

Supplemental materials**Psychosis spectrum symptoms among individuals with schizophrenia-associated copy number variants and evidence of cerebellar correlates of symptom severity***Table of contents*

Supplemental figures and tables	2-22
Figure S1. Age-stratified prevalence rates of florid or attenuated psychotic symptoms in 3q29Del.....	2
Table S1. Demographic and relevant clinical information for the HC and 22q11.2Del samples and comparison with 3q29Del.....	3
Table S2. Correlations between SIPS symptom ratings and age at visit, stratified by symptom domain and diagnostic group.....	4
Figure S2. Scatter plots of the relationship between SIPS symptom ratings and age at visit, stratified by symptom domain and diagnostic group.....	4
Table S3. Sex-adjusted means and standard errors of SIPS ratings after log-transformation.....	5
Figure S3. Sex-specific SIPS ratings among DS groups and HCs.....	6
Table S4. Demographic, clinical, and MRI-derived volumetric information for the 3q29Del subsample with available neuroimaging and SIPS data.....	7
Figure S4. Comparison of hemisphere-specific ROI volumes in the 3q29Del subsample with available neuroimaging and SIPS data.....	8
Table S5. Extended linear regression results: The effect of cerebellar volumetric measures on domain-specific SIPS symptom ratings in 3q29Del and polynomial modeling of age.....	9-17
Figure S5. Sensitivity analysis: The relationship between cerebellar cortex volume and positive symptom severity in 3q29Del after removal of an extreme data point.....	18
Table S6. Extended logistic regression results: The effect of cerebellar volumetric measures on the probability of a psychotic disorder or APSS diagnosis in 3q29Del and polynomial modeling of age.....	19-21
Table S7. Nested comparison of diagnostic and dimensional phenotypes in 3q29Del participants with versus without posterior fossa arachnoid cyst and mega cisterna magna findings.....	22
Supplemental methods	23-24
Assessment of attenuated and florid psychotic symptoms.....	23
Detailed cognitive assessment protocols.....	23
Extended methods for structural magnetic resonance imaging.....	24
Extended statistical analyses.....	24
List of R packages used for statistical analyses.....	24-25
References for supplemental methods.....	25-26
Supplemental results	26
Sex differences in SIPS ratings within groups.....	26

Supplemental figures and tables

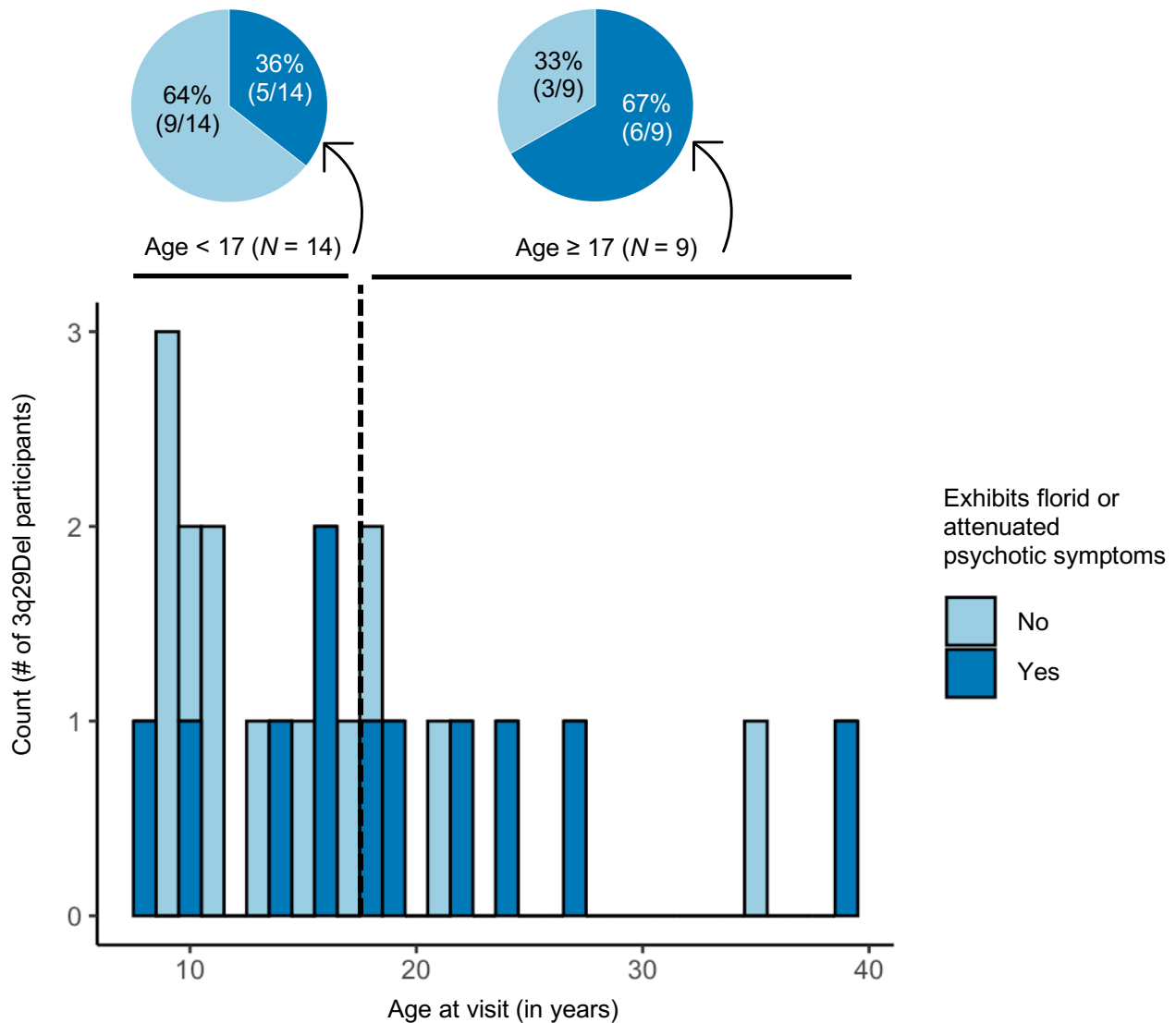


Figure S1. Age-stratified prevalence rates of florid or attenuated psychotic symptoms in 3q29Del. 48% of the 3q29Del sample (11/23) meets diagnostic criteria for psychosis or exhibits one or more clinically significant attenuated positive symptoms (i.e., at least one positive symptom rated three or higher on the SIPS). Based on published demographic data on CHR groups in the general population, the mean age at ascertainment is between 17 and 18 years. Thus, the majority of the 3q29Del participants in the present sample (61%, 14/23) fall below the estimated age for typical onset of both psychosis and the psychosis prodrome (dashed line on histogram marks age 17 years). When stratified by age-group (age < 17 years vs age ≥ 17 years), our results indicate that the prevalence rate of psychosis or clinically significant attenuated positive symptoms among 3q29Del participants younger than age 17 is 36% (5/14), whereas this rate is 67% (6/9) among those aged 17 and older. *Abbreviations:* Deletion, Del; Structured Interview for Psychosis-Risk Syndromes, SIPS; clinical high risk, CHR.

	3q29Del (N = 23)	HC (N = 279)	22q11.2Del (N = 31)	Pairwise comparisons with 3q29Del	
				3q29Del vs. HC	3q29Del vs. 22q11.2Del
Age (in years)					
<i>M</i> ± <i>SD</i>	16.94 ± 8.24	20.08 ± 4.761	19.75 ± 4.12	<i>p</i> -value ^a = 0.003**	<i>p</i> -value ^a = 0.02*
Median	15.89	20.04	19.10		
[Range]	[8.08 – 39.12]	[12.07 – 34.41]	[13.85 – 29.77]		
Sex, N (%)					
Male	14 (61%)	141 (51%)	14 (45%)	<i>p</i> -value ^b = 0.46	<i>p</i> -value ^b = 0.39
Female	9 (39%)	138 (49%)	17 (55%)		
Race, N (%)					
White	20 (87%)	152 (55%)	13 (76%)	<i>p</i> -value ^c = 0.006**	<i>p</i> -value ^c = 0.02*
Black or African American	0 (0%)	48 (17%)	3 (18%)		
Asian	0 (0%)	30 (11%)	1 (6%)		
More than one race	3 (13%)	29 (10%)	0 (0%)		
Other (e.g., Native Hawaiian or Pacific Islander)	0 (0%)	20 (7%)	0 (0%)		
Ethnicity, N (%)					
Hispanic/Latino	1 (4%)	50 (18%)	2 (11%)	<i>p</i> -value ^c = 0.14	<i>p</i> -value ^c = 0.58
Non-Hispanic/Latino	22 (96%)	229 (82%)	17 (89%)		
General cognitive abilities					
<i>M</i> ± <i>SD</i>	73.74 ± 11.98	111.00 ± 14.16	73.71 ± 15.91	<i>p</i> -value ^a = < 0.001***	<i>p</i> -value ^a = 0.88
Median	75	112	71		
[Range]	[46 – 99]	[72 – 139]	[46 – 109]		

Table S1. Demographic and relevant clinical information for the HC and 22q11.2Del samples and comparison with 3q29Del. ^a Wilcoxon rank sum test with continuity correction, ^b Pearson's chi-squared test with Yates' continuity correction. ^c Fisher's exact test (due to smaller subsamples). There were no significant differences in the sex or ethnicity compositions of the 3q29Del vs HC groups, or the 3q29Del vs 22q11.2Del groups (p 's > 0.05). There was a significant age difference between the 3q29Del vs HC groups ($p \leq 0.01$), and the 3q29Del vs 22q11.2Del groups ($p \leq 0.05$), with the 3q29Del group being younger on average. There were significant differences in the race composition of the 3q29Del vs HC groups ($p \leq 0.01$), and the 3q29Del vs 22q11.2Del groups ($p \leq 0.05$), with the 3q29Del group having a larger fraction of participants identifying as "White". There was no significant difference between the 3q29Del vs 22q11.2Del groups in general cognitive ability ($p > 0.05$). As expected, there was a significant difference between the 3q29Del vs HC groups in general cognitive ability ($p \leq 0.001$), with 3q29Del participants exhibiting lower scores than HCs on average. For general cognitive abilities, HC $N = 262$ due to missing data. For race and ethnicity, 22q11.2Del $N = 17$ and $N = 19$, respectively due to missing data. Percentages reflect fraction of participants with complete data. Two-tailed significance levels: p -value ≤ 0.001 '***', p -value ≤ 0.01 '**', p -value ≤ 0.05 '*', p -value ≤ 0.1 '†'. Abbreviations: healthy control, HC; deletion, Del; mean, *M*; standard deviation, *SD*.

