

UNIVERSITY OF PAVIA - IUSS School for Advanced Studies Pavia

**Department of Brain and Behavioral Sciences (DBBS)
MSc in Psychology, Neuroscience and Human Sciences**



**UNIVERSITÀ
DI PAVIA**



IUSS

Analysis of Protein-Altering Variants in Motor-Based Autism Subgroups: Implications for Understanding Pathophysiology

Supervisor: Prof. Sara Mascheretti

Co-Supervisor: Dr. Alessandro Crippa

Thesis written by
Mattia Diana 531059

Academic year: 2024/2025

Contents

Abstract	4
1 Introduction	5
1.1 Historical & Epidemiological Background	5
1.2 Diagnosis & Symptomatology	10
2 Materials and Methods	19
2.1 Variants Identification	20
2.2 Gene Sets Definition	21
2.3 Evaluation of Gene Sets Variants Abundance	21
2.4 Gene Set Variants Enrichment Analysis	22
2.5 Gene Sets and Subjects Clustering	22
2.6 Gene Expression in the Brain During Neurodevelopmental Stages	22
2.7 Genes Spatio-Temporal Co-Expression in the Developing Brain	23
2.8 Genes Protein–Protein Interactions	23
2.9 Incidence of Autism-Related Genes	23
2.10 Statistical Analysis	24
3 Results	24
3.1 Gene Sets Showing Differential PAV Enrichment Between Motor-Based Autism Subgroups	26
3.2 Genes in the Modules Are Expressed in the Brain in the Developmental Period	30
3.3 Identification of Extended Modules Through Spatio-Temporal Co-Expression and Protein–Protein Interactions	31
3.4 SFARI Genes Are Overrepresented in the Original & Extended Modules .	32
4 Discussion	36

4.1	TC1 — Sensory Visual Dysfunction (SVD)	37
4.2	TC2 — Peripheral Inflammation (PI)	38
4.3	TC3 — Neurodevelopment & Behavioral Dysregulation (N-BD)	39
4.4	TC4 — Cellular Pruning Mediated by Neuroinflammation (CPMN)	40
4.5	TC5 — Neurogenetic Transcriptional Regulation & Movement Alteration (NTR-MA)	41
4.6	Integration Across Modules: Spatio-Temporal co-Expression and Protein–Protein Interactions	43
4.7	SFARI Genes are Overrepresented in the Original and Extended Modules	44
4.8	Limitations	45
5	Conclusion	47
	Appendix	49
	References	50

Abstract

Autism spectrum disorder (ASD) is highly heterogeneous, with motor functioning being a core yet underexplored domain possibly influencing phenotypic variability. Understanding the genetic underpinnings remains a significant challenge. In the following study, we investigated whether a multi-step, data-driven analysis could identify gene sets with differing loads of protein-altering variants (PAVs) between two subgroups of autistic children stratified by high versus low motor risk as assessed via the Developmental Coordination Disorder Questionnaire (DCDQ). From the larger project "Progetto Mosaico", a total of 62 children (mean age = 8.24 years) were chosen, with DNA extracted from blood or saliva samples. Gene set variant enrichment analysis was conducted to identify sets with significantly different PAV incidence between groups. These sets were clustered into functional modules, and their spatio-temporal brain expression was examined using the BrainSpan Atlas of the Developing Human Brain. Modules were further extended by incorporating genes co-expressed or physically interacting with the original clusters, and enrichment for ASD susceptibility genes was evaluated. Results revealed distinct modules associated with neurodevelopmental processes, synaptic communication, motor-related pathways and other biological processes. These modules displayed specific expression patterns across developmental stages, supporting their potential contribution to phenotypic variability within ASD. Overall, our findings highlight the relevance of motor-related genetic architecture in shaping autism heterogeneity and suggest that non-core motor features can serve as objective anchors for subtyping, providing novel insights into the biological mechanisms underlying ASD diversity.

Keywords: Autism Spectrum Disorder (ASD); motor functioning; heterogeneity; protein-altering variants (PAVs); gene set enrichment analysis.

1 Introduction

1.1 Historical & Epidemiological Background

Autism Spectrum Disorder (ASD) is a condition included within the broader category of Neurodevelopmental Disorders (NDD), characterized by deficits in social interaction and excessive, restricted and repetitive behaviors (American Psychiatric Association, 2022). The use of the term *spectrum* reflects the highly heterogeneity across its phenotypic presentation, on distinct levels of analysis both cultural and social, as well as at a deeper level of genetic and neurophysiology (Lord et al., 2020; Lord et al., 2022; Masi et al., 2017).

The first use of the term as a diagnostic label, is attributed to Kanner (1943) after conducting a study on a group of 11 children aged 2-10 years, whose behavior exhibited recurring characteristics within the group, such as *hyper-sensitivity to stimuli* or *delayed echolalia*. Over time, scientific research has substantially advanced our understanding of this intricate topic, leading to the overcoming of outdated theories such as Bettelheim's "*Refrigerator Mother*" hypothesis, which erroneously attributed the child's difficulties to the mother's emotional coldness (Bettelheim, 1999), in favor of a more comprehensive and multifaceted description of the phenomenon (Jacob et al., 2019).

Moreover, there is an ongoing debate among different stakeholders about whether autism should be considered a medical disability or a different manifestation of human behavior and a diversity in their neurophysiology - a concept classified as *neurodiversity*. On a basic etymological level, neurodiversity may be defined as "*the idea that the human brain can be wired in a variety of ways, none of them inherently superior*" (Antze, 2010). On this topic, the discussion refers to the prominent neurodiversity movement and autistic advocacy, which argue for a de-pathologized view of autism and instead emphasize a social identity grounded in rights-based advocacy and a distinct social model of development (Milton, 2017). Autism thus becomes increasingly associated with well-established conceptualizations of individual differences within developmental science, rather than being framed solely as a pathological category (Green, 2023), consequently moving further away from the over-medicalization and a deficit-centered view, that the Western biomedical community tends to cherish (Anderson-Chavarria, 2021). At the same time, many autistic individuals have profound needs and are vulnerable to harm and marginalization, and many

struggle with their difficulties, feeling excluded from society (Lord et al., 2022), leading some to an inquiry of a complete cure or at least something to ease their symptomatology. Since being autistic comes with a plethora of issues such as, a significant increase in depression and suicide rates (Casten et al., 2023; Hedley & Uljarević, 2018; Hirvikoski et al., 2016; O'Halloran et al., 2022), co-occurrence to other psychiatric conditions (e.g: schizophrenia spectrum disorders, affective disorders, anxiety disorders) (American Psychiatric Association, 2022; Jokiranta-Olkonien et al., 2016; Lai et al., 2019) as well as physiological ones (e.g: epilepsy and gastrointestinal disorders) (Davignon et al., 2018) or intellectual disabilities (Clark & Adams, 2022), having the correct accommodation or something that could "cure" this, might be quite appealing to some.

It is crucial to emphasize, that our objective should lie in the middle of these discussions, focusing on actually listening to autistic voices, rather than strictly following a classical medical strategy and widen our approach, taking in consideration unbiased observations and therapeutical techniques in order to destigmatize and to promote a better understanding of ASD people. To paraphrase Professor Davide Liccione (2019), autistic individuals possess a "*way of being in the world*" that is totally different from non-autistic ones. Therefore, we should not try to impose the "typical" way of existing that corresponds to the experiences of the majority of people. On the contrary, we should try to understand these different modalities of existence, thus allowing for better mediation between the world of an autistic person and that of other people.¹

Before delving deeper into the main topic, it is essential to establish the underlying framework, in order to ensure a comprehensive understanding of the topic. Following the *DSM-V-TR* (2022) definition, NDDs are a group of conditions with onset during the developmental period which typically manifest early in childhood. They are characterized by developmental deficits or differences in brain processes that result in impairments in different domains (E.g.: *personal, social, academic, or occupational*). In addition to ASD, among them we can find other disorders, such as *Attention Deficit/Hyperactivity Disorder (ADHD)*, *Specific Learning Disorder (SLD)* and *Language Disorder (LD)*. While, NDDs exhibit significant heterogeneity regarding their clinical features, etiology, responses to treatment, and outcomes, grouping them under the same classification proves to be clinically useful for both assessment and intervention (Thapar et al., 2017).

¹Due to lack of time and space, I will not go into detail about the perspective described by professor Liccione, for further information, please refer to his work.

Beyond clinical complexity, trying to capture the societal dimension is equally challenging. Consequently, epidemiological data on NDDs often display substantial variability, especially when considering that they encompass a broad group of disorders rather than a single, uniform condition. Some estimates suggest a prevalence of 2-4% of the world population (Sabariego-Navarro et al., 2022), while others indicate global fluctuations between 4.70% and 88.50% (Francés et al., 2022). The authors of this review explain that these vast discrepancies may be attributed to differences in diagnostic criteria, assessment tools, and socio-contextual factors, which have led to inconsistencies across studies and countries. The inconsistency in prevalence rates could also be significantly influenced by the common occurrence of comorbidities among NDDs. For example, although ASD has been associated with different NDDs, ADHD is consistently recognized as one of the most common comorbid disorders, with a prevalence ranging from 50% to 70% (Rong et al., 2021). It is well established that NDDs frequently co-occur (Lundström et al., 2015; Thapar et al., 2017), and although the etiology of this co-occurrence is not well understood, some observations have implicated shared genetic liability between neurodevelopmental conditions (Hegemann et al., 2024). The overlapping symptomatology between these different conditions often complicates the diagnostic process, leading to the potential over- or under- estimation of specific disorders (Castro et al., 2026).

ASD's data follow a similar trend. One of the first epidemiological reports was recorded by Victor Lotter (1966) in the United Kingdom, where he identified a prevalence of 4.5 cases per 10,000 children. Over time, numerous studies have reported widely varying results. Globally, the prevalence of ASD is estimated to be around 1% (Castro et al., 2026; Zeidan et al., 2022), although the figures vary depending on region and methodology. In their meta-analysis, Salari et al., (2022) found that the estimated global prevalence was 0.6%, with regional variations of 0.4% in Asia, 1% in the Americas and Africa, 0.5% in Europe, and up to 1.7% in Australia. In the United States, data from the Autism and Developmental Disabilities Monitoring show a substantial increase in prevalence among 8-year-old children, from 6.7 per 1,000 in 2000 (ADDM & CDC, 2007) to 18.5 per 1,000 in 2016 (Maenner et al., 2020), within the 11 monitored sites.² More recent analysis shows that the prevalence was 23.0 per 1,000 (*one in 44*) children aged 8 years, with a

²Locations available on <https://www.cdc.gov/autism/addm-network/sites.html>

male-to-female ratio of 4.2:1 (Maenner et al., 2021).

In Italy, epidemiological evidence shows trends that appear to be consistent with those of other western countries. A study carried out in the province of L'Aquila using data from the Autism Register of the Regional Reference Center for Autism (CRRRA), reported an overall crude prevalence of 0.95% for individuals born between 2001 and 2015. Notably, their results reveal that the cumulative prevalence of autism diagnosis remains largely stable over time, suggesting that observed changes are primarily due to earlier diagnosis rather than an actual increase in cases (Valenti et al., 2019). Similarly, another study conducted in the province of Pisa estimated a prevalence of 1.15% school-aged children (*7-9 years old*), aligning with the European average. However, when only certified diagnosis were taken into consideration the rate was approximately 0.79%, demonstrating how diagnostic access and reporting systems used during studies can influence the prevalence estimates (Narzisi et al., 2018). Moving towards the national level, the Italian Ministry of Health in collaboration with the Istituto Superiore di Sanità and the Ministry of Education reported a prevalence of 1.30% children aged 7–9 years between 2016 and 2018 (Scattoni et al., 2023) with a male-to-female ratio (4.4:1) consistent with data from the U.S. (Baio et al., 2018; Maenner et al., 2021). More importantly, Scattoni and colleagues found regional variability among the country, with higher rates in central Italy and lower rates in the south. As explained by the writers, this pattern may be influenced by socio-economic and environmental factors, as well as disparities in healthcare access (Scattoni et al., 2023). Overall, these findings suggest that Italy's prevalence data for ASD are comparable to international figures, yet there exists some variability that persists across regions and age groups, underlying the need for consistent national surveillance systems and harmonized diagnostic procedures.

Furthermore, beyond geographical and methodological differences, another critical issue concerning the persistent sex bias in autism diagnosis can be observed from the aforementioned data. Even though other studies estimate a lower prevalence (3:1) (Loomes et al., 2017), the sex difference remains notable. While the reason for the male preponderance it's not completely understood, it is likely to be multifactorial. The literature about this topic appears to be quite contradictory, as some studies observe sex differences in the symptomatology (Werling & Geschwind, 2013), while others don't (Nasca et al., 2020). Nevertheless, a point of convergence among different stakeholders

is that autistic males tend to exhibit higher levels of externalizing behaviors—such as aggression, hyperactivity, and especially restricted or repetitive behaviors—compared to autistic females (Breach & Lenz, 2023; Wiggins et al., 2021). This, combined with generally subtler socio-communicative difficulties among women (Lai et al., 2011) may be one of the reason for this underrepresentation.

Another reason for this effect may reside in the so called camouflaging, defined as one's ability to mask their typical autistic behaviors that are considered socially unacceptable and/or performatively offer socially acceptable behaviors (Lai et al., 2017). Examples of camouflaging include: increasing eye contact during conversation, using learned phrases or pre-prepared jokes in conversation, mimicking other's social behavior, imitating facial expressions or gestures, and learning and following social scripts (Lai & Baron-Cohen, 2015). This pattern seems to be more present in women (Rynkiewicz et al., 2016), and might be a result of the higher social pressure that women go through. In addition, adolescent autistic females have been shown to be better at using subtle language skills compared to males and the use of more vivid gestures (Corbett et al., 2021). Taken together, these elements contribute to a diagnostic framework that remains insufficiently sensitive to the female autism phenotype (Hull et al., 2020). Importantly, this disproportionate under-identification also results in less research conducted in this group to address the issue of sex/gender-related differences in functioning or their unique clinical manifestation.

Furthermore, significant limitation concerns the inconsistent findings surrounding the characterization of the female ASD phenotype (Young et al., 2018). In most studies, the women included have received an ASD diagnosis based on criteria primarily developed from the male presentation of the condition (Tubío-Funqueiriño et al., 2021). As a result, the available evidence on the female ASD phenotype may fail to capture potential behavioral and symptom-related differences that diverge from the traditional ASD profile (Hiller et al., 2014). Nonetheless, progress has been made in addressing these challenges. For example, clinical services specifically tailored for autistic women have begun to emerge, along with the development of assessment tools designed to analyze autistic traits in women. These include the Questionnaire for Autism Spectrum Conditions (*Q-ASC*), which incorporates more gender-sensitive criteria (Simcoe et al., 2018), and the Camouflaging Autistic Traits Questionnaire (*CAT-Q*), which evaluates camouflaging behaviors (Hull et al., 2019).

1.2 Diagnosis & Symptomatology

According to the *DSM-V-TR* (2022), as of today, the latest edition of the manual, the diagnostic criteria for ASD are defined as follows:

A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by all of the following, currently or by history (*examples are illustrative, not exhaustive; see text*):

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following, currently or by history (*examples are illustrative, not exhaustive; see text*):

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g., simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).

2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior (e.g., extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).
3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).
4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual developmental disorder (intellectual disability) or global developmental delay. Intellectual developmental disorder and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual developmental disorder, social communication should be below that expected for general developmental level.

