UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR
E DO DESENVOLVIMENTO

Tiago Fernando Chaves

VARIAÇÕES DO NÚMERO DE CÓPIAS E LONGOS TRECHOS CONTÍGUOS DE HOMOZIGOSE DETECTADOS POR MICROARRAY CROMOSSÔMICO (CMA) EM PACIENTES COM DISTÚRBIOS DO NEURODESENVOLVIMENTO E/OU ANOMALIAS CONGÊNITAS.

Florianópolis
2023

Tiago Fernando Chaves

VARIAÇÕES DO NÚMERO DE CÓPIAS E LONGOS TRECHOS CONTÍGUOS DE HOMOZIGOSE DETECTADOS POR MICROARRAY CROMOSSÔMICO (CMA) EM PACIENTES COM DISTÚRBIOS DO NEURODESENVOLVIMENTO E/OU ANOMALIAS CONGÊNITAS.

Tese de doutorado submetida ao Programa de Pós-graduação em Biologia Celular e do Desenvolvimento da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Biologia Celular e do Desenvolvimento.

Orientadora: Profa. Dra. Yara Costa Netto Muniz
Coorientadora: Profa. Dra. Angelica Francesca Maris

Florianópolis

2023

Chaves, Tiago Fernando
Variações do número de cópias e longos trechos contíguos de homozigose detectados por microarray cromossômico (CMA) em pacientes com distúrbios do neurodesenvolvimento e/ou anomalias congênitas. / Tiago Fernando Chaves ; orientadora, Yara Costa Netto Muniz, coorientadora, Angelica Francesca Maris, 2023.
206 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento, Florianópolis, 2023.

Inclui referências.

1. Biologia Celular e do Desenvolvimento. 2. Distúrbios do Neurodesenvolviment. 3. Transtorno do Espectro do Autismo. 4. Deficiência Intelectual. 5. Neurogenética. I. Muniz, Yara Costa Netto . II. Maris, Angelica Francesca . III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biologia Celular e do Desenvolvimento. IV. Título.

Tiago Fernando Chaves

VARIAÇÕES DO NÚMERO DE CÓPIAS E LONGOS TRECHOS CONTÍGUOS DE HOMOZIGOSE DETECTADOS POR MICROARRAY CROMOSSÔMICO (CMA) EM PACIENTES COM DISTÚRBIOS DO NEURODESENVOLVIMENTO E/OU ANOMALIAS CONGÊNITAS.

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 30 de junho de 2023, pela banca examinadora composta pelos seguintes membros:

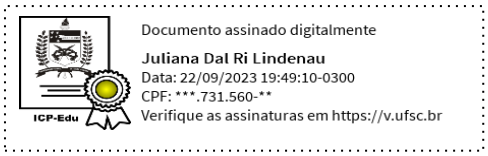
Prof. Geison de Souza Izidio, Dr.
Universidade Federal de Santa Catarina

Profª Ana Cristina Victorino Krepischi, Drª
Universidade de São Paulo

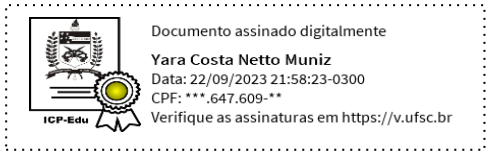
Prof Aparecido Divino da Cruz, Dr.
Pontifícia Universidade Católica de Goiás

Profª Louise Lapagesse de Camargo Pinto, Drª
Hospital Infantil Joana de Gusmão

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Doutor em Biologia Celular E Do Desenvolvimento.



Coordenação do Programa de Pós-Graduação



Profª. Yara Costa Netto Muniz, Drª.
Orientador(a)

Florianópolis, 2023.

Este trabalho é dedicado aos meus avós, Maria Erotildes e João Pedro Chaves (***in memorian***).

AGRADECIMENTOS

Primeiramente um imenso agradecimento à Universidade Federal de Santa Catarina (UFSC) pela oportunidade concedida para minha formação como doutor e por todas as experiências vividas durante a pós-graduação.

Um agradecimento muito especial aos mais pais, Roseli e Dair, e minha irmã Luara por sempre terem me dado forças e apoio, principalmente no momento de sair da minha pequena e humilde cidade natal para construir essa jornada longe de casa. Um agradecimento ao demais familiares e amigos.

Agradeço à minha orientadora Profª Drª Yara Costa Netto Muniz e coorientadora Profª. Dra. Angélica Francesca Maris, por toda a confiança, apoio, críticas construtivas e por todos os ensinamentos transmitidos ao longo desse tempo, muito obrigado!

Ao meu colega e amigo, Luan Freitas de Oliveira, companheiro de laboratório, por toda a ajuda e incentivo que você sempre me ofereceu. Aos demais membros da equipe de pesquisa, Dr. Jorge Barbato, Drª. Gisele de Lucca, Drª Louise Lapagesse de Camargo Pinto e Drª Pricila Bernardi por terem conseguido um espaço nas suas agendas, para colaborar e responder os questionários necessários para que este estudo pudesse se realizar, muito obrigado.

Obrigado a todos os professores da Pós-graduação em Biologia Celular e do Desenvolvimento pelas disciplinas ministradas e pelos conhecimentos transmitidos. Agradeço aos Representantes Discentes do PPG-BCD das gestões 2018-2019 e 2019-2021, pelos momentos compartilhados, desafios superados e seus aprendizados. Também, aos colegas das diferentes disciplinas, pela parceria ao longo do período de pós-graduação. Gostaria de agradecer imensamente a CAPES pelo apoio financeiro concedido através da bolsa, importante para execução deste trabalho e minha permanência nesta cidade.

Agradeço em especial aos membros da banca, Prof. Dr. Aparecido Divino da Cruz, Profª. Drª Ana Cristina Victorino Krepischi, Prof. Dr. Geison de Souza Izidio e Profª. Drª Talita Da Silva Jeremias, por cordialmente aceitarem avaliarem essa imensa tese. Considerações que com certeza serão muito bem-vindas para a melhoria deste trabalho. Obrigado a todos que de certa forma contribuíram para a realização deste trabalho.

“ARS ET SCIENTIA”

RESUMO

Nas últimas décadas, a citogenética molecular experimentou avanços significativos na investigação genética dos Distúrbios do Neurodesenvolvimento (DsND), como a Deficiência Intelectual (DI) e/ou o Transtorno do Espectro do Autismo (TEA), que afetam em torno de 3-4% da população mundial e podem estar associados à presença de anomalias congênitas (ACs) e/ou características dismórficas (DF) (dismorfismos - sindrômicos). Em especial, os exames genéticos baseados em microarranjos cromossômicos, conhecidos como CMA (Chromosomal Microarrays), tornaram-se uma ferramenta essencial na avaliação das Variações no Número de Cópias (CNVs), que incluem microdeleções e microduplicações, em pacientes afetados por DsND. As CNVs envolvem ganhos e perdas de material genômico e podem abranger regiões que contêm genes expressos no cérebro ou que desempenham um papel crucial no neurodesenvolvimento. Essas alterações na dosagem genética podem ter um papel significativo na patogênese de distúrbios do desenvolvimento. As plataformas modernas de CMA com alta densidade combinam sondas de oligonucleotídeos para a detecção de CNVs com sondas para Polimorfismos de Nucleotídeo Único (SNPs). Isso permite a identificação de regiões de Homozigose Cromossômica Completa ou Segmentar, conhecidas como Longos Trechos Contíguos de Homozigose (LCSH, Long Contiguous Stretches of Homozygosity). A presença destes LCSH pode ser indicativa de Dissomia Uniparental (UPD), consanguinidade, endogamia, características populacionais específicas, bem como eventos de reparo replicativo de DNA. Este estudo teve como principais objetivos: (1) Revisar a literatura quanto à origem dos DsND, ao papel das CNVs como fator etiológico para os DsND, às aplicações clínicas de várias plataformas de CMAs sua aplicação mundial e às taxas de diagnóstico através do CMA; (2) Analisar, interpretar e caracterizar CNVs, identificar a taxa diagnóstica do CMA em um coorte de indivíduos afetados DsND e/ou ACs do sul do Brasil, com foco nos casos com TEA, bem como avaliar as frequências fenotípicas e as implicações dos LCSHs detectados por exames de CMA. Para o objetivo (1), no repositório PUBMED/MEDLINE, foi analisada a literatura publicada entre 2010 e 2023 por palavras chave no título ou resumo. Dentre os 189 estudos de coorte encontrados, 84 publicações que preencheram os critérios estabelecidos foram selecionadas, mostrando uma taxa diagnóstica média de 16,8% para detecção de CNVs patogênicas por CMA. Esses estudos contemplam diferentes populações e etnias, com as maiores amostras em países do hemisfério norte. Apesar do sequenciamento de Exoma e os exames por CMA serem recomendados como testes de primeira linha na investigação dos DsND, na realidade econômica de países menos desenvolvidos o uso da citogenética clássica ainda é predominante, quando não é o único recurso. Para o objetivo (2), foi realizada uma análise retrospectiva de arquivos de leitura de CMAs das plataformas da Affymetrix CytoScan®HD [41%] ou CytoScan® 750K [59%], assim como foram coletados e analisados os dados clínicos disponíveis de 1.012 indivíduos afetados. Estes eram sua maioria crianças com DsND e/ou ACs, cujos exames foram solicitados por geneticistas e neurologistas para fins de diagnóstico. A análise revelou 206 CNVs consideradas patogênicas (PCNV), incluindo 132 deleções e 74 duplicações, fornecendo uma resposta diagnóstica para 17% dos indivíduos da coorte. Além disso, 12% dos pacientes apresentaram variantes raras de significado clínico incerto (VUS), em alguns casos consideradas possivelmente patogênicas (LPCNV), porém sem uma conclusão diagnóstica, como principal CNV clinicamente relevante. Para 71% da coorte nenhuma CNV clinicamente

relevante foi encontrada. Considerados com crescente relevância dentre os DsND, os indivíduos com TEA representaram cerca de um terço dos casos analisados (333). Analisados como sub-coorte, a taxa de diagnóstico para TEA foi de 10%, com variações dependendo da presença de características dismórficas (16%) ou da ocorrência "isolada" do TEA (7%). Utilizando a capacidade destas plataformas de reconhecer regiões homozigotas, LCSH (≥3 Mbp) foram analisados tanto no contexto de seu possível significado patogênico, quanto para identificar LCSH potencialmente derivados de haplótipos ancestrais da população. Esta análise foi possível para 953 CMAs, revelando que LCSH estavam presentes em 91% dos exames. Em 11,5% dos indivíduos, sugeriram consanguinidade do primeiro ao quinto grau entre seus progenitores, com possível impacto clínico, e em 2,8%, revelaram uma potencial UPD. Aqueles LCSH encontrados com frequência de 5% ou mais na coorte foram considerados LCSH comuns na população em geral, permitindo delinear 10 regiões como sendo potenciais haplótipos ancestrais de relevância clínica baixa. Em relação aos fenótipos clínicos desta coorte, as principais indicações para o CMA foram atraso no desenvolvimento (56%), DI (33%), TEA (33%), e características dismórficas (56%). Certos fenótipos clínicos foram associados a uma maior probabilidade de indicar um portador de uma CNV patogênica. As CNVs raras encontradas, assim como as LCSH que possivelmente representem UPDs foram caracterizadas e registradas junto com os principais fenótipos apresentados por cada paciente. As LCSH consideradas potenciais haplótipos ancestrais segregando na população do sul do Brasil, foram registradas para auxiliar na interpretação e priorização de LCHS que efetivamente tem uma maior relevância clínica. Este estudo representa o maior conjunto de dados de CMA em uma coorte de pacientes com DsND e/ou ACs na região sul do Brasil até o momento. Este trabalho destaca a importância da interpretação de CNVs e LCSH na prática clínica, inclusive no contexto do TEA.

Palavras-chave: Microarray cromossômico, CNVs, deficiência intelectual, transtorno do espectro do autismo, distúrbios do neurodesenvolvimento, SNPs, LCSH.

ABSTRACT

In recent decades, molecular cytogenetics has seen significant advances in genetic research on Neurodevelopmental Disorders (NDDs), such as Intellectual Disability (ID) and/or Autism Spectrum Disorder (ASD), which affect around 3-4% of the global population and can be associated with the presence of congenital anomalies (CAs) and/or dysmorphic features (DF) (dysmorphisms - syndromic). In particular, genetic tests based on chromosomal microarrays, known as CMA (Chromosomal Microarrays), have become an essential tool in the assessment of Copy Number Variations (CNVs), including microdeletions and microduplications, in patients affected by NDDs. CNVs involve gains and losses of genomic material and can encompass regions containing genes expressed in the brain or playing a crucial role in neurodevelopment. These changes in genetic dosage can play a significant role in the pathogenesis of developmental disorders. Modern CMA platforms with high density combine oligonucleotide probes for CNV detection with probes for Single Nucleotide Polymorphisms (SNPs). This allows the identification of regions of Complete or Segmental Chromosomal Homozygosity, known as Long Contiguous Stretches of Homozygosity (LCSH). The presence of these LCSH may be indicative of Uniparental Disomy (UPD), consanguinity, endogamy, specific population characteristics, as well as replicative DNA repair events. For goal (1), in the PUBMED/MEDLINE repository, the literature published between 2010 and 2023 was analyzed by keywords in the title or abstract. Among the 189 cohort studies found, 84 publications that met the established criteria were selected, demonstrating an average diagnostic rate of 16.8% for the detection of pathogenic CNVs by CMA. These studies encompass different populations and ethnicities, with the largest samples from northern hemisphere countries. Despite Exome sequencing and CMA tests being recommended as first-line tests in the investigation of NDDs, in the economic reality of less developed countries, classical cytogenetics is still predominant, if not the only resource. For objective (2), a retrospective analysis of CMA reading files from the Affymetrix CytoScan®HD [41%] or CytoScan® 750K [59%] platforms was performed, and clinical data were collected and analyzed for 1,012 affected individuals. The majority of these individuals were children with NDDs and/or CAs, whose tests were requested by geneticists and neurologists for diagnostic purposes. The analysis revealed 206 CNVs considered pathogenic, including 132 deletions and 74 duplications, providing a diagnostic answer for 17% of the individuals in the cohort. Furthermore, for 12% of patients the only clinically relevant CNV were Variants of Uncertain Significance (VUS), in some cases interpreted as likely pathogenic (LPCNV), however not considered a clear diagnostic conclusion. For 71% of the cohort, no clinically relevant CNVs were found. With a growing significance among NDDs, individuals with ASD represented about one-third of the cases analyzed (333). When analyzed as a sub-cohort, the diagnostic rate for ASD was 10%, with variations depending on the presence of dysmorphic features (16%) or when considered as having "isolated" ASD (7%). Using the capability of these platforms to recognize homozygous regions, LCSH (≥3 Mbp) were analyzed both in the context of their potential pathogenic significance and to identify LCSH that could potentially be derived from ancestral population haplotypes. This analysis was feasible for 953 CMAs, revealing that LCSH were present in 91% of the tests. In 11.5% of individuals, they suggested consanguinity from the first to the fifth degree among the parents, with potential clinical implications, and in 2.8% they revealed a potential Uniparental Disomy (UPD). Those LCSH found with a frequency of 5% or more in the cohort were considered common LCSH in the general population, allowing the

delineation of 10 regions as ancestral haplotypes of potentially low clinical relevance. Regarding the clinical phenotypes of this cohort, the primary indications for CMA were developmental delay (56%), ID (33%), ASD (33%), and dysmorphic features (56%). Certain clinical phenotypes were associated with a higher likelihood of indicating a carrier of a pathogenic CNV. The rare CNVs found, as well as the LCSH that may represent UPDs, were characterized and recorded along with the main phenotypes presented by each patient. LCSH considered potential ancestral haplotypes segregating in the population of southern Brazil were recorded to assist in the interpretation and prioritization of LCSH that indeed have greater clinical relevance. This study represents the largest dataset of CMA in a cohort of patients with NDDs and/or CAs in the southern region of Brazil to date. This work highlights the importance of interpreting CNVs and LCSH in clinical practice, including in the context of ASD.

Keywords: Chromosomal microarray, CNVs, intellectual disability, autism spectrum disorder, neurodevelopmental disorders, SNPs, LCSH.

LISTA DE FIGURAS

CAPÍTULO I

CAPÍTULO II

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ACMG – Colégio Americano de Genética Médica (*American College of Medical Genetics*);

ACs - Anomalias congênitas;

ADHD = *Attention-deficit/hyperactivity disorder;*

Array-CGH - Hibridização Genômica Comparativa por microchip (*array Comparative Genomic Hybridization*);

ASD – *Autism Spectrum Disorder*;

BAC - Cromossomos Artificiais Bacterianos (*Bacterial Artificial Chromosomes*);

BIR - Replicação Induzida Por Quebra (*break induced replication*);

CDC – Centro de Controle e Prevenção de Doenças;

CGH CGH - Hibridização Genômica Comparativa;

ChAS - *Chromosome Analysis Suite*;

CMA - Microarray Cromossômico (*Chromosomal microarray*);

CNV - Variações do Número de Cópias (*Copy Number Variation*);

DD – *Developmental Delay*;

DF – *Dysmorphic Features*;

DI - Deficiência Intelectual;

DNA - Ácido Desoxirribonucleico;

DSM - Manual Diagnóstico e Estatístico de Transtornos Mentais (*Diagnostic and Statistical Manual of Mental Disorders*);

DsND - Distúrbios do desenvolvimento cerebral;

F = *Female*,

FD – *Facial Dysmorphia*;

GRCh - Consórcio Genoma Referência (*Genome Reference Consortium*);

Hg - Genoma Humano (*Human Genome*);

HIJG - Hospital Infantil Joana de Gusmão;

IBD - Regiões Idênticas Por Descendência (*Identical By Descente*);

ID – Intellectual Disability (not specified)

ISCA- *International Standards for Cytogenomic Arrays*;

IUGR = *Intrauterine growth restriction;*

Kpb - Kilo pares de base (mil pares bases);

LCR - Repetições de Baixa Cópia (*low-copy repeats*);

LCSH - Longos Trechos Contíguos De Homozigose (*Long Contiguous Stretches Of Homozygosity*);

LDO = *learning difficulty only (no ID)*;

LOH - Perda de Heterozigose (*Loss of Heterosygosity*);

LPCNV – *Likely Pathogenic Copy Number Variations*;

M = *Male*;

MildID = *Mild Intellectual Disability*;

ModID = *Moderate Intellectual Disability*;

Mpb - Mega pares de base (milhões de pares de bases);

NAHR - Recombinação homóloga não alélica (*nonallelic homologous recombination*);

NCBI - Centro Nacional para a Informação Biotecnológica;

NDD – *Neurodevelopmental Disorder*;

NGS - Sequenciamento de Nova Geração (*Next-Generation Sequencing*);

NHEJ - Junção de Extremidades Não-Homólogas (*non-homologous end joining*);

OMIM - *Online Mendelian Inheritance in Man*;

PCNV – *Pathogenic Copy Number Variations*;

PubMed - *US National Library of Medicine National Institutes of Health*;

SC - Santa Catarina;

SevID = *Severe Intellectual Disability*;

SLD = *Speech and/or language delay or impairment*;

SNP - Polimorfismos de Nucleotídeo Único;

SNV – Variantes de Nucleotídeo Único;

TCLE - Termo de Consentimento Livre e Esclarecido;

TEA - Transtorno Do Espectro Do Autismo;

UCSC - Universidadade da Califórnia Santa Cruz;

UFSC - Universidade Federal de Santa Catarina;

UPD - Disomia Uniparental (*uniparental disomy*);

VUS - Variantes de Significado Clínico Incerto (*Variants of Uncertain clinical Significance*);

SUMÁRIO

1 INTRODUÇÃO

1.1 DISTÚRBIOS DO NEURODESENVOLVIMENTO.

Os distúrbios do neurodesenvolvimento (DsND) ou simplesmente distúrbios do desenvolvimento, são condições heterogêneas, complexas e de difícil conceituação, mas de forma geral se referem um grupo de condições que afetam o correto desenvolvimento do cérebro e do sistema nervoso (THAPAR; COOPER; RUTTER, 2017).

Atualmente, os DsND acarretam, em sua maioria, em deficiência intelectual (DI) e/ou transtorno do espectro do autismo (TEA), afetando em torno de 3–4% da população mundial (CAPPUCCIO et al., 2016; PEREIRA et al., 2014).

No entanto, em sua definição, os DsND também englobam outra gama ampla de distúrbios neurológicos e psiquiátricos que são clínica e causalmente díspares; como síndromes genéticas raras, paralisia cerebral, anomalias neurais, esquizofrenia, Transtorno de Déficit de Atenção/Hiperatividade (TDAH), distúrbios motores do neurodesenvolvimento, transtornos específicos de aprendizagem e epilepsia (THAPAR, [s.d.])

Os DsND geralmente são observados e diagnosticados na infância e podem ter um impacto significativo no desenvolvimento do indivíduo afetado, resultando em limitações funcionais nas suas atividades cotidianas, como no autocuidado, receptividade e linguagem expressiva, aprendizagem, mobilidade, autodireção, capacidade de vida independente, autossuficiência econômica' (D'AMOURS et al., 2014).

Tais distúrbios, quando isolados, são denominados não sindrômicos e, quando associados à presença de dismorfismos ou de anomalias congênitas (ACs) aparentes, são denominados sindrômicos (ABOU JAMRA et al., 2011).

A DI é uma parte importante dos DsND e varia em gravidade, com diferentes graus de comprometimento cognitivo. Pode ser leve, moderada, grave ou severa, afetando a capacidade do indivíduo de aprender e realizar atividades cotidianas (BOAT et al., 2015). Os fatores genéticos desempenham um papel significativo na etiologia da DI, com várias alterações genéticas associadas a diferentes graus de comprometimento cognitivo. Fatores ambientais também desempenham um papel

importante, como exposição a toxinas e deficiências nutricionais durante a gravidez (MICHELSON; CLARK, 2020; NAIR et al., 2022; SHAFFER, 2005).

Além da DI, o TEA é outra condição abrangida pelos DsND. O TEA é caracterizado por dificuldades na interação social, comunicação e comportamentos repetitivos (DSM-5 DIAGNOSTIC CLASSIFICATION, 2013). Sua prevalência tem aumentado globalmente e tem uma forte base genética, embora os fatores ambientais também desempenhem um papel (KOGAN et al., 2018; MAENNER et al., 2020). A comorbidade entre DI e TEA é observada em muitos casos, destacando a complexidade dos DsND. O diagnóstico e a compreensão dessas condições são essenciais para orientar o tratamento e o apoio apropriados (ROSELLÓ et al., 2014).

As relações dos DsND com alterações estruturas submicroscópicas no genoma, como fator etiológico causal, são complexas e multifacetadas. Tanto a DI quanto o TEA têm bases genéticas significativas (DEVLIN; SCHERER, 2012; GAUGLER et al., 2014; VISSERS et al., 2003).

Dentro das variações estruturais no genoma humano, destacam-se as Variações no Número de Cópias (CNVs, *Copy Number Variation*), que podem envolver alterações na quantidade de cópias de um trecho do DNA envolvendo um ou mais genes. Essas variações podem variar em tamanho, desde dezenas de bases até megabases, e podem afetar pequenas porções de um gene, como regiões exônicas ou regulatórias, até grandes segmentos cromossômicos contendo centenas de genes (NOWAKOWSKA, 2017).

As CNVs podem ter impactos significativos na dosagem gênica e na intensidade da sinalização molecular, pois a quantidade de proteína produzida está diretamente relacionada ao número de cópias do gene. Existem dois principais efeitos relacionados às CNVs (RICE; MCLYSAGHT, 2017):

1. **Haploinsuficiência** - Isso ocorre quando um gene, quando em apenas uma cópia funcional (por exemplo, devido a uma microdeleção ou inativação da outra cópia), não produz proteína suficiente para desempenhar sua função normal. Isso pode resultar em problemas fenotípicos.

2. **Triplossensibilidade** - Esse efeito ocorre quando genes são sensíveis à superexpressão devido a um aumento no número de cópias causado por uma

microduplicação. Isso pode comprometer a função normal desses genes e levar a efeitos fenotípicos.

A sensibilidade à dosagem gênica desempenha um papel importante na patogenicidade das CNVs, determinando se uma CNV específica causa uma anormalidade fenotípica (DAVOLI et al., 2013; RICE; MCLYSAGHT, 2017).

As CNVs também podem afetar a função de genes de outras maneiras, como levar à deleção ou interrupção de genes, o que pode expor mutações recessivas presentes no alelo do cromossomo homólogo ou alterar a comunicação entre alelos. Portanto, as CNVs têm o potencial de causar uma ampla gama de efeitos fenotípicos, desde traços adaptativos até problemas graves, incluindo letalidade embrionária (CAPPUCCIO et al., 2016; COOK; SCHERER, 2008; ZARREI et al., 2015).

Atualmente, a técnica de Microarray Cromossômico (CMA, *Chromosomal Microarray*), é amplamente utilizada para a detecção de CNVs no genoma humano. Esta técnica permite a hibridização de pequenos fragmentos de um genoma teste com sondas isoladas em micropoços fixados em uma matriz sólida, utilizando o mesmo princípio de comparação de fluorescência (LAY-SON et al., 2015; RETTERER et al., 2015).

Na pesquisa genética e na prática clínica, identificar e interpretar o efeito de CNVs, especialmente em relação aos DsND, pode ser desafiador. Isso ocorre porque a correlação entre fenótipo e CNVs pode ser subjetiva, influenciada por fatores como origem étnica e ambiente. Além disso, novas CNVs patogênicas continuam sendo descobertas em diversas categorias de doenças.

A interpretação das CNVs, de acordo com as diretrizes da American College of Medical Genetics and Genomics (ACMG), é um processo crucial na avaliação clínica de pacientes com suspeita de fatores genéticos (RIGGS et al., 2020).

A ACMG estabeleceu diretrizes específicas para ajudar na interpretação das CNVs detectadas em indivíduos afetados. Essas diretrizes consideram diversos fatores, incluindo a relevância clínica das CNVs, seu tamanho, impacto nos genes, histórico de doenças em famílias, bem como a existência de estudos de associação genética e evidências funcionais.

A interpretação das CNVs pode resultar em diferentes classificações, de acordo com as diretrizes da ACMG (RIGGS et al., 2020):

1. **Patogênicas** - CNVs que têm alta probabilidade de causar doença e estão associadas a fenótipos clínicos específicos.

2. **Possivelmente patogênicas** - CNVs que têm uma associação menos clara com a doença, mas ainda podem contribuir para o quadro clínico do paciente.

3. **Variantes de significado clínico incerto** - CNVs que não têm evidências suficientes para serem consideradas patogênicas ou possivelmente patogênicas, mas também não podem ser consideradas benignas.

4. **Benignas** - CNVs que são consideradas improváveis de causar doenças ou ter relevância clínica.

A interpretação precisa das CNVs é fundamental para orientar o aconselhamento genético, o diagnóstico e o tratamento de pacientes com doenças genéticas. É importante ressaltar que a interpretação das CNVs muitas vezes requer uma abordagem multidisciplinar, envolvendo geneticistas clínicos, especialistas em bioinformática e geneticistas moleculares, bem como a integração de dados clínicos e genômicos para uma avaliação completa e precisa (QUINTELA et al., 2017).

A pesquisa sobre CNVs é importante porque elas desempenham um papel na variabilidade genética e fenotípica não patológica entre os indivíduos, mas também podem estar associadas ao desenvolvimento de várias patologias, incluindo distúrbios no desenvolvimento físico e neurológico.

A inclusão de sondas para Polimorfismo de Nucleotídeo Único (SNPs, *Single Nucleotide Polymorphisms*) em plataformas de CMA permite a detecção de regiões com ausência de heterozigose (AOH, *Absence of heterozygosity*). Estas incluem os longos trechos contíguos de homozigose (LCSH, *Long Contiguous Stretches Of Homozygosity*), que se refere a regiões onde todas as cópias cromossômicas do trecho são homozigotas, também conhecidos como corridas de homozigose (ROH), assim como regiões no DNA onde, devido a uma deleção, há perda de heterozigosidade (LOH) (KEARNEY, 2012; LI et al., 2006).

Os LCSHs são resultados de diversos fatores, como consanguinidade, endogamia, forças evolutivas e eventos de reparo após quebras cromossômicas. Dependendo de sua quantidade e tamanho no genoma, podem indicar

consanguinidade, dissomia uniparental ou simplesmente representar regiões de baixa recombinação, nas quais certos haplótipos se tornam comuns em uma população (CHIA, 2016; IOUROV et al., 2015).

Embora ainda relativamente pouco estudados em comparação com as CNVs, os LCSHs são uma das variações genômicas mais comuns em humanos (CONRAD et al., 2010; IOUROV et al., 2015; SIMÓN-SÁNCHEZ et al., 2012; VORSANOVA et al., 2010). Eles fornecem informações sobre diversidade genética, podem ser relevantes para estudos de associação genética de traços complexos e foram associados à patogênese de doenças. Vários estudos relacionaram LCSHs encontrados como possível causa de DsND (GAMSIZ et al., 2013; GANDIN et al., 2015; IOUROV et al., 2015).

Quando há excesso de LCSHs que englobam genes de doença recessiva, estudos evidenciaram que a quantidade de homozigosidade no genoma parece modular o grau de comprometimento cognitivo (GANDIN et al., 2015). Da mesma forma, a quantidade de LCSHs revela parentesco e aumenta a probabilidade de herança de transtornos monogênicos recessivos (CHIA; LEISA; CHIA, 2016).

Embora a detecção de LCSHs por si só não leve a um desfecho clínico ou diagnóstico, muitas vezes eles revelam um alvo para investigações futuras e podem ter significância benigna. A interpretação de LCSHs pode ser um desafio para os laboratórios diagnósticos, mas o registro de LCSHs comuns na população, que geralmente não são de relevância clínica, pode facilitar a análise e orientar a investigação para eventos raros e genes candidatos (CHIA, 2016). A análise de LCSHs não só traz uma melhor compreensão do fundo genético e evolução histórica do indivíduo, como também fornece informações auxiliares para os projetos de mapeamento de associação de genes de doenças, informações para detecção de aberração cromossômica e ajuda na interpretação dos resultados (YANG et al., 2011).

A compreensão das CNVs, LCSHs e sua relação com os DsND desempenha um papel fundamental na pesquisa e diagnóstico dessas condições, oferecendo informações essenciais para identificar causas genéticas e desenvolver estratégias na investigação etiológica dos DsND.

2 OBJETIVOS

2.1 GERAL

O objetivo geral deste estudo é aprofundar o conhecimento sobre os Distúrbios do Neurodesenvolvimento (DsND) através da expansão da amostragem, investigação do papel da consanguinidade, taxas e casos potenciais de dissomia uniparental (UPD) e identificação de haplótipos de LCSHs comuns na população, visando estabelecer padrões e limiares para análise de resultados de *Microarrays Cromossômicos* (CMA).

2.2 ESPECÍFICOS

- Revisar a literatura, identificando as possíveis origem dos DsND, o papel das CNVs como fator etiológico, as aplicações clínicas de várias plataformas de CMAs, as aplicações e taxas de diagnósticas mundiais em estudos de coortes com DsND, investigadas com CMA;

- Analisar e caracterizar as CNVs encontradas no grupo de indivíduos investigados de acordo com seu tipo, localização, tamanho, genes presentes, e presença na população;

- Ampliar ou consolidar os dados obtidos no mestrado sobre os fenótipos mais frequentes na coorte e aqueles mais associadas à detecção de CNVs patogênicas;

- Consolidar os dados obtidos durante o mestrado sobre a taxa e o papel da consanguinidade nessa coorte;

-Identificar e examinar padrões de LCSHs, que sugerirem dissomia uniparetal (UPD);

-Identificar e investigar quais são os haplótipos ancestrais de LCSHs comuns nessa coorte e sua frequência;

- Identificar os pacientes com potenciais mutações autossômicas recessivas, que tenham maior probabilidade diagnóstica com uma investigação por análise de

sequenciamento exoma, para auxiliar os médicos no prosseguimento da investigação diagnóstica;

- Com os resultados das análises, propor novas regiões e/ou genes candidatos causais dos DsND.

3 MATERIAL E MÉTODOS

3.1 REVISÃO DA LITERATURA

Para revisarmos a literatura quanto à origem dos DsND, ao papel das CNVs como fator etiológico para os DsND, às aplicações clínicas de várias plataformas de CMAs e à aplicação mundial e às taxas de diagnóstico através do CMA, realizamos uma revisão bibliográfica no repositório PUBMED/MEDILINE. Pesquisamos os termos "CMA" ou "*Chromosomal Microarray*" ou "*array-CGH*" e "*Neurodevelopmental disorders*" ou "*intellectual disability*" ou "ID" ou "*autism spectrum disorder*" ou "ASD" ou "*developmental delay*" e "*Cohort*" em títulos ou resumo, publicados entre 2010 a 2023.

3.2 ASPECTOS ÉTICOS

O projeto foi submetido [ao] e aprovado pelo Comitê de Ética em Pesquisa do Hospital Infantil Joana de Gusmão (HIJG) de Florianópolis-SC, Brasil, sob o parecer Nº 2.339.104 (Anexo A), respeitando as diretrizes e critérios estabelecidos na Resolução 466/12 do Conselho Nacional de Saúde.

Médicos(as) geneticistas, pediatras e neurologistas do Hospital Infantil Joana de Gusmão (HIJG), Hospital Universitário Polydoro Ernani de São Thiago (HU/UFSC) e de clínicas particulares foram convidados a colaborar com a pesquisa através de um Termo de Consentimento Livre e Esclarecido (TCLE) (Anexo B). Foram recrutados os pacientes que realizaram o exame de CMA entre 2013 a 2019, que previamente haviam consentindo e autorizando a divulgação de todos os resultados identificados no exame em publicações científicas mediante o Termo de Consentimento Informado do Laboratório Neurogene (Anexo C), e que aceitaram participar voluntariamente, através da assinatura do TCLE do projeto (Anexo D) apresentado a eles durante suas consultas rotineiras no HIJG e no HU. Nos casos em que os pais ou responsáveis legais dos pacientes entre 6 a 18 anos e intelectualmente capazes consentiram na participação da pesquisa, o Termo de Assentimento foi apresentado e devidamente

explicado aos pacientes (Anexo E e F), assegurando seu direito de participação ou recusa na pesquisa, independente do aceite dos pais ou responsáveis legais.

Para aqueles casos em que os pacientes já não eram mais acompanhados pelos serviços de atendimento do HIJG e do HU, o contato para convite à participação na pesquisa procedeu-se através de comunicação telefônica pelo médico responsável ou com o auxílio de membros da equipe de pesquisa supervisionados pelo médico. Nas ocasiões em que, por algum motivo (óbito, mudança de cidade, perda de contato), não era possível o contato com o paciente, assumimos, mediante a Justificativa Da Ausência Do TCLE (Anexo G), o compromisso de, ao utilizar dados e/ou informações coletadas nos prontuários dos participantes da pesquisa, assegurar a confidencialidade de tais dados e a privacidade dos envolvidos.

3.3 AMOSTRA

Foram selecionados aproximadamente 1020 exames de *Microarrays Cromossômicos (CMA)*, solicitados por médicos geneticistas e neurologistas do Hospital Infantil Joana de Gusmão, Hospital Universitário Professor Polydoro Ernani de São Thiago e clínicas particulares de Florianópolis, realizados através do laboratório Neurogene, com consentimento dos pacientes ou dos seus responsáveis (através do Termo de Consentimento Livre e Esclarecido) e disponibilidade dos médicos para participar da pesquisa. Trata-se de estudo retrospectivo-prospectivo. Além dos 430 exames provenientes do estudo anterior do laboratório (Trabalho de mestrado do candidato, 2018), foram incluídos, cerca de 590 exames de pacientes com DsND realizados no período de 2016 a 2019. Para os 430 exames anteriores, aplicou-se uma reanálise à luz dos novos conhecimentos na área.

3.4 ANÁLISES GENÔMICAS

As plataformas de *Microarray,* utilizadas nos exames realizados pelo Laboratório Neurogene são a CytoScan® 750K, 750.000 marcadores genômicos (550.000 sondas para detectar CNVs e aproximadamente 200.000 SNPs) e a CytoScan™ HD, que oferecem mais de 2,7 milhões de marcadores genômicos (1,95

milhões de sondas distribuídas pelo genoma para detectar CNVs, e 750 mil sondas para SNPs), ambas da Affymetrix (Thermo Fisher Scientific, Waltham, MA). Os arquivos resultantes da leitura dos arrays foram analisados com o software Chromosome Analysis Suite (ChAS) Affymetrix®, com base na versão do genoma humano de fevereiro de 2009 (GRCh37/hg19), dando-se ênfase para as CNVs raras, para as quais a relação genótipo-fenótipo não estava clara.

Para se tentar fazer a relação fenótipo-genótipo, com o intuito de correlacionar possíveis genes causais, bem como correlacionar fenótipos preditivos a CNVs patogênicas, foi coletada a descrição clínica detalhada dos indivíduos afetados junto aos médicos através de um formulário pré-estabelecido.

3.4.1 Análise de CNVs -Viés Clinico

Em nossa análise, consideramos para fins de interpretação as CNVs com tamanhos maiores que 100 Kb para deleções e maiores que 150 Kb para duplicações, ambas com, no mínimo, 50 marcadores, de acordo com as recomendações da ACMG (REHDER et al., 2013). Para analisar e interpretar as CNVs quanto à sua função, efeitos de dosagem (através de estudos de superexpressão ou haploinsuficiência conhecida) e efeitos das mutações, utilizamos ferramentas de bioinformática, analisando os bancos de dados públicos, como o ISCA (International Standard Cytogenomic Array), DGV (Database of Genomic Variants), OMIM (Online Mendelian Inheritance in Man), DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources), e bancos de dados privados, como o CAGDB (Cytogenomics Array Group CNV Database). A maioria desses recursos é acessível através do navegador genômico da Universidade de Santa Cruz, EUA (UCSC Genome Browser).

Todas as variáveis - como localização, tipo e tamanho de CNV, classificação da CNV, descrições clínicas (fenótipos), dados de exames relevantes (cariótipo, X-frágil etc.), faixa etária, idade de realização do exame, sexo do paciente e quantidade - foram compiladas em uma planilha simples do Excel, para posterior aplicação de análises estatísticas utilizando o software R (versão 3.4.2, The R Foundation for Statistical Computing). O objetivo era compreender a média e o desvio padrão do tamanho das CNVs, a frequência dos fenótipos clínicos, a taxa de diagnóstico do estudo, a idade média e a proporção de cada sexo na coorte estudada, a frequência

de alterações genômicas em cada cromossomo e a relação entre fenótipos individuais ou grupos de fenótipos clínicos com a interpretação das CNVs. Além disso, buscamos identificar indicativos que pudessem auxiliar na seleção dos pacientes mais adequados para a submissão ao CMA como teste de primeira linha.

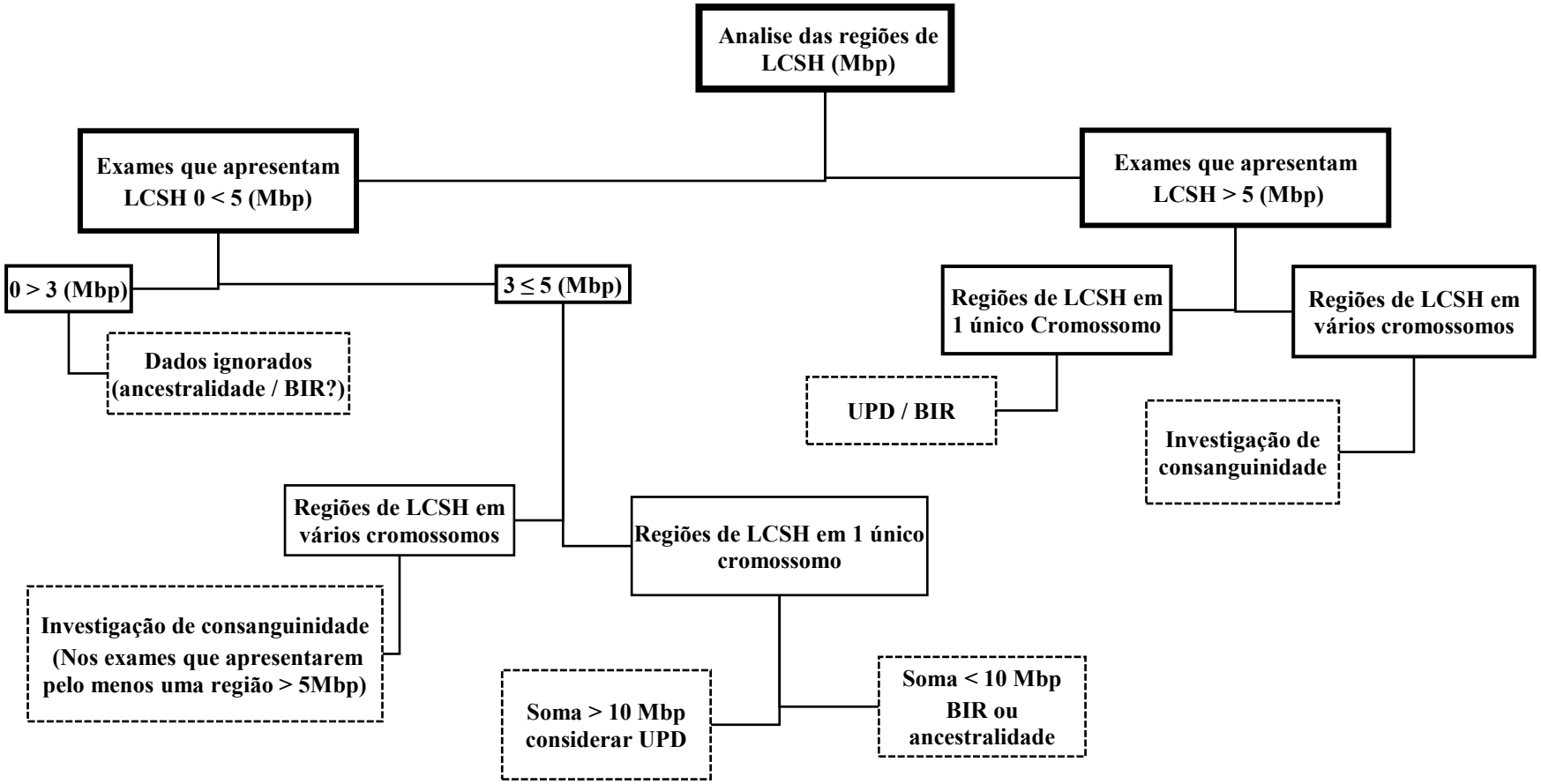
Para estabelecer a relação fenótipo-genótipo, com o objetivo de correlacionar possíveis genes causais, foi coletada junto aos médicos uma descrição clínica detalhada dos indivíduos afetados, por meio de um formulário preestabelecido (Apêndice A). Esse formulário foi elaborado para obter informações abrangentes sobre o quadro clínico de cada paciente, incluindo histórico de exames físicos, moleculares, metabólicos, genéticos, comportamentais e o uso de medicações.

3.4.2 Seleção e análise dos LCSHs

A análise dos exames que apresentaram LCSHs seguiu a metodologia descrita em [3]. O Fluxograma 1 ilustra o fluxo da análise de amostras contendo LCSHs.

Os arquivos resultantes foram analisados usando o software Chromosome Analysis Suite (ChAS) Affymetrix®, que é baseado na sequência do genoma de referência do banco de dados da Universidade da Califórnia, Santa Cruz (https://genome.ucsc.edu/cgi-bin/ hgGateway). Usamos a versão do genoma humano de fevereiro de 2009 (GRCh37/hg19) para a análise. Para focar em dados clinicamente relevantes e evitar confusão na análise, aplicamos um limite de ≥ 3 megapares de bases (Mbp) para a análise de LCSH. Esse limite é normalmente usados em investigações clínicas, ao contrário de estudos de base populacional, onde o limite de corte geralmente é consideravelmente mais baixo [1]. Incluímos todos os participantes que possuíam LCSHs que atendiam aos critérios acima, independentemente de apresentarem uma variante do número de cópias patogênica (CNV).

Fluxograma 01- Análise dos exames que apresentaram regiões de LCSH



LCSH (longos trechos contínuos de homozigose), BIR (replicação induzida por quebra) e ancestralidade (homozigosidade por haplóstipos ancestrais)

Fonte- Autor, 2023

3.4.2.1 Análise da consanguinidade

Para verificar a frequência de consanguinidade na coorte, aplicou-se a metodologia baseada no trabalho de Kearney, Kearney e Conlin (2011). Nos padrões de homozigose que sugeriram consanguinidade, foram somadas todas as regiões de LCSH >3 Mb distribuídas nos cromossomos, excluindo-se as LCSHs localizadas nos cromossomos sexuais, dividindo-se o somatório total em Mb pelos 2.881 Mb de tamanho do genoma autossômico (GRCh37/hg19) e correlacionando-se a porcentagem obtida com o coeficiente de endogamia (F), que é de 25% (1/4 - pai-filho/irmão-irmã), 12,5% (1/8 - tio-sobrinha/tia-sobrinho), 6,5% (1/16 - primos de primeiro grau/meio-tio-sobrinha), 3,12% (1/32 - primos de primeiro grau afastado), 1,5% (1/64 - primos de segundo grau), 0,5% (1/128 – primos de terceiro grau).

3.4.2.2 Dissomia uniparental

Nos exames em que foi constatada a presença de um ou mais LCSHs restrito(s) a um único cromossomo autossômico com tamanho ou soma (no caso de múltiplos LCSHs no mesmo cromossomo) ≥ 10 Mb, foi considerada uma possível dissomia uniparental (isodissomia).

Como o software ChAS reconhece regiões com perda de heterozigosidade (LOH), que incluem não apenas regiões homozigóticas, mas também regiões hemizigóticas geradas por uma deleção, todos os casos que tinham LOHs ≥ 10 Mbp de tamanho (ou soma dos tamanhos) em um único cromossomo autossômicos, independentemente da presença de mais cromossomos com LOHs, foram revisados manualmente, para eliminar o efeito de confusão de eventuais regiões hemizigóticas para chamar LSCHs e, finalmente, um UPD.

3.4.2.3 Análise das LCSH mais frequentes

Selecionamos e analisamos as corridas de CMAs para as citobandas, que mais frequentemente apresentaram regiões com LCSH ≥ 3 Mbp em um cromossomo autossômico e aquelas presentes em mais de 5% dos indivíduos foram consideradas

LCSHs comuns. Esse percentual foi escolhido porque a frequência de ≥ 1%, que é o limite usual para definir polimorfismos comuns de SNPs em uma população, não foi considerada aplicável aqui por se tratar de uma coorte afetada. Outros também escolheram o mesmo limiar (ou inferior) para considerar o LCSH encontrado em uma coorte afetada como variação comum, provavelmente sem significado clínico para sua análise (282-286). Portanto, ao fazê-lo, acreditamos ter uma margem de segurança adequada para selecionar LCSHs comuns devido a haplótipos ancestrais, e não devido a consanguinidade ou outros mecanismos relacionados à patogênese.

Para delinear uma posição genômica mais precisa para as LCSH mais frequentes, as seções de homozigose compartilhada foram sobrepostas e suas posições genômicas obtidas com base na mediana de seu início e fim.

3.4.2.3.1 Bioinformática nas análises de LCSHs.

Para análise de consanguinidade e análise comparativa de LSCH entre os casos, bem como para UPD, todos os LOHs encontrados no CHAS para cada caso foram copiados com identificação codificada, compilados em planilhas de Excel para análise.

Para uma análise mais adequada e precisa, o processo foi automatizado, e todos os LCSHs encontrados na coorte foram importados para o Google Colab e manipulados usando a linguagem de programação Python. As bibliotecas usadas para manipulação e análise de dados foram Pandas e NumPy (para cálculos numéricos). O código utilizado para a análise está disponível na página GitHub do projeto: https://github.com/tiagochavo87/LCSH_analysis

4 RESULTADOS

Os resultados desta tese foram organizados na forma de artigos científicos e compreendem dois capítulos:

5.1 **CAPÍTULO I** - Copy number variations as a causal genetic factor in neurodevelopmental disorders: a review of clinical applications.

5.2 **CAPÍTULO II** - A cohort study of individuals with neurodevelopmental disorders and/or congenital anomalies investigated by high-resolution chromosomal microarrays in southern Brazil: *The significance of autism spectrum disorder*.

4.1 CAPÍTULO I - COPY NUMBER VARIATIONS AS A CAUSAL GENETIC FACTOR IN NEURODEVELOPMENTAL DISORDERS: *A REVIEW OF CLINICAL APPLICATIONS*.

Variações Do Número De Cópias Como Fator Genético Causal Nos Distúrbios Do Neurodesenvolvimento: ***Uma revisão das aplicações clínicas.***

Na última década, a citogenética molecular deu um salto na investigação genética de Distúrbios do Neurodesenvolvimento (DsND), como a deficiência intelectual (DI) e/ou transtorno do espectro do autismo (TEA), que afetam cerca de 3-4% da população mundial. Em particular, as variações do número de cópias (CNVs) (microdeleções e microduplicações cromossômicas), detectadas por Microarranjos Cromossômicos (CMAs), desempenham um papel importante na etiologia dos DsND. Revisamos (1) a origem dos DsND, (2) o papel das CNVs como fator etiológico para os DsND, (3) as aplicações clínicas de várias plataformas de CMAs e (4) a aplicação mundial e as taxas de diagnóstico através do CMA, analisamos a literatura publicada entre 2010 e 2023, utilizando os termos: CMA ou Chromosomal Microarray ou array-CGH; e Distúrbios do Neurodesenvolvimento ou Deficiência Intelectual ou DI ou transtorno do espectro do autismo ou TEA ou atraso no desenvolvimento; e Coorte, em títulos ou resumo. Dos 189 estudos de coorte encontrados, 84 casos foram selecionados, mostrando uma taxa diagnóstica média de 16,8% para detecção de CNVs patogênicas por microarranjos cromossômicos. Encontramos uma média diagnóstica de 16% com base em 84 estudos transversais que usaram CMAs para investigar a etiologia dos DNs. Esses estudos publicados apresentam amostras de diferentes populações e etnias, com as maiores amostras em países do hemisfério norte. Apesar dos CMAs serem recomendados como teste de primeira linha na investigação de DsND e, mais recentemente, até mesmo o uso de dados de sequenciamento de nova geração (NGS) para avaliar CNVs, na realidade econômica de países menos desenvolvidos o uso da citogenética clássica ainda é predominante, quando não é o único recurso.

Palavras-chave: *Microarrays*, deficiência intelectual, transtornos do espectro autista, transtornos do neurodesenvolvimento, microarranjos cromossômicos.

COPY NUMBER VARIATIONS AS A CAUSAL GENETIC FACTOR IN NEURODEVELOPMENTAL DISORDERS: A review of clinical applications

CHAVES, TIAGO FERNANDO[1A*]; MUNIZ, YARA COSTA NETTO[2A] AND MARIS, ANGELICA FRANCESCA[3B].

[1] BIOLOGIST, PHD STUDENT IN CELL BIOLOGY AND DEVELOPMENT, tiagochavo@msn.com (CORRESPONDING AUTHOR);
[2] BIOLOGIST, PHD IN GENETICS, PROFESSOR OF THE DEPARTMENT OF CELL BIOLOGY, EMBRYOLOGY AND GENETICS.
[3] BIOLOGIST, PHD IN GENETICS AND MOLECULAR BIOLOGY,
[A]LABORATÓRIO DE POLIMORFISMOS GENÉTICOS (LAPOGE) - UNIVERSIDADE FEDERAL DE SANTA CATARINA – FLORIANÓPOLIS –SC.
[B]CHILDREN'S HOSPITAL JOANA DE GUSMÃO – FLORIANÓPOLIS – SC, BRAZIL.

EMAIL: tiagochavo@msn.com

ABSTRACT.

In the last decade, molecular cytogenetics has taken a leap forward in the genetic investigation of Neurodevelopmental Disorders (NDDs), such as intellectual disability (ID) and/or autism spectrum disorder (ASD), which affect around 3-4% of the world population. In particular copy number variations (CNVs) (chromosomal microdeletions and microduplications), accessed through Chromosomal Microarrays (CMAs), were found to play an important role in the etiology of NDs. We review (1) the origin of NDs, (2) the role of CNVs as an etiological factor for NDs, (3) the clinical applications of various platforms of CMAs and (4) worldwide application and diagnostic rates trough CMA, analyzing the literature published between 2010 and 2023, using the terms: CMA or Chromosomal Microarray or array-CGH; and Neurodevelopmental Disorders or Intellectual Disability or ID or autism spectrum disorder or ASD or developmental delay; and Cohort, in titles or abstract. Of the 189 cohort studies found, 84 cases were selected, showing an average diagnostic rate of 16.8% for detection of pathogenic CNVs by chromosomal microarray. We found a diagnostic mean of 16.8% based on 84 cross-sectional studies that used CMAs to investigate the etiology of NDs. These published studies present sampling of different populations and ethnicities, with the largest samples being in countries in the northern hemisphere. Despite the CMAs being recommended as a first-line test in the investigation of NDs and more recently even the use of next generation sequencing (NGS) data to evaluate CNVs, in the economic reality of less developed countries the use of classical cytogenetics is still predominant, when not the only resource.