SIPS Domains	Pearson's correlation coefficients between age and symptom ratings			Pairwise comparisons of group-specific correlation coefficients	
	HC (N = 279)	22q11.2Del (N = 31)	3q29Del (N = 23)	3q29Del vs. HC	3q29Del vs. 22q11.2Del
Positive Symptom Domain	-0.11 ^a	0.19	0.40 ^a	z = 2.31 p-value = 0.02	z = 0.78 p-value = 0.43
Negative Symptom Domain	-0.15 ^{*a}	0.02	0.37 ^a	z = 2.33 p-value = 0.02	z = 1.26 p-value = 0.21
Disorganization Symptom Domain	-0.08	0.15	0.15	z = 1.00 p-value = 0.32	z = 0.00 p-value = 1.00
General Symptom Domain	-0.03	0.25	0.26	z = 1.28 p-value = 0.20	z = 0.04 p-value = 0.97

Table S2. Correlations between SIPS symptom ratings and age at visit, stratified by symptom domain and diagnostic group. To assess the relationship between age and SIPS symptom ratings, Pearson correlations were calculated between age and SIPS domain scores in each study group, separately. Correlation coefficients were positive and highest in magnitude for 3q29Del, positive and modest in 22q11.2Del, but negative in HCs. This is consistent with the group differences in the mean and range of age. The 3q29Del group is younger and encompasses an age-range that is typically associated with an increase in symptom severity. The 22q11.2Del and HC groups are older, and the age range in HCs extends into young adulthood when psychotic-like experiences typically decline. Given the group variation in sample size, power for detecting significant correlations between age and symptom severity was lowest in 3q29Del, highest for the HCs, and moderate in 22q11.2Del. Fisher's *r*-to-*Z* transformation was subsequently performed to test the significance of between-group differences in these correlation coefficients. The last two columns of the above table present results from these pairwise comparisons in 3q29Del vs HC, and in 3q29Del vs 22q11.2Del. Correlations between age and positive and negative symptom domain ratings showed a significant difference between 3q29Del and HC groups (p 's ≤ 0.05). *Significant group-specific Pearson correlation ($p \leq 0.05$). ^a Significant difference between groups in correlation coefficients, based on Fisher's *r*-to-*Z* transformation ($p \leq 0.05$). For the positive symptom domain total, 22q11.2Del $N = 30$ due to missing data. For the general symptom domain total, HC $N = 278$ due to missing data. *Abbreviations:* Structured Interview for Psychosis-Risk Syndromes, SIPS; healthy controls, HC; deletion, Del.

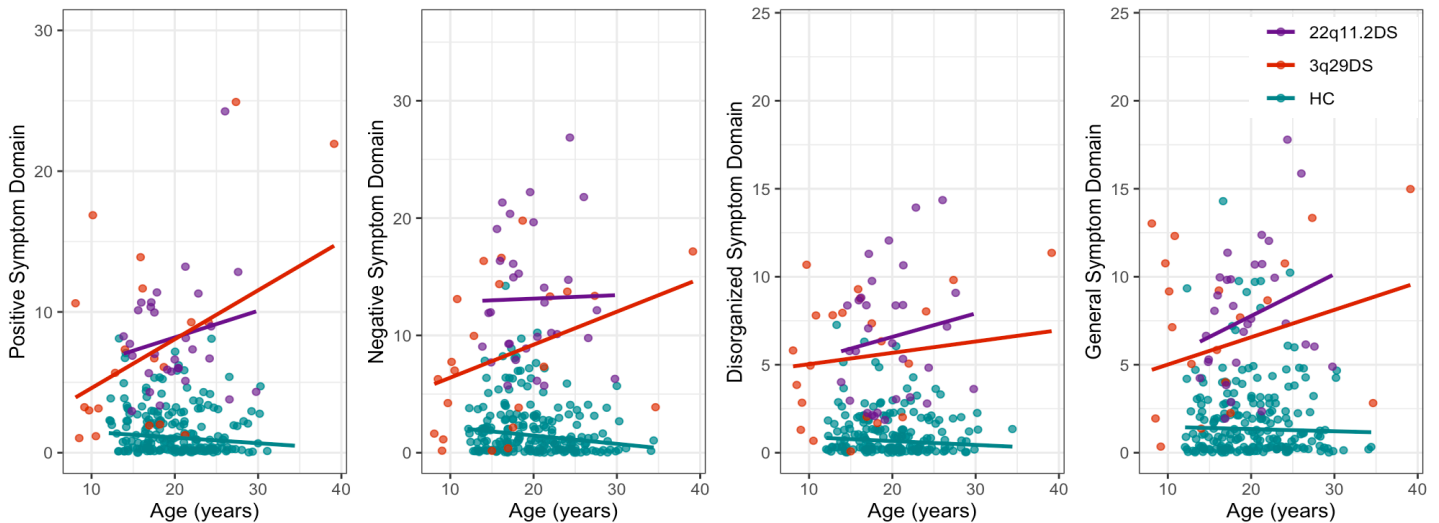
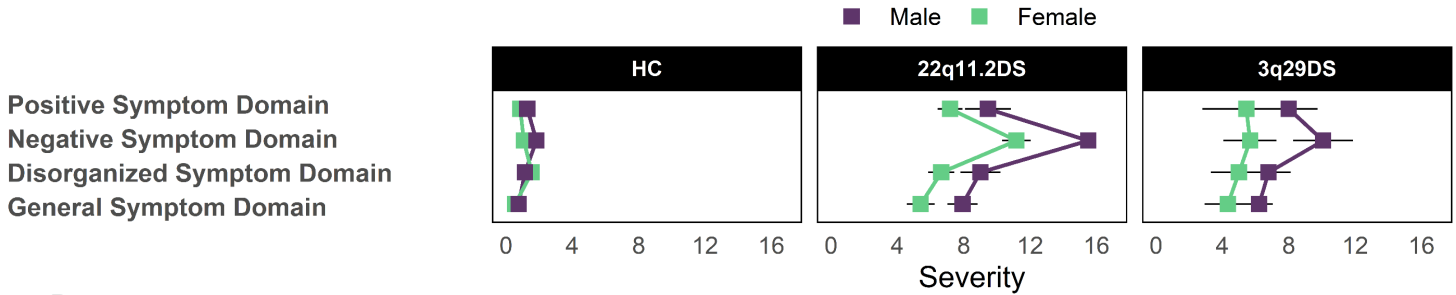


Figure S2. Scatter plots of the relationship between SIPS symptom ratings and age at visit, stratified by symptom domain and diagnostic group. The scatterplots above are provided for visual evaluation of the correlation results reported in Table S2. Solid lines represent the line of linear fit for each group. A slight jitter was systematically added to each plot to minimize overplotting. *Abbreviations:* Structured Interview for Psychosis-Risk Syndromes, SIPS; healthy controls, HC; deletion syndrome, DS.

SIPS symptom domains and items	HC (N = 279)		22q11.2Del (N = 31)		3q29Del (N = 23)	
	Adjusted mean	Std. Error	Adjusted mean	Std. Error	Adjusted mean	Std. Error
Positive Symptom Domain	0.21	0.02	0.93	0.05	0.71	0.06
P1. Unusual Thought Content	0.07	0.01	0.39	0.03	0.26	0.04
P2. Suspiciousness	0.07	0.01	0.39	0.03	0.20	0.03
P3. Grandiosity	0.05	0.01	0.21	0.03	0.16	0.03
P4. Perceptual Abnormalities	0.06	0.01	0.40	0.03	0.36	0.03
P5. Disorganized Communication	0.04	0.01	0.41	0.03	0.37	0.03
Negative Symptom Domain	0.26	0.02	1.12	0.05	0.81	0.06
N1. Social Anhedonia	0.07	0.01	0.56	0.03	0.29	0.03
N2. Avolition	0.07	0.01	0.45	0.03	0.23	0.04
N3. Expression of Emotion	0.03	0.01	0.35	0.03	0.27	0.03
N4. Experience of Emotion and Self	0.02	0.01	0.27	0.02	0.13	0.03
N5. Ideational Richness	0.07	0.01	0.54	0.03	0.40	0.04
N6. Occupational Functioning	0.09	0.01	0.43	0.04	0.35	0.04
Disorganization Symptom Domain	0.15	0.01	0.83	0.04	0.71	0.05
D1. Odd Behavior or Appearance	0.02	0.01	0.34	0.02	0.27	0.03
D2. Bizarre Thinking	0.01	0.01	0.19	0.02	0.18	0.02
D3. Trouble with Focus and Attention	0.11	0.01	0.55	0.03	0.51	0.04
D4. Personal Hygiene	0.02	0.01	0.30	0.02	0.14	0.03
General Symptom Domain	0.57	0.04	2.03	0.13	1.60	0.15
G1. Sleep Disturbance	0.11	0.01	0.37	0.04	0.23	0.04
G2. Dysphoric Mood	0.10	0.01	0.49	0.04	0.24	0.04
G3. Motor Disturbances	0.05	0.01	0.38	0.03	0.37	0.03
G4. Impaired Tolerance to Normal Stress	0.06	0.01	0.37	0.03	0.40	0.04

Table S3. Sex-adjusted means and standard errors of SIPS ratings after log-transformation. The log-transformed sex-adjusted means and standard errors for both DS groups and HCs are reported above. For P5 and the positive symptom domain total, 22q11.2Del $N = 30$ due to missing data from one 22q11.2Del participant. For G2 and the general symptom domain total, HC $N = 278$ due to missing data from one HC. *Abbreviations:* Structured Interview for Psychosis-Risk Syndromes, SIPS; healthy control, HC; deletion, Del; standard, std.

