Table 1. Sample characteristics of the study cohorts

Demographic and phenotypic characteristics of individuals included in the SPARK discovery cohort and SSC validation cohort. Continuous variables are presented as mean (standard deviation), and categorical variables as number (percentage). Age at registration was converted from months to years for presentation. Race categories were harmonized into mutually exclusive categories using a predefined hierarchical classification to facilitate comparisons between cohorts. Between-cohort differences were evaluated using Wilcoxon rank-sum tests for continuous variables, Pearson's chi-square test for sex, and an overall Pearson's chi-square test across race categories for race. SCQ, Social Communication Questionnaire; DCDQ, Developmental Coordination Disorder Questionnaire; RBS-R, Repetitive Behavior Scale-Revised; VABS, Vineland Adaptive Behavior Scales; CBCL, Child Behavior Checklist.

Variables	SPARK	SSC	*P* value
Sample size, N	6,355	2,717	
Female, N (%)	1525 (24.0%)	366 (13.5%)	1.70E-29
Age in years, mean (SD)	9.18 (3.43)	8.98 (3.54)	1.78E-06
Race, N (%)			2.61E-47
African American	345 (5.4%)	125 (4.6%)	
Asian	132 (2.1%)	149 (5.5%)	
Native American	32 (0.5%)	30 (1.1%)	
Native Hawaiian	7 (0.1%)	10 (0.4%)	
White	4940 (77.7%)	2166 (79.7%)	
Other race	219 (3.4%)	147 (5.4%)	
More than one race	680 (10.7%)	90 (3.3%)	
SCQ summary score, mean (SD)	21.39 (7.37)	20.02 (6.44)	5.35E-20
RBS-R total score, mean (SD)	35.73 (21.39)	27.14 (17.41)	3.83E-77
DCDQ total score, mean (SD)	38.04 (12.75)	38.65 (11.90)	2.93E-03
Vineland Adaptive Behavior Scales, mean (SD)			
Adaptive Behavior Composite (ABC)	73.18 (14.23)	73.21 (12.01)	5.21E-02
Communication	74.84 (18.85)	77.12 (14.50)	1.36E-01
Daily Living Skills	76.50 (16.99)	76.45 (13.81)	1.56E-01
Socialization	69.61 (17.49)	71.02 (12.55)	8.64E-01
Mood or anxiety disorders, n (%) [§]	2610 (41.1%)	717 (26.4%)	3.63E-40
ADHD, n (%) [§]	3274 (51.5%)	415 (15.3%)	5.09E-227

† SCQ, Social Communication Questionnaire; DCDQ, Developmental Coordination Disorder Questionnaire; RBS-R, Repetitive Behavior Scale-Revised; CBCL, Child Behavior Checklist; ADHD, attention-deficit/hyperactivity disorder.

‡ Continuous variables were compared between cohorts using Wilcoxon rank-sum tests where as categorical variables were compared by Pearson's chi-square.

§ In SPARK, these measures were obtained from binary medical-history surveys. In SSC, mood or anxiety disorders were defined as a CBCL Anxious/Depressed, Affective Problems, or Anxiety Problems T-score greater than 70, and ADHD was defined as a CBCL ADHD Problems T-score greater than 70. Medical-history measures were compared between cohorts using Pearson chi-square tests.

Table 2. Clinical characteristics of ASD support subtypes in the SPARK discovery and SSC validation cohorts.

Subtype 1, Subtype 2, and Subtype 3 correspond to the High-support, Behavioral-support, and Low-support subtypes, respectively. Values are presented as n (%) or mean [standard deviation] as appropriate. Clinical characteristics include language ability, cognitive performance, social communication symptoms (SCQ), motor coordination (DCDQ), repetitive behaviors (RBS-R), adaptive functioning (Vineland measures), and behavioral or psychiatric features. Within each cohort, differences across the three subtypes were evaluated using Pearson's χ^2 tests for categorical variables and Kruskal–Wallis tests for continuous variables. For behavioral and psychiatric measures, SPARK values represent the number and percentage of participants with parent-reported medical features, whereas SSC values represent CBCL-derived quantitative symptom scores because comparable case-control information was unavailable for a large proportion of SSC participants. The consistency of subtype-specific clinical profiles across SPARK and SSC supports the preservation of High-support, Behavioral-support, and Low-support subtype characteristics in the independent validation cohort.

