2019

CTCF variants in 39 individuals with a variable neurodevelopmental disorder broaden the mutational and clinical spectrum

Enrico D. H. Konrad

Marwan Shinawi

Emily Fassi

et al.

ARTICLE | Genetics inMedicine

Open

CTCF variants in 39 individuals with a variable neurodevelopmental disorder broaden the mutational and clinical spectrum

A full list of authors and affiliations appears at the end of the paper.

Purpose: Pathogenic variants in the chromatin organizer *CTCF* were previously reported in seven individuals with a neurodevelopmental disorder (NDD).

Methods: Through international collaboration we collected data from 39 subjects with variants in *CTCF*. We performed transcriptome analysis on RNA from blood samples and utilized *Drosophila melanogaster* to investigate the impact of *Ctcf* dosage alteration on nervous system development and function.

Results: The individuals in our cohort carried 2 deletions, 8 likely gene-disruptive, 2 splice-site, and 20 different missense variants, most of them de novo. Two cases were familial. The associated phenotype was of variable severity extending from mild developmental delay or normal IQ to severe intellectual disability. Feeding difficulties and behavioral abnormalities were common, and variable other findings including growth restriction and cardiac defects were observed. RNA-sequencing in five individuals identified 3828 deregulated genes enriched for known NDD genes and biological processes such as transcriptional regulation. *Ctcf* dosage alteration in *Drosophila* resulted in impaired gross neurological functioning and learning and memory deficits.

Conclusion: We significantly broaden the mutational and clinical spectrum of *CTCF*-associated NDDs. Our data shed light onto the functional role of CTCF by identifying deregulated genes and show that Ctcf alterations result in nervous system defects in *Drosophila*.

Genetics in Medicine (2019) 21:2723–2733; https://doi.org/10.1038/s41436-019-0585-z

Keywords: CTCF; neurodevelopmental disorders; *Drosophila melanogaster*; intellectual disability; chromatin organization

INTRODUCTION

CCCTC-binding factor CTCF is one of the most important chromatin organizers in vertebrates. It is crucial for orchestrating the three-dimensional chromatin structure by intra- and interchromosomal loop formation and by contributing to the organization of topologically associated domains.[1,2] Additionally, it is involved in many chromatin regulating processes, including gene regulation, blocking of heterochromatin spreading, imprinting,[3] and X-inactivation.[4] CTCF is able to bind various DNA motifs with different specific combinations of its 11 C2H2-type zinc fingers[5] and has over 30,000 binding sites in the human genome.[6]

In 2013 Gregor et al. identified a larger de novo deletion (280 kb, 8 genes), two intragenic likely gene-disruptive (LGD) variants (c.375dupT, p.(Val126Cysfs*14); c.1186dupA, p.[Arg496Lysfs*13]) and one missense variant (c.1699C > T, p.[Arg567Trp]) in *CTCF* in four individuals with variable intellectual disability (ID), microcephaly, and growth retardation[7] (mental retardation, autosomal dominant 21, MRD21, MIM 615502). In three of these individuals RNA-sequencing was performed, which revealed a broad deregulation of genes involved in cellular response to extracellular stimuli.[7] Since that initial report, only a few other case studies on intragenic pathogenic *CTCF* variants have been published. A de novo frameshift variant was identified in a girl with developmental delay, short stature, severe microcephaly, heart defects, and various other anomalies,[8] and very recently, two frameshifting and the same missense variant as in Gregor et al.[7] were reported in three Chinese individuals with ID, feeding difficulties, and microcephaly.[9] Additionally, two larger deletions of 1.1 and 1.6 Mb, respectively, encompassing *CTCF*, were identified in individuals with developmental delay and growth impairment.[10] Further studies reported the identification of *CTCF* variants in large cohorts with various clinical indications but did not provide detailed clinical information.[11–13]

We now report on 39 additional individuals with variants in *CTCF*, further delineating the mutational and clinical spectrum of *CTCF*-related neurodevelopmental disorders (NDD). By RNA-sequencing we confirm a broad deregulation

Correspondence: Christiane Zweier (christiane.zweier@uk-erlangen.de)
These authors contributed equally: Hilde Van Esch, Christiane Zweier

Submitted 3 April 2019; accepted: 6 June 2019
Published online: 26 June 2019

of genes in five affected individuals. By utilizing *Drosophila melanogaster* we could demonstrate that Ctcf is crucial for learning and memory in a dosage-sensitive manner, thus confirming loss of function or haploinsufficiency as the driving force in *CTCF*-related NDDs.

MATERIALS AND METHODS

Patients and patient material

Personal communication with colleagues following the initial report,[7] searching the DECIPHER database,[14,15] and using GeneMatcher[16] enabled us to collect clinical and mutational details on 39 individuals with variants in *CTCF*. Testing (for individual methods see Table S1) in collaborating centers was performed either in the setting of routine diagnostic testing without the requirement for institutional ethics approval or within research settings approved by the ethical review board of the respective institutions (University Erlangen-Nuremberg, University of Tartu, UZ Leuven, Northern Ostrobothnia Hospital District, Cambridge South REC, Republic of Ireland REC; see also Table S1). Informed consent for publication of patient photos and for testing and for publication of mutational and clinical data was obtained from the individuals, or their parents or legal guardians. RNA from blood lymphocytes was collected with the PAXgene system (PreAnalytiX, BD and Qiagen, Hombrechtikon, Switzerland) from three individuals with missense variants and two individuals with LGD variants plus from two individuals with unclear splice-site or nonsense variants. RNA samples from four healthy female and four healthy male young adults were used as controls.

Structural analysis for selected variants was performed as described in the Supplementary Methods.

Transcriptome analysis

Quality control of RNA samples was performed using a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). Library preparation was performed using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA).

Libraries were subjected to single-end sequencing (101 bp) on a HighSeq-2500 platform (Illumina). The obtained reads were converted to .fastq format and demultiplexed using bcl2fastq v2.17.1.14. Quality filtering was performed using cutadapt v. 1.15;[17] then reads were mapped against the human reference genome (Ensembl GRCh37, release 87) using the STAR aligner v. 2.5.4a,[18] and a STAR genome directory created by supplying the Ensembl gtf annotation file (release 87). Read counts per gene were obtained using featureCounts program v. 1.6.1[19] and the Ensembl gtf annotation file. Subsequent analyses were performed using R version 3.5.0.[20] In particular, differential expression analysis was performed with the DESeq2 package v.1.20.0.[21] Additionally, the lfc shrink function of the apeglm package was used to control for expression changes in lowly expressed genes.[22] Gene lists were filtered for protein coding genes with a base mean count of at least 10. Differentially expressed (DE) genes were identified by filtering for padj ≤0.01. DE genes with a log2foldchange >0 or <0 were considered as "upregulated" and "downregulated," respectively. Enrichment of GO Terms among DE genes was analyzed utilizing the PANTHER version 14.0[23] enrichment analysis tool on the Gene Ontology[24,25] homepage with the following settings: test: "Fisher's Exact"; correction: "Calculate false discovery rate"; and GO-Slim terms as annotation data set. A list of all expressed genes within controls and the affected individuals was used as background.