A Domain-wise ratings



B Item-wise ratings

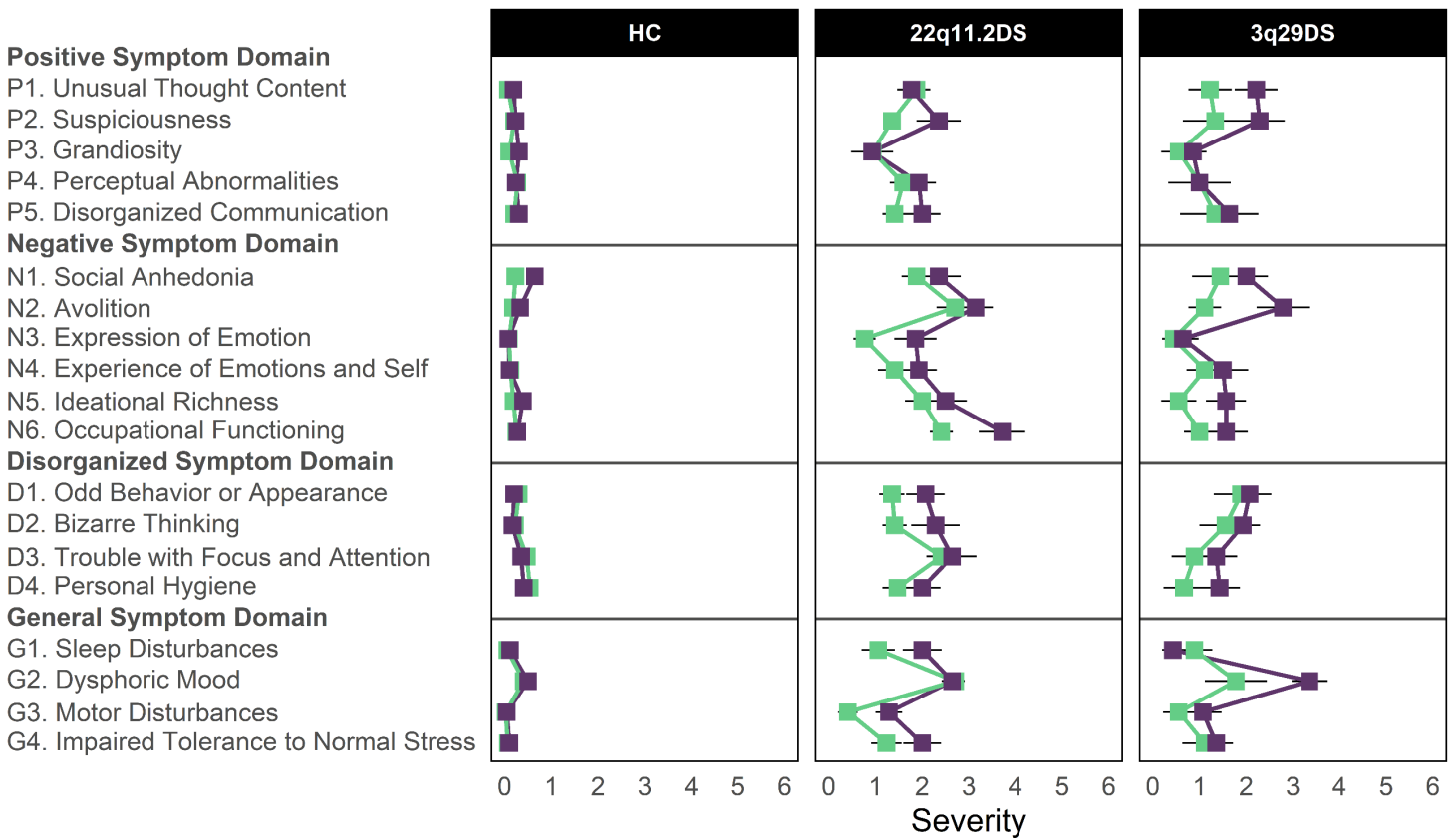


Figure S3. Sex-specific SIPS ratings among DS groups and HCs. Sex differences were investigated in SIPS domain totals (panel **A**) and item-specific SIPS ratings (panel **B**) within each group. Plots reflect the unadjusted means of individual SIPS ratings for each group. Standard error bars are shown. For P5 and the positive symptom domain total, 22q11.2Del $N = 30$ due to missing data from one 22q11.2Del participant. For G2 and the general symptom domain total, HC $N = 278$ due to missing data from one HC. *Abbreviations:* Structured Interview for Psychosis-Risk Syndromes, SIPS; healthy control, HC; deletion, Del; deletion syndrome, DS.

3q29Del sub-sample with available neuroimaging and SIPS data (N = 17)

Demographic characteristics

Age (in years)	
<i>M ± SD</i>	18.13 ± 9.03
Median [Range]	15.89 [8.08 – 39.12]
Sex, N (%)	
Male	10 (59%)
Female	7 (41%)
Race, N (%)	
White	15 (88%)
More than one race	2 (12%)
Ethnicity, N (%)	
Hispanic/Latino	1 (6%)
Non-Hispanic/Latino	16 (94%)

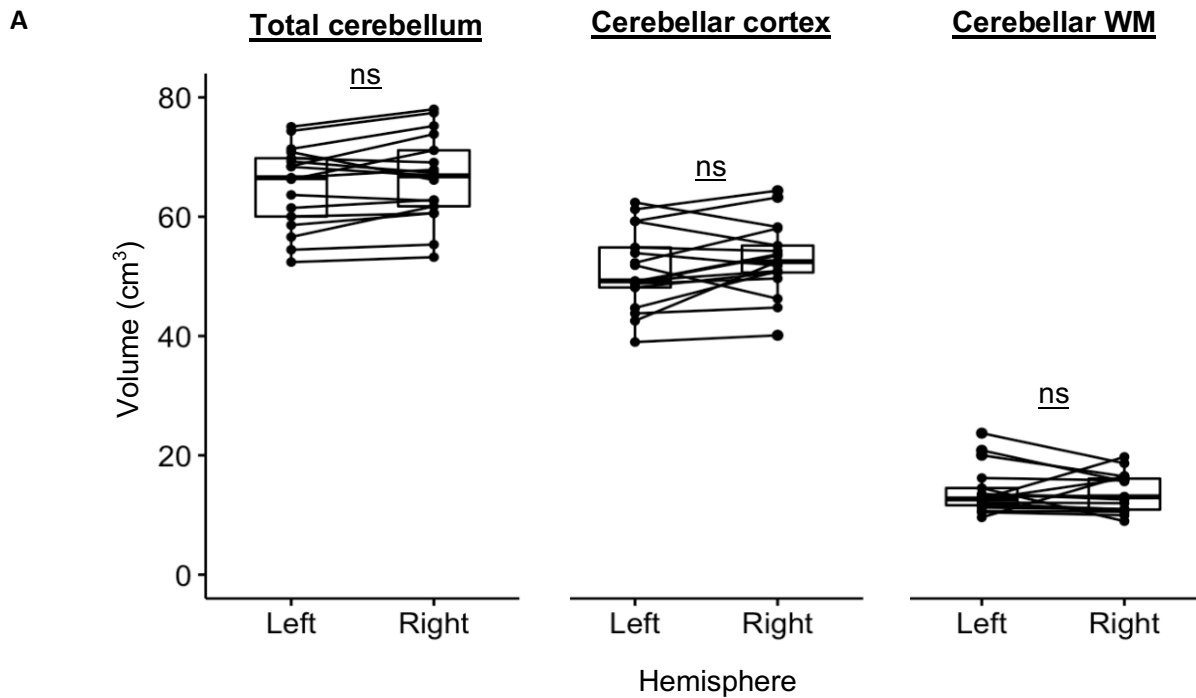
Clinical characteristics

Psychotic disorder dx., N (%)	3 (18%)
Attenuated psychotic symptom syndrome dx., N (%)	3 (18%)
Positive symptom domain total, SIPS	
<i>M ± SD</i>	7.65 ± 7.86
Median [Range]	6.00 [0 – 25]
Negative symptom domain total, SIPS	
<i>M ± SD</i>	8.18 ± 5.53
Median [Range]	7.00 [0 – 17]
Disorganization symptom domain total, SIPS	
<i>M ± SD</i>	5.41 ± 3.81
Median [Range]	5.00 [0 – 11]
General cognitive abilities	
<i>M ± SD</i>	71.94 ± 12.02
Median [Range]	73 [46 – 94]
Antipsychotic usage, N (%)	2 (12%)

Neuroimaging measures

Total cerebellum volume (cm³)	
<i>M ± SD</i>	131.62 ± 13.82
Median [Range]	135.23 [105.61 – 153.09]
Cerebellar cortex volume (cm³)	
<i>M ± SD</i>	103.96 ± 12.21
Median [Range]	102.46 [79.11 – 125.60]
Cerebellar white matter volume (cm³)	
<i>M ± SD</i>	27.66 ± 6.25
Median [Range]	26.18 [20.42 – 42.37]
Estimated total intracranial volume (cm³)	
<i>M ± SD</i>	1384.30 ± 108.26
Median [Range]	1350.52 [1236.90 – 1574.65]

Table S4. Demographic, clinical, and MRI-derived volumetric information for the 3q29Del subsample with available neuroimaging and SIPS data. *Abbreviations:* Deletion, Del; magnetic resonance imaging, MRI; Structured Interview for Psychosis-Risk Syndromes, SIPS; mean, *M*; standard deviation, *SD*; diagnosis, *dx.*