	SPARK				SSC			
Variables	**Subtype 1 (HSN)**	**Subtype 2 (EBSN)**	**Subtype 3 (LSN)**	***P* value**	**Subtype 1 (HSN)**	**Subtype 2 (EBSN)**	**Subtype 3 (LSN)**	***P* value**
Sample size, N (%)	1962 (30.9)	1531 (24.1)	2862 (45.0)		702 (25.8)	513 (18.9)	1502 (55.3)	
Female, N (%)	491 (25.0)	347 (22.7)	687 (24.0)	2.69E-01	127 (18.1)	78 (15.2)	161 (10.7)	6.34E-06
Age, mean (SD)	107.37 (42.87)	108.48 (40.15)	112.98 (40.22)	2.77E-05	117.08 (45.47)	107.33 (41.24)	103.64 (40.78)	6.28E-10
Language level, N (%) †				4.97E-300				1.75E-138
No words	354 (18.0)	29 (1.9)	22 (0.8)		104 (14.8)	1 (0.2)	1 (0.1)	
Uses single words	308 (15.7)	84 (5.5)	61 (2.1)		162 (23.1)	40 (7.8)	18 (1.2)	
Combines or uses three words	377 (19.2)	208 (13.6)	170 (5.9)		436 (62.1)	472 (92.0)	1483 (98.7)	
Uses longer sentences	923 (47.0)	1210 (79.0)	2609 (91.2)		—	—	—	
Cognitive test score, N (%)				4.96E-297				1.50E-258
≤ 24	187 (9.5)	51 (3.3)	19 (0.7)		71 (10.1)	4 (0.8)	0 (0.0)	
25 – 39	165 (8.4)	59 (3.9)	33 (1.2)		200 (28.5)	24 (4.7)	4 (0.3)	
40 – 54	227 (11.6)	62 (4.0)	41 (1.4)		147 (20.9)	60 (11.7)	39 (2.6)	
55 – 69	346 (17.6)	161 (10.5)	109 (3.8)		109 (15.5)	106 (20.7)	120 (8.0)	
70 – 79	318 (16.2)	218 (14.2)	268 (9.4)		68 (9.7)	88 (17.2)	186 (12.4)	
80 – 89	228 (11.6)	218 (14.2)	307 (10.7)		45 (6.4)	57 (11.1)	246 (16.4)	
90 – 109	221 (11.3)	240 (15.7)	637 (22.3)		47 (6.7)	114 (22.2)	502 (33.4)	
110 – 119	126 (6.4)	208 (13.6)	521 (18.2)		8 (1.1)	35 (6.8)	166 (11.1)	
120 – 129	79 (4.0)	164 (10.7)	445 (15.5)		3 (0.4)	14 (2.7)	116 (7.7)	
≥ 130	65 (3.3)	150 (9.8)	482 (16.8)		4 (0.6)	11 (2.1)	123 (8.2)	
SCQ summary score, mean (SD)	24.65 (6.27)	24.99 (5.82)	17.24 (6.63)	<1E-308	23.44 (5.46)	24.29 (5.26)	16.95 (5.52)	1.32E-173
DCDQ total score, mean (SD)	31.26 (10.57)	34.84 (10.38)	44.40 (12.23)	<1E-308	31.54 (9.75)	34.99 (10.46)	43.22 (11.19)	1.75E-219
RBS-R total score, mean (SD)	36.44 (19.93)	58.47 (17.79)	23.08 (11.71)	7.22E-303	26.90 (13.52)	49.76 (16.07)	19.52 (11.91)	2.48E-118
Vineland Adaptive Behavior Scales, mean (SD)								
Adaptive Behavior Composite (ABC)	58.60 (11.64)	72.61 (6.43)	83.48 (9.09)	<1E-308	59.83 (7.69)	70.00 (6.66)	80.55 (8.79)	<1E-308
Communication	56.19 (18.22)	75.18 (9.53)	87.44 (10.75)	<1E-308	61.78 (10.72)	74.42 (8.67)	85.21 (11.10)	9.25E-294
Daily Living Skills	59.79 (12.85)	76.41 (11.05)	88.01 (11.77)	<1E-308	62.09 (9.07)	73.45 (9.54)	84.18 (10.79)	3.23E-298
Socialization	52.57 (15.00)	68.87 (11.61)	81.70 (10.36)	<1E-308	58.07 (8.24)	67.15 (8.25)	78.41 (9.59)	0.00E+00
Mood or anxiety disorders, N (%) §	688 (35.1)	797 (52.1)	1125 (39.3)	1.88E-24	175 (24.9)	230 (44.8)	312 (20.8)	1.06E-25
ADHD, N (%) §	905 (46.1)	912 (59.6)	1457 (50.9)	2.10E-14	106 (15.1)	143 (27.9)	166 (11.1)	6.84E-19

