1	Altered gene expression profiles impair the nervous system development in
2	individuals with 15q13.3 microdeletion
3	
4 5 6 7 8	Marek B. Körner <sup>1, 2</sup> , Akhil Velluva <sup>1, 3</sup> , Linnaeus Bundalian <sup>2</sup> , Maximilian Radtke <sup>2</sup> , Chen-Ching Lin <sup>4</sup> , Pia Zacher <sup>2, 5</sup> , Tobias Bartolomaeus <sup>2</sup> , Anna Kirstein <sup>6</sup> , Achmed Mrestani <sup>1, 7</sup> , Nicole Scholz <sup>1</sup> , Konrad Platzer <sup>2</sup> , Anne-Christin Teichmann <sup>2</sup> , Julia Hentschel <sup>2</sup> , Tobias Langenhan <sup>1</sup> , Johannes R. Lemke <sup>2</sup> , Antje Garten <sup>6</sup> , Rami Abou Jamra <sup>2*#</sup> , Diana Le Duc <sup>2, 3*#</sup>
9 10	<sup>1</sup> Rudolf Schönheimer Institute of Biochemistry, Division of General Biochemistry, Medical Faculty, Leipzig University, 04103 Leipzig, Germany
11	<sup>2</sup> Institute of Human Genetics, University of Leipzig Medical Center, 04103 Leipzig, Germany
12 13	<sup>3</sup> Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103, Leipzig, Germany
14 15	<sup>4</sup> Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, 11221 Taipei, Taiwan
16	<sup>5</sup> Epilepsy Center Kleinwachau, 01454 Radeberg, Germany
17 18	<sup>6</sup> Pediatric Research Center, University Hospital for Children and Adolescents, Leipzig University, 04103 Leipzig, Germany.
19	<sup>7</sup> Department of Neurology, University of Leipzig Medical Center, 04103 Leipzig, Germany
20	
21	* Equal contribution
22	<sup>#</sup> Correspondence to <u>diana_leduc@eva.mpg.de</u> and <u>Rami.AbouJamra@medizin.uni-leipzig.de</u>
23	
24	

## 25 Abstract

Background: The 15q13.3 microdeletion has pleiotropic effects ranging from
apparently healthy to severely affected individuals. The underlying basis of the variable
phenotype remains elusive.

Methods: We analyzed gene expression using blood from 3 individuals with 15q13.3 microdeletion and brain cortex tissue from 10 mice Df[h15q13]/+. We assessed differentially expressed genes (DEGs), protein-protein interaction (PPI) functional modules, and gene expression in brain developmental stages.

33 **Results:** The deleted genes' haploinsufficiency was not transcriptionally compensated, 34 suggesting a dosage effect may contribute to the pathomechanism. DEGs shared between 35 tested individuals and a corresponding mouse model show a significant overlap including 36 genes involved in monogenic neurodevelopmental disorders. Yet, network-wide 37 dysregulatory effects suggest the phenotype is not caused by a singular critical gene. A 38 significant proportion of blood DEGs, silenced in adult brain, have maximum expression 39 during the prenatal brain development. Based on DEGs and their PPI partners we identified 40 altered functional modules related to developmental processes, including nervous system 41 development.

42 **Conclusions:** We show that the 15q13.3 microdeletion has a ubiquitous impact on the 43 transcriptome pattern, especially dysregulation of genes involved in brain development. The 44 high phenotypic variability seen in 15q13.3 microdeletion could stem from an increased 45 vulnerability during brain development, instead of a specific pathomechanism.

Keywords: 15q13.3; Copy number variants; Transcriptomics; Protein-protein
interaction networks; Nervous System Development.

### 48 Introduction

49 Individuals with 15q13.3 microdeletion (OMIM #612001) show clinical manifestations ranging from no obvious symptoms to severe intellectual disability, 50 neuropsychiatric disorders, and epilepsy<sup>1</sup> (Fig. 1 A, B). The most common 15q13.3 deletion 51 spans 2 Mb and includes eight RefSeq genes (CHRNA7, FAN1, TRPM1, KLF13, OTUD7A, 52 MTMR10, ARHGAP11B, and MIR211, Fig. 1A)<sup>2</sup>. Many functional and association studies 53 54 inquired which gene(s) encompassed by the deleted region could be responsible for the 55 phenotype. However, the results were as variable as the clinical manifestation and different groups proposed multiple candidates (CHRNA7<sup>3,4</sup>, OTUD7A<sup>5,6</sup>, FAN1<sup>7</sup>, ARHGAP11B<sup>8</sup>, 56 TRPM1<sup>9</sup>, KLF13<sup>10</sup>), to explain the symptoms (Fig. 1A). A mouse model of the 15q13.3 57 58 microdeletion syndrome (Df[h15q13]/+) shows manifestations similar to affected humans 59 ranging from attention deficits to impaired behavior and disrupted prefrontal cortex 60 processing <sup>11</sup>. Thus, the microdeletion effect is stable across species and clearly impairs 61 nervous system function.

62 Recently, it was suggested that many disrupted biological pathways such as Wnt 63 signaling or ribosome biogenesis may be involved in the molecular mechanism underlying the disease rather than singular dosage-affected genes<sup>12</sup>. While Zhang et al. used a multiomics 64 65 approach to identify perturbed biological processes, the multiple employed analyses showed 66 disagreement with respect to the pathomechanism. Also, no shared dysregulated genes were identified between human induced pluripotent stem cells (iPSCs) and mouse cortex <sup>12</sup>. This 67 68 could be a result of the *in vitro* setup and neuronal differentiation protocols, which generally impact gene expression profiles <sup>13</sup>. We thus sought to inquire gene expression profiles in 69 70 subjects with 15q13.3 microdeletion in a native/in vivo state.