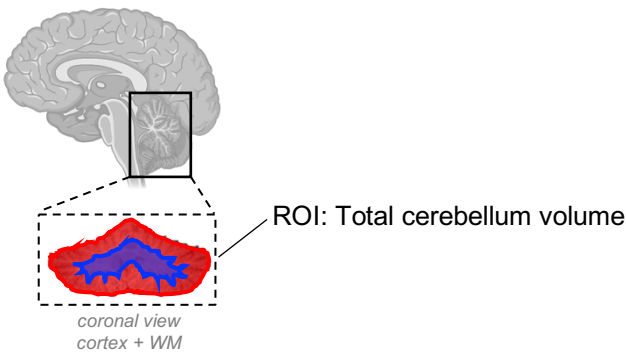
B

ROI (cm ³)	Left Hemisphere	Right Hemisphere	Difference (Left - Right)	<i>p</i> -value
Total cerebellum volume				
Mean ± SD	65.15 ± 6.86	66.46 ± 7.22	-1.31 ± 2.75	0.07
Median (range)	66.54 (52.39 – 75.08)	66.85 (53.22 – 78.01)	-1.35 (-5.35 – 4.70)	
Cerebellar cortex volume				
Mean ± SD	51.14 ± 6.74	52.82 ± 6.13	-1.69 ± 4.11	0.11
Median (range)	49.24 (38.99 – 62.39)	52.47 (40.12 – 64.37)	-1.65 (-9.92 – 5.61)	
Cerebellar WM volume				
Mean ± SD	14.02 ± 3.97	13.64 ± 3.21	0.37 ± 3.60	0.13
Median (range)	12.69 (9.60 – 23.72)	13.05 (8.96 – 19.74)	0.40 (-7.05 – 5.56)	

Figure S4. Comparison of hemisphere-specific ROI volumes in the 3q29Del subsample with available neuroimaging and SIPS data. A) Box plots showing the paired distribution of left and right hemispheric ROI volumes in 3q29Del ($N = 17$). **B)** Descriptive statistics and results of paired-samples tests for inter-hemispheric differences in ROI volumes. In cases where assumptions of paired-samples t-test were violated, the non-parametric Wilcoxon signed-rank test was used. There were no significant volumetric differences between the two hemispheres of any of the ROIs investigated in this study ($p > 0.05$). Hence, left and right hemispheric volumes were added to derive a single bilateral volume for each ROI in downstream analyses. *Abbreviations:* Deletion, Del; Structured Interview for Psychosis-Risk Syndromes, SIPS; magnetic resonance imaging, MRI; region of interest, ROI; white matter, WM; standard deviation, SD; not significant, ns.

Table S5. Extended linear regression results: The effect of cerebellar volumetric measures on domain-specific SIPS symptom ratings in 3q29Del and polynomial modeling of age. Multiple linear regression models include sex, age, age² (when appropriate), and/or eICV (when appropriate) as covariates. Heteroskedasticity-robust Wald tests were performed to sequentially compare simpler models (model 1) to more complex models (model 2) to identify the best fitting polynomial function of age for each ROI. Results indicate that the addition of a quadratic age term in model 2 does not yield a significantly better fit to the data than model 1 (p 's > 0.05); hence we favor parsimony and base our statistical inferences on model 1 (highlighted in blue). In **A-J** heteroskedasticity-robust estimates are reported along with Wald statistics to avoid bias in variance-covariance matrices, given observed violations of statistical assumptions for ordinary least squares regression. In all models, the main effect of cerebellar volume is highlighted in grey for clarity. Schematic illustrations for total cerebellum, cerebellar cortex, and cerebellar white matter are provided below their corresponding regression tables. 3q29Del $N = 17$. *Abbreviations:* 3q29 deletion, 3q29Del; Structured Interview for Psychosis-Risk Syndromes, SIPS; region of interest, ROI; unstandardized coefficient estimate, b ; confidence interval, CI ; degrees of freedom, DF ; standard error, SE ; estimated total intracranial volume, eICV; variance inflation factor, VIF. Contrast coding: reference level for the sex variable is female. Two-tailed significance levels: p -value ≤ 0.001 '***', p -value ≤ 0.01 '**', p -value ≤ 0.05 '*', p -value ≤ 0.1 '†'.

A. Outcome: Positive Symptom Domain Total					
Degree of polynomial	Explanatory variables	b	CI (95%)	SE b	p -value
Linear (Model 1) Best-fit	Intercept	33.62	-11.49 – 78.73	20.88	0.13
	Age (years)	0.26	-0.24 – 0.75	0.23	0.28
	Sex [Male]	8.88	2.48 – 15.27	2.96	0.01**
	Total Cerebellum Volume (cm³)	-0.27	-0.61 – 0.06	0.16	0.10†
	R^2 / R^2 adjusted	0.34 / 0.19			
	Robust Wald test	F-statistic = 4.02 on 3 and 13 DF, p -value = 0.03*			
Quadratic (Model 2)	Intercept	40.16	-8.33 – 88.66	22.26	0.10†
	Age	-0.34	-2.14 – 1.47	0.83	0.69
	Age ²	0.01	-0.03 – 0.06	0.02	0.51
	Sex [Male]	8.54	2.47 – 14.62	2.79	0.01**
	Total Cerebellum Volume (cm³)	-0.28	-0.64 – 0.08	0.16	0.11
	R^2 / R^2 adjusted	0.36 / 0.15			
Robust Wald test	F-statistic = 4.07 on 4 and 12 DF, p -value = 0.03*				



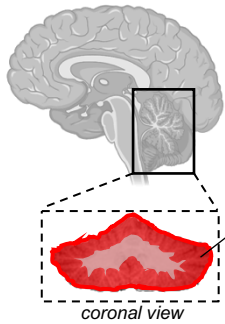
Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

Table S5 continued...

B. Outcome: Positive Symptom Domain Total					
<i>Degree of polynomial</i>	<i>Explanatory variables</i>	<i>b</i>	<i>CI (95%)</i>	<i>SE b</i>	<i>p-value</i>
Linear (Model 1) Best-fit	Intercept	43.39	4.75 – 82.04	17.89	0.03*
	Age (years)	0.18	-0.22 – 0.58	0.18	0.36
	Sex [Male]	9.70	4.21 – 15.19	2.54	0.002**
	Cerebellar Cortex Volume (cm³)	-0.43	-0.78 – -0.07	0.16	0.02*
	R ² / R ² adjusted	0.49 / 0.37			
	Robust Wald test	F-statistic = 11.07 on 3 and 13 DF, <i>p-value</i> < 0.001***			
Quadratic (Model 2)	Intercept	42.69	5.29 – 80.09	17.17	0.03*
	Age	0.49	-1.47 – 2.45	0.90	0.60
	Age ²	-0.01	-0.05 – 0.04	0.02	0.74
	Sex [Male]	10.22	3.08 – 17.35	3.27	0.01**
	Cerebellar Cortex Volume (cm³)	-0.45	-0.87 – -0.03	0.19	0.04*
	R ² / R ² adjusted	0.49 / 0.32			
Robust Wald test	F-statistic = 5.95 on 4 and 12 DF, <i>p-value</i> = 0.01**				



Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

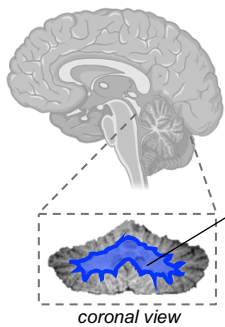
Final model: Model 1

C. Outcome: Positive Symptom Domain Total						
	<i>Explanatory variables</i>	<i>b</i>	<i>CI (95%)</i>	<i>SE b</i>	<i>p-value</i>	<i>VIF</i>
Secondary model with eICV correction	Intercept	70.23	31.12 – 109.34	17.95	0.002**	
	Age (years)	0.21	-0.23 – 0.64	0.20	0.32	1.20
	Sex [Male]	10.18	5.48 – 14.89	2.16	< 0.001***	1.43
	eICV (cm ³)	-0.03	-0.05 – -0.01	0.01	0.01**	1.37
	Cerebellar Cortex Volume (cm³)	-0.29	-0.56 – -0.02	0.12	0.03*	1.99
	R ² / R ² adjusted	0.61 / 0.48				
Robust Wald test	F-statistic = 28.78 on 4 and 12 DF, <i>p-value</i> < 0.001***					

VIF > 5 was treated as indicative of a problematic amount of collinearity.

Table S5 continued...

D. Outcome: Positive Symptom Domain Total					
Degree of polynomial	Explanatory variables	<i>b</i>	CI (95%)	SE <i>b</i>	<i>p</i> -value
Linear (Model 1) Best-fit	Intercept	-8.00	-20.09 – 4.10	5.60	0.18
	Age (years)	0.34	-0.07 – 0.76	0.19	0.10 [†]
	Sex [Male]	3.65	-4.27 – 11.57	3.66	0.34
	Cerebellar White Matter Volume (cm³)	0.27	-0.19 – 0.73	0.21	0.23
	R ² / R ² adjusted	0.26 / 0.09			
Robust Wald test	F-statistic = 11.58 on 3 and 13 DF, <i>p</i> -value < 0.001***				
Quadratic (Model 2)	Intercept	-11.11	-66.04 – 43.82	25.21	0.67
	Age	0.55	-2.90 – 4.01	1.58	0.73
	Age ²	-0.00	-0.08 – 0.07	0.04	0.89
	Sex [Male]	3.64	-4.75 – 12.03	3.85	0.36
	Cerebellar White Matter Volume (cm³)	0.31	-0.70 – 1.33	0.47	0.51
	R ² / R ² adjusted	0.26 / 0.01			
Robust Wald test	F-statistic = 7.10 on 4 and 12 DF, <i>p</i> -value = 0.004**				



ROI: Cerebellar white matter volume

coronal view

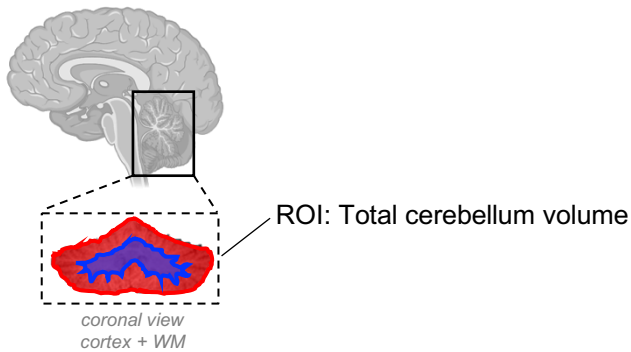
Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

Table S5 continued...