Immunofluorescence and fractionation

Immunofluorescence and fractionation to assess subcellular localization of CTCF are described in detail in the Supplementary Material and Methods.

Fly lines and conditions

Flies were kept on standard food containing cornmeal, sugar, yeast, and agar. Tissue-specific knockdown or overexpression was attained with the UAS/GAL4 system.[26] Crosses were carried out at 28 °C to induce knockdown and at 25 °C to induce overexpression (and knockdown for the courtship conditioning assay). The *Ctcf*-RNAi line (VDRC 108857/KK) and the respective control (VDRC 60100) were obtained from the Vienna Drosophila Resource Center (VDRC, Vienna, Austria),[27] as well as the control (VDRC 60000) used for the hypomorphic *Ctcf* mutant line (BL 21162). This line and the GAL4-driver stocks (BL 4414, *actin*-GAL4 for ubiquitous expression; BL 8816, *D42*-GAL4 for motoneurons; BL 8184, *DJ757*-GAL4 expressed in muscle; BL 7415, *repo*-GAL4 expressed in glia; BL 8765, *elav*-GAL4/Cyo expressed pan-neuronal) and the control line for UAS-*Ctcf* (BL 24749) were obtained from the Bloomington *Drosophila* Stock Center (BDSC, Bloomington, IN, USA). The *Ctcf* mutant line was isogenized by backcrossing it with vdrc 60000 for at least 7 generations. UAS-*Ctcf* was obtained from the Zurich ORFeome Project (FlyORF, Zurich, Switzerland).[28] The driver lines for class IV dendritic arborization (da) neurons (UAS-*Dcr*-2;*477*-GAL4,UAS-*mCD8::GFP*;*ppk*-GAL4/Tm3sb), mushroom body (UAS-Dcr-2;247-GAL4), and pan-neuronal expression (*elav*-GAL4;*elav*-GAL4) were assembled from colleagues. We confirmed *Ctcf* knockdown and overexpression by quantitative reverse transcription polymerase chain reaction (RT-PCR) (Fig. S1).

Analyses of larval neuromuscular junctions and dendritic arborization neurons

Analysis of type 1b neuromuscular junctions (NMJs) of larval muscle 4 as well as analysis of dendritic arborization (da) neurons were performed as described previously[29] (for detailed description, see Supplementary Methods).

Bang sensitivity and climbing assay

Climbing behavior and bang sensitivity testing was performed as described elsewhere[29] and as in more detail in the Supplementary Methods upon pan-neuronal (*elav*-GAL4), motoneuronal (*D42*-GAL4), and/or glial (*repo*-GAL4) manipulation.

Courtship conditioning paradigm

The courtship conditioning paradigm assays were performed as described previously[29] with the mutant *Ctcf* line and upon mushroom body (UAS-*Dcr*-2;*247*-GAL4) specific knockdown or overexpression of *Ctcf*. Flies were kept at 25 °C and 70% humidity at a 12:12 light–dark cycle. Virgin males were trained individually by pairing them with premated females. Learning and short-term memory were tested immediately or 1 hour after a training period of 1 hour, respectively. Long-term memory was tested 24 hours after a training period of 14 hours. The courtship index (CI), the percentage of time each male spent courting a nonreceptive female, was manually assessed from 10-minute movies. By comparing the CI of naïve and trained males a learning index was calculated: LI = (CInaive − CItrained) / CInaive. Differences between LIs of control and mutant flies were statistically compared by a randomization test with 10,000 bootstrap replicates with a custom R script.[30]

RESULTS

Mutational spectrum encompasses large deletions, likely gene-disruptive, and missense variants

The identified aberrations in *CTCF* include two large deletions in 16q22.1, encompassing *CTCF* plus 26 (arr [GRCh37] 16q22.1[67588276_68173459]×1) or 43 (arr [GRCh37] 16q22.1[67345851_68899546]×1) neighboring genes, respectively. In addition, we identified six frameshifting variants (including a deletion of exon 8); two nonsense variants; two variants in the splice acceptor site of exon 4, for one of which in-frame deletion of exon 4 was demonstrated (Fig. S2a,b); and 20 different missense variants (Fig. **1**, Table S1). Regarding the latter, seven amino acid residues were recurrently affected: p.Arg342 was twice exchanged to Cys and to His in three individuals from one family; p.Arg368 was altered twice to Cys and once to His; p.His373 once to Asp and once to Pro; and p.(Arg377His), p.(Pro378Leu), and p.(Arg448Gln) were observed twice, respectively (Fig. **1**, Table S1). p.(Arg567Trp) has been reported in two individuals previously.[7,9] All missense variants were located in exons encoding one of the 11 zinc fingers. Of note, the majority (6/8) of LGD variants were located in exons 3 and 4, just upstream of or at the start of the zinc finger domains. Most variants were shown to be de novo. In six cases one or both parents were not available for testing, and two cases were familial with the variant being transmitted from a presumably healthy mother to her affected son or from an affected father to his twin children. In the first family, the nonsense variant p.(Arg654*) was located rather C-terminal, therefore we confirmed nonsense-mediated messenger RNA (mRNA) decay by RT-PCR (Fig. S2c). Language barrier prevented a proper evaluation of the mother's cognitive capabilities, and though not indicated by the read distribution in DNA from blood, a tissue-specific mosaicism in her could not be excluded. In the second family, structural analysis of the p.(Arg342His) variant indicated a negative effect on CTCF stability, mildly weaker compared with the p.(Arg342Cys) variant observed de novo in two individuals (Fig. S3a–c). One missense variant was mosaic with an alternative allele frequency of approximately 13%. Apart from the previously published p.(Arg567Trp) variant, all missense variants were novel, affected highly conserved amino acids, and were predicted to be deleterious by several in vitro prediction programs (Table S1). The p.(Arg368His) variant occurred once in gnomAD, while all other variants were absent. Several residues of identified missense variants, i.e., p.(Arg278Leu), p.(Asp529Asn), p.(Arg567Trp), carried one other amino acid exchange in gnomAD (Table S1). Comparing the recurrent pathogenic variant p.(Arg567Trp) with p.(Arg567Gln) from gnomAD by structural modeling indicated an only mild impairment of polar interaction with the phosphate backbone of the DNA for the Arg567Gln variant, while the Arg567Trp variant resulted in a significantly reduced binding affinity (Fig. S3d–f).