According to this description of symptoms, in ASD, the primary deficits mainly fall into two distinct areas: social interaction and communication (criterion A) and motor skills and interests (criterion B). Respectively, we can observe behaviors such as the inability to uniformly use gaze, facial expressions, and gestures to regulate social interaction; repetitive and stereotyped use of words or phrases out of context; compulsive adherence to behavioral routines or rituals and repetitive interests that are unusual in both content

and nature (Barone, 2019). These deficits can already be observed in children from the second year of life. Interestingly, these two seemingly diverse aspects, actually interact with one another.

The potential role of motor skills in ASD

Motor skills in children at risk for developing ASD are associated with delays in communication abilities at 18 months of age (Bhat et al., 2012), as well as difficulties in processing gaze direction and facial expressions at ages 5-7 (Leonard & Hill, 2014). Similarly, limited abilities in gross motor skills (*i.e., movement involving large muscle groups*) have been shown to correlate with poorer interpersonal coordination (Fitzpatrick et al., 2017) and a reduced participation in social activities in children and adolescents (Jarus et al., 2011). While both of these aspects strongly characterizes autistic individuals, albeit with different levels of impact on daily functioning, among the scientific community there has traditionally been greater attention to the social domain than to the motor one. Most studies have focused on high-level *theory of mind* abilities (*i.e., the capability to impute mental states to oneself and to others, usually referred to as “mindreading” or “mentalizing”*) (Heyes & Frith, 2014; Koster-Hale & Saxe, 2013), and only in recent years has the motor dimension begun to receive increased consideration.

Among motor-related features, motor stereotypies (MS) represent some of the most commonly observed behaviors in early childhood (*typically before age three*) and occur in autistic and non-autistic children, and vary in intensity from mild to safety-concerning (Melo et al., 2020). They consist of habitual rhythmic and repetitive movements of fingers, hands, or other body parts (e.g, bilateral flapping or rotating hand movements, finger flicking, jumping, arm waving, or head nodding) (Chawarska et al., 2025). MS are defining symptoms of stereotypic movement disorder (SMD) and occur frequently with a variety of neurodevelopmental conditions, including ASD and intellectual disability (ID), other neurodevelopmental conditions, as well as in children without frank social or cognitive concerns (Fernandez et al., 2023; Keating et al., 2023; Larkin et al., 2017; Uljarević et al., 2017). Within this framework, MS, as well as other specific motor symptoms, may indicate a common denominator across different disorders, encouraging the development of a transdiagnostic investigations aimed at improving diagnostic precision to inquiry the possibility of a shared pathophysiology (Pacheco et al., 2022; Walther & Mittal,

2022). Moreover, large cohort analyses (e.g., SPARK) report that 85–87% of autistic children score in the at-risk range for motor impairment on parent-report screening, and population-based research has found similarly high rates (Bhat, 2020; Bhat et al., 2022; Ketcheson et al., 2021; Licari et al., 2020).

Beyond MS, autistic children commonly present difficulties in visuomotor coordination, balance, gait, fine-motor dexterity, and motor planning across childhood and adolescence, with limited improvement over time (Bhat, 2020). Importantly, motor measures, both clinician-observed and parent-reported, predict core ASD features (*social-communication delays, repetitive behavior severity*) and co-occurring outcomes (*language, cognitive, and functional delays*), even after accounting for age, sex, and, to some extent, cognitive level (Bhat, 2021; Bhat et al., 2022; Ketcheson et al., 2021). The etiology of these phenomena remains uncertain, but current evidence suggests that biological and genetic mechanisms play a key role. Indeed, neurobiological and genetic evidence indicates shared mechanisms underlying these atypical features. For instance, animal and human studies implicate the basal ganglia and cerebellar dysfunction, as well as disruptions in *corticostriatal-thalamo-cortical circuits* (Canales & Graybiel, 2000; Dean et al., 2022; Péter et al., 2017). Furthermore, volumetric striatal alterations and imbalanced corticostriatal connectivity have also been reported (Abbott et al., 2018; Augustine et al., 2021; Langen et al., 2009; Mahone et al., 2016). In addition, dysregulated dopaminergic and GABAergic signaling in relevant brain regions contributes to MS expression, as well as other atypical motor behaviors (Harris et al., 2016). Moreover, different studies suggests the presence of familial aggregation and genetic associations that further support the biological basis of these effects (Cantor et al., 2018; Evans et al., 2017; Fernandez et al., 2023).

Beside these diverse clinical/empirical observation, the study of motor function could offer several advantages for ASD research. Mainly, the motor system can be more easily assessed in comparison to other complex cognitive function, given the limited demands of motor task (Crippa et al., 2021). Lastly, motor abilities can be quantified precisely using methods such as kinematic analysis, offering therefore more objective measures in comparison with questionnaires or qualitative behavioral observation (R. B. Wilson et al., 2018). For instance, Crippa et al., (2015), used a Support Vector Machine (SVM) model to analyze data from a reach-to-grasp task in preschool children with ASD, finding that

seven of the 17 kinematic features measured were sufficient to classify autism with 96.7% accuracy. In light of these different evidence reported by the scientific community, and being motor impairments pervasive across the spectrum, different authors suggest to posit a more central diagnostic role and even considering motor impairment as an ASD specifier to better integrate motor assessment and intervention into clinical care (Belmonte, 2022; Chawarska et al., 2025; Pacheco et al., 2022; Walther & Mittal, 2022).

Noticeably, within most of the literature cited in this work, a trend appears among them: their differentiation criteria for ASD is primarily focused on how the core symptoms of social-communication (SC) and restricted repetitive behavior (RRB) manifest amidst the autistic population, rather than non-core features such as language, intellectual, motor, and adaptive functioning (as an acronym 'LIMA'). As observed by Happé and Frith (2020), this core-symptom-centered perspective may have been adequate in the past; however, today ASD is mainly characterized by people with varying and/or subtler core SC and RRB issues, but whom are much more likely to not have profound symptoms in LIMA. In light of the shifting perspective, some authors (Mandelli, Severino, et al., 2024) suggests that autism should be conceptualized as a set of distinct autisms primarily differentiated by the non-core LIMA features. Starting from the already mentioned debate on the terminology (*disability or diversity*) they propose an initial distinction between two highly differentiated early developmental and behavioral types: a form of profound disability (Type I) and a form of difference (Type II). These two types are hypothesized to diverge most prominently across development. By bringing together these three defining dimensions -disability versus difference over development- they introduce the AUTISMS-3D framework (A3D). Quoting directly the authors:

"The A3D model suggests that when one focuses on the non-core LIMA features, autism can be thought of as at least two different types that contrast disability versus difference over development."³

The key aspects of this model are the distinction of the two autism types, but mainly how these characteristics develop over time. Another way that it separates itself from other approaches is not relying on subjective expert clinical judgement. Although expert clinical judgment remains the gold standard for diagnosing autism itself, A3D predicts

³Mandelli, Severino, et al., 2024, p. 2

that, when it comes to identifying subtype distinctions, those based on LIMA features will manifest as robust statistical differences that unsupervised, data-driven models can detect with remarkably high generalizability. Without going into too many details, what we can draw from this is that the motor domain, as well as the other traits, are getting more attention showing how they can be helpful in diagnostic procedures, in scientific research and reducing ASD heterogeneity. In line with this view, empirical evidence shows that motor functioning allows the identification of distinct autism subgroups that differ systematically in developmental trajectories. Such findings support the A3D assumption that non-core LIMA features capture meaningful dimensions of variability within the autism spectrum and can therefore serve as objective anchors for distinguishing subtle and intricate differences over development (Mandelli, Landi, et al., 2024).

Etiopathological aspects of ASD

As briefly mentioned before, ASD etiology represents one of the most intricate points of research on this topic. Although still not completely understood, some aspects are undeniably linked to autism, such as genetic factors. Even though genetics is not the only factor involved -for instance, environmental influences also contribute- it undoubtedly plays a key role in its representation. In fact, the heritability of the condition is quite high, with estimates ranging between 64% and 91% (Bai et al., 2019; Huguet et al., 2016), with some studies reporting variability between males and females (*around 87.0% for males and 75.7% for females*) (Sandin et al., 2024). Over the past years, the use of genetic analyses as a lens to investigate ASD has gained increasing popularity. For instance, different authors analyzed the possible genetic link between ASD and other comorbidities/symptoms. As stated before in this work, ASD individuals are more susceptible to suicidal ideations and, even among the general population, suicide risk is partly genetic in its nature, with heritability estimates ranging from 17% to 55% (Vaquero-Lorenzo & Vasquez, 2020). However, a study conducted by Williams and colleagues (2021) highlights a potential overlap between the biology of suicide and autism, identifying several mutations in the NRXN1 gene, which are known to be a risk factor for autism (Kasem et al., 2018; Südhof, 2017; Williams et al., 2019). Furthermore, thanks to continuous advancement in biotechnology, hundreds of de-novo and rare inherited variants have been identified as major contributors to individual autism risk (Choi & An, 2021; De Rubeis et al., 2014;

Satterstrom et al., 2020), and some studies report paternal origins (Huguet et al., 2013).

To clarify, de-novo variants are defined as:

*A term used to describe a change in the DNA sequence of a gene that is seen for the first time in a person and has not appeared in previous generations.*⁴

Nevertheless, several studies have shown that also common genetic variants -those frequently found in the general population- contribute substantially to autism heritability, albeit with a smaller effect compared to de novo and rare inherited variants (Gaugler et al., 2014; Grove et al., 2019). Consequently, it's safe to assume that autism is highly heritable and polygenic, with both rare and common variants playing crucial roles (Cirnigliaro et al., 2023). Over the last decade, research has increasingly exploited this genetic heterogeneity to dissect phenotypic diversity, linking specific genotypes to distinct clinical features of autism (Bertelsen et al., 2021; Di Giovanni et al., 2023; Jeste & Geschwind, 2014). Regardless of these findings, it's important to note that, even though it has been observed that the burden of de novo mutations affecting genes expressed in the brain is higher in individuals with ASD, none of the identified genes individually account for more than 1% of ASD individuals (Huguet et al., 2013). Many of the genes associated with ASD are typically involved in common biological processes and pathways, such as regulation of gene expression (*e.g., chromatin modification and alternative splicing*), brain development, synaptic neuronal communication, and the formation and function of neurons and synapses (Duda et al., 2018; Shen et al., 2020).

One of the most widely used approach to study gene expressions, namely individual gene analysis (IGA), focuses on individual genes and their significance between two groups of samples compared (Nam & Kim, 2008). Given the complexity and heterogeneity of this disorder, one of the main challenges to tackle when designing studies through the genetic lens, is the need of thousands of participants in order to try to identify the functional consequences of autism-related genes. Another issue may stand in the difficulties in analyzing single genes individually and their expression on the macroscopic scale. IGA typically yields a list of altered genes from a cutoff threshold (Nam & Kim, 2008), however this gives rise to some limitations; first, the cutoff threshold is usually chosen arbitrarily, which in term affects the final result of IGA (Breslin et al., 2004), second, many genes with

⁴Definition available at:

<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/de-novo-variant>

a minor/moderate expression changes are discarded because of the chosen cutoff value, which leads to a reduction in statistical prowess (Ben-Shaul et al., 2005; Breslin et al., 2004).

Considering these different important aspects, some alternative approach arised in order to overcome the aforementioned limitations. With this interest in mind, gene set analysis (GSA) becomes applicable to different types of genetical studies. GSA shifts the focus from single genes to networks of genes converging on biologically relevant processes. Gene sets are composed of genes sharing a common function, defined based on various criteria such as membership in certain biological pathways or being co-expressed together under a certain condition (Maleki et al., 2020; Subramanian et al., 2005), which are gathered into different databases.⁵ The primary aim of gene set analysis is to identify enrichment (*more expression activity*) or depletion (*less expression activity*) of expression levels of a given gene set (Maleki et al., 2020).

Following the previous studies by Scaccabarozzi et al., (2025), this work explores whether a multi-step unbiased analysis could identify gene sets relevant to autism motor-subtyping and functionally characterise them in terms of biological processes. To better understand our base framework and the methodology utilized in our study, it would be reasonable to further introduce the foremost research. Scaccabarozzi and colleagues investigated if data-driven analysis could identify sets of genes with differential loads of protein-altering variants (PAVs) between two subgroups of autistic children stratified by intelligence quotient (IQ). Dividing a sample of 71 autistic children aged 3–12 years into a higher IQ subgroup (IQ >80, n=43) and a lower IQ subgroup (IQ ≤80, n=28), the authors screened participants' DNA through targeted next-generation sequencing of approximately 5,200 disease-associated genes and performed a gene set variants enrichment analysis (GSVEA) across nearly 30,000 functionally characterised gene sets. Significant gene sets were hierarchically clustered into functional modules, whose brain expression profiles were characterised using the BrainSpan Atlas of the Developing Human Brain, and each module was further extended by incorporating genes showing both physical protein-protein interactions and significant spatio-temporal co-expression in the developing brain.

⁵Examples of database: MSigDB (Subramanian et al., 2005), GeneSigDB (Culhane et al., 2012), and GeneSetDB (Araki et al., 2012)

Their analysis yielded 38 significant gene sets (FDR $q < 0.05$), clustering into four functional modules involving *ion cell communication*, *neurocognition*, *gastrointestinal function*, and *immune system*, each displaying distinct brain expression patterns across neurodevelopmental stages. When extended through co-expressed interactors, high-confidence SFARI autism susceptibility genes were significantly overrepresented within these modules (OR=2.296, $p=0.0001$), providing preliminary evidence that autism phenotypic variability likely originates from multiple interacting biological pathways rather than independent disrupted cascades.

Building on this framework, since we intended to evaluate gene sets' cumulative effects, we considered all PAVs, regardless of both their frequency in the population and their effect on the proteins. In order to achieve it, we posed different objectives. Firstly, we divided the clinical sample of autistic children into two subgroups based on their standardized scores of high/low motor risk. Using a data-driven enrichment analysis we then identified gene sets with diverse load of PAVs between participants. These acquired gene sets were clustered into modules based on their PAVs enrichment and the biological processes that best characterized them. Subsequently, we named them to facilitate explanations and analysis. Next, to investigate the functional relevance of the identified modules, we examined their expression patterns across brain regions and developmental stages using the BrainSpan Atlas of the Developing Human Brain⁶ and, lastly, we assessed whether genes within the modules, as well as those in the extended set of co-expressed interacting genes, were enriched for candidate genes strongly associated with autism susceptibility based on data from the Simons Foundation Autism Research Initiative database⁷, which continuously integrates genetic evidence from multiple research studies.

⁶Available at <https://www.brainspan.org/>

⁷Available at <https://www.simonsfoundation.org/>

2 Materials and Methods

Participants were recruited at the Child Psychopathology Unit of Scientific Institute, IRCCS Eugenio Medea (Bosisio Parini, Italy), over a 36-month period between July 2016 and July 2019, as a part of a larger study (“Progetto Mosaico”; Crippa et al., 2021). Autistic children were admitted to inpatient/outpatient units of our institute either for assessment or for rehabilitation programs. From this pool of participants, a total of 62 autistic children (mean age = 8.24 years, SD = 1.94) were selected for inclusion in the present study based on the availability of Developmental Coordination Disorder Questionnaire data. All participants had been previously diagnosed at E. Medea institute on the basis of a consensus “best estimate” DSM-5 clinical diagnostic process informed by, but not dependent on, scores on the Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) (Lord et al., 2012).

