Human Genetics (2023) 142:1201–1213
https://doi.org/10.1007/s00439-022-02482-5

ORIGINAL INVESTIGATION

Overlapping pathogenic de novo CNVs in neurodevelopmental disorders and congenital anomalies impacting constraint genes regulating early development

Seyed Ali Safizadeh Shabestari[1] · Nasna Nassir[1] · Samana Sopariwala[2] · Islam Karimov[3] · Richa Tambi[1] · Binte Zehra[1] · Noor Kosaji[1] · Hosneara Akter[4] · Bakhrom K. Berdiev[1] · Mohammed Uddin[1,5]

Received: 21 June 2022 / Accepted: 21 August 2022 / Published online: 16 November 2022

Abstract

Neurodevelopmental disorders (NDDs) and congenital anomalies (CAs) are rare disorders with complex etiology. In this study, we investigated the less understood genomic overlap of copy number variants (CNVs) in two large cohorts of NDD and CA patients to identify de novo CNVs and candidate genes associated with both phenotypes. We analyzed clinical microarray CNV data from 10,620 NDD and 3176 CA cases annotated using Horizon platform of GenomeArc Analytics and applied rigorous downstream analysis to evaluate overlapping genes from NDD and CA CNVs. Out of 13,796 patients, only 195 cases contained 218 validated de novo CNVs. Eighteen percent (31/170) de novo CNVs in NDD cases and 40% (19/48) de novo CNVs in CA cases contained genomic overlaps impacting developmentally constraint genes. Seventy-nine constraint genes (10.1% non-OMIM entries) were found to have significantly enriched genomic overlap within rare de novo pathogenic deletions (P value = 0.01, OR = 1.58) and 45 constraint genes (13.3% non-OMIM entries) within rare de novo pathogenic duplications (P value = 0.01, OR = 1.97). Analysis of spatiotemporal transcriptome demonstrated both pathogenic deletion and duplication genes to be highly expressed during the prenatal stage in human developmental brain (P value = 4.95 X 10^{-6}). From the list of overlapping genes, *EHMT1*, an interesting known NDD gene encompassed pathogenic deletion CNVs from both NDD and CA patients, whereas *FAM189A1*, and *FSTL5* are new candidate genes from non-OMIM entries. In summary, we have identified constraint overlapping genes from CNVs (including de novo) in NDD and CA patients that have the potential to play a vital role in common disease etiology.

Introduction

Neurodevelopmental disorders (NDDs) and congenital anomalies (CAs) are commonly reported as a collection of rare disorders with a strong genetic basis (Casanova et al. 2018; Akter et al. 2021). NDDs are characterized by disruptions in tightly coordinated events of brain development that hinder achieving emotional, cognitive, and motor developmental milestones (Parenti et al. 2020). For example, gene mutations that occur in synaptic proteins, neurexin 1 (*NRXN1*) and *SHANK3* have been associated with the development of autism in early childhood (Walsh et al. 2008). NDDs constitute attention deficit hyperactivity disorder (ADHD), intellectual disability (ID), communication disorders, epilepsy, and autism spectrum disorder (ASD) (Mullin et al. 2013; Hu et al. 2014; Nassir et al. 2021). In contrast, CAs include a broad range of visible abnormalities of body structure or function that exist at birth with a prenatal origin (World Health Organization 2020). CAs are a broad umbrella of disorders consisting of congenital heart defects (CHDs), microcephaly, and dysmorphic features such as cleft palate among others (Dolk et al. 2010; Duncan and Chodirker 2011; Kaminsky et al. 2011; DeSilva et al. 2016; Ameen et al. 2018).

Although there is no universally accepted phenotypic criteria to differentiate these two broad pathologies (Sugranyes et al. 2011; Owoeye et al. 2013; Toufaily et al. 2018),

✉ Mohammed Uddin
mohammed.uddin@mbru.ac.ae

1 Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, UAE

2 University of Guelph, Toronto, ON, Canada

3 University of Bremen, Bremen, Germany

4 Genetics and Genomic Medicine Centre, NeuroGen Healthcare, Dhaka, Bangladesh

5 GenomeArc Inc, Toronto, ON, Canada

they are further categorized into different disease entities based on phenotype since there are no specific biomarkers to diagnose or differentiate between different NDDs and CAs (American Psychiatric Association 2013). However, they are strongly interlinked through their phenotypic pathogenesis and complications. For instance, patients with CHDs are at higher risk of developing NDDs, with a 20% chance of progression for mild CHD patients, and a higher than 50% probability for severe cases (Marino et al. 2012). This is likely a result of poor defected blood flow to the brain that compromises oxygen delivery, in turn affecting brain development (Perles et al. 2015; Ta-Shma et al. 2018).

To better understand the overlap of such phenotypes, it is important to investigate the underlying genomic interrelations. There is a host of genomic disorders (Bragin et al. 2014; Uddin et al. 2016) related to large structural variants that often present phenotypes that manifest with different disorders. For example, 15q13.3 microdeletion syndrome manifests in epilepsy, autism, and schizophrenia (Uddin et al. 2018) with varying frequency. These phenotypically overlapped genomic regions are comprised of genes that are highly constraint and might be involved in regulating different pathways related to multiple phenotypes. NDD and CA are phenotypically distinct yet co-occur often among rare disorders. For example, there are genomic regions that have been reported from both NDD and CA cases such as 22q11.2 microdeletion syndrome (McDonald-McGinn et al. 2015), which in some cases develop congenital heart diseases as a primary phenotype. However, there are also cases of 22q11.2 microdeletion syndrome with no apparent congenital anomalies (Rozas et al. 2019). Therefore, it is interesting to identify these overlapping regions from different phenotypes and delineate the pathways as the constraint genes underlying these overlapping phenotypic co-morbidities are still largely unknown.

In this study, we used phenotypically characterized large NDD and CA cohort data to identify the genes within the overlapping genomic regions impacted by de novo and rare CNVs. By applying pathway analysis and using human developmental transcriptome data, we found these genes to be associated with altered neural connectivity and selective tissue formation. Identification of the shared pathogenic mechanisms between NDDs and CAs will assist in effective diagnosis and targeted therapeutics.

Materials and methods

Sample details

Clinical microarray data were collected from an NDD cohort ($n = 10{,}620$) (Uddin et al. 2016) with unrelated patients reported with phenotypes of autism spectrum disorder, language/speech delays, developmental delay, learning disability, mental retardation, seizures, or hypotonia. Our second cohort comprised of unrelated cases ($n = 3176$) consisting of a heterogeneous population carrying rare CAs with phenotypes of dysmorphic features, cleft palate, congenital heart defect, hypoplastic right heart, microcephaly, and (multiple) congenital anomalies (Uddin et al. 2016). These two cohorts were recruited from SickKids hospital (total cases $n = 8929$; NDD cases $n = 7107$; CA cases $n = 1822$) in Toronto, and Credit Valley Hospital (total cases $n = 4867$; NDD cases $n = 3513$; CA cases $n = 1354$) in Mississauga, respectively (Fig. 1, Table 1). Both cohorts consisted of a heterogenous population with similar geographic and socioeconomic backgrounds. Inclusion criteria for NDD samples comprised of the presence of any neurodevelopmental disorder as the primary phenotype which were documented by diagnostic behavior, phenotypes, and chromosomal microarray analysis. Regarding CAs cohort, the primary phenotype was reported as multiple congenital anomalies and congenital heart defects. There also exists the possibility that the CA patients might have some degree of NDD phenotype as a secondary manifestation that may have been under-reported.

Chromosomal microarray analysis

A circular binary segmentation algorithm (Olshen et al. 2004) was applied on obtained clinical microarray data from both hospitals using International Standards for Cytogenomic Arrays ISCA 180 K comparative genomic hybridization array (aCGH) to detect large CNVs. To compare individual probe intensities, we used a pool of ten samples for reference. Each sample variant was annotated by employing numerous tools, including ANNOVAR (Wang et al. 2010) and Horizon platform of GenomeArc Analytics. The clinical laboratory geneticist manually annotated pathogenicity (pathogenic, likely pathogenic, variant of unknown significance (VUS), likely benign) of each CNV applying American College of Medical Genetics (ACMG) guidelines (Kearney et al. 2011). CNVs smaller than 10 Kb and larger than 10 Mb were excluded from analysis. The original dataset had de novo variant information, where parent DNA was accessible (Uddin et al. 2016).