E. Outcome: Negative Symptom Domain Total					
<i>Degree of polynomial</i>	<i>Explanatory variables</i>	<i>b</i>	<i>CI (95%)</i>	<i>SE b</i>	<i>p-value</i>
Linear (Model 1) Best-fit	Intercept	-1.46	-36.03 – 33.11	16.00	0.93
	Age (years)	0.28	0.05 – 0.52	0.11	0.02*
	Sex [Male]	2.76	-4.21 – 9.74	3.23	0.41
	Total Cerebellum Volume (cm³)	0.02	-0.25 – 0.30	0.13	0.86
	R ² / R ² adjusted	0.24 / 0.07			
Robust Wald test	F-statistic = 3.90 on 3 and 13 DF, <i>p-value</i> = 0.03*				
Quadratic (Model 2)	Intercept	-4.68	-42.10 – 32.74	17.17	0.79
	Age	0.57	-0.78 – 1.93	0.62	0.37
	Age ²	-0.01	-0.04 – 0.02	0.01	0.66
	Sex [Male]	2.93	-4.70 – 10.55	3.50	0.42
	Total Cerebellum Volume (cm³)	0.03	-0.25 – 0.30	0.13	0.84
	R ² / R ² adjusted	0.25 / 0.002			
Robust Wald test	F-statistic = 2.39 on 4 and 12 DF, <i>p-value</i> = 0.12				



Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

Table S5 continued...

F. Outcome: Negative Symptom Domain Total					
Degree of polynomial	Explanatory variables	<i>b</i>	CI (95%)	SE <i>b</i>	<i>p</i> -value
Linear (Model 1) Best-fit	Intercept	-2.15	-29.68 – 25.37	12.74	0.87
	Age (years)	0.29	0.04 – 0.54	0.12	0.03*
	Sex [Male]	2.71	-3.78 – 9.19	3.00	0.38
	Cerebellar Cortex Volume (cm³)	0.03	-0.24 – 0.30	0.13	0.79
	R ² / R ² adjusted	0.24 / 0.07			
	Robust Wald test	F-statistic = 3.54 on 3 and 13 DF, <i>p</i> -value = 0.05*			
Quadratic (Model 2)	Intercept	-2.69	-29.62 – 24.24	12.36	0.83
	Age	0.53	-0.97 – 2.03	0.69	0.46
	Age ²	-0.01	-0.04 – 0.03	0.02	0.74
	Sex [Male]	3.10	-4.99 – 11.19	3.71	0.42
	Cerebellar Cortex Volume (cm³)	0.02	-0.29 – 0.33	0.14	0.90
	R ² / R ² adjusted	0.25 / -0.001			
Robust Wald test	F-statistic = 2.12 on 4 and 12 DF, <i>p</i> -value = 0.14				

Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

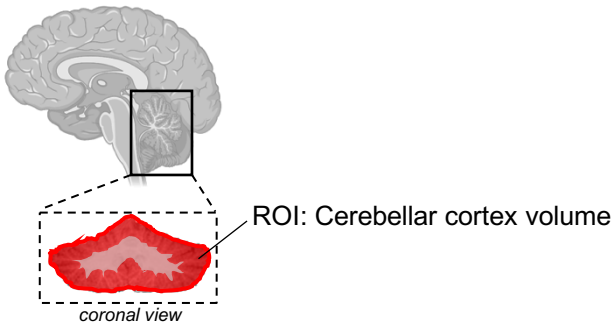


Table S5 continued...

G. Outcome: Negative Symptom Domain Total					
Degree of polynomial	Explanatory variables	<i>b</i>	CI (95%)	SE <i>b</i>	<i>p</i> -value
Linear (Model 1) Best-fit	Intercept	1.95	-8.00 – 11.91	4.61	0.68
	Age (years)	0.28	0.02 – 0.54	0.12	0.04*
	Sex [Male]	3.19	-2.15 – 8.53	2.47	0.22
	Cerebellar White Matter Volume (cm³)	-0.02	-0.40 – 0.36	0.18	0.91
	R ² / R ² adjusted Robust Wald test	0.24 / 0.06 F-statistic = 3.37 on 3 and 13 DF, <i>p</i> -value = 0.05*			
Quadratic (Model 2)	Intercept	-4.62	-42.17 – 32.94	17.24	0.79
	Age	0.72	-1.57 – 3.00	1.05	0.51
	Age ²	-0.01	-0.06 – 0.04	0.02	0.69
	Sex [Male]	3.17	-2.52 – 8.87	2.62	0.25
	Cerebellar White Matter Volume (cm³)	0.08	-0.64 – 0.79	0.33	0.82
	R ² / R ² adjusted Robust Wald test	0.25 / 0.003 F-statistic = 2.14 on 4 and 12 DF, <i>p</i> -value = 0.14			

Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

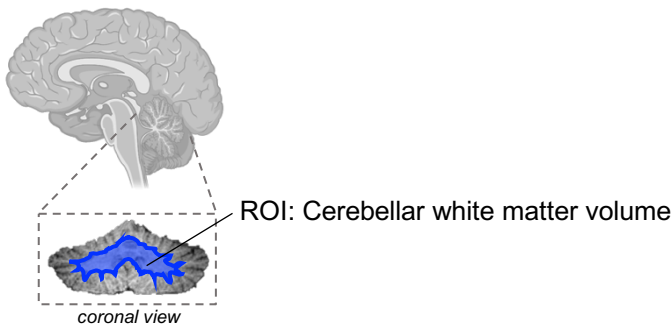
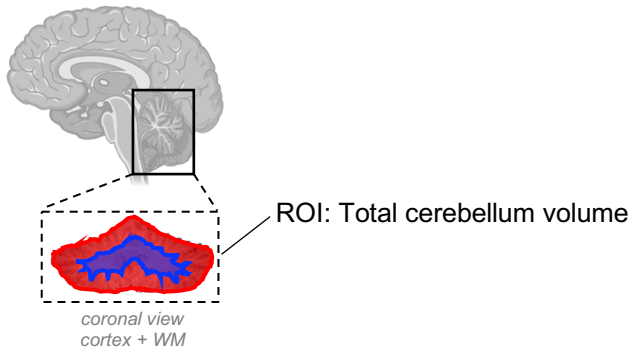


Table S5 continued...

H. Outcome: Disorganization Symptom Domain Total					
Degree of polynomial	Explanatory variables	<i>b</i>	CI (95%)	SE <i>b</i>	<i>p</i> -value
Linear (Model 1) Best-fit	Intercept	-1.89	-26.83 – 23.05	11.54	0.87
	Age (years)	0.08	-0.15 – 0.31	0.11	0.45
	Sex [Male]	1.81	-4.64 – 8.26	2.98	0.55
	Total Cerebellum Volume (cm³)	0.04	-0.18 – 0.25	0.10	0.72
	R ² / R ² adjusted	0.12 / -0.09			
Robust Wald test	F-statistic = 0.70 on 3 and 13 DF, <i>p</i> -value = 0.57				
Quadratic (Model 2)	Intercept	1.80	-23.62 – 27.23	11.67	0.88
	Age	-0.25	-1.35 – 0.85	0.50	0.63
	Age ²	0.01	-0.02 – 0.03	0.01	0.50
	Sex [Male]	1.62	-5.17 – 8.41	3.12	0.61
	Total Cerebellum Volume (cm³)	0.03	-0.19 – 0.25	0.10	0.76
	R ² / R ² adjusted	0.14 / -0.14			
Robust Wald test	F-statistic = 0.91 on 4 and 12 DF, <i>p</i> -value = 0.49				



Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

Table S5 continued...

I. Outcome: Disorganization Symptom Domain Total					
Degree of polynomial	Explanatory variables	<i>b</i>	CI (95%)	SE <i>b</i>	<i>p</i> -value
Linear (Model 1) Best-fit	Intercept	9.57	-10.29 – 29.42	9.19	0.32
	Age (years)	0.04	-0.22 – 0.30	0.12	0.74
	Sex [Male]	3.14	-1.61 – 7.88	2.20	0.18
	Cerebellar Cortex Volume (cm³)	-0.06	-0.25 – 0.12	0.08	0.45
	R ² / R ² adjusted Robust Wald test	0.13 / -0.06 F-statistic = 0.74 on 3 and 13 DF, <i>p</i> -value = 0.55			
Quadratic (Model 2)	Intercept	10.14	-12.14 – 32.41	10.22	0.34
	Age	-0.21	-1.58 – 1.16	0.63	0.74
	Age ²	0.01	-0.02 – 0.04	0.01	0.67
	Sex [Male]	2.72	-3.54 – 8.97	2.87	0.36
	Cerebellar Cortex Volume (cm³)	-0.05	-0.29 – 0.19	0.11	0.68
	R ² / R ² adjusted Robust Wald test	0.15 / -0.13 F-statistic = 1.14 on 4 and 12 DF, <i>p</i> -value = 0.39			

Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

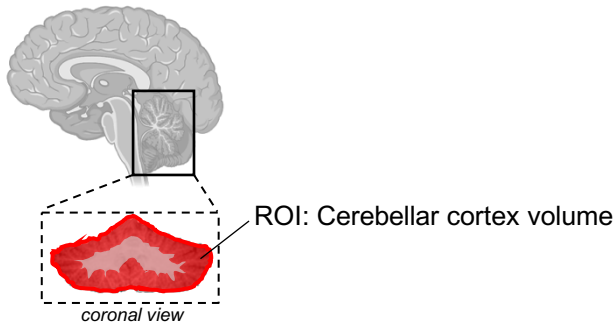
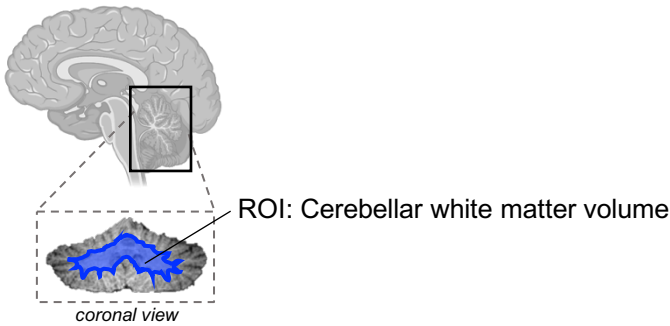


Table S5 continued...

