



Article

<https://doi.org/10.1038/s44220-023-00056-6>

Prognostic early snapshot stratification of autism based on adaptive functioning

Received: 3 August 2022

Accepted: 23 March 2023

Published online: 4 May 2023

Check for updates

Veronica Mandelli^{1,2}, Isotta Landi², Elena Maria Busuoli^{1,2}, Eric Courchesne³, Karen Pierce³ & Michael V. Lombardo^{1,2}✉

A major goal of precision medicine is to predict prognosis based on individualized information at the earliest possible points in development. Using early snapshots of adaptive functioning and unsupervised data-driven discovery methods, we uncover highly stable early autism subtypes that yield information relevant to later prognosis. Data from the National Institute of Mental Health Data Archive (NDA) ($n = 1,098$) was used to uncover three early subtypes (<72 months) that generalize with 96% accuracy. Outcome data from NDA ($n = 2,561$; mean age, 13 years) also reproducibly clusters into three subtypes with 99% generalization accuracy. Early snapshot subtypes predict developmental trajectories in non-verbal cognitive, language and motor domains and are predictive of membership in different adaptive functioning outcome subtypes. Robust and prognosis-relevant subtyping of autism based on early snapshots of adaptive functioning may aid future research work via prediction of these subtypes with our reproducible stratification model.

Autism is a clinical consensus label based on early difficulties in the domains of social-communication and restricted repetitive behaviors¹. Although the label of autism helps maximize consensus and reliability amongst clinical diagnostic judgments based on behavior, it may be less useful for many other important clinical and translational research objectives, such as honing in on differential biology, outcomes and treatment responses^{2,3}. With a view towards applying precision medicine^{4,5} to the field of autism, we should aim to move closer to labels that have higher utility for these types of objectives. As a step in this direction, a recent *Lancet* commission has proposed a call to action for more precise labels, such as ‘profound autism’, to identify the most profoundly affected individuals, who require extra services and support⁶. Thus, a first-level distinction in the autism population should be made that separates out autistic individuals characterized by ‘disability’ versus ‘difference’ in developmental outcomes.

In this Article we characterize autism subtypes within the domain of adaptive functioning. There are several important reasons for why subtyping based on adaptive functioning may be important. First, adaptive functioning is a pivotal domain with high ecological validity and

predictive power for explaining later life outcomes (for example, later independent living, educational attainment and employment)^{7–10}, and is also associated with services and unmet needs¹¹. Variability in adaptive functioning in the autism population is considerable, ranging from very profoundly affected individuals to those within typically developing (TD) norms^{12–14}. Thus, distinctions within adaptive functioning are clearly needed to separate clinically meaningful and outcome-sensitive heterogeneity in autism. Second, adaptive functioning can be quickly measured throughout the lifespan with standardized clinical assessment tools such as the Vineland Adaptive Behavior Scales (VABS)^{15,16}. This tool offers the ability to carry out quick, repeated and affordable assessments of an individual throughout their lifespan and can be deployed in multiple settings. Other advantages of the VABS are age-normalized scores and the ability to interpret minimal clinically significant change¹². Third, changing adaptive functioning has become one of the key objectives for intervention research¹². Thus, stratification models that can provide useful subtype labels may be important for facilitating advances in personalizing interventions. Fourth, adaptive functioning can be disentangled from an individual’s level

¹Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy. ²Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy. ³Autism Center of Excellence, Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. ✉e-mail: michael.lombardo@it.it

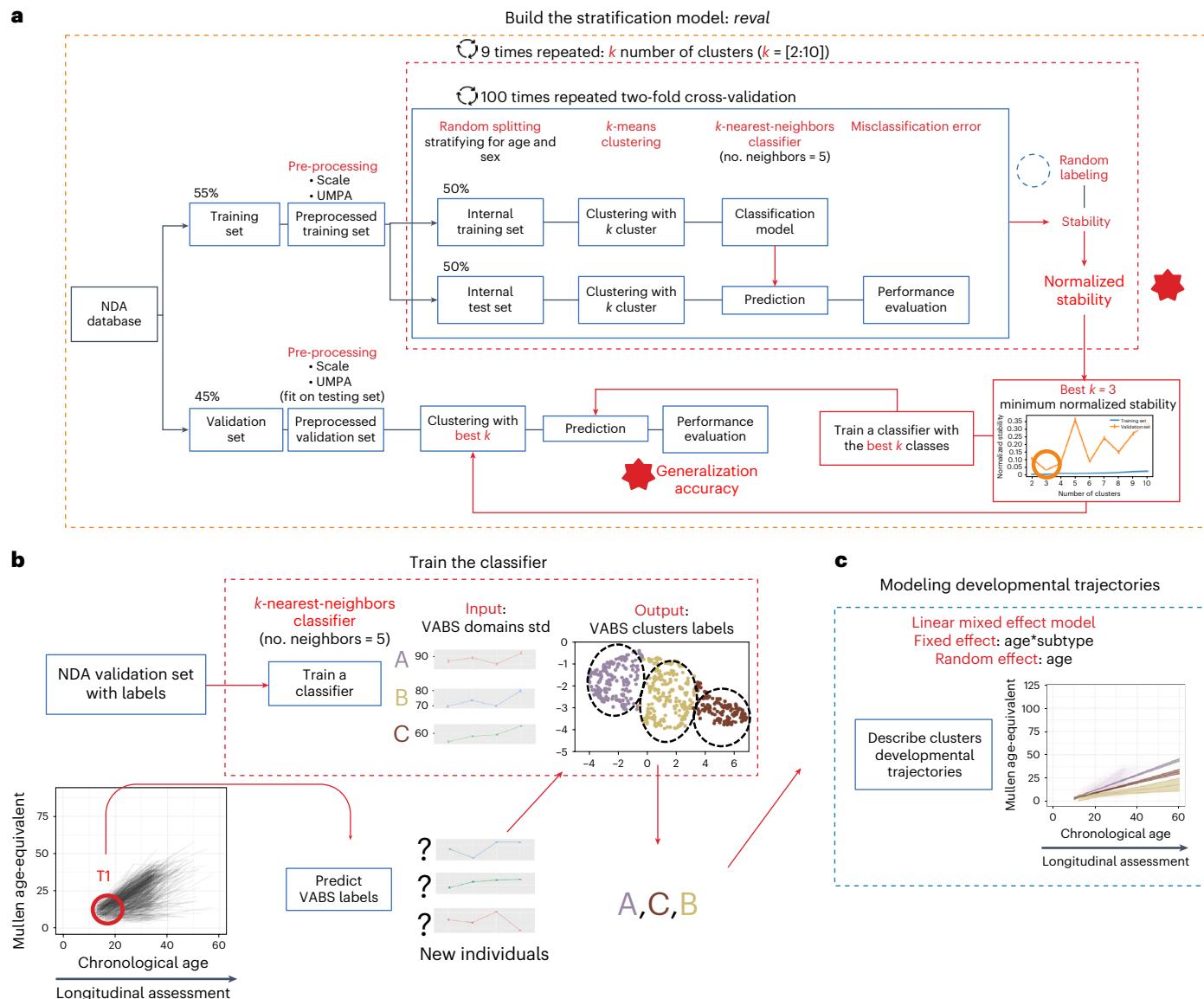


Fig. 1 | Schematic outlining the data analysis plan. **a–c**, The dataset used for initial data-driven discovery and validation through stability-based clustering is the NDA dataset (<https://nda.nih.gov>). NDA data are fed through the pipeline shown in **a**, which illustrates the *reval* algorithm pipeline. Once a robust and

highly generalizable classifier is built from the NDA dataset, we apply that snapshot prediction model to the UCSD ACE longitudinal dataset. **b,c**, The analysis pipeline for applying the subtype prediction model (**b**) and then modeling developmental trajectories (**c**).

of intellectual functioning¹³. Although there is a relationship between adaptive functioning and large differences in intellectual functioning (for example, contrasting individuals with IQ < 70 versus IQ ≥ 70), amongst those with IQ ≥ 70, the variability in adaptive functioning is still considerable, ranging from highly affected individuals to individuals within the normative range for their age¹². As individuals become older, the potential discrepancy between IQ and adaptive functioning can widen^{13,17}. Thus, subtyping based on adaptive functioning may be able to capture real-world, clinically meaningful variability between individuals, even within the range of intact intellectual functioning.

