Research Paper

Ancestral Variations of the *PCDHG* Gene Cluster Predispose to Dyslexia in a Multiplex Family

Teesta Naskar [a,e], Mohammed Faruq [b], Priyajit Banerjee [c], Massarat Khan [d], Rashi Midha [a,1], Renu Kumari [b,f], Subhashree Devasenapathy [a], Bharat Prajapati [a], Sanghamitra Sengupta [e], Deepti Jain [c], Mitali Mukerji [b], Nandini Chatterjee Singh [a,*], Subrata Sinha [a,g,*]

[a] *Cellular and Molecular Neuroscience Division, National Brain Research Centre, Manesar, Gurgaon, Haryana 122051, India*
[b] *Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology, New Delhi 110007, India*
[c] *Regional Centre for Biotechnology, NCR Biotech Science Cluster, Faridabad, Haryana 121001, India*
[d] *Maharashtra Dyslexia Association, Mumbai, Maharashtra 400088, India*
[e] *Department of Biochemistry, University of Calcutta, Kolkata, West Bengal 700019, India*
[f] *Academy of Scientific and Innovative Research, CSIR-Institute of Genomics & Integrative Biology (AcSIR-IGIB), New Delhi 110020, India*
[g] *Department of Biochemistry, All India Institute of Medical Sciences, New Delhi 110029, India*

ARTICLE INFO

ABSTRACT

Dyslexia is a heritable neurodevelopmental disorder characterized by difficulties in reading and writing. In this study, we describe the identification of a set of 17 polymorphisms located across 1.9 Mb region on chromosome 5q31.3, encompassing genes of the *PCDHG* cluster, *TAF7*, *PCDH1* and *ARHGAP26*, dominantly inherited with dyslexia in a multi-incident family. Strikingly, the non-risk form of seven variations of the *PCDHG* cluster, are preponderant in the human lineage, while risk alleles are ancestral and conserved across Neanderthals to non-human primates. Four of these seven ancestral variations (c.460A > C [p.Ile154Leu], c.541G > A [p.Ala181Thr], c.2036G > C [p.Arg679Pro] and c.2059A > G [p.Lys687Glu]) result in amino acid alterations. p.Ile154Leu and p.Ala181Thr are present at EC2: EC3 interacting interface of γA3-PCDH and γA4-PCDH respectively might affect trans-homophilic interaction and hence neuronal connectivity. p.Arg679Pro and p.Lys687Glu are present within the linker region connecting trans-membrane to extracellular domain. Sequence analysis indicated the importance of p.Ile154, p.Arg679 and p.Lys687 in maintaining class specificity. Thus the observed association of *PCDHG* genes encoding neural adhesion proteins reinforces the hypothesis of aberrant neuronal connectivity in the pathophysiology of dyslexia. Additionally, the striking conservation of the identified variants indicates a role of *PCDHG* in the evolution of highly specialized cognitive skills critical to reading.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Reading is a specific, advanced cognitive activity of humans. However individuals with dyslexia experience varying degrees of difficulty in performing this skill despite adequate intelligence or education and absence of neurological illness or sensory deficits (Peterson and Pennington, 2015). Worldwide epidemiological data suggests that the prevalence of dyslexia is approximately 5–12% (Shaywitz et al., 2007), while in India it is reported to be 9–11% (Mogasale et al., 2012). Dyslexia is known to have a strong neurodevelopmental origin, as a result of aberrations in neuronal migration and connectivity, as elucidated in a number of studies involving postmortem brains and neuroimaging (Paulesu et al., 1996; Skeide et al., 2015). Studies on postmortem brains of individuals with dyslexia have shown specific histological anomalies including ectopias and heterotopias resulting from abnormal neuronal migration (Galaburda and Kemper, 1979; Galaburda et al., 1985). Functional magnetic resonance imaging (f-MRI) and positron emission tomography (PET) scan studies have reported the neural correlates of acquired skills like reading and writing (Paulesu et al., 2014). These studies attributed dyslexia to poor or delayed neuronal maturation and disrupted functional connectivity of neurons.

This neurodevelopmental disorder also has a strong genetic component which could be heterogeneous in nature (Mascheretti et al., 2017). The heritability of dyslexia, as well as its constituent sub-phenotypes, has been shown by a number of twin and family-based studies (Fisher et al., 2002; Schumacher et al., 2007). Genome-wide studies using complex pedigrees have reiterated the heritability of dyslexia sub-phenotypes and mapped several genomic loci designated as *DYX1-DYX9* includes candidate genes like *ROBO1* (Hannula-Jouppi et al.,

2005), *KIAA0319* (Paracchini et al., 2006), *DCDC2* (Meng et al., 2005), *DYX1C1* (Tapia-Paez et al., 2008), mostly implicated in neurite outgrowth, neural connectivity, migration and development. In addition to the genes of these *DYX* loci, many other genes such as *FOXP2*, *CNTNAP2* (Peter et al., 2011), *SLC2A3* (Roeske et al., 2011), *GRIN2B* (Mascheretti et al., 2015), *CEP63* (Einarsdottir et al., 2015) and *PCDH11X* (Veerappa et al., 2013) have been shown to be associated with dyslexia.

The studies so far indicate that dyslexia is likely to be a collection of many different endophenotypes resulting in multiple molecular and cellular pathologies. However, the basic molecular underpinnings of this disability are still elusive. Therefore, for a better understanding of the pathophysiology of reading (dis)ability, an effort to identify novel susceptible genes to the disorder, we have investigated the genetic basis of dyslexia inheritance, applying whole exome sequencing and genome wide SNP array, in a three generational family from a highly endogamous group from Western India.

We identified 17 variations present at or adjacent to the protocadherin gamma (*PCDHG*) gene cluster that co-segregated with dominantly inherited dyslexia in the family being studied. The clustered protocadherins play important roles in several steps of neural morphogenesis and connectivity. Remarkable features of the γ-PCDH proteins, including their extensive molecular diversity, enriched synaptic localization, isoform specific homophilic adhesion, cell specific expression pattern, dendritic expression and spine morphogenesis suggest their indispensable role in the development and maintenance of neural circuits and their functional maturity and connectivity (Chen and Maniatis, 2013; Schreiner and Weiner, 2010; Kostadinov and Sanes, 2015).

Clustered α and β-*PCDH* as well several non-clustered *PCDH* have been reported to be associated with many neurodevelopmental disorders like autism spectrum disorder (ASD) (Anitha et al., 2013), schizophrenia (Jiang et al., 2017), epilepsy (Cooper et al., 2016), intellectual disability. In an Indian family based study, a genome wide scan identified copy number variations of *PCDH11X*, a non-cluster protocadherin as being associated with dyslexia (Veerappa et al., 2013).

In the present study, identification of multiple variations co-segregated with dyslexia like a single haplotype block provides mechanistic insights into the disease pathophysiology. In addition, the presence of the variations on extracellular domains of γPCDHs along with the importance of p.Ile154Leu and p.Ala181Thr in trans-homophilic interactions strengthen the hypothesis of aberrant neuronal connectivity in the pathophysiology of dyslexia and could guide to generate physiologically relevant cellular and animal models. Interestingly, the striking evolutionary conservation of seven of these dyslexia associated variants, including four non-synonymous amino acid changes (c.460A > C [p.Ile154Leu], c.541G > A [p.Ala181Thr], c.2036G > C [p.Arg679Pro] and c.2059A > G [p.Lys687Glu]), indicates their evolutionary significance in the development of cognitive substrates underlying the unique human ability to read.

2. Materials and Methods

2.1. Participants

An extended family KA25 with familial dyslexia were identified for this study (Fig. 1). Twenty members out of twenty six, were included in this study. The study was approved by the Institutional Human Ethics Committee of National Brain Research Centre, Manesar, India and signed informed consent was obtained from all the participants in accordance with the Declaration of Helsinki. In case of children, the signed informed consent was obtained from their parents. Members of the family were in the age range of 4–78 years; while tests for reading were administered to the 7–70 years age group. Except for two participants (IV-8 and IV-9), all others had a graduate degree and had received at least 15 years of academic education in English. They all reported English as their preferred and proficient language and hence languages

and reading assessments were carried out in English, using Dyslexia Assessments for Languages of India (DALI) (see Web Resource). DALI is a standardized and validated battery of assessments developed by the National Brain Research Centre, India and is available in four Indian languages namely Hindi, Marathi, Kannada and English. Non-verbal intelligence quotient was assessed using Standard Progressive Matrices (Raven, 2000). Participants were interviewed individually to ascertain reading history, difficulties during schooling and performances in their remedial classes for those who had undergone the remedial program.

2.2. Whole Exome Sequencing

Whole exome sequencing was performed for individuals II-1 II-9, III-1, III-2, III-3, III-4, III-8, III-9, IV-1, IV-2, IV-3, IV-6 and IV-7 on Illumina Hi-seq 2000. For each sample, 2 μg of non-degraded high molecular weight genomic DNA was used by following manufacturer's protocol.

Bioinformatics analyses and quality check of sequence reads were done through genome reassembly pipeline of NGS toolkit (Patel and Jain, 2012). After ensuring quality, raw sequence reads of two end-sequenced read files were mapped to human indexed reference genome file (Grch37/hg19) by using Bowtie2 (Langmead and Salzberg, 2012). Variant calling from (Sequence Alignment Map) SAM aligned file was performed by using SAMTools (Li et al., 2009) following conversion of SAM file to BAM (Binary Alignment/Map format) file using 'samtools view' parameter, sorting of BAM file using 'samtools sort' parameter, indexing of BAM file using 'samtools index' parameter, generating variation BCF (binary) file using 'samtoolsmpileup' and converting BCF (binary) file to VCF (text) variation file using bcftools. To recalibrate the base quality score as well the local alignment around insertions and deletions was done by GATK31 method. After passing the data QC (80% coverage, >25× depth), total 156,294 variations were found to be shared among all thirteen individuals.