Keywords: Microarrays, intellectual disability, autism spectrum disorders, neurodevelopmental disorders, chromosomal microarrays.

NEURODEVELOPMENTAL DISORDERS

Neurodevelopmental disorders (NDDs) are heterogeneous conditions, complex, ambiguous and difficult to conceptualize. However, in clinical practice and research, the term ND is consolidated as a reference to a group of characteristics that compromises normal development of the central nervous system [1, 2]. Their clinical signs usually are perceived in childhood before puberty, often since the initial postnatal period and sometimes even detected during prenatal care (intrauterine) [1, 3].

NDDs mainly causes phenotypes such as intellectual disability (ID) and/or autism spectrum disorder (ASD), neurological impairments that affect around 3-4% of the world population [4, 5]. Clinical conditions may be present as isolated phenotypic manifestations, being called non-syndromic, or may be associated with the presence of apparent dysmorphisms or congenital anomalies (CAs), defined as syndromic [6].

Intellectual Disability

ID is a condition characterized by deficits in cognitive functioning, usually accessed by an intelligence quotient test with an IQ <70, and in adaptive behavior, beginning in childhood. There are several indices of cognitive impairment, such as profound ID (IQ <20), severe ID (IQ <34) moderate ID (IQ 35 - 49) and mild ID (IQ 50 - 70). For a long time, this clinical aspect was referred to as mental retardation, a term later considered inappropriate and currently in disuse [7].

The etiology of ID involves heterogeneous causes, such as environmental, sociocultural, neurodevelopmental, and genetic factors, often correlated with each other, during a "critical time window", as illustrated in Figure 1, in which different factors affect neurodevelopment. The timing of gene expression, combined with environmental factors, determines brain development, especially during the intrauterine phase [8]. The degree of cognitive impairment can often be related to the casual factor of ID, with perinatal asphyxia, prenatal infections or strokes, as well as gross chromosomal imbalances being more common in cases where intellectuality is most compromised [9].

Genetic factors account for about 4 to 15% of mild ID and 20% to almost 50% of more severe ID [10]. When a history of parental consanguinity is present, autosomal

recessive mutations predominate [11]. Considering hereditary ID in general, the fragile X syndrome (OMIM), caused by amplification of a trinucleotide repeat (CGG) in the gene FMR1, is the most prevalent monogenic disorder in mild to severe ID [12].

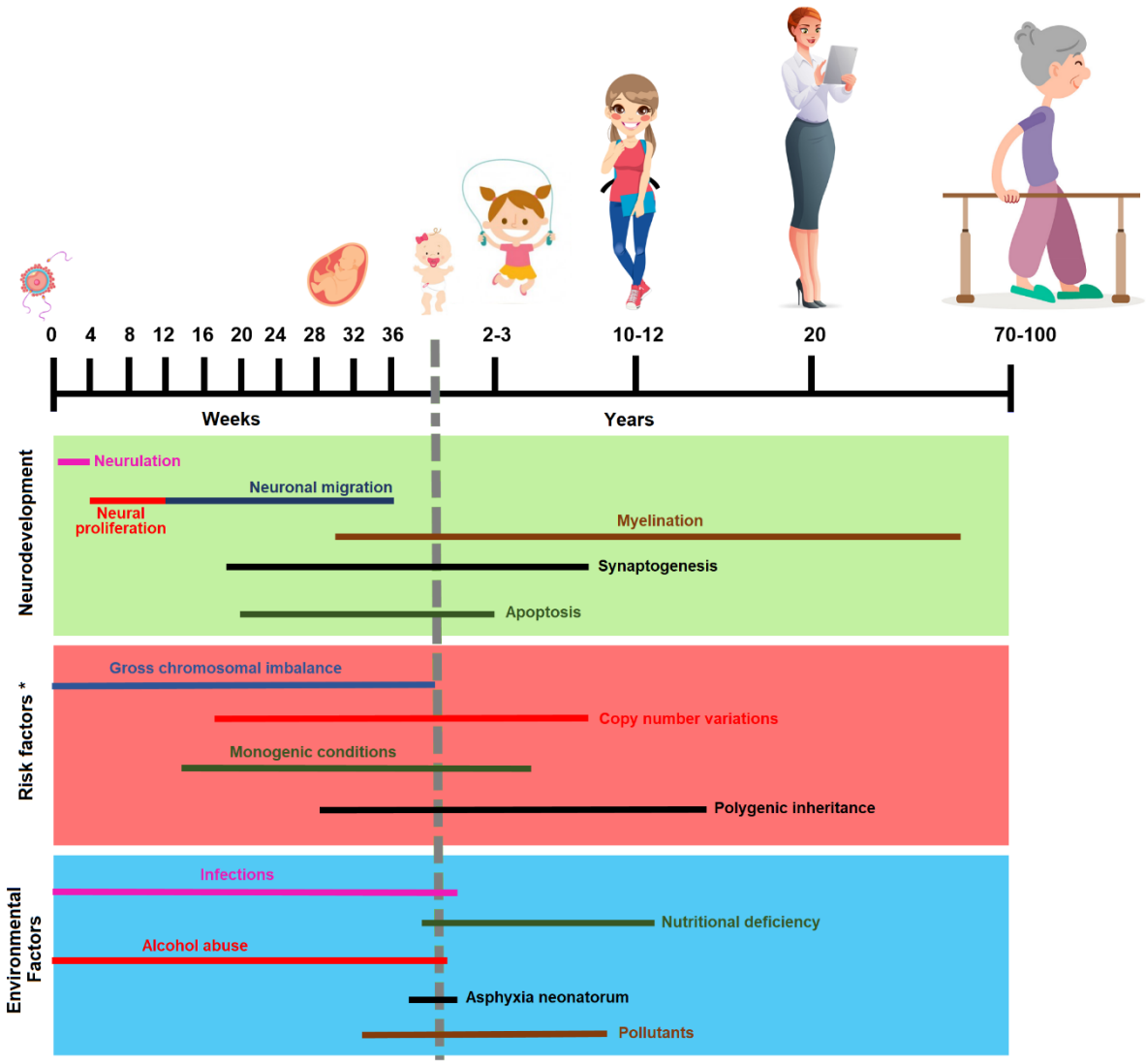


Figure 1 Schematic representation of stages of Neurodevelopment, related to genetic and environmental factors and their time window. Modified from Chiurazzi & Pirozzi, 2016.[9]. *For genetic factors as etiologic causes of ID, the onset of clinical signs or the time of detection is shown.

Chromosomal aberrations (microscopic and submicroscopic) are one of the most important genetic etiologies of ID and represent about 25% of the cases. With trisomy 21 alone accounting for more than 7% of these cases, other typical chromosomal changes include the trisomy's of chromosomes 18 and 13, chromosome X imbalances, like X monosomy, and recurrent microdeletions, primarily involving chromosomal regions 1p36, 2q37, 4p16, 5p15, 7q11.2, 8p23.1, 8q23q24, 9q34.3, 11p11.2, 15q11.2, 16p13.3, 17p13.3, 17p11.2, 17q21.3, 18q23, 22q11.2, 22q13 [13].

Some environmental factors can act in a pervasive manner, depending on the gestational phase, dose or exposure time. Environmental chemicals, pollutants, drugs, smoking, and excessive ionizing radiation can lead to mutations in germ cell lines, that can affect individual genes, that can cause aneuploidies, euploidies and structural chromosomal changes via chromosomal breakage and DNA repair. DNA repair mechanisms can also result in submicroscopic deletions or duplications in the chromosomes, resulting in altered copy number of gene-containing DNA stretches, often causing expression changes in neurodevelopment-related genes. Similarly, maternal exposure during pregnancy to drug use, alcohol abuse, infections, and metabolic conditions such as iodine deficiency and nutritional deficiency (mother and fetus), diabetes or phenylketonuria, are environmental etiologies with varying effects [9].

The diagnosis of ID is considered only after the age of five, since the available cognitive assessment tools are generally not applicable before this age. For children under five the term developmental delay (DD) is used, considering the possibility of a mere delay which may eventually be overcome [11].

Autism Spectrum Disorder

Considered the second most common neurodevelopmental disorder after ID, Autism Spectrum Disorder (ASD) affects approximately 1 in every 54 individuals [14]. According to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V)[15], the hallmarks of ASD are deficits in social communication and interaction, restricted or repetitive behaviors or interests that negatively impact their ability to function in school, work, and other areas of life. ASD include individuals with and without ID, and comprises the previously separate diagnostics of Autism, Asperger's Disorder and Invasive Developmental Disorder, since they are considered to be several manifestations of the same condition (DSM-V 2013). It is diagnosed in male subjects at a 4:1 ratio, its incidence (or diagnosis) has increased noticeably over the years [12]. Asperger's syndrome (AS) is a subtype of ASD where linguistic and cognitive development is preserved and represent the most prominent feature of autism (deficits in social communication and interaction, restricted or repetitive behavior’s) [16].

ASD is considered one of the most heritable neurodevelopmental disorders, with an estimated heritability of 83 to 93%. In approximately 5 - 15% of those affected,

especially those accompanied with ID, de novo copy number variations (CNVs) or rare single nucleotide variants (with a population frequency of < 0.05%) can be identified in clinically relevant genes. It is noteworthy that even being of primarily genetic etiology, in about 80% of cases it is still not possible to clarify the changes involved. To date, more than 100 ASD susceptibility genes with widely indeterminate expressiveness and penetrance variants are known, but the pathogenetic mechanistic is still unclear [16–18].

Congenital Anomalies and other Frequent Comorbidities of NDDs

Congenital anomalies (CAs), which are often the cause or accompany NDDs, consist of malformations of the normal anatomical pattern present at birth [19]. Their presence seriously influences the quality of life of those affected and is a major cause of child mortality worldwide. In developed countries, such as the United States, CAs account for around 20% of child deaths [18]. In Europe, data from the European Surveillance of Congenital Anomalies showed a perinatal mortality rate of around 9.3 newborns with CAs per 10,000 births between 2008 and 2012 [20]. In Latin American countries, CAs are between the second and fifth causes of death in children under one year of age [21], and in Brazil, more specifically, they are the second cause. Worldwide estimates indicate that about 1 in 33 live births has at least one CAs, leading to a total of 3.2 million newborns with CAs per year [22].

The etiology of CAs involves heterogeneous causes, environmental and genetic factors, often correlated with each other, which lead to impairment of differentiation and development of tissues, organs or body patterns (limbs, face, ears, eyes, etc ...). Estimates suggest that about 10% of CAs are the result of the detrimental effect of environmental factors on fetal development, i.e., the effects of teratogens. The term teratogen defines any environmental factor that can produce an CAs. Genetic factors such as autosomal or X-linked genetic inheritance, chromosomal imbalance or de novo mutations represent 10 to 25% of the known causes of CAs and 65% remain without clear etiology, involving polygenic, multifactorial genetic factors (gene-environment interactions), stochastic developmental abnormalities or synergistic interactions with teratogens as the cause [23].

Other less frequent phenotypes associated or not with DI, ASD and CAs, also characterize NDDs, such as epilepsy that affects approximately 1% of the population.

This condition has an estimated genetic etiology of 40%, and the phenotype is part of the non-primary clinical condition of more than 200 Mendelian disorders, in addition to its association with non-Mendelian disorders (complex diseases) and chromosomal disorders affecting genes related to neuronal ion channels [24, 25].

Affected with NDDs due to their often-compromised adaptive skills, limited intellectual capacity, and frequent association with other comorbidities, impact the daily lives of family members responsible for their care, as well as needing greater support from public health systems [3, 26].

Proper diagnosis is necessary for the clinical follow-up of individuals with NDDs and to provide the family with appropriate genetic counseling to prevent the risk of recurrence. [23] Due to the high clinical and genetic heterogeneity, studying and diagnosing NDDs is one of the most complex fields in health, as both genetic and environmental factors, alone or in combination, can play an important role in their pathogenesis [24, 25].

Since the scope of this review is the relationship between copy number variations (CNVs) and neurodevelopmental disorders, we will focus on the Chromosomal Microarray (CMA) and Next Generation Sequencing (NGS) technologies, since, as reinforced by Savatt and Myers (2021)[27], it is crucial that pediatric clinicians understand these now common tests and their role in the care of children with neurodevelopmental disorders.

COPY NUMBER VARIATIONS (CNVs)

The genetic sequence of the human genome, as well as that of all biological organisms, is constantly changing, and stochastic mutations, through often imperceptible phenotypic changes, are targets of natural selection. This process leads to a possibility of evolution and adaptation. In particular, the genetic material of humans is similarly shared between individuals around the world, and estimates point to a difference in SNPs in only about 0.1% of the DNA sequence between people. Detection, analysis and study approaches for this simple fraction were considered responsible for the genetic variation among individuals, widely used for phenotypic correlations and susceptibility to certain pathologies [24]. However, the study of microduplications and microdeletions in the human genome using the Comparative

Genome Hybridization techniques demonstrated that normal individuals, that is, individuals not considered to be affected clinically by idiopathic genetic factors, carry deletions and/or duplications in their DNA, including gene-containing genomic stretches. These CNVs vary in size from dozens of bases (>50 bp) to megabases (Mb) within a single human genome, differentiating human genomes by more than 1.2%, due to CNVs, an unexpected source of hitherto unknown variation to be explored in addition to SNPs [22, 23].

In one of the pioneering studies relating CNV to a phenotype, Bridges [25] described the duplication of the Bar gene in Drosophila melanogaster, associated with the small-sized eye phenotype, called the bar-eye (“bar”). Currently, CNVs have been highlighted as an abundant form of variation in the human genome whose study has increasingly demonstrated its association with genetic and phenotypic diversity [28, 29]. Recent work by Zarrei et al. [30], based on 55 studies of CNVs detected at high resolution in unaffected individuals, demonstrated that approximately 100 genes can be eliminated homozygously in the human genome without leading to apparent phenotypic effects.

If, on the one hand, CNVs are abundant sources of genomic variation, on the other hand, they are also important genetic factors in NDDs. CNVs can affect the function of genes in several ways: (1) the deletion or interruption of one or more genes can cause functional loss due to haploinsufficiency of the remaining allele(s) or by unmasking recessive mutations or can alter communication between alleles by suppressing regulatory elements; (2) duplications can alter the expression of dominant or recessive alleles. In either case, there is the potential to cause relevant clinical phenotypes [5, 31]. There is a continuous spectrum of phenotypic effects of CNVs, from adaptive traits to embryonic lethality [30].

Identifying and interpreting the contribution of each variation to neurodevelopment has been a challenge both in research and in clinical diagnostic laboratories. There is a limit where the correlation of the phenotype with CNV starts to be more subjective, as well as the contribution of other factors, such as ethnic origins or environmental factors, make interpretation difficult, especially when these CNVs are rare and there are no recurrences of association with the phenotype [30]. Essentially, new pathogenic CNVs continue to be described for different classes of diseases [32].

Molecular mechanisms that originate CNVs

Deletions, duplications, triplications, additional amplifications, CNVs, absence of heterozygosity (AOH), copies of neutral inversions, insertions and translocations within the DNA, are examples of structural variations that involve interrupting the structure of the double strand of DNA, resulting in changes that encompass many base pairs (bp), unlike variations like SNPs that result in a change restricted to one or a few bases. These structural variations are the consequences of mechanisms based on recombination and replication, involved in the synthesis, replication, recombination and repair of genetic material [33].

Recent publications addressing the molecular mechanisms of DNA replication, recombination and repair in humans have contributed to a better understanding of how structural variations can occur in a specific genomic region and lead to changes in gene expression, both locally or elsewhere in the genome, due to its restructuring [34–38].

One of the ways to theorize the potential mechanisms for the formation of genomic rearrangements involved in the formation of a variation in a specific locus is to evaluate the characteristics of the genomic region in question regarding its organization, observing the presence of repeated sequences, their relative orientation, size, density and distribution. According to Lander et al. [39], repetitive sequences account for approximately 50% of the entire human genome and consist of moving elements, repetitions of simple sequences, repetitive tandem sequences (consecutive, typical of centromeres, telomeres, short arms of acrocentric chromosomes and groupings of ribosomal genes) and, in particular, low-copy repeats (LCRs). Also known as segmental duplications, LCRs are elements from 10 to 400 Kpb in length, with a sequence identity greater than 95%, that comprise up to 5% of the genome, being present in two or more copies in the haploid reference genome [40].

For the formation of a CNV, initially it is necessary to have one or more breaks in the DNA, followed by the junction of genomic sequences, with gain or loss of DNA stretches. This can happen due to uneven crossover between non-allelic sequences or by repair mechanisms [30, 40].

Recurrent CNVs

Many CNVs that are found recurrently originated a long time ago, being inherited from one generation to another, thus being typical of certain populations or sub-populations. It is even possible that some are related to typical phenotypic variations among populations. However, what draws attention are some CNVs that are not inherited and are often shared by unrelated individuals. These probably have their origin in rearrangements prone to recurrence by mechanisms involving repetitive sequences. Recurrent CNVs can be inherited and are typical of certain populations or sub-populations, while non-inherited CNVs may be caused by mechanisms such as Nonallelic Homologous Recombination (NAHR) [41].

Nonallelic Homologous Recombination - NAHR

The NAHR (Nonallelic Homologous Recombination), also known as ectopic recombination (crossover off the correct position), occurs during the crossover in meiosis between direct homologous sequences paralogs (generated by duplication during evolution) that are not in allelic positions.

The LCRs are quite prone to NAHR, especially among LCRs in the same orientation (direct repeats). The affected regions may contain dose sensitive genes and give rise to a pathogenic CNV after an NAHR [41]. With preserved break points, the NAHR produces recurring and sometimes reciprocal deletions and duplications, the frequency of which depends on the distance and length of the CSFs involved.

Currently, over 40 non-overlapping genomic loci associated with NAHR are known to result in microdeletion and microduplication syndromes related to genomic disorders [42]. Lupski [43], estimated the prevalence of recurrent CNVs at birth in the DiGeorge-Velo cardiofacial (del22q11.2), Williams- Beuren (del7q11.23) and Smith-Magenis (del17p11.2) syndromes at 1/4,000, 1/10,000 and 1/25,000, respectively.

Non-recurring CNVs

Non-recurrent CNVs are most often associated with complex structural rearrangements within or among chromosomes and have unique sizes due to the random location of their break and junction points. They are usually CNVs of exclusive genomic size in the individual, which makes CNVs challenging for clinical interpretation [44].

Eventually, non-recurring CNVs can also be delimited by LCRs. Carvalho et al [33], exemplify the formation of a complex rearrangement of CNVs mediated by inverse parallel repetitions, whose final product consisted of an inverted and interspersed triplication and with duplicated genomic segments, described as DUP-TRP/INV- DUP. The same authors report the formation of non-recurring CNVs involving genomic regions with tandem LCRs, but point out that most pathogenic non-recurrent CNVs are formed in regions without LCRs.

Non-recurring CNVs of different sizes that occur at a given locus in the genome of individuals who share similar phenotypes can be aligned to reveal the smallest overlapping region, delimiting candidate genes for the manifested phenotype [36].

The most common mechanism of formation of non-recurrent CNVs is by Non-Homologous End Joining (NHEJ).

Non-Homologous End Joining – NHEJ

Non-Homologous End Joining is the main repair mechanism for chromosomal breaks (double-strand breaks in DNA) that occur in human cells, mainly as a result of oxidative damage and ionizing radiation, and is active throughout the cellular cycle. It repairs DNA via a multiprotein-complex whose composition depends on the nature of the double break, essentially recognizing and mending broken DNA ends [45][41]. The repair by NHEJ is considered mutagenic, requiring in about 60% of the amendments, only the alignment of the double chains with at least 1 or 2 bp of microhomology between the ends. Besides being responsible for unique CNVs, NHEJ also generates small deletions and insertion of random nucleotides at the junction of the break points. Most pathogenic non-recurrent CNVs in NDDs have microhomologies of 2-33 bp in length at these junctions. In addition, the insertion of short segments (<100 bp) homologous to close genomic regions has been observed in up to 35% of non-recurrent CNV junctions [46, 47].

For a more detailed and comprehensive discussion of structural variation in the human genome, we suggest reading Carvalho and Lupski (2016) [33].

GENETIC SCREENING AND DIAGNOSTICS OF CNVs IN NDDs.

Since the introduction of chromosomal banding techniques in the late 1960s cytogenetic tests for screening and diagnosing NDDs are constantly evolving and developing [48]. Conventional karyotype analysis has been a useful tool in genetic diagnosis, being the gold standard cytogenetics test for a long time, detecting whole or partial chromosome losses or duplications and eventual structural changes. Considered a low resolution test, the classical karyotyping can only detect gains or losses of DNA that are larger than 5-10 Mpb, depending on the chromosomal region, and the technique used, thus resulting in a low diagnostic rate, identifying genetic changes in only about 3% of individuals with neurodevelopmental disorders [7, 49].

To identify genetic syndromes involving deletions and duplications below 5-10 Mbp resolution, various techniques have been developed. Among them, Fluorescence in Situ Hybridization (FISH) is widely used, which involves the use of fluorescent genomic probes complementary to the known deleted or duplicated DNA region in the investigated syndrome. These probes, along with a control probe to another region, are labeled with fluorescent markers and then hybridized to the DNA of the patient's cells, which are placed on a microscope slide [50]. Subsequently, in a fluorescence microscopy analysis, metaphase cells are examined for the presence or absence of fluorescent emissions from the probes, which allows the identification of deletions/duplications in sizes of 2-5 Mb. However, this technique requires a clear diagnostic hypothesis and knowledge of the exact sequence to be investigated. The development and application of this technique marked the beginning of molecular cytogenetics [51].

In 1992 a new method for detecting chromosomal imbalances emerged, called Chromosomal Comparative Genomic Hybridization (CGH), where entire chromosomes of a test genome, colored with fluorescence are hybridized (combined) with a reference genome marked with fluorescence of a distinct color (usually the fluorescent molecules used are red and green). When the DNA of the complementary chromosomes of test and reference genomes hybridize together equally the resulting fluorescence is yellow. When either green or red fluorescence predominates at a specific chromosomal region, then this stretch of DNA is over or underrepresented, allowing to infer duplications or deletions in one genome regarding the other. The classic CGH has an average resolution of 10 - 20 Mbp and, even though it is not able to detect minor chromosomal changes, it has been widely used in several types of

cancer, in the search for duplications and/or deletions in tumor DNA as it allows direct comparison of the genome of normal versus tumor cells from one individual [33].

Comparative genomic hybridization by arrays (array-CGH or aCGH) also known as Chromosomal microarray (CMA) or simply genomic microarrays, emerged as a development of the classical CGH, allowing to detect CNVs of much smaller size. It allows the hybridization of small fragments of a test genome to probes of a reference genome that are fixed in microwells to a solid matrix (microchip), using the same CGH comparison principle (Figure 2).

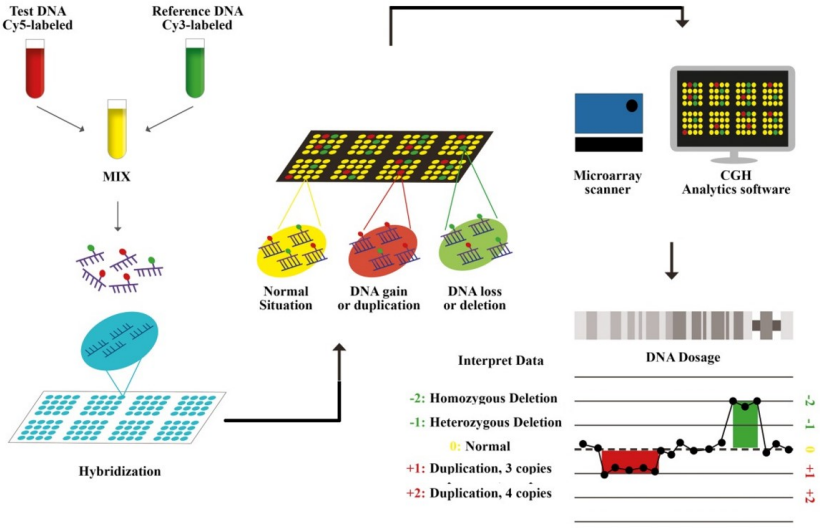


Figure 2- Array comparative genomic hybridization - array-CGH.
Modified from Colaianni; Mazzei and Cavallaro, 2016 [51].

There is however an important limitation of the microarray in relation to the conventional karyotype. Because of the quantitative nature of this test, CMA does not detect balanced changes, as can occur in some inversions and translocations, and cannot specify where in the genome the additional DNA is located in the event of dosage gain. Also, CMA fails to detect the often-complex chromosomal rearrangements that may underlie an apparently simple duplication or deletion, which occurs even more often when more than one significant CNV is found in the genome [53, 54].

Technology based on Chromosomal Microarray

Initially, the genomic microarray technology used probes whose matrix was made from DNA segments cloned into vectors such as Bacterial Artificial Chromosomes (BACs) and Yeast Artificial Chromossomes (YACs). These DNA clones were relatively large, with sizes of approximately 150 and 200 Kpb, generating an intense hybridization signal with a high signal-to-noise ratio. Targeting mostly subtelomeric, pericentric and other regions involved with recurrent microdeletion or microduplication syndromes already characterized, in the remaining genome the clones (probes) were arranged to cover intervals of about 1 Mbp or larger. Although this resolution does not allow to detect minor or unknown microdeletions/microduplications, the diagnostic yield of this technique is 7% to 11% in cases already screened with classical cytogenetic analysis [9, 55].

With the improvement of the array technology, arrays based on oligonucleotide probes emerged and are currently the predominant arrays in the detection of CNVs, having replaced the BAC/YAC -based arrays in clinical practice. The great advantage of the oligonucleotide CMA is the small size of the probes which vary between ~25 to 85 bp, allowing a higher resolution, depending on the number and spacing (density) of the probes used, increases the diagnostic yield in 15% to 20% compared to conventional karyotyping [48, 56, 57].

Currently, commercially available high-density oligonucleotide arrays have available three different technologies: (1) Chromosomal Microarray (CMA) with non-polymorphic oligonucleotides, (2) genotyping arrays of Single Nucleotide Polymorphisms (SNP arrays) and (3) platforms which combine both, CMA and SNP arrays.

As of 2011, the use of CMA has been clinically recommended as the preferred cytogenetic diagnostic test for patients with NDDs, including delayed neuropsychomotor development.

CMA Platforms

The main companies that work with commercial oligonucleotide platforms are Agilent, Illumina and Affymetrix (Thermo-Fisher Scientific).

Agilent Technologies® and has several array-CGH oligonucleotide platforms (1x1M, 2x400K, 4x180K and 8x60K), for example the 8x60K platform has eight areas with ~ 60,000 probes in each area, with probes dispersed throughout the entire genome

(Agilent Technologies, Santa Clara, USA), with an average spacing of 40 Kpb between the probes.

Illumina has the CytoChip Oligo 2x105K (2 x ~105k probes) and CytoChip ISCA microarray platforms, which have a design developed with a focus on neurodevelopmental disorders.

One of the most modern CMAs platforms is Affymetrix CytoScan®HD (Affymetrix, Santa Clara, USA), currently purchased by Thermo Fisher, which offers more than 2.7 million genomic markers, including 1.95 million distributed probes by the genome to detect CNVs, in addition to 750 thousand probes for single nucleotide polymorphisms (SNPs). On this platform the average probe spacing for the RefSeq genes is 400 bp, with 96% of the genes represented by probe coverage.

Another Affymetrix platform is CytoScan® 750K Arrays, with lower resolution and probes empirically selected from CytoScan® HD, has 750,000 genomic markers, consisting of 550,000 unique non-polymorphic probes and approximately 200,000 SNPs. For both, the recommendation as a parameter of analysis is 150 Kpb for gains and 100 Kpb for genomic losses, according to the recommendations of the ACMG (American College of Medical Genetics) [58].

Technology based on New Generation Sequencing (NSG)

Another major advance in molecular diagnostics and genetic testing has been the advent and enhancement of next-generation sequencing (NGS) technologies, such as targeted sequencing [59, 60], whole-exome sequencing (WES) and whole-genome sequencing (WGS) [36, 59, 61, 62].

NGS involves sequencing millions of small DNA fragments in parallel, while WGS sequences each of the three billion bases in the human genome multiple times to provide high depth and accuracy. Bioinformatics analysis is then used to align and piece together these fragments using the human reference genome as a guide. NGS can be used to sequence entire genomes, as in WGS, or restricted to specific areas, including the coding regions of all 22,000 genes (WES) or a small number of individual genes [63]. That revolutionized the ability to simultaneously analyze multiple genes accurately and efficiently and has made it possible for laboratories and clinical research centers to create multigene panels for many clinical indications [64, 65].

Most commercially available massively parallel sequencing platforms are based on the sequencing by synthesis (SBS) concept. This method involves the incorporation of nucleotides using enzymes and detection schemes, allowing the platform to collect data during the synthesis step. The enzymatic synthesis occurs on a model template, enabling high-throughput sequencing of millions of DNA fragments in parallel [66]. However, currently, there are several different NGS platforms using different sequencing technologies, the detailed discussion of which is beyond the scope of this article, as the main purpose of NGS is the analysis of single nucleotide variants (SNVs), also known as point substitutions or mutations and the detection of insertion or deletion events ("indels"), in which one or more nucleotides were added or deleted compared to the reference genome [66, 67].

Here, we will focus on the different technological features of the NGS that enable the detection of CNVs.

Detecting structural variation and CNVs using NGS data requires different bioinformatics algorithms than SNV calls.

Paired-end mapping – The method is based on comparing the average insert size between the actual sequenced reading pairs with the expected size based on a reference genome. However, this method cannot be applied to detect CNVs in regions of low complexity with segmental duplication. An advantage of the paired-end mapping method for reading pair is that it detects both CNVs and rearrangements (translocations and inversions) [68, 69].

Split read - The method uses end-of-pair sequencing reads, where only one read of the pair has a reliable mapping and the other fully or partially fails to map to the genome. Unmapped reads are a potential source of single base pair level breakpoints [69].

Read depth - They are based on the assumption that there is a correlation between the depth of coverage of a genomic region and the number of copies of the region. Read depth methods can be categorized into three classes: single sample, paired case/control samples, and a large sample population. The advantage of the read depth method is the ability to detect large CNVs and predict the actual number of copies; counterpoint, this method cannot detect the exact breakpoint or detect rearrangements [69, 70].

De novo Assembly – In this method, a contig/scaffold is first generated, which is then compared with the reference genome to discover structural variation.

Eukaryotic genomes contain a significant fraction of segmental repeats and duplications, which makes these methods less accurate and more complex, performing poorly in these complex regions [69].

Among these four strategies, RD-based methods are the most used, as they can theoretically infer CNVs of any size and deduce the corresponding copy numbers [71].

In practice, in routine CNV analysis, many clinical laboratories that are using NGS data to detect CNVs and generally use a combination of 2 or more methods.

It should be noted that, although in the last decade, the chromosomal microarray (CMA) has been considered the first-line clinical diagnostic test for individuals with NDDs or congenital anomalies, as it allows a diagnostic yield of approximately 15-20% [72], recent studies with WES, suggest a diagnostic yield of 30-50%, because with the new analysis methods it detects from single nucleotide mutations to CNV [61, 73, 74]. Therefore, the American College of Medical Genetics and Genomics (ACMG) now strongly recommend the application of Exome as a first line test in patients with NDDs [75].

CNVS INTERPRETATION AND CLASSIFICATION

Interpretation and classification of CNVs

The interpretation of CNVs by CMAs has been expanding the range of known mutations for many clinical genetic disorders. Initially, platforms and software for processing raw data by default exclude the most common CNVs, which are generally related to population variations, contributing to a more careful interpretation.

In the clinical interpretation of the treated data, most laboratories approach a flow of three steps aiming to characterize each detected CNV:

(1) initially seek to compare the detected CNVs with control data sets (genomic variations in unaffected individuals, often parents of patients, or individuals whose purpose of the microarray examination was otherwise biased) from their internal databases, national databases and international ones deposited in online databases,

thus eliminating frequent variations and also the less frequent, but still considered “common” in human beings [76].

Another form of comparison with control is the parental test to look for inherited small and rare CNVs, identifying particular family variants, not previously seen in patient cohorts or in a control group, in which case additional family samples such as from unaffected siblings are still possible [77]. If a CNV that is detected in the affected individual is also seen in an unaffected parent or sibling, it is less likely to be pathogenic. However, in the case of comparison of familial variants, it is important not to exclude these CNVs from the next stages of analysis, since there are reports of hereditary pathogenic CNVs that may be of variable expression/penetrance or may be a clue for investigating a recessive mutation in the remaining allele in the affected individual [78];

(2) Comparing the CNVs detected with data already found and interpreted by other affected individuals, both in their own databases and with databases of cohorts of affected individuals. An important detail at this stage is the relative size of CNVs, large CNVs (microscopically visible through classical cytogenetics, used for decades in clinical practice) are very often associated with phenotypic consequences as they encompass a significant number of genes, in addition to being the most characterized in cohorts affected individuals.

(3) finally, the last step consists of a thorough analysis of the remaining CNVs, considering whether it includes unique sequences, if it contains genes, regulatory regions or if it is composed of repetitive elements and/or pseudogenes.

When CNV covers genes, the content of the gene and its expressed product are considered in terms of their potential clinical associations, as well as data on their sensitivity to dosage. In the case of CNVs involving a partial gene (deleting or duplicating only one/a few exons, promoter regions, intronic regions), it is important to analyze whether this structural variation in the gene affects its product, both in the conformation of the protein or molecule, as in its amount. Both analyzes are based on the role of the gene and its predicted or characterized function in model organisms in studies published in the scientific literature. Finally, no conclusions about pathogenicity can be drawn until the variant is well characterized in more affected individuals [76].

In contrast to frequent or polymorphic CNVs, which occur in the population with a frequency above 1%, some CNVs are characterized as rare CNVs when they occur at a frequency below 1%. Both types of CNVs occur both in the normal population and in affected individuals, but pathogenic changes are usually rare or very rare [79].

For an adequate interpretation of CNV and to define the main clinical manifestations associated with microduplications of microdeletions found, the phenotype-genotype correlation is crucial. Therefore, it is essential for the requesting physician to describe the patient's clinical and dysmorphological phenotypes as accurately as possible [77, 80]. The biggest challenge is still to distinguish, through the analysis of CNVs, benign variations from those that may have clinical relevance [81].

Recently, in a joint effort between the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen), aiming to promote greater consistency and transparency in the interpretations of CNVs, a series of technical recommendations were published, in which they categorize the most relevant evidence for the classification of a CNV like genomic content, dose sensitivity predictions, predicted functional effect, clinical overlap with patients in the medical literature, evidence from case and control databases, and inheritance patterns for individual CNVs [82].

After nearly a decade from the first guidelines in 2011, the updated guidelines classify the CNVs into one of five categories: benign, likely benign, a variant of uncertain significance, likely pathogenic, or pathogenic.

Furthermore, to minimize the divergences between laboratories and subjectivities in the interpretations of the CNVs, a semi-quantitative scoring system based on points was developed, in which a relative weight was assigned to each evidence. For gains (duplications) and losses (deletions), separate scoring metrics were developed, due to the distinct properties and differences of the CNVs. Based on genomic content, dosage sensitivity predictions, predicted functional effects, clinical overlap with patients in the medical literature, evidence from case and control databases, and inheritance patterns for individual CNVs, the scoring system categorizes evidence as "very strong" if it receives ≥ 0.90 points, "strong" with 0.45 points, "moderate" with 0.30 points and "evidence" with ≤ 0.15 points. CNVs with an endpoint value ≥0.99 are interpreted as pathogenic, while CNVs with values between 0.90 and 0.98 are considered likely to be pathogenic. Variants of uncertain significance (VUS) have a wider range of scores corresponding to points between -0.89 and 0.89.

CNVs with refuted evidence with scores between -0.90 and -0.98, or ≤-0.99 are considered probably benign and benign, respectively.

DIAGNOSTIC RATE OF NDDs IN CNVs STUDIES

In order to review the diagnostic mean in cohort studies that evaluate CNVs as genetic etiologies of NDDs, we set a bibliographic review in the PUBMED/MEDILINE repository, in which the following terms were searched: ((((CMA [Title/Abstract]) OR (chromosomal Microarray [Title/Abstract])) OR (array-CGH [Title/Abstract])) and ((((((Neurodevelopmental disorders [Title/Abstract]) OR (intellectual disability [Title/Abstract])) OR (ID [Title/Abstract])) OR (autism spectrum disorder [Title/Abstract])) OR (ASD [Title/Abstract])) OR (developmental delay [Title/Abstract])) and (Cohort)), in publications between 2010 to 2023.

As a result, 189 indexed publications. After a thorough reading and evaluation, in which we excluded review publications, case descriptions, investigations of syndromes or specific genomic regions, duplicate results and a study without a clear description of the number of individuals in the sample and/or diagnostic rate, we had selected 84 studies published with the diagnostic rate of different ethnicities, populations, age groups and with the most diverse NDDs. Table 1 exemplifies the various studies of CNVs carried out in cohorts of affected individuals around the world and their diagnostic rates.

To assess the mean diagnostic rate of CMA in these studies, outliers in detection rate of pathogenic CNVs were identified (robust nonlinear regression [Q = 1%]), and excluded (1 study).

Table 1 - The 84 published studies that used CMA for diagnostic testing in the search for genetic etiology in a cohort of affected individuals and fulfilled the selection criteria.

Study	Cohort	CMA Platforms	Sample	Detection rate of pathogenic+ CNVs in %.
Ezugha et al, (2010)[83]	Children With NDDs or Epilepsy from Philadelphia (USA)	BACs and custom microarray with 1.8 million SNPs.	82	21
Shen et al, (2010)[84]	Patients with ASD from Boston (USA)	244k [A], Affymetrix 500k or v5.0 SNPs-array	848	7
Bartnik et al,	Patients with epilepsy and NDDs	V8.0 and V8.1 OLIGO (180 K)	102	23.5

Study	Cohort	CMA Platforms	Sample	Detection rate of pathogenic+ CNVs in %.
(2012)[85]	(Poland)	(custom-designed)		
Iourov et al, (2012)[86]	Children with NDDs (Russia)	slightly modified Constitutional BACs Chip®4.0 Perkin Elmer	54	28
Howell et al, (2013)[87]	Patients with neurological conditions (Australia)	HumanCytoSNP-12 v2.1C or 2.7 M arrays [B]	215	8.8
Shoukier et al, (2013)[88]	Children with unexplained DD/ID (Germany)	244K or SurePrint G3 Human [A]	342	13.2
Kashevarova et al, (2014)[89]	Russian patients with ID	44K e 60 K. [A]	79	28
Bartnik et al, (2014)[90]	Polish patients with NDDs	V8 OLIGO 180k (custom-designed)	256	19.8
Preiksaitiene et al, (2014)[91]	Lithuanians with NDDs	105k e 400k [A]	211	13.7
Roselló et al, (2014)[80]	Children with NDDs (Spain)	44K [A]	246	29.7
Asadollahi et al, (2014)(ASADOLLAHI et al., 2014)	Patients with NDDs of European origin	SNP Array 6.0, Cytogenetics 2.7 and CytoScan® HD [B]	714	11
Henderson et al (2014)[93]	CMAs from Johns Hopkins Hospital and Kennedy Krieger Institute (USA)	HumanQuad610 BeadChip or Human Omni 1M BeadChip [C]	1.780	12.7
Pereira et al, (2014)[94]	Patients with ID from central Brazil	CytoScan® HD [B]	83	21.7
Chong et al, (2014)[95]	Patients with NDDs (Hong Kong)	Custom-designed 44 K and ISCA designed 180 K oligonucleotide microarrays	105	19
Pfundt et al, (2015)[96]	North Americans with NDDs	CytoScan DX (Platform similar to CytoScan® HD)[B]	960	13.8
Coutton et al, (2015) [7]	French children with mild ID	180K [A]	66	21
Lay-son et al, (2015)[56]	Patients with NDDs from Santiago (Chile)	CytoScan® HD [B]	40	25
Al-Qattan et al, (2015)[97]	Cohort of DD/ID (Saudi Arabia - consanguinity rate of 56%)	SNP 6.0 array, Cyto-V2, CytoScan® HD [B]	183	21
Eriksson et al, (2015)[98]	Children with ASD (Sweden)	180-k custom design (Oxford Gene Technology Begbroke, Oxfordshire, UK)	162	8.6
Naseer et al, (2015)[99]	Epilepsy patients with and without ID (Saudi Arabia)	Sure print G3 Hmn CGH 2x 400K [A]	22	36.4
Xu et al, (2016)[100]	Patients with ASD/DD/ID recruited from Duke Children's Hospital (USA)	BlueGenome CytoChip v2 BAC array [C]; v6.0 SNP array and CytoScan® HD [B]	115	22.6
Ho et al (2016)[101]	CMA to a CLIA-licensed laboratory for etiological diagnosis (USA)	FSDX plus (N = 5487) and CytoScan® HD [B] (N = 1172)	5.487	9.2
Wolfe et al, (2016)[102]	Adults with ID presenting with comorbid psychiatric disorders	NimbleGen 135K	202	11

Study	Cohort	CMA Platforms	Sample	Detection rate of pathogenic+ CNVs in %.
	(England)			
Siu et al, (2016)[103]	Chinese ASD patients	NimbleGen CGX-135K oligonucleotide arrays	68	11.8
D'Arrigo et al, (2016)(D'ARRIGO et al., 2016)	Patients with NDDs from Milan (Italy)	4X180 [A] or ISCA 4x180 K Cytochip, BlueGenome	329	16
Quintela et al, (2017)[105]	Galician patients (Spain)	Cytogenetics Whole-Genome 2.7 M and CytoScan® HD [B]	573	13.6
Anazi et al, (2017)[106]	Individuals with ID (Saudi Arabia)	CytoScan® HD [B]	178	27
Faundes et al, (2017)[107]	Chilean patients with NDDs	8×60K [A], ISCA v2 [A]	224	19.2
Berg et al (2017)[108]	Children with early-life epilepsies (USA)	Does not cite	188	17
Di Gregorio et al, (2017)[109]	Patients with DD/ID (Italy)	60K [A]	1015	16
Heide et al, (2017)[110]	Patients with corpus callosum abnormality and ID (France)	370CNV-Quad, cytoSNP-12, HumanOmniExpress-24 Illumina	149	13.3
Hnoonual et al, (2017)[111]	Thai patients with ASD	Infinium CytoSNP-850K v1.1 Beadchip [C]	114	6.1
Mak et al, (2017)[112]	Chinese children with ASD	NimbleGen-CGX-135k or Agilent-CGX 60k oligonucleotide array	258	3.5
Lintas et al, (2017)[113]	19 Italians multiplex families with ASD	180K [A]	41	36.6
Peycheva et al, (2018)(133)	Bulgarian patients with epilepsy and ID	SurePrint G3 Unrestricted CGH ISCA v2, 4 × 180 K [A]	92	15.2
Xu et al, (2018)[114]	Han Chinese children with NDDs	CytoScan® HD [B]	434	13.6
Thygesen et al, (2018)[115]	Adults with ID and comorbid psychiatric disorders (Catalonia, Spain; Leuven, Belgium; and England, UK)	400 K [A],CytoSure ISC[A] oligoarray set, NimbleGen 135K and Cytoscan750K [B]	599	13
Papuc et al, (2018)[116]	Patients with epileptic encephalopathies and NDDs from Zurich (Switzerland)	CytoScan® HD [B]	63	10
Fan et al, (2018)[117]	Chinese patients with DD/ID	CytoScan® HD and CytoScan 750K [B]	710	28
Lee et al, (2018)[118]	Patients with DD/ID from Seoul (South Korea)	1×244K, 4×180K, or 8×60K [A]	649	16.9
Napoli et al, (2018)[119]	Individuals with essential ASD (Italy)	4 × 180K [A]	133	9
Chan et al, (2018)[120]	Children with ID from Hong Kong	PerkinElmer CGXTM v2 60K arrays	138	11.6
Maini et al, (2018)[121]	Children with NDDs (Italy)	8x60K oligochips	339	20.6
Homma et al, (2018)[122]	Children with NNDs, dysmorphisms and short stature from São Paulo (Brazil)	whole-genome 180 K platform, SNP array CytoSNP-850K BeadChip or CytoScan HD [A]	229	13

Study	Cohort	CMA Platforms	Sample	Detection rate of pathogenic+ CNVs in %.
Sys et al, (2018)[123]	Children diagnosed with ASD (Belgium)	60K [A]	311	8.9
Viñas-Jornet et al, (2018)[124]	Adult with mild-moderate ID and co-morbid psychiatric disease (país?)	400K [A]	86	12.8
Munnich et al, (2019)[125]	Atypical and/or syndromic ASD patients with moderate to severe ID (France)	60K [A]	388	8.8
Coppola et al, (2019)[126]	European patients with epilepsy (Paises?)	Several arrays [A] and HumanCNV [C]	1097	12.9
Chaves et al, (2019)[53]	Patients with NDDs from the south of Brazil	CytoScan® HD and CytoScan 750K [B]	420	18
Han et al, (2019)[127]	Patients with NDDs (South Korea)	SurePrint G3 Human CGH Microarray 8 × 60 K kit	65	38.4
Jang et al, (2019)[128]	Patients with NDDs (South Korea)	SurePrint G3 Human CGH Microarray 8 X 60K	617	19.8
Lindstrand et al, (2019)[129]	Patients with NDDs (Sweden)	4 × 180K custom oligonucleotide microarray	100	12
Wayhelova et al, (2019)(WAYHELOVA et al., 2019)	Patients with NDDs (Czech Republic)	SurePrint G3 Human and Cytosure ISCA 4X180K UPD array	542	17.7
Micleaa et al, (201 9)[131]	Romanian patients with obesity and NDDs	Infinium OmniExpress-24 BeadChip array	36	22.2
Capkova et al, (2019)[132]	Patients with ASD (Czech Republic)	Cytoscan HD [A] or CytoSNP-12 [C]	92	8.1
Monteiro et al, (2019)[133]	Portuguese children and adolescents with ASD	4 × 180K [A]	253	11.5
Pinheiro et al, (2020)[134]	Patients with NDDs (Portugal)	4x180K [A]	215	23.3
Calderoni et al, (2020)[135]	Females with idiopathic ASD from Italy	8×60K [A]	90	22.2
Mohamed et al, (2020)[136]	Unrelated patients with NDDs (Saudi Arabia)	CytoScan® HD [B]	169	11
Espeche et al, (2020)[137]	Patients with DI (Argentina)	ISCA v2 8 × 60 K [A]	133	12
Farooqi et al, (2020)[138]	Reinterpretation of CMA from UT Southwestern institutional	customized CMA	998	10
Yuan et al, (2021)[139]	Pediatric patients with NDDs (China)	CytoScan® HD [B], 244k/180k/60k [A] or custom-designed chip, SNP chip [C]	9.782	21.37
Milone et al, (2021)[140]	Patients with NDDs (Italy)	44K, 60K, 180K [A] and 6.0 Chip [B]	593	16
Zacher et al, (2021)[141]	Adult/elderly with NDDs and epilepsy (Germany)	CytoSNP-850K v 1.1. and 1.2	150	16
Yang et al, (2021)[142]	Korean children with DD/ID (South Korea)	CytoScan 750K [A]	308	18.5

Study	Cohort	CMA Platforms	Sample	Detection rate of pathogenic+ CNVs in %.
Moirangthem et al, (2021)[143]	Patients with disorders with overgrowth and ID from North India	CytoScan® HD [B]	18	11.1
Volo et al, (2021)[144]	Italian cohort of patients with ID, MCA and ASD	BACs array-CGH and a-CGH with oligonucleotides	343	17.8
Dai et al, (2021)[145]	Chinese patients with ID	CytoScan® HD [B]	105	46.67*
Yang et al, (2021)[146]	Patients with epilepsy (China)	CytoScan 750K [B]	102	16.7
Costa et al, (2021)[147]	Patients with ASD (Brazil)	180K [A]	122	4.1
Perovic et al, (2022)[148]	Serbian patients with CAs and/or NDDs	SurePrint G3 Human CGH Microarray 8 × 60K, and SurePrint G3 Human CGH +SNP Microarray 4 × 180K [A]	430	16.3
Chehbani et al, (2022)[149]	Children with ASD (Tunisia)	CGH SurePrint G3 Microarray 4 × 180 K Kit [A]	98	11.2
Nassir et al, (2022)[150]	Patients with NDDs (United Arab Emirates)	Does not cite	98	5.3
Liu et al, (2022)[151]	Han Chinese patients with NDDs	SNP Array 6.0 or CytoScan® HD [B]	402	20.9
Lee et al, (2022)[152]	ASD patients (Taiwan)	GeneChip Genome-Wide Human SNP array 6.0 [B]	80	33.8
Leite, et al, (2022)[153]	Patients with NDDs from Central Brazil	CytoScan® HD [B]	83	32.5
Su et al, (2022)[154]	Retrospective study: Patients with NDDs from Colorado (USA)	CytoChip-180K Oligo and CytoSNP-850K [C]	4040	12.9
Krepischi et al, (2022)[155]	Brazilian children with NDDs	Infinium CytoSNP 850K BeadChip [C] and CytoScan 750K Array [B]	5.788	19.5
Tolezano et al, (2022)[156]	patients with NDD and microcephaly associated or not with other congenital anomalies (Brazilian Patients)	60 K or 180 K whole-genome	185	18.4
Streață et al,(2022)[157]	Romanian patients with syndromic or non-syndromicglobal developmental delay /ID,	SurePrint G3 v2 8 × 60K (141 patients), 4 × 180K (37 patients) [A], and CytoSure ISCA V2 CGH 8 × 60K array (193 patients)	202	22.3
Lengyel et al. (2022)[158]	Hungarian patients with NDDs and/or congenital anomalies	NimbleGen Array (CGX 1.4 M) (30 patients; qChip Post (60 K; 5 patients) and 180 K oligo-array (18 patients)[A], CytoScan Optima (300 K; 5 patients), CytoScan 750 K (17 patients), and CytoScan HD (3 patients) [B]	78	37.2
Sheth et al, (2023)[159]	patient-parent trios of Indian origin diagnosed with ASD	CytoScan Optima 300K [B]	110	2.9
Sandoval-Talamantes et al,(2023)[160]	individuals over 3 years of age, who met DSM-5 diagnostic criteria for ASD (Spain)	SurePrint G3 CGH 8 × 60 K [A]	212	13
Akter et al, (2023)[161]	DDNs patients with ASD (India)	Global Screening Array-24 + v1.0 [C]	212	12.3

(NDDs) Neurodevelopmental disorders, (DD) developmental delay, (ID) Intellectual disability, (CAs) Congenital anomalies, (ASD) Autism spectrum disorder, (CMA) Chromosomal microarrays. (A) Agilent Technologies, Santa Clara, CA, USA, (B) Affymetrix, Santa Clara, CA, USA, (C) Illumina, San Diego, CA, USA. +Detection rate including pathogenic and potentially pathogenic CNVs. * Outliers.

Our review involving 84 cohort studies of affected individuals with NDDs investigated using CMA technology (of which 1 were excluded because they were considered outliers), averaged 16.8% (95% CI: 15.07-18.54) detection rate of pathogenic CNVs. Figure 03, compared to the detection rate of pathogenic CNVs before and after excluding outliers.

