The Psychosis Spectrum in 22q11.2 Deletion Syndrome Is Comparable to That of Non-Deleted Youths

Supplemental Information

Supplemental Methods and Materials

Sample Selection and Matching

The 22q11.2 deletion syndrome (22q11DS) and non-deleted (ND) groups were not directly ascertained from psychiatric help-seeking populations. However, help-seeking individuals were not excluded. An equal proportion of participants in each group (around two-thirds) had sought psychiatric or psychological help for any reason in their lifetimes, including contacts with psychiatrists, psychologists, and school counselors. Most participants with 22q11DS were recruited through the 22q and You Center at the Children's Hospital of Philadelphia (CHOP). All have molecularly confirmed deletion of the 22q11.2 region; methods included fluorescent in situ hybridization, multiplex ligation-dependent probe amplification, array comparative genomic hybridization, and microarray. Recent publications detail participants' neuropsychiatric phenotype (1-5).

ND controls were drawn from a subset of the Philadelphia Neurodevelopmental Cohort (PNC) selected to return for in-depth phenotyping. The ND group was chosen to have a comparable proportion of individuals categorized as psychosis-spectrum compared to the 22q11DS group in order to enable comparison of psychosis features between the two groups. As further detailed (6), the PNC was recruited from the general pediatric population at CHOP; youths (N=9,498) were first assessed in person by highly trained and supervised clinical research coordinators who administered a structured interview evaluating psychopathology. Threshold and subthreshold psychosis symptoms were assessed with the PRIME Screen-Revised (7), Kiddie-Schedule for Affective Disorders and Schizophrenia (K-SADS) psychosis section (8), and selected questions from the Structured Interview for Prodromal Syndromes (SIPS) (9, 10). Participants were then selected based on the presence or absence of psychosis-spectrum symptoms

(including both threshold and prodromal/subthreshold) to undergo more in-depth follow-up assessments described in the main methods section as well as below (6). Socioeconomic status (SES) was estimated for each household based on the median yearly household income in the subject's zip code, as reported by the American Community Survey (11).

For both samples, exclusion criteria included serious ongoing medical problems, inability to provide assent or informed consent, and moderate to severe intellectual disability. Severity of intellectual disability was evaluated based on clinical IQ testing when available (n=86 of 150 subjects with 22q11DS) and/or the Wide Range Achievement Test 4 reading subtest (12). IQ results reported by participants and parents and checked in clinical records were obtained with the Wechsler Adult Intelligence Scale (forms I, III, and IV), Wechsler Intelligence Scale for Children (forms III and IV), and the Wechsler Preschool and Primary Scale of Intelligence. Formal IQ testing was not incorporated into study procedures due to concerns for overburdening participants; these IQ tests were administered by clinical neuropsychologists, with expertise in assessment of 22q11DS, to whom participants had been referred as a part of separate medical or educational assessments. Individuals with significant intellectual disability (estimated IQ<70) were excluded because they were likely to have little insight into psychiatric phenomena and findings would have limited generalizability to the general population. Study procedures were conducted while the participants were medically stable and ambulatory. No changes were made in the participants' treatment. The Institutional Review Boards of the University of Pennsylvania and CHOP approved all procedures. Informed consent/assent was obtained from each participant and accompanying parent.

Nine participants with 22q11DS and seven ND used antipsychotic medication within six months of the assessment. Of 119 subjects with 22q11DS for whom family history (FH) was available, five subjects likely had a first-degree family member with psychosis, and three subjects had FH that was possibly positive. FH was available for 129 ND subjects, with seven probably positive for psychosis, five possibly positive, and four likely positive for psychosis secondary to medical condition or intoxication. Among subjects with 22q11DS, two subjects are monozygotic twin brothers, two subjects have fathers

with 22q11DS (not enrolled), and one subject has a mother with 22q11DS (not enrolled). ND subjects had no known FH of neurogenetic disorder based on interview and electronic medical records.

Supplemental Figure S1 summarizes the selection of 300 matched subjects. At the time of this analysis, 344 individuals with 22q11DS had been ascertained, with 189 completed quality assurance. Of those, 150 fell within the age range (9-24 years) for which ND controls were available, including 94 individuals with psychosis-spectrum and 56 without psychosis-spectrum. There were 467 assessments of ND individuals available. Matching was conducted using an optimizing algorithm created at Mayo Clinic (13) in SAS programming language (14).

Instruments and Measures

The semi-structured interviews were administered by Bachelor's, Master's and Doctoral level interviewers who underwent formal training conducted by a doctoral level faculty member (MEC) with extensive experience and training in semi-structured interview assessment and diagnosis of psychotic and sub-psychotic symptoms. The K-SADS and SIPS were modified to produce parent versions by substituting third-person pronouns and inserting the term "your child." They were administered as follows: 9-10 year olds received parent interviews alone, 11-17 year olds received both parent and proband interviews, and adults 18-24 years old received only proband interviews. Additional collateral interviews were administered to parents of most adult subjects with 22q11DS, based on their availability. A detailed description and analysis of our administration of the SIPS to participants with 22q11DS was recently published (1).