Control dataset

In this study, data from 9692 unrelated samples (Uddin et al. 2016) with no known psychiatric history have been used as population control (Fig. 1). This was collected from several major population-scale studies that utilized high-resolution microarray platforms. Illumina 1 M from the Study of Addiction Genetics and Environment (SAGE) (Bierut et al. 2010) and the Health, Aging, and Body Composition (HABC) (Coviello et al. 2012) assayed 4347

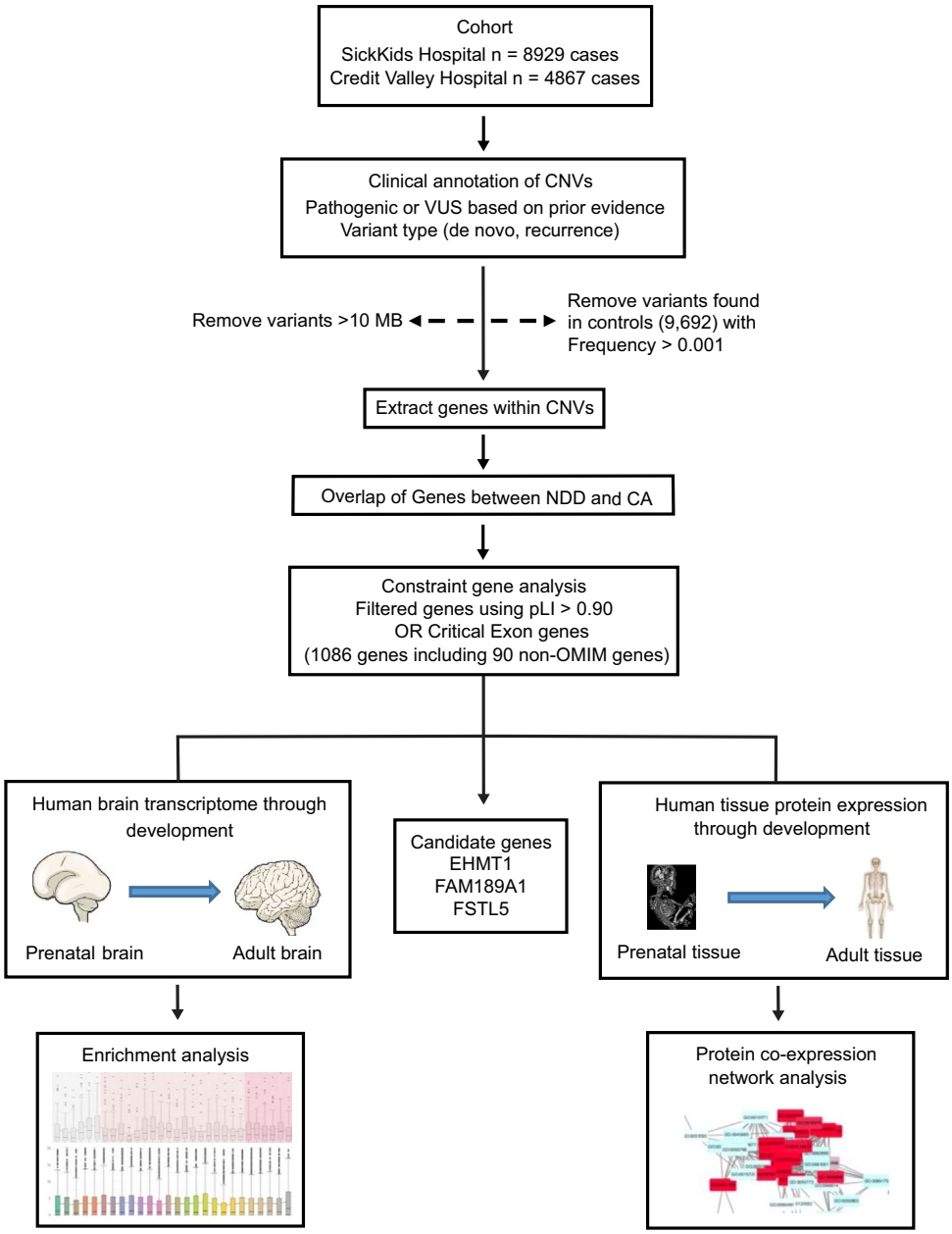

Fig. 1 Schematic of study framework to identify and analyze overlapping (NDD and CA), constraint, candidate, and non-OMIM genes. Steps to identify candidate set of phenotypically relevant genes impacted by rare copy number variations (CNVs) (pathogenic/VUS; deletion/duplication) in neurodevelopmental disorder and congenital anomaly cases. Initial filtering was carried out with control variants (frequency) and constraint filtering using CE or pLI measures, and then performing enrichment in developmental human (prenatal, early childhood, and adult) brain transcriptome (RNA sequencing). *CNVs* copy number variations, *NDD* neurodevelopmental disorder, *VUS* variant of unknown significance

Table 1 Demographics of the two cohorts specifying the breakdown in number of patients by phenotype, hospital, and gender

	SickKids Hospital	Credit Valley Hospital
Neurodevelopmental disorders		
Male	4862	2417
Female	2245	1096
Congenital anomalies		
Male	958	763
Female	865	590

control samples; Illumina Omni 2.5 M from the Cooperative Health Research in the Region of Augsburg KORA projects (Verhoeven et al. 2013) and Collaborative Genetic Study of Nicotine Dependence (COGEND) (Bierut et al. 2007) assayed 2988 control samples; and Affymetrix 6.0 from the PopGen project (Krawczak et al. 2006) and the Ottawa Heart Institute (Stewart et al. 2009) assayed 2357 control samples. Using a high-resolution control will allow us to improve false positive calls from the ISCA low resolution case cohorts and will provide convincing association signals.

Gene set curation and overlap analysis

We used the GRCh37/hg19 build and unique coding sequence (CDS) ids for identifying regions of the DNA that encode for proteins, and removing repeats or duplicates, to analyze our data. CNVs were interpreted based on probable clinical significance or pathogenicity, variant type (deletion, duplication), inheritance (familial, de novo), gender (male, female), phenotype (NDD, CA), gene density and content (Additional file 1: Suppl. Fig. 1). First, we extracted the genes from the control dataset with frequency > 0.001 using the respective CDS ids, and similarly, we extracted the genes from the respective CNVs using the CDS ids. Subsequently, all gene overlaps between the control gene lists and patient gene lists were removed. In addition, the remaining genes extracted from CNVs based on gender, pathogenicity, and type were compiled (Fig. 2). We performed Fisher's exact test (FET) using the R package (GeneOverlap) to measure statistical significance (P value < 0.05).

Proteomic and multi-tissue transcriptome expression analysis

We used proteomic data from human protein expression studies at different developmental stages and expression data from multiple tissues to further analyze the genes that were extracted from NDD CNVs and had no overlap with genes from the CA CNVs. Proteomic and multi-tissue transcriptome datasets are described in detail in the following section.

Proteomic data analysis

To analyze protein expression levels at two developmental stages in human tissues, we used high-resolution genome-wide Fourier transform mass spectrometry data (downloaded from the Human Proteome Map) (Kim et al. 2014), including in-depth proteomic profiling of 30 histologically normal human samples: 7 fetal tissues (heart, liver, gut, ovary, testis, brain, and placenta), and 18 adult tissues (frontal cortex, spinal cord, retina, heart, liver, ovary, testis, lung, adrenal, gallbladder, pancreas, kidney, esophagus, colon, rectum, urinary bladder, and prostate), and 6 hematopoietic adult cells (B cells, CD4 cells, CD8 cells, NK cells, monocytes, and platelets) (Additional file 1: Suppl. Fig. 5) (Kim et al. 2014). For processing the data, fragmentation (high–high mode) was applied using the high-resolution Fourier transform mass spectrometers, identifying the proteins encoded by 17,294 genes, which accounts for 84% of annotated protein-coding human genes (Kim et al. 2014). For measuring protein expression, we used spectral counts per gene per sample. We performed Fisher's exact t-test for the overlapped NDD and CA gene lists with CE or pLI enrichment using the R package (GeneOverlap) to measure statistical significance.