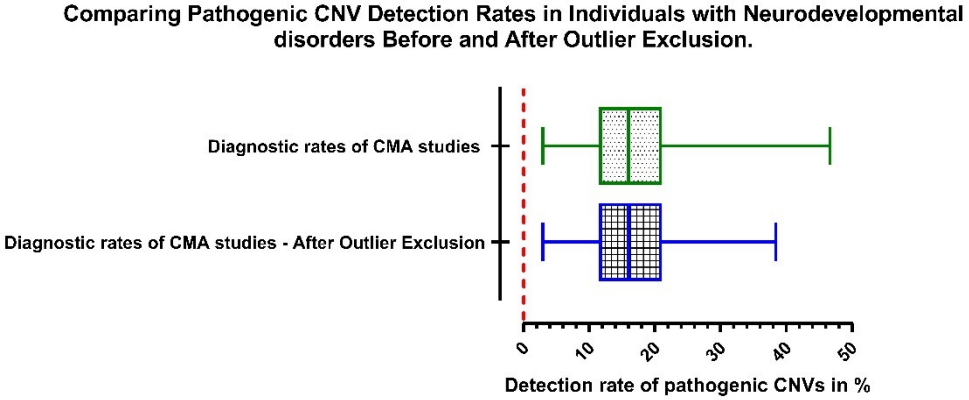
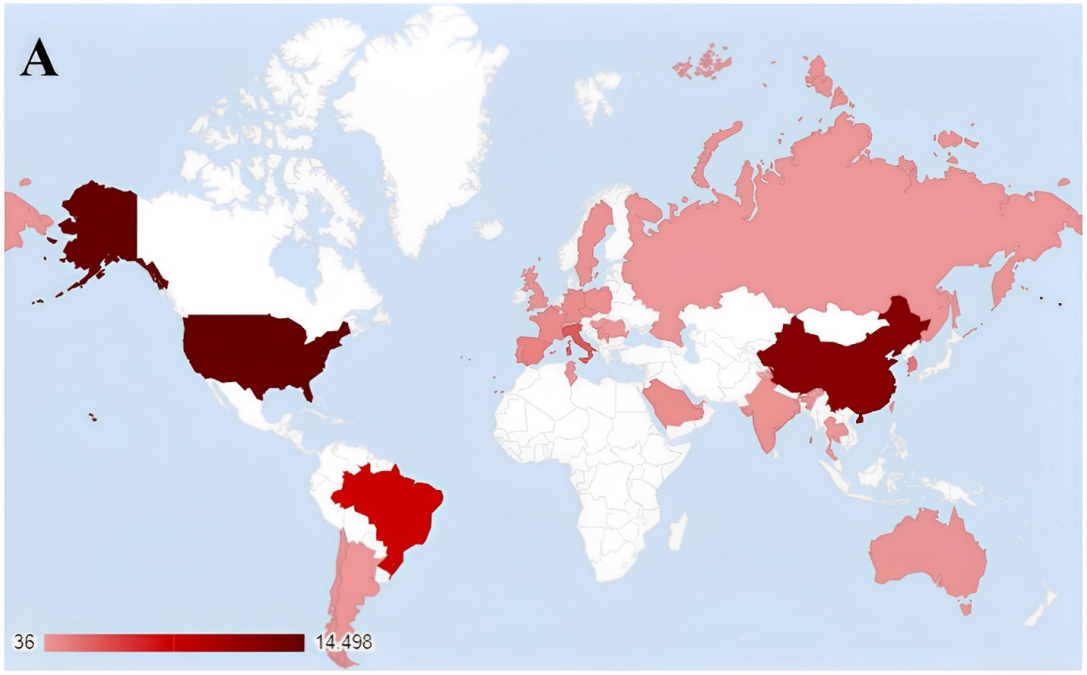


Figure 03: Comparing Pathogenic CNV Detection Rates in Individuals with NDDs Before and After Outlier Exclusion. This figure illustrates the results of our comprehensive review involving 84 cohort studies of individuals affected by NDDs, where CMA technology was employed for investigation. After excluding 1 outlier studies, the averaged pathogenic CNV detection rate was found to be 16.8% (95% CI: 15.07-18.54). This figure visually presents the contrast between pathogenic CNV detection rates before and after outlier removal

In figure 04A we can see the comparison of the size (by country of origin of the cohort) and of the different cohorts sampled in diagnostic studies of pathogenic CNVs through CMAs in individuals with DNNs.

Figure 04

Comparison of sample size in number of individuals by cohort country in studies of pathogenic CNVs.



Average detection rate of pathogenic CNVs in % per sampled cohort (Countries).

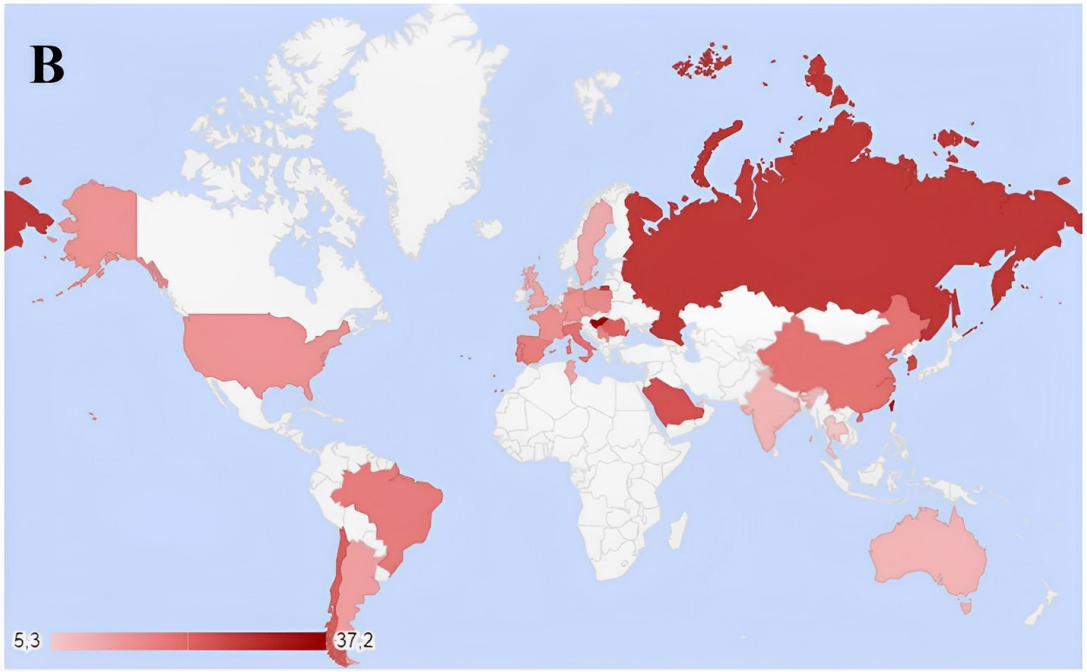


Figure 04A: Cohort Size and Geographic Distribution in Pathogenic CNV Diagnostic Studies. This illustrates cohort size differences by country in diagnostic studies of pathogenic CNVs using CMA for individuals with DDNs. Developed countries like the USA and China have larger cohorts, as do Europe and the northern hemisphere, while less developed regions like South America and Africa have

smaller cohorts. Latin America primarily uses classic karyotyping for genetic testing. **Figure 04B**: Average Pathogenic CNV Detection Rate Across Sampled Cohorts. This figure displays the average pathogenic CNV detection rate, as a percentage per cohort (countries), after excluding outliers from the diagnostic analysis. Most studies fall within a 14% to 25% diagnostic range, indicating consistent detection rates across different cohorts.

As we can see (figure 04A), the most economically developed countries, such as the USA and China, are the ones with the largest cohorts of individuals with NDDs studied for genetic diagnosis, through the CMA, as well as there are significant samples in Europe and the northern hemisphere. In contrast to a low sample of economically less developed countries, such as countries in the southern hemisphere, as in South America and the African continent. In Latin America classic karyotyping is still the predominant genetic test in clinical practice [53].

In figure 04B we can see the average detection rate of pathogenic CNVs in % per sampled cohort (Countries), excluding outliers for the diagnostic rate. We can note that most studies are within a diagnostic range of 14 to 25%.

The variance in the diagnostic rate of pathogenic CNVs, detected through the CMA, occurs for different platforms used, the amount and spacing between probes, the coverage of the entire genome, as well as the analytical power of each platform, may compromise the size and location of diagnostically important CNVs.

Another important factor in the detection rate of pathogenic CNVs was the knowledge added in the last decade about the pathogenic potential of new CNVs, as new studies were being developed, new knowledge about the genotype x phenotype relationship became clearer, thus increasing the efficiencies in the interpretation of detected CNVs.

We also cannot fail to highlight that the diagnostic strategy, whether using CMA as a first-line test in the diagnosis of NDDs, or as a complementary test, after screening the classic cytogenetic test, such as karyotype or Fragile-X, thus influencing the final efficiency rate of the CMA. Many of these cohort studies, due to their inclusion of multi-center diagnostic and research centers, employ mixed and/or diverse genetic screening diagnostic strategies, thereby becoming a limitation for diagnostic rate comparison.

CONCLUSIONS AND PERSPECTIVES

In recent years, there has been a significant advancement in the field of genomic analysis, particularly in the use of CMA technologies for the detection and evaluation of structural variations in the genome. This has led to an increase in cohort studies of neuroaffected individuals, from diverse populations, aimed at understanding the underlying pathophysiological pathways affected by CNVs.

Despite the growing availability and use of CMA as a first-line diagnostic test in the investigation of neurodevelopmental disorders, the average diagnostic rate of CNVs remains at 16.8% in cohorts of patients from different populations and ethnicities. However, it is worth noting that most of these studies have been conducted in the Northern Hemisphere, while in developing countries, such as those predominant in the Southern Hemisphere, classical cytogenetics is still used as a first-line diagnostic test in the investigation of neurodevelopmental disorders.

Therefore, there is a need for further research on CNVs and their implications in diverse populations, particularly in developing countries, in order to increase the efficiency and accuracy of diagnosis and management of neurodevelopmental disorders. Additionally, ongoing research and advancements in technology are expected to improve the detection and interpretation of CNVs, providing valuable insights into the genotype-phenotype relationships and pathophysiological mechanisms underlying these disorders.

LIST OF ABBREVIATIONS

NDDs – Neurodevelopmental disorders
ID - Intellectual disability
ASD - Autism spectrum disorder
CMA - Chromosomal microarrays
CNV - Copy number variants
NGS - Next Generation Sequencing
ACs - Congenital Anomalies
DGV - Database of Genomic Variant
OMIM - Online Mendelian Inheritance in Man
DECIPHER - Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources
VOUS – Variant(s) of uncertain clinical significance
DD- Development delay
CAs - Congenital anomalys

IUGR - Intrauterine growth restriction
Mbp - Mega base pairs
LCSH - Long continuous stretches of homozygosity

DECLARATIONS

Consent for publication

Not applicable

Availability of data and material

Not applicable.

Competing interests

Not applicable.

Funding

TFC were supported with a scholarship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Brazil. The funding body provided financial support but had no other role in the design of the study and collection, analysis, and interpretation of data or in writing of the manuscript.

Authors' contributions

TFC conducted the entire study as part of the doctoral work, under the guidance of YCM and co-supervision of AFM. TFC: They were responsible for the study design, data collection, analysis and interpretation of results, analysis of statistical data. All wrote and reviewed the manuscript.

All authors read and approved the final manuscript.

Acknowledgements

We thank CAPES for their financial support with a post-graduate grant to TFC.

REFERENCES

1. Thapar A, Cooper M, Rutter M. Neurodevelopmental disorders. The Lancet Psychiatry. 2017;4:339–46.
2. Schaaf CP, Betancur C, Yuen RKC, Parr JR, Skuse DH, Gallagher L, et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. Nat Rev Genet. 2020;21:367–76.
3. Thibaut F. Psychiatric disorders: neurodevelopmental disorders, neurodegenerative disorders, or both? Dialogues Clin Neurosci. 2018;20:251.
4. International Classification of Functioning, Disability and Health World Health Organization Geneva ICF ii WHO Library Cataloguing-in-Publication Data International classification of functioning, disability and health : ICF. 2001.
5. Cappuccio G, Vitiello F, Casertano A, Fontana P, Genesio R, Bruzzese D, et al. New insights in the interpretation of array-CGH: autism spectrum disorder and positive family history for intellectual disability predict the detection of pathogenic variants. Ital J Pediatr. 2016;42:39.
6. Abou Jamra R, Wohlfart S, Zweier M, Uebe S, Priebe L, Ekici A, et al. Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. Eur J Hum Genet. 2011;19:1161–6.
7. Coutton C, Dieterich K, Satre V, Vieville G, Amblard F, David M, et al. Array-CGH in children with mild intellectual disability: a population-based study. Eur J Pediatr. 2015;174:75–83.
8. Mandy W, Lai M-C. Annual Research Review: The role of the environment in the developmental psychopathology of autism spectrum condition. J Child Psychol Psychiatry. 2016;57:271–92.
9. Chiurazzi P, Pirozzi F. Advances in understanding – genetic basis of intellectual disability. F1000Research. 2016;5:599.
10. Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med. 2014;16:176–82.
11. Kvarnung M, Nilsson D, Lindstrand A, Korenke GC, Chiang SCC, Blennow E, et al. A novel intellectual disability syndrome caused by GPI anchor deficiency due to homozygous mutations in PIGT. J Med Genet. 2013;50:521–8.
12. Stone WL, Los E. Fragile X Syndrome. StatPearls Publishing; 2017.

13. Miclea D, Peca L, Cuzmici Z, Pop IV. Genetic testing in patients with global developmental delay/intellectual disabilities. A review. Clujul Med. 2015;88:288.
14. Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ. 2020;69:1–12.
15. Cooper R. Diagnostic and Statistical Manual of Mental Disorders (DSM). Knowl Organ. 2017. https://doi.org/10.1016/j.psychres.2011.06.006.
16. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L, et al. FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. Cell. 2015;162:375–90.
17. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev. 2012;22:229–37.
18. Betancur C, Coleman M. Etiological heterogeneity in autism spectrum disorders : role of rare variants The Neuroscience of Autism Spectrum Disorders. 2014.
19. Chen L, Zhou W, Zhang L, Zhang F. Genome architecture and its roles in human copy number variation. Genomics Inform. 2014;12:136–44.
20. EUROCAT. http://www.eurocat-network.eu/accessprevalencedata/keypublichealthindicators. Accessed 22 Nov 2017.
21. Xu J, Murphy SL, Kochanek KD, Bastian BA. National Vital Statistics Reports, Volume 64, Number 2, 02/16/2016. 2016.
22. Mazzu-Nascimento T, Melo DG, Morbioli GG, Carrilho E, Vianna FSL, Da Silva AA, et al. Teratogens: A public health issue – A Brazilian overview. Genet Mol Biol. 2017;40:387–97.
23. Frías JL, Gilbert-Barness E. Human Teratogens: Current Controversies. Adv Pediatr. 2008;55:171–211.
24. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7:85–97.
25. Bridges CB. THE BAR "GENE" A DUPLICATION. Science (80-). 1936;83:210–1.
26. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007;8:639–46.

27. Savatt JM, Myers SM. Genetic Testing in Neurodevelopmental Disorders. Front Pediatr. 2021;9:526779.
28. Uddin M, Thiruvahindrapuram B, Walker S, Wang Z, Hu P, Lamoureux S, et al. A high-resolution copy-number variation resource for clinical and population genetics. Genetics in Medicine. 2015;17:747–52.
29. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464:704–12.
30. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet. 2015;16:172–83.
31. Cook EH, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. Nature. 2008;455 October:919–23.
32. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. Expert Rev Mol Med. 2010;12:e8.
33. Carvalho CMBB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nature Publishing Group; 2016.
34. Ricard G, Molina J, Chrast J, Gu W, Gheldof N, Pradervand S, et al. Phenotypic Consequences of Copy Number Variation: Insights from Smith-Magenis and Potocki-Lupski Syndrome Mouse Models. PLoS Biol. 2010;8:e1000543.
35. Gheldof N, Witwicki RM, Migliavacca E, Leleu M, Didelot G, Harewood L, et al. Structural Variation-Associated Expression Changes Are Paralleled by Chromatin Architecture Modifications. PLoS One. 2013;8:e79973.
36. Harel T, Lupski JR. Genomic disorders 20 years on—mechanisms for clinical manifestations. Clinical Genetics. 2018;93:439–49.
37. Gu S, Szafranski P, Akdemir ZC, Yuan B, Cooper ML, Magriñá MA, et al. Mechanisms for Complex Chromosomal Insertions. PLoS Genet. 2016;12:e1006446.
38. Baskin B, Stavropoulos DJ, Rebeiro PA, Orr J, Li M, Steele L, et al. Complex genomic rearrangements in the dystrophin gene due to replication-based mechanisms. Mol Genet genomic Med. 2014;2:539–47.
39. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.
40. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends Genet. 2002;18:74–82.

41. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet. 2006;38:1038–42.
42. Weise A, Mrasek K, Klein E, Mulatinho M, Llerena JC, Hardekopf D, et al. Microdeletion and microduplication syndromes. J Histochem Cytochem. 2012;60:346–58.
43. Lupski JR. Genomic rearrangements and sporadic disease. Nat Genet. 2007;39:S43–7.
44. Stankiewicz P, Pursley AN, Cheung SW. Challenges in clinical interpretation of microduplications detected by array CGH analysis. American Journal of Medical Genetics, Part A. 2010;152:1089–100.
45. Reid DA, Rothenberg E. Repair of chromosomal breaks by NHEJ. Oncotarget. 2015;6:15730–1.
46. Pannunzio NR, Li S, Watanabe G, Lieber MR. Non-homologous end joining often uses microhomology: implications for alternative end joining. DNA Repair (Amst). 2014;17:74–80.
47. Liu P, Carvalho CM, Hastings P, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. Curr Opin Genet Dev. 2012;22:211–20.
48. Manning M, Hudgins L. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet Med. 2010;12:742–5.
49. Sagoo GS, Mohammed S, Barton G, Norbury G, Ahn JW, Ogilvie CM, et al. Cost Effectiveness of Using Array-CGH for Diagnosing Learning Disability. Appl Health Econ Health Policy. 2015;13:421–32.
50. Oostlander A, Meijer G, Ylstra B. Microarray-based comparative genomic hybridization and its applications in human genetics. Clin Genet. 2004;66:488–95.
51. Crotwell PL, Hoyme HE. Advances in whole-genome genetic testing: from chromosomes to microarrays. Curr Probl Pediatr Adolesc Health Care. 2012;42:47–73.
52. Colaianni V, Mazzei R, Cavallaro S. Copy number variations and stroke. Neurological Sciences. 2016;37:1895–904.
53. Chaves TF, Baretto N, Oliveira LF de, Ocampos M, Barbato IT, Anselmi M, et al. Copy Number Variations in a Cohort of 420 Individuals with Neurodevelopmental Disorders From the South of Brazil. Sci Rep. 2019;9:1–20.

54. Wapner RJ, Martin CL, Levy B, Ballif BC, Eng CM, Zachary JM, et al. Chromosomal Microarray versus Karyotyping for Prenatal Diagnosis. N Engl J Med. 2012;367:2175.
55. Karampetsou E, Morrogh D, Chitty L. Microarray Technology for the Diagnosis of Fetal Chromosomal Aberrations: Which Platform Should We Use? J Clin Med. 2014;3:663–78.
56. Lay-Son G, Espinoza K, Vial C, Rivera JC, Guzm??n ML, Repetto GM. Chromosomal microarrays testing in children with developmental disabilities and congenital anomalies. J Pediatr (Rio J). 2015;91:189–95.
57. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genet Med. 2013;15:901–9.
58. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am J Hum Genet. 2010;86:749–64.
59. Soehn AS, Rattay TW, Beck-Wödl S, Schäferhoff K, Monk D, Döbler-Neumann M, et al. Uniparental disomy of chromosome 16 unmasks recessive mutations of FA2H /SPG35 in 4 families. Neurology. 2016;87:186–91.
60. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011;70:863–85.
61. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med. 2018;20:1122–30.
62. Sobreira N, Schiettecatte F, Boehm C, Valle D, Hamosh A. New Tools for Mendelian Disease Gene Identification: PhenoDB Variant Analysis Module; and GeneMatcher, a Web-Based Tool for Linking Investigators with an Interest in the Same Gene. Hum Mutat. 2015;36:425–31.
63. Behjati S, Tarpey PS. What is next generation sequencing? Arch Dis Child Educ Pract Ed. 2013;98:236.
64. Lindy AS, Stosser MB, Butler E, Downtain-Pickersgill C, Shanmugham A, Retterer K, et al. Diagnostic outcomes for genetic testing of 70 genes in 8565

patients with epilepsy and neurodevelopmental disorders. Epilepsia. 2018;59:1062–71.
65. Cogliati F, Giorgini V, Masciadri M, Bonati MT, Marchi M, Cracco I, et al. Pathogenic Variants in STXBP1 and in Genes for GABAa Receptor Subunities Cause Atypical Rett/Rett-like Phenotypes. Int J Mol Sci. 2019;20.
66. Nishio S, Usami S. The Clinical Next-Generation Sequencing Database: A Tool for the Unified Management of Clinical Information and Genetic Variants to Accelerate Variant Pathogenicity Classification. Hum Mutat. 2017. https://doi.org/10.1002/humu.23160.
67. Liu Z, Zhu L, Roberts R, Tong W. Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We? Trends in Genetics. 2019;35:852–67.
68. Zhong Y, Xu F, Wu J, Schubert J, Li MM. Application of Next Generation Sequencing in Laboratory Medicine. Ann Lab Med. 2021;41:25.
69. Pirooznia M, Goes F, Zandi PP. Whole-genome CNV analysis: Advances in computational approaches. Front Genet. 2015;6 MAR:138.
70. Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. Genet Med. 2014;16:176–82.
71. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: Features and perspectives. BMC Bioinformatics. 2013;14:1–16.
72. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am J Hum Genet. 2010;86:749–64.
73. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet Med. 2019;21:2413–21.
74. Álvarez-Mora MI, Sánchez A, Rodríguez-Revenga L, Corominas J, Rabionet R, Puig S, et al. Diagnostic yield of next-generation sequencing in 87 families with neurodevelopmental disorders. Orphanet J Rare Dis. 2022;17:60.

75. Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2021;23:2029–37.
76. de Leeuw N, Dijkhuizen T, Hehir-Kwa JY, Carter NP, Feuk L, Firth H V, et al. Diagnostic interpretation of array data using public databases and internet sources. Hum Mutat. 2012;33:930–40.
77. Vermeesch JR, Brady PD, Sanlaville D, Kok K, Hastings RJ. Genome-wide arrays: Quality criteria and platforms to be used in routine diagnostics. Hum Mutat. 2012;33:906–15.
78. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011;13:777–84.
79. Valsesia A, Macé A, Jacquemont S, Beckmann JS, Kutalik Z. The growing importance of CNVs: New insights for detection and clinical interpretation. Frontiers in Genetics. 2013;4 MAY:92.
80. Roselló M, Martínez F, Monfort S, Mayo S, Oltra S, Orellana C. Phenotype profiling of patients with intellectual disability and copy number variations. Eur J Paediatr Neurol. 2014;18:558–66.
81. Palmer E, Speirs H, Taylor PJ, Mullan G, Turner G, Einfeld S, et al. Changing interpretation of chromosomal microarray over time in a community cohort with intellectual disability. Am J Med Genet Part A. 2014;164:377–85.
82. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22:245–57.
83. Ezugha H, Anderson CE, Marks HG, Khurana D, Legido A, Valencia I. Microarray Analysis in Children With Developmental Disorder or Epilepsy. Pediatr Neurol. 2010;43:391–4.
84. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, et al. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics. 2010;125.

85. Bartnik M, Szczepanik E, Derwińska K, Wiśniowiecka-Kowalnik B, Gambin T, Sykulski M, et al. Application of array comparative genomic hybridization in 102 patients with epilepsy and additional neurodevelopmental disorders. Am J Med Genet Part B Neuropsychiatr Genet. 2012;159B:760–71.
86. Iourov IY, Vorsanova SG, Kurinnaia OS, Zelenova MA, Silvanovich AP, Yurov YB. Molecular karyotyping by array CGH in a Russian cohort of children with intellectual disability, autism, epilepsy and congenital anomalies. Mol Cytogenet. 2012;5.
87. Howell KB, Kornberg AJ, Harvey AS, Ryan MM, Mackay MT, Freeman JL, et al. High resolution chromosomal microarray in undiagnosed neurological disorders. J Paediatr Child Health. 2013;49:716–24.
88. Shoukier M, Klein N, Auber B, Wickert J, Schröder J, Zoll B, et al. Array CGH in patients with developmental delay or intellectual disability: are there phenotypic clues to pathogenic copy number variants? Clin Genet. 2013;83:53–65.
89. Kashevarova AA, Nazarenko LP, Skryabin NA, Salyukova OA, Chechetkina NN, Tolmacheva EN, et al. Array CGH analysis of a cohort of Russian patients with intellectual disability. Gene. 2014;536:145–50.
90. Bartnik M, Nowakowska B, Derwińska K, Wiśniowiecka-Kowalnik B, Kedzior M, Bernaciak J, et al. Application of array comparative genomic hybridization in 256 patients with developmental delay or intellectual disability. J Appl Genet. 2014;55:125–44.
91. Preiksaitiene E, Molytė A, Kasnauskiene J, Ciuladaite Z, Utkus A, Patsalis PC, et al. Considering specific clinical features as evidence of pathogenic copy number variants. J Appl Genet. 2014;55:189–96.
92. Asadollahi R, Oneda B, Joset P, Azzarello-Burri S, Bartholdi D, Steindl K, et al. Original article: The clinical significance of small copy number variants in neurodevelopmental disorders. J Med Genet. 2014;51:677.
93. Henderson LB, Applegate CD, Wohler E, Sheridan MB, Hoover-Fong J, Batista DAS. The impact of chromosomal microarray on clinical management: a retrospective analysis. Genet Med. 2014;16:657–64.
94. Pereira RR, Pinto IP, Minasi LB, De Melo AV, Da Cruz E Cunha DM, Cruz AS, et al. Screening for intellectual disability using high-resolution CMA technology in a retrospective cohort from Central Brazil. PLoS One. 2014;9:e103117.

95. Chong WWS, Lo IFM, Lam STS, Wang CC, Luk HM, Leung TY, et al. Performance of chromosomal microarray for patients with intellectual disabilities/developmental delay, autism, and multiple congenital anomalies in a Chinese cohort. Mol Cytogenet. 2014;7.
96. Pfundt R, Kwiatkowski K, Roter A, Shukla A, Thorland E, Hockett R, et al. Clinical performance of the CytoScan Dx Assay in diagnosing developmental delay/intellectual disability. Genet Med. 18.
97. Al-Qattan SM, Wakil SM, Anazi S, Alazami AM, Patel N, Shaheen R, et al. The clinical utility of molecular karyotyping for neurocognitive phenotypes in a consanguineous population. Genet Med. 2015;17:719–25.
98. Eriksson MA, Liedén A, Westerlund J, Bremer A, Wincent J, Sahlin E, et al. Rare copy number variants are common in young children with autism spectrum disorder. Acta Paediatr. 2015;104:610–8.
99. Naseer MI, Faheem M, Chaudhary AG, Kumosani TA, Al-Quaiti MM, Jan MM, et al. Genome wide analysis of novel copy number variations duplications/deletions of different epileptic patients in Saudi Arabia. BMC Genomics. 2015;16 Suppl 1 Suppl 1.
100. Xu Q, Goldstein J, Wang P, Gadi IK, Labreche H, Rehder C, et al. Chromosomal Microarray Analysis in Clinical Evaluation of Neurodevelopmental Disorders-Reporting a Novel Deletion of SETDB1 and Illustration of Counseling Challenge. Pediatr Res. 2016;80:371.
101. Ho KS, Twede H, Vanzo R, Harward E, Hensel CH, Martin MM, et al. Clinical Performance of an Ultrahigh Resolution Chromosomal Microarray Optimized for Neurodevelopmental Disorders. Biomed Res Int. 2016;2016.
102. Wolfe K, Strydom A, Morrogh D, Carter J, Cutajar P, Eyeoyibo M, et al. Chromosomal microarray testing in adults with intellectual disability presenting with comorbid psychiatric disorders. Eur J Hum Genet. 2017;25:66.
103. Siu W-K, Lam C-W, Mak CM, Lau ET-K, Tang MH-Y, Tang W-F, et al. Diagnostic yield of array CGH in patients with autism spectrum disorder in Hong Kong. Clin Transl Med. 2016;5.
104. D'Arrigo S, Gavazzi F, Alfei E, Zuffardi O, Montomoli C, Corso B, et al. The Diagnostic Yield of Array Comparative Genomic Hybridization Is High Regardless of Severity of Intellectual Disability/Developmental Delay in Children. J Child Neurol. 2016;31:691–9.

105. Quintela I, Eirís J, Gómez-Lado C, Pérez-Gay L, Dacruz D, Cruz R, et al. Copy number variation analysis of patients with intellectual disability from North-West Spain. Gene. 2017;626:189–99.
106. Anazi S, Maddirevula S, Faqeih E, Alsedairy H, Alzahrani F, Shamseldin HE, et al. Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol Psychiatry. 2017;22:615–24.
107. Faundes V, Santa María L, Morales P, Curotto B, Alliende MA, Faundes V, et al. Microarrays in 236 patients with neurodevelopmental disorders and congenital abnormalities. Rev Med Chil. 2017;145:854–61.
108. Berg AT, Coryell J, Saneto RP, Grinspan ZM, Alexander JJ, Kekis M, et al. Early-Life Epilepsies and the Emerging Role of Genetic Testing. JAMA Pediatr. 2017;171:863.
109. Di Gregorio E, Riberi E, Belligni EF, Biamino E, Spielmann M, Ala U, et al. Copy number variants analysis in a cohort of isolated and syndromic developmental delay/intellectual disability reveals novel genomic disorders, position effects and candidate disease genes. Clin Genet. 2017;92:415–22.
110. Heide S, Keren B, Billette de Villemeur T, Chantot-Bastaraud S, Depienne C, Nava C, et al. Copy Number Variations Found in Patients with a Corpus Callosum Abnormality and Intellectual Disability. J Pediatr. 2017;185:160-166.e1.
111. Hnoonual A, Thammachote W, Tim-Aroon T, Rojnueangnit K, Hansakunachai T, Sombuntham T, et al. Chromosomal microarray analysis in a cohort of underrepresented population identifies SERINC2 as a novel candidate gene for autism spectrum disorder. Sci Rep. 2017;7.
112. Mak ASL, Chiu ATG, Leung GKC, Mak CCY, Chu YWY, Mok GTK, et al. Use of clinical chromosomal microarray in Chinese patients with autism spectrum disorder-implications of a copy number variation involving DPP10. Mol Autism. 2017;8.
113. Lintas C, Picinelli C, Piras IS, Sacco R, Brogna C, Persico AM. Copy number variation in 19 Italian multiplex families with autism spectrum disorder: Importance of synaptic and neurite elongation genes. Am J Med Genet Part B Neuropsychiatr Genet. 2017;174:547–56.
114. Xu M, Ji Y, Zhang T, Jiang X, Fan Y, Geng J, et al. Clinical Application of Chromosome Microarray Analysis in Han Chinese Children with Neurodevelopmental Disorders. Neurosci Bull. 2018;34:981.

115. Thygesen JH, Wolfe K, McQuillin A, Viñas-Jornet M, Baena N, Brison N, et al. Neurodevelopmental risk copy number variants in adults with intellectual disabilities and comorbid psychiatric disorders. Br J Psychiatry. 2018;212:287.
116. Papuc SM, Abela L, Steindl K, Begemann A, Simmons TL, Schmitt B, et al. The role of recessive inheritance in early-onset epileptic encephalopathies: a combined whole-exome sequencing and copy number study. Eur J Hum Genet. 2019;27:408.
117. Fan Y, Wu Y, Wang L, Wang Y, Gong Z, Qiu W, et al. Chromosomal microarray analysis in developmental delay and intellectual disability with comorbid conditions. BMC Med Genomics. 2018;11.
118. Lee JS, Hwang H, Kim SY, Kim KJ, Choi JS, Woo MJ, et al. Chromosomal Microarray With Clinical Diagnostic Utility in Children With Developmental Delay or Intellectual Disability. Ann Lab Med. 2018;38:473.
119. Napoli E, Russo S, Casula L, Alesi V, Amendola FA, Angioni A, et al. Array-CGH Analysis in a Cohort of Phenotypically Well-Characterized Individuals with “Essential” Autism Spectrum Disorders. J Autism Dev Disord. 2018;48:442–9.
120. Chan PYT, Luk HM, Lee FMY, Lo IFM. Genetic profile and clinical application of chromosomal microarray in children with intellectual disability in Hong Kong. Hong Kong Med J = Xianggang yi xue za zhi. 2018;24:451–9.
121. Maini I, Ivanovski I, Djuric O, Caraffi SG, Errichiello E, Marinelli M, et al. Prematurity, ventricular septal defect and dysmorphisms are independent predictors of pathogenic copy number variants: a retrospective study on array-CGH results and phenotypical features of 293 children with neurodevelopmental disorders and/or multiple congenital anomalies. Ital J Pediatr. 2018;44.
122. Homma TK, Krepischi ACV, Furuya TK, Honjo RS, Malaquias AC, Bertola DR, et al. Recurrent Copy Number Variants Associated with Syndromic Short Stature of Unknown Cause. Horm Res Paediatr. 2018;89:13–21.
123. Sys M, Van Den Bogaert A, Roosens B, Lampo A, Jansen A, Wouters S, et al. Can clinical characteristics be criteria to perform chromosomal microarray analysis in children and adolescents with autism spectrum disorders? Minerva Pediatr. 2018;70:225–32.
124. Viñas-Jornet M, Esteba-Castillo S, Baena N, Ribas-Vidal N, Ruiz A, Torrents-Rodas D, et al. High Incidence of Copy Number Variants in Adults with Intellectual Disability and Co-morbid Psychiatric Disorders. Behav Genet. 2018;48:323–36.

125. Munnich A, Demily C, Frugère L, Duwime C, Malan V, Barcia G, et al. Impact of on-site clinical genetics consultations on diagnostic rate in children and young adults with autism spectrum disorder. Mol Autism. 2019;10.
126. Coppola A, Cellini E, Stamberger H, Saarentaus E, Cetica V, Lal D, et al. Diagnostic implications of genetic copy number variation in epilepsy plus. Epilepsia. 2019;60:689.
127. Han JY, Jang W, Park J, Kim M, Kim Y, Lee IG. Diagnostic approach with genetic tests for global developmental delay and/or intellectual disability: Single tertiary center experience. Ann Hum Genet. 2019;83:115–23.
128. Jang W, Kim Y, Han E, Park J, Chae H, Kwon A, et al. Chromosomal Microarray Analysis as a First-Tier Clinical Diagnostic Test in Patients With Developmental Delay/Intellectual Disability, Autism Spectrum Disorders, and Multiple Congenital Anomalies: A Prospective Multicenter Study in Korea. Ann Lab Med. 2019;39:299–310.
129. Lindstrand A, Eisfeldt J, Pettersson M, Carvalho CMB, Kvarnung M, Grigelioniene G, et al. From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. Genome Med. 2019;11.
130. Wayhelova M, Smetana J, Vallova V, Hladilkova E, Filkova H, Hanakova M, et al. The clinical benefit of array-based comparative genomic hybridization for detection of copy number variants in Czech children with intellectual disability and developmental delay. BMC Med Genomics. 2019;12.
131. Micleaa D, Al-Khzouza C, Osan S, Bucerzan S, Cret V, Popp RA, et al. Genomic study via chromosomal microarray analysis in a group of Romanian patients with obesity and developmental disability/intellectual disability. J Pediatr Endocrinol Metab. 2019;32:667–74.
132. Capkova P, Srovnal J, Capkova Z, Staffova K, Becvarova V, Trkova M, et al. MLPA is a practical and complementary alternative to CMA for diagnostic testing in patients with autism spectrum disorders and identifying new candidate CNVs associated with autism. PeerJ. 2019;6.
133. Monteiro S, Pinto J, Mira Coelho A, Leão M, Dória P. Identification of Copy Number Variation by Array-CGH in Portuguese Children and Adolescents Diagnosed with Autism Spectrum Disorders. Neuropediatrics. 2019;50:367–77.

134. Pinheiro MI, Silva C, Lourenço L, Gonçalves D, Dória S, Guardiano M, et al. Array-CGH: Importance in the study of developmental delays in pediatrics. Rev Neurol. 2020;71:171–6.
135. Calderoni S, Ricca I, Balboni G, Cagiano R, Cassandrini D, Doccini S, et al. Evaluation of Chromosome Microarray Analysis in a Large Cohort of Females with Autism Spectrum Disorders: A Single Center Italian Study. J Pers Med. 2020;10:1–18.
136. Ali MAM, Hassan AM, Saafan MA, Abdelmagid AA. Additive Diagnostic Yield of Homozygosity Regions Identified During Chromosomal microarray Testing in Children with Developmental Delay, Dysmorphic Features or Congenital Anomalies. Biochem Genet. 2020;58:74–101.
137. Espeche LD, Solari AP, Mori MÁ, Arenas RM, Palomares M, Pérez M, et al. Implementation of chromosomal microarrays in a cohort of patients with intellectual disability at the Argentinean public health system. Mol Biol Rep. 2020;47:6863–78.
138. Farooqi MS, Figueroa S, Gotway G, Wang J, Luu HS, Park JY. Reinterpretation of Chromosomal Microarrays with Detailed Medical History. J Pediatr. 2020;222:180-185.e1.
139. Yuan H, Shangguan S, Li Z, Luo J, Su J, Yao R, et al. CNV profiles of Chinese pediatric patients with developmental disorders. Genet Med. 2021;23:669–78.
140. Milone R, Cesario C, Goldoni M, Pasquariello R, Fusilli C, Giovannetti A, et al. Correlating Neuroimaging and CNVs Data: 7 Years of Cytogenomic Microarray Analysis on Patients Affected by Neurodevelopmental Disorders. J Pediatr Genet. 2021;10:292.
141. Zacher P, Mayer T, Brandhoff F, Bartolomaeus T, Le Duc D, Finzel M, et al. The genetic landscape of intellectual disability and epilepsy in adults and the elderly: a systematic genetic work-up of 150 individuals. Genet Med. 2021;23:1492.
142. Yang EH, Shin YB, Choi SH, Yoo HW, Kim HY, Kwak MJ, et al. Chromosomal Microarray in Children With Developmental Delay: The Experience of a Tertiary Center in Korea. Front Pediatr. 2021;9:690493.
143. Moirangthem A, Mandal K, Saxena D, Srivastava P, Gambhir PS, Agrawal N, et al. Genetic heterogeneity of disorders with overgrowth and intellectual disability: Experience from a center in North India. Am J Med Genet Part A. 2021;185:2345–55.
144. Palka Bayard de Volo C, Alfonsi M, Morizio E, Guaciali-Franchi P, Mohn A, Chiarelli F. A 343 Italian cohort of patients analysed with array-comparative genomic

hybridization: unsolved problems and genetic counselling difficulties. J Intellect Disabil Res. 2021;65:863–9.
145. Dai L, Zhang D, Wu Z, Guan X, Ma M, Li L, et al. A Tiered Genetic Screening Strategy for the Molecular Diagnosis of Intellectual Disability in Chinese Patients. Front Genet. 2021;12.
146. Yang M, Xu B, Wang J, Zhang Z, Xie H, Wang H, et al. Genetic diagnoses in pediatric patients with epilepsy and comorbid intellectual disability. Epilepsy Res. 2021;170:106552.
147. Costa CIS, da Silva Montenegro EM, Zarrei M, de Sá Moreira E, Silva IMW, de Oliveira Scliar M, et al. Copy number variations in a Brazilian cohort with autism spectrum disorders highlight the contribution of cell adhesion genes. Clin Genet. 2022;101:134–41.
148. Perovic D, Damnjanovic T, Jekic B, Dusanovic-Pjevic M, Grk M, Djuranovic A, et al. Chromosomal microarray in postnatal diagnosis of congenital anomalies and neurodevelopmental disorders in Serbian patients. J Clin Lab Anal. 2022;36.
149. Chehbani F, Tomaiuolo P, Picinelli C, Baccarin M, Castronovo P, Scattoni ML, et al. Yield of array-CGH analysis in Tunisian children with autism spectrum disorder. Mol Genet Genomic Med. 2022;10.
150. Nassir N, Sati I, Al Shaibani S, Ahmed A, Almidani O, Akter H, et al. Detection of copy number variants and genes by chromosomal microarray in an Emirati neurodevelopmental disorders cohort. Neurogenetics. 2022;23:137–49.
151. Liu Y, Lv Y, Zarrei M, Dong R, Yang X, Higginbotham EJ, et al. Chromosomal microarray analysis of 410 Han Chinese patients with autism spectrum disorder or unexplained intellectual disability and developmental delay. NPJ Genomic Med. 2022;7.
152. Lee CL, Chuang CK, Tu RY, Chiu HC, Lo YT, Chang YH, et al. Increased Diagnostic Yield of Array Comparative Genomic Hybridization for Autism Spectrum Disorder in One Institution in Taiwan. Medicina (B Aires). 2022;58.
153. da Cunha Leite AJ, Pinto IP, Leijsten N, Ruiterkamp-Versteeg M, Pfundt R, de Leeuw N, et al. Diagnostic yield of patients with undiagnosed intellectual disability, global developmental delay and multiples congenital anomalies using karyotype, microarray analysis, whole exome sequencing from Central Brazil. PLoS One. 2022;17:e0266493.

154. Su M, Page S, Haag M, Swisshelm K, Hennerich D, Graw S, et al. Clinical utility and cost-effectiveness analysis of chromosome testing concomitant with chromosomal microarray of patients with constitutional disorders in a U.S. academic medical center. J Genet Couns. 2022;31:364–74.
155. Krepischi AC V., Villela D, da Costa SS, Mazzonetto PC, Schauren J, Migliavacca MP, et al. Chromosomal microarray analyses from 5778 patients with neurodevelopmental disorders and congenital anomalies in Brazil. Sci Reports 2022 121. 2022;12:1–11.
156. Tolezano GC, Bastos GC, da Costa SS, Freire BL, Homma TK, Honjo RS, et al. Burden of Rare Copy Number Variants in Microcephaly: A Brazilian Cohort of 185 Microcephalic Patients and Review of the Literature. J Autism Dev Disord 2022. 2022;:1–32.
157. Streață I, Caramizaru A, Riza A-L, Șerban-Sosoi S, Pîrvu A, Cara M-L, et al. Pathogenic Copy Number Variations Involved in the Genetic Etiology of Syndromic and Non-Syndromic Intellectual Disability-Data from a Romanian Cohort. Diagnostics (Basel, Switzerland). 2022;12.
158. Lengyel A, Pinti É, Pikó H, Kristóf Á, Abonyi T, Némethi Z, et al. Clinical evaluation of rare copy number variations identified by chromosomal microarray in a Hungarian neurodevelopmental disorder patient cohort. Mol Cytogenet. 2022;15:1–13.
159. Sheth F, Shah J, Jain D, Shah S, Patel H, Patel K, et al. Comparative yield of molecular diagnostic algorithms for autism spectrum disorder diagnosis in India: evidence supporting whole exome sequencing as first tier test. BMC Neurol. 2023;23:292.
160. Sandoval-Talamantes AK, Mori MÁ, Santos-Simarro F, García-Miñaur S, Mansilla E, Tenorio JA, et al. Chromosomal Microarray in Patients with Non-Syndromic Autism Spectrum Disorders in the Clinical Routine of a Tertiary Hospital. Genes (Basel). 2023;14.
161. Akter H, Rahman MM, Sarker S, Basiruzzaman M, Islam MM, Rahaman MA, et al. Construction of copy number variation landscape and characterization of associated genes in a Bangladeshi cohort of neurodevelopmental disorders. Front Genet. 2023;14:955631.

4.2 CAPÍTULO II - A COHORT STUDY OF INDIVIDUALS WITH NEURODEVELOPMENTAL DISORDERS AND/OR CONGENITAL ANOMALIES INVESTIGATED BY HIGH-RESOLUTION CHROMOSOMAL MICROARRAYS IN SOUTHERN BRAZIL: The Significance of Autism Spectrum Disorder.

Estudo de Coorte de Indivíduos com Distúrbios do Desenvolvimento Cerebral e Anomalias Congênitas Investigados por Microarray Cromossômico de Alta Resolução no Sul do Brasil: *A relevância do Transtorno do Espectro Autismo***.**

Microarray cromossômico (CMA) é referência na avaliação de variações do número de cópias (CNVs) de indivíduos com distúrbios do neurodesenvolvimento (DsND), como deficiência intelectual (DI) e/ou transtorno do espectro do autismo (TEA), que afetam cerca de 3-4% da população mundial. As plataformas modernas para CMA também incluem sondas para polimorfismos de nucleotídeo único (SNPs) que detectam regiões homozigóticas no genoma, como longos trechos contíguos de homozigose (LCSH), que resultam de homozigose cromossômica completa ou segmentar e podem ser indicativos de dissomia uniparental (UPD), endogamia, características populacionais, bem como eventos replicativos de reparo de DNA. Neste estudo retrospectivo, analisamos os arquivos de leitura de CMA solicitados por geneticistas e neurologistas para fins de diagnóstico, juntamente com os dados clínicos disponíveis. Nossos objetivos foram interpretar as Variações no Número de Cópias (CNVs) e avaliar as frequências e implicações dos Longos Trechos Contíguos de Homozigose (LCSH) detectados pelas plataformas Affymetrix CytoScan®HD (41%) ou 750K (59%) em 1.012 pacientes do sul do Brasil. A maioria dos pacientes era composta por crianças com DsND, e/ou anomalias congênitas (CAs). Um total de 206 CNVs foram interpretadas como patogênicas, incluindo 132 deleções e 74 duplicações, encontradas em 17% dos pacientes da coorte e em todos os cromossomos. Além disso, 12% apresentaram variantes raras de significado clínico incerto, incluindo LPCNVs, como a única CNV clinicamente relevante. Dentro do contexto dos DsND o TEA possui importância particular, devido à sua prevalência crescente e suas crescentes repercussões para indivíduos, famílias e comunidades. O TEA foi uma das fenotipagens clínicas, quando não a principal razão para a realização dos testes, em cerca de um terço da coorte, e esses pacientes foram

analisados como uma subcoorte. Considerando apenas os pacientes com TEA, a taxa de diagnóstico foi de 11%, dentro da faixa relatada na literatura (8-21%). Foi mais alta (16%) quando associada a características dismórficas e menor (7%) para TEA "isolado" (sem DI e sem características dismórficas). Em 953 CMAs de toda a coorte, os LCSH (≥3 Mbp) foram analisados não apenas no contexto de seu potencial significado patogênico, mas também foram explorados para identificar LCSH comuns na população do sul do Brasil. O CMA revelou pelo menos um LCSH em 91% dos pacientes. Em cerca de 11,5% dos pacientes, o LCSH sugeriu consanguinidade do primeiro ao quinto grau, com maior probabilidade de impacto clínico, e em 2,8%, revelaram uma provável Dissomia Uniparental (UPD). Os LCSH encontrados com frequência de 5% ou mais foram considerados LCSH comuns na população em geral, permitindo-nos delinear 10 regiões como potencialmente representando haplótipos ancestrais de significado clínico negligenciável. As principais indicações para o CMA foram atraso no desenvolvimento (56%), DI (33%), TEA (33%) e características sindrômicas (56%). Alguns fenótipos nessa população podem ser preditivos de uma maior probabilidade de indicar um portador de uma CNV patogênica. Aqui, apresentamos o maior relatório de dados de CMA em uma coorte com DsND e/ou CAs do Sul do Brasil. Caracterizamos as CNVs raras encontradas junto com os principais fenótipos apresentados por cada paciente e demonstramos a importância e utilidade da interpretação de LCSH nos resultados do CMA que incorporam SNPs, bem como ilustramos o valor do CMA para investigar CNVs no TEA.

Palavras-chave: Autismo, Anomalias Congênitas, LCSH, Variações No Número De Cópias, transtornos do neurodesenvolvimento, Microarranjos Cromossômicos, Brasil.

A COHORT STUDY OF INDIVIDUALS WITH NEURODEVELOPMENTAL DISORDERS AND/OR CONGENITAL ANOMALIES INVESTIGATED BY HIGH-RESOLUTION CHROMOSOMAL MICROARRAYS IN SOUTHERN BRAZIL: THE SIGNIFICANCE OF AUTISM SPECTRUM DISORDER**.**

Tiago Fernando Chaves[1,a], Maristela Ocampos[2,c], Ingrid Tremel Barbato[3,c], Louise Lapagesse de Camargo Pinto[4,d], Gisele Rozone de Luca[5,d], Jorge Humberto Barbato Filho[6,c], Priscila Bernardi[7,e], Yara Costa Netto Muniz[8,a] and Angelica Francesca Maris[9,b c].

1 Biologist, PhD student in cell biology and development, tiagochavo@msn.com (corresponding author);
2 Biologist, PhD in biotechnology and molecular biology, m_ocampos@yahoo.com;
3 Biologist and MSc in chemical engineering, ingridtbarbato@gmail.com;
4 Medical geneticist, PhD in child and adolescent health, lapagesselou@gmail.com;
5 Medical geneticist and pediatrician, giselegenetica@gmail.com;
6 Medical neuropediatricia, jorgebarbato@yahoo.com.br;
7 Medical geneticist, pricila.bernardi@gmail.com;
8 Professor of the Department of Cell Biology, Embryology and Genetics, Biologist, PhD in genetics, yara.muniz@ufsc
9 Professor of the Department of Cell Biology, Embryology and Genetics (Ret.) and genetic counselor (volunteer), PhD in Genetics and Molecular Biology, afmaris@gmail.com, angelica.maris@ufsc.br (corresponding author).
a Laboratório De Polimorfismos Genéticos (LAPOGE) - Universidade Federal de Santa Catarina – Florianópolis –SC.
b Universidade Federal de Santa Catarina – Florianópolis –SC.
c Laboratory Neurogene – Florianópolis – SC.
d Children's Hospital Joana de Gusmão – Florianópolis – SC.
e University Hospital Professor Polydoro Ernani de São Thiago, Florianópolis – SC, Brazil.
Email: tiagochavo@msn.com; afmaris@gmail.com

ABSTRACT

Chromosomal microarray (CMA) is the reference in evaluation of copy number variations (CNVs) in individuals with neurodevelopmental disorders (NDDs), such as intellectual disability (ID) and/or autism spectrum disorder (ASD), which affect around 3-4% of the world's population. Modern platforms for CMA, also include probes for single nucleotide polymorphisms (SNPs) that detect homozygous regions in the genome, such as long contiguous stretches of homozygosity (LCSH). These regions result from complete or segmental chromosomal homozygosis and may be indicative of uniparental disomy (UPD), inbreeding, population characteristics, as well as replicative DNA repair events. In this retrospective study, we analyzed CMA reading files requested by geneticists and neurologists for diagnostic purposes along with available clinical data. Our objectives were interpreting CNVs and assess the frequencies and implications of LCSH detected by Affymetrix CytoScan®HD (41%) or 750K (59%) platforms in 1,012 patients from the south of Brazil. The patients were mainly children with NDDs and/or congenital anomalies (CAs). A total of 206 CNVs were interpreted as pathogenic, including 132 deletions and 74 duplications, were found in 17% of the patients of the cohort and across all chromosomes. Further 12% presented rare variants of uncertain clinical significance, including LPCNVs, as the only clinically relevant CNV. Within the realm of NDDs, ASD carries a particular importance, owing to its escalating prevalence and its growing repercussions for individuals, families, and communities. ASD was one of the clinical phenotypes, when not the main reason for referral to testing, for about one-third of the cohort and these patients were further analyzed as a sub-cohort. Considering only the patients with ASD, the diagnostic rate was 11%, within the range reported in the literature (8-21%). It was higher (16%) when associated with dysmorphic features and lower (7%) for "isolated" ASD (without ID and without dysmorphic features). In 953 CMAs of the whole cohort, LCSH (≥3 Mbp) were analyzed not only in the context of their potential pathogenic significance but were also explored to identify common LCSH in the south Brazilians population. CMA revealed at least one LCSH in 91% of the patients. For about 11.5% of patients, the LCSH suggested consanguinity from the first to the fifth degree, with a greater probability of clinical impact, and in 2.8%, they revealed a

putative UPD. LCSH found at a frequency of 5% or more were considered common LCSH in the general population, allowing us to delineate 10 regions as potentially representing ancestral haplotypes of neglectable clinical significance. The main referrals to the CMA were developmental delay (56%), DI (33%), ASD (33%) and syndromic features (56%). Some phenotypes in this population may be predictive of a higher probability of indicating a carrier of a pathogenic CNV. Here we present the largest report of CMA data in a cohort with NDDs and/or CAs from the South of Brazil. We characterize the rare CNVs found along with the main phenotypes presented by each patient and show the importance and usefulness of LCSH interpretation in CMA results that incorporate SNPs, as well as we illustrate the value of CMA to investigate CNV in ASD.

Keywords: autism, congenital anomalies, LCSH, copy number variations, neurodevelopmental disorders, chromosomal microarrays, Brazil.

BACKGROUND

Neurodevelopmental disorders (NDDs) predominantly encompass developmental delay (DD), intellectual disability (ID), and/or autism spectrum disorders (ASD), impacting approximately 3-4% of the global population [1, 2]. These conditions are classified as non-syndromic when they occur in isolation and syndromic when they co-occur with dysmorphisms or evident congenital anomalies (CAs)[3].