All participants also underwent an IQ evaluation, with specific assessment tools chosen according to participants’ developmental level among the Griffiths Mental Development Scales (Griffiths, 1984), Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) (Wechsler, 2002), and Wechsler Intelligence Scale for Children-IV (WISC-IV) (Wechsler, 2012). Exclusion criteria included: well-defined genetic disorder, use of medication affecting the central nervous system, the presence of significant sensory impairment (e.g., blindness, deafness), abnormalities detected by MRI, and suffering from chronic or acute medical illness. All participants were drug-naïve. All of the participants’ parents or legal guardians signed the informed written consent before the children’s participation.

Participants’ motor skills were assessed using the Movement Assessment Battery for Children 2 (*MABC2*; (Biancotto et al., 2017)), the Developmental Coordination Disorder Questionnaire (*DCDQ*; (B. N. Wilson et al., 2007)), and the kinematic analysis of a reach-to-drop task (Crippa et al., 2015; Forti et al., 2011).

Among the different investigated features, in the current analysis, participants were divided into two subgroups based on parental report of motor coordination difficulties as assessed by the DCDQ. Specifically, children scoring below the 10th percentile were classified as at risk for Developmental Coordination Disorder (*DCD-risk*), whereas those scoring at or above this threshold were classified as not at risk (*NO DCD-risk*). The DCDQ is a parent questionnaire originally developed for estimating gross and fine motor skill

difficulties in children aged 5–15 years. It consists of 15 items investigating different motor subdomains, such as ball skills, complex motor coordination, and fine/general motor skills. For each item, parents rate the children's degree of motor coordination on a 5-point scale comparing it with same-age peers. The DCDQ yields a total score, with higher scores indicating better motor functioning, and high internal consistency (0.89) has been reported (B. N. Wilson et al., 2009). Age-specific cutoffs define the 10th percentile threshold: a score below 47 for children aged 5–7 years 11 months, below 56 for children aged 8–9 years 11 months, and below 58 for children aged 10–15 years.

Blood or saliva samples from participants were obtained in-house at the Molecular Biology Laboratory of our Institute for DNA extraction. Blood was collected in tubes with EDTA. Genomic DNA was extracted from blood samples through the GenElute Blood Genomic DNA kit (Sigma). DNA from each sample was eluted in 50 μ L of 10 mM Tris (*pH*9.0)/0.5mM EDTA and stored at -20°C. The Oragene OG-500 kit (DNA Genotek) was used to collect saliva samples. Genomic DNA was extracted from saliva samples by precipitation in ethanol following the manufacturer's protocol, resuspended in 50 μ L of 10 mM Tris (*pH*9.0)/1mM EDTA and stored at -20°C.

Lastly, information about the familial socioeconomic status (SES) was collected using the Hollingshead scale for parental employment (Hollingshead, 1975). It is crucial to note, before proceeding further, that certain data points are absent, due to the inability to conduct the requisite tests during the initial data collection phase.

2.1 Variants Identification

Targeted next-generation sequencing was employed to screen participants' DNA. The SureSelect Focused Exome panel was selected for its targeted design, which restricts sequencing to approximately 5,200 disease-associated genes. Library preparation was carried out from genomic DNA using the SureSelect enrichment system (Agilent Technologies), and sequencing was performed on the NextSeq platform following the manufacturer's specifications (Illumina, San Diego, CA, USA). Alignment of sequenced reads to the reference target regions and variant calling were carried out with the BWA enrichment application, which incorporates GATK for variant calling, on BaseSpace Onsite (Illumina, San Diego, CA, USA). Calls identified as erroneous were subsequently removed. Given that spurious variant calls were retrieved for female participants, the Y

chromosome was excluded from downstream analyses.

Functional annotation of identified variants was performed with ANNOVAR (Wang et al., 2010), using a refGene database built by converting the human reference genome (GRCh37/hg19) GFF3 and FASTA files via the gff3toGenePred package (Grüning et al., 2018). Variants were classified as protein-altering (PAVs) if they were located within the exonic regions of the panel and annotated as indels, stop codon gains, stop codon losses, or non-synonymous substitutions. To quantify variants for each participant, we applied two complementary filtering criteria: (i) variant frequency in the general population (rare variants defined as absent from, or with frequency below 0.01 in, the gnomAD database; (Karczewski et al., 2020)); and (ii) clinical pathogenicity (variants classified as Pathogenic or Likely Pathogenic according to ACMG-AMP guidelines, as retrieved from the InterVar database; (Li & Wang, 2017)).

2.2 Gene Sets Definition

A total of 29,227 gene sets were assembled by querying four publicly available biological databases: Gene Ontology (Ashburner et al., 2000) (version 2019-03), the Human Phenotype Ontology (Köhler et al., 2019) (version 2019-03), the KEGG Pathway Database (Kanehisa et al., 2010) (release 2019-02-25), and the Reactome Pathway Database (Fabregat et al., 2018) (Physical Entity Identifier mapping files, updated to 2019-03). Each gene set groups together genes sharing a common biological function, pathway membership, or disease annotation.

2.3 Evaluation of Gene Sets Variants Abundance

To account for differences in zygosity across participants, the number of variants V attributable to a gene set S within a group of subjects G was computed as:

$$V = \frac{\left(\sum_{i \in G} \sum_{g \in S_A} V_{ig} \right) \left(\sum_{i \in G} \sum_{g \in S_A} 2L_g \right) + \left(\sum_{i \in G} \sum_{g \in S_X} V_{ig} \right) \left(\sum_{i \in G} \sum_{g \in S_A} N_i L_g \right)}{\sum_{i \in G} \sum_{g \in S_A} 2L_g + \sum_{i \in G} \sum_{g \in S_A} N_i L_g} \quad (1)$$

V_{ig} denotes the number of alternative alleles detected for subject i in gene g , L_g is the coding length of gene g , and N_i equals 2 for female subjects and 1 for male subjects. The summations range over all subjects in G and all genes in the set S , partitioned into autosomal genes (S_A) and X-linked genes (S_X). An enrichment score for a given group or individual subject was then defined as the variant count derived from Equation (1), normalized by the total number of variants computed across all genes in the exome panel.

2.4 Gene Set Variants Enrichment Analysis

The gene set variants enrichment analysis (GSVEA) aimed to determine whether the abundance of PAVs in each gene set differed significantly between the DCD-risk and NO DCD-risk subgroups. For each gene set S , a Fisher’s exact test was applied to the following contingency table:

	Variants in gene set	Variants not in gene set
DCD-risk	V_{G1}	$V_{A1} - V_{G1}$
NO DCD-risk	V_{G2}	$V_{A2} - V_{G2}$

where V_{G1} and V_{G2} are computed via Equation (1) for the specific gene set, while V_{A1} and V_{A2} are computed over all genes in the exome panel. When testing multiple gene sets simultaneously, p -values were corrected for multiple comparisons using false discovery rate (FDR) estimation, and q -values are reported throughout.

2.5 Gene Sets and Subjects Clustering

Hierarchical clustering was applied uniformly throughout the analysis using the Ward algorithm (Murtagh & Legendre, 2014) on Euclidean distances. The optimal number of clusters was selected in each case via silhouette analysis (Rousseeuw, 1987).

2.6 Gene Expression in the Brain During Neurodevelopmental Stages

The spatial and temporal expression profiles of genes in the identified modules were characterized using the ABAEnrichment R package (version 1.20.0; (Grote et al., 2016)), which interfaces with the BrainSpan Atlas of the Developing Human Brain (Allen Institute

for Brain Science, retrieved June 2026). The BrainSpan Atlas organizes data across five neurodevelopmental stages: prenatal, infancy (0–2 years), childhood (3–11 years), adolescence (12–19 years), and adulthood (>19 years). Given the early-onset nature of autism, the present analyses were restricted to the first four stages.

2.7 Genes Spatio-Temporal Co-Expression in the Developing Brain

For every gene included in the exome panel, an expression vector was constructed by extracting, from the BrainSpan Atlas, the expression measurements recorded across all available brain structures during the first four developmental stages. Co-expression between any two genes was quantified using the Pearson correlation coefficient (r) applied to their respective expression vectors. Associated p -values were corrected for multiple testing (FDR), and gene pairs were considered co-expressed when both $r > 0.7$ and $q < 0.05$.

2.8 Genes Protein–Protein Interactions

Physical protein–protein interactions between genes were retrieved from the BioGRID database (Stark et al., 2006). Any gene sharing at least one documented physical interaction with a gene already belonging to one of the modules was classified as a direct interactor. Each module was then expanded by incorporating genes satisfying both the co-expression criterion defined above and the physical interaction criterion, yielding a set of extended modules.

2.9 Incidence of Autism-Related Genes

To evaluate the relevance of the identified modules to autism susceptibility, we examined the proportion of genes annotated with a gene score of 1 (High Confidence) in the SFARI database (Simons Foundation, retrieved June 2026) within both the original modules and their extended networks. Statistical significance was assessed via Fisher’s exact tests, considering the full exome panel as the background gene set; where multiple comparisons were involved, Bonferroni correction was applied.

2.10 Statistical Analysis

Between-group comparisons of continuous demographic and clinical variables were performed using independent-samples t -tests. Where Levene's test indicated a violation of the equal variances assumption ($p < .05$), Welch's correction was applied to the degrees of freedom. The chi-squared test (χ^2) was used for the categorical variable sex. All statistical tests were two-tailed, and a significance threshold of $\alpha = 0.05$ was adopted. Where multiple comparisons were performed, p -values were corrected using the False Discovery Rate (FDR) procedure and reported as q -values. Statistical analyses were performed in R (version 4.1; R Core Team, 2013).

3 Results

Table 1 provides the descriptive statistics and between-group comparisons for the two subgroups of autistic children classified as DCD-risk ($n = 43$) and NO DCD-risk ($n = 19$), as well as for the total sample ($N = 62$; mean age = 8.24 years, $SD = 1.94$). The sample was predominantly male (54 males, 8 females), consistent with the established sex ratio reported in the epidemiological literature on ASD (Maenner et al., 2021; Scattoni et al., 2023). The two subgroups did not differ significantly in age ($t = -1.138$, $p = .260$), IQ ($t = -0.445$, $p = .658$), SES ($t = 0.486$, $p = .629$), or ADOS comparison scores ($t = -0.530$, $p = .599$). The distribution of males and females did not vary significantly between groups ($\chi^2 = 0.138$, $p = .711$), indicating comparable demographic characteristics across the two subgroups.

Overall, the absence of significant between-group differences in the main demographic and clinical variables confirms that the two subgroups are comparable with respect to potential confounders, allowing for the interpretation of any genetic findings as more specifically related to the motor risk stratification criterion.

	Group	Total		Mean	SD	Statistic	P-value
		M	F				
GENDER	NO DCD-risk	17	2			0.138 ^b	0.711
	DCD-risk	37	6				
AGE (years)	NO DCD-risk	19		7.82	2.35	-1.138 ^a	0.260
	DCD-risk	43		8.43	1.74		
IQ	NO DCD-risk	18		86.78	25.68	-0.445 ^a	0.658
	DCD-risk	42		89.52	20.11		
SES	NO DCD-risk	17		60.88	18.64	0.486 ^a	0.629
	DCD-risk	42		58.33	18.10		
ADOS (comparison)	NO DCD-risk	15		6.13	1.85	-0.530 ^a	0.599
	DCD-risk	36		6.42	1.70		

Table 1. Descriptive statistics and between-group comparisons subdivided per DCD risk. *Note.* SD = Standard Deviation; IQ = Intelligence Quotient; SES = Socioeconomic Status; ADOS = Autism Diagnostic Observation Schedule. ^aIndependent samples t-test (Welch's correction applied where Levene's test was significant, $p < .05$); ^bChi-square test. H_a : NO DCD-risk $\neq$ DCD-risk. All tests two-tailed, $\alpha = .05$.

3.1 Gene Sets Showing Differential PAV Enrichment Between Motor-Based Autism Subgroups

Following this preliminary check, a gene set variants enrichment analysis (GSVEA) was performed across approximately 29,000 functionally characterized gene sets, assembled from four publicly available biological databases: Gene Ontology (GO; Ashburner et al., 2000), the Human Phenotype Ontology (HPO; Köhler et al., 2019), the KEGG Pathway Database (Kanehisa et al., 2010), and the Reactome Pathway Database (Fabregat et al., 2018). The goal of the analysis was to determine whether the abundance of protein-altering variants (PAVs) observed in a given gene set differed significantly between the DCD-risk and NO DCD-risk subgroups.

Among the considered gene sets, 33 showed a significantly different incidence of PAVs between the two subgroups of autistic participants (FDR, $q < 0.05$). For each subject, the proportion of PAVs in each significant gene set — the enrichment score — was calculated. An optimal hierarchical clustering procedure (Ward algorithm applied to Euclidean distances; Murtagh and Legendre, 2014) was then applied to the resulting matrix at both the gene set and subject level (see Figure 1). The significant gene sets clustered into five terms' clusters (TC), labelled according to the biological processes and phenotypic features most representative of each cluster's composition as follows:

1. **Sensory Visual Dysfunction** (*SVD*)
2. **Peripheral Inflammation** (*PI*)
3. **Neurodevelopment & Behavioral Dysregulation** (*N-BD*)
4. **Cellular Pruning Mediated by Neuroinflammation** (*CPMN*)
5. **Neurogenetic Transcriptional Regulation & Movement Alteration** (*NTR-MA*)

Three subject clusters (SC) were additionally identified at the participant level. The data-driven identification of these five modules was based on the initial top-down splitting of our sample in two subgroups on the basis of their measures of DCD- risk/no risk. With a bottom-up approach, we also explored whether the modules' enrichment effectively identified clusters of participants with phenotypic differences. Table 2 reports all 33 significant gene sets organized by terms' cluster, along with the originating ontology database and associated p-value.

Cluster	Term	Ontology	<i>p</i> -value
TC1	Congenital stationary cone dysfunction	HPO	6.08×10^{-6}
TC2	Chronic otitis media	HPO	6.33×10^{-6}
TC3	Head-banging	HPO	6.27×10^{-8}
	Sleep-wake inversion	HPO	6.27×10^{-8}
	Broad face	HPO	8.55×10^{-8}
	Speech apraxia	HPO	9.05×10^{-8}
	Sound sensitivity	HPO	1.28×10^{-6}
	Circadian Clock	Reactome	1.52×10^{-6}
	Negative regulation of multicellular organism growth	GO-BP	2.88×10^{-6}
	Sleep-wake cycle disturbance	HPO	2.93×10^{-6}
	Abnormality of chromosome segregation	HPO	1.27×10^{-5}
	Corticospinal tract hypoplasia	HPO	1.44×10^{-5}
	Aplasia/Hypoplasia of the corticospinal tracts	HPO	1.44×10^{-5}
	Expressive language delay	HPO	1.95×10^{-5}
	Hyperacusis	HPO	2.13×10^{-5}
	Velopharyngeal insufficiency	HPO	2.14×10^{-5}
	Delayed eruption of primary teeth	HPO	2.26×10^{-5}
	Self-mutilation	HPO	2.65×10^{-5}
TC4	Complement-mediated synapse pruning	GO-BP	3.03×10^{-5}
	Vertebrate eye-specific patterning	GO-BP	3.03×10^{-5}
TC5	Dyssynergia	HPO	2.72×10^{-10}
	Fetal cystic hygroma	HPO	2.72×10^{-10}
	Abnormality of the dentate nucleus	HPO	1.88×10^{-9}
	Atrophy of the dentate nucleus	HPO	5.18×10^{-9}
	Regulation of PTEN gene transcription	Reactome	1.86×10^{-7}
	Nuclear matrix	GO-CC	2.48×10^{-7}
	Oromandibular dystonia	HPO	2.57×10^{-7}
	RNA polymerase II transcription factor binding	GO-MF	2.65×10^{-7}
	Nuclear periphery	GO-CC	3.96×10^{-7}

Cluster	Term	Ontology	<i>p</i> -value
TC5	Increased nuchal translucency	HPO	1.52×10^{-6}
	Hyperintensity of cerebral white matter on MRI	HPO	3.54×10^{-6}
	Cystic hygroma	HPO	1.87×10^{-5}
	Truncal ataxia	HPO	1.99×10^{-5}

Table 2. Significant gene sets (33 total) organized by terms' cluster, ontology, and *p*-value. *Note.* TC = Terms' Cluster; HPO = Human Phenotype Ontology; GO-BP/CC/MF = Gene Ontology Biological Process/Cellular Component/Molecular Function. TC1: *Sensory Visual Dysfunction* (SVD); TC2: *Peripheral Inflammation* (PI); TC3: *Neurodevelopment & Behavioral Dysregulation* (N-BD); TC4: *Cellular Pruning Mediated by Neuroinflammation* (CPMN); TC5: *Neurogenetic Transcriptional Regulation & Movement Alteration* (NTR-MA).