J. Outcome: Disorganization Symptom Domain Total					
Degree of polynomial	Explanatory variables	<i>b</i>	CI (95%)	SE <i>b</i>	<i>p</i> -value
Linear (Model 1) Best-fit	Intercept	-3.96	-10.97 – 3.05	3.24	0.24
	Age (years)	0.05	-0.13 – 0.24	0.09	0.54
	Sex [Male]	1.34	-3.31 – 5.98	2.15	0.54
	Cerebellar White Matter Volume (cm³)	0.28	-0.06 – 0.61	0.15	0.09[†]
	R ² / R ² adjusted Robust Wald test	0.29 / 0.13 F-statistic = 6.19 on 3 and 13 DF, <i>p</i> -value = 0.01**			
Quadratic (Model 2)	Intercept	-11.07	-31.63 – 9.49	9.44	0.26
	Age	0.53	-0.68 – 1.75	0.56	0.36
	Age ²	-0.01	-0.04 – 0.02	0.01	0.40
	Sex [Male]	1.32	-3.53 – 6.17	2.22	0.56
	Cerebellar White Matter Volume (cm³)	0.38	-0.11 – 0.87	0.22	0.11
R ² / R ² adjusted Robust Wald test	0.32 / 0.10 F-statistic = 4.51 on 4 and 12 DF, <i>p</i> -value = 0.02*				

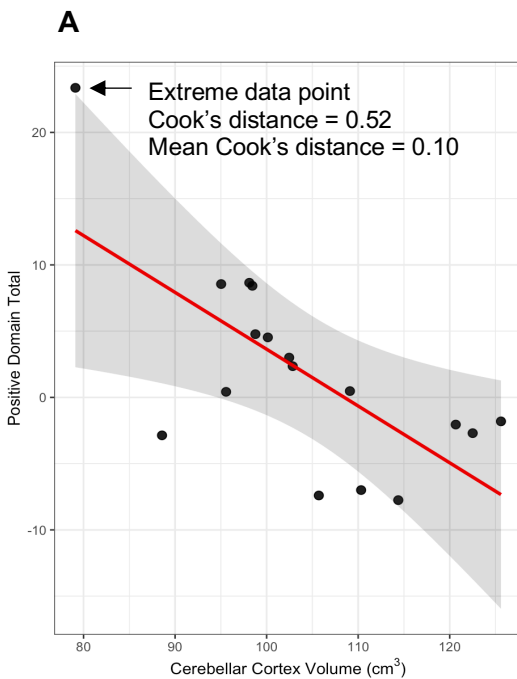


Model 1 vs Model 2 – Robust Wald Test:

	Residual DF	DF	F-value	Pr(>F)
1	13			
2	12	1	0	0.99

Final model: Model 1

Figure S5. Sensitivity analysis: The relationship between cerebellar cortex volume and positive symptom severity in 3q29Del after removal of an extreme data point. **A)** Predictor effect plot showing the relationship between cerebellar cortex volume and SIPS positive symptom ratings (domain total) in $N = 17$ 3q29Del participants. Predicted values of positive symptom ratings (y-axis) were computed from the multiple linear regression model in Table S5B, while age and sex were held fixed. Visual inspection of this figure suggests that one datapoint indicated by the black arrow may have high leverage, although it follows the general trend observed in the remaining data. To compute the influence exerted by this single datapoint on the predicted outcome, we calculated its Cook's distance. The mean Cook's distance in the present sample was 0.10, while the Cook's distance for the interrogated participant was 0.52. As values greater than 4 times the mean Cook's distance of the sample may be classified as influential, we performed a sensitivity analysis by removing this datapoint from the sample. Note that this participant is the only female participant who meets diagnostic criteria for psychosis in the 3q29Del neuroimaging dataset. **B)** Multiple linear regression results in $N = 16$ 3q29Del participants. Heteroskedasticity-robust estimates are reported along with Wald statistics to avoid bias in the variance-covariance matrix, given observed violations of statistical assumptions for ordinary least squares regression. The main effect of cerebellar cortex volume on positive symptom severity is highlighted in grey for clarity. Results of this sensitivity analysis indicate that the coefficient of determination (R^2) remains relatively stable after removal of this extreme datapoint. Although statistical power is reduced due to slightly smaller sample size, there was still a significant inverse relationship between cerebellar cortex volume and positive symptom severity, while correcting for age and sex ($p \leq 0.05$). *Abbreviations:* 3q29 deletion, 3q29Del; Structured Interview for Psychosis-Risk Syndromes, SIPS; unstandardized coefficient estimate, b ; confidence interval, CI ; degrees of freedom, DF ; standard error, SE . Contrast coding: reference level for the categorical sex variable is female. Two-tailed significance levels: p -value ≤ 0.001 '***', p -value ≤ 0.01 '**', p -value ≤ 0.05 '*', p -value ≤ 0.1 '+



B

Regression analysis after removal of one extreme data point:

Outcome: Positive Symptom Domain Total ($N = 16$)

<i>Explanatory variables</i>	<i>b</i>	<i>CI (95%)</i>	<i>SE b</i>	<i>p-value</i>
Intercept	26.91	-1.44 – 55.26	13.01	0.06 [†]
Age (years)	0.15	-0.24 – 0.54	0.18	0.42
Sex [Male]	9.87	3.60 – 16.14	2.88	0.005**
Cerebellar Cortex Volume (cm^3)	-0.28	-0.54 – -0.01	0.12	0.04*
R^2 / R^2 adjusted	0.50 / 0.37			
Robust Wald test	F-statistic = 4.86 on 3 and 12 DF, p -value = 0.02*			

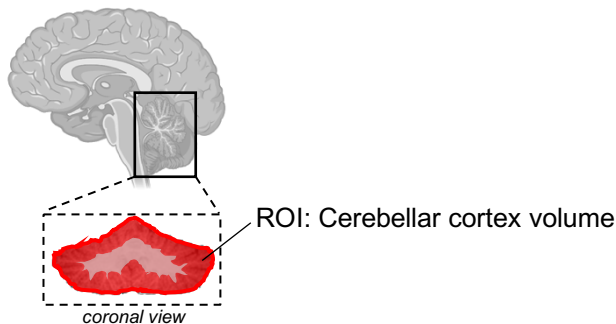
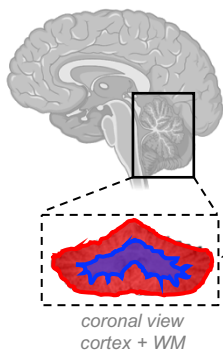


Table S6. Extended logistic regression results: The effect of cerebellar volumetric measures on the probability of a psychotic disorder or APSS diagnosis in 3q29Del and polynomial modeling of age. A-C) Logistic regression models include sex, age, and/or age² (when appropriate) as covariates. Analysis of variance (ANOVA) with likelihood ratio test was performed to sequentially compare simpler models (model 1) to more complex models (model 2) to identify the best fitting polynomial function of age for each ROI. Results indicate that the addition of a quadratic age term in model 2 does not yield a significantly better fit to the data than model 1 in any of our analyses (p 's > 0.05); hence we favor parsimony and base our statistical inferences on model 1 (highlighted in blue). In all models, the main effect of cerebellar volume is highlighted in gray for clarity. 3q29Del $N = 17$. *Abbreviations:* 3q29 deletion, 3q29Del; attenuated psychotic symptom syndrome, APSS; region of interest, ROI; odds ratio, OR; confidence interval, CI; degrees of freedom, DF; standard error, SE. Contrast coding: reference level for the sex variable is female. Two-tailed significance levels: p -value ≤ 0.001 '***', p -value ≤ 0.01 '**', p -value ≤ 0.05 '*', p -value ≤ 0.1 '†'.

A. Binary outcome: Presence / absence of a psychotic disorder or APSS diagnosis					
<i>Degree of polynomial</i>	<i>Explanatory variables</i>	<i>OR</i>	<i>CI (95%)</i>	<i>SE</i>	<i>p-value</i>
Linear (Model 1) Best-fit	Intercept	2.77	0.00 – 4028197.83	18.89	0.88
	Age (years)	1.09	0.96 – 1.29	0.08	0.25
	Sex [Male]	4.34	0.24 – 173.78	6.86	0.35
	Total Cerebellum Volume (cm³)	0.97	0.87 – 1.08	0.05	0.56
	Tjur's R ²	0.14			
Quadratic (Model 2)	Intercept	7.64	0.00 – 67707088.23	57.13	0.79
	Age	0.97	0.50 – 1.90	0.32	0.93
	Age ²	1.00	0.99 – 1.02	0.01	0.73
	Sex [Male]	3.98	0.22 – 162.76	6.29	0.38
	Total Cerebellum Volume (cm³)	0.97	0.87 – 1.08	0.05	0.57
	Tjur's R ²	0.15			



ROI: Total cerebellum volume

Model 1 vs Model 2 – ANOVA with likelihood ratio test:

	Residual DF	Residual Deviance	DF	Deviance	Pr(> Chi)
1	13	19.40			
2	12	19.28	1	0.12	0.83

Final model: Model 1

Table S6 continued...