With strong genetic underpinnings, ASD holds great significance within the realm of NDDs due to its high prevalence and increasing impact on individuals, families, and communities. The disorder's heterogeneity spans a wide spectrum of symptoms and severity, usually accompanied by co-occurring conditions, being characterized by impairment in social interaction and communication. According to the Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition (DSM-V), we can understand the deficits in social interactions and social communications of individuals with ASD based on three aspects: socio-emotional reciprocity; non-verbal communicative behaviors used for social interaction, development, maintenance and understanding of relationships; and restricted behaviors, such as repetitive patterns exhibited as movements, use of objects or speech, unalterable routines or ritualized behaviors (verbal or non-verbal), fixation on singular interests, and abnormal response to variations in sensory aspects of the environment [4]. Based on common deficits, the DSM-5 defines the current diagnosis of ASD that now, along with those of autistic disorder (classical autism), also incorporates the diagnoses of childhood disintegrative disorder, pervasive developmental disorder without other specification, and Asperger's syndrome.

Sometimes ASD is the main diagnosis, sometimes it is comorbid to other NDDs such as ID, frequent in the autistic spectrum. It can also be present in syndromic conditions when apparent dysmorphic features (DF) or CAs are present [3].

It is estimated that ASD presents a heritability between 0.5 to 0.9% [5, 6]. A recent review covering 74 studies with 30,212,757 participants concluded an estimated global prevalence of ASD of 0.6%. It is highest in America (1%), Africa (1%) and

Australia (1.7%) [7]. The prevalence of ASD worldwide has increased in recent decades, for example in the USA, the Centers for Disease Control and Prevention reported that the overall prevalence of ASD was 1,5% in 2010, 1.4% in 2012, 1.7 in 2014 and 1.9 in 2016, 2.3 in 2018 (CDC). The overall prevalence of ASD in Europe, Asia has also been gradually increasing [8, 9]. In Brazil, as well as in Latin America in general, epidemiological data on the prevalence of ASD are scarce. A single study carried out in the Southeast region of Brazil in 2011, found an estimated prevalence of 0,3% [10], however, it is believed to be an underestimation due to methodological issues. If we apply the prevalence of 1% estimated for the American population to the Brazilian population (214 million), ASD should affect approximately 2 million individuals [11].

Genetic and/or genomic factors such as single nucleotide polymorphisms (SNPs) and CNVs [12–17] have been suggested as the etiological cause in 50-60% of cases of ASD [18]. The SFARI Gene [6], one of the leading and constantly updated genetic databases on ASD, associates 1,262 genes and 2,290 CNVs, including those with rare frequency, to the condition (data from December 2022).

CNVs are structural variations in the DNA that involve gains or losses of large segments of genetic material (from hundreds to several million base pairs) that may be inherited or occur spontaneously during the formation of egg or sperm cells and can affect gene dosage, causing loss of function, haploinsufficiency, or overexpression of genes [19]. Specific CNVs have been shown to cause or increase the likelihood of developing certain NDDs such as ID, ASD, schizophrenia, as well as CAs. However, most people with CNVs do not have developmental disorders and for many CNVs related to disorders the presence of the CNV *per se* does not implicate necessarily the presence of the disorder, because their penetrance and expression is impacted by other genetic and/or by environmental factors, which makes their interpretation challenging.

For over a decade, Chromosomal microarray (CMA) technologies have been clinically recommended as the primary cytogenetic diagnostic test for investigating patients with NDDs [20] and in 2020 the ACMG reinforced this statement, along with a more detailed guidance on interpreting results [21].

Most modern microarray platforms along with genome-wide oligonucleotide probes (depending on the CMA design) also integrate high-density SNP probes, that test for single nucleotide changes in DNA sequences, allowing to detect regions of homozygosity that can be associated with disease or other traits like ancestry.

Long contiguous stretches of homozygosity (LCSHs) are relatively common in the general population and can occur due to the chance of unions among individuals with a common ancestor, in these cases they rarely are related to disease, likely characterizing regions of low recombination in the genome [22, 23]. However, larger LCSHs can also reveal consanguinity among parents, uniparental disomy (UPD) or homologous recombinational DNA repair events and therefore be associated with an increased risk for certain genetic disorders, particularly those caused by recessive genetic mutations. In population studies, the minimal thresholds for calling LCSH are usually set around 0.5–1.0 Mbp, while in clinical analysis, minimal thresholds are more conservatively set at 3–10 Mbp [24].

The presence of multiple large LCSH ≥ 5 Mbp, distributed throughout several chromosomes suggests consanguinity between the individual's biological parents, increasing the chance of inheritance of recessive monogenic disorders. However, when large LCSH(s), reside in only one chromosome, this can reflect correction of meiotic or early post meiotic errors that resulted in total or partial uniparental disomy (UPD). UPD occurs when a person receives the two copies of a chromosome, or part of a chromosome, from only one parent [25]. The two copies can be of maternal (UPDmat) or paternal (UPDpat) origin. An UPD is not necessarily pathogenic, however it is an important cause of genetic disease because several genes suffer genomic imprinting, which silences one allele of the chromosomal pair in a gender-specific manner and a series of imprinting disorders cause NDs associated with ID, autistic behavior, DD and seizures, like the Angelman's syndrome (UPD (15) pat), Prader-Willi syndrome (UPD (15) mat), Beckwith-Wiedemann syndrome (UPD (11) pat), Silver-Russell syndrome (UPD (7) mat), Temple syndrome (UPD (14) mat) and Kagami-Ogata syndrome (UPD (14) pat) [26]. Even when not affecting imprinted genes, the UPD can uncover recessive mutations in the uniparental homozygotic regions, for which the sole transmitting parent of this region was heterozygous.

Whole chromosome UPDs can arise as consequence of the correction of a meiotic segregation error that resulted in a monosomic or a trisomic zygote, by duplicating the only chromosome present in the monosomic zygote or by losing one of the exceeding chromosomes in case of trisomy. In the monosomy rescue both chromosomes of the pair will be from only one progenitor and completely homozygous (isodisomic) whether in the trisomy rescue the UPD only occurs when the two chromosomes that were retained are from the same progenitor. In later case they can be totally isodisomic when the meiotic non-disjunction of the two sister chromatids occurred in meiosis II, however, when the meiotic error occurred in meiosis I, because of the homologous chromosomal recombination they will be partially iso/heterodisomic (one or more LSCHs on the chromosome) or completely heterodisomic (not originating homozygous regions) since the outer sister chromatids do not recombine [27, 28]. Segmental UPDs can have complex causes, like rescue of a partial trisomy caused by translocated chromosomes, DNA double-strand breaks or others involving a replicative DNA repair mechanism [28–32].

The aims of this study included establishing the overall diagnostic rate of CMA in our settings, to verify the contribution of LCSH, the significance of patients with ASD phenotypes, to see if there is a difference in the diagnostic yield when considering only those with ASD phenotypes, and to provide detailed genetic data of known causal CNVs and/or of other rare, possibly causal, CNVs identified in the cohort.

METHODS

Ethical Aspects

The research project was reviewed and approved by the Research Ethics Committee of the Hospital Infantil Joana de Gusmão, the children's hospital in Florianópolis-SC, Brazil, under the reference number 2,339,104, and was conducted in accordance with the guidelines and criteria set forth in resolution No. 466/12 of the Brazilian National Health Council. Patients or their caregivers provided informed

consent to participate in the study. In cases where it was not possible to contact the patient for justifiable reasons (such as loss of contact information), the data was still used, and a Justification of Absence of Consent was signed by the research team. The team committed to maintaining the confidentiality and privacy of the patients whose data and/or information was collected in the records.

Cohort

The aim of this study was to investigate a significant cohort with developmental disorders from South Brazil. We collected a total of 1120 chromosomal microarray (CMA) read files that were performed by the Laboratório Neurogene in Florianópolis, Santa Catarina, Brazil, upon request by medical geneticists and neurologists for investigative/diagnostic purposes, primarily from the Joana de Gusmão Children's Hospital, but also from MDs from the University Hospital Professor Polydoro Ernani de São Thiago and from private clinics in Florianópolis, State of Santa Catarina, between 2013 and 2019. These include also 420 previously published cases[28]. Furthermore, 68 out of 1120 cases were excluded because they belonged to unaffected family members and 40 cases were excluded from the statistics of developmental disorders due to insufficient clinical information. The analyzed sample, therefore, consists of CMA read files and available clinical data from 1,012 patients, primarily children with neurodevelopmental disorders, from southern Brazil.

For analysis of the significance of ASD in our cohort, we established a sub- where we included every patient of the cohort where the clinical phenotype specifically mentioned ASD, autistic disorder (classical autism), childhood disintegrative disorder, pervasive developmental disorder without other specification or Asperger's syndrome as the main reason for referral to testing or as one of the phenotypes of a broader spectrum. We call "syndromic autism" those patients that had dysmorphic features/congenital anomalies (accompanied or not by intellectual disability) mentioned within their clinical phenotypes. In non-syndromic cases we have autism with intellectual disability and what we call "isolated autism", which would be the non-syndromic autism without intellectual disability. The ASD sub-cohort refers to 333

patients from the south of Brazil, of which 134 are part of a previously published study [33], for which CMA reading files and clinical data were available.

Collection of clinical data

To establish a correlation between the phenotype and potential causal genes, we gathered the required phenotypic/clinical data in the exam request form and, when possible, supplemented with direct information by their medical doctors. This was done through a questionnaire that asked information about the individual's clinical presentation, behavior, history of physical exams, previous genetic and metabolic tests results, and prescription medication. No new appointments were arranged with the patients for this study, and clinicians retrieved most of the data from their medical records.

Genomic analysis

The investigative CMA platforms used were CytoScan® 750K (59%) and CytoScan™ HD (41%) and the resulting files were analysed using the Chromosome Analysis Suite (ChAS) Affymetrix® software, which is based on the reference genome sequence of the University of California, Santa Cruz database (https-//genome.ucsc.edu/cgi-bin/hgGateway) using the human genome version of February 2009 (GRCh37/hg19). The analysis was retrospective, with the use of the CMA runs obtained from a clinical diagnostic laboratory, with previous consent of the patients.

Typically, the filter criteria for interpreting CNVs for diagnostic purposes are sizes larger than 100 Kbp for deletions and larger than 150 Kbp for duplications, both containing at least 50 markers, according to ACMG recommendations [18, 19]. However, since this is a research study, that aims to identify potential new genes involved in developmental disturbances, we reduced the filter parameters to >10 Kbp for deletions and for duplications, both with at least 10 markers. To interpret the CNVs, we followed the latest recommendations of the ACMG and the Clinical Genome Resource [21].

CNVs interpretation and classification

To interpret CNVs, regarding their function, dosage effects (known haploinsufficiency or overexpression studies) and effects of mutations, the UCSC Genome Browser with integrated databases was widely used, mainly ClinVar (NCBI), DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans using Ensembles Resources), DGV (Database of Genomic Variants), OMIM (Online Mendelian Inheritance in Man), ISCA (International Standard Cytogenomic Array), dbGaP (Database of Genotypes and Phenotype), dbVAR (Database of Large Scale Genomic Variants), ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations), PUBMED (Public Medline), ClinGen (Clinical Genome Resource), MGI (Mouse Genome Informatics Database, from The Jackson Laboratory), SFARI (Simons Foundation Autism Research Initiative) and the private database CAGdb (Cytogenomics Array Group CNV Database). We also use the interpretation of genomic variants as a tool, the Franklin platform [20], based on Artificial Intelligence, for classification and interpretation based on scores, as suggested by Riggs and coworkers [5].

The variants were classified into four types according to clinical interpretation as benign variants, variants of uncertain significance (VUS), likely pathogenic VUS (LPCNVs), or pathogenic variants (PCNVs), and the result in each case was assigned based on the CNVs of greatest clinical relevance detected in the genome of the patients [5].

Variables like location, type and size of each CNV, the CNV classification, number of CNVs detected for each individual, age, gender, clinical descriptions (phenotypes), previous genetic testing results (karyotype, fragile X, etc.), and other relevant known clinical data to which we had access, were compiled (with coded identification) into simple Excel sheet for data handling with the R software (version 3.4.2, the R Foundation for statistical Computing). This was done to understand the phenotypic frequency, the diagnostic rates, the average age and the gender distribution in the cohort, the frequency of genomic changes in each chromosome and

to find if there are any phenotypic clues related to a higher diagnostic probability by CMA (predictive phenotypes of a higher chance to be related to a pathogenic CNV), that possibly in the future could allow to select the cases that would benefit the most using CMA as a first-line test in settings of financial shortage.

Statistics

In the study, in addition to the descriptive biostatistical analysis, the univariate analysis (Fisher's test) was applied to identify eventual predictive phenotypes for a higher diagnostic result (greater chance of having a pathogenic CNV). To compare the mean sizes, amounts of covered genes and quantities of covered OMINs genes in the CNVs, by type of CNV found, multivariate analysis such as mean comparison test (Tukey's Multiple test) was applied. A *p*-value less than 0.05 was considered statistically significant.

Selection and analysis of LCSH

The analysis and selection of LCSH followed the methodology outlined in Chaves & coworkers (2019), applying a threshold of ≥ 3 megabase pairs (Mbp) for the LCSH analysis. This threshold is typically used in clinical investigations, as opposed to population-based studies, where the cut-off threshold is usually considerably lower [8]. All participants who had LCSHs satisfying the above criteria were included, regardless of whether they had or not a pathogenic CNV.

Automation of LCSHs analyses.

For investigation of consanguinity and comparative LCSH analysis among cases as well as for calling potential UPD, all the LCSH reported in ChAS for each case were copied with coded identification and compiled into Excel sheets.

For a more adequate and precise analysis the process was automatized and all LCSHs found in the cohort were imported into Google Colab and manipulated using the Python programming language. The libraries used for data manipulation and

analysis were Pandas and NumPy (for numerical computations). The code used for the analysis is available on the project's GitHub page: https://github.com/tiagochavo87/LCSH_analysis

Analysis of consanguinity

The frequency of consanguinity in the cohort was calculated according to Kearney, Kearney and Conlin (2011). In short, when the homozygous patterns suggested inbreeding, all the regions of homozygosity ≥ 3 Mbp distributed throughout the chromosomes were added, with exception of the LCSH located on the sex chromosomes; the total sum in Mbp being divided by the size of the autosomal genome, 2.881 Mbp (GRCh37 / hg19). The percentage obtained was correlated with the inbreeding coefficient (F), which is: 25% (first grade; 1/4 - parent/child or full siblings), 12.5% (1/8 – second grade: half siblings; uncle/niece or aunt/nephew; double first cousins; grandparent/grandchild), 6% (1/16 - third grade: first cousins), 3% (1/32 fourth grade: first cousins once removed), 1.5% (1/64 – fifth grade: second cousins), < 0.5% (1/128 - seventh grade: third cousins) [8]. Kearney and co-workers emphasized that this is a crude calculation, likely to represent an underestimate of the actual homozygous proportion because of the applied threshold of LCSHs over 3 Mbp and because the CMAs may not have SNP probes in certain regions like the acrocentric short arms and the centromeric regions. On the other hand, depending on the degree of inbreeding in the population, these correlations eventually could overestimate the direct kinship relation of the proband.

Uniparental Disomy (UPD)

When only LCSHs 3 to < 5 Mbp were present in the genome, but in one single autosomal chromosome the sum of two or three LCSHs (< 5 Mbp) exceeded 10 Mbp, the homozygous regions were considered a potential isodisomy resulting from a uniparental disomy (UPD) event that underwent previous recombination. When one or more LCSH over 5 Mbp was present in a single chromosome with a size or sum (in the case of multiple LCSHs) ≥ 10 Mbp, it was considered a potential UPD (regardless of

eventual LCSHs ≤ 5 Mbp on other chromosomes). When more chromosomes had LCSHs over 5 Mbp, it was not considered a potential UPD case [28].

The ChAS software does not recognize homozygosity, but the absence of heterozygosity named there as loss of heterozygosity (LOH). This includes hemizygous regions generated by a larger deletion. Therefore, all cases with LOHs ≥ 10 Mbp in size on a single autosomal chromosome, regardless of the presence of an additional chromosome with LOH(s) over 10 Mbps in size (or sum of sizes), were manually reviewed, to eliminate the confounding effect of eventual hemizygous regions to call LSCHs and ultimately an UPD.

Analysis of the most frequent LCSH

Of the 953 files available for LCSH analysis we selected the 917 microarrays for the cytobands that most frequently showed regions with LCSH ≥ 3 Mbp on an autosomal chromosome, and those LCSHs present in more than 5% of individuals were considered common LCSH. This percentage was chosen because the frequency of ≥ 1%, which is the usual threshold to define common polymorphisms of SNPs in a population, was not considered applicable here because this is an affected cohort. Also, others have chosen the same threshold (or lower) to consider LCSH found in an affected cohort as a common variation, probably not having clinical significance for their analysis [34–38]. Hence, in doing so we believe to have an adequate safety margin for selecting common LCSH due to ancestral haplotypes rather than due to consanguinity or other pathogenesis-related mechanisms.

To delineate a more accurate genomic position for the most frequent LCSH, the shared homozygosis sections were superimposed, and their genomic positions obtained based on the median of their beginning and end.

RESULTS

Out of the 1,012 cases, 615 (61%) were male and 397 (39%) were female, with ages ranging from 0 to 55 years, and a mean age of 10 years (median = 7.15, standard deviation = 10.2).

Previous karyotyping results were available for 182 patients, with 122 normal and 60 abnormal results for which CMA was requested to identify the specific sequences involved. However, for most patients no information about previous genetic assessments was available.

From the 1,012 microarrays, a total of 7,150 CNVs which fulfilled the filtering criteria were selected; 3,747 duplications and 3,403 deletions which were interpreted and classified into benign CNVs, pathogenic CNVs (PCNVs), variants of uncertain clinical significance (VUS) and likely pathogenic CNVs (LPCNVs).

Phenotypic Characterization

Out of the 1,012 cases, four were excluded from the phenotypic characterization due to the unavailability of clinical data.

The cohort is mostly characterized by individuals with neurodevelopmental impairment (85%), and 83% of cases had ID and/or DD. In 56% of cases only DD was present while ID was described in 33%. It should be noted that 420 (42%) were under 5 years of age, which is below the age range for intellectual disability diagnosis.

Phenotypic Characterization for cases with ASD

Cases with ASD represent 33% of our cohort, these 333 cases, 77 (23%) were under 5 years of age, below the age for diagnosis of ID, and of these, 17 (22%) had DF. Of the other 256 individuals 5 years or older, 68 had ID, of which 36 also had DF; 43 had only DF, and 145 had "isolated" autism (see cohort in methodology).

Of the 262 male cases, 59 (53%) were below age 5, the diagnostic age for ID, and of these 12 had DF. Of the 203 male cases, aged 5 or more, 53 presented ID, and of these 29 had DF, whereas 150 (74%) had no ID of which 30 presented DF and 120 presented what we call “isolated” autism.

Of the 71 female ASD cases, 18 (25%) were under age 5, and of these 5 had DF. Of the 53 females aged 5 or more, 15 had ID, and of these 7 also had DF, 38 (72%) had ASD without ID, 13 of them with dysmorphic features (DF) and 25 of them presenting what we call “isolated” autism.

In figure 1 we summarize the phenotypic characterization of the cases that presented ASD in the cohort.

Figure 1 - Summary of the phenotypic characterization of the cohort:

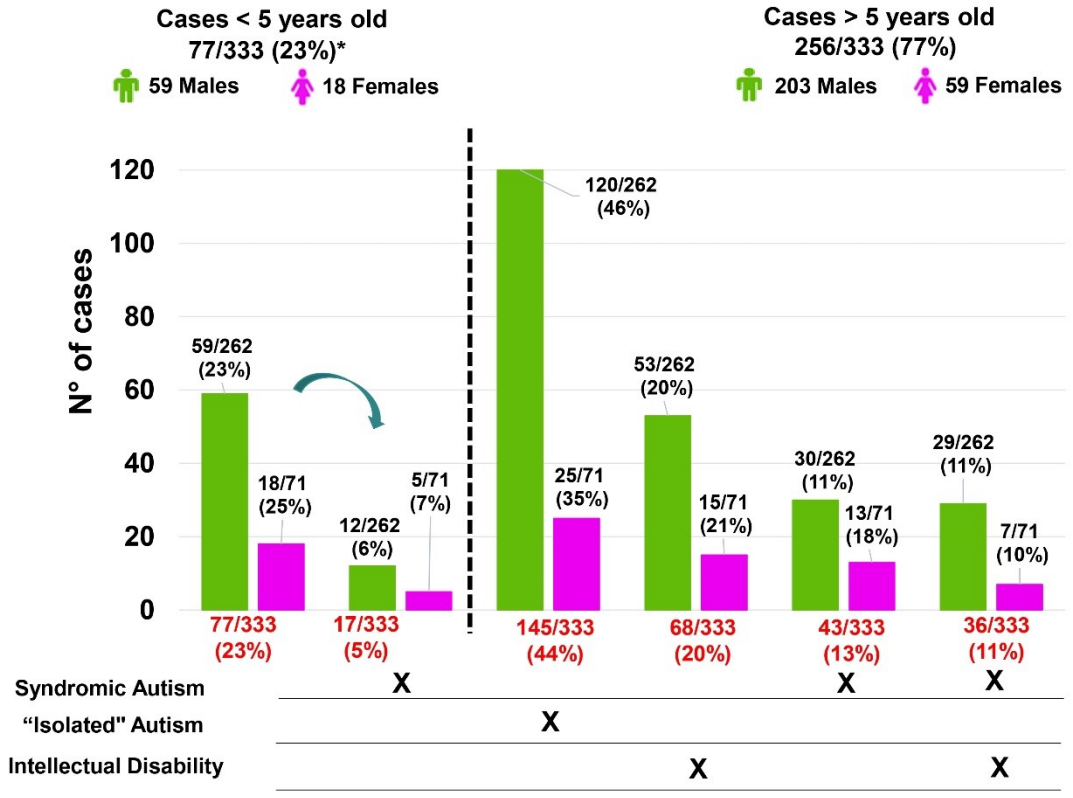


Figure 1: Phenotypic summary of the cohort, showing the number of cases in each category. The arrow indicates a deriving subcategory. Above each percentage are the numbers used to derive this percentage.

Other phenotypes

In addition to the main neurodevelopmental phenotypes, most individuals have syndromic features (56%) such as congenital anomalies or malformations or atypical (dysmorphic) facial features (47% of the cohort). Psychiatric or behavioral problems, variations in height or body weight were less frequent accompanying phenotypes.

The phenotypic characteristics recorded in our cohort are listed in Table 1.

Table 1 - The clinical characteristics recorded for patients with negative (only benign CNVs) and pathogenic (only PCNV) CMA results[#].

SIGNS / SYMPTOMS	In the cohort (N= 1008)	Negative (N =706)[#]	Pathogenic (N =175)[#]	*p*-value	Odds ratio
CHARACTERISTICS					
Obesity	3% (33)	2% (17)	5% (8)	0.076	0.46
Low weight	5% (55)	2% (34)	9% (16)	**0.010***	**0.44**
Abnormal growth	3% (29)	3% (21)	3% (5)	1	1.04
Short stature	10% (104)	9% (67)	14% (23)	0.05	0.60
Slender build	3% (34)	3% (20)	5% (8)	0.233	0.61
Prenatal problems	4% (36)	3% (23)	4% (6)	0.817	0.95
NEURODEVELOPMENT	85% (854)	85% (600)	83% (146)	0.639	1.12
Developmental delay	56% (569)	53% (377)	70% (119)	**0.0003*****	**0.53**
Motor development delay	8% (85)	7% (46)	12% (20)	**0.036***	**0.54**
Deafness or hearing loss	3% (31)	3% (19)	4% (7)	0.218	0.58
Speech and language delay and/or dyslalia	21% (216)	21% (151)	26% (44)	0.224	0.79
Difficulty of learning	6% (60)	7% (47)	4% (9)	0.603	1.32
Intellectual disability	33% (330)	31% (216)	41% (69)	**0.014***	**0.65**
Mild	4% (37)	3% (24)	2% (4)	-	-
Moderate	2% (16)	2% (11)	2% (4)	-	-
Severe	2% (19)	2% (11)	2% (4)	-	-
Not Specified	26% (258)	24% (170)	34% (57)	-	-
Intellectual disability and/or developmental delay	83% (834)	65% (456)	76% (129)	**0.025***	**0.65**
BEHAVIORAL		-	-		
Behavioral changes (Obsessive-compulsive disorder, attention deficit hyperactivity disorder, self and hetero-aggression, behavior disorder, psychosis)	12% (122)	11% (79)	14% (23)	0.509	0.83
Autism Spectrum Disorder	33% (333)	36% (255)	20% (34)	**0.0001****	**2.18**
CONGENITAL MALFORMATION(S) AND/OR	56% (563)	-	-		

SIGNS / SYMPTOMS	In the cohort (N= 1008)	Negative (N =706)[#]	Pathogenic (N =175)[#]	*p*-value	Odds ratio
CHARACTERISTICS					
DYSMORPHISM(S)					
FACIAL MALFORMATIONS/ DYSMORPHISMS	47% (471)	43% (305)	65% (110)	**0.0001****	**0.42**
Long face	2% (23)	-	-		
Wide face	1% (11)	-	-		
Narrow face	1% (5)	-	-		
Triangular face	1% (9)	-	-		
Asymetrical face	1% (12)	-	-		
Cleft palate	3% (31)	-	-		
Micrognathia	2% (21)	-	-		
Mouth/Lips (unusual)	5% (49)	-	-		
Microcephaly (Craniosynostosis included)	7% (67)	-	-		
Macrocephaly	3% (35)	-	-		
Ears (dysmorphic)	7% (70)	-	-		
Eyes (unusual)	14% (140)	-	-		
Forehead (unusual)	1% (7)	-	-		
Eyebrows (unusual)	1% (8)	-	-		
Nose (unusual)	4% (42)	-	-		
Hair (unusual)	2% (18)	-	-		
Not Specified	22% (218)	-	-		
OTHER CONGENITAL MALFORMATIONS		-	-		
Musculoskeletal (scoliosis, diaphragmatic hernia, vertebral anomaly)	4% (42)	4% (29)	2% (4)	0.830	1.21
Upper limb anomalies	8% (79)	6% (40)	15% (25)	**0.0003*****	**0.36**
Lower limb anomalies	8% (83)	6% (45)	15% (25)	**0.0015*****	**0.41**
Heart anomalies and malformations	8% (79)	7% (48)	12% (20)	**0.018***	**0.51**
Gastrointestinal anomalies and malformations	4% (44)	4% (25)	6% (10)	0.1955	0.61
Genitourinary anomalies and malformations	4% (44)	4% (26)	9% (15)	**0.004****	**0.38**
NEUROLOGIC ABNORMALITY	24% (239)	22% (155)	29% (50)	0.071	0.70
Epilepsy	6% (62)	6% (42)	5% (8)	0.856	1.17
Ataxia	2% (18)	1% (10)	2% (4)	0.495	0.61
Hypotonia	7% (70)	7% (51)	8% (14)	0.746	0.90
Abnormal brain structure	11% (112)	10% (72)	14% (24)	0.177	0.71

SIGNS / SYMPTOMS	In the cohort (N= 1008)	Negative (N =706)[#]	Pathogenic (N =175)[#]	*p*-value	Odds ratio
CHARACTERISTICS					
Seizures	6% (61)	5% (37)	6% (10)	0.850	0.91
ENDOCRINOLOGICAL ABNORMALITIES	4% (39)	3% (21)	5% (8)	0.340	0.64
CUTANEOUS ABNORMALITIES (hyper and hypopigmentation, hemangioma, freckles, café-au-lait spots and others)	3% (29)	2% (16)	4% (7)	0.192	0.56
HEMATOLOGIC ABNORMALITIES	2% (19)	2% (14)	1% (2)	0.751	1.75

Comparison groups diagnosed with pathogenic CNVs (diagnosed) versus the groups without clinically relevant CNVs (no CNVs or only benign CNVs). Cases where VUS and LPCNVs was the most relevant finding (128 individuals) were not considered in the correlation, because they represent inconclusive diagnosis.

Diagnostic rate and interpretation of CNVs

Within our cohort of 1.012 individuals, we identified 358 rare CNVs (VUS, LPCNVs and PCNVs), of which 203 were interpreted as pathogenic and were present in 170 individuals, representing 17% of the cohort. The description of the PCNVs and clinical phenotypes of the carrier patients are listed in Table 2, Table 3 and the previously published are listed in Chaves & coworkers [33].

Table 2 - Pathogenic CNVs (PCNV) found in the cohort without ASD. Does not include cases previously published in [33].

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#497	Del	arr[GRCh37] 2p13.2p13.1(72,707,781-73,680,438)x1	973	12	EXOC6B, SPR	Abnormal growth, DD and psychiatric disorder	M/-	ND		-
#501	Del	arr[hg19] Xq21.1-q28(84,107,007-155,233,098)x1	71,126	401	445 OMINs	DD, DI, hypoplastic fourth toe, phalangeal malformation	F/ 2 PCNVs	ND	(46,X,der(X)(~q21-qter)	
#501	Dup	arr[hg19] 11q13.1q25(65,446,446-134,937,416)x3	69,490	550	729 OMINs	DD, DI, hypoplastic fourth toe, phalangeal malformation	F/ 2 PCNVs	ND	(46,X,der(X)(~q21-qter)	
#527	Del	arr[GRCh37] 5p15.33p14.1(113,577-27,590,026)x1	27,476	151	61 OMINs	Short stature, agenesis of the corpus callo sum, DD, cardiomyopathy and laryngeal web	F/ potential UPD5	ND	46,XX,del(5)(~p13-p15)	Cri du Chat¨syndrome
#533	Del	arr[GRCh37] 16q24.2q24.3(87,939,406-89,481,546)x1	1,542	38	ANKRD11	Obesity, DD, SLD, ID	F/-	ND	-	KBG syndrome
#542	Dup	arr[GRCh37] 17p12(14,083,055-15,503,234)x3	1,420	15	PMP22 (*601097)	Charcot-Marie-Tooth like phenotype	F/-	ND	-	Charcot-Marie-Tooth type 1A – CMT1A
#549	Del	arr[hg19] 8q22.2-q22.3(101,265,736-104,749,739) x1	3,484	53	SLC25A3Z (*610815),GRHL2 (*608576)	Gross motor delay	M/-	ND	-	-

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#552	Dup	arr[GRCh37] Xq13.2q21.32(71,947,354-93,480,817)x3	21,533	84	49 OMINs	Fetal loss	F/-	ND	46,XX,dup (X)(q13q22)	-
#568	Del	arr[GRCh37] 3p22.3p22.2(32,654,188-36,570,368)x1	3,916	15	TRIM71, CRTAP and GLB1	Hypotonia, poor head support, does not eye track, long face, downslanted palpebral fissures, stridor, glossoptosis, widely-spaced nipples, dolichocephaly, micrognathia hyperconvex toenail, spatulate terminal phalanges, inguinal hernia, macrocephaly, tall stature	F/-	ND	-	-
#574	Del	arr[hg19] Xq28(154,154,958-155,233,731)x1	1,079	21	RAB39B (*300774)	umbilical cord hernia, macrocephaly	F/-	ND	45X,der(X)t(X;15)(q28;q11.2),-15	-
#581	Del	arr[GRCh37] 1p36.33p36.23(849,467-7,883,834)x1	7,034	162	GABRD, GLB1 (*611458), CRTAP (*605497)	Newborn, hydrocephalus with mild to moderate colpocephay of the lateral ventricles, w/o intracranial hypertension, microtia grade I, overlapping toes, broad forehead, bulbous nose, in ICU under	F/-	ND	-	¨Chromosome 1p36 Deletion Syndrome

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
						mechanical ventilation.				
#606	Del	arr[hg19] 5q35.2q35.3 (175,416,095-177,439,550)x1	2,023	51	NSD1 (*606681)	large hands and feet, plagiocephaly, DD, axial hypotonia, dilated Cardiopathy, inguinal hernia.	F/-	ND	-	Sotos syndrome
#611	Del	arr[hg19] 1p36.23-p36.33(849,466-2,040,693)x1	2,040	66	GABRD (137163), MMP23B (603321)	DD, ID, DF, history of pre-maturity with macrocephaly.	F/-	ND	-	Chromosome 1p36 deletion syndrome
#620	Dup	arr[GRCh37] 15q11.2q13.3(22,770,422-32,915,723)x4	10,145	179	GABRB3(* 137192)	SevID, short stature	F/-	ND	46,XX +mar	Síndrome de Tetrassomia Parcial do 15q
#626	Del	arr[GRCh37] 21q22.12q22.3(37,742,853-42,805,421)x1	5,063	49	DYRK1A (600855), KCNJ6 (600877)	DD, microcephaly, abnormal pinna, long phalanges and widely-spaced nipples.	F/-	ND	-	DYRK1A-related intellectual disability syndrome
#632	Del	arr[GRCh37] 22q11.21(18,636,750-21,800,471)x1	3,164	91	TBX1 (602054)	quadriparesis with hypotonia, multiple CAs, DD	M/-	ND	-	22q11.2 deletion syndrome
#646	Del	arr[GRCh37] 13q34(114,036,741-115,107,733)x1	1,071	22	CHAMP1	Leopard syndrome like frekles, mildID, alopecia, tremor	F/1 VUS	ND	-	-
#651	Del	arr[GRCh37] 5p15.33p15.2(113,577-13,142,487)x1	13,029	103	SLC9A3	DD, ID, seizures, glaucoma	M/ 2 PCNVs	ND	46,XY, add(5)(p15)	Cri du Chat syndrome (#123450)

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#651	Dup	arr[GRCh37] 13q31.1q34(87,535,468-115,107,733)x3	27,572	201	81 OMINs	DD, ID, seizures, glaucoma	M/ 2 PCNVs	ND	46,XY, add(5)(p15)	-
#652	Del	arr[GRCh37] 6q25.3q27(159,427,416-170,914,297)x1	11,487	102	48 OMINs	microcephaly, DD, ID, FD	F/-	ND	46,XX/46,XX del(6)(q13q15)?	-
#653	Dup	arr[GRCh37] 21q11.2q22.3(15,016,487-48,093,361)x3	10,072	369	167 OMINs	microcephaly, microtia, DD	M/-	ND	47,XY,del(9)(˜p21-pter),+extra	Partial 21q trisomy
#664	Del	arr[GRCh37] 17p11.2(17,121,644-20,187,953)x1	3,066		RAI1 (*607642)	DD, microtia, ventricular septal defect, FD.	M/-	ND	-	Smith-Magenis syndrome
#668	Del	arr[GRCh37] 15q11.2q13.1(23,290,788-28,704,050)x1	5,4,13	126	SNRPN (*182279)	DD, generalized hypotonia, dyslalia, ID, binge eating, motor restlessness, low vision, almond-shaped eyes	F/-	ND	-	Prader Willi syndrome
#676	Del	arr[GRCh37] 6q25.1q27(152,345,416-170,914,297)x1	18,569	146	ARID1B (*614556)	micrognathia, glossoptosis, agenesis of the corpus callosum, DD	F/-	ND	46,XX del(6)(q2-)	-
#678	Del	arr[GRCh37] 7q11.23(72,701,099-74,141,494)x1	1,440	30	ELN (*130160)	DD, hypotonia, bone abnormalities	M/-	ND	-	Williams Beuren syndrome

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#687	Del	arr[GRCh37] 5p15.33p15.2(113,577-13,142,487)x1	13,029 ,	103	CTNND2(*604275)	DD, intraventricular communication, axial polidactilia, FD	F/ 2 PCNVs	ND	46,XX, add(5)(p15)	Cri du Chat syndrome
#687	Dup	arr[GRCh37] 13q31.1q34(87,535,468-115,107,733)x3	27,572	201	CHAMP1	DD, intraventricular communication, axial polidactilia, FD	F/ 2 PCNVs	ND	46,XX, add(5)(p15)	-
#689	Dup	arr[GRCh37] 8p23.3p22(158,049-14,095,397)x3	13,937	152	TNKS1 (*603303), SOX7 (*612202), GATA4 (*600576)	Intraventricular communiCAstion, patent foramen ovale, aortic stenosis, pulmonary stenosis, inguinal hernia, long palpebral fissures, DD, ID	F/-	ND	46,XX,14pst+	8p23.3p22 duplication syndrome
#698	Del	arr[GRCh37] 11p14.2p13(26,997,314-33,491,623)x1	6,494	47	WT1 (*607102), PAX6 (*607108), BDNF (*113505)	ID, aniridia, cryptorchidism, micropenis, hypogonadism, myopia, FD	M/-	ND	-	WAGR syndrome
#709	Del	arr[GRCh37] 1p36.33p36.22(849,467-11,465,408)x1	10,616	196	GABRD (137163), PRKCZ (176982), SKI (164780)	Syndromic features (not specified)	F/-	ND	-	Chromosome 1p36 deletion syndrome
#712	Del	arr[GRCh37] 2q37.2q37.3(235,913,632-242,782,258)x1	6,869	86	HDAC4 (605314), GPR35 (602646)	Sagittal craniostenosis, motor delay, SLD, short fingers, FD	F/-	ND	-	Chromosome 2q37 deletion syndrome
#713	Del	arr[GRCh37] Xq26.2(132,496,732-133,293,329)x0	797	3	GPC4 (300168), GPC3 (300037)	Suspected Beckwith Wiedemann syndrome	M/-	ND	-	Simpsom-Gobali-Behmel syndrome

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#720	Del	arr[GRCh37] 12q13.3q14.1(57,075,559-58,481,772)x1	1,406	56	KIF5A, MARS	Hypotonia, ModID, psychiatric disorder, elongated face, oblique palpebral fissures, abnormal lip shape	F/-	ND	-	-
#735	Del	arr[GRCh37] 3q22.1q25.2(132,936,742-152,466,305)x1	19,530	147	ZIC4 (*608948), ZIC1 (*600470)	Multiple CAs	M/-	ND	-	-
#739	Dup	arr[GRCh37] 8q24.13q24.3(125,496,812-146,295,771)x3	20,799	219	117 OMINs	DD, ID, interatrial communication, preauricular appendage, FD	F/2 PCNVs	ND	46,XX +mar	-
#739	Dup	arr[GRCh37] 22q11.1q11.21(17,277,402-20,729,389)x3	3,452	85	44 OMINs	DD, ID, interatrial communication, preauricular appendage, FD	F/2 PCNVs	ND	46,XX +mar	Chromosome 22q11.2 microduplication syndrome
#740	Del	arr[hg19] 9q22.33q33.1(102,245,320-119,845,528)x1	17,600	136	ZNF462	Short stature, global DD, microcephaly, FD	M/-	ND	-	-
#746	Del	arr[GRCh37] 1p36.31-p22(9,580,727-11,784,118)x1	2,203	38	PIK3CD, MTOR	Failure to thrive, DD, ID, conduct disorder, FD, prominent nasal septum and retrognathia	F/-	ND	-	Chromosome 1p36 deletion syndrome
#760	Del	arr[GRCh37] 12p13.2p12.3(11,867,287-15,360,229)x1	3,493	46	GRIN2B (138252)	DD, ID	F/-	ND	-	-

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#761	Del	arr[hg19] Xp22.31(6,455,149-8,135,644)x1	1,680		STS (*300747)	Growth delay, DD	F/-	ND	-	-
#786	Del	arr[GRCh37] 14q32.33(105,213,585-106,328,827)x1	1,115	34	NUDT14 (*609219), BRF1 (*604902), PACS2 (*610423), MTA1 (*603526)	Global DD, hypotonia, SLD, hypotonic face, mesofacial hypoplasia, oblique palpebral fissures, tapered nasal bridge, smooth philtrum	F/-	ND	46,XX add[14](q32.3)	-
#786	Dup	arr[GRCh37] 18q21.31q23(54,211,852-78,013,728)x3	23,802	122	PIGN (*606097)	Global DD, hypotonia, SLD, hypotonic face, mesofacial hypoplasia, oblique palpebral fissures, tapered nasal bridge, smooth philtrum	F/-	ND	46,XX add[14](q32.3)	Partial 18q trissomy
#795	Del	arr[GRCh37] 15q11.2q12(23,214,984-25,778,351)x1	2,563	110	14 OMINs	4 yrs,scoliosis, protruded ears, thin upper lip, tapered fingers, DD.	M/-	ND	-	Prader Willi or Angelman syndrome.
#801	Del	arr[GRCh37] Xq22.1q22.3(101,083,092-105,991,325)x1	4,908	57	PLP1	FD, hypothyroidism, low weight, DD	F/-	ND	-	-
#807	Dup	arr[GRCh37] Xp22.33p22.13(878,067-19,071,519)x2	18,193	125	90 OMINs	DD, hypotonia, long hands and feet	M/-	ND	-	-

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#811	Dup	arr[GRCh37] 12p13.33p12.1(173,787-21,971,638)x3	21,798	312	217 OMINs	FD, strabismus, bulbous nose, epicanthus, dysphagia	M/-	ND	46,XY,add(18)(~p11.32)	-
#825	Del	arr[GRCh37] 2q31.1q31.3(175,599,084-181,995,668)x1	6,397	57	HOXD10 (*142984)	ASD, camptodactily of 2nd fingers, deformity of 5th fingers, DD, dyslalia	M/-	ND	-	Chromosome 2q31.2 microdeletion syndrome
#826	Del	arr[GRCh37] 20q13.13(47,150,102-49,526,051)x1	2,376	38	ADNP, KCNB1	SLD, congenital cardiopathy, nephropathy, cognitive delay and motor delay	F/-	ND	-	-
#828	Dup	arr[GRCh37] 1q21.1q21.2(146,106,724-147,830,830)x3	1,724	24	GJA5 (121013), GJA8 (600897)	ID, strabismus and protruding ears Brother of case #829	M/ 1 VUS	ND	-	Chromosome 1q21.1 duplication syndrome
#829	Dup	arr[GRCh37] 1q21.1q21.2(146,106,724-147,830,830)x3	1,724	24	GJA5 (121013), GJA8 (600897)	ID, strabismus and protruding ears. Brother of case #828	M/ 1 VUS	ND	-	Chromosome 1q21.1 duplication syndrome
#848	Del	arr[GRCh37] 16p11.2(29,591,327-30,190,029)x1	599	22	32 OMINs	Temporal lobe epilepsy, in clusters.	F/-	ND	-	Chromosome 16p.11.2 deletion syndrome
#861	Dup	arr[GRCh37] 8p23.3p21.1(158,049-27,502,930)x3 pat	27,345	263	129 OMINs	Low weight, elongated face, synophrys and pointed fingers	M/ 2 PCNVs	Father with partial trisomy of chr 8 and monosomy of chr 12 karyotype	-	-

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#861	Del	arr[GRCh37] 12p13.33p13.31(173,787-5,952,112)x1 pat	5,778	65	40 OMINs	Low weight, elongated face, synophrys and pointed fingers	M/ 2 PCNVs	Father with partial trisomy of chr 8 and monosomy of chr 12 karyotype	-	-
#862	Dup	arr[GRCh37] Xp11.23p11.22(48,224,463-52,841,006)x3	4,617	102	78 OMINs	FD, camptodactyly, nasal voice, SLD, ID e finger-like thumb	F/-	ND	-	Chromosome Xp11.23p11.22 duplication syndrome
#875	Del	arr[hg19] 4q28.3q31.21(136,216,198-14,1932,587)x1	5,716	-	MAML3 (608991)	Failure to thrive, DD, SLD, ADHD, hypermetropia	F/-	ND	-	-
#927	Del	arr[GRCh37] 15q11.2q13.1(23,620,192-28,704,050)x1	5,084	120	UBE3A (*601623)	Suspected Angelman syndrome	M/ 2 VUS	ND	-	Angelman syndrome
#944	Del	arr[GRCh37] 7q11.23(72,692,113-74,136,633)x1	1,445	30	ELN (*130160)	Suspected Williams-Beuren syndrome. Father with tuberous sclerosis	M/-	ND	-	Williams Beuren syndrome
#949	Del	arr[GRCh37] 2p14(64,103,858-67,815,028)x1	3,711	34	SLC1A4, SPRED2	Suspected genetic condition	M/-	ND	-	2p14 microdeletion syndrome
#953	Del	arr[GRCh37] 6p25.3p25.2 (156974-4009868)x1	3,853	41	FOXC1, RIPK1	Behavioural issues, FD, altered dentinogenesis, low weight, low height.	M/-	ND	45,XX,der(6)t(6;14)(p25;q11.2) – 14	6p25 deletion syndrome

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#960	Del	arr[GRCh37] 17q21.31(43,710,150-44,214,816)x1	505	10	KANSL1(*612852).	DD, SLD, dyslalia, ID, global hypotonia, no sphincter control	F/-	ND	-	Koolen de Vries syndrome
#961	Dup	arr[GRCh37] 12p13.31p13.33(1,73,786-34,759,042)x3	33,639	397	-	DF, cleft palate, anal fistula, imperforate anus, hypertrophic CAsrdiomyopathy, hypodense area affecting the cortiCAsl region, pigmentary lesion in the right eye.	M/-	ND	-	Trisomy of the Chromosome 12p¨
#985	Del	arr[GRCh37] 15q24.1q24.2 (72,965,465-76,073,450)x1	3,108	70	SIN3A	Prognathism, clinodactyly, DD, partial syndactyly	F/-	ND	-	Chromosome 15q24 Microdeletion syndrome
#993	Del	arr[GRCh37] 7q11.23(72,723,370-74,136,633)x1	1,413	30	ELN	DD, ligamentous laxity, tricuspid reflux, pulmonary stenosis.	F/-	ND	-	**Williams Beuren syndrome**
#995	Del	arr[hg19]8p23.1-8p23.3(158,048-6,999,114)x1	6,886	48	ARHGEF10	Failure to thrive, FD, hydrocephalus, thin corpus callosum, ventricular ectasia.	F/ 3 PCNVs	ND	46,XX,add(8)(?-pter)	Recombinant chromosome 8 syndrome or San Luis Valle syndrome
#995	Dup	arr[hg19]8p12-p23.1 (11,935023-31,833,216)x3	29,487	181	-	Failure to thrive, FD, hydrocephalus, thin corpus callosum, ventricular ectasia.	F/ 3 PCNVs	ND	46,XX,add(8)(?-pter)	Recombinant chromosome 8 syndrome or San Luis Valle syndrome

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#995	Dup	arr[hg19] 8q24.3 (144,794,838-146,295,771)x3	1,501	77	-	Failure to thrive, FD, hydrocephalus, thin corpus callosum, ventricular ectasia.	F/ 3 PCNVs	ND	46,XX,add(8)(?-pter)	St. Louis Valley syndrome
#1005	Del	arr[hg19] 16p11.2 (29,580,020-30,190,029)x1	610	32	ALDOA (*103850)	DD, SLD, FD	F/-	ND	-	Chromosome 16p.11.2 deletion syndrome
#1025	Dup	arr[GRCh37] 4q31.23q35.2(149,377,750-190,957,460)x3	41,580	210	110 OMINs	FD, deafness	F/-	ND	46,XX add(4) ~ q35	Chromosome 4q31.23q35.2 trisomy
#1033	Dup	arr[hg19] 22q11.21 (18,916,842-21,461,017)x3	2,544	79	TBX1(*602054)	DD, SLD, ID, seizures and ADHD	M/-	ND	-	22q11.21 duplication syndrome
#1034	Del	arr[hg19] 13q22.3(78,451,099-78,483,275)x1	32	2	EDNRB (*131244)	Deafness, iris heterochromia, white lock. Sister of #1035	F/-	ND	-	Waardenburg syndrome type 4A
#1035	Del	arr[hg19] 13q22.3(78,451,099-78,483,275)x1	32	2	EDNRB (*131244)	Deafness, iris heterochromia, white lock Sister of #1034	F/-	ND	-	Waardenburg syndrome type 4A
#1042	Del	arr[hg19] 22q11.21 (18,648,855-21,800,471)x1	3,151	97	TBX1(*602054)	ataxia, DD, FD, epiCAsnthus and oblique palpebral fissures	M/-	ND	-	Di George syndrome
#1047	Del	r[hg19] 2q37.1q37.3(235,387,296-242,782,258)x1	7,395	89	HDAC4 (*605314)	DD, motor delay, ligamentous laxity	F/-	ND	-	Chromosome 2q37 deletion syndrome

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#1048	Del	arr[hg19] 13q22.1q31.1(74,541,519-81,906,132)X1	7,365	48	EDNRB (*131244), SPRY2 (*602466)	DD, motor delay, ID, FD, pectus exCAsvatum	F/-	ND	-	-
#1059	Del	arr[hg19] 5q22.1q23.3 (110,061,210-128,612,286)x1	18,551	104	APC (*611731)	valvular and supravalvular stenosis, single kidney on the right, failure to thrive	F/-	ND	-	-
#1068	Del	arr[GRCh37] 15q11.2(22,770,422-23,277,436)x1	507	7	TUBGCP5 (*608147), CYFIP1 (*606322), NIPA2 (*608146), NIPA1 (*608145)	Failure to thrive, microcephaly, DD, SLD, ID, psychomotor agitation, ASD, congenital CAsrdiopathy, bilateral ectopic testis, right hemiparesis, altered MRI	M/-	ND	-	Chromosome 15q11.2 deletion syndrome
#1074	Dup	arr[GRCh37] 7q31.32q33(122,745,868-136,171,005)x3	13,425	125	LEP (*164160)	DD, SLD, ID, FD, enlarged thumbs, hirsutism (legs, arms, back), short stature, microcephaly	F/-	ND	-	7q31.32q33 partial trisomy
#1080	Dup	arr[GRCh37] 10p15.3p13(100,048-14,753,970)x3	14,654	124	53 OMINs	DD, speaks few words, FD, seizures agitation, lack of concentration, oblique palbebral	F/-	ND	-	-

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
						fissures, flat feet, brachydactyly, vitiligo(?) Mother has a transloCAstion.				
#1080	Del	arr[GRCh37] 15q26.3(100,995,007-102,429,040)x1	1,434	22	10 OMINs	DD, speaks few words, FD, seizures agitation, lack of concentration, oblique palbebral fissures, flat feet, brachydactyly, vitiligo(?) Mother has a translocation.	F/-	ND	-	-
#1080	Dup	arr[GRCh37] 10p15.3p13(100,048_14,753,970)x3	14,654	124	53 OMINs	DD, speaks few words, DF, seizures, restlessness and lack of concentration, oblique palpebral fissures, flat feet, brachydactyly, investigation of vitiligo. Mother has a translocation.	F/-	ND	-	-
#1083	Del	arr[GRCh37] 8q23.1q24.11(109,456,954-118,350,705)x1	8,894	25	TRPS1(*604386)	Suspected Trichorhinophalangeal syndrome.	F/-	ND	-	Trichorhinophalangeal syndrome (TRPS).
#1109	Dup	arr[GRCh37] 4q31.21q35.2(141,800,494-190,957,460)x3	49,157	254	136 OMINs	Test was made to identify additional material on chromosome 3	M/ 2 VUS	ND	Mentions Karyotype with additional material on chromosome 3	-

Table of Pathogenic CNVs (PCNV)

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#1113	Del	arr[GRCh37] 22q11.21(20,716,877-21,800,471)x1	1,084	31	LZTR1 (600574), SERPIND1 (142360)	Discrete DD, dyslalia, microphthalmia, bilateral leucoma, bilateral exophthalmos, oblique palpebral fissures, telecanthus, limitations in elbow extension, left equinovarus foot	F/-	ND	-	Distal 22q11.2 deletion syndrome
#1117	Del	arr[hg19] 13q14.13q21.33(47,060,617-72,679,280)x1	25,619	124	51 OMINs	motor delay, dolichocephaly, high palate, strabismus	M/-	ND	46,XY,del(13)(q22q31)	Chromosome 13q14 deletion syndrome

Pathogenic CNVs (PCNVS) found by CMA in the cohort, with the number of genes present in the region, listing some the relevant genes and available phenotypes for each individual. Dup = Duplication, Del = Deletion, CAs = congenital anomaly, DD = developmental delay, mildID = mild intellectual disability, ModID = moderate intellectual disability, SevID = severe intellectual disability, ASD = Autism spectrum disorder, FD = facial dysmorphism, SLD = speech and/or language delay or impairment, IUGR = intrauterine growth restriction, ADHD = Attention-deficit/hyperactivity disorder, LD = learning difficulty, ND = not determined. F = Female, M = Male. VUS = (CNV of uncertain significance). LPCNVs = likely pathogenic CNVs. Under column Gender/"other info": Patient may have 2 PCNVs, or additionally 1 VUS or 1 LPCNVs. All PCNVs are listed in this table. VUS and LPCNVs are listed in another table.