We report results from twelve tasks of the Penn Computerized Neurocognitive Battery (CNB), grouped into four cognitive domains: (1) Executive Function: Penn Conditional Exclusion Test, Penn Continuous Performance Test, Penn Letter N-Back Test; (2) Episodic Memory: Penn Word Memory Test, Penn Facial Memory Test, Visual Object Learning Test; (3) Complex Cognition: Penn Verbal Reasoning Test, Penn Matrix Reasoning Task, Penn Line Orientation Test; (4) Social Cognition: Penn Emotion Identification Test, Penn Emotion Differentiation Test, Penn Age Differentiation Test (Supplemental

Table S1). Accuracy and speed are recorded for these tasks in addition to 2 others assessing motor speed alone. Here, we present data relating to assessments of accuracy alone. Comparisons were also made with efficiency scores, which are normalized means of accuracy and speed; they did not differ from results reflecting accuracy alone and are not presented in the interest of brevity.

Scoring and Consensus Diagnosis

Following the completed assessments, narrative case summaries were constructed to integrate information from proband and collateral interviews. These were presented and ratings were finalized by consensus from at least two doctoral level clinicians with expertise in psychosis and child psychopathology; consensus ratings were established based on standardized anchors. Determination of participants' status as "psychosis-prone" and "psychotic" also occurred during consensus case conference.

Participants were considered "psychosis-prone" if they met one of the following criteria: [1] One or more clinically significant "positive" symptom rated 3-5 on the SOPS (unusual thought content, suspiciousness/persecutory ideas, grandiose ideas, perceptual abnormalities, disorganized communication); [2] Two or more clinically significant "negative" or "disorganized" symptoms rated 3-6 on the SOPS (social anhedonia, avolition, expression of emotions, experience of emotions and self, ideational richness, occupational functioning, odd behavior or appearance, bizarre thinking, impaired attention, impairment in personal hygiene). "Negative" and "disorganized" symptoms were included at a higher threshold because they have been predictive of conversion to psychosis in the general population, but may be less specific than positive symptoms. Specifically, the North American Prodrome Longitudinal Study (NAPLS) found that disorganized communication and overall disorganized symptoms were significantly higher in participants with clinical high risk (CHR) who later transitioned to psychosis compared to high risk individuals who were either in remission or continued to be at risk but without transitioning to psychosis at their 2-year follow-up (15). Additionally, CHR subjects who converted to psychosis at one year were found to have higher ratings on each of the six negative SOPS items at baseline (16). In a separate measure based on the Canoon-Spoor Premorbid Adjustment Scale, adolescent

social dysfunction was also predictive of conversion to psychosis (17). The majority of psychosis-prone individuals in this sample displayed positive symptomatology (83% of psychosis-prone 22q11DS and 82% of psychosis-prone ND).

Data Analysis and Differential Item Functioning

Continuous clinical and demographic variables compared using Student's t-tests included: age, mean onset, SOPS total, GAF, reading proficiency, and education (Table 1). Fisher's exact test was used for categorical variables: psychosis-spectrum, sex, race, and comorbidity (Table 1 and Figure 3). P-values were adjusted for multiple comparisons using the Holm method (18, 19). Reading proficiency measurements from the Wide Range Achievement Test 4 reading subtest are standardized with respect to age against a large population sample aged 5 – 94 years (mean=100; standard deviation=15; test re-test reliability coefficients=0.78 – 0.90) (12).