Previous longitudinal work has attempted to identify subtypes based on differential trajectories of the VABS over the first two decades of life^{14,17–19}. Although this work is immensely important for describing how different types of individual develop in terms of adaptive functioning, it cannot be utilized for early stratification in important clinical contexts such as intervention, because rich longitudinal information is not known about participants in such studies. A gap is present whereby there is a key need to be able to stratify individuals in

developmental-and outcome-sensitive ways based on single snapshots of information at early stages in development. If there was a tool that allowed for highly robust and reproducible subtyping based on early single snapshots of adaptive functioning, this would potentially fill this gap and lead to further insights about how to predict treatment response and outcome in such individuals. Given this potential, we aimed to develop a stratification model that allows for data-driven discovery of robust and reproducible VABS subtypes based on a single snapshot of early VABS scores. We then show how subtypes are useful for predicting subsequent adaptive functioning outcome subtypes and developmental trajectories in non-verbal cognitive, language and motor domains.

Results

Identification of adaptive functioning subtypes

In our first analysis, we sought to test whether unsupervised data-driven stratifications could be made in autism based on early snapshots (<72 months) of adaptive functioning using the VABS. Our analysis approach

applies stability-based relative clustering validation to identify data-driven clusters that are stable and reproducible in independent datasets (Fig. 1). We found that a three-cluster solution is unequivocally the best cluster solution that minimizes normalized cluster stability. This model produces very high generalization accuracy (96%) in independent data (for example, the held-out National Institute of Mental Health Data Archive (NDA) Validation set; Fig. 2a and Supplementary Data 1). An equal male bias is present in each of the three subtypes ($\chi^2(2) = 3.38$, $P = 0.19$). Importantly, the three-subtype solution heavily deviates from the null hypothesis that the data originate from a single multivariate Gaussian distribution ($P = 9.99 \times 10^{-5}$). Plots of the Uniform Manifold Approximation and Projection (UMAP) reduced data show evidence of three distinct peaks, one for each of the three subtypes (Fig. 2b). To better describe this three-subtype solution, we next plotted VABS scores for each subtype and in each VABS domain (Fig. 2c,d). We see a clear distinction between the subtypes in terms of ability—high, medium to low—that is preserved across each VABS domain. Although the subtype distributions overlap to some extent, the size of differences between subtypes is typically quite large (for example, Cohen's $d > 1$) for each pairwise subtype comparison (Fig. 2c,d), and these effect sizes are robustly preserved in independent training and validation sets.

We next applied the same stratification approach to a much older NDA outcome-relevant cohort ($n = 2,561$; 6–61 years; mean age of 13 years). A three-subtype solution emerges from this dataset with 99% generalization accuracy and strong rejection of the null hypothesis that data originate from a single multivariate Gaussian distribution ($P = 9.99 \times 10^{-5}$; Fig. 3a,b). One of the subtypes can be considered an extreme outlier subtype, because the scores are at floor levels near 20 and have hardly any variability around this floor (Fig. 3c,d). This subtype is relatively small ($n = 79$) and comprises ~3% of all individuals in this older cohort. The remaining two subtypes can be described as relatively high or low and are both relatively large and equal in size (high, $n = 1,444$; low, $n = 1,038$).

Continuity of early snapshot subtypes at outcome

Although early development is quite variable, later-life outcomes tend to be much more stable^{6,18}. Studies examining predictors of later-life adaptive functioning have shown that better adaptive functioning earlier in life predicts better later-life outcomes¹⁴. Thus, the presence of a high and low group at early ages may be prognostically suggestive of good versus poor later-life outcome. However, the presence of a third subtype between these two (the medium subtype) may suggest that this subtype is more uncertain regarding their later-life good versus poor outcomes. To test these predictions, we examined the correspondence of subtype labels in a subset of $n = 130$ individuals present in both the NDA early snapshot and older outcome datasets. As predicted, individuals in the early high or low subtypes are highly probable of remaining in that same high or low subtype at outcome (high = 82%, 95% CI = 69–93%; low = 84%, 95% CI = 71–94%; high and low combined = 83%, 95% CI = 75–91%). In contrast, the early snapshot medium subtype is much more ambiguous with respect to later subtype outcomes, with ~57% (95% CI = 43–71%) moving to the high outcome subtype, and the remaining 43% (95% CI = 28–56%) moving to the low outcome subtype (Fig. 3e). Corroborating this finding as well as previous work¹⁴, longitudinal VABS data show relatively flat or slightly declining standardized-score group-level trajectories throughout the time period up to 72 months and with no differences in trajectories between the subtypes (Supplementary Fig. 1 and Supplementary Data 2 and 3).

Developmental trajectories in adaptive functioning subtypes

In our next set of analyses we examined whether autism early adaptive functioning subtypes are distinctions that are sensitive to differential trajectories in non-verbal cognitive, language and fine motor domains, as measured by the Mullen Scales of Early Learning (MSEL) through the first five and a half years of life. In longitudinal NDA data,

we find age*subtype interactions throughout MSEL visual reception (VR), expressive language (EL), receptive language (RL) and fine motor (FM) subscales. These differences in developmental trajectories are most pronounced for the low subtype when compared with the high or medium subtypes, whereas differences between the high and medium subtypes are less strong (Fig. 4a and Supplementary Data 2 and 3). Because clustering was applied to the first timepoint of the NDA data, these results are not independent of the clustering procedure and may be biased. Thus, in the next analysis we tested for subtype differences in the MSEL developmental trajectories using a large and completely independent dataset—the University of California, San Diego Autism Center of Excellence (UCSD ACE) dataset ($n = 1,185$). Again, we discover age*subtype interactions across all MSEL subscales. These effects are driven by age*subtype interactions for all pairwise between-subtype comparisons except for the FM subscale, where strong differences were much less apparent (Fig. 4b and Supplementary Data 2 and 3). Including sex as a covariate in our longitudinal models resulted in nearly identical results (Supplementary Data 2). These results provide strong independent replication of the subtype MSEL trajectory differences. Overall, these results indicate that autism adaptive functioning subtypes isolated from early snapshots from the VABS are predictive of later trajectory differences across a range of developmental domains such as non-verbal cognitive ability, language and fine motor skills.

Comparing subtype versus normative models

Our clustering analysis provides evidence that autism is not one homogeneous population with respect to early or later-life adaptive functioning. We have also shown that early snapshot adaptive functioning subtypes are outcome-relevant and developmentally sensitive to variability in non-verbal cognitive, language and motor trajectories. However, is the early snapshot subtyping model better than other competing models at explaining developmental trajectories? To answer this question we compared the early snapshot subtyping model to a normative model that uses typical-development defined age-standardized norm cutoffs for adaptive functioning on the VABS (VABS norm). The VABS norm model uses 1 and 2 standard deviation cutoffs below the mean to create three subtypes. These subtype labels are then used in longitudinal MSEL models of the UCSD ACE data (Supplementary Data 2 and 3). Model comparison Akaike information criteria (AIC) statistics were computed for both the *reval* subtype and VABS norm models. We find that the VABS norm model produces lower AIC values than the *reval* subtype model and with ΔAIC values greater than 10. This indicates that traditionally defined ‘disability’ subtypes from VABS norms predict developmental trajectories as well or better than the data-driven *reval* autism subtype model on its own. We next created a ‘hybrid’ model based on the combination of the *reval*/autism subtype and VABS norm labels (Fig. 5a,b and Supplementary Data 2 and 3). This hybrid model produces the lowest AIC values, indicative of a much better model than either *reval* subtyping or the VABS norm model alone ($\Delta\text{AIC} > 10$; Fig. 5c). Thus, improved utility for predicting cognitive and motor developmental trajectories could be facilitated through the combination of both a data-driven autism-specific subtyping approach and information regarding where the child stands relative to typically developing adaptive functioning norms.

