Aus der Klinik für Psychiatrie und Psychotherapie
der Universität München

Direktor: Prof. Dr. med. Peter Falkai

Predictive models in psychiatry: State of the art and future directions investigating cortical folding of the brain

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Rachele Sanfelici
aus Mantova (Italien)

2022

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. med. Nikolaos Koutsouleris
Mitberichterstatter:	Prof. Dr. Michael Ingrisch
	PD Dr. Matthias Brendel
	Prof. Dr. Franziska Dorn
Mitbetreuung durch den promovierten Mitarbeiter:	Dr. Dominic B. Dwyer
Dekan:	Prof. Dr. med. Thomas Gundermann
Tag der mündlichen Prüfung:	21.03.2022

Affidavit

Eidesstattliche Versicherung

Sanfelici, Rachele
Name, Vorname

Ich erkläre hiermit an Eidesstatt, dass ich die vorliegende Dissertation mit dem Titel:

Predictive models in psychiatry: State of the art and future directions investigating cortical folding of the brain

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, den 25.03.2022
Ort, Datum

Rachele Sanfelici
Unterschrift Doktorandin bzw. Doktorand

List of contents

Abbreviations

BAC: Balanced Accuracy

CHR: Clinical High Risk for psychosis

ICD: International Classification of Diseases

MRI: Magnetic Resonance Imaging

NNMF: Non-Negative Matrix Factorization

PSC: Patterns of Structural Covariance

SVM: Support Vector Machine

Publication list

Paper I

Sanfelici R.*, Dwyer D.*, Antonucci, L.A., Koutsouleris N. (2020). **Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view of the state of the art.** Biological Psychiatry, 88(4): 349-360. *These authors contributed equally.

ISI Web of Knowledge: Biological Psychiatry

Impact factor 2020: 13.382
5-year impact factor 2020: 14.101
Ranked 7th of 156 psychiatry journals and 11th of 273 neuroscience journals

Paper II

Sanfelici R., Ruef A., Antonucci LA., Penzel N., Sotiras A., Dong MS., Urquijo-Castro M., Wenzel J., Kambeitz-Ilankovic L., Hettwer MD., Ruhrmann S., Chisholm K., Riecher-Rössler A., Falkai P., Pantelis C., Salokangas RKR., Lencer R., Bertolino A., Kambeitz J., Meisenzahl E., Borgwardt S., Brambilla P., Wood SJ., Upthegrove R., Schultze-Lutter F., Koutsouleris N.*, Dwyer DB.*, and the PRONIA Consortium. (2021). **Novel gyrification networks reveal links with psychiatric risk factors in early illness.** Cerebral Cortex, published online on September 14th 2021. https://doi.org/10.1093/cercor/bhab288. *These authors contributed equally.

ISI Web of Knowledge: Cerebral Cortex

Impact factor 2020: 5.375
5-year impact factor 2020: 6.108
Ranked 68th of 273 neuroscience journals

1. Contribution to the publications

1.1 Contribution to Paper I

In this work we present a systematic review and meta-analysis of diagnostic (i.e., distinguishing CHR from healthy individuals) and prognostic models (prediction of transition to psychosis or functioning) based on machine learning and Cox regression methods. I explored the complex areas of both machine learning and early recognition in psychiatry during my PhD, so that the knowledge I gained on these two topics allowed the production of an informed and informative meta-analysis. For this paper, I conducted an extensive online research of pertinent manuscripts following the PRISMA guidelines[1] using PubMed and Scopus search engines. I thoroughly screened in total 1103 articles following inclusion/exclusion criteria agreed with co-authors. I was responsible for conceptualization of the methodological approach in light of the main aims of our study, i.e.: I) definition of predictive models including not only transition, but also functional outcomes, II) focus on models developed using established machine learning methods, which have a realistic applicability in clinical practice, and III) investigation of models' performance and the potential influence of data modality, algorithm used and validation procedures. I drafted the whole manuscript, was primarily involved in the revision process and finalized the published article.

1.2 Contribution to Paper II

This work has been conducted within the international, large-scale European project PRONIA (www.pronia.eu) carried in 10 European early recognition centres. I worked as a psychologist for the project in the LMU psychiatric clinic—the main coordinating centre of the study. I was directly involved in the recruitment, neuropsychological testing, MRI scanning, interview, evaluation and differential diagnosis of patients with affective and psychosis spectrum disorders. I supervised and conducted follow-up examinations (in total 8 through 3 years for each participant) for around 50 patients and healthy controls. I conducted extensive neuroimaging pipeline testing (CAT12, FreeSurfer) and implementation of MRI quality control techniques in order to establish the most stable methods for brain surface reconstruction both for my project and for the whole consortium. Within a fruitful collaboration with Prof. Sotiras from the USA (Washington University) I learned and implemented a novel multivariate method (e.g., Non-Negative Matrix Factorization[2]) on my sample of study. I executed, under supervision, multiple multi- and univariate analyses on neuroimaging and clinical data from, in total, 1105 individuals from the PRONIA cohort. In parallel, I got

acquainted in the literature on the research field of interest (i.e., gyrifcation in psychiatry), while also collecting evidence on more basic biological mechanisms of cortical folding and disruptions thereof in other neurological pathologies. Furthermore, I was responsible for concepts and hypotheses generation, critical discussion and conclusions driven by the study's results. The manuscript, including tables, figures, supplementary material and full reference list, was entirely written by me and improved thanks to the support of supervisors.

2. Introduction

Psychosis is one of the most burdening psychiatric disorders, as measured by economic loss, morbidity, and mortality worldwide[3]. In the past two decades, the concepts of early recognition, early intervention and precision psychiatry have been introduced to try to detect potential risk pathways and prevent disease development[4]. Cutting-edge methods, such as machine learning, have been of central importance in the enduring attempt to construct personalized prognoses and have led to the development of several risk calculator models. Precision psychiatry needs, however, further basic investigation of potential endophenotypes (i.e., genetic/biological markers of a disease) in order to feed models with informative data for prediction.

To this extent, we present our complementary research based on I) a meta-analysis of the published machine learning-models for prediction in at-risk patients (Paper I), and II) investigation of cortical brain folding, or gyrification, as a potential marker for psychosis development or functional outcome (Appendix) and its broader role in psychopathology (Paper II).

2.1 Early recognition in psychiatry

The Clinical High Risk (CHR) concept describes a clinical condition characterized by sub-threshold psychotic symptoms and cognitive disturbances. This paradigm has facilitated research into the clinical underpinnings of help-seeking individuals potentially at risk for developing psychosis[5]. However, the actual transition rates based solely on the CHR readout have still been particularly low[6,7], suggesting that the symptomatology of risk alone is not able to detect the majority of transitions to the overt disease.

Therefore, research has been trying to discover and understand further biological, clinical and biographical risk factors able to both early detect a predisposition to the disease and also predict its development. For instance, findings have shown that CHR individuals experience more environmental adverse events[8], show hematological alterations[9] and differ from their healthy counterparts in the morphology[10], electrophysiology[11] and resting-state, as well as task-related function of their brain[12]. The complexity of the CHR state calls also for powerful methods, which are able to deal with the high dimensionality of the data at hand and, at the same time, enable a subject-specific risk estimation. To this extent, methodological proceedings have enabled an historical shift of paradigm by introducing machine learning to the field and suggesting a realistic future for personalized predictive psychiatry.

Machine learning is an area of artificial intelligence, which uses advanced algorithms that account for the multivariate structure of large, multimodal datasets (e.g., patients' cognitive, clinical, biological and sociodemographic information) to detect specific patterns, or structure, in the data[13]. Algorithms *learn* these patterns and are tuned to recognize the same structure in new, unseen data, so that models generalize to independent datasets. This multivariate pattern recognition framework can enable both more precise diagnoses (e.g., a classification between psychopathologies or between patients and healthy individuals) and prognoses (e.g., a prediction of disease development or functional outcome). Hence, models constructed with machine learning could be applied in psychiatric care to support clinicians' expertise and help them take critical therapeutic decisions. Research in the past two decades has leveraged the potential of machine learning and has produced a number of predictive models (or risk calculators) for at-risk individuals based on clinical, cognitive, and brain imaging data[14–17] reaching over 80% accuracy. However, still no published model has been applied in real-life clinical practice, mainly because of the still unknown degree of their overall accuracy and reliability.

To clarify the translational potential of the machine learning algorithms, we systematically reviewed and meta-analyzed all available diagnostic and prognostic models for CHR individuals based on machine learning methods (Paper I). Our results showed a relatively good accuracy of models overall and, importantly, a comparable performance between those based on clinical information (e.g., symptoms) and those based on biological information (e.g., brain morphology). Additionally, one important future direction emerging from our study was that further basic research on potential *biomarkers* (i.e., biological signs of risk of disease development) is of central importance to improve models' performance.

One family of biomarkers focuses on structural and functional brain properties, usually analyzed using Magnetic Resonance Imaging (MRI). Structural MRI has already offered the opportunity to detect disruptions in brain volume or density both in first-episode psychosis and at-risk persons[10], and differences between those who develop the disease and those who do not[18]. These findings could be important to promote the use of neurological information as a supplemental diagnostic and prognostic instrument in clinical practice. However, cortical brain volume is known to be influenced by several internal and external confounding factors like drug consumption, antipsychotic medication, plasticity mechanisms or lifestyle characteristics[18,19], potentially shadowing the unique underlying disease effects. As such, more stable measures may be required if predictions from brain MRI measures are to be used in machine learning pipelines.

Gyrification could be such a stable candidate because it is the convolutional property of the human brain cortex, which is known to be mostly genetically driven[20], much less sensitive to external factors and to change during lifetime only slightly[21]. Hence, this morphological measure might be very informative of early neurodevelopmental processes and disruptions thereof, possibly underlying psychiatric diseases or impaired functional outcome. However, further investigation of gyrification is required before it can be used as a potential predictor of disease.

2.2 Gyrification

The cortical folding process is tightly linked to early neurodevelopment because it begins around the third semester of fetal life and peaks at about 2 years post conception[20]. Gyrification is genetically determined[22], and evidence shows that several complex processes play a role in the formation of the individual cortical morphology (e.g., biological and biomechanical forces, as well as anabolic and metabolic processes[23,24]). The importance of these structural cortical differences for human behavior is supported by severe cognitive impairments in gyrencephalic malformations[25] and folding abnormalities in several diseases accompanied by cognitive dysfunctions (e.g., schizophrenia[26], autism[27] or Williams syndrome[28]). The intuitive link between the complexity of the convoluted cortex and cognition has been also validated both in animals (e.g., species with increased gyrification show higher cognitive abilities,[29]) and in humans[30,31].

In mental diseases, gyrification abnormalities have been found in affective and non-affective psychotic syndromes[32,33], depression[34] and even before the first manifestation of psychosis[35]. Some evidence shows that at-risk individuals differ in their gyrification patterns from their healthy counterparts and even that folding aberrations might be predictive of a transition to the overt disease[35]. Nevertheless, results remain inconclusive and inconsistent[36], possibly because of methodological limitations in dealing with a high dimensional data space and the still understudied field of gyrification itself.

One the one hand, traditional statistical methods used to analyze gyrification (e.g., general linear models) are based on assumptions and attempt modelling the data following a-priori hypotheses, thereby potentially overseeing multidimensional and interconnected gyrification patterns. On the other hand, traditional statistics focuses on group-level differences allowing only descriptive conclusions and not testing the single-subject predictive potential of gyrification. Multivariate methods like machine learning enable individual predictions and might be more suited to complex neuroimaging data[13]. The little available

evidence on gyrification-based predictive models shows that cortical folding can predict negative symptoms trajectories[37] and that disorganized folding networks are predictive of psychosis transition in CHR individuals[38]. However, to the best of our knowledge, still no specific investigation of the role of cortical gyrification in prediction of transition to psychosis or of functional outcome based on machine learning exists.

In a first step towards a translational gyrification model, we therefore investigated the hypothesis that gyrification would predict transition or functional outcomes by using machine learning methods in 158 CHR patients (Appendix). Our results showed that gyrification could not predict either outcome category significantly above chance level. These negative findings suggest either I) a further methodological limitation, or II) that cortical folding is not specifically predictive of psychotic episodes, but rather plays a greater role in neurodevelopmental insults influencing psychiatric diseases regardless of diagnostic category.

To disentangle these speculations, we further explored the role of cortical gyrification in psychiatric risk (Paper II) by:

I) using a novel and advanced statistical method that could address the challenges faced when dealing with high dimensional data that were incompletely addressed with standard gyrification pipelines, and

II) focusing on transdiagnostic disease processes in order to determine whether gyrification abnormalities crossed diagnostic boundaries to broadly influence functional outcomes (i.e., as opposed to specifically influencing outcomes in a psychosis risk group).

2.3 Methodological proceedings in gyrification research

The high dimensionality of brain gyrification is usually handled with the use of traditional brain atlases based on coarse anatomical characteristics (e.g., borders between folds and gyri[39,40]), whereby the assumption that folding patterns follow observable surface boundaries must not necessarily be met.

One alternative approach is to shift to an investigation of the cortical structural co-variance. The concept of co-variation of structural brain morphology has been widely recognized in the last two decades and expresses the phenomenon of inter-individual cortical differences co-varying with other, topologically distinct, brain regions[41,42]. Structural covariance is highly heritable, relates to behavioral variation in the population, and is thought to reflect coordinated developmental processes[42]. Seed- and network-based analyses or Principal Component Analyses have been the most popular techniques to investigate structural

covariance, yielding important insights into psychopathologies[43–46]. Only few studies have investigated gyrification structural networks in schizophrenia[47,48] and high-risk populations[38], highlighting the potential of this measure of inter-individual variation to better identify the underpinnings of psychiatric endophenotypes. One equally promising approach to investigate covariance has been newly proposed by Sotiras and colleagues[49], who applied Non-Negative Matrix Factorization (NNMF) in order to detect patterns of structural covariance of cortical thickness in a healthy population. NNMF is an unsupervised multivariate technique, which captures a sparse, parts-based representation of the data[2,50]. This method is particularly useful in the neuroimaging context for two main reasons: first, it is able to aggregate variance in a parcellation-like way, while also accounting for the multivariate nature of cortical features; second, NNMF allows subdividing covariance at different resolutions, which reflects the hierarchical and modular organization of the human brain cortex. Sotiras and colleagues[49] demonstrated the importance of cortical thickness-based covariance for the understanding of healthy coordinated cortical development. Investigating gyrification co-variation might shed light on even earlier developmental mechanisms, potentially reflecting the abnormal maturational processes leading to psychopathology. The solutions generated from the analyses could also be further used in machine learning pipelines in the future.

2.4 Transdiagnostic disease processes

In the last decades, psychiatric care has been evolving towards a more process-based, transdiagnostic approach, as opposed to the traditional diagnose-oriented one[51]. On the one hand, the trans-nosological nature of symptoms and comorbidities has been widely recognized; on the other hand, research has been pointing to common genetic, neurobiological, as well as pathophysiological underpinnings of major psychiatric diseases[52–54]. A transdiagnostic framework might be based on dysfunctions shared across diseases (for instance cognitive or functioning disabilities), which, in turn, might be caused by similar insults during early neurodevelopment. A more in-depth understanding of these risk factors might be of great value for the development of both more precise machine learning models, as well as tailored early transdiagnostic interventions[55].

Gyrification might be especially valuable for this challenge because of its neurodevelopmental nature and because folding abnormalities have been found across several disorders[56]. Nevertheless, transdiagnostic gyrification and its link to putative common disease manifestations—especially in the early phases of disease when diagnostic borders are more subtle—are still highly understudied.

Therefore, we aimed first at investigating data-driven structural covariance patterns of cortical gyrification in a healthy population (N=318) using NNMF, in order to overcome potential methodological limitations in the field. Further, we used a large clinical sample including individuals with a first episode of psychosis, a first episode of depression, and CHR (N=713) to investigate how patterns of gyrification are expressed in psychopathology, and whether they relate to similarities or differences between patients (Paper II). We found that patients' gyrification differed from that of healthy individuals, and that patterns were highly comparable across diagnostic categories. Furthermore, folding abnormalities were linked to commonly disrupted psychological mechanisms such as cognition and global functioning, and not to disease-specific symptoms.

Our results support the hypothesis of neurodevelopmental insults affecting the folding of the cerebral cortex and leading to psychopathological manifestations shared by typically distinct diagnostic categories. This transdiagnostic nature and the lack of associations between gyrification and specific symptoms suggests that gyrification abnormalities might be too unspecific to signal the manifestation of a psychotic episode or functional outcomes after one year, and thus might not be predictive when integrated in machine learning models, as we found in our analyses (Appendix). In fact, psychosis might be caused by more complex interactions of events, including a range of environmental factors[48], that are not captured by cortical folding. Nevertheless, gyrification might add important information within multivariate predictive models (Paper I) by expressing early insults on a neurobiological level, which signal common features of mental illness such as cognitive or functioning impairments—as we demonstrated in our study (Paper II).

In order to successfully build diagnostic and prognostic models, which can be integrated in psychiatric clinical practice, research must thus necessarily further pursue the challenge of understanding the neurobiological mechanisms leading to pathology. A deeper investigation of biomarkers linked to very early neurodevelopmental processes such as gyrification might be very useful to shed light on transdiagnostic features underlying psychiatric diseases and hence contribute to a broader conceptualization of risk in psychiatry.

3. Zusammenfassung:

Das Psychoserisikosyndrom ermöglicht die Untersuchung phänotypischer und mechanistischer Faktoren, die das Risiko junger Menschen beeinflussen eine Psychose zu entwickeln - eine der belastendsten psychiatrischen Erkrankungen weltweit[3]. Die Erforschung von Biomarkern spielt in der Früherkennung von Psychosen eine große Rolle[58]. Biomarker sind biologische/physiologische oder klinische Variablen, die das Risiko eines möglichen Übergangs von einem Psychose-Risiko-Syndrom in eine manifeste Psychose reflektieren oder z.B. mit Änderungen des Funktionsniveaus assoziiert sind. Mit Hilfe von Biomarkern und fortschrittlichen statistischen Methoden, wie Maschinellem Lernen (machine learning)[13], konnten zahlreiche multivariate diagnostische und prädiktive Modelle entwickelt werden, die in Zukunft den klinischen Alltag mittels personalisierter Vorhersagen effizienter gestalten könnten. Um machine learning-Modelle auf die psychiatrische Versorgung zu übertragen, müssen jedoch zwei entscheidende Forschungszweige parallel verfolgt werden: I) Nachweis der Wirksamkeit, Zuverlässigkeit und Replizierbarkeit bestehender prädiktiver Modelle und II) die Suche nach weiteren aussagekräftigen Biomarkern, die in der personalisierten Psychiatrie eingesetzt werden können.

In der vorliegenden Arbeit stellen wir uns dieser Herausforderung, indem wir I) eine systematische Review und Meta-Analyse veröffentlichter diagnostischer und prognostischer Modelle für Psychoserisikosyndrome durchführen (Paper I), II) die Rolle der Hirngyrifizierung als potentiellen Biomarker für Risikopersonen untersuchen (Appendix) und III) die Bedeutung der Gyrifizierung im weiteren Rahmen psychiatrischer Erkrankungen und deren Risiko erforschen (Paper II).

Unsere systematische Review zeigte, dass machine learning-basierte diagnostische und prognostische Modelle für Risikopersonen grundsätzlich eine gute Genauigkeit (67-78% Sensitivität und 77-78% Spezifität) zeigen, unabhängig von den verwendeten Datenmodalitäten oder dem gewählten Algorithmus. Hohe Heterogenität in den Studien und ein Publikationsbias könnten jedoch unsere Ergebnisse beeinflusst haben, so dass eine eindeutige Schlussfolgerung schwer zu ziehen ist.

Um die Rolle der Hirngyrifizierung als möglichen Biomarker in Hochrisikopatienten zu untersuchen, entwickelten wir machine learning Modelle zur Prädiktion des Funktionsniveaus einerseits und der Transition in eine klinisch manifeste Psychose andererseits, welche Ergebnisse knapp über dem Zufallsniveau erreichten (max. ausgeglichene Genauigkeit 53,4%). Dies deutet darauf hin, dass die Rolle der Gyrifizierung nicht spezifisch für das Psychoserisiko ist, sondern mit neurologischen Entwicklungsprozessen

zusammenhängen könnte, die ein breiteres Spektrum psychiatrischer Erkrankungen betreffen.

Um diese Hypothese zu untersuchen, analysierten wir die strukturelle Kovarianz der Gyrifizierung in einer großen transdiagnostischen Patientenpopulation, bestehend aus Patienten mit einer ersten psychotischen Episode, depressiven Patienten und Hochrisikopatienten, im Vergleich mit einer gesunden Kontrollpopulation. Hierbei zeigte sich eine reduzierte Gyrifizierung in der Patientenpopulation, welche mit entwicklungsbedingten Risikofaktoren (Neurokognition und Funktionsfähigkeit) assoziiert war, jedoch nicht mit dem Schweregrad der Symptome korrelierte. Diese Faltungsanomalien könnten somit das Korrelat früher fehlerhafter neuronaler Entwicklungsprozesse sein, die die Vulnerabilität für psychiatrische Erkrankungen erhöhen.

Wie unsere Ergebnisse zeigen, ist der Weg zu belastbaren prognostischen Modellen in der psychiatrischen Diagnostik noch lang und erfordert weitere Grundlagenforschung. Hirnmorphologische Maße wie die Gyrifizierung können ein besseres Verständnis entscheidender Mechanismen neuronaler Entwicklungsprozesse ermöglichen, die einem breiten Spektrum psychiatrischer Erkrankungen zu Grunde liegen.

4. Abstract (English):

The Clinical High Risk (CHR) has enabled research into phenotypic and mechanistic factors highlighting the potential risk for young individuals to develop psychosis—one of the most burdening psychiatric conditions worldwide[3]. Detection of risk for psychosis has been supported by so-called biomarkers, i.e., biological readouts of risk, which could potentially signal a transition to the overt disease or negative functional outcomes[58]. Several multivariate diagnostic or predictive models based on biomarkers have been developed using advanced methods like machine learning[13], producing personalized predictions, which could support everyday clinical decisions. However, in order to translate machine learning models to psychiatric care, two crucial research directions need to be followed in parallel to: I) prove the efficacy, reliability and replicability of existing predictive models, and II) further investigate particularly meaningful biomarkers which can be employed in personalized psychiatry. In the presented work, we pursued this challenge by: I) Conducting a systematic review and meta-analysis of published diagnostic and prognostic models for CHR (Paper I); II) Investigating the potential role of brain gyrification as a biomarker for at-risk individuals (Appendix); and III) Further exploring the significance of gyrification in extended psychiatric etiology and risk (Paper II).

In the meta-analysis, we discovered that machine learning- models for CHR individuals showed relatively good accuracy (67-78% sensitivity and 77-78% specificity) and all models worked equally well, irrespective of data analyzed or algorithm chosen. High heterogeneity throughout studies and a publication bias could have affected our results, so that we could not draw definite conclusions.

Machine learning models constructed on gyrification in at-risk individuals could predict functional outcome or transition to psychosis only slightly above chance level (max. balanced accuracy 53.4%). These results suggested that the role of gyrification in risk might not necessarily be specific, but rather linked to neurodevelopmental processes affecting a wider range of psychiatric diseases (i.e., transdiagnostically).

To investigate the transdiagnostic neurodevelopmental hypothesis, we analyzed gyrification structural covariance in a large transdiagnostic population of first episode psychosis and depression and CHR individuals. Our results revealed reduced gyrification in patients compared to healthy controls, which was associated with developmentally mediated risk factors (i.e., neurocognition and functioning), but not current symptoms. Hence, these cortical folding abnormalities might reflect early neurodevelopmental insults that increase individuals' vulnerability to psychiatric disorders.

Taken together, the road to usable prognostic models for psychiatry is still long and requires further basic research. Brain morphological measures such as gyrification might facilitate a better understanding of crucial neurodevelopmental mechanisms potentially influencing a broader spectrum of psychiatric diseases.

5. Paper I

Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: A meta-analytic view on the state of the art.

Psychosis risk syndromes have been extensively investigated in the past two decades with the aim of predicting and possibly preventing transition to the overt disorder in help-seeking individuals. Novel statistical methods like machine learning and Cox proportional hazard regression have been crucial to develop personalized models able to diagnose risk for psychosis and predict a future outcome in these individuals based on different data modalities (e.g., neurocognitive or neuromorphological characteristics). However, despite their great potential, these models have still not been translated into clinical practice.

To shed light on the current state of published machine learning- and Cox regression-based diagnostic and prognostic models, we thoroughly reviewed the literature and conducted a meta-analysis on accuracy performances. We investigated different methodological approaches and data modalities, specifically focusing on performance differences between clinical (i.e., based on symptoms, cognition and environmental factors) and biological models (i.e., constructed on brain morphology and function).

We selected 44 articles, including in total 3707 individuals for prognostic and 1052 for diagnostic studies. Psychosis risk syndromes could be relatively accurate diagnosed (78% sensitivity and 77% specificity), while prognostic models reached overall a sensitivity and specificity of 67% and 78%, respectively. Machine learning models gained a 10% higher sensitivity compared to those using Cox regression, however validation techniques also vastly differed between the two approaches. These results were not moderated by the type of data modality, the algorithm used, or the at-risk population studied. Importantly, we detected a publication bias for prognostic studies, which points to inflated results reported by studies with smaller sample sizes.

Our results showed comparable performance between clinical and biological models, which calls for improvement in basic research on brain markers of disease. Further, findings may be affected by I) heterogeneity in the field, including definitions of clinical populations, data domains and machine learning algorithms used, and II) degree of methodological validity, reliability and generalizability. These factors might hinder the translation of diagnostic and prognostic models to clinical practice and need to be thoroughly taken into consideration in future research.

Archival Report

Biological Psychiatry

Individualized Diagnostic and Prognostic Models for Patients With Psychosis Risk Syndromes: A Meta-analytic View on the State of the Art

Rachele Sanfelici, Dominic B. Dwyer, Linda A. Antonucci, and Nikolaos Koutsouleris

ABSTRACT

BACKGROUND: The clinical high risk (CHR) paradigm has facilitated research into the underpinnings of help-seeking individuals at risk for developing psychosis, aiming at predicting and possibly preventing transition to the overt disorder. Statistical methods such as machine learning and Cox regression have provided the methodological basis for this research by enabling the construction of diagnostic models (i.e., distinguishing CHR individuals from healthy individuals) and prognostic models (i.e., predicting a future outcome) based on different data modalities, including clinical, neurocognitive, and neurobiological data. However, their translation to clinical practice is still hindered by the high heterogeneity of both CHR populations and methodologies applied.

METHODS: We systematically reviewed the literature on diagnostic and prognostic models built on Cox regression and machine learning. Furthermore, we conducted a meta-analysis on prediction performances investigating heterogeneity of methodological approaches and data modality.

RESULTS: A total of 44 articles were included, covering 3707 individuals for prognostic studies and 1052 individuals for diagnostic studies (572 CHR patients and 480 healthy control subjects). CHR patients could be classified against healthy control subjects with 78% sensitivity and 77% specificity. Across prognostic models, sensitivity reached 67% and specificity reached 78%. Machine learning models outperformed those applying Cox regression by 10% sensitivity. There was a publication bias for prognostic studies yet no other moderator effects.

CONCLUSIONS: Our results may be driven by substantial clinical and methodological heterogeneity currently affecting several aspects of the CHR field and limiting the clinical implementability of the proposed models. We discuss conceptual and methodological harmonization strategies to facilitate more reliable and generalizable models for future clinical practice.

Keywords: Biomarkers, Clinical psychobiology, Machine learning, Predictive psychiatry, Psychosis, Translational medicine

https://doi.org/10.1016/j.biopsych.2020.02.009

Psychotic disorders are among the most disabling mental illnesses and represent one of the top 20% causes of socioeconomic burden worldwide (1). Therefore, psychiatric research has substantially invested in better early detection strategies for these disorders (2). The clinical high risk (CHR) concept (3) describes a mental state characterized by subthreshold psychotic symptoms that differ quantitatively in their intensity from those of a full-blown psychosis (Supplement and Table 1). The CHR paradigm has become a well-established clinical avenue to early detect and potentially treat the psychosis high-risk states. Based on the CHR paradigm, researchers have investigated the nature of the prepsychotic phase from both pathophysiological and epidemiological perspectives (4,5). However, these efforts have been challenged by a constantly declining incidence rate of psychosis among CHR patients (4,6), with roughly one third of not-transitioned CHR cases still experiencing subthreshold symptoms, psychosocial impairments (7), and lower level of quality of life (8). Thus, the CHR designation delineates a mental condition that is burdensome per se and, in addition, is associated with a known set of comorbidities (e.g., depression, substance abuse, anxiety disorders) (9). Therefore, predictive psychiatry has gradually broadened its scope from detecting disease transition to encompassing adverse outcomes more broadly [e.g., functional deficits (10), treatment response (11), persisting negative symptoms (12), psychiatric comorbidities (13)].

Considering that clinical CHR instruments alone detect only about 47% of transitions after 3 years (14), efforts have been made to identify potential risk factors for psychosis in several symptomatological and biological readouts, or biomarkers, of the disorder (15) so that individualized prognostication may be enhanced. The presence of environmental adverse events (16), cognitive impairments (17), neuromorphological (18), and electrophysiological (19) and hematological (20) alterations, as well

as resting-state (21) and task-related (22) neural activity and connectivity anomalies, has been consistently reported in people at risk for psychosis compared with healthy individuals. Some of these phenotypes have been associated with both disease course and transition to the overt disease (4). Therefore, the identification of reliable markers able to distinguish between at-risk and healthy populations may be potentially useful in clinical practice to monitor disease development and treatment outcome (23) and to obviate time-consuming CHR assessments. The two prevailing statistical approaches to address the challenge of single-subject prediction are machine learning (ML) methods (e.g., support vector machine, LASSO [least absolute shrinkage and selection operator] regression, random forest), which can handle large databases and different data domains (24,25), and Cox proportional hazard regression, a form of multivariate survival analysis (26) able to investigate time-to-conversion trajectories. Recent research applying these methods has produced prognostic models able to stratify CHR patients into different risk classes according to their pretest risk enrichment (27) or a set of combined predictors (28,29), or to predict patients' functional outcomes based on different data modalities with performance accuracies of up to 83% (10,30). Despite the great potential of these models, their applicability is still hindered by the methodological heterogeneity in the field. Indeed, CHR patients are identified by several clinical instruments and are characterized by subtypes with different levels of risk (14). Moreover, models' generalizability has been assessed through discrepant validation strategies across studies, ranging from the less replicable (i.e., single-site cross-validation [CV]) to the most robust (i.e., validation to external samples) (25). Thus, methodological approaches still lack standardized validation strategies testing clinical applicability under real-world conditions. One way to tackle these issues is to use a meta-analytic approach to quantitatively investigate models' performance across different outcomes, algorithms, and data modalities. Although important contributions to this goal have been made (5,29,31), to the best of our knowledge, the field is still lacking such an analysis. Investigating the field's heterogeneity would allow a comprehensive assessment of accuracy and validity of the existing diagnostic and prognostic models, an important prerequisite for establishing reliable tools for psychosis risk quantification in clinical care.