Table of Pathogenic CNVs (PCNV)

Table 3 - Pathogenic CNVs found in the ASD Cohort. Includes ASD cases of the cohort previously published [33].

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#15	Del	arr[hg19] 16p11.2(28,689,085-29,043,863)x1	355	18	SH2B1	DD, ASD	M/ Affected brother (#16)	ND		Distal 16p11.2 Deletion Syndrome
#16	Del	arr[hg19] 16p11.2(28,689,085-29,388,495)x1	362	18	SH2B1	DD, ASD	M/ Affected brother (#15)	ND		Distal 16p11.2 Deletion Syndrome
#52	Del	arr[hg19] 22q13.33(50,788,193-51,115,526)x1	327	18	SHANK3	SevID, ASD, motor difficulties, FD, CAs and epilepsy.	M/	ND		Phelan-McDermid Syndrome
#66	Dup	arr[hg19] 15q25.1q26.3(80,304,866-102,429,040)x3	22,124	175	IGFR1, AKAP13, CPEB1, NTRK3, WDR73	SevID, ASD, convulsions, SLD, hyperactivity, CAs (one kidney) and FD	M/ -	ND		-
#69	Del	arr[hg19] 16p12.2p11.2(21,405,327-29,388,495)x1	7,983	82	SH2B1	MildID, ASD, SLD, hyperactivity and FD	M/ -	ND		-
#70	Dup	arr[hg19] 7q11.23(72,732,834-74,155,067)x3	1,422	27	WBSCR27, WBSCR28	ModID, ASD and hyperactivity	M/ -	ND		Williams-Beuren Region Duplication Syndrome
#76	Dup	arr[hg19] 7q11.23(72,556,215-74,245,599)x3	1,689	34	WBSCR27, WBSCR28	MildID, ASD	M/ -	ND		Williams-Beuren Region Duplication Syndrome
#77	Del	arr[hg19] 15q13.2q13.3(31,073,735-32,446,830)x1	1,373	9	CHNA7	MildID, ASD and hyperactivity	M/ -	ND		-
#148	Dup	arr[hg19] Xp22.3q28(1-247,249,719)x3 ou arr(X)x3	155,270	-	-	DD, ASD and schizophrenia	F/-	ND		Triple X Syndrome
#184	Del	arr[hg19] 15q11.2q13.1(22,770,421-28,823,722)x1	6,053	121	UBE3A, SNRPN	DD, ID, epilepsy, ASD and ADHD	M/-	ND		Angelman Syndrome
#235	Dup	arr[hg19] 17p11.2(16,591,260-20,473,937)x3	3,882	68	RAI	Slender build, DD, SLD, ModID, ASD and FD	F/-	ND		Potocki-Lupski Syndrome

Table of Pathogenic CNVs found in the ASD Cohort.

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#255	Dup	arr[hg19] 22q11.21q11.23(18,493,187-24,313,652)x3	5,820	125	TBX1	DD, ASD and FD	M/-	ND		22q11.21 Duplication Syndrome
#345	Del	arr[hg19] 14q32.2q32.31(100,095,248-102,755,064)x1	2,660	117	PEGS (DLK1 and RTL1), MEGS (MEG3 and MEG8)	Low weight, short stature, prematurity, IUGR, ataxia, scoliosis, DD, SLD, SevID, ASD, FD and early puberty	F/-	ND		Temple Syndrome
#385	Del	arr[hg19] 21q22.12q22.2(35,834,713-39,831,660)x1	3,997	32	DYRK1A	Convulsions, DD, ID, SevID, ASD, cardiomyopathy, CAs (abnormal external genitalia) and thrombocytopenia	M/-	ND		21q22.12 Microdeletion Syndrome
#416	Del	arr[hg19] 18q21.32q23(58,921,746-78,013,728)x1	19,092	75	PIGN	Obesity, CASs, DD, ID, deafness, ASD, FD, and thrombocytopenia	M/-	ND		18 q21.32-qter Deletion Syndrome
#443	Dup	arr[hg19] 22q12.3q13.1(35,888,588-38,692,765)x4	2,804	59	45 OMIMs	Low weight, short stature, DD, SLD, ASD, behavioral disorder DF and mongolian spots	M/-	ND		-
#455	Dup	arr[hg19] Yp11.31p11.2-Yq11.23(2,650,140-28,799,937)x2	26,149	486	39 OMIMs	ASD and tall stature	M/-	ND	47, XY, +mar	XYY-Region Syndrome
#470	Del	arr[hg19] 2q37.3(238,092,121-242,782,258)x1	4,690	73	HDAC4	Asperger's syndrome	F/-	ND		2q37.3 Microdeletion Syndrome
#511	Dup	arr[hg19] 2q11.2(99,222,915-101,919,539)x3	2,696	29	13 OMIMs	ASD, ID, tall stature, CAs and FD	M/1 of 2 pCNVs	ND	47,XY+mar(64%)/48,XY, ++mar(6%)	-
#511	Dup	arr[hg19] 2q11.1q11.2(95,327,873-98,719,140)x4	3,391	52	24 OMIMs	ASD, ID, tall stature, CAs and FD	M//1 of 2 pCNVs	ND	47,XY+mar(64%)/48,XY, ++mar(6%)	-
#586	Del	arr[hg19] 15q21.3(57,289,688-57,510,425)x1	221	1	TCF12	ASD, hyperactivity and FD (Asymmetric facies)	M/-	ND		-
#594	Dup	arr[hg19] 1q32.3q41-1q43q44(212,011,806-249,181,598)x3	36,743	581	169 OMIMs	ASD, ID, CAs and FD	F/1 of 2 pCNVs	ND	46,XX, add(22)(q13)	1q32.3-qterm trisomy
#594	Del	arr[hg19] 22q13.31q13.33(47,771,299-51,197,766)x1	3,426	49	29 OMIMs	ASD, ID, CAs and FD	F/1 of 2 pCNVs	ND	46,XX, add(22)(q13)	Phelan-Mcdermid Syndrome

Table of Pathogenic CNVs found in the ASD Cohort.

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
#667	Del	arr[hg19] 18q12.3(42,453,211-42,988,420)x1	535	3	SETBP1 ,SLC14A2	ASD	M/-	ND		Intellectual developmental disorder, autosomal dominant 29
#714	Del	arr[hg19] 16p11.2(29,591,326-30,190,029)x1	598	31	20 OMIMs	Asperger's syndrome.	M/-	ND		Chromosome 16p.11.2 Deletion Syndrome
#737	Dup	arr[hg19] 16p13.3p12.3(85,880-18,242,713)x3	18,156	342	CREBBP	ASD, FD and CAs	F/1 of 2 pCNVs	ND		Partial Trisomy 16p13.3 Syndrome
#737	Del	arr[hg19] Xq27.3q28(145,443,311-155,233,098)x1	9,723	167	FMR1, AFF2	ASD, FD and CAs	F/1 of 2 pCNVs	ND		-
#751	Del	arr[hg19] 18q12.2q21.1(36,210,635-44,530,609)x1	8,319	28	SETBP1	ASD, FD and CAs	M/-	ND		18q Deletion Syndrome
#791	Del	arr[hg19] 14q12(29,197,241-29,514,397)x1	317	4	FOXG1	ASD, DD, SLD, FD, CAs and seizures	F/-	ND		FOXG1 Syndrome
#809	Dup	arr[hg19] Xq28(153,123,879-153,621,056)x2	497	21	MECP2	ASD and CAs	M/-	ND		MECP2 Duplication Syndrome
#853	Del	arr[hg19] 16p11.2(29,591,326-30,176,508)x1	585	31	20 OMIMs	ASD, SLD, FD, dyslalias and motor difficulties	M/-	ND		Chromosome 16p.11.2 Deletion Syndrome
#873	Del	arr[hg19] 13q33.2q34(105,020,842-115,107,733)x1	10,086	86	EFNB2, LIG4, SOX1	ASD, ID, CAs, FD and microcephaly	F/-	ND		Distal 13q Deletion Syndrome
#913	Dup	arr[hg19] 15q24.1q24.2(72,899,646-75,567,198)x3	2,667	52	32 OMIMs	ASD, FD and CAs	M/-	ND		-
#970	Dup	arr[hg19] 1q21.1q21.2(146,106,723-147,830,830)x3	1,724	56	SATB2	ASD	F/-	ND		1q21.1 Microduplication Syndrome
#1026	Dup	arr[hg19] 2q33.1(200,182,545-201,185,809)x3	1,003	8	SATB2	ASD, ID and DF	M/ potencial UPD:	ND		-

Table of Pathogenic CNVs found in the ASD Cohort.

Case	PCNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	Some of the Relevant Genes	Phenotype	Gender/other info	Inheritance	Karyotype	Syndrome
							22q13.1q13.33 (13.2 Mbp; 37,977,281-51,157,531)			
#1050	Del	arr[hg19] 9p24.3p24.1(208,454-5,222,238)x1	5,013	27	DMRT1, DMRT2, DMRT3	ASD, ID, pectus excavatum and FD	M/-	ND		-
#1100	Del	arr[hg19] 15q13.2q13.3(31,098,690-32,444,261)x1	1,346	18	CHRNA7	Asperger's syndrome	F/-	ND		15q13.3 Microdeletion Syndrome
#1107	Del	arr[hg19] 9p24.3p22.3(208,454-15,424,987) x1	15,216	137	NFIB, FREM1	ASD	M/-	ND	46, XY, del(9)(~p22.2-pter)	9p Deletion Syndrome

Pathogenic CNVs found by CMA in the cohort with ASD, with the number of genes present in the region, listing the most relevant genes and phenotypes for each individual. Dup = Duplication, Del = Deletion, CAs = congenital anomaly, DD = developmental delay, MildID = Mild intellectual disability, ModID = moderate intellectual disability, SevID = severe intellectual disability, ASD = Autism spectrum disorder, FD = facial dysmorphism, SLD = speech and/or language delay or impairment, IUGR = intrauterine growth restriction, ADHD = Attention-deficit/hyperactivity disorder, LDO = learning difficulty only, LD = Learning disability, ND = not determined. F = Female, M = Male. 1 of 2 pCNVs = 1 of 2 patogenic CNVs from one individual.

Table of Pathogenic CNVs found in the ASD Cohort.

Among the 170 individuals with pathogenic CNVs of the whole cohort of 1012 cases, including those previously published by Chaves & coworkers [33], 26 carried more than one PCNV. 19 of them were carriers of 2 PCNVs (cases #33, #47, #61, #127, #251, #331, #332, #372, #407, #501, #511, #594, #651, #687, #737, #739, #786, #861, and #1080). Additionally, 7 cases had 3 pathogenic CNVs (cases #151, #188, #196, #219, #270, #392, and #995). In 3 cases (#81, #255, and #331), a pathogenic CNV was accompanied by VUS.

Out of the 204 pathogenic CNVs, 119 were deletions, resulting in only one copy of the involved sequence, except for case #713. The deletion in this case involved a genomic region of the boy's single X sex chromosome. And six cases (#81, #255, #331, #646, #927 and #1109), along with a pathogenic deletion, also presented VUS.

The other 74 pathogenic CNVs were duplications, which usually result in a total of three copies of the involved sequence, but in eight males (#24, #25, #116, #151, #30, #455, #807 and #809) involved a relevant region of a sex chromosome and resulted in two copies (the main reason for pathogenicity is the fact that in males none of the duplicated copies on X undergoes inactivation, which it does in females) and in five cases (#306, #422, #443, #511 and #620) the CNV found was in a state of four copies. Figure 2 illustrates the frequency and number of pathogenic CNVs found per chromosome.

Figure 2 -Circle ideogram with the pathogenic CNVs and VUS* detected in our study.

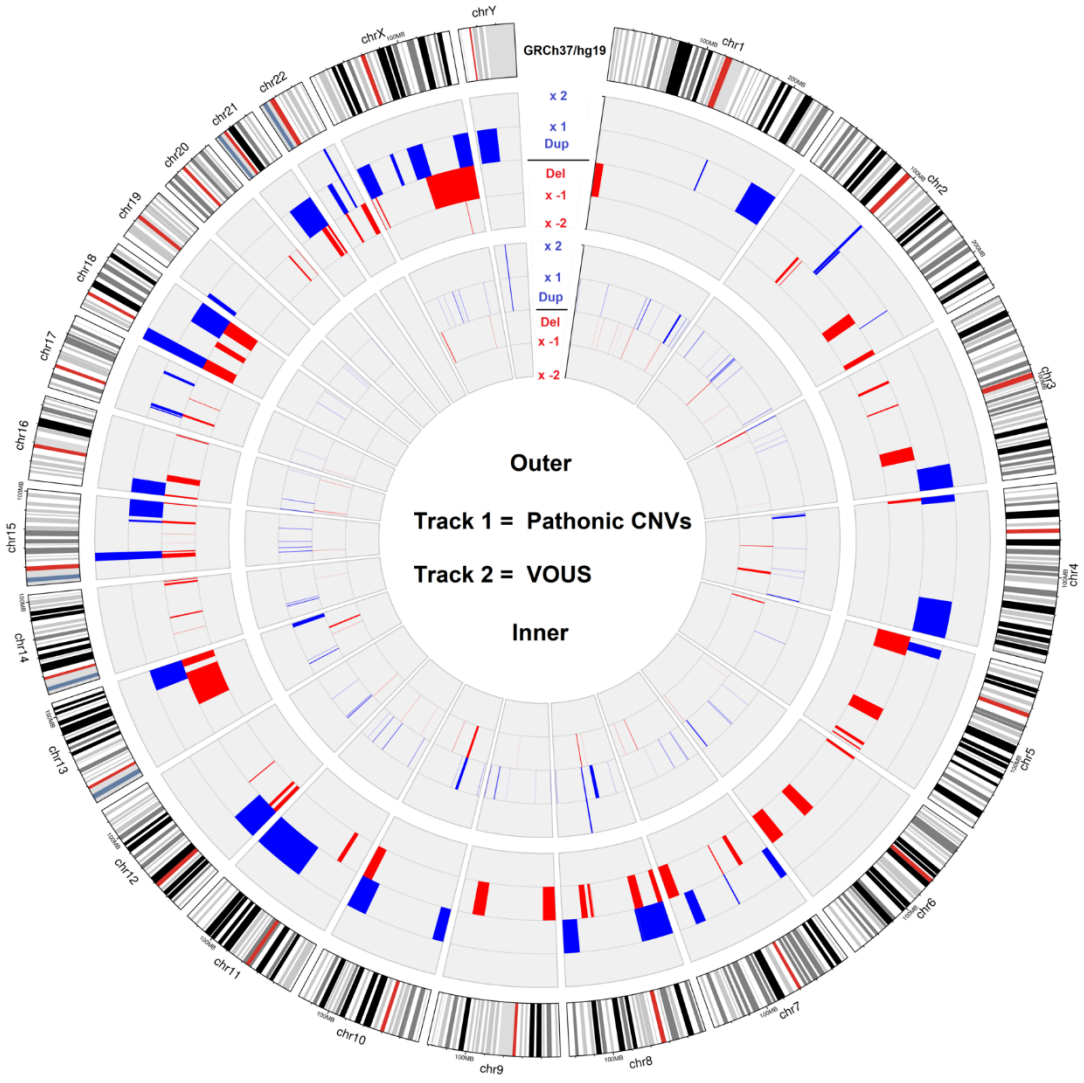


Figure 2: The circle ideogram plot show in the outer track (track1) the genomic positions of all pathogenic CNVs (pCNVs) found in the study per human chromosome. *The VUS (including the LPCNVs) are plotted in inner track (track 2). The bars in blue represent duplication, either for x1 or x2 additional copies. The bars in red represent deletions in x1 (heterozygous) or x2 (homozygous) copies.

Pathogenic CNVs were found on all chromosomes (see supplementary information 1 - Pathogenic CNVs per chromosome), with sizes from 32 Kbp to 71 Mbp (SD = 9992, Mean = 8365) and contained 1 to 581 genes per PCNV (SD = 93, Mean

= 87), of which 1 to 87 (SD = 13, Mean = 9) are genes cited in the OMIM database (OMIM genes) (see supplementary information 2).

Univariate analysis (Fisher's test) indicated the predictive phenotypes for a higher diagnostic outcome (greater chance of having a pathogenic CNV) in our cohort with DNNs: Developmental delay (*p*-value = <0.001, OR = 0.53); Autism Spectrum Disorder (*p*-value = <0.001, OR = 2.18); Facial Malformations/Dysmorphisms (*p*-value = <0.001, OR = 0.42); Upper limb anomalies (*p*-value = <0.001, OR = 0.36); Lower limb anomalies (*p*-value = 0.001, OR = 0.41); genitourinary anomalies and malformations (*p*-value = 0.004, OR = 0.38); Low weight (*p*-value = 0.01, OR = 0.44); Intellectual disability (*p*-value = 0.014, OR = 0.65); Heart anomalies and malformations (*p*-value = 0.018, OR = 0.51); ID or DD (*p*-value = 0.025, OR = 0.65) and Motor development delay (*p*-value = 0.036, OR = 0.54). There was no significantly higher diagnostic result by CMA for the other phenotypes (see supplementary information 3).

Following the scoring system, another 155 rare CNVs were interpreted as 141 Variants of uncertain significance (VUS) (Supplementary Table 1) and 14 as Likely Pathogenic CNVs (LPCNVs) (Table 4), these being the main findings in 13% of the cohort. Of these, 102 are duplications and 53 are deletions. In cases #635, #658, #929 2 VUS were detected and in cases #649, #937, 3 VUS.

These variants were found on most chromosomes except for 21 and 22 (see supplementary information 1 - VUS per chromosome), with sizes from 30 Kbp to 8 Mbp (SD = 1266, Mean = 802) and contained 1 to 87 genes (SD = 13, Mean= 9), of which 1 to 38 (SD = 5 Mean = 5) are genes cited in the OMIM database (OMIM genes) (see supplementary information 2). Figure 2 illustrates the frequency and amount of VUS per chromosome (in track 2). Fourteen VUS, according to the scoring system were found to be LPCNVs (Table 4)

Table 4 Likely Pathogenic CNVs found in the cohort.

Case	CNV	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#1015	Dup	arr[hg19] 1q21.3(153,568,824-154,833,332)x3	1,264	39	26	GATAD2B	ASD, ID and obesity	F/-
#1127	Del	arr[hg19]2q31.2(179,396,924-179,629,278)x1	232	2	2	TTN (*188840)	ASD, epilepsy	M/-
#513	Dup	arr[hg19] 10q11.22q11.23(46,252,072-51,903,756)x3	5,652	61	-	-	ASD and ID	F/-
#519	Del	arr[GRCh37] 9q21.2(79,995,119-80,139,559)x1	144	3	2	VPS13A (605978), GNA14 (604397)	MildID, ADHD	F/-
#547	Dup	arr[GRCh37] 8q12.1q12.3(56,379,919-63,866,456)x3	7,487	43	21	CHD7 (*611238)	ptosis, extrahepatic portosystemic shunt type Ib, patent foramen ovale, left ventricular hypertrophy	M/-
#596	Del	arr[hg19] 8q22.2 (100,067,471-100,622,400)x1	555	3		VPS13B (*607817)	DD, obesity, ID, anxiety,diabetes mellitus	F/-
#597	Del	arr[GRCh37] 12p11.23(27,316,348-27,796,495)x1	480	7	3	PFFIBP1 (*603141)	recurrent otitis, seizure, precocious puberty, SLD, broad forehead, long eyelashes	F/
#633	Del	arr[hg19] 6q26(162,374,660-162,738,968)x1	364	1	1	PARK2	ASD	M/
#823	Del	arr[hg19] 5q34q35.1(165,498,746-169,954,911)x1	4,456	42	29	KCNMB1 (603951)	DD, speech disorder, short frenulum, low weight, short stature, FD, speech delay, consanguineous parents.	F/-
#829	Del	arr[GRCh37] 5p15.31p15.2(9,090,338-11,635,988)x1	2,545	20	8	-	ID, strabismus, protruding ears brother of case #828	M/ + 1 PCNVCNV

Table of Likely Pathogenic CNVs (LPCNVs).

#833	Del	arr[hg19] 1q21.1(145,252,423-145,888,926)x1	637	24	12	-	ASD	M/-
#847	Del	arr[hg19] 1p12p11.2(120,527,347-120,617,367)x1	90	1	1	NOTCH2	ASD and microcephaly	M/-
#852	Del	arr[hg19] 2q13(110,498,141-110,980,295)x1	482	11	3	NPHP1 (*607100)	Auditory processing disorder, LD, microcephaly	M/-
#956	Del	arr[hg19] 14q22.1q22.2(52,412,733-54,387,154)x1	1,974	14	4	ACTR2, Rab-1A	Low weight, short stature, broad forehead, triangular face, everted lips, ogival palate, congenital cardiopathy, SLD	F/-

Likely Pathogenic CNVs (LPCNVs), found in the cohort, with the number of genes present in the region, listing some of the relevant genes and available phenotypes for each case. Dup = Duplication, Del = Deletion, CAs = Congenital Anomalies, DD = Developmental Delay, ID = non-specified Intellectual Disability, mildID = mild Intellectual Disability, ModID = moderate Intellectual Disability, SevID = severe Intellectual Disability, ASD = Autism Spectrum Disorder, FD = Facial Dysmorphisms, SLD = speech and/or language delay/impairment, IUGR = Intrauterine growth restriction, ADHD = Attention-deficit/hyperactivity disorder, LD = learning difficulty, ASD = Autism spectrum disorder, F = Female, M = Male

Table of Likely Pathogenic CNVs (LPCNVs).

All other CNVs were interpreted as either common genetic polymorphisms or benign variants found in all chromosomes, with sizes that varied from 10 Kbp to 24 Mbp (SD= 586, Mean= 298) and contained zero to 227 genes (SD= 8, Mean= 3), of which zero to 144 (SD= 4 Mean= 1) are genes cited in the OMIM database (OMIM genes) (see supplementary information 2).

Diagnostic rate and interpretation of CNVs for cases with ASD

When analyzing separately the 333 CMAs from patients where ASD (including all definitions of the spectrum) was cited as the main reason for referral or as one of several phenotypes of the patient, a total of 3,259 CNVs that met the filtering criteria were detected. Of those 1,494 were duplications and were 1,765 deletions, most of them interpreted as benign. In 35 CMAs no CNVs meeting the filtering criteria were detected. The frequency of the most relevant type of CNV found in each case in the whole cohort and the sub-cohort with ASD is illustrated in figure 3 (A1 and A2). The proportional contribution of each type of CNV per subclass of ASD is illustrated in figure 3B.

In 11% of cases (35/333) we identified a total of 38 rare CNVs that were interpreted as pathogenic (Table 3), 22 deletions and 15 duplications. The particularities of cases #511, #594 and #737, with 2 PCNVs, cases #455 (Y Chromosome), #809 (X chromosome) and cases #443 and #511 (PCNV in a four-copy state) were mentioned before.

In the ASD sub-cohort pathogenic CNVs were found on 14 of the 22 human chromosomes (1, 2, 7, 9, 13, 14, 15, 16, 17, 18, 21, 22, X and Y), with sizes from 221 Kbp to 22 Mbp (SD = 5561, Mean = 4926) and contained 1 to 342 genes (SD = 63, Mean = 60), of which 1 to 83 (SD = 32, Mean = 29) are genes cited in the OMIM database (genes OMIM) (see supplementary information 3).

For individuals affected with syndromic ASD (with dysmorphic features) the diagnostic rate was higher than for the whole ASD cohort, confirmed by univariate analysis 16% (p=0.02, OR 2.43, for pathogenic CNVs) (figure 3C).

Figure 3

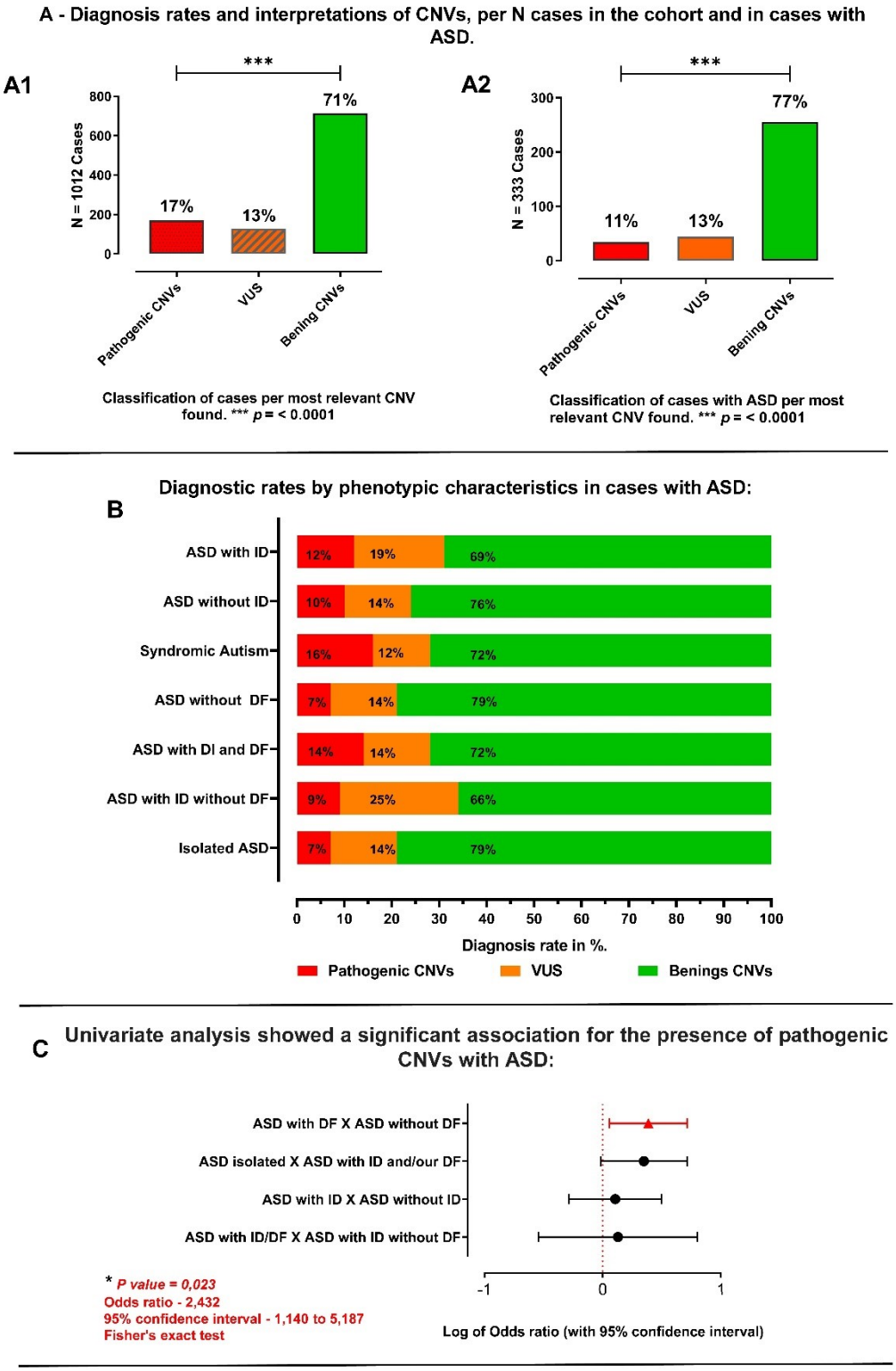


Figure 3: A1: Classification of cases per most relevant CNV found. **A2**: Classification of cases per most relevant CNV found in cases with ASD. **B**: Diagnostic rates per ASD phenotypic category. ASD: Autism spectrum disorder, ID: Intellectual Disability, DF: Dysmorphic Features (syndromic), Classical Autism (Including ASD cases high functioning isolated ASD), Isolated ASD: ASD without ID and without DF. **C**: *Odds ratios* for pathogenic CNVs in the classes of phenotypes. *Odds ratios* shown in log2 scale.

In cases with ASD, DF and ID, the diagnostic rate was 14%, and for ASD with ID, but without DF, it was 12%. For "isolated" ASD, the diagnosis dropped to 7%.

In the 39 cases < 5 years, 5 (13%) had pathogenic CNVs and 6 had only VUS.

For 13% (44/333) of the cases, VUS, which are also rare CNVs, were the only relevant findings, totaling 48 CNVs, 20 deletions and 28 duplications (Supplementary table 1). These variants also were found on most chromosomes, except for chromosomes 4,5,12,18,19,20,21 and 22, with sizes from 10 Kbp to 5.6 Gbp (SD = 1032 Kbp, Mean = 700 Kbp) and contained 1 to 61 genes (SD = 12, Mean= 9), of which 1 to 26 (SD = 5 Mean = 4) are genes cited in the OMIM database (OMIM genes) (see supplementary information 3). In tracks 3 and 4 of the circus ideogram graph (see supplementary information 4), the VUS found per chromosome are plotted.

Four of these VUS (in cases #513, #633, #833 and #1127) were subclassified as LPCNVs, currently without convincing evidence (Table 4).

All other CNVs were interpreted as either benign or common genetic polymorphisms, submicroscopic variants found in all chromosomes, with sizes that varied from 10 Kbp to 24 Gbp (SD = 870, Mean = 228) and contained zero to 181 genes (SD = 9, Mean = 3), of which zero to 96 (SD = 4 Mean = 1) are genes cited in the OMIM database (OMIM genes) (see supplementary information 3).

Long Contiguous Stretches of Homozygosity in the samples

In total, 953 CMA results whose files were available and accessible for the LCSHs study were analyzed. The majority (91%) of CMAs had at least one autosomal LCSH (≥ 3 Mbp), resulting in a total of 3445 LCSH identified in 865 individuals. Only 88 CMAs did not show any LCSH (≥ 3 Mbp). Of the total, 59% (565/953) had only LCSH below 5 Mbp, while 31% (300/953) had one or more LCSH ≥ 5 Mbp.

LCSH leading to suspected UPD

In 27 individuals (~2.8%) of the 953 CMA analyzed, which include 11 previously published cases [28] the LCSH suggested a potential UPD (Table 5 and Fig. 4).

Table 5: Cases with potential UPDs, where a single autosomal chromosome presented LCSH(s) over 3 Mbp, that that alone or in addition of LCSHs ≥3 Mbp reached a size of ≥10 Mbp with no other LSCH over 5 Mbp on any other autosomal chromosome.

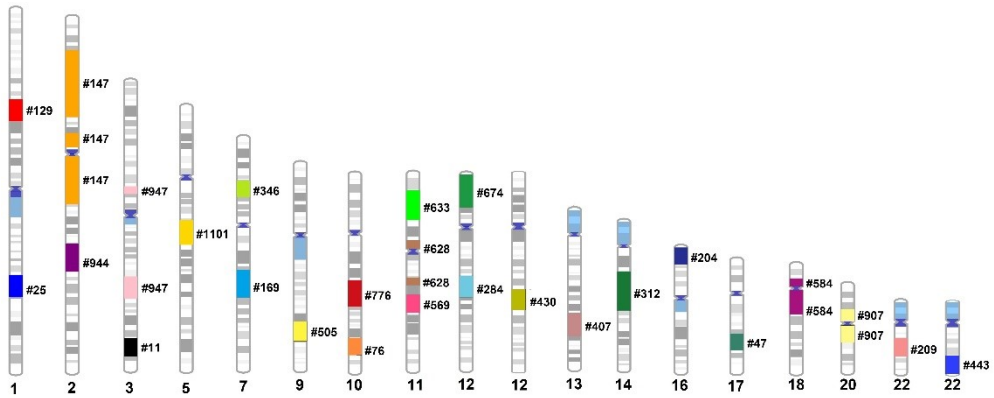
Case	Chr	UPD segment (Isodisomy)	Size Mbp	Other Findings	Phenotype
#25*	1	1q25.3q31.3 (182,537,598-197,949,082)	15.4	Parental origin unknown. 1 PCNV on chrX (causal) *	Male, 16 yrs., DD, ID, SLD, FD, obesity
#129*	1	1p31.3p31.1 (61,620,929-76,755,163)	15.1	Parental origin unknown Without rare CNVs	Male, 4 yrs., DD, SLD, ASD
#147*	2	2p12p11.2 (9.9 Mbp; 79,211,952-89,129,064) & 2q11.1q14.3 (33 Mbp; 95,341,387-128,342,675) & 2p24.1p14 (45.9 Mbp; 22,170,065-68,067,589)	88.8	Parental origin unknown Without rare CNVs	Male, 4 yrs., DD, ASD
#944	2	2q24.1q31.1(155,368,924-174,708,199)	19.3	Parental origin unknown. 1 PCNV on chr 7 (causal) Table 2	Suspected Williams-Beuren Syndrome
#11	3	3q26.32q28(176,695,771-189,044,675)	12.3	Parental origin unknown Without rare CNVs.	Male, 6 yrs., bilateral cleft lip/palate, iris coloboma, blepharophimosis, camptodactyly, patent ductus arteriosus in the past, spina bifida, cerebral ventricle asymmetry.
#947	3	3p13p12.3(5.3 Mbp; 72,016,624-77,325,155) & 3q22.2q25.1(15.4 Mbp; 133,992,740-149,438,082)	20.8	Parental origin unknown Without rare CNVs	Fem, 8 mo., IUGR, DD, FD macrocephaly, short stature, small hands/feet, hypoplastic external genitalia
#1101	5	5q14.1q15(77,967,561-94,997,034)	17	Parental origin unknown Without rare CNVs	Male, 6 years, ASD
#169*	7	7q21.13q31.1(90,678,991-109,653,423)	19	Parental origin unknown 1 PCNV on chr 18 (causal)*	Fem, 9 yrs., FD, learning difficulties, short stature, ophthalmopathies
#346*	7	7p14.3p14.1 (29,374,797-40,699,189)	10.6	Parental origin unknown Without rare CNVs	Male, 15 yrs., DD, severe ID, epilepsy, short stature, absent speech, gastroesophageal reflux and cerebellar atrophy
#833	8	8q13.3q22.1(70,942,228-94,406,882)	23.4	Parental origin unknown 1 LPCNVs on chr 1 Table 3	Male, 2 yrs. 8 mo., ASD.
#505	9	9q31.2q33.1(108,394,893-122,047,673)	13.6	Parental origin unknown Without rare CNVs	Female, 12 years, DD, ID, SLD

Case	Chr	UPD segment (Isodisomy)	Size Mbp	Other Findings	Phenotype
#76*	10	10q25.2q26.13 (112,544,654-124,513,498)	12	1 PCNV on chr 7 (causal)*	Male, 12 yrs., DD, mild ID, ASD, FD
#776	10	10q22.1q23.31(72,616,063-91,065,521)	18.5	Parental origin unknown Without rare CNVs.	Fem, 5 yrs., ASD. Likewise affected sister and brother with ASD, with unremarkable microarray results.
#569	11	11q14.1q21(83,339,664-95,895,139)	12.6	Parental origin unknown Without rare CNVs	Fem, DI, DF, microcephaly, atopic dermatites.
#633	11	11p15.3p13(11,473,107-32,068,176)	20.6	Parental origin unknown 1 LPCNVs on chr 6 (Table 3)	Male 5 yrs., ASD
#628	11	11p11.2p11.12(5.7 Mbp; 45,853,773-51,550,787) & 11q13.4q13.5(5.2 Mbp; 71,543,708-76,752,248)	10.9	Parental origin unknown Without rare CNVs	Fem, 9 yrs., ASD
#674	12	12p13.33p12.1(257,936-22,766,988)	22.5	Parental origin unknown Without rare CNVs	Male, 8 yrs., macroglossia, protruding tongue, laryngeal alterations, closure of the posterior pharynx, laryngotracheomalacia, possible Di George syndrome, peripheral pulmonary stenosis
#284	12	12q15q21.31(69,859,080-84,755,083)	14.9	Parental origin unknown Without rare CNVs Normal karyotype	Male, 10 yrs., DD, ID, FD, obesity, SLD, hypotonia, high palate, clinodactyly, long 2nd and 3rd toes, foot polydactyly, unilateral cryptorchidism, retinitis pigmentosa
#430	12	12q21.2q21.33(78,736,693-92,566,637)	13.8	Parental origin unknown Without rare CNVs	Fem, 4 yrs., DD, FD, short stature, protruding ears, low vision, retinal spot, intracranial calcifications
#407	13	13q22.1q31.3(75,078,803-92,192,744)	17.1	Parental origin unknown, half-sister with Down syndrome, 46,XX, add (21)(q22.3) 2 causal PCNVs on chr 3 and 21 (Table PCNVs)	FD, palatine cleft, upslanting palpebral fissures, Low weight, DD, abnormal growth, seizures, neuropathies, congenital cardiopathy, atrial and ventricular septal defects
#312*	14	14q13.2q23.2 (36,397,727-64,565,981)	28.1	Parental origin unknown 1 PCNV on chr. 22 (causal)*	Male, 11 yrs., SDL, learning disability, FD, abnormal brain structure
#204*	16	16p13.3p13.13 (12.5 Mbp; 89,560-12,548,052)	12.5	Parental origin unknown Without rare CNVs Normal Karyotype	Fem, 2 yrs., IUGR, oligohydramnios, low birth weight, low stature, hypotonia, camptodactyly, DD, SLD, trigonocephaly, epicanthus, downslanting palpebral fissures, atrial septal defect
#47*	17	17q22q24.2 (53,332,043-65,633,600)	12.3	Parental origin unknown	Fem, 8 yrs., FD, abnormal eyelashes, widow's peak,

Case	Chr	UPD segment (Isodisomy)	Size Mbp	Other Findings	Phenotype
				Normal karyotype 1 mosaic PCNV on chr X, (contribution)	supernumerary nipple, short stature, anomalies of upper and lower limbs
#584	18	18p11.22p11.21(5.2 Mbp; 9,990,161-15,143,714) & 18q11.1q12.2(17.5 Mbp; 18,540,834-36,061,962)	22.7	Parental origin unknown 1 VUS chr 4	Male, 1 year and 10 months, DD and macrocephaly
#907	20	20q11.21q13.11(12.5 Mbp; 29,510,307-42,027,093) & 20p12.1p11.1(8.8 Mbp; 17,489,413-26,266,313)	21.3	Parental origin unknown Without rare CNVs	Male, DD, deafness, ocular anomalies and oral cleft.
#209	22	22q12.1q13.1 (26,504,838-40,021,614)	13.5	Parental origin unknown Without rare CNVs	Male, 5 yrs. 8mo., DD, SLD, ID.
#443*	22	22q13.1q13.33 (37,977,281-51,157,531)	13.2	Parental origin unknown 1 PCNV of 2.8 Mbp (x 4), partially overlapping with this probable UPD.*	Male, 2 yrs., low weight, short stature, FD, DD, mongolian spots, poor ear development, SLD, ASD, disturbed behavior, agressive

Identified in previous work *(CHAVES et al., 2019)[33].

Figure 4 - Chromosomal distribution of the 27 cases with LCSH (single or sum) ≥ 10 Mbp restricted to one chromosome, suggesting putative UPDs.



Consanguinity

Analysis of LCSH distributed across multiple chromosomes indicated some degree of inbreeding in 36.5% (348/953) of cases, with over 24% suggesting seventh- to sixth-degree parentage (as third cousins); 7.2%, fifth grade (eg, second cousins); 1.8%, fourth grade (distant first cousins); 1.8%, third degree (first cousin; half-uncle with niece); 0.6%, second-degree (half-siblings, uncle-niece, double cousins) and in

two cases (0.2%) parental kinship suggested incest as it is a coefficient of first-degree inbreeding [father (mother) /daughter (son), full siblings].

Clinically more relevant first-to-fifth-degree kinship was suggested by ~11.5% of cases.

Table 6 - Details the results referring to the 4.3% of cases that suggested kinship from first to fourth grade.

Cases	∑ *of* LCSH (Mbp)	Possible parental relationship	Degree of kinship	Coefficient of inbreeding (F)	LCSH (IBD) expected not tested (~%)
#194	760	father (mother) / daughter (son); complete siblings	First	0.264	25
#834	1.053			0.37	25
#271	334	half-brothers; uncle (aunt)/niece (nephew); double first cousins; grandfather/granddaughter	Second	0.116	12,5
#1068	403			0.14	12,5
#918	285			0.10	12,5
#297	314			0.109	12,5
#380	346			0.121	12,5
#220	402			0.139	12,5
#187	196	first cousins	Third	0.068	6
#275	225			0.078	6
#395	136			0.047	6
#412	123			0.043	6
#413	162			0.056	6
#419	181			0.063	6
#354	193			0.067	6
#364	165			0.057	6
#540	196			0.068	6
#645	238			0.082	6
#730	137			0.047	6
#754	204			0.070	6
#766	136			0.04	6
#823	183			0.063	6
#910	248			0.086	6
#1088	227			0.079	6
#1103	239			0.082	6
#157	62	first cousins once removed	fourth	0.022	3
#273	110			0.038	3
#287	96			0.033	3
#311	82			0.028	3
#378	93			0.032	3
#412	123			0.042	3
#506	68			0.023	3
#546	73			0.025	3
#612	88			0.030	3
#614	81			0.028	3

Cases	∑ *of* LCSH (Mbp)	Possible parental relationship	Degree of kinship	Coefficient of inbreeding (F)	LCSH (IBD) expected not tested (~%)
#663	90			0.031	3
#676	106			0.036	3
#770	75			0.026	3
#789	123			0.042	3
#806	74			0.025	3
#905	66			0.023	3
#1011	79			0.027	3

LCSH with frequency ≥ 5%

Due to the scarcity of information about common LCSH in the Brazilian population in previous work we decided to explore the data from this affected cohort to identify frequent LCSH in the population of Santa Catarina, which we consider to potentially be non-causal for the developmental issues of the patients [26], and now we revise the findings with a larger sample.

The frequency of 5% or more to consider a recurrent LCSH as a common finding in the population of southern Brazil was decided on an empirical basis. This threshold was established to ensure a significant safety margin compared to the 1% threshold used for considering a Single Nucleotide Polymorphism (SNP) as a common variant in the population. This choice was made because analyzing an affected population can introduce bias. However, it is still possible that certain autozygous haplotypes act in conjunction with other genetic variations to manifest the phenotype.

The LCSH identified as frequent, potentially representing regions of low recombination that can maintain ancestral haplotypes identical by descent, are shown in Table 7 and Fig. 5.

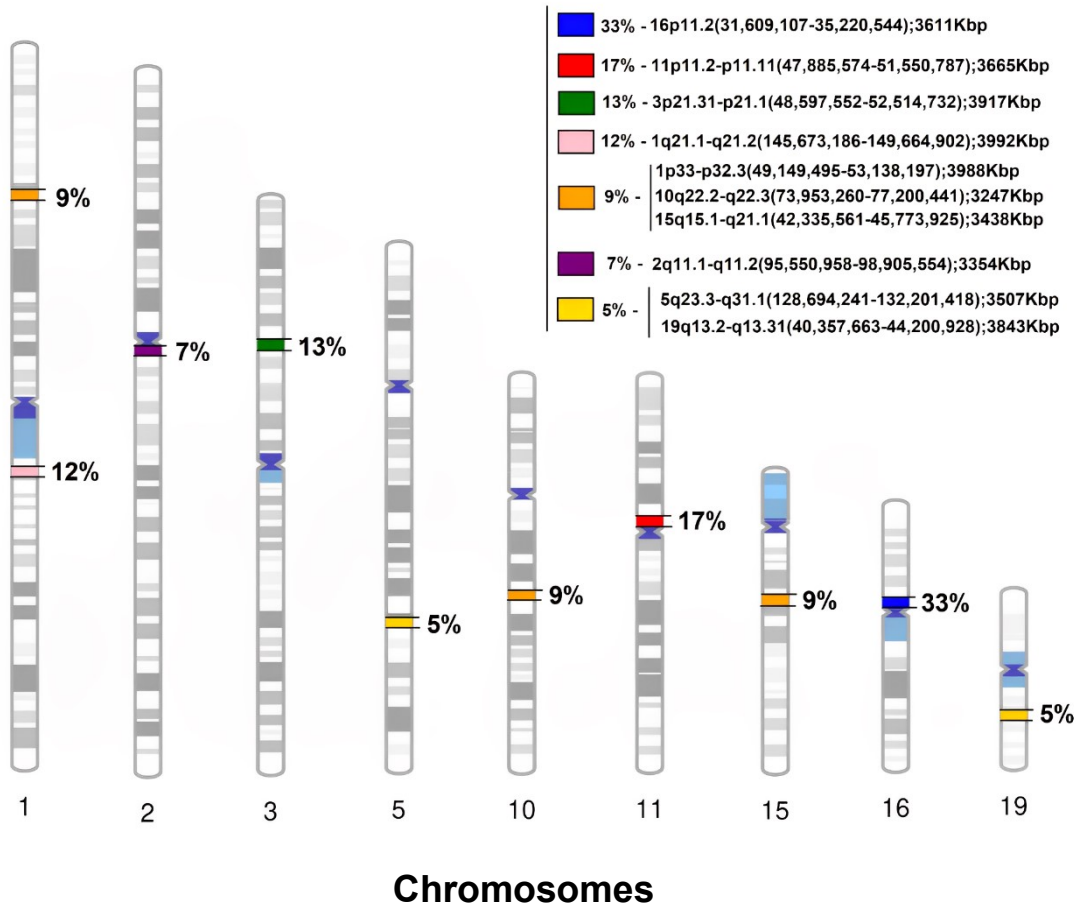
Table 7 - Regions of LCSH considered common (frequency ≥ 5%) identified among 917 CMA results.

Frequencies	Chr/Cytobands	Initial Position	Final Position	Size (Kbp)
33	16p11.2[a,b,c,d,f]	31609107	35220544	3.611
17	11p11.2-p11.11[a,b,c,d]	47885574	51550787	3.665
13	3p21.31-p21.1[a,b,c,d,e]	48597552	52514732	3.917

9	1p33-p32.3[a,c,d]	49149495	53138197	3.988
9	15q15.1-q21.1[a,d]	42335561	45773925	3.438
9	10q22.2-q22.3[a,d]	73953260	77200441	3.247
7	2q11.1-q11.2[a,c,d]	95550958	98905554	3.354
12	1q21.1-q21.2[a,c,d]	145673186	149664902	3.992
5	19q13.2-q13.31	40357663	44200928	3.843
5	5q23.3-q31.1[c]	128694241	132201418	3.507

When the beginning and/or end of the cytobands were variable, a linear position was obtained based on the median of the beginning or end. All analyses, as well as linear positions, were based on the human reference genome, version GRCh37/hg19. a Chaves et al. 2019. b Wang et al. 2015. c Kearney H. M. (Personal communication, 2017). d Sanchez P. (Personal communication, 2017). e Pajusalu et al. 2015. f Neta et al. 2022. The underlined LCSH was only found in our study.

Figure 5 - Visualization of the chromosomal locations of the LCSHs in autosomal chromosomes considered common (frequency ≥ 5%) identified among 917 CMA results.



DISCUSSION

This expanded retrospective cohort study involved 1012 patients with neurodevelopmental disorders (NDDs) and congenital anomalies (CAs) from the state of Santa Catarina. A total of 206 pathogenic copy number variations (CNVs) were identified in 170 individuals, resulting in a diagnostic yield of 17%. This diagnostic yield is almost the same as the 18% obtained in our first study [33] and within the range of 15% to 20% of the diagnostic rate reported in the literature for patients with NDDs [33, 39–51].

It is important to highlight that out of the 173 cases with pathogenic CNVs, 32 cases had a previous abnormal karyotype result, which prompted the CMAs to identify the DNA sequences involved. Excluding the 32 cases with known abnormal karyotypes, the diagnostic rate drops to 14%. The chromosomal microarray (CMA) was essential in discovering altered sequences in abnormal karyotype results, offering unexpected insights into differences compared to what a karyotype suggests. The CMA allows for scrutiny, and sometimes it reveals deletions in chromosomes where the karyotype suggests additions or additions when the karyotype suggested deletions.

In our previous work, we extensively discussed the usefulness of traditional karyotyping as a complement to CMA results, exemplified by 17 cases with altered chromosomal results and a PCNV [33]. We can only underscore the importance of having both classical karyotype results and CMA results. They provide valuable clues about the processes leading to pathogenic changes and are crucial for genetic counselling. Unfortunately, as CMA testing becomes more prevalent, traditional karyotyping is performed less frequently, even though it should at least be conducted for the child and parents when results indicate a pathogenic CNV or a potential UPD.

CNVs

Our analysis revealed pathogenic CNVs across all human chromosomes, with more than one causative variant identified in 15% of individuals. Deletions accounted

for the majority (64%) of all detected pathogenic variants, consistent with the findings of others [52], whereas for VUS the deletions represented only 34%.

Our findings indicate a higher incidence of pathogenic variants on chromosomes 1, 3, 19, and X, with 17, 16, 15, and 18 PCNVs, respectively. This contrasts with the results of previous studies [23–26] (see Supplementary information 1- Pathogenic CNVs per chromosome).

The sizes of the PCNVs, the number of genes they covered, and the number of OMIM genes associated with these CNVs to those of the VUS and non-causative (benign) CNVs, show a statistically significant difference with *P<0.0001* (according to Tukey's Multiple test) (Figure 3 -A1 and Supplementary information 2). This is comprehensible, since larger CNVs, with more genes, in particular with more genes related to disease or known to drive important cellular processes will have a higher impact, which tends to be greater for absence of gene copies than for their excess.

As depicted in the circus ideogram (Figure 02), pathogenic CNVs tend to be situated near telomeres in most chromosomes. This is expected since subtelomeric regions are prone to rearrangements, given that only one chromosomal breakpoint is required to initiate a submicroscopic abnormality [53].

Pathogenic CNVs are also known as recurrent and non-recurrent. While non-recurrent pathogenic CNVs occur sporadically in the genome, with probable origins in replication errors or DNA repair mechanisms, they cover different gene contents and consequently present variable phenotypes [52–54]. Recurrent pathogenic CNVs, in turn, are associated with known and characterized microdeletion and microduplication syndromes. Recurrence of these CNVs is mediated by non-allelic homologous recombination between locus-specific low copy repeats (LCRs) [55, 56].

We have identified a total of 71 individuals with known syndromes that are associated with 72% of pathogenic CNVs. Among them, the most common were Angelman/Prader Willi syndrome, Di George syndrome (0.7%), 1p36 deletion syndrome (0.6%), 16p11.2 deletion syndrome, and Cri Du Chat syndrome (0.5%) (Supplementary Table 2).

Phenotypic characterization

Characterizing phenotypes is a crucial step in investigating the genetic etiologies of developmental disorders, helping to identify the role of the genes involved, as Moeschler and Shevell's (2014)[57] emphasized in their systematic review about the investigation of children with global developmental delay and intellectual disability.

In our cohort, the phenotypic characterization revealed a predominance of phenotypes related to NDs, accounting for 85% of cases, similar to findings reported by others [52, 56, 57], with 83% of the individuals presenting ID and/or DD. In 56% of cases, only DD was present, while ID was mentioned for 33%.

It's worth noting that 42% of the cohort was under 5 years of age, which is below the typical age range for diagnosing ID and eventual deficits are diagnosed as DD. Nevertheless, even considering that many individuals with DD are not necessarily intellectually deficient, it is still possible to estimate the prevalence of Intellectual Disability (ID) by including individuals with both DD and ID, because it is known that most individuals with DD in early childhood will later receive a diagnosis of ID [58].

Along with major neurodevelopmental phenotypes, many individuals exhibit syndromic features (56%), such as congenital anomalies or malformations, and most (47% of all) had atypical facial appearance (facial dysmorphism). Other comorbidities, such as psychiatric or behavioural problems, and variations in physical parameters, like height or body weight, were less frequently reported.

With a larger sample than in our previous study, the univariate analysis confirmed our first findings, showing a significant association for the presence of pathogenic CNVs with autism spectrum disorders, facial malformations/dysmorphisms and genitourinary anomalies/malformations. Obesity and short stature, that were significantly related as second relevant phenotypes when the cohort was smaller [33], lost their significance in the now larger sample. Now developmental delay, intellectual disability, limb anomalies, low weight, heart anomalies/malformations and motor development delay gained in significance (see Supplementary Information 3).

However, even with such an extended sample, there is not one phenotype or group of neurodevelopmental or malformation phenotypes with sufficiently robust evidence as to justify a preferential CMA testing decision. Additionally, we are aware

of our limitations in obtaining standardized phenotype data. This is mainly because there is no standardized phenotype collection and annotation among medical doctors, most of whom are not geneticists and have limited access to genetic tests for follow-up genome sequencing or mutation investigation.

In the State of Santa Catarina, which has approximately 7 million inhabitants, there are only a few (about five) medical geneticists, most of whom practice in Florianópolis, the state capital. Consequently, many patients come from distant areas or are referred for testing by medical doctors outside the main city, without the opportunity to consult with a medical geneticist. A comprehensive and standardized reassessment in all cases, which is currently beyond our capabilities, would be crucial for confidently confirming the phenotype findings and, not to mention, aiding in the interpretation of the CNVs found.