Finally, to aid future research applications, we developed an application that will allow users to input VABS scores and receive subtype labels as output (https://landii.shinyapps.io/vineland_statification_proj/). The intent behind this tool is to allow the field to immediately begin using these subtype labels for a priori experimental design in future studies. Furthermore, at an individualized level, the application may be useful for giving expectations behind developmental progress of an individual, given their subtype. For example, using the beta coefficients from the developmental trajectory models, we can express the expected rate of growth per month in terms of age-equivalent scores for each domain on the MSEL (Supplementary Data 4).

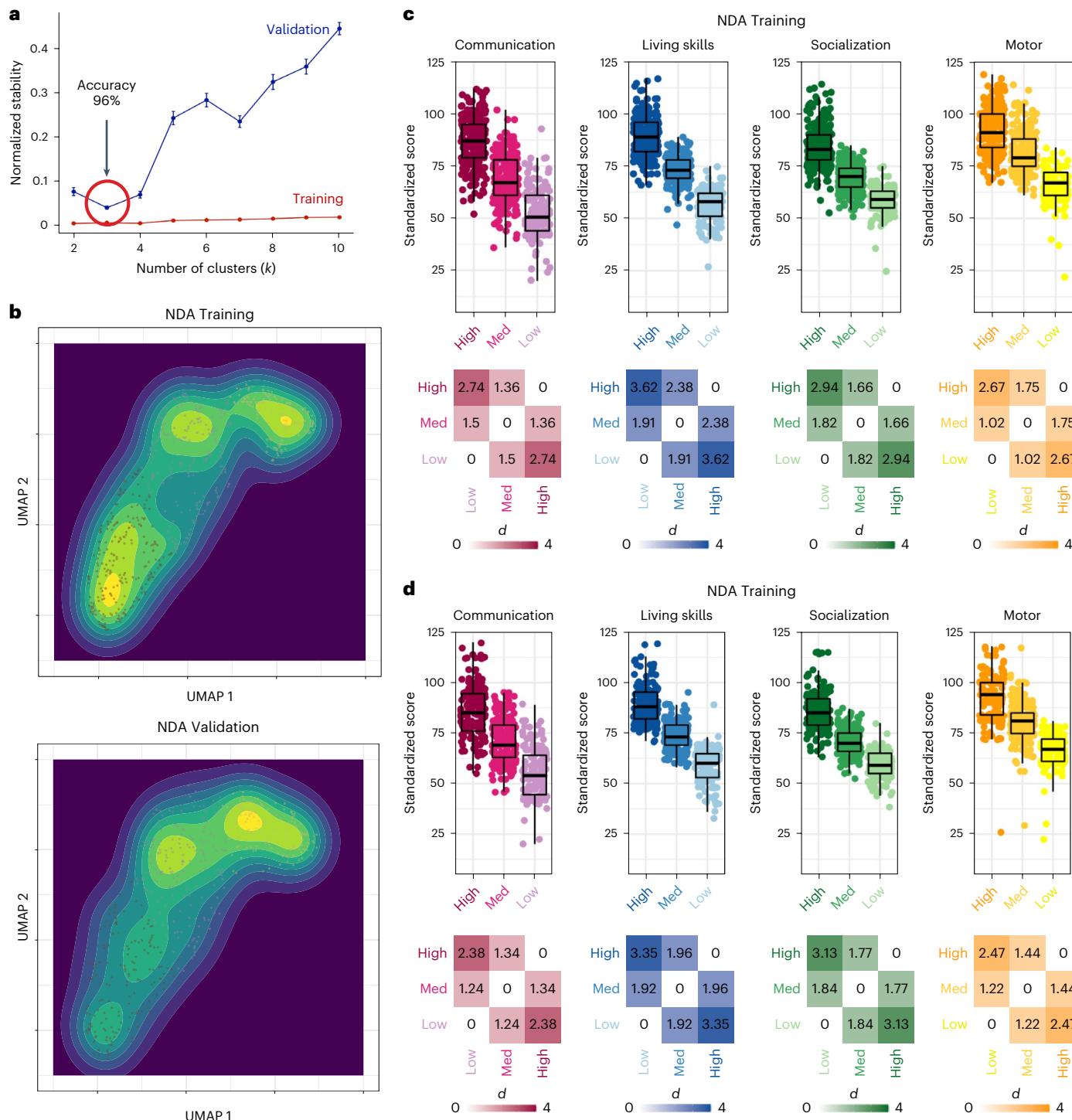


Fig. 2 | Unsupervised data-driven discovery of early snapshot autism adaptive functioning subtypes. This figure shows the results of stability-based relative clustering validation applied to the early snapshot NDA dataset. **a**, Normalized stability plot. Error bars in this plot represent the 95% confidence interval for normalized stability from repeated cross-validation. The optimal cluster solution k is the one that minimizes the normalized stability. It is clear that $k = 3$ is the optimal solution, minimizing the normalized cluster stability. In an independent validation set, this $k = 3$ solution generalizes with 96% accuracy. **b**, Plots of UMAP-reduced data input into the clustering for both the NDA Training (top) and

Validation (bottom) sets. The individuals are colored by subtype over the density plots. **c,d**, Graphical description of the subtypes across each VABS domain, with scatter boxplots and heatmaps depicting the pairwise standardized effect size difference (Cohen's d) between the subtypes; plots are shown for both the NDA Training set (**c**) and Validation (**d**) set. The boxplots show the interquartile range (IQR; first quartile, 25th percentile; third quartile, 75th percentile), and the whiskers indicate $Q_1 - (1.5 \times IQR)$ or $Q_3 + (1.5 \times IQR)$. The line within the boxplot represents the median. The sample sizes for all panels in this figure were $n = 603$ for the NDA Training set and $n = 495$ for the NDA Validation set.

Discussion

In this study we aimed to identify developmental- and outcome-sensitive stratifications of the early autism spectrum based on adaptive

functioning profiles. Standardized clinical assessment of adaptive functioning via the VABS offers a simple, easy-to-administer and rigorous way to obtain such early snapshots of stratified adaptive functioning