2.3. Variant Prioritization

The variants were prioritized on the basis of dominant inheritance pattern of the disorder in this family. Therefore all 156,294 variations from whole exome sequencing, were filtered out for the risk alleles that were either heterozygous or homozygous in affected individuals against homozygous non-risk genotypes in unaffected individuals. As per the dominant model, it was assumed that one copy of the risk allele was sufficient to develop the disorder.

2.4. Genome Wide SNP Scan

Each DNA sample from the family KA25 was genotyped for 719665 SNP marker using Illumina Human Omni express12v1-1. 1 μg DNA per sample was used for the fragmentation process followed by PCR enrichment for SNP. Initial genomic data scan was performed by using iScan (Illumina). Variants were annotated by ANNOVAR (Wang et al., 2010) and after generating the base call files all the individual files were merged and processed in Genome Studio 1.7 where a final file was generated and analyzed.

2.5. Multiple Sequence Alignment

To investigate conservation of each identified variations within the Neanderthal genome (Prüfer et al., 2017) and primate groups, we have performed the multiple sequence alignment of the flanking sequences of each SNPs using NCBI blast tool, clustal-w and t-coffee (Notredame et al., 2000). Results were generated using ESPript (Gouet et al., 1999). Primate sequences were collected from both UCSC and NCBI database. Neanderthal genome sequences were obtained from the 'Ancient Genome Browser' (see Web Resource) of the Department of Evolutionary Genetics, Max Planck Institute of Evolutionary Anthropology, Leipzig, Germany and also from the UCSC Genome Browser.

Fig. 1. Variations co-segregating with familial dyslexia following dominant inheritance pattern. Pedigree of the extended family KA25. Black filled symbols indicate individuals with dyslexia; white symbols indicate asymptomatic individuals. Symbols with question marks indicate undiagnosed/unknown dyslexia status, 'n' within diamond shaped box indicates unknown lineage information and NA marked individuals have been excluded from the study due to unavailability. Generations are marked with Roman numbers on the left of the image and individuals are counted from left to right. Names of the 17 variations are written on the extreme left of the image and the genotypes of these variants are written under each members of family. Genotypes enclosed in the box indicate the risk haplotype. Individuals marked with asterisk were selected for exome sequencing while the SNP array was performed for entire family.

We have also considered the Multiz Alignment 100 vertebrates from UCSC Genome Browser. Whole protein alignment for γA4-PCDH, γA3-PCDH and γB2-PCDH were performed within primate lineages.

2.6. Homology Model Construction

The template (PDB code: 4ZI9) used for homology modelling of the γA4Pcdh and γA3Pcdh was selected on the basis of resolution, homology and trans dimeric orientation. 4ZI9 describing the structure of mouse γA4PCDH was found to be the best possible template which contains EC1–3. Wild type and mutant human γA4PCDH and γA3PCDH EC1–3 was modeled using Discovery Studio 3.5 (Dassault Systèmes BIOVIA, Discovery Studio Modelling Environment). Sequence of the 4ZI9 structure was aligned against the target sequence to identify the matched regions. Based on the atomic coordinates of the template, homology model of the monomeric target protein was constructed. Dimeric model was then constructed by structural superimposition of the monomeric model on the template (PDB code: 4ZI9). The model was then energy minimized using the steepest descent (Max steps 100) with the CHARM force field. Constructed models were verified by Ramachandran plot in Coot (Emsley and Cowtan, 2004) and figures were generated by DS Visualizer (BIOVIA). Electrostatic potential surface was calculated by APBS plugin in PyMol (DeLano, 2002). Multiple sequence alignment was performed by T-Coffee server and results were generated using ESPript (Gouet et al., 1999).

2.7. Brain Region Specific Gene Expression Profile of PCDHG

Brain region specific gene expression profile of the genes of *PCDHG* cluster was used to generate a heatmap by using Excel utilities and Python script. The raw microarray data of developing human brain was obtained from the Allen Brain Atlas (Miller et al., 2014). The original dataset (expression.csv) was divided into three categories: prenatal; postnatal till 4 years; and adults. In all cases, the average expression value (statistical mean) of each gene corresponding to the brain region, was considered to generate the heatmap on which genes were put on the X axis while Y axis contains the brain regions. The expression value was scaled from 5 to 8. The level of gene expression was measured with RPKM (Reads per Kilobase Million) unit.

2.8. eQTL Analysis of the Identified Variants

Computed expression quantitative trait loci (e-QTL) results from the genotype tissue expression (GTEx) generated eQTL dataset were analyzed for the identified variations (GTEx Consortium, 2017; Pirinen et al., 2015). Box plot representation of the rank normalized gene expression data of individual SNPs were downloaded from the GTEx portal.

3. Results

3.1. Ascertainment of the Dyslexia Status

The participants in the family (KA25) under study, who scored < 1.0 SD below the mean in at least two tests were diagnosed as Dyslexic (Table S1). In addition, the performance of each individual was compared with their previous clinical records and history. Out of twenty family members, sixteen furnished past clinical records. Subjects II-9, III-1, III-3, III-5, III-8, III-10, IV-2, IV-6 and IV-9 were considered as dyslexics on the basis of their performances score on current test battery as well past clinical records, self-reports, history of remediation etc. IV-7 was considered to have dyslexia because of his authenticated and extensive clinical records, although the current test battery could not differentiate him from the unaffected members. II-1, III-12 and IV-1 were categorized to have dyslexia on the basis of past clinical records and self-reported history as they were not available for the current

assessment even though their DNA was available for genotyping. III-2, III-4, III-11 and IV-3 were classified as normal readers on the basis of their current assessment records, past clinical history and self-reports while III-9 was categorized according to self-reports and past clinical diagnosis. II-6 was too old and IV-8 was too young to assess and hence not included in determining genetic association even though their DNA was available.

3.2. Variations of the Protocadherin Gamma Gene Cluster Are Co-Segregated with Dyslexia

KA25, a three generation multiplex family from a highly endogamous community, presented the possibility of identifying variations associated with dyslexia by whole exome sequencing. A total of 22 single nucleotide variations were identified to be associated with dyslexia by following the dominant pattern of inheritance. Out of 156294 variants, these 22 variants were either homozygous or heterozygous for risk allele among all affected individuals while homozygous for non-risk allele among all unaffected individuals. These variants mapped to chromosomes 1, 2, 3, 5, 9, 12, 14, 15 and 21 (Table S2). Thirteen of these variations were in a strong linkage disequilibrium (LD) with each other and located on chromosome 5q31.3 (Table 1, Figs. 1 & 2). 12 variations were validated using Sanger sequencing in all members of the family KA25 as well as in unrelated normal readers ($n = 56$) from the same ethnic group (Table S3). However, genotype of rs62378403 could not be validated with Sanger sequencing and hence not included in further analysis.

Ten out of twelve variations were mapped to protocadherin gamma (*PCDHG*) gene cluster in chr5q31.3. The PCDHG cluster proteins are the putative trans-synaptic recognition molecules. Genes of *PCDHG* cluster are predominantly expressed in the developing human brain, especially in the regions important for cognition and learning (Fig. S1). They comprise of 22 homologous variable exons arrayed tandemly, with each variable exon following three constant exons and stochastically express transmembrane protein isoforms with similar functions. Each of these isoforms consists of six extracellular cadherin (EC) repeats followed by transmembrane helix and C-terminal intracellular domain (Wu and Maniatis, 1999). All the identified variants were located on the variable exons (Fig. 2), which encode extracellular domains of the respective proteins. Within chr5q31.3, we have also identified variants, rs7730, located in TATA-Box Binding Protein Associated Factor 7 (*TAF7*), an intronless gene, flanked by protocadherin beta and gamma clusters that have diverse function in transcription initiation (Gegonne et al., 2006) and rs853158, located in *ARHGAP26* (Fig. 2). *ARHGAP26* encodes Rho GTPase activating protein 26 which belongs to the family of GTPase regulator associated with FAK (GRAF1) and is abundant in neonatal brain (Lucken-Ardjomande Hasler et al., 2014).

For a cross genomic validation of the identified loci, we performed genome wide SNP genotyping and a total of 719,665 variations were analyzed for prioritization using the same procedure previously employed in the exome sequencing variation prioritization process. A total of 21 single nucleotide variations across chromosome 2, 3, 4, 5, 7, 9, 14 were identified to segregate in an autosomal dominant manner (Fisher's exact test $P < 0.0001$) (Table S4) with dyslexia in the family KA25.

We have investigated the variations present on common overlapping genomic regions of exome sequencing and genome wide genotyping. Thus variations located on chr5q31.3, were considered further. Among these, rs13359820 was present in *TAF7*, while non-coding variants rs10491311 and rs13361997 mapped to the *PCDHG* cluster. rs6888135 mapped to protocadherin 1 (*PCDH1*) gene which is located between the *PCDHG* gene cluster and *ARHGAP26*, while rs153149 was present in the intergenic region of *PCDH1* and *ARHGAP26*. rs153149 lying within the CTCF/ cohesion (Rad21) contact domain (chr5:141439092-141439557) and mapped to the regulatory sites of glucosamine-6-phosphate deaminase 1 gene *GNPDA1* (Fig. S2). *PCDH1*, a non-cluster protocadherin, is also largely expressed in the nervous

Table 1
List of the identified variations co-segregated with dyslexia with SNP consequences.