Our aim was to review the literature on ML-based and Cox regression–based diagnostic models (i.e., discriminating CHR individuals from healthy individuals) and prognostic models (i.e., predictive approaches for transition or negative outcomes). Furthermore, we performed a meta-analysis of models' performance, with the aim of investigating the effects of 1) data modality, 2) type of algorithm, and 3) validation paradigms. We expected that our results would elucidate the complexity of methods and data domains currently used in the predictive analytics arm of CHR research. This will facilitate a deeper understanding of the state of the art within the field and may clarify the bottlenecks impeding clinical translation.

METHODS AND MATERIALS

Literature Search

We conducted a systematic search of published original articles in English through June 30, 2019, using a range of search terms in PubMed and Scopus as well as reference lists of the included articles (Supplement). We selected studies that reported prognostic or diagnostic models constructed using ML or Cox proportional hazard regression. Concerning diagnostic models, we included only those that used healthy control subjects (HCs) as a reference group to enlarge the sample size by selecting comparable classification models across studies. CHR included patients with a psychosis risk syndrome categorized as CHR, ultra high risk (UHR), or at-risk mental states (Table 1) as well as those with a familial risk (FR) or 22q11.2 deletion syndrome (22q11.2DS). Studies were included if measures of performance accuracy were reported (i.e., true positives [TP], false positives [FP], true negatives [TN], and false negatives [FN]) or if they could be extracted. Results of the literature search are illustrated in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flowchart (32) (Figure S1).

A comprehensive list of all variables extracted by each study is reported in the Supplement (second section). Performance accuracy measures used for analyses comprised TP, FN, TN, FP, sensitivity (SE) [TP/(TP + FN)], and specificity (SP) [TN/(TN + FP)].

Data Analysis

The meta-analysis of diagnostic models was conducted following previous work (33). Extracted SE and SP were converted to a confusion matrix tabulated across studies. Publication bias was assessed with both overall diagnostic odds ratio and SE. The Deeks *et al.* (34) method was used to account for biases associated with unequal proportions of TP and TN cases (Supplement).

Models were built using the bivariate random effects modeling of Reitsma *et al.* (2005) (35) in the mada R package (version 0.5.8), which permits the analysis of SE and SP separately by explicitly accounting for correlations between each measure, incorporating precision estimates arising from sample size differences (i.e., more precision with higher weight), and modeling normal distributions of each with a random effects approach. This bivariate method was used to produce summary estimates of SE, SP, and confidence intervals (CIs) that were used in forest plots, in addition to the analysis of moderators using mixed modeling. Moderators were age, sex, data modality, algorithm, presence of CV, type of CHR, being a multisite study, and year of publication. For prognostic studies, we also investigated follow-up time and prognostic target. Moderator analyses were conducted if a minimum of 10 models for variable were available to decrease the standard error and maximize power in case of high between-study variance (36) and to control for sample size and CV scheme—the latter factor overlapping with algorithm used. Results were corrected for false discovery rate. Likelihood ratios and diagnostic odds ratios were produced using a Markov chain Monte Carlo approach within the mada toolbox. All analyses were conducted with R (version 3.6.0).

RESULTS

The systematic literature search detected 881 articles, from which 44 were considered eligible after screening for exclusion criteria, for a total of 12 diagnostic models (Table 2 and

Table 1. Definitions of Different Psychosis Risk Syndromes Commonly Referred to as CHR States and Descriptions of the Abbreviations and Respective Clinical Diagnostic Instruments

Concept	Description	Instruments
CHR	Clinical high risk: psychosis risk syndrome operationalized by UHR, BS, or both diagnostic criteria	All instruments below
ARMS	At-risk mental state: same as the CHR state	
UHR	Ultra high risk: psychosis risk syndrome described by the fulfillment of APS, BLIP, or GRDS criteria	SIPS, SOPS, CAARMS
APS	Attenuated psychotic symptoms: subthreshold psychotic symptoms	
BLIPS	Brief limited intermittent psychotic symptoms: full-blown psychotic symptoms present for a maximum of a week	
GRDS	Genetic risk and deterioration syndrome: family history of psychosis or schizotypal personality and drop in functioning or sustained low functioning[a]	
BS	Basic symptoms: subjective disturbances of cognitive, affective, and perceptive nature	BSABS
COGDIS	Cognitive disturbances: 9 BS describing disturbances of cognitive nature	SPI-A/SPI-CY
COPER	Cognitive-perceptive symptoms: 10 BS describing disturbances of a cognitive-perceptual nature	
UPS	Unspecific prodromal symptoms: unspecific attenuated symptoms characterizing a low-risk state	BSIP

BSABS, Bonn Scale for the Assessment of Basic Symptoms; BSIP, Basel Screening Instrument for Psychosis; CAARMS, Comprehensive Assessment of the At-Risk Mental State; SIPS, Structured Interview for the Prodromal Syndrome; SOPS, Scale of Prodromal Symptoms; SPI-A/SPI-CY; Schizophrenia Proneness Instrument–Adult version/Schizophrenia Proneness Instrument–Child and Youth version.

[a]Drop in functioning is described 1) in the CAARMS as a Social and Occupational Functioning Assessment Scale (SOFAS) score ≤30% compared with the previous functioning, within the last year, and for at least 1 month and 2) in the SIPS/SOPS as a 30% decrease in the Global Assessment of Functioning scale score from premorbid baseline. A sustained low functioning is defined only in the CAARMS as a SOFAS score ≤50 in the past year or longer.

Figure S1) and 32 prognostic models (Table 3 and Figure S1). The final sample comprised 3707 patients for prognostic studies (mean age = 20.41 years; ∼58% male), of which 320 (∼9%) were CHR patients investigated for nontransition outcomes (mean age = 19.25 years; 56% male) and 1052 were used for diagnostic classification (mean age = 23.42 years; ∼59% male), of which 480 (45%) were HCs. In addition, 26 studies used ML (all diagnostic studies) and 18 were conducted with Cox regression (Tables 2 and 3 and Table S1).

Meta-analytic Results

CHR individuals could be classified against HCs with an overall SE of 78% (95% CI = 73%–83%) and an SP of 77% (95% CI = 68%–84%), while across all prognostic models SE reached 67% (95% CI = 63%–70%) and SP reached 78% (95% CI = 73%–82%). Prognostic studies showed a publication bias (R^2 = .26, $p < .001$), whereas diagnostic studies did not (R^2 = .07, $p > .05$) (Figure S2). Performances of both models' categories are illustrated in two summary receiving operating characteristic curves (Figures 1 and 2) and forest plots (Figures 3 and 4). Within diagnostic models, moderator effects of type of CHR and algorithm, data modality, presence of CV, and being a multisite study were not investigated because less than 10 models per factor were available (36). We found no effects of moderator variables in either application domain ($p > .10$) (Table S2) even when splitting the sample based on CV (Supplement).

Table 2. Summary of Diagnostic Studies Included in the Current Meta-analysis

Study	CHR Type	Data Modality	Algorithm	Outcome	SE	FPR
Bendfeldt *et al.* (37)	UHR, UPS	Biological: fMRI	SVM	Diagnosis	74	0.42
Guo *et al.* (39)	FR	Biological: fMRI	SVM	Diagnosis	60	0.6
Koutsouleris *et al.* (41)	UHR, BS	Clinical: cognition	SVM	Diagnosis	96	0.2
Koutsouleris *et al.* (43)	UHR, BS	Biological: sMRI	SVM	Diagnosis	89	0.2
Liu *et al.* (53)	FR	Biological: fMRI	SVM	Diagnosis	72	0.14
Pettersson-Yeo *et al.* (47)	UHR	Biological: sMRI	SVM	Diagnosis	80	0.27
Scariati *et al.* (48)	22q11.2DS	Biological: fMRI	SVM	Diagnosis	81	0.12
Studerus *et al.* (99)	UHR, UPS	Clinical: cognition	Random forest	Diagnosis	73	0.23
Tylee *et al.* (49)	22q11.2DS	Biological: DTI	SVM	Diagnosis	85	0.18
Valli *et al.* (50)	UHR	Biological: sMRI	SVM	Diagnosis	68	0.24
Wang *et al.* (42)	UHR	Biological: fMRI	SVM	Diagnosis	82	0.31
Zhu *et al.* (46)	UHR	Biological: fMRI	SVM	Diagnosis	72	0.53

22q11.2DS, 22q11.2 deletion syndrome; BS, basic symptoms; CHR, clinical high risk; DTI, diffusion tensor imaging; fMRI, functional magnetic resonance imaging; FPR, false positive rate; FR, familial risk; SE, sensitivity; sMRI, structural magnetic resonance imaging; SVM, support vector machine; UHR, ultra high risk; UPS, unspecific prodromal symptoms.

Table 3. Summary of Prognostic Studies Included in the Current Meta-analysis

Study	CHR Type	Data Modality	Algorithm	Outcome	SE	FPR
Amminger *et al.* (11)	UHR	Biological: lipids	GPC	Functioning	83	0.25
Bedi *et al.* (54)	UHR	Clinical: speech	Convex Hull	Transition	100	0
Buchy *et al.* (100)	UHR	Clinical: substance use	Cox regression	Transition	69	0.19
Cannon *et al.* (101)	UHR	Clinical: symptoms, family risk, functioning	Cox regression	Transition	67	0.47
Cannon *et al.* (28)	UHR	Multimodal: symptoms, environment, genetic, cognition	Cox regression	Transition	67	0.28
Carrión *et al.* (74)	UHR	Multimodal: symptoms, environment, genetic, cognition	Cox regression	Transition	58	0.27
Chan *et al.* (56)	UHR, UPS	Biological: serum	LASSO regression	Transition	89	0.34
		Clinical: positive symptoms			78	0.4
		Multimodal: serum, symptoms			89	0.21
Cornblatt *et al.* (58)	UHR	Multimodal: clinical, demographics, cognition	Cox regression	Transition	60	0.03
Das *et al.* (55)	UHR, UPS	Biological: cortical gyrification	Randomized trees	Transition	66	0.03
de Wit *et al.* (30)	UHR, BS	Biological: sMRI, gyrification	SVM	Functioning	67	0.25
		Clinical: disorganized speech			76	0.25
		Multimodal: sMRI, clinical, combination			68	0.19
DeVylder *et al.* (102)	UHR	Clinical: disorganized communication	Cox regression	Functioning	58	0.4
Dragt *et al.* (64)	UHR	Clinical: disorganized communication	Cox regression	Transition	50	0.09
Francesconi *et al.* (59)	UHR	Clinical: thought content, ToM, processing, NSS	Cox regression	Transition	67	0.03
Fusar-Poli *et al.* (60)	UHR-BLIPS	Clinical: disorganizing symptoms	LASSO Cox regression	Transition	24	0.37
Gothelf *et al.* (38)	22q11.2DS	Biological: sMRI	SVM	Transition	90	0
Hoffman *et al.* (73)	UHR	Clinical: cognition	Cox regression	Transition	89	0.11
Kambeitz-Ilankovic *et al.* (40)	UHR, BS	Biological: cortical surface area	SVM	Functioning	79	0.15
Koutsouleris *et al.* (41)	UHR, BS	Clinical: cognition	SVM	Transition	80	0.25
Koutsouleris *et al.* (10)	UHR, BS	Biological: sMRI	SVM	Functioning (role)	67	0.53
		Clinical: functioning			61	0.25
		Multimodal: sMRI and functioning			59	0.3
Koutsouleris *et al.* (10)	UHR, BS	Biological: sMRI	SVM	Functioning (social)	80	0.28
		Clinical: functioning			70	0.16
		Multimodal: sMRI and functioning			83	0.18
Koutsouleris *et al.* (44)	UHR, BS	Biological: sMRI	SVM	Transition	76	0.15
Lavoie *et al.* (71)	UHR	Biological: blood antioxidant	Cox regression	Transition	91	0.33
Mechelli *et al.* (45)	UHR	Clinical: disorders of thought content, attenuated positive symptoms, functioning	SVM	Transition	69	0.39
				Functioning	63	0.37
Michel *et al.* (61)	UHR, BS	Clinical: SIPS, SPI-A, cognition	Cox regression	Transition	57	0.45
Nieman *et al.* (62)	UHR, BS	Multimodal: symptoms and ERPs	Cox regression	Transition	78	0.12
Perkins *et al.* (20)	UHR	Biological: blood plasma analytes	Greedy algorithm	Transition	60	0.1
Ramyead *et al.* (57)	UHR, UPS	Biological: EEG	LASSO	Transition	58	0.17
Ruhrmann *et al.* (65)	UHR, BS	Clinical: symptoms, sleep, schizotypy, functioning, education	Cox regression	Transition	42	0.02
Tarbox *et al.* (66)	UHR	Clinical: alogia, anhedonia/asociality, suspiciousness	Cox regression	Transition	62	0.39
Thompson *et al.* (67)	UHR	Clinical: unusual thought content, functioning, family history, functional decline	Cox regression	Transition	30	0.11
van Tricht *et al.* (68)	UHR	Biological: EEG	Cox regression	Transition	46	0.13
van Tricht *et al.* (72)	UHR, BS	Biological: EEG	Cox regression	Transition	83	0.21

Table 3. Continued

Study	CHR Type	Data Modality	Algorithm	Outcome	SE	FPR
Zarogianni *et al.* (51)	UHR, BS	Multimodal: sMRI and cognition	SVM	Transition	63	0.16
Zarogianni *et al.* (52)	FR	Biological: sMRI	SVM	Transition	76	0.23
		Multimodal: sMRI and cognition			100	0.17

22q11.2DS, 22q11.2 deletion syndrome; BLIPS, brief limited intermittent psychotic symptoms; BS, basic symptoms; CHR, clinical high risk; EEG, electroencephalography; ERP, evoked response potential; FPR, false positive rate; FR, familial risk; GPC, Gaussian process classification; LASSO, least absolute shrinkage and selection operator; NSS, neurological soft signs; SE, sensitivity; SIPS, Structured Interview for Prodromal Syndromes; sMRI, structural magnetic resonance imaging; SPI-A, Schizophrenia Proneness Instrument–Adult version; SVM, support vector machine; ToM, theory of mind; UHR, ultra high risk; UPS, unspecific prodromal symptoms.

Effect of Algorithm Choice

A total of 19 ML studies (73%) employed a support vector machine algorithm (10,30,37–53), while the rest used Gaussian process (11) or convex hull classification (54), randomized trees (55), greedy algorithm (20), random forest (5), or LASSO regression (56,57). All ML models were computed with CV, whereas studies using Cox regression applied bootstrapping (28,58–62), reported apparent results (i.e., the model is tested in the same sample from which it was derived) (63–68), or lacked a validation procedure. Among the cross-validated studies, 58% applied leave-one-out CV, 3 of which nested and 7 of which used *k*-fold CV (3 in its repeated nested form). Only 1 study applied a leave-site-out CV (10), that is, a form of internal–external validation (69). Within prognostic studies, we found a main effect of CV/algorithm on SE (p = .009; χ^2_2 = 6.96, p = .031); that is, cross-validated ML models reached a higher SE (71%, 95% CI = 67%–74%) than Cox regression ones (61%, 95% CI = 54%–68%) (Figure 4).

Effect of Data Modality

Diagnostic models included the use of functional (37,39,43,47,48) and structural (46,50,70) magnetic resonance imaging (MRI) and diffusion tensor imaging (49), and behavioral models were based on neurocognitive functions (42,43).

Models for prediction of transition to psychosis involved blood-based (20,56,71), electrophysiological (57,68,72), and neuroanatomical data using white and/or gray matter volume (38,44,51) or gyrification measures (55). Clinical models were trained on prodromal positive and negative symptoms, functioning, and family risk associated with functional decline; the neurocognitive modality was based on executive functions and verbal IQ (41) or speech features (54,73). Multimodal approaches included different combinations of clinical, neuropsychological, and demographic variables as well as genetic risk (28,51,52,54,74). One model was built on P300 amplitude from event-related potentials and sociopersonal adjustment (62). Functional outcomes were predicted with neuroanatomical (63,9,19) and blood-based biomarkers (11), and 2 studies combined clinical and MRI measures (10,30). There were no effects of data modality on SE (p = .172) or false positive rate (p = .606) (Table S2).

Effect of Sample Characteristics

Performance accuracies were not influenced by age and sex of individuals ($p > .10$) (Table S2). CHR in 86% of the studies fulfilled the UHR criteria (75), while 6 models were based on the genetic risk syndromes 22q11.2DS (38,48,49) or FR (39,52,53). Because of this imbalance, we could not statistically test the effects of this variable, yet results did not change when excluding patients with 22q11.2DS and FR (Supplement).

Furthermore, individuals differed in their outcome definitions. Poor functional outcome was defined on the Global Assessment of Functioning scale (GAF) (cutoff: 70) (40), the Social and Occupational Functioning Assessment Scale (score ≤50) (45), the GAF modified version (76) defining nonresilience through a cutoff of ≤65 (30), or the Global Functioning social/role scale (<8) (10). In one case (11), treatment response was operationalized as an increase of ≥15 points in the GAF. There were no significant effects on SE or false positive rate driven by

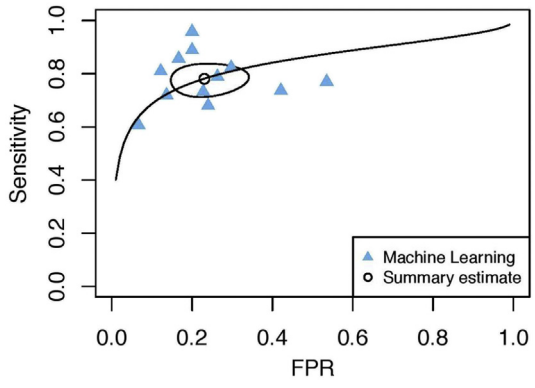


Figure 1. Summary receiver operating characteristic curve of diagnostic studies. FPR, false positive rate.

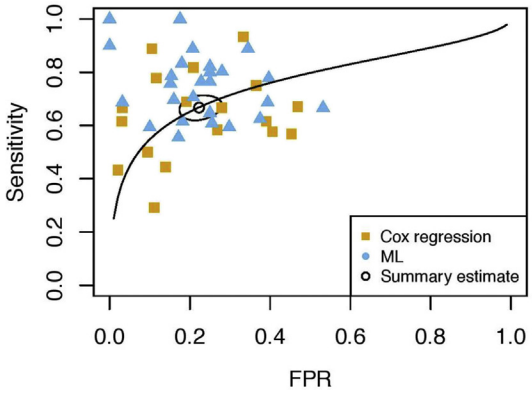


Figure 2. Summary receiver operating characteristic curve of prognostic studies. FPR, false positive rate; ML, machine learning.

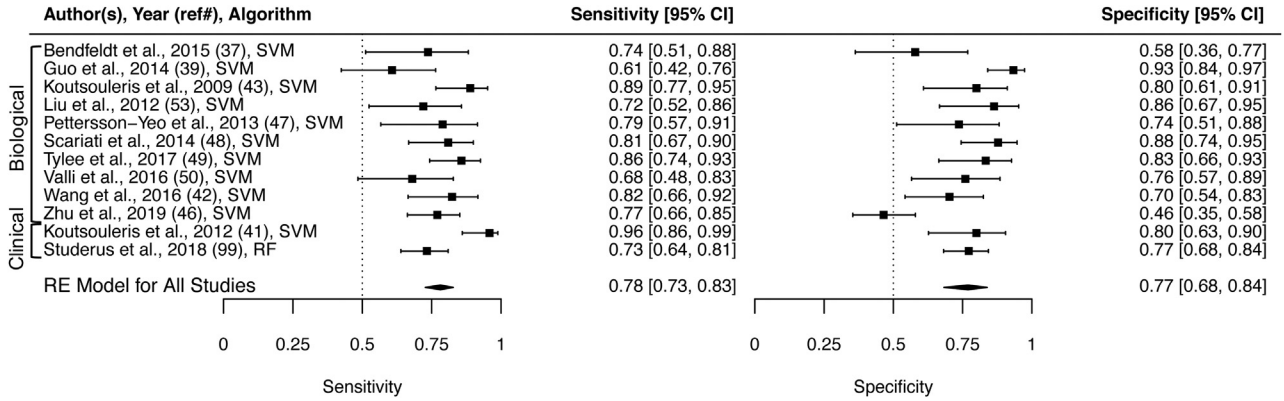


Figure 3. Forest plot of sensitivity and specificity for all diagnostic studies divided by data modality. CI, confidence interval; RE, random effects; RF, random forest; SVM, support vector machine.

prognostic target (p = .570 or .085, respectively) or the duration of time-to-follow-up examination (p = .637 or .305, respectively).

DISCUSSION

We conducted a systematic review and meta-analysis on 44 studies reporting prognostic and diagnostic models for a total of 3707 and 572 CHR individuals, respectively, with the aim to quantitatively assess their accuracy, validity, and heterogeneity. Our results point to good model performance overall and to a higher SE of ML models compared with Cox regression in prognostic studies. This effect was fully collinear with that of CV, mainly due to the complete overlap of this factor with algorithm type. Notably, there were no significant effects of data modality, CHR or CV type, prognostic target, or any other potential confounding variables (e.g., age distribution, sex, year of publication, follow-up interval time) on accuracy performance in our data. It is noteworthy that in prognostic studies we observed a publication bias, that is, the tendency for studies with smaller sample sizes to report higher, and potentially inflated, prediction accuracies (77). This might have affected our results (77) so that we cannot draw robust conclusions from our meta-analytical findings.

Methodological Differences and Pitfalls

Prognostic models employing ML outperformed those using Cox regression by 10% SE. This finding may have resulted from a complex interplay of cohort-related and methodological heterogeneity. Notably, there was a complete overlap between the statistical method chosen and implementation of CV, that is, all ML models were cross-validated, while only 6 Cox regression studies applied bootstrapping as the validation procedure. Because the choice of a reliable validation method strongly determines both performance and generalizability of models (25), this methodological discrepancy may have biased our findings. Validation issues were also present in studies employing ML for prognostic modeling. First, 53% of these studies applied CV without nesting and repetitions, which is known to generate overoptimistic results due to high variability and information leakage between training and testing data during model optimization (78). The extended use of this validation scheme may explain the higher SE found in ML studies.

Second, several Cox regression studies included in this meta-analysis either did not report probability thresholds or chose a priori optimal thresholds from the data. While ML's lack of homogeneous thresholds is mainly handled via CV schemes averaging performances across folds and repetitions, the use of p values or data-derived thresholds without a proper training–test separation might have inflated Cox regression models' performance (63).

Third, preprocessing approaches varied across studies. In 3 cases, for instance, prognostic features were derived from univariate group comparisons or by applying principal component analysis outside the CV scheme (20,43,53), which is a known source of information leakage, because variance from the training sample data is carried into the test sample (25). One model was constructed on a nonrandom sampling of the training set (49), while another model classified patients at UHR from HCs based on the brain pattern shared by patients at UHR and with first-episode psychosis (46). These approaches, as well as the use of stepwise methods in Cox regression models, entail sample-driven variance and, therefore, could lead to good predictive performance, but arguably they should be tested for generalizability in an external dataset. Valuable alternatives are literature-based feature selection and embedded feature optimization, where the intrinsic optimal feature configuration is learned by the model itself (79).

It should be noted that some of the studies included in our meta-analytic contribution had very low sample sizes. One study had $N < 20$, while 2 diagnostic and 21 prognostic models had, respectively, less than 20 CHR individuals or CHR with poor outcome. Findings from these studies might be consistent with literature demonstrating a publication bias toward increased accuracy with reduced sample size (80), possibly caused by overfitting. This indicates the need for future ML research to employ larger, preferably multisite samples for both diagnostic and prognostic purposes (80).

Taken together, these issues may mirror the heterogeneity of methodological procedures within the field. Arguably, the

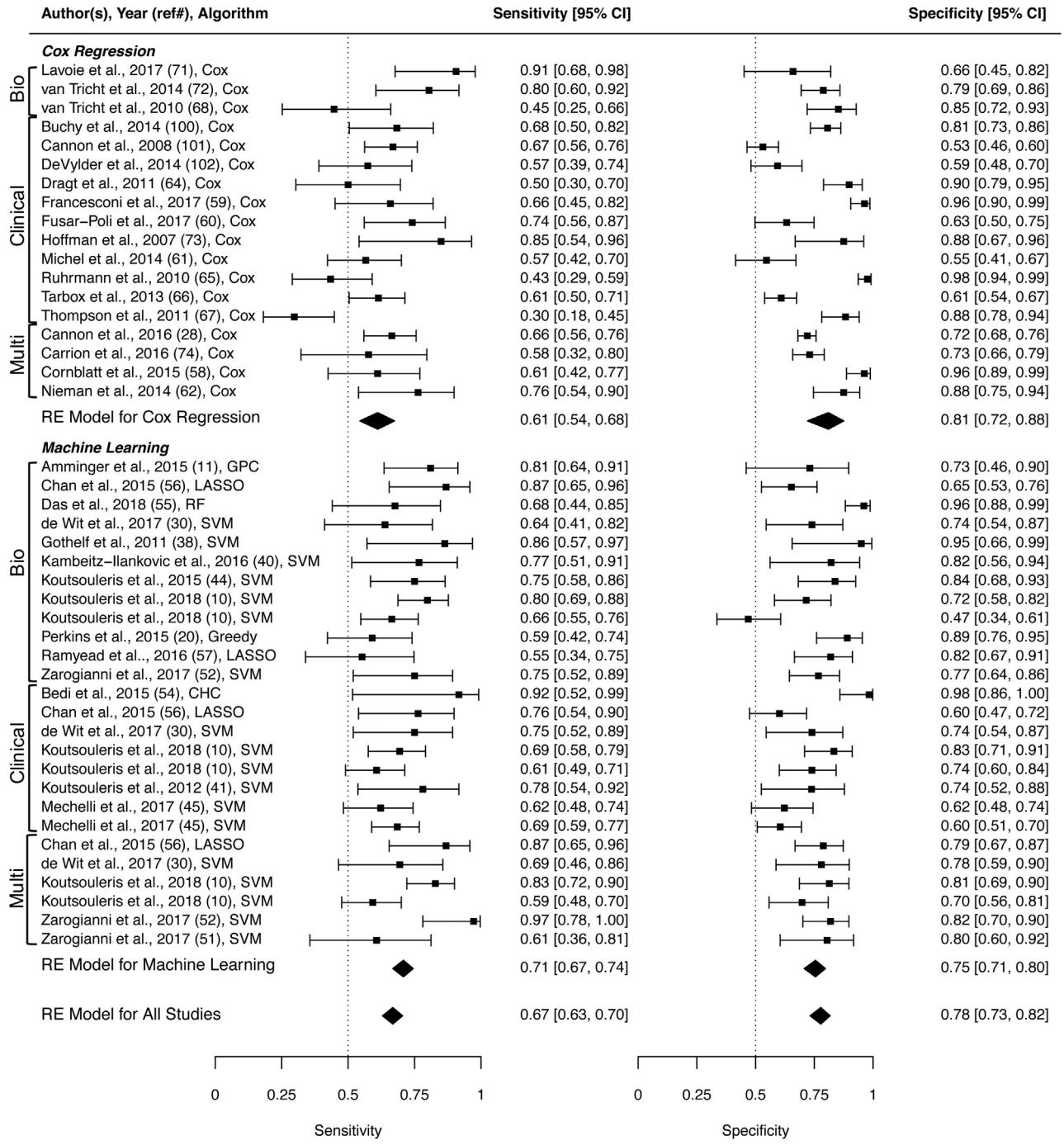


Figure 4. Forest plot of sensitivity and specificity for all prognostic studies divided by algorithm and data modality. CHC, convex hull classification; CI, confidence interval; GPC, Gaussian process classifier; LASSO, least absolute shrinkage and selection operator regularized regression; RE, random effects; RF, random forest; SVM, support vector machine.

application of ML techniques to diagnosis and prognosis in psychiatry is still relatively young (24), so conventions and standard operating procedures facilitating model comparability and replicability have not become generally accepted. Our findings highlight the urgency to develop such guidelines for the construction of prognostic and diagnostic models (81). As indicated in Table 4, the most important ones are 1) the implementation of repeated nested CV, internal–external, or external validation schemes and 2) the full and strict embedding of all preprocessing or feature engineering procedures

Table 4. Conceptual and Methodological Guidelines for Construction of Diagnostic and Predictive Models Implementable in Real-Life Clinical Practice

Guidelines	Practical Suggestions
Conceptual Guidelines	
Harmonization of the CHR definition and diagnostic instruments	Create a harmonized early recognition instrument that encompasses those at-risk definitions and criteria from the existing diverse inventories that parsimoniously delineate the CHR state and also are predictive of its adverse outcomes
Broaden the scope of prediction to nontransition outcomes	Harmonize social and occupational outcomes, pharmacological and nonpharmacological treatment response criteria, and definitions of persistence or remission of symptoms and use these end points in future predictive studies
Methodological Guidelines	
Increase in sample size	Facilitate collaborative science approaches that enable the harmonization of end-point definitions and the external validation of predictive models Get access to open-source databases
Study design harmonization	Employ reliable methodologies (CV and external validation are recommended); avoid leave-one-out CV; implement *k*-fold CV Embed all preprocessing or feature engineering procedures within the chosen CV scheme Enforce preregistration processes (as in clinical trials) to facilitate monitoring of standardized data acquisition, model discovery, and validation plan
Common modeling platforms and open-source model libraries	Large-scale, consortium-wide international model benchmarking

CHR, clinical high risk; CV, cross-validation.

within the CV scheme. Researchers, funding organizations, and journals should support efforts to standardize approaches, favoring the importance of thorough validation over model performance per se.

Type of Data Modality

Overall, most models were constructed using biological (44%) and clinical (38%) data, with only 10 prognostic models based on more than one data modality. Most diagnostic models used MRI data (83%), whereas prognostic models showed a higher variability. Prognostic models of psychosis transition included molecular, neuroanatomical, electrophysiological, neuropsychological, and clinical data modalities, most of the latter trained on prodromal positive and negative symptoms, functioning, and FR associated with functional decline. We found no significant differences in predictive accuracy when comparing data modalities within and between algorithms.

This result may mirror a real lack of significant differences in biomarker type when distinguishing the CHR state from the norm or predicted outcome. However, because only 4 prognostic studies tested the relative and combined predictive ability of different data modalities on the same individuals (10,30,52,56), and because data modalities are overall under- or overrepresented, the currently available studies do not allow this conclusion to be drawn. Further research directly comparing performance across data modalities, followed by meta-analytic evaluation, is warranted.