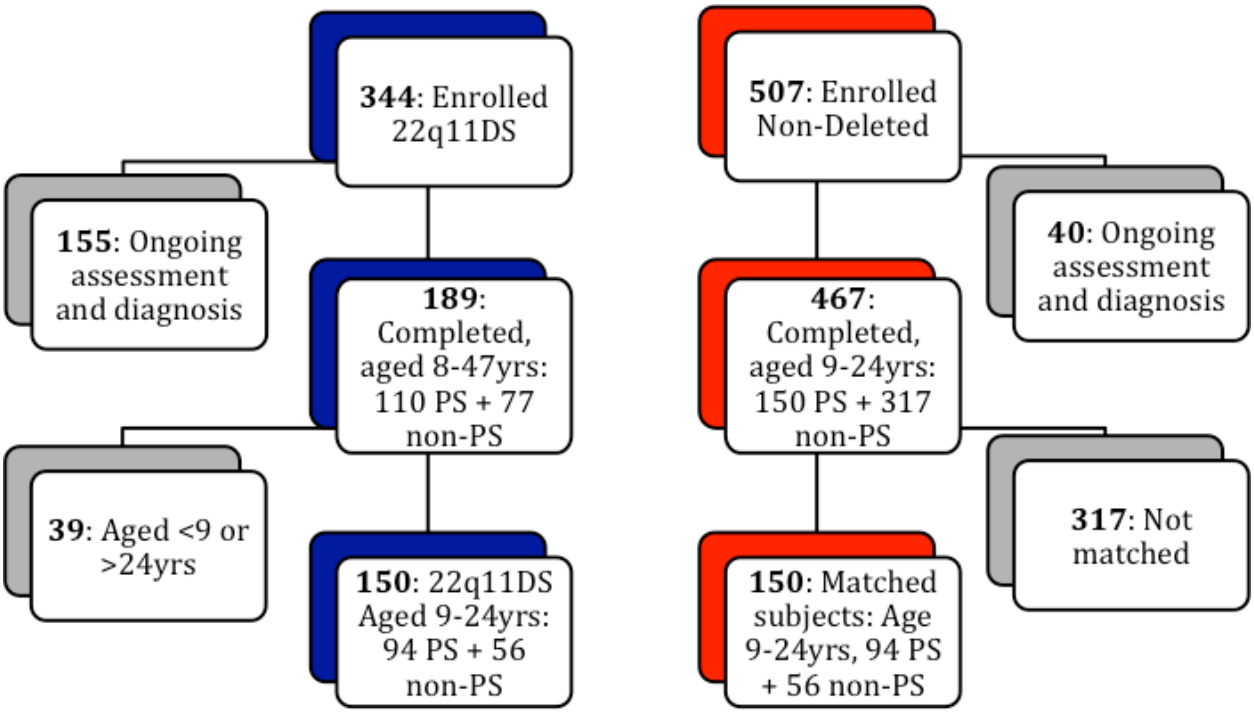
Differential item functioning (DIF), also called "item bias" occurs when two groups of interest (e.g. sex, race, etc.) have different probabilities of endorsing a certain response on that item *even after holding overall rating level constant* (20). A classic example from the 1977 SAT asked test takers to decipher the analogy, "Decoy:Duck," with potential responses: "Net:Butterfly," "Web:Spider," "Lure:Fish," "Lasso:Rope," and "Detour:Shortcut." Dorans and Kulick (1983) found that this item was biased against females because it required knowledge of hunting and fishing, two activities that are traditionally more common for males in most cultures, especially at that time. Thus, for example, if you have a group of males of exactly average ability in analogical reasoning and a group of females also of exactly average ability, the males will still have a higher probability of answering the item correctly (21).

In this study, we apply DIF to address the following: when holding constant the overall burden of psychosis-spectrum symptoms (as reflected by the total number of clinically significant SOPS items), do individuals with 22q11DS differ from ND in the probability of endorsing clinically significant (≥3) symptoms on any SOPS subscales? Scores were dichotomized at the level of clinical significance (0-2 vs. 3-6). All analyses were conducted in R using the *difR* package (18, 22). There are many methods for

detecting DIF (23), and we used four common methods available in most psychometric and structural equation modeling software: [1] Transformed Item Difficulties (24), [2] Mantel-Haenszel (25), [3] Standardization (26), and [4] Logistic Regression (27).

Supplemental Table S2 (below) shows the results of the DIF analyses using the four methods listed above. As is often the case, the four methods do not provide identical results, which is a reason to report multiple methods. For eleven out of thirteen items, there was complete agreement across methods—i.e. either none of them detected DIF, or all did. For the remainder, there was disagreement among the methods, and we opted for a cutoff of at least three methods detecting DIF before making the claim that the item was indeed functioning differently in 22q individuals. These items (bolded in Supplemental Table S2) are p1, p2, n2, n5, d2, and g4. DIF was also calculated for the Caucasian subsample for which DIF in N5 and G4 continued to be significant. Likely due to loss of power, there was no longer significant DIF for p1, p2, n2, or d2.

Statistical softwares used for these analyses included STATA and R, including the *difR* package (18, 22, 28).



Supplemental Figure S1. Selection of 22q11DS and Non-Deleted Subjects – There were 150 subjects with 22q11DS and completed interviews and diagnoses who fell within the age range of ND controls (9 – 24 years). ND subjects were matched to those with 22q11DS on age and sex and enriched for psychosis symptoms to allow for comparison of psychosis features between the two groups. Both 22q11DS and ND groups were recruited from non-help-seeking community or general medical clinic sources; each includes 150 subjects, 94 of which are PS and 56 who are not. *22q11DS*=22q11.2 deletion syndrome; *ND*=non-deleted; *PS*=psychosis-spectrum including individuals with psychosis and psychosis-proneness; *Yrs*=years