One major challenge of transcriptomics in a clinical setting is tissue-specific gene expression <sup>14,15</sup> and the fact that most of the times the only accessible tissue to probe is blood <sup>16</sup>. In our previous work we showed, however, that known genes for neurodevelopmental disorders are not necessarily expressed in the adult brain and that genes which are relevant during embryonic development of the central nervous system can be silenced at a later timepoint <sup>14</sup>. Thus, although not regarded as a representative tissue, blood transcriptomics has the potential to reveal aspects missed in other tissues or iPSCs.

In the present study, we analyzed the changes in gene expression profiles in the blood of three individuals with heterozygous microdeletion 15q13.3 and intellectual disability associated with epilepsy. We identified a significant overlap (*p*-value = 0.02) of 68

differentially expressed genes (DEGs) between humans and mouse (Df[h15q13]/+) cortex. The gene ontology (GO) category most significantly enriched with DEGs was "nervous system development–GO:0007399" and DEGs in blood, which are not expressed in the adult brain, revealed maximum expression levels in the prenatal stage of brain development. The disrupted gene expression profile could lead to an increased vulnerability in the early stages of nervous system development.

87 Materials and Methods

88 *Ethics approval* 

89 This study was approved and monitored by the ethics committee of the University of

90 Leipzig, Germany (224/16-ek and 402/16-ek).

## 91 Chr15q13.2q13.3 microdeletion individuals and mouse model (Df[h15q13]/+)

92 Three individuals with diagnosed heterozygous 15q13.2q13.3 microdeletions were previously described <sup>17</sup>. For all individuals we performed Illumina TruSight One Panel and 93 94 microarray analysis. Patient 1 is a male with а deletion of 2.56 Mbp 95 (15q13.2q13.3(30366247\_32927476)x1). The second patient is female and carries a deletion 96 on chromosome 15 of 1.57 Mbp (15q13.2q13.3(30936285\_32514341)x1). Patient 3 is a male, 97 has a deletion of 2.14 Mbp on chromosome 15 (15q13.2q13.3(30371774 32514341)x1). He 98 was additionally diagnosed with a maternally inherited splicing-variant in the remaining 99 TRPM1 allele and a heterozygous de novo point mutation in MITF. Pathogenic variants in 100 MITF gene cause albinism, which was also clinically diagnosed in this individual. His 101 ophthalmological phenotype (severe myopia, astigmatism, and pendular nystagmus) were 102 clinical symptoms of the autosomal recessive TRPM1 phenotype (one allele being deleted as 103 part of the 15q13.3 microdeletion and the other allele carrying the maternally inherited 104 splicing variant). In sum, he suffered from a complex combined phenotype with neurologic 105 symptoms attributed to the 15q13.3 microdeletion. Blood RNA samples were taken from the 106 three individuals (two males and one female, aged 27-63 years) and four control subjects 107 (two males and two females, aged 20–52 years).

To identify molecular changes which are consistent across species we used the data generated by Gordon and colleagues (GSE129891)<sup>18</sup>. We analyzed transcriptomes from cerebral cortex tissue of mice with heterozygous deletions on mouse chromosome 7qC syntenic to human 15q13.3<sup>11,18</sup>.

### 112 **RNA extraction and sequencing**

RNA was extracted from PAXgene blood samples using PAXgene Blood RNA Kit
(Qiagen). RNA sequencing (RNA-seq) libraries were prepared using TruSeq RNA Library
Prep Kit v2 (Illumina, San Diego, CA) and sequenced on an Illumina NovaSeq platform with
151 bp paired-end reads.

# 117 Differential Gene Expression (DEG) analysis

118 RNA-seq reads were mapped to the human genome assembly hg38 with STAR (version 2.6.1d) <sup>19</sup>. We computed the transcript levels with htseq-count (version 0.6.0) <sup>20</sup>. 119 120 From GSE129891 we analyzed read counts of ten wild type and ten (Df[h15q13]/+) mouse 121 cerebral cortex samples. Genes with a sum of less than 10 reads in all samples together were 122 excluded from further analysis. Differential expression of genes was determined with the R package DESeq2 (version 1.30.1)<sup>21</sup>, which uses the Benjamini-Hochberg method to correct 123 for multiple testing  $^{22}$ . Genes were considered to be significantly differentially expressed if p-124 125 adj < 0.05. To check clustering of RNA-sequencing samples of subjects and controls, a 126 principal component analysis (PCA) was performed with the R package pcaExplorer (version 2.6.0)<sup>23</sup>. RNA count data were variance stabilized transformed and the 500 most variant 127 128 genes (top n genes) were selected for computing the principal components.

# 129 Expression of DEGs in different developmental stages

Expression data of DEGs were obtained from PTEE (version 1.1)<sup>14</sup> for the adult brain 130 131 cortex. Genes expressed at a low level in adult brain tissue may be expressed at a higher level 132 in the developing brain and therefore, could still play a significant role in neurodevelopment. 133 Hence, for DEGs expressed <1.5 TPM in adult brain cortex (according to PTEE), expression 134 levels in different developmental stages were obtained from the R package ABAEnrichment (version 1.20.0)<sup>24</sup> for the whole brain. We used a Tukey's HSD test to determine whether this 135 136 group of DEGs (<1.5 TPM in adult brain cortex) displays a significantly different expression 137 profile between the brain developmental stages.

138 DEGs which are expressed <1.5 TPM in adult brain cortex but >1.5 RPKM in prenatal 139 stage of the whole brain and reach their maximum of expression in the prenatal stage were 140 selected and further analyzed for GO enrichment.

### 141 **DEGs involved in NDDs**

A list of genes, which are known to play a role in NDDs, was obtained from PTEE <sup>14</sup>. DEGs of 15q13.2q13.3 microdeletion patients were compared to the list of NDD genes, to determine DEGs that could contribute to the neurological symptoms observed in those patients. The significance for enrichment of DEGs with NDD genes was calculated using a binomial test in R <sup>25</sup>.