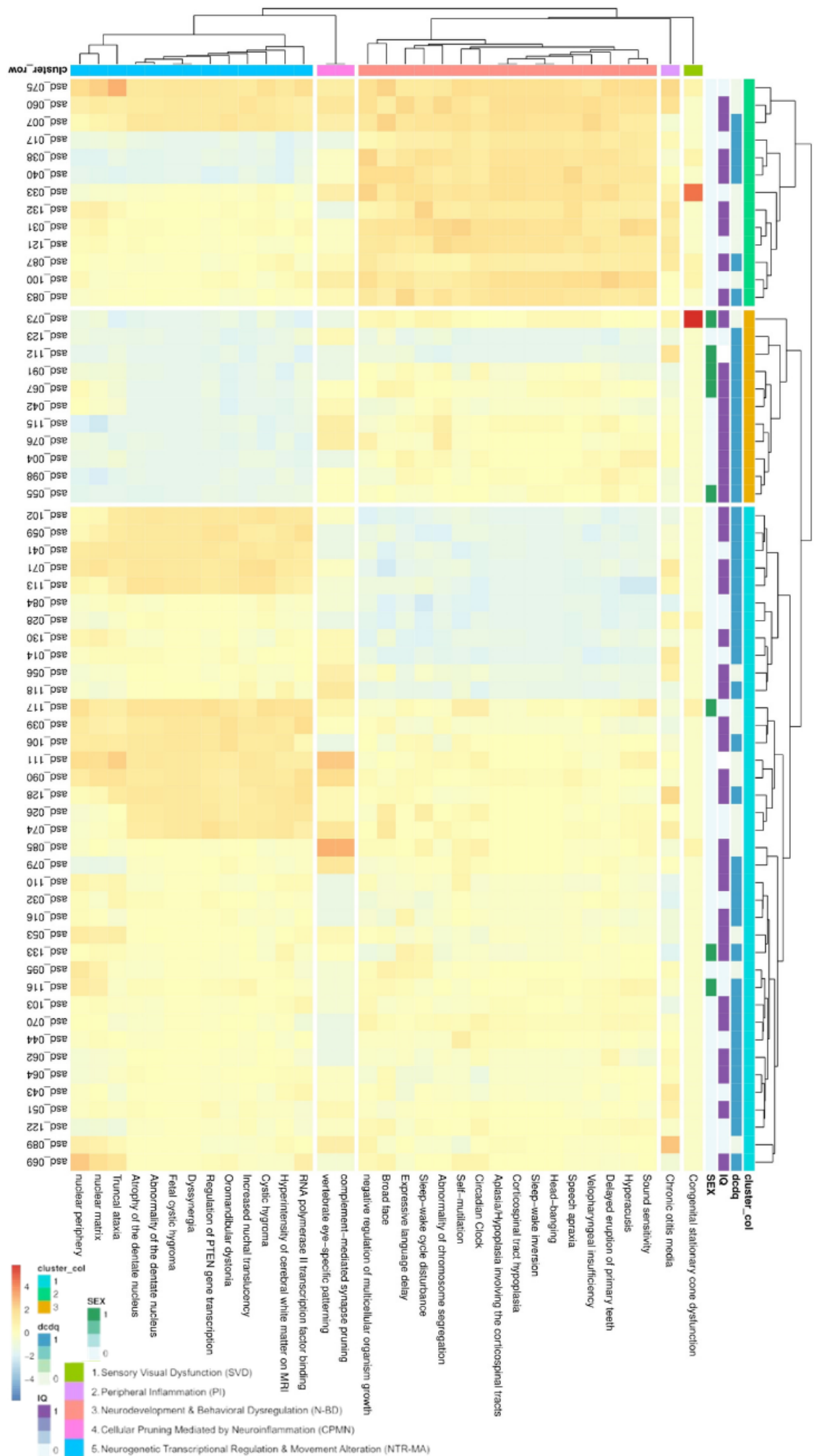


Figure 1. Heatmap showing the incidence of protein-altering variants (PAVs) for each subject (columns) in the gene sets (rows) differentially enriched in PAVs between the two subgroups of autistic children. Data-driven optimal hierarchical clustering procedure applied to both gene sets and subjects revealed five terms' clusters (TC) and three subjects' clusters (SC).

3.2 Genes in the Modules Are Expressed in the Brain in the Developmental Period

To investigate whether the genes included in the identified modules are expressed in the brain across neurodevelopmental stages, we leveraged the spatial gene expression data provided by the BrainSpan Atlas of the Developing Human Brain (Allen Institute for Brain Science, retrieved June 2026). Given the early-onset nature of autism, analyses were restricted to the first four developmental stages listed in the BrainSpan Atlas: prenatal, infancy (0–2 years), childhood (3–11 years), and adolescence (12–19 years).

It should be noted that the ABAEnrichment analysis requires a minimum number of genes per module to compute reliable enrichment statistics. TC1 (SVD) is represented by a single gene (*RPGR*) and TC4 (CPMN) by only three genes (*CIQA*, *C3*, *ITGAM*). Given this limited gene set size, the test lacks the statistical power to detect enrichment, and these two modules were therefore excluded from the brain expression analysis.

The analysis revealed that genes in three of the five identified modules — TC2 (PI), TC3 (N-BD), and TC5 (NTR-MA) — showed significant enrichment ($\text{FWER} < 0.05$) in specific brain structures across at least one developmental stage. An overview of the significantly enriched brain structures for each module and developmental stage is provided in **Table 4**. Where noted, FWER values slightly above the significance threshold ($0.05 < \text{FWER} \leq 0.06$) are also reported, as they may reflect genuine biological signals attenuated by the limited gene set or sample size of the present study.

Genes in the TC3 (N-BD) module displayed broad and persistent expression across sensorimotor, prefrontal, temporal, and subcortical regions from the prenatal period through adolescence, with particularly sustained enrichment in the primary somatosensory and motor cortices and the mediodorsal thalamus. These structures are directly implicated in the corticospinal and behavioural phenotypes characterizing this cluster. In contrast, genes in the TC5 (NTR-MA) module were predominantly enriched during the prenatal period across motor and frontal cortical regions, with a selective

re-emergence of significant enrichment in the cerebellar cortex during adolescence, consistent with the prolonged maturation of the cerebellum and the motor coordination phenotypes defining this cluster. Genes in the TC2 (PI) module showed no significant enrichment in any brain structure across developmental stages, with the sole exception of the enrichment in the inferior parietal cortex during adolescence (FWER = 0.0409).

3.3 Identification of Extended Modules Through Spatio-Temporal Co-Expression and Protein–Protein Interactions

To explore the functional interplay between the identified modules more extensively, each module was expanded by selecting genes that both physically interact at the protein level with module members and display a significant degree of spatio-temporal co-expression in the developing brain. Physical protein–protein interactions were retrieved from the BioGRID database (Stark et al., 2006), and co-expression was quantified via Pearson correlation coefficient applied to BrainSpan expression vectors restricted to the first four developmental stages. Candidate interactors were required to satisfy both criteria simultaneously: a Pearson correlation coefficient $r > 0.7$ and a Benjamini–Hochberg corrected $q < 0.05$, computed over all gene pairs within the exome panel. The resulting extended modules and their sizes relative to the original gene sets are reported in **Table 3**.

Module	Original size	Extended size
SVD	1	2
PI	69	153
NBD	112	351
CPMN	3	8
NTR_MA	186	669
Total	371	1183

Table 3. Original and extended sizes for each module and overall totals.

The five extended modules comprised ($n = 1183$) module-gene assignments. After removing genes shared across multiple modules, a final set of ($n = 891$) unique genes was obtained. Then they were pooled into a unified gene set, and a pairwise spatio-temporal co-expression matrix was computed across all gene pairs. Hierarchical clustering — average linkage applied to $1 - r$ distances — identified three genes' clusters (GC) based on their levels of co-expression (**Figure 2**). Genes in GC1 ($n = 350$) and GC2 ($n = 538$) displayed high within-cluster and low between-cluster co-expression, forming two internally coherent groups with limited mutual correlation. Genes in GC3 ($n = 3$) showed a more diffuse, intermediate co-expression profile, both internally and with respect to the other two clusters.

3.4 SFARI Genes Are Overrepresented in the Original & Extended Modules

To evaluate the relevance of the identified gene sets to autism susceptibility, we examined the proportion of high-confidence autism-related genes — defined as genes annotated with a gene score of 1 in the SFARI database (Simons Foundation, retrieved June 2026)— within both the original modules and their extended networks, using the full exome panel as the background reference for all Fisher's exact tests.

The original modules comprised 371 gene-module assignments. Because 28 genes were shared across multiple modules, they were counted only once, yielding a total of 343 unique genes, of which 86 (25.07%) were annotated in the SFARI high-confidence category, a proportion significantly higher than expected under random sampling from the exome panel ($OR = 2.593$, $p < 0.001$). When considering the extended modules ($n = 891$), 195 genes (21.88%) carried a high-confidence SFARI annotation, and the enrichment remained significant ($OR = 2.438$, $p < 0.001$).

Examining the distribution of SFARI genes across the three co-expression clusters, with Bonferroni correction applied for multiple comparisons, revealed differential enrichment patterns. GC2 showed significant overrepresentation of high-confidence SFARI genes (137 out of 538, 25.47%; $OR = 2.825$, $p_{\text{Bonf}} < 0.001$), identifying this cluster as the primary locus of autism susceptibility gene enrichment within the extended network. GC1 showed a non-significant trend in the same direction (57 out of 350, 16.29%; $OR = 1.418$, $p_{\text{Bonf}} = 0.069$), while GC3 (only three genes) did not reach significance.

	TC2 (PI)				TC3 (N-BD)				TC5 (NTR-MA)			
	Pre-natal	Infant 0-2y	Child 3-11y	Adol 12-19y	Pre-natal	Infant 0-2y	Child 3-11y	Adol 12-19y	Pre-natal	Infant 0-2y	Child 3-11y	Adol 12-19y
A1C primary auditory cortex (core)					0.0124		0.0230	0.0529				
AMY amygdaloid complex					0.0515		0.0537					
CBC cerebellar cortex						0.0324	0.0064	0.0073	0.0572			0.0474
DFC dorsolateral prefrontal cortex								0.0491				
DLTC dorsolateral temporal neocortex					0.0370		0.0557					
FCx frontal neocortex									0.0306			
HIP hippocampus							0.0333					
IPC inferior parietal cortex				0.0409	0.0209		0.0474	0.0415				
ITC inferolateral temporal cortex					0.0554		0.0302	0.0551	0.0549			
M1C primary motor cortex					0.0068		0.0372	0.0214	0.0409			
MD mediodorsal nucleus of thalamus						0.0277	0.0272	0.0167				
OFC orbital frontal cortex							0.051	0.0446				
PCx parietal neocortex					0.0047							
PFC prefrontal cortex					0.0383							

	TC2 (PI)				TC3 (N-BD)				TC5 (NTR-MA)			
	Pre-natal	Infant 0-2y	Child 3-11y	Adol 12-19y	Pre-natal	Infant 0-2y	Child 3-11y	Adol 12-19y	Pre-natal	Infant 0-2y	Child 3-11y	Adol 12-19y
S1C primary somatosensory cortex					0.0043	0.0197	0.0433	0.0340				
STC superior temporal cortex					0.0141		0.0137	0.0299	0.0464			
STR striatum							0.0403					
TCx temporal neocortex					0.0397							
V1C primary visual cortex					0.0252		0.0208	0.0386	0.0239			
VFC ventrolateral prefrontal cortex					0.0166		0.0174	0.0460				

Table 4. Sampled brain structures are listed on the left, and investigated age ranges are reported in columns.

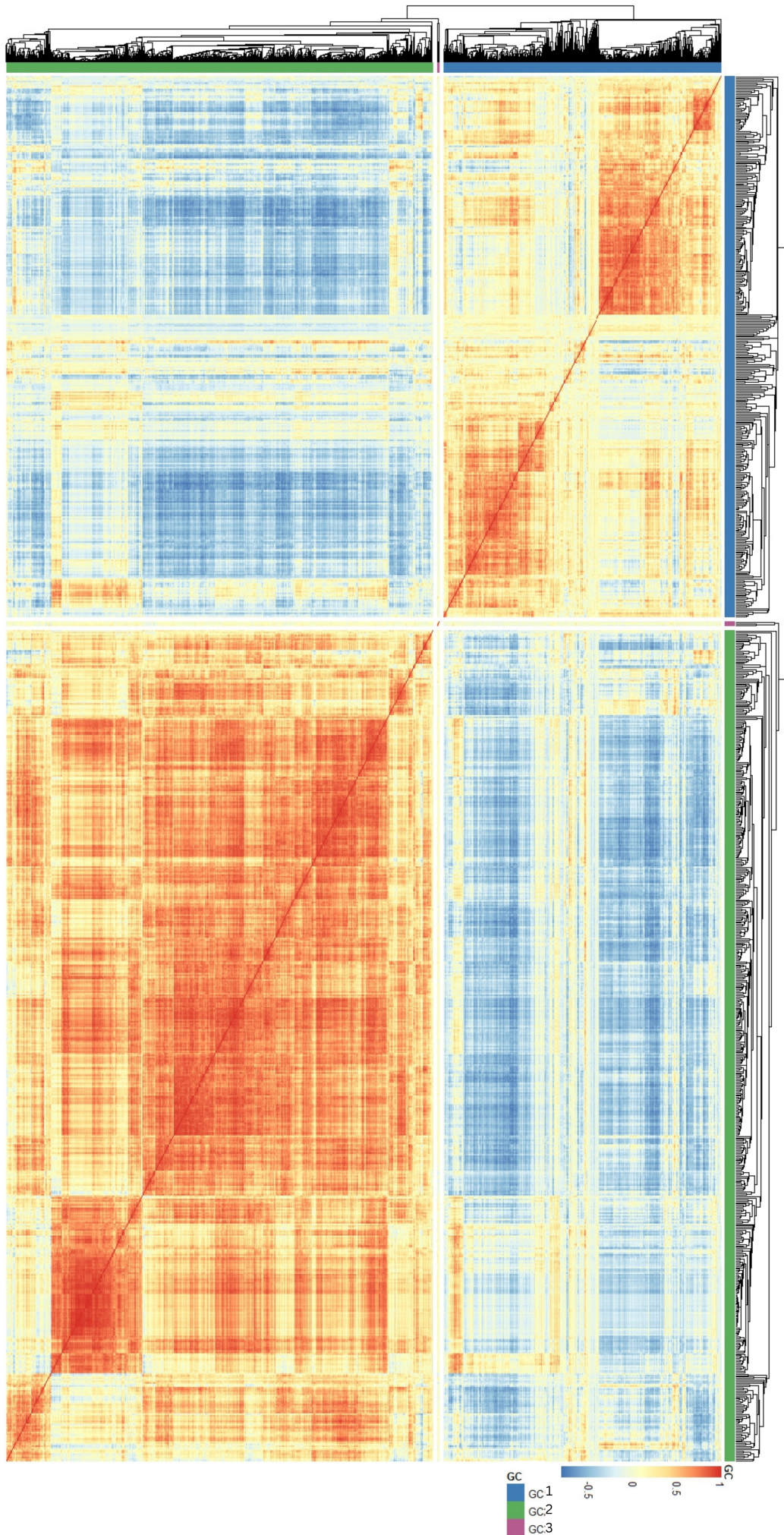


Figure 2. Heatmap showing levels of spatio-temporal co-expression in terms of correlation (color scale) among the total group of genes in the five extended modules. Data-driven hierarchical clustering revealed three genes' clusters (GCs) based on the levels of co-expression.