Supplemental Table S1: Penn Computerized Neurocognitive Battery

Cognitive Domain	Task	Task Description
Executive Function	Penn Conditional Exclusion Test	Measures abstraction and mental flexibility by assessing ability to detect and adjust to changing rules
	Penn Continuous Performance Test	Measures attention with displays of vertical and horizontal lines with objective of identifying when complete letters and numbers are formed
	Penn Letter N-Back Test	Measures working memory with consecutive displays of numbers and letters with objective of identifying currently present (0-back), previously present (1-back) or from 2 stimuli ago (2-back)
Episodic Memory	Penn Word Memory Test	Twenty words are presented then recognition is assessed when words are represented with distractors
	Penn Facial Memory Test	Twenty faces are presented then recognition is assessed when faces are represented with distractors
	Visual Object Learning Test	Twenty Euclidean shapes are presented then recognition is assessed when shapes are represented with distractors
Complex Cognition	Penn Verbal Reasoning Test	Language-mediated cognition measured with series of analogy problems patterned after Educational Testing Service factor-referenced test kit
	Penn Matrix Reasoning Task	Nonverbal reasoning measured with matrix reasoning problems used in Raven's Progressive Matrices Test and the Matrix Reasoning subscale of the WAIS-III
	Penn Line Orientation Test	Spatial ability measured by presenting lines with different lengths and orientation with objective of rotating one line to match the orientation of the other
Social Cognition	Penn Emotion Identification Test	Assesses ability to identify 5 emotion states (happiness, sadness, anger, fear, neutral) in 40 faces
	Penn Emotion Differentiation Test	Assesses ability to differentiate emotional intensity in 2 faces showing the same emotion
	Penn Age Differentiation Test	Assesses ability to differentiate age in 2 neutral faces morphed to reflect different ages

Note: Cognition was assessed with the Penn Computerized Neurocognitive Battery (CNB), which included 12 tasks assessing 4 cognitive domains. WAIS=Wechsler Adult Intelligence Scale.

Supplemental Table S2: Differential Item Functioning Detection Methods

	DIF-Detection Method				
Item	***T.I.D.***	***M-H***	***Stand.***	***Logistic***	***#DIF***
p1	**DIF**	**DIF**	**DIF**	**DIF**	**4 out of 4**
p2	**NoDIF**	**DIF**	**DIF**	**DIF**	**3 out of 4**
p3	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
p4	NoDIF	NoDIF	NoDIF	DIF	1 out of 4
p5	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
n1	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
n2	**NoDIF**	**DIF**	**DIF**	**DIF**	**3 out of 4**
n3	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
n4	NoDIF	NoDIF	NoDIF	DIF	1 out of 4
n5	**NoDIF**	**DIF**	**DIF**	**DIF**	**3 out of 4**
n6	NoDIF	NoDIF	NoDIF	DIF	1 out of 4
d1	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
d2	**NoDIF**	**DIF**	**DIF**	**DIF**	**3 out of 4**
d3	NoDIF	NoDIF	NoDIF	DIF	1 out of 4
d4	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g1	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g2	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g3	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g4	**DIF**	**DIF**	**DIF**	**DIF**	**4 out of 4**

Note: Four methods for assessing DIF were used to address whether individuals with 22q11DS endorse differently compared to ND when compared at equivalent levels of overall ratings. Results suggest DIF exists for 6 of 19 items when using a threshold of 3 or more tests indicating DIF. These include P1, P2, N2, N5, D2, and G4.

DIF=differential item functioning; T.I.D.=transformed item difficulties; M-H=Mantel-Haenszel; Stand.=standardization; Logistic=logistic regression; p1=unusual thought content/delusional ideas; p2=suspiciousness/persecutory thinking; p3=grandiosity; p4=perceptual abnormalities/hallucinations; p5=disorganized communication; n1=social anhedonia; n2=avolition; n3=expression of emotions; n4=experience of emotions; n5=ideational richness; n6=occupational functioning; d1=odd Behavior or appearance; d2=bizarre thinking; d3=trouble with focus and attention; d4=personal hygiene; g1=sleep disturbances; g2=dysphoric mood; g3=motor disturbances; g4=impaired tolerance to normal stress

Supplemental Table S3: Differential Item Functioning Detection Methods for Caucasian Subsamples

	DIF-Detection Method				
Item	***T.I.D.***	***M-H***	***Stand.***	***Logistic***	***#DIF***
p1	NoDIF	NoDIF	DIF	NoDIF	1 out of 4
p2	NoDIF	NoDIF	DIF	NoDIF	1 out of 4
p3	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
p4	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
p5	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
n1	NoDIF	NoDIF	DIF	NoDIF	1 out of 4
n2	NoDIF	NoDIF	DIF	NoDIF	1 out of 4
n3	NoDIF	NoDIF	NoDIF	DIF	1 out of 4
n4	NoDIF	NoDIF	NoDIF	DIF	1 out of 4
n5	**DIF**	**DIF**	**DIF**	**DIF**	**4 out of 4**
n6	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
d1	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
d2	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
d3	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
d4	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g1	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g2	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g3	NoDIF	NoDIF	NoDIF	NoDIF	0 out of 4
g4	**DIF**	**DIF**	**DIF**	**DIF**	**4 out of 4**