According to the guidelines of the American College of Medical Genetics and Genomics (ACMG), variants in 36 individuals were considered as pathogenic or likely pathogenic, while three variants (p.[Arg278Leu], p.[Ser360Arg], p.[Asp529Asn]) remained of unknown significance due to an atypical phenotype or lack of segregation testing (Table S1). While structural modeling for the p.(Asp529Asn) variant indicated that it might be tolerated without significantly affecting the CTCF structure (Fig. S3d), it suggested a decreased DNA-binding affinity for the p.(Arg278Leu) variant (Fig. S3g–h) and impaired DNA base pair recognition for the p.(Ser360Arg) variant (Fig. S3i–k).

Clinical spectrum is highly variable

All but one (due to young age) of the 36 affected individuals with intragenic *CTCF* aberrations classified as pathogenic or likely pathogenic presented with developmental delay and/or ID. Cognitive impairment was extremely variable, ranging from learning difficulties, normal IQ, and attending mainstream school or graduating from college in seven individuals to severe ID in three individuals. Walking age ranged from 12 months to three years, and age of first words from 12–18 months to lack of speech at age 12 years. Older individuals with active speech usually spoke in sentences. Intrauterine growth restriction or low birth measurements were reported in ten individuals, and failure to thrive or feeding difficulties, often requiring tube feeding, occurred in 23 individuals. Postnatal short stature was noted in 6 and (borderline) microcephaly in 12 individuals. Tall stature and/or obesity were observed in three individuals. Behavioral anomalies such as autistic features, attention deficit and hyperactivity or aggressivity were common and reported in 24 individuals. Cardiac defects occurred in 11 individuals, and palatal anomalies such as cleft palate or high palate were present in 12 individuals. Conductive and/or sensorineuronal hearing loss were reported in 10 and vision anomalies in 15 individuals. Recurrent urinary tract, airway, or middle ear infections were reported in 14 individuals. Febrile or nonfebrile seizures or teeth anomalies occurred in four

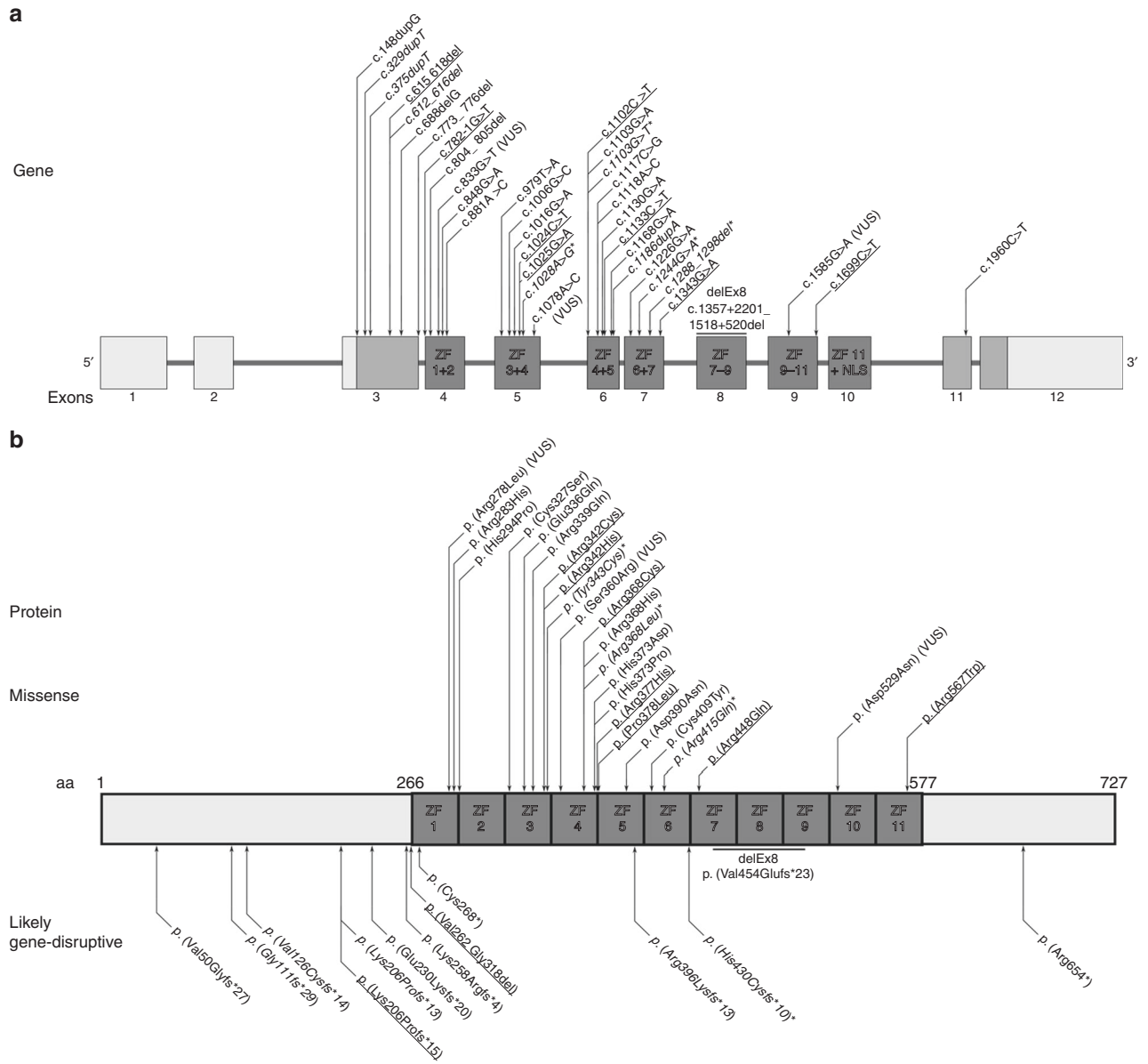

Fig. 1 Variants identified in *CTCF*. **a** Schematic drawing of the *CTCF* gene with known and novel variants. Noncoding exons are displayed in light gray, coding exons in medium gray, and exons encoding zinc fingers in dark gray. **b** Schematic drawing of the CTCF protein. Missense or in-frame variants are displayed above and likely gene-disruptive variants below the scheme. Nonbold variants and p.(Arg567Trp) were published previously with or *without clinical details.[7–9, 11–13] Underlined variants are recurrent. All identified pathogenic or likely pathogenic in-frame or missense variants affect one of the 11 zinc finger domains.

individuals, respectively. Furthermore, nonspecific magnetic resonance image (MRI) anomalies and minor skeletal anomalies were reported (Table 1, Table S1). Minor facial dysmorphisms were frequently noted but did not point to a recognizable, typical facial gestalt (Fig. 2, Table S1). Individuals 1 and 2 with larger deletions presented with a similar phenotype with mild to moderate ID and without growth and other major abnormalities (Table S1). Individual 1 carrying the 1.5-Mb deletion is at high risk of developing gastric cancer because *CDH1* maps within the deleted fragment. No obvious genotype–phenotype correlations between LGD and missense variants or between intragenic variants and larger deletions could be delineated (Table 1).