## 147 GO enrichment

Gene ontology enrichment analysis was performed with the R package GOfuncR (version 1.14.0) for up- and down-regulated DEGs <sup>26</sup>. GO nodes with a family wise error rate (FWER) <0.05 were considered significantly enriched. To check for unspecific GO enrichment analysis results, the four control subjects were split in two groups and differential gene expression and GO enrichment analyses were performed for those two control groups.

## 153 Identification of activated/inactivated DEG-interacted functional modules

154 We investigated the activity of DEG-interacted functional modules to elucidate the 155 roles of DEGs in 15q microdeletion. The DEG-interacted network was constructed by the 156 DEGs and their interacting partners in the human protein interaction network (PIN), which was obtained from the InBio Map database<sup>27</sup>. A DEG-interacted functional module is a 157 158 subnetwork of the DEG-interacted network formed by genes annotated by the same biological processes. The functional annotations of genes were obtained from GO<sup>28,29</sup>, and only the 159 160 annotations supported by experiments were used in this study. Additionally, to ensure the 161 DEGs' participation and functional association among genes, all the functional modules were 162 required to contain at least one DEG and one interaction. To determine if the member genes 163 of the tested functional module were overrepresented at the top of the entire ranked gene list, we performed the gene set enrichment analysis (GSEA)<sup>30</sup> for evaluating each module's 164 activity and inactivity separately. To assess the activity (inactivity), the entire gene list was 165 166 ranked downward (upward) by the fold change of genes between the 15q microdeletion and 167 controls. We then calculated the enrichment score (ES) for each functional module by 168 walking down the ranked gene list. The ES of functional module *f* is defined as below:

$$ES_f = \max(S_i), i \in \{1, \dots, N\}$$

$$S_i = \sum_{j=0}^{i} p \times \frac{1}{N_f} - (1-p) \times \frac{1}{(N-N_f)}, \text{ where } p = \begin{cases} 0\\ 1, if gene \ i \ \in f \end{cases}$$

169 where  $S_i$  is the score of gene *i*, *i* is ordered by fold change,  $N_f$  is the number of genes in 170 the tested functional module, *N* is the number of total ranked genes, and *p* is a binary 171 parameter. To estimate the significance of  $ES_f$ , we produced 1,000 scores  $ES_{rand}$  calculated 172 from 1,000 randomly permutated gene lists. Then, we denoted the standard score *z*, which was 173 defined as below, as the activity or inactivity of functional module *f*.

$$z = \frac{ES_f - \mu}{\sigma}$$

174 where  $\mu$  and  $\sigma$  are respectively the mean and standard deviation of 1,000 *ES<sub>rand</sub>*. 175 Finally, the functional modules possessing *z* of activity greater than two and *z* of inactivity 176 less than zero were defined as activated; and the functional modules with *z* of inactivity 177 greater than two and *z* of activity less than zero were defined as inactivated. The discovered 178 activated or inactivated functional modules were further summarized/clustered by the 179 REVIGO <sup>31</sup> algorithm with similarity  $\geq$  0.9 that was calculated from Resnik <sup>32</sup> algorithm and 180 visualized using the treemap package <sup>33</sup>.

181 To predict key transcription factors and cofactors that drive transcriptomic differences 182 between microdeletion individuals and controls we used Mining Algorithm for GenetIc 183 Controllers (MAGIC), which leverages ENCODE ChIP-seq data to look for statistical 184 enrichment of transcription factors and cofactors in genes and flanking regions <sup>34</sup>.

#### 185 Results

# 186

# Transcriptional changes in 15q13.3 microdeletion individuals and Df[h15q13]/+ mice

187 To study the effects of the 15q13.3 microdeletion on transcriptional regulation, we 188 performed RNA-seq from three individuals carrying a heterozygous 15q13.3 microdeletion 189 (Fig. 1) and four control subjects. Further, to identify robust changes across species and 190 tissues, we analyzed cerebral cortex tissue from ten mice (Df[h15q13]/+) reported by Gordon 191 et al.<sup>18</sup>. While in the 15q13.3 microdeletion subjects we identified 2,334 genes (adjusted pvalue < 0.05) with altered expression levels compared to controls (Supplementary Table S1), 192 193 only genes within the deleted region withstood multiple testing correction in the mouse 194 (Supplementary Table S1). This could be related to a difference in syntemy between the 195 mouse and human chromosomal regions, to the high interindividual variability of our 196 subjects, or to the generally mild impact on gene expression with genes not reaching the 197 dysregulation threshold necessary to withstand conservative multiple testing correction. The immediate result of applying multiple testing correction is that the probability a true effect 198 may be rejected will increase <sup>35</sup>. To control for false positives, but also to avoid erroneously 199 rejecting real effects we decided to focus on DEGs shared between human and mouse. We, 200 201 thus, considered genes with uncorrected p-value < 0.05 in the mouse and identified 68 shared 202 genes between the two species and different tissues (Supplementary Table S1). To test 203 whether the number of overlapping genes is higher than expected by chance we performed 204 100,000 random samplings considering a total of 20,000 genes. This yielded a p-value of 0.02 205 suggesting the overlap is significant. By contrast, when we considered DEGs among controls, 206 only six genes were shared with the (Df[h15q13]/+) mouse model, which is an amount 207 expected to occur by chance (*p*-value=0.94 from 100,000 simulations).

208 To check whether the gene dosage affects gene expression, we identified genes located 209 in the deleted site, which are expressed in blood and brain cortex (Supplementary Fig. S1). Four genes have an expression higher than 1.5 TPMs in brain cortex <sup>14</sup> (FAN1, MTMR10, 210 KLF13, OTUD7A) of which MTMR10 and KLF13 are also highly expressed in blood 211 212 (Supplementary Fig. S1). These genes were significantly downregulated in both human and 213 mouse samples (Supplementary Table S1). Moreover, although FAN1 and OTUD7A display 214 low expression levels in blood (Supplementary Fig. S1), they were also significantly 215 differentially expressed in our blood transcriptome analysis (Supplementary Table S1).