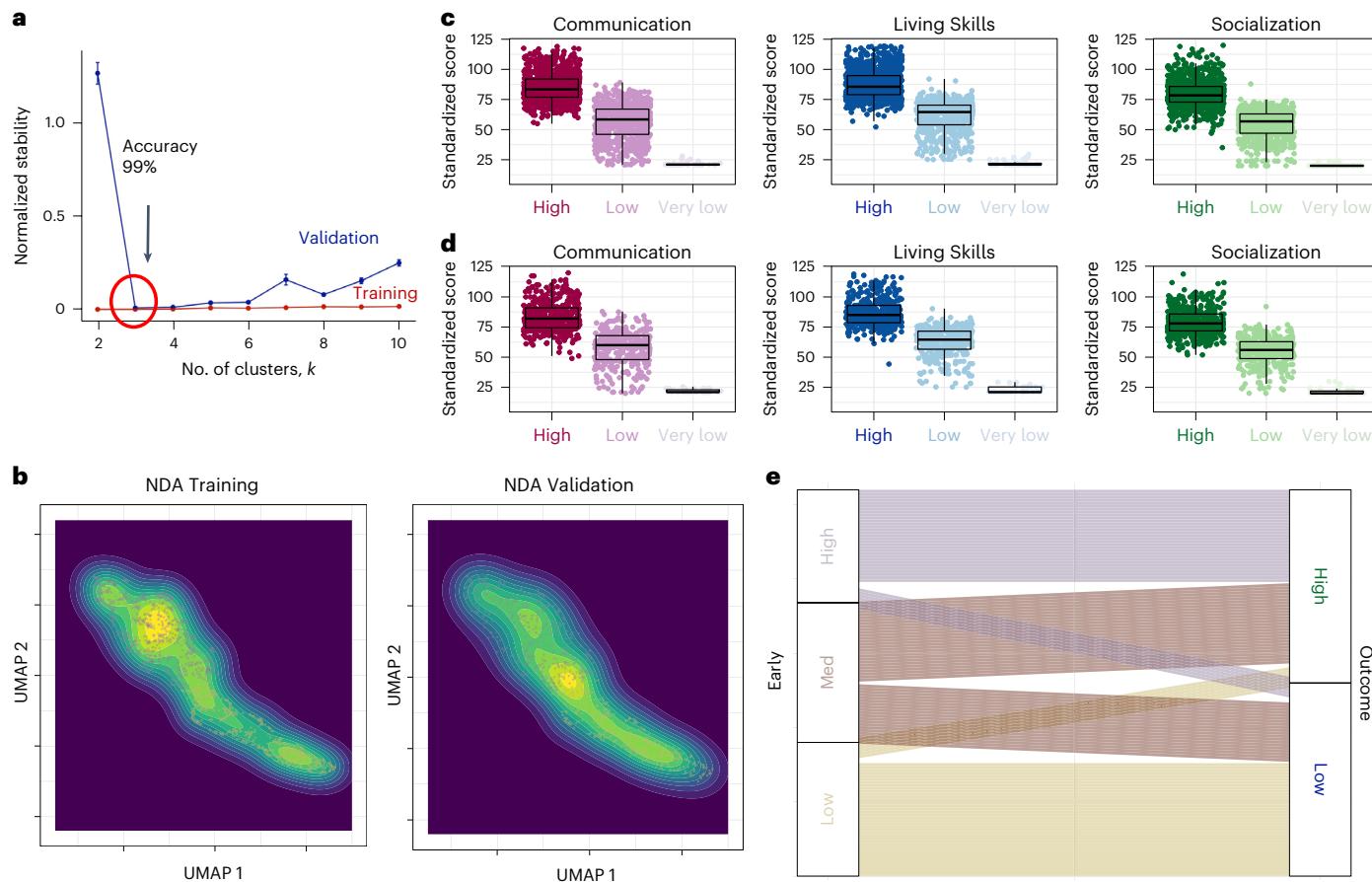


Fig. 3 | Later life outcome stratification. **a**, Normalized stability plot, showing that the best clustering solution is $k = 3$. This solution achieves 99% accuracy when generalizing to the validation set. Error bars in this plot represent the 95% confidence interval for normalized stability from repeated cross-validation. **b–d**, UMAP scatter boxplots (**c,d**) and density plots (**b**) showing the high and low subtypes. The boxplots indicate IQR (first quartile, 25th percentile; third quartile, 75th percentile), and the whiskers indicate $Q_1 - (1.5 \times IQR)$ or $Q_3 + (1.5 \times IQR)$.

The line within the boxplot represents the median. The plots for NDA Training (**c**) and NDA Validation (**d**) describe the subtypes for each of the three VABS domains (communication, living skills and socialization). **e**, Alluvial plot of the $n = 130$ subset of individuals that were in both early snapshot (left) and later life outcome (right) cohorts. The sample sizes for **a–d** were $n = 1,715$ for the NDA Training set and $n = 846$ for the NDA Validation set.

profiles. We found the presence of robust and stable clusters that highly deviate from a single multivariate Gaussian null distribution. This result indicates that autism is not one homogeneous population with respect to adaptive functioning. In early development, autism can be described by three subtypes that split into high, medium and low adaptive functioning strata. Similarly, at older ages (for example, mean age 13 years in the outcome cohort), autism can also be robustly clustered into three subtypes. One of these subtypes is an extreme outlier with very low VABS scores. The other two subtypes correspond to relatively high versus low subtypes. We also discovered that some of the early snapshot subtypes hold information sensitive to later outcomes. Both early high and low subtypes are highly likely to remain in those same subtypes at later ages. In contrast, the early medium subtype is more uncertain in terms of later outcome subtype membership. This result indicates that early snapshot stratification can result in labels that are outcome-sensitive, especially if an individual falls into the early snapshot high or low subtypes. For individuals falling into the early snapshot medium subtype, it may be pertinent for future work to examine what might be the predictors that nudge these individuals into different outcome subtypes (for example, the influence of different kinds of intervention, comorbid attention deficit hyperactivity disorder, executive functioning, environmental or educational influences)^{17,20–22}.

The discovered subtypes also hold predictive information about trajectories outside adaptive functioning, such as the non-verbal

cognitive, language and motor domains measured by the MSEL. Evidence of age*subtype interactions suggests that the subtypes show substantial differences in the steepness of their slopes over age (for example, rate of growth), with the most accelerated development present in the high group, and the more incremental and slowest growth exhibited by the low group. This result suggests that our subtyping approach yields robust and reproducible subtype labels that are developmentally sensitive outside of adaptive functioning and which could be utilized to inform expectations about prognosis in these domains. Finally, while the subtype model explains substantial variance in developmental trajectories, the model could be further refined by simple knowledge of where an individual stands with respect to TD adaptive functioning norms. Combining the knowledge of both TD-defined norms and autism-specific norms for subtyping, we constructed a hybrid model, where labels are informed by where an individual is relative to both TD and autism norms. This hybrid model was proven to be the best of all models at predicting variance in non-verbal cognitive, language and motor trajectories. Therefore, one key utility of our data-driven autism-specific subtyping approach is that it provides useful labels that, when combined with VABS norms about typical development, can prominently enhance our precision at estimating expected growth for autistic individuals in each subtype.

It is important to underscore that our clustering approach allows us to not only identify the optimal number of clusters, but also

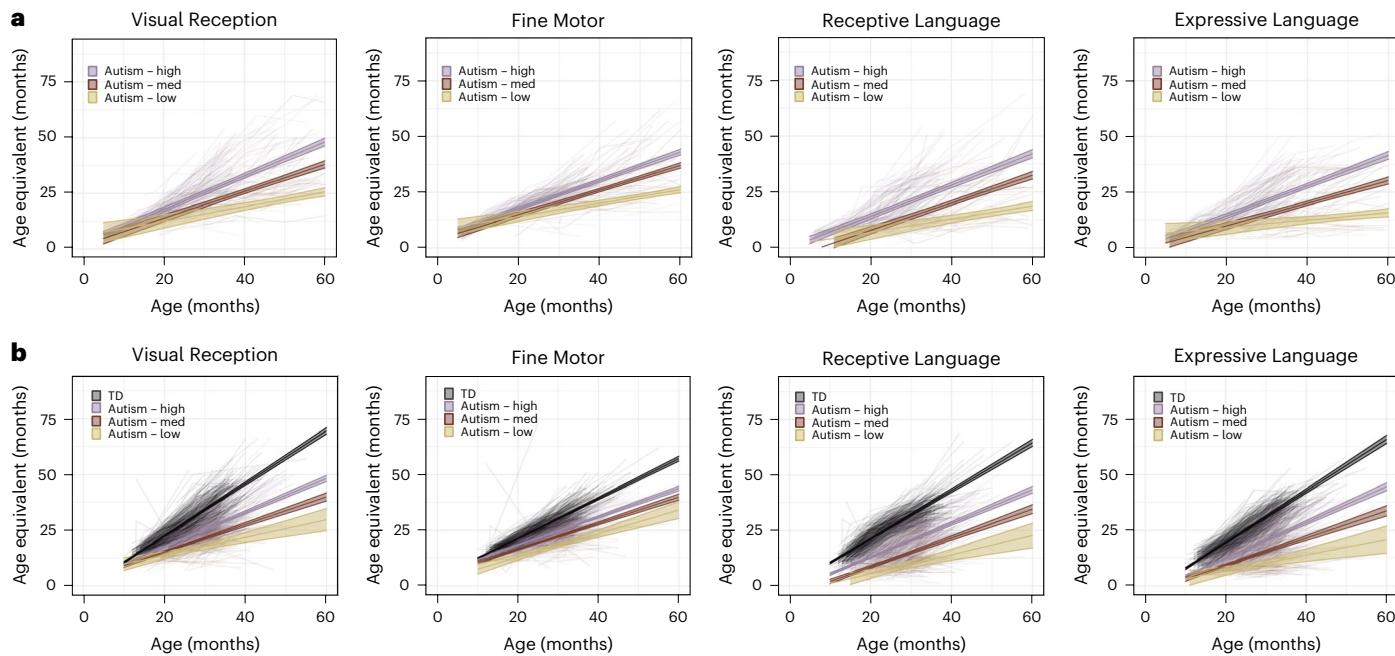


Fig. 4 | Differences in developmental trajectories among autism adaptive functioning subtypes in non-verbal cognitive, language and fine motor domains measured by the MSEL. **a,b,** Spaghetti plots of the developmental trajectories for MSEL VR, EL, RL and FM subscales for NDA (**a**) and UCSD ACE (**b**) datasets. Individual trajectories are shown in the background as transparent