Alternatively, our results may reflect the complexity of the multifaceted architecture of psychosis risk (82), which might be only partly captured by single data modalities. Indeed, a neuroanatomical biomarker might be informative for genetically or pathophysiologically driven mechanisms given that genes' effect may be closer to brain than to behavior (83); on the other hand, neurocognitive performance might explain more environmentally driven variance relating, for example, to socioeconomic status (84). Hence, a multimodal approach may be a viable way to reconcile and leverage information from single risk domains. Powerful new methodologies able to combine multiple sources of data, such as similarity network fusion (85), might be suitable for this purpose. Indeed, research has shown that a combination of clinical variables and structural brain imaging data might represent a promising multimodal framework for psychosis prediction (10,23,31). Along these lines, Schmidt *et al.* (29) devised a 3-stage sequential testing paradigm, which in theory reaches nearly perfect positive predictive value when individuals are tested on one multimodal modality (i.e., clinical and electroencephalography) and two biological data modalities (i.e., structural MRI and blood based). However, these findings are simulated, have not been confirmed in empirical studies yet, and did not follow a thorough meta-analytical approach like the one implemented here.

Alternatively, similar performance of tested data modalities may have resulted from the variability induced by higher-order algorithm–data validation interactions. To thoroughly compare models originating from different data spaces, methodological consensus guidelines are urgently needed in the precision psychiatry field. A strict cross-study standardization, in terms of both data definitions and algorithm implementations, may shed light on real phenotypic and neurobiological differences and thus lead to unique insights into the pathology of emerging psychosis.

At-Risk State/Sample Differences

Another source of heterogeneity affecting our results may be due to clinical sample definitions. Most of the at-risk individuals in our sample fulfilled the UHR criteria, while a minority (5.7%) had an FR or a 22q11.2DS diagnosis, which prevented us from quantitatively estimating the effects of risk group designation. However, it is noteworthy that two of the instruments operationalizing UHR criteria (i.e., SIPS [Structured Interview for Prodromal Syndromes] and CAARMS [Comprehensive Assessment of At-Risk Mental States])

include a genetic risk group (i.e., the genetic risk deterioration syndrome) and that two studies in our sample included FRs and deletion syndrome patients with subthreshold psychotic symptoms (49,52). This diagnostic overlap might create, on the one hand, a further source of variability and, on the other, a tangible bridge to the well-known heterogeneity among CHR individuals. This issue was tackled by a recent study (14) that provided evidence of a differential risk level within the subcategories of the CHR construct. Hence, further research should put effort into revising the CHR paradigm toward a more parsimonious definition based on one gold-standard clinical instrument and clear-cut biological underpinnings.

Furthermore, in our sample, criteria to define transition to psychosis or poor functional outcome differed both in their operationalization and in the threshold used within a specific diagnostic instrument. Another issue in the variability of outcome definition is dichotomization of continuous variables such as GAF and global functioning, which has proven to be a potential source of bias in prognostic models (63,86). It is noteworthy that 1 study (45) addressed this point by conducting an additional analysis to investigate the continuous nature of functioning by using a support vector regression algorithm. The predictability of nontransition outcomes in at-risk individuals is still relatively unexplored. Therefore, there is a need for clinical consensus on relevant nontransition outcomes and how they should be assessed. Additionally, adopting adaptive risk models, which capture the high extent of variability of symptoms and risk factors over time (87), may tackle this complexity and provide more precise measurements of developing negative outcomes, as proposed by digital phenotyping approaches (88).

Notably, CHR populations differ not only in their clinical picture but also along demographic and sociocultural dimensions (89). For instance, American CHR individuals are usually younger (~16–18 years) than their European counterparts (~22–24 years). Interestingly, recent research has shown that neuroanatomical development and risk for developing psychosis are interconnected (90,91). This evidence might also reveal neurobiological processes leading to neurocognitive changes in the CHR state (92,93). Overall, our findings suggest that the gestalt of the CHR state might be successfully modeled only if multiple behavioral and neurobiological moderators are conjointly considered using standardized multivariate methods, thereby fully embracing the complexity of this risk paradigm.

Limitations

Our meta-analysis was driven by the primary aim to evaluate the potential applicability of diagnostic and prognostic models in real-life clinical practice. Therefore, we focused only on the two currently prevailing methodological approaches (i.e., ML and Cox regression). Importantly, we might have missed significant results by excluding other more traditional statistical methods such as logistic regression (15,63), which has often been implemented for prognostic purposes (15,63), eventually showing higher performance than ML (94). Nevertheless, ML approaches enable the investigation of the intrinsic complexity of specific data types (e.g., brain features) and are devised for better generalizability.

Another limitation might be the lack of investigation into symptomatology, treatment, substance use, or additional comorbidities, which was due to missing or inconsistent information for several studies. Indeed, already in patients with first-episode psychosis, antipsychotic treatment has been shown to have neuroanatomical effects (95), and continuous cannabis use has been shown to lead to worse outcomes (96). It is also plausible that the high variability of symptoms and clinical comorbidities in the CHR population (13) has further introduced spurious variance in our analyses.

Furthermore, the CHR paradigm has proven to have intrinsic limitations. On the one hand, its predictive power might be partly driven by the so-called pretest risk enrichment; that is, the assessment of at-risk criteria in a specific constellation of help-seeking individuals (97,98). On the other hand, it might not capture the full extent of risk in the population, as a recent study pointed out by reporting that most transitions occurred in patients with an unclear psychiatric diagnosis or no CHR status (9). Because most prognostic models have been developed for the CHR state, their usefulness outside of this category should be intensively investigated.

Lastly, given the heterogeneity of our data and the publication bias detected, our meta-analysis is inherently limited to a description of, not an ultimate decision on, which diagnostic and prognostic models are sufficiently reliable to be applied in clinical settings.

Conclusions

A comprehensive paradigm shift is required to enable the clinical application of diagnostic and prognostic models for the CHR state. First, the field requires study design harmonization, which demands reliable methodological approaches such as CV or external validation to ensure generalizability. An approach to enhance the studies' potential for real-life implementation could be a preregistration process similar to clinical trials, during which their validity in terms of standardized data acquisition, model discovery, and validation could be monitored. Furthermore, large-scale international model benchmarking at the level of external model validation can be achieved only by constructing common modeling platforms and open source model libraries. The National Institute of Mental Health's Harmonization of At-Risk Multisite Observational Networks for Youth (HARMONY) is a first step in the above direction. Consortium-wise coordinated work will also allow strategic methodological testing; that is, controlled comparison of algorithms, preprocessing and feature optimization pipelines, and multiple data modalities (for an overview of conceptual and methodological guidelines, see Table 4). Multimodal ML carries the challenging responsibility to better disentangle the complex architecture of psychosis risk within a clinical consensus environment. This should involve efforts to unify the CHR definition, both theoretically and practically, and also to embrace relevant nontransition outcomes to broaden the prognostic scope. Future studies are warranted to investigate whether harmonizing procedures within precision psychiatry will lead to more reliable and reproducible translational research in the field.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by a EU-FP7-HEALTH grant for the project "PRONIA" (Personalized Prognostic Tools for Early Psychosis Management) (Grant No. 602152) and by the National Institute of Mental Health (NIMH) for the project "HARMONY" (Harmonization of At Risk Multisite Observational Networks for Youth) (Grant No. MH081928). PRONIA, BMBF (Federal Ministry of Education and Research), and the Max Planck Society funded RS.

NK received honoraria for two lectures from Otsuka. He has a patent issued related to adaptive pattern recognition for psychosis risk modeling (U.S. patent 20160192889A1). RS received honoraria for one lecture from Lundbeck. The other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychiatry and Psychotherapy (RS, DBD, LAA, NK), Ludwig-Maximilian-University Munich, Max Planck Institute of Psychiatry Munich (NK), Munich, and Max Planck School of Cognition (RS), Leipzig, Germany; and Department of Education, Psychology, and Communication (LAA), University of Bari "Aldo Moro," Bari, Italy.

RS and DBD contributed equally to this work.

Address correspondence to Nikolaos Koutsouleris, M.D., Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Nussbaumstr. 7, D-80336 Munich, Germany; E-mail: nikolaos.koutsouleris@med.uni-muenchen.de.

Received Sep 30, 2019; revised Jan 25, 2020; accepted Feb 6, 2020.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2020.02.009.

REFERENCES

1. Vigo D, Thornicroft G, Atun R (2016): Estimating the true global burden of mental illness. Lancet Psychiatry 3:171–178.
2. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW (2018): The science of prognosis in psychiatry: A review. JAMA Psychiatry 75:1280–1288.
3. Fusar-Poli P, Borgwardt S, Bechdolf A, Addington J, Riecher-Rössler A, Schultze-Lutter F, *et al.* (2013): The psychosis at risk state: A comprehensive state-of-the-art review. JAMA Psychiatry 70:107–120.
4. Riecher-Rössler A, Studerus E (2017): Prediction of conversion to psychosis in individuals with an at-risk mental state: A brief update on recent developments. Curr Opin Psychiatry 30:209–219.
5. Studerus E, Ramyead A, Riecher-Rössler A (2017): Prediction of transition to psychosis in patients with a clinical high risk for psychosis: A systematic review of methodology and reporting. Psychol Med 47:1163–1178.
6. Fusar-Poli P, Bonoldi I, Yung AR, Borgwardt S, Kempton MJ, Valmaggia L, *et al.* (2012): Predicting psychosis. Arch Gen Psychiatry 69:220–229.
7. Beck K, Andreou C, Studerus E, Heitz U, Ittig S, Leanza L, Riecher-Rössler A (2019): Clinical and functional long-term outcome of patients at clinical high risk (CHR) for psychosis without transition to psychosis: A systematic review. Schizophr Res 210:39–47.
8. Fusar-Poli P, Rocchetti M, Sardella A, Avila A, Brandizzi M, Caverzasi E, *et al.* (2015): Disorder, not just state of risk: Meta-analysis of functioning and quality of life in people at high risk of psychosis. Br J Psychiatry 207:198–206.
9. Fusar-Poli P, Rutigliano G, Stahl D, Davies C, Bonoldi I, Reilly T, McGuire P (2017): Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. JAMA Psychiatry 74:493–500.
10. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, *et al.* (2018): Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: A multimodal, multisite machine learning analysis. JAMA Psychiatry 75:1156–1172.
11. Amminger GP, Mechelli A, Rice S, Kim SW, Klier CM, McNamara RK, *et al.* (2015): Predictors of treatment response in young people at ultra-high risk for psychosis who received long-chain omega-3 fatty acids. Transl Psychiatry 5:3–9.
12. Yung AR, Nelson B, McGorry PD, Wood SJ, Lin A (2019): Persistent negative symptoms in individuals at ultra high risk for psychosis. Schizophr Res 206:355–361.
13. Rutigliano G, Valmaggia L, Landi P, Frascarelli M, Cappucciati M, Sear V, *et al.* (2016): Persistence or recurrence of non-psychotic comorbid mental disorders associated with 6-year poor functional outcomes in patients at ultra high risk for psychosis. J Affect Disord 203:101–110.
14. Fusar-Poli P, Cappucciati M, Borgwardt S, Woods SW, Addington J, Nelson B, *et al.* (2016): Heterogeneity of psychosis risk within individuals at clinical high risk: A meta-analytical stratification. JAMA Psychiatry 73:113–120.
15. Addington J, Farris M, Stowkowy J, Santesteban-Echarri O, Metzak P, Kalathil MS (2019): Predictors of transition to psychosis in individuals at clinical high risk. Curr Psychiatry Rep 21:39.
16. Fusar-Poli P, Tantardini M, De Simone S, Ramella-Cravaro V, Oliver D, Kingdon J, *et al.* (2017): Deconstructing vulnerability for psychosis: Meta-analysis of environmental risk factors for psychosis in subjects at ultra high-risk. Eur Psychiatry 40:65–75.
17. Seidman LJ, Shapiro DI, Stone WS, Woodberry KA, Ronzio A, Cornblatt BA, *et al.* (2016): Association of neurocognition with transition to psychosis: Baseline functioning in the second phase of the North American Prodrome Longitudinal Study. JAMA Psychiatry 73:1239–1248.
18. Gifford G, Crossley N, Fusar-Poli P, Schnack HG, Kahn RS, Koutsouleris N, *et al.* (2017): Using neuroimaging to help predict the onset of psychosis. NeuroImage 145:209–217.
19. Perez VB, Woods SW, Roach BJ, Ford JM, McGlashan TH, Srihari VH, Mathalon DH (2014): Automatic auditory processing deficits in schizophrenia and clinical high-risk patients: Forecasting psychosis risk with mismatch negativity. Biol Psychiatry 75:459–469.
20. Perkins DO, Jeffries CD, Addington J, Bearden CE, Cadenhead KS, Cannon TD, *et al.* (2015): Towards a psychosis risk blood diagnostic for persons experiencing high-risk symptoms: Preliminary results from the NAPLS project. Schizophr Bull 41:419–428.
21. Anticevic A, Haut K, Murray JD, Repovs G, Yang GJ, Diehl C, *et al.* (2015): Association of thalamic dysconnectivity and conversion to psychosis in youth and young adults at elevated clinical risk. JAMA Psychiatry 72:882–891.
22. Antonucci LA, Penzel N, Pergola G, Kambeitz-Ilankovic L, Dwyer D, Kambeitz J, *et al.* (2020): Multivariate classification of schizophrenia and its familial risk based on load-dependent attentional control brain functional connectivity. Neuropsychopharmacology 45:613–621.
23. Antonucci LA, Pergola G, Pigoni A, Dwyer D, Kambeitz-Ilankovic L, Penzel N, *et al.* (2020): A pattern of cognitive deficits stratified for genetic and environmental risk reliably classifies patients with schizophrenia from healthy controls. Biol Psychiatry 87:697–707.
24. Bzdok D, Meyer-Lindenberg A (2018): Machine learning for precision psychiatry: Opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 3:223–230.
25. Dwyer DB, Falkai P, Koutsouleris N (2018): Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol 14:91–118.
26. Cox DR (1972): Regression models and life-tables. J R Stat Soc Ser B: Methodological 34:187–220.
27. Fusar-Poli P, Rutigliano G, Stahl D, Schmidt A, Ramella-Cravaro V, Hitesh S, McGuire P (2016): Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. JAMA Psychiatry 73:1260–1267.
28. Cannon TD, Yu C, Addington J, Bearden CE, Cadenhead KS, Cornblatt BA, *et al.* (2016): An individualized risk calculator for research in prodromal psychosis. Am J Psychiatry 173:980–988.
29. Schmidt A, Cappucciati M, Radua J, Rutigliano G, Rocchetti M, Dell'Osso L, *et al.* (2017): Improving prognostic accuracy in subjects at clinical high risk for psychosis: Systematic review of predictive

models and meta-analytical sequential testing simulation. Schizophr Bull 43:375–388.
30. de Wit S, Ziermans TB, Nieuwenhuis M, Schothorst PF, van Engeland H, Kahn RS, *et al.* (2017): Individual prediction of long-term outcome in adolescents at ultra-high risk for psychosis: Applying machine learning techniques to brain imaging data. Hum Brain Mapp 38:704–714.
31. Strobl EV, Eack SM, Swaminathan V, Visweswaran S (2012): Predicting the risk of psychosis onset: Advances and prospects. Early Interv Psychiatry 6:368–379.
32. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009): Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 6:1000097.
33. Kambeitz J, Cabral C, Sacchet MD, Gotlib IH, Zahn R, Serpa MH, *et al.* (2017): Detecting neuroimaging biomarkers for depression: A meta-analysis of multivariate pattern recognition studies. Biol Psychiatry 82:330–338.
34. Deeks JJ, Macaskill P, Irwig L (2005): The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58:882–893.
35. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH (2005): Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982–990.
36. Borenstein M, Hedges LV (2009): Introduction to Meta-analysis. Hoboken, NJ: John Wiley.
37. Bendfeldt K, Smieskova R, Koutsouleris N, Klöppel S, Schmidt A, Walter A, *et al.* (2015): Classifying individuals at high-risk for psychosis based on functional brain activity during working memory processing. NeuroImage Clin 9:555–563.
38. Gothelf D, Hoeft F, Ueno T, Sugiura L, Lee AD, Thompson P, Reiss AL (2011): Developmental changes in multivariate neuroanatomical patterns that predict risk for psychosis in 22q11.2 deletion syndrome. J Psychiatr Res 45:322–331.
39. Guo S, Palaniyappan L, Yang B, Liu Z, Xue Z, Feng J (2014): Anatomical distance affects functional connectivity in patients with schizophrenia and their siblings. Schizophr Bull 40:449–459.
40. Kambeitz-Ilankovic L, Meisenzahl EM, Cabral C, von Saldern S, Kambeitz J, Falkai P, *et al.* (2016): Prediction of outcome in the psychosis prodrome using neuroanatomical pattern classification. Schizophr Res 173:159–165.
41. Koutsouleris N, Davatzikos C, Bottlender R, Patschurek-Kliche K, Scheuerecker J, Decker P, *et al.* (2012): Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. Schizophr Bull 38:1200–1215.
42. Wang S, Wang G, Lv H, Wu R, Zhao J, Guo W (2016): Abnormal regional homogeneity as potential imaging biomarker for psychosis risk syndrome: A resting-state fMRI study and support vector machine analysis. Sci Rep 6:27619.
43. Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, *et al.* (2009): Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch Gen Psychiatry 66:700–712.
44. Koutsouleris N, Riecher-Rössler A, Meisenzahl EM, Smieskova R, Studerus E, Kambeitz-Ilankovic L, *et al.* (2015): Detecting the psychosis prodrome across high-risk populations using neuroanatomical biomarkers. Schizophr Bull 41:471–482.
45. Mechelli A, Lin A, Wood S, McGorry P, Amminger P, Tognin S, *et al.* (2017): Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. Schizophr Res 184:32–38.
46. Zhu F, Liu Y, Liu F, Yang R, Li H, Chen J, *et al.* (2019): Functional asymmetry of thalamocortical networks in subjects at ultra-high risk for psychosis and first-episode schizophrenia. Eur Neuropsychopharmacol 29:519–528.
47. Pettersson-Yeo W, Benetti S, Marquand AF, Dell'Acqua F, Williams SCR, Allen P, *et al.* (2013): Using genetic, cognitive and multi-modal neuroimaging data to identify ultra-high-risk and first-episode psychosis at the individual level. Psychol Med 43:2547–2562.
48. Scariati E, Schaer M, Richiardi J, Schneider M, Debbané M, Van De Ville D, Eliez S (2014): Identifying 22q11.2 deletion syndrome and psychosis using resting-state connectivity patterns. Brain Topogr 27:808–821.
49. Tylee DS, Kikinis Z, Quinn TP, Antshel KM, Fremont W, Tahir MA, *et al.* (2017): Machine-learning classification of 22q11.2 deletion syndrome: A diffusion tensor imaging study. Neuroimage Clin 15:832–842.
50. Valli I, Marquand AF, Mechelli A, Raffin M, Allen P, Seal ML, McGuire P (2016): Identifying individuals at high risk of psychosis: Predictive utility of support vector machine using structural and functional MRI data. Front Psychiatry 7:52.
51. Zarogianni E, Storkey AJ, Borgwardt S, Smieskova R, Studerus E, Riecher-Rössler A, Lawrie SM (2019): Individualized prediction of psychosis in subjects with an at-risk mental state. Schizophr Res 214:18–23.
52. Zarogianni E, Storkey AJ, Johnstone EC, Owens DGC, Lawrie SM (2017): Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. Schizophr Res 181:6–12.
53. Liu M, Zeng L-L, Shen H, Liu Z, Hu D (2012): Potential risk for healthy siblings to develop schizophrenia: Evidence from pattern classification with whole-brain connectivity. NeuroReport 23:265–269.
54. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, *et al.* (2015): Automated analysis of free speech predicts psychosis onset in high-risk youths. npj Schizophr 1:15030.
55. Das T, Borgwardt S, Hauke DJ, Harrisberger F, Lang UE, Riecher-Rössler A, *et al.* (2018): Disorganized gyrification network properties during the transition to psychosis. JAMA Psychiatry 75:613–622.
56. Chan MK, Krebs MO, Cox D, Guest PC, Yolken RH, Rahmoune H, *et al.* (2015): Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. Transl Psychiatry 5: e601.
57. Ramyead A, Studerus E, Kometer M, Uttinger M, Gschwandtner U, Fuhr P, Riecher-Rössler A (2016): Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. World J Biol Psychiatry 17:285–295.
58. Cornblatt BA, Auther A, Mclaughlin D, Olsen RH, John M, Christoph U, *et al.* (2015): Psychosis prevention: A modified clinical high risk perspective from the recognition and prevention (RAP) program. Am J Psychiatry 172:986–994.
59. Francesconi M, Minichino A, Carrión RE, Delle Chiaie R, Bevilacqua A, Parisi M, *et al.* (2017): Psychosis prediction in secondary mental health services: A broad, comprehensive approach to the "at risk mental state" syndrome. Eur Psychiatry 40:96–104.
60. Fusar-Poli P, Cappucciati M, De Micheli A, Rutigliano G, Bonoldi I, Tognin S, *et al.* (2017): Diagnostic and prognostic significance of brief limited intermittent psychotic symptoms (BLIPS) in individuals at ultra high risk. Schizophr Bull 43:48–56.
61. Michel C, Ruhrmann S, Schimmelmann BG, Klosterkötter J, Schultze-Lutter F (2014): A stratified model for psychosis prediction in clinical practice. Schizophr Bull 40:1533–1542.
62. Nieman DH, Ruhrmann S, Dragt S, Soen F, Van Tricht MJ, Koelman JHTM, *et al.* (2014): Psychosis prediction: Stratification of risk estimation with information-processing and premorbid functioning variables. Schizophr Bull 40:1482–1490.
63. Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* (2014): Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. PLoS Med 11:1001744.
64. Dragt S, Nieman DH, Veltman D, Becker HE, van de Fliert R, de Haan L, Linszen DH (2011): Environmental factors and social adjustment as predictors of a first psychosis in subjects at ultra high risk. Schizophr Res 125:69–76.
65. Ruhrmann S, Schultze-Lutter F, Salokangas RKR, Heinimaa M, Linszen D, Dingemans P, *et al.* (2010): Prediction of psychosis in adolescents and young adults at high risk: Results from the

prospective European Prediction of Psychosis Study. JAMA Psychiatry 67:241–251.
66. Tarbox SI, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Perkins DO, *et al.* (2013): Premorbid functional development and conversion to psychosis in clinical high-risk youths. Dev Psychopathol 25:1171–1186.
67. Thompson A, Nelson B, Yung A (2011): Predictive validity of clinical variables in the "at risk" for psychosis population: International comparison with results from the North American Prodrome Longitudinal Study. Schizophr Res 126:51–57.
68. Van Tricht MJ, Nieman DH, Koelman JHTM, Van Der Meer JN, Bour LJ, De Haan L, Linszen DH (2010): Reduced parietal P300 amplitude is associated with an increased risk for a first psychotic episode. Biol Psychiatry 68:642–648.
69. Steyerberg EW, Harrell FE (2016): Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 69:245–247.
70. Koutsouleris N, Gaser C, Bottlender R, Davatzikos C, Decker P, Jäger M, *et al.* (2010): Use of neuroanatomical pattern regression to predict the structural brain dynamics of vulnerability and transition to psychosis. Schizophr Res 123:175–187.
71. Lavoie S, Berger M, Schlögelhofer M, Schäfer MR, Rice S, Kim SW, *et al.* (2017): Erythrocyte glutathione levels as long-term predictor of transition to psychosis. Transl Psychiatry 7:6–10.
72. van Tricht MJ, Ruhrmann S, Arns M, Müller R, Bodatsch M, Velthorst E, *et al.* (2014): Can quantitative EEG measures predict clinical outcome in subjects at clinical high risk for psychosis? A prospective multicenter study. Schizophr Res 153:42–47.
73. Hoffman RE, Woods SW, Hawkins KA, Pittman B, Tohen M, Preda A, *et al.* (2007): Extracting spurious messages from noise and risk of schizophrenia-spectrum disorders in a prodromal population. Br J Psychiatry 191:355–356.
74. Carrión RE, Cornblatt BA, Burton CZ, Tso IF, Auther AM, Adelsheim S, *et al.* (2016): Personalized prediction of psychosis: External validation of the NAPLS-2 psychosis risk calculator with the EDIPPP project. Am J Psychiatry 173:989–996.
75. Yung AR, McGorry PD (1996): The initial prodrome in psychosis: Descriptive and qualitative aspects. Aust N Z J Psychiatry 30:587–599.
76. Hall RCW (1995): Global Assessment of Functioning: A modified scale. Psychosomatics 36:267–275.
77. Sterne JAC, Egger M, Smith GD (2001): Investigating and dealing with publication and other biases in meta-analysis. Br J Med 323:101–105.
78. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B (2017): Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. NeuroImage 145B:166–179.
79. Snoek J, Larochelle H, Adams RP (2012): Practical Bayesian optimization of machine learning algorithms. arXiv:1206.2944.
80. Schnack HG, Kahn RS (2016): Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. Front Psychiatry 7:50.
81. Poldrack RA, Huckins G, Varoquaux G (2019): Establishment of best practices for evidence for prediction: A review [published online ahead of print Nov 27]. JAMA Psychiatry.
82. Howes OD, Murray RM (2014): Schizophrenia: An integrated sociodevelopmental-cognitive model. Lancet 383:1677–1687.
83. Bertolino A, Blasi G (2009): The genetics of schizophrenia. Neuroscience 164:288–299.
84. Cuesta MJ, Sánchez-Torres AM, Cabrera B, Bioque M, Merchán-Naranjo J, Corripio I, *et al.* (2015): Premorbid adjustment and clinical correlates of cognitive impairment in first-episode psychosis: The PEPsCog Study. Schizophr Res 164:65–73.
85. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, *et al.* (2014): Similarity network fusion for aggregating data types on a genomic scale. Nat Methods 11:333–337.
86. Royston P, Moons KGM, Altman DG, Vergouwe Y (2009): Prognosis and prognostic research: Developing a prognostic model. BMJ 338:b604.
87. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, Reininghaus U (2018): Experience sampling methodology in mental health research: New insights and technical developments. World Psychiatry 17:123–132.
88. Insel TR (2018): Digital phenotyping: A global tool for psychiatry. World Psychiatry 17:276–277.
89. Li H, Shapiro DI, Seidman LJ (2019): Handbook of Attenuated Psychosis Syndrome Across Cultures: International Perspectives on Early Identification and Intervention. Li H, Shapiro DI, Seidman LG, editors. Basel, Switzerland: Springer Nature.
90. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, *et al.* (2018): Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. JAMA Psychiatry 75:960–968.
91. Chung Y, Addington J, Bearden CE, Cadenhead K, Cornblatt B, Mathalon DH, *et al.* (2019): Adding a neuroanatomical biomarker to an individualized risk calculator for psychosis: A proof-of-concept study. Schizophr Res 208:41–43.
92. Kambeitz-Ilankovic L, Haas SS, Meisenzahl E, Dwyer DB, Weiske J, Peters H, *et al.* (2019): Neurocognitive and neuroanatomical maturation in the clinical high-risk states for psychosis: A pattern recognition study. NeuroImage Clin 21:101624.
93. Koutsouleris N, Davatzikos C, Borgwardt S, Gaser C, Bottlender R, Frodl T, *et al.* (2014): Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders. Schizophr Bull 40:1140–1153.
94. Perlis RH (2013): A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. Biol Psychiatry 74:7–14.
95. Lesh TA, Tanase C, Geib BR, Niendam TA, Yoon JH, Minzenberg MJ, *et al.* (2015): A multimodal analysis of antipsychotic effects on brain structure and function in first-episode schizophrenia. JAMA Psychiatry 72:226–234.
96. Schoeler T, Petros N, Di Forti M, Klamerus E, Foglia E, Murray R, Bhattacharyya S (2017): Poor medication adherence and risk of relapse associated with continued cannabis use in patients with first-episode psychosis: A prospective analysis. Lancet Psychiatry 4:627–633.
97. Fusar-Poli P, Schultze-Lutter F, Cappucciati M, Rutigliano G, Bonoldi I, Stahl D, *et al.* (2016): The dark side of the moon: Meta-analytical impact of recruitment strategies on risk enrichment in the clinical high risk state for psychosis. Schizophr Bull 42:732–743.
98. Fusar-Poli P (2017): Why ultra high risk criteria for psychosis prediction do not work well outside clinical samples and what to do about it. World Psychiatry 16:212–213.
99. Studerus E, Corbisiero S, Mazzariello N, Ittig S, Leanza L, Egloff L, *et al.* (2018): Can neuropsychological testing facilitate differential diagnosis between at-risk mental state (ARMS) for psychosis and adult attention-deficit/hyperactivity disorder (ADHD)? Eur Psychiatry 52:38–44.
100. Buchy L, Perkins D, Woods SW, Liu L, Addington J (2014): Impact of substance use on conversion to psychosis in youth at clinical high risk of psychosis. Schizophr Res 156:277–280.
101. Cannon TD, Cadenhead K, Cornblatt B, Woods SW, Addington J, Walker E, *et al.* (2008): Prediction of psychosis in youth at high clinical risk: A multisite longitudinal study in North America. Arch Gen Psychiatry 65:28–37.
102. DeVylder JE, Muchomba FM, Gill KE, Ben-David S, Walder DJ, Malaspina D, Corcoran CM (2014): Symptom trajectories and psychosis onset in a clinical high-risk cohort: The relevance of subthreshold thought disorder. Schizophr Res 159:278–283.

Individualized Diagnostic and Prognostic Models for Patients With Psychosis Risk Syndromes: A Meta-Analytic View on the State of the Art

SUPPLEMENTARY INFORMATION

Literature search

A comprehensive search of the published literature was conducted in the search engine Scopus using the following search string: ALL (((""risk for psychosis" OR "risk of psychosis" OR "at risk mental state*" OR "psychosis like" OR "risk state for psychosis" OR "brief limited intermitted psycho*" OR “psychosis prodrome” OR "psychosis risk syndrome") AND (“biomarker*” OR "predict*" OR "classif*" OR "outcome" OR "prognos*" OR "diagnos*" OR "transit*" OR "psychosis development" OR "psychosis onset") AND ("machine learning" OR "support vector machine" OR "SVM" OR "multivariate pattern recognition" OR "multivariate pattern analys*" OR "multivariate regression" OR "cox regression" OR "cox hazard")) AND LANGUAGE(english) AND PUBYEAR > 2000. Articles were considered only until June 30th 2019. The search in PubMed was more specific in order to enlarge the articles’ catchment pool, and thus included also the following search terms: "at risk for psychosis", "high risk for psychosis", "high risk of psychosis", "at risk mental state*", "ARMS", "ultra high risk for psychosis, "UHR", "clinical high risk for psychosis”, "psychosis like", "clinical high risk state for psychosis", "CHR" "familial high risk", "genetic risk", "brief limited intermitted psycho* symptoms", "BLIPS", “psychosis prodrome”, "psychosis risk syndrome", “neuroimaging”, "neuroanatomic*", "gray matter", "grey matter", "white matter", "cortical", "brain", "magnetic resonance imaging", "MRI", "functional MRI", "fMRI", "resting state", "diffusion tensor imaging", "DTI", "EEG", “neuropsychol*”, "neurocogni*", "cognit*", "attentio*", "language", "linguist*", "memory", "task switch*", "social cognit*", "intelligen*", "education*", “genetic*”, "polygenic", "polygenic risk score*", "stress*", "trauma*", "childhood trauma", "fatty acids", "obstetric complication*", "schizotypy", "substance use", and "environm*".

