

Cluster Analysis of Short Sensory Profile Data Reveals Sensory-based Subgroups in Autism Spectrum Disorder

Ariel Maia Lyons-Warren (✉ lyonswar@bcm.edu)

Baylor College of Medicine <https://orcid.org/0000-0002-8584-5525>

Ying-Wooi Wan

Baylor College of Medicine

Research

Keywords: Sensory processing, SSP, Autism Spectrum Disorder, Phenotyping, Neurodevelopmental disorders, Genetic variant, Heterogeneity

Posted Date: March 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-248609/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at International Journal of Molecular Sciences on October 27th, 2023. See the published version at <https://doi.org/10.3390/ijms232113030>.

Abstract

Background: Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairment in social interactions and communication. Additional features include restricted, repetitive patterns of behaviors, and differences in sensory processing. The clinical presentation of patients with ASD is heterogeneous, likely reflecting multiple underlying etiologies. Heterogeneity in presentation and treatment response are barriers to development of precise therapeutic approaches. Therefore, identification of clinically meaningful subgroups within ASD is critical to develop targeted interventions. We hypothesized that sensory features can be used to identify clinically recognizable subgroups with shared underlying etiologies.

Methods: Subjects included 378 individuals with a clinical diagnosis of ASD who contributed Short Sensory Profile (SSP) data assessing the frequency of sensory behaviors and whole genome sequencing results to the Autism Speaks' MSSNG database. To determine if the SSP could be used to subgroup individuals with ASD, we performed cluster analysis on responses to all 38 questions, followed by an independent cluster analysis using only a subset of questions selected specifically to assay hyper- and hypo-sensitivity to sensory stimulation. Cross-validation of the resulting clusters determined the final subgroups. To test for shared underlying etiologies, we correlated variant frequency across subgroups for each of 24,896 genes. Variant frequency included any variation in each gene regardless of the type of variant. To be significantly associated with a subgroup, a gene variant frequency had to be greater than four standard deviations (SD) from the mean frequency for all subgroups and 3 SD different from each subgroup.

Results: We identified seven distinct sensory-based ASD subgroups. Subgroup 1, characterized by atypical scores in all sensory areas, was not associated with any genes. Subgroups 2, 4 and 6 were significantly associated with four to six genes each. Subgroups 3, 5 and 7 were enriched for 126, 12 and 50 genes, respectively.

Limitations: This study was performed using retrospective data that did not include other phenotypic data such as age, comorbidities, or measures of disease severity. All those likely contribute to the variability of the identified subgroups

Conclusions: These results support the use of sensory features to identify ASD subgroups with shared genetic mechanisms.

Background

Autism spectrum disorder (ASD) is a group of pervasive neurodevelopmental disorders defined by the presence of language delay, impairment in social interactions and repetitive, restricted behaviors (Masi et al., 2017). Despite a high prevalence of at least 1.5% and increasing family and societal costs (Lyll et al., 2017, Baxter et al., 2015, Kogan et al., 2008), there are currently no targeted pharmacological treatments for ASD (Pandina et al., 2020). A major barrier to development of targeted treatments is the heterogeneity

in disease presentation seen across patients with ASD (Betancur et al., 2011, Jeste et al., 2014). Varied clinical presentations likely reflect diverse underlying molecular mechanisms. Similarly, diversity in patient groups may obscure efficacy in treatment trials. Thus, it is necessary to develop a standardized approach to distinguish different subgroups of patients with ASD based on both clinical features and underlying molecular/genetic etiologies (Ingram et al, 2008, Jacob et al., 2019, Wolfers et al., 2019).

Sensory processing differences are a common and debilitating component of the diagnostic criteria for ASD (Posar and Visconti 2018). As such, sensory-based subgrouping of ASD patients is promising in the classification of the heterogeneous ASD patients (Lane et al., 2014). Notably, sensory profiles where the specific pattern of sensory features is present in a certain patient population have been applied to other neurodevelopmental disorders such as Angelman, Cornelia de Lange and Fragile X syndromes (Heald et al, 2020) as well as Phelan-McDermid Syndrome and *SYNGAP-1* Associated intellectual disability (Lyons-Warren et al, 2020). Further, sensory differences have been correlated with other features of ASD such as level of functioning (Klintwall et al., 2011). Given the strong evidence supporting the use of sensory features to identify subgroups within ASD, it is not surprising that previous studies have used cluster analysis to identify sensory-based subgroups (reviewed in Simpson et al., 2019). However, despite multiple studies, no consensus about ASD subgrouping has been reached, reflecting the heterogeneity of ASD as well as the variety of different sensory survey tools used. Finally, no previous proposed sensory-based subgroups have been linked to underlying molecular mechanisms or genetic variants.

Here, we capitalize on Short Sensory Profile (SSP) and whole genome sequencing (WGS) data from 378 patients with ASD, collected as part of the Autism Speaks' MSSNG Database (<https://research.mssng.org>). We report seven sensory-based ASD subgroups with shared underlying genetic mechanisms, paving the way toward better classification of this heterogeneous patient population and development of targeted therapeutics.

Methods

The aim of this study was to identify subgroups of patients with ASD based on their sensory phenotypes and to determine if these subgroups shared underlying genetic mechanisms. Towards this aim, this study utilized Short Sensory Profile (SSP) data and whole genome sequencing (WGS) results for 378 subjects, of which 80% (303/378) were male, obtained from the Autism Speaks' MSSNG Database. Additional information including age or spectrum of impairment for these individuals was not available. The Institutional Review Board for human studies at Baylor College of Medicine reviewed this study and granted an exemption based on the absence of individually identifying information in the database.

Procedures

The SSP (McIntosh et al. 1999) is a 38-question parent survey that quantifies frequency of sensory behaviors using a 5-point likert scale (1 = always, 5 = never). SSP data have previously been used to assess sensory features in many neurodevelopmental disorders (Mangeot et al., 2001; Carr et al., 2010;

John et al., 2010; Bruni et al., 2010) including ASD (Rogers et al., 2003; Tomchek and Dunn 2007; Ashburner et al., 2008; Chen et al., 2009). Questions are divided into seven sections termed sensory areas: Tactile Sensitivity, Taste/Smell Sensitivity, Movement Sensitivity, Underresponsive/Seeks Sensation, Auditory Filtering, Low Energy/Weak, and Visual/Auditory Sensitivity. Scores are summed in each area and compared to standardized scores for typically developing children. Subjects are categorized as typically responding, probably different, or definitely different.

In order to more directly assay hyper- and hypo-sensitivity, we extracted questions that asked specifically about hyper- or hypo-sensitivity in each of the five main sensory domains using best clinical judgement (Table 1). Two of the sensory areas in the SSP evaluate features outside the five main sensory domains of hearing, vision, touch, taste, and smell. For example, the Movement Sensitivity category asks how often the child “fears falling or heights.” These questions were not included in the hyper- and hypo-sensitivity sub-analysis. Similarly, some questions regarding primary senses cannot be extrapolated to hyper- or hypo-sensitivity, such as “Rubs or scratches out a spot that has been touched.” Scores for the subset of questions within each sensory domain were summed and assigned as typically responsive or hyper/hypo responsive using adjusted cut-offs based on the SSP validated data (McIntosh et al. 1999). Cluster analyses performed using both the full SSP and this subset of questions. (Fig. 1)

Data Analysis

First, histogram visualization of the SSP responses in each of the seven sensory areas and the four sensory domains allowed evaluation of the range of responses for all 378 subjects. One subject was excluded from further analysis due to incomplete data (Fig. 1). Next, we performed two independent cluster analyses on SSP responses from 377 individuals using the responses to all 38 questions or the subset of responses focused on hyper-and hypo-sensitivity (Table 1). These datasets will be referred to as “38Question” or “Hyper/hypo sensitivity” respectively. The cluster analysis was done using K-mean clustering on bootstrap samples and then using consensus clusters from the bootstraps as the final clusters. Final clusters for K = 3–10 were visually compared for each data set. To cross-validate the data, we compared subgroups identified by each clustering method to each other and only included the 222 subjects in overlapping clusters that could be defined using both methods in the final analysis.