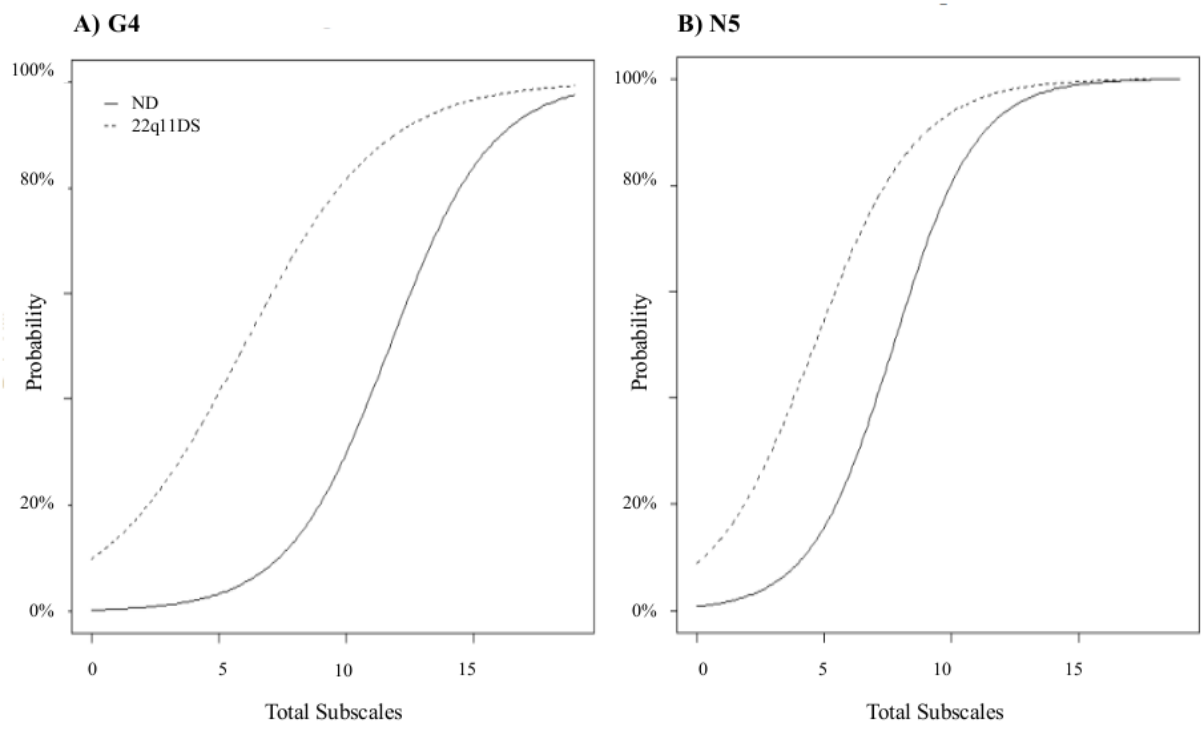
Note: Four methods for assessing DIF were used to address whether Caucasian individuals with 22q11DS endorse differently compared to Caucasian ND when compared at equivalent levels of overall ratings. Results suggest DIF exists for 2 of 19 items (N5 and G4) when using a threshold of 3 or more tests indicating DIF. DIF is no longer significant for N2, P1, P2, or D2 in the Caucasian subsamples. However, non-significant trends are consistent with that of the total sample: individuals with 22q11DS appeared possibly more likely to endorse significant N2, and less likely to endorse P1, P2, and D2 (graphs not included for non-significant trends).

DIF=differential item functioning; T.I.D.=transformed item difficulties; M-H=Mantel-Haenszel; Stand.=standardization; Logistic=logistic regression; p1=unusual thought content/delusional ideas; p2=suspiciousness/persecutory thinking; p3=grandiosity; p4=perceptual abnormalities/hallucinations; p5=disorganized communication; n1=social anhedonia; n2=avolition; n3=expression of emotions; n4=experience of emotions; n5=ideational richness; n6=occupational functioning; d1=odd Behavior or appearance; d2=bizarre thinking; d3=trouble with focus and attention; d4=personal hygiene; g1=sleep disturbances; g2=dysphoric mood; g3=motor disturbances; g4=impaired tolerance to normal stress

Supplemental Table S4: Measures for Caucasian Subsamples

Variable	22q11DS	ND	p	d
n	130	52	-	-
Mean Age (yrs ± SD)	15.5 ± 4.3	15.7 ± 3.8	N.S.	-0.05
Psychosis Spectrum (%)	77 (59%)	23 (44%)	-	-
Psychosis-Prone	67 (52%)	19 (37%)	N.S.	-
Psychosis	10 (8%)	4 (8%)	N.S.	-
Mean Onset (yrs ± SD)	11.6 ± 4.2	11.6 ± 4.2	N.S.	0.00
SIPS Total (score ± SD)	19.8 ± 13.8	11.4 ± 10.0	<0.001	0.66
GAF (score ± SD)	63.3 ± 13.9	76.0 ± 14.0	<0.001	-0.91
Sex				
Male (%)	78 (60%)	35 (67%)	N.S.	-
Female (%)	52 (40%)	17 (33%)	N.S.	-
Race				
Caucasian (%)	130 (100%)	52 (100%)	-	-
Reading Proficiency (±SD)	90.3 ± 13.0	105.9 ± 14.6	<0.001	-1.16
Education (yrs ± SD)				
Proband	8.1 ± 3.6	8.9 ± 3.9	N.S.	-0.22
Mother	15.1 ± 2.3	15.7 ± 2.4	N.S.	-0.26
Estimated income (±SD)	76,655 ± 30,357	74,315 ± 29,078	N.S.	0.08