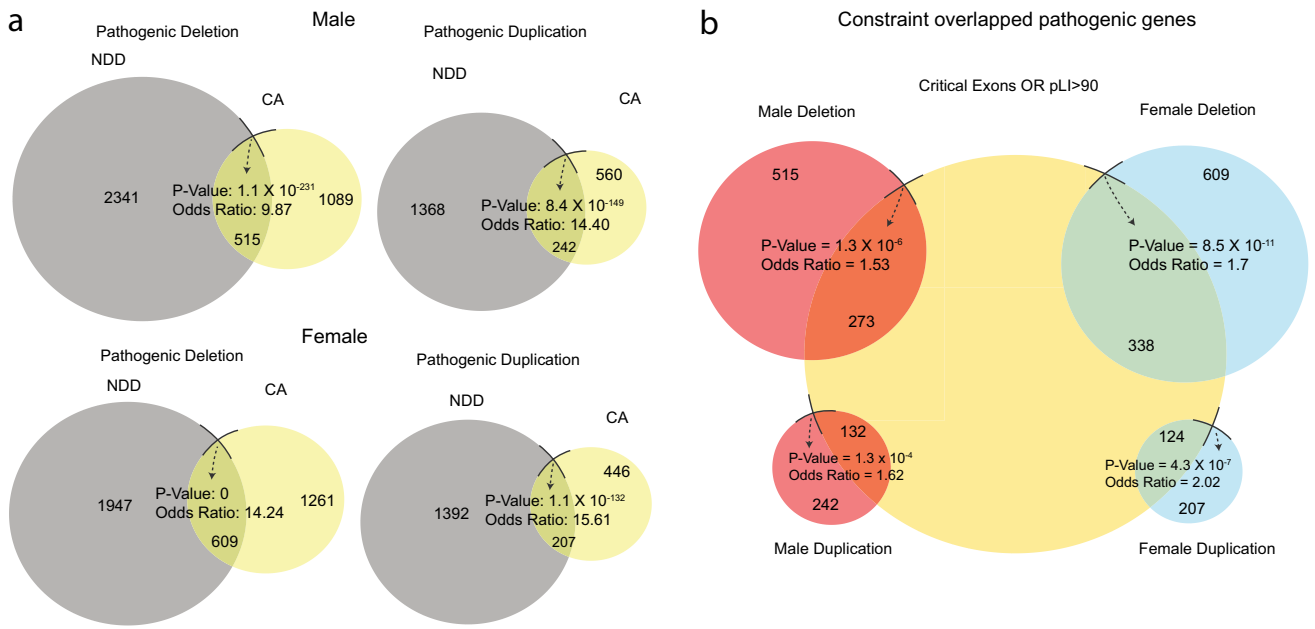

Fig. 2 Diagrams displaying the significance of the overlap between neurodevelopmental disorder and congenital anomaly cases. **a** Venn diagram displaying the significance (FET: P value and Odds Ratio) of overlapped pathogenic gene lists between neurodevelopmental disorder (NDD) and congenital anomaly (CA) CNVs in males and females, before filtering with constraint measures (CE and pLI). **b** Venn diagram displaying the significance (FET, P value and Odds Ratio) of overlap between constraint (CE or pLI) gene list and the respective genes extracted from NDD and CA CNVs present in males and females, respectively. *FET* Fisher's exact test

Multi-tissue transcriptome analysis

We measured expression levels (in triplicate) using Affymetrix GeneChip Human Exon 1.0 ST array (Gardina et al. 2006) and transcriptomes from cerebellum, breast, heart, liver, muscle, kidney, thyroid, pancreas, prostate, spleen, and testis, removing probes prone to multiple hybridizations. We used the Robust Multi-array Average (RMA) algorithm (Irizarry et al. 2003) to subtract the background signal and normalized the log2 expression values for each exon. Expression of 16,713 RefSeq genes was surveyed in all 11 tissues. A log2-transformed intensity threshold of ≥ 6 to define the expression (Kang et al. 2011) was used to detect 16,411 genes with at least one exon expressed in a tissue sample. Reads per kilobase of transcript per million (RPKM) was used as the expression unit for exons from the mapped reads (Additional file 1: Suppl. Fig. 6).

The CNVs chosen for proteomic and multi-tissue transcriptome data were genes from NDD pathogenic deletion CNVs that were not overlapped with CA gene list.

Constraint gene analysis and data filtering

We have defined 'constraint genes' in our analysis if a gene present in both NDD and CA CNVs had a significant overlap with either critical exon (CE) or pLI ≥ 0.9 (probability of being Loss of Function intolerant). CE and pLI filtering methods are described in detail in the following section.

Spatiotemporal expression data from human brain and critical exons (CE)

Critical exons are highly expressed exons with low mutation burden. For this project, we have recalculated critical exon matrix based on our previous work (Uddin et al. 2016). For deleterious genes that harbor de novo mutations, critical exons were significantly enriched in individuals with ASD relative to their siblings without ASD (Uddin et al. 2014). We utilized these highly specific set of genes (critical exon genes) derived from computing exon level spatiotemporal RNA-seq expression of 388 tissue samples (derived from 42 different brain donors). RPKM was used as the expression unit for exons from the mapped reads. The selection of donors was made to include at least two sex and aged-matched donors, and each developmental period: prenatal (8–37 weeks post-conception), early childhood (10 months to 15 years), and adulthood (> 17 years). We derived the expression data of 16 brain regions within 3 developmental periods for each donor (Fig. 3 and Additional file 1: Suppl. Fig. 7). We used gnomAD to identify the non-synonymous rare (< 0.01 frequency) mutation burden. An exon is categorized as 'critical exon' if its expression is high (> 75th percentile) and gnomAD population non-synonymous mutation burden is low (< 75th percentile) compared to the entire dataset. A gene is considered a 'critical exon gene' if one or more exons were annotated as 'critical exon' for at least 50 RNA-seq brain samples.

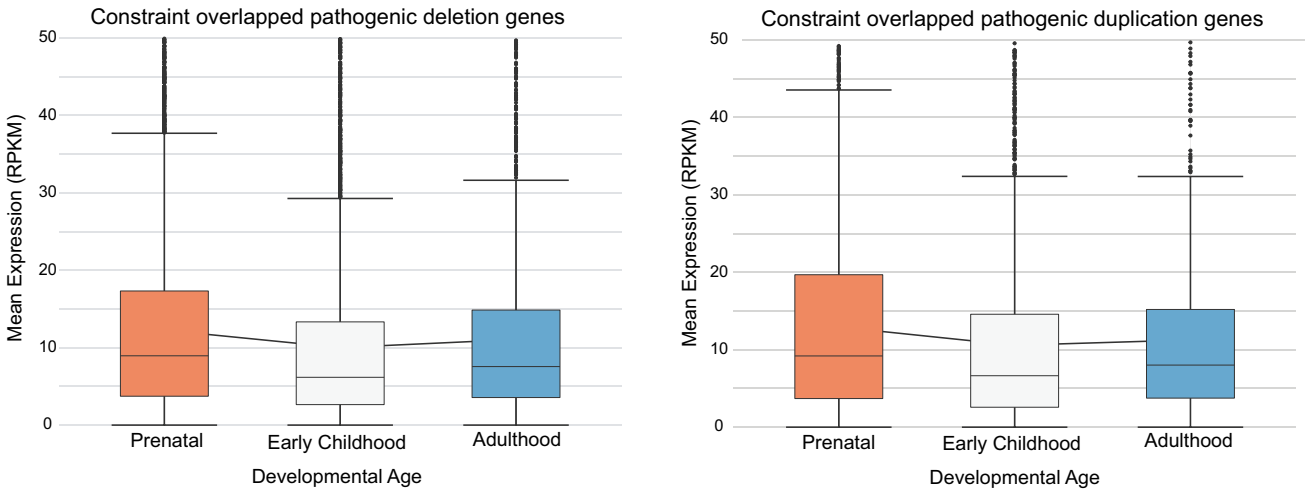

Fig. 3 Developmental transcriptomics of constraint pathogenic genes. Boxplots displaying the expression of constraint overlapped (NDD and CA) genes extracted from pathogenic deletion and duplication CNVs in developmental brain transcriptome data of prenatal (8–37 weeks post-conception), early childhood (10 months to 15 years), and adult subjects (> 17 years) displayed on the x-axis. Y-axis represents normalized gene expression in Reads per kilobase per million (RPKM) units. Boxplots showing median, interquartile range (IQR) with whiskers adding IQR to the 1st and 3rd quartile, and the line connecting the three boxes is comparing the mean expression of the three developmental periods

pLI

As a second filtering criteria for constrained genes, we obtained pLI scores from Exac database to identify the tolerance of a susceptible gene to loss of function and pLI ≥ 0.9 are extremely LoF intolerant (Lek et al. 2016). We performed Fisher's exact *t*-test for the overlapped gene lists (NDD and CA) with CE or pLI enrichment using the R package (GeneOverlap) to measure statistical significance.