ASD Cases

For the 333 cases of cohort who were diagnosed within the ASD, the ages ranged from a few months to 34 years, with a male predominance of 3,7:1. This is interesting, because when considering the male to female ratio of the whole cohort, the proportion is 1.55:1 and when the cases that mention ASD phenotypes in the clinical description are excluded, the male to female ratio is 1.1:1. We are aware that the cases did not underwent a standardized clinical assessment for ASD. However, the ratio of about 4M:F is well established in the literature, and has led to specific reviews on gender differences in ASD [59–64].

Based on the clinical data which we could obtain, 29% of the individuals (79 aged 5 or more; 17 under 5 years of age) of our ASD cohort also had dysmorphic features (DF), a term that we used to include facial dysmorphia and/or congenital anomalies. When DF were present, we considered them to be syndromic ASD cases, that could have ID or not.

Like the diagnosis of ASD, the diagnosis of ID did not follow a standardized protocol. Some individuals underwent detailed cognitive tests, and others were diagnosed by doctors based of several criteria, this can be seen on tables 1 and 2,

where in most cases only ID is mentioned, without the degree of the ID (mild, moderate, severe). Within the 256 individuals with ASD aged 5 or more, 68 (27%) had some degree of ID. Isolated ASD, which we use to define the non-syndromic patients without ID, comprised 44% (145/333) of the cohort.

According to Rosti et al (2014)[65], approximately 75% of ASD were essential (non-syndromic) cases, whereas 25% are syndromic. Lovrečić et al (2018)[66], reported a proportion of 41% of isolated ASD, 41% with DD and 19% with complex (syndromic) phenotypes when studying a cohort of 150 ASD cases.

There are wide differences within the published prevalence of ID among autistic individuals, Chiurazzi et al (2020)[67] mentions a coexistence of 70% of cases with ASD with ID, while 40% of cases with ID have ASD [68]. The Autism and Developmental Disabilities Monitoring Network (ADDM) funded by the CDC, states that about one third of individuals (35.2%) of the ASD spectrum also have some degree of ID (CDC – Autism Spectrum Disorder, last reviewed December 15, 2022).

There are gender differences among the subclasses of ASD. Whereas the male:female ratio for the whole ASD cohort is 3.8:1, for syndromic ASD it is 2.9:1. In syndromic ASD with ID it is 4.1:1; syndromic ASD w/o ID, 2.3:1. For non-syndromic with ID it is 3:1, and for isolated Autism (non-syndromic w/o ID) it is 4.8:1.

CNVs were found in 90% of the 333 CMAs analysed, and 38 CNVs interpreted as pathogenic were detected in 35 cases with ASD, resulting in a diagnostic yield of 11%, lower than the diagnostic rate for the whole cohort (17%), but within the range of 8 to 22% cited in the literature for other ASD cohorts [16, 66, 69–81]. And without the ASD cases, the diagnostic rate of the cohort increases to 20%.

Within the 35 cases with pathogenic CNVs, 4 were among the 9 patients that had previous abnormal karyotype results, for which the CMA test was requested to identify the DNA sequences involved. Excluding the 4 cases with known abnormal karyotypes, the diagnostic rate drops to 9%, however, the diagnostic yield was considered 10% because the CMA was essential to discover the altered sequences in the abnormal karyotype results.

Recurrent and rare CNVs in ASD

The pathogenic CNVs found in this study and the reported phenotypes of the respective patients are detailed in Table 3. We highlight the genetic syndromes involved with these alterations, which were identified in our cohort, in addition to the most common syndromes in ASD, which involve the chromosomal regions 15q11-q13, 16p11.2 and 22q11.2 [82–88], such as the 15q13.3 Microdeletion Syndrome (#612001), Chromosome 16p.11.2 Deletion Syndrome (OMIM# 611913 ; n= 2), Distal 16p11.2 Deletion Syndrome (#613444) (in 2 cases), Distal 22q11.2 Microduplication Syndrome (# 608363) and Angelman/Prader-Willi Syndrome (*600162).

Also rarer syndromes like 1q21.1 Microduplication Syndrome (#612475), 2q37.3 Microdeletion Syndrome (#600430), Williams-Beuren Region Duplication Syndrome (#609757, n=2), 9p Deletion Syndrome (#158170), Distal 13q Deletion Syndrome (#613884), Temple Syndrome (#616222), Partial Trisomy 16p13.3 Syndrome, Potocki-Lupski Syndrome (#610883), Distal Chromosome 18q Deletion Syndrome (#601808), 18q Deletion Syndrome (#601808), Schinzel Giedion Syndrome (#269150), 21q22.12 Microdeletion Syndrome, 22q13 microdeletion/Phelan-McDermid syndrome (OMIM# 606232; n= 2), MECP2 Duplication Syndrome (#300260), Triple X Syndrome and XYY Region Syndrome have been associated to ASDs.

Among the pathogenic CNVs detected in our study, the ones with the highest frequency in the literature, based on data from the SFARI bank, are the 16p11.2 microdeletion (108 entries), followed by the duplication of 7q11.23 (85 entries), the 16p13 microduplication. 3p12.3 (73 entries), the Xq28 microduplication (59 entries), the 15q11.2q13.1 microdeletion (56 entries), the 22q13.33 microduplication (54 entries), and the 17p11.2 microduplication (45 entries). And identical to the findings of Cheng et al (2015) [89], in our study chromosomes 15, 16 and 22 together contributed to more than 25% of pathogenic CNVs.

Among the rarer findings, based on the SFARI database we have: Case #66, carrying a 22 Mbp microduplication at 15q25.1q26.3(80,304,866-102,429,040), with no SFARI entry for the locus; Case #345, a 2.7 Mbp microdeletion at 14q32.2q32.31(100,095,248-102,755,064), with two entries; the case #385, with a 4 Mbp microdeletion at 21q22.12q22.2(35,834,713-39,831,660), with only one entry; Case #443, carrying a heterozygous microduplication (x4) of 2.8 Mbp at

22q12.3q13.1(35,888,588-38,692,765), with two entries for duplication and 4 for locus deletion; Case #455, which is a 26 Mbp duplication in Yp11.31p11.2-Yq11.23(2,650,140-28,799,937), with 6 entries from a single study [87]; In case #751, with an 8.3 Mpb microdeletion at 18q12.2q21.1(36,210,635-44,530,609), with a single entry; Case #873, a 10 Mpb microdeletion at 13q33.2q34(105,020,842-115,107,733), with 11 entries. And case #1107, with altered karyotype, as previously mentioned, presented a deletion of 15 Mbp in 9p24.3p22.3(208,454-15,424,987), with two entries, one deletion and one duplication.

When it comes to submicroscopic chromosomal alterations, both deletion and duplication of CNVs can result in decreased gene expression by gene disruption, whether gene duplications can also lead to overexpression of genes.

As discussed by Milen et al [60], the detection and interpretation of recurrent CNVs, which are often associated with ASD, facilitates post-test genetic counseling, since one can safely conclude the genetic etiology by associating the CNVs with the clinical characteristics of the patient. In most cases, particularly when the parents are unaffected, it is more likely that pathogenic CNVs have their "de novo" origins. This occurs due to events such as errors during meiotic recombination, early illegitimate mitotic recombinations, or due to repairs to DNA double-stranded breaks during the first divisions of embryonic cells [90].

On the other hand, pathogenic CNVs can also originate from the consequences of a balanced chromosomal translocation in the genome of the parents, according to Nowakowska et al (2016) [91], it is advisable to test the parents of individuals with large pathogenic CNVs, through the classic karyotype, since that balanced translocations cannot be identified by CMA and carry a high risk of recurrence.

Influence of dysmorphic features and/or ID in the diagnostic rate

Although the diagnostic rate for several phenotypic groups was higher than the 11% of diagnostic rate found in the ASD cohort, only the diagnostic yield of 16% for syndromic ASD was confirmed as significant by univariate analysis (p = < 0,05, OR = 2,43)(Figure 3C).

Several studies have investigated the diagnostic yield of CMAs and genome sequencing techniques in cohorts with neurodevelopmental disorders and, even though with a large diagnostic variation when whole genome or exome sequencing is applied, syndromic patients tend to have significatively higher probability for a positive diagnostic result [33, 92, 93]. Specifically for ASD, the mean diagnostic yield is usually lower than for a typical neurodevelopmental cohort. However, among autism subtypes, higher diagnostic usually occurs when ASD is syndromic accompanied with other features and is syndromic (or complex) ASD [69, 94].

LCSHs

In 2006, Li et al, (2006)[34], indicated that LCSH were more common in the human genome than was considered at the time and that they could have an impact on many fields of genetic studies. We now know that LCSH are one of the most common types of genomic traits in humans, being observed throughout the human genome as a consequence of inbreeding or evolutionary forces [22, 26, 95–97].

Previously we described the analysis LCSHs in 430 cases that are part of this cohort [28]. Now, considering the whole cohort, we found that 91% of the individuals have at least one autosomal LCSH ≥ 3 Mbp as revealed by their CMAs tests.

Potential UPDs were found in 2.8% of the CMAs of the cohort, similar to the 2.6% we found in or previous work [28]. The frequency of potential or confirmed UPDs found among published cohorts varies largely among studies. Investigating 214,915 trios, from the 23andMe sequencing dataset, representing a non-clinical general population, the authors found 105 cases of UPD estimating that UPD occurs with an overall prevalence rate of roughly 1 in 2,000 births or 0,05% [98]. The frequency of UPDs found in studies that used exome sequencing of patient-parent trios of large clinical populations for all sorts of genetic conditions is higher and oscillates between 0,2 and 0,6% [99–101]. The investigation for UPDs with whole genome sequencing of 164 parent-child trios in a more selected cohort, an Irish cohort with rare disorders, found 3 UPDs a frequency of 1.8% [100].

Using CMA platforms with distinct SNP density and in clinical populations with distinct ethnic backgrounds, the reported potential UPD rate oscillates from 1 to over 4% [23, 101–103].

We want to emphasize once again that CMA technology can only detect UPD regions in cases of isodisomy; it cannot identify UPDs with total heterodisomy. In a complete UPD, whether it's isodisomic, iso/heterodisomic, or entirely heterodisomic, both homologous chromosomes will exhibit the gender-specific imprinting of the sole transmitting parent across their entire length. It's also important to remember that long, uninterrupted stretches of homozygosity may also result from homologous repair through a breakage-induced DNA replication mechanism, which, in contrast, can originate segmental UPDs [104].

When considering the processes that lead to UPD, it's worth noting that among the 27 cases with LCSH suggesting a potential UPD, eight also had PCNVs that were either considered responsible or partially responsible for their clinical conditions. Additionally, three presented VUS, including two with LPCNVs.

One exception is case #584, which had a PCNV spanning 2.8 Mbp (x4) and overlapped with approximately 1 Mbp of the homozygous region associated with the putative UPD, whose complex origin hints to a real segmental UPD. All other CNVs were located on chromosomes unrelated to the identified UPD. We did not detect any traces of mosaicism involving the affected chromosome in any of the cases, which could have suggested a trisomy rescue.

When a potential UPD is found on one of the chromosomes related to imprinting disorders, like chromosomes 6, 7, 11, 14, 15 or 20, and the phenotype of the patient fits the potential imprinting disorder phenotype, the follow-up is straightforward [60]. However, most often the UPDs are on chromosomes without imprinted regions and sequencing of the isodisomic region should be considered because it often unmasks a homozygous deleterious variant inherited from a heterozygous parent.

Out of the 27 potential UPD cases identified in our study (Table 5 and Fig. 4), only seven were associated with chromosomes known for imprinting disorders [105]. Cases #169 and #346 on chromosome 7, as well as case #312 on chromosome 14, have been previously discussed [28]. Among the cases with potential UPD-like LCSH patterns on chromosome 11, case #633 has a PCNV identified as the causal factor for

its clinical condition, and cases #569 and #628 do not exhibit the hallmark phenotypes typically associated with Beckwith–Wiedemann overgrowth syndrome caused by UPD(11)pat or Silver-Russel Syndrome caused by UPD(11)mat. The same is true for case #907 on chromosome 20, whose available phenotypes do not correlate at all with the imprinting disorders of these chromosome.

Consanguinity

Approximately 24% of the CMAs revealed an LCSH pattern suggesting a distant familial connection (sixth or seventh degree) among the parents of patients affected by NDs. As we've previously mentioned, these findings may be indicative of regional immigration patterns and intermarriage among immigrants in southern Brazil. When the relationship suggested by the LCSH is distant and more associated with the endogamous characteristics of the population, the likelihood of clinical significance decreases.

More significant is the fact that in 11.5% of the CMAs, the LCSHs indicated a first to fifth-degree parental relationship between the parents. These cases are more likely to have a clinical impact because the closer the parentage, the higher the proportion of shared alleles, increasing the risk of inheriting two copies of an autosomal recessive (AR) mutation [24]. We provide an in-depth discussion of the impacts and relevance of these findings in a previous publication [28].

As shown in Table 6, two patients exhibit homozygosity, indicating potential first-degree relatedness among their parents. These results are communicated to the referring physicians by the diagnostic laboratory. It is the responsibility of these physicians to follow the appropriate protocols for these cases.

For one patient (#1068) where a second-degree relatedness is suggested among his parents (Table 6) a PCNV was identified in chr 15 (Table 2). This patient presents a complex syndromic phenotype that extends beyond the typical manifestations usually associated with this deletion, which are mainly related to ASD, DD and behavioural issues, suggesting the participation of a causal autosomal recessive development gene.

LCSH considered common (frequency ≥ 5%)

As extensively discussed in Chaves et al (2019) [28], identifying and knowing the most common (recurrent) LCSH allows us to focus the analysis on the most clinically significant LCSH. Following the same reasoning and criteria of our initial study, in this new analysis, we have identified 10 LCSH ≥ 3 Mbp occurring at a frequency of 5% or higher, thus considering these LCSH as a possible common variation in our population.

All LCSH, except for 19q13.2-q13.31 (40,357,663-44,200,928), which was identified as frequent in our dataset (Table 7) have been previously recognized as common LCSH by other research groups in clinical investigations involving patients with developmental disorders [28, 35–38, 103], including our previous work. These LCSH are typically considered low recombination regions, representing blocks of ancestral haplotypes, and are generally interpreted as potentially non-pathogenic.

Wang et al. (2015)[36] identified several of these regions as recurrent LCSH without clinical relevance in a cohort of patients with NDDs, including unaffected parents. Kearney HM [38] reported them as findings occurring at a frequency > 5% in CMA readings (CytoScan® HD, Affymetrix) from affected individuals. Sanchez P [37] in an analysis of a cohort of 278 affected Hispanics reported LCSH as common when their frequency exceeded 3% in CMA samples (CytoScan® HD, Affymetrix). Neta et al. (2022)[103] reported the region we found on chromosome 16 as occurring at a frequency of 12.7% in a cohort of 100 patients with ID and/or ASD from the Midwest region of Brazil. Pajusalu et al. (2015)[35] reported similar findings to ours on chromosomes 3 and 11 as recurrent LCSH with frequencies of 9.3% and 6%, respectively, using a minimum cutoff size of 5 Mbp, in the investigation of 2110 consecutive Estonian patients (including prenatal care and parenting samples).

In our previous research, we identified as common the regions 6p22.2p22.1 (26,340,871-30,006,805) and 20q11.21q11.23 (31,940,638-36,081,725), also reported as common by Sanchez P [37], Kearney HM [38], and Pajusalu et al. (2015)[35], as well as 7q11.22q11.23 (71,997,278 -76,128,151), that had no prior report. However,

they were not confirmed at a frequency ≥ 5% in this larger sample. Conversely, our previous study did not identify 5q23.3-q31.1 (128,694,241-132,201,418), also found by Kearney HM [38], as frequent. However, in the larger cohort this LCSH now shows up at a frequency above 5%.

We found no previous report of the LCSH in 19q13.2q13.31 (40,357,663-44,200,928) that we identified now. This homozygous region is not associated with any genes known to have an imprinting pattern in humans [106]. It encompasses 148 known genes, out of which 81 are listed in OMIM, including five genes related to autosomal recessive (AR) disorders: Charcot-Marie-Tooth Disease, Type 4F (#614895), Maple Syrup Urine Disease (#248600), Neurodevelopmental disorder with hypotonia, neuropathy, and deafness (#617519), Ethylmalonic Encephalopathy (#602473), and Agammaglobulinemia 3 (#613501).

The LCSH considered frequent and common in the current study not only support the findings and discussions of our previous research but also raise the possibility that our threshold of considering LCSHs only at a frequency ≥ 5% could be too conservative. It might be a relatively safe alternative to consider a lower threshold, such as LCSHs with a frequency above 4% or 3%, as Sanchez P [37] did.

CONCLUSIONS

In this retrospective study, we present the largest report of microarray chromosome data (CMA) in a cohort with neurodevelopmental disorders (NDDs) and/or congenital anomalies (CAs) from Southern Brazil. We achieved a diagnosis rate of 17%, consistent with the literature (15-20%). We characterized the rare copy number variations (CNVs) that we identified and associated them with the main phenotypes presented by each patient. The interpretation of CNVs is challenging and relies on information such as frequency and characterization in affected populations, typically obtained from cohort studies with significant sample sizes.

The primary reasons for referring individuals to CMA testing in this study were developmental delay/intellectual disability and autism spectrum disorder, often

accompanied by syndromic features like congenital anomalies or dysmorphic features. Certain phenotypes have been shown to predict a higher likelihood of carrying a pathogenic CNVs.

For the cases with the ASD, although our diagnostic yield of 10% for ASD is within the range reported in the literature (8-21%), it is higher (16%) when it is syndromic, associated with dysmorphic features, and lower (7%) for "isolated" ASD.

Among the 953 CMAs analysed for contiguous stretches of homozygosity (LCSH), we observed 27 large LCSH (≥ 10 Mbp, ranging from 10.6 to 88.8 Mbp) on a single autosome, suggesting a potential frequency of uniparental disomy (UPD) of 2.8%. However, the limitations of CMA underestimate the true UPD rate, as it can only suggest its presence when uniparental isodisomy is detected. The absence of methylation tests hinders confirming these findings as real UPDs and distinguishing between complete and segmental UPDs.

Regarding consanguinity, the analysis of LCSHs indicated a possible descent from first- to fifth-degree relatives in approximately 11.5% of the cohort. This information is crucial for genetic counseling, as close relatives pose an empirical risk of recurrence, potentially due to autozygous autosomal recessive (AR) mutations. In cases with affected siblings, the analysis of regions that are identical by descent (IBD) can assist in identifying the target region for investigation, particularly when employing whole-exome sequencing (WES).

We identified 10 LCSHs with a frequency above 5% in individuals with NDs. Nine of these LCSH had previously been reported as common variants by other research groups, suggesting that they are likely normal population variants in Santa Catarina. It might be possible that our threshold of considering LCSHs only at a frequency ≥ 5% could be too conservative. While valuable for prioritizing clinically relevant LCSHs for analysis, a clinical contribution of this homozygous regions cannot be completely ruled out.

Overall, analysing LCSHs detected by CMA with high SNP density provides valuable information to aid in the investigation of neurodevelopmental disorders. However, these findings are mostly theoretical and suggestive, serving as guidelines for further investigations such as methylation analysis, targeted gene sequencing, or WES.

LIST OF ABBREVIATIONS

NDs – Neurodevelopmental disorders
ID - Intellectual disability
ASD - Autism spectrum disorder
CMA - Chromosomal microarrays
CNV - Copy number variants
DGV - Database of Genomic Variant
OMIM - Online Mendelian Inheritance in Man
DECIPHER - Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources
VOUS – Variant(s) of uncertain clinical significance
DD- Development delay
CA - Congenital anomaly
IUGR - Intrauterine growth restriction
Mbp - Mega base pairs
LCSH - Long continuous stretches of homozygosity

DECLARATIONS

Ethics approval and consent to participate

The project was submitted and approved by the Research Ethics Committee of the Hospital Infantil Joana de Gusmão, the children hospital of Florianópolis-SC, Brazil, under the Nr 2,339,104, and respects the guidelines and criteria established by the resolution 466/12 of the Brazilian National Health Council. Patients or their caregivers signed the Informed Consent Form to participate in the study. In cases in which it was not possible to contact the patient for any justifiable reason (loss of contact information, mainly) the data was used and a Justification of Absence of Consent was signed by the research team, ensuring the commitment to maintain confidentiality and privacy of the patients whose data and/or information was collected in the records.

Consent for publication

Not applicable

Availability of data and material

The datasets used and/or analyzed during the current study can be requested from the corresponding author on reasonable request. However, since the patients or

their caregivers signed an Informed Consent Form specifying that the data will be used only for the present study, their use for another study necessarily implies a new submission to the ethics committee of the Hospital Infantil Joana de Gusmão and depends on a new approval.

Competing interests

Not applicable.

Funding

TFC were supported with a scholarship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Brazil. The funding body provided financial support but had no other role in the design of the study and collection, analysis, and interpretation of data or in writing of the manuscript.

Authors' contributions

TFC conducted the entire study as part of their doctoral (PhD) work, under guidance from YCNM and AFM. TFC: Were responsible for the study design, data collection, analysis, analysis of statistical and data interpretation of results in LCSHs analysis. TFC, MO, and ITB for data interpretation of results in CNVs. The manuscript was written by TFC, and reviewed by YCNM and AFM.

GRL, JHBF, LLCP, PB: Are collaborating physicians in the study, requested the exams and participated with the clinical aspects of the cases.

All authors read and approved the final manuscript.

Acknowledgements

We thank all families who have agreed to participate in this study. We thank CAPES for their financial support with a post-graduate grant to TFC.

REFERENCES

1. International Classification of Functioning, Disability and Health World Health Organization Geneva ICF ii WHO Library Cataloguing-in-Publication Data International classification of functioning, disability and health : ICF. 2001.

2. Cappuccio G, Vitiello F, Casertano A, Fontana P, Genesio R, Bruzzese D, et al. New insights in the interpretation of array-CGH: autism spectrum disorder and positive family history for intellectual disability predict the detection of pathogenic variants. Ital J Pediatr. 2016;42:39.

3. Abou Jamra R, Wohlfart S, Zweier M, Uebe S, Priebe L, Ekici A, et al. Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. Eur J Hum Genet. 2011;19:1161–6.

4. DSM-5 Diagnostic Classification. In: Diagnostic and Statistical Manual of Mental Disorders. American Psychiatric Association; 2013.

5. Data & Statistics on Autism Spectrum Disorder | CDC. https://www.cdc.gov/ncbddd/autism/data.html. Accessed 25 Nov 2022.

6. SFARI Gene - Welcome. https://gene.sfari.org/. Accessed 27 Nov 2022.

7. Salari N, Rasoulpoor S, Rasoulpoor S, Shohaimi S, Jafarpour S, Abdoli N, et al. The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis. Ital J Pediatr 2022 481. 2022;48:1–16.

8. Idring S, Lundberg M, Sturm H, Dalman C, Gumpert C, Rai D, et al. Changes in Prevalence of Autism Spectrum Disorders in 2001–2011: Findings from the Stockholm Youth Cohort. J Autism Dev Disord. 2015;45:1766–73.

9. Qiu S, Lu Y, Li Y, Shi J, Cui H, Gu Y, et al. Prevalence of autism spectrum disorder in Asia: A systematic review and meta-analysis. Psychiatry Res. 2020;284:112679.

10. Paula CS, Fombonne E, Gadia C, Tuchman R, Rosanoff M. Autism in Brazil: perspectives from science and society. Rev Assoc Med Bras. 2011;57:2–5.

11. Brentani H, Polanczyk GV, Miguel EC. Brazil and Autism. Encycl Autism Spectr Disord. 2021;:734–44.

12. Rylaarsdam L, Guemez-Gamboa A. Genetic Causes and Modifiers of Autism Spectrum Disorder. Front Cell Neurosci. 2019;13.

13. Tammimies K, Li D, Rabkina I, Stamouli S, Becker M, Nicolaou V, et al. Association between Copy Number Variation and Response to Social Skills Training in Autism Spectrum Disorder. Sci Rep. 2019;9.

14. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. Am J Hum Genet. 2014;94:677.
15. Hippolyte L, Maillard AM, Rodriguez-Herreros B, Pain A, Martin-Brevet S, Ferrari C, et al. The Number of Genomic Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. Biol Psychiatry. 2016;80:129–39.
16. Calderoni S, Ricca I, Balboni G, Cagiano R, Cassandrini D, Doccini S, et al. Evaluation of Chromosome Microarray Analysis in a Large Cohort of Females with Autism Spectrum Disorders: A Single Center Italian Study. J Pers Med. 2020;10:1–18.
17. Lee CL, Chuang CK, Tu RY, Chiu HC, Lo YT, Chang YH, et al. Increased Diagnostic Yield of Array Comparative Genomic Hybridization for Autism Spectrum Disorder in One Institution in Taiwan. Medicina (B Aires). 2022;58.
18. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46:881–5.
19. Song T, Xu Y, Li Y, Jia L, Zheng J, Dang Y, et al. Detection of submicroscopic chromosomal aberrations by chromosomal microarray analysis for the prenatal diagnosis of central nervous system abnormalities. J Clin Lab Anal. 2020;34:e23434.
20. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am J Hum Genet. 2010;86:749–64.
21. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22:245–57.
22. Yang HC, Chang LC, Huggins RM, Chen CH, Mullighan CG. LOHAS: loss-of-heterozygosity analysis suite. Genet Epidemiol. 2011;35:247–60.
23. Alabdullatif MA, Al Dhaibani MA, Khassawneh MY, El-Hattab AW. Chromosomal microarray in a highly consanguineous population: diagnostic yield, utility of regions of homozygosity, and novel mutations. Clin Genet. 2011;91:616–22.

24. Kearney HM, Kearney JB, Conlin LK. Diagnostic implications of excessive homozygosity detected by SNP-Based microarrays: Consanguinity, uniparental disomy, and recessive single-gene mutations. Clin Lab Med. 2011;31:595–613.
25. Spence JE, Perciaccante RG, Greig GM, Willard HF, Ledbetter DH, Hejtmancik JF, et al. Uniparental disomy as a mechanism for human genetic disease. Am J Hum Genet. 1988;42:217–26.
26. Iourov IY, Vorsanova SG, Korostelev SA, Zelenova MA, Yurov YB. Long contiguous stretches of homozygosity spanning shortly the imprinted loci are associated with intellectual disability, autism and/or epilepsy. Mol Cytogenet. 2015;8:77.
27. Hunter N. Meiotic Recombination: The Essence of Heredity. Cold Spring Harb Perspect Biol. 2015;7.
28. Chaves TF, Oliveira LF, Ocampos M, Barbato IT, De Luca GR, Barbato Filho JH, et al. Long contiguous stretches of homozygosity detected by chromosomal microarrays (CMA) in patients with neurodevelopmental disorders in the South of Brazil. BMC Med Genomics. 2019;12.
29. Gardner RJ., Sutherland GR, Shaffer LG. Chromosome Abnormalities and Genetic Counseling. Oxford University Press; 2011.
30. Engel E. A fascination with chromosome rescue in uniparental disomy: Mendelian recessive outlaws and imprinting copyrights infringements. European Journal of Human Genetics. 2006;14:1158–69.
31. Kotzot D. Complex and segmental uniparental disomy updated. J Med Genet. 2008;45:545–56.
32. Liehr T. Cytogenetic contribution to uniparental disomy (UPD). Mol Cytogenet. 2010;3:8.
33. Chaves TF, Baretto N, Oliveira LF de, Ocampos M, Barbato IT, Anselmi M, et al. Copy Number Variations in a Cohort of 420 Individuals with Neurodevelopmental Disorders From the South of Brazil. Sci Rep. 2019;9:1–20.
34. Li L-H, Ho S-F, Chen C-H, Wei C-Y, Wong W-C, Li L-Y, et al. Long contiguous stretches of homozygosity in the human genome. Hum Mutat. 2006;27:1115–21.

35. Pajusalu S, Žilina O, Yakoreva M, Tammur P, Kuuse K, Mölter-Väär T, et al. The Diagnostic Utility of Single Long Contiguous Stretches of Homozygosity in Patients without Parental Consanguinity. Mol Syndromol. 2015;6:135–40.
36. Wang J-C, Ross L, Mahon LW, Owen R, Hemmat M, Wang BT, et al. Regions of homozygosity identified by oligonucleotide SNP arrays: evaluating the incidence and clinical utility. Eur J Hum Genet. 2015;23:663–71.
37. Sanchez P. Common AOH Hispanics. (personal communication). 2017.
38. Kearney HM. Common AOH blocks-Fullerton (personal communication). 2017.
39. Roselló M, Martínez F, Monfort S, Mayo S, Oltra S, Orellana C. Phenotype profiling of patients with intellectual disability and copy number variations. Eur J Paediatr Neurol. 2014;18:558–66.
40. Bartnik M, Szczepanik E, Derwińska K, Wiśniowiecka-Kowalnik B, Gambin T, Sykulski M, et al. Application of array comparative genomic hybridization in 102 patients with epilepsy and additional neurodevelopmental disorders. Am J Med Genet Part B Neuropsychiatr Genet. 2012;159B:760–71.
41. Coutton C, Dieterich K, Satre V, Vieville G, Amblard F, David M, et al. Array-CGH in children with mild intellectual disability: a population-based study. Eur J Pediatr. 2015;174:75–83.
42. Preiksaitiene E, Molyte A, Kasnauskiene J, Ciuladaite Z, Utkus A, Patsalis PC, et al. Considering specific clinical features as evidence of pathogenic copy number variants. Journal of Applied Genetics. 2014;55:189–96.
43. Lay-Son G, Espinoza K, Vial C, Rivera JC, Guzm??n ML, Repetto GM. Chromosomal microarrays testing in children with developmental disabilities and congenital anomalies. J Pediatr (Rio J). 2015;91:189–95.
44. Chen CP, Hwang KS, Su HY, Lin SP, Su YN, Chern SR, et al. Prenatal diagnosis and molecular cytogenetic characterization of a de novo interstitial duplication of 14q (14q31.3→q32.12) associated with abnormal maternal serum biochemistry. Taiwan J Obstet Gynecol. 2013;52:125–8.
45. García-Acero M, Suárez-Obando F, Gómez-Gutiérrez A. CGH analysis in Colombian patients: findings of 1374 arrays in a seven-year study. 2018;11:46.

46. Pereira RR, Pinto lP, Minasi LB, De Melo AV, Da Cruz E Cunha DM, Cruz AS, et al. Screening for intellectual disability using high-resolution CMA technology in a retrospective cohort from Central Brazil. PLoS One. 2014;9:e103117.
47. Krepischi AC V., Villela D, da Costa SS, Mazzonetto PC, Schauren J, Migliavacca MP, et al. Chromosomal microarray analyses from 5778 patients with neurodevelopmental disorders and congenital anomalies in Brazil. Sci Reports 2022 121. 2022;12:1–11.
48. Pfundt R, Kwiatkowski K, Roter A, Shukla A, Thorland E, Hockett R, et al. Clinical performance of the CytoScan Dx Assay in diagnosing developmental delay/intellectual disability. Genet Med. 18.
49. Anazi S, Maddirevula S, Faqeih E, Alsedairy H, Alzahrani F, Shamseldin HE, et al. Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol Psychiatry. 2017;22:615–24.
50. Kashevarova AA, Nazarenko LP, Skryabin NA, Salyukova OA, Chechetkina NN, Tolmacheva EN, et al. Array CGH analysis of a cohort of Russian patients with intellectual disability. Gene. 2014;536:145–50.
51. Fan Y, Wu Y, Wang L, Wang Y, Gong Z, Qiu W, et al. Chromosomal microarray analysis in developmental delay and intellectual disability with comorbid conditions. BMC Med Genomics. 2018;11.
52. Perovic D, Damnjanovic T, Jekic B, Dusanovic-Pjevic M, Grk M, Djuranovic A, et al. Chromosomal microarray in postnatal diagnosis of congenital anomalies and neurodevelopmental disorders in Serbian patients. J Clin Lab Anal. 2022;36.
53. Luo Y, Hermetz KE, Jackson JM, Mulle JG, Dodd A, Tsuchiya KD, et al. Diverse mutational mechanisms cause pathogenic subtelomeric rearrangements. Hum Mol Genet. 2011;20:3769.
54. Peng G, Zhou Q, Chai H, Wen J, Zhao H, Taylor HS, et al. Estimation on risk of spontaneous abortions by genomic disorders from a meta-analysis of microarray results on large case series of pregnancy losses. Mol Genet Genomic Med. 2023;00:e2181.
55. Chen C-P, Chang S-Y, Wang L-K, Chang T-Y, Chern S-R, Wu P-S, et al. Prenatal diagnosis of a familial 15q11.2 (BP1-BP2) microdeletion encompassing TUBGCP5, CYFIP1, NIPA2 and NIPA1 in a fetus with ventriculomegaly, microcephaly and

intrauterine growth restriction on prenatal ultrasound. Taiwan J Obstet Gynecol. 2018;57:730–3.
56. Yuan H, Shangguan S, Li Z, Luo J, Su J, Yao R, et al. CNV profiles of Chinese pediatric patients with developmental disorders. Genet Med. 2021;23:669–78.
57. Moeschler JB, Shevell M, Saul RA, Chen E, Freedenberg DL, Hamid R, et al. Clinical Report: Comprehensive Evaluation of the Child With Intellectual Disability or Global Developmental Delays. Pediatrics. 2014;134:e903.
58. Uddin M, Thiruvahindrapuram B, Walker S, Wang Z, Hu P, Lamoureux S, et al. A high-resolution copy-number variation resource for clinical and population genetics. Genetics in Medicine. 2015;17:747–52.
59. Werling DM, Geschwind DH. Sex differences in autism spectrum disorders. Curr Opin Neurol. 2013;26:146.
60. Ferri SL, Abel T, Brodkin ES. Sex Differences in Autism Spectrum Disorder: a Review. Curr Psychiatry Rep. 2018;20.
61. Beggiato A, Peyre H, Maruani A, Scheid I, Rastam M, Amsellem F, et al. Gender differences in autism spectrum disorders: Divergence among specific core symptoms. Autism Res. 2017;10:680–9.
62. Williams OOF, Coppolino M, Perreault ML. Sex differences in neuronal systems function and behaviour: beyond a single diagnosis in autism spectrum disorders. Transl Psychiatry. 2021;11.
63. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. J Am Acad Child Adolesc Psychiatry. 2017;56:466–74.
64. James WH, Grech V. Potential explanations of behavioural and other differences and similarities between males and females with autism spectrum disorder. Early Hum Dev. 2020;140.
65. Rosti RO, Sadek AA, Vaux KK, Gleeson JG. The genetic landscape of autism spectrum disorders. Dev Med Child Neurol. 2014;56:12–8.
66. Lovrečić L, Rajar P, Volk M, Bertok S, Gnidovec Stražišar B, Osredkar D, et al. Diagnostic efficacy and new variants in isolated and complex autism spectrum disorder using molecular karyotyping. J Appl Genet. 2018;59:179–85.

67. Chiurazzi P, Kiani AK, Miertus J, Barati S, Manara E, Paolacci S, et al. Genetic analysis of intellectual disability and autism. Acta Bio Medica Atenei Parm. 2020;91 Suppl 13:1–5.
68. Xiong J, Chen S, Pang N, Deng X, Yang L, He F, et al. Neurological diseases with autism spectrum disorder: Role of ASD risk genes. Front Neurosci. 2019;13 APR.
69. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. JAMA - J Am Med Assoc. 2015;314:595–903.
70. Zarrei M, Burton CL, Engchuan W, Young EJ, Higginbotham EJ, MacDonald JR, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. NPJ Genomic Med. 2019;4.
71. Costa CIS, da Silva Montenegro EM, Zarrei M, de Sá Moreira E, Silva IMW, de Oliveira Scliar M, et al. Copy number variations in a Brazilian cohort with autism spectrum disorders highlight the contribution of cell adhesion genes. Clin Genet. 2022;101:134–41.
72. Sys M, Van Den Bogaert A, Roosens B, Lampo A, Jansen A, Wouters S, et al. Can clinical characteristics be criteria to perform chromosomal microarray analysis in children and adolescents with autism spectrum disorders? Minerva Pediatr. 2018;70:225–32.
73. Chehbani F, Tomaiuolo P, Picinelli C, Baccarin M, Castronovo P, Scattoni ML, et al. Yield of array-CGH analysis in Tunisian children with autism spectrum disorder. Mol Genet Genomic Med. 2022;10.
74. Yang EH, Shin YB, Choi SH, Yoo HW, Kim HY, Kwak MJ, et al. Chromosomal Microarray in Children With Developmental Delay: The Experience of a Tertiary Center in Korea. Front Pediatr. 2021;9:944.
75. Cheng SSW, Chan KYK, Leung KKP, Au PKC, Tam WK, Li SKM, et al. Experience of chromosomal microarray applied in prenatal and postnatal settings in Hong Kong. Am J Med Genet Part C Semin Med Genet. 2019;181:196–207.
76. Munnich A, Demily C, Frugère L, Duwime C, Malan V, Barcia G, et al. Impact of on-site clinical genetics consultations on diagnostic rate in children and young adults with autism spectrum disorder. Mol Autism. 2019;10.

77. Lok WY, Kong CW, Hui SYA, Shi MM, Choy KW, To WK, et al. Chromosomal abnormalities and neurological outcomes in fetal cerebral ventriculomegaly: a retrospective cohort analysis. Hong Kong Med J. 2021;27:428–36.
78. Harris HK, Sideridis GD, Barbaresi WJ, Harstad E. Pathogenic yield of genetic testing in autism spectrum disorder. Pediatrics. 2020;146.
79. Capkova P, Srovnal J, Capkova Z, Staffova K, Becvarova V, Trkova M, et al. MLPA is a practical and complementary alternative to CMA for diagnostic testing in patients with autism spectrum disorders and identifying new candidate CNVs associated with autism. https://doi.org/10.7717/peerj.6183.
80. Ho KS, Twede H, Vanzo R, Harward E, Hensel CH, Martin MM, et al. Clinical Performance of an Ultrahigh Resolution Chromosomal Microarray Optimized for Neurodevelopmental Disorders. Biomed Res Int. 2016;2016.
81. Roberts JL, Hovanes K, Dasouki M, Manzardo AM, Butler MG. Chromosomal Microarray Analysis of Consecutive Individuals with Autism Spectrum Disorders or Learning Disability Presenting for Genetic Services. Gene. 2014;535:70.
82. Chang YS, Lin CY, Huang HY, Chang JG, Kuo HT. Chromosomal microarray and whole-exome sequence analysis in Taiwanese patients with autism spectrum disorder. Mol Genet Genomic Med. 2019;7:996.
83. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong Association of De Novo Copy Number Mutations with Autism. Science. 2007;316:445.
84. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S, et al. Novel Submicroscopic chromosomal abnormalities detected in Autism Spectrum Disorder. Biol Psychiatry. 2008;63:1111.
85. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. Am J Hum Genet. 2008;82:477.
86. Alhazmi S, Alzahrani M, Farsi R, Alharbi M, Algothmi K, Alburae N, et al. Multiple Recurrent Copy Number Variations (CNVs) in Chromosome 22 Including 22q11.2 Associated with Autism Spectrum Disorder. Pharmgenomics Pers Med. 2022;15:705.
87. Wenger TL, Miller JS, DePolo LM, De Marchena AB, Clements CC, Emanuel BS, et al. 22q11.2 duplication syndrome: elevated rate of autism spectrum disorder and need for medical screening. Mol Autism. 2016;7.

88. Budisteanu M, Papuc SM, Streata I, Cucu M, Pirvu A, Serban-Sosoi S, et al. The Phenotypic Spectrum of 15q13.3 Region Duplications: Report of 5 Patients. Genes (Basel). 2021;12.
89. Li M, Long C, Yang L. Hippocampal-Prefrontal Circuit and Disrupted Functional Connectivity in Psychiatric and Neurodegenerative Disorders. Biomed Res Int. 2015;2015:1–10.
90. Wang J, Fan HC, Behr B, Quake SR. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell. 2012;150:402–12.
91. Bartnik M, Nowakowska B, Derwińska K, Wiśniowiecka-Kowalnik B, Kedzior M, Bernaciak J, et al. Application of array comparative genomic hybridization in 256 patients with developmental delay or intellectual disability. J Appl Genet. 2014;55:125–44.
92. Oğuz S, Arslan UE, Kiper P, Alikaşifoğlu M, Boduroğlu K, Utine GE. Diagnostic yield of microarrays in individuals with non-syndromic developmental delay and intellectual disability. J Intellect Disabil Res. 2021;65:1033–48.
93. Martínez F, Caro-Llopis A, Roselló M, Oltra S, Mayo S, Monfort S, et al. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. J Med Genet. 2017;54:87–92.
94. Ho KS, Wassman ER, Baxter AL, Hensel CH, Martin MM, Prasad A, et al. Chromosomal Microarray Analysis of Consecutive Individuals with Autism Spectrum Disorders Using an Ultra-High Resolution Chromosomal Microarray Optimized for Neurodevelopmental Disorders. Int J Mol Sci. 2016;17.
95. Wiszniewska J, Bi W, Shaw C, Stankiewicz P, Kang S-HL, Pursley AN, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. Eur J Hum Genet. 2014;22:79–87.
96. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature. 2010;464:704–12.
97. Vorsanova SG, Yurov YB, Soloviev I V, Iourov IY. Molecular cytogenetic diagnosis and somatic genome variations. Curr Genomics. 2010;11:440–6.

98. Nakka P, Pattillo Smith S, O'Donnell-Luria AH, McManus KF, Agee M, Auton A, et al. Characterization of Prevalence and Health Consequences of Uniparental Disomy in Four Million Individuals from the General Population. Am J Hum Genet. 2019;105:921–32.

99. Wang L, Liu P, Bi W, Sim T, Wang X, Walkiewicz M, et al. Contribution of uniparental disomy in a clinical trio exome cohort of 2675 patients. Mol Genet genomic Med. 2021;9.

100. Molloy B, Jones ER, Linhares ND, Buckley PG, Leahy TR, Lynch B, et al. Uniparental disomy screen of Irish rare disorder cohort unmasks homozygous variants of clinical significance in the TMCO1 and PRKRA genes. Front Genet. 2022;13:945296.

101. Sasaki K, Mishima H, Miura K, Yoshiura K. Uniparental disomy analysis in trios using genome-wide SNP array and whole-genome sequencing data imply segmental uniparental isodisomy in general populations. Gene. 2013;512:267–74.

102. Bruno DL, Ganesamoorthy D, Schoumans J, Bankier A, Coman D, Delatycki M, et al. Detection of cryptic pathogenic copy number variations and constitutional loss of heterozygosity using high resolution SNP microarray analysis in 117 patients referred for cytogenetic analysis and impact on clinical practice. J Med Genet. 2009;46:123–31.

103. Neta AR de C, Pinto IP, Da Cruz AS, Da Cruz AD, Minasi LB. Longos trechos contíguos em homozigose identificados por análise cromossômica por microarranjos em uma população com deficiência intelectual e transtorno do espectro autista do Brasil Central. Conjecturas. 2022;2022:355–67.

104. Sakofsky CJ, Malkova A. Break induced replication in eukaryotes: mechanisms, functions, and consequences. Crit Rev Biochem Mol Biol. 2017;52:395–413.

105. Chao Y, Qin Y, Zou X, Wang X, Hu C, Xia F, et al. Promising therapeutic aspects in human genetic imprinting disorders. Clin Epigenetics 2022 141. 2022;14:1–19.

106. Jirtle R, Randy L. Jirtle. Geneimprint : Genes. 2012. http://www.geneimprint.com/site/genes-by-species. Accessed 24 Feb 2018.

SUPPLEMENTARY TABLE 1

Table of VUS found in the cohort (INCLUDING CASES PUBLISHED IN chaves, et al, 2019).

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#1	Dup	arr[hg19] 2p24.1(23,982,758-24,813,485)x3	831	18	7	ATAD2B, UBXN2A	MildID, overweight	F/-
#5	Dup	arr[hg19] 6q15(89,917,335-90,485,874)x3	568	7	4	GABRR1, GABRR2	MID, Mot Dif and hyperactivity	**M**
#6	Dup	arr[hg19] 2q37.2q37.3(236,733,535-237,355,774)x3	622	4	2	AGAP1, GBX2	DD, convulsions and FD	**F**
#7	Dup	arr[hg19] 1q44(246,324,898-246,688,599)x3	363	2	1	SMYD3	ASD, motor difficulties, convulsions and FD	M/-
#13	Del	arr[hg19] 11q14.1(84,050,388-84,415,990)x1	365	1	1	DLG2	ASD, LDO, motor difficulties, FD and SLD	M/-
#19	Dup	arr[hg19] 8q21.13(82,061,218-84,515,685)x4	2,454	10	6	IMPA1	DD, FD, gastroschisis bladder exstrophy, hydronephrosis and Abnormal growth	**M/ Affected brother (#18)**
#21	Del	arr[hg19] 2q13(110,504,318-111,365,996)x1	861	16	3	NPHP1	ID	**M**
#32	Dup	arr[hg19] 20q13.33	200	9	5	KCNQ2, CHRNA1	Convulsions, low weight, prematurity, FD,	
#40	Del	arr[hg19] 14q24.2(73,590,938-73,776,190)x1	185	4	2	PSEN1, NUMBP1	ASD and SLD	M/-
#43	Del	arr[hg19] 16q23.2(80,260,131-80,701,060)x1	440	2	1	DYNLRB2, CDYL2	MildID, ASD, motor difficulties, SLD, hyperactivity and FD	M/-
#50	Del	arr[hg19] 13q12.12(60,425,635-60,688,042)x1	262	25	2	SGCG, SACS	MID	**F**
#58	Dup	arr[hg19] 11q22.3(102,946,063-103,827,049)x3	880	4	2	DYNC2H1	DD, LDO, Mot Dif and FD	**M**
#64	Dup	arr[hg19] 9q34.3(139,381,821-140,086,032)x3	704	48	28	NOTCH1	DD, SLD, ID and FD	**M**
#81	Dup	arr[hg19] 16p13.3(549,826-1,449,862)x3	900	45	26	CACNA1H	SLD, convulsions and FD	**M/PCNV**
#82	Dup	arr[hg19] 4q35.2(188,106,543-189,797,261)x3	1,691	5	1	ZFP42	DD and SLD	**M**

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#86	Del	arr[hg19] 13q21.2(60,425,635-60,688,042)x1	262	2	1	DIAPH3	DD and LDO	F/-
#89	Dup	arr[hg19] 9p24.3(319,876-517,446)x3	198	2	2	DOCK8 , KANK1	ASD, SLD, motor difficulties and FD	F/-
#109	Dup	arr[hg19] 4q31.1(139,758,054-139,988,340)x3	230	2	1	CCRN4L	DD and FD	**M**
#112	Dup	arr[hg19] 9p13.3(34,211,157-34,395,294)x3	184	5	3	UBAP1, NUDT2	SID, Aut, convulsions, SLD, mot dif and FD	**M**
#117	Dup	arr[hg19] 19q13.33(48,206,212-48,431,081)x3	224,869	25	7	CORD2		**F/-**
#136	Dup	arr[hg19] 4q28.1q28.2(128,789,028-128,891,808)x3	102,78	3	2	PLK4	Low weight, short stature, IUGR, FD, thin hair	**F/-**
#138	Dup	arr[hg19] 6p21.2(37,609,169-37,868,513)x3	259,344	2	2	MDGA1	Prematurity, DD, polydactyly, aggression,	**M/-**
#144	Del	arr[hg19] 8q13.1q13.2(67,999,679-68,190,627)x1	190,948	2	2	CSPP1	DD, SLD, ID and FD	**F/-**
#178	Dup	arr[hg19] 11q23.3(117,000,284-117,312,611)x3	312,327	10	7	DSCAML1, CEP164, BACE1	ASD, FD, macrocephaly	M/-
#180	Del	arr[hg19] 16p13.3(6,243,228-6,835,898)x1	592,67	1	1	RBFOX1	DD, hypothyroidism	**M/-**
#215	Del	arr[hg19] 3q26.33(179,508,262-179,621,954)x1	113,692	1	1	PEX5L	Motor delay, DD, ID, ASD and ADHD	M/-
#223	Dup	arr[hg19] 15q24.1(72,838,805-73,581,757)x3	742,952	8	4	BBS4	Short stature, IUGR, DD, MID and FD	**M/*3Pv**
#223	Dup	arr[hg19] 3p26.3(255,645-1,510,822)x3	1,255,177	2	2	CTN6, CHL1	Short stature, IUGR, DD, MID and FD	**M/*3Pv**
#223	Dup	arr[hg19] 6q25.3(156,488,875-158,534,725)x3	2045,85	9	4	SNX9, ARID1B	Short stature, IUGR, DD, MID and FD	**M/*3Pv**
#245	Dup	arr[hg19] 14q12(26,490,666-27,520,832)x3	1,030,166	2	1	NOVA1	Obesity, encephalopathy, CAs, DD and FD	**F/-**
#248	Del	arr[hg19] 10q23.1(87,392,282-87,791,684)x1	399,402	1	1	GRID1	Abnormal brain structure, DD,	**M/-**
#255	Del	arr[hg19] 10q23.1(87,691,467-87,843,627)x1	152,16	1	1	GRID1	DD	**M/*2Pv**
#268	Del	arr[hg19] 2q13(110,504,318-111,365,996)x1	861,678	16	3	NPH1	ASD	M/-
#276	Dup	arr[hg19] Xq26.2(130,672,818-130,967,726)x3	294,908	2	3	KAL1	DD, FD, cardiomyopathy, thyroid dysfunction and myopia	**F/-**

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#278	Dup	arr[hg19] 19q13.42(54,201,711-54,420,807)x3	219,096	39	9	MIR, NLRP12	Epilepsy, abnormal brain structure	**F/-**
#290	Dup	arr[hg19] 2q13(110,496,601-110,983,418)x3	486,817	14	3	NPHP1	Genetic counseling	**M/-**
#294	Dup	arr[hg19] 2q13(110,498,141-110,980,295)x3	482,154	14	3	NPHP1	DD, ID, FD and	**F/-**
#299	Dup	arr[hg19] 17q11.2(28,952,286-29,150,025)x3	197,739	4	1	CRLF3	DD, ASD and behavioural disorder	M/-
#309	Del	arr[hg19] 17p13.1(6,949,507-7,217,381)x1	267,874	16	15	-	Short stature, DD, ID, FD and microcephaly	**M/-**
#311	Dup	arr[hg19] 1p31.3(61,699,736-62,125,970)x3	426,234	2	1	NFIA	Obesity, CAs, DD,	**F/-**
#319	Dup	arr[hg19] 16p13.3(1,252,411-1,404,818)x3	152,407	9	8	5 OMIMs	Anal imperforation, onfalocele and cloacal exstrophy	**F/-**
#331	Dup	arr[hg19] 4p16.3p16.2(4,025,257-4,618,896)x3	593,639	7	3	NSG1	DD, epilepsy and FD	**M/*Pv**
#336	Dup	arr[hg19] 1q25.3(183,589,206-183,827,325)x3	238,119	3	3	ARPC5, APOBEC4, RGL1	DD and FD	**F/-**
#342	Del	arr[hg19] 3p24.2(24,376,230-24,492,572)x1	116,342	1	1	THRB	DD, Bilateral hearing impairment and FD	**F/-**
#346	Del	arr[hg19] 7q31.1(111,485,313-111,922,531)x1	437,218	2	2	DOCK4	Low weight, slender build, motor delay, DD, SLD, SevID and ASD.	M/-
#354	Dup	arr[hg19] 9q33.1(118,409,943-119,207,073)x3	797,13	4	3	NOC2L	Consanguineous parents, quadriparesis, DD, FD and ostium secundum	**M/-**
#359	Dup	arr[hg19] 5q14.1(80,019,759-80,535,750)x3	515,991	6	3	MSH3, RASGRF2, CKMT2	Convulsions, LDO, MID	**F/-**
#360	Del	arr[hg19] 1p31.1(72,257,666-72,499,784)x1	242,118	2	1	NEGR1	Convulsions, LDO, F	**F/-**
#369	Dup	arr[hg19] 12p11.22p11.21(30,175,955-31,570,927)x3	1,394	9	3	IPO8, CAPRIN2, DDX11	Abnormal brain structure	**M/-**
#383	Dup	arr[hg19] 10q11.23(51,250,417-51,755,110)x3	504,693	7	4	PARG , MSMBP , NCOA4 , TIMM23	Convulsions, DD, SLD, ASD, behavioural disorder and gluten intolerance	M/-