Note: 22q11DS=22q11.2 deletion syndrome; ND=non-deleted; SOPS=scale of prodromal symptoms; GAF=global assessment of function; SD=standard deviation; Yrs=years; d=Cohen's d for effect size; N.S.=non-significant

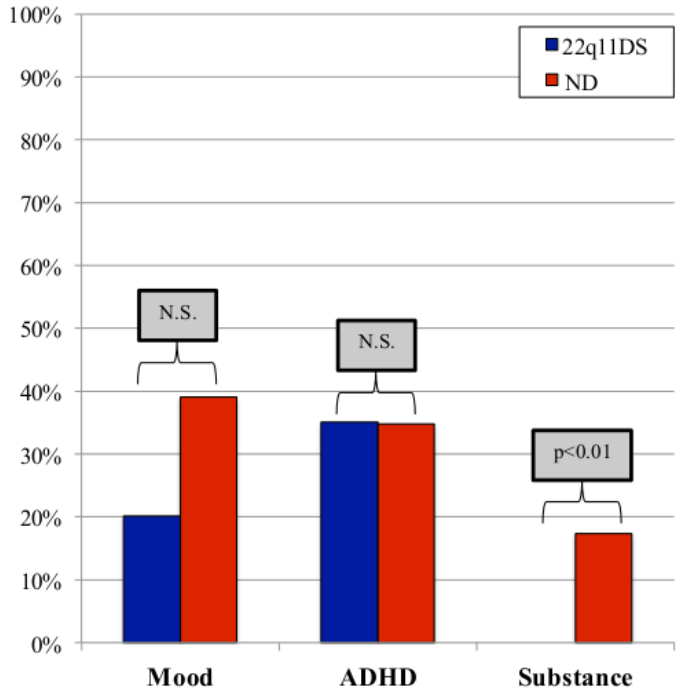


Supplemental Figure S2*:* SOPS Subscales with Significant Differential Item Functioning in Caucasian Subsamples – Compared to Caucasian ND at the same total level of symptomatology, Caucasian individuals with 22q11 are more likely to endorse clinically significant impairment in stress tolerance (G4) and ideational richness (N5). Y-axis illustrates probability of endorsing clinically significant symptoms for that subscale. X-axis represents the number of total clinically significant subscales endorsed by a participant. 22q11DS=22q11.2 deletion syndrome; ND=non-deleted; SOPS=scale of prodromal symptoms