Position	Gene name	SNP Id	Source	Nucleotide change	Amino acid change	SNP consequence
Chr5:140698165	*TAF7*	rs7730	Exome seq	c.*397T > C	–	3′ UTR variant
Chr5:140701730	*TAF7*	rs13359820	SNP array	c.-2119T > C	–	Intron variant
Chr5:140717739	*PCDHGA1*	rs10491311	SNP array	c.2421 + 5067T > C	–	Intronic
Chr5:140724060	*PCDHGA3*	rs11575948	Exome seq	c.460A > C	p.Ile154Leu	Non-synonymous missense
Chr5:140731408	*PCDHGB1*	rs3749777	Exome seq	c.1581A > G	p.Thr527	Synonymous
Chr5:140735215	*PCDHGA4*	rs11575949	Exome seq	c.541G > A	p.Ala181Thr	Non-synonymous missense
Chr5:140736474	*PCDHGA4*	rs17097226	Exome seq	c.1800T > G	p.Thr600	Synonymous
Chr5:140741673	*PCDHGB2*	rs73265834	Exome seq	c.1971G > C	p.Thr657	Synonymous
Chr5:140741738	*PCDHGB2*	rs62378417	Exome seq	c.2036G > C	p.Arg679Pro	Non-synonymous missense
Chr5:140741761	*PCDHGB2*	rs57735633	Exome seq	c.2059A > G	p.Lys687Glu	Non-synonymous missense
Chr5:140744395	*PCDHGA5*	rs57308563	Exome seq	c.498T > C	p.Ser166	Synonymous
Chr5:140755387	*PCDHGA6*	rs62378422	Exome seq	c.1737C > T	p.Pro579	Synonymous
Chr5:140787850	*PCDHGB6*	rs62378453	Exome seq	c.81C > G	p.Pro27	Synonymous
Chr5:140863674	*PCDHGC3*	rs13361997	SNP array	c.2430 + 5561A > C	–	Synonymous
Chr5:141254063	*PCDH1*	rs6888135	SNP array	c.40 + 3725G > T	–	Intronic
Chr5:141439358	Intergenic b/w *MRPL11P2* and *NDF1P1*	rs153149	SNP array	g.141439358A > G	–	Non-coding
Chr5:142605172	*ARHGAP26*	rs853158	Exome seq	c.*3161T > C	–	Downstream variant

system, and plays multiple roles during tissue-specific and circuit-specific neuronal development viz. establishment of specific synaptic connections, neuronal migration and maintenance of adult hippocampal circuitry (Redies et al., 2008; Kim et al., 2010). Taken together, 17 variations (12 from exome sequencing and 5 from SNP array) located across 1.9 Mb regions (140698165–142,605,172 bp) of chromosome 5q31.3 (Figs. 1 & 2) were identified to be co-segregating with dyslexia in the family KA25. Four non-synonymous *PCDHG* variations result in single amino acid alterations, which are rs11575948; c.460A > C (p.Ile154Leu) in γA3-PCDH; rs11575949; c.541G > A (p.Ala181Thr) in γA4-PCDH. rs62378417; c.2036G > C (p.Arg679Pro) and rs5773563;

c.2059A > G (p.Lys687Glu) in γB2-PCDH. The identified variations on chr5q31.3 were considered further based on the possible functional consequences of non-synonymous alterations, along with the evolutionary significance and significant e-QTL effect of non-coding variants.

3.3. Alternative Alleles are Associated with Reduced In-Vivo Expression in Basal Ganglia

Among the seventeen identified variants, three non-coding variants such as rs7730, rs13359820, rs10491311 are reported as expression quantitative trait loci (eQTL) with substantial effect sizes and associated
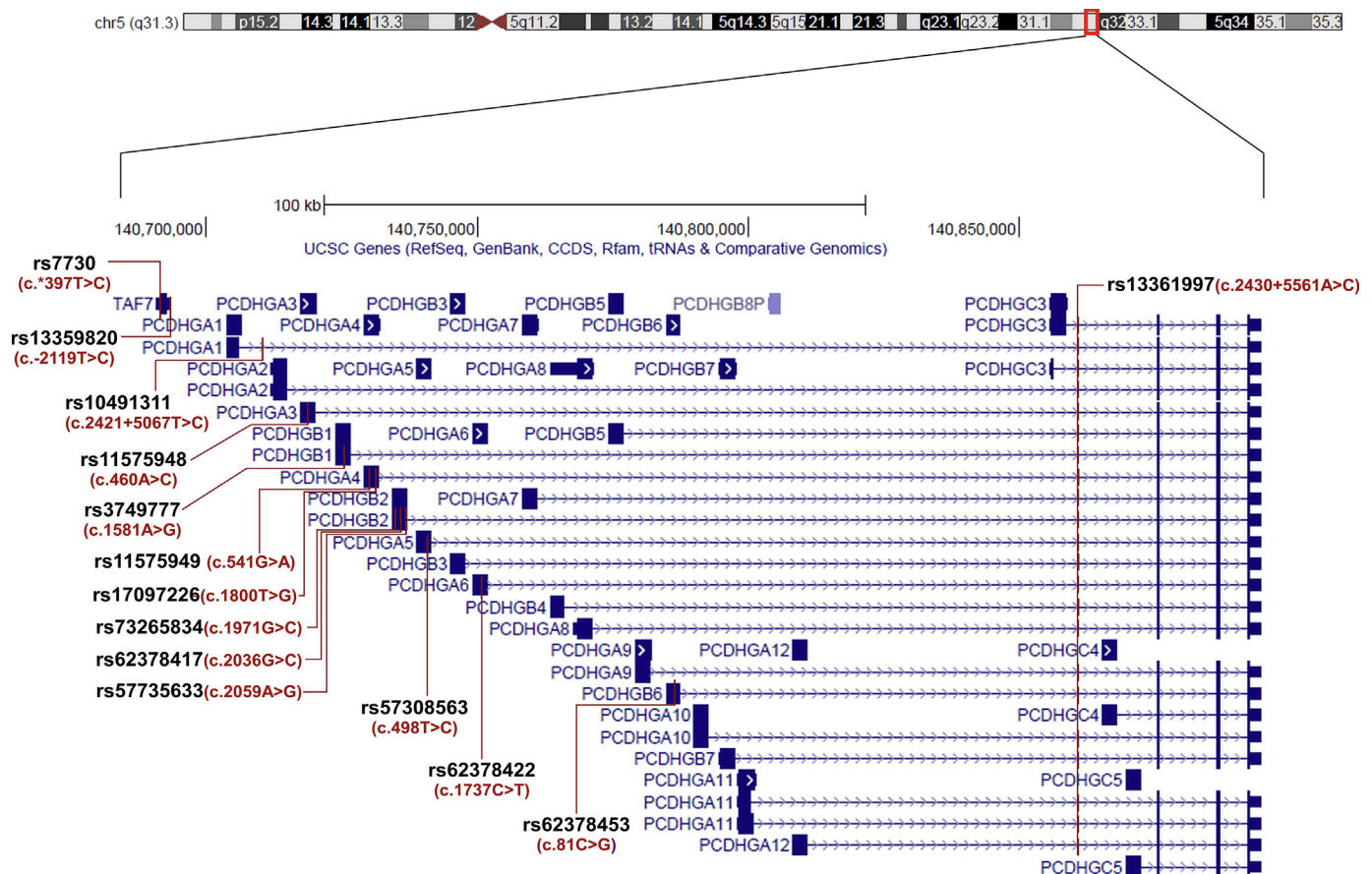
Fig. 2. Overview of variations in chr5q31.3. Schematic diagram includes *PCDHG* gene cluster of chromosome 5q31.3 from approximately 140,690,000–140,890,000 bp, drawn to scale (with coordinates according to GRCh37/hg19 taken from UCSC genome browser). First exons of each isoform genes are depicted as filled blue box. SNPs indicated with red colored lines.
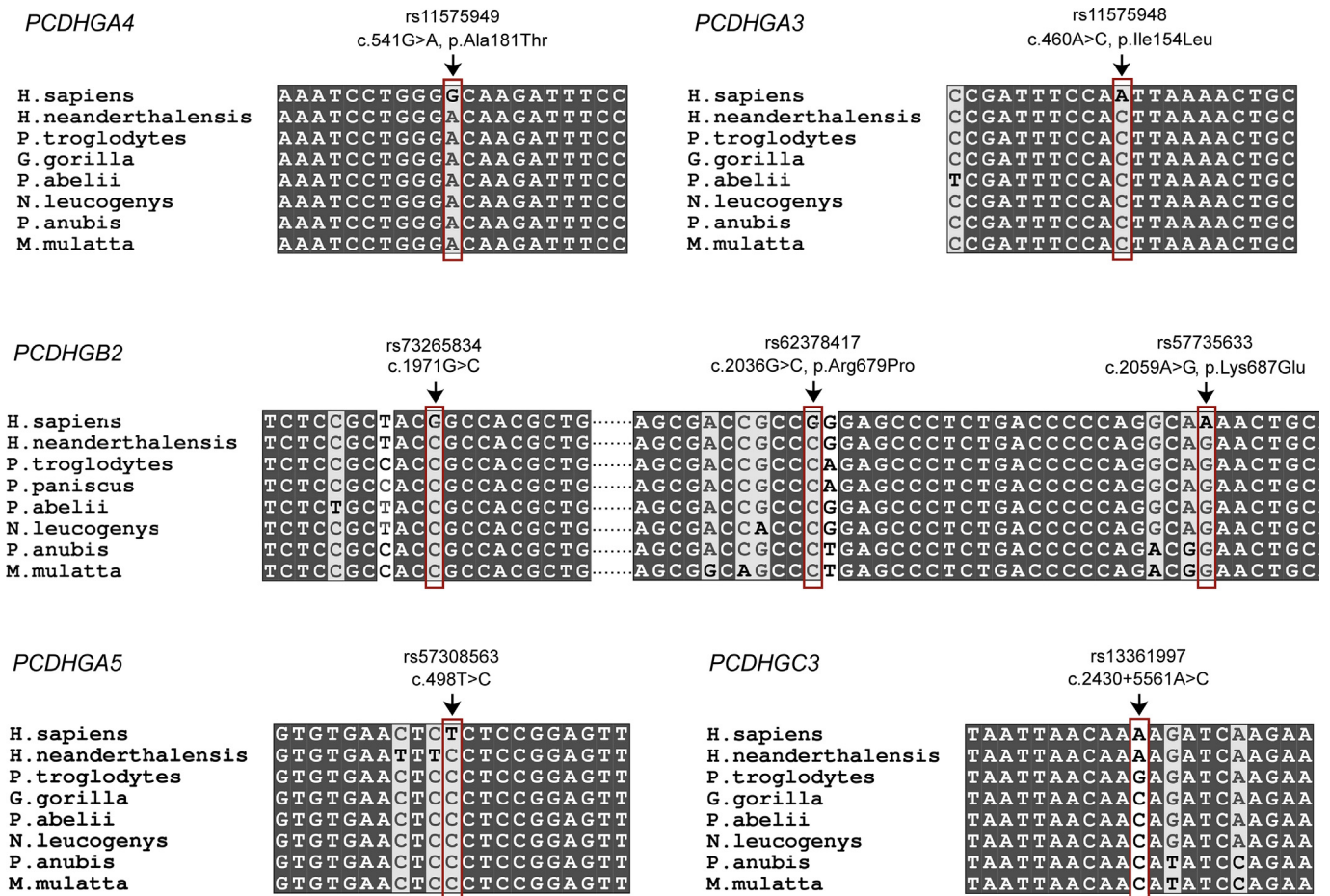
Fig. 3. Incidence of lineage specific variations of PCDHG gene cluster. Figure depicts the multiple sequence alignments of seven variations throughout primates to Neanderthal to human lineage. The identified variations where the risk allele is the ancestral while the non-risk allele is the preponderant human form, are marked with arrow. The respective alleles are bordered with a red-brown box.