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#384	Dup	arr[hg19] 10q21.1(59,984,568-60,285,875)x3	301,307	5	5	IPMK, CISD1, UBE2D1	Motor delay and chronic encephalopathy	**M/-**
#384	Dup	arr[hg19] 18q22.3(72,755,482-73,023,597)x3	268,115	3	1	TSHZ1	Motor delay and chronic encephalopathy	**M/-**
#397	Dup	arr[hg19] 16p12.2(21,817,921-22,431,357)x3	613,436	9	3	UQCRC2, EEF2K, CDR2	DD, ASD and FD	M/-
#401	Dup	arr[hg19] 2q11.1(95,733,867-96,279,208)x3	545,341	8	3	ZNF2, MRPS5, KCNIP3	ASD, DF	F/-
#423	Dup	arr[hg19] 12q21.31(80,559,698-80,918,615)x3	358,917	2	2	OTOGL, PTPRQ	CAs, ID and FD	**F/-**
#444	Del	arr[hg19] 16p13.3(6,644,079-6,675,606)x1	31	1	1	RBFOX1	ASD	**M/-**
#456	Dup	arr[hg19] 8q24.3(144,262,042-144,486,369)x3	224	8	4	GPIHBP1	ASD and SLD	M/-
#477	Dup	arr[hg19] 7p15.3(24,133,960-24,671,640)x3	538	2	2	NPY, MPP6	ASD and ID	M/-
#492	Dup	arr[hg19] Yq11.221q11.222(19,563,599-21,028,944)x4	1,465	12	3	XKRY, HSFY1, CDY2A	ASD and ID	M/ 46, XY, inv(9)
#500	Del	arr[hg19]10q23.2(88,466,260-88,577,094)x1	111	2	2	LDB3 (605906), BMPR1A (601299)	Multiple CAs, right periorbital oedema, thumbs with distal implantation, nail hypoplasia, absence of distal phalanx, bilateral 5th finger hypoplasia, single unilateral palmar fold, clubfoot to the right, omphalocele, pulmonary cardiopathy	F/-
#567	Dup	arr[hg19]16p13.3(1,052,880-1,268,271)x3	215	4	3	C1QTNF8 (*614147), CACNA1H (*607904)	Short stature, disturbed behaviour, long palpebral fissures, long eyelashes, long philtrum, anteverted nostrils, posteriorly rotated ears, short fingers, finger pads, partial syndactyly, wide hallux, dysplastic nails of 2nd and 5th toes, hyperlordosis.	M/-
#577	Dup	arr[hg19]4q28.1(126,172,904-126,474,296)x3	301	2	1	FAT4 (612411)	DD, CNS malformation, cerebellar hypoplasia, lack of CP growth	M/-

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#584	Dup	arr[hg19]4q21.1q21.21(78,742,412-78,989,716)x3	247	2	2	MRPL1 (611821), FRAS1 (607830)	DD, macrocephaly, FD	M/-
#592	Dup	arr[hg19] 9q33.1(119,677,859-120,377,754)x3	700	2	1	ASTN2	ASD, ID, SLD and hemiparesis	F/-
#595	Dup	arr[hg19] 14q12(32,110,535-32,560,537)x3	450	4	2	NUBPL(*613621)	ASD, seizures	M/-
#602	Del	arr[hg19]Xq25(126,962,874-127,374,779)x1	412	1	1	ACTRT1 (*300487)	Suspected Alagille syndrome	P/-
#607	Dup	arr[hg19] 8q24.3(143,610,752-143,933,329)x3	322	15	9	SLURP1 (*606119)	Macrocephaly, advanced bone age, skin spots, ectrodactilia	F/-
#635	Dup	arr[hg19]16p12.2(21,841,354-22,442,007)x3	600	13	4	UQCRC2 (*191329)	Atrophy, polymicrogyria, lissencephaly, microcephaly, interatrial communication, DD	F/ 2VUS
#635	Dup	arr[hg19]2q21.1(131,502,025-131,970,782)x3	468	5	3	ARHGEF4 (*605216)	Atrophy, polymicrogyria, lissencephaly, microcephaly, interatrial communication, DD	F/ 2VUS
#644	Dup	arr[hg19]15q15.3(43,868,571-43,977,181)x3	108	5	4	STRC (606440)	DD, suspected Susac syndrome	M/-
#646	Dup	arr[hg19]13q33.2q34(105,943,388-114,027,457)x3	8,084	68	21	-	Leopard syndrome like frekles, mildID, alopecia, tremor	F/ +1PCNV

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#649	Dup	arr[hg19] 7q21.3(96,630,285-96,773,686)x3	143	3	3	DLX5 (600028)	Multiple CAs, cardiopathy, bilateral thumb agenesis, vertical talus feet, microtia, suspected TAR or VACTERL	F/ 3VUS
#649	Dup	arr[hg19] Xq21.1(76,959,491-77,184,107)x3	224	4	4	COX7B (300885)	Multiple CAs, cardiopathy, bilateral thumb agenesis, vertical talus feet, microtia, suspected TAR or VACTERL	F/ 3VUS
#649	Dup	arr[hg19] Xp11.3 (44,597,231-44,820,429)x3	223	2	2	KDM6A (*300128)	Multiple CAs, cardiopathy, bilateral thumb agenesis, vertical talus feet, microtia, suspected TAR or VACTERL	F/ 3VUS
#658	Dup	arr[hg19] 9q22.31(94,665,153-94,812,523)x3	147	2	2	ROR2, SPTLC1	ASD	M/2VUS
#658	Dup	arr[hg19] 13q12.3(31,764,937-32,190,263)x3	425	1	1	B3GLCT	ASD	M/2VUS
#661	Dup	arr[hg19] 6q14.3(87,410,108-87,911,748)x3	502	3	3	HTR1E, CGA, ZNF292	ASD	M/Brother with ASD #659 (negative)
#671	Dup	arr[hg19] 14q12(29,221,762-29,459,916) x3	238	4	1	FOXG1 (*164874)	Epilepsy, DD, suspected Dravet syndrome (SCN1A)	F/-
#695	Del	arr[hg19] 16p13.3(6,887,840-6,966,572)x1	79	1	1	RBFOX1	ASD	M/-
#716	Dup	arr[hg19] 3p21.31 (45779135-46005169)x3	226	6	6	SLC6A20 (*605616)	Disturbed behavior, SLD, FD, Karyotype: 46,XY,22ps +	M/-
#727	Dup	arr[hg19] 2q13(110,504,318-111,365,996)x3	862	16	3	NPHP1	ASD, DD, abnormal growth, FD, CAs, macrocephaly and hirsutism	F/-
#738	Dup	arr[hg19] 15q13.3(31,999,631-32,444,043)x3	444	1	1	CHRNA7	ASD, ID, FD, SLD, abnormal growth, dyslalia, motor difficulties and ADHD	M/-

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#741	Dup	arr[hg19]13q12.12(23,473,290-24,979,038)x3	1,506	15	8	-	Multiple comorbidities (not specified)	M/-
#742	Dup	arr[hg19] 14q12(32,110,535-32,600,382)x3	490	4	2	NUBPL (*613621)	Difficult-to-control epilepsy, cerebral palsy, DD, overweight, frontal temporal cortical dysplasia	M/-
#744	Dup	arr[hg19] 17p13.3(791,201-977,024)x3	186	5	3	TIMM22 (*607251) e ABR (*600365)	ASD, hypertension, dyslipidemia	M/-
#756	Del	arr[hg19]6p22.3(17,808,245-17,978,255)x1	170,011	1	1	KIF13A (605433)	Deformed toes, clubbed fingers, ptosis, proeminent nose, kyphosis, short stature, DD	M/-
#758	Dup	arr[hg19]15q13.3(32,003,538-32,931,921)x3	928	12	8	CHRNA7 (*118511)	DD, hypotonia, strabismus,	F/-
#761	Del	arr[hg19] Xp22.31(6,455,149-8,135,644)x1	1,680	5	5	STS (*300747)	Failure to thrive, DD	F/-
#768	Dup	arr[hg19] 1p36.32(3,311,950-3,589,407)x3	277	8	6	PRDM16 (*605557)	Arched palate, hair line in M, wide-spaced nipples, heart fremitus, cubitus valgus, growth delay	F/-
#769	Dup	arr[hg19]18q21.1(44,515,228-44,898,600)x3	383	6	4	KATNAL2 (*614697)	dystonia, dysarthria, hyperactivity, mild ID	F/-
#777	Del	arr[hg19]7q34(141,799,147-142,047,384)x1	248	5	1	MGAM (154360)	congenital deafness	M/ brother of #778

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#777	Dup	arr[hg19]7p15.2(25,994,970-26,238,444)x3	244	2	2	NFE2L3 (604135), HNRNPA2B1 (600124)	congenital deafness	M/ brother of #778
#778	Del	arr[hg19]7q34(141,799,147-142,047,384)x1	248	5	1	MGAM (154360)	congenital deafness	M/ brother of #777
#778	Dup	arr[hg19]7p15.2(25,994,970-26,238,444)x3	244	2	2	NFE2L3 (604135), HNRNPA2B1 (600124)	congenital deafness	M/ brother of #777
#785	Dup	arr[hg19]9p24.3(208,455-336,687)x3	128	2	1	DOCK8 (611432)	single palmar fold, inverted nipples, bilateral cryptorchidism, inguinal hernia, growth delay	M/-
#790	Del	arr[hg19] 1p13.3(108,700,187-108,962,439)x1	262	4	2	SLC25A24 (608744)	cleft palate, DD, FD, hyperactivity	M/-
#792	Del	arr[hg19] 1p32.1(59,848,744-60,044,036)x1	195	1	1	FGGY (*611370)	Mild DD, generalised epilepsy, mild tricuspid insufficiency, right lung hypoplasia	F/-
#798	Del	arr[hg19] 4q13.3q21.1(74,958,696-76,339,793)x1	1,381	13	7	-	consanguineous parents, pre and post-natal short stature, microcephaly, hypertrophic cardiomyopathy, oblique palpebral fissures	F/-
#810	Dup	arr[hg19] 7p11.2(57,233,082-57,906,704)x3	674	4	-	-	ASD	M/-
#814	Dup	arr[hg19] 15q11.2(22,770,421-23,288,350)x3	518	8	4	NIPA1	ASD	M/-

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#816	Del	arr[hg19] 8q21.13(80,370,864-82,083,687)x1	1,713	10	5	-	DD, cardiac malformation, global hypotonia, signs of speech apraxia	M/-
#828	Del	arr[hg19]5p15.31p15.2(9,090,338-11,635,988)x1	2,545	20	8	-	ID, strabismus, protruding ears brother of case #829	M/ +1 PCNV
#829	Del	arr[hg19]5p15.31p15.2(9,090,338-11,635,988)x1	2545	20	8	-	ID and DF	M/ Brother of case #828 + 1 PCNV
#830	Del	arr[hg19] 11p14.2p14.1(26,997,314-27,233,664)x1	236	3	1	BBOX1	ASD	M/-
#830	Del	arr[hg19] 11p14.2p14.1(26,997,314-27,233,664)x1	236	3	1	BBOX1	ASD	M/-
#835	Del	arr[hg19]16p13.3(6,887,841-7,013,897)x1	126	1	1	RBFOX1(*605104)	ModID, DD, divergent strabismus, scoliosis	F/-
#835	Dup	arr[hg19]11p15.2(14,589,177-15,240,408)x3	651	6	6	-	ModID, DD, divergent strabismus, scoliosis	F/-
#854	Dup	arr[hg19] 1q21.1(145,369,184-145,988,238)x3	619	24	12	RBM8A (605313), PEX11B (603867)	Developmental regression, seizures, suspected Rett syndrome.	F/-
#855	Dup	arr[hg19] 14q12(26,490,666-27,520,832)x3	1,057	6	1	NOVA1 (*602157)	ASD	F/-
#870	Dup	arr[hg19] 12p11.1(34,065,100-34,802,951)x3	738	2	1	ALG10 (*603313)	Ataxic hypotonic cerebral palsy, cannot walk, absent speech, seizures	F/-

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#876	Dup	arr[hg19] Xq13.1(69,207,741-69,317,932)x3	110	3	3	EDA (300451)	Macrocephaly, broad forehead, sialorrhea, limb hemihypertrophy, joint laxity, suspected Sotos syndrome	F/-
#880	Del	arr[hg19] 16p13.3(6,967,626-6,998,141)x1	30,5	1	1	RBFOX1 (* 605104)	ASD	M/-
#904	Del	arr[hg19] 7q31.1(110,863,879-111,283,978)x1	420	1	1	IMMP2L	ASD and DD	M/-
#914	Dup	arr[hg19]3p14.3(57,370,900-57,746,279)x3	375	7	4	-	ID, microcephaly, cleft lip, foot anomaly, small hands, ichthyosis vulgaris	F/-
#918	Dup	arr[hg19]10p15.1(4,620,457-5,104,391)x3	484	7	4	-	ADHD, ID	M/-
#927	Dup	arr[hg19]15q11.2(22,770,422-23,615,769)x3	845	4	4	-	Suspected Angelman syndrome	M/ 1 PCNV
#927	Dup	arr[hg19] 15q11.2(22,770,421-23,615,769)x3	845	16	4	-	Suspected Angelman syndrome	M/ 1 PCNV
#928	Del	arr[hg19] 1q21.2(147,723,034-147,830,830)x1	108	2	1	NBPF8	ASD, FD, SLD, CAs and macrocephaly	M/-
#929	Dup	arr[hg19] 8p21.3(19,775,847-20,071,770)x3	296	3	3	LPL (*609708)	Fetal losses with malformation	F/ 2 VUS
#929	Dup	arr[hg19] 11q14.1(77,492,774-78,509,705)x3	1,017	16	3	-	Fetal losses with malformation	F/ 2 VUS

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#937	Del	arr[hg19] 1q41(219,090,413-219,734,998)x1	645	4	1	LYPLAL1	ASD	M/3VUS
#937	Dup	arr[hg19] 1q41q42.12(221,795,858-225,581,420)x3	3,786	29	13	WDR26	ASD	M/3VUS
#937	Dup	arr[hg19] Xq21.1(76,993,262-77,392,096)x2	399	7	7	-	ASD	M/3VUS
#941	Dup	arr[hg19] Xq28(152,927,530-152,991,389)x2	64	4	4	ABCD1	ASD, MildID, SLD and hyperkinetic disorder	M/-
#942	Dup	arr[hg19]Xp21.2(31,116,865-31,359,757)x3	243	1	1	DMD (*300377).	DD, SLD, hiperactivity, oppositional defiant disorder,seizures	F/-
#954	Del	arr[hg19] 16p13.3(3,315,091-3,432,025)x1	117	6	3	-	ASD, DD, Abnormal growth, FD, CAs, abnormal brain structure, motor difficulties and epilepsy	M/-
#959	Dup	arr[hg19] 2q12.2q12.3(106,873,992-108,480,894)x3	1,607	9	4	-	Suspected genetic condition	F/-
#984	Del	arr[hg19] 13q22.2q31.1(76,555,343-80,068,154)x1	3,513	24	12	-	ASD, motor agitation and use of corrective lenses	M/-
#1006	Del	arr[hg19] 15q11.2(22,770,421-23,214,655)x1	444	6	4	CYFIP1, NIPA2	ASD	M/-
#1021	Dup	arr[hg19]3p24.1(30577491-30775657)x3	198	2	2	TGFBR2 (*190182)	microcephaly and micrognathia	M/-

Case	VUS	Microarray Nomenclature	Size (Kbp)	N° of Genes	N° of Genes in OMIM	Important Genes	Phenotype	Gender /Notes
#1053	Del	arr[hg19] 7q31.1(111,483,652-111,922,578)x1	439	4	2	DOCK4 e ZNF277	ASD, midID, obesity and gynecomastia	M/-
#1057	Dup	arr[hg19] 8q12.1(56,717,039-57,240,751)x3	523	10	7	CHCHD7 (*611238)	short stature, limb anomalies	F/-
#1088	Dup	arr[hg19] 1q43(237,843,614-238,415,416)x3	572	3	2	RYR2 (180902), ZP4 (613514)	DD, ataxia, psychomotor agitation	M/-
#1091	Dup	arr[hg19] Xq11.2q12(64,008,668-64,866,293)x2	858	4	3	ZC4H2	ASD	M/-
#1099	Dup	arr[hg19] 16q24.1(84,439,871-84,676,492)x3	237	4	2	ATP2C2 (*613082)	Obesity, ASD, Karyotype: 46,XY inv3(q13q26)	F/-
#1108	Dup	arr[hg19] 15q11.2(22,770,421-23,082,237)x3	312	6	4	NIPA1 (*608145)	poor motor coordination, DD, SLD, anxiety	F/-
#1109	Dup	arr[hg19]4p16.1(7,136,707-11,214,146)x3	4,077	47	17	-	Additional material on chromosome 3.	M/ +1 PCNV
#1119	Dup	arr[hg19] 2p13.2(71,586,200-71,730,470)x3	144	2	2	ZNF638 (614349), DYSF (603009).	ID, syndromic, epilepsy, scoliosis	F/-
#1120	Dup	arr[hg19] 15q21.1(45,225,243-45,806,154)x3	581	26	12	-	ASD and MildID	M/-
#1127	Del	arr[hg19]2q31.2(179,396,924-179,629,278)x1	232	2	2	TTN	ASD and epilepsy	M/

Copy Number Variants of Unknown Significance (VUS) found in the cohort, with the number of genes present in the region, listing the most relevant genes and phenotypes for each individual. Dup = Duplication, Del = Deletion, CAs = Congenital Anomalies, DD = Developmental Delay,

ID = Unspecified intellectual disability, MildID = Mild Intellectual Disability, ModID = Moderate Intellectual Disability, SevID = Severe Intellectual Disability, ASD = Autism Spectrum Disorder, and FD = Facial Dysmorphisms, SLD = speech and/or language delay or impairment, IUGR = Intrauterine growth restriction, ADHD = Attention-deficit/hyperactivity disorder, LDO = learning difficulty only (no ID), ASD = Autism spectrum disorder, F = Female, M = Male. LPCNVs = likely pathogenic CNVs. *2V = Patients with 2 VUS. *3V = Patients with 3 VUS.

SUPPLEMENTARY TABLE 2

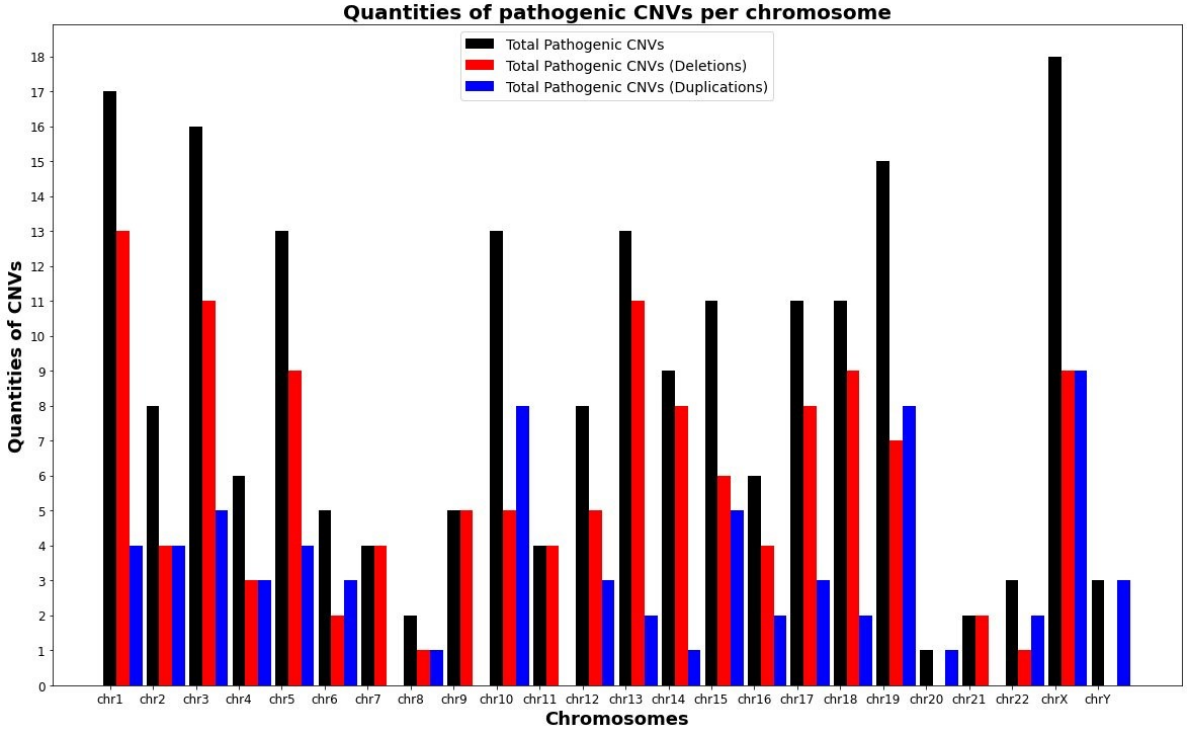
Prevalence of Syndromes Associated with Pathogenic CNVs.

Identified syndromes	Frequency in the cohort (%)	Frequencies in pathogenic CNVs (%)
Angelman/Prader Willi syndrome	0,69	3,38
Di George syndrome	0,69	3,38
1p36 deletion syndrome	0,59	2,90
16p11.2 deletion syndrome	0,49	2,42
Cri Du Chat syndrome	0,49	2,42
Inverted 8p duplication/deletion syndrome	0,49	2,42
1q21.1 Microduplication Syndrome	0,39	1,93
22q11.2 duplication syndrome	0,39	1,93
Distal 13q deletion Syndrome	0,39	1,93
partial trisomy 7q31.32q33	0,39	1,93
PhelanMcDermid Syndrome	0,39	1,93
Waardenburg syndrome	0,39	1,93
Williams Beuren Syndrome	0,39	1,93
2q37 Microdeletion Syndrome	0,29	1,45
Koolen de Vries syndrome	0,29	1,45
partial trisomy 19p13	0,29	1,45
Terminal 21q del	0,29	1,45
Trissomia Parcial do 21q	0,29	1,45
15q11q13 duplication syndrome	0,20	0,97
15q13.3 Microdeletion Syndrome	0,20	0,97
18p deletion syndrome	0,20	0,97
18q deletion syndrome	0,20	0,97
2q31.1 microdeletion syndrome	0,20	0,97
6p25 deletion syndrome	0,20	0,97
9p Deletion Syndrome	0,20	0,97
ATR16 syndrome	0,20	0,97
Distal 16p11.2 deletion Syndrome	0,20	0,97
partial 13q monosomy syndrome	0,20	0,97
Partial Trisomy Distal 4q	0,20	0,97
Temple syndrome	0,20	0,97
Trisomy 12p	0,20	0,97
8p23.1 deletion syndrome	0,10	0,48
10q26 deletion syndrome	0,10	0,48
15q24 deletion syndrome	0,10	0,48
16p13.3 microduplication syndrome	0,10	0,48
18 q21.32qter deletion syndrome	0,10	0,48
21q22.12 microdeletion syndrome	0,10	0,48
3q29 Microduplication syndrome	0,10	0,48

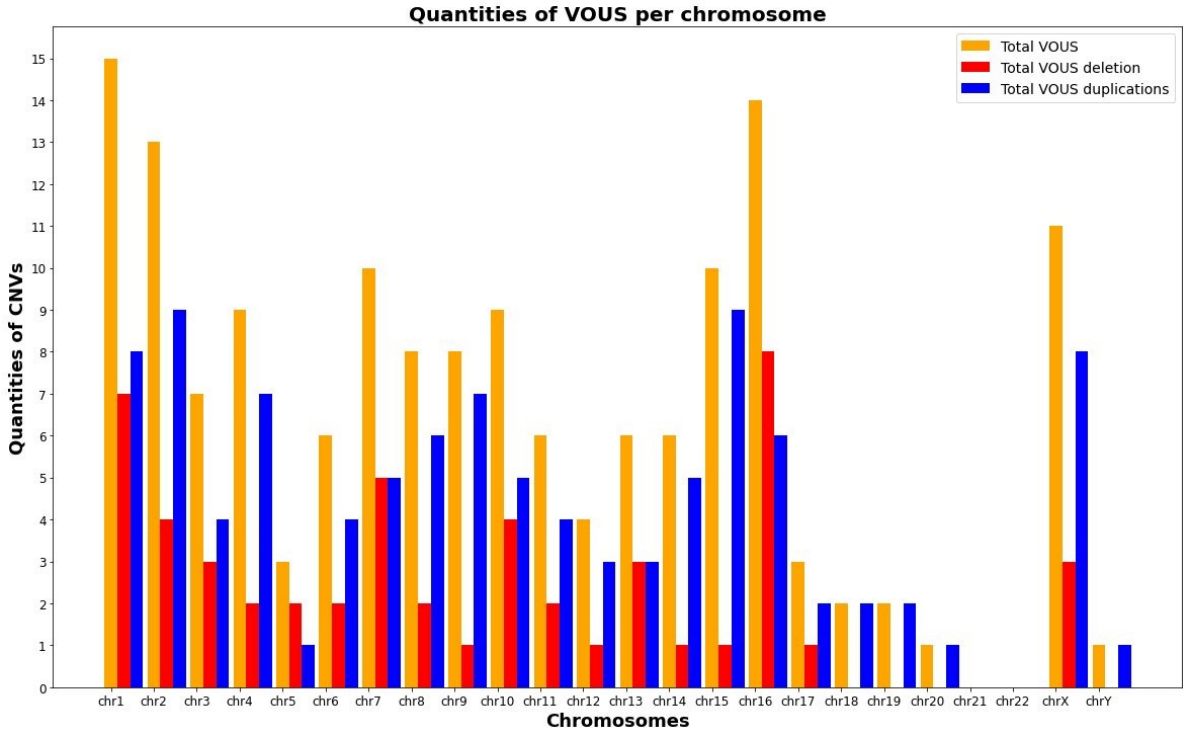
Identified syndromes	Frequency in the cohort (%)	Frequencies in pathogenic CNVs (%)
7q3436 deletion syndrome	0,10	0,48
8p interstícial deletion including p12 syndrome	0,10	0,48
CharcotMarieTooth disease type 1A – CMT1A (	0,10	0,48
CoffinSiris syndrome	0,10	0,48
Cohen syndrome	0,10	0,48
distal trisomy 10q syndrome	0,10	0,48
Distal trisomy 18q	0,10	0,48
Distal trisomy 3q	0,10	0,48
Distal trisomy 8p	0,10	0,48
DYRK1A related intellectual disability syndrome	0,10	0,48
Greig syndrome	0,10	0,48
KBG syndrome	0,10	0,48
Keutel syndrome	0,10	0,48
MECP2 Duplication Syndrome	0,10	0,48
partial 18p deletion syndrome	0,10	0,48
Partial Trisomy 16p13.3 Syndrome	0,10	0,48
partial trisomy 18q	0,10	0,48
partial trisomy 5p14.3p15.31	0,10	0,48
Rett syndrome	0,10	0,48
Schinzel Giedion Syndrome	0,10	0,48
SimpsonGolabiBehmel Syndrome	0,10	0,48
Síndrome de San Luis Valley	0,10	0,48
Síndrome Tricorrinofalangeana (TRPS).	0,10	0,48
SmithMagenis Syndrome	0,10	0,48
Sotos Syndrome	0,10	0,48
tetrasomy 18p11.21p11.32	0,10	0,48
Triple X Syndrome	0,10	0,48
WAGR syndrome	0,10	0,48
WolfHirschhorn syndrome	0,10	0,48
Xq26.3, Xq27.3q28 and Xq28 duplication syndromes	0,10	0,48
Xq27.3q28 and Xq28 duplication syndrome	0,10	0,48
Xq28 duplication syndrome	0,10	0,48
XYYRegion Syndrome	0,10	0,48
Total syndromes identified	**14,51**	**71,50**

Note: Values represent the prevalence of syndromes in individuals with idenified pathogenic CNVs.

SUPPLEMENTARY INFORMATION 1



Pathogenic CNVs per chromosome

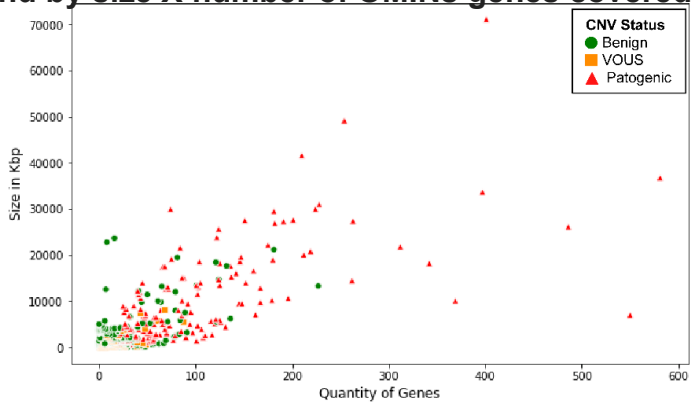


VUS per chromosome.

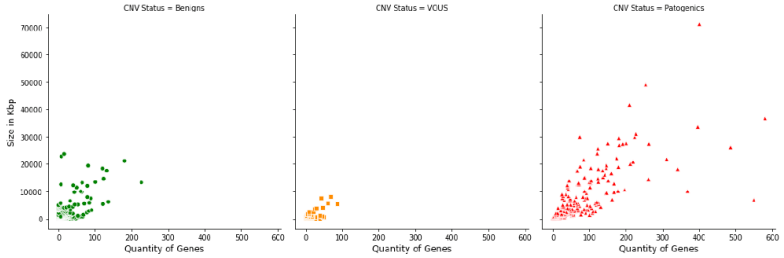
SUPPLEMENTARY INFORMATION 2

Comparison of CNVs interpreted, by size X number of genes covered (A & B) and by size X number of OMINs genes covered (C & D):

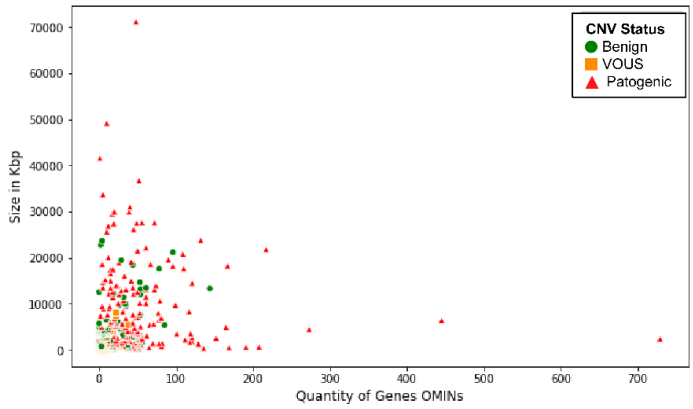
A



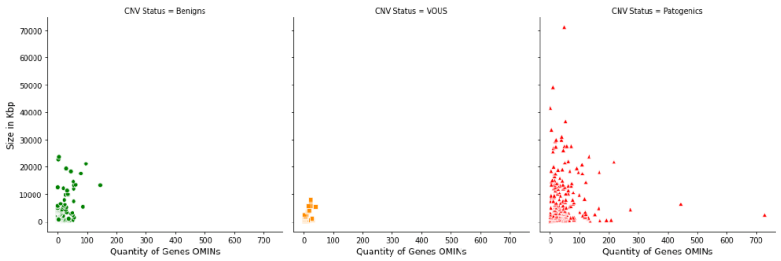
B



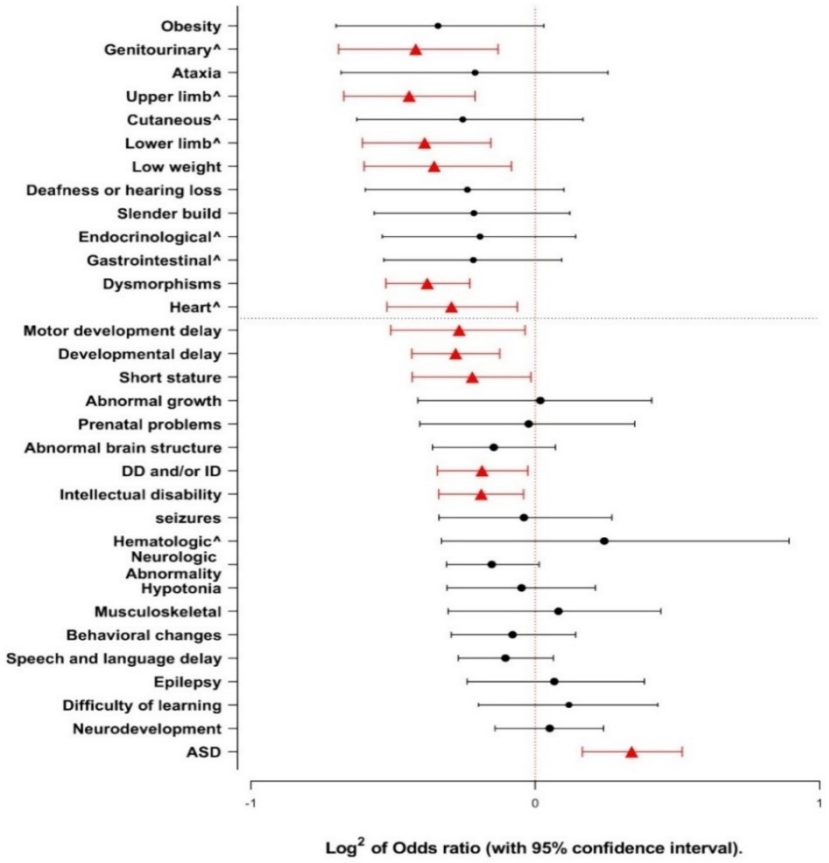
C



D

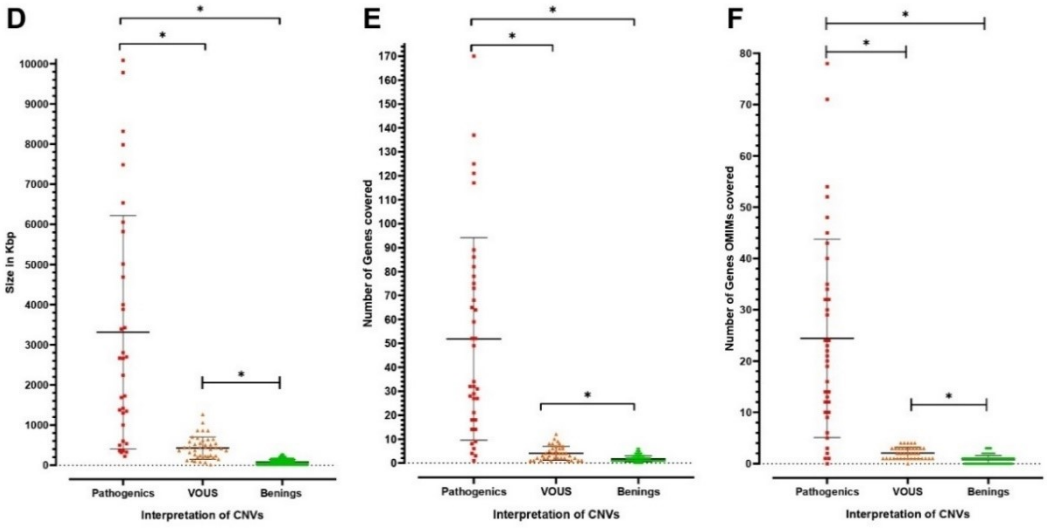


SUPPLEMENTARY INFORMATION 3



. Odds ratios shown in log2 scale. Odds ratios with a *p*-value < 0.05, two tailed were displayed in red, while others were shown in black..^: anomalies and malformations, ID: Intellectual disability; DD: Developmental delay and ASD: Autism spectrum disorder.

Comparison of CNVs interpreted in cases with ASD, by size (D), number of genes covered (E) and number of OMINs genes covered (F):



SUPPLEMENTARY INFORMATION 4

Circus ideogram with pathogenic CNVs found in ASD cases, plotted by chromosome.

Figure 2

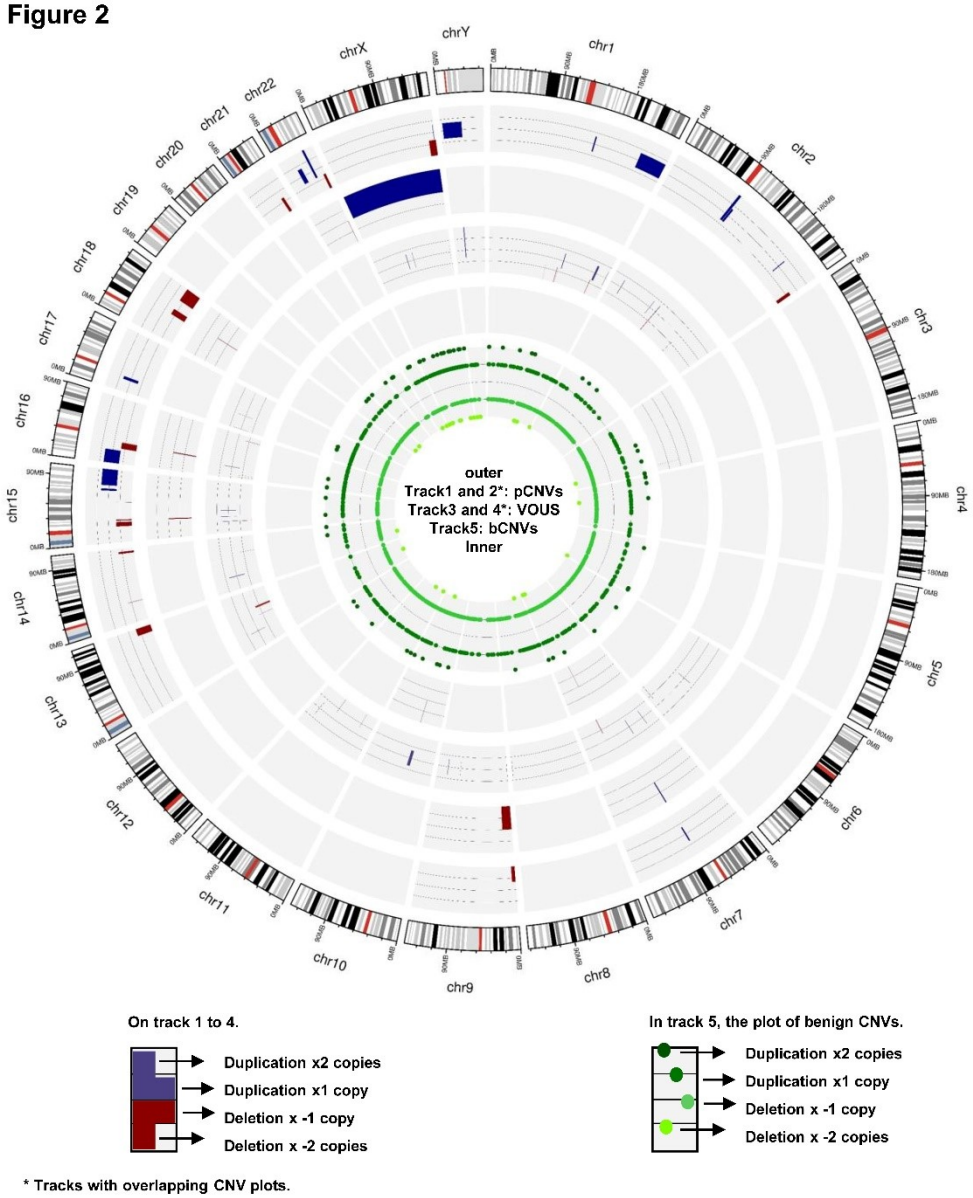


Figure 2: The circle ideogram plot shows in the ouhermost tracks the genomic positions of all pathogenic CNVs (pCNVs) found in the study per human chromosome, wich are plotted in two tracks, the first and the second to allow visualization of overlapping pCNVs. The VUS are plotted in the third and fourth tracks (also two tracks, for overlapping VUS). The bars in blue represent duplication, either for x1 or x2 additional copies. The bars in red represent deletions in x1 (heterozygous) or x2 (hemozygous) copies. At the center of track 5 are plotted all the CNVs interpreted as benign (bCNVs) detected in the study, these in turn, are presented by dots in shades of green, which represent different states of copies (duplication x1 or x2 copies and deletions x1 or

5 CONSIDERAÇÕES FINAIS

Nossa revisão bibliográfica envolvendo 84 estudos de coorte de indivíduos afetados com Distúrbios do Neurodesenvolvimento (DsND) investigados por Microarray Cromossômico (CMA), obteve uma média de 16,8% (95% CI: 15,07-18,54) taxa de detecção de CNVs patogênicas e possivelmente patogênicas. Sendo que as maiores amostragens de coortes estudas referem-se aos países economicamente mais desenvolvidos do hemisfério norte, contrastando com uma baixa amostragem de países economicamente menos desenvolvidos do hemisfério sul, como na América do Sul.

Demostramos aqui que o CMA é uma ferramenta essencial para decifrar as sequências envolvidas nas anormalidades estruturais detectadas pela análise cromossômica clássica, assim como pacientes com resultados anormais do CMA devem ter seus cromossomos analisados, permitindo uma visão dos mecanismos que deram origem à anormalidade genética, o que é relevante para o aconselhamento genético.

Neste estudo retrospectivo apresentamos o maior relato de dados de CMA em uma coorte com DsND e/ou Anomalia Congênitas (AC) do Sul do Brasil. Obtivemos uma taxa de diagnóstico de 17 % quando analisamos a coorte, compatível com a literatura (15–20%). Para os casos com transtorno do espectro autista (TEA), o rendimento diagnóstico foi de 11%, sendo maior quando sindrômico e menor (7%) para TEA "isolado", taxas compatíveis com na literatura (8-21%) para coortes de TEA.

Caracterizamos as raras CNVs encontradas junto com os principais fenótipos apresentados por cada paciente, uma vez que, a interpretação CNVs é desafiadora e depende muito de informações, como frequência e caracterização em populações afetadas, que normalmente são obtidas de estudos de coorte com tamanhos de amostra significativos.

As principais razões para encaminhar indivíduos para o teste de CMA neste estudo foram atraso no desenvolvimento/deficiência intelectual e o TEA, na maioria das vezes, com a presença de características sindrômicas, como anomalias ou malformações congênitas ou dismórficas. Sendo que alguns fenótipos demostram ser

preditivos de uma maior probabilidade de carregar uma CNV patogênica como: o Atraso no desenvolvimento (AD)(*p*-value = <0,001, OR = 0,53); TEA (*p*-value = <0,001, OR = 2,18); Malformações/Dismorfismos Faciais (*p*-value = <0,001, OR = 0,42); anomalias do membro superior (*p*-value = <0,001, OR = 0,36); anomalias de membros inferiores (*p*-value = 0,001, OR = 0,41); anomalias e malformações geniturinárias (*p*-value = 0,004, OR = 0,38); Baixo peso (*p*-value = 0,01, OR = 0,44); DI (*p*-value = 0,014, OR = 0,65); Anomalias e malformações cardíacas (*p*-value = 0,018, OR = 0,51); DI ou AD (*p*-value = 0,025, OR = 0,65) e atraso no desenvolvimento motor (*p*-value = 0,036, OR = 0,54).

Para os 953 CMAs, em que os longos trechos contíguos de homozigose (LCSH) foram analisados, 27 grandes LCSH ≥ 10 megabases (Mpb), variando de 10,6 a 88,8 Mpb, foram observadas em um único autossomo, sugerindo potencial de dissomia uniparental (UPD) em uma frequência de 2,8%. As limitações da CMA resultam na subestimação da taxa de UPD, pois só é possível sugerir sua presença quando a isodissomia uniparental é detectada. No entanto, ausência de testes de metilação impede a confirmação desses achados como UPDs reais e a diferenciação entre UPDs completos e segmentares.

Quanto a consanguinidade, a análise dos LCHSs sugeriram descendência de parentes de primeiro a quinto grau, em aproximadamente 11,5% da coorte. Essas informações são essenciais para o aconselhamento genético, uma vez que há risco empírico de recorrência, nos casos de parentesco próximo, deve ser considerar a possibilidade de uma mutação autossômica recessiva autozigótica. Se houver irmãos adicionais afetados, a análise das regiões idênticas por descendência pode auxiliar na identificação da região alvo a ser investigada por meio do sequenciamento completo do exoma (WES), se necessário.

Identificamos aqui uma série de 10 LCSHs em frequência ≥ 5% em indivíduos com DsND. Nove dessas LCSHs já haviam sido relatadas anteriormente como variações comuns por outros grupos de pesquisa, sugerindo que são prováveis variantes populacionais normais em Santa Catarina. Essas informações são valiosas para priorizar LCSHs com maior probabilidade de ter relevância clínica para análise. No entanto, não podem ser completamente descartados.

Por fim, a análise dos LCSHs detectados por CMA com alta densidade de SNPs pode fornecer informações valiosas para auxiliar na investigação de distúrbios do neurodesenvolvimento. No entanto, esses achados permanecem principalmente teóricos e sugestivos, podendo ser norteadores para novas investigações, como análise de metilação, sequenciamento de genes direcionados ou WES.

6 REFERÊNCIAS

ABOU JAMRA, R. et al. Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. **European Journal of Human Genetics**, *[S. l.]*, v. 19, n. 11, p. 1161–1166, 2011. DOI: 10.1038/ejhg.2011.98. Disponível em: http://www.nature.com/doifinder/10.1038/ejhg.2011.98.

ASADOLLAHI, Reza et al. Original article: The clinical significance of small copy number variants in neurodevelopmental disorders. **Journal of Medical Genetics**, *[S. l.]*, v. 51, n. 10, p. 677, 2014. DOI: 10.1136/JMEDGENET-2014-102588. Disponível em: /pmc/articles/PMC4173859/. Acesso em: 27 ago. 2022.

BOAT, Thomas F.; WU, Joel T.; DISORDERS, Committee to Evaluate the Supplemental Security Income Disability Program for Children with Mental; POPULATIONS, Board on the Health of Select; BOARD ON CHILDREN, Youth, and Families; MEDICINE, Institute of; EDUCATION, Division of Behavioral and Social Sciences and; THE NATIONAL ACADEMIES OF SCIENCES, Engineering, and Medicine. Clinical Characteristics of Intellectual Disabilities. *[S. l.]*, 2015. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK332877/. Acesso em: 5 maio. 2023.

CAPPUCCIO, Gerarda et al. New insights in the interpretation of array-CGH: autism spectrum disorder and positive family history for intellectual disability predict the detection of pathogenic variants. **Italian journal of pediatrics**, *[S. l.]*, v. 42, p. 39, 2016. DOI: 10.1186/s13052-016-0246-7.

CHAVES, T. F. et al. Copy Number Variations in a Cohort of 420 Individuals with Neurodevelopmental Disorders From the South of Brazil. **Scientific Reports**, *[S. l.]*, v. 9, n. 1, 2019. a. DOI: 10.1038/s41598-019-54347-z.

CHAVES, Tiago Fernando; OLIVEIRA, Luan Freitas; OCAMPOS, Maristela; BARBATO, Ingrid Tremel; DE LUCA, Gisele Rozone; BARBATO FILHO, Jorge Humbeto; DE CAMARGO PINTO, Louise Lapagesse; BERNARDI, Pricila; MARIS, Angelica Francesca. Long contiguous stretches of homozygosity detected by chromosomal microarrays (CMA) in patients with neurodevelopmental disorders in the South of Brazil. **BMC Medical Genomics**, *[S. l.]*, v. 12, n. 1, p. 50, 2019. b. DOI: 10.1186/s12920-019-0496-5. Disponível em: https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-019-0496-5. Acesso em: 20 mar. 2019.

CHIA, Nisbet Nicole Leisa. **DNA Copy Number Variation in a Cohort of Healthy Australian Women**. 2016. Australian National University, *[S. l.]*, 2016.

CHIA, Nisbet Nicole Leisa Nisbet; LEISA, Nicole; CHIA, Nisbet Nicole Leisa Nisbet. **DNA Copy Number Variation in a Cohort of Healthy Australian Women**. 2016. *[S. l.]*, 2016.

CONRAD, Donald F. et al. Origins and functional impact of copy number variation in the human genome. **Nature**, *[S. l.]*, v. 464, n. 7289, p. 704–12, 2010. DOI: 10.1038/nature08516.

COOK, Edwin H.; SCHERER, Stephen W. Copy-number variations associated with neuropsychiatric conditions. **Nature**, *[S. l.]*, v. 455, n. October, p. 919–923, 2008. DOI: 10.1038/nature07458.

D'AMOURS, Guylaine et al. SNP arrays: comparing diagnostic yields for four platforms in children with developmental delay. **BMC Medical Genomics**, *[S. l.]*, v. 7, n. 1, p. 70, 2014. DOI: 10.1186/s12920-014-0070-0.

D'ARRIGO, Stefano et al. The Diagnostic Yield of Array Comparative Genomic Hybridization Is High Regardless of Severity of Intellectual Disability/Developmental Delay in Children. **Journal of child neurology**, *[S. l.]*, v. 31, n. 6, p. 691–699, 2016. DOI: 10.1177/0883073815613562. Disponível em: https://pubmed.ncbi.nlm.nih.gov/26511719/. Acesso em: 29 ago. 2022.

DAVOLI, Teresa; XU, Andrew Wei; MENGWASSER, Kristen E.; SACK, Laura M.; YOON, John C.; PARK, Peter J.; ELLEDGE, Stephen J. Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. **Cell**, *[S. l.]*, v. 155, n. 4, p. 948–962, 2013. DOI: 10.1016/J.CELL.2013.10.011.

DEVLIN, Bernie; SCHERER, Stephen W. Genetic architecture in autism spectrum disorder. **Current Opinion in Genetics & Development**, *[S. l.]*, v. 22, n. 3, p. 229–237, 2012. DOI: 10.1016/j.gde.2012.03.002. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/22463983. Acesso em: 23 out. 2017.

DSM-5 Diagnostic Classification. *In*: **Diagnostic and Statistical Manual of Mental Disorders**. [s.l.] : American Psychiatric Association, 2013. DOI: 10.1176/appi.books.9780890425596.x00DiagnosticClassification. Disponível em: http://psychiatryonline.org/doi/10.1176/appi.books.9780890425596.x00DiagnosticClassification. Acesso em: 3 fev. 2018.

GAMSIZ, Ece D. et al. Intellectual disability is associated with increased runs

of homozygosity in simplex autism. **American journal of human genetics**, *[S. l.]*, v. 93, n. 1, p. 103–9, 2013. DOI: 10.1016/j.ajhg.2013.06.004.

GANDIN, Ilaria et al. Excess of runs of homozygosity is associated with severe cognitive impairment in intellectual disability. **Genetics in medicine : official journal of the American College of Medical Genetics**, *[S. l.]*, v. 17, n. 5, p. 396–9, 2015. DOI: 10.1038/gim.2014.118.

GAUGLER, Trent et al. Most genetic risk for autism resides with common variation. **Nature genetics**, *[S. l.]*, v. 46, n. 8, p. 881–885, 2014. DOI: 10.1038/NG.3039. Disponível em: https://pubmed.ncbi.nlm.nih.gov/25038753/. Acesso em: 28 nov. 2022.

IOUROV, Ivan Y.; VORSANOVA, Svetlana G.; KOROSTELEV, Sergei A.; ZELENOVA, Maria A.; YUROV, Yuri B. Long contiguous stretches of homozygosity spanning shortly the imprinted loci are associated with intellectual disability, autism and/or epilepsy. **Molecular cytogenetics**, *[S. l.]*, v. 8, p. 77, 2015. DOI: 10.1186/s13039-015-0182-z. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/26478745. Acesso em: 17 jan. 2017.

KEARNEY, Hutton M. **SNP allele analysis: UPD and then some**. 2012. Disponível em: https://www.mscience.com.au/upload/pages/acw/hutton-kearney_melb-cytogenomics-workshop_snps_dec2012.pdf. Acesso em: 22 set. 2017.

KOGAN, Michael D. et al. The Prevalence of Parent-Reported Autism Spectrum Disorder Among US Children. **Pediatrics**, *[S. l.]*, v. 142, n. 6, 2018. DOI: 10.1542/PEDS.2017-4161. Disponível em: https://pubmed.ncbi.nlm.nih.gov/30478241/. Acesso em: 27 nov. 2022.

LAY-SON, Guillermo; ESPINOZA, Karena; VIAL, Cecilia; RIVERA, Juan C.; GUZM??N, Mar??a L.; REPETTO, Gabriela M. Chromosomal microarrays testing in children with developmental disabilities and congenital anomalies. **Jornal de Pediatria**, *[S. l.]*, v. 91, n. 2, p. 189–195, 2015. DOI: 10.1016/j.jped.2014.07.003.

LI, Ling-Hui et al. Long contiguous stretches of homozygosity in the human genome. **Human Mutation**, *[S. l.]*, v. 27, n. 11, p. 1115–1121, 2006. DOI: 10.1002/humu.20399.