Variant Extraction and Analysis

WGS results within the MSSNG database are stored in Google Query. All variants that were annotated to a gene and had a minimum Combined Annotation Dependent Depletion (CADD) score of 15 were extracted from this database. In order to be exploratory, a lower threshold that included both deleterious and common variants was used. This list was then used to call variants from each subject. Called variants were aggregated at a gene level for each subject meaning a variant was marked observed if at least one variant annotated to the gene was found regardless of the variant type. Using this method, 24,896 genes for which at least one patient had a variant were identified. For each gene, the variant frequency was calculated as the percent of subjects in each cluster with variant(s) in that gene. Variant frequency did not follow a normal distribution; therefore the nonparametric test Kruskal-Wallis was used to compare

frequencies across clusters. In order to reduce the number of genes to a manageable size in an unbiased fashion, genes for which the variant frequency differed across clusters were prioritized using a cut-off chosen based on the distribution of variant frequencies. Specifically, 58.44% (14,550/24,896) of genes had a variant frequency less than 10% in all clusters and were excluded. To determine if variant frequency was statistically associated with a given cluster, the mean difference in variant frequency for each gene for each cluster was plotted. Genes with variant frequency more than 4SD above or below the mean variant frequency in the other six clusters were evaluated further. Within this subset, genes with variant frequency more than 3SD above or below the variant frequency of each other cluster were considered significantly associated with that subgroup. (Fig. 1)

Results

Short Sensory Profile Responses in non-syndromic ASD are heterogeneous

The primary goal of this study was to use sensory features to identify subgroups within ASD. In order to create subgroups, there must be variation within the group. Therefore, we first looked at the distribution of SSP scores within our cohort. Consistent with prior reports (Rogers et al., 2003; Tomchek and Dunn 2007; Ashburner et al., 2008; Chen et al., 2009), a wide range of scores in each of the seven sensory areas assayed by the SSP was observed (Fig. 2A). Most histograms skewed towards higher values associated with typical scoring ranges. However, the score distributions for under-responsiveness-seeks sensation and auditory filtering exhibited a bell shape distribution with the peak in mid-range scores suggesting most subjects were probably different from typically developing children in those areas (Fig. 2A). Only eight subjects (2%) scored in the typically performing range (Fig. 2A, green shading) for all sensory areas. Similarly, only 24 subjects (6%) scored in the definitely different range (Fig. 2A, blue shading) for all sensory areas. Scoring in the probably or definitely different range for all sensory areas was slightly more common, seen in 54 subjects (14%). Overall, the scores were broadly distributed suggesting sufficient variation to allow for clustering.

Meaningful subgrouping requires that the features being grouped are not random. Therefore, we next asked if all possible combinations of sensory features were present. To answer this question, data were simplified into binary categories in which all subjects were classified as typically performing or different from typical performance including probably different and definitely different using standardized, previously published ranges from a group of 1,037 children without disabilities (McIntosh et al. 1999). Using this method, only 82 out of 128 possible combinations of sensory areas were present in the cohort suggesting that sensory changes are not random and that certain sensory areas tend to be affected together. In fact, the 10 most common patterns described 162 subjects (43%) (Table 2). 45 patterns were rare, being seen in three or fewer subjects each.

Broad heterogeneity was also present in the distribution of hyper- and hypo-sensitivity scores (Fig. 2B) derived from a subset of the SSP responses (Table 1). Notably, many subjects were both hyper and

hyposensitive in the modalities of touch (47, 12.4%), hearing (156, 41%) and vision (35, 9%). Consistent with the analysis of the seven sensory areas defined by the SSP, only 26 subjects (7%) were hypersensitive in all sensory modalities. 32 subjects (8.4%) were neither hyper- nor hypo-sensitive in any sensory modality. Overall, scores in the subset analysis were similarly broadly distributed suggesting sufficient variation to allow for clustering.

Cluster analysis optimizes at seven sensory based subgroups

Having determined that SSP scores in this cohort were sufficiently broadly distributed to allow for clustering, we performed cluster analysis on bootstrap samples for 100 iterations using K-means clustering on responses from 377 subjects. The K-mean clustering was run with K = 3 to K = 10 on bootstrap samples for 100 iterations and the consensus clusters for each K were used as the final cluster for subsequent analyses. This cluster analysis was separately performed using both responses to all 38 questions (Fig. 3A) and the hyper/hypo sensitivity subset (Fig. 3B). Visual inspection of the consensus heatmaps illustrated that K = 3 provided the most consistent clustering. However, these groupings could be described by “affected in all areas,” “affected in no areas” and “all other subjects” and thus were not clinically relevant. Overall, we determined six clusters represented the most reproducible and consistent number of subgroups (Fig. 3).

Cluster analysis of responses to all 38 questions in the SSP identified six subgroups (Fig. 4A). Group 1 (N = 72) was characterized by high percentage of subjects who scored probably or definitely different in all seven areas. Similarly, group 2 (N = 72) was characterized by predominantly typical performance in six of the seven sensory areas. For group 2, subjects scored probably different only in area seven (visual/auditory sensitivity). Each of the remaining four subgroups exhibited a unique pattern (Fig. 4A). For example, group 3 (N = 64) was defined by subjects scoring in the definitely different range in all areas except three and six (movement sensitivity and low energy/weak). In contrast, group 4 (N = 79) was defined by subjects scoring in the definitely different range only in areas four and five (under-responsiveness/seeks sensation and auditory filtering) paired with probably different scoring in area seven (visual/auditory sensitivity). Group 5 (N = 63) was associated with isolated scoring in the definitely different range for areas two and five (taste/smell sensitivity and auditory filtering) as well as probably different scores for area seven. Finally, group 6 (N = 27) was associated with scores in the definitely different range for areas five and six (auditory filtering and low energy) with mixed scores in areas one through four as well as probably different scores for area seven. Consistent with the distribution of scores in this cohort (Fig. 2), subjects scoring probably or definitely different in areas five (auditory filtering) and seven (visual/auditory sensitivity) were seen in all six clusters. Thus, cluster analysis identified a cohort of subjects likely to be different in all or almost all areas, subjects mostly scoring in the typically developing range and four unique clusters with distinct patterns of sensory features.

Next, we performed the same analysis using the subset of questions evaluating hyper and hypo sensitivity (Table 2, Fig. 4B). Notably, group 1 (N = 92), characterized by increased hyper- and hypo-

sensitivity in all sensory modalities, was similar to group 1 in the 38-question analysis. Similarly, group 4 (N = 51) was characterized by the absence of hyper- or hypo-sensitivity in all sensory modalities and group 3 (N = 60) was characterized by isolated auditory hyposensitivity. These two groups corresponded to 38-question cluster 2 phenotypically and had the greatest overlap in subjects with 38-question group 2. The remaining three groups were defined by distinct patterns of hyper- and hypo-sensitivity. Specifically, group 2 (N = 74) was characterized by auditory hyper- and hypo-sensitivity. Group 5 (N = 58) was characterized by hypo-sensitivity in the auditory and tactile domains. Finally, group 6 (N = 42) was characterized by auditory hypo-sensitivity and hyper-sensitivity to taste. (Fig. 4B) Similar to the 38-question analysis, five of the six groups included auditory hypo-sensitivity reflecting the high prevalence of auditory behaviors in this cohort. Thus, both cluster analyses identified a cohort of subjects that were very affected, a group mostly scoring in the typically developing or unaffected range and at least three unique clusters with distinct patterns of sensory features.

Having identified phenotypically distinct subgroups using two methods, the subgroups were validated by evaluating the degree of overlap between the two methods. Not unexpectedly given the corresponding phenotypic features, there was notable overlap between the groups identified in each method (Fig. 5). However, 38-question group 2 was divided between hyper/hypo sensitivity groups 3 and 4. Similarly, hyper/hypo sensitivity group 1 had the greatest overlap with both 38-question groups 1 and 6. The overlap between the two methods best identified seven subgroups (Fig. 5, highlighted boxes) comprising 222 subjects.

Thus, we concluded that sensory features can reliably be used to generate clinically distinct subgroups within a heterogeneous population of non-syndromic ASD patients. Further, we conclude that in the MSSNG cohort, there are seven distinct subgroups.

Subgroups are associated with genetic variants

Having successfully identified seven sensory based subgroups, we next asked if these subgroups predicted shared molecular mechanisms as indicated by correlation with genetic variants using whole genome sequencing (Fig. 1). The analysis included variants with at least a CADD score of 15 and annotation to a gene from the MSSNG database which were aggregated to a gene level for each patient for further analysis. Whole genome sequencing results were not available for seven of the subjects coming from five of the seven subgroups. In total, we identified 24,896 genes for which at least one patient had an annotated variant.

In order to ask if variants in a given gene were associated with a subgroup, we calculated variant frequency within each subgroup for each of these genes. The mean variant frequency was 9.14% (Fig. 6A). Most genes had a low variant frequency, with the mean variant frequency at 2SD being only 30%. Importantly, because all variants were collapsed, this analysis is looking at both common and rare variant frequency and thus a low variant frequency is not unexpected. Many (41.5%) genes had a low variant frequency in *all* subgroups meaning less than 10% of subjects in the subgroup had a variant. In fact, the mean variant frequency for each subgroup was less than 13% (range 7.1% – 12.2%). Group 3

had the highest mean variant frequency at 12.2% which was significantly higher than all other groups (Kruskal-Wallis, $p < 2.2e-16$).