§ In SPARK, these measures were obtained from binary medical-history surveys. In SSC, mood or anxiety disorders were defined as a CBCL Anxious/Depressed, Affective Problems, or Anxiety Problems T-score greater than 70, and ADHD was defined as a CBCL ADHD Problems T-score greater than 70. Medical-history measures were compared between subtypes using Pearson chi-square tests.

† Language level was obtained at individual registration in SPARK and from the ADI-R overall language-level item in SSC. Because the original response categories differed between cohorts, conceptually corresponding categories are aligned for presentation but are not directly equivalent; SPARK further distinguishes combining three words from using longer sentences.

Table 3. Correlations of ML-derived ASD subtype scores and symptom measures in the SPARK discovery and SSC validation cohorts.

Three continuous subtype scores were derived for each participant, corresponding to the High-support-need (HSN), Behavioral-support-need (EBSN), and Low-support-need (LSN) dimensions. For each subtype score, the association with each clinical measure is summarized either as a correlation coefficient (r) or as the mean difference (Δ) between binary groups, depending on the variable type, with the corresponding statistical test indicated in the table. Clinical measures include language ability, cognitive performance, social communication symptoms (SCQ), motor coordination (DCDQ), repetitive behaviors (RBS-R), adaptive functioning (Vineland measures), and behavioral or psychiatric features. Because comparable binary clinical history variables were unavailable for most SSC participants, behavioral and psychiatric measures in SSC are represented by quantitative CBCL-derived symptom scores. Associations were evaluated independently in the SPARK discovery cohort and the SSC validation cohort.

	SPARK						SSC					
Variables	**Subtype 1 Score (HSN)**	***P* value**	**Subtype 2 score (EBSN)**	***P* value**	**Subtype 3 Score (LSN)**	***P* value**	**Subtype 1 Score (HSN)**	***P* value**	**Subtype 2 score (EBSN)**	***P* value**	**Subtype 3 Score (LSN)**	***P* value**
Language level (r , ANOVA)	-0.5351	< 2E-16	0.2735	< 2E-16	0.4934	< 2E-16	-0.5655	< 2E-16	0.2992	< 2E-16	0.4888	< 2E-16
Cogitive test score (r , ANOVA)	-0.4874	< 2E-16	0.0986	3.26E-15	0.5137	< 2E-16	-0.6914	< 2E-16	0.1033	6.84E-08	0.6619	< 2E-16
SCQ summary score (r , Pearson)	0.3757	< 2E-16	0.1861	< 2E-16	−0.5710	< 2E-16	0.4511	< 2E-16	0.2667	< 2E-16	−0.6076	< 2E-16
RBS-R total score (r , Pearson)	0.1048	< 2E-16	0.4859	< 2E-16	−0.6032	< 2E-16	0.1116	5.43E-09	0.5801	< 2E-16	−0.5205	< 2E-16
DCDQ total score (r , Pearson)	−0.4526	< 2E-16	−0.0754	1.74E-09	0.5528	< 2E-16	−0.4694	< 2E-16	−0.0875	4.89E-06	0.5161	< 2E-16
Vineland Adaptive Behavior Scales (r , Pearson)												
Adaptive Behavior Composite (ABC)	−0.8505	< 2E-16	0.1185	< 2E-16	0.8717	< 2E-16	−0.8634	< 2E-16	−0.0303	1.14E-01	0.8577	< 2E-16
Communication	−0.8131	< 2E-16	0.1824	< 2E-16	0.8195	< 2E-16	−0.8097	< 2E-16	0.0251	1.91E-01	0.7886	< 2E-16
Daily Living Skills	−0.7909	< 2E-16	0.1148	4.26E-20	0.7968	< 2E-16	−0.7988	< 2E-16	−0.0077	6.90E-01	0.7826	< 2E-16
Socialization	−0.7773	< 2E-16	0.1215	< 2E-16	0.8065	< 2E-16	−0.7883	< 2E-16	−0.0472	1.39E-02	0.8036	< 2E-16
CBCL T-score on ADHD (r , Pearson)	0.1353	5.97E-15	0.1925	< 2E-16	−0.2848	< 2E-16	0.1154	1.61E-09	0.2448	< 2E-16	-0.2391	< 2E-16
CBCL T-score on Anxious or Depressed (r , Pearsc	−0.1151	3.33E-11	0.2695	< 2E-16	−0.0640	2.32E-04	−0.1234	1.10E-10	0.2711	< 2E-16	-0.0162	4.00E-01
ADHD diagnosis (Δ, t-test) §	−0.0223	5.61E-06	0.0493	< 2E-16	0.0007	8.90E-01	0.0158	4.70E-02	0.0669	< 2E-16	-0.0649	5.82E-15
Mood or anxiety disorders diagnosis (Δ, t-test) §	−0.0273	1.76E-08	0.0516	< 2E-16	−0.0046	3.56E-01	0.0105	1.17E-01	0.0588	< 2E-16	-0.0571	6.32E-16