4 Discussion

In the present study, we identified five gene modules showing differential loads of protein-altering variants (PAVs) between two subgroups of autistic children stratified by motor risk, as assessed via the Developmental Coordination Disorder Questionnaire (DCDQ). Building directly on the multi-step, data-driven framework proposed by (Scaccabarozzi et al., 2025), we extended the original IQ-based stratification to a motor-based one, in line with the growing recognition that non-core LIMA features — language, intellectual, motor, and adaptive functioning — may represent objective, quantifiable anchors for parsing autism heterogeneity (Mandelli, Severino, et al., 2024). Importantly, motor coordination difficulties are not a peripheral or merely co-occurring feature of autism: as outlined in the Introduction, they emerge early in development, predict core social-communicative and repetitive behavior severity even after accounting for age, sex, and cognitive level (Bhat, 2021; Bhat et al., 2022), and rest on a plausible neurobiological substrate involving cerebellar and corticostriatal circuits (Canales & Graybiel, 2000; Dean et al., 2022). This view is further reinforced by recent evidence that an autism subtype defined by poor motor skills also displays measurably enhanced motor noise at the kinematic level (Mandelli, Landi, et al., 2024), supporting the notion that motor functioning is not merely a downstream correlate of autism severity but a dimension with its own identifiable behavioral and, plausibly, biological signature. The five modules identified here — Sensory Visual Dysfunction (SVD), Peripheral Inflammation (PI), Neurodevelopment & Behavioral Dysregulation (N-BD), Cellular Pruning Mediated by Neuroinflammation (CPMN), and Neurogenetic Transcriptional Regulation & Movement Alteration (NTR-MA) — span a markedly different and arguably more circuit-specific biological landscape than the four modules (ion cell communication, neurocognition, gastrointestinal function, and immune system) reported by (Scaccabarozzi et al., 2025) for IQ-based stratification. This divergence is itself an informative finding: it suggests that different non-core dimensions of the autism phenotype may be underpinned by partially

distinct, although likely interacting, genetic architectures, rather than collapsing onto a single unitary biological axis of "general severity."

4.1 TC1 — Sensory Visual Dysfunction (SVD)

Interpretation of this module should be considered highly preliminary, as TC1 was represented by a single significant gene set and, correspondingly, by a single gene; strictly speaking, it does not constitute a multi-gene biological module in the same sense as the other four clusters, and the label “module” is used here only for consistency of terminology across clusters. The corresponding gene set — congenital stationary cone dysfunction ($p = 6.08 \times 10^{-6}$) — and, correspondingly the single underlying gene, *RPGR* (Retinitis Pigmentosa GTPase Regulator). *RPGR* encodes a protein localized to the connecting cilium of photoreceptors and is predominantly associated with X-linked retinitis pigmentosa and cone-rod dystrophies. At first glance, the emergence of a retinal ciliopathy gene within a motor-based stratification may appear biologically distant from the phenotype under investigation. However, primary cilia are not confined to the photoreceptor; they play a well-established role in neuronal migration, axon guidance, and broader aspects of brain development (Louvi & Grove, 2011), and ciliary dysfunction has been repeatedly implicated across neurodevelopmental conditions. Taken together with the well-documented overlap between sensory processing atypicalities — visual processing among them — and motor coordination difficulties in autism (Bhat, 2020), the presence of this cluster within a motor risk stratification framework may reflect a biologically plausible link between sensory and motor domains; however, the present dataset includes no direct measures of visual function, and the connection between *RPGR*-related ciliary biology and the motor phenotype studied here is consistent with, but not directly demonstrated by, the present data. The evidence for this cluster is necessarily preliminary given its minimal size. Coherently, TC1 could not be carried forward to the brain expression analysis: a single-gene module simply does not provide the ABAEnrichment procedure with sufficient statistical power to detect reliable enrichment, and *RPGR* itself is predominantly expressed in retinal photoreceptors rather than in the broader brain structures sampled by the BrainSpan Atlas. This is consistent with the biological nature of the gene rather than a contradiction of the finding: a peripheral sensory gene need not and, mechanistically, would not be expected to show enrichment in bulk cortical or subcortical RNA-seq data.

4.2 TC2 — Peripheral Inflammation (PI)

TC2 was anchored by a single significant gene set — chronic otitis media ($p = 6.33 \times 10^{-6}$) — yet encompassed 69 unique genes in its original form, expanding to 153 genes once extended through co-expression and protein–protein interaction criteria. Despite being represented by a single phenotypic term, the gene composition of this module is markedly immunological, including key regulators of adaptive and innate immunity, B-cell signalling, and inflammatory cascades. The label *Peripheral Inflammation* should be regarded as a heuristic summary of the module’s predominant gene composition rather than a direct demonstration of inflammatory dysregulation in the present sample: the cluster is statistically anchored by a single HPO term (chronic otitis media), and an equally defensible label might have emphasized recurrent infection susceptibility rather than inflammation per se. The selection of a peripheral, infection-related phenotype as the label for an immune module is consistent with a substantial and converging body of evidence implicating immune dysregulation in autism, both as a candidate etiological mechanism — for example through maternal immune activation — and as a peripheral endophenotype of the condition itself (Meltzer & Van de Water, 2017; Ormstad et al., 2018). The brain expression analysis offers a particularly interesting — and, in our view, informative — pattern for this module. With the sole exception of a borderline enrichment in the inferior parietal cortex during adolescence (FWER = 0.0409), genes in TC2 showed no significant enrichment in any brain structure across the four developmental stages considered. This absence of cortical or subcortical enrichment mirrors almost exactly the pattern reported for the immune system module (Scaccabarozzi et al., 2025), and the convergence between two independent stratification criteria — IQ in the original study, motor risk here — strengthens rather than weakens the interpretation that immune-related genetic susceptibility in autism may act predominantly through systemic, rather than brain-localized, pathways. As noted for the analogous immune module (Scaccabarozzi et al., 2025), the BrainSpan Atlas reflects expression profiles measured in typically developing individuals via bulk tissue RNA-seq; it is plausible that the presence of PAVs in our autistic participants shifts the typical expression pattern of these genes, producing the immune transcriptome alterations reported elsewhere in autistic post-mortem brain tissue (Garbett et al., 2008; Voineagu et al., 2011), or, alternatively, that the relevant mechanism is mediated through peripheral-to-central signalling routes such as the gut–brain axis, where

the immune system acts as a crucial intermediary (Azhari et al., 2019; Powell et al., 2017). Although speculative, and not directly tested with the present data, these findings may be compatible with broader immune-brain interaction models proposed in ASD. Within the present motor-based framework, peripheral inflammation may contribute to coordination difficulties indirectly, for instance by affecting cerebellar and basal ganglia maturation during critical windows of synaptic refinement, a hypothesis that future longitudinal or neuroimaging-informed work could more directly test.

4.3 TC3 — Neurodevelopment & Behavioral Dysregulation (N-BD)

TC3 was the broadest cluster in terms of the number of significant gene sets, comprising terms spanning core and associated autistic features — head-banging ($p = 6.27 \times 10^{-8}$), self-mutilation ($p = 2.65 \times 10^{-5}$), sleep-wake inversion and disturbance ($p = 6.27 \times 10^{-8}$ and $p = 2.93 \times 10^{-6}$, respectively) — alongside motor and communication-related terms, including speech apraxia ($p = 9.05 \times 10^{-8}$), corticospinal tract hypoplasia and aplasia/hypoplasia of the corticospinal tracts (both $p = 1.44 \times 10^{-5}$), and expressive language delay ($p = 1.95 \times 10^{-5}$). Sensory features were also represented, with sound sensitivity ($p = 1.28 \times 10^{-6}$) and hyperacusis ($p = 2.13 \times 10^{-5}$), together with the Reactome term *Circadian Clock* ($p = 1.52 \times 10^{-6}$). The co-occurrence of sleep disturbance, self-injurious behavior, speech apraxia, and corticospinal abnormalities within a single cluster is not incidental: many of these terms converge on *RAI1* (Retinoic Acid Induced 1), a dosage-sensitive transcription factor whose haploinsufficiency causes Smith-Magenis Syndrome, a neurodevelopmental condition characterized by intellectual disability, profound sleep disturbance, self-injurious behavior, and autistic features. The recurrence of *RAI1* across such phenotypically diverse gene sets — from circadian and behavioral terms to corticospinal and speech-related ones — suggests that this gene may function as a hub linking multiple facets of the cluster’s biological signature, rather than each term reflecting an independent genetic mechanism. Other genes contributing to TC3 reinforce this interpretation: *GRIN2A*, involved in glutamatergic neurotransmission and implicated in speech and language disorders; *MECP2*, the Rett syndrome gene; *UBE3A*, associated with Angelman syndrome; and *ARID1B*, a chromatin remodelling gene consistently and robustly implicated in autism. Coherently with this composition, the brain expression analysis revealed that TC3 genes were broadly and persistently

enriched from the prenatal period through adolescence across sensorimotor, prefrontal, temporal, and subcortical regions, with particularly sustained enrichment in the primary somatosensory and motor cortices and the mediodorsal nucleus of the thalamus. These structures map onto the corticospinal and behavioral phenotypes that define the cluster, and the persistence of the signal across nearly the entire developmental window investigated is compatible with the broad, multi-domain nature of the syndromes (Smith-Magenis among them) associated with its core genes; this expression enrichment, however, reflects statistical co-occurrence with relevant brain structures and HPO phenotype terms, not a demonstrated causal pathway from gene to symptom. Within a motor risk framework, TC3 therefore appears compatible with capturing a transdiagnostic neurodevelopmental liability — spanning sleep, self-regulation, speech, and corticospinal integrity — of which motor coordination difficulties may represent one clinically salient but not isolated expression.

4.4 TC4 — Cellular Pruning Mediated by Neuroinflammation (CPMN)

The label *Cellular Pruning Mediated by Neuroinflammation* is retained here for consistency with the nomenclature established in the Results chapter; however, the name itself already embeds a mechanistic claim — pruning *mediated by* neuroinflammation — that the underlying three-gene module cannot, on its own, support. A more neutral description, such as “complement-associated synaptic processes,” may better reflect the actual evidentiary basis of this cluster, and readers should interpret the label accordingly, as a descriptive heuristic rather than as a validated mechanism. Given the extremely limited module size — only three genes — biological interpretations of TC4 should be considered exploratory, and the following discussion is intended to outline biological plausibility rather than to support a confirmed mechanism. TC4 comprised only two significant gene sets — complement-mediated synapse pruning and vertebrate eye-specific patterning (both $p = 3.03 \times 10^{-5}$) — underpinned by three genes: *CIQA*, *C3*, and *ITGAM*. Despite its minimal size, the cluster carries a coherent and well-supported biological narrative. *CIQA* and *C3* are core components of the classical complement cascade, extensively implicated in activity-dependent synaptic pruning during neurodevelopment, with C1q tagging weak or supernumerary synapses for elimination by microglia (Stephan

et al., 2012). *ITGAM* (encoding CD11b) is a microglial surface receptor central to phagocytosis and neuroinflammatory signalling. Aberrant complement-mediated pruning has been proposed as a mechanism contributing to atypical synaptic connectivity in autism (Filipello et al., 2018), and a process so intimately tied to the elimination and refinement of synaptic connections is mechanistically plausible as a contributor to motor circuit maturation, given how extensively cerebellar and corticostriatal circuits depend on activity-dependent pruning during early development. As with TC1, the extremely limited gene set size (only three genes) precluded reliable brain expression testing via ABAEnrichment: the module was excluded from this analysis for lack of statistical power. This limitation is compounded by a biological one, again consistent with — rather than contradicting — the finding itself. *CIQA*, *C3*, and *ITGAM* are expressed in a highly cell-type-specific manner, predominantly in microglia, whose transcriptional signature is diluted and not adequately captured by bulk tissue RNA-seq data such as that underlying the BrainSpan Atlas (Scaccabarozzi et al., 2025). A complement- and microglia-driven pruning module would therefore be expected to evade detection in bulk-tissue expression atlases regardless of its true biological relevance, underscoring the need for single-cell or microglia-specific expression resources in future replications of this analysis.