Pathway enrichment analysis

We performed the enrichment analysis of the most significant gene overlaps from the respective type of CNVs to determine the major pathways in which the constrained genes were expressed. We scanned the KEGG pathway database which comprises of an assembly of the up-to-date interactions, reactions, and relations of molecular networks (http://www.genome.jp/kegg/pathway.html) and GO database (http://geneontology.org/) to identify all the pathways in which five or more genes (from the constraint gene set) were expressed. Only the pathways having more than 50 genes and less than 1000 genes were considered for this analysis. We called a gene set enriched if it overlapped between our gene set and the KEGG-GO pathway database with significance (Fischer Exact Test (FET)). The pathways were identified by their unique KEGG ID and name. The significant pathways ($P < 0.05$) with a false discovery rate (FDR) < 0.01 were used to construct the pathway network map using Cytoscape (https://cytoscape.org/) for visualization.

Results

Overlapping genes extracted from de novo CNVs in NDD and CA cases

After analyzing a total of 13,796 patients from both cohorts, 195 patients (151 NDD cases (77% of de novo cases) and 44 CA cases (23% of de novo cases)) contained a total of 218 validated de novo CNVs from which 50 de novo CNVs overlapped between the NDD and CA cases (18% (31/170) of de novo NDD CNVs and 40% (19/48) of de novo CA CNVs have overlapping genomic regions impacting developmentally constraint genes) (Additional file 2: Suppl. Table 1). The phenotypes of the de novo CNVs that overlapped were developmental delay, multiple congenital anomalies, and autism (Additional file 1: Suppl. Fig. 4, Additional file 2: Suppl. Table 1). After filtering small (CNVs < 10 Kb) and large (CNVs ≥ 10 Mb) variants, 126 pathogenic CNVs (97 NDD and 29 CA) (Additional file 1: Suppl. Fig. 1b), and 80 VUS CNVs (65 NDD and 15 CA) were retained (Additional file 1: Suppl. Fig. 2a). Larger CNVs (size range of 1–5 Mb) were most prevalent compared to smaller CNVs (Range < 1 Mb). NDD pathogenic CNVs were present across 55 male and 36 female, and CA pathogenic CNVs across 15 male and 14 female cases.

Genes per de novo variants averaged mostly in the 1–50 kb range, comprising more than 70% of the number of genes in each respective exonic variant size category (Additional file 1: Suppl. Fig. 1b, Additional file 1: Suppl. Fig. 2a). After filtering out the genes from the de novo CNVs by overlapping with control gene set, we discovered 138 de novo genes to be impacted by at least one pathogenic deletion (P value = 2.87×10^{-90}, OR = 14.69) in CNVs containing both NDD and CA cases. Similarly, significant overlap of 72 genes from the de novo pathogenic duplications (P value = 9.7×10^{-62}, OR = 22.12) were found. The overlap of de novo VUS deletion gene set with control were not significant (P value = 0.14, OR = 6.41), so is the overlap of genes from de novo NDD and CA VUS duplications after filtering with control genes (P value = 1, OR = 0); therefore, we decided to provide two separate results: (1) unfiltered and (2) filtered using the controls and the pre-determined criteria. In the unfiltered analysis, there was a significant overlap of 13 de novo VUS deletion genes (P value = 5.65×10^{-16}, OR = 38.67) between the NDD and CA cases but not among the duplications (P value = 0.08, OR = 2.51). Whereas after filtering, there was no significant overlap between the union of de novo VUS deletion gene sets (NDD and CA) (post-filtering with controls) and pLI and CE filters (as described in the methods) (P value = 0.43, OR = infinite). Similarly, there was no significance between the overlap of genes from de novo NDD and CA VUS duplications after filtering (P value = 1, OR = 0).

Constraint overlapping genes from the pathogenic NDD and CA de novo CNVs

After constraint gene filtering analysis, significant overlap was observed for 79 (10.1% non-OMIM entries) de novo pathogenic deletion CNV affected genes (P value = 0.01, OR = 1.58) and 45 (13.3% non-OMIM entries) de novo pathogenic duplication affected genes (P value = 0.01, OR = 1.97) (Additional file 2: Suppl. Table 2a/2b).

Overlapping genes in NDD and CA CNVs across gender (pre-filtering)

Male

Five hundred and fifteen genes with at least 1 exon impacted by pathogenic deletion CNVs in male show significant (P value = 1.1×10^{-231}, OR = 9.87) overlap between the genes from NDD and CA cases (Fig. 2a). Similarly, significant overlap was also observed for 242 genes from pathogenic

duplications in male (P value = 8.4×10^{-149}, OR = 14.40) (Fig. 2a). The VUS gene lists had 112 intersected genes from (deletions that showed significant (P value = 1.2×10^{-58}, OR = 9.53) the overlap between the NDD and CA cases in male (Additional file 1: Suppl. Figure 3a). Significant overlap was also observed for 320 VUS duplications in male (P value = 1.1×10^{-99}, OR = 4.47) (Additional file 1: Suppl. Figure 3a).

Female

Six hundred and nine genes with at least 1 exon impacted by pathogenic deletions in female show significant (P value < 0, OR = 14.24) overlap from NDD and CA cases (Fig. 2a). Similarly, significant overlap was also observed for 207 genes within pathogenic duplication CNVs in female (P value = 1.1×10^{-132}, OR = 15.61) (Fig. 2a). The overlapped VUS gene lists had 64 intersected genes with at least one exon impacted by deletions (P value = 5.8×10^{-31}, OR = 7.60) between the NDD and CA cases in female (Additional file 1: Suppl. Figure 3a). Significant overlap was also observed for 168 genes within VUS duplications in female (P value = 9.8×10^{-42}, OR = 4.06) (Additional file 1: Suppl. Fig. 3a).

Constraint genes within the overlapped NDD and CA cases across gender (post-filtering)

Male

After applying CE and pLI constraint gene thresholds (detailed in Methods), 273 overlapped genes with at least 1 exon impacted by pathogenic deletions were found in male (P value = 1.30×10^{-6}, OR = 1.53) with both NDD and CA cases (Fig. 2b). Significant overlap was also observed for 132 constraint genes impacted by pathogenic duplications in male (P value = 1.3×10^{-4}, OR = 1.62) (Fig. 2b). After constraint filtering of the overlapped VUS gene lists, 46 genes with at least 1 exon impacted by deletions showed no significant (P value = 0.67, OR = 0.93) overlap between the NDD and CA cases (Additional file 1: Suppl. Fig. 3b). Significant overlap was observed for 320 constraint genes impacted by VUS duplications (P value = 6.9×10^{-4}, OR = 1.37) (Additional file 1: Suppl. Figure 3b).

Female

After filtering the constraint overlapped gene lists, 338 genes with at least 1 exon impacted by pathogenic (OR = 1.69) deletions showed significant (P value = 8.5×10^{-11}) overlap between the genes from NDD and CA cases in female (Fig. 2b). Similarly, significant overlap was also observed for 124 intersected genes from constraint pathogenic duplications (P value = 4.3×10^{-7}, OR = 2.02) (Fig. 2b). After constraint filtering of the overlapped VUS gene lists, 64 genes with at least 1 exon impacted by deletions showed no significant (P value = 0.11, OR = 1.84) overlap between the NDD and CA cases (Additional file 1: Suppl. Figure 3b). Significant overlap was observed for 89 intersected genes from constraint VUS duplications (P value = 4.5×10^{-3}, OR = 1.52) (Additional file 1: Suppl. Fig. 3b).

X-chromosome analysis

Among the CNVs that impacted constraint overlapped gene lists the X-chromosome were pathogenic duplication and VUS duplication CNVs found in both males and females (Additional File 2: Suppl. Table 3). There were 11 pathogenic duplication CNVs (7 NDD, 4 CA CNVs) in males and 4 pathogenic duplication CNVs (2 NDD, 2 CA CNVs) in females, and 13 VUS duplication CNVs (7 NDD, 6 CA CNVs) in males and 4 VUS duplication CNVs (1 NDD, 3 CA CNVs) in females (Additional File 2: Suppl. Table 3). There were no details available on the specific phenotypes other than the broad category.