To screen for enrichment of variants in a given subgroup, we calculated the average difference in variant frequency. For example, 62% of subjects in subgroup 6 had at least one variant in *SHANK1*. In contrast, the other subgroups had variant frequencies ranging from 29–33% for an average of 31%. Thus, the average difference in variant frequency was 30.6% ($p < 0.0005$ for a gene to have a difference in variant frequency of this magnitude) indicating variants in *SHANK1* were 30.6% more common in subgroup 6 compared to other subgroups. In contrast, the mean frequency of these variants as listed in gnomAD is 0.88%. Plotting the distribution of the difference in variant frequency for each gene revealed that the top 1% of genes had a difference in variant frequency of 21.68% or greater (Fig. 6B).

It is not known what percent difference in variant frequency is clinically significant. Therefore, we considered multiple thresholds based on standard deviations with a goal of being more inclusive and ultimately used dual criteria to ensure that all associated genes associated with just one subgroup. Four standard deviations (20.7% increased or decreased variant frequency from the mean of all subgroups) was the initial cut-off. The second threshold accounted for differences in subgroup distribution in that any genes that were less than three standard deviations from any other subgroup were also excluded. Following these criteria, some groups associated with a small number of genes including six genes with increased variant frequency in group 2, four genes with increased or decreased variant frequency in group 4, and six genes with increased or decreased frequency in group 6. In contrast, group 3 associated with 126 genes with increased variant frequency, 12 genes with increased or decreased variant frequency associated with group 5, and 50 genes with increased or decreased variant frequency associated with group 7. Notably, no genes associated with group 1, characterized by difficulties in all sensory areas (Fig. 6C).

We were particularly interested in the three subgroups associated with a small number of genes. Subgroup 2, characterized by differences in sensory areas four (under-responsiveness) and five (auditory filtering) as well as hyper- and hypo- auditory sensitivity, was associated with variants in *FUT8*, *GGT2*, *KCNQ1*, *NALCN-AS1*, *PHF12*, and *SLC25A23* (Table 3). Five of these genes are protein coding genes, although their functions vary. Subgroup 4, characterized by largely typical performance in all areas for both methods, associated with variants in *CACNA2D3*, *CUX1*, *SMG6*, and *SOBP* (Table 3). Three of these four genes negatively correlated meaning subjects were *less* likely to have variants. Finally, subgroup 6, characterized by definitely different scores in areas two and five (taste/smell sensitivity and auditory filtering) as well as hyper-sensitivity to taste and auditory hypo-sensitivity, associated with variants in *FNDC3B*, *GRIK4*, *LINC01798*, *NPR3*, *SHANK1*, and *SPAG7* (Table 3). Notably, the first four of these genes negatively correlated. Further, *SHANK1* was the most strongly associated with this subgroup and is part of a gene family previously identified as being related to ASD (Sala et al., 2015, Sato et al., 2012). The other four subgroups were associated with larger numbers of genes (Supplementary Table 1).

Discussion

Sensory features can be used to identify unique subgroups in ASD

The primary aim of this study was to identify sensory-based ASD subgroups that shared underlying molecular mechanisms based on correlation with genetic variants. Our results demonstrate that cluster analysis can be used to identify distinct subgroups with unique sensory phenotypes. Specifically, in addition to a subgroup categorized by involvement of many sensory processing differences and a group characterized by minimal differences in sensory areas, we identified five subgroups with unique sensory patterns associated with specific gene variants. Cluster analysis easily discriminated these sensory-based subgroups, proving feasibility for this type of analysis. Further, these results provide the first evidence for a possible molecular link to sensory-based subgroups in ASD. Overall, these data support the idea that sensory features are strong candidates to guide patient classification in treatment trials.

Using sensory features to resolve heterogeneity in ASD has been proposed previously (reviewed in Simpson et al., 2019, Lane et al., 2011, Ausderau et al 2014, Ben-Sasson et al 2008, Lane et al 2014, Lane et al, 2011, Liss et al 2006, Little et al., 2017, Tomcheck et al, 2018). Most studies use parental surveys such as the SSP or the SSP-2. Notably, there is little consensus across studies other than a “very affected” group and a “minimally affected” group. The significant heterogeneity in ASD likely contributes to these challenges. To address the lack of reproducibility in clustering results, we analyzed our data using two separate approaches and saw internal validity. However, future collaborations using the same clustering method on multiple data sets will be helpful in identifying universal sensory subgroups. Further, most previous studies used broad sensory based descriptors such as “sensory adaptive” or “sensory over responsivity” rather than looking specifically at individual sensory domains (Simpson et al., 2019). We chose to focus on hyper- and hypo-sensitivity because these qualities are more easily quantified using direct sensory testing in both patients and mouse models. Future analysis using surveys targeting hyper- and hypo-sensitivity may help to reduce noise in patient datasets.

Genes associated with each sensory subgroup are relevant to ASD

Identifying less heterogeneous subgroups within ASD has been proposed as a potential key to developing targeted treatments (Jeste et al., 2014, Genovese and Butler 2020). Subgrouping using various clinical features has been done (Eaves et al., 1994, Qi et al., 2020) but it has been increasingly recognized that the genetic signature of each subgroup will be important (Betancur 2011, Hu and Bi 2020). While sensory based subgroups in ASD have been proposed (reviewed in Simpson et al., 2019), this study is the first to correlate these sensory based subgroups with genetic variants.

Correlation with genes with variants for each of the seven identified subgroups can be divided into two categories. Subgroups 2, 4 and 6 were each associated with variants in a small number of genes (Table 3) suggesting that a single, genetic mechanism maybe strongly associated with those phenotypes. For example, subgroup 4, characterized by the absence of any hyper- or hypo-sensitivity with only

probable difference in the domain of under-responsiveness / seeks sensation, was associated with increased variants in *SOBP*. *SOBP* is a nuclear zinc finger protein associated with intellectual disability (Birk et al., 2010). In mice, the homologue *Jxc1* was shown to cause deafness due to disrupted cochlear development (Chen et al., 2008). Interestingly, *Jxc1* mRNA expression was also seen in the retina and olfactory epithelium suggesting a link to multiple peripheral sensory systems (Chen et al., 2008). Subgroup 4 was also associated with decreased likelihood of variants in three other genes (Table 3). In turn, it may be equally important to look at what genes are not disrupted. Negative correlation may be of particular value for subgroups with more typical sensory features. Many of the associated genes have been previously associated with neurological disorders. For example, subgroup 2, characterized by auditory hyper- and hypo-sensitivity consistent with atypical scores for auditory filtering and a different from typical children in the area of under-responsiveness / seeks sensation, was most strongly associated with variants in *FUT8* (Table 3). *FUT8* is a fucosyltransferase essential for neurodevelopment that has also been studied in relation to the immune system and cancer (Schneider et al., 2017). Biallelic mutations in *FUT8* were reported in seven individuals with neurological impairment (Ng et al., 2018, Ng et al., 2020). Similarly, subgroup 6, characterized by auditory hypo-sensitivity paired with auditory filtering scores in the different from typically developing children range as well as taste hyper-sensitivity paired with definitely different taste and smell sensitivity scores, was associated with variants in *SHANK1*. The *SHANK1* knockout mouse is a model for ASD (Jiang and Ehlers 2013, Sungur et al., 2018), although we are not aware of any studies evaluating taste in patients or mice with these mutations. Thus, all of these genes are strong candidates to identify a shared molecular mechanism contributing to the specific sensory phenotype for that subgroup.

The other four clusters identified in our analysis were associated with increased variants in 12 or more genes. One interpretation is that there is too much noise in our dataset, and thus we were not able to refine to a single gene associated with that phenotype. Future studies will be needed to functionally evaluate these groups of genes to determine if they mediate a shared pathway. Alternatively, it is possible that a combination of genes contributes to the phenotype through multiple functional pathways creating a genetic signature. Further analysis of the genetic variants in these subgroups will be necessary to elucidate the specific mechanistic connection for those subgroups.

How can patients be both hyper and hypo sensitive?