with the expression of *PCDHGA1* gene (Fig. S3). According to Genotype Tissue Expression (GTEx) RNA sequence dataset, the alternative alleles of these variations exhibit significantly reduced expression ($P = 5.9 \times 10^{-6}$ for rs7730 and rs10491311; $P = 1 \times 10^{-5}$ for rs13359820) mostly in the caudate (basal ganglia) regions of the brain (Table S5). In basal ganglia, each of the alternative alleles is associated with reduced expression of *PCDHGA1* gene. There are only 2 homozygous alternative genotypes ($N = 2$) and 31 heterozygous genotypes ($N = 31$) for both rs7730 and rs10491311 SNVs. Similarly, there are 3 homozygous alternative genotypes ($N = 3$) and 30 heterozygous genotypes ($N = 30$) for rs13359820. Therefore heterozygous forms of these variations have a significantly lower expression compared to homozygous reference and minimal expression for homozygous mutant (Fig. S3). Previously, imaging studies have implicated basal ganglia in playing a crucial role in the hyperactivation of the brain's caudate region, a phenomenon comes with the consequence with dyslexia (Hoeft et al., 2007; Krishnan et al., 2016). Such preceding observations prompt a relevant association of these variations with reading difficulties.

3.4. Evolutionary Characteristics of Identified Protocadherin Variations

Seven of the identified variations (rs11575948, rs11575949, rs57308563, rs73265834, rs62378417, rs57735633, rs13361997) of chr5q31.3, have several evolutionary characteristics. Multiple sequence alignment revealed that the alternative form of each of these loci are conserved within non-human primates whereas the wild type forms are preponderant in human (Fig. 3). The ancestral form of each locus was part of the risk haplotype and had a ubiquitous presence across

all affected members, while the human specific alleles of these seven loci were omnipresent in all the unaffected members of the family. Interestingly alternative forms of six of these ancestral variants (rs11575948, rs11575949, rs57308563, rs73265834, rs62378417, rs57735633) are also found in Neanderthal genome. Furthermore, comparative genomic analysis of 100 vertebrates (Multiz Align) according to UCSC genome database shows that the wild type non-risk form c.2036G of (rs62378417) *PCDHGB2* is exclusively present only in humans and c.2430 + 5561A of (rs13361997) *PCDHGC3* is exclusive to only in hominins (humans and Neanderthals) but none of the other 100 vertebrates.

The four amino acid altering changes follow the same conserved pattern in the corresponding protein sequences. p.Ile154Leu and p.Ala181Thr are present in the EC2 domain of γA3PCDH and γA4PCDH respectively (Fig. 4a and b). Multiple sequence alignment of the γA3PCDH and γA4PCDH in orthologs showed that the residue Ile154 is Leu in all except γA3PCDH of humans (Fig. 4c) and Ala181 is Thr in all the primates except humans (Fig. 4d).

3.5. Functional Characterization of Identified Alterations through Homology Modelling

3.5.1. Homology Models of γA3PCDH and γA4PCDH and Implications of Observed Alterations

The clustered PCDH proteins contain six extracellular (EC) cadherin repeats (EC1–6) with similar structures. p.Ile154Leu in γA3-PCDH and p.Ala181Thr in γA4PCDH are present in the EC2 domains of respective proteins (Fig. 4a and b). Cell-cell recognition of PCDH involves EC1–4
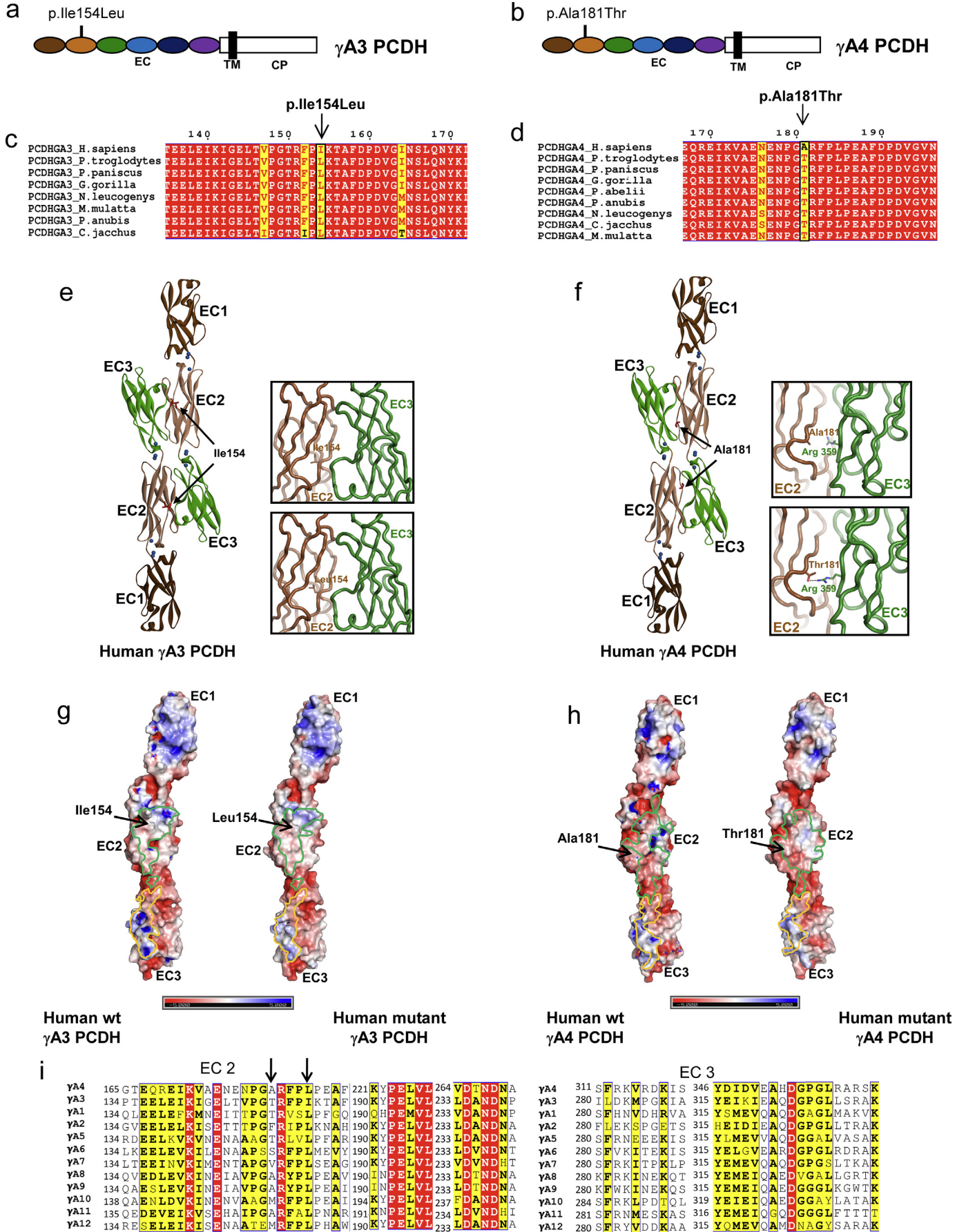
interface as shown structurally and experimentally through mutagenesis (Rubinstein et al., 2015; Goodman et al., 2016a; Nicoludis et al., 2015). Since crystal structures of human PCDH are not currently available, in order to analyze the possible effect of mutation of the residues p.Ile154Leu in γA3-PCDH and p.Ala181Thr in γA4PCDH, both of which lie in the EC2 domain (Fig. 4e & f), homology based models were constructed.