Expression of constraint NDD genes in developmental brain and multi-tissue transcriptome and proteome

Analysis of the developmental brain transcriptome data demonstrated prenatal expression to be the highest for both constraint overlapped pathogenic deletion (P value = 4.95×10^{-6}) and duplication genes (P value = 0.01), followed by adulthood, and early childhood, respectively (Fig. 3). Differential proteomic tissue expression demonstrated that the adult testis and adult retina have the highest expression in pathogenic deletion and duplication genes, respectively (Additional file 1: Suppl. Fig. 5). Differential transcriptomic tissue expression was non-specific for NDD pathogenic deletion and duplication genes (Additional file 1: Suppl. Fig. 6).

Candidate gene-specific mutation data

We identified 1086 constrained genes whose mutation might contribute to NDD and CA phenotypes. Three unique candidate genes, *EHMT1*, *FAM189A1*, and *FSTL5*, were chosen, with the former selected from the overlapped pathogenic deletion genes list, identified in the respective CNVs less than 1 Mb. *EHMT1* was found in four deletion CNVs (the highest frequency from our CNV data) and this CNV was considered pathogenic with respect to both NDD and CA phenotype (Frega et al. 2019). The remaining two novel candidate genes were selected from the significant overlapped gene lists after filtering with CE and pLI > 0.90

that contained no Online Mendelian Inheritance in Man (OMIM) entries (total of 90 unique genes impacted by 90 CNVs) (Additional file 2: Suppl. Table 2a) and had the highest number of gene-specific CNVs in the literature. We reviewed additional cohorts in DECIPHER (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources) and publications (Pubmed) (Additional file 2: Suppl. Table 4).

Candidate gene: *EHMT1*

Pathogenic deletions in our cohort within chromosome region 9q34.3 affected the gene, euchromatic histone methyltransferase 1 (*EHMT1*) (Fig. 4a). The EHMT1 protein is known to control brown adipose cell fate and is an essential brown adipose tissue (BAT)-enriched lysine methyltransferase in the PRDM16 transcriptional complex (Ohno et al. 2013). From clinical cohorts (DECIPHER), we found enrichment of CNVs less than 10 Kb affecting *EHMT1* among cases (6 deletions, 1 triplication, and 46 single-nucleotide variants, including 33 de novo) (Fig. 4a/d) (Additional file 2: Suppl. Table 4a). Schaefer et al. reported that knock-out EHMT1 -/- mice decreased euchromatic H3K9 methylation in the forebrain and upregulation of neuronal and non-neuronal genes, especially affecting those involved in developmental stage-dependent gene expression (Schaefer et al. 2009). Moreover, the KO mice displayed defects in learning and memory, and demonstrated *EHMT1* to be a key regulator of transcriptional homeostasis of cognition and adaptive behavior (Schaefer et al. 2009).

Candidate gene: *FAM1891A*

After enriching the overlapped gene lists with critical exons and pLI, we formulated a non-OMIM gene list from which *FAM189A1* (family with sequence similarity 189 member A1) (Fig. 4b/e) was the only de novo gene that contained at least two gene-specific CNVs (a total of five deletions) (Fig. 4b/e). It was also present in the de novo pathogenic deletion list (Additional file 2: Suppl. Table 2b), and in both male and female overlapped gene lists (Additional file 2: Suppl. Table 2c and d). In clinical cohorts (DECIPHER), we found enrichment of CNVs less than 1 Mb affecting *FAM189A1* among cases (five deletions and five duplications) (Additional file 2: Suppl. Table 4b). The gene is expressed in the pancreatic tissue (specialized epithelial cells) and thyroid gland with single-cell type specificity in the neuronal cells of the brain (Human Protein Atlas (http://proteinatlas.org)) (Uhlén et al. 2015). In a study conducted by Murray et al. on genome-wide association between individuals with life-threatening arrhythmia and normal controls in the span of at least 3 years (Murray et al. 2012), the highest P value of 5.0×10^{-6} and odds ratio of 2.02 were located in the gene *FAM189A1*.

Candidate gene: *FSTL5*

FSTL5 (Follistatin-like 5) is the other non-OMIM entry candidate gene within the overlapped gene lists that are enriched with critical exons and pLI > 0.9 (Additional file 2: Suppl. Table 2a). It was only identified in the female pathogenic deletion overlapped gene lists. From published data and in clinical cohorts (DECIPHER), we found enrichment of CNVs less than 1 Mb affecting *FSTL5* among cases (three deletions and two duplications) (Fig. 4c/f) (Additional file 2: Suppl. Table 4c). FSTL5 is hypothesized to be an extracellular protein with roles in enabling calcium ion binding activity and cell differentiation [provided by Alliance of Genome Resources, Apr 2022 (Agapite et al. 2020)]. The gene is expressed in retina and brain according to the Human Protein Atlas (http://proteinatlas.org) (Uhlén et al. 2015). Studies have shown an array of functions for FSTL5 in the human body, ranging from inhibiting the progression of hepatocellular carcinoma (Zhang et al. 2015, 2020; Li et al. 2018) to being a marker of poor prognosis in Non-WNT/Non-SHH medulloblastoma (Remke et al. 2011).

Pathways enriched in NDD and CA gene sets

The genes in constraint overlapped (NDD and CA) de novo pathogenic deletion CNVs are enriched in important pathways such as cellular DNA repair, cellular junction organization, and methyl transferase activity (Fig. 5). NDD pathogenic deletion genes were enriched in biological pathways that include transmembrane ion transport, photoreceptor cilium activity, and organ system development (Additional file 1: Suppl. Fig. 8). However, the constraint genes in overlapped pathogenic deletion CNVs were significantly enriched in chemical synaptic transmission, catabolic activity, morphogenesis and differentiation (Additional file 1: Suppl. Fig. 9). These enriched pathways demonstrate the involvement of the overlapped genes in both NDD- and CA-related pathways.

Discussion

The complex molecular interaction of genes may underly the phenotypic heterogeneity that may impact various developmental pathways [37]. In this study, we have used two large cohorts of NDD and CA patients to identify de novo variants and their associated constraint genes. We identified a core set of overlapping constraint genes that can help explain the complex molecular etiology of NDD and CA. Our result shows that these constraint genes (i)