Supplemental Table S5: Cognition and Relationships with Clinical Items

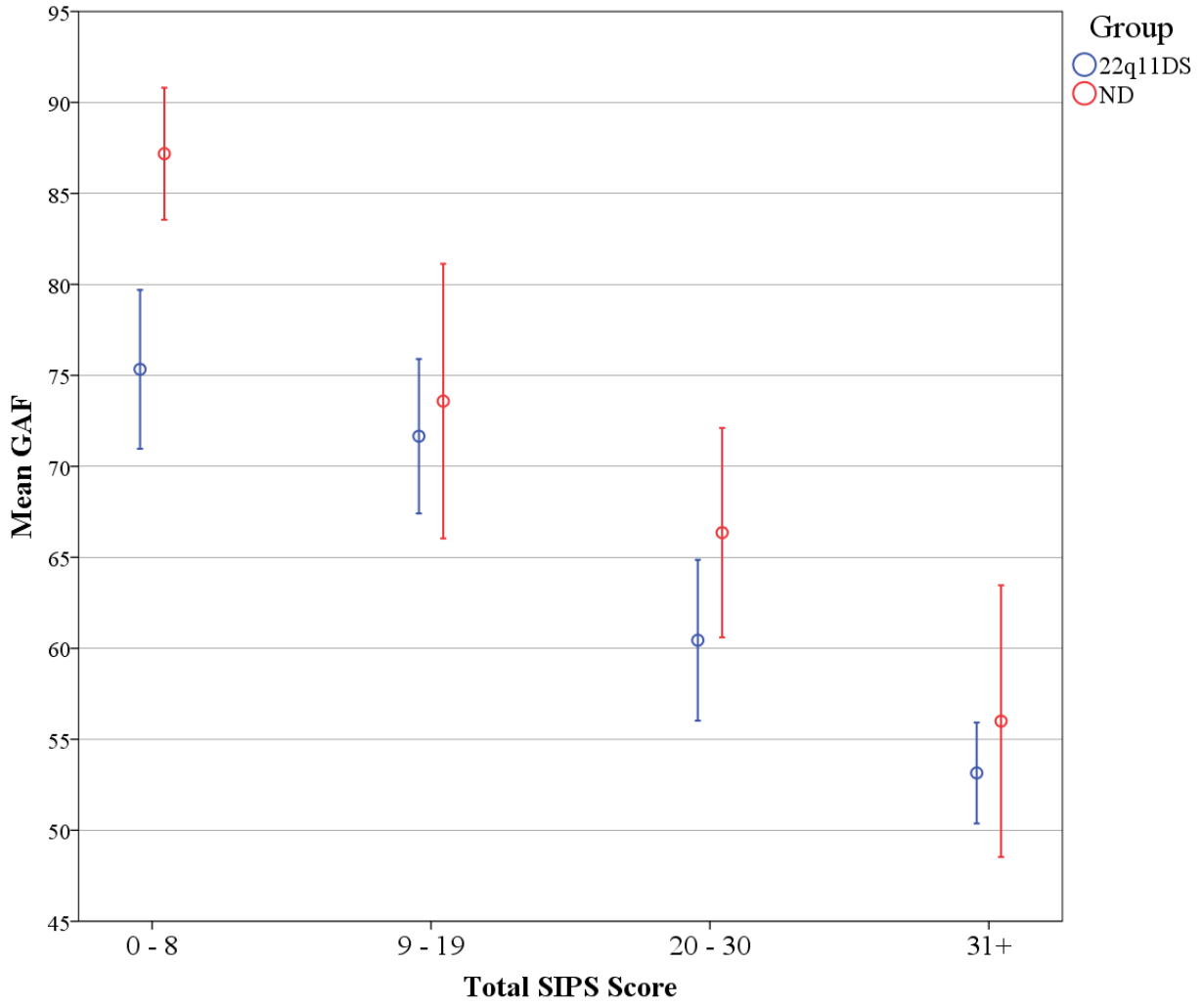
SOPS Item	Cognitive Domain	P-Value	Coefficient
P1	All Domains	0.95	0.01
	Executive Function	0.95	-0.04
	Episodic Memory	1.00	0.00
	Complex Cognition	0.90	0.03
	Social Cognition	0.29	0.12
P2	All Domains	0.95	0.01
	Executive Function	0.95	0.03
	Episodic Memory	1.00	-0.01
	Complex Cognition	0.90	0.01
	Social Cognition	0.29	0.10
N2	All Domains	0.95	-0.02
	Executive Function	0.95	-0.01
	Episodic Memory	1.00	-0.04
	Complex Cognition	0.90	0.08
	Social Cognition	0.48	0.05
N5	All Domains	0.00	-0.24
	Executive Function	0.04	-0.21
	Episodic Memory	0.23	-0.15
	Complex Cognition	0.01	-0.28
	Social Cognition	0.48	-0.05
D2	All Domains	0.95	-0.01
	Executive Function	0.95	-0.04
	Episodic Memory	1.00	0.03
	Complex Cognition	0.90	0.02
	Social Cognition	0.81	0.01
G4	All Domains	0.95	0.00
	Executive Function	0.95	-0.02
	Episodic Memory	1.00	0.02
	Complex Cognition	0.90	0.05
	Social Cognition	0.30	0.08

Note: Relationships between cognition and SOPS items were evaluated using separate multiple linear regressions of the ordinal SOPS item score on each neurocognitive domain, while covarying for group (22q11DS vs. ND). P-values and coefficients for the correlations are reported with coefficients corresponding to the absolute change in SOPS score for each 1 standard deviation change in cognition. P-values were adjusted for multiple comparisons. P1=unusual thought content/delusional ideas; P2=suspiciousness/persecutory thinking; N2=avolition; N5=ideational richness; D2=bizarre thinking; G4=impaired tolerance to normal stress



Supplemental Figure S3: Comorbidities of Psychosis-Spectrum Caucasian Subsamples - Mood disorders and ADHD occur in similar prevalence, but individuals with 22q11DS who are psychosis-spectrum are less likely to have comorbid substance disorders ($p<0.01$). Prevalence of these comorbidities in psychosis-spectrum Caucasian subsamples are as follows: Mood disorders – 20% in 22q11DS vs. 39% in ND; ADHD – 35% in 22q11DS vs. 35% in ND; Substance abuse or dependence – 0% in 22q11DS vs. 17% in ND. Prevalence of comorbidities were compared using two-sided Student's t-test with significance threshold set at $p=0.05$.

22q11DS=22q11.2 deletion syndrome; ND=non-deleted; N.S.=non-significant; Mood=mood disorders including major depression, bipolar disorder, dysthymia, and unspecified depressive and mood disorders; ADHD=attention deficit hyperactivity disorder; Substance=abuse or dependence on alcohol or illicits including hallucinogens, opioids, anxiolytics, and cocaine.



Supplemental Figure S4: Global Assessment of Function and Total SOPS Score for Caucasian Subsamples – Mean GAF is plotted against total SOPS score for Caucasian 22q11DS (blue) and Caucasian ND participants (red). Total SOPS score was segmented into 4 groups, ranging from 0-10, 11-19, 20-30, and 31 and above. In addition, GAF was linearly regressed on total SOPS score, group, and group x total SOPS score for the Caucasian subsamples. A significant interaction was found between group and total SOPS score ($p<0.01$). As illustrated here, it appears that GAF is lower in individuals with 22q11DS, but only for those with low SOPS scores. This is consistent with results in the total sample.