We next focused on shared DEGs between 15q13.3 microdeletion individuals and the mouse model. Variants in eight of these genes (*PHIP*, *KAT6A*, *VPS13B*, *GPAA1*, *CHD7*,

FIBP, KMT2C, AP1S1) are known causes for monogenic NDD <sup>36</sup>. There are six genes which 218 219 have a gene ontology (GO) annotation related to gene expression (ZFP57, EDA, KAT6A, 220 CD46, PIK3R3) and also six genes related to brain development (CHD7, ITGA4, MYLIP, 221 PAFAH1B3, SIRT2, B4GALT2). Interestingly, we could also identify components of the 222 major histocompatibility complex, class II to be dysregulated in both blood and brain 223 (Supplementary Table S1), which may reflect a disturbed inflammatory or immune process.

224

# Molecular pathways affected by transcriptome alterations in 15q13.3 microdeletion

225 To identify molecular pathways that may be affected by the gene expression profiles 226 we performed GO enrichment analysis followed by protein-protein-interaction (PPI) networks, as previously described  $^{37,38}$ . Using the mouse data, we identified general GO 227 228 categories like cellular components or developmental processes to be enriched with DEGs 229 (Supplementary Table S2). For human subjects, there were two less general GO terms which 230 were most significantly enriched with DEGs: "nervous system development" (GO:0007399, 231 *p*-value after family wise error rate (FWER) multiple correction = 0.036) with 32 associated 232 genes and "DNA binding" (GO:0003677, p-value FWER multiple correction = 0.046) with 233 183 associated genes (Table 1, Supplementary Table S2).

234 To better delineate molecular pathways, involved in the copy-number variant (CNV) 235 pathomechanism, we further focused on identifying the functional modules formed by DEGs 236 from human subjects and their PPI partners. This revealed that most nodes clustered in 237 cellular processes like metabolic pathways, signaling, or cellular components (Fig. 2, 238 Supplementary Table S2). Since regulation of gene expression appeared to be perturbed, we 239 tested if the gene expression profile matches dysregulation of one or more transcription factors based on ENCODE Chip-seq data <sup>34</sup>. This analysis revealed no significant enrichment 240 241 for genes associated with a known transcription regulator, suggesting that a single gene 242 cannot explain the observed expression profile. Interestingly, inactivated functional modules 243 clusters are mainly involved in immune response and regulation of gene expression (Fig. 2B, 244 Supplementary Table S2). Oligodendrocyte differentiation and development appear to be 245 affected, which together with the "positive regulation of neuron death" (Fig. 2A, B, 246 Supplementary Table S2) could explain the impaired nervous system development. To check 247 whether those molecular pathways are specifically identified in the microdeletion individuals, 248 we analyzed differential gene expression between two control groups. GO terms significantly 249 enriched with DEGs of the control groups were mostly related to immune response, and to a 250 much lesser extent to gene expression regulation (Supplementary Table S2). Thus, the

identification of those molecular pathways in the individuals bearing 15q13.3 microdeletion may not be related to the deleted region, but rather to the analyzed tissue. In contrast, we did not identify any GO terms related to nervous system development in the controls. This supports our hypothesis that the effect on pathways related to nervous system development in the affected individuals is a result of the microdeletion.

We showed that affected genes were enriched in pathways related to nervous system development (Table 1) and that PPIs influence apoptosis and neuron death (Fig. 2B). This prompted us to inquire all DEGs that have known Mendelian associations with monogenic NDD. We identified 252 of the DEGs to be related to monogenic intellectual disability (Fig. 2C, Supplementary Table S1). The number of genes is significantly higher than expected by chance (*p*-value binomial test = 0.003), which could suggest an underlying polygenic effect that leads to an increased risk for a neurodevelopmental disorder.

## 263 Dysregulated genes in 15q13.3 microdeletion individuals expressed in the developing brain

264 Further, we asked whether blood DEGs, which are not expressed in the adult brain 265 cortex, may have been expressed in the developing brain. We identified 358 DEGs, which are expressed in blood but not in the adult brain. We used the ABAEnrichment package in R<sup>24</sup> to 266 267 check the expression levels of these genes during the different stages of brain development 268 (Supplementary Table S3). For 245 of the 358 genes, we could retrieve expression levels from 269 the Allen Brain Atlas. Our analysis revealed that for DEGs, which are silenced in the adult 270 brain, there is a significant enrichment for the ones with a maximum expression level in the 271 developing brain (p-value = 0.04, Fig. 3). A GO enrichment analysis of the 53 genes showed 272 that several of these genes are involved in chromosome organization during cell division, but 273 also identified the e.g. DRAXIN gene to be dysregulated, which is involved in the 274 development of spinal cord (Supplementary Table S3).

# 275 Discussion

The 15q13.3 microdeletion is associated with pleiotropic effects and has been described in a wide spectrum of clinical contexts ranging from apparently healthy individuals to some severely affected with ID, epilepsy or even schizophrenia (Fig. 1B)<sup>2</sup>. It is difficult to dissect the mechanisms contributing to the nervous system developmental disturbance mostly because of the limitations of *in vitro* approaches aiming to reproduce human brain development <sup>10</sup>. Thus, the etiology of the 15q13.3 microdeletion's range of hypervariable symptoms remains elusive.

To understand how dysregulation of gene expression contributes to 15q13.3 microdeletion pathomechanisms, we aimed to circumvent artefacts introduced by conventional *in vitro* approaches. Thus, we analyzed transcriptional profiles from three individuals with 15q13.3 microdeletion in an *in vivo* state in blood, which is an easily accessible tissue. To identify genes, which are robustly dysregulated we additionally analyzed brain cortex tissue from a 15q13.3 microdeletion mouse model.