One of the most surprising findings from our study was the presence of both hyper- and hypo-sensitivity within the same sensory domain in the modalities of touch (47, 12.4%), hearing (156, 41%) and vision (35, 9%). Further, hyper/hypo sensitivity group 2 was characterized by selective auditory hyper- and hypo-sensitivity. Similar seemingly contradictory sensory features have been reported in syndromic ASD, such as the simultaneous presence of both seeking and avoiding sensory features in Phelan-McDermid Syndrome and *SYNGAP1*-ID as measured on the SSP-2 (Lyons-Warren et al., 2020). One possible explanation is that circuit disruptions leading to these phenotypes affect sensory processing rather than pure sensory detection as supported by difficulties in filtering and integration in patients with ASD (DePape et al., 2012, Stevenson et al., 2014). While some sensory processing does occur in the periphery,

the presence of mixed hyper- and hypo-sensitivity phenotypes points to central mechanisms such as global feedback (Noriega 2008) or cortical circuit dysfunction (Gustafsson 1997) that have been proposed to underly sensory abnormalities in ASD. In contrast, in mouse models of non-syndromic ASD, peripheral circuit disruptions have been shown to underly olfactory deficits (Huang et al., 2019). In contrast, in mouse models of syndromic and non-syndromic ASD, peripheral circuit disruptions have been shown to underly olfactory deficits with some models demonstrating both hyper- and hypo-sensitivity depending on the task (Lyons-Warren et al., 2021). As ASD is heterogeneous and likely represents the confluence of many different mechanisms, both central and peripheral circuits maybe involved in different patient subgroups. Alternatively, Carroll et al recently proposed that phenotypic differences could result from a combination of both central and peripheral mechanisms (Carroll et al., 2020). Thus, it will be important to understand the prevalence of the combined hyper- and hypo-sensitivity subgroup and the frequency of this presentation for each of the sensory domains.

Limitations

This study has several limitations. Most importantly, it is a retrospective study which greatly limited the available data. For example, while MSSNG has whole genome sequences for over 10,000 individuals with ASD, Short Sensory Profile Responses with whole genome sequencing was available for only 364 individuals. Thus, the final number of subjects in each cluster was relatively small. Further, additional information about these subjects such as their age or ADOS scores was not available. Prior studies have suggested that sensory processing differences are associated with level of function (Klintwall et al., 2011, Aykan et al., 2020). Future prospective studies will be needed to validate the subgroups identified in this study and to expand our understanding of the associated clinical features for each of these subgroups.

Another limitation stemming from the retrospective nature of this study is that the analysis presented here was done using data obtained from the Short Sensory Profile (SSP). The SSP evaluates sensory behaviors in seven domains evaluating 'sensory processing patterns' and 'self-regulation strategies' which is a more global or holistic view of sensory processing (Dunn 1997, Dunn 2014, Little et al., 2016). While the SSP is well validated and is very useful for clinically understanding the impact of sensory changes in ASD, it does not directly test for hyper- and hypo-sensitivity which are potentially easier to link to circuit mechanisms. We addressed this limitation by using a subset of responses focusing on hyper- and hypo-sensitivity (Table 1). However, our subset of questions has not been validated and a different cut-off value likely would have changed clustering outcomes. We are reassured by the high degree of phenotypic overlap between the two methods. However, future studies should use direct measure of sensory ability and more targeted questionnaires to validate the subgroups identified.

Finally, our findings represent only a correlation between sensory-based subgroups and genes with variants. This study collapsed all variants with a CADD score greater than 15 and thus included potentially non-deleterious variants but did not include frameshifts or deletions. Further, no distinction was made between heterozygous or homozygous variants. This approach was taken in order to identify broad genetic patterns associated with the identified subgroups. As variant frequency was compared

between subgroups, even if the individual variants themselves are not predicted to be deleterious, the difference between subgroups is still relevant. However, functional studies will be necessary to test if the identified genes are causative of the sensory changes,

Conclusions

Using Short Sensory Profile (SSP) data from 377 individuals with non-syndromic ASD Spectrum Disorder (ASD) obtained through the MSSNG database, we identified seven sensory-based subgroups. These subgroups exhibited high impairment across multiple domains, absence of hyper- or hypo-sensitivity, or discreet combinations of sensory features. Further, each of these subgroups associated with an increase in variants within a relatively small group of genes. We conclude that sensory features can be used to identify subgroups within ASD with shared patterns of genetic variants. These results represent the first step towards identifying mechanistically linked subgroups important for targeted drug development and evaluating efficacy in therapeutic trials.

Abbreviations

ASD = Autism Spectrum Disorder

SSP = Short Sensory Profile

CADD = Combined Annotation Dependent Depletion

MSSNG = name of the Autism Speaks database

ADOS = Autism Diagnostic Observation Schedule

Declarations

Ethics approval and consent to participate

The Institutional Review Board for human studies at Baylor College of Medicine reviewed this study and granted an exemption based on the absence of individually identifying information in the database.

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed during the current study are available in the MSSNG repository, <https://research.mss.ng>. Restrictions apply to the availability of these data, which were used under

license for the current study, and so are not publicly available. Data are however available from Autism Speaks following appropriate request for access.

Competing interests

The authors declare that they have no competing interests

Funding

This study was funded by NINDS (1K12NS098482) award to A.L-W. The project described was also supported by IDDRG Grant Number 1P50HD103555-01 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development, RP170387 from the Cancer Prevention and Research Institute of Texas, and NINDS (R01HL147064-01A1) which support Y-W.W.

Authors' contributions

Both authors contributed equally to this manuscript including experimental design, data analysis and interpretation of results. A.L-W. conceived the research question, provided clinical expertise including identifying hyper- and hypo-sensitivity related questions and critical analysis of identified genes, and wrote the first draft of the manuscript. Y-W.W. extracted all data from the database, provided statistical expertise including performing the cluster analysis and contributed to writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to acknowledge the resources of MSSNG (www.mss.ng), Autism Speaks and The Centre for Applied Genomics at The Hospital for Sick Children, Toronto, Canada. We also thank the participating families for their time and contributions to this database, as well as the generosity of the donors who supported this program.

Authors' information (optional)

Not applicable

References

1. Ausderau KK, Furlong M, Sideris J, Bulluck J, Little LM, Watson LR, Boyd BA, Belger A, Dickie VA, Baranek GT. Sensory subtypes in children with autism spectrum disorder: latent profile transition analysis using a national survey of sensory features. *J Child Psychol Psychiatry*. 2014

- Aug;55(8):935-44. doi: 10.1111/jcpp.12219. Epub 2014 Mar 27. PMID: 25039572; PMCID: PMC4159682.
2. Ashburner J, Ziviani J, Rodger S. Sensory processing and classroom emotional, behavioral, and educational outcomes in children with autism spectrum disorder. *Am J Occup Ther.* 2008 Sep-Oct;62(5):564-73. doi: 10.5014/ajot.62.5.564. PMID: 18826017.
 3. Aykan S, Gürses E, Tokgöz-Yılmaz S, Kalaycıoğlu C. Auditory Processing Differences Correlate With Autistic Traits in Males. *Front Hum Neurosci.* 2020 Oct 7;14:584704. doi: 10.3389/fnhum.2020.584704. PMID: 33192419; PMCID: PMC7588834.
 4. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. *Psychol Med.* 2015 Feb;45(3):601-13. doi: 10.1017/S003329171400172X. Epub 2014 Aug 11. PMID: 25108395.
 5. Ben-Sasson A, Cermak SA, Orsmond GI, Tager-Flusberg H, Kadlec MB, Carter AS. Sensory clusters of toddlers with autism spectrum disorders: differences in affective symptoms. *J Child Psychol Psychiatry.* 2008 Aug;49(8):817-25. doi: 10.1111/j.1469-7610.2008.01899.x. Epub 2008 May 21. PMID: 18498344.
 6. Betancur C. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. *Brain Res.* 2011;1380: 42–77.
 7. Birk E, Har-Zahav A, Manzini CM, Pasmanik-Chor M, Kornreich L, Walsh CA, Noben-Trauth K, Albin A, Simon AJ, Colleaux L, Morad Y, Rainshtein L, Tischfield DJ, Wang P, Magal N, Maya I, Shoshani N, Rechavi G, Gothelf D, Maydan G, Shohat M, Basel-Vanagaite L. SOBP is mutated in syndromic and nonsyndromic intellectual disability and is highly expressed in the brain limbic system. *Am J Hum Genet.* 2010 Nov 12;87(5):694-700. doi: 10.1016/j.ajhg.2010.10.005. Epub 2010 Oct 28. PMID: 21035105; PMCID: PMC2978971.
 8. Bruni M, Cameron D, Dua S, Noy S. Reported sensory processing of children with Down syndrome. *Phys Occup Ther Pediatr.* 2010 Nov;30(4):280-93. doi: 10.3109/01942638.2010.486962. PMID: 20735195.
 9. Carr JL, Agnihotri S, Keightley M. Sensory processing and adaptive behavior deficits of children across the fetal alcohol spectrum disorder continuum. *Alcohol Clin Exp Res.* 2010 Jun;34(6):1022-32. doi: 10.1111/j.1530-0277.2010.01177.x. Epub 2010 Apr 5. PMID: 20374212.
 10. Carroll L, Braeutigam S, Dawes JM, Krsnik Z, Kostovic I, Coutinho E, Dewing JM, Horton CA, Gomez-Nicola D, Menassa DA. Autism Spectrum Disorders: Multiple Routes to, and Multiple Consequences of, Abnormal Synaptic Function and Connectivity. *Neuroscientist.* 2020 May 22:1073858420921378. doi: 10.1177/1073858420921378. Epub ahead of print. PMID: 32441222.
 11. Chen YH, Rodgers J, McConachie H. Restricted and repetitive behaviours, sensory processing and cognitive style in children with autism spectrum disorders. *J Autism Dev Disord.* 2009 Apr;39(4):635-42. doi: 10.1007/s10803-008-0663-6. Epub 2008 Nov 18. PMID: 19015969.
 12. Chen Z, Montcouquiol M, Calderon R, Jenkins NA, Copeland NG, Kelley MW, Noben-Trauth K. Jxc1/Sobp, encoding a nuclear zinc finger protein, is critical for cochlear growth, cell fate, and