GAF=global assessment of function; SOPS=scale of prodromal symptoms; 22q11DS=22q11.2 deletion syndrome; ND=non-deleted.

Supplemental References

1. Tang SX, Yi JJ, Moore TM, Calkins ME, Kohler CG, Whinna DA, *et al*. (2014a): Subthreshold psychotic symptoms in 22q11.2 deletion syndrome. *J Am Acad Child Adolesc Psychiatry* 53: 991-1000.e2.

2. Tang SX, Yi JJ, Calkins ME, Whinna DA, Kohler CG, Souders MC, *et al*. (2013b): Psychiatric disorders in 22q11.2 deletion syndrome are prevalent but undertreated. *Psychol Med* : 1-11.

3. Gur RE, Yi JJ, McDonald-McGinn DM, Tang SX, Calkins ME, Whinna D, *et al*. (2014c): Neurocognitive development in 22q11.2 deletion syndrome: comparison with youth having developmental delay and medical comorbidities. *Mol Psychiatry* 19: 1205-1211.

4. Yi JJ, Calkins ME, Tang SX, Kohler CG, McDonald-McGinn DM, Zackai EH, *et al*. (2015d): Impact of psychiatric comorbidity and cognitive deficit on function in 22q11.2 deletion syndrome. *J Clin Psychiatry* 76: e1262-e1270.

5. Yi JJ, Tang SX, McDonald-McGinn DM, Calkins ME, Whinna DA, Souders MC, *et al*. (2014e): Contribution of congenital heart disease to neuropsychiatric outcome in school-age children with 22q11.2 deletion syndrome. *Am J Med Genet B Neuropsychiatr Genet* 165: 137-147.

6. Calkins ME, Moore TM, Merikangas KR, Burstein M, Satterthwaite TD, Bilker WB, *et al*. (2014f): The psychosis spectrum in a young US community sample: findings from the Philadelphia Neurodevelopmental Cohort. *World Psychiatry* 13: 296-305.

7. Miller TJ, Cicchetti D, Markovich PJ, McGlashan TH, Woods SW. (2004g): The SIPS screen: a brief self-report screen to detect the schizophrenia prodrome. *Schizophr Res* 70: 78.

8. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, *et al*. (1997h): Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36: 980-988.

9. Miller TJ, McGlashan TH, Rosen JL, Somjee L, Markovich PJ, Stein K, *et al*. (2002i): Prospective diagnosis of the initial prodrome for schizophrenia based on the Structured Interview for Prodromal Syndromes: preliminary evidence of interrater reliability and predictive validity. *Am J Psychiatry* 159: 863-865.

10. Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, *et al*. (2003j): 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.

11. Median Household Income [2006-2010]. Zip Code Characteristics. Retrieved from: http://www.psc.isr.umich.edu/dis/census/Features/tract2zip; Retrieved on January 21, 2016.

12. Wilkinson GS, Robertson GJ. (2006a): Wide Range Achievement Test (WRAT4). *Psychological Assessment Resources, Lutz* .

13. Bergstralh E, Kosanke J. (April, 2014): vmatch.

14. SAS version 9.4. SAS Institute, Carey, NC, USA.

15. Addington J, Liu L, Buchy L, Cadenhead KS, Cannon TD, Cornblatt BA, *et al*. (2015b): North American Prodrome Longitudinal Study (NAPLS 2): The Prodromal Symptoms. *J Nerv Ment Dis* 203: 328-335.

16. Piskulic D, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Heinssen R, *et al*. (2012c): Negative symptoms in individuals at clinical high risk of psychosis. *Psychiatry Res* 196: 220-224.

17. Tarbox SI, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Perkins DO, *et al*. (2013d): Premorbid functional development and conversion to psychosis in clinical high-risk youths. *Dev Psychopathol* 25: 1171-1186.

18. (2014): R. R core team, Vienna, Austria.

19. Holm S. (1979f): A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* : 65-70.

20. Osterlind SJ, Everson HT. (2009b): *Differential item functioning.* Sage Publications, .

21. Dorans NJ, Kulick E. (1983c): Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach. *ETS Research Report Series* 1983: i-14.

22. Magis D, Beland S, Raiche G. (2012f): difR: Collection of methods to detect dichotomous differential item functioning (DIF) in psychometrics. *R package version* 4.

23. Zumbo BD. (2007g): Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language assessment quarterly* 4: 223-233.

24. Angoff WH, Ford SF. (1971h): Item-Race interaction on a test of scholastic aptitude. *ETS Research Bulletin Series* 1971: i-24.

25. Holland PW, Thayer DT. (1988i): Differential item performance and the Mantel-Haenszel procedure. *Test validity* : 129-145.

26. Dorans NJ, Kulick E. (1986j): Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of educational measurement* : 355-368.

27. Swaminathan H, Rogers HJ. (1990k): Detecting differential item functioning using logistic regression procedures. *Journal of Educational measurement* : 361-370.

28. STATA SE version 12.0. Statacorp LP, College Station, TX.