289 We initially checked for dosage effects of the genes included in the microdeletion 290 (Fig. 1A). This revealed that there are four genes with high expression in brain cortex (Supplementary Fig. S1<sup>14</sup>): FAN1, MTMR10, KLF13, OTUD7A, all of which showed 291 292 significant down-regulation in blood (Supplementary Table S1) of our subjects, as well as 293 mouse brain cortex, confirming that a gene dosage effect of the microdeletion contributes to 294 the transcriptional dysregulation. Other genes included in the typical deletion region: *TRPM1*, 295 CHRNA7, MIR211, ARHGAP11B, display low expression levels in brain cortex and blood 296 (Supplementary Fig. S1) and were not significantly differentially expressed in the 15q13.3 297 microdeletion individuals. While *MIR211* is a microRNA, which was not sequenced probably 298 as a result of the library preparation protocol and ARHGAP11B is a human specific gene  $^{36}$ , 299 Trpm1 and Chrna7 were down-regulated in mouse brain cortex (Supplementary Table S1), 300 further supporting the importance of the haploinsufficiency.

301 We next focused on dysregulated genes across the two different tissues in the human 302 subjects and the mouse model. One of the shared down-regulated genes is CHD7, which is 303 frequently associated with CHARGE syndrome and has been shown to be highly relevant for neuronal differentiation and brain development <sup>39</sup>. CDH7, but also other shared dysregulated 304 305 genes like KMT2C and KAT6A are involved in chromatin remodeling and hence in regulation 306 of gene expression. This is in accordance with the findings of Zhang *et al.*, who described a 307 global epigenomic reprogramming of iPSCs from 15q13.3 microdeletion individuals<sup>12</sup>. 308 However, in their approach they were not able to identify driving factors, potentially 309 secondary to the bias induced by *in vitro* cultivation. This could also explain why they do not 310 identify any shared dysregulations with the mouse model and their multiomics approach 311 yielded a rather low correlation level among the multiple analyses.

To identify affected molecular pathways we performed a GO analysis of DEGs. This showed a significant enrichment of DEGs that are involved in DNA binding (Table 1). Moreover, an analysis of functional modules formed by DEGs and their PPI partners confirmed that gene expression regulation is affected (Fig. 2B, Supplementary Table 2). Yet,

based on ENCODE data <sup>34</sup>, we did not identify any transcription factor that could explain the observed transcriptional profile. This is in accordance with the observation of Zhang *et al.* that the disease-relevant impact of the 15q13.3 microdeletion is probably caused by the combinatorial effects of several genes, rather than a single "master" gene. Our analysis showed network-wide dysregulatory effects and explains why knockout models of singular genes encompassed in the deletion could not fully recapitulate the phenotype <sup>3–10</sup>.

322 The GO analysis also revealed that DEGs show a significant enrichment in the 323 nervous system development category (Table 1). Indeed, we could show that the set of 324 dysregulated genes contained a significant proportion (p-value = 0.003) of genes which have 325 been related to monogenic NDD (Fig. 2C, Supplementary Table S1). The PPI functional 326 module analysis identified more specific developmental processes of the nervous system like 327 oligodendrocyte differentiation, axon and neuron projection development, as well as "positive 328 regulation of neuron death" to be affected (Fig. 2A, B, Supplementary Table S2). This aligns 329 with experimental data from Df[h15q13]/+ mice, which shows that both loss of OTUD7A and CHRNA7 contribute to dendrite outgrowth defects <sup>3,6</sup>, and also with the recently described 330 involvement of *Klf13* in the development of cortical interneurons.<sup>10</sup> 331

To determine whether other DEGs silenced in the adult human brain might have played a role in the nervous system development, we used the Allen Brain Atlas data, which provides brain gene expression data during different developmental stages. <sup>24</sup> This showed that a significant number of genes found to be differentially expressed in blood, but silenced in the adult brain had maximum expression levels in the prenatal stage (Fig. 3).

337 Our data suggests that network-wide dysregulatory effects contribute to 15q13.3 338 microdeletion pathomechanisms. There are several lines of evidence that indicate a disturbed 339 nervous system development, suggesting the severity of 15q13.3 microdeletion individuals' 340 symptoms is probably determined in the early embryonic stages. The identification of 341 dysregulated genes clustering in inflammatory and immune pathways may be related to the 342 analyzed tissue, namely blood. However, since we identified components of the major 343 histocompatibility complex, class II to be dysregulated in the mouse brain, we cannot rule out 344 that immune insults could contribute to the increased vulnerability of 15q13.3 microdeletion-345 bearing offspring.

A major limitation of our study is the small cohort, in which individual-characteristic gene expression levels, which were not caused by the microdeletion, can have a big impact on the analysis. We attempted to circumvent this by comparing our data to the mouse model.

However, a larger cohort and potentially the analysis of an additional tissue like skin, could further refine the analysis and unveil which gene network is mostly responsible for the pathomechanism. This is crucial for directing future efforts to minimize the severity of the phenotype.

### 353 Acknowledgements

We are grateful that our patients and their families agreed to participate in this study. We thank Sandra Schinkel, Kathleen Lehmann, and Sophie Behrendt for their great technical assistance and Rigo Schulz for server support. We are grateful to Torsten Schöneberg for his input and help to draw Fig. 1.

## 358 Author contributions

359 MK performed gene expression analyses, contributed to the design of the study and 360 writing of the manuscript. AV and LB supported bioinformatic analyses and performed GO 361 enrichment analyses. MR performed wet lab work and contributed to the design of the study. 362 CCL performed PPI analyses and contributed to manuscript writing. PZ recruited patients, 363 performed phenotyping, and coordinated the genetic diagnosis. TB, AT, KP, and JH 364 performed genetic diagnosis and contributed to the writing of the manuscript. AK, AM, NS, 365 TL, KP, JRL, and AG contributed to analysis design, data interpretation, and writing of the 366 manuscript. RAJ and DLD designed the study, coordinated the contact to patients, contributed 367 to genetic diagnosis, gene expression analysis, and writing of the manuscript.