B. Binary outcome: Presence / absence of a psychotic disorder or APSS diagnosis					
<i>Degree of polynomial</i>	<i>Explanatory variables</i>	<i>OR</i>	<i>CI (95%)</i>	<i>SE</i>	<i>p-value</i>
Linear (Model 1) Best-fit	Intercept	72.82	0.00 – 125831051.98	458.80	0.50
	Age (years)	1.07	0.93 – 1.29	0.08	0.36
	Sex [Male]	6.87	0.43 – 389.88	10.95	0.23
	Cerebellar Cortex Volume (cm³)	0.93	0.81 – 1.04	0.06	0.24
	Tjur's R ²	0.20			
Quadratic (Model 2)	Intercept	67.04	0.00 – 130020106.90	427.82	0.51
	Age	1.10	0.54 – 2.35	0.39	0.78
	Age ²	1.00	0.98 – 1.02	0.01	0.93
	Sex [Male]	7.19	0.38 – 436.11	12.10	0.24
	Cerebellar Cortex Volume (cm³)	0.93	0.80 – 1.05	0.06	0.26
	Tjur's R ²	0.20			

Model 1 vs Model 2 – ANOVA with likelihood ratio test:

	Residual DF	Residual Deviance	DF	Deviance	Pr(> Chi)
1	13	18.24			
2	12	18.24	1	0.01	0.94

Final model: Model 1

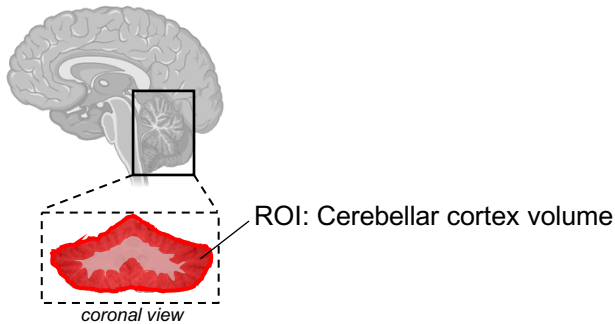


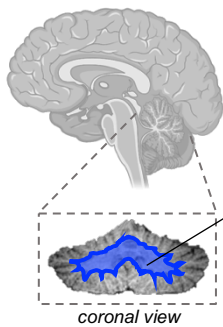
Table S6 continued...

C. Binary outcome: Presence / absence of a psychotic disorder or APSS diagnosis					
<i>Degree of polynomial</i>	<i>Explanatory variables</i>	<i>OR</i>	<i>CI (95%)</i>	<i>SE</i>	<i>p-value</i>
Linear (Model 1) Best-fit	Intercept	0.00	0.00 – 1.12	0.01	0.17
	Age (years)	1.13	0.97 – 1.42	0.11	0.21
	Sex [Male]	2.00	0.17 – 37.92	2.58	0.59
	Cerebellar White Matter Volume (cm³)	1.15	0.92 – 1.57	0.15	0.31
	Tjur's R ²	0.20			
Quadratic (Model 2)	Intercept	0.00	0.00 – 103.47	0.00	0.24
	Age	1.30	0.57 – 3.34	0.56	0.54
	Age ²	1.00	0.98 – 1.02	0.01	0.73
	Sex [Male]	1.90	0.16 – 35.91	2.45	0.62
	Cerebellar White Matter Volume (cm³)	1.17	0.89 – 1.66	0.18	0.28
	Tjur's R ²	0.20			

Model 1 vs Model 2 – ANOVA with likelihood ratio test:

	Residual DF	Residual Deviance	DF	Deviance	Pr(> Chi)
1	13	18.44			
2	12	18.32	1	0.12	0.73

Final model: Model 1



ROI: Cerebellar white matter volume

3q29Del sub-sample with available neuroimaging and SIPS data (N = 17)			
	Participants <u>without</u> PFAC / MCM findings (N = 9)	Participants <u>with</u> PFAC / MCM findings (N = 8)	Test statistics
Diagnostic evaluation of psychotic symptoms			
Psychosis or APSS dx., n (%)			
Absent	6 (66.67%)	5 (62.50%)	OR = 1.19, 95% CI = 0.11 – 13.47, <i>p</i> -value ^a = 1
Present	3 (33.33%)	3 (37.50%)	
Psychosis dx., n (%)			
Absent	6 (66.67%)	8 (100%)	OR = 0, 95% CI = 0.00 – 2.54, <i>p</i> -value ^a = 0.21
Present	3 (33.33%)	0 (0%)	
APSS dx., n (%)			
Absent	9 (100%)	5 (62.50%)	OR = INF, 95% CI = 0.52 – INF, <i>p</i> -value ^a = 0.08 [†]
Present	0 (0%)	3 (37.50%)	
Dimensional evaluation of psychotic symptoms			
SIPS Positive Symptom Domain			
Mean ± SD	7.56 ± 9.53	7.75 ± 6.11	<i>W</i> = 42.00, <i>p</i> -value ^b = 0.60
Median [Range]	3 [0 – 25]	8 [1 – 17]	
SIPS Negative Symptom Domain			
Mean ± SD	7.56 ± 6.21	8.88 ± 4.97	<i>t</i> = 0.48, DF = 15, <i>p</i> -value ^c = 0.64
Median [Range]	6 [0 – 17]	7.5 [2 – 16]	
SIPS Disorganization Symptom Domain			
Mean ± SD	6.00 ± 4.56	4.75 ± 2.92	<i>t</i> = -0.66, DF = 15, <i>p</i> -value ^c = 0.52
Median [Range]	8 [0 – 11]	5 [1 – 9]	

Table S7. Nested comparison of diagnostic and dimensional phenotypes in 3q29Del participants with versus without posterior fossa arachnoid cyst and mega cisterna magna findings. ^a Fisher's exact test, ^b Wilcoxon rank sum test with continuity correction, ^c Student's two sample t-test. Results from diagnostic evaluation of psychotic symptoms indicate that there is no statistically significant difference between 3q29Del participants with versus without PFAC/MCM findings in the prevalence of psychotic disorder and/or APSS diagnoses (*p*'s > 0.05). There was a trend-level association (*p* ≤ 0.1) between the odds of meeting diagnostic criteria for APSS and having a PFAC/MCM finding, which may warrant consideration in future studies with larger sample sizes. Results from dimensional evaluation of psychotic symptoms indicate that there was no statistically significant difference between 3q29Del participants with versus without PFAC/MCM findings in SIPS domain totals for positive, negative or disorganization symptoms (*p*'s > 0.05). Non-parametric test statistics are reported in cases where the data do not meet parametric assumptions. Note that given well-established challenges in differentiating PFACs from MCM using conventional MRI (i.e., shared characteristics in appearance), we considered the prevalence rate of these two neuroanatomical findings jointly. *Abbreviations:* 3q29 deletion, 3q29Del; diagnosis, dx; posterior fossa arachnoid cyst, PFAC; mega cisterna magna, MCM; attenuated psychotic symptom syndrome, APSS; Structured Interview for Psychosis-Risk Syndromes, SIPS; standard deviation, SD; degrees of freedom, DF; odds ratio, OR; infinity, INF. Two-tailed significance levels: *p*-value ≤ 0.001 '***', *p*-value ≤ 0.01 '**', *p*-value ≤ 0.05 '*', *p*-value ≤ 0.1 '†'.

Supplemental methods

Assessment of attenuated and florid psychotic symptoms

In both DS samples and HCs, the SIPS semi-structured clinical interview was administered by trained personnel, with two main aims: 1) to determine the presence of psychosis, present or past, and 2) to assess the severity of attenuated symptoms (Miller et al., 2003). The 19 items in the SIPS are grouped within four domains: positive, negative, disorganization, and general. Each item is rated on a scale from 0 (*Absent*) to 6 (*Severe and Psychotic* for positive symptoms, *Extreme* for others), with a rating of 3 (*Moderate*) indicating clinical significance. The intensity of experience, any associated distress and impairment, and insight (for positive symptoms) were considered for each rating. Item ratings were summed to produce a total score for each domain. The onset and worsening of symptoms were collected for 3q29Del to determine the presence of a psychosis-risk syndrome. The basis of ratings for 3q29Del and 22q11.2Del also relied on the reports of guardians when present. Overall, the instrument maintains excellent inter-rater reliability (Miller et al., 2003) and is comparable to the performance of another valid instrument in the CHR field (Fusar-Poli et al., 2015). For adolescents and adults with 3q29Del meeting criteria for a psychotic disorder, the Structured Clinical Interview for DSM-5 (SCID-5) (First et al., 2015) specified the diagnosis.

All participants provided written consent to participate in each respective project; in cases when the participant was a minor, a parent or legal guardian provided consent and the minor provided assent. The informed consent process and study procedures were in accordance with Emory University Institutional Review Board for the 3q29Del and 22q11.2Del samples. For NAPLS2, the study procedures were approved by the review boards of each individual site, including Emory University.

Cognitive assessment protocols

The 3q29Del sample: Participants who were 18-years-old or older completed the Vocabulary, Similarities, Block Design and Matrix Reasoning subtests of the second edition of the Wechsler Abbreviated Scale of Intelligence (WASI-II) (Wechsler, 2011). The scores from all four subsets were then combined to form a Full Scale Intelligence Quotient (FSIQ) score to estimate general cognitive ability. Participants who were between seven and 17-years-old completed the Verbal Similarities, Word Definitions, Sequential and Quantitative Reasoning and Matrices subtests of the second edition (school-age form) of the Differential Ability Scales (DAS-II) (Elliot, 2007). Younger participants completed the Verbal Comprehension, Naming Vocabulary, Picture Similarities and Matrices subtests of the second edition (early years form) of the Differential Ability Scales (DAS-II) (Elliot, 2007). The scores from all four subsets were then combined to form a General Conceptual Ability (GCA) composite score to estimate general cognitive ability. See Murphy et al. (2018) for detailed study protocol.

The 22q11.2Del sample: Participants who were 17-years-old or older completed the Vocabulary, Arithmetic, Similarities, and Block Design subtests of the third edition of the Wechsler Adult Intelligence Scale (WAIS-III) (Wechsler, 1997). Younger participants completed the equivalent subtests in the third edition of the Wechsler Intelligence Scale for Children (WISC-III) (Wechsler, 1991). As participants completed only a portion of each intelligence scale, an estimation of intellectual functioning hinged upon performance on these subtests. Sattler (2002) identified short-form combinations of individual subtests that could reliably estimate intellectual functioning. The most valid short form for each intelligence scale was selected for this sample from these combinations; see Appendices A and C in Sattler (2002) for reliability and validity coefficients and conversion tables. For the WISC-III, the best performing short form included the sum of Arithmetic, Vocabulary and Block Design ($r = 0.88$), whereas for the WAIS-III the short form included all four administered subtests ($r = 0.92$). The sum of the scaled scores within each combination were used to derive the estimated full scale deviation quotients (Sattler, 2002).

Healthy controls: All participants were administered two subtests from the first edition of the Wechsler Abbreviated Scale of Intelligence (WASI-I): Vocabulary and Block Design (Wechsler, 1999). The four-subtest WASI was developed from the Wechsler intelligence scales in order to briefly approximate FSIQ, as it relies on its own two- and four-subtest short forms. Two-subtest estimates of FSIQ from the WASI strongly correlate with estimates from the WAIS-III among neurotypical samples ($r = 0.87$) (Wechsler, 1999). Performance on these two subtests (Vocabulary and Block Design) were utilized to estimate FSIQ.