MAENNER, Matthew J. et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. **MMWR. Surveillance Summaries**, *[S. l.]*, v. 69, n. 4, p. 1–12, 2020. DOI: 10.15585/MMWR.SS6904A1. Disponível em:

https://www.cdc.gov/mmwr/volumes/69/ss/ss6904a1.htm. Acesso em: 25 fev. 2023.

MICHELSON, David Joshua; CLARK, Robin Dawn. Optimizing Genetic Diagnosis of Neurodevelopmental Disorders in the Clinical Setting. **Clinics in Laboratory Medicine**, *[S. l.]*, v. 40, n. 3, p. 231–256, 2020. DOI: 10.1016/J.CLL.2020.05.001.

NAIR, R.; CHEN, M.; DUTT, A. S.; HAGOPIAN, L.; SINGH, A.; DU, M. Significant regional inequalities in the prevalence of intellectual disability and trends from 1990 to 2019: a systematic analysis of GBD 2019. **Epidemiology and Psychiatric Sciences**, *[S. l.]*, v. 31, 2022. DOI: 10.1017/S2045796022000701. Disponível em: /pmc/articles/PMC9805697/. Acesso em: 22 abr. 2023.

NOWAKOWSKA, Beata. Clinical interpretation of copy number variants in the human genome. **Journal of applied genetics**, *[S. l.]*, v. 58, n. 4, p. 449–457, 2017. DOI: 10.1007/s13353-017-0407-4. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/28963714. Acesso em: 21 nov. 2017.

PEREIRA, Rodrigo Roncato et al. Screening for intellectual disability using high-resolution CMA technology in a retrospective cohort from Central Brazil. **PLoS ONE**, *[S. l.]*, v. 9, n. 7, p. e103117, 2014. DOI: 10.1371/journal.pone.0103117.

QUINTELA, Inés et al. Copy number variation analysis of patients with intellectual disability from North-West Spain. **Gene**, *[S. l.]*, v. 626, p. 189–199, 2017. DOI: 10.1016/J.GENE.2017.05.032. Disponível em: https://www.sciencedirect.com/science/article/pii/S0378111917303657?via%3Dihub. Acesso em: 8 out. 2019.

RETTERER, Kyle et al. Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. **Genetics in Medicine**, *[S. l.]*, v. 17, n. 8, p. 623–629, 2015. DOI: 10.1038/gim.2014.160.

RICE, Alan M.; MCLYSAGHT, Aoife. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. **Nature Communications**, *[S. l.]*, v. 8, p. 14366, 2017. DOI: 10.1038/ncomms14366. Disponível em: http://www.nature.com/doifinder/10.1038/ncomms14366. Acesso em: 25 jan. 2019.

RIGGS, Erin Rooney et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). **Genetics in Medicine**, *[S. l.]*, v. 22, n. 2, p. 245–257, 2020. DOI: 10.1038/s41436-019-0686-8. Disponível em:

https://pubmed.ncbi.nlm.nih.gov/31690835/. Acesso em: 20 jul. 2020.

ROSELLÓ, Mónica; MARTÍNEZ, Francisco; MONFORT, Sandra; MAYO, Sonia; OLTRA, Silvestre; ORELLANA, Carmen. Phenotype profiling of patients with intellectual disability and copy number variations. **European Journal of Paediatric Neurology**, *[S. l.]*, v. 18, n. 5, p. 558–566, 2014. DOI: 10.1016/j.ejpn.2014.04.010. Disponível em: http://www.sciencedirect.com/science/article/pii/S1090379814000622?via%3Dihub. Acesso em: 18 jan. 2018.

SHAFFER, Lisa G. American College of Medical Genetics guideline on the cytogenetic evaluation of the individual with developmental delay or mental retardation. **Genetics in medicine : official journal of the American College of Medical Genetics**, *[S. l.]*, v. 7, n. 9, p. 650–654, 2005. DOI: 10.1097/01.gim.0000186545.83160.1e. Disponível em: http://www.ncbi.nlm.nih.gov/pubmed/16301868. Acesso em: 1 dez. 2017.

SIMÓN-SÁNCHEZ, Javier et al. Cooperative genome-wide analysis shows increased homozygosity in early onset Parkinson's Disease. **PLoS ONE**, *[S. l.]*, v. 7, n. 3, p. e28787, 2012. DOI: 10.1371/journal.pone.0028787.

THAPAR, Anita. Rutter's child and adolescent psychiatry. *[S. l.]*, [s.d.]. Disponível em: https://www.wiley.com/en-jp/Rutter%27s+Child+and+Adolescent+Psychiatry%2C+6th+Edition-p-9781118381960. Acesso em: 22 abr. 2023.

THAPAR, Anita; COOPER, Miriam; RUTTER, Michael. Neurodevelopmental disorders. **The Lancet Psychiatry**, *[S. l.]*, v. 4, n. 4, p. 339–346, 2017. DOI: 10.1016/S2215-0366(16)30376-5.

VISSERS, Lisenka E. L. M. et al. Array-Based Comparative Genomic Hybridization for the Genomewide Detection of Submicroscopic Chromosomal Abnormalities. **American Journal of Human Genetics**, *[S. l.]*, v. 73, n. 6, p. 1261, 2003. DOI: 10.1086/379977. Disponível em: /pmc/articles/PMC1180392/. Acesso em: 5 maio. 2023.

VORSANOVA, S. G.; YUROV, Y. B.; SOLOVIEV, I. V; IOUROV, I. Y. Molecular cytogenetic diagnosis and somatic genome variations. **Current genomics**, *[S. l.]*, v. 11, n. 6, p. 440–6, 2010. DOI: 10.2174/138920210793176010.

WAYHELOVA, Marketa et al. The clinical benefit of array-based comparative genomic hybridization for detection of copy number variants in Czech children with

intellectual disability and developmental delay. **BMC medical genomics**, *[S. l.]*, v. 12, n. 1, 2019. DOI: 10.1186/S12920-019-0559-7. Disponível em: https://pubmed.ncbi.nlm.nih.gov/31337399/. Acesso em: 29 ago. 2022.

YANG, Hsin Chou; CHANG, Lun Ching; HUGGINS, Richard M.; CHEN, Chun Houh; MULLIGHAN, Charles G. LOHAS: Loss-of-heterozygosity analysis suite. **Genetic Epidemiology**, *[S. l.]*, v. 35, n. 4, p. 247–260, 2011. DOI: 10.1002/gepi.20573.

ZARREI, Mehdi; MACDONALD, Jeffrey R.; MERICO, Daniele; SCHERER, Stephen W. A copy number variation map of the human genome. **Nature Reviews Genetics**, *[S. l.]*, v. 16, n. 3, p. 172–183, 2015. DOI: 10.1038/nrg3871. Disponível em: http://www.nature.com/doifinder/10.1038/nrg3871. Acesso em: 21 nov. 2017.

7 APÊNDICE A - FORMULÁRIO CLÍNICO

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CCB - Laboratório de Neurogenética do Desenvolvimento

FORMULÁRIO CLÍNICO

FICHA CLÍNICA DE PACIENTES ENCAMINHADOS PARA CGH *ARRAY*

1.**Número do CGH *array*** *:* _____ 2.Faixa etária do Paciente: _____

3.Sexo: ( ) Masculino ( ) Feminino

4.Peso: _____ kg 5.Estatura: _____cm 5.1 ( ) Crescimento longilíneo

4.1 ( ) Obesidade 5.2 ( ) Baixa estatura

4.2 ( ) Baixo peso 5.3 ( ) Crescimento Anormal

Obs:_____

6.HISTÓRIA PERINATAL

6.1 ( ) Prematuridade 6.2 ( ) Crescimento intra uterino restrito (CIUR)

6.3 ( ) Artéria Umbilical única 6.4 ( ) Oligoidrâmnio 6.5 ( ) Polidrâmnio

6.6 ( ) Asfixia Neonatal

6.7 Teste do Pézinho ( ) Sim ( ) Não Qual? ( ) Básico ( ) Ampliado ( ) Expandido

( ) Alterado:_____

7.EXAMES

7.1 ( ) Cariótipo: ( ) Normal ( ) Alterado

7.2 ( ) FISH: ( ) Normal ( ) Alterado

7.3 ( ) Metilação do DNA, Prader-Willi/Angelmann:

( ) Normal ( ) Alterado

7.4 ( ) X-frágil ( ) Normal ( ) Alterado

7.5 ( ) Metabólicos ( ) Normal ( ) Alterado

7.6 ( ) CGH*array* ( ) Normal ( ) Alterado

Obs:

8.Presença conhecida de casos de **DI** na família ( ) Não ( ) Sim. Descrever detalhes/parentesco no verso.

9.Presença conhecida de casos de **Transtorno psiquiátrico** na família ( ) Não ( ) Sim

10.Presença conhecida de casos de **Anomalias Congênitas** na família ( ) Não ( ) Sim

11.NEUROLÓGICO

11.1 ( ) Ataxia 11.2 ( ) Distonia 11.3 ( ) Coréia 11.4 ( ) Hipotonia

11.5 ( ) Defeito no Tubo Neural 11.6 ( ) Convulsões:_____

11.7 ( ) Estrutura Cerebral Anormal:_____

11.8 Outros:_____

12.MUSCULOESQUELÉTICO

12.1( ) Escoliose 12.2 ( ) Hérnia Diafragmática 12.3 ( ) Anomalia Vertebral Qual?_____

12.4 ( ) Anomalias Membros Superiores Qual?_____

12.5 ( ) Anomalias Membros Inferiores Qual?_____

12.6 Mãos ( ) Polidactilia ( ) Sindactilia uni ou bilateral, dedos envolvidos?_____

12.7 Pés ( ) Polidactilia, ( ) Sindactilia uni ou bilateral, dedos envolvidos?_____

12.8 Outros:_____

13.DESENVOLVIMENTO FÍSICO E MOTOR

13.1 ( ) Atraso no desenvolvimento não especificado 13.2 ( ) Atraso motor grosseiro

13.3 ( ) Atraso motor fino 13.4 Outros:_____

14.NEURODESENVOLVIMENTO

14.1 ( ) Atraso de Desenvolvimento Neuropsicomotor (abaixo de 5 anos)

14.2 ( ) Atraso de Fala 14.3 ( ) Dificuldade de aprendizado apenas

14.4 ( ) Deficiência Intelectual ( ) Leve ( ) Moderada ( ) Severa

14.5 ( ) Epilepsia

14.6 Outros:_____

15.COMPORTAMENTAL

15.1 ( ) Autismo 15.2 ( ) Síndrome de Asperger 15.3 ( ) Transtorno Obsessivo Compulsivo

15.4 ( ) Esquizofrenia 15.7 ( ) TDAH 15.5 ( ) Déficit de Atenção sem hiperatividade

15.8 Outros:_____

16.DISMORFIAS

16.1 ( ) Dismorfias Não Especificas:_____

16.2 ( ) Dismorfias Faciais: ( ) Face Assimetrica ( ) Grosseira ( ) Pequena ( ) Estreita ( ) Alongada ( ) Ampla ( ) Quadrada ( ) Plana ( ) Triangular ( ) Hipoplasia Maxilar ( ) Outros:_____

16.3 ( ) Lábio Leporino 16.4 ( ) Fenda Platina

16.5 Cranio: ( ) Craniosinostose ( ) Macrocefalia ( ) Microcefalia ( ) Plagiocefalia ( ) Braquicefalia 16.8 ( ) Orelhas: _____

16.9 ( ) Olhos: ( ) Hipertelorismo ( ) Hipotelorismo ( ) Epicanto ( ) Cílios invertidos ( ) Fenda ocular obliqua para cima ( ) Fenda ocular obliqua para baixo ( ) Outros (coloração, manchas, etc.)_____

16.10 ( ) Boca/Lábios:_____

16.11 ( ) Testa:_____

16.12 ( ) Sobrancelha:_____

16.13 ( ) Nariz:_____

16.14 ( ) Cabelo:_____

16.15 Circunferência da Cabeça:_____cm

16.16 ( ) Outros:_____

17.CARDÍACO

17.1 ( ) Defeito Do Septo Atrial 17.2 ( ) Defeito Do Septo Ventricular

17.3 ( ) Tetralogia de Fallot 17.4 ( ) Coarctação da aorta

17.5 ( ) Outra Anomalia/Malformação Cardíaca_____

18.GASTROINTESTINAL

18.1 ( ) Gastrosquise 18.2 ( ) Estenose Pilórica 18.3 ( ) Onfalocele

18.4 ( ) Megacólon 18.5 ( ) Outra Anomalia ou Malformação

Gastrointestinal:_____

19.GENITURINÁRIO

19.1 ( ) Genitália Ambígua 19.2 ( ) Malformação Renal

19.3 ( ) Hipospadia 19.4 ( ) Hidronefrose

19.5 ( ) Obstrução Uretral 19.6 ( ) Genitália Externa Anormal

19.7 ( ) Outra Anomalia ou Malformação

Genital:_____

20. ( ) ANORMALIDADES ENDÓCRINAS:_____

21.CUTÂNEO

21.1 ( ) Hiperpigmentação 21.2 ( ) Hipopigmentação 21.3 ( ) Hemangioma

21.4 Outros (nevos, sardas, manchas café-au-lait):_____

22. ALTERAÇÕES HEMATOLÓGICAS:_____

23. USO DE MEDICAMENTO CONTROLADO: ( ) Não ( ) Sim

_____

_____

24. MOTIVO DO ENCAMINHAMENTO:

_____

_____

Responsável pela coleta:_____

Médico:_____

Local:_____

Data: ___/___/_____

OUTRAS OBSERVAÇÕES:

_____

_____

_____

8 ANEXO A – PARECER DO COMITÊ DE ÉTICA EM PESQUISA

HOSPITAL INFANTIL JOANA DE GUSMÃO/ SES -SC

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: ANÁLISE E INTERPRETAÇÃO DAS VARIAÇÕES GENÔMICAS DETECTADAS POR CGH ARRAY EM PACIENTES COM DISTÚRBIOS DO DESENVOLVIMENTO.

Pesquisador: Angelica Francesca Maris

Área Temática: Genética Humana:
(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP;);

Versão: 2

CAAE: 74841617.0.0000.5361

Instituição Proponente: Hospital Infantil Joana de Gusmão/ SES - SC

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 2.339.104

Apresentação do Projeto:

A análise de CNVs no genoma humano com a tecnologia do CGH array é uma área de estudo relativamente recente, com um universo a ser explorado. A possibilidade de colaboração com um laboratório comercial de genética, com um hospital público e com um grupo de médicos que trabalhem com pacientes com distúrbios do desenvolvimento abre possibilidades que a academia, isolada, não fornece. A interpretação clínica de CNVs encontradas em pacientes com DDC ainda é um desafio e depende, em grande parte, das informações de frequência e detecção em coortes com amostragem significativa, como a que será analisada neste estudo. O enriquecimento dos bancos de dados compartilhados sustenta ou renova a compreensão sobre as variações do que é normal ou patogênico, possibilitando uma constante revisão, principalmente das alterações de patogenicidade incerta, as quais podem passar a ter outras interpretações à luz dos novos conhecimentos. Em particular pretende-se também explorar as possibilidades das novas plataformas de CGH array detectarem regiões de homozigose como alvos potenciais para encontrar mutações autossômicas recessivas ou síndromes de imprinting quando há indícios de consanguinidade ou UPD. Este trabalho vem a somar no esclarecimento da etiologia dos DDCs e contribuir para a interpretação de VOUs, buscando a colaboração de pesquisadores, médicos,diretores de laboratórios clínicos e

Endereço: Rui Barbosa, nº 152
Bairro: Agronômica **CEP:** 88.025-301
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3251-9092 **Fax:** (48)3251-9092 **E-mail:** cephijg@saude.sc.gov.br

bioinformatas. A coleta de dados clínicos com os médicos será realizada no HIJG, no HU. Para se tentar fazer a relação fenótipo-genótipo, no intuito de correlacionar possíveis genes causais, será coletada uma descrição clínica detalhada dos indivíduos afetados junto aos médicos através de um formulário pré-estabelecido (Apêndice A). Somente participarão médicos que assinarem o TCLE (Apêndice G). Buscar-se-ão informações sobre o quadro clínico de cada paciente, bem como seu histórico de exames físicos, moleculares, metabólicos, genéticos, comportamentais e uso de medicações. A identidade dos indivíduos não será coletada nos formulários; cada exame é identificado por um código personalizado. O laboratório juntamente com a equipe de pesquisa repassa este código com o nome do respectivo paciente diretamente ao médico, os pesquisadores trataram as análises de dados
apenas pelo código do exame. Para a coleta dos dados clínicos, será fornecido ao médico o nome e o código do paciente, referente ao exame em questão, o médico optar por responder o questionário diretamente (identificado pelo código do exame), com base no prontuário ou repassará as informações necessárias para preenchimento do formulário para um membro de
nossa equipe que poderá ir no consultório dele(a).

I)AMOSTRA: Resultados de aproximadamente 800 exames de CGH array, solicitados por médicos do Hospital Infantil Joana de Gusmão (HIJG), Hospital Universitário Professor Polydoro Ernani de Santhiago (HU) e clínicas particulares de Florianópolis, realizados através do laboratório Neurogene.

Trata-se de estudo retrospectivo-prospectivo. Além dos 109 exames provenientes do estudo
anterior do laboratório, serão analisados, aproximadamente, mais 700 exames de pacientes com DDC realizados no período de 2014 a 2019. Para os exames de 2013, será feita uma
reanálise. Médicos(as) do HIJG, HU e de clínicas particulares serão convidados a colaborar com a pesquisa através de um TCLE que esclarecera sua forma de participação na pesquisa, os pacientes que já realizaram o exame de CGH array serão convidados a participar da pesquisa através de TCLE apresentando a eles durante suas consultas rotineiras no HIJG e no HU, outro TCLE será apresentado em convite para aqueles que forem encaminhados para realização do exame de CGH array por solicitação de seus médicos. Nos casos em que os pacientes entre 6 à 18 anos, o Termo de Assentimento será apresentado aos pacientes, assegurando seu direito de participação ou recusa na pesquisa. Nos casos em que os pacientes já não são mais acompanhados pelo hospital, o contato para convite da pesquisa será realizado

através de telefone pelo médico responsável. Quando não for possível o contato com o paciente por algum motivo, a Justificativa Da Ausência Do TCLE assegurar a confidencialidade e a privacidade dos mesmos.

II) DADOS CLÍNICOS: A coleta de dados clínicos com os médicos será realizada no HIJG, no HU. Será coletada uma descrição clínica detalhada dos indivíduos afetados junto aos médicos através de um formulário pré-estabelecido. Buscar-se-ão informações sobre o quadro clínico de cada paciente, bem como seu histórico de exames. A identidade dos indivíduos não será coletada nos formulários; cada exame é identificado por um código personalizado. O laboratório repassa este código com o nome do respectivo paciente diretamente ao médico, os pesquisadores trataram as análises de dados apenas pelo código do exame. Será fornecido ao médico o nome e o código do paciente, referente ao exame em questão, o médico optar por responder o questionário diretamente, com base no prontuário ou por entrevista com um membro da equipe no consultório dele(a).

III)COLETA E ANÁLISES GENÔMICAS:
As coletas dos resultados dos exames de CGH array serão realizadas no Laboratório Neurogene, serão coletados arquivos digitais resultantes de leitura dos exames de CGH array e analisados com o software da Affymetrix®, dando-se ênfase para as CNVs raras (patogênicas ou VOUs).

IV)Análise de CNVs: Para analisar e interpretar as CNVs quanto a sua função, efeitos de dosagem e efeitos de mutações, serão utilizados bancos de dados públicos, privados e ferramentas de bioinformática. Todas as variáveis como: localização, tipo e tamanho de CNV, classificação da CNV, descrições clínicas (fenótipos), dados de exames, faixa etária, sexo do paciente, serão compiladas em folha simples de Excel e aplicadas estatísticas,
a fim de compreender a média e o desvio padrão do tamanho das CNVs, a frequência fenotípica, a taxa diagnóstica do estudo, a idade média e a porcentagem de cada sexo na coorte estudada, frequência de alterações genômicas em cada cromossomo e a relação do fenótipo ou grupos de fenótipos clínicos com a interpretação das CNVs, para estabelecer se há indícios que possibilitem reconhecer os pacientes mais indicados para submissão ao CGH array como teste de primeira linha. V)Análise de LCSH: A análise dos exames que apresentarem regiões de LCSH seguirá o Fluxograma 1. VI)Análise da consanguinidade: Para verificar a frequência de consanguinidade na coorte, será utilizada a metodologia baseada no

trabalhado de Kearney, Kearney e Conlin (2011).

Objetivo da Pesquisa:

Objetivo Primário:

Pretende-se interpretar as variações genômicas detectadas por CGH array em pacientes com distúrbios do neurodesenvolvimento, bem como levantar e discutir a taxa diagnostica do exame, o papel da consanguinidade nessa coorte, contribuir com o enriquecimento de informações diagnosticas disponíveis, e propor genes candidatos aos DDC.

Objetivos Secundários:

1) Analisar e caracterizar as CNVs encontradas no grupo de indivíduos investigados de acordo com seu tipo, localização, tamanho, genes presentes, e presença na população;

2) Tentar estabelecer relação fenótipo-genótipo;

3) Verificar se há características clínicas ou fenotípicas associados à maior detecção de CNVs patogênicas, delineando critérios para a seleção de pacientes a serem investigados com CGH array como teste de primeira linha;

4) Analisar a taxa e o papel da consanguinidade nessa coorte;

5) Com os resultados das análises, propor novas regiões e/ou genes candidatos causativos de DDC;

6) Identificar pacientes com potenciais mutações autossômicas recessivas para sugerir aos médicos a investigação por análise de exoma.

Avaliação dos Riscos e Benefícios:

Riscos:

As identidades dos médicos participantes dificilmente serão sigilosas, uma vez que são poucos os profissionais que solicitam estes exames, porém não serão citados nominalmente em publicações. Esta pesquisa tomará todo o cuidado possível para garantir o sigilo dos pacientes cujos dados são coletados – por isto a proposição da forma de realizar os questionários (vide metodologia). Porém, sabe-se que sempre existe a possibilidade de quebra de sigilo, mesmo que involuntário.

Benefícios:

Devido ao caráter acadêmico da pesquisa, o participante não terá nenhum benefício direto, no entanto os resultados desse estudo poderão no futuro proporcionar esclarecimentos sobre os aspectos de algumas alterações genômicas, estabelecendo com maior confiabilidade a taxa
diagnóstica obtida com este exame e contribuindo para o aconselhamento genético de resultados semelhantes. Caso na reanálise das alterações dos resultados do array CGH de um de seus

pacientes surjam dados novos sobre alguma alteração que a época do exame era de significado desconhecido, e que possam auxiliar no diagnóstico etiológico do paciente, iremos comunicar este fato a você.

Comentários e Considerações sobre a Pesquisa:

O estudo é relevante do ponto de vista social pelo conhecimento a ser gerado. O pesquisador apresentou informações que o credencia tecnicamente a executar o protocolo de pesquisa.

Considerações sobre os Termos de apresentação obrigatória:

Trata-se de resposta às pendências elencadas pelo CEP-HIJG, Número do Parecer: 2.285.889.

Conclusões ou Pendências e Lista de Inadequações:

(1) Os documentos referentes aos Termos de Concordância dos serviços de genética do HU e do HIJG estão assinados pela Pesquisadora/Orientadora quando devem ser assinados pelas respectivas chefias de serviço. Solicita-se adequação. Bem como apresentar os termos de concordância de todas as clínicas/instituições publicas e privadas participantes;

Resposta: Os documentos referentes aos Termos de Concordância dos serviços seguem as diretrizes dos documentos necessários para submissão de projetos de pesquisa que consta na pagina virtual do comitê de ética do HIJG, como “Documento de concordância do orientador e do serviço onde a pesquisa será realizada (anexo 3)”. Nesse sentido entendemos que é
necessária tanto a assinatura do orientador, quanto a dos responsáveis dos serviços dos locais que ocorreram as coletas de dados. Já previamente anexado ao referido projeto, estão os documentos de concordância do serviço do HIJG e do Laboratório Neurogene, devidamente assinados pela orientadora e pesquisadora responsável pelo projeto, pela Gerência Técnica do HIJG e pela responsável técnica da Neurogene. No entanto a concordância do serviço do HU apenas está assinada pela orientadora e pesquisadora responsável, pois no momento da submissão do referido projeto, o mesmo estava em análise pelo comitê de ética interna do HU, o qual concordou com coleta de dados na sua instituição conforme declaração assinada pela Gerente de Ensino e Pesquisa do HU-UFSC-EBSERH (Novo anexo “CONCORDANCIA_SERVICO_HU”). Conforme solicitado pelo CEP anexamos
a concordância do coordenador do programa de pós-graduação para que a pesquisa seja desenvolvida no Laboratório de Neurogenética da UFSC coordenado pela professora e pesquisadora responsável Dra. Angelica Francesca Maris, vinculada ao Programa de Pós-graduação em Biologia Celular e do Desenvolvimento como professora colaboradora (Novo anexo

(3) Conforme exposto na metodologia, solicita-se constar nos Termos de Consentimento Livre e Esclarecido que o médico do participante irá contribuir com a pesquisa informando os dados clínicos destes, que constam nos prontuários de atendimento permitindo que o participante se manifeste favorável ou não disponibilização desses dados.

Resposta: Conforme solicitado pelo CEP adicionamos a frase "O médico(a) que solicitou o exame de CGH array é participante dessa pesquisa e ele(a) irá contribuir informando e nos repassando seus dados clínicos que constam em seus prontuários de atendimento" no item PROCEDIMENTOS dos TCLEs.

Anexamos novamente como

"TCLE_PARTICIPANTES_ENCAMINHADOS_CGH_ARRAY_V2" e

"TCLE_PARTICIPANTES_REALIZARAM_CGH_ARRAY_V2".

Análise: Pendência atendida.

Considerações Finais a critério do CEP:

Conforme preconizado na Resolução 466/2012, XI.2, item d, cabe ao pesquisador elaborar e apresentar os relatórios parciais e final.

Assim sendo, o(a) pesquisador(a) deve enviar relatórios parciais semestrais da pesquisa ao CEP (a partir de ABRIL/2017) e relatório final quando do seu encerramento.

Um modelo deste relatório está disponibilizado no site http://www.saude.sc.gov.br/hijg/cep/deveresdopesquisador.htm

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_983177.pdf	22/09/2017 15:59:01		Aceito
Outros	Resposta_as_Pendencias.pdf	22/09/2017 15:57:04	TIAGO FERNANDO CHAVES	Aceito
Outros	CONCORDANCIA_COORDENADOR_PPGBCD.pdf	22/09/2017 15:52:17	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO_ASSNTIMENTO_6_a_11.pdf	21/09/2017 17:04:32	TIAGO FERNANDO CHAVES	Aceito

HOSPITAL INFANTIL JOANA DE GUSMÃO/ SES -SC

Plataforma Brasil

Continuação do Parecer: 2.339.104

TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO_ASSNTIMENTO_EXAMES_REALIZADOS_12_18.pdf	21/09/2017 17:02:42	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO_ASSENTIMENTO_ENCAMINHAMENTO_EXAMES_12_18.pdf	21/09/2017 17:02:33	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_PARTICIPANTES_REALIZARAM_CGH_ARRAY_V2.pdf	21/09/2017 17:02:24	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_PARTICIPANTES_ENCAMINHADOS_CGH_ARRAY_V2.pdf	21/09/2017 17:01:27	TIAGO FERNANDO CHAVES	Aceito
Outros	concodancia_servico_HU.pdf	21/09/2017 17:00:17	TIAGO FERNANDO CHAVES	Aceito
Projeto Detalhado / Brochura Investigador	PROJETO_DETALHADO.pdf	30/08/2017 17:11:53	TIAGO FERNANDO CHAVES	Aceito
Outros	CONCORDANCIA_SERVICO_NEUROGENE_ORIENTADOR.pdf	30/08/2017 15:44:32	TIAGO FERNANDO CHAVES	Aceito
Outros	CONCORDANCIA_SERVICO_NEUROGENE.pdf	30/08/2017 15:38:21	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	JUSTIFICATIVA_AUSENCIA_TCLE.pdf	30/08/2017 15:32:46	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO_ASSNTIMENTO_EXAMES_REALIZADOS.pdf	30/08/2017 15:25:54	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TERMO_ASSENTIMENTO_ENCAMINHAMENTO_EXAMES.pdf	30/08/2017 15:25:38	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_PARTICIPANTES_REALIZARAM_CGH_ARRAY.pdf	30/08/2017 15:25:17	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_PARTICIPANTES_ENCAMINHADOS_CGH_ARRAY.pdf	30/08/2017 15:25:02	TIAGO FERNANDO CHAVES	Aceito
TCLE / Termos de Assentimento /	TCLE_MEDICOS_COLABORADORES.pdf	30/08/2017 15:24:42	TIAGO FERNANDO CHAVES	Aceito

Endereço: Rui Barbosa, nº 152
Bairro: Agronômica **CEP:** 88.025-301
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3251-9092 **Fax:** (48)3251-9092 **E-mail:** cephijg@saude.sc.gov.br

HOSPITAL INFANTIL JOANA DE GUSMÃO/ SES -SC

Continuação do Parecer: 2.339.104

Justificativa de Ausência	TCLE_MEDICOS_COLABORADORES.pdf	30/08/2017 15:24:42	TIAGO FERNANDO CHAVES	Aceito
Outros	DECLARACAO_DIRECAO_HIJG.pdf	30/08/2017 15:23:46	TIAGO FERNANDO CHAVES	Aceito
Outros	CONCORDANCIA_SERVICO_HIJG.pdf	30/08/2017 15:20:51	TIAGO FERNANDO CHAVES	Aceito
Folha de Rosto	FOLHA_DE_ROSTO.pdf	30/08/2017 15:16:07	TIAGO FERNANDO CHAVES	Aceito
Outros	Instrumento_Coleta_Dados.pdf	28/08/2017 11:06:55	TIAGO FERNANDO CHAVES	Aceito
Outros	Declaracao_p_Publicacao.pdf	28/08/2017 11:01:36	TIAGO FERNANDO CHAVES	Aceito
Outros	concodancia_servico_HU_Orientador.pdf	28/08/2017 10:54:46	TIAGO FERNANDO CHAVES	Aceito
Outros	Concordancia_do_servico_HIJG_Orientador.pdf	28/08/2017 10:51:03	TIAGO FERNANDO CHAVES	Aceito
Outros	Compromisso_etico.pdf	28/08/2017 10:42:08	TIAGO FERNANDO CHAVES	Aceito
Outros	Carta_Encaminhamento_CEP.pdf	28/08/2017 10:40:35	TIAGO FERNANDO CHAVES	Aceito
Outros	Documento_Comprobatorio.pdf	28/08/2017 10:30:20	TIAGO FERNANDO CHAVES	Aceito
Declaração de Instituição e Infraestrutura	DECLARACAO_DA_INSTITUICAO.pdf	22/08/2017 13:20:54	TIAGO FERNANDO CHAVES	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

FLORIANOPOLIS, 19 de Outubro de 2017

Assinado por:
Vanessa Borges Platt
(Coordenador)

Endereço: Rui Barbosa, nº 152
Bairro: Agronômica **CEP:** 88.025-301
UF: SC **Município:** FLORIANOPOLIS
Telefone: (48)3251-9092 **Fax:** (48)3251-9092 **E-mail:** cephijg@saude.sc.gov.br

9 ANEXO B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO – MÉDICOS COLABORADORES

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Projeto de Pesquisa: "Análise e interpretação das variações genômicas detectadas por CGH array em pacientes com distúrbios do desenvolvimento."

PESQUISADORA RESPONSÁVEL: Angélica Francesca Maris

Informações: Pesquisadores da Universidade Federal der Santa Catarina estão realizando uma pesquisa para entender como alterações detectadas por CGH *array* estão relacionadas com o quadro clínico e fenotípico do paciente, qual a taxa diagnóstica deste exame em Santa Catarina, assim como verificar se há informações relevantes sobre a variabilidade genômica da população do Estado de Santa Catarina. Como parte deste estudo, você está sendo **convidado** a responder a **um questionário** que pretende coletar dados clínicos e fenotípicos sobre pacientes seus que realizaram o exame de CGH *array*. Sua participação é **opcional** e somente estaremos coletando informações para os casos onde os pacientes ou seus responsáveis legais **autorizaram,** através da assinatura de um termo, o laboratório Neurogene a publicar os resultados desde que dados pessoais sejam mantidos em sigilo. Através do Laboratório Neurogene, que realiza os exames, os dados são inseridos em bancos de dados públicos como o DECIPHER ou mais restritos como o CAGDB. Os resultados de nossa pesquisa poderão fazer parte de trabalhos de conclusão de curso, mestrado ou doutorado e se intenciona publicá-los em periódicos especializados. Deixamos claro que sua participação é **voluntária**. A equipe agradece antecipadamente a sua colaboração e estará a sua disposição para esclarecer dúvidas antes e durante o desenvolvimento da pesquisa. Para isto você pode contatar a pesquisadora responsável, Dra. Angelica Francesca Maris através dos telefones **(48)3721-9887**, **(48)99119-9760** ou e-mail: **afmaris@gmail.com**

Procedimentos: Se você concorda em participar desta pesquisa, você terá que responder um **questionário** semi-estruturado, de quatro páginas, que poderá ser respondido diretamente por você, sobre dados clínicos (alterações de exames neurológicos, metabólicos, genéticos, histórico perinatal, de desenvolvimento, medicações, presença de histórico familiar) e fenotípicos (dismorfologias, por exemplo) do paciente em questão. Não serão solicitados nomes ou dados de identificação do paciente. Cada exame é identificado por um código personalizado. O laboratório juntamente com os pesquisadores repassaram este código com o nome do respectivo paciente diretamente a você, enquanto os pesquisadores utilizaram apenas o código do exame no questionário. Você poderá optar por responder o questionário (identificado pelo código do exame) diretamente ou um membro de nossa equipe poderá ir a seu local de atendimento. Você responderão então ao questionário em entrevista, repassando as informações necessárias com base no prontuário. Os pesquisadores irão tratar a identidade com padrões profissionais de sigilo.

Riscos: As identidades dos médicos participantes dificilmente serão sigilosas, uma vez que são poucos os profissionais que solicitam estes exames, porém não serão citados nominalmente em publicações. Esta pesquisa tomará todo o cuidado possível para garantir o sigilo dos participantes cujos dados são coletados – por isto

a proposição da forma de realizar os questionários. Porém, sabe-se que sempre existe a possibilidade de quebra de sigilo, mesmo que involuntário. Responder aos questionários pessoalmente ou por entrevista demanda tempo e isto pode se tornar um fator de aborrecimento.
Custos: Você não precisará pagar nada para fazer parte desse estudo.
Benefícios: Devido ao carácter acadêmico da pesquisa, o participante não terá nenhum benefício direto, no entanto os resultados desse estudo poderão no futuro proporcionar esclarecimentos sobre os aspectos de algumas alterações genômicas, estabelecendo com maior confiabilidade a taxa diagnóstica obtenível com este exame e contribuindo para o aconselhamento genético de resultados semelhantes. Caso na reanálise das alterações dos resultados do CGH *array* de um de seus pacientes participantes da pesquisa, surjam dados novos sobre alguma alteração que a época do exame era de significado desconhecido, e que possam auxiliar no diagnóstico etiológico do participante, iremos comunicar este fato a você.

Este termo de consentimento encontra-se impresso em duas vias, sendo que uma cópia será arquivada pelo pesquisador responsável, e a outra será fornecida ao (a) Sr (a).

Pesquisador responsável:

_____

DECLARAÇÃO DE CONSENTIMENTO

Eu, _____, fui esclarecido(a) sobre a pesquisa "Análise e interpretação das variações genômicas detectadas por CGH array em pacientes com distúrbios do desenvolvimento."

Florianópolis, _____ de _____ de _____.

Assinatura:_____ CRM:_____/____

Em caso de dúvidas com respeito aos aspectos éticos deste estudo, você poderá consultar o:
Comitê de Ética em Pesquisa do Hospital Infantil Joana de Gusmão.
Florianópolis- Santa Catarina.
TELEFONE (48) 3251 9092
EMAIL: cephijg@saude.sc.gov.br

10 ANEXO C – TERMO DE CONSENTIMENTO NEUROGENE

NEUROGENE LABORATÓRIO S/S LTDA
Rua Alves de Brito, 346, sala 201- CEP 88015-440
Centro – Florianópolis – SC
Fone/Fax: (48)-30281286 (48)-3223-0229 (48)-7812-1130
www.neurogene.com.br

Termo de Consentimento Informado

Eu,_____

_____residente e domiciliado em

_____

portador do RG:_____e CPF:_____

nascido (a) em _____/____/_____, abaixo assinado (a), concordo de livre e espontânea vontade em fornecer amostra de sangue minha e/ou de meu (minha) filho (a) para a realização do exame de array – CGH.

Declaro que obtive todas as informações necessárias, bem como todos os eventuais esclarecimentos quanto as dúvidas por mim apresentadas, Entendo que:

1 - O objetivo deste exame é verificar a presença de perdas ou ganhos de material genômico na amostra submetida.
2 - O exame é necessário para que uma possível causa da doença genética em questão seja identificada mas não há nenhuma certeza de que a causa seja encontrada.
3 Esse exame não oferece tratamento terapêutico.
4 Este exame não identifica alterações cromossômicas equilibradas, mutações de ponto (gênicas) ou alterações no DNA mitocondrial
5 Se for identificada alguma variante cromossômica não descrita será realizado exame dos pais para comparação, com custo adicional.
6 Embora esse exame seja confiável e realizado com controle de qualidade, pode não ser informativo em alguns casos.

1- Eu _____ autorizo a divulgação no laudo final de todos os resultados identificados no exame, ou seja, TODOS os achados genomicos serão divulgados e não somente os relacionados com a clinica do paciente.

2- Eu _____autorizo a divulgação dos resultados em publicações científicas, desde que dados pessoais sejam mantidos em sigilo.

3 – Eu NÃO autorizo a divulgação dos resultados em publicações científicas.

ASSINATURA:-

_____

11 ANEXO D – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO - PARTICIPANTES

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

PROJETO DE PESQUISA: "Análise e interpretação das variações genômicas detectadas por CGH *array* em pacientes com distúrbios do desenvolvimento".

PESQUISADORA RESPONSÁVEL: Angélica Francesca Maris

Este projeto de pesquisa está sendo conduzido por uma equipe de pesquisa coordenada pela Dra. Angelica Francesca Maris do Laboratório de Neurogenética do Desenvolvimento do Departamento de Biologia Celular, Embriologia e Genética da Universidade Federal de Santa Catarina, juntamente com o Serviço de Genética do Hospital Infantil Joana e Gusmão, estamos tentando entender melhor como alterações detectadas pelo exame genético CGH *array* estão relacionadas com as dificuldades, doenças ou malformações que levam os médicos a considerar uma causa genética. Pretendemos averiguar em quantas vesses é possível achar a causa dessas alterações realizando o exame de CGH *array* no estado de Santa Catarina, assim como verificar se há informações importantes sobre a diversidade genética dessa população.

Como **você ou seu filho(a) já realizaram** CGH *array*, você ou seu filho(a) estão sendo convidados(as) a participar de modo opcional dessa pesquisa. **Opcional significa que você pode se recusar a participar deste pesquisa, sem qualquer prejuízo ou constrangimento em relação ao atendimento que recebem junto ao(s) seu(s) médicos(as).**

Este estudo é voluntário, e incluirá somente pessoas que escolherem fazer parte da mesma. **Por favor, decida com calma** e peça ao responsável pela pesquisa ou a equipe da pesquisa para explicar a você quaisquer palavras neste documento que você não compreenda e antes de assinar certifique-se de que tudo foi respondido a sua satisfação. Sinta-se livre para discutir a informação deste documento com seus amigos, família ou com seu médico.

PROCEDIMENTOS: Se você concordar em participar desta pesquisa, ou que seu filho(a) participe desta pesquisa, vamos estudar os resultados do seu teste de CGH

12 ANEXO E – TERMO DE ASSENTIMENTO – 6 A 11 ANOS

TERMO DE ASSENTIMENTO

Termo de assentimento para criança e adolescente (maiores de 6 anos e menores de 11 anos)

Assentimento Informado para participar da pesquisa: "Análise E Interpretação Das Variações Genômicas Detectadas Por CGH *Array* Em Pacientes Com Distúrbios Do Desenvolvimento".

Olá! Tudo bem contigo?

Eu sou aluno(a) da professora Angelica Francesca Maris da Universidade Federal de Santa Catarina e estou aqui para te convidar a participar de um estudo que estou fazendo sobre as mudanças no DNA mostradas nos exames (CGH *array)*. Quero estudar crianças e adolescentes, que assim como você, fizeram esse exame médico para ver uma pequena mudança no DNA.

O DNA que estudamos fica dentro do sangue de cada pessoa e é semelhante a "letrinhas" que "dizem" como devemos crescer, a cor dos nossos olhos, cor do cabelo e como cada um é. Esse exame (CGH *array)* que **você fez ou ira realizar** mostra se há algumas mudanças nessas "letrinhas".

Eu quero entender melhor o que isso pode causar a você e a essas outras crianças, e também tentar descobrir porque essas mudanças acontecem.

Seus pais já concordaram em participar desse estudo e já assinaram um papel parecido com este, mas isso não quer dizer que você tem que participar. Você pode escolher se vai deixar que eu fale com seu médico e estude os seus exames! Para isso, você pode conversar com a sua família e comigo antes de escolher se vai participar ou não.

Outras crianças que, assim como você, realizaram o exame CGH *array*, também irão participar deste estudo.

Você não precisa participar do estudo se não quiser, é um direito seu e você nem ninguém terão nenhum problema se desistir.

O estudo será realizado no Hospital Infantil Joana de Gusmão ou no Hospital Universitário Professor Polydoro Ernani de São Thiago, onde você e outras crianças

já foram atendidas por médicos. Você não precisará fazer mais nenhuma consulta para participar.

Se você quiser fazer parte desse estudo, eu entrarei em contato com seu médico, conversarei com ele sobre os sintomas que ele observou em você, vou ter acesso aos exames que você já fez, enfim, saberei de todas as coisas que ele sabe e anotou sobre você. Mas lembre-se, eu não contarei a ninguém!

E se você decidir participar, eu tomarei todos os cuidados para que nem o seu nome e nem o nome de sua família apareça em nenhum lugar. Apenas nossa equipe de estudo poderá ver seus exames, falar com o(a) seu médico(a) e estudar as doenças que você pode ter.

Mesmo com todo o cuidado que tomaremos há a possibilidade de que outras pessoas saibam que você está participando desse estudo. E você poderá sentir vergonha nisso.

Você não ganhara nada para participar desse estudo, mas também não terá nenhum gasto.

Quando terminarmos o estudo, poderemos mostrar ele em reuniões cientificas, revistas de cientistas e poderão ser entregue a outros cientistas, porém não falarei o seu nome, nem de seus familiares.

Se você tiver alguma dúvida, você pode me perguntar. Caso aconteça algo errado, ou você desista de participar desse estudo, você pode nos procurar pelos telefones (48)3721-9887, (48)99119-9760 da pesquisadora Angelica Francesca Maris.

Ou você poderá pedir informações diretamente ao **Comitê de Ética e Pesquisa** do Hospital Infantil Joana de Gusmão pelo telefone 48 – 32519092

Um Comitê de Ética em Pesquisa em Seres Humanos (CEP) é composto por um grupo de pessoas que estão trabalhando para garantir que seus direitos como participante de pesquisa sejam respeitados. Ele tem a obrigação de avaliar se a pesquisa foi planejada e se está sendo executada de forma ética. Se você achar que a pesquisa não está sendo realizada da forma como você imaginou ou que está sendo prejudicado de alguma forma, você pode entrar em contato com o CEP do Hospital Infantil Joana de Gusmão pelo telefone 48 – 32519092 ou pelo email: cephijg@saude.sc.gov.br . Você pode inclusive fazer a reclamação sem se identificar, se preferir.

Você que decide se ira participar. E ninguém ficará bravo ou desapontado com você caso você diga não. **A ESCOLHA É SUA.** Você pode pensar nisto e falar depois se você quiser. Ou você pode dizer sim agora e mudar de idéia depois e tudo continuará bem.

Eu entendi que o estudo é sobre os resultados nos exames de CGH *array* que podem mostrar mudanças no meu DNA e se há algo errado nele. Também entendi que se eu aceitar participar do estudo, os resultados dos meus exames médicos serão escritos juntamente com as mudanças do meu DNA. **EU ACEITO PARTICIPAR DESSE ESTUDO.**

ESTE TERMO DE CONSENTIMENTO ENCONTRA-SE IMPRESSO EM DUAS VIAS, sendo que uma cópia será arquivada pelo pesquisador responsável, e a outra será fornecida ao (a) Sr (a).

_____

NOME E ASSINATURA DO CRIANÇA/ADOLESCENTE DATA

_____

NOME E ASSINATURA PAIS/RESPONSÁVEIS DATA

_____

NOME E ASSINATURA DO INVESTIGADOR DATA

13 ANEXO F – TERMO DE ASSENTIMENTO – 12 A 18 ANOS

TERMO DE ASSENTIMENTO

Termo de assentimento para criança e adolescente (maiores de 12 anos e menores de 18 anos)

Assentimento Informado para participar da pesquisa: "Análise E Interpretação Das Variações Genômicas Detectadas Por CGH *Array* Em Pacientes Com Distúrbios Do Desenvolvimento".

PESQUISADORA RESPONSÁVEL: Angélica Francesca Maris

Olá, sou pesquisador(a) orientado(a) da professora e pesquisadora responsável Dra. Angelica Francesca Maris da Universidade Federal de Santa Catarina e estou fazendo uma pesquisa sobre alterações no DNA mostradas nos exames de CGH *array*. O DNA é nossa identidade biológica, uma sequência de nucleotídeos semelhantes a "letrinhas" que ficam dentro do sangue e esses nucleotídeos que "dizem" ao organismo em desenvolvimento como se construir. Esse exame (CGH *array)* mostra se há algumas diferenças maiores no DNA.

Você está sendo convidado(a) para participar da pesquisa "Análise E Interpretação Das Variações Genômicas Detectadas Por CGH *Array* Em Pacientes Com Distúrbios Do Desenvolvimento" .

Seus pais permitiram que você participe mas você pode conversar com eles ou comigo antes de aceitar.

Queremos saber se os resultados do exame genético CGH *array*, **que você realizou**, para ver se há alterações nessa sequencia semelhante a "letrinhas" do seu DNA. Outras crianças que, assim como você, apresentaram essas alterações e realizaram o exame CGH *array*, também irão participar desta pesquisa.

Você não precisa participar da pesquisa se não quiser, é um direito seu e você nem ninguém terão nenhum problema se desistir.

A pesquisa será feita no Hospital Infantil Joana de Gusmão ou no Hospital Universitário Professor Polydoro Ernani de São Thiago, onde você e outras crianças já foram atendidas por médicos. Você não precisará fazer mais nenhuma consulta para participar. Se você decidir participar, nós iremos usar as informações que os

médicos já colheram de você para realizar a nossa pesquisa. Se você decidir participar, a equipe de pesquisa tomará todos os cuidados para que nem o seu nome e nem o nome de sua família apareça em nenhum lugar. Essa pesquisa é sigilosa, isso quer dizer que somente nossa equipe poderá analisar os seus exames, falar com o(a) seu médico(a) e estudar os sintomas que você tem. Mesmo com todo o cuidado que tomaremos há a possibilidade de que outras pessoas saibam que você está participando dessa pesquisa.

Você não terá nenhum benefício, e não receberá nada para participar dessa pesquisa, mas há possibilidades de que no futuro conheça-se mais sobre as alterações que ocorrem no seu DNA.

Quando terminarmos a pesquisa, poderemos mostrar os resultados em reuniões cientificas, revistas de cientistas e esses resultados poderão ser publicados, porém não divulgaremos o seu nome, nem de seus familiares.

Se você tiver alguma dúvida, você pode me perguntar. Caso aconteça algo errado, ou você desista de participar dessa pesquisa, você pode nos procurar pelos telefones (48)3721-9887, (48)99119-9760 da pesquisador/a Angelica Francesca Maris.

Ou você poderá pedir informações diretamente ao **Comitê de Ética e Pesquisa** do Hospital Infantil Joana de Gusmão pelo telefone 48 – 32519092

> *Um Comitê de Ética em Pesquisa em Seres Humanos (CEP) é composto por um grupo de pessoas que estão trabalhando para garantir que seus direitos como participante de pesquisa sejam respeitados. Ele tem a obrigação de avaliar se a pesquisa foi planejada e se está sendo executada de forma ética. Se você achar que a pesquisa não está sendo realizada da forma como você imaginou ou que está sendo prejudicado de alguma forma, você pode entrar em contato com o CEP do Hospital Infantil Joana de Gusmão pelo telefone 48 – 32519092 ou pelo email: cephijg@saude.sc.gov.br . Você pode inclusive fazer a reclamação sem se identificar, se preferir.*

Sua participação é voluntária. E ninguém ficará bravo ou desapontado com você caso você diga não. **A ESCOLHA É SUA.** Você pode pensar nisto e falar depois se você quiser. Ou você pode dizer sim agora e mudar de idéia depois e tudo continuará bem.

Eu entendi que a pesquisa é sobre os resultados nos exames de CGH *array* que podem mostrar as alterações no meu DNA e se há algo modificado nele. Também compreendi que se eu concordar em fazer parte dessa pesquisa significa que as informações de meus exames médicos serão descritas juntamente com as alterações do meu DNA. **EU ACEITO PARTICIPAR DESSA PESQUISA.**

ESTE TERMO DE CONSENTIMENTO ENCONTRA-SE IMPRESSO EM DUAS VIAS, sendo que uma cópia será arquivada pelo pesquisador responsável, e a outra será fornecida ao (a) Sr (a).

_____

NOME E ASSINATURA DO CRIANÇA/ADOLESCENTE DATA

_____

NOME E ASSINATURA PAIS/RESPONSÁVEIS DATA

_____

NOME E ASSINATURA DO INVESTIGADOR DATA

14 ANEXO G – JUSTIFICATIVA DA AUSÊNCIA DO TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

JUSTIFICATIVA DA AUSÊNCIA DO TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Conforme assegura a Resolução 466/12, do Conselho Nacional de Saúde, em seu IV artigo que versa sobre o Consentimento Livre e Esclarecido:

> "O respeito devido à dignidade humana exige que toda pesquisa se processe após consentimento livre e esclarecido dos participantes, indivíduos ou grupos que, por si e/ou por seus representantes legais manifestem a sua anuência à participação na pesquisa".

Entretanto, a Resolução 466/2012 reconhece algumas situações especiais (IV.7 e IV.8) em que o TCLE pode ser dispensado, devendo o pesquisador (a) apresentar justificativa com as causas da impossibilidade de obtê-lo. Essa justificativa deve ser apresentada em documento anexo ao projeto de pesquisa e o CEP julgará sua pertinência.

Eu, Angelica Francesca Maris, pelo presente termo, solicito ao Comitê de Ética em Pesquisa do Hospital Infantil Joana de Gusmão (HIJG) a DISPENSA do Termo de Consentimento Livre e Esclarecido, em razão da pesquisa intitulada "Análise E Interpretação Das Variações Genômicas Detectadas Por CGH *Array* Em Pacientes Com Distúrbios Do Desenvolvimento", apresentar caráter retrospectivo para aqueles participantes que já realizaram o exame de CGH *array* anterior ao início do estudo, por se tratar de levantamento de dados clínicos (alterações de exames neurológicos, metabólicos, genéticos, histórico perinatal, de desenvolvimento, medicações, presença de histórico familiar) e fenotípicos (dismorfologias, por exemplo) para auxiliar na interpretação de variações genéticas dos referidos participantes. Essas informações serão coletadas por questionários juntos aos prontuários de pacientes que já não estão mais sendo acompanhados pelo serviço de genética do referido hospital, residem em outras localidades ou por algum motivo estão incontactáveis, os quais nomes e identificações serão mantidos em sigilo, em conformidade com o que prevê os termos da Resolução 466/12 do Conselho Nacional de Saúde e da declaração assinada pela direção do HIJG, autorizando a realização da pesquisa.

Assumo mediante este Termo, o compromisso de, ao utilizar dados e/ou informações coletadas no(s) prontuários do(s) participantes (s) da pesquisa, assegurar a confidencialidade e a privacidade dos mesmos.

Florianópolis, 25 de agosto de 2017.

Angelica Francesca Maris

Prof.ª Dra. Angelica Maris
Lab Neurogenética do Desenvolvimento - CCB/UFSC
SIAPE nº 1807224