Individuals 38 and 39 with variants of unknown significance due to lack of parental samples also presented with similar phenotypes, while in contrast, individual 37 with an unclear de novo variant did not show any developmental delay or cognitive impairment, but showed immune deficiency. Although recurrent infections are common in individuals with *CTCF* variants, it remains unclear if the severe proneness to infections in individual 37 might be caused by the *CTCF* variant alone.

Table 1 Main clinical features in individuals with pathogenic or likely pathogenic variants in *CTCF*

	LGD variants $n = 8$	Missense/in-frame variants $n = 26$	Published variants[a] $n = 7$	Total intragenic	Large deletions[b] $n = 5$
IUGR/SGA	2/8	8/25	5/7	37%	0/5
Feeding difficulties/FTT	6/8	17/23	7/7	78%	3/5
Short stature	1/8	5/25	5/7	27%	2–3/5[c]
Microcephaly	3/7	7–9/22	7/7	52%	2–3/5[c]
Walking age	14–32 months	12–36 months	14 months to >3.5 years		23 months to >4.2 years
Age first words	12 months–4 years	12 mo—no speech	12 months to >3.5 years		18 months to >4.2 years
Unspecified DD/ID	0/8	7/25	1/7	20%	0/5
Mild/borderline DD/ID	5/8	9/25	3/7	42%	2/5
Moderate ID	3/8	6/25	1/7	25%	3/5
Severe ID	0/8	3/25	2/7	12%	0/5
Behavioral anomalies	5/6	19/23	4/7	77%	4/5
Seizures	1/8	3/23	0/7	10%	1/5
Cardiac defects	2/6	9/23	3/7	38%	0/5
Palatal anomalies	3/8	9/23	2/7	36%	0/5
Hearing loss	0/8	10/23	0/7	26%	1/5
Vision anomalies	6/8	9/19	6/7	61%	3/5
Recurrent infections	2/7	12/23	3/7	45%	1/5

DD developmental delay, *FTT* failure to thrive, *ID* intellectual disability, *IUGR* intrauterine growth restriction, *LGD* likely gene-disruptive, *SGA* small for gestational age.
[a]See refs. [7–9].
[b]Two patients from this study plus three published ones.[7–10]
[c]One patient was treated with growth hormones.

Missense variants in *CTCF* do not affect its subcellular localization

To investigate a possible effect of missense variants on the subcellular localization and DNA-binding capacities of CTCF, we performed immunofluorescence upon overexpression of eight different mutant *CTCF* constructs in HeLa cells, respectively. All tested variants resulted in a nuclear localization with traces of protein in the cytosol, thus closely resembling wild-type CTCF and indicating that missense variants do not alter gross subcellular localization or distribution (Fig. S4a). Fractionating lymphoblastoid cell lines from two affected individuals into subcellular compartments did not indicate differences in the distribution of mutant CTCF compared with a healthy control (Fig. S4b).

Transcriptome analysis reveals a broad deregulation of genes

RNA-sequencing on RNA from blood cells of five individuals with pathogenic *CTCF* variants and eight healthy controls confirmed similarly decreased *CTCF* expression for the two LGD variants (log2 fold changes = −0.94 and −1.07, respectively), indicating nonsense-mediated mRNA decay as previously confirmed for two other frameshifting variants.[7] In accordance with previous findings,[7] at least two of the three individuals with missense variants had only mildly decreased *CTCF* levels compared with controls (log2 fold changes p.[Arg368Cys]: −0.83; p.[Arg342Cys]: −0.02; p.[Glu336Gln]: −0.2) (Table S2). Affected individuals clustered with each other as did controls regarding differentially expressed genes (Fig. **3a**).

There were 3828 genes that were differentially expressed between affected individuals and controls; 1667 were upregulated and 2161 were downregulated (Table S2). Regarding differentially expressed genes, affected individuals behaved more similarly within their mutational group (likely gene-disruptive variants vs. missense variants) (Fig. **3a**). Nevertheless, we found significant overlap between differentially expressed genes in the two individuals with LGD variants and the three individuals with missense variants ($p < 1 \times 10^{-16}$, hypergeometric test), indicating similar consequences of various types of pathogenic variants on gene regulation.

To investigate the role of differentially expressed genes in neurodevelopmental function and dysfunction, we calculated enrichment of known NDD-associated genes (SysID, status November 2018, Table S2) among the deregulated genes. This was significant for both the upregulated and the downregulated genes ($p = 4.13 \times 10^{-4}$ and 2.2×10^{-16}, respectively, chi-square test).

Downregulated genes were enriched for Gene Ontology (GO) terms such as transcription-related processes and regulation of biological processes (Fig. **3b**, Table S2). Upregulated genes were enriched for translational processes and ribonucleoprotein and ribosomal processes (Fig. **3c**, Table S2).

Fig. 2 Facial gestalt of individuals with pathogenic variants in *CTCF*. Note minor facial dysmorphism such as a long face (e.g., I5, I8, I11, I27), a prominent forehead (e.g., I4, I8, I18, I33, I35) or a bulbous nasal tip and long palpebral fissures in several individuals, but no overall distinct, recognizable facial gestalt.

Ctcf dosage alterations in *Drosophila melanogaster* impair gross neurological functioning and complex learning and memory behavior

Given that *CTCF* haploinsufficiency leads to disease, we utilized *Drosophila melanogaster* as a model organism to investigate the consequences of altered *Ctcf* dosage on nervous system development and function. First, we investigated whether altered levels of *Ctcf* resulted in altered development and morphology of larval neuromuscular junctions (NMJs). These large synapses represent an established model system for vertebrate glutamatergic synapses.[31] We did not observe any significant or consistent alterations regarding NMJ area, length, or number of synaptic boutons or active zones in the mutant, or upon pan-neuronal knockdown or overexpression (Fig. S5a–c). As a significant decrease of dendritic branching has been reported in a conditional Ctcf knockout mouse,[32] we examined the large dendritic arborization (da) neurons in *Drosophila* larvae. We did not observe any significant alterations regarding number or length of dendrites upon knockdown of Ctcf in these neurons compared with controls, while overexpressing Ctcf resulted in a mild increase in total length (Fig. S6). Also testing for bang sensitivity, a model for seizure susceptibility,[33] did not result in aberrant phenotypes upon pan-neuronal knockdown or overexpression or in the mutant condition (Fig. S5d–f).