The protein sequences for the human γA4PCDH and γA3PCDH were obtained from NCBI database. EC1-EC3 of the human wild-type and mutant (Ile154Leu) γA3PCDH (Fig. 4e), as well as the human wild type and mutant (Ala181Thr) γA4PCDH (Fig. 4f), were independently modeled using high-resolution (1.7 Å) crystal structure of mouse γA1protocadherin EC1–3 dimer (PDB code: 4ZI9) as the template. The template shares 79% homology and 61% identity to *Homo sapiens* γA4 and γA3-PCDH EC1–3. The extracellular cadherin domains consist of Greek-key ß-sandwich motif and are arranged in tandem with the linker region between them. Each of the linker regions contains three conserved Ca^{+2} binding sites (Nagar et al., 1996). We predict that all linkers between the EC domains will be occupied by Ca^{+2} ions since the calcium binding motif is conserved in the human proteins. A conserved disulfide bond is present between Cys127 and Cys133 in γA4PCDH and Cys96 and Cys102 γA3PCDH models. Similar to the template, the models display an antiparallel arrangement of the two monomers for the formation of dimer such that EC2 from one protocadherin interacts with EC3 from the other. This arrangement is important for the formation of specific trans-homophilic interactions (Goodman et al., 2016b). The models of the wild-type and altered γA4PCDH dimer shows good stereochemistry with main chain conformations for 95.8% and 94.9% of amino acids being present in the most favored region of the Ramachandran plot respectively. The reliability of the wild-type and altered dimer γA3PCDH was also analyzed and 95.8% and 96.4% residues were found to be in the most favorable region of the Ramachandran plot respectively. Overall, our models show canonical features of the other protocadherins and can be utilized for the analysis of the mutations.

The dimer model was also constructed for both γA3PCDH and γA4PCDH based on the template dimer structure. The interface residues were identified from the model. These residues are present in the EC2 and EC3 domains. This region overlaps with the specificity determining region of EC2–3 of mouse Pcdh and is involved in the trans-homophilic interactions (Goodman et al., 2016a; Nicoludis et al., 2015). Ile154 and Ala181 are present in the loop between beta strand 1 (β1) and 2 (β2) of EC2 which interacts with the last beta strand (β7) of EC3. There is no major structural change between the wild-type and the mutant proteins. In case of the γA4PCDH mutant Ala181Thr, a small hydrophobic residue (Ala) is replaced by larger hydrophilic residue (Thr). Ala181 does not participate in the trans-homophilic interactions in the wildtype model. On the other hand, Thr181 shows interaction with Arg359 of the trans protomer in the mutant model, (Fig. 4f insert view). Thus, the models reveal that the mutation converts this residue from a solvent-accessible to being part of interacting interface. In the structure of mouse γA1Pcdh (template), although there is a Thr present at position 181 it does not participate in trans protomer interaction due to absence of Arg359 (replaced by Met) instead, it forms intra protomer hydrogen bond with Thr147. Ile154Leu mutation in γA3PCDH does not result in any major structural changes and the analysis of interface using COCOMAPS webserver (Vangone et al., 2011) shows similar buried

surface area between wild-type γA3PCDH (1407 Å2) and mutant dimer (1454 Å2). Residue Ile154 is pointing toward the hydrophobic core of the EC repeat, its mutation to Leu in our models does not perturb protein folding or the stability of the core structure. The ability of this mutation to allosterically alter the interaction interface cannot be ruled out completely.

Electrostatic potential surface calculations were carried out to assess the change in the charge as well as the shape of the surface in the wild-type and mutant proteins (Fig. 4g and h). The analysis reveals that there are changes in the surface properties on mutation. It is possible that these mutations may result in subtle changes in the specificity and or affinity of the trans-homophilic interactions in protocadherins. In addition, these mutations can also exert an indirect effect through allostery to alter the strength of cell-cell adhesion.

Interestingly, multiple sequence alignment of EC2 and EC3 region of 12 different γAPCDH isoforms of humans show that the interaction interface is not conserved (Fig. 4i). This is expected, as this region is important for distinguishing one isoform from another. Any mutation within this region can result in alteration of specificity of the homophilic interactions. p.Ile154Leu and p.Ala181Thr are present within this region of γA3PCDH and γA4PCDH respectively. Interestingly, residue 154 is leucine in all the γA isoforms except γA3PCDH (Ile154). Yet these altered residues are strictly conserved between the orthologs of a particular isoform.

3.5.2. Implication of mutation in the linker region of human γB2PCDH

The mutation p.Arg679Pro and p.Lys687Glu in the human γB2PCDH lie in the linker region between EC6 and transmembrane domain of the protocadherin gamma (Fig. 5a). Although EC6 subdomain is involved in cis-interactions, the role of the linker region is not clear. Currently, there is no structural information available on the linker region between the EC6 and the transmembrane domain from any protocadherins subgroup or the cadherin superfamily. This region is not present in any of the constructs crystallized so far and hence it was not modeled. Arg to Pro and Lys to Glu conversions are otherwise drastic mutations and may affect folding or protein stability in general. These mutations can also result in the change in the rigidity of the linker region and may affect the formation of the stable cis and/or trans interactions.

Structure prediction of the linker region using the Predict Protein server (https://www.predictprotein.org/home) shows that this stretch has a small helix at the N and C terminus and unstructured region in-between. In fact, structure prediction of all the PcdhγB sequences yields the same result. Multiple sequence alignment of this region shows that Pro and Glu are highly conserved within primate orthologs (Fig. 5b) as well as between different isoforms of humans (Fig. 5c). Arg and Lys possibly define the identity of γB2PCDH class of protocadherin. Once these residues are mutated to Pro and Glu respectively, the identity of this class could be compromised.

In vitro studies suggest that the γPCDH can undergo cleavage by metalloproteinases near the membrane to release the ectodomain (EC1–6) (Reiss et al., 2006; Haas et al., 2004). The cleavage site for these proteases is present in this linker region. Alterations within this region may affect the binding and cleavage by these metalloproteinases.

Fig. 4. Alterations of γA4-PCDH and γA3-PCDH are present in the interacting interface of specificity determining region. (a) Schematic diagram of γA4-PCDH and (b) γA3-PCDH depicting the position of the amino acid alteration. Extracellular domain (EC), transmembrane domain (TM) and cytoplasmic domain (CP) are labeled. (c) The portion of the sequence alignment among different orthologs of γA3PCDH and (d) γA4PCDH showing interfacing region of EC2: EC3. (e) Homology model of the trans dimer of the wild type γA3PCDH (EC1–3) and (f) wild type γA4PCDH (EC1–3). The structures are depicted in ribbon and bound calcium ions are seen as cyan spheres. The residues Leu154 in γA3PCDH and Ala181 in γA4PCDH are shown in red stick. The inset shows enlarged view of the wild type proteins at the top and mutants in the lower panel. The amino acids are shown as sticks and hydrogen bond is shown as a black broken line. (g) Electrostatic surface potential of wild type and altered γA3PCDH and (h) wild type and altered γA4PCDH wild type and altered colored according to the bar underneath. The trans homophilic interaction interface is highlighted using green (EC2) and yellow (EC3) outline. Sites of alterations are marked with arrow in both cases. (i) Alignment of the interfacing region in the homology models among human γPCDH isoforms showing sequence variability within the interaction interface of EC2 and EC3.
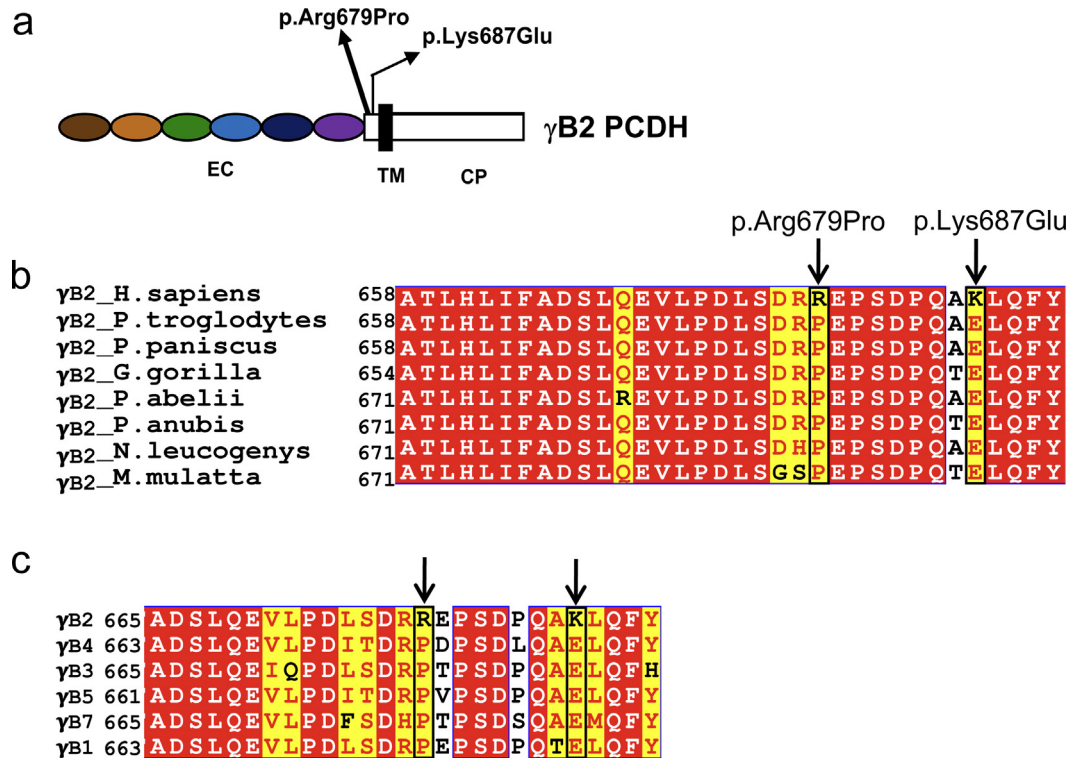
Fig. 5. Species specific isoform identity of p.Arg679Pro and p.Lys687Glu. (a) Schematic diagram of γB2PCDH depicting the position of the amino acid alteration. Extracellular domain (EC), transmembrane domain (TM) and cytoplasmic domain (CD) are labeled. (b) Multiple sequence alignment among the loop region of γB2PCDH primate orthologs showing human specific variations by arrows. (c) Multiple sequence alignment of loop region connecting EC6 and TM domain of human γB2PCDH among different isoforms. Altered residues are shown with arrow.