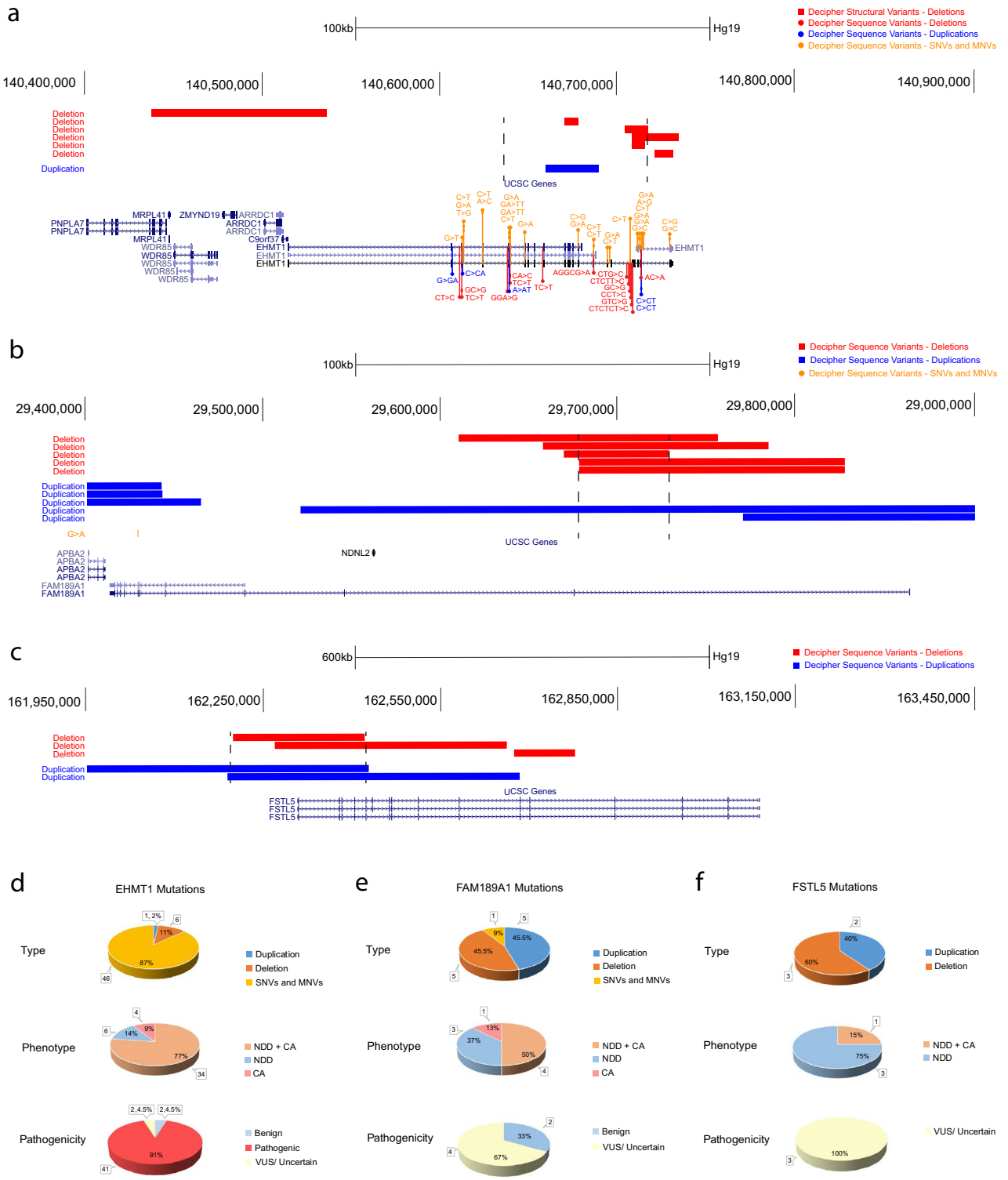

Fig. 4 Mutational landscape in candidate genes *EHMT1*, *FAM189A1*, and *FSTL5*. **a** The breakpoints of pathogenic deletion (red) and duplication (blue) variants impacting *EHMT1,* and nearby genes reported in DECIPHER. **b** The breakpoints of pathogenic deletion (red) and duplication (blue) variants impacting *FAM189A1,* and nearby genes reported in DECIPHER. **c** The breakpoints of pathogenic deletion (red) and duplication (blue) variants impacting *FSTL5* and nearby genes reported in DECIPHER. **d** The pie charts describe the count and percentage of *EHMT1* variant mutations reported in DECIPHER with the following variables: type, phenotype, and pathogenicity. **e** The pie charts describe the count and percentage of *FAM189A1* variant mutations reported in DECIPHER. **f** The pie charts describe the count and percentage of *FSTL5* variant mutations reported in DECIPHER. DECIPHER, DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources; SNV, single-nucleotide variant; MNV, multi-nucleotide variant

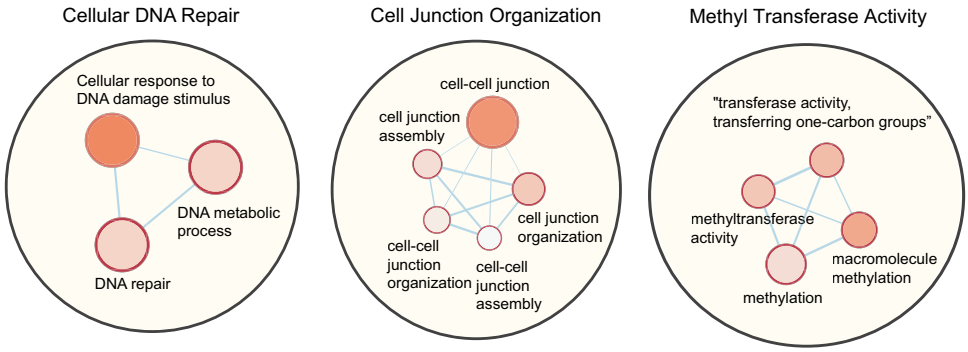

Fig. 5 Pathway network map of constraint overlapped (NDD and CA) genes. Pathway network analysis of constraint overlapped genes from de novo pathogenic deletion CNVs enriched clusters drawn using Cytoscape. The analysis of overlapped loss pathogenic genes with significant pathways (P value < 0.05) with a false-discovery rate (FDR) < 0.01. The color gradient and size of the node represented the P value and odds ratio, respectively

are significantly impacted by de novo pathogenic CNVs reported in both NDD and CA; (ii) are highly upregulated in prenatal stage of brain; (iii) are involved in developmental cellular pathways; and (iv) provide a list of candidate genes that may not be captured from individual cohort analysis (NDD or CA), rather captured in combined analysis.

The simultaneous presence of both phenotypes of NDDs and CAs has been described recently. Of significance, Fregaue et al. (2016) reported that proximal deletions of 1p36 or haploinsufficiency of the *RERE* gene, found in 10 subjects were strongly associated with the onset of both NDD and CA phenotypes, and this was also observed in RERE-deficient mice and zebrafish. Furthermore, Jordan et al. studied nine individuals with NEBDEH that had partial deletions or deleterious sequence variants in *RERE* (Jordan et al. 2018). CHARGE syndrome, a differential to RERE-related disorders, indicated to involve both NDDs and CAs, is reported to be caused by de novo mutations in the *CHD7* gene with a prevalence of 1 in 10,000 births (Jordan et al. 2018). Clinical features include coloboma, heart defects, choanal atresia, retarded growth and development, genital abnormalities, ear anomalies, and distinguishing features from RERE-related disorders are the presence of semicircular canal defects or tracheoesophageal fistulas in CHARGE syndrome patients (Hsu et al. 2014).

Our study offers an initial comparison of the two sexes to detect genes and variants in NDD and CA phenotypes for all the autosomes and X-chromosome. By assessing 13,796 sequenced patients, we identified 217 CNVs with enrichment of de novo variants in patients with NDDs and CAs, irrespective of gender. Comparatively, de novo mutations in these variants were greater in males than in females. Male de novo pathogenic deletion variants contained nine more CNVs larger than 5 Mb compared to females and that may explain the increased prevalence of de novo CNVs in males.

Comparing all sets of gene overlaps (from male and female CNVs), we identified the NDD and CA overlap of genes extracted from pathogenic deletion CNVs in female to be the most significant (P value = 0) and the overlap of genes extracted from pathogenic duplication CNVs in female to have the highest odds ratio (OR = 15.61). Similarly, among the overlapped genes from pathogenic CNVs that underwent constraint filtering using CE or pLI, the gene overlap from the pathogenic deletion variants in female were the most significant (P value = 8.5×10^{-11}) and the gene overlap from pathogenic duplication CNVs in female held the highest odds ratio (OR = 2.02).

Constraint genes are highly upregulated in prenatal period which shows their importance in early neurogenesis and organ development. Out of the constraint overlapped genes (1086 genes), we shortlisted three genes, one with possible studied pleiotropism of NDDs and CAs: *EHMT1*, which causes Kleefstra syndrome, known to harbor heterozygous intragenic *EHMT1* pathogenic variants from heterozygous deletions at chromosome 9q34.3 (Yatsenko et al. 2009; Willemsen et al. 2012). This syndrome involves NDD characteristics of childhood atonia, autistic-like features, intellectual disability, and CA characteristics of distinctive facial features (Cormier-Daire et al. 2003; Stewart et al. 2004; Kleefstra et al. 2005; Yatsenko et al. 2005), heart defects, renal/urologic defects, and genital defects in males among others. It is reported that both genders are equally affected and with some indication of genotype–phenotype correlation in 9q34.3 deletions affected by pathogenic variants that are smaller in size (< 1 Mb) (Yatsenko et al. 2009; Kleefstra et al. 2009; Willemsen et al. 2012). However, the grouping of several haplo-insufficient genes generates a pathological phenotype, and a direct causal relationship of phenotype to an individual gene cannot be ascertained. The other candidate genes, *FAM189A1* and *FSTL5*, were selected from the non-OMIM gene list that was curated from the constraint

filtered 1089 candidate genes. Our study suggests the possible roles of these genes in NDDs and CAs that have no OMIM entries.

Establishing genotype and phenotype correlation is complex (Uddin et al. 2019), especially for constraint genes that are reported in multiple distinct phenotypes (Woodbury-Smith et al. 2017a, b). Future development of artificial intelligence coupled with deep phenotypic information might improve the delineation of constraint genes that may underly the etiology of NDD and CA. We have demonstrated the multi-faceted use of different types of molecular data from the human brain tissue to interpret and identify candidate genes for NDD and CA disorders, from pathogenic variants and VUS. One of the limitations of our study might be the under-reporting of phenotypes between the cases of the two cohorts, as some of the CA cases might have later developed NDD symptoms which cannot be captured in a retrospective cohort without reevaluating the patient status. Our assessable approach considering the reported phenotypes enables the indexing of genes affected by respective CNVs for a possible role in neurodevelopmental disorders and congenital anomalies.