4.5 TC5 — Neurogenetic Transcriptional Regulation & Movement Alteration (NTR-MA)

TC5 was the largest and most statistically robust cluster identified in the present analysis, anchored by terms reflecting structural and functional abnormalities of the cerebellum and related motor pathways: dyssynergia ($p = 2.72 \times 10^{-10}$), abnormality and atrophy of the dentate nucleus ($p = 1.88 \times 10^{-9}$ and $p = 5.18 \times 10^{-9}$, respectively), truncal ataxia ($p = 1.99 \times 10^{-5}$), and oromandibular dystonia ($p = 2.57 \times 10^{-7}$). Alongside these unambiguously motor phenotypes, the cluster also comprised terms related to transcriptional regulation — RNA polymerase II transcription factor binding ($p = 2.65 \times 10^{-7}$) and regulation of *PTEN* gene transcription ($p = 1.86 \times 10^{-7}$) — and cellular compartment terms such as nuclear matrix ($p = 2.48 \times 10^{-7}$) and nuclear periphery ($p = 3.96 \times 10^{-7}$), together with the neuroimaging finding of hyperintensity of cerebral white matter on MRI ($p = 3.54 \times 10^{-6}$). This is, of all five clusters identified in the present study, the one that appears particularly relevant to the motor-based stratification

employed in the present study, given its direct and unambiguous link to the motor stratification criterion itself; its emergence as the largest and most significant module supports the relevance of the DCD-risk stratification within this exploratory sample, but does not by itself demonstrate that this module explains motor impairment, since no direct relationship between gene-level variation and individual motor performance was tested here. This should not, however, be read as independent validation of motor risk as a stratifying variable more generally, which would require replication in an independent cohort. Key genes underlying TC5 include *ATNI*, whose variants cause dentatorubral-pallidoluysian atrophy, a condition marked by cerebellar ataxia and involuntary movements; *ATPIA3*, encoding a sodium-potassium ATPase subunit already linked both to autism and to movement disorders such as alternating hemiplegia of childhood (Schmunk & Gargus, 2013); *CTNNB1*, a core component of the Wnt/ β -catenin pathway governing synaptic development; and *GSK3B*, a kinase regulating neuronal differentiation that has been repeatedly implicated in autism-related signalling pathways. *PTEN* regulation also emerged among the cluster's significant terms; however, this and the other genes above should be read as illustrative components of a broader, network-level signature rather than as evidence that any single gene drives the cluster's association with motor risk — the relevant unit of interpretation throughout this section remains the module, not any individual gene. The brain expression results for TC5 are compatible with, and arguably add plausibility to, this genetic profile: genes were predominantly enriched during the prenatal period across motor and frontal cortical regions, with a notable and selective re-emergence of enrichment in the cerebellar cortex during adolescence. Cerebellar maturation is known to be markedly protracted relative to most other brain structures, extending well into the second decade of life; the re-appearance of a cerebellar signal specifically during adolescence is therefore consistent with the prolonged developmental trajectory of this structure and with the motor coordination phenotypes that anchor the cluster. Taken together, TC5 is consistent with the possibility of a biologically coherent module in which cerebellar and corticospinal motor circuitry, transcriptional regulation, and chromatin organization converge, echoing the established role of cerebellar transcription factors in motor learning and broader neurodevelopmental trajectories (Scaccabarozzi et al., 2025); given the exploratory and cross-sectional nature of the present analysis, this convergence should be regarded as hypothesis-generating

rather than as established mechanism.

4.6 Integration Across Modules: Spatio-Temporal co-Expression and Protein–Protein Interactions

Beyond characterizing each module in isolation, we extended our investigation by considering, for every gene within the five modules, its direct physical interactors that also displayed significant spatio-temporal co-expression in the developing brain. This extension increased the total gene count from 371 genes in the original modules to 1183 genes in the extended modules (Table 3), with the relative expansion being most marked for NTR-MA (from 186 to 669 genes) and PI (from 69 to 153 genes), and comparatively limited — as expected given their minimal starting size — for SVD (from 1 to 2 genes) and CPMN (from 3 to 8 genes). When all genes across the five extended modules were pooled and clustered according to their pairwise co-expression profiles, three genes' clusters (GC) emerged. GC1 ($n = 350$) and GC2 ($n = 538$) each displayed high within-cluster and comparatively low between-cluster co-expression, indicating that they represent two internally coherent — and at least partially distinct — biological programmes operating across brain development, whereas GC3 ($n = 3$) showed a diffuse, intermediate co-expression profile that precludes any strong functional interpretation given its negligible size. This three-cluster structure is broadly reminiscent of the pattern reported by (Scaccabarozzi et al., 2025) in their own extended network analysis. The observed network architecture should not be interpreted as evidence of direct biological interaction between the genes involved, but rather as a statistical indication of potential functional convergence: spatio-temporal co-expression and physical protein–protein interaction are correlational and database-derived measures, and neither establishes that the genes in question act together within the same causal pathway in our sample. With this caveat in mind, the pattern reinforces the more general hypothesis — echoed across both studies — that the genetic architecture of autism heterogeneity is unlikely to consist of independent, modular cascades acting in isolation. Rather, genes belonging to ostensibly distinct functional categories (immune, synaptic, transcriptional, motor) appear to be densely and specifically interconnected at both the level of spatio-temporal brain co-expression and of physical protein–protein interaction, which suggests the possibility that the phenotypic variability captured by motor risk stratification emerges from the joint,

and possibly compensatory or synergistic, action of multiple interacting pathways rather than from a single dominant mechanism, although this interpretation remains a statistical inference rather than a demonstrated biological fact.

4.7 SFARI Genes are Overrepresented in the Original and Extended Modules

This enrichment does not constitute independent validation of the gene-set variant enrichment analysis itself, since the modules were derived from an unbiased, autism-agnostic exome panel rather than from a list of known autism genes; nonetheless, it should be read as convergent evidence supporting the biological relevance of the identified modules, not as proof that the modules capture autism-specific biology beyond what SFARI already describes. A central aim of the present work, following (Scaccabarozzi et al., 2025), was to determine whether the gene sets identified through an entirely data-driven, motor-based stratification would independently recover genes already known to be associated with autism susceptibility. The results are consistent with this convergence. Among the 343 genes comprising the original modules, 86 (25.07%) carried a high-confidence SFARI annotation, a proportion markedly — and significantly — higher than would be expected by chance against the full exome panel background ($OR = 2.593$, $p < 0.001$). This is a numerically stronger enrichment than that reported for the original IQ-based modules by (Scaccabarozzi et al., 2025), where high-confidence SFARI genes in the (un-extended) modules did not reach significance ($OR = 1.534$, $p = 0.2063$), although the two estimates come from different samples and stratification criteria and should not be compared as if drawn from a single, directly contrastable test. The enrichment remained robust and significant when extending the gene set through co-expression and protein–protein interaction criteria (195 of 891 genes, 21.88%; $OR = 2.438$, $p < 0.001$), in this respect mirroring the finding from the original study, where SFARI enrichment became significant specifically at the extended-network level ($OR = 2.296$, $p = 0.0001$). Taken together, these results are consistent with the possibility that motor risk taps into genetic variation that is already, and independently, recognized as relevant to autism susceptibility — in line with the broader view that motor coordination difficulties are not a peripheral epiphenomenon of the condition but rather a phenotypic expression that may be related to autism-relevant genetic variation (Bhat,

2021; Bhat et al., 2022); given the exploratory, single-cohort design of the present study, this association should be treated as preliminary and hypothesis-generating rather than as evidence that motor risk indexes genetic liability more directly than IQ.

Looking specifically at the distribution of SFARI genes across the three coexpression derived genes' clusters, GC2 emerged as the primary locus of autism susceptibility gene enrichment within the extended network (137 of 538 genes, 25.47%; OR = 2.825, Bonferroni-corrected $p < 0.001$), whereas GC1 showed only a non-significant trend in the same direction (57 of 350, 16.29%; OR = 1.418, Bonferroni-corrected $p = 0.069$) and GC3 — comprising only three genes — could not be reliably interpreted given its negligible size (1 of 3, 33.33%; OR = 3.549, Bonferroni-corrected $p = 0.981$). This asymmetry between GC1 and GC2 indicates that the convergence between our motor-derived gene modules and established autism-risk genes is not uniformly distributed across the extended network but is instead concentrated within a specific, densely co-expressed sub-network, echoing the analogous — though not numerically identical — pattern reported by (Scaccabarozzi et al., 2025), in which SFARI enrichment was similarly concentrated within one of three co-expression clusters and, within it, specifically over-represented in the extended neurocognition and gastrointestinal networks. While the present study's modules differ in biological theme from those of the original IQ-based stratification, the recurring observation — across two independent stratification criteria and two independent samples — that autism susceptibility genes cluster preferentially within specific, highly co-expressed sub-networks of a broader extended gene set, rather than being evenly distributed throughout it, is consistent with the possibility that distinct phenotypic dimensions of autism (IQ, motor functioning) may each relate to partially separable, yet richly interconnected, regions of the same underlying genetic and developmental landscape; given the limited sample size and the absence of a replication cohort, this remains a tentative interpretation that would need to be tested directly in future work.

4.8 Limitations

Several limitations of the present study warrant consideration. First and foremost, the sample size ($N = 62$, with only 19 participants in the NO DCD-risk subgroup) is modest, and the resulting findings must therefore be regarded as preliminary, in line with the

analogous caveat raised by (Scaccabarozzi et al., 2025) for their own sample of 71 participants. The unequal distribution of participants across the two motor subgroups, combined with the markedly skewed male-to-female ratio of the overall sample (54 males, 8 females) consistent with established epidemiological patterns in autism, further limits the generalizability of the present findings and precludes any meaningful exploration of sex-specific genetic effects. Second, as in the original study, the genetic analysis was restricted to a targeted exome panel of approximately 5,200 disease-associated genes rather than the whole exome or genome; while this restriction did not prevent the recovery of biologically coherent and SFARI-enriched modules, a broader genomic survey could plausibly identify additional, currently undetected, motor-relevant gene sets. Third, two of the five modules identified here — SVD and CPMN — were represented by only one and three genes, respectively. While both proved biologically interpretable, their minimal size precluded reliable brain expression testing and necessarily limits the strength of any conclusion that can be drawn about their role in motor risk stratification; these should be considered hypothesis-generating rather than confirmatory findings. Fourth, brain expression profiles were derived from the BrainSpan Atlas of the Developing Human Brain, which reflects bulk tissue RNA-seq in typically developing individuals. As discussed above for both TC2 and TC4, this approach is intrinsically limited in its capacity to detect genuine biological signals arising from highly cell-type-specific expression (e.g., microglia) or from genuinely altered, autism-specific expression patterns that diverge from the typical developmental trajectory captured by the atlas. Fifth, motor risk stratification in the present study relied exclusively on a parent-report measure, the DCDQ, despite the availability of additional, objective motor measures within the broader project from which this sample was drawn (e.g., the Movement Assessment Battery for Children-2 and kinematic indices). Although the DCDQ is well-validated and widely used (B. N. Wilson et al., 2009), parent-report instruments are inherently more subject to reporting bias than clinician-administered or kinematic measures, such as the reach-to-drop task also collected within the broader "Progetto Mosaico" cohort (Crippa et al., 2015), which future work should incorporate to corroborate and refine the present genotype–motor phenotype associations. Sixth, by considering only the cumulative load of PAVs within each gene set, the present approach — as in (Scaccabarozzi et al., 2025) — does not differentiate between variants of differing functional severity or pathogenicity within the same gene

set, a simplification that may obscure finer-grained genotype–phenotype relationships. Finally, and perhaps most importantly, the present findings have not been tested in an independent cohort. The small sample size alone does not fully capture this limitation: even a larger single-cohort study would remain vulnerable to sample-specific artifacts that only an independent replication can rule out. Replication in independent cohorts will therefore be necessary to determine the stability and generalizability of the identified modules before they can be considered anything more than candidate findings.

5 Conclusion

The present study extended the multi-step, data-driven framework developed by (Scaccabarozzi et al., 2025) from an IQ-based to a motor-based stratification of autistic children, and identified five candidate gene modules — SVD, PI, N-BD, CPMN, and NTR-MA — showing differential loads of protein-altering variants between children at high versus low risk for developmental coordination difficulties; these should be understood as candidate modules emerging from an exploratory, unbiased pipeline, not as an established genetic architecture of motor subtypes in autism. These modules spanned biological processes ranging from ciliary and sensory function to peripheral immune regulation, broad neurodevelopmental and behavioral dysregulation, complement-mediated synaptic pruning, and — most prominently — cerebellar and corticospinal transcriptional and motor regulation. Three of the five modules showed distinct, biologically plausible patterns of spatio-temporal brain expression, and, once extended through physical protein–protein interactions and co-expression in the developing brain, all modules were collectively and significantly enriched for high-confidence autism susceptibility genes from the SFARI database, with this enrichment concentrated within a specific, densely co-expressed sub-network of genes.

Taken together, these findings are consistent with, and extend, the hypothesis already advanced by (Scaccabarozzi et al., 2025): that autism phenotypic heterogeneity may arise from multiple, interacting genetic pathways rather than from isolated, independent cascades. The present results suggest that this pattern may hold not only when autism subgroups are defined by cognitive functioning, but also when they are defined by a non-core, historically under-investigated dimension such as motor coordination, although this parallel rests on two small, non-replicated samples analyzed with the same exploratory

pipeline and should be interpreted cautiously. This is broadly compatible with the theoretical shift — articulated in the LIMA framework and the AUTISMS-3D model (Mandelli, Severino, et al., 2024) — toward conceptualizing autism heterogeneity through non-core features that may serve as potential objective anchors, quantifiably assessed, rather than through core symptom severity alone. Motor coordination difficulties, in this sense, emerge from the present analysis not necessarily as a secondary or incidental correlate of autism, but as a phenotypic dimension consistent with the possibility that it relates to a biologically plausible and developmentally meaningful genetic architecture; confirming this would require replication in independent, adequately powered samples, and the present findings should be regarded as hypothesis-generating rather than confirmatory.

Future research should aim to replicate these findings in larger and independent cohorts, ideally leveraging publicly available sequencing resources (e.g., SPARK) and whole-exome or whole-genome data, and should incorporate objective, clinician administered or kinematic motor measures alongside parent-report instruments. Combining motor-based genetic stratification with other LIMA dimensions — language, intellectual, and adaptive functioning — within the same analytic framework could further clarify whether distinct non-core features map onto distinct, partially overlapping, or entirely shared genetic architectures, ultimately contributing to a more biologically grounded and clinically actionable understanding of autism subtyping.

Taken together, these findings provide preliminary evidence that motor-based stratification may help reduce phenotypic heterogeneity in autism and generate biologically meaningful hypotheses for future research, although replication in larger independent samples remains essential.

Appendix

Language Choices

In the context of autism and other sensitive topics, language functions as a powerful tool that can influence both thought and perception. The use of respectful and accurate terminology is essential—not only to promote effective communication among professionals, but also to ensure that individuals are represented thoughtfully, avoiding the reinforcement of stigmatizing or harmful stereotypes.⁸ The ongoing discussion concerning the most appropriate terminology primarily contrasts two linguistic approaches: Identity-First Language (IFL; e.g., *autistic individuals*) and Person-First Language (PFL; e.g., *individuals with autism*). Empirical research indicates that preferences differ across stakeholder groups: autistic adults generally favor IFL, whereas professionals are more likely to use person-first constructions (Taboas et al., 2022). Although there is no universal consensus, a clear trend has emerged toward adopting more inclusive and person-respecting language, reflecting a broader awareness of how members of the autism community wish to be described. Acknowledging the validity of both perspectives, this paper adopts Identity-First Language.

Ethical Declaration

The present research was approved by the Ethics Committee of our institute “Comitato Etico IRCCS E. Medea—Sezione Scientifica Associazione La Nostra Famiglia” and was therefore performed in accordance with the ethical standards set forth in the 1964 Declaration of Helsinki and later amendments.