- patterning of the organ of corti. *J Neurosci*. 2008 Jun 25;28(26):6633-41. doi: 10.1523/JNEUROSCI.1280-08.2008. PMID: 18579736; PMCID: PMC2556235.
13. DePape AM, Hall GB, Tillmann B, Trainor LJ. Auditory processing in high-functioning adolescents with Autism Spectrum Disorder. *PLoS One*. 2012;7(9):e44084. doi: 10.1371/journal.pone.0044084. Epub 2012 Sep 12. PMID: 22984462; PMCID: PMC3440400.
 14. Dunn W. The impact of sensory processing abilities on the daily lives of young children and their families: A conceptual model. *Infants Young Child*. 1997;9:23–35.
 15. Dunn, W. (2014) *Sensory Profile-2*. Pearson Publishing, San Antonio, TX.
 16. Eaves LC, Ho HH, Eaves DM. (1994) Subtypes of autism by cluster analysis. *J Autism Dev Disord*; 24(1):3-22.
 17. Genovese A, & Butler MG (2020). Clinical Assessment, Genetics, and Treatment Approaches in Autism Spectrum Disorder (ASD). *International journal of molecular sciences*, 21(13), 4726.
 18. Gustafsson L. Inadequate cortical feature maps: a neural circuit theory of autism. *Biol Psychiatry*. 1997 Dec 15;42(12):1138-47. doi: 10.1016/s0006-3223(97)00141-8. PMID: 9426884.
 19. Heald, M., Adams, D., and Oliver, C. (2020) Profiles of atypical sensory processing in Angelman, Cornelia de Lange and Fragile X syndromes. *Journal of Intellectual Disability Research*, 64: 117–130.
 20. Hu VW, Bi C. Phenotypic Subtyping and Re-analyses of Existing Transcriptomic Data from Autistic Proband Families Reveal Differentially Expressed and ASD Trait-Associated Genes. *Front Neurol*. 2020 Nov 12;11:578972. doi: 10.3389/fneur.2020.578972. PMID: 33281715; PMCID: PMC7689346.
 21. Huang TN, Yen TL, Qiu LR, Chuang HC, Lerch JP, Hsueh YP. Haploinsufficiency of autism causative gene *Tbr1* impairs olfactory discrimination and neuronal activation of the olfactory
 22. system in mice. *Mol Autism*. 2019 Feb 11;10:5. doi: 10.1186/s13229-019-0257-5. PMID: 30792833; PMCID: PMC6371489
 23. Ingram DG, Takahashi TN, Miles JH. Defining autism subgroups: a taxometric solution. *J Autism Dev Disord*. 2008 May;38(5):950-60. doi: 10.1007/s10803-007-0469-y. PMID: 17985224.
 24. Jacob S, Wolff JJ, Steinbach MS, Doyle CB, Kumar V, Elison JT. Neurodevelopmental heterogeneity and computational approaches for understanding autism. *Transl Psychiatry*. 2019 Feb 4;9(1):63. doi: 10.1038/s41398-019-0390-0. PMID: 30718453; PMCID: PMC6362076.
 25. John AE, Mervis CB. Sensory modulation impairments in children with Williams syndrome. *Am J Med Genet C Semin Med Genet*. 2010 May 15;154C(2):266-76. doi: 10.1002/ajmg.c.30260. PMID: 20425786; PMCID: PMC2997471.
 26. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol*. 2014;10: 74–81.
 27. Kogan MD , Strickland BB , Blumberg SJ , Singh GK , Perrin JM , van Dyck PC (2008). A national profile of the health care experiences and family impact of autism spectrum disorder among children

- in the United States, 2005-2006 . Pediatrics 122 , e1149 -e1158 .10.1542/peds.2008-1057 19047216
28. Klintwall L, Holm A, Eriksson M, Carlsson LH, Olsson MB, Hedvall A, Gillberg C, Fernell E. Sensory abnormalities in autism. A brief report. Res Dev Disabil. 2011 Mar-Apr;32(2):795-800. doi: 10.1016/j.ridd.2010.10.021. Epub 2010 Dec 15. PMID: 21111574.
 29. Lane, A., Dennis, S. J., & Geraghty, M. E. (2011). Brief report: Further evidence of sensory subtypes in autism. Journal of Autism and Developmental Disorders, 41(6), 826–831. <https://doi.org/10.1007/s10803-010-1103-y>.
 30. Lane AE, Molloy CA, Bishop SL. Classification of children with autism spectrum disorder by sensory subtype: a case for sensory-based phenotypes. Autism Res. 2014 Jun;7(3):322-33. doi: 10.1002/aur.1368. Epub 2014 Mar 17. PMID: 24639147.
 31. Little LM, Dean E, Tomchek SD, Dunn W. Classifying sensory profiles of children in the general population. Child Care Health Dev. 2017 Jan;43(1):81-88. doi: 10.1111/cch.12391. Epub 2016 Aug 21. PMID: 27545764.
 32. Liss M, Saulnier C, Fein D, Kinsbourne M. Sensory and attention abnormalities in autistic spectrum disorders. Autism. 2006 Mar;10(2):155-72. doi: 10.1177/1362361306062021. PMID: 16613865.
 33. Lyall, K., Croen, L., Daniels, J., Fallin, M. D., Ladd-Acosta, C., Lee, B. K., Park, B. Y., Snyder, N. W., Schendel, D., Volk, H., Windham, G. C., & Newschaffer, C. (2017). The Changing Epidemiology of Autism Spectrum Disorders. Annual review of public health, 38, 81–102. <https://doi.org/10.1146/annurev-publhealth-031816-044318>
 34. Lyons-Warren AM, McCormack MC, Holder Jr JL. Brief Report: Sensory processing phenotypes in Phelan-McDermid Syndrome and SYNGAP1-related Intellectual Disability medRxiv 2020.11.30.20241315; doi: <https://doi.org/10.1101/2020.11.30.20241315>
 35. Lyons-Warren AM, Herman I, Hunt PK, Arenkiel B. Olfactory deficits in neurodevelopmental disorders: from mouse to human. 2021 Neuroscience and Biobehavioral Reviews *in press*
 36. Mangeot SD, Miller LJ, McIntosh DN, McGrath-Clarke J, Simon J, Hagerman RJ, Goldson E. Sensory modulation dysfunction in children with attention-deficit-hyperactivity disorder. Dev Med Child Neurol. 2001 Jun;43(6):399-406. doi: 10.1017/s0012162201000743. PMID: 11409829.
 37. A. Masi, M.M. DeMayo, N. Glozier, A.J. Guastella. An overview of autism spectrum disorder, heterogeneity and treatment options Neurosci. Bull. (2017)
 38. McIntosh, D.N., Miller, L.J., & Shyu, V. (1999). Development and validation of the Short Sensory Profile. In W. Dunn (Ed.), Sensory Profile manual (pp. 59–73). San Antonio, TX: Psychological Corporation.
 39. Ng BG, Dastsooz H, Silawi M, Habibzadeh P, Jahan SB, Fard MAF, Halliday BJ, Raymond K, Ruzhnikov MRZ, Tabatabaei Z, Taghipour-Sheshdeh A, Brimble E, Robertson SP, Faghihi MA, Freeze HH. Expanding the molecular and clinical phenotypes of FUT8-CDG. J Inherit Metab Dis. 2020 Jul;43(4):871-879. doi: 10.1002/jimd.12221. Epub 2020 Feb 23. PMID: 32049367; PMCID: PMC7359201.