# 368 Funding

This study is funded by the Else Kröner-Fresenius-Stiftung 2020\_EKEA.42 to DLD and the German Research Foundation SFB 1052 project B10 to DLD and AG. DLD is funded through the "Clinician Scientist Programm, Medizinische Fakultät der Universität Leipzig". Open Access funding enabled and organized by Projekt DEAL.

#### 373 Data availability

374 RNA sequencing reads and expression profiles have been submitted to the Gene 375 Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under number accession 376 GSE197903. The code analyzing has deposited used for data been under 377 https://github.com/akhilvelluva/15q13.3.

378

379

# 380 References

- van Bon, B. W., Mefford, H. C. & de Vries, B. B. *15q13.3 Microdeletion*.
   *GeneReviews*® (University of Washington, Seattle, 2015).
- Lowther, C., Costain, G., Stavropoulos, D. J., Melvin, R., *et al.* Delineating the
  15q13.3 microdeletion phenotype: A case series and comprehensive review of the
  literature. *Genetics in Medicine* vol. 17 149–157 (2015).
- Gillentine, M. A., Yin, J., Bajic, A., Zhang, P., *et al.* Functional Consequences of
   CHRNA7 Copy-Number Alterations in Induced Pluripotent Stem Cells and Neural
   Progenitor Cells. *Am. J. Hum. Genet.* **101**, 874–887 (2017).
- 4. Hoppman-Chaney, N., Wain, K., Seger, P. R., Superneau, D. W., *et al.* Identification of
  single gene deletions at 15q13.3: Further evidence that CHRNA7 causes the 15q13.3
  microdeletion syndrome phenotype. *Clin. Genet.* 83, 345–351 (2013).
- 392 5. Yin, J., Chen, W., Chao, E. S., Soriano, S., *et al.* Otud7a Knockout Mice Recapitulate
  393 Many Neurological Features of 15q13.3 Microdeletion Syndrome. *Am. J. Hum. Genet.*394 102, 296–308 (2018).
- 395 6. Uddin, M., Unda, B. K., Kwan, V., Holzapfel, N. T., *et al.* OTUD7A Regulates
  396 Neurodevelopmental Phenotypes in the 15q13.3 Microdeletion Syndrome. *Am. J. Hum.*397 *Genet.* 102, 278–295 (2018).
- Ionita-Laza, I., Xu, B., Makarov, V., Buxbaum, J. D., *et al.* Scan statistic-based
  analysis of exome sequencing data identifies FAN1 at 15q13.3 as a susceptibility gene
  for schizophrenia and autism. *Proc. Natl. Acad. Sci. U. S. A.* 111, 343–348 (2014).
- 401 8. Florio, M., Albert, M., Taverna, E., Namba, T., *et al.* Human-specific gene
  402 ARHGAP11B promotes basal progenitor amplification and neocortex expansion.
  403 Science (80-.). 347, 1465–1470 (2015).
- 404 9. Hori, T., Ikuta, S., Hattori, S., Takao, K., *et al.* Mice with mutations in Trpm1, a gene
  405 in the locus of 15q13.3 microdeletion syndrome, display pronounced hyperactivity and
  406 decreased anxiety-like behavior. *Mol. Brain* 14, 1–16 (2021).
- Malwade, S., Gasthaus, J., Bellardita, C., Andelic, M., *et al.* Identification of
  Vulnerable Interneuron Subtypes in 15q13.3 Microdeletion Syndrome Using SingleCell Transcriptomics. *Biol. Psychiatry* (2021) doi:10.1016/j.biopsych.2021.09.012.

- 11. Nilsson, S. R. O., Celada, P., Fejgin, K., Thelin, J., *et al.* A mouse model of the
  15q13.3 microdeletion syndrome shows prefrontal neurophysiological dysfunctions
  and attentional impairment. *Psychopharmacology (Berl).* 233, 2151–2163 (2016).
- Zhang, S., Zhang, X., Purmann, C., Ma, S., *et al.* Network Effects of the 15q13.3
  Microdeletion on the Transcriptome and Epigenome in Human-Induced Neurons. *Biol. Psychiatry* 89, 497–509 (2021).
- 416 13. Solomon, E., Davis-Anderson, K., Hovde, B., Micheva-Viteva, S., *et al.* Global
  417 transcriptome profile of the developmental principles of in vitro iPSC-to-motor neuron
  418 differentiation. *BMC Mol. Cell Biol.* 22, (2021).
- 419 14. Velluva, A., Radtke, M., Horn, S., Popp, B., *et al.* Phenotype-tissue expression and
  420 exploration (PTEE) resource facilitates the choice of tissue for RNA-seq-based clinical
  421 genetics studies. *BMC Genomics* 22, 802 (2021).
- Frésard, L., Smail, C., Ferraro, N. M., Teran, N. A., *et al.* Identification of rare-disease
  genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* 25,
  911–919 (2019).
- 425 16. Curry, P. D. K., Broda, K. L. & Carroll, C. J. The Role of RNA-Sequencing as a New
  426 Genetic Diagnosis Tool. *Curr. Genet. Med. Rep.* 9, 13–21 (2021).
- 427 17. Zacher, P., Mayer, T., Brandhoff, F., Bartolomaeus, T., *et al.* The genetic landscape of
  428 intellectual disability and epilepsy in adults and the elderly: a systematic genetic work429 up of 150 individuals. *Genet. Med.* 23, 1492–1497 (2021).
- 430 18. Gordon, A., Forsingdal, A., Klewe, I. V., Nielsen, J., *et al.* Transcriptomic networks
  431 implicate neuronal energetic abnormalities in three mouse models harboring autism and
  432 schizophrenia-associated mutations. *Mol. Psychiatry* 26, 1520–1534 (2021).
- 433 19. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., *et al.* STAR: Ultrafast universal
  434 RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).
- 435 20. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high436 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- 437 21. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
  438 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, (2014).
- 439 22. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., *et al.* Controlling the false discovery
  440 rate in behavior genetics research. in *Behavioural Brain Research* vol. 125 279–284