⁸For a list of terms/expression that are likely to cause offense, see: <https://www.autismeurope.org/about-autism/acceptable-language>

References

- Abbott, A. E., Linke, A. C., Nair, A., Jahedi, A., Alba, L. A., Keown, C. L., Fishman, I., & Müller, R. A. (2018). Repetitive behaviors in autism are linked to imbalance of corticostriatal connectivity: A functional connectivity mri study. *Social Cognitive and Affective Neuroscience*, 13(1), 32–42.
- ADDM & CDC. (2007). Prevalence of autism spectrum disorders: Autism and developmental disabilities monitoring network, six sites, united states, 2000. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, 56(1), 1–11.
- Allen Institute for Brain Science. (n.d.). Brainspan atlas of the developing human brain [Retrieved June, 2026]. <https://www.brainspan.org/>
- American Psychiatric Association. (2022). Neurodevelopmental disorders. In *Diagnostic and statistical manual of mental disorders* (5th, text rev.).
- Anderson-Chavarria, M. (2021). The autism predicament: Models of autism and their impact on autistic identity. *Disability & Society*, 37(8), 1321–1341.
- Antze, P. (2010). On the pragmatics of empathy in the neurodiversity movement. In *Ordinary ethics: Anthropology, language, and action* (pp. 310–327).
- Araki, H., Knapp, C., Tsai, P., & Print, C. (2012). Genesetdb: A comprehensive meta-database, statistical and visualisation framework for gene set analysis. *FEBS Open Bio*, 2, 76–82.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29.
- Augustine, F., Nebel, M. B., Mostofsky, S. H., Mahone, E. M., & Singer, H. S. (2021). Aberrant prefrontal cortical-striatal functional connectivity in children with primary complex motor stereotypies. *Cortex*, 142, 272–282.
- Azhari, A., Azizan, F., & Esposito, G. (2019). A systematic review of gut-immune-brain mechanisms in autism spectrum disorder. *Developmental Psychobiology*, 61(5), 752–771.
- Bai, D., Yip, B. H. K., Windham, G. C., et al. (2019). Association of genetic and environmental factors with autism in a 5-country cohort. *JAMA Psychiatry*, 76(10), 1035–1043.
- Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorodny, W., Robinson Rosenberg, C., White, T., Durkin, M. S., Imm, P., Nikolaou, L., Yeargin-Allsopp, M., Lee, L. C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., . . . Dowling, N. F. (2018). Prevalence of autism spectrum disorder among children

- aged 8 years – autism and developmental disabilities monitoring network, 11 sites, united states, 2014. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, 67(6), 1–23.
- Barone, L. (Ed.). (2019). *Manuale di psicologia dello sviluppo*. Carocci Editore.
- Belmonte, M. K. (2022). Motor symptoms in the asd diagnostic criteria: A conservative perspective. *Autism Research*, 15(9), 1582–1584.
- Ben-Shaul, Y., Bergman, H., & Soreq, H. (2005). Identifying subtle interrelated changes in functional gene categories using continuous measures of gene expression. *Bioinformatics*, 21(7), 1129–1137.
- Bertelsen, N., Landi, I., Bethlehem, R., Seidlitz, J., Busuoli, E., Mandelli, V., & et al. (2021). Imbalanced social-communicative and restricted repetitive behavior subtypes of autism spectrum disorder exhibit different neural circuitry. *Communications Biology*, 4(1), 574.
- Bettelheim, B. (1999). *La fortezza vuota: L'autismo infantile e la nascita del sé* [Traduzione di A. Pandolfi]. Garzanti.
- Bhat, A., Galloway, J., & Landa, R. (2012). Relation between early motor delay and later communication delay in infants at risk for autism. *Infant Behavior & Development*, 35(4), 838–846.
- Bhat, A. (2020). Is motor impairment in autism spectrum disorder distinct from developmental coordination disorder? a report from the spark study. *Physical Therapy*, 100(4), 633–644.
- Bhat, A. (2021). Motor impairment increases in children with autism spectrum disorder as a function of social communication, cognitive and functional impairment, repetitive behavior severity, and comorbid diagnoses: A spark study report. *Autism Research*, 14, 202–219.
- Bhat, A., Boulton, A. J., & Tulskey, D. S. (2022). A further study of relations between motor impairment and social communication, cognitive, language, functional impairments, and repetitive behavior severity in children with asd using the spark study dataset. *Autism Research*, 15(6), 1156–1178.
- Biancotto, M., Guicciardi, M., Pelamatti, G. M., Santamaria, T., & Zoia, S. (2017). *Movement assessment battery for children-second edition. standardizzazione italiana*. Giunti O.S. Psychometrics.
- Breach, M. R., & Lenz, K. M. (2023). Sex differences in neurodevelopmental disorders: A key role for the immune system. *Current Topics in Behavioral Neurosciences*, 62, 165–206.

- Breslin, T., Edén, P., & Krogh, M. (2004). Comparing functional annotation analyses with catmap. *BMC Bioinformatics*, 5(1), 193.
- Canales, J. J., & Graybiel, A. M. (2000). A measure of striatal function predicts motor stereotypy. *Nature Neuroscience*, 3(4), 377–383.
- Cantor, R. M., Navarro, L., Won, H., Walker, R. L., Lowe, J. K., & Geschwind, D. H. (2018). Asd restricted and repetitive behaviors associated at 17q21.33: Genes prioritized by expression in fetal brains. *Molecular Psychiatry*, 23(4), 993–1000.
- Casten, L. G., Thomas, T. R., Doobay, A. F., Foley-Nicpon, M., Kramer, S., Nickl-Jockschat, T., & Michaelson, J. J. (2023). The combination of autism and exceptional cognitive ability is associated with suicidal ideation. *Neurobiology of Learning and Memory*, 197, 107698.
- Castro, K., Riesgo, R., & Gadia, C. (2026). Autism spectrum disorder: Overdiagnosis or a new pandemic? *Jornal de Pediatria*, 102(Supplement 1), 101439.
- Chawarska, K., Fernandez, T. V., Milgramm, A., Volkmar, F., Torres-Viso, M., Powell, K., Verneti, A., Macari, S., & Morgan, C. (2025). Motor stereotypies in toddlers with and without autism: A transdiagnostic dimension. *Journal of Child Psychology and Psychiatry*.
- Choi, L., & An, J. (2021). Genetic architecture of autism spectrum disorder: Lessons from large-scale genomic studies. *Neuroscience & Biobehavioral Reviews*, 128, 244–257.
- Cirnigliaro, M., Chang, T., Arteaga, S., Pérez-Cano, L., Ruzzo, E., Gordon, A., & et al. (2023). The contributions of rare inherited and polygenic risk to asd in multiplex families. *Proceedings of the National Academy of Sciences*, 120(31), e2215632120.
- Clark, M., & Adams, D. (2022). Resilience in autism and intellectual disability: A systematic review. *Review Journal of Autism and Developmental Disorders*, 9, 39–53.
- Corbett, B. A., et al. (2021). Camouflaging in autism: Examining sex-based and compensatory models in social cognition and communication. *Autism Research*, 14(1), 127–142.
- Crippa, A., Craig, F., Busti Ceccarelli, S., Mauri, M., Grazioli, S., Scionti, N., Cremascoli, A., Ferrante, C., Visioli, C., Marzocchi, G. M., Molteni, M., & Nobile, M. (2021). A multimethod approach to assessing motor skills in boys and girls with autism spectrum disorder. *Autism*, 25(5), 1481–1491.
- Crippa, A., Salvatore, C., Perego, P., et al. (2015). Use of machine learning to identify children with autism and their motor abnormalities. *Journal of Autism and Developmental Disorders*, 45, 2146–2156.

- Culhane, A. C., Schröder, M. S., Sultana, R., Picard, S. C., Martinelli, E. N., Kelly, C., Haibe-Kains, B., Kapushesky, M., St Pierre, A.-A., Flahive, W., Picard, K. C., Gusenleitner, D., Papenhausen, G., O'Connor, N., Correll, M., & Quackenbush, J. (2012). Genesigdb: A manually curated database and resource for analysis of gene expression signatures. *Nucleic Acids Research*, 40(D1), D1060–D1066.
- Davignon, M. N., Qian, Y., Massolo, M., & Croen, L. A. (2018). Psychiatric and medical conditions in transition-aged individuals with asd. *Pediatrics*, 141(suppl 4), S335–S345.
- De Rubeis, S., He, X., Goldberg, A., Poultney, C., Samocha, K., Ercument Cicek, A., & et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515(7526), 209–215.
- Dean, S. L., Tochen, L., Augustine, F., Ali, S. F., Crocetti, D., Rajendran, S., Blue, M. E., Mahone, E. M., Mostofsky, S. H., & Singer, H. S. (2022). The role of the cerebellum in repetitive behavior across species: Childhood stereotypies and deer mice. *Cerebellum (London, England)*, 21(3), 440–451.
- Di Giovanni, D., Enea, R., Di Micco, V., Benvenuto, A., Curatolo, P., & Emberti Gialloreti, L. (2023). Using machine learning to explore shared genetic pathways and possible endophenotypes in autism spectrum disorder. *Genes*, 14(2), 313.
- Duda, M., et al. (2018). Brain-specific functional relationship networks inform autism spectrum disorder gene prediction. *Translational Psychiatry*, 8(1), 56.
- Evans, D. W., Uljarević, M., Lusk, L. G., Loth, E., & Frazier, T. (2017). Development of two dimensional measures of restricted and repetitive behavior in parents and children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(1), 51–58.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The reactome pathway knowledgebase. *Nucleic Acids Research*, 46(D1), D649–D655.
- Fernandez, T. V., Williams, Z. P., Kline, T., Rajendran, S., Augustine, F., Wright, N., et al. (2023). Primary complex motor stereotypies are associated with de novo damaging dna coding mutations that identify kdm5b as a risk gene. *PLoS One*, 18(10), e0291978.
- Filipello, F., Morini, R., Corradini, I., Zerbi, V., Canzi, A., Michalski, B., Erreni, M., Markicewicz, M., Wang, Y., Mortara, L., Tonon, G., Colombo, M., Bagri, A., Vannella, K. M., Fahnestock, M., Paolicelli, R. C., & Matteoli, M. (2018). The microglial innate immune receptor trem2 is required for synapse elimination and normal brain connectivity. *Immunity*, 48(5), 979–991.

- Fitzpatrick, P., Romero, V., Amaral, J. L., et al. (2017). Social motor synchronization: Insights for understanding social behavior in autism. *Journal of Autism and Developmental Disorders*, 47, 2092–2107.
- Forti, S., Valli, A., Perego, P., Nobile, M., Crippa, A., & Molteni, M. (2011). Motor planning and control in autism. a kinematic analysis of preschool children. *Research in Autism Spectrum Disorders*, 5(2), 834–842.
- Francés, L., Quintero, J., Fernández, A., Ruiz, A., Caules, J., Fillon, G., Hervás, A., & Soler, C. V. (2022). Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: A systematic review in accordance with the PRISMA criteria. *Child and Adolescent Psychiatry and Mental Health*, 16(1), 27.
- Garbett, K., Ebert, P. J., Mitchell, A., Lintas, C., Manzi, B., Mirnics, K., & Persico, A. M. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiology of Disease*, 30(3), 303–311.
- Gaugler, T., Klei, L., Sanders, S., Bodea, C., Goldberg, A., Lee, A., & et al. (2014). Most genetic risk for autism resides with common variation. *Nature Genetics*, 46(8), 881–885.
- Green, J. (2023). Debate: Neurodiversity, autism and healthcare. *Child and Adolescent Mental Health*, 28, 438–442.
- Griffiths, R. (1984). *The abilities of young children: A comprehensive system of mental measurement for the first eight years of life*.
- Grote, S., Prüfer, K., Kelso, J., & Dannemann, M. (2016). ABAEnrichment: An R package to test for gene set expression enrichment in the adult and developing human brain. *Bioinformatics*, 32(20), 3201–3203.
- Grove, J., Ripke, S., Als, T., Mattheisen, M., Walters, R., Won, H., & et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, 51(3), 431–444.
- Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., Valieris, R., Köster, J., & Bioconda Team. (2018). Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nature Methods*, 15(7), 475–476.
- Happé, F., & Frith, U. (2020). Annual research review: Looking back to look forward - changes in the concept of autism and implications for future research. *Journal of Child Psychology and Psychiatry*, 61(3), 218–232.

- Harris, A. D., Singer, H. S., Horska, A., Kline, T., Ryan, M., Edden, R. A., & Mahone, E. M. (2016). Gaba and glutamate in children with primary complex motor stereotypies: An 1h-mrs study at 7t. *AJNR. American Journal of Neuroradiology*, 37(3), 552–557.
- Hedley, D., & Uljarević, M. (2018). Systematic review of suicide in autism spectrum disorder: Current trends and implications. *Current Developmental Disorders Reports*, 5, 65–76.
- Hegemann, L., Corfield, E. C., Askelund, A. D., Allegrini, A. G., Askeland, R. B., Ronald, A., Ask, H., St Pourcain, B., Andreassen, O. A., Hannigan, L. J., & Havdahl, A. (2024). Genetic and phenotypic heterogeneity in early neurodevelopmental traits in the norwegian mother, father and child cohort study. *Molecular Autism*, 15(1), 25.
- Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science*, 344(6190), 1243091.
- Hiller, R. M., Young, R. L., & Weber, N. (2014). Sex differences in autism spectrum disorder based on dsm-5 criteria: Evidence from clinician and teacher reporting. *Journal of Abnormal Child Psychology*, 42(8), 1381–1393.
- Hirvikoski, T., Mittendorfer-Rutz, E., Boman, M., Larsson, H., Lichtenstein, P., & Bölte, S. (2016). Premature mortality in autism spectrum disorder. *The British Journal of Psychiatry*, 208(3), 232–238.
- Hollingshead, A. B. (1975). *Four factor index of social status*. Yale University.
- Huguet, G., Benabou, M., & Bourgeron, T. (2016). The genetics of autism spectrum disorders. In P. Sassone-Corsi & Y. Christen (Eds.), *A time for metabolism and hormones*. Springer.
- Huguet, G., Ey, E., & Bourgeron, T. (2013). The genetic landscapes of autism spectrum disorders. *Annual Review of Genomics and Human Genetics*, 14, 191–213.
- Hull, L., Lai, M.-C., Baron-Cohen, S., Allison, C., Smith, P., Petrides, K., & Mandy, W. (2019). Gender differences in self-reported camouflaging in autistic and non-autistic adults. *Autism*, 23(4), 942–953.
- Hull, L., Petrides, K., & Mandy, W. (2020). The female autism phenotype and camouflaging: A narrative review. *Review Journal of Autism and Developmental Disorders*, 7, 306–317.
- Jacob, S., Wolff, J. J., Steinbach, M. S., Doyle, C. B., Kumar, V., & Elison, J. T. (2019). Neurodevelopmental heterogeneity and computational approaches for understanding autism. *Translational Psychiatry*, 9(1), 63.

- Jarus, T., Lourie-Gelberg, Y., Engel-Yeger, B., & Bart, O. (2011). Participation patterns of school-aged children with and without dcd. *Research in Developmental Disabilities*, 32(4), 1323–1331.
- Jeste, S., & Geschwind, D. (2014). Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nature Reviews Neurology*, 10(2), 74–81.
- Jokiranta-Olkonien, E., Cheslack-Postava, K., Sucksdorff, D., & et al. (2016). Risk of psychiatric and neurodevelopmental disorders among siblings of probands with autism spectrum disorders. *JAMA Psychiatry*, 73(6), 622–629.
- Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research*, 38(suppl_1), D355–D360.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, 2(3), 217–250.
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434–443.
- Kasem, E., Kurihara, T., & Tabuchi, K. (2018). Neurexins and neuropsychiatric disorders. *Neuroscience Research*, 127, 53–60.
- Keating, J., Van Goozen, S., Uljarevic, M., Hay, D., & Leekam, S. R. (2023). Restricted and repetitive behaviors and their developmental and demographic correlates in 4–8-year-old children: A transdiagnostic approach. *Frontiers in Behavioral Neuroscience*, 17, 1085404.
- Ketcheson, L. R., Pitchford, E. A., & Wentz, C. F. (2021). The relationship between developmental coordination disorder and concurrent deficits in social communication and repetitive behaviors among children with autism spectrum disorder. *Autism Research*, 14, 804–816.
- Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gouridine, J.-P., Gargano, M., Harris, N. L., Matentzoglou, N., McMurphy, J. A., et al. (2019). Expansion of the human phenotype ontology (HPO) knowledge base and new applications. *Nucleic Acids Research*, 47(D1), D1018–D1027.
- Koster-Hale, J., & Saxe, R. (2013). Theory of mind: A neural prediction problem. *Neuron*, 79(5), 836–848.
- Lai, M.-C., Kasse, C., Besney, R., & et al. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *Lancet Psychiatry*, 6, 819–829.