The impact of *Ctcf* dosage alteration on gross neurological function was addressed with the climbing assay, which is based on the negative geotaxis, an innate behavior to climb up after being tapped down.[34] Flies with knockdown of *Ctcf* in all neurons or specifically in motoneurons showed a significant impairment in their climbing abilities while it was not altered in flies with glial knockdown. Pan-neuronal overexpression of Ctcf did not result in impaired climbing behavior, but overexpression in motoneurons or particularly in glia cells did (Fig. **4a, c**).

Finally, we tested the consequences of *Ctcf* dosage alterations on complex learning and memory processes utilizing the courtship condition paradigm.[35] Ubiquitous low levels of *Ctcf* in the hypomorphic mutant as well as knockdown of *Ctcf* specifically in the mushroom body, the fly

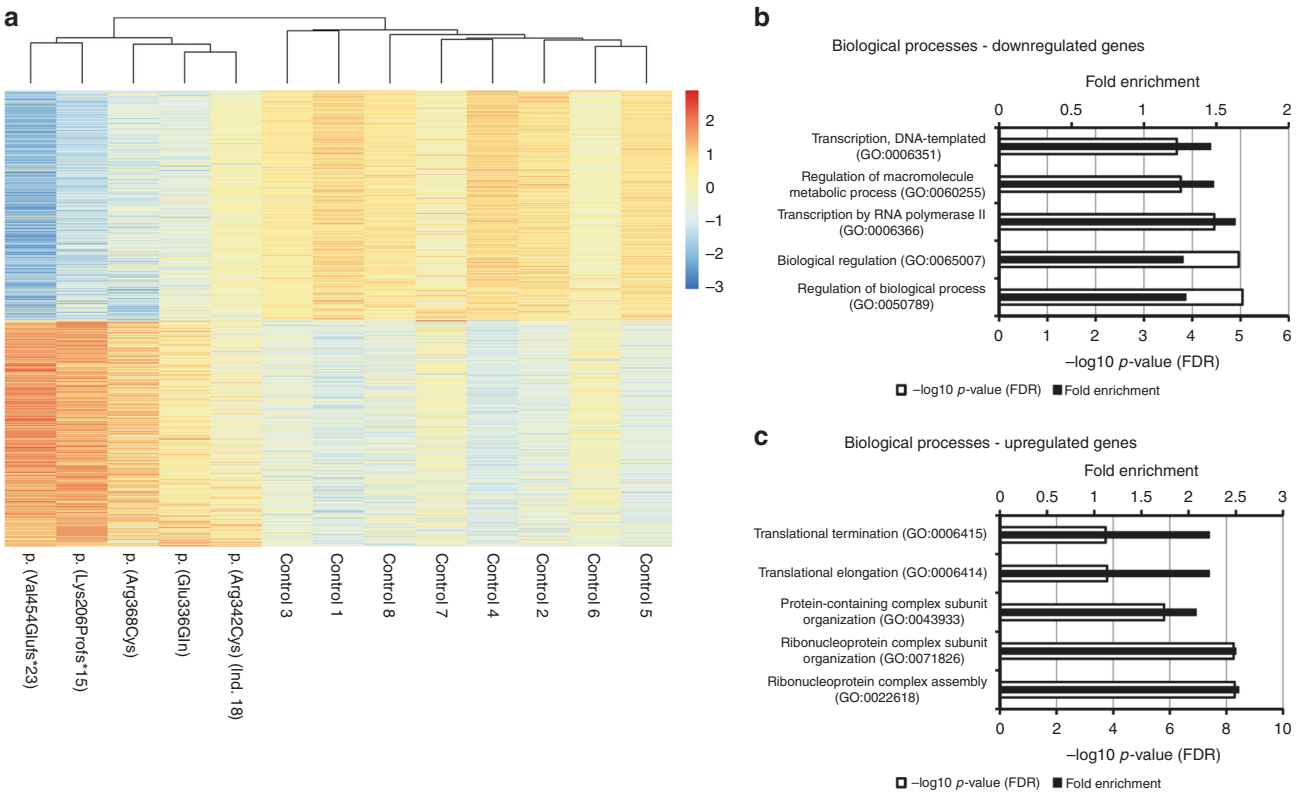

Fig. 3 Transcriptome analysis in *CTCF*-deficient individuals. a Heatmap displaying the 3828 differentially expressed genes between five affected individuals and eight healthy controls and depicting the expression similarity between individuals with corrected *p* value <0.01; 2161 genes were downregulated and 1667 genes were upregulated in the affected individuals. Individuals are clustered by Euclidean distance, and genes are sorted by adjusted *p* value. Blue, downregulated; red, upregulated. **b**, **c** Gene Ontology (GO) term analysis on **b** downregulated and **c** upregulated genes depicting the top five GO terms for biological processes *FDR* false discovery rate.

center for learning and memory, resulted in significantly impaired learning and short-term memory compared with the respective controls. Overexpression of *Ctcf* in the mushroom body did not result in significantly altered learning and short-term memory, but in significantly reduced long-term memory (Fig. 4**d–f**), indicating a crucial role of correct *Ctcf* dosage for complex learning and memory processes.

DISCUSSION

By assembling data on additional 36 individuals with pathogenic or likely pathogenic variants in *CTCF* we considerably contribute to the delineation of the mutational and clinical spectrum of *CTCF*-associated NDDs. The phenotype associated with *CTCF* aberrations is highly variable. All individuals presented with developmental delay and/or a variable degree of learning or cognitive difficulties, extending into (low) normal IQ ranges. The mild end of the phenotypic spectrum is reflected by two familial cases in our cohort. As also true for other, potentially mild NDDs, autosomal dominant inheritance has to be considered, particularly in trio exome approaches when primarily searching for de novo variants. Our observations might also imply that due to a possibly very mild phenotype or incomplete penetrance, very rare presence of a variant in public databases such as ExAC or gnomAD is not necessarily an exclusion criterion for its pathogenicity.

Feeding anomalies and failure to thrive, as well as behavioral anomalies, were the most frequently associated symptoms. In addition, postnatal short stature and microcephaly, as well as a whole range of different birth malformations and anomalies, including cardiac defects, cleft palate, hearing loss, vision anomalies, recurrent infections, and muscular hypotonia were noted. Despite this wide range of anomalies and despite noticing minor facial dysmorphism in many of the affected individuals, we consider the phenotype not sufficiently distinct as to be clinically recognizable.