lines, and subtype trajectories (light purple, high; maroon, medium; cream, low; black, TD) are shown in solid lines with 95% confidence bands around them. Sample sizes: $n = 701$ for **a** (autism – high, $n = 287$; autism – medium, $n = 274$; autism – low, $n = 140$); $n = 1,185$ for autistic (autism – high, $n = 702$; autism – medium, $n = 410$; autism – low, $n = 73$) and $n = 689$ for TD children for **b**.

describes how stable and generalizable that solution is to unseen new datasets. This three-cluster early snapshot model is 96% accurate, and the outcome subtype model is 99% accurate. Given this high level of generalization accuracy combined with the breadth of the dataset in which it was defined (for example, NDA data; $n > 1,000$), this result allows for high levels of confidence that these are stable and highly robust subtypes present in the autism population. This high level of confidence in generalization allows for immediate application of this subtyping model in new research. This immediate impact and ability to reuse our stratification model is novel with respect to most studies using unsupervised data-driven techniques. A large limitation of most studies using traditional clustering techniques is that those studies are descriptive of what occurs in the dataset in question, but give no indication about the generalizability of this solution in new unseen data, nor do they provide a reproducible easy-to-use stratification tool for the field to immediately apply in future research. Our approach immediately translates unsupervised data-driven discoveries into supervised knowledge that can be used in future studies to further progress the field.

One unique facet of our work is based on the idea of identifying developmental- and outcome-sensitive subtype information from early single snapshots of adaptive functioning in autistic individuals. The enhanced prognostic sensitivity of our subtyping approach has many potential high-value research applications that could accelerate future work. One example of this potential could be in a context like treatment research. For example, a specific behavioral early intervention that shows a small on-average effect in all autistic toddlers may show a much more pronounced effect in a specific early snapshot adaptive functioning subtype. In another example, a clinical trial of a pharmacological treatment may show little to no effect when analyzed in a traditional type of case–control model. Such clinical trial data could be re-analyzed after pretreatment adaptive functioning stratification to examine whether the drug has a differential effect on a specific subset of individuals that show very different pretreatment

adaptive functioning profiles. These are just a few of the many ways to deploy our reproducible stratification model in future research to help accelerate progress in the domain of personalized treatment research. Outside early behavioral and pharmacological intervention research, there may be other potential high-value research applications for our stratification approach that investigate how more supportive and personalized systems of care around an individual can better support optimal growth and outcomes for diagnosed individuals and their families. It is theorized that the adaptive potential of autistic individuals can vary considerably with the degree of environmental support within an individual’s systems of care⁶. Therefore, it will be of key importance in future work to put these ideas to the test, and early snapshot adaptive functioning stratification is one way in which such hypotheses could be immediately testable. Finally, one further high-value impact of our work could be within the area of mental health and quality of life. Both are key areas that need further research, as many autistic individuals suffer from a myriad of other mental health issues and a poorer quality of life^{23–25}. These issues also extend beyond the diagnosed individual and can affect the larger family unit (for example, caregivers)^{26,27}. It will be of high importance for future work to better understand how mental health and quality of life in affected individuals and their families can be better facilitated by such stratification approaches. A key overall research direction here could be based on how systems of care around an individual could be better adapted based on knowledge about subtype stratification to enhance mental health and quality of life.

There are some important caveats and limitations to highlight about the current work. First, it should be noted that this early snapshot approach is distinct from the type of information gleaned from other work that tries to subtype based on differential trajectories on the VABS over the first two decades of life^{14,17–19}. The latter attempts to utilize known trajectories to derive subtypes. Although such models are fit for optimally explaining within-individual and between-individual variance in trajectories, the opportunity for using it as an early prognostic tool is limited because full trajectory data on individuals are required

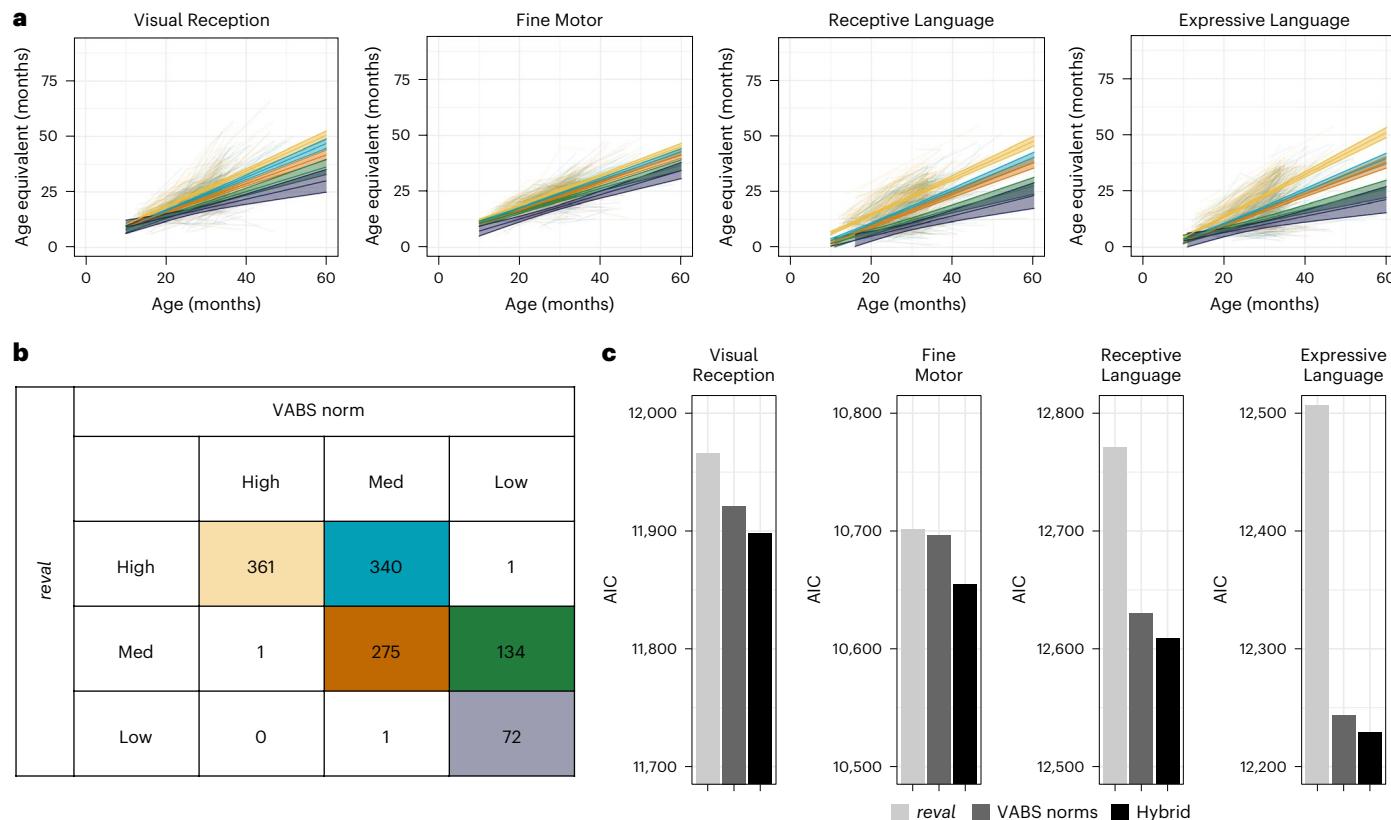


Fig. 5 | Model comparison between the autism-specific *reval* subtyping, VABS normative model and a hybrid model. The hybrid model was built by combining the adaptive functioning subtype labels obtained using *reval* with those obtained by stratifying using the VABS norm model. **a**, Spaghetti plots of the developmental trajectories for MSEL VR, EL, RL and FM subscales for the five hybrid subtypes. Individual trajectories are shown in the background as transparent lines, and subtype trajectories are shown as solid lines with 95% confidence bands around them. **b**, Confusion matrix. The subtype colors in **a**

are indicative of the subtypes shown in the cells of the confusion matrix (**b**), and sample sizes for the plots in **a** are indicated in the cells in **b**. This confusion matrix indicates where individuals fall with respect to either the *reval* or VABS norm subtyping models. **c**, AIC model comparison statistics for comparing the *reval* (light gray), VABS norm (dark gray) and hybrid (black) models. AIC is shown on the y-axis, and the models on the x axis. The model with the lowest AIC is considered best.