When reporting results, we followed the listed rules in order to allow comparability with other studies and further analyses: (I) In case of multimodal models (i.e., a combination of predictive variables), we additionally reported results from the single data modalities only if these could be similarly categorized to those in other studies; (II) neuroimaging models were reported only according to whole-brain analyses’ results (1); (III) in case several accuracy results were presented based on different features in the same data modality, we calculated and reported the performance’s average following previous research (2). The same

approach was used for one study (3) reporting results at different thresholds. In case there was more than 20% sample overlap between studies, we selected the most recent study with the biggest dataset. However, we included models constructed on the same sample if either the prognostic target or the data modality differed between them. Studies were excluded if the follow-up data was not complete and thus transition rate based on assumptions. One study needed to be excluded (4) because focusing on a sample of individuals undergoing CHR examination, thus not yet categorized as at-risk.

Data extraction

The following variables were extracted: title, authors, year of publication, demographic information (i.e., sample size, sex, age, at-risk category and diagnostic instruments for the risk status), type of data modality (i.e., biological, including neuroimaging and metabolites; clinical, including symptoms, functioning, neuropsychological and environmental measures; multimodal, a combination of different biomarker types), algorithm type, and cross-validation (CV) scheme, when applied. Prognostic target (i.e., transition or functioning) and time to follow-up were extracted for prognostic studies. Inconsistencies in the extraction results were discussed between R.S., D.D. and N.K.

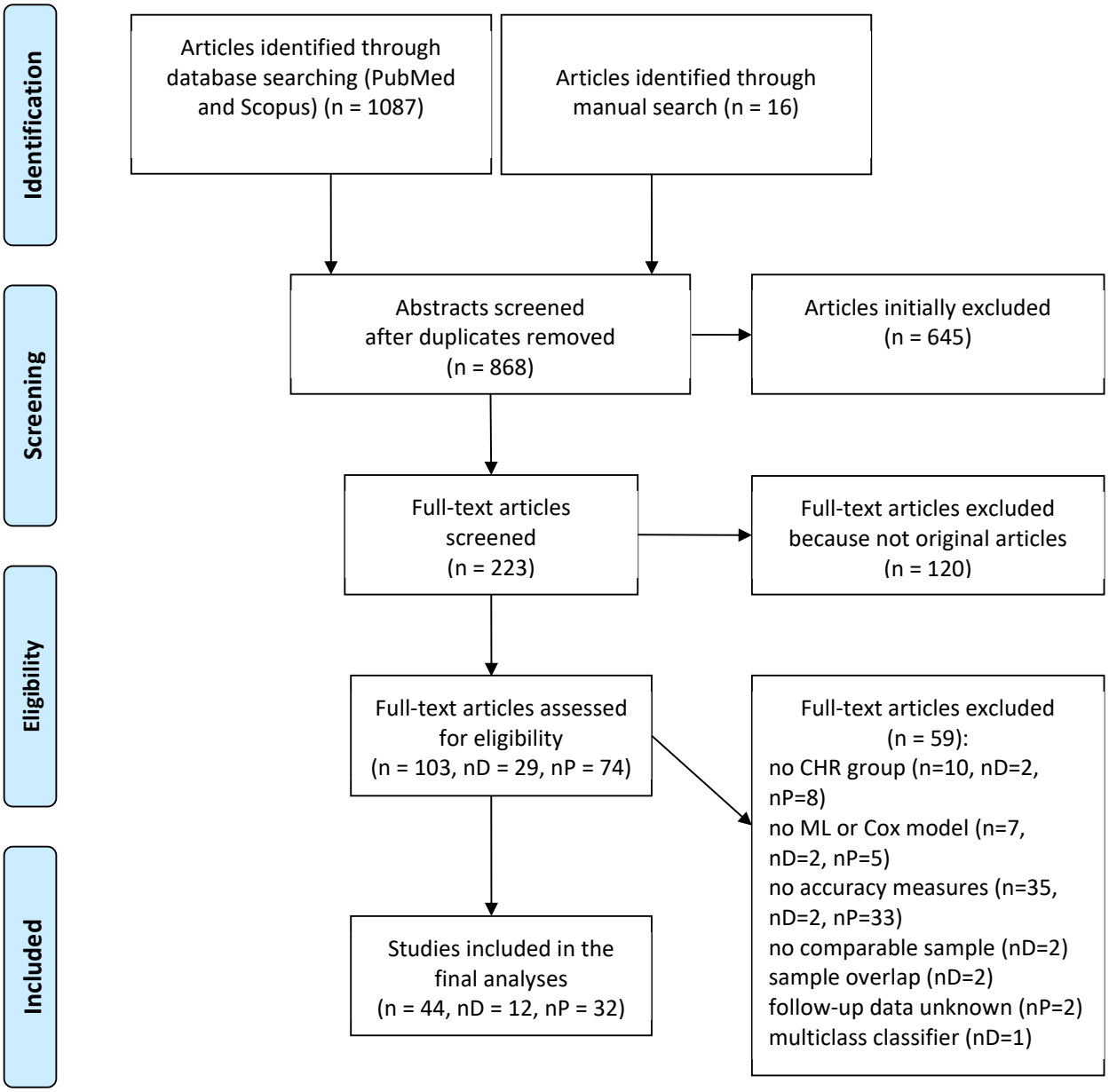


Figure S1: Flowchart illustrating details of the systematic literature search following the PRISMA guidelines (19)**.** Abbreviations: nD: number of diagnostic articles; nP: number of prognostic articles

Prognostic models																		
Study	**Non-converters**			**Converters**			**CHR type**	**CHR criteria**	**Data modality**	**Algorithm**	**Cross-Validation**	**Outcome**	**Follow-up (months)**	**SE**	**SP**	**BAC**	**PPV**	**NPV**	**PSI**
	N	**Age y**	**Males**	**N**	**Age y**	**Males**													
Amminger 2015	12	16.7	n.r.	28	15.8	n.r.	UHR	SIPS	Biological	GPC	LOOCV	Functioning	12	83	75	79	89	65	54
Bedi 2015	29	21.2	66%	5	22.2	80%	UHR	SIPS	Clinical	Convex Hull	LOOCV	Transition	30	100	100	100	100	100	100
Buchy 2014	141	19.8	58%	29	19.7	48%	UHR	SIPS	Clinical	Cox regression	No	Transition	48	69	81	75	43	33	35
Cannon 2008	209	n.r.	n.r.	82	n.r.	n.r.	UHR	SIPS	Clinical	Cox regression	No	Transition	30	67	53	60	34	80	16
Cannon 2016	512	n.r.	n.r.	84	n.r.	n.r.	UHR	SIPS	Multimodal	Cox regression	No	Transition	24	67	72	70	28	93	21
Carrión 2016	164	n.r.	n.r.	12	n.r.	n.r.	UHR	SIPS	Multimodal	Cox regression	No	Transition	24	58	73	66	14	96	10
									Biological					89	66	78	45	95	40
Chan 2015	58	22	61%	18	20	57%	UHR, UPS	CAARMS	Clinical	LASSO regression	10-fold CV	Transition	24	78	60	69	38	90	27
									Multimodal					89	79	84	57	96	53
Cornblatt 2015	66	n.r.	n.r.	26	n.r.	n.r.	UHR	SIPS	Multimodal	Cox regression	No	Transition	36	60	97	79	89	86	75
Das 2018	63	24	78%	16	26	50%	UHR, UPS	BSIP	Biological	Randomized trees	5-fold CV	Transition	20	66	97	82	85	92	77
									Biological					67	75	71	66	76	42
de Wit 2017	24	15.4	76%	17	15.9	58%	UHR, BS	SIPS, SPI-A	Clinical	SVM	LOOCV	Functioning	72	76	75	76	68	82	50
									Multimodal					68	81	75	72	78	50
DeVylder 2014	74	20.1	81%	26	20	76%	UHR	SIPS	Clinical	Cox regression	No	Functioning	30	58	60	59	34	80	14
Dragt 2011	53	18.9	62%	19	20.3	74%	UHR	SIPS	Clinical	Cox regression	No	Transition	36	50	91	71	67	84	50
Francesconi 2017	95	24.4	52%	21	24.1	57%	UHR	CAARMS	Clinical	Cox regression	No	Transition	36	67	97	82	83	93	76
Fusar-Poli 2017	525	n.r.	n.r.	185	n.r.	n.r.	UHR-BLIPS	SIPS	Clinical	LASSO Cox regression	k-fold CV	Transition	48	24	93	58	53	94	30
Gothelf 2011	9	n.r.	n.r.	10	n.r.	n.r.	22q11.2DS	-	Biological	SVM	LOOCV	Transition	60	90	100	95	100	90	90
Hoffman 2007	9	n.r.	n.r.	19	n.r.	n.r.	UHR	SIPS	Clinical	Cox regression	No	Transition	24	89	89	89	80	94	74

Kambeitz-Ilankovic 2016	13	23.2	62%	14	23.5*	79%	UHR, BS	SIPS, SPI-A	Biological	SVM	Nested LOOCV	Functioning	48	79	85	82	85	79	64
Koutsouleris 2012	20	25.8	70%	15	22.8	73%	UHR, BS	SIPS, SPI-A	Clinical	SVM	Repeated double CV	Transition	48	80	75	78	71	83	54
									Biological					67	47	57	65	49	14
Koutsouleris 2018	47	24.4	47%	69	23.7	52%	UHR, BS	SIPS, SPI-A	Clinical	SVM	Nested LSOCV	Functioning	12	61	75	68	78	57	35
									Multimodal					59	70	65	74	54	28
									Biological					80	72	76	79	73	52
Koutsouleris 2018	50	24.5	48%	66	23.6	52%	UHR, BS	SIPS, SPI-A	Clinical	SVM	Nested LSOCV	Functioning	12	70	84	77	85	68	53
									Multimodal					83	82	83	86	79	64
Koutsouleris 2015	33	24.6	61%	33	25	73%	UHR, BS	SIPS, SPI-A	Biological	SVM	Repeated double CV	Transition	53	76	85	81	84	78	61
Lavoie 2017	21	16.4	29%	15	14.9	33%	UHR	CAARMS	Biological	Cox regression	No	Transition	84	91	67	79	66	91	58
Mechelli 2017	99	19.4	49%	99	19.5	49%	UHR	CAARMS	Clinical	SVM	LOOCV	Transition	90	69	61	65	64	66	30
	48	19.7	46%	48	19.7	46%						Functioning		63	63	63	63	63	26
Michel 2014	53	25.3	64%	44	24.1	66%	UHR, BS	SIPS, SPI-A	Clinical	Cox regression	No	Transition	24	57	55	56	51	61	12
Nieman 2014	43	19	63%	18	20.3	72%	UHR, BS	SIPS, SPI-A	Multimodal	Cox regression	No	Transition	36	78	88	83	73	91	64
Perkins 2015	40	19.5	63%	32	19.2	70%	UHR	SIPS	Biological	Greedy algorithm	5-fold CV	Transition	24	60	90	75	83	74	57
Ramyead 2016	35	25.8	66%	18	26.7	56%	UHR, UPS	BSIP	Biological	LASSO	Repeated nested CV	Transition	36	58	83	71	64	79	43
Ruhrmann 2010	146	n.r.	n.r.	37	n.r.	n.r.	UHR, BS	SIPS, SPI-A	Clinical	Cox regression	No	Transition	18	42	98	70	84	87	71
Tarbox 2013	192	17.9	61%	78	18.4	55%	UHR	SIPS	Clinical	Cox regression	No	Transition	30	62	61	62	39	80	19
Thompson 2011	63	19.3	49%	41	19.5	49%	UHR	CAARMS	Clinical	Cox regression	No	Transition	28	30	89	60	64	68	30
van Tricht 2010	43	19.3	67%	18	20.4	72%	UHR	SIPS	Biological	Cox regression	No	Transition	36	46	87	67	60	79	39
van Tricht 2014	91	22.0	64%	22	21.8	64%	UHR, BS	SIPS, SPI-A	Biological	Cox regression	No	Transition	18	83	79	81	49	95	44
Zarogianni 2017	22	23.9	56%	13	26.8	33%	UHR, BS	SIPS, SPI-A	Multimodal	SVM	nested LOOCV	Transition	48	63	84	74	70	79	49

Zarogianni 2017[b]	57	20.1	65%	17	20.8	35%	FR	-	Biological	SVM	nested LOOCV	Transition	77	76	77	77	50	91	41
									Multimodal					100	83	92	64	100	64
Classification models																			
Study	**Healthy Controls**			**CHR**			**CHR type**	**CHR instruments**	**Data modality**	**Algorithm**	**Cross-Validation**	**Outcome**		**SE**	**SP**	**BAC**	**PPV**	**NPV**	**PSI**
	N	**Age y**	**Males**	**N**	**Age y**	**Males**													
Bendfeldt 2015	19	n.r.	n.r.	19	n.r.	n.r.	UHR, UPS	BSIP	Biological	SVM	LOOCV	Diagnosis	-	74	58	66	64	69	33
Guo 2014	60	27.2	58%	28	25.8	54%	FR	-	Biological	SVM	LOOCV	Diagnosis	-	60	94	77	56.4	3.6	66
Koutsouleris 2012	48	26	60%	30	24.7	67%	UHR, BS	SIPS, SPI-A	Clinical	SVM	Repeated double CV	Diagnosis	-	96	80	88	88	93	81
Koutsouleris 2009	25	n.r.	n.r.	45	n.r.	76%, 62%	UHR, BS	SIPS, SPI-A	Biological	SVM	5-fold CV	Diagnosis	-	89	80	85	89	80	69
Liu 2012	25	25.5	55%	22	25.6	56%	FR	-	Biological	SVM	LOOCV	Diagnosis	-	72	86	79	85	73	58
Pettersson-Yeo 2013	19	23.3	47%	19	22.4	47%	UHR	CAARMS	Biological	SVM	LOOCV	Diagnosis	-	80	73	77	75	78	53
Scariati 2014	41	18.2	49%	42	18.2	57%	22q11.2DS	-	Biological	SVM	LOOCV	Diagnosis	-	81	88	85	87	82	69
Studerus 2018	101	25	57%	101	25.4	70%	UHR, UPS	BSIP	Clinical	Random forest	10-fold CV	Diagnosis	-	73	77	75	76	74	50
Tylee 2017	56	21	53%	30	20.9	54%	22q11.2DS	-	Biological	SVM	5-fold CV	Diagnosis	-	85	82	84	90	75	64
Valli 2016	25	25.1	56%	25	23.8	72%	UHR	CAARMS	Biological	SVM	LOOCV	Diagnosis	-	68	76	72	74	70	44
Wang 2016	37	20.8	49%	34	21.5	62%	UHR	CAARMS	Biological	SVM	LOOCV	Diagnosis	-	82	69	76	71	81	51
Zhu 2019	74	21.4	51%	71	22	58%	UHR	SIPS	Biological	SVM	5-fold CV	Diagnosis	-	77	47	62	60	66	26

Table S1: Summary of the studies included in the present meta-analysis. Abbreviations: 22q11.2DS: 22q11.2 deletion syndrome, ARMS: At Risk Mental State, BAC: balanced accuracy, BS: basic symptoms, BSIP: Basel Screening Instrument for Psychosis, CAARMS: Comprehensive Assessment of at Risk Mental States, CHR: Clinical High Risk, CV: cross-validation, FR: familial risk, GPC: Gaussian process classification, LASSO: least absolute shrinkage and selection operator, LOOCV: leave-one-out cross-validation, LSOCV: leave-one-site-out cross-validation, NPV: negative predictive value, n.r.: not reported, PPV: positive predictive value, PSI: prognostic summary index, SE: sensitivity, SIPS: Structure Interview for Psychosis-Risk Syndromes, SP: specificity, SPI-A: Schizophrenia Proneness Instrument-Adult version, SVM: support vector machine, UHR: Ultra High Risk, UPS: unspecific prodromal symptoms.

Clinical characteristics of CHRs

CHRs in about 86% of the studies fulfilled the UHR criteria (5), operationalized by either the Structured Interview for Psychosis–Risk Syndromes (SIPS (6)), the Comprehensive Assessment of at Risk Mental States (CAARMS (7)), or the Basel Screening Instrument for Psychosis (BSIP (8)) (see (9) for a comprehensive review). Of those studies, 34% also included individuals fulfilling basic symptoms (BS, based on the Schizophrenia Proneness Instruments (10) or the Bonn Scale for the Assessment of Basic Symptoms (BSAB-S, (11)) or unspecific prodromal symptoms (UPS) based on the BSIP.

Moreover, in order to maximize our meta-analytical sample size, we included three studies on 22q11.2 deletion syndrome (12–14) and three on persons with familial risk for psychosis (3, 15, 16), i.e., having a first degree relative with a psychotic disorder. 22q11.2 deletion syndrome is a genetic disease caused by a chromosomic deletion and leading to both hereditary disfunctions (e.g., congenital heart disease) and later-onset behavioral and psychiatric illnesses, one of which being psychosis (17). Moderator analyses for type of CHR (i.e., either clinically assessed versus FR or based on different clinical subgroups) were not conducted because of the sample's heterogeneity.

Investigation of publication bias

Publication bias was investigated by calculating the effective sample size (ESS) and plotting the inverse square root of this term against the logarithm of the DOR (lnDOR); the bias can be then assessed with a regression between the measures, weighted by the ESS. Publication bias on SE was investigated using funnel plots with both the inverse square root of ESS and the raw sample size.

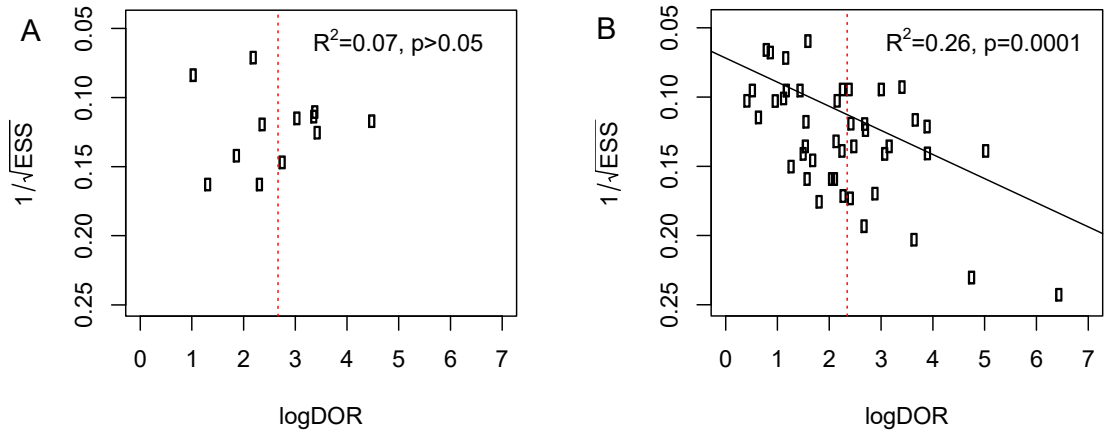


Figure S2: Funnel plots for publication bias. Plots show the effect of the inverted square root of the effective sample size (ESS) against the logarithm of diagnostic odds ratios (logDOR). A: classification studies; B: prognostic studies.

Supplementary analyses

Sample-split based on Cross-Validation

Within prognostic studies, we found a significant interaction effect of sample size and cross-validation (CV) on sensitivity (p = .009, see main results in the manuscript). Because the CV factor completely overlapped with the models' algorithm (i.e., all ML studies performed CV, while no study using Cox did), we decided to further investigate heterogeneity by splitting the sample of prognostic models based on application of CV. We computed bivariate regressions with the following moderators: age, sex, year of publication, time to follow-up and prognostic target (the latter two for ML models only). Analyses of the effects of data modality or CHR type, type of CV scheme or being a multisite study could not be computed because there were less than 10 models for each variable (18). The only effect surviving correction for multiple comparisons was that of time to follow-up examination on sensitivity performance when controlling for sample size in Cox, or not cross-validated, models (p = .003, $\chi^2_{(2)} = 6.85$, p = 0.03, see Table S2 and Figure S3).

Classification studies										
	Sensitivity					False Positive Rate (1-Specificity)				
Covariates	**Estimate**	**z**	**p**	**95%-CI**		**Estimate**	**z**	**p**	**95%-CI**	
Age (years)	-0.092	-1.072	0.1247	-0.164	0.025	-0.07	-1.449	0.284	-0.26	0.076
Sex (% males)	0.02	1.168	0.243	0	0.052	-0.012	-0.452	0.651	-0.063	0.04
Year of publication	0.058	0.3	0.765	-0.319	0.434	-0.264	-1.271	0.204	-0.671	0.143
Prognostic studies										
	Sensitivity					False Positive Rate (1-Specificity)				
Covariates	**Estimate**	**z**	**p**	**95%-CI**		**Estimate**	**z**	**p**	**95%-CI**	
Age (years)	-0.013	-0.44	0.66	-0.07	0.045	-0.035	-0.825	0.409	-0.118	0.048
Sex (% males)	-0.007	-0.782	0.434	-0.025	0.01	-0.016	-1.333	0.183	-0.04	0
Year of publication	0.03	0.843	0.399	-0.04	0.101	0.032	0.598	0.55	-0.073	0.137
Data modality (clinical)	-0.263	-1.367	0.172	-0.639	0.114	0.146	0.516	0.606	-0.408	0.699
Algorithm/CV	0.451	2.627	0.009	0.115	0.788	0.345	1.34	0.18	-0.16	0.85
Follow-up interval (months)	0.002	0.471	0.637	-0.005	0.008	0.005	1.025	0.305	-0.005	0.015
Prognostic target	-0.112	-0.568	0.57	-0.496	0.273	0.487	1.721	0.085	-0.068	1.043
Multisite studies	-0.005	-0.028	0.978	-0.354	0.344	-0.306	-1.095	0.274	-0.852	0.241
Machine learning/CV studies										
	Sensitivity					False Positive Rate (1-Specificity)				
Covariates	**Estimate**	**z**	**p**	**95%-CI**		**Estimate**	**z**	**p**	**95%-CI**	
Age (years)	-0.008	-0.217	0.828	-0.079	0.063	0.006	2.303	0.021*	0.001	0.011
Sex (% males)	-0.014	-1.304	0.192	-0.034	0	-0.024	-2.119	0.034*	-0.045	0
Year of publication	-0.097	-1.328	0.184	-0.24	0.046	-0.05	-0.591	0.554	-0.216	0.116
Follow-up interval (months)	-0.001	-0.235	0.814	-0.007	0.002	0.002	0.51	0.61	-0.006	0.009

Prognostic target	-0.095	-0.503	0.615	-0.463	0.274	0.236	1.07	0.284	-0.197	0.669
Cox regression/not CV studies										
	Sensitivity					**False Positive Rate (1-Specificity)**				
Covariates	**Estimate**	**z**	**p**	**95%-CI**		**Estimate**	**z**	**p**	**95%-CI**	
Age (years)	-0.017	-0.349	0.727	-0.112	0.078	-0.037	-0.42	0.675	-0.207	0.134
Sex (% males)	-0.006	-0.326	0.745	-0.043	0.03	0.002	0.082	0.935	-0.051	0.05
Year of publication	0.087	2.031	0.042*	0.003	0.172	0.053	0.0633	0.527	-0.111	0.217
Follow-up interval (months)	0.029	2.983	0.003**	0.01	0.047	0.012	0.772	0.44	-0.019	0.043
Multisite studies	0.263	0.769	0.442	-0.408	0.935	-0.694	-1.141	0.254	-1.886	0.498

Table S2: Moderator analyses. *not significant after False Discovery Rate (FDR) correction; **significant after FDR correction. Abbreviations: CHR: Clinical High Risk; CV: cross-validation.

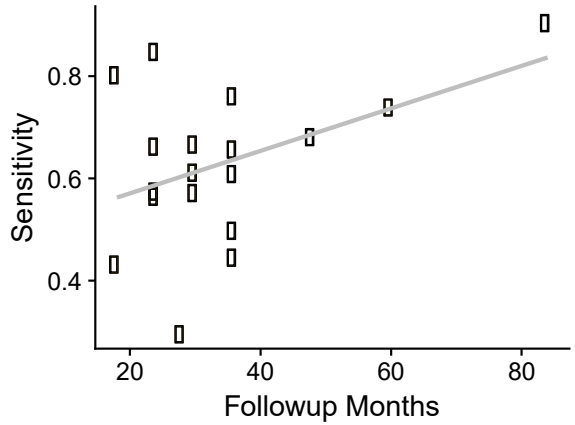


Figure S3: Effects of time-to-follow-up (in months) on sensitivity in models using Cox regressions.

Exclusion of 22q11.2DS and FR

22q11.2DS and subjects with FR are conceptually different types of risk definitions and are either potentially separate syndromes (e.g., 22q11.2DS) or are a single component of a larger clinical high-risk definition (e.g., familial risk). To investigate the effects of their exclusion on the meta-analytic results, we excluded them and repeated the bias analysis and receiver operator characteristic (ROC) curves (Figures S4 and S5). Results demonstrated an R squared change of 0.01 for both diagnostic (Figure S4, A) and prognostic (Figure S4, B) studies. ROC curves demonstrated minimal changes in sensitivity and false-positive rates.

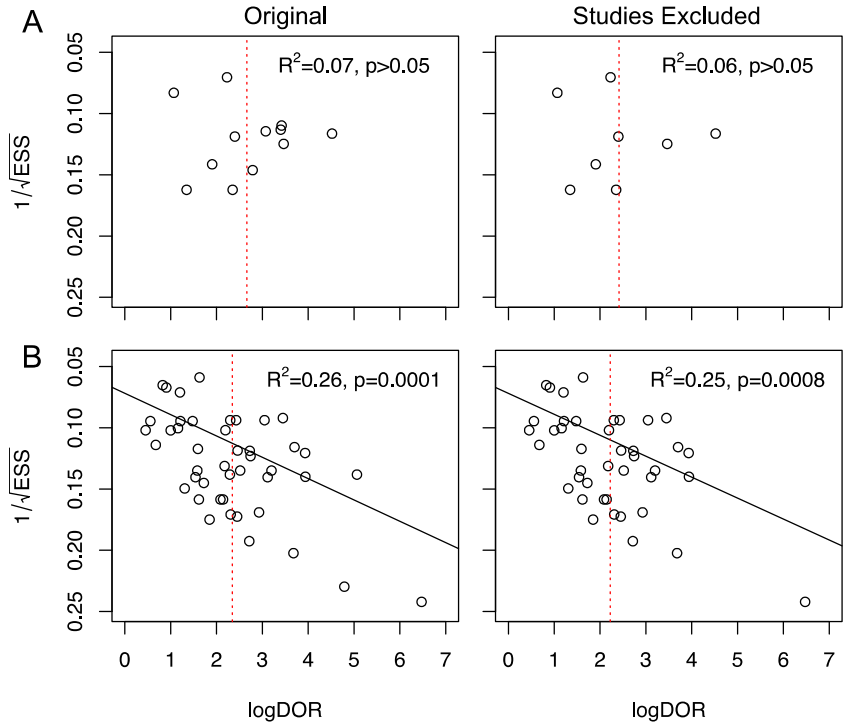


Figure S4: Comparison of original bias plots and plots after exclusion of studies related to 22q.11 syndrome and familial risk. A) diagnostic bias minimally changes between the original analysis (left) and when the studies are excluded (right); B) prognostic bias also minimally changes.

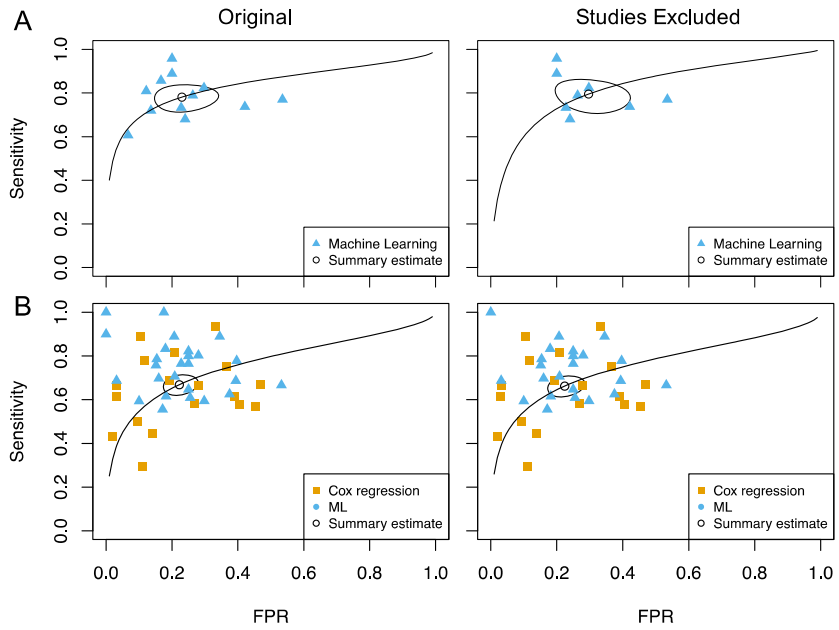


Figure S5: Comparison of receiver operating characteristic (ROC) curves after exclusion of 22q11 and familial risk studies. A) Diagnostic ROC curves demonstrate minimal change when the original analysis (left) is compared with the analysis after study exclusion (right). B) Prognostic ROC curves demonstrate minimal change after study exclusion.

Supplementary References

1. Bendfeldt K, Smieskova R, Koutsouleris N, Klöppel S, Schmidt A, Walter A, *et al.* (2015): Classifying individuals at high-risk for psychosis based on functional brain activity during working memory processing. *NeuroImage Clin*. 9: 555–563.
2. Kambeitz J, Cabral C, Sacchet MD, Gotlib IH, Zahn R, Serpa MH, *et al.* (2017): Detecting Neuroimaging Biomarkers for Depression: A Meta-analysis of Multivariate Pattern Recognition Studies. *Biol Psychiatry*. 82: 330–338.
3. Guo S, Palaniyappan L, Yang B, Liu Z, Xue Z, Feng J (2014): Anatomical distance affects functional connectivity in patients with schizophrenia and their siblings. *Schizophr Bull*. 40: 449–459.
4. Fusar-Poli P, Rutigliano G, Stahl D, Schmidt A, Ramella-Cravaro V, Hitesh S, McGuire P (2016): Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. *JAMA Psychiatry*. 73: 1260–1267.
5. Yung AR, McGorry PD (1996): The initial prodrome in psychosis: Descriptive and

qualitative aspects. *Aust N Z J Psychiatry*. 30: 587–599.

6. Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Ventura J, McFarlane W, *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.
7. Yung AR, Yung AR, Yuen HP, Mcgorry PD, Lisa J, Kelly D, *et al.* (2005): Mapping the onset of psychosis : the Comprehensive Assessment of At-Risk Mental States. 8674.
8. Riecher-Rössler A, Gschwandtner U, Aston J, Borgwardt S, Drewe M, Fuhr P, *et al.* (2007): The Basel early-detection-of-psychosis (FEPSY)-study - Design and preliminary results. *Acta Psychiatr Scand*. doi: 10.1111/j.1600-0447.2006.00854.x.
9. Fusar-Poli P, Borgwardt S, Bechdolf A, Addington J, Riecher-Rössler A, Schultze-Lutter F, *et al.* (2013): The psychosis at risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry*. 70: 107–120.
10. Schultze-Lutter F, Ruhrmann S, Fusar-Poli P, Bechdolf A, G. Schimmelmann B, Klosterkotter J (2012): Basic Symptoms and the Prediction of First-Episode Psychosis. *Curr Pharm Des*.
11. Klosterkötter J, Gross G, Huber G, Wieneke A, Steinmeyer EM, Schultze-Lutter F (1997): Evaluation of the “Bonn Scale for the assessment of basic symptoms - BSABS” as an instrument for the assessment of schizophrenia proneness: A review of recent findings. *Neurol Psychiatry Brain Res*.
12. Gothelf D, Hoeft F, Ueno T, Sugiura L, Lee AD, Thompson P, Reiss AL (2011): Developmental changes in multivariate neuroanatomical patterns that predict risk for psychosis in 22q11.2 deletion syndrome. *J Psychiatr Res*. 45: 322–331.
13. Scariati E, Schaer M, Richiardi J, Schneider M, Debbané M, Van De Ville D, Eliez S (2014): Identifying 22q11.2 Deletion Syndrome and Psychosis Using Resting-State Connectivity Patterns. *Brain Topogr*. 27: 808–821.
14. Tylee DS, Kikinis Z, Quinn TP, Antshel KM, Fremont W, Tahir MA, *et al.* (2017): NeuroImage : Clinical Machine-learning classi fi cation of 22q11 . 2 deletion syndrome : A di ff usion tensor imaging study.
15. Liu M, Zeng L-L, Shen H, Liu Z, Hu D (2012): Potential risk for healthy siblings to develop schizophrenia: evidence from pattern classification with whole-brain connectivity. *Neuroreport*. 23.
16. Zarogianni E, Storkey AJ, Johnstone EC, Owens DGC, Lawrie SM (2017): Improved individualized prediction of schizophrenia in subjects at familial high risk, based on

neuroanatomical data, schizotypal and neurocognitive features. *Schizophr Res*. 181: 6–12.

17. McDonald-McGinn DM, Sullivan KE, Marino B, Philip N, Swillen A, Vorstman JAS, *et al.* (2015): 22Q11.2 Deletion Syndrome. *Nat Rev Dis Prim*. 1.
18. Borenstein M, Hedges L V (2009): *Introduction to Meta-analysis*. Hoboken, NY: John Wiley & Sons;
19. Moher D, Liberati A, Tetzlaff J, Altman DG (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*.

6. Paper II

Novel gyrification networks reveal links with psychiatric risk factors in early illness

Evidence of altered gyrification has been found in clinical populations diagnosed with schizophrenia, depression, and psychosis risk states. Such findings may reflect a developmental-related, transdiagnostic, signature, but this hypothesis has not been investigated yet and existing studies may be methodologically limited.

Thus, we aimed to derive gyrification-specific covariance maps in order to investigate associations with symptoms, cognition, and functioning in a sample of individuals in early illness stages. A recently introduced, data-driven method, Orthogonal Projective Non-Negative Matrix Factorization, delineated gyrification-based Patterns of Structural Covariance (PSC) in 308 healthy controls. The PSC-map was applied to a sample of patients with recent onset psychosis or depression, and clinical high-risk for psychosis (N=713). Gyrification differences compared to controls were determined, and associations with diagnosis, symptoms, cognition, and functioning were investigated using linear models.

We detected 18 PSCs in controls, the majority of which were externally validated in an independent healthy sample (N=84). PSCs differed between patients and controls in temporal-insular, lateral occipital, and lateral fronto-parietal areas ($p^{FDR}<0.01$). Gyrification abnormalities were observable in high-risk, psychotic, and early depression patients. Altered cortical folding demonstrated associations with cognitive domains and role functioning, but not with symptomatology.

Our findings highlight a sparse representation of cortical gyrification in controls, which is altered in early psychiatric illnesses and high-risk individuals and is not associated with symptom severity. A neurodevelopmentally-linked signature was suggested by relationships with cognition and lifetime role functioning. Further studies are required to delineate how and to what extent gyrification might add important information within predictive models by expressing early insults at a neurobiological level, which signal common features of mental illness.

Cerebral Cortex, 2021;00: 1–12

https://doi.org/10.1093/cercor/bhab288
Original Article

ORIGINAL ARTICLE

Novel Gyrification Networks Reveal Links with Psychiatric Risk Factors in Early Illness

Rachele Sanfelici[1,2], Anne Ruef[1], Linda A. Antonucci[1,3], Nora Penzel[4], Aristeidis Sotiras[5], Mark Sen Dong[1], Maria Urquijo-Castro[1], Julian Wenzel[4], Lana Kambeitz-Ilankovic[1,4], Meike D. Hettwer[2], Stephan Ruhrmann[4], Katharine Chisholm[6,7], Anita Riecher-Rössler[8], Peter Falkai[1,9], Christos Pantelis[10], Raimo K. R. Salokangas[11], Rebekka Lencer[12,13], Alessandro Bertolino[3], Joseph Kambeitz[4], Eva Meisenzahl[14], Stefan Borgwardt[13,15], Paolo Brambilla[16,17], Stephen J. Wood[18,19,20], Rachel Upthegrove[6,21], Frauke Schultze-Lutter[14,22,23], Nikolaos Koutsouleris[1,9,24,†], Dominic B. Dwyer[1,†] and the PRONIA Consortium[‡]

[1]Department of Psychiatry and Psychotherapy, Ludwig-Maximilian University, Munich, 80336, Germany, [2]Max Planck School of Cognition, Leipzig, 04103, Germany, [3]Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, 70124, Italy, [4]Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, 50937, Germany, [5]Department of Radiology and Institute of Informatics, Washington University in St. Luis, st. Luis, MO63110, USA, [6]Institute for Mental Health, University of Birmingham, Birmingham, B15 2TT, UK, [7]Department of Psychology, Aston University, Birmingham, B4 7ET, UK, [8]Medical Faculty, University of Basel, Basel, 4051, Switzerland, [9]Max-Planck Institute of Psychiatry, Munich, 80804, Germany, [10]Melbourne Neuropsychiatry Centrem University of Melbourne & Melbourne Health, Melbourne, 3053, Australia, [11]Department of Psychiatry, University of Turku, Turku, 20700, Finland, [12]Department of Psychiatry and Psychotherapy, University of Münster, Münster, 48149, Germany, [13]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, 23538, Germany, [14]Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, 40629, Germany, [15]Department of Psychiatry (Psychiatric University Hospital, UPK), University of Basel, Basel, 4002, Switzerland, [16]Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Grande Ospedale Maggiore Policlinico, Milano, 20122, Italy, [17]Department of Pathophysiology and Transplantation, University of Milan, Milan, 20122, Italy, [18]Centre for Youth Mental Health, University of Melbourne, Melbourne, 3052, Australia, [19]Orygen, Melbourne, 3052, Australia, [20]School of Psychology, University of Birmingham, Birmingham, B15 2TT, UK, [21]Early Intervention Service, Birmingham Women's and Children's NHS foundation Trust, Birmingham, B4 6NH, UK, [22]Department of Psychology and Mental Health, Faculty of Psychology, Airlangga University, Surubaya, 60286, Indonesia, [23]University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, 3000, Switzerland and [24]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, UK

Address correspondence to Nikolaos Koutsouleris, Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Nussbaumstr. 7, D-80336 Munich, Germany. Email: nikolaos.koutsouleris@med.uni-muenchen.de

†Nikolaos Koutsouleris and Dominic B. Dwyer have equally contributed.

‡See PRONIA consortium author list.

Abstract

Adult gyrification provides a window into coordinated early neurodevelopment when disruptions predispose individuals to psychiatric illness. We hypothesized that the echoes of such disruptions should be observed within structural gyrification networks in early psychiatric illness that would demonstrate associations with developmentally relevant variables rather than specific psychiatric symptoms. We employed a new data-driven method (Orthogonal Projective Non-Negative Matrix Factorization) to delineate novel gyrification-based networks of structural covariance in 308 healthy controls. Gyrification within the networks was then compared to 713 patients with recent onset psychosis or depression, and at clinical high-risk. Associations with diagnosis, symptoms, cognition, and functioning were investigated using linear models. Results demonstrated 18 novel gyrification networks in controls as verified by internal and external validation. Gyrification was reduced in patients in temporal-insular, lateral occipital, and lateral fronto-parietal networks ($p^{FDR} < 0.01$) and was not moderated by illness group. Higher gyrification was associated with better cognitive performance and lifetime role functioning, but not with symptoms. The findings demonstrated that gyrification can be parsed into novel brain networks that highlight generalized illness effects linked to developmental vulnerability. When combined, our study widens the window into the etiology of psychiatric risk and its expression in adulthood.

Key words: clinical high risk, cortical folding, depression, psychosis, structural covariance

Introduction

Gyrification (i.e., the degree of cortical folding) is a fundamental property of the human brain, which primarily arises from a complex interplay of both genetic and biological (Borrell 2018), as well as biomechanical factors (Kroenke and Bayly 2018) acting during early cortical development that peaks at week 66/80 post-conception (Llinares-Benadero and Borrell 2019). Recent evidence also shows that glial cells formation and migration in the cortex and subcortical white matter influences to a great extent the formation of cortical folds (Rash et al. 2019). Associations with structural and functional connectivity reflect these interactions (White and Hilgetag 2011), while also highlighting the contribution of environmental factors that influence a protracted developmental course extending throughout adolescence (Cao et al. 2017). Beyond the consistently homogeneous primary gyri (i.e., convex ridges) and sulci (i.e., concave grooves) (Lohmann et al. 2008), whose morphology is only impacted by severe neurodevelopmental disorders (Barkovich 2010), this continued complex interplay is reflected in more fine-grained folding patterns showing high interindividual variability (Glasser et al. 2017) that is related to psychiatric illness (Guo et al. 2015).

Pertinent current questions relate to the morphology of gyrification abnormalities in mental illness, their psychiatric manifestation, and their origin. For example, both increased and decreased gyrification have been found in established psychoses (Palaniyappan et al. 2011; Nanda et al. 2014; Matsuda and Ohi 2018) and across depression and bipolar disorder (Depping et al. 2018). While there may be relationships between gyrification and illness-specific psychiatric symptoms (Matsuda and Ohi 2018), transdiagnostic cognitive disturbances with a putative neurodevelopmental basis could also mediate these relationships (Cao et al. 2017; Popovic et al. 2020). This hypothesis is supported by associations between gyrification and cognitive ability across species (Pillay and Manger 2007; Gautam et al. 2015; Gregory et al. 2016), the coexistence of cognitive and gyrification abnormalities in neurodevelopmental disorders (Kippenhan et al. 2005; Wallace et al. 2013) and adults born very preterm (Papini et al. 2020), and by relationships found between fronto-temporal folding and cognition in affective psychoses (Rodrigue et al. 2018). Neurodevelopmental contributions are also suggested by folding differences in individuals at clinical high-risk for psychosis (Sasabayashi et al. 2017) who develop a first episode (Das et al. 2018) and by the effects of perinatal stress on altered gyrification and mood disturbances (Mareckova et al. 2020). Notably, no study has yet investigated how gyrification relates to other clinical aspects usually impaired in mental illnesses, such as everyday functioning across the lifetime (i.e., social and occupational). Late adolescence and early adulthood might be the most informative timeframes to investigate such transdiagnostic cognitive and functioning phenomena because of the high comorbidity rate of symptoms, especially in the initial phases of psychiatric illness (Musliner et al. 2019; Thapar and Riglin 2020). However, existing results remain inconclusive, differing both quantitatively and qualitatively across studies (Matsuda and Ohi 2018).

A potential reason for inconsistent results could be the use of techniques that cannot harness the highly interconnected gyrification system of the brain, such as traditional brain atlases, which are based on coarse anatomical characteristics, or mass univariate vertex-wise approaches. Whole-brain structural covariance methods (Alexander-Bloch et al. 2013; Evans 2013) might be more powerful in this regard because they are known to produce cortical maps that are highly heritable, related to behavioral variation, and have their origin in coordinated developmental processes (Alexander-Bloch et al. 2013). Seed-based techniques have been the most popular in this domain where an existing brain atlas is used to discover relationships between nodes (Bassett et al. 2008; Yeh et al. 2010; Van Den Heuvel et al. 2013; Wang et al. 2016), yielding insights regarding gyrification in patients with established schizophrenia

(Palaniyappan et al. 2015, 2016). However, these studies involve an assumption that traditional atlases (e.g., the Desikan-Killiany) are valid for gyrification networks and this may not be the case. Hypothesis-free, data-driven techniques may thus be best suited to obtain gyrification-specific maps. A preliminary study has been conducted using principal components analysis (PCA) (Das et al. 2018), but this technique is limited both in terms of replicability and by its fuzzy representation of variance, which cannot define clear boundaries between components (Sotiras et al. 2015).

In this study, we first aimed to delineate gyrification covariance patterns in healthy control individuals using a novel Non-Negative Matrix Factorization (NNMF) technique, following research investigating cortical thickness (Sotiras et al. 2017). NNMF produces a parsimonious "alphabet" of clearly separated and well-defined variance components (Yang and Oja 2010; Sotiras et al. 2015), thus overcoming limitations of other dimensionality reduction methods by enhancing results' interpretability. We then aimed to assess developmental and sex effects on data-driven components, before investigating gyrification disruptions in early stages of psychosis and depression (i.e., recent-onset psychosis, ROP, and recent-onset depression, ROD) and in clinical high-risk (CHR) individuals. Merging these three unique patient populations allowed us to investigate whether potential gyrification abnormalities were shared across clinical manifestations and/or comorbidities. Further, we explored the specificity of the abnormalities in patients in relationship to diagnosis, symptom severity (i.e., psychosis, depression, and subjective cognitive disturbances), cognition, and occupational and social functioning. We expected to find novel gyrification covariance patterns in controls that demonstrated shared spatial abnormalities across clinical groups and were associated with cognition and functioning.

Materials and Methods

Participants

A total of 413 healthy controls and 901 patients with ROP, ROD, or CHR were recruited within the PRONIA study (www.pronia.eu), an international longitudinal project conducted across seven European sites (Supplementary Information, Koutsouleris et al. 2018). Previous work on the PRONIA sample investigated gray matter volume for diagnostic and prognostic purposes (Koutsouleris et al. 2018, 2020; Upthegrove et al. 2020) and its associations with childhood trauma (Popovic et al. 2020). A subsample of 329 controls and 754 patients was selected from five out of seven sites (Munich, Cologne, Basel, Turku, and Udine) based on the availability of surface-based neuroimaging data and the need to match sex and age during site correction (see Supplementary Information). The two excluded sites (Milan and Birmingham) were used as external validation (Supplementary Information). Following magnetic resonance imaging (MRI) processing and quality control (i.e., 45 surfaces visually inspected and 21 excluded, see Supplementary Information), 308 controls and 713 patients constituted the final sample (Supplementary Figure 1; Table 1). All participants provided written informed consent and the study protocol was approved by each ethical committee.

Neuropsychological and Clinical Assessments

Patients underwent a comprehensive neuropsychological battery (Koutsouleris et al. 2018, Supplementary Information and Supplementary Table 8). To reduce multiple testing and build interpretable neurocognitive factors, we derived six cognitive domain scores and a global score of cognition following a similar approach to MATRICS (Nuechterlein et al. 2008). Data were checked for operator errors and outliers (i.e., 3 SD away from mean), and verbal learning scores were harmonized between the sites (Supplementary Information). Social cognition, working memory, speed of processing, verbal learning, reasoning, and attention summary scores were calculated. After standardization, a score of global cognition was computed by calculating the aggregate average across the six cognitive scores. In patients, social and role functioning were measured using the Global Functioning: Social Scale (GF:S) and Global Functioning: Role Scale (GF:R)(Cornblatt et al. 2007). Symptoms were evaluated using the Beck Depression Inventory (BDI)(Beck and Steer 1984), the Structured Interview for Psychosis-Risk Syndromes (SIPS)(Miller et al. 2003), from which the sum score was calculated separately for the positive (P), negative (N), disorganized (D), and general (G) symptoms items, and the "cognitive disturbances" items (COGDIS) from the Schizophrenia Proneness Instrument-Adult version (SPI-A(Schultze-Lutter et al. 2007); Supplementary Information).

MRI Data Acquisition and Processing

Participants were scanned using 3 T MRI scanners except for one site (Milan), which used a 1.5 T machine (Koutsouleris et al. 2018, Supplementary Table 2). Structural scans were visually inspected for motion artifacts and neuroanatomical abnormalities. The FreeSurfer software package (v6.0.0, https://surfer.nmr.mgh.harvard.edu/) was used to reconstruct the cortical surfaces from the T1-weighted structural MRI scans (Dale A.M., Fischl B. 1999; Fischl 2012, Supplementary Information). Quality control consisted in targeted inspection of the Euler number (Rosen et al. 2018) distribution in the sample followed by exclusion of scans (Supplementary Figure 2, Supplementary Information). Gyrification was calculated vertex-wise on each 3D cortical mesh using the local gyrification index (LGI, Schaer et al. 2012). This measure represents a ratio of the buried surface compared to a flat surface and its values range from 1, i.e., low gyrification, to 5, i.e., high gyrification. The LGI meshes were resampled to the fsaverage6 template (40.962 vertices per hemisphere) in order to reduce dimensionality, and then smoothed with a 5 mm Gaussian filter kernel (as usual for LGI, e.g., Sasabayashi et al. 2017).

Site Effects

The ComBat harmonization technique was used to mitigate site effects (Johnson et al. 2007). ComBat is an empirical Bayesian framework that removes site-effects variance while retaining age and sex effects to specifically investigate them in further analyses. To also maintain disease effects in the patient sample, ComBat was applied to the resampled and smoothed LGI-meshes in the controls before the estimates derived from the correction were applied, without modification, to the patients (Supplementary Information).

Non-Negative Matrix Factorization

We applied the orthonormal projective variant of NNMF (opNNMF, Sotiras et al. 2015), following Sotiras and colleagues (Sotiras et al. 2017), to the ComBat-corrected LGI maps of controls

Table 1 Sociodemographic and clinical information of the analyzed sample

						PAT study-groups		HC vs. PAT	
	HC	Pooled PAT	CHR	ROD	ROP	F/X [2]	p	t/X [2]/z	p
N (%)	308	713	224 (31.4)	226 (31.7)	263 (36.9)	21.331	**0.006**		
Age, mean (SD)	25.7 (6.1)	25.2 (5.9)	24.2 (6.0)	25.4 (5.6)	25.8 (5.9)	4.623	0.010	1.231	0.219
Sex, females (%)	183 (59.4)	326 (45.7)	113 (50.4)	110 (48.7)	103 (39.2)	7.366	0.025	16.132	<**0.001**
Handedness, mean (SD)[1]	76.2 (44.5)	68.8 (54.0)	66 (49.7)	73 (57.5)	67.8 (54.0)	0.848	0.429	2.011	0.045
Ethnicity, N (%)									
Caucasian	283 (92.8)	584 (86.4)	186 (83.0)	195 (86.3)	203 (77.2)	19.814	**0.003**	9.566	0.023
Asian	4 (1.3)	22 (3.3)	3 (1.3)	2 (0.9)	17 (6.5)				
African	1 (0.3)	11 (1.6)	3 (1.3)	2 (0.9)	6 (2.3)				
other	17 (5.6)	59 (8.7)	18 (8.0)	15 (6.6)	26 (9.9)				
Urbanicity, N (%)									
> 500 000	108 (35.4)	322 (47.9)	74 (33.0)	123 (54.4)	125 (47.5)	33.670	<**0.001**	13.901	**0.003**
100 000–500 000	83 (27.2)	158 (23.5)	57 (25.4)	31 (13.7)	70 (26.6)				
10 000–100 000	73 (23.9)	126 (18.8)	55 (24.6)	39 (17.3)	32 (12.2)				
<10 000	41 (13.4)	66 (9.8)	23 (10.3)	21 (9.3)	22 (8.4)				
Education, years (SD)	15.5 (3.3)	14.0 (5.4)	13.5 (3.2)	14.4 (3.0)	14.2 (7.8)	1.704	0.183	4.343	<**0.001**
Employment [2], N (%)									
Managers	1 (0.2)	9 (2.2)	3 (1.3)	1 (0.4)	5 (1.9)	19.501	0.244	39.910	<**0.001**
Professionals	12 (2.0)	67 (16.1)	19 (8.5)	25 (11.1)	23 (8.7)				
Technicians	133 (21.8)	78 (18.7)	14 (6.3)	35 (15.5)	29 (11.0)				
Clerical support workers	111 (18.2)	21 (5.03)	7 (3.1)	6 (2.7)	8 (3.0)				
Service and sales workers	36 (5.9)	91 (21.8)	30 (13.4)	28 (12.4)	33 (12.5)				
Agricoltural, forestry, fishery	121 (19.8)	3 (0.7)	0 (0)	2 (0.9)	1 (0.4)				
Craft, trades workers	5 (0.8)	47 (11.3)	13 (5.8)	14 (6.2)	20 (7.6)				
Plant/machine operators	52 (8.5)	10 (2.4)	4 (1.8)	3 (1.3)	3 (1.4)				
Elementary occupations	13 (2.1)	91 (21.8)	29 (12.9)	19 (8.4)	43 (16.3)				
Never employed, N (%)	126 (20.7)	237 (33.2)	143 (63.8)	149 (65.9)	184 (70.0)	2.146	0.342	0.587	0.444
Relationship status, N(%)									
single	136.0 (44.6)	474 (70.2)	147 (65.6)	144 (63.7)	183 (69.6)	4.909	0.555	73.620	<**0.001**
married	17.0 (5.6)	40 (5.9)	9 (4.0)	15 (6.6)	16 (6.1)				
partnership	149.0 (48.9)	148 (21.9)	51 (22.8)	51 (22.6)	46 (17.5)				
separated/divorced	3.0 (1.0)	13 (1.9)	3 (1.3)	4 (1.8)	6 (2.3)				
Functioning, mean (SD)									
GF:R current	8.5 (0.7)	5.5 (1.7)	5.6 (1.7)	5.9 (1.7)	4.9 (1.7)	45.390	<**0.001**	23.467	<**0.001**
GF:R L past year	8.2 (0.8)	5.1 (1.7)	5.4 (1.6)	5.5 (1.6)	4.5 (1.7)	58.595	<**0.001**	23.572	<**0.001**
GF:R H past year	8.6 (0.7)	7.0 (1.4)	7.0 (1.1)	7.4 (1.3)	6.7 (1.6)	25.429	<**0.001**	17.809	<**0.001**
GF:R H lifetime	8.6 (0.7)	7.9 (0.9)	7.9 (0.8)	8.1 (0.9)	7.8 (1.0)	18.798	<**0.001**	12.044	<**0.001**
GF:S current	8.5 (0.7)	6.0 (1.4)	6.2 (1.3)	6.2 (1.3)	5.6 (1.5)	31.338	<**0.001**	23.073	<**0.001**
GF:S L past year	8.2 (0.8)	5.5 (1.4)	5.8 (1.3)	5.8 (1.3)	5.0 (1.6)	43.528	<**0.001**	23.426	<**0.001**
GF:S H past year	8.6 (0.7)	7.0 (1.3)	7.1 (1.2)	7.2 (1.2)	6.7 (1.3)	16.401	<**0.001**	19.371	<**0.001**
GF:S H lifetime	8.7 (0.7)	7.8 (0.9)	7.8 (0.9)	7.9 (0.9)	7.8 (0.9)	4.829	0.089	14.282	<**0.001**
Symptoms, mean (SD)									
SIPS-P	NA	8.9 (7.2)	8.1 (2.0)	1.8 (4.3)	15.7 (5.5)	612.727	<**0.001**	NA	NA
SIPS-N	NA	10.4 (6.6)	10.1 (5.6)	9.9 (6.6)	11.1 (7.4)	2.226	0.109	NA	NA
SIPS-D	NA	3.7 (3.4)	3.4 (2.1)	2.2 (2.8)	5.2 (4.2)	50.861	<**0.001**	NA	NA
SIPS-G	NA	7.6 (4.1)	7.8 (3.7)	8.0 (4.0)	7.1 (4.5)	2.810	0.061	NA	NA
COGDIS sum	NA	7.6 (8.0)	9.5 (4.8)	3.0 (6.9)	9.9 (9.3)	60.923	<**0.001**	NA	NA
BDI	NA	23.6 (12.1)	25.4 (12.1)	25.7 (11.2)	20.2 (12.1)	14.467	<**0.001**	NA	NA
Cognition, mean (SD)									
Social cognition[3]	19.4 (2.1)	18.7 (2.5)	18.9 (2.2)	19.3 (2.1)	18.1 (2.9)	12.955	<**0.001**	4.144	<**0.001**
Working memory[4]	17.9 (3.8)	16.0 (4.0)	16.6 (4.2)	16.4 (4.0)	15.0 (3.6)	12.280	<**0.001**	7.093	<**0.001**
Verbal fluency[5]	25.8 (5.7)	22.1 (6.6)	22.8 (6.7)	23.4 (6.3)	20.2 (6.3)	16.129	<**0.001**	8.381	<**0.001**
TMT-A	27.5 (9.0)	31.7 (12.1)	31.5 (10.9)	29.1 (11.4)	34.1 (13.2)	9.673	<**0.001**	−5.278	<**0.001**
DSST	65.7 (10.4)	57.0 (13.5)	59.7 (11.9)	61.5 (12.6)	50.6 (13.2)	49.093	<**0.001**	9.932	<**0.001**
Verbal learning[6]	61.1 (6.9)	56.8 (9.5)	58.2 (8.5)	58.6 (8.4)	54.0 (10.5)	17.602	<**0.001**	7.076	<**0.001**
Reasoning[7]	21.2 (3.2)	19.9 (4.1)	20.4 (3.6)	20.5 (3.7)	18.8 (4.5)	12.112	<**0.001**	4.721	<**0.001**
CPT-IP (N correct)	274.4 (13.0)	265.8 (17.3)	268.6 (18.3)	269.7 (14.0)	259.9 (17.3)	23.632	<**0.001**	7.716	<**0.001**
CPT-IP (N errors)	10.3 (6.0)	11.4 (7.0)	12.1 (6.5)	10.8 (7.6)	11.4 (6.8)	1.676	0.188	−2.375	0.018
Medication, N (%)									
Antipsychotics (AP)	NA	104 (14.6)	14 (6.3)	5 (2.2)	85 (32.3)	299.510	<**0.001**	NA	NA
Antidepressants (AD)	NA	146 (20.5)	54 (24.1)	83 (36.7)	9 (3.4)			NA	NA
AP + AD	NA	38 (5.3)	11 (4.9)	16 (7.1)	11 (4.2)			NA	NA
Anxiolytics/Sedatives (Anx)	NA	13 (1.8)	6 (2.7)	6 (2.7)	1 (0.4)			NA	NA
AP + Anx	NA	65 (9.1)	6 (2.7)	0 (0)	59 (22.4)			NA	NA
AD+Anx	NA	49 (6.9)	10 (4.5)	34 (15.0)	5 (1.9)			NA	NA
AP + AD+Anx	NA	28 (3.9)	10 (4.5)	4 (1.8)	14 (5.3)			NA	NA

Bold values represent significance after correction for multiple comparisons (False Discovery Rate, FDR < 0.01). Abbreviations: GF: Global Functioning Role (R) or Social (S), [0:10], higher scores indicate better functioning; L: lowest; H: highest; y: year; SIPS-P and N: Structured Interview for Prodromal Syndromes-Positive or Negative scale. Displayed are composite scores for all four subscales (i.e., sum of five Positive, six Negative, four Disorganized symptoms, and four items for General psychopathology) [0:6], higher scores indicate higher symptom severity; COGDIS: Schizophrenia Proneness Instrument, Cognitive Disturbances (sum of nine items); BDI: Beck's Depression Inventory; TMT-A: Trail-Making Test-A; DSST: Digit-Symbol Substitution Test; CPT-IP: Continuous Performance Test-Identical Pairs; HC: Healthy Control; PAT: patients; CHR: Clinical High Risk; ROD: Recent Onset Depression; ROP: Recent Onset Psychosis. [1]: Edinburgh Handedness score, [−100;100], higher scores indicate more pronounced right-handedness; [2]: ISCO, International Standard Classification of occupation (www.ilo.org/public/english/bureau/stat/isco/isco08) [1–9]; [3]: measured with Diagnostic Analysis of Nonverbal Accuracy (DANVA); [4]: measured with forward + backward Digit Span; [5]: correct words in Verbal Fluency (phonemic); [6]: sum of 5-Rey Auditory Verbal Learning Test repetitions; [7]: Matrix Reasoning subtest from the Wechsler Adult Intelligence Scale. Notes: cognition scores are those used to build the seven cognitive domains (refer to text).

to detect Patterns of Structural Covariance (PSCs, see Supplementary Information). opNNMF is an unsupervised multivariate method that deconstructs a data matrix based on a nonnegative combination of its parts, or components. This method is particularly useful in the neuroimaging context for two main reasons: first, unlike PCA, it can aggregate variance in a parcellation-like way, in line with the hierarchical and modular organization of the human brain cortex. Second, it can generalize well to unseen data (Sotiras et al. 2015). To establish the optimal number of PSCs, we quantified the PSC goodness of fit using the reconstruction error (i.e., the amount of reconstructed variance in the data not explained by the components; Sotiras et al. 2017) and PSC generalizability using split-half analyses (i.e., applying NNMF independently on two randomly generated subsamples and calculating the components' degree of overlap) (Supplementary Information, Supplementary Figure 6). The NNMF component solution was also externally validated in the two held-out sites by calculating the inner product between all components to identify maximal overlap and by visual inspection (Supplementary Information). The PSC maps were then applied to patients and the mean gyrification values for each component were extracted from both patients and controls (Supplementary Information, Supplementary Figure 1).

Investigation of Group and Clinical and Neuropsychological Associations

The associations between age, sex, and site and PSC gyrification values for controls and patients were first estimated using correlations, t-tests, and analyses of variance. Group effects (controls vs. patients, independent variable) were then investigated by fitting linear models for each PSC (dependent variable) with age and sex as covariates. We further investigated quadratic age effects and age-by-group interaction effects (Supplementary Information) and explored the contribution of each PRONIA study-group (i.e., ROP, ROD, CHR) (Supplementary Information). Supplemental analyses explored the potential confounding effect of scan quality (Supplementary Information).

PSCs showing a significant group effect (False Discovery Rate, FDR-correction $P < 0.01$, following Gregory et al. 2016) were then investigated using linear models to determine relationships with symptoms, cognition, and functioning in the patient sample. Supplementary analyses were conducted to test the potential confounding effects of MRI scan quality, education, and premorbid intelligence quotient (IQ) in order to investigate cognitive and functioning specificity (Supplementary Information). The same analyses were also performed in the control sample to investigate potential disease-independent associations between PSCs, cognition, and functioning. All analyses ran in Matlab (version R2020a) and results were FDR corrected ($P < 0.01$).

Data Availability

The code and models that support the findings of the current study are available on request from the corresponding author. The datasets generated and analyzed are not publicly available due to data restriction policies defined in the participants' signed informed consent.