Previous studies have pointed to haploinsufficiency of *CTCF* as the main disease-causing mechanism.[7,8,10] An earlier hypothesis suggested a possible genotype–phenotype correlation between missense and LGD variants but was based on a very small number of affected individuals.[7] Our extensive follow-up study does not indicate any clear correlation between nature and location of variants and clinical presentation. Furthermore, differentially expressed genes significantly overlapped between individuals with either missense or LGD variants, and we did not observe effects on gene regulation or gross subcellular localization, though

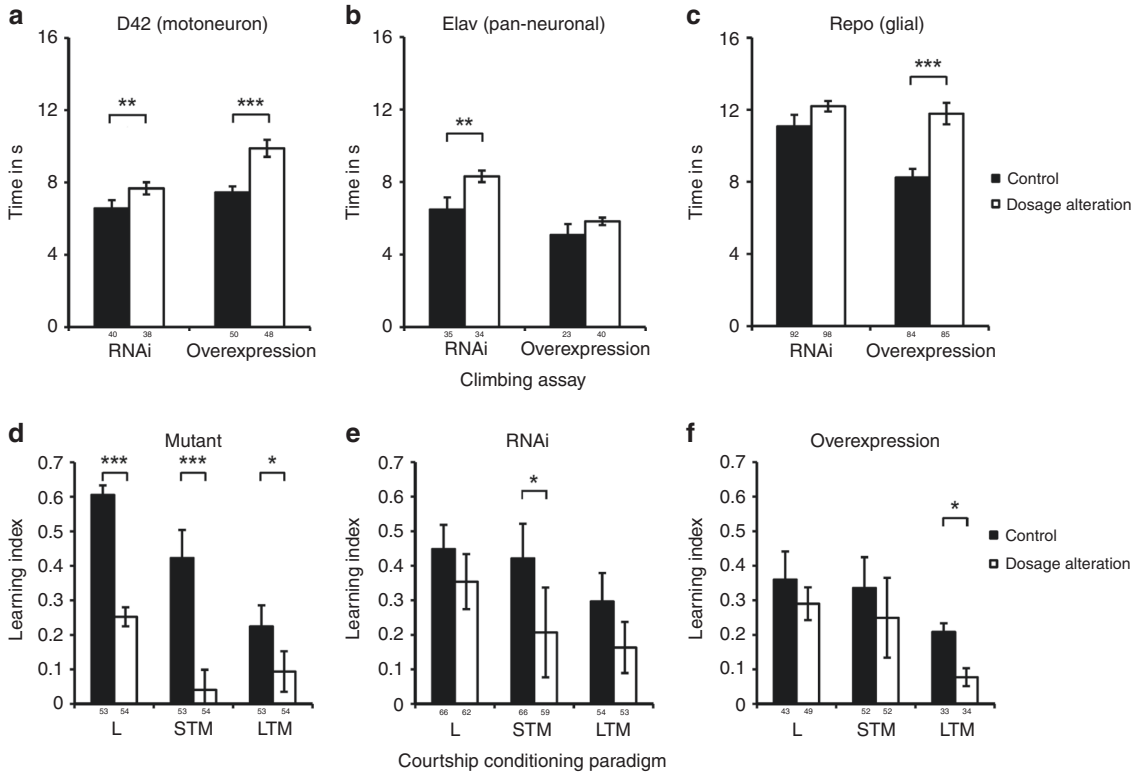

Fig. 4 Impact of *Ctcf* dosage alteration on fly gross neurological functioning and learning and memory. a–c Flies with knockdown of *Ctcf* in **a** motoneurons and **b** all neurons but not in **c** glia showed significant locomotor impairment in the climbing assay. Pan-neuronal overexpression **b** had no effect, but **c** glial and **a** motoneuronal driven overexpression of *Ctcf* resulted in significant locomotor impairment. The numbers below the columns represent the samples in groups of ten flies tested per genotype. **d** In the courtship conditioning paradigm, the hypomorphic *Ctcf* mutant line showed highly significant impairment of learning (L) and short-term memory (STM) and significant impairment of long-term memory (LTM). **e** In accordance, mushroom body–driven (UAS-*Dcr-2*;*247*-GAL4) knockdown of *Ctcf* resulted in significantly reduced short-term memory (STM). Similar tendencies were observed for L and LTM, but not reaching significance. **f** Overexpression of *Ctcf* in the mushroom body did not result in a significant effect on L and STM, but in a significant impairment of long-term memory. Graphs display mean values of the number of animals indicated below the columns per genotype from at least three independent test days. Differences between learning indexes of control and mutant flies were statistically compared by a randomization test with 10,000 bootstrap replicates with a custom R script.[30] Error bars depict SEM. Asterisks indicate statistical significance (*p* value: *≤0.05, **≤0.01, ***≤0.001).

subtle changes would still be possible. Both the previous and all newly identified missense variants are located within several of the 11 zinc finger domains, most likely resulting in loss of function by impaired DNA binding, as also indicated by structural modeling for several variants. Of note, the neurodevelopmental phenotype in individuals with larger deletions encompassing *CTCF* is not markedly different or more severe than in individuals with intragenic missense or LGD variants. This indicates that no other dosage-sensitive genes relevant for neurodevelopment are located nearby and further implies (functional) haploinsufficiency of *CTCF* resulting from all kinds of pathogenic variants as the most likely disease mechanism.

Considering the crucial and extremely broad role of CTCF in chromatin organization and regulation it is still very surprising that its haploinsufficiency or loss of function can result in such a relatively mild phenotype. Nevertheless, transcriptome analysis showed a broad deregulation of genes. In line with a previous report[7] and with gene expression profiles in conditional knockout mice,[32] we detected more downregulated than upregulated genes, further supporting a primarily activating role of CTCF for target genes involved in the pathomechanism of *CTCF*-related disorders. Differentially expressed genes were enriched for biological processes and for general ribosomal and transcriptional processes. They were also enriched for known NDD-associated genes, emphasizing the role of CTCF in neurodevelopmental function and dysfunction. CTCF-deficient mice die in early implantation stages,[36] highlighting its essential role in embryonic development. Conditional knockout in postmitotic projection neurons or in the developing brain resulted in postnatal growth retardation, abnormal behavior, dendritic arborization anomalies, and death within one month or in deficits in neuroprogenitor differentiation resulting in microcephaly and perinatal death, respectively.[32,37] These observations support a crucial role of CTCF in neurodevelopment, but the early lethality prevents investigating the role of CTCF for learning and memory processes in vivo. We chose *Drosophila* as a model because existence of a viable hypomorphic mutant and tissue-specific knockdown and overexpression could circumvent lethality. In contrast to

observations of reduced dendritic arborization and spine numbers in mice,[32,37] we did not observe alterations in *Drosophila* larval dendritic arborization neurons or neuromuscular junctions upon knockdown of *Ctcf* and only mildly significant changes of dendritic length upon overexpression. Such discordant findings might be related to the fact that, in contrast to mammals, *Drosophila* possesses three other insulator binding proteins next to Ctcf that could take over some of the Ctcf functions.[38] This might be the reason why Ctcf is less crucial for setting up genome organization or global gene expression during *Drosophila* embryogenesis and development, despite its many binding sites throughout the fly genome.[39,40] In accordance with a more important role in adult flies[39] we observed climbing deficits upon pan-neuronal or motoneuronal knockdown and upon motoneuronal and glial overexpression. The marked phenotype upon glial overexpression might point to a so-far underestimated role of Ctcf in glia, but would require further experimental follow-up. In accordance with cognitive deficits in humans with pathogenic *CTCF* variants, we observed learning and memory deficits in the hypomorphic mutant line and upon mushroom body–specific knockdown or overexpression. These observations confirm the importance of proper *Ctcf/CTCF* dosage for learning and memory processes.