for the model to make label predictions. In contrast, our approach attempts to identify future predictive information via subtypes derived from single snapshots of VABS scores at early ages. Thus, the research applications for this niche are considerable and somewhat different from the other approaches, which are more fit to describing subtypes based on observed trajectories. Another caveat to underscore is that the current work is primarily relevant to the population from which the data are sampled—the autism population as represented in Western developed countries. Future work is needed to examine how well the current results and models generalize outside of this context. Finally, it is important to stress that our model was not developed for use outside research settings. It is not a tool that should be used in real-world clinical applications or to make key real-world decisions. The primary value and intent behind developing such a model is to aid future research.

In conclusion, we have demonstrated that a single snapshot of early adaptive functioning from the VABS can be utilized to predict highly robust and reproducible data-driven subtype labels that are informative about differential outcomes in adaptive functioning as well as different developmental trajectories in areas like non-verbal cognitive ability, language and motor behavior. Because our clustering approach allows for immediate translation of unsupervised data-driven discoveries into supervised knowledge (for example, a classifier), this work stands out in being able to allow others in the field to directly utilize our subtyping model in future work. Our stratification application (https://landii.shinyapps.io/vineland_statification_proj/) allows individuals to insert their own VABS data and obtain subtype labels

and developmental predictions about such subtypes as output. We hope that this work enables future research discoveries and is useful to help clinicians achieve initial expectations about prognosis based on limited early snapshots of adaptive functioning that are likely to be routine parts of initial clinical assessments.

Methods

All work reported here was approved by the Province of Trento Azienda Provinciale per i Servizi Sanitaria (APSS) ethical committee under protocol IIT EMN-755816-002-AUTISMS. Data collection for the UCSD ACE dataset was approved by the Institutional Review Board at the University of California, San Diego. Parents provided written informed consent according to the Declaration of Helsinki and were paid for their participation. All data analyzed in this work were anonymized.

NDA dataset

To first identify robust and reproducible early snapshot subtypes using unsupervised data-driven methods, we utilized large publicly available phenotypic data from NDA. Querying NDA in March 2020, we identified children with a diagnosis of autism at 6–72 months with at least one set of either a VABS II parent and caregiver form, VABS II survey form or VABS-3^{15,16} (Fig. 1a). After data merging and cleaning, $n=1,098$ individuals from $n=48$ different originating datasets remained for further downstream analysis. A subset of these individuals also had longitudinal VABS data ($n=410$). For the cross-sectional early snapshot clustering analyses, only the earliest timepoint was used from

these individuals. However, in later follow-up analyses, we utilized the longitudinal subset for examining developmental trajectories. We also extracted another $n = 2,561$ individuals from the NDA dataset with VABS scores at older ages (6–61 years). An identical data-driven clustering approach was used for this older ‘outcome’ cohort. See the Supplementary Methods for more details about the NDA dataset and Supplementary Data 5 for a list of NDA IDs and collections from which these data were taken.

UCSD ACE dataset

A second independent dataset was also utilized, with which we applied the NDA early snapshot subtype prediction model to predict subtype labels based on early VABS scores. This second dataset was longitudinal and allowed for an independent test of longitudinal hypotheses about the subtype model’s sensitivity to detect differences in developmental trajectories. This longitudinal dataset was collected at the UCSD ACE and comprises $n = 1,185$ autistic children and $n = 689$ TD children aged 9–72 months at their first VABS assessment (T1) (Supplementary Data 6). UCSD ACE was sampled and ascertained through a combination of population-based autism risk screening at 12 months²⁸ and community referrals. Longitudinal assessments took place approximately every six months from intake until outcome assessments at around four to five years of age. Clinical diagnosis of autism was made at UCSD ACE by expert clinicians at the outcome assessment and was aided by a full battery of tests including VABS, MSEL²⁹ and the Autism Diagnostic Observation Schedule (ADOS-2)³⁰. These diagnoses have been shown to be highly stable even from very early ages³¹. Intake VABS scores from UCSD ACE were utilized to predict subtype labels from the NDA prediction model and then subsequent longitudinal modeling was implemented on the full set of UCSD ACE VABS or MSEL scores to test for trajectory differences. Work utilizing this dataset was approved by the Institutional Review Board at the University of California, San Diego. Parents provided written informed consent according to the Declaration of Helsinki and were paid for their participation. More details about the UCSD ACE dataset are provided in the Supplementary Methods.

Measures

VABS. The VABS is a widely used standardized and semi-structured parent interview for assessing adaptive functioning in typical and clinical developmental populations throughout the lifespan. For children aged ≤ 72 months, the VABS assesses adaptive functioning in four major domains: communication, daily living skills, socialization and motor skills. For individuals beyond 72 months, only the communication, daily living skills and socialization domains are used. Standardized scores for each domain are computed to indicate where an individual scores relative to typically developing age-appropriate norms, whereby for each standardized score, the mean is 100 and the standard deviation is 15. For model comparison purposes, we utilized these typically developing norms to construct VABS normative subtypes using cutoffs of 1 and 2 standard deviations below the mean, applied to the Adaptive Behavior Composite score (ABC), computed using all four VABS domains (communication, daily living skills, socialization and motor).

MSEL. The MSEL is a standardized developmental test that can be administered from birth to 68 months of age that assesses the development of non-verbal cognitive, language and motor skills. Four of the five MSEL subscales were utilized in this work: VR, EL, RL and FM. In this work we examined growth over time in age-equivalent scores for each of these subscales. MSEL data were available for both the NDA and UCSD ACE datasets. For evaluating MSEL developmental trajectories, we utilized MSEL age-equivalent scores.

Stability-based relative clustering validation. Unsupervised data-driven clustering was achieved using our Python library (*reval*; https://github.com/IIT-LAND/reval_clustering) for implementing

stability-based relative clustering validation^{32,33} (see Supplementary Methods for more details and Fig. 1). To reproduce this analysis please see our code, which is deposited at https://github.com/IIT-LAND/vineland_subtyping. Our choice of clustering and classification algorithms throughout the application of *reval* was *k*-means clustering and *k*-nearest-neighbors classification. For the NDA early snapshot dataset (6–72 months), we used a 55/45 training-validation split, while ensuring that age and sex were balanced across this split. Due to the much larger sample size of the older NDA outcome dataset, we used a 67/33 training-validation split. After the data were split into training and validation sets, we implemented preprocessing steps such as batch correction for different VABS versions (see Supplementary Methods for more details), imputing missing values with a *k*-nearest neighbors imputation algorithm (`sklearn.impute.KNNImputer`), scaled data to a mean of 0 and standard deviation of 1 (`sklearn.preprocessing.StandardScaler`), and then applied UMAP³⁴ dimensionality reduction ($n_{neighbors} = 30$, $min_dist = 0.0$, $n_{components} = 2$, $random_state = 42$, metric = Euclidean). All preprocessing steps fitted on the training set were then applied to the validation set. After preprocessing, we apply a twofold internal cross-validation scheme to identify the optimal number of clusters (*k*-range, 2–10) that minimizes normalized cluster stability. This internal cross-validation scheme was repeated 100 times with different cross-validation splits to ensure robustness. Finally, a grid search procedure was utilized to select the optimal hyperparameters (for example, $k = 5$ neighbors for the *k*-nearest-neighbor classifier) utilized throughout. The final optimal *k* identified from the training set was applied to the validation set for *k*-means clustering, then a $k = 5$ *k*-nearest-neighbor classifier was fit to the training set and then applied to predict clustering labels on the validation set. The generalization accuracy on the validation set was then computed by comparing the classifier’s predicted labels to the actual clustering labels identified in the validation set.