4. Discussion

In this study, by using two complementary genome analysis methods, we have identified a set of 17 single nucleotide polymorphisms on chr5q31.3 comprising the *PCDHG* gene cluster along with *TAF7*, *PCDH1* and *ARHGAP26* as co-segregating with dyslexia in an autosomal dominant manner (Figs. 1 & 2).

All the identified *PCDHG* variants that co-segregated with dyslexia are present in the variable exons which encode extracellular domain. It is noteworthy that extracellular domains, especially EC2 and EC3 domains of γ-PCDH are enriched with positively selected codons, probably responsible for remarkable diversity required for neuronal connections in the brain (Wu, 2005). The most striking observation was that the seven identified variants including the four non-synonymous changes, associated with dyslexia are ancestral and while their non-risk counterparts are preponderant in humans (Fig. 3). Especially, the presence of dyslexia associated variants in Neanderthal genome, provide an important indication regarding the specific cognitive attributes of modern humans. So far, the only other gene where the risk associated ancestral variants are linked to a speech related disorder is *FOXP2*, a transcription factor, expressed in the basal ganglia (Scharff and Petri, 2011). Of the non-synonymous ancestral variants identified in our study, the p. Ile154Leu and p.Ala181Thr are present in the EC2: EC3 trans-homophilic interaction interface (Fig. 4e & 4f) and our modelling study indicate a remarkable interaction specificity where a single allelic change can perturb homophilic interactions and hence the interacting network (Fig. 4f). The change in charge distribution, as well as shape of the surface for Ile154Leu and Ala181Thr, could also affect the specificity and/or affinity of the trans-homophilic interaction of protocadherins (Fig. 4g & 4h). Interestingly Ile154 is present in the interaction specific recognition domain of human γA3PCDH and might be important in conferring species specific isoform identity of the recognition region (Fig. 4c & 4i). These results reinforce the unique specificity of protocadherin gamma and their importance in building and maintaining the precise

neuronal connection in the brain. By combining modelling studies and sequence analysis we suggest a role for human specific residues in the evolution of more exquisite levels of class specific interactions.

We also made similar observations in γBPCDH with respect to p. Arg679Pro and p.Lys687Glu variants. In this case, presence of the alternate alleles (Pro and Glu) which are conserved throughout γB-PCDH paralogs (Fig. 5c) as well non-human primate orthologs (Fig. 5b), could compromise the species specific isoform identity of these variations in γB2-PCDH. The p.Arg679Pro and p.Lys687Glu alterations are present in the linker region between EC6 and transmembrane domain of γB2PCDH, which is probably the metalloprotease ADAM10 recognition site. Ectodomain shedding by ADAM10 followed by γ-secretase mediated proteolysis (Reiss et al., 2006, Haas et al., 2004) regulates the downstream signaling pathway involving adhesion kinases FAK, Pyk2, PKC, MARCKs and Rho GTPases that ultimately promote dendritic arborization (Suo et al., 2012; Garrett et al., 2012; Molumby et al., 2016). Thus the variations that predominantly present on the extracellular domain of γPCDH, could generate distinct yet diverse malfunctions of the entire sub-cellular ensemble associated with the γPCDH molecular cascade, which could ultimately affect in neuronal circuit formation.

Learning to read is a complex process and requires an organized coordination and accurate, rapid and timely integration of different neural systems relevant to cognitive and sensory process. The fundamental functions of information acquisition and processing by the brain necessitates the correct wiring of neural circuitry during development (Stiles and Jernigan, 2010). Functional neural circuit construction requires a specific and organized regulation of cell-cell interactions in almost at all developmental stages, including neuronal differentiation, neuronal migration, axon outgrowth, dendrite arborization and synapse formation and stabilization (Tau and Peterson, 2010; Weiner et al., 2013). Cell-cell recognition through cell adhesion molecules is central to establishing this coordination; as cell-type specific surface molecules provide unique cellular surface identities and molecular diversity through their extracellular interactions that ultimately determine the formation of

precise neural circuitry (Takeichi, 2007; Shapiro and Colman, 1999). Any mistake, error or mutation that leads to the formation of incorrect or altered neuronal connections can result in a number of neurodevelopmental disorders. Protocadherins are the largest cell adhesion proteins and γPCDH act as neural glue and play important roles in formation and maintenance of the neural circuit (Hasegawa et al., 2017; Weiner et al., 2013). The regulation of the gene expression of these large cluster proteins (PCDH) is underlying in the two CTCF/cohesion mediated chromatin contact domains (CCD) named α and β/γ domain. These two domains are enriched with several CTCF binding sites (CBS). The expression of each *PCDH* gene isoform is regulated by the convergent oriented CTCF binding site (CBS) that form loops between the promoters and enhancers (Guo et al., 2015). The identified non coding variants rs7730, rs13359820 and rs10491311 are located within the beta/gamma topological domain and therefore might affect the gene expression of any members of *PCDHG* gene cluster. This prediction could explain the significant e-QTL effect of these three non-coding variants on *PCDHGA1* gene expression (Fig. S3, Table S5). Another variation rs153149 is present within the CTCF binding site and would be predicted to involve to the regulation of gene *GNPDA1*. Therefore, all these identified non-coding and coding variations could affect the action of different genes both at the level of gene expression and protein interaction.

KA25, the family we have studied, appears to have a highly susceptible genetic background of chr5q31.3 for developing dyslexia. The identified amino acid alterations may not generate any drastic differences but their consequent changes could affect the connectivity of the brain. This is in concordance with dyslexia being more a result of quantitative changes that affect the ability to read effectively rather than a severe neurodevelopmental anomaly.

Therefore, considering the important associations of *PCDHG* with neurodevelopment, a substantial interpretation of this study may be that the presence of all the alternative forms collectively morph the genomic landscape to exert a combinatorial phenotype that manifests as dyslexia. Notwithstanding the highly conserved nature of the *PCDHG* gene, the observed variations support the presumption that the identified variations follow a co-evolutionary pattern with the evolution of the brain itself within the primate lineage and endow humans with the unique advantage in the process of reading. We conclude that such lineage specificity might underlie evolutionary changes in the human lineage integral to the development of neuronal networks essential for reading which is a cognitive skill unique to humans. Therefore, our results suggest a potential link between *PCDHG* with reading and highlight an essential relevance in unraveling the genetic bias that leads to the development of skill reading.

Authors' Contribution

SS and NCS conceived and directed the study. MM and MF coordinated exome sequencing and SNP array experiments where TN and RK performed the experiments. Data analysis was performed by TN and MF. DJ and PB directed and performed the molecular modelling of the study as well prepared figures for the same. NCS, MK, RM coordinated and performed the behavioral assessment. TN, SS, MM, DJ, PB, SSG and NCS wrote the manuscript. TN, SD and BP have done the bioinformatics analysis. All the analysis and results were discussed with all the authors and they have approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi. org/10.1016/j.ebiom.2017.12.031.

References

Anitha, A., Thanseem, I., Nakamura, K., Yamada, K., Iwayama, Y., Toyota, T., Iwata, Y., Suzuki, K., Sugiyama, T., Tsujii, M., Yoshikawa, T., Mori, N., 2013. Protocadherin α (PCDHA) as a novel susceptibility gene for autism. J. Psychiatry Neurosci. 38 (3): 192–198. https://doi.org/10.1503/jpn.120058.

Chen, W., Maniatis, T., 2013. Clustered protocadherins. Development 140 (16): 3297–3302. https://doi.org/10.1242/dev.090621.

Cooper, S., Jontes, J., Sotomayor, M., 2016. Structural determinants of adhesion by Protocadherin-19 and implications for its role in epilepsy. eLife 5. https://doi.org/10.7554/eLife.18529.

DeLano, W.L., 2002. Unraveling hot spots in binding interfaces: progress and challenges. Curr. Opin. Struct. Biol. 12 (1):14–20. https://doi.org/10.1016/S0959-440X(02)00283-X.