40. Ng BG, Xu G, Chandy N, Steyermark J, Shinde DN, Radtke K, Raymond K, Lebrilla CB, AlAsmari A, Suchy SF, Powis Z, Fageih EA, Berry SA, Kronn DF, Freeze HH. Biallelic Mutations in FUT8 Cause a Congenital Disorder of Glycosylation with Defective Fucosylation. *Am J Hum Genet.* 2018 Jan 4;102(1):188-195. doi: 10.1016/j.ajhg.2017.12.009. PMID: 29304374; PMCID: PMC5777984.
41. Noriega G. A neural model for compensation of sensory abnormalities in autism through feedback from a measure of global perception. *IEEE Trans Neural Netw.* 2008 Aug;19(8):1402-14. doi: 10.1109/TNN.2008.2000447. PMID: 18701370.
42. Pandina G, Ring RH, Bangerter A, Ness S. Current Approaches to the Pharmacologic Treatment of Core Symptoms Across the Lifespan of Autism Spectrum Disorder. *Psychiatr Clin North Am.* 2020 Dec;43(4):629-645. doi: 10.1016/j.psc.2020.08.003. PMID: 33126999.
43. Posar A, Visconti P. Sensory abnormalities in children with autism spectrum disorder. *J Pediatr (Rio J).* 2018 Jul-Aug;94(4):342-350. doi: 10.1016/j.jped.2017.08.008. Epub 2017 Nov 4. PMID: 29112858.
44. Rogers SJ, Hepburn S, Wehner E. Parent reports of sensory symptoms in toddlers with autism and those with other developmental disorders. *J Autism Dev Disord.* 2003 Dec;33(6):631-42. doi: 10.1023/b:jadd.0000006000.38991.a7. PMID: 14714932.
45. Sala C, Vicidomini C, Bigi I, Mossa A, Verpelli C. Shank synaptic scaffold proteins: keys to understanding the pathogenesis of autism and other synaptic disorders. *J Neurochem.* 2015 Dec;135(5):849-58. doi: 10.1111/jnc.13232. Epub 2015 Sep 3. PMID: 26338675.
46. Sato D, Lionel AC, Leblond CS, Prasad A, Pinto D, Walker S, O'Connor I, Russell C, Drmic IE, Hamdan FF, Michaud JL, Endris V, Roeth R, Delorme R, Huguet G, Leboyer M, Rastam M, Gillberg C, Lathrop M, Stavropoulos DJ, Anagnostou E, Weksberg R, Fombonne E, Zwaigenbaum L, Fernandez BA, Roberts W, Rappold GA, Marshall CR, Bourgeron T, Szatmari P, Scherer SW. SHANK1 Deletions in Males with Autism Spectrum Disorder. *Am J Hum Genet.* 2012 May 4;90(5):879-87. doi: 10.1016/j.ajhg.2012.03.017. Epub 2012 Apr 12. PMID: 22503632; PMCID: PMC3376495.
47. Schneider M, Al-Shareffi E, Haltiwanger RS. Biological functions of fucose in mammals. *Glycobiology.* 2017 Jul 1;27(7):601-618. doi: 10.1093/glycob/cwx034. PMID: 28430973; PMCID: PMC5458543.
48. Simpson K, Adams D, Alston-Knox C, Heussler HS, Keen D. Exploring the Sensory Profiles of Children on the Autism Spectrum Using the Short Sensory Profile-2 (SSP-2). *J Autism Dev Disord.* 2019;49(5):2069-2079. doi:10.1007/s10803-019-03889-2
49. Stevenson RA, Siemann JK, Schneider BC, Eberly HE, Woynaroski TG, Camarata SM, Wallace MT. Multisensory temporal integration in autism spectrum disorders. *J Neurosci.* 2014 Jan 15;34(3):691-7. doi: 10.1523/JNEUROSCI.3615-13.2014. PMID: 24431427; PMCID: PMC3891950.
50. Tomchek SD, Dunn W. Sensory processing in children with and without autism: a comparative study using the short sensory profile. *Am J Occup Ther.* 2007 Mar-Apr;61(2):190-200. doi: 10.5014/ajot.61.2.190. PMID: 17436841.

51. Tomchek SD, Little LM, Myers J, Dunn W. Sensory Subtypes in Preschool Aged Children with Autism Spectrum Disorder. *J Autism Dev Disord*. 2018 Jun;48(6):2139-2147. doi: 10.1007/s10803-018-3468-2. PMID: 29417432.
52. Qi S, Morris R, Turner JA, Fu Z, Jiang R, Deramus TP, Zhi D, Calhoun VD, Sui J. Common and unique multimodal covarying patterns in autism spectrum disorder subtypes. *Mol Autism*. 2020 Nov 18;11(1):90. doi: 10.1186/s13229-020-00397-4. PMID: 33208189; PMCID: PMC7673101.
53. Wolfers T, Floris DL, Dinga R, van Rooij D, Isakoglou C, Kia SM, Zabihi M, Llera A, Chowdanayaka R, Kumar VJ, Peng H, Laidi C, Batalle D, Dimitrova R, Charman T, Loth E, Lai MC, Jones E, Baumeister S, Moessnang C, Banaschewski T, Ecker C, Dumas G, O'Muircheartaigh J, Murphy D, Buitelaar JK, Marquand AF, Beckmann CF. From pattern classification to stratification: towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neurosci Biobehav Rev*. 2019 Sep;104:240-254. doi: 10.1016/j.neubiorev.2019.07.010. Epub 2019 Jul 19. PMID: 31330196.

Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

Figures

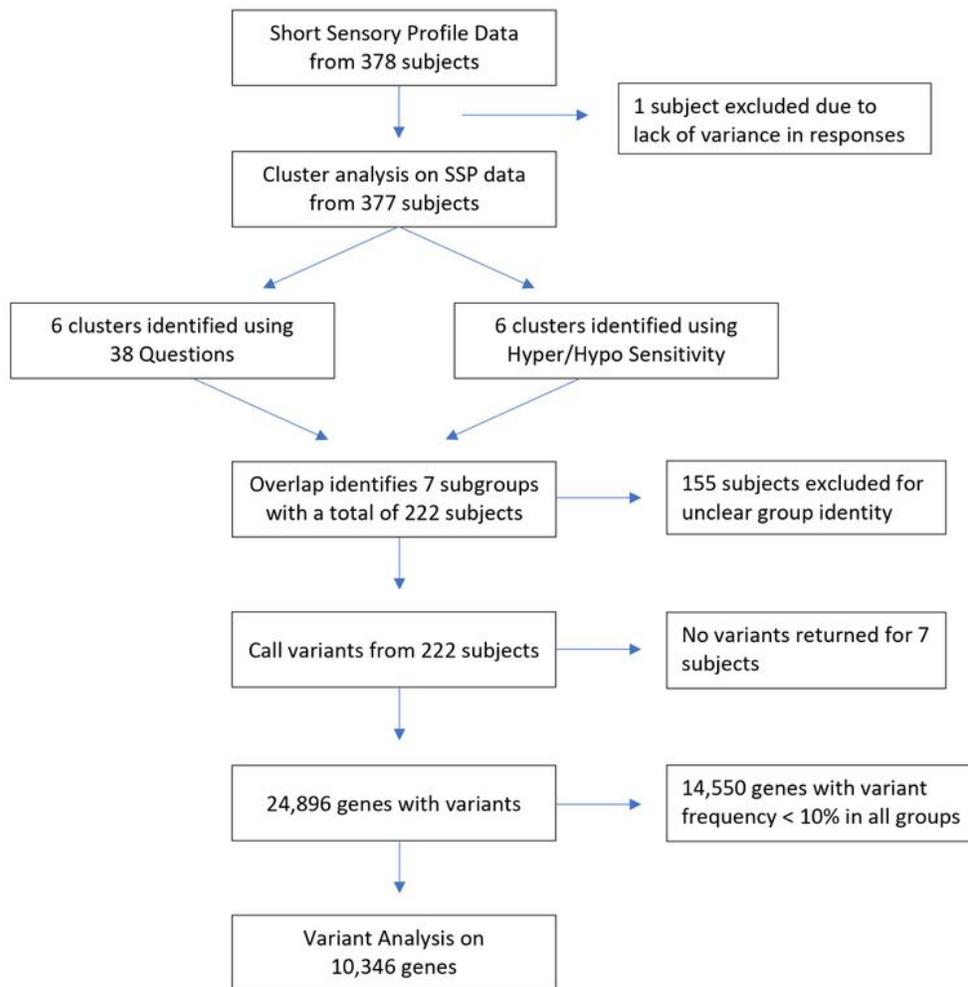


Figure 1

Method flowchart Short sensory profile (SSP) data for 378 patients was obtained from MSSNG. After evaluating the quality of the data, 377 patients were included in the cluster analyses which both revealed six clusters. Overlap from these two methods yielded seven subgroups comprised of 222 subjects, of which 215 were included in the variant analysis.