§ In SPARK, these measures were obtained from binary medical-history surveys. In SSC, mood or anxiety disorders were defined as a CBCL Anxious or Depressed, Affective Problems, or Anxiety Problems T-score greater than 70, and ADHD was defined as a CBCL ADHD Problems T-score greater than 70.

Table 4. Associations between ML-derived ASD subtype scores and support needs, intervention utilization, special education and medication in SPARK and SSC.

Associations were evaluated using logistic regression models with survey-derived binary outcomes as dependent variables and continuous subtype scores as predictors. Each support-need category was analyzed as a separate binary outcome. Models were adjusted for sex, age, household income, father's highest education, mother's highest education, and the first ten ancestry principal components. Odds ratios (ORs) represent the change in the odds of each outcome per 0.1-unit increase in the corresponding subtype score. Corresponding 95% confidence intervals (CIs) and P values are reported. Analyses were performed separately in SPARK and SSC according to the availability of outcome measures. *P < 0.05; **P < 0.01; ***P < 0.001.

Cohort	Category	Type of intervention	Subtype Score	OR (95% CI) †	*P* value §
SPARK	Support Needs	Minimal to no support	Subtype 1 (HSN)	0.56 (0.51–0.60)	1.06E-44***
			Subtype 2 (EBSN)	0.78 (0.73–0.82)	5.5E-18***
			Subtype 3 (LSN)	1.90 (1.77–2.03)	5.25E-75***
		Some support	Subtype 1 (HSN)	0.76 (0.73–0.78)	6.76E-60***
			Subtype 2 (EBSN)	1.10 (1.06–1.13)	1.39E-08***
			Subtype 3 (LSN)	1.38 (1.34–1.43)	1.06E-82***
		Substantial support	Subtype 1 (HSN)	1.20 (1.17–1.24)	3.43E-31***
			Subtype 2 (EBSN)	1.12 (1.09–1.16)	3.15E-12***
			Subtype 3 (LSN)	0.75 (0.73–0.78)	2.99E-61***
		Very substantial support	Subtype 1 (HSN)	1.63 (1.56–1.71)	3.71E-98***
			Subtype 2 (EBSN)	0.77 (0.73–0.81)	1.6E-23***
			Subtype 3 (LSN)	0.43 (0.40–0.47)	1.35E-112***
	Intervention	Speech/language therapy	Subtype 1 (HSN)	1.21 (1.17–1.25)	1.71E-30***
			Subtype 2 (EBSN)	0.96 (0.93–0.99)	2.09E-02*
			Subtype 3 (LSN)	0.82 (0.79–0.84)	1.95E-34***
		Medication	Subtype 1 (HSN)	1.06 (1.02–1.09)	7.89E-04***
			Subtype 2 (EBSN)	1.10 (1.06–1.14)	6.48E-08***
			Subtype 3 (LSN)	0.88 (0.85–0.91)	1.67E-14***
		Counseling/therapy with a psychologist or similar	Subtype 1 (HSN)	0.81 (0.78–0.84)	3.63E-34***
			Subtype 2 (EBSN)	1.23 (1.19–1.27)	1.21E-30***
			Subtype 3 (LSN)	1.10 (1.07–1.14)	8.52E-09***