Einarsdottir, E., Svensson, I., Darki, F., Peyrard-Janvid, M., Lindvall, J., Ameur, A., Jacobsson, C., Klingberg, T., Kere, J., Matsson, H., 2015. Mutation in CEP63 co-segregating with developmental dyslexia in a Swedish family. Hum. Genet. 134 (11–12): 1239–1248. https://doi.org/10.1007/s00439-015-1602-1.

Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60 (12):2126–2132. https://doi.org/10.1107/S0907444904019158.

Fisher, S., Francks, C., Marlow, A., MacPhie, I., Newbury, D., Cardon, L., Ishikawa-Brush, Y., Richardson, A., Talcott, J., Gayán, J., Olson, R., Pennington, B., Smith, S., DeFries, J., Stein, J., Monaco, A., 2002. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. Nat. Genet. 30 (1):86–91. https://doi.org/10.1038/ng792.

Galaburda, A., Kemper, T., 1979. Cytoarchitectonic abnormalities in developmental dyslexia: a case study. Ann. Neurol. 6 (2):94–100. https://doi.org/10.1002/ana.410060203.

Galaburda, A., Sherman, G., Rosen, G., Aboitiz, F., Geschwind, N., 1985. Developmental dyslexia: four consecutive patients with cortical anomalies. Ann. Neurol. 18 (2): 222–233. https://doi.org/10.1002/ana.410180210.

Garrett, A., Schreiner, D., Lobas, M., Weiner, J., 2012. γ-Protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. Neuron 74 (2):269–276. https://doi.org/10.1016/j.neuron.2012.01.028.

Gegonne, A., Weissman, J., Zhou, M., Brady, J., Singer, D., 2006. TAF7: a possible transcription initiation check-point regulator. Proc. Natl. Acad. Sci. U. S. A. 103 (3):602–607. https://doi.org/10.1073/pnas.0510031103.

Goodman, K., Rubinstein, R., Thu, C., Bahna, F., Mannepalli, S., Ahlsén, G., Rittenhouse, C., Maniatis, T., Honig, B., Shapiro, L., 2016a. Structural basis of diverse homophilic recognition by clustered α- and β-protocadherins. Neuron 90 (4):709–723. https://doi.org/10.1016/j.neuron.2016.04.004.

Goodman, K., Rubinstein, R., Thu, C., Mannepalli, S., Bahna, F., Ahlsén, G., Rittenhouse, C., Maniatis, T., Honig, B., Shapiro, L., 2016b. γ-Protocadherin structural diversity and functional implications. eLife 5. https://doi.org/10.7554/eLife.20930.001.

Gouet, P., Courcelle, E., Stuart, D., Metoz, F., 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15 (4):305–308. https://doi.org/10.1093/bioinformatics/15.4.305.

GTEx Consortium, 2017. Genetic effects on gene expression across human tissues. Nature 550 (7675):204–213. https://doi.org/10.1038/nature24277.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M., Ren, B., Krainer, A., Maniatis, T., Wu, Q., 2015. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell 162 (4):900–910. https://doi.org/10.1016/j.cell.2015.07.038.

Haas, I., Frank, M., Veron, N., Kemler, R., 2004. Presenilin-dependent processing and nuclear function of protocadherins. J. Biol. Chem. 280 (10):9313–9319. https://doi.org/10.1074/jbc.M412909200.

Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kääriäinen, H., Kere, J., 2005. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. PLoS Genet. 1 (4):0467–0474. https://doi.org/10.1371/journal.pgen.0010050.

Hasegawa, S., Kobayashi, H., Kumagai, M., Nishimaru, H., Tarusawa, E., Kanda, H., Sanbo, M., Yoshimura, Y., Hirabayashi, M., Hirabayashi, T., Yagi, T., 2017. Clustered protocadherins are required for building functional neural circuits. Front. Mol. Neurosci. 10. https://doi.org/10.3389/fnmol.2017.00114.

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J., McMillon, G., Kolchugina, G., Black, J., Faizi, A., Deutsch, G., Siok, W., Reiss, A., Whitfield-Gabrieli, S., Gabrieli, J., 2007. Functional and morphometric brain dissociation between dyslexia and reading ability. Proc. Natl. Acad. Sci. U. S. A. 104 (10):4234–4239. https://doi.org/10.1073/pnas.0609399104.

Jiang, Y., Loh, Y., Rajarajan, P., Hirayama, T., Liao, W., Kassim, B., Javidfar, B., Hartley, B., Kleofas, L., Park, R., Labonte, B., Ho, S., Chandrasekaran, S., Do, C., Ramirez, B., Peter, C., CW, J., Safaie, B., Morishita, H., Roussos, P., Nestler, E., Schaefer, A., Tycko, B., Brennand, K., Yagi, T., Shen, L., Akbarian, S., 2017. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. Nat. Genet. 49 (8): 1239–1250. https://doi.org/10.1038/ng.3906.

Kim, S., Mo, J., Han, S., Choi, S., Han, S., Moon, B., Rhyu, I., Sun, W., Kim, H., 2010. The expression of non-clustered protocadherins in adult rat hippocampal formation and the connecting brain regions. Neuroscience 170 (1):189–199. https://doi.org/10.1016/j.neuroscience.2010.05.027.

Kostadinov, D., Sanes, J., 2015. Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. eLife 4. https://doi.org/10.7554/eLife.08964.

Krishnan, S., Watkins, K., Bishop, D., 2016. Neurobiological basis of language learning difficulties. Trends Cogn. Sci. 20 (9):701–714. https://doi.org/10.1016/j.tics.2016.06.012.

Langmead, B., Salzberg, S., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4):357–359. https://doi.org/10.1038/nmeth.1923.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25 (16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lucken-Ardjomande Hasler, S., Vallis, Y., Jolin, H., McKenzie, A., McMahon, H., 2014. GRAF1a is a brain-specific protein that promotes lipid droplet clustering and growth, and is enriched at lipid droplet junctions. J. Cell Sci. 127 (21):4602–4619. https://doi.org/10.1242/jcs.147694.

Mascheretti, S., Facoetti, A., Giorda, R., Beri, S., Riva, V., Trezzi, V., Cellino, M., Marino, C., 2015. GRIN2B mediates susceptibility to intelligence quotient and cognitive impairments in developmental dyslexia. Psychiatr. Genet. 25 (1):9–20. https://doi.org/10.1097/YPG.0000000000000068.

Mascheretti, S., De Luca, A., Trezzi, V., Peruzzo, D., Nordio, A., Marino, C., Arrigoni, F., 2017. Neurogenetics of developmental dyslexia: from genes to behavior through brain neuroimaging and cognitive and sensorial mechanisms. Transl. Psychiatry 7 (1), e987. https://doi.org/10.1038/tp.2016.240.

Meng, H., Smith, S., Hager, K., Held, M., Liu, J., Olson, R., Pennington, B., DeFries, J., Gelernter, J., O'Reilly-Pol, T., Somlo, S., Skudlarski, P., Shaywitz, S., Shaywitz, B., Marchione, K., Wang, Y., Paramasivam, M., LoTurco, J., Page, G., Gruen, J., 2005. DCDC2 is associated with reading disability and modulates neuronal development in the brain. Proc. Natl. Acad. Sci. U. S. A. 102 (47):17053–17058. https://doi.org/10.1073/pnas.0508591102.

Miller, J., Ding, S., Sunkin, S., Smith, K., Ng, L., Szafer, A., Ebbert, A., Riley, Z., Royall, J., Aiona, K., Arnold, J., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R., Dee, N., Dolbeare, T., Facer, B., Feng, D., Fliss, T., Gee, G., Goldy, J., Gourley, L., Gregor, B., Gu, G., Howard, R., Jochim, J., Kuan, C., Lau, C., Lee, C., Lee, F., Lemon, T., Lesnar, P., McMurray, B., Mastan, N., Mosqueda, N., Naluai-Cecchini, T., Ngo, N., Nyhus, J., Oldre, A., Olson, E., Parente, J., Parker, P., Parry, S., Stevens, A., Pletikos, M., Reding, M., Roll, K., Sandman, D., Sarreal, M., Shapouri, S., Shapovalova, N., Shen, E., Sjoquist, N., Slaughterbeck, C., Smith, M., Sodt, A., Williams, D., Zöllei, L., Fischl, B., Gerstein, M., Geschwind, D., Glass, I., Hawrylycz, M., Hevner, R., Huang, H., Jones, A., Knowles, J., Levitt, P., Phillips, J., Šestan, N., Wohnoutka, P., Dang, C., Bernard, A., Hohmann, J., Lein, E., 2014. Transcriptional landscape of the prenatal human brain. Nature 508 (7495):199–206. https://doi.org/10.1038/nature13185.

Mogasale, V.V., Patil, V., Patil, N., Mogasale, V., 2012. Prevalence of specific learning disabilities among primary school children in a south Indian city. Ind. J. Pediatr. 79 (3):342–347. https://doi.org/10.1007/s12098-011-0553-3.

Molumby, M., Keeler, A., Weiner, J., 2016. Homophilic protocadherin cell-cell interactions promote dendrite complexity. Cell Rep. 15 (5):1037–1050. https://doi.org/10.1016/j.celrep.2016.03.093.

Nagar, B., Overduin, M., Ikura, M., Rini, J., 1996. Structural basis of calcium-induced E-cadherin rigidification and dimerization. Nature 380 (6572):360–364. https://doi.org/10.1038/380360a0.

Nicoludis, J., Lau, S., Schärfe, C., Marks, D., Weihofen, W., Gaudet, R., 2015. Structure and sequence analyses of clustered protocadherins reveal antiparallel interactions that mediate homophilic specificity. Structure 23 (11):2087–2098. https://doi.org/10.1016/j.str.2015.09.005.