Although stability-based relative clustering validation in *reval* tells us about the stability of clustering solutions, it does not test whether the actual solution is indicative of true clusters. Therefore, we followed up on the *reval* analysis by using the *SigClust* library in R to test whether the observed clustering solution significantly differs from the null hypothesis that the data originate from a single multivariate Gaussian distribution³⁵.

Longitudinal data analyses. To test for subtype differences in developmental trajectories, we utilized longitudinal data present in NDA and a second independent dataset, UCSD ACE. After *reval* identified the optimal clustering solution that generates high levels of generalization accuracy, we built a prediction model (that is, a *k*-nearest-neighbors classifier) that learns the subtype distinctions from NDA data, then applied this model to both NDA and UCSD ACE longitudinal datasets. For these prediction models we utilized the earliest VABS observation per each individual, after identical preprocessing steps as those described before implementation of *reval* (for example, imputation, scaling and UMAP dimensionality reduction). For testing developmental trajectory differences, we utilized linear mixed-effect models (that is, the *lmer* function in the *lme4* R library). The dependent variable in these models was either VABS domain standardized scores or MSEL subscale age-equivalent scores. Fixed effects were specified as age, subtype and the age*subtype interaction, and the random effect of age was modeled within-subject with random intercepts and slopes. Because multiple comparisons were made, we use false discovery rate (FDR) correction at a level of $q < 0.05$. Follow-up tests for specific pairwise group comparisons were made to decompose and describe the age*subtype interactions, and similar FDR control for multiple comparisons was used.

Model comparisons. To evaluate the utility of the VABS early snapshot subtyping model against other types of models, we compared

how well a model using VABS typically developing norms using the ABC score (VABS norm; for example, 1 and 2 standard deviation cut-offs) could predict trajectories. The ABC score was computed using all four subscales, motor included, for the early snapshot cohort (<72 months), and using only the communication, daily living skills and socialization for the outcome cohort (>72 months). We also examined a hybrid model combining the *revall* subtype labels with the VABS norm model (that is, hybrid label = *revallLabel_VABSNormLabel*). Models were compared by evaluating how well they explain variance in longitudinal MSEL analyses, with the AIC being used as the model comparison statistic. Models with lower AIC scores are considered better and we also computed AIC difference scores (Δ AIC) as a quantitative indicator of just how much better the best model is than the comparison model. Δ AIC > 10 indicates that the comparison model has little to no support for being as good as the best model with the lowest AIC³⁶.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this Article.

Data availability

All data utilized in this work can be found in the National Institute of Mental Health Data Archive (NDA; <https://nda.nih.gov>). NDA global unique identifiers (GUID) and collection IDs are provided in Supplementary Data 5 to identify individuals used in all analyses on the NDA dataset. For the UCSD ACE data, the data are available on NDA through collection IDs 9, 2466, 2968, 2290 and 2115.

Code availability

Analysis code to reproduce the analyses and figures is available on GitHub (https://github.com/IIT-LAND/vineland_subtyping). The *revall* Python library can be found on GitHub (https://github.com/IIT-LAND/revall_clustering) and the documentation at <https://revall.readthedocs.io>. The Shiny app that allows users to input their own VABS data and get subtype predictions can be found at https://landit.shinyapps.io/vineland_stabilization_proj/.

References

1. Lai, M.-C., Lombardo, M. V. & Baron-Cohen, S. *Autism. Lancet* **383**, 896–910 (2014).
2. Lombardo, M. V. & Mandelli, V. Rethinking our concepts and assumptions about autism. *Front. Psychiatry* **13**, 903489 (2022).
3. Lombardo, M. V., Lai, M.-C. & Baron-Cohen, S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol. Psychiatry* **24**, 1435–1450 (2019).
4. Collins, F. S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
5. Kapur, S., Phillips, A. G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).
6. Lord, C. et al. The Lancet Commission on the future of care and clinical research in autism. *Lancet* **S0140-6736**, 01541–01545 (2021).
7. de Bildt, A., Sytema, S., Kraijer, D., Sparrow, S. & Minderaa, R. Adaptive functioning and behaviour problems in relation to level of education in children and adolescents with intellectual disability. *J. Intellect. Disabil. Res.* **49**, 672–681 (2005).
8. Farley, M. A. et al. Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism Res.* **2**, 109–118 (2009).
9. Taylor, J. L. & Mailick, M. R. A longitudinal examination of 10-year change in vocational and educational activities for adults with autism spectrum disorders. *Dev. Psychol.* **50**, 699–708 (2014).
10. Clarke, E. B., McCauley, J. B. & Lord, C. Post-high school daily living skills in autism spectrum disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **60**, 978–985 (2021).
11. Taylor, J. L. & Henninger, N. A. Frequency and correlates of service access among youth with autism transitioning to adulthood. *J. Autism Dev. Disord.* **45**, 179–191 (2015).
12. Chatham, C. H. et al. Adaptive behavior in autism: minimal clinically important differences on the Vineland-II. *Autism Res.* **11**, 270–283 (2018).
13. Tillmann, J. et al. Investigating the factors underlying adaptive functioning in autism in the EU-AIMS Longitudinal European Autism Project. *Autism Res.* **12**, 645–657 (2019).
14. Szatmari, P. et al. Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA Psychiatry* **72**, 276–283 (2015).
15. Sparrow, S., Balla, D., Cicchetti, D. V. & Doll, E. A. *Vineland-II Scales Of Adaptive Behavior* (American Guidance Service, 2005).
16. Sparrow, S. S., Cicchetti, D. V. & Saulnier, C. *Vineland-3: Vineland Adaptive Behavior Scales* (Pearson, 2016).
17. Bal, V. H., Kim, S.-H., Cheong, D. & Lord, C. Daily living skills in individuals with autism spectrum disorder from 2 to 21 years of age. *Autism* **19**, 774–784 (2015).
18. Pickles, A., Anderson, D. K. & Lord, C. Heterogeneity and plasticity in the development of language: a 17-year follow-up of children referred early for possible autism. *J. Child Psychol. Psychiatry* **55**, 1354–1362 (2014).
19. Baghdadli, A. et al. Adaptive trajectories and early risk factors in the autism spectrum: a 15-year prospective study. *Autism Res.* **11**, 1455–1467 (2018).
20. Szatmari, P. et al. Association of child and family attributes with outcomes in children with autism. *JAMA Netw. Open* **4**, e212530 (2021).
21. Chandler, S. et al. Pathways to adaptive functioning in autism from early childhood to adolescence. *Autism Res.* **15**, 1883–1893 (2022).
22. Pugliese, C. E. et al. Longitudinal examination of adaptive behavior in autism spectrum disorders: influence of executive function. *J. Autism Dev. Disord.* **46**, 467–477 (2016).
23. Lai, M.-C. et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *Lancet Psychiatry* **6**, 819–829 (2019).
24. van Heijst, B. F. C. & Geurts, H. M. Quality of life in autism across the lifespan: a meta-analysis. *Autism* **19**, 158–167 (2015).
25. Ayres, M. et al. A systematic review of quality of life of adults on the autism spectrum. *Autism* **22**, 774–783 (2018).
26. Bonis, S. Stress and parents of children with autism: a review of literature. *Issues Ment. Health Nurs.* **37**, 153–163 (2016).
27. Vasilopoulou, E. & Nisbet, J. The quality of life of parents of children with Autism Spectrum Disorder: a systematic review. *Res. Autism Spectr. Disord.* **23**, 36–49 (2016).
28. Pierce, K. et al. Detecting, studying and treating autism early: the one-year well-baby check-up approach. *J. Pediatr.* **159**, 458–465 (2011).
29. Mullen, E. *Mullen Scales of Early Learning* (American Guidance Service, 1995).
30. Lord, C. et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223 (2000).
31. Pierce, K. et al. Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatr.* **173**, 578–587 (2019).