In summary, our study extensively broadens the clinical and mutational spectrum of *CTCF*-related NDDs. A high number of deregulated genes from transcriptome analyses and gross neurological as well as learning and memory deficits in *Drosophila melanogaster* provide insights into the role of proper CTCF/Ctcf dosage for postnatal neurological and cognitive function and dysfunction.

URLs

Gene Ontology http://geneontology.org/ Homologene https://www.ncbi.nlm.nih.gov/homologene OMIM https://www.omim.org/ PANTHER http://pantherdb.org/ Primer3 http://primer3.ut.ee/ SysID database https://sysid.cmbi.umcn.nl/ UCSC Genome Browser https://genome.ucsc.edu/

SUPPLEMENTARY INFORMATION

The online version of this article (https://doi.org/10.1038/s41436-019-0585-z) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

We thank all individuals and families for participating in this study. We especially thank Laila Distel and Christine Suchy for excellent technical assistance; André Reis, Arif Ekici, and Fulvia Ferrazzi at the next-generation sequencing core facility at the Institute of Human Genetics in Erlangen; and Felix Engel for help with the confocal microscope, which was supported by the German Research Foundation (INST 410/91-1 FUGG). C.Z. is supported by grants from the German Research Foundation (ZW184/1-2, ZW184/3-1, and 270949263/GRK2162) and by the Interdisciplinary Center for Clinical Research in Erlangen (E26 and ELAN-Fonds). H.V.E. is a clinical investigator of FWO Vlaanderen. K.Õ. and S.P. received support from Estonian Research Council grants PUT355, PRG471, and PUTJD827. M.H.W. is supported by T32GM007748. This study makes use of data generated by the DECIPHER community. Funding for the project was provided by the Wellcome Trust. The DDD Study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003), a parallel funding partnership between Wellcome and the Department of Health, and the Wellcome Sanger Institute (grant number WT098051). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. The Broad Center for Mendelian Genomics (UM1 HG008900) is funded by the National Human Genome Research Institute with supplemental funding provided by the National Heart, Lung, and Blood Institute under the Trans-Omics for Precision Medicine (TOPMed) program and the National Eye Institute. Please also see Supplementary Acknowledgements.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. Gerasimova TI, Byrd K, Corces VG. A chromatin insulator determines the nuclear localization of DNA. Mol Cell. 2000;6:1025–1035.
2. Ji X, Dadon DB, Powell BE, et al. 3D chromosome regulatory landscape of human pluripotent cells. Cell Stem Cell. 2016;18:262–275.
3. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000;405:482–485.
4. Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT. CTCF, a candidate trans-acting factor for X-inactivation choice. Science. 2002;295:345–347.
5. Klenova EM, Nicolas RH, Paterson HF, et al. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. Mol Cell Biol. 1993;13:7612–7624.
6. Bao L, Zhou M, Cui Y. CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. Nucleic Acids Res. 2008;36(Database issue):D83–87.
7. Gregor A, Oti M, Kouwenhoven EN, et al. De novo mutations in the genome organizer CTCF cause intellectual disability. Am J Hum Genet. 2013;93:124–131.
8. Bastaki F, Nair P, Mohamed M, et al. Identification of a novel CTCF mutation responsible for syndromic intellectual disability—a case report. BMC Med Genet. 2017;18:68.
9. Chen F, Yuan H, Wu W, et al. Three additional de novo CTCF mutations in Chinese patients help to define an emerging neurodevelopmental disorder. Am J Med Genet C Semin Med Genet. 2019;181:218–225.
10. Hori I, Kawamura R, Nakabayashi K, et al. CTCF deletion syndrome: clinical features and epigenetic delineation. J Med Genet. 2017;54:836–842.
11. Meng L, Pammi M, Saronwala A, et al. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. JAMA Pediatr. 2017;171:e173438.
12. Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18:696–704.
13. Willsey AJ, Fernandez TV, Yu D, et al. De novo coding variants are strongly associated with Tourette disorder. Neuron. 2017;94:486–99 e489.
14. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–438.

15. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84:524–533.
16. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36:928–930.
17. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetJournal. 2011;17:10–12.
18. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
19. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–930.
20. R Core Team. R: a language and environment for statistical computing. Vienna, Austria.: R Foundation for Statistical Computing; 2015.
21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
22. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2018 Nov 3; https://doi.org/10.1093/bioinformatics/bty895 [Epub ahead of print].
23. Mi H, Huang X, Muruganujan A, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 2017;45(D1):D183–D189.
24. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–29.
25. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic Acids Res. 2017;45(D1):D331–D338.
26. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993;118:401–415.
27. Dietzl G, Chen D, Schnorrer F, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature. 2007;448:151–156.
28. Bischof J, Bjorklund M, Furger E, Schertel C, Taipale J, Basler K. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development. 2013;140:2434–2442.
29. Straub J, Konrad EDH, Gruner J, et al. Missense variants in RHOBTB2 cause a developmental and epileptic encephalopathy in humans, and altered levels cause neurological defects in Drosophila. Am J Hum Genet. 2018;102:44–57.
30. Gregor A, Kramer JM, van der Voet M, et al. Altered GPM6A/M6 dosage impairs cognition and causes phenotypes responsive to cholesterol in human and Drosophila. Hum Mutat. 2014;35:1495–1505.
31. Koh YH, Gramates LS, Budnik V. Drosophila larval neuromuscular junction: molecular components and mechanisms underlying synaptic plasticity. Microsc Res Tech. 2000;49:14–25.
32. Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. Cell Rep. 2012;2:345–357.
33. Kuebler D, Tanouye MA. Modifications of seizure susceptibility in Drosophila. J Neurophysiol. 2000;83:998–1009.
34. Palladino MJ, Hadley TJ, Ganetzky B. Temperature-sensitive paralytic mutants are enriched for those causing neurodegeneration in Drosophila. Genetics. 2002;161:1197–1208.
35. Siegel RW, Hall JC. Conditioned responses in courtship behavior of normal and mutant Drosophila. Proc Natl Acad Sci USA. 1979;76:3430–3434.
36. Moore JM, Rabaia NA, Smith LE, et al. Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos. PLoS ONE. 2012;7:e34915.
37. Watson LA, Wang X, Elbert A, Kernohan KD, Galjart N, Berube NG. Dual effect of CTCF loss on neuroprogenitor differentiation and survival. J Neurosci. 2014;34:2860–2870.
38. Moon H, Filippova G, Loukinov D, et al. CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. EMBO Rep. 2005;6:165–170.
39. Gambetta MC, Furlong EEM. The insulator protein CTCF is required for correct Hox gene expression, but not for embryonic development in Drosophila. Genetics. 2018;210:129–136.
40. Schwartz YB, Linder-Basso D, Kharchenko PV, et al. Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res. 2012;22:2188–2198.