Notredame, C., Higgins, D., Heringa, J., 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. J. Mol. Biol. 302 (1):205–217. https://doi.org/10.1006/jmbi.2000.4042.

Paracchini, S., Thomas, A., Castro, S., Lai, C., Paramasivam, M., Wang, Y., Keating, B., Taylor, J., Hacking, D., Scerri, T., Francks, C., Richardson, A., Wade-Martins, R., Stein, J., Knight, J., Copp, A., LoTurco, J., Monaco, A., 2006. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. Hum. Mol. Genet. 15 (10):1659–1666. https://doi.org/10.1093/hmg/ddl089.

Patel, R., Jain, M., 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7 (2):e30619. https://doi.org/10.1371/journal.pone.0030619.

Paulesu, E., Frith, U., Snowling, M., Gallagher, A., Morton, J., Frackowiak, R., Frith, C., 1996. Is developmental dyslexia a disconnection syndrome? Brain 119 (1):143–157. https://doi.org/10.1093/brain/119.1.143.

Paulesu, E., Danelli, L., Berlingeri, M., 2014. Reading the dyslexic brain: multiple dysfunctional routes revealed by a new meta-analysis of PET and fMRI activation studies. Front. Hum. Neurosci. 8. https://doi.org/10.3389/fnhum.2014.00830.

Peter, B., Raskind, W.H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V., Wijsman, E., Brkanac, Z., 2011. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. J. Neurodev. Disord. 3 (1):39–49. https://doi.org/10.1007/s11689-010-9065-0.

Peterson, R., Pennington, B., 2015. Developmental dyslexia. Annu. Rev. Clin. Psychol. 11 (1):283–307. https://doi.org/10.1146/annurev-clinpsy-032814-112842.

Pirinen, M., Lappalainen, T., Zaitlen, N., Dermitzakis, E., Donnelly, P., McCarthy, M., Rivas, M., 2015. Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics 31 (15):2497–2504. https://doi.org/10.1093/bioinformatics/btv074.

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., Dannemann, M., Nelson, B., Key, F., Rudan, P., Kućan, Ž., Gušić, I., Golovanova, L., Doronichev, V., Patterson, N., Reich, D., Eichler, E., Slatkin, M., Schierup, M., Andrés, A., Kelso, J., Meyer, M., Pääbo, S., 2017. A high-coverage Neandertal genome from Vindija Cave in Croatia. Science 358 (6363): 655–658. https://doi.org/10.1126/science.aao1887.

Raven, J., 2000. The Raven's progressive matrices: change and stability over culture and time. Cogn. Psychol. 41 (1):1–48. https://doi.org/10.1006/cogp.1999.0735.

Redies, C., Heyder, J., Kohoutek, T., Staes, K., Van Roy, F., 2008. Expression of protocadherin-1 (Pcdh1) during mouse development. Dev. Dyn. 237 (9): 2496–2505. https://doi.org/10.1002/dvdy.21650.

Reiss, K., Maretzky, T., Haas, I., Schulte, M., Ludwig, A., Frank, M., Saftig, P., 2006. Regulated ADAM10-dependent ectodomain shedding of γ-protocadherin C3 modulates cell-cell adhesion. J. Biol. Chem. 281 (31):21735–21744. https://doi.org/10.1074/jbc.M602663200.

Roeske, D., Ludwig, K., Neuhoff, N., Becker, J., Bartling, J., Bruder, J., Brockschmidt, F., Warnke, A., Remschmidt, H., Hoffmann, P., Müller-Myhsok, B., Nöthen, M., Schulte-Körne, G., 2011. First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. Mol. Psychiatry 16 (1):97–107. https://doi.org/10.1038/mp.2009.102.

Rubinstein, R., Thu, C., Goodman, K., Wolcott, H., Bahna, F., Mannepalli, S., Ahlsen, G., Chevee, M., Halim, A., Clausen, H., Maniatis, T., Shapiro, L., Honig, B., 2015. Molecular logic of neuronal self-recognition through protocadherin domain interactions. Cell 163 (3):629–642. https://doi.org/10.1016/j.cell.2015.09.026.

Scharff, C., Petri, J., 2011. Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. Phil. Trans. R. Soc. B Biol. Sci. 366 (1574):2124–2140. https://doi.org/10.1098/rstb.2011.0001.

Schreiner, D., Weiner, J., 2010. Combinatorial homophilic interaction between -protocadherin multimers greatly expands the molecular diversity of cell adhesion. Proc. Natl. Acad. Sci. U. S. A. 107 (33):14893–14898. https://doi.org/10.1073/pnas.1004526107.

Schumacher, J., Hoffmann, P., Schmal, C., Schulte-Korne, G., Nothen, M., 2007. Genetics of dyslexia: the evolving landscape. J. Med. Genet. 44 (5):289–297. https://doi.org/10.1136/jmg.2006.046516.

Shapiro, L., Colman, D., 1999. The diversity of cadherins and implications for a synaptic adhesive code in the CNS. Neuron 23 (3):427–430. https://doi.org/10.1016/S0896-6273(00)80796-5.

Shaywitz, S., Gruen, J., Shaywitz, B., 2007. Management of dyslexia, its rationale, and underlying neurobiology. Pediatr. Clin. N. Am. 54 (3):609–623. https://doi.org/10.1016/j.pcl.2007.02.013.

Skeide, M., Kirsten, H., Kraft, I., Schaadt, G., Müller, B., Neef, N., Brauer, J., Wilcke, A., Emmrich, F., Boltze, J., Friederici, A., 2015. Genetic dyslexia risk variant is related to neural connectivity patterns underlying phonological awareness in children. NeuroImage 118:414–421. https://doi.org/10.1016/j.neuroimage.2015.06.024.

Stiles, J., Jernigan, T., 2010. The basics of brain development. Neuropsychol. Rev. 20 (4): 327–348. https://doi.org/10.1007/s11065-010-9148-4.

Suo, L., Lu, H., Ying, G., Capecchi, M., Wu, Q., 2012. Protocadherin clusters and cell adhesion kinase regulate dendrite complexity through Rho GTPase. J. Mol. Cell Biol. 4 (6):362–376. https://doi.org/10.1093/jmcb/mjs034.

Takeichi, M., 2007. The cadherin superfamily in neuronal connections and interactions. Nat. Rev. Neurosci. 8 (1):11–20. https://doi.org/10.1038/nrn2043.

Tapia-Paez, I., Tammimies, K., Massinen, S., Roy, A., Kere, J., 2008. The complex of TFII-I, PARP1, and SFPQ proteins regulates the DYX1C1 gene implicated in neuronal migration and dyslexia. FASEB J. 22 (8):3001–3009. https://doi.org/10.1096/fj.07-104455.

Tau, G., Peterson, B., 2010. Normal development of brain circuits. Neuropsychopharmacology 35 (1):147–168. https://doi.org/10.1038/npp.2009.115.

Vangone, A., Spinelli, R., Scarano, V., Cavallo, L., Oliva, R., 2011. COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes. Bioinformatics 27 (20):2915–2916. https://doi.org/10.1093/bioinformatics/btr484.

Veerappa, A., Saldanha, M., Padakannaya, P., Ramachandra, N., 2013. Genome-wide copy number scan identifies disruption of PCDH11X in developmental dyslexia. Am. J. Med. Genet. B Neuropsychiatr. Genet. 162 (8):889–897. https://doi.org/10.1002/ajmg.b.32199.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38 (16):e164. https://doi.org/10.1093/nar/gkq603.

Weiner, J., Jontes, J., Burgess, R., 2013. Introduction to mechanisms of neural circuit formation. Front. Mol. Neurosci. 6. https://doi.org/10.3389/fnmol.2013.00012.

Wu, Q., 2005. Comparative genomics and diversifying selection of the clustered vertebrate protocadherin genes. Genetics 169 (4):2179–2188. https://doi.org/10.1534/genetics.104.037606.

Wu, Q., Maniatis, T., 1999. A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell 97 (6):779–790. https://doi.org/10.1016/S0092-8674(00)80789-8.

Web resources

1000 genome, d. http://www.1000genomes.org.
dbSNP, d. http://www.ncbi.nlm.nih.gov/projects/SNP.
UCSC Genome Browser, d. https://genome.ucsc.edu.
Ancient Genome Browser, d. https://bioinf.eva.mpg.de/jbrowse.

OMIM, d. http://www.omim.org.
NCBI, d. http://www.ncbi.nlm.nih.gov.
NCBI Blast, d. http://blast.ncbi.nlm.nih.gov/Blast.cgi.
NCBI Primer Blast, d. http://www.ncbi.nlm.nih.gov/tools/primer-blast.
IGV, d. http://www.broadinstitute.org/igv.
GenomeStudio, d. http://www.illumina.com/techniques/microarrays/array-data-analysis experimental-design/genomestudio.html.
ANNOVAR, d. http://www.openbioinformatics.org/annovar.
PolyPhen-2, d. http://genetics.bwh.harvard.edu/pph2.
SIFT, d. http://sift.bii.a-star.edu.sg.
Clustal Omega, d. http://www.ebi.ac.uk/Tools/msa/clustalo.
T-coffee, d. http://tcoffee.crg.cat.
PDB, d. http://www.rcsb.org/pdb/home/home.do.
Predict Protein server, d. https://www.predictprotein.org/home.
GTEx Portal, d. https://www.gtexportal.org/home/.
Allen Brain Atlas, d. http://www.brain-map.org/.
DALI, d. http://pib.nic.in/newsite/PrintRelease.aspx?relid=128722 http://14.139.62.11/DALI/details.php.