Results

Controls and patients differed in their sex and urbanicity distribution, educational and employment level, relationship status, and all functioning and cognitive domains, except for error numbers in the continuous performance test (CPT; P values$^{FDR} < 0.003$, Table 1). No difference was found for age, handedness, and ethnicity (P values$^{FDR} > 0.02$). The three clinical study groups differed from each other in terms of ethnicity, urbanicity, all functioning subscales (except GF:S highest lifetime), positive and depressive symptoms, COGDIS, cognitive abilities (except in CPT error numbers) and medication intake (P values$^{FDR} < 0.006$). Between the three study-groups, age, sex, handedness, education, employment and relationship status, and negative and general symptoms were similar (P values$^{FDR} > 0.01$). Site distribution of the pooled sample and sociodemographic information are reported in Supplementary Table 1.

Patterns of Structural Covariance

A total of 18 components best fit the gyrification data (Fig. 1, Supplementary Information). PSCs were spatially distinct and differed from structural and functional brain parcellations (Supplementary Information, Supplementary Figures 4 and 5). Figure 2 shows the cortical ribbon coverage when all PSCs are projected on the brain (see Supplementary Figure 8 for an individual representation of components). In the two completely held-out sites, 72% of components exhibited an inner product above 0.5 (Supplementary Figure 10) and were visually highly similar (Supplementary Figure 9). The mean inner product across PSC solutions was no different to the discovery sample split-half analysis result used to define the component solution (i.e., 0.622 and 0.623, respectively, $P = 0.493$; Supplementary Information). Post hoc analyses revealed significant associations of gyrification in controls across all PSCs with age (r ranging [−0.18;−0.55], $p^{FDR} < 0.01$) and sex (t range [1.94;7.11], $p^{FDR} < 0.01$) with no site differences (all $p^{FDR} > 0.01$, Supplementary Table 5). In patients, similar demographic relationships were found (age: r range [−0.143;−0.475], $p^{FDR} < 0.001$; sex: t range [2.10;10.70], $p^{FDR} < 0.01$); three PSCs showed site effects (left PSC 6 and bilateral 17; Supplementary Table 6). Quadratic effects of age on gyrification were detected for five PSCs (i.e., PSC 3 and 9, and right PSC 18, Supplementary Information), however, the additional explained variance when adding the age quadratic term in linear models was minimal (Supplementary Information). No interactions between age and group were found in the pooled patient sample or within study-groups. No associations between PSCs in the patient sample and medication were found.

Effect of Group

A total of 14 gyrification components showed significant differences between controls and patients (models adjusted R^2 [0.123;0.289]; maximum partial correlation coefficient for the specific effect of group −0.157, models' p^{FDR} [2.07^{E-29};1.57^{E-75}]; coefficients' t-values from −2.98 to −5.08; coefficients' p^{FDR} [0.003;4.60^{E-07}], Supplementary Table 9). This pattern included, bilaterally, a temporal-insular area (PSC1 and 2, 5 left and 10 right and 12), a lateral occipital area (PSC 6 left and 5 right), and a lateral fronto-parietal area (PSC 7 left and 3 right). Within the right hemisphere, fronto-parietal (PSC 9, 13 and 18) and cingular gyrus areas (PSC 16) also differed between controls and patients (Fig. 3). These results were not influenced by MRI scan quality (Supplementary Information, Supplementary Table 10). The pattern of reduced gyrification in patients could almost completely be replicated in the external validation sample

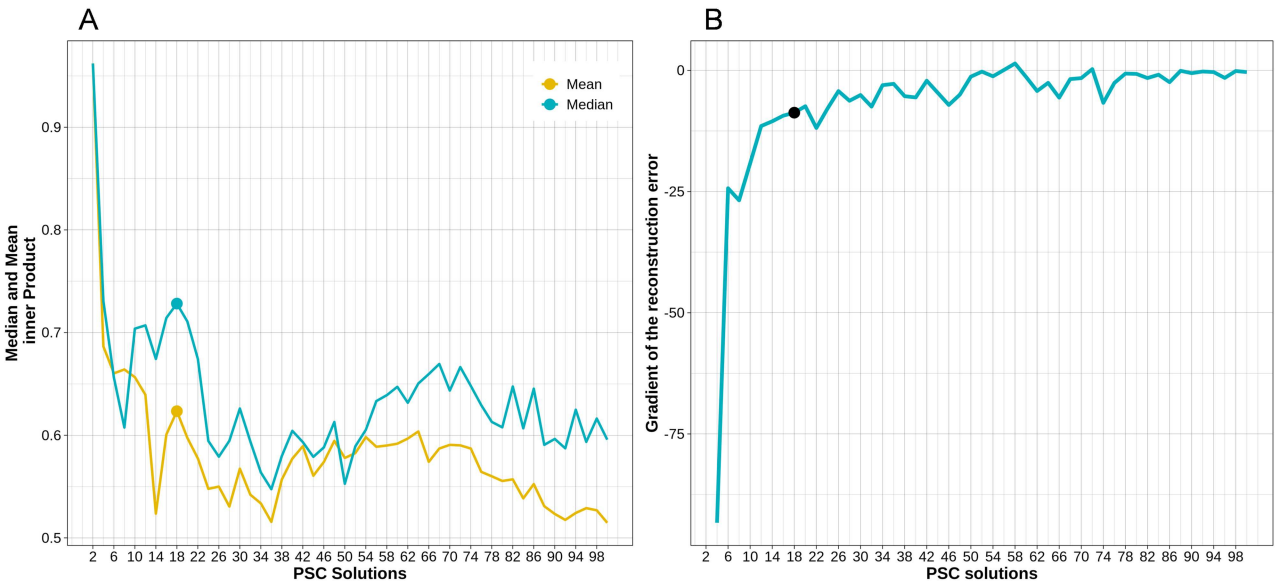


Figure 1. Split-half reproducibility (A) and reconstruction error (B) analyses. A: The median and mean inner product of the two independently estimated sets of components in the split-halves (y-axis) is represented as a function of the number of components estimated (x-axis). B: The gradient of the reconstruction error is reported as a function of the number of components estimated. Abbreviations: PSC: Pattern of Structural Covariance.

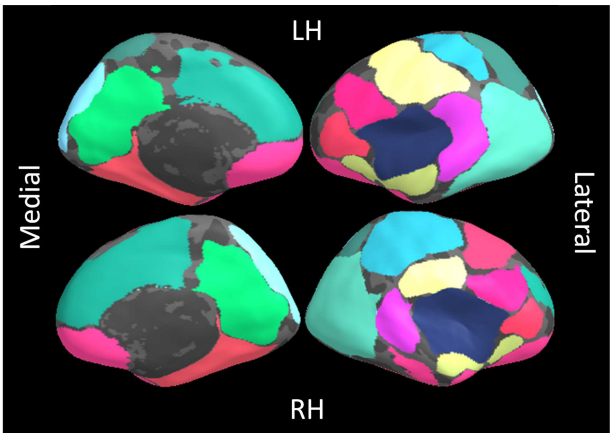


Figure 2. 18 gyrification-based Patterns of Structural Covariance (PSC) overlapped on an inflated brain surface. RH: right hemisphere, LH: left hemisphere. The medial part of the brain is displayed on the left, the lateral on the right. Same colors correspond to the same PSC mapping bilaterally.

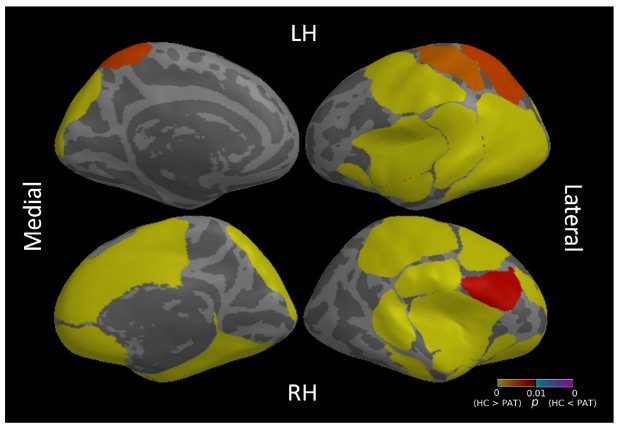


Figure 3. Effects of group (Healthy Controls, HC vs patients, PAT) resulting from the comparison of 29 gyrification Patterns of Structural Covariance (PSCs) between HC and PAT. The color table represents *P* values for each ROI. Only *P* values lower than 0.01 (uncorrected) are shown. ROIs are overlaid on an inflated common surface. RH: right hemisphere; LH: left hemisphere. The medial part of the brain is displayed on the left, the lateral on the right.

(Supplementary Information, Supplementary Table 11). Post hoc analyses of significant components demonstrated significance across all groups at a threshold of $p^{FDR} < 0.05$, though marginal increases in effect size coupled with moderate significance increases were apparent for ROP and CHR for some PSCs (Supplementary Information, Supplementary Table 21). There were no differences between the ROD, ROP, and CHR groups in pair-wise comparisons (Supplementary Information).

Associations with Cognitive and Clinical Measures

In patients, significant associations between 12/14 PSCs were found for cognition and functioning domains (Fig. 4, Supplementary Information, Supplementary Tables 12, 16). For cognition, four temporal components (i.e., superior temporal gyrus, pars orbitalis and triangularis, and insula: PSC 1, 2 and 12) explained up to 3.8% of variance for working memory, speed of processing, reasoning, and global cognition (r range [0.116;0.196], $p^{FDR} < 0.002$). Additionally, working memory was associated with the right angular gyrus and lateral occipital lobe (PSC 10 and 5, respectively $r = 0.151$, $p^{FDR} < 0.001$ and $r = 0.125$, $p^{FDR} = 0.001$), and global cognition with the left angular and medial temporal gyrus (PSC 5, $r = 0.122$, $p^{FDR} = 0.003$). Results were not explained by either quality of MRI scans (Supplementary Information, Supplementary Table 13), or the patients' educational level (Supplementary Information, Supplementary Table 14). Cognitive associations were found at a threshold of $p^{FDR} < 0.05$ when controlling for premorbid IQ (r range [0.109;0.135], $p^{FDR} < 0.008$,

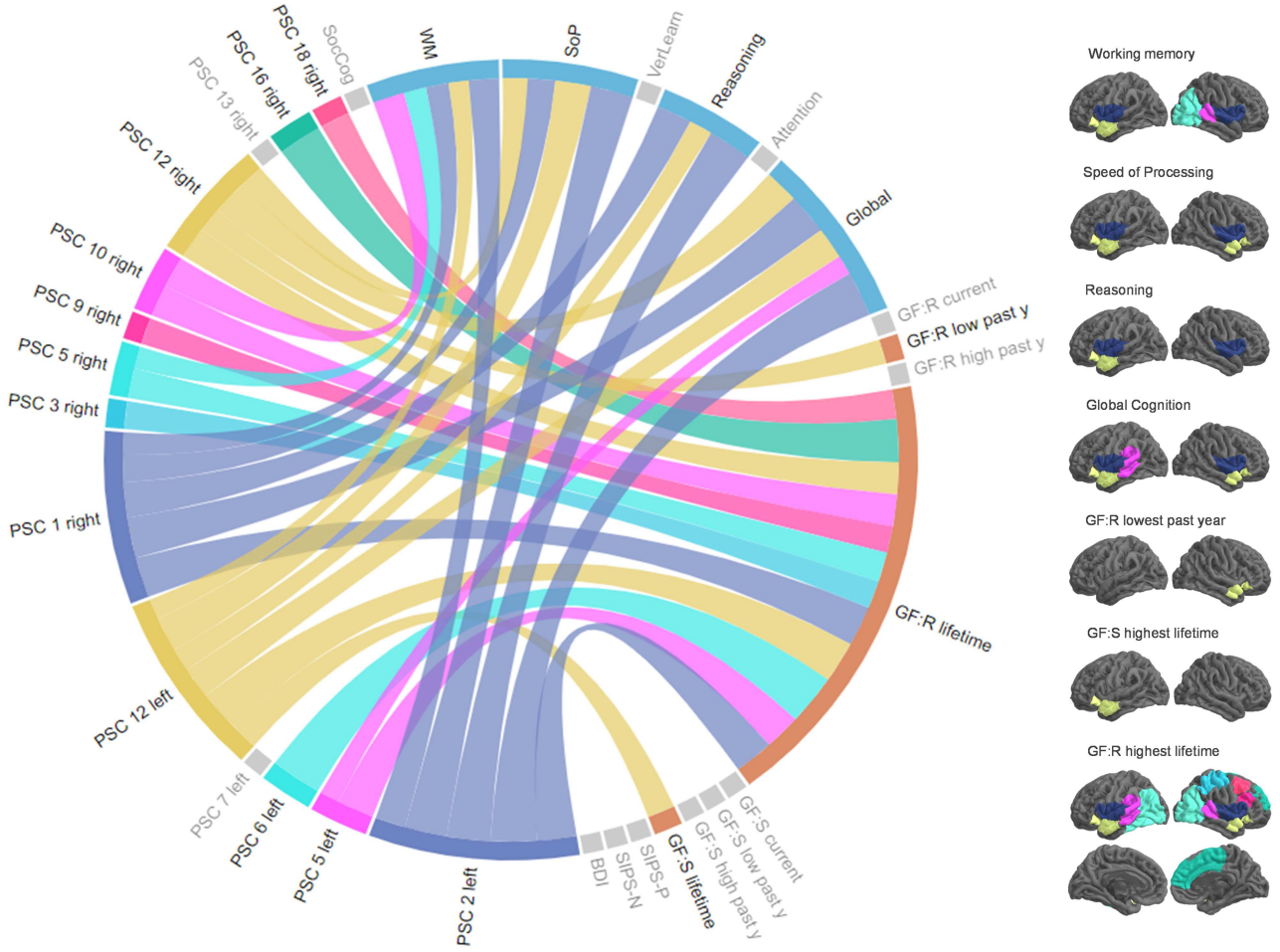


Figure 4. Associations between gyrification components and functioning and neurocognitive domains. The circular plot (left side) represents associations between components and neurocognitive (in blue) and functioning (in orange) domains, with thicker lines reflecting higher effect sizes/lower *P* values (FDR correction $P < 0.01$). Gray domains are those not significant in the analyses. The right panel displays the components associated with functioning and neurocognitive domains overlaid on a common cortical surface. Colors on the brains correspond to those in the plot. Abbreviations: PSC: Pattern of Structural Covariance; SocCog: social cognition; WM: working memory; SoP: speed of processing; VerLearn: verbal learning; Global: global cognition; GF:R/S: global functioning: Role/Social; lifetime: the highest functioning lifetime; low: lowest; high: highest; y: year; SIPS-P/N: Structured Interview for the Prodromal Syndrome, Positive/Negative symptoms; BDI: Beck Depression Inventory.

Supplementary Information, Supplementary Table 15, Supplementary Figure 11).

In the functioning domain, the left and right PSC 12 explained approximately 1.7% of variance of the patients' highest social functioning lifetime and the lowest role functioning in the past year ($p^{FDR} < 0.001$), respectively. Remaining significant associations were found specifically in the highest role functioning lifetime for 12 PSCs (r range [0.137;0.195], all $p^{FDR} < 0.001$, Fig. 4, Supplementary Table 16). When controlling for premorbid IQ, three PSCs survived at a threshold of $p^{FDR} < 0.01$, while results were overlapping for $p^{FDR} < 0.05$ (Supplementary Information, Supplementary Table 18, Supplementary Figure 11). These relationships were not completely explained by educational level (Supplementary Information, Supplementary Table 17) or MRI scan quality (Supplementary Information, Supplementary Table 19). Associations with the cognitive and functioning domains could not be replicated in the validation cohort—potentially because of the smaller sample size. We found no significant associations with symptoms (Supplementary Table 20). In controls, no associations survived multiple comparisons correction, neither in the cognitive, nor in the functioning domain (all $p^{FDR} > 0.01$, Supplementary Information, Supplementary Table 22).

Discussion

Novel and distinct gyrification covariance components were found, with a specific pattern of abnormality in early mental illness that was evident across all diagnostic groups and was not associated with symptoms. As hypothesized, the components showed rather small but significant associations with cognitive domains and functioning, which were not completely explained by premorbid intelligence, suggesting illness-related effects. While both age and sex effects were observed, we did not detect a specific interaction with diagnosis but did find increases of effect size when compared to controls in the psychosis and psychosis-prone study groups. These results point to transdiagnostic neurodevelopmental effects on gyrification occurring before late adolescence, which may precede the onset of illness.

Gyrification components were stable, replicable to some extent in new sites, demonstrated hemispheric symmetry, and

were substantially novel. Quantitative comparisons indicated differences to traditional atlases (Supplementary Information, Supplementary Figure 4) and newer functional parcellations (Thomas Yeo et al. 2011) (Supplementary Figure 5), while qualitative comparisons indicated differences to thickness-based PSCs (Sotiras et al. 2017) and genetically derived structural brain networks (Chen et al. 2013). When compared to commonly used atlases in previous gyrification work (e.g., Desikan-Killiany), PSCs critically crossed gyral boundaries highlighting the importance of data-driven methods for gyrification research as opposed to traditional atlases. Hemispheric symmetry indicated shared structural and functional relationships, while some important exceptions highlighted known lateralized areas, e.g., Broca's area (Knecht 2000; Toga and Thompson 2003). Expected linear negative relationships with age were found for the components based on hypotheses related to cortical restructuring processes during adolescence and in adult life (Klein et al. 2014; Cao et al. 2017), which occur against a background of early neurodevelopmental gyrogenesis (Llinares-Benadero and Borrell 2019). Given our restricted age window, the results reinforce research highlighting the importance of ongoing gyrification during adolescence and adulthood.

Decreased gyrification was found in patients, covering temporal, parieto-frontal, occipital, and insular areas bilaterally, and parieto-frontal and cingulate areas on the right hemisphere. The results are in line with previous research using traditional vertex-wise or atlas-based techniques, where hypogyrification has been detected in the temporal, insular, and frontal areas in schizophrenia (Palaniyappan and Liddle 2012), depression (Depping et al. 2018), and first-episode psychosis (Palaniyappan et al. 2013). Similar folding abnormalities were also found in a recent study investigating individuals at high-risk for psychosis (Sasabayashi et al. 2017), implicating bilateral supramarginal gyri, rostral middle frontal gyri, lateral occipital gyri, in addition to specific associations with the right pre- and postcentral gyri, superior frontal gyri, paracentral lobules, and cingulate. Our findings also add to this research field by providing a clearly defined gyrification covariance map that could be used in further studies, enabling a finer investigation of cortical geometry in both healthy and clinical populations. Furthermore, results also highlight a transdiagnostic signature of abnormal cortical folding, with slightly more evident disruptions in early psychosis and its risk, potentially suggesting a more distinctive neurodevelopmental component in this diagnostic group than in mood disorders (Craddock and Owen 2010). Altered gyrification in patients was not specific to current symptomatology, which has been reported in depression (Depping et al. 2018) and in cases of rare genetic disorders demonstrating disrupted gyrification with no symptoms (Caverzasi et al. 2019).

Of the several mechanisms, which are thought to be responsible for the gyrogenesis (e.g., axonal tension, differential growth, synaptogenesis; Llinares-Benadero and Borrell 2019), the involvement of radial glial cells during neurodevelopment (Rash et al. 2019) might play a particularly relevant role in psychopathology (Kato et al. 2017; Dietz et al. 2020). Disruptions of progenitor glial cells' differentiation might in fact impact both patterns of cortical convolution and neurotransmission, which, in turn, might reflect pathophysiological manifestations of abnormal development. This cascade might start from gene expression levels involved in both pre- and post-natal cortex development, which have shown to be associated with cortical thickness profiles across disorders (Patel et al. 2021). Recent studies also point to a shared genetic liability across psychiatric disorders (Musliner et al. 2019), which supports our transdiagnostic findings of disrupted gyrification. Our findings thus reinforce the need to further investigate the neurobiological underpinnings of cortical folding networks throughout the lifespan within and across psychiatric disorders.

A bilateral cortical area covering the insula, pars orbitalis, triangularis, and opercularis (i.e., left PSC 2 and right 1) was also associated with multiple cognitive domains in patients—working memory, speed of processing, reasoning, and global cognition—with a mediation effect of premorbid IQ. These results are in line with limited previous research in established illness showing increased cortical curvature in schizophrenia related to lower premorbid and current intelligence (Jessen et al. 2019). Rather weak multivariate associations were also found between gyrification in frontotemporal areas (left pars orbitalis and triangularis, right pars opercularis and orbitalis, and insula) and general cognitive ability in bipolar, schizoaffective, and schizophrenic patients (Rodrigue et al. 2018). Interestingly, similar gyrification–cognition relationships detected in this study have been shown in healthy midlife individuals, where positive associations are found between gyrification and executive functions in lateral frontal cortex (Gautam et al. 2015), global cognition and superior temporal gyrus, as well as insular cortex and postcentral gyrus (Lamballais et al. 2020). Moreover, positive but weak associations with general intelligence have been demonstrated in children and adolescents (Gregory et al. 2016; Chung et al. 2017; Mathias et al. 2020). Of note, we found no cognitive associations in healthy young adults. On the one hand, it is possible that interindividual differences in gyrification–cognition relationships exist also for this different age range, but are too weak to be detectable, whereas they may be observable in psychiatric disorders due to being more pronounced as a result of illness and neurodevelopmental insults. On the other hand, associations found could point to unstable relationships between cortical convolution and cognitive abilities overall (Mathias et al. 2020). These might for instance be detectable only for subtypes of the population (Dwyer et al. 2020), or, although less plausible, only for specific cognitive domains. Further research is warranted to investigate whether and to what extent gyrification may mirror cognition in the healthy brain by comparing a broad spectrum of cognitive constructs in multicenter cohorts covering the full lifespan.

The consistent involvement of Broca's area (and its contralateral equivalent) across cognitive domains might be due to its recognized double functional nature integrating both language-specific and multiple-demand networks participating in attention, working memory, planning and fluid intelligence (Fedorenko and Blank 2020). Bilateral insula associations may indicate both its involvement in feeling states and saliency, and its indirect connection with cognitive processes controlled prevalently in subregions of the prefrontal cortex (Namkung et al. 2017). The insular role both in emotional states and in motivational/cognitive processes might be one potential explanation for its disruptions found across a number of different psychiatric disorders (Goodkind et al. 2015; Namkung et al. 2017). Interestingly, the lack of insular relationships with symptoms in our study might point to gyrification abnormalities, which appear to be specific to the cognitive domain.

For the first time, we also detected a distributed pattern of gyrification-based PSCs (including fronto-temporal and lateral occipital areas bilaterally and pre-frontal, parietal and cingular areas on the right hemisphere) that was specifically correlated with the highest lifetime role functioning in the patient sample. Importantly, our results were only partly explained

by premorbid intelligence, thus suggesting a specific cortical folding signature linked to the individual functioning and not driven by general cognition. When combined with the cognitive findings, these results may highlight the effects of abnormal gyrification on premorbid cognition and functioning over the lifetime that confer transdiagnostic risk to psychiatric illness during critical periods in adolescence and young adulthood. Such vulnerability is likely to be mediated by disruptions in coordinated, or synchronized, developmental processes, caused by pleiotropic genetic mechanisms, environmental factors, or insults before and during gyrogenesis (Alexander-Bloch et al. 2013). This hypothesis is supported by the lack of disrupted developmental pathways in our patient sample and suggests that future studies need to focus on younger at-risk samples (e.g., genetic risk samples) to identify candidate gyrification mechanisms that may be amenable to developmental interventions. Further research exploring which sulci are more evolutionary driven and therefore less prone to plastic reorganization (Schmitt et al. 2008; Rollins et al. 2020) would also be indicated in addition to multimodal covariance studies to assess the developmental interactions of different morphological measures (e.g., volume, density, and thickness).

Our study has some limitations. First, as in most factor analyses, the number of components chosen to best represent the variance decomposition, although normally supported by stability metrics, was ultimately selected by the investigators. Many other solutions might as well be explanatory of gyrification patterns. However, in our sample, a smaller number of PSCs showed a lower grade of resolution (Supplementary Information), with similar PSC being fused in spreader components. Second, the variance explained in our clinical and neuropsychological models was low, in line with previous findings (Mathias et al. 2020) and, in general, research in psychology (Schäfer and Schwarz 2019). Third, our replication might have been hindered either by residual site effects (potentially caused by coil differences), or the sample size of the controls. Although the PSCs were promisingly partly generalizable to an independent subsample, further replication of our gyrification components in larger samples is required to validate their future use. Future investigation of the role of other psychopathologies in the transdiagnostic signature observed in our study is also highly warranted. Finally, even though our study lacks a deep biological investigation of cortical folding from a mechanistic perspective, it nevertheless represents an important step forward in the field that could lead to mechanistic studies (e.g., in combination with diffusion tensor imaging).

Our work adds knowledge to previous gyrification research by establishing a novel covariance map, whose components revealed cortical folding abnormalities in early psychosis, risk-states and, to a lesser extent, depression, irrespective of the psychopathological features. Relationships between cortical folding and cognition suggested a neurodevelopmental origin that was supported by a further association with role functioning. When combined, our results highlight the importance of studying gyrification before late adolescence to delineate genetic, developmental, and environmental mechanisms that potentially influence this morphological measure and put adolescents and young adults at risk of mental illness.

Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

Funding

This work was supported by PRONIA: a Collaboration Project funded by the European Union under the 7th Framework Program under grant agreement number 602152. BMBF and Max Planck Society (grant agreement number M526300) funded R.S. Structural European Funding of the Italian Minister of Education (Attraction and International Mobility—AIM—action, grant agreement No 1859959) funded L.A.A. NIH/NIA supported A.S. (grant R01AG06710). National Health and Medical Research Council Senior Principal Research Fellowship (grants 628386 and 1105825) and European Union–National Health and Medical Research Council (grant 1075379) supported C.P., S.R., A.R-R., and N.K. reported receiving grants from the European Union (EU) during the conduct of the study.

Notes

The PRONIA consortium: Principal investigator and primary contact is Prof. Nikolaos Koutsouleris (nikolaos.koutsouleris@med.uni-muenchen.de). PRONIA consortium members listed here performed the screening, recruitment, rating, examination, and follow-up of the study participants and were involved in implementing the examination protocols of the study, setting up its information technological infrastructure, and organizing the flow and quality control of the data analyzed in this article between the local study sites and the central study database.

Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany: Shalaila Haas, Alkomiet Hasan, Claudius Hoff, Ifrah Khanyaree, Aylin Melo, Susanna Muckenhuber-Sternbauer, Yanis Köhler, Ömer Öztürk, Nora Penzel, David Popovic, Adrian Rangnick, Sebastian von Saldern, Rachele Sanfelici, Moritz Spangemacher, Ana Tupac, Maria Fernanda Urquijo, Johanna Weiske, Antonia Wosgien, Camilla Krämer.

Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany: Karsten Blume, Dominika Julkowski, Nathalie Kaden, Ruth Milz, Alexandra Nikolaides, Mauro Seves, Silke Vent, Martina Wassen.

Department of Psychiatry (Psychiatric University Hospital, UPK), University of Basel, Switzerland: Christina Andreou, Laura Egloff, Fabienne Harrisberger, Ulrike Heitz, Claudia Lenz, Letizia Leanza, Amatya Mackintosh, Renata Smieskova, Erich Studerus, Anna Walter, Sonja Widmayer.

Institute for Mental Health & School of Psychology, University of Birmingham, United Kingdom: Chris Day, Sian Lowri Griffiths, Mariam Iqbal, Mirabel Pelton, Pavan Mallikarjun, Alexandra Stainton, Ashleigh Lin, Paris Lalousis.

Department of Psychiatry, University of Turku, Finland: Alexander Denissoff, Anu Ellilä, Tiina From, Markus Heinimaa, Tuula Ilonen, Päivi Jalo, Heikki Laurikainen, Antti Luutonen, Akseli Mäkela, Janina Paju, Henri Pesonen, Reetta-Liina Säilä, Anna Toivonen, Otto Turtonen.

Department of Psychiatry (Psychiatric University Hospital LVR/HHU Düsseldorf), University of Düsseldorf, Germany: Sonja Botterweck, Norman Kluthausen, Gerald Antoch, Julian Caspers, Hans-Jörg Wittsack.

Department of Psychiatry and Psychotherapy, and Department of Child Adolescence Psychiatry, Psychotherapy and Psychosomatics, University of Muenster, Germany:

Marian Surmann, Udo Dannlowski, Olga Bienek, Georg Romer.

General Electric Global Research Inc., USA.

Ana Beatriz Solana, Manuela Abraham, Timo Schirmer.
Workgroup of Paolo Brambilla, University of Milan, Italy:

- Department of Neuroscience and Mental Health, Fondazione IRCCS Ca′ Granda Ospedale Maggiore Policlinico, University of Milan, Milan, Italy: Carlo Altamura, Marika Belleri, Francesca Bottinelli, Adele Ferro, Marta Re
- Programma2000, Niguarda Hospital, Milan: Emiliano Monzani, Maurizio Sberna
- San Paolo Hospital, Milan: Armando D'Agostino, Lorenzo Del Fabro
- Villa San Benedetto Menni, Albese con Cassano (CO): Giampaolo Perna, Maria Nobile, Alessandra Alciati

Workgroup of Paolo Brambilla at the University of Udine, Italy

- Department of Medical Area, University of Udine, Udine, Italy: Matteo Balestrieri, Carolina Bonivento, Giuseppe Cabras, Franco Fabbro
- IRCCS Scientific Institute "E. Medea", Polo FVG, Udine: Marco Garzitto, Sara Piccin

Conflict of interest: R.S., R.U., and N.K. report educational fees from Lundbeck/Otsuka, outside the submitted work. R.U. reports receiving personal fees from Sunovion Pharmaceuticals, Inc, outside the submitted work. A.B. has received lecture fees from Otsuka, Janssen, Lundbeck, and consultant fees from Biogen. C.P. participated in advisory boards for Janssen-Cilag, AstraZeneca, Lundbeck, and Servier and received honoraria for talks presented at educational meetings organized by AstraZeneca, Janssen-Cilag, Eli Lilly, Pfizer, Lundbeck, and Shire, outside the submitted work. N.K. and E.M. reported having a patent to US20160192889A1 licensed. No other conflicts of interest were reported.

References

Alexander-Bloch A, Giedd JN, Bullmore E. 2013. Imaging structural co-variance between human brain regions. *Nat Rev Neurosci*. 14:322–336.

Barkovich AJ. 2010. Current concepts of polymicrogyria. *Neuroradiology*. 52:479–487.

Bassett DS, Bullmore E, Verchinski BA, Mattay VS, Weinberger DR, Meyer-Lindenberg A. 2008. Hierarchical organization of human cortical networks in health and schizophrenia. *J Neurosci*. 28:9239–9248.

Beck AT, Steer RA. 1984. Internal consistencies of the original and revised beck depression inventory.*J Clin Psychol*. 40:1365–1367.

Borrell V. 2018. How cells fold the cerebral cortex. *J Neurosci*. 38:776–783.