32. Lange, T., Roth, V., Braun, M. L. & Buhmann, J. M. Stability-based validation of clustering solutions. *Neural Comput.* **16**, 1299–1323 (2004).
33. Landi, I., Mandelli, V. & Lombardo, M. V. reval: A Python package to determine best clustering solutions with stability-based relative clustering validation. *Patterns (N. Y.)* **2**, 100228 (2021).
34. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2020).
35. Liu, Y., Hayes, D. N., Nobel, A. & Marron, J. S. Statistical significance of clustering for high-dimension, low-sample size data. *J. Am. Stat. Assoc.* **103**, 1281–1293 (2008).
36. Burnham, K. P. & Anderson, D. R. Multimodel inference: understanding AIC and BIC in model selection. *Soc. Methods Res.* **33**, 261–304 (2004).

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 755816 (ERC Starting Grant to M.V.L.). Data from UCSD ACE were collected with support from the following grants: NIMH R01-MH080134 (K.P.), NIMH R01-MH104446 (K.P.), NFAR grant (K.P.), NIMH Autism Center of Excellence grant P50-MH081755 (E.C. and K.P.), NIMH R01-MH036840 (E.C.), NIMH R01-MH110558 (E.C.), NIMH U01-MH108898 (E.C.) and NIDCD R01-DC016385 (E.C., K.P. and M.V.L.).

Author contributions

Conceptualization was provided by M.V.L., V.M. and I.L. Methodology was provided by M.V.L., I.L. and V.M. Software was contributed by I.L. and V.M. Formal analysis was carried out by M.V.L., V.M. and I.L. Investigations were carried out by M.V.L., I.L., V.M., E.M.B., E.C. and K.P. Data curation was performed by M.V.L., I.L., V.M., E.C. and K.P. The original draft was prepared by M.V.L., V.M., I.L., E.M.B., E.C. and K.P. Review and editing was carried out by M.V.L., V.M., I.L., E.M.B., E.C. and K.P. Visualization was provided by M.V.L., V.M. and I.L. Supervision was provided by M.V.L. Project

administration was carried out by M.V.L. Funding acquisition was provided by M.V.L., E.C. and K.P.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44220-023-00056-6>.

Correspondence and requests for materials should be addressed to Michael V. Lombardo.

Peer review information *Nature Mental Health* thanks Weizhao Lu, Antonia San José Cáceres and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection
Data analysis	Analysis code to reproduce the analyses and figures can be found on GitHub (https://github.com/IIT-LAND/vineland_subtyping). The reval Python library (version 1.0) can be found on GitHub (https://github.com/IIT-LAND/reval_clustering) and the documentation can be found here: https://reval.readthedocs.io . Python version 3.8 or higher and R version 4 or higher is necessary to reproduce the analysis, as indicated on the GitHub repo. Version numbers for R libraries such as SigClust are not specified by the developer and thus can be considered as v1.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data utilized in this work can be found on the National Institute of Mental Health Data Archive (NDA; <https://nda.nih.gov>). NDA global unique identifiers (GUID)

and collection IDs can be found Supplementary Data 5 to identify individuals used in all analyses on the NDA dataset. For the UCSD ACE data, this data is available on NDA through collection IDs 9, 2466, 2968, 2290, and 2115.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Sex has been used in the manuscript and has been included in data analyses.

Population characteristics

To first identify robust and reproducible early snapshot subtypes using unsupervised data-driven methods, we utilized large publicly available phenotypic data from the National Institute of Mental Health Data Archive (NDA). Querying NDA in March 2020, we identified children with a diagnosis of autism ages 6-72 months with at least one set of either Vineland Adaptive Behavior Scales (VABS) II parent and caregiver form, VABS-II survey form, or VABS-3. After data merging and cleaning, n=1,098 individuals from n=48 different originating datasets remained for further downstream analysis. A subset of these individuals also had longitudinal VABS data (n=410). For the cross-sectional early snapshot clustering analyses, only the earliest time-point was used from these individuals. However, in later follow-up analyses, we utilized the longitudinal subset for examining developmental trajectories. We also extracted another n=2,561 individuals from NDA with VABS scores at older ages (6-61 years). An identical data-driven clustering approach was used for this older ‘outcome’ cohort.

A second independent dataset was also utilized, whereby we applied the NDA early snapshot subtype prediction model to predict subtype labels based on early VABS scores. This second dataset was longitudinal and allowed for an independent test of longitudinal hypotheses about the subtype model’s sensitivity to detect differences in developmental trajectories. This longitudinal dataset was collected at the University of California, San Diego Autism Center of Excellence (UCSD ACE) and comprises n=1,185 autistic children and n=689 typically-developing (TD) children, aged 9-72 months at their first Vineland assessment (T1). UCSD ACE was sampled and ascertained through a combination of population-based autism risk screening at 12 months and community referrals. Longitudinal assessments took place approximately every 6 months from intake until outcome assessments around 4-5 years of age. Clinical diagnosis of autism was made at UCSD ACE by expert clinicians at the outcome assessment and was aided by a full battery of tests including VABS, the Mullen Scales of Early Learning (MSEL), and the Autism Diagnostic Observation Schedule (ADOS-2). These diagnoses have been shown to be highly stable even from very early ages. Intake VABS scores from UCSD ACE were utilized to predict subtype labels from the NDA prediction model and then subsequent longitudinal modeling was implemented on the full set of UCSD ACE VABS or MSEL scores to test for trajectory differences. Work utilizing this dataset was approved by the Institutional Review Board at University of California, San Diego. Parents provided written informed consent according to the Declaration of Helsinki and were paid for their participation.

Recruitment

UCSD ACE was sampled and ascertained through a combination of population-based autism risk screening at 12 months and community referrals.

Ethics oversight

All work reported here was approved by the Province of Trento Azienda Provinciale per i Servizi Sanitaria (APSS) ethical committee under protocol IIT EMN-755816-002-AUTISMS. Data collection for the UCSD ACE dataset was approved by the Institutional Review Board at University of California, San Diego. Parents provided written informed consent according to the Declaration of Helsinki and were paid for their participation. All data analyzed in this work was anonymized.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample size calculation was performed. All available data from NDA and UCSD ACE was used to maximize sample size as much as possible and to give the widest possible spread across the entire autism population. This strategy was utilized to maximize statistical power as much as possible and to make for the most generalizable results to the entire autism population.

Data exclusions

If more than one VABS subscale domain was missing, the subject was dropped from the analysis

Replication

Stratification analyses were split into train and validation sets for replication. Longitudinal analyses were replicated in NDA and UCSD ACCE datasets.

Randomization

Randomization into diagnostic groups (i.e. TD vs autism) is not possible for this study, because diagnosis cannot be randomized. However, within the stratification analyses in autism, randomization was utilized to divide data into independent training and validation sets, and to

subsequently split the training set into independent training and test sets for cross validation. For these randomizations, each originating NDA dataset was randomly split while ensuring that other covariates such as age and sex were balanced across the splits.

Blinding

Randomized allocation of data to training and validation sets as well as internal training and test sets was blind to final subtype labels.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|-------------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | Antibodies |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | Animals and other organisms |
| <input checked="" type="checkbox"/> | Clinical data |
| <input checked="" type="checkbox"/> | Dual use research of concern |

Methods

- | | |
|-------------------------------------|------------------------|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | ChIP-seq |
| <input checked="" type="checkbox"/> | Flow cytometry |
| <input checked="" type="checkbox"/> | MRI-based neuroimaging |