Extended methods for structural magnetic resonance imaging

T1- and T2-weighted structural magnetic resonance imaging (MRI) was performed in 17 3q29Del participants with available SIPS data, ages 8.08-39.12 (mean \pm SD = 18.13 \pm 9.03 years), at the Emory University Center for Systems Imaging Core. Images were acquired on a Siemens Magnetom Prisma 3T scanner in the sagittal plane using a 32-channel Prisma head coil with 3D MPRAGE and SPACE sequences, after completion of a training protocol using a mock scanner. To avoid bias and variability associated with hand-drawn measurements, we performed automatic segmentation of acquired images by using FreeSurfer. As recommended, two trained evaluators (ES, LL) inspected the MRI scans to determine whether technical problems (e.g., motion artifacts) or notable pathology (e.g., arachnoid cysts) interfere with segmentation quality; all scans included in this study passed quality control. Note that when compared with manual delineation, segmentations obtained by the FreeSurfer software were found by others to yield well-defined boundaries for the cerebellum comparable in accuracy to manual labeling; a previous quantitative evaluation of robustness and accuracy indicated an excellent dice similarity index, as well as high recall and precision values for cerebellar segmentation results acquired by FreeSurfer (Fischl et al., 2002; Lee et al., 2015). In this framework, cerebellar cortex refers to the tightly folded outer mantle that mostly contains gray matter, while cerebellar white matter refers to the inner core that mostly contains myelinated nerve fibers.

Extended statistical analyses

In between-group analyses, since we observed several violations of the statistical assumptions required for ANCOVA (partially due to a likely floor effect in the HC sample), all SIPS scores were log-transformed to improve normality and homogeneity of variances. Age and sex were considered as potential covariates prior to assessing the variation in psychotic symptom profiles between groups. We ran an exploratory Pearson correlation analysis between age and SIPS domain scores separately by diagnostic group and performed Fisher's r -to- Z transformation to assess the significance of the difference between resulting coefficients. Additionally, we examined sex differences among 3q29Del participants as well as within the comparison groups, using ANCOVAs by symptom domain and then item-wise, while adjusting for age. In sex-specific analyses, assumption violations were observed only in the HC group, hence log-transformation was applied only to HC scores in these analyses. Note that since general symptoms are considered non-specific to psychosis (e.g., impaired tolerance to stress), they were excluded from downstream neuroimaging analyses in the 3q29Del sample.

List of R packages used for statistical analyses

Pearson's chi-squared tests, Fisher's exact tests, t-tests (paired and unpaired) and Wilcoxon signed-rank tests were performed using the standard R *chisq.test()*, *fisher.test()*, *t.test()* and *wilcox.test()* functions. Analyses of variance (ANOVA) and analyses of covariance (ANCOVA) were performed by the standard R *anova()* function. Partial eta-squared was used to assess effect sizes via the *eta_squared()* function from the *effectsize* package (<https://CRAN.R-project.org/package=effectsize>). Pairwise comparisons for 3q29Del vs. HCs and 3q29Del vs. 22q11.2Del were performed to assess the relative severity of attenuated symptoms using the *glht()* function from the *multcomp* package (<https://CRAN.R-project.org/package=multcomp>). The standard R *cor.test()* function was used to perform Pearson's

correlations. Fisher's r -to- Z transformation for exploratory correlation analyses between age and symptom domain scores was performed by using the `cocor.indep.groups()` function from the `cocor` package (<https://CRAN.R-project.org/package=cocor>). The multiple linear regression and binary logistic regression models reported in the neuroimaging arm of the study were performed via the standard R `lm()` and `glm()` functions (family=binomial, link=logit), respectively. Wald statistics were calculated by the `waldtest()` function from the `lmtest` package (<https://CRAN.R-project.org/package=lmtest>). Heteroscedasticity-robust estimates were calculated using the `vcovHC()` function from the `sandwich` package (<https://CRAN.R-project.org/package=sandwich>). Diagnostics plots for linear regression were inspected using the R base function `plot()`. Shapiro-Wilk tests were performed using the standard R `shapiro.test()` function to check for assumptions of normality. Breusch-Pagan and Levene's test were performed using the `bptest()` function from the `lmtest` package (<https://CRAN.R-project.org/package=lmtest>) and the `levene_test()` function from the `rstatix` package (<https://CRAN.R-project.org/package=rstatix>) to check for homogeneity of variances, respectively. Graphics were generated by the `ggplot2` (<https://CRAN.R-project.org/package=ggplot2>), `ggpubr` (<https://cran.r-project.org/package=ggpubr>) and `jtools` (<https://cran.r-project.org/package=jtools>) packages.

References for supplemental methods

Elliott CD (2007): *Manual for the Differential Abilities Scales, 2nd ed.* San Antonio, TX: Harcourt Assessment.

First MB, Williams JB, Karg RS, Spitzer RL (2015): Structured clinical interview for DSM-5—Research version (SCID-5 for DSM-5, research version; SCID-5-RV). Arlington, VA: American Psychiatric Association.1-94.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C *et al.* (2002): Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3): 341-355.

Fusar-Poli P, Cappucciati M, Rutigliano G, Schultze-Lutter F, Bonoldi I, Borgwardt S, *et al.* (2015): At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction. *World Psychiatry*. 14:322-332.

Lee DK, Yoon U, Kwak K, Lee JM (2015): Automated Segmentation of Cerebellum Using Brain Mask and Partial Volume Estimation Map. *Comput Math Methods Med* 2015: 167489.

Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, *et al.* (2003): Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull*. 29:703-715.

Murphy MM, Lindsey Burrell T, Cubells JF, Espana RA, Gambello MJ, Goines KCB, *et al.* (2018): Study protocol for The Emory 3q29 Project: evaluation of neurodevelopmental, psychiatric, and medical symptoms in 3q29 deletion syndrome. *BMC Psychiatry* 18:183.

Sattler JM (2002): *Assessment of children: Cognitive applications.* San Diego, CA: Jerome M. Sattler, Publisher, Inc.

Wechsler D (1991): *Manual for the Wechsler Intelligence Scale for Children-III.* San Antonio, TX: The Psychological Corporation.

Wechsler D (1997): *Manual for the Wechsler Adult Intelligence Scale-III.* San Antonio, TX: The Psychological Corporation.

Wechsler D (1999): *Manual for the Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Wechsler D (2011): *Manual for the Wechsler Abbreviated Scale of Intelligence, 2nd ed.* Bloomington, MN: Pearson.

Supplemental results

Sex differences in SIPS ratings within groups

We examined within-group sex differences for the three diagnostic groups. Statistical analyses included a set of ANCOVAs adjusted for age; values for HCs were log-transformed given assumption violations within this group. As these analyses were conducted to explore whether there were sex differences in symptoms similar to those observed in research on CHR and psychotic patients, correction for multiple comparisons was not made. The unadjusted means and standard errors for each group are presented in Fig. S3.

The differences in domain scores between males and females with 3q29Del were nonsignificant for positive [$F(1,20) = 1.60, p = 0.22$], disorganization [$F(1,20) = 1.90, p = 0.18$], and general [$F(1,20) = 1.13, p = 0.30$] symptoms. Sex did explain variation in negative symptoms [$F(1, 20) = 4.85, p = 0.04, \eta^2_p = 0.20$], as males exhibited more severe symptoms (age-adjusted mean \pm SD = 10.42 ± 5.58) than females (age-adjusted mean \pm SD = 5.12 ± 5.61). Similarly, among the 22q11.2Del group, there was no sex difference in positive [$F(1, 27) = 2.40, p = 0.13$], disorganization [$F(1, 28) = 4.19, p = 0.05$], or general [$F(1, 28) = 3.07, p = 0.09$] symptoms. Only negative symptoms differed by sex [$F(1, 28) = 5.10, p = 0.03, \eta^2_p = 0.15$], with males with 22q11.2Del exhibiting more severe symptoms (age-adjusted mean \pm SD = 15.50 ± 5.31) than females (age-adjusted mean \pm SD = 11.20 ± 5.32). Sex differences among HCs evidenced a similar pattern, with no sex differences for disorganization [$F(1, 276) = 1.21, p = 0.27$], or general [$F(1, 276) = 2.96, p = 0.09$] symptoms, and negative symptoms were rated higher in males (log-transformed age-adjusted mean \pm SD = 0.311 ± 3.08) than females (log-transformed age-adjusted mean \pm SD = 0.21 ± 3.10) [$F(1, 276) = 6.94, p = 0.01, \eta^2_p = 0.02$]. However, sex differences among HCs were observed in positive symptoms [$F(1, 276) = 3.93, p = 0.05, \eta^2_p = 0.01$], with greater severity among males (log-transformed age-adjusted mean \pm SD = 0.25 ± 2.79) versus females (log-transformed age-adjusted mean \pm SD = 0.18 ± 2.78).

Sex differences were also tested in item-wise ratings within each group, while adjusting for age. For 3q29Del, only one item, impaired ideational richness, marginally varied by sex, with males rated higher [$F(1,20) = 4.71, p = 0.04, \eta^2_p = 0.19$]. In contrast, males were rated significantly higher than females with 22q11.2Del on ratings for suspiciousness [$F(1,20) = 4.32, p = 0.05, \eta^2_p = 0.13$], social anhedonia [$F(1,28) = 6.06, p = 0.02, \eta^2_p = 0.18$], experience of emotion and self [$F(1,20) = 5.43, p = 0.03, \eta^2_p = 0.16$], and bizarre thinking [$F(1,20) = 6.22, p = 0.02, \eta^2_p = 0.18$]. Lastly, for HCs, males were rated higher than females on grandiosity [$F(1,276) = 10.27, p = 0.001, \eta^2_p = 0.04$], disorganized communication [$F(1,276) = 6.02, p = 0.01, \eta^2_p = 0.02$], and occupational functioning [$F(1,276) = 7.77, p = 0.014, \eta^2_p = 0.03$].