441		(Behav Brain Res, 2001).
442	23.	Marini, F. & Binder, H. PcaExplorer: An R/Bioconductor package for interacting with
443		RNA-seq principal components. BMC Bioinformatics 20, (2019).
444	24.	Grote, S., Prüfer, K., Kelso, J. & Dannemann, M. ABAEnrichment: An R package to
445		test for gene set expression enrichment in the adult and developing human brain.
446		<i>Bioinformatics</i> <b>32</b> , 3201–3203 (2016).
447	25.	Team, R. C. R: A language and environment for statistical computing. (2013).
448	26.	Grote, S. GOfuncR: Gene ontology enrichment using FUNC. R Packag. version 1.5.1
449		(2018).
450	27.	Li, T., Wernersson, R., Hansen, R. B., Horn, H., et al. A scored human protein-protein
451		interaction network to catalyze genomic interpretation. Nat. Methods 14, 61-64 (2016).
452	28.	Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., et al. Gene ontology: Tool for
453		the unification of biology. Nature Genetics vol. 25 25-29 (2000).
454	29.	Carbon, S., Douglass, E., Dunn, N., Good, B., et al. The Gene Ontology Resource: 20
455		years and still GOing strong. Nucleic Acids Res. 47, D330–D338 (2019).
456	30.	Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., et al. Gene set
457		enrichment analysis: A knowledge-based approach for interpreting genome-wide
458		expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550 (2005).
459	31.	Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. Revigo summarizes and visualizes
460		long lists of gene ontology terms. PLoS One 6, (2011).
461	32.	Resnik, P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its
462		Application to Problems of Ambiguity in Natural Language. J. Artif. Intell. Res. 11,
463		95–130 (1999).
464	33.	Tennekes, M. Package 'treemap' Type Package Title Treemap Visualization. (2021).
465	34.	Roopra, A. MAgIC: A tool for predicting transcription factors and cofactors driving
466		gene sets using ENCODE data. PLoS Comput. Biol. 16, e1007800 (2020).
467	35.	Groenwold, R. H. H., Goeman, J. J. & Le Cessie, S. Multiple testing: When is many
468		too much? Eur. J. Endocrinol. 184, E11-E14 (2021).
469	36.	Xing, L., Kubik Zahorodna, A., Namba, T., Pinson, A., et al. Expression of
470		human specific ARHGAP11B in mice leads to neocortex expansion and increased

- 471 memory flexibility. *EMBO J.* **40**, (2021).
- 472 37. Jäger, E., Schulz, A., Lede, V., Lin, C.-C., *et al.* Dendritic Cells Regulate GPR34
  473 through Mitogenic Signals and Undergo Apoptosis in Its Absence. *J. Immunol.* 196,
  474 2504–2513 (2016).
- 475 38. Le Duc, D., Lin, C. C., Popkova, Y., Yang, Z., *et al.* Reduced lipolysis in lipoma
  476 phenocopies lipid accumulation in obesity. *Int. J. Obes.* 45, 565–576 (2021).
- Feng, W., Kawauchi, D., Körkel-Qu, H., Deng, H., *et al.* Chd7 is indispensable for
  mammalian brain development through activation of a neuronal differentiation
  programme. *Nat. Commun.* 8, 1–14 (2017).
- 480 40. Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., *et al.* The mutational
  481 constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443
  482 (2020).

483

484

## 485 Figure legends

486 Fig. 1. Overview of the 15q13.3 locus and symptoms associated with the 487 microdeletion A. Schematic representation of the 15q13.3 microdeletion region. Protein 488 coding genes within the region are shown beneath chromosome 15. The color legend corresponds to the pLI score as a measure of loss-of-function deleteriousness <sup>40</sup>. Underlined 489 490 genes have been considered candidates that are responsible for the observed phenotypes (CHRNA7<sup>3,4</sup>, OTUD7A<sup>5,6</sup>, FAN1<sup>7</sup>, ARHGAP11B<sup>8</sup>, TRPM1<sup>9</sup>, KLF13<sup>10</sup>) B. Individuals with 491 492 15q13.3 microdeletion display a heterogenous phenotype which can range from normal 493 development to severe intellectual disability (ID) or neurodevelopmental disorders (NDD). A 494 delineation of the phenotype based on 246 cases revealed predominantly neurologic 495 symptoms of which ID, epilepsy, and neuropsychiatric disorders are most prominent<sup>2</sup>.

496 Fig. 2. Functional modules representation. A. Activated functional modules clusters in 497 individuals with 15q13.3 microdeletion. Functions that could influence nervous system 498 development are clustered in oligodendrocyte differentiation under the "cellular 499 differentiation" category. The size of the boxes is proportional to the activation level of the 500 module. B. Inactivated functional modules clusters in individuals with 15q13.3 microdeletion. 501 These include processes relevant for neuron development. The size of the boxes is

502 proportional to the inactivation level of the module. C. 252 of the DEGs are related to 503 monogenic neurodevelopmental disorders (NDD). The number of genes is significantly 504 higher than expected by chance (p-value binomial test = 0.003).

Fig. 3. Inquiry of DEGs which are not expressed in the adult brain cortex. Based on the Allen Brain Atlas these genes show a significant enrichment for genes with highest expression level in the prenatal stage (p-value adult vs. prenatal stage = 0.04).

508

509 **Table 1.** Enriched overrepresented GO terms in DEGs of 15q13.3 individuals. FWER:

510 family-wise error rate corrected p-value; #genes: number of DEGs involved in the function.