Figure 2

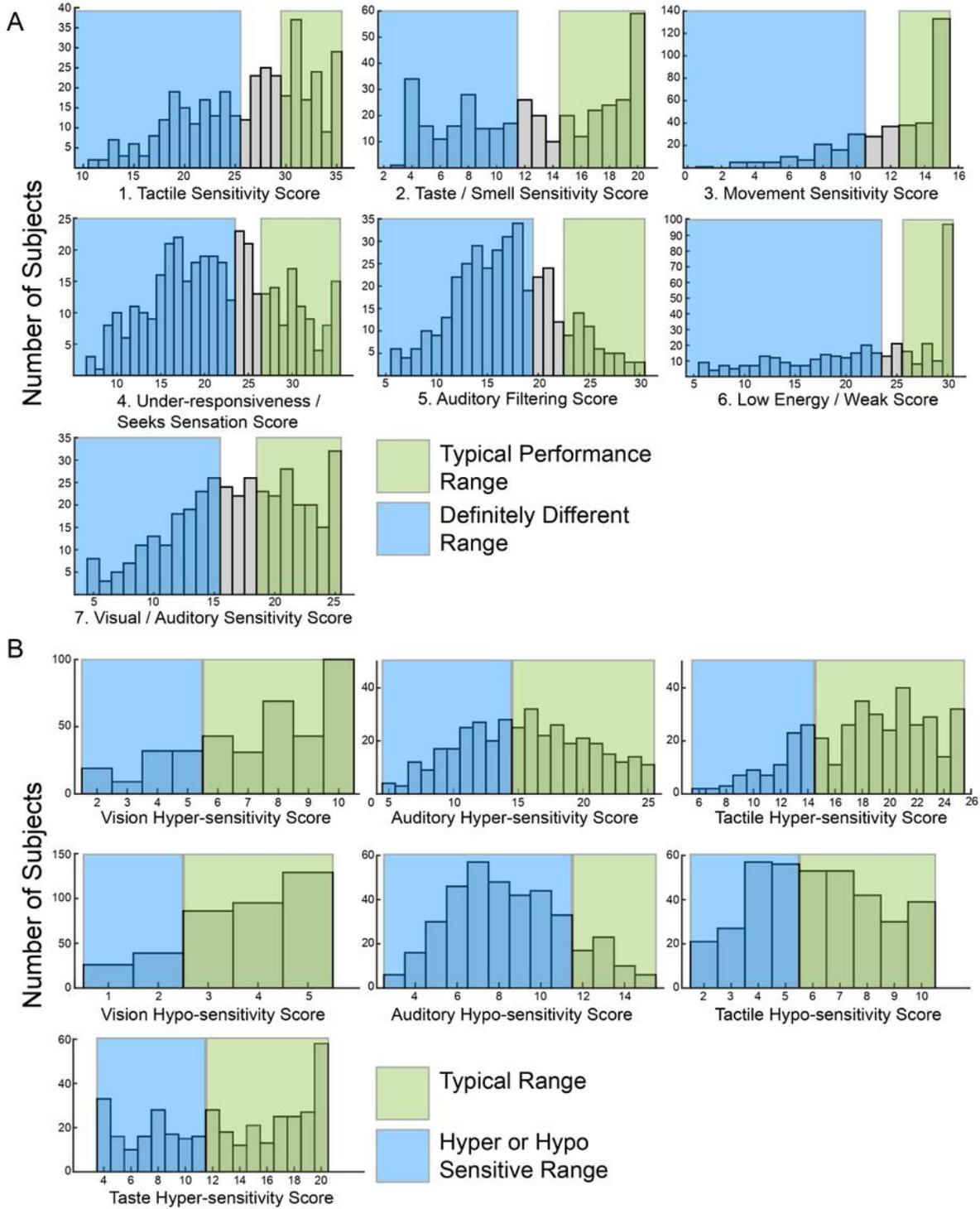


Figure 2

Broad distribution of Short Sensory Profile responses Histograms showing the number of subjects with each raw score for (A) each of the seven sensory areas defined by the Short Sensory Profile and (B) hyper- or hypo-sensitivity for each sensory domain of vision, hearing and touch as well as hyper-sensitivity to taste. No questions in the Short Sensory Profile evaluate hypo-sensitivity to taste or olfactory sensitivities.

Figure 3

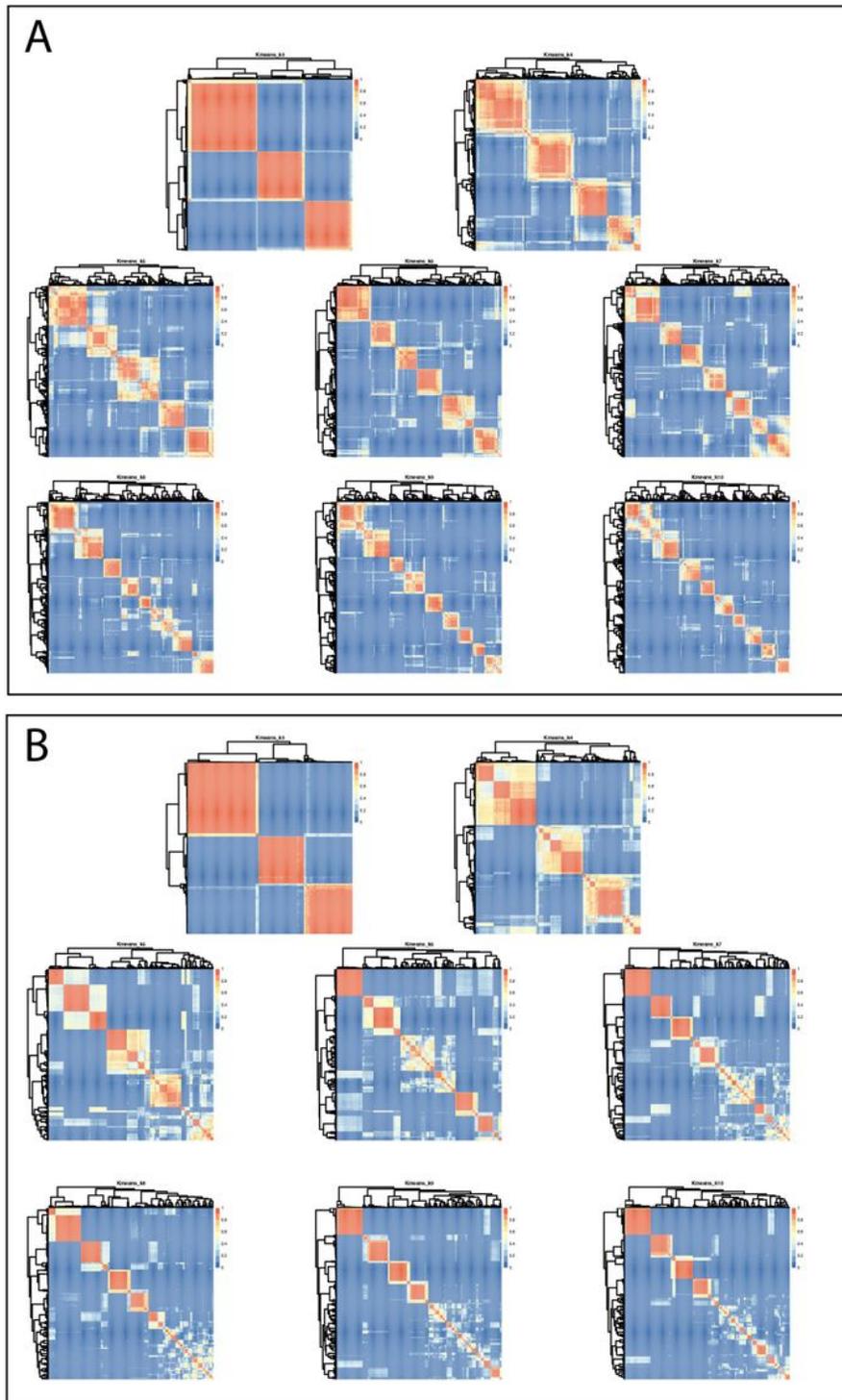


Figure 3

Iterative cluster analysis reveals optimal repeat clustering when using k means = 6 Heatmaps showing overlap for clusters generated using K = 3-10 based on input from (A) all 38 questions in the SSP and (B) subset analysis focusing on Hyper/Hypo-sensitivity. Red indicates 100% consensus and blue indicates no consensus. Lighter shading indicates variation across iterations and thus less consensus.