Cao B, Mwangi B, Passos IC, Wu MJ, Keser Z, Zunta-Soares GB, Xu D, Hasan KM, Soares JC. 2017. Lifespan Gyrification trajectories of human brain in healthy individuals and patients with major psychiatric disorders. *Sci Rep*. 7:1–8.

Caverzasi E, Battistella G, Chu SA, Rosen H, Zanto TP, Karydas A, Shwe W, Coppola G, Geschwind DH, Rademakers R, et al. 2019. Gyrification abnormalities in presymptomatic c9orf72 expansion carriers. *J Neurol Neurosurg Psychiatry*. 90:1005–1010.

Chen CH, Fiecas M, Gutiérrez ED, Panizzon MS, Eyler LT, Vuoksimaa E, Thompson WK, Fennema-Notestine C, Hagler DJ, Jernigan TL, et al. 2013. Genetic topography of brain morphology. *Proc Natl Acad Sci U S A*. 110:17089–17094.

Chung YS, Hyatt CJ, Stevens MC. 2017. Adolescent maturation of the relationship between cortical gyrification and cognitive ability. *Neuroimage*. 158:319–331.

Cornblatt BA, Auther AM, Niendam T, Smith CW, Zinberg J, Bearden CE, Cannon TD. 2007. Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophr Bull*. 33:688–702.

Craddock N, Owen MJ. 2010. The Kraepelinian dichotomy - going, going… but still not gone. *Br J Psychiatry*. 196:92–95.

Dale AM, Fischl B, SML. 1999. Cortical surface-based analysis. *Neuroimage*. 194:179–194.

Das T, Borgwardt S, Hauke DJ, Harrisberger F, Lang UE, Riecher-Rössler A, Palaniyappan L, Schmidt A. 2018. Disorganized gyrification network properties during the transition to psychosis. *JAMA Psychiatry*. 75:613–622.

Depping MS, Thomann PA, Wolf ND, Vasic N, Sosic-Vasic Z, Schmitgen MM, Sambataro F, Wolf RC. 2018. Common and distinct patterns of abnormal cortical gyrification in major depression and borderline personality disorder. *Eur Neuropsychopharmacol*. 28:1115–1125.

Dietz AG, Goldman SA, Nedergaard M. 2020. Glial cells in schizophrenia: a unified hypothesis. *The Lancet Psychiatry*. 7:272–281.

Dwyer DB, Kalman JL, Budde M, Kambeitz J, Ruef A, Antonucci LA, Kambeitz-Ilankovic L, Hasan A, Kondofersky I, Anderson-Schmidt H, et al. 2020. An investigation of psychosis subgroups with prognostic validation and exploration of genetic underpinnings: the PsyCourse study. *JAMA Psychiatry*. 77:523–533.

Evans AC. 2013. Networks of anatomical covariance. *Neuroimage*. 80:489–504.

Fedorenko E, Blank IA. 2020. Broca's area is not a natural kind. *Trends Cogn Sci*. 24:270–284.

Fischl B. 2012. FreeSurfer. *Neuroimage*. 62:774–781.

Gautam P, Anstey KJ, Wen W, Sachdev PS, Cherbuin N. 2015. Cortical gyrification and its relationships with cortical volume, cortical thickness, and cognitive performance in healthy mid-life adults. *Behav Brain Res*. 287:331–339.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, et al. 2017. Europe PMC funders group Europe PMC funders author manuscripts a multi-modal parcellation of human cerebral cortex. *Nature*. 536:171–178.

Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, Ortega BN, Zaiko YV, Roach EL, Korgaonkar MS, et al. 2015. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*. 72:305–315.

Gregory MD, Kippenhan JS, Dickinson D, Carrasco J, Mattay VS, Weinberger DR, Berman KF. 2016. Regional variations in brain gyrification are associated with general cognitive ability in humans. *Curr Biol*. 26:1301–1305.

Guo S, Iwabuchi S, Balain V, Feng J, Liddle P, Palaniyappan L. 2015. Cortical folding and the potential for prognostic neuroimaging in schizophrenia. *Br J Psychiatry*. 207: 458–459.

Jessen K, Rostrup E, Mandl RCW, Nielsen MO, Bak N, Fagerlund B, Glenthoj BY, Ebdrup BH. 2019. Cortical structures and their clinical correlates in antipsychotic-naïve schizophrenia patients before and after 6 weeks of dopamine D 2/3 receptor antagonist treatment. *Psychol Med*. 49:754–763.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 8:118–127.

Kato TA, Myint AM, Steiner J. 2017. Editorial: minding glial cells in the novel understandings of mental illness. *Front Cell Neurosci*. 11:10–12.

Kippenhan JS, Olsen RK, Mervis CB, Morris CA, Kohn P, Meyer-Lindenberg A, Berman KF. 2005. Genetic contributions to human Gyrification: sulcal morphometry in Williams syndrome. *J Neurosci*. 25:7840–7846.

Klein D, Rotarska-Jagiela A, Genc E, Sritharan S, Mohr H, Roux F, Han CE, Kaiser M, Singer W, Peter JU. 2014. Adolescent brain maturation and cortical folding: evidence for reductions in gyrification. *PLoS One*. 9:e84914.

Knecht S. 2000. Handedness and hemispheric language dominance in healthy humans. *Brain*. 12:2512–2518.

Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, Popovic D, Oeztuerk O, Haas SS, Weiske J, et al. 2020. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry*. 78:195–209.

Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, Paolini M, Chisholm K, Kambeitz J, Haidl T, et al. 2018. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry*. 75:1156–1172.

Kroenke CD, Bayly PV. 2018. How forces fold the cerebral cortex. *J Neurosci*. 38:767–775.

Lamballais S, Vinke EJ, Vernooij MW, Ikram MA, Muetzel RL. 2020. Cortical gyrification in relation to age and cognition in older adults. *Neuroimage*. 212:116637.

Llinares-Benadero C, Borrell V. 2019. Deconstructing cortical folding: genetic, cellular and mechanical determinants. *Nat Rev Neurosci*. 20:161–176.

Lohmann G, Von Cramon DY, Colchester ACF. 2008. Deep sulcal landmarks provide an organizing framework for human cortical folding. *Cereb Cortex*. 18:1415–1420.

Mareckova K, Miles A, Andryskova L, Brazdil M, Nikolova YS. 2020. Temporally and sex-specific effects of maternal perinatal stress on offspring cortical gyrification and mood in young adulthood. *Hum Brain Mapp*. 41:4866–4875.

Mathias SR, Knowles EEM, Mollon J, Rodrigue A, Koenis MMC, Alexander-Bloch AF, Winkler AM, Olvera RL, Duggirala R, Göring HHH, et al. 2020. Minimal relationship between local Gyrification and general cognitive ability in humans. *Cereb Cortex*. 30:3439–3450.

Matsuda Y, Ohi K. 2018. Cortical gyrification in schizophrenia: current perspectives. *Neuropsychiatr Dis Treat*. 14:1861–1869.

Miller TJ, Mcqlashan TH, Rosen JL, Cadenhead K, Ventura J, Mcfarlane W, Perkins DO, Pearlson QD, Woods SW. 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.

Musliner KL, Mortensen PB, McGrath JJ, Suppli NP, Hougaard DM, Bybjerg-Grauholm J, Bækvad-Hansen M, Andreassen O, Pedersen CB, Pedersen MG, et al. 2019. Association of Polygenic Liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population. *JAMA Psychiatry*. 76:516–525.

Namkung H, Kim SH, Sawa A. 2017. The insula: an underestimated brain area in clinical neuroscience, psychiatry, and neurology. *Trends Neurosci*. 40:200–207.

Nanda P, Tandon N, Mathew IT, Giakoumatos CI, Abhishekh HA, Clementz BA, Pearlson GD, Sweeney J, Tamminga CA, Keshavan MS. 2014. Local gyrification index in Probands with psychotic disorders and their first-degree relatives. *Biol Psychiatry*. 76:447–455.

Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, Essock S, Fenton WS, Frese FJ, Gold JM, et al. 2008. The MATRICS consensus cognitive battery, part 1: test selection, reliability, and validity. *Am J Psychiatry*. 165:203–213.

Palaniyappan L, Liddle PF. 2012. Differential effects of surface area, gyrification and cortical thickness on voxel based morphometric deficits in schizophrenia. *Neuroimage*. 60:693–699.

Palaniyappan L, Mallikarjun P, Joseph V, White TP, Liddle PF. 2011. Folding of the prefrontal cortex in schizophrenia: regional differences in gyrification. *Biol Psychiatry*. 69:974–979.

Palaniyappan L, Marques TR, Taylor H, Handley R, Mondelli V, Bonaccorso S, Giordano A, McQueen G, DiForti M, Simmons A, et al. 2013. Cortical folding defects as markers of poor treatment response in first-episode psychosis. *JAMA psychiatry*. 70:1031–1040.

Palaniyappan L, Marques TR, Taylor H, Mondelli V, Reinders AATS, Bonaccorso S, Giordano A, DiForti M, Simmons A, David AS, et al. 2016. Globally efficient brain organization and treatment response in psychosis: a connectomic study of Gyrification. *Schizophr Bull*. 42:1446–1456.

Palaniyappan L, Park B, Balain V, Dangi R, Liddle P. 2015. Abnormalities in structural covariance of cortical gyrification in schizophrenia. *Brain Struct Funct*. 220:2059–2071.

Papini C, Palaniyappan L, Kroll J, Froudist-Walsh S, Murray RM, Nosarti C. 2020. Altered cortical Gyrification in adults who were born very preterm and its associations with cognition and mental health. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 5:640–650.

Patel Y, Parker N, Shin J, Howard D, French L, Thomopoulos SI, Pozzi E, Abe Y, Abé C, Anticevic A, et al. 2021. Virtual histology of cortical thickness and shared neurobiology in 6 psychiatric disorders. *JAMA Psychiatry*. 78:47–63.

Pillay P, Manger PR. 2007. Order-specific quantitative patterns of cortical gyrification. *Eur J Neurosci*. 25:2705–2712.

Popovic D, Ruef A, Dwyer DB, Antonucci LA, Eder J, Sanfelici R, Kambeitz-Ilankovic L, Oztuerk OF, Dong MS, Paul R, et al. 2020. Traces of trauma: a multivariate pattern analysis of childhood trauma, brain structure, and clinical phenotypes. *Biol Psychiatry*. 88:829–842.

Rash BG, Duque A, Morozov YM, Arellano JI, Micali N, Rakic P. 2019. Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum. *Proc Natl Acad Sci U S A*. 116:7089–7094.

Rodrigue AL, McDowell JE, Tandon N, Keshavan MS, Tamminga CA, Pearlson GD, Sweeney JA, Gibbons RD, Clementz BA. 2018. Multivariate relationships between cognition and brain anatomy across the psychosis Spectrum. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 3:992–1002.

Rollins CPE, Garrison JR, Arribas M, Seyedsalehi A, Li Z, Chan RCK, Yang J, Wang D, Liò P, Yan C, et al. 2020. Evidence in cortical folding patterns for prenatal predispositions to hallucinations in schizophrenia. *Transl Psychiatry*. 10:387.

Rosen AFG, Roalf DR, Ruparel K, Blake J, Seelaus K, Villa LP, Ciric R, Cook PA, Davatzikos C, Elliott MA, et al. 2018. Quantitative assessment of structural image quality. *Neuroimage*. 169:407–418.

Sasabayashi D, Takayanagi Y, Takahashi T, Koike S, Yamasue H, Katagiri N, Sakuma A, Obara C, Nakamura M, Furuichi A, et al. 2017. Increased occipital Gyrification and development

of psychotic disorders in individuals with an at-risk mental state: a Multicenter study. *Biol Psychiatry*. 82:737–745.

Schaer M, Cuadra MB, Schmansky N, Fischl B, Thiran J-P, Eliez S. 2012. How to measure cortical folding from MR images: a step-by-step tutorial to compute local Gyrification index. *J Vis Exp*. 59:e3417.

Schäfer T, Schwarz MA. 2019. The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases. *Front Psychol*. 10:1–13.

Schmitt JE, Lenroot RK, Wallace GL, Ordaz S, Taylor KN, Kabani N, Greenstein D, Lerch JP, Kendler KS, Neale MC, et al. 2008. Identification of genetically mediated cortical networks: a multivariate study of pediatric twins and siblings. *Cereb Cortex*. 18:1737–1747.

Schultze-Lutter F, Addington J, Ruhrmann S, Klosterkötter J. 2007. *Schizophrenia Proneness Instrument, Adult Version (SPI-A)*. Roma: Giovanni Fioriti Ed.

Sotiras A, Resnick SM, Davatzikos C. 2015. Finding imaging patterns of structural covariance via non-negative matrix factorization. *Neuroimage*. 108:1–16.

Sotiras A, Toledo JB, Gur RE, Gur RC, Satterthwaite TD, Davatzikos C. 2017. Patterns of coordinated cortical remodeling during adolescence and their associations with functional specialization and evolutionary expansion. *Proc Natl Acad Sci*. 114:3527–3532.

Thapar A, Riglin L. 2020. The importance of a developmental perspective in psychiatry: what do recent genetic-epidemiological findings show. *Mol Psychiatry*. 25:1631–1639.

Thomas Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, et al. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*. 106:1125–1165.

Toga AW, Thompson PM. 2003. Mapping brain asymmetry. *Nat Rev Neurosci*. 4:37–48.

Upthegrove R, Lalousis P, Mallikarjun P, Chisholm K, Griffiths SL, Iqbal M, Pelton M, Reniers R, Stainton A, Rosen M, et al. 2020. The psychopathology and neuroanatomical markers of depression in early psychosis. *Schizophr Bull*. 47:249–258.

Van Den Heuvel MP, Sporns O, Collin G, Scheewe T, Mandl RCW, Cahn W, Goni J, Pol HEH, Kahn RS. 2013. Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry*. 70:783–792.

Wallace GL, Robustelli B, Dankner N, Kenworthy L, Giedd JN, Martin A. 2013. Increased gyrification, but comparable surface area in adolescents with autism spectrum disorders. *Brain*. 136:1956–1967.

Wang T, Wang K, Qu H, Zhou J, Li Q, Deng Z, Du X, Lv F, Ren G, Guo J, et al. 2016. Disorganized cortical thickness covariance network in major depressive disorder implicated by aberrant hubs in large-scale networks. *Sci Rep*. 6:1–12.

White T, Hilgetag CC. 2011. Gyrification and neural connectivity in schizophrenia. *Dev Psychopathol*. 23:339–352.

Yang Z, Oja E. 2010. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Trans Neural Networks*. 21: 734–749.

Yeh PH, Zhu H, Nicoletti MA, Hatch JP, Brambilla P, Soares JC. 2010. Structural equation modeling and principal component analysis of gray matter volumes in major depressive and bipolar disorders: differences in latent volumetric structure. *Psychiatry Res – Neuroimaging*. 184:177–185.

Novel gyrification networks reveal links with psychiatric risk factors in early illness

Supplementary Information

SUPPLEMENTAL METHODS

Study sample

Participants were healthy controls and patients recruited in 7 European cities out of 5 countries for the longitudinal European project PRONIA (www.pronia.eu). Specifics about recruitment procedures and assessments have been described elsewhere (Koutsouleris et al. 2018).

Briefly, general inclusion criteria were age between 15 and 40 years, sufficient language skills for participation as well as capacity to provide informed consent/assent. General exclusion criteria were an IQ below 70, current or past head trauma with loss of consciousness (> 5 minutes), current or past known neurological or somatic disorders potentially affecting the structure or functioning of the brain, current or past alcohol dependence, or polysubstance dependence within the past six months, and any medical indication against MRI.

Specific controls' exclusion criteria were any current or past DSM-IV axis disorder, a positive familial history (1st degree relatives) for affective or non-affective psychoses and an intake of psychotropic medications or drugs more than 5 times/year and in the month before inclusion. Patients comprised persons with a clinical high-risk state for psychosis (CHR), a recent onset psychosis (ROP) and a recent onset depression (ROD). CHR were included if they fulfilled either cognitive disturbances criteria assessed using the Schizophrenia Proneness Instrument (SPI-A,(Schultze-Lutter et al. 2007, 2012)) and/or 2) ultra-high-risk (UHR) criteria for psychosis based on the Structured Interview for Psychosis-Risk Syndromes (SIPS,(Miller et al. 2003)). CHR exclusion criteria were antipsychotic medication for more than 30 cumulative days at or above minimum dosage of the "first-episode psychosis" range of DGPPN S3 ("Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e. V.", German Association for Psychiatry, Psychotherapy and Psychosomatics) guidelines and any intake of antipsychotic medication within the past 3 months before clinical baseline assessments at or above the same minimum dosage. ROP participants had to meet the following two criteria: I) affective or non-affective psychotic episode according to criteria of the Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR) fulfilled in the past 3 months and II) onset of psychosis within past 24 months. ROP were excluded if their antipsychotic medication intake was longer than 90 cumulative days with a daily dose rate at or above the cited minimum dosage. ROD inclusion criteria were I) DSM-IV-TR major depressive episode (lifetime), II) major depressive disorder criteria fulfilled within past three months and III) duration of first depressive episode no longer than 24 months. Specific ROD exclusion criteria were: 1) more than 1 major depressive episode, 2) antipsychotic medication for > 30 cumulative days at or above the cited minimum dosage, and III) any intake of antipsychotic medication within the past 3 months before baseline assessments at or above the cited minimum dosage.

All participants provided written informed consent and the study protocol was approved by each ethical committee.

The full dataset at the start of analyses included 1492 persons (454 controls subjects and 1038 patients), of which 423 controls and 951 patients being recruited in the original 7 sites (i.e., Munich, Basel, Cologne, Birmingham, Milan, Turku and Udine) and the remaining 118 in three sites, which newly became part of the PRONIA consortium (i.e., Bari, Münster and Düsseldorf). We focused on the original PRONIA 7 sites-sample in order to maximize sample size for each site. At the time the present analyses started, neuroimaging data was available for 413 controls and 901 patients. In the controls, in order to match persons for age and sex and keep a relatively similar sample size across the sites, we further excluded participants from Milan and Birmingham; the first center because of the significantly smaller number of controls (i.e., 33 controls) and the second because of the significantly lower age range in the sample compared to Munich and Turku (i.e., mean age 25.12, SD 5.98). However, we used these two controls

held-out samples to externally validate our components solution (see below). In following this strategy, we effectively balanced the sample sizes, age and sex ranges across centers without enforcing a one-to-one matching strategy. The matched sample comprised thus 329 controls. In order to enable the external application of ComBat for site correction based on the controls sample (see ComBat section below for more information on this approach), patients recruited in the two mentioned PRONIA sites were excluded as well. Hence, 754 patients were included (see Supplementary Figure 1). We had to exclude further participants (21 controls and 41 patients) after running the cortical surface reconstruction pipeline (see MRI data processing section) and the neuroimaging quality control assessments (see MRI Quality control procedures section) because of poor image quality which led to FreeSurfer reconstruction errors or gyrification calculation failure (see details in next sections). Hence, our final sample was of 308 controls and 713 patients: 224 CHR, 263 ROP and 226 ROD (Supplementary Figure 1).

Information on sociodemographics, functioning and cognition for patients and controls and, additionally, clinical variables for patients (the latter subdivided by study-group) is listed in Table 1. Additionally, distribution of the sample in sites and respective age, sex, handedness and education characteristics can be found in Supplementary Table 1. Between sites, there were no significant age ($F_{(307)}$=1.27, p=0.28), sex ($\chi^2_{(9)}$=5.41, p=0.25) handedness ($F_{(280)}$=0.32, p=0.87) or education ($F_{(282)}$=0.78, p=0.54) differences in controls, while patients showed a significant difference in education only ($F_{(661)}$=2.96, p=0.02).

MRI data acquisition

When setting up the PRONIA study, we decided to generate an MRI database that would represent the MR scanner sequence heterogeneity encountered in clinical real-world. A minimal harmonization protocol was used that required the PRONIA sites to only 1) acquire isotropic or nearly isotropic voxel sizes of preferably 1 mm resolution, 2) set the Field Of View (FOV) parameters accordingly to guarantee the full 3D coverage of the brain including all parts of the cerebellum, and 3) define the relaxation time (TR) and echo time (TE) as well as other imaging parameters in a way that would maximize the contrast between cortical ribbon and the white matter and enhance the signal-to-noise ratio in the images. Supplementary Table 2 reports the acquisition parameters for all scanners of all PRONIA sites. At each site, all images were visually inspected for neuroanatomical abnormalities by experienced neuroradiologists, automatically defaced, and anonymized using an in-house FreeSurfer-based script prior to data centralization.

MRI data processing and gyrification calculation

All images were processed with the FreeSurfer software package (v. 6.0.0, https://surfer.nmr.mgh.harvard.edu/). The pipeline consists of the following main steps: bias field correction, labeling of the white matter and splitting the two hemispheres, removal of cerebellum and brainstem, formation of a triangular mesh that covers the external boundary of white and grey matter relying on voxel intensities, deformation and expansion of the mesh towards the pial surface and topology correction to fix structural abnormalities. In addition to the FreeSurfer default parameters, we added the following flags: -cw256, -3T, -multistrip and -clean-bm. -cw256 was included for images which have a FOV > 256 (in the PRONIA case, Turku), in order to conform the images to dimensions of 256^3. The -3T flag enables two specific options in recon-all for images acquired with a 3T scanner: 3T-specific Non-Uniform intensity correction parameters are used in the normalization stage, and the Schwartz 3T atlas is used for Talairach alignment. The -multistrip option was chosen to optimize the pre flooding height used by the watershed algorithm during the skull stripping step to find a boundary between the brain and skull. The FreeSurfer mri_watershed program uses a default pre flooding

height of 25 percent and, in general, a larger number will make the algorithm more conservative (i.e., if part of the brain has been removed), while a smaller one more aggressive (i.e., part of the skull has been left behind). This flag instructs the algorithm to run several different watershed thresholds simultaneously (i.e., 5,10,20,30) and to choose automatically the one that produces the best result. Finally, the -clean-bm flag enables the overwriting of the new edits on the old brainmask.mgz volume (https://surfer.nmr.mgh.harvard.edu/fswiki/).

After quality control assessments (see next section), gyrification meshes (163.842 vertices per hemisphere) were extracted using the Local Gyrification Index (LGI) approach (Schaer et al. 2012), resampled to the fsaverage6 surface (40.962 vertices per hemisphere) and smoothed with a 5mm full width and a half maximum (FWHM) isotropic Gaussian filter kernel. The Free-Surfer cortical reconstruction failed for 10 controls and 20 patients, due to artefacts which hindered the correct distinction between white and grey matter. The LGI calculation failed for 5 controls and 6 patients, hence in total 15 controls and 26 patients needed to be excluded from the final sample.

MRI quality control procedures

Quality control procedures consisted in extraction and investigation of the Euler number distribution for both the controls and the patient sample (Supplementary Figure 2). The Euler number is a quality measure provided by FreeSurfer, which gives a reliable estimation of the complexity of the cortical reconstruction and effectively detects outliers (Rosen et al. 2018). The highest Euler number would be 1, with the larger negative deviation from the unity meaning a more complex cortical surface, hence a higher chance that the reconstruction is flawed. However, it has been shown that the Euler number also highly correlates with cortical thickness per se (5), which, in the case of gyrification, could lead to an exclusion of just more complex (i.e., gyrified) cortices. We thus visually inspected all cortical surfaces deviating ≤ than 2 SD from the mean in order to detect only those, which did not reach a good quality standard according to the ADNI guidelines (http://adni.loni.usc.edu/methods/mri-tool/). Eight cortical surfaces for controls and 37 for patients, whose Euler number was lower than 2 SD from the mean, were visually inspected and, in total, 6 from controls and 15 from patients excluded.

ComBat

To tackle the site effects issue, which is highly common and expected in large consortia such as PRONIA, we applied ComBat (‘‘combating batch effects when combining batches”). ComBat is an empirical Bayes estimation method originally implemented to harmonize genetic data from different recruiting sites (Johnson et al. 2007). ComBat has been applied to several neuroimaging data modalities (e.g., (Fortin et al. 2017, 2018)) and has been proved efficient, especially in disentangling site effects from variance of biological nature (e.g., age, sex or disease), which can be retained. ComBat was implemented in Matlab (version R2020a; https://github.com/Jfortin1/ComBatHarmonization) and was applied to the resampled and smoothed LGI cortical maps in the control population using default settings. After the correction, no site-effects were detected in the resulting Patterns of Structural Covariance (PSC) (Supplementary Table 4). One of the advantages of this approach is that it gives the opportunity to retain biological variance in the sample (e.g., age and sex) while modelling and correcting for site effects. Disease effects could potentially also be modelled and spared from the correction, nevertheless, to lower the chances of losing any psychopathological variance, we decided to follow a more conservative approach: The estimates derived from the application of ComBat to the controls sample were applied, separately and with no modifications, to the patient sample.

Non-Negative Matrix Factorization and extraction of gyrification values

The orthonormal projective variant of Non-Negative Matrix Factorization (opNNMF) results in a sparse representation of the data—in our study of gyrification-based brain cortical meshes. The projectivity constraint was chosen specifically for its capacity to further enhance the sparseness of the PSCs by forcing every component to capture a distinct source of variation in the data—this leads to a distributed representation of gyrification across the brain. This property is especially valuable in the neuroimaging field, where a parts-based cortical map supports interpretation of, for instance, known distinct brain areas associated with specific functions. Moreover, each individual in the analysis' sample is assigned with a vector of K weights, or loading coefficients, (with K being the number of components chosen), representing the relative contribution of the Kth component to the whole cortical gyrification reconstruction for that subject.

Weights are, however, expressed in arbitrary units, which hinders potential interpretation regarding the underlying gyrification. Therefore, we implemented a straightforward approach to extract the original gyrification values for our 18-PSC solution in order to 1) relate to more interpretable values, and 2) to facilitate replicating the components by just sharing the maps (applicable to any gyrification mesh) and not needing to apply the models. First, the 18 resulting PSC maps (40.962 vertices for each component) were thresholded using the following calculation: (maximum value – minimum value) / 2. This threshold is also used for visualization of components, so that the resulting clusters completely overlap with their visual representation. This procedure was followed separately for the left and right hemisphere, as the PSCs do not exactly map in the same regions bilaterally.

Four components for the left (PSC 1, 10, 13 and 18) and three for the right hemisphere (PSC 2, 6 and 7) were asymmetrical, i.e., the structural covariance is concentrated on only one hemisphere. In fact, the surface mesh coverage after thresholding was much smaller for these components than the other PSCs (i.e., less than 10% of the full map compared to the largest PSC), which pointed to a weaker representation of covariance on the hemisphere. To assure a spatial balance throughout all components, as well as to avoid increasing the number of comparisons, these PSCs were excluded from further analyses.

The remaining binarized maps were used to calculate the mean gyrification values for each PSC across all controls, similar to the known parcellation, or Region of Interest (ROI) approaches. The ROI maps could then be applied to the patient sample, so that each individual was assigned, in total, 29 PSC-gyrification values (14 for the left and 15 for the right hemisphere), which could be compared in terms of absolute gyrification values. We investigated the associations between these PSCs and the original PSC-derived loading coefficients in the controls in order to determine if loading patterns were positively correlated with absolute gyrification values. Supplementary Table 3 shows correlations between the 18 loading coefficients-PSC and the 18 derived gyrification-PSCs for left and right hemisphere. All correlations were significant after correction for multiple comparisons (all p-values <0.0001), with r ranging from 0.218 to 1 for the left and 0.728 to 1 for the right hemisphere.

Neurocognitive battery

The neurocognitive assessment included following tests: Continuous-Performance Test – Identical Pairs (adapted tablet version, CPT-IP;(Cornblatt et al. 1988)), Diagnostic Analysis of Non-Verbal Accuracy (adapted tablet version, DANVA;(Nowicki and Duke 1994)), Auditory Digit Span forward and backward (adapted from the PEBL battery, DS;(Wechsler et al. 2008)), Digit-Symbol- Substitution Test (BACS battery, DSST;(Wechsler et al. 2008)), Rey Auditory Verbal

Learning Test (RAVLT; (McMinn et al. 1988)), Rey-Osterrieth complex figure (ROCF;(Osterrieth 1944)), Salience Attribution Task (adapted version, SAT;(Roiser et al. 2010)), Self-Ordered Pointing Task (adapted version, SOPT;(Ross et al. 2007)), Trail-Making Test A and B (TMT-A/-B;(Sánchez-Cubillo et al. 2009)), Verbal Fluency test phonemic and semantic (VF;(Harrison et al. 2000)) and the Matrix Reasoning and Vocabulary subtests from the Wechsler Adult Intelligence Scale, 4th ed. (WAIS-IV;(Wechsler et al. 2008)). Moreover, as the RAVLT was not available in Finnish, the revised version of the Hopkins Verbal Learning Test (HVLT-R;(Benedict et al. 2003)) was included in Turku's neuropsychological battery (Supplementary Table 8). Premorbid Intelligence Quotient (IQ) was estimated using the Vocabulary subtest from the WAIS-IV. For individuals with foreign backgrounds who were not able to complete the test, we administered the General Knowledge subtest from the WAIS-IV.

Harmonization of verbal learning scores

As already introduced, verbal learning in the PRONIA consortium is mainly measured with the RAVLT (McMinn et al. 1988), with Turku being the only site administering the HVLT-R (Benedict et al. 2003). A harmonization of the two scales presents 3 main challenges represented by the following differences between the two tests: 1) number of items (12 for HVLT-R and 15 for RAVLT), 2) number of trials (3 repetitions for HVLT-R and 5 for RAVLT), and 3) a semantic subgrouping for HVLT-R, but not for RAVLT. To harmonize measures, we recruited 36 healthy volunteers from 5 out of 7 PRONIA sites (i.e., Munich, Milan, Udine, Cologne and Birmingham, mean age 23.17, SD 6.22, 61.1% males) and administered both tests, counterbalancing the order, within at least 2 hours. Not to lose any information from the RAVLT test and relying on the known floor effects shown after its third trial (Tierney et al. 1994), we calculated the sum of the 5 RAVLT-repetitions and built a linear regression model as such: RAVLT-sum5 = a * HVLT-R-sum3 + b, with a = 25.51264904, slope and b = 1.191092045, intercept of the regression. Using this translator, the HVLT-R data from Turku's participants could be transformed in the sum of RAVLT's 5 repetitions.