511 For genes included in the nodes refer to Supplementary Table S2.

GO ID	Ontology	GO term	raw <i>p</i> - value	FWER	#genes	
GO:0007399	biological process	nervous system development	1.93E-05	0.036	32	
GO:0003677	molecular function	DNA binding	0.0001	0.046	183	

512

513





в

A NADH dehydrogenase complex assembly			oly	aerobic respiration					antigen processing and presentation of exogenous peptide antigen via MHC class II			T cell activation i in immune res		
										II	nmu	ne	ce	
												– nega nec	ative re rosis f	egula factoi
proton transmembrane transport		e tRNA metab			c process	ribo subui	somal large hit biogenesis		regulation of B cell receptor signaling pathway			natural killer co		er cel
С	el	lu	lar r	ne	tabo	olisn	n							
			a bic	llpha–a osynth	amino acio etic proce	d ss			re	gulation of ruffle		posit of DNA polyr	tive reg A-dired nerase	gulati cted e acti
nucleoside triphosph biosynthetic proces		ohate ess mitochondri			al gene expression		RNA localization to Cajal body		positive regulation of protein localization to chromosome, telomeric region		protein localization to nucleoplastic			
medRxiv preprint doi: https://do	bi.org/10.	1101/202 ew) is the	22.04.08.22273231 e author/funder. who	; this versio o has grant	n posted April 16, ed medRxiv a licer	2022. The copyrigh	nt holde	er for this preprint						
<b>B</b> organelle fission	Gol	gi orç	ganization	nucleo tra	cytoplasmic ansport	vacuo organiza	ole atior	macropha derived foam ce differentia	age d ell tion	nucleic acid transport	cellular ı cytokin	respons e stimul	e to us	re int beta
nuclear division	nega cata	ative of ce abolie	regulation ellular c process	P- ass	-body sembly	regulatio chroma assembly disassen	n of tin y or nbly	FRNA secondar structure unwindin	ry e ng	nucleus organization protein	regulatio	un of appaB	positi regulati	ive on of
	T cell r		cell migration o		ear pore nization	clathrin co assembl		regulation of cell size	of	localization to cytoplasmic stress granule	signalii	ng	C NIK/NF-kappa Signaling	
cell junction assembly	plasma membrane pos bounded cell			or positive of	gani e regulation cellular	zatio regulation protein	OI	of transport		maintenance of DNA repeat elements	defense response to Gram-positive		positive regulation of pri–miRNA transcription by RNA polymerase II	
	mc	proje orpho	ection ogenesis	carb metabo	ohydrate blic process	localization cell periph	n to Iery	receptor internalizatio	on	telomerase catalytic core complex assembly	Dacteri			
positive regulation of organelle organization	0	ve: organ	sicle nization	replicatic chroma	DNA n-independent atin assembly	carbohydr transpoi	ate rt	regulation of telomerase catalytic core complex assem	f e nbly	membrane invagination	regulati canonica signaling p	on of al Wnt pathway	transn recept tyrosi sig	nemb tor pr ine king gnalin
			regulation	of	positive	e regulatior		ositive regulati	ion	regulation of	ماليا	or e	ian	ali
regulation of mRN/ metabolic process	INA ess		RNA splicin		of mRN	A metaboli ocess	c o	of angiogenes	is	gene silencing by miRNA	positive re			cepto
regulati	0	n	of g	en	e ex	pres	SS	pri-miRNA	f ov	regulation of translation, ncRNA-mediated	of insulin r signaling p	eceptor bathway	sig path JAł	gnalir hway K–ST
histone modificatio	posttranscription regulation o gene expressi			of ssion	mRNA pr	a catabolic ocess	RNA polymeras regulation of DI recombination		<u>se II</u> NA n	positive regulation of transcription from RNA polymerase II promoter in response to hypoxia	regulation of transforming growth factor beta receptor signaling pathway		regu epider facto	ulatior mal g r rece

nvolved onse	h	definitive emopoiesis	regulation o transduction class med	of signal n by p53 diator	positive regulation of calcium–mediated signaling	
res	bo	nse	cellu	lar sig	naling	
ion of tumor production		regulation of natural killer cell chemotaxis	regulation o TORC2 signa	of TC	DRC2 signaling	
chemo	otaxis		cellular	differ	entiation	C DEGs
on DNA ⁄ity	prote into	ein insertion membrane	oligodendro differentiat	ocyte tion	regulation of definitive erythrocyte differentiation	
niz	axone	on eme assembly	resp negative regul DNA damage re signal transdu p53 class me	onse t damag lation of esponse, ction by ediator	CODNA CODNA Tesponse to DNA amage checkpoin signaling	2,334
gulation erleukin i produc	of 1–1 ction	positive regulation of cell migration	peptidyl–serine phosphorylation	protein monoubiquitinati	positive regulation <sup>on</sup> of protein phosphorylation	252 <i>p</i> -value=0.0
pos regula cytok	elper 1	regulation of T-helper 1 type immune response	protein acylation <b>rotein</b>	N-terminal protein amino acid modification <b>proc</b>	glycoprotein biosynthetic process <b>essing</b> rotein processing	1,452
positive inflamm	e regulat atory res	tion of sponse	acetylation	process a	myloid precursor protein catabolic process	NDD gene
rane btein lase 7 <b>ng</b> via AT	phospholip phosphory cellu gas pmeosta	ve neuror projection on of projection death	response to calcium ion abolism cyclic nucleotide metabolic process glucose metabolic process regulation of synaptic vesicle exocytosis	endoth develop epith pulmona develop negative r of apoptoti <b>apop</b> positive re of apoptoti	elium oment elium of endothelial cell proliferation ry valve oment egulation c process tosis tosis	







**Supplementary Fig. S1.** Expression levels of genes located in the 15q13.3 microdeletion region. TPM (transcript per kilobase million mapped reads) values of gene expression levels are depicted for brain cortex tissue and whole blood. The dashed lines represent 1.5 TPM. Expression values were obtained from PTEE (https://bioinf.eva.mpg.de/PTEE/).