		Social skills group	Subtype 1 (HSN)	0.91 (0.87–0.94)	2.42E-08***
			Subtype 2 (EBSN)	1.11 (1.07–1.15)	1.27E-08***
			Subtype 3 (LSN)	1.05 (1.02–1.09)	2.63E-03**
	Special Education	Self-contained classroom for ASD - full-time	Subtype 1 (HSN)	1.38 (1.33–1.43)	1.64E-61***
			Subtype 2 (EBSN)	0.87 (0.83–0.91)	3.90E-10***
			Subtype 3 (LSN)	0.69 (0.66–0.72)	1.98E-61***
		Other self-contained classroom - full-time	Subtype 1 (HSN)	1.21 (1.15–1.27)	1.03E-14***
			Subtype 2 (EBSN)	1.02 (0.96–1.07)	5.69E-01
			Subtype 3 (LSN)	0.78 (0.74–0.83)	3.45E-19***
		1:1 classroom aide	Subtype 1 (HSN)	1.20 (1.16–1.24)	3.08E-29***
			Subtype 2 (EBSN)	1.00 (0.97–1.04)	7.79E-01
			Subtype 3 (LSN)	0.79 (0.76–0.82)	1.41E-43***
		Private school for ASD	Subtype 1 (HSN)	1.24 (1.18–1.31)	1.11E-15***
			Subtype 2 (EBSN)	0.93 (0.88–1.00)	4.09E-02*
			Subtype 3 (LSN)	0.79 (0.75–0.84)	4.16E-15***
		Speech therapy	Subtype 1 (HSN)	1.30 (1.25–1.35)	6.16E-39***
			Subtype 2 (EBSN)	0.97 (0.93–1.00)	5.19E-02
			Subtype 3 (LSN)	0.78 (0.75–0.81)	1.52E-40***
		Occupational therapy	Subtype 1 (HSN)	1.28 (1.24–1.33)	1.55E-41***
			Subtype 2 (EBSN)	0.99 (0.96–1.02)	6.14E-01
			Subtype 3 (LSN)	0.77 (0.74–0.79)	1.58E-50***
		Physical therapy	Subtype 1 (HSN)	1.24 (1.20–1.28)	8.71E-40***
			Subtype 2 (EBSN)	0.97 (0.93–1.00)	4.42E-02*
			Subtype 3 (LSN)	0.76 (0.74–0.79)	1.25E-51***
		Social skills group	Subtype 1 (HSN)	0.90 (0.87–0.93)	4.26E-10***
			Subtype 2 (EBSN)	1.13 (1.09–1.17)	5.3E-13***
			Subtype 3 (LSN)	1.05 (1.02–1.08)	2.77E-03**
SSC	Medication use	Medications for Attention Deficit Disorder	Subtype 1 (HSN)	0.91 (0.85–0.98)	9.23E-03**
			Subtype 2 (EBSN)	1.21 (1.12–1.31)	1.83E-06***
			Subtype 3 (LSN)	1.03 (0.97–1.10)	3.52E-01
		Antidepressants	Subtype 1 (HSN)	1.08 (1.00–1.18)	5.54E-02
			Subtype 2 (EBSN)	1.18 (1.07–1.30)	9.67E-04***
			Subtype 3 (LSN)	0.86 (0.79–0.93)	1.89E-04***
		Antiepileptics or anti-seizure medications	Subtype 1 (HSN)	1.33 (1.20–1.49)	1.81E-07***
			Subtype 2 (EBSN)	0.93 (0.80–1.07)	3.01E-01
			Subtype 3 (LSN)	0.73 (0.65–0.82)	1.29E-07***
		Sedatives or sleeping pills	Subtype 1 (HSN)	1.31 (1.15–1.49)	5.22E-05***
			Subtype 2 (EBSN)	0.94 (0.79–1.11)	4.45E-01
			Subtype 3 (LSN)	0.72 (0.63–0.83)	2.67E-06***
		Mood stabilizers or anti-psychotics	Subtype 1 (HSN)	1.28 (1.18–1.38)	2.48E-09***
			Subtype 2 (EBSN)	1.08 (0.98–1.19)	1.06E-01
			Subtype 3 (LSN)	0.71 (0.65–0.77)	5.66E-16***

§ *P < 0.05; **P < 0.01; ***P < 0.001.