Enrico D. H. Konrad, MS[1], Niels Nardini, MS[1], Almuth Caliebe, MD[2], Inga Nagel, PhD[2,3], Dana Young, MS[4], Gabriella Horvath, MD PhD[4], Stephanie L. Santoro, MD[5], Christine Shuss, MS[5], Alban Ziegler, MD[6], Dominique Bonneau, MD PhD[6], Marlies Kempers, MD PhD[7], Rolph Pfundt, PhD[7], Eric Legius, MD PhD[8], Arjan Bouman, MD[9], Kyra E. Stuurman, MD[9], Katrin Õunap, MD PhD[10,11,12], Sander Pajusalu, MD PhD[10,11,13], Monica H. Wojcik, MD[12,14], Georgia Vasileiou, PhD[1], Gwenaël Le Guyader, MD PhD[15], Hege M. Schnelle, MD[16], Siren Berland, MD[16], Evelien Zonneveld-Huijssoon, MD PhD[17], Simone Kersten, MS[7], Aditi Gupta, PhD[18], Patrick R. Blackburn, PhD[19], Marissa S. Ellingson, MS[19], Matthew J. Ferber, PhD[19], Radhika Dhamija, MBBS[20], Eric W. Klee, PhD[18], Meriel McEntagart, MD[21], Klaske D. Lichtenbelt, MD[22], Amy Kenney, MS[23], Samantha A. Vergano, MD[23], Rami Abou Jamra, MD[24], Konrad Platzer, MD[24], Mary Ella Pierpont, MD PhD[25], Divya Khattar, MD[26], Robert J. Hopkin, MD[26], Richard J. Martin, MD[27], Marjolijn C. J. Jongmans, MD PhD[7], Vivian Y. Chang, MD MS[28,29], Julian A. Martinez-Agosto, MD PhD[30,31], Outi Kuismin, MD[32,33,34,35], Mitja I. Kurki, PhD[32,36,37], Olli Pietiläinen, PhD[37,38], Aarno Palotie, MD PhD[32,36,37,39,40],

Timothy J. Maarup, MD[41], Diana S. Johnson, BM MD[42], Katja Venborg Pedersen, MD PhD[43], Lone W. Laulund, MD[44], Sally A. Lynch, MD[45], Moira Blyth, MBChB DM[46], Katrina Prescott, MBBS PhD[46], Natalie Canham, MB, ChB[47], Rita Ibitoye, MD[47], Eva H. Brilstra, MD[22], Marwan Shinawi, MD[48], Emily Fassi, MS[48], DDD Study[49], Heinrich Sticht, PhD[50], Anne Gregor, PhD[1], Hilde Van Esch, MD PhD[51] and Christiane Zweier, MD PhD[1]

[1]Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany. [2]Institute of Human Genetics, Universitätsklinikum Schleswig Holstein Campus Kiel and Christian-Albrechts-Universität, Kiel, Germany. [3]Institute of Experimental and Clinical Pharmacology, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany. [4]Adult Metabolic Diseases Clinic, Vancouver General Hospital, Vancouver, BC, Canada. [5]Division of Molecular and Human Genetics, Nationwide Children's Hospital, Columbus, OH, USA. [6]Département de Biochimie et Génétique, CHU Angers et Mitolab INSERM 1083—CNRS 6015, Angers, France. [7]Department of Human Genetics, Radboud University Medical Center and Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands. [8]Department of Human Genetics, KU Leuven and Center for Human Genetics, University Hospital Leuven, KU Leuven, Leuven, Belgium. [9]Department of Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. [10]Department of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia. [11]Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia. [12]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [13]Department of Genetics, Yale University School of Medicine, New Haven, CT, USA. [14]Divisions of Genetics and Genomics and Newborn Medicine, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA. [15]Service de Génétique Clinique, CHU de Poitiers, Poitiers, France. [16]Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway. [17]Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [18]Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA. [19]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. [20]Department of Clinical Genomics, Mayo Clinic, Scottsdale, AZ, USA. [21]South West Thames Regional Genetics Centre, St. George's Healthcare NHS Trust, St. George's, University of London, London, UK. [22]Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. [23]Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Norfolk, VA, USA. [24]Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany. [25]Department of Pediatrics, Division of Genetics and Metabolism, University of Minnesota, Minneapolis, MN, USA. [26]Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [27]The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [28]Division of Pediatric Hematology-Oncology, Department of Pediatrics, David Geffen School of Medicine, Los Angeles, CA, USA. [29]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA, USA. [30]Division of Medical Genetics, Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. [31]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. [32]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [33]PEDEGO Research Unit, University of Oulu, Oulu, Finland. [34]Medical Research Center, Oulu University Hospital, University of Oulu, Oulu, Finland. [35]Department of Clinical Genetics, Oulu University Hospital, Oulu, Finland. [36]Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [37]The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [38]Department of Stem Cell and Regenerative Biology, University of Harvard, Cambridge, MA, USA. [39]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [40]Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. [41]Department of Genetics, Kaiser Permanente, Los Angeles, CA, USA. [42]Sheffield Children's Hospital, Sheffield, UK. [43]Department of Clinical Genetics, Odense University Hospital, Odense, Denmark. [44]Department of Pediatrics, Odense University Hospital, Odense, Denmark. [45]University College Dublin and Temple Street Children's Hospital, Dublin, Ireland. [46]Yorkshire Regional Genetics Service, Chapel Allerton Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK. [47]North West Thames Regional Genetics Service, Northwick Park Hospital, Harrow, UK. [48]Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA. [49]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [50]Institute of Biochemistry, Emil-Fischer Center, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany. [51]Center for Human Genetics, University Hospital Leuven, KU Leuven, Leuven, Belgium