It may be possible to develop effective targeted treatment with single-cell analysis of overlapping NDD and CA genes in the same way that the identification of cancer subtypes could enable and guide cancer treatments (Parsons et al. 2008; Chapman et al. 2011; Herbst et al. 2014). This study is a proof-of-concept for how massive amounts of CNV and transcriptomic data can be used to expand existing knowledge and bring precision medicine to treating NDD and CA cases with overlapping genes and CNVs. Further studies to understand the functional regulation of the candidate genes may help in targeted therapeutics and timely interventions throughout development to mitigate the effects of different genomic alterations.

Conclusion

We have incorporated multi-dimensional transcriptome data from different sources to understand the genetic overlap of NDDs and CAs. We observed that those different mutations may be implicated in a molecular subtype of NDD and CA. By applying an integrative framework, we examined the convergence of clinical mutations onto specific disease-related pathways. The comprehensive analytical framework in our work can be utilized to uncover functional elements for other genetic diseases, enhancing their risk assessment. The overlap of molecular subtypes of NDD and CA risk genes to brain tissue cell types and pathways will be vital for the future development of effective combined diagnosis of NDD and CA and aid in therapeutics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00439-022-02482-5.

Acknowledgements We thank Dimitri J. Stavropoulos and Marsha Speevak for their assistance in data compilation and the families for their participation in research and genomic studies.

Author contributions SASS and MU conceptualized the study and designed the experiments. SASS, NN, SS, IK, RT, BZ, NK, HA, BB, and MU did critical analysis and review. SASS, BB, NN, and MU contributed into writing the manuscript. All the authors contributed to critical review and editing of the manuscript and approved the submitted manuscript. All the authors read and approved the final manuscript.

Funding This work was supported, in whole or in part, by the Al Jalila Foundation, internal grant awards from Mohammed Bin Rashid University of Medicine and Health Sciences (MBRU)—College of Medicine (MBRU-CM-RG2018-04, MBRU-CM-RG2018-05, MBRU-CM-RG2020-02, and MBRU-CM-RG2020-12); Sandooq Al Watan Research & Development Grant (SWARD-F2018-002); AlMahmeed Collaborative Research Awards (ALM1801, ALM20-0074); and Al Jalila Foundation Grant (AJF201763). Dr. Nasna Nassir was supported by MBRU Post-Doctoral Fellow Award (MBRU-PD-2020-02). Dr. Richa Tambi was supported by MBRU Post-Doctoral Fellow Award (MBRU-PD-2020-04).

Availability of data and materials The datasets supporting the conclusions of this article are included within the article (and its Additional files).

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical approval and consent to participate This original study (Uddin et al. 2016) has been approved by The Hospital for Sick Children research ethics board, REB # 1000030304 and by the College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences (MBRU-IRB-2017-004).

Consent for publication Not applicable.

References

Akter H, Hossain MS, Dity NJ et al (2021) Whole exome sequencing uncovered highly penetrant recessive mutations for a spectrum of rare genetic pediatric diseases in Bangladesh. NPJ Genom Med 61(6):1–9. https://doi.org/10.1038/s41525-021-00173-0

Ameen SK, Alalaf SK, Shabila NP (2018) Pattern of congenital anomalies at birth and their correlations with maternal characteristics in the maternity teaching hospital, Erbil city, Iraq. BMC Pregnancy Childbirth. https://doi.org/10.1186/S12884-018-2141-2

American Psychiatric Association (2013) Diagnostic and statistical manual of DSM-5^{TM}

Bierut LJ, Madden PAF, Breslau N et al (2007) Novel genes identified in a high-density genome wide association study for nicotine dependence. Hum Mol Genet 16:24–35. https://doi.org/10.1093/hmg/ddl441

Bierut LJ, Agrawal A, Bucholz KK et al (2010) A genome-wide association study of alcohol dependence. Proc Natl Acad Sci USA 107:5082–5087. https://doi.org/10.1073/pnas.0911109107

Bragin E, Chatzimichali EA, Wright CF et al (2014) DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res 42:D993–D1000. https://doi.org/10.1093/NAR/GKT937

Casanova EL, Gerstner Z, Sharp JL et al (2018) Widespread Genotype-Phenotype Correlations in Intellectual Disability. Front Psychiatry 9:535. https://doi.org/10.3389/FPSYT.2018.00535/BIBTEX

Chapman PB, Hauschild A, Robert C et al (2011) Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 364:2507–2516. https://doi.org/10.1056/NEJMOA1103782

Cormier-Daire V, Molinari F, Rio M et al (2003) Cryptic terminal deletion of chromosome 9q34: a novel cause of syndromic obesity in childhood? J Med Genet 40:300–303. https://doi.org/10.1136/JMG.40.4.300

Coviello AD, Haring R, Wellons M et al (2012) A genome-wide association meta-analysis of circulating sex hormone-binding globulin reveals multiple loci implicated in sex steroid hormone regulation. PLoS Genet 8:e1002805. https://doi.org/10.1371/journal.pgen.1002805

DeSilva M, Munoz FM, Mcmillan M et al (2016) Congenital anomalies: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 34:6015. https://doi.org/10.1016/J.VACCINE.2016.03.047

Dolk H, Loane M, Garne E (2010) The prevalence of congenital anomalies in Europe. Adv Exp Med Biol 686:349–364. https://doi.org/10.1007/978-90-481-9485-8_20

Duncan AMV, Chodirker B (2011) Use of array genomic hybridization technology for constitutional genetic diagnosis in Canada. Paediatr Child Health 16:211. https://doi.org/10.1093/PCH/16.4.211

Frega M, Linda K, Keller JM et al (2019) (2019) Neuronal network dysfunction in a model for Kleefstra syndrome mediated by enhanced NMDAR signaling. Nat Commun 101(10):1–15. https://doi.org/10.1038/s41467-019-12947-3

FregeauKim BBJ, Hernández-García A et al (2016) De Novo mutations of RERE cause a genetic syndrome with features that overlap those associated with proximal 1p36 deletions. Am J Hum Genet 98:963–970. https://doi.org/10.1016/J.AJHG.2016.03.002

Gardina PJ, Clark TA, Shimada B et al (2006) Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. BMC Genomics 7:1–18. https://doi.org/10.1186/1471-2164-7-325/FIGURES/8

Herbst RS, Soria JC, Kowanetz M et al (2014) Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature 515:563–567. https://doi.org/10.1038/NATURE14011

Hsu P, Ma A, Wilson M et al (2014) CHARGE syndrome: a review. J Paediatr Child Health 50:504–511. https://doi.org/10.1111/JPC.12497

Hu WF, Chahrour MH, Walsh CA et al (2014) The diverse genetic landscape of neurodevelopmental disorders. Ann Rev Genom Hum Gent 15:195–213. https://doi.org/10.1146/ANNUREV-GENOM-090413-025600

Irizarry RA, Bolstad BM, Collin F et al (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31:e15. https://doi.org/10.1093/NAR/GNG015

Jordan VK, Fregeau B, Ge X et al (2018) Genotype–phenotype correlations in individuals with pathogenic RERE variants. Hum Mutat 39:666. https://doi.org/10.1002/HUMU.23400

Kaminsky EB, Kaul V, Paschall J et al (2011) An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med 13:777–784. https://doi.org/10.1097/GIM.0B013E31822C79F9

Kang HJ, Kawasawa YI, Cheng F et al (2011) Spatio-temporal transcriptome of the human brain. Nature 478:483–489. https://doi.org/10.1038/nature10523

Kearney HM, Thorland EC, Brown KK et al (2011) American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 13:680–685. https://doi.org/10.1097/GIM.0B013E3182217A3A

Kim MS, Pinto SM, Getnet D et al (2014) A draft map of the human proteome. Nat 509:575–581. https://doi.org/10.1038/nature13302

Kleefstra T, Smidt M, Banning MJG et al (2005) Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. J Med Genet 42:299–306. https://doi.org/10.1136/JMG.2004.028464

Kleefstra T, Van Zelst-Stams WA, Nillesen WM et al (2009) Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. J Med Genet 46:598–606. https://doi.org/10.1136/JMG.2008.062950

Krawczak M, Nikolaus S, von Eberstein H et al (2006) PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Public Health Genomics 9:55–61. https://doi.org/10.1159/000090694