† Each 0.1-unit increase in the subtype score was associated with OR-fold higher odds of the outcome.

Table 5. Associations of de novo and copy number variant burdens with continuous ASD subtype scores in the SPARK and SSC cohorts.

Associations of rare variant burden with continuous Low-Support-Needs (LSN), Emotional-Behavioral-Support-Needs (EBSN), and High-Support-Needs (HSN) scores were evaluated using separate linear regression models in the SPARK and SSC cohorts. For de novo variants, logit-transformed subtype scores were modeled as the outcomes, with the total number of de novo variants (nDNV) or loss-of-function de novo variants (nLoF) included as the predictor of interest. For copy number variants (CNVs), raw subtype scores were modeled as the outcomes, with total CNV, deletion (DEL), or duplication (DUP) burden included as the predictor of interest. All models were adjusted for sex, age, and the first ten ancestry principal components. Regression coefficients (Beta), standard errors (SE), and P values are reported. Beta represents the change in the corresponding subtype score, on the modeled scale, per additional variant. Analyses were performed separately for each subtype score and variant type. Nominally significant associations ($P < 0.05$) are indicated by an asterisk.

Cohort	Variant type †	Subtype score	Beta	SE	*P* value §
SPARK	DNV	HSN	4.02E-02	1.87E-02	3.21E-02*
		EBSN	-5.69E-03	1.66E-02	7.32E-01
		LSN	-3.39E-02	1.90E-02	7.55E-02
	LoF DNV	HSN	4.96E-02	5.90E-02	4.01E-01
		EBSN	6.39E-02	5.22E-02	2.21E-01
		LSN	-7.07E-02	5.99E-02	2.38E-01
	Total CNV	HSN	3.09E-03	1.25E-03	1.34E-02*
		EBSN	-9.80E-05	1.19E-03	9.34E-01
		LSN	-2.74E-03	1.29E-03	3.34E-02*
	DEL	HSN	2.47E-03	2.07E-03	2.33E-01
		EBSN	3.13E-04	1.97E-03	8.74E-01
		LSN	-1.40E-03	2.14E-03	5.14E-01
	DUP	HSN	4.43E-03	1.77E-03	1.26E-02*
		EBSN	-4.28E-04	1.69E-03	8.00E-01
		LSN	-4.52E-03	1.83E-03	1.37E-02*
SSC	DNV	HSN	-5.16E-03	1.12E-02	6.45E-01
		EBSN	6.80E-03	9.46E-03	4.72E-01
		LSN	2.40E-03	1.18E-02	8.39E-01
	LoF DNV	HSN	2.91E-02	3.23E-02	3.68E-01
		EBSN	4.50E-02	2.73E-02	9.99E-02
		LSN	-3.78E-02	3.41E-02	2.68E-01
	Total CNV	HSN	-7.18E-04	7.62E-04	3.46E-01
		EBSN	1.66E-03	6.45E-04	1.02E-02*
		LSN	-3.93E-04	8.10E-04	6.27E-01
	DEL	HSN	-1.07E-03	8.42E-04	2.03E-01
		EBSN	1.59E-03	7.14E-04	2.61E-02*
		LSN	1.02E-04	8.95E-04	9.10E-01
	DUP	HSN	9.06E-04	1.82E-03	6.19E-01
		EBSN	2.05E-03	1.54E-03	1.84E-01
		LSN	-2.73E-03	1.94E-03	1.59E-01

§ *P < 0.05.

† **DNV**, *de novo* variant; **LoF**, loss-of-function variant; **CNV,** copy number variant; **DEL**, copy number deletion; **DUP**, copy number duplication.