- Lai, M.-C., Lombardo, M., & Baron-Cohen, S. (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*, 21(6), 690–702.
- Lai, M.-C., Lombardo, M. V., Pasco, G., Ruigrok, A. N. V., Wheelwright, S. J., Sadek, S. A., & et al. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PLoS ONE*, 6(6), e20835.
- Lai, M.-C., & Baron-Cohen, S. (2015). Identifying the lost generation of adults with autism spectrum conditions. *Lancet Psychiatry*, 2(11), 1013–1027.
- Langen, M., Schnack, H. G., Nederveen, H., Bos, D., Lahuis, B. E., de Jonge, M. V., van Engeland, H., & Durston, S. (2009). Changes in the developmental trajectories of striatum in autism. *Biological Psychiatry*, 66(4), 327–333.
- Larkin, F., Meins, E., Centifanti, L. C., Fernyhough, C., & Leekam, S. R. (2017). How does restricted and repetitive behavior relate to language and cognition in typical development? *Development and Psychopathology*, 29(3), 863–874.
- Leonard, H., & Hill, E. (2014). The impact of motor development on typical and atypical social cognition and language: A systematic review. *Child and Adolescent Mental Health*, 19(3), 163–170.
- Li, Q., & Wang, K. (2017). InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *The American Journal of Human Genetics*, 100(2), 267–280.
- Licari, M. K., Alvares, G. A., Varcin, K., Evans, K. L., Cleary, D., Reid, S. L., Glasson, E. J., Bebbington, K., Reynolds, J. E., Wray, J., & Whitehouse, A. J. O. (2020). Prevalence of motor difficulties in autism spectrum disorder: Analysis of a population-based cohort. *Autism Research*, 13, 298–306.
- Liccione, D. (2019). *Psicoterapia cognitiva neuropsicologica*. Bollati Boringhieri.
- Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56(6), 466–474.
- Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, 6(1), 5.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., & Gotham, K. (2012). *Autism diagnostic observation schedule, volume ados-2 (2nd ed.)* Western Psychological Services.

- Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., de Vries, P. J., Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., . . . McCauley, J. B. (2022). The lancet commission on the future of care and clinical research in autism. *The Lancet*, 399(10321), 271–334.
- Lotter, V. (1966). Epidemiology of autistic conditions in young children. *Social Psychiatry*, 1, 124–135.
- Louvi, A., & Grove, E. A. (2011). Cilia in the cns: The quiet organelle claims center stage. *Neuron*, 69(6), 1046–1060.
- Lundström, S., Reichenberg, A., Melke, J., Råstam, M., Kerekes, N., Lichtenstein, P., Gillberg, C., & Anckarsäter, H. (2015). Autism spectrum disorders and coexisting disorders in a nationwide swedish twin study. *Journal of Child Psychology and Psychiatry*, 56, 702–710.
- Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., Christensen, D. L., Wiggins, L. D., Pettygrove, S., Andrews, J. G., Lopez, M., Hudson, A., Baroud, T., Schwenk, Y., White, T., Rosenberg, C. R., Lee, L. C., Harrington, R. A., Huston, M., & Dietz, P. M. (2020). Prevalence of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, united states, 2016. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, 69(4), 1–12.
- Maenner, M. J., Shaw, K. A., Bakian, A. V., Bilder, D. A., Durkin, M. S., Esler, A., Furnier, S. M., Hallas, L., Hall-Lande, J., Hudson, A., Hughes, M. M., Patrick, M., Pierce, K., Poynter, J. N., Salinas, A., Shenouda, J., Vehorn, A., Warren, Z., Constantino, J. N., . . . Cogswell, M. E. (2021). Prevalence and characteristics of autism spectrum disorder among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, united states, 2018. *Morbidity and Mortality Weekly Report. Surveillance Summaries*, 70(11), 1–16.
- Mahone, E. M., Crocetti, D., Tochen, L., Kline, T., Mostofsky, S. H., & Singer, H. S. (2016). Anomalous putamen volume in children with complex motor stereotypies. *Pediatric Neurology*, 65, 59–63.
- Maleki, F., Ovens, K., Hogan, D. J., & Kusalik, A. J. (2020). Gene set analysis: Challenges, opportunities, and future research. *Frontiers in Genetics*, 11, 654.
- Mandelli, V., Landi, I., Ceccarelli, S. B., Molteni, M., Nobile, M., D’Ausilio, A., Fadiga, L., Crippa, A., & Lombardo, M. V. (2024). Enhanced motor noise in an autism subtype with poor motor skills. *Molecular Autism*, 15(1), 36.

- Mandelli, V., Severino, I., Eyler, L., Pierce, K., Courchesne, E., & Lombardo, M. V. (2024). A 3d approach to understanding heterogeneity in early developing autisms. *medRxiv*.
- Masi, A., DeMayo, M. M., Glozier, N., & Guastella, A. J. (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience Bulletin*, 33(2), 183–193.
- Melo, C., Ruano, L., Jorge, J., Pinto Ribeiro, T., Oliveira, G., Azevedo, L., & Temudo, T. (2020). Prevalence and determinants of motor stereotypies in autism spectrum disorder: A systematic review and meta-analysis. *Autism*, 24(3), 569–590.
- Meltzer, A., & Van de Water, J. (2017). The role of the immune system in autism spectrum disorder. *Neuropsychopharmacology*, 42(1), 284–298.
- Milton, D. (2017). *A mismatch of salience: Explorations of the nature of autism from theory to practice*. Pavilion Press.
- Murtagh, F., & Legendre, P. (2014). Ward’s hierarchical agglomerative clustering method: Which algorithms implement Ward’s criterion? *Journal of Classification*, 31(3), 274–295.
- Nam, D., & Kim, S.-Y. (2008). Gene-set approach for expression pattern analysis. *Briefings in Bioinformatics*, 9(3), 189–197.
- Narzisi, A., Posada, M., Barbieri, F., Chericoni, N., Ciuffolini, D., Pinzino, M., Romano, R., Scattoni, M. L., Tancredi, R., Calderoni, S., & Muratori, F. (2018). Prevalence of autism spectrum disorder in a large italian catchment area: A school-based population study within the asdeu project. *Epidemiology and Psychiatric Sciences*, 29, e5.
- Nasca, B. C., Lopata, C., Donnelly, J. P., & et al. (2020). Sex differences in externalizing and internalizing symptoms of children with asd. *Journal of Autism and Developmental Disorders*, 50, 3245–3252.
- O’Halloran, L., Coey, P., & Wilson, C. (2022). Suicidality in autistic youth: A systematic review and meta-analysis. *Clinical Psychology Review*, 93, 102144.
- Ormstad, H., Bryn, V., Saugstad, O. D., Skjeldal, O., & Maes, M. (2018). Role of the immune system in autism spectrum disorders. *CNS & Neurological Disorders – Drug Targets*, 17(7), 489–495.
- Pacheco, J., Garvey, M. A., Sarampote, C. S., Cohen, E. D., Murphy, E. R., & Friedman-Hill, S. R. (2022). Annual research review: The contributions of the rdcc research framework on understanding the neurodevelopmental origins, progression and treatment of mental illnesses. *Journal of Child Psychology and Psychiatry*, 63(4), 360–376.

- Péter, Z., Oliphant, M. E., & Fernandez, T. V. (2017). Motor stereotypies: A pathophysiological review. *Frontiers in Neuroscience*, *11*, 171.
- Powell, N., Walker, M. M., & Talley, N. J. (2017). The mucosal immune system: Master regulator of bidirectional gut–brain communications. *Nature Reviews Gastroenterology & Hepatology*, *14*(3), 143–159.
- R Core Team. (2013). *R: A language and environment for statistical computing* [Retrieved from <http://www.r-project.org/>].
- Rong, Y., Yang, C.-J., Jin, Y., & Wang, Y. (2021). Prevalence of attention-deficit/hyperactivity disorder in individuals with autism spectrum disorder: A meta-analysis. *Research in Autism Spectrum Disorders*, *83*, 101759.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.
- Rynkiewicz, A., Schuller, B., Marchi, E., et al. (2016). An investigation of the ‘female camouflage effect’ in autism using a computerized ados-2 and a test of sex/gender differences. *Molecular Autism*, *7*, 10.
- Sabariego-Navarro, M., Fernández-Blanco, Á., Sierra, C., & Dierssen, M. (2022). Neurodevelopmental disorders: 2022 update. *Free Neuropathology*, *3*, 8.
- Salari, N., Rasoulpoor, S., Rasoulpoor, S., & et al. (2022). The global prevalence of autism spectrum disorder: A comprehensive systematic review and meta-analysis. *Italian Journal of Pediatrics*, *48*, 112.
- Sandin, S., Yip, B. H. K., Yin, W., et al. (2024). Examining sex differences in autism heritability. *JAMA Psychiatry*, *81*(7), 673–680.
- Satterstrom, F., Kosmicki, J., Wang, J., Breen, M., De Rubeis, S., An, J., & et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, *180*(3), 568–584.e23.
- Scaccabarozzi, G., Fumagalli, L., Mambretti, M., Giorda, R., Villa, M., Busti Ceccarelli, S., Villa, L., Mani, E., Nobile, M., Molteni, M., Pozzoli, U., & Crippa, A. (2025). Potentially damaging variants’ analysis in autism subgroups uncovers early brain-expressed gene modules relevant to autism pathophysiology.
- Scattoni, M. L., Fatta, L. M., Micai, M., et al. (2023). Autism spectrum disorder prevalence in Italy: A nationwide study promoted by the ministry of health. *Child and Adolescent Psychiatry and Mental Health*, *17*, 125.

- Schmunk, G., & Gargus, J. J. (2013). Channelopathy pathogenesis in autism spectrum disorders. *Frontiers in Genetics*, 4, 222.
- Shen, L., Liu, X., Zhang, H., Lin, J., Feng, C., & Iqbal, J. (2020). Biomarkers in autism spectrum disorders: Current progress. *Clinica Chimica Acta*, 502, 41–54.
- Simcoe, S. M., Brownlow, C., Garnett, M. S., Rynkiewicz, A., & Attwood, T. (2018). Profiling autism symptomatology: An exploration of the q-asc parental report scale in capturing sex differences in autism. *Journal of Autism and Developmental Disorders*, 48(2), 389–403.
- Simons Foundation. (n.d.). Simons foundation [Retrieved June, 2026]. <https://www.simonsfoundation.org/>
- Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: A general repository for interaction datasets. *Nucleic Acids Research*, 34(suppl_1), D535–D539.
- Stephan, A. H., Barres, B. A., & Stevens, B. (2012). The complement system: An unexpected role in synaptic pruning during development and disease. *Annual Review of Neuroscience*, 35, 369–389.
- Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., & et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545–15550.
- Südhof, T. C. (2017). Synaptic neuroligin complexes: A molecular code for the logic of neural circuits. *Cell*, 171(4), 745–769.
- Taboas, A., Doepke, K., & Zimmerman, C. (2022). Preferences for identity-first versus person-first language in a us sample of autism stakeholders. *Autism*, 27(2), 565–570.
- Thapar, A., Cooper, M., & Rutter, M. (2017). Neurodevelopmental disorders. *The Lancet Psychiatry*, 4(4), 339–346.
- Tubío-Figueirinho, M., Cruz, S., Sampaio, A., et al. (2021). Social camouflaging in females with autism spectrum disorder: A systematic review. *Journal of Autism and Developmental Disorders*, 51, 2190–2199.
- Uljarević, M., Arnott, B., Carrington, S. J., Meins, E., Fernyhough, C., McConachie, H., & Leekam, S. R. (2017). Development of restricted and repetitive behaviors from 15 to 77 months: Stability of two distinct subtypes? *Developmental Psychology*, 53(10), 1859.

- Valenti, M., Vagnetti, R., Masedu, F., Pino, M. C., Rossi, A., Scattoni, M. L., & et al. (2019). Register-based cumulative prevalence of autism spectrum disorders during childhood and adolescence in central Italy. *ebph*, 16(4).
- Vaquero-Lorenzo, C., & Vasquez, M. A. (2020). Suicide: Genetics and heritability. *Current Topics in Behavioral Neurosciences*, 46, 63–78.
- Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., & Geschwind, D. H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351), 380–384.
- Walther, S., & Mittal, V. A. (2022). Motor behavior is relevant for understanding mechanism, bolstering prediction, and improving treatment: A transdiagnostic perspective. *Schizophrenia Bulletin*, 48(4), 741–748.
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164.
- Wechsler, D. (2002). *Wechsler preschool and primary scale of intelligence (wppsi-iii)* (3rd ed.). Psychological Corporation.
- Wechsler, D. (2012). *Wechsler intelligence scale for children-iv Italian edition*. Organizzazioni Speciali.
- Werling, D. M., & Geschwind, D. H. (2013). Sex differences in autism spectrum disorders. *Current Opinion in Neurology*, 26(2), 146–153.
- Wiggins, L. D., Rubenstein, E., Windham, G., Barger, B., Croen, L., Dowling, N., Giarelli, E., Levy, S., Moody, E., Soke, G., Fields, V., & Schieve, L. (2021). Evaluation of sex differences in preschool children with and without autism spectrum disorder enrolled in the study to explore early development. *Research in Developmental Disabilities*, 112, 103897.
- William, N., Reissner, C., Sargent, R., Darlington, T. M., DiBlasi, E., Li, Q. S., Keeshin, B., Callor, W. B., Ferris, E., Jerominski, L., Smith, K. R., Christensen, E. D., Gray, D. M., Camp, N. J., Missler, M., Williams, M. E., & Coon, H. (2021). Neurexin 1 variants as risk factors for suicide death. *Molecular Psychiatry*, 26(12), 7436–7445.
- Williams, S. M., An, J. Y., Edson, J., & et al. (2019). An integrative analysis of non-coding regulatory DNA variations associated with autism spectrum disorder. *Molecular Psychiatry*, 24, 1707–1719.

- Wilson, B. N., Crawford, S. G., Green, D., Roberts, G., Aylott, A., & Kaplan, B. J. (2009). Psychometric properties of the revised developmental coordination disorder questionnaire. *Physical & Occupational Therapy in Pediatrics*, 29(2), 182–202.
- Wilson, B. N., Kaplan, B. J., Crawford, S. G., & Roberts, G. (2007). Cross-cultural adaptation, reliability and predictive validity of the italian version of developmental coordination disorder questionnaire (dcdq). *European Journal of Paediatric Neurology*, 18, 267–272.
- Wilson, R. B., McCracken, J. T., Rinehart, N. J., et al. (2018). What's missing in autism spectrum disorder motor assessments? *Journal of Neurodevelopmental Disorders*, 10, 33.
- Young, H., Oreve, M. J., & Speranza, M. (2018). Clinical characteristics and problems diagnosing autism spectrum disorder in girls. *Archives de Pédiatrie*, 25(6), 399–403.
- Zeidan, J., Fombonne, E., Scolah, J., Ibrahim, A., Durkin, M. S., Saxena, S., Yusuf, A., Shih, A., & Elsabbagh, M. (2022). Global prevalence of autism: A systematic review update. *Autism Research*, 15(5), 778–790.