Lek M, Karczewski KJ, Minikel EV et al (2016) (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nat 536:285–291. https://doi.org/10.1038/nature19057

Li C, Dai L, Zhang J et al (2018) Follistatin-like protein 5 inhibits hepatocellular carcinoma progression by inducing caspase-dependent apoptosis and regulating Bcl-2 family proteins. J Cell Mol Med 22:6190. https://doi.org/10.1111/JCMM.13906

Marino BS, Lipkin PH, Newburger JW et al (2012) Neurodevelopmental outcomes in children with congenital heart disease: Evaluation and management a scientific statement from the american heart association. Circulation 126:1143–1172

McDonald-McGinn DM, Sullivan KE, Marino B et al (2015) 22q11.2 deletion syndrome. Nat Rev Dis Prim 1:15071. https://doi.org/10.1038/NRDP.2015.71

Mullin AP, Gokhale A, Moreno-De-Luca A et al (2013) (2013) Neurodevelopmental disorders: mechanisms and boundary definitions from genomes, interactomes and proteomes. Transl Psychiatry 312(3):e329–e329. https://doi.org/10.1038/tp.2013.108

Murray SS, Smith EN, Villarasa N et al (2012) Genome-wide association of implantable cardioverter-defibrillator activation with life-threatening arrhythmias. PLoS One. https://doi.org/10.1371/JOURNAL.PONE.0025387

Nassir N, Bankapur A, Samara B et al (2021) Single-cell transcriptome identifies molecular subtype of autism spectrum disorder impacted by de novo loss-of-function variants regulating glial cells. Hum Genomics 15:1–16. https://doi.org/10.1186/S40246-021-00368-7/FIGURES/6

Ohno H, Shinoda K, Ohyama K et al (2013) (2013) EHMT1 controls brown adipose cell fate and thermogenesis through the PRDM16 complex. Nat 504:163–167. https://doi.org/10.1038/nature12652

Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5:557–572. https://doi.org/10.1093/biostatistics/kxh008

Owoeye O, Kingston T, Scully PJ et al (2013) Epidemiological and clinical characterization following a first psychotic episode in major depressive disorder: comparisons with schizophrenia and bipolar I disorder in the cavan-monaghan first episode psychosis study (CAMFEPS). Schizophr Bull 39:756–765. https://doi.org/10.1093/SCHBUL/SBT075

Parenti I, Rabaneda LG, Schoen H, Novarino G (2020) Neurodevelopmental disorders: from genetics to functional pathways. Trends Neurosci 43:608–621. https://doi.org/10.1016/J.TINS.2020.05.004

Parsons DW, Jones S, Zhang X et al (2008) An integrated genomic analysis of human glioblastoma multiforme. Science 321:1807–1812. https://doi.org/10.1126/SCIENCE.1164382

Perles Z, Moon S, Ta-Shma A (2015) A human laterality disorder caused by a homozygous deleterious mutation in MMP21. J Med Genet 52:840–847. https://doi.org/10.1136/jmedgenet-2015-103336

Remke M, Hielscher T, Korshunov A et al (2011) FSTL5 is a marker of poor prognosis in non-WNT/non-SHH medulloblastoma. J Clin Oncol 29:3852–3861. https://doi.org/10.1200/JCO.2011.36.2798

Rozas MF, Benavides F, León L, Repetto GM (2019) Association between phenotype and deletion size in 22q11.2 microdeletion syndrome: systematic review and meta-analysis. Orphanet J Rare Dis 14:1–9. https://doi.org/10.1186/S13023-019-1170-X/FIGURES/5

Schaefer A, Sampath SC, Intrator A et al (2009) Control of cognition and adaptive behavior by the GLP/G9a epigenetic suppressor complex. Neuron 64:678–691. https://doi.org/10.1016/J.NEURON.2009.11.019/ATTACHMENT/65281776-2250-4F3E-BB6C-BABB8714B193/MMC1.PDF

Stewart DR, Huang A, Faravelli F et al (2004) Subtelomeric deletions of chromosome 9q: a novel microdeletion syndrome. Am J Med Genet A 128A:340–351. https://doi.org/10.1002/AJMG.A.30136

Stewart AFR, Dandona S, Chen L et al (2009) Kinesin family member 6 variant Trp719Arg does not associate with angiographically defined coronary artery disease in the Ottawa heart genomics study. J Am Coll Cardiol 53:1471–1472

Sugranyes G, Kyriakopoulos M, Corrigall R et al (2011) Autism spectrum disorders and schizophrenia: meta-analysis of the neural correlates of social cognition. PLoS One 6:e25322. https://doi.org/10.1371/JOURNAL.PONE.0025322

Ta-Shma A, Hjeij R, Perles Z et al (2018) Homozygous loss-of-function mutations in MNS1 cause laterality defects and likely male infertility. PLoS Genet. https://doi.org/10.1371/journal.pgen.1007602

Toufaily MH, Westgate MN, Lin AE, Holmes LB (2018) Causes of congenital malformations. Birth Defects Res 110:87–91. https://doi.org/10.1002/BDR2.1105

Uddin M, Tammimies K, Pellecchia G et al (2014) Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. Nat Genet 46:742–747. https://doi.org/10.1038/ng.2980

Uddin M, Pellecchia G, Thiruvahindrapuram B et al (2016) Indexing effects of copy number variation on genes involved in developmental delay. Sci Rep 61(6):1–12. https://doi.org/10.1038/srep28663

Uddin M, Unda BK, Kwan V et al (2018) OTUD7A regulates neurodevelopmental phenotypes in the 15q13.3 microdeletion syndrome. Am J Hum Genet 102:278–295. https://doi.org/10.1016/J.AJHG.2018.01.006

Uddin M, Wang Y, Woodbury-Smith M (2019) Artificial intelligence for precision medicine in neurodevelopmental disorders. NPJ Digit Med 21(2):1–10. https://doi.org/10.1038/s41746-019-0191-0

Uhlén M, Fagerberg L, Hallström BM et al (2015) Proteomics. Tissue-based map of the human proteome. Science. https://doi.org/10.1126/SCIENCE.1260419

Verhoeven VJM, Hysi PG, Wojciechowski R et al (2013) Genome-wide meta-analyses of multiancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. Nat Genet 45:314–318. https://doi.org/10.1038/ng.2554

Walsh CA, Morrow EM, Rubenstein JLR (2008) Autism and brain development. Cell 135:396. https://doi.org/10.1016/J.CELL.2008.10.015

Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. https://doi.org/10.1093/NAR/GKQ603

Willemsen MH, Vulto-Van Silfhout AT, Nillesen WM et al (2012) Update on Kleefstra syndrome. Mol Syndromol 2:202. https://doi.org/10.1159/000335648

Woodbury-Smith M, Deneault E, Yuen RKC et al (2017a) Mutations in RAB39B in individuals with intellectual disability, autism spectrum disorder, and macrocephaly. Mol Autism 8:1–10. https://doi.org/10.1186/S13229-017-0175-3/FIGURES/3

Woodbury-Smith M, Nicolson R, Zarrei M et al (2017b) Variable phenotype expression in a family segregating microdeletions of the NRXN1 and MBD5 autism spectrum disorder susceptibility genes. NPJ Genomic Med 21(2):1–8. https://doi.org/10.1038/s41525-017-0020-9

World Health Organization (2020) Birth defects surveillance a manual for programme managers, 2nd edn. World Health Organization

Yatsenko SA, Cheung SW, Scott DA et al (2005) Deletion 9q34.3 syndrome: genotype-phenotype correlations and an extended deletion in a patient with features of Opitz C trigonocephaly. J Med Genet 42:328–335. https://doi.org/10.1136/JMG.2004.028258

Yatsenko SA, Brundage EK, Roney EK et al (2009) Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3 microdeletion syndrome. Hum Mol Genet 18:1924–1936. https://doi.org/10.1093/HMG/DDP114

Zhang D, Ma X, Sun W et al (2015) Down-regulated FSTL5 promotes cell proliferation and survival by affecting Wnt/β-catenin signaling in hepatocellular carcinoma. Int J Clin Exp Pathol 8:3386

Zhang DY, Lei JS, Sun WL et al (2020) Follistatin Like 5 (FSTL5) inhibits epithelial to mesenchymal transition in hepatocellular carcinoma. Chin Med J (engl) 133:1798. https://doi.org/10.1097/CM9.0000000000000847