transgene expression before testing. For larval olfactory assays, larvae were allowed to hatch on fly food containing 100µM RU486 and stay there for 5-days before testing with the larval memory assays during late third instar larval stage.

Courtship suppression assay. Flies were raised at 25°C before collecting the F1 males and females separately. For courtship assays, naïve male (3-4 day old) and a mated female (3-4-day old virgin female mated one day before training) were transferred into the courtship chamber and the courtship was observed for 60mins. Courtship Index (CI) is defined as fraction of time a male spent in courtship activity during 10 min observation period. Initial 10mins ($CI_{initial}$) and final 10mins (CI_{final}) were carefully

monitored for the time the test male spends in courting the mated female. After 2-hours, the trained male was subsequently transferred to another courtship chamber with an immobilized virgin tester female and the courtship was monitored immediately for 10mins (CI_{test}). As a sham-trained control, the tester male was isolated for 60mins in the courtship chamber before testing towards the tester female (CI_{sham}). Courtship index, learning index and memory indexes were calculated as previously reported[12,13]:

1. Learning index = $CI_{final}/CI_{initial}$
2. Memory Index = CI_{test}/mean CI_{sham}

Larval olfactory-mediated memory tests. Larval memory tests were performed as per established protocols[14,15]. Briefly, assay plates were filled with a thin layer of 2.5% agarose (Sigma Aldrich Cat. No.#A5093, St. Louis, MO, Unites States) with or without a gustatory reinforcer. We used 2M sodium chloride (Sigma Aldrich Cat. No.#S7653, St. Louis, MO, Unites States), 2M fructose (Sigma Aldrich Cat. No.#47740, St. Louis, MO, Unites States) as reinforcers. Olfactory stimuli provided using 10µl amyl acetate (AM; Acros Organics, via Thermo Fisher, New Jersey, United States, Cat No#AC149182500 diluted 1:250 in paraffin oil, Sigma Aldrich Cat. No.#18512, St. Louis, United States) and benzaldehyde (BA; Sigma Aldrich Cat. No.#418099, St. Louis, United States). Odorants were placed in specific containers with perforated lids to minimize evaporation. First group of 30 larvae were placed on petri dish containing sodium chloride as a negative reinforcer or fructose as a positive reinforcer and AM as an olfactory stimulator. After 5 min, these larvae were transferred to pure agarose petri dish containing BA. A second group of larvae received the reciprocal training. After three cycles of training, larvae were transferred to another agarose plate before further testing. Memory was tested by placing larvae onto test plates that contain positive or negative gustatory reinforcers with AM and BA placed on opposite sides. After 5 min, individuals on both sides were counted and preference indexes (PREF) and

performance indexes (PI) were calculated as mentioned below. Negative PI numbers indicate aversive associative learning, whereas positive PI numbers reflect appetitive associative learning:

1. (1a) $PREF_{AM+/BA}$ = (# AM—# BA) / # TOTAL
2. (1b) $PREF_{AM/BA+}$ = (# AM—# BA) / # TOTAL
3. (2) PI = ($PREF_{AM+/BA}$–$PREF_{AM/BA+}$) / 2test

Climbing Assay. Motor function in *Drosophila* was quantified as shown previously[16,17]. Briefly, flies were collected and aged at 25℃ for three weeks before testing. Using a negative geotaxis assay, where flies (10 per vial) were challenged to climb a vertical height of 10cm in an empty fly food vial. This trial is repeated 10 times per vial with a gap of one minute between trials and approximately 150 flies were analyzed for each genotype. The percentage of flies that crossed the 10cm mark was determined and plotted as % climbing pass rate.

References to the Supplemental Information

1. Wiel, L. *et al.* MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat* **40**, 1030-1038 (2019).
2. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
3. Braun, J.E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol Cell* **44**, 120-33 (2011).
4. Primer3web. *http://primer3.ut.ee/ (Accessed: 13th June 2019)* (2019).
5. Gaudet, M., Fara, A.-G., Beritognolo, I. & Sabatti, M. . Allele-specific PCR in SNP genotyping. in Single Nucleotide Polymorphisms. . *Methods in Molecular Biology (ed. Komar, A.)* (2009).
6. de Brouwer, A.P., van Bokhoven, H. & Kremer, H. Comparison of 12 reference genes for normalization of gene expression levels in Epstein-Barr virus-transformed lymphoblastoid cell lines and fibroblasts. *Mol Diagn Ther* **10**, 197-204 (2006).
7. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
8. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
9. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).
10. Karczewski. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv* doi:10.1101/531210, 39 (2019).
11. GTEX Portal. *GTEX project* Available at: *https://gtexportal.org/home/.* (Accessed: 23rd April 2019).
12. Ejima, A. & Griffith, L.C. Measurement of Courtship Behavior in Drosophila melanogaster. *CSH Protoc* **2007**, pdb prot4847 (2007).
13. Ejima, A. & Griffith, L.C. Assay for courtship suppression in Drosophila. *Cold Spring Harb Protoc* **2011**, pdb prot5575 (2011).
14. Apostolopoulou, A.A., Widmann, A., Rohwedder, A., Pfitzenmaier, J.E. & Thum, A.S. Appetitive associative olfactory learning in Drosophila larvae. *J Vis Exp* (2013).
15. Widmann, A. *et al.* Genetic Dissection of Aversive Associative Olfactory Learning and Memory in Drosophila Larvae. *PLoS Genet* **12**, e1006378 (2016).
16. Collart, M.A. The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip Rev RNA* **7**, 438-54 (2016).
17. Kamachi, Y. & Kondoh, H. Sox proteins: regulators of cell fate specification and differentiation. *Development* **140**, 4129-4144 (2013).