Figure 4

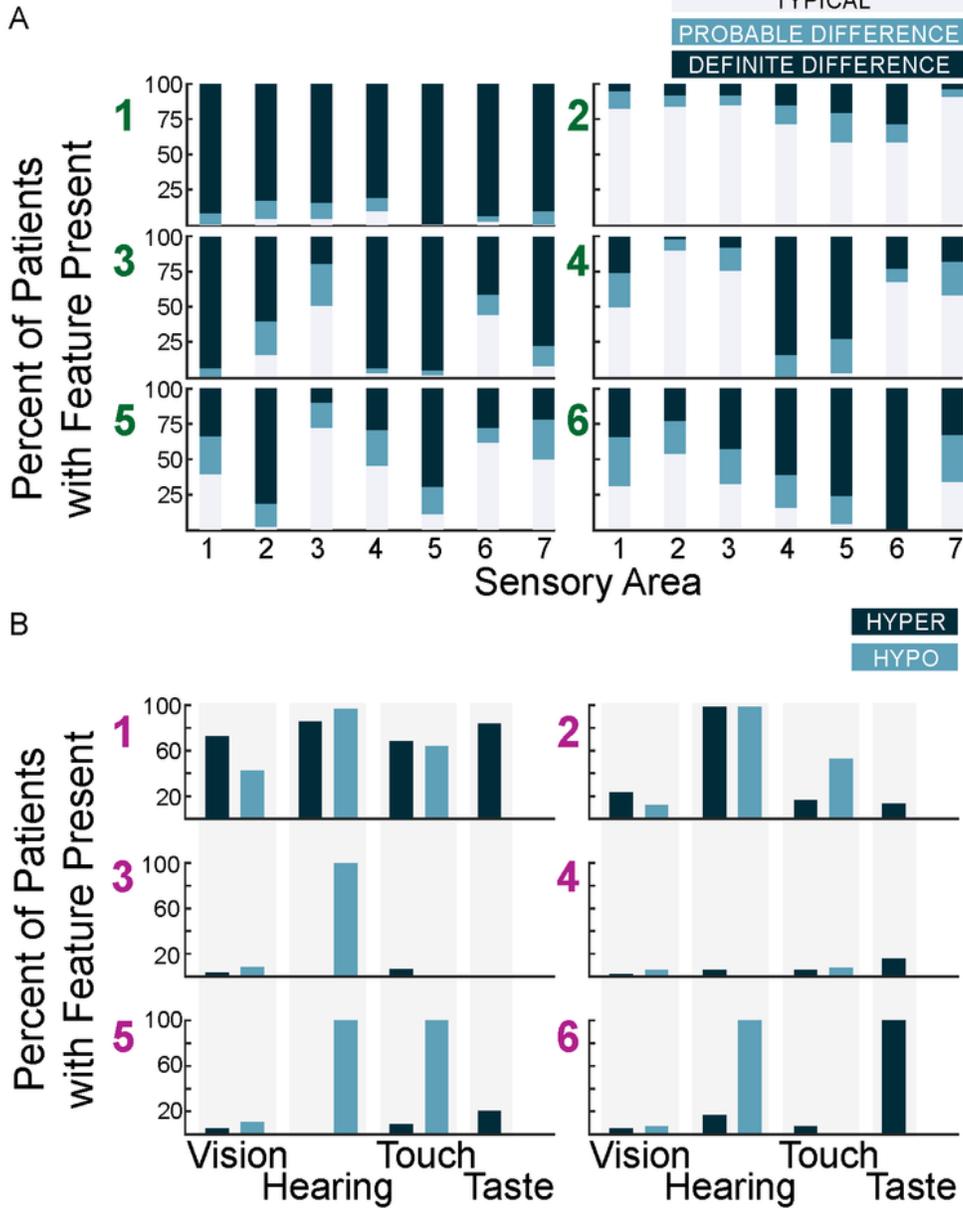


Figure 4

Subgroup features Histograms showing the number of subjects with each sensory feature for six subgroups identified in (A) 38 question analysis and (B) Hyper/Hypo sensitivity analysis.

Figure 5

38 Question Sensory Areas

		1	2	3	4	5	6
Hyper/Hypo Sensitivity	1	54	0	1	15	5	17
	2	13	2	17	31	5	6
	3	1	23	16	13	6	1
	4	0	38	2	4	7	0
	5	4	6	28	9	10	1
	6	0	3	0	7	30	2

Figure 5

Consensus clusters Grid showing the number of subjects in each cluster combination. The six subgroups identified through the 38 Question analysis are plotted on the horizontal access and the six subgroups identified through the hyper/hypo sensitivity analysis are plotted on the vertical access. Areas of significant overlap used to determine the final consensus subgroups are shaded in yellow.

Figure 6

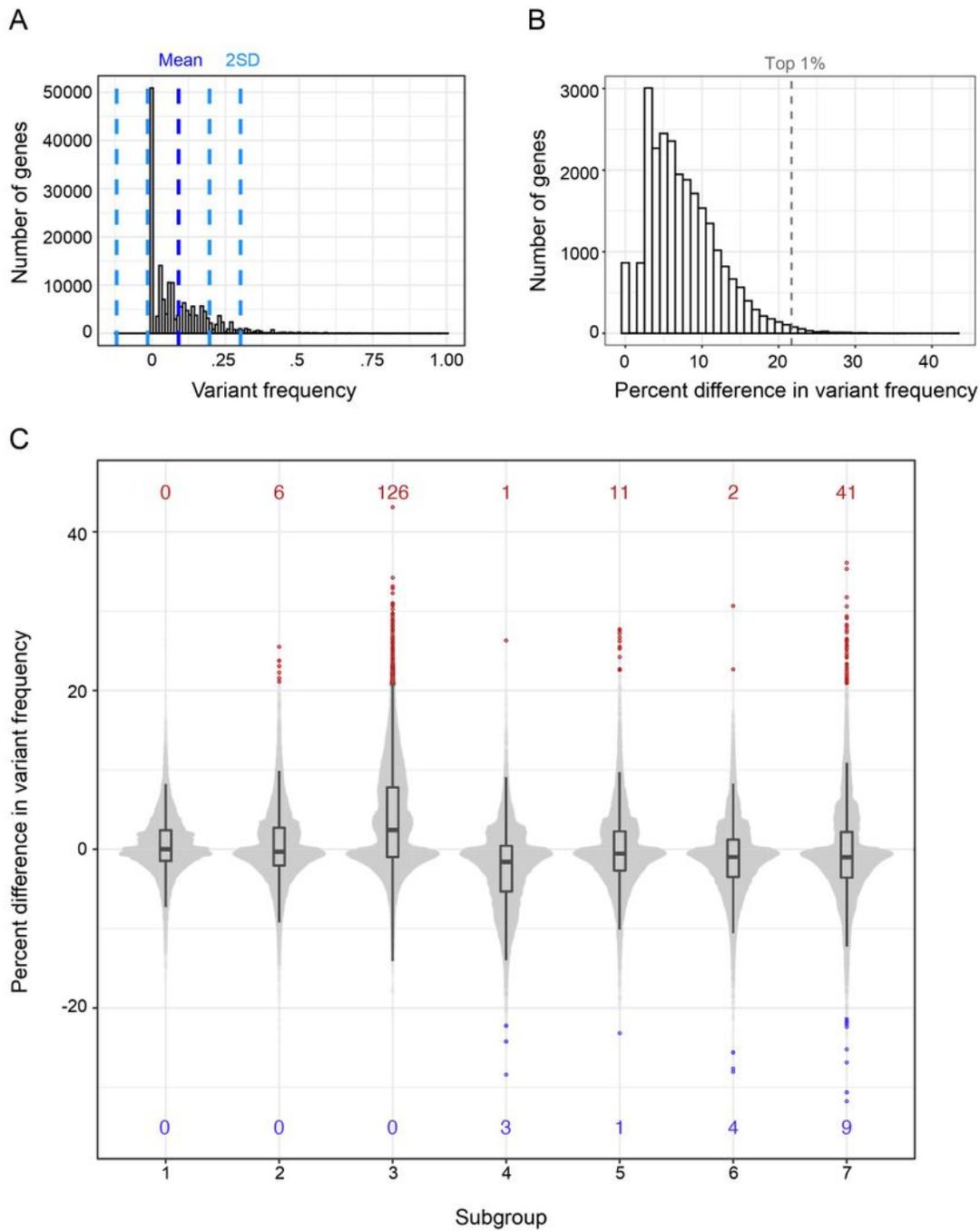


Figure 6

Average difference in variant frequency (A) Histogram showing the distribution and standard deviation (blue lines) of variant frequency for all genes for all subgroups. (B) Histogram showing the distribution of the difference in variant frequency for each gene. Difference in variant frequency calculated as the maximum frequency minus the mean of the frequency in the other 6 subgroups. Only 1% of genes had a difference in variant frequency greater than 21% (gray dashed line). (C) For each subgroup 1-7, the

percent difference of each gene relative to the mean of the variant frequency in the other six subgroups. Note that subgroup three is skewed slightly higher. The number of genes meeting the criteria for significant association due to increased variant frequency is listed at the top and the number of genes meeting criteria for significant association due to decreased variant frequency is shown at the bottom of the chart.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [LyonsWarrenetalSupplementary.pdf](#)
- [LyonsWarrenetalTables.pdf](#)