Neuropsychological tests used to construct the cognitive scores

In order to avoid multiple testing, we calculated 6 main cognitive domain scores and one composite score for global cognition following a highly comparable approach to that of the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS (Nuechterlein et al. 2008)) recommendations. Available PRONIA neuropsychological tests (see Supplementary Table 8 and below for detailed description) were used to calculate 6 out of 7 cognitive domains from MATRICS: social cognition, working memory, speed of processing, verbal learning, reasoning and attention. We did not include visual learning (as in the original MATRICS) in the cognitive scores' computation because no test in the PRONIA neuropsychological battery could be compared to either the Neuropsychological Assessment Battery, shape learning subtest (Zgaljardic and Temple 2010), or the Brief Visuospatial Memory Test-revised (Zgaljardic and Temple 2010).
Social cognition deviated from the tests selected in MATRICS and was calculated using the DANVA. Working memory was computed by summing the forward and backward trials of the ADS—a very similar test to the digit sequencing subtest of the BACS included in MATRICS. To calculate a composite score of speed of processing, we relied on three tests also used in MATRICS and averaged two graphomotor tests (i.e., the TMT-A and the DSS) and one verbal test (i.e., semantic VF, correct words in 60 seconds). To calculate verbal learning, we took the sum of the first 3 RAVLT trials, in order to enhance similarity with the HVLT-R in MATRICS, which only has three repetitions. Reasoning was assessed with the Matrix subtest from WAIS IV, raw scores, while Attention using the CPT-IP, as in MATRICS. After Z-score transformation

of the cognitive domains based on the controls sample, we computed a composite score of Global Cognition by calculating the aggregate average across the 6 standardized scores.

The PRONIA neuropsychological tests used for the calculation of the sores were the following:

1. *Diagnostic Analysis of Non-Verbal Accuracy, Affective Faces trial* (DANVA-2-AF;(Nowicki and Duke 1994)): Simple socio-cognitive test, consisting in the recognition of 24 emotional face expressions (*i.e.*, classified as: 'Happy', 'Sad', 'Angry', or 'Fearful'). DANVA-2-AF was administered and scored with tablet support, resulting in a 0-24 total score (number of correct responses).

2. *Auditory Digit Span, Forward & Backward trials* (ADS-F&B;(Wechsler et al. 2008)): Test of verbal short-term memory and verbal working memory. In a first, *forward*, set of trials participants listened to sequences of numbers, progressively increasing in length (16 trials, from two to nine digits), and they had to repeat them in the same order to the examiner. In a second, *backward*, set of repetitions in reverse order was requested (14 trials, from two to eight digits). Numbers were presented at one digit per second rate by a recorded male voice, while the examiner registered responses on the tablet during administration with automatic scoring and stimuli presentation. Each condition was interrupted in case of two errors on a list of the same length. ADS-F&B resulted in a total score, corresponding to the number of correct responses (theoretical range: 0-30), and in a score for the *backward* trials only (0-14), intended as verbal working memory measure.

3. *Verbal Fluency, Phonemic & Semantic trials* (VF-P&S; (Harrison et al. 2000)): Test of verbal fluency. Participants were asked to produce as many words as possible within one minute. In the *phonemic* condition words should begin with a given letter (dependent on assessment language); in the *semantic* condition names of 'animals' were requested. The produced words were audio recorded and a rater registered the number of correct responses, of repetitions, and of errors (*e.g.*, fantastic animals) after the administration. The total test scores correspond to the number of correct words produced in each condition.

4. *Rey Auditory Verbal Learning Test* (RAVLT;(McMinn et al. 1988)): Auditory verbal learning task consisting in learning a list of 15 semantically unrelated words (*List-A*) in five consecutive repetitions (*immediate memory* trials), and in their retention over a 30-minutes time interval (*delayed memory* trial). Before each *immediate* memory trial, List-A was presented at one word per second rate by a recorded male voice. After the fifth trial, a different list (*List-B*) was presented and recalled (*Interference* trial), then participants were requested to recall List-A again, without re-presenting it (*Post-interference* trial). Since RAVLT was not available in Finnish, the revised version of the *Hopkins' Verbal Learning Test* (HVLT-R;(Benedict et al. 2003)) was included in the Finnish version of the neuropsychological battery. HVLT-R includes a single 12-word list, repeated three times instead of five, without interference and post-interference trials. To avoid interferences, only non-verbal tests from the neuropsychological battery were administered before the *delayed* memory trial. The examiner registered responses on the tablet during administration, with automatic scoring number of *correct* words, *repetitions*, *out-of-lists* words, and *interferences* (*i.e.*, words from incorrect list; for RAVLT only). To summarize verbal learning performances in the tests, the total number of correct words in *immediate* memory trials were considered, together with a measure of slope in performances. Also, we considered the number of correct words produced after the first presentation of List-A (and of List-B, for RAVLT only) as measure of short-term verbal memory, while that for *delayed* memory trial (and the post-interference one, for RALVT only) was intended as a measure of long-term verbal memory.

5. *Trail Making Task, A & B trials* (TM-A&B;(Sánchez-Cubillo et al. 2009)): Test of processing speed, also requiring sequencing, graphomotor capacity, visual attention, search ability and

flexibility. The test started with a simpler request (consisting in connecting in sequence 25 numbered stimuli scattered on a sheet of paper; part *A*), followed by a more complex one (in which participants needed to concurrently consider two different sequences, numbers and letters; part *B*). TM-A&B was administered in paper-pencil format with tablet support (i.e., to register the time taken to complete each part). During the test, the examiner registered any *error* (*e.g.*, connect two wrong stimuli) or rules *violation* (e.g., lift the pencil from the paper), immediately correcting participants. The main score of the test was the difference in seconds needed to complete part B and part A (switch time linked to task complexity). Time needed to complete part A was also considered, as measure of processing speed and graphomotor skills in a simple task.

6. *Continuous Performance Test, Identical Pairs version* (CPT-IP;(Cornblatt et al. 1988)): Test of selective and sustained visual attention. The participants were asked to watch at a series of 300 four-digits number, presented at the rate of one per second, and to respond as fast as possible when the number presented is identical to the preceding one. Stimuli were constituted by: *targets* (numbers identical to the preceding one; 61 trials), *distractors* (numbers similar, but not identical, to preceding one; *e.g.*, composed by the same digits in a different order; 59 trials), and *fillers* (number not-related to the preceding one; 180 trials). Different kinds of stimuli were randomly distributed in six 50-trials intervals. CPT-IP was administered and scored with tablet support. Reaction time and response list were recorded, so that each response was classifiable as: *Hit response* (response to a target), *Commission error* (false alarm response to a distractor), *Random/Distraction error* (false alarm response to a filler), *Missing response* (no-response to a target) and *Correct rejection* (no-response to a distractor or a filler). The test provided also a *Reaction time* (RT) measure, taken for Hit responses, and two *Sensitivity index* (d') responses, for Hits on Commission errors on and for Hits on Random/Distraction errors. For RT and both d', we also considered 50-trials slopes in performances.

7. *Digit Symbol Substitution Test* (DSST;(Wechsler et al. 2008)): Test of sustained attention, working memory, and processing speed. The participants were presented with a table univocally associating nine symbols to as many digits, together with an answer sheet listing 110 symbols, randomly arranged by rows. The task required writing the correct digit below each symbol, proceeding in order. The time limit was fixed at 90 seconds. The DSST was administered in paper-and-pencil, using tablet to give the stop signal. Test score was the difference between the number of correct responses and that of errors.

Description of clinical instruments

The associations between PSCs and clinical variables were investigated by choosing instruments which describe 1) positive, negative, disorganization and general symptoms, 2) subjective cognitive disturbances, and 3) depressive symptoms. For this purpose, we took the SIPS (Miller et al. 2003), the 'cognitive disturbances' subscale (COGDIS) derived from the SPI-A (Schultze-Lutter et al. 2007), containing 9 basic symptoms that describe subjective disturbances of cognitive nature, and the sum score of the Beck's Depression Inventory (BDI, (Beck and Steer 1984)) self-rating instrument.
The SIPS contains 5 main positive (P), 6 negative (N), 4 disorganization (D) and 4 general (G) symptoms items. The choice of this questionnaire was based on its transferability also to CHR subjects (while for instance the PANSS is considered to be less sensitive instrument in this study population (Schultze-Lutter et al. 2013)). Because ceiling effects might be expected when taking single items only, as the items' scale ranges from 1 to 6—6 being the full-psychotic level—we calculated composite scores by summing the main items. As a sanity-check, we investigated the distribution of SIPS-P composite scores in the patient population, noticing no ceiling effects (Supplementary Figure 3).

SUPPLEMENTAL ANALYSES

Selection of the best-fitting PSC solution

To evaluate the NNMF performance, we calculated the residual, or gradient, of the reconstruction error for a range of PSC resolutions (i.e., even numbers from 2 to 100). The reconstruction error decreases monotonically with higher number of components, while its gradient signalizes the point where a relatively smaller additional decrease is observed, thus representing the PSC solution, which fits most of the variance in the data (Figure 1B).

To test the reproducibility of the components obtained, we randomly split the controls sample in two halves with similar age, sex and site distribution (age: $t_{(306)}=0.5$, p=0.56; sex: $\chi^2_{(3)}=0.65$, p=0.41; site: $\chi^2_{(9)}=4.9$, p=0.29) and reconstructed 2 to 100 components for each half, independently. The degree of spatial overlap was calculated using the inner product (mean and median shown in Figure 1A). The median and mean inner product showed 2 main peaks at the 2- and 18-PSC solution. The inner product tends to be higher when less and larger components are estimated, because the spatial overlap is, by chance, stronger. Thus, we excluded the 2-PSC solution and investigated the 18-PSC one further. For completeness, both the 2-PSC and the 10-PSC (an additional, though lower peak in the reproducibility plot, Figure 1A) are represented in Supplementary Figure 7. When 2 components are estimated, most of the bilateral lateral cortex and the medial temporal lobe, perirhinal and enthorinal cortex, parahippocampal and fusiform gyrus are represented in PSC 1, while the second component reflects the asymmetrical tendency towards the right hemisphere we found in our chosen solution and in further analyses. The 10-PSC solution highly resembles some of our 18 gyrification components, advocating for the stability of the method. Supplementary Figure 6 shows the 18-components solution for both split-halves samples. Noteworthy, most of the components are almost indistinguishable between the two samples, with few exceptions: PSC 1 and 2 in split1 and 10 and 6 in split2 map the same area asymmetrically, while PSC 1 in split2 and PSC 3 in split2 are bilateral. PSC 17 in split2 is unique. Supplementary Figure 8 represents the 18 PSCs. Our gyrification components diverge from the known gyral and sulcal patterns of traditional atlases (see *Comparison with structural and functional atlas* below and Supplementary Figure 4 and 5).

Comparison with structural and functional atlases

In order to investigate whether our gyrification-based PSCs are associated with I) a known and often used structural atlas based on gyral and sulcal morphometry and II) established networks reflecting the cortical organization based on resting-state functional connectivity, we calculated the spatial overlap between our PSCs and, respectively, the 35 parcels from the Desikan-Killiany atlas (9), and the 17 functional networks identified by Yeo et al. (10). The spatial overlap was estimated by binarizing both the PSC and the 35 parcels and 17-networks maps and calculating the percentage of overlapping vertices.

Supplementary Figure 4 shows the spatial overlap between the 18 PSC and the 35 parcels of the Desikan-Killiany atlas. We notice high overlap between small Desikan-parcels and the PSCs. However, the majority of the gyrification components shows wider surface coverage and thus only partially coincide with the parcels. For instance, parcel 10 overlaps for 92.4% with PSC4 because of the low area covered compared to the PSC. Yeo networks, reflecting the cortical organization based on resting-state functional connectivity, overlap to a lesser extent with the 18 PSC (Supplementary Figure 5). This might result from the intrinsic nature of

functional connectivity, which comprises multiple brain areas for each network, whereas the PSCs are sparser and locally distinguishable. This double comparison suggests that our gyrification-driven parcellation of the cortex deviates from the known structural and functional atlases.

External validation

In order to externally validate our 18-PSCs solution, we applied the NNMF pipeline on the healthy controls recruited in the two held-out sites (i.e., Birmingham and Milan, N=84 cortical surfaces, which went through the LGI calculation and the quality control procedure). We first applied ComBat as for the main analyses in the pooled controls sample to the resampled and smoothed LGI-meshes of the two held-out sites, retaining effects of age and sex. Notably, these two subsamples presented both strong demographic (i.e., age effects, $t = -3.703$, $p<0.001$) and scanner differences, as Milan and Birmingham used a 1.5T and a 3T scanner, respectively. Therefore, ComBat was not completely effective in controlling for site effects. Second, we ran the NNMF pipeline with the same settings as for the pooled analyses, selecting specifically the 18-PSC solution.

We compared the original and replication PSCs qualitatively (Supplementary Figure 9) and quantitatively using the inner product calculation, following the approach used to choose the best fitting PSC solution (see section *Selection of the best fitting PSC solution*). The inner product expresses the degree on spatial overlap of the gyrification components and thus is an objective criterion to evaluate replication's performance.

Results show that 13 PSCs out of 18 (i.e., 72%) showed an inner product above 0.5 (i.e., [0.52:0.97]), while the rest 5 remain under the 0.5 range (i.e., [0.34:0.47]). The mean maximum inner product for all PSCs was 0.62, which was in line with our split-half analyses results leading to the choice of the PSC solution (i.e., mean inner product of 0.623; $t_{(17)}=-0.01$, $p=0.493$; Figure 1, main manuscript). Of note, the inner product estimates the exact overlap in space of components' values on the cortical ribbon, hence, in our case, it was potentially unlikely to reach high precision when taking into consideration the high dimensionality of the cortical mesh—i.e., ~80.000 vertices. However, the qualitative comparisons of both the split-half analyses (Supplementary Figure 6) and the PSC-replication (Supplementary Figure 9) show highly similar components, suggesting the strength of the NNMF technique in detecting a highly comparable underlying signature in independent gyrification data.

To test whether the pattern of reduced gyrification found in the patients of the discovery sample could also been replicated in the external sample, we used the ROI-approach described in the section *Non-Negative Matrix Factorization and extraction of gyrification values*. In brief, we extracted the 18 gyrification components (calculated in the discovery sample) in the replication dataset—i.e., 84 healthy controls and 147 patients—and subsequentially ran linear models to predict the effect of group (i.e., healthy controls vs. patients) controlling for age and sex. Results are presented in Supplementary Table 11 and show a very similar pattern of reduced gyrification in patients from the external validation sample as compared to healthy controls from the same sample. Specifically, 12 out of 14 PSCs showing significant differences between healthy and psychiatric individuals ($p^{FDR}<0.01$, in bold) overlap with the original PSCs showing group effects in the discovery sample, while 4 more PSCs in the external validation sample show a group effect. Effect sizes ranging from [-0.185:-0.294] are slightly stronger than in the original analyses. These results are promising and suggest that patients from independent clinical cohorts show a highly comparable pattern of lower gyrification in NNMF-based structural covariance components.

We could not find any significant associations between replication-PSCs and cognitive domains, potentially because of both the smaller sample size in the replication dataset, and the rather weak effects found on gyrification related to behavioral measures, as reported in recent literature (Mathias et al. 2020) and empirically found in our study.

Associations with age and sex

Age and sex associations on the 18 PSC-loading coefficients for controls were investigated using correlations and t-tests and are shown in Supplementary Table 4. Furthermore, in order both to relate to interpretable gyrification patterns, and to investigate associations in the patient sample, age and sex effects were estimated on the 29 gyrification-PSCs derived from the 18-PSC solution (Supplementary Table 5 and 6).

We explored the potential impact of non-linear age effects on gyrification by augmenting the linear models (prediction of each PSC with group—controls vs. patients—age and site as covariates) with age^2 and comparing the original with the augmented models using ANOVAs (p-values corrected FDR<0.01). Following a similar approach, we investigated whether there were differential developmental trajectories in gyrification in patients compared to controls, and whether these differences were specific to study-groups. To test this hypothesis, interaction effects of age-by-group (controls vs. patients) and age-by-study-groups (CHR, ROD, ROP) were added to the linear models and the augmented models were compared to the original ones.

Quadratic age significantly influenced five (bilateral PSC 3 and 9 and right PSC 18) out of 29 linear models (all five $p^{FDR}<0.006$). However, the gained explained variance when considering quadratic models was neglectable, with a R^2 gain ranging from 0.3% to 0.7%, while the specific effect of group for those models originally showing group effects (i.e., right PSC 3, 9 and 18) was maintained (all p^{FDR} for the group variable <0.0006). We found no interactions of age with group, neither when considering the pooled patient sample, nor when investigating the study-groups (all p^{FDR} resulting from ANOVAs>0.3).

Potential confounding factors

Associations between gyrification-PSC and cognitive and clinical measures in the patient sample could be potentially explained by other influencing factors, such as quality of MRI scans or the patients' educational level. Specifically, the Euler number is a measure of the cortical complexity resulting from surface reconstruction and is thought to be a proxy of scan quality (Rosen et al. 2018). However, it has also been shown that the Euler number correlates highly and heterogeneously with cortical thickness *per se*, age and symptoms in patient populations (Rosen et al. 2018; Zabihi et al. 2019). As gyrification is also an indirect measure of cortical complexity, the Euler number might in fact be too collinear with many measures of interest in our study. In order to corroborate this hypothesis, we investigated correlations between the Euler number and our PSCs both in the healthy and the patient sample, as well as individuals' age, because a good scan quality measure should mirror mainly MRI scan abnormalities and not be associated with other biological variance of interest. Supplementary Table 7 depicts very strong associations between the Euler number and both the gyrification-based PSCs (all p-values < 0.0002, with r ranging from -0.208 to -0.436) and age, in patients ($r=0.165$, $p=9.01E^{-06}$) and controls ($r=0.322$, $p=6.78E^{-09}$). In light of these results, we conclude that the use of the Euler number as a scan quality proxy might be restricted to an initial flagging of the extreme deviant cortical cortices. Indeed, this measure is strongly associated with several variables of interest (i.e., age, disease, and gyrification) and thus should be used cautiously in specific analyses.

Effects of MRI scan quality were therefore investigated by taking 1) the weighted average Image Quality Rating (IQR), i.e., a CAT12 toolbox (http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf)-based quality index combining NCR (Noise Contrast Ratio), ICR (Inhomogeneity Contrast Ratio) and RES (RMS resolution), 2) the Signal to Noise Ratio and 3) the Contrast to Noise Ratio, both the latter ones also resulting from the CAT12 volume reconstruction pipeline. Education was measured by summing the person's total years of education. We repeated statistical analyses augmenting the linear models with the three measures of MRI scan quality and years of education as terms in the regressions.

Results for the linear models investigating effect of group (i.e., controls vs. patients) remained comparable to the original results (Supplementary Table 9) after controlling for each of the other 3 quality measure in the linear model (Supplementary Table 10), with 1) same results for the IQR, 2) 92.8% overlap for SNR, and 3) the CNR case detecting more specific relationships (i.e., four PSCs survived correction for multiple comparisons), yet all other PSCs showing trends to significance (Supplementary Table 10).
We also investigated whether MRI quality measures interfere with the cognition and functioning results, by controlling the linear models for each of the 3 quality measures cited. For the cognition domain, as can be observed in Supplementary Table 14, results for models with the three quality measures were highly comparable to the original ones (Supplementary Table 12), with two additional significant associations (i.e., right PSC 1 and PSC 12 with attention for SNR), and one association not surviving multiple comparisons correction in the CNR-linear models, yet showing a clear trend (i.e., left PSC 12 with working memory, p=0.004). Similarly, results within the functioning domain when correcting for scan quality (Supplementary Table 19) were highly overlapping with the original models (Supplementary Table 16). The main detected differences were 1) significance for the left PSC 6 in the linear models with both CNR and SNR to predict GF:R highest past year, and 2) the right PSC 12 did not survive multiple comparison correction in the relationship with GF:R lowest past year, though showing a clear trend (p=0.004 for CNR and 0.001 for SNR). Hence, we can conclude that MRI scan quality did not significantly affect relationships between PSC and cognitive/clinical variables.

In order to investigate whether the patients' educational level might explain associations found between our PSCs and cognitive or functioning domains, we investigated correlations between patients' PSCs and education, and repeated cognition- and functioning-linear models adding years of education as an additional covariate. Results showed that, in patients, education was not correlated with any of the PSCs (all p>0.004, FDR-corrected). Linear models' results are reported in Supplementary Table 14 for cognition and Supplementary Table 17 for functioning and depict an almost complete overlap with our original results. Only one additional PSC (i.e., right PSC 12) was significant for working memory and attention and one PSC (the right PSC 9) lost its significance in the functioning model (p=0.002, all p-values FDR-corrected p<0.01 and all linear models' overall p-values were significant).

We further investigated the influence of premorbid intelligence (calculated using IQ, see supplemental section 'Neurocognitive battery') on the associations found between PSCs and both cognitive and functioning domains. Post-hoc partial correlations (correcting for age, sex and IQ) were calculated between the 14 PSCs showing a group effect and cognitive and functioning domains. When correcting multiple comparisons with the stringent threshold used throughout analyses (p^{FDR}<0.01), no cognitive associations survived, while relationships with functioning were more specific, i.e., 3 out of the original 12 PSCs were associated with GF:R highest lifetime (left PSC 6 and 12 and right PSC 16, p<0.0002, Supplementary Table 18). However, when using a more lenient correction threshold of p^{FDR}<0.05, four temporal PSCs (left 1, right 2 and bilateral 12) were associated with speed of processing, reasoning and global cognition (p<0.008, Supplementary Table 15), while all functioning results highly resembled the original analyses, except for the right PSC 9, showing only a trend to significance (p=0.008,

Supplementary Table 18). An overview of all analyses on clinical and neuropsychological associations, as well as a comparison with the models investigating the role of IQ in these relationships, is depicted in Supplementary Figure 10.

Investigation of clinical study groups

The linear models on effect of group revealed a set of PSCs, which differed between controls and the patient group. However, it might be plausible that the effects observed were mainly driven by a specific group of patients, for instance those experiencing the most burdening symptoms. In PRONIA, patients have been recruited following three study-specific diagnostic categories (see Study sample) and are divided in ROD, CHR and ROP. We investigated this hypothesis by fitting 3 post-hoc linear models for the 14 PSCs showing a group effect (dependent variables) with group (i.e., controls vs. patients) and age and sex as covariates and correcting each time for 2 out of the 3 study-groups, thus investigating effects of the left-out study group on the pooled group effect.

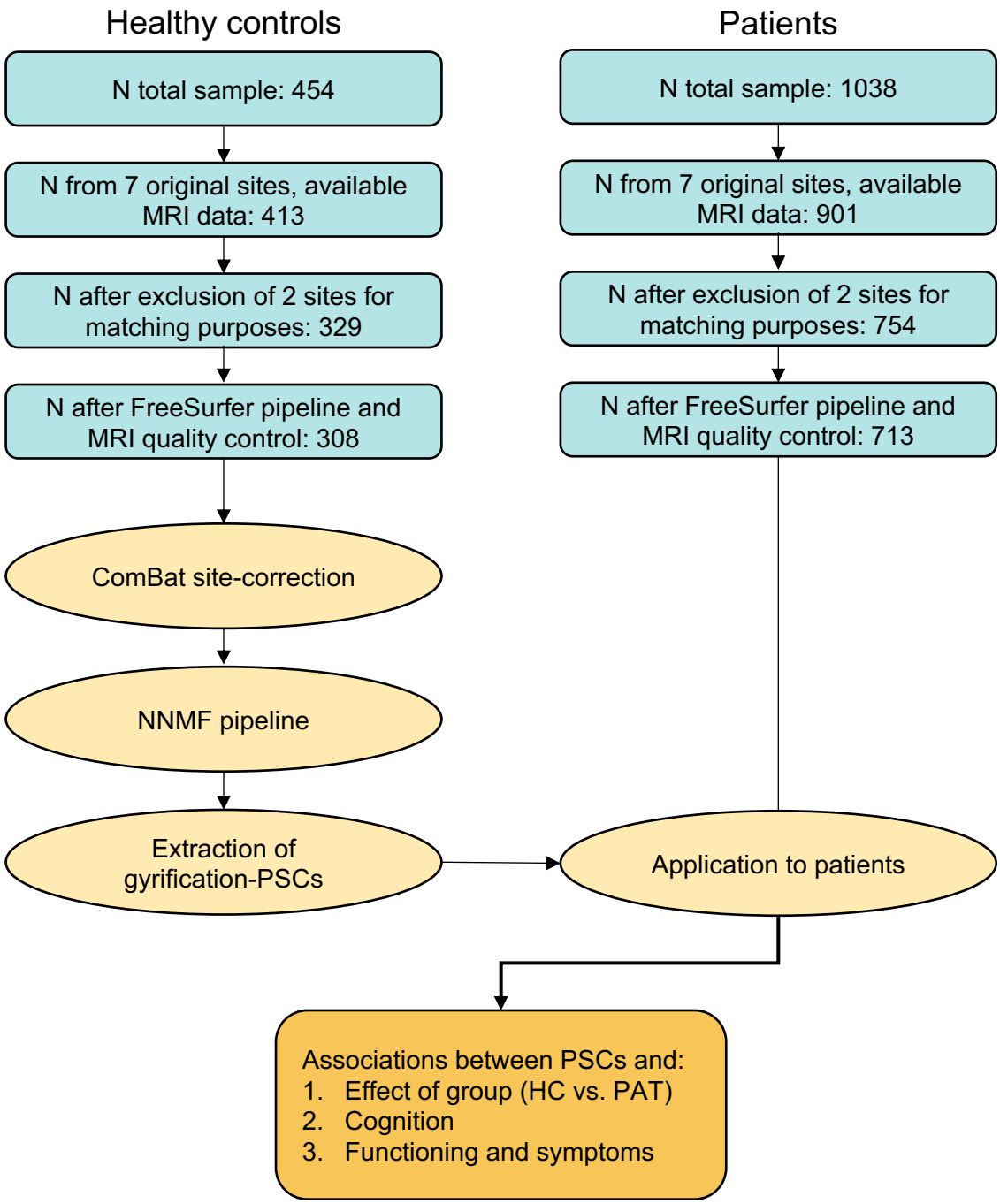
Results can be observed in Supplementary Table 21. Individuals with a first episode psychosis showed the most significant and spread effects (partial r range [-0.148:-0.059], p range [$0.01:6^{E-6}$]), with 11 significant PSCs out of the 14 originally detected (p^{FDR} <0.01). In CHR effects were comparable (partial r range [-0.134:-0.048]), though less PSCs survived correction (p range [$0.013:1^{E-5}$]). In the first episode depression group, no PSCs survived correction for multiple comparisons at p^{FDR} <0.01, however effects could be detected at the more lenient p^{FDR}<0.05 correction threshold for all 14 PSCs (partial r range [-0.105:-0.071], p range [$0.023:7^{E-4}$]).

Furthermore, we investigated differences between the three study-groups in the 14 PSCs fitting a linear model for the effect of study-group (independent variable) on each PSC (dependent variable) and correcting for age and sex. Results showed no significant differences between ROD, CHR and ROP (p values uncorrected range: [0.848:0.094]. Hence, although we notice a staging effect of disease in the psychosis and psychosis-prone study groups, this pattern suggests that each study-group alone expresses an overlapping disease signature, and that the pooled patient group effect cannot be completely explained by one psychopathology alone.

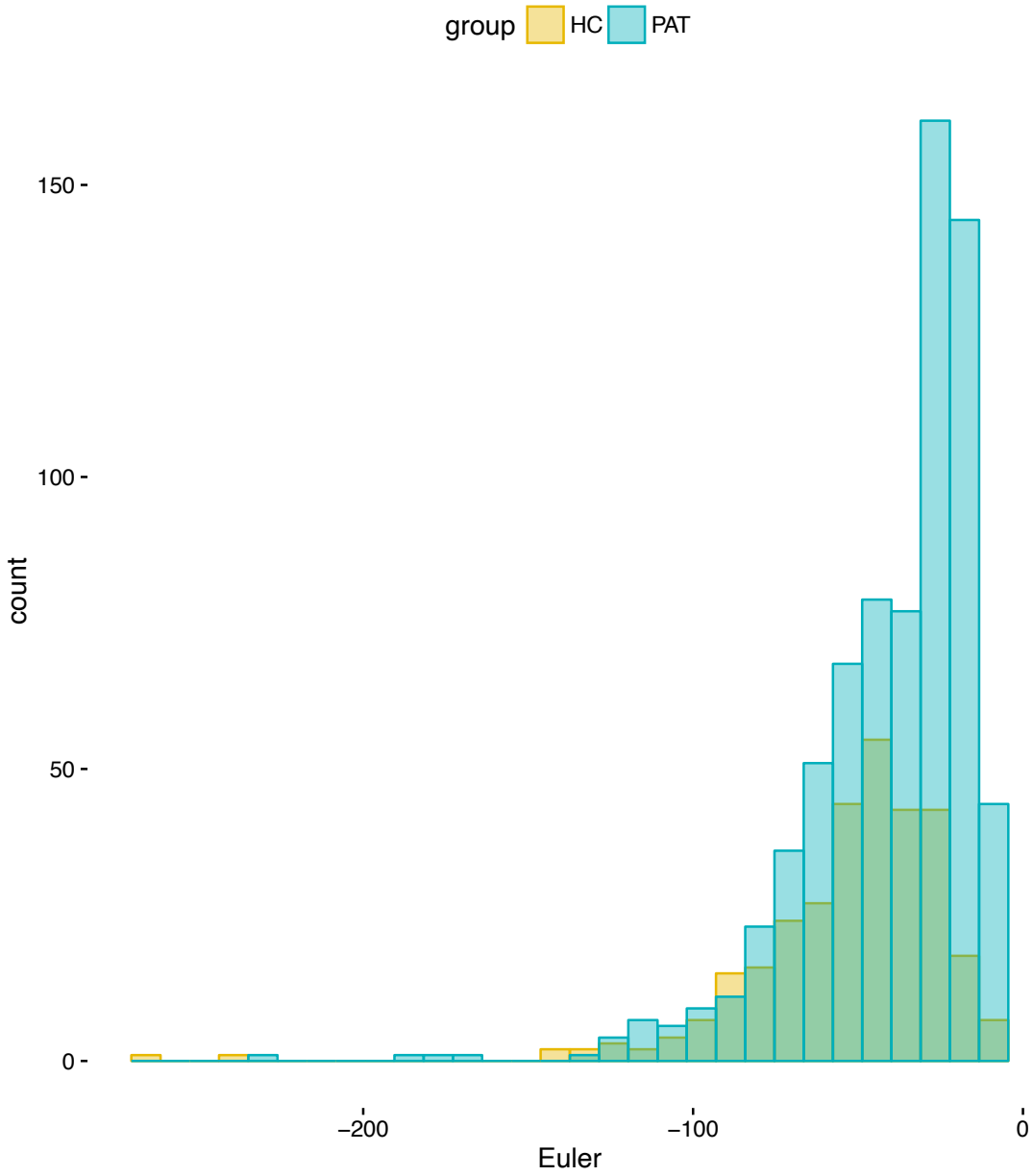
Associations between PSCs and clinical and cognitive measures in healthy controls

We repeated all statistical analyses ran in the transdiagnostic patient group in the controls sample to investigate potential disease-independent associations between the PSCs and the same cognitive and clinical measures investigated in patients. The linear models included age, sex and each of the 14 PSCs that was identified as significant in the group analysis as independent variables to predict each 7 cognitive domains (dependent variables), and the 8 GF-subscales. Results were FDR-corrected for multiple comparisons (p-FDR<0.01) and none of the associations survived correction. A few trends for associations in the cognitive domain could be observed (Supplementary Table 22); those between the left PSC 2 and the right PSC 1 and global cognition resemble those found in the patient sample, while other not significant results are not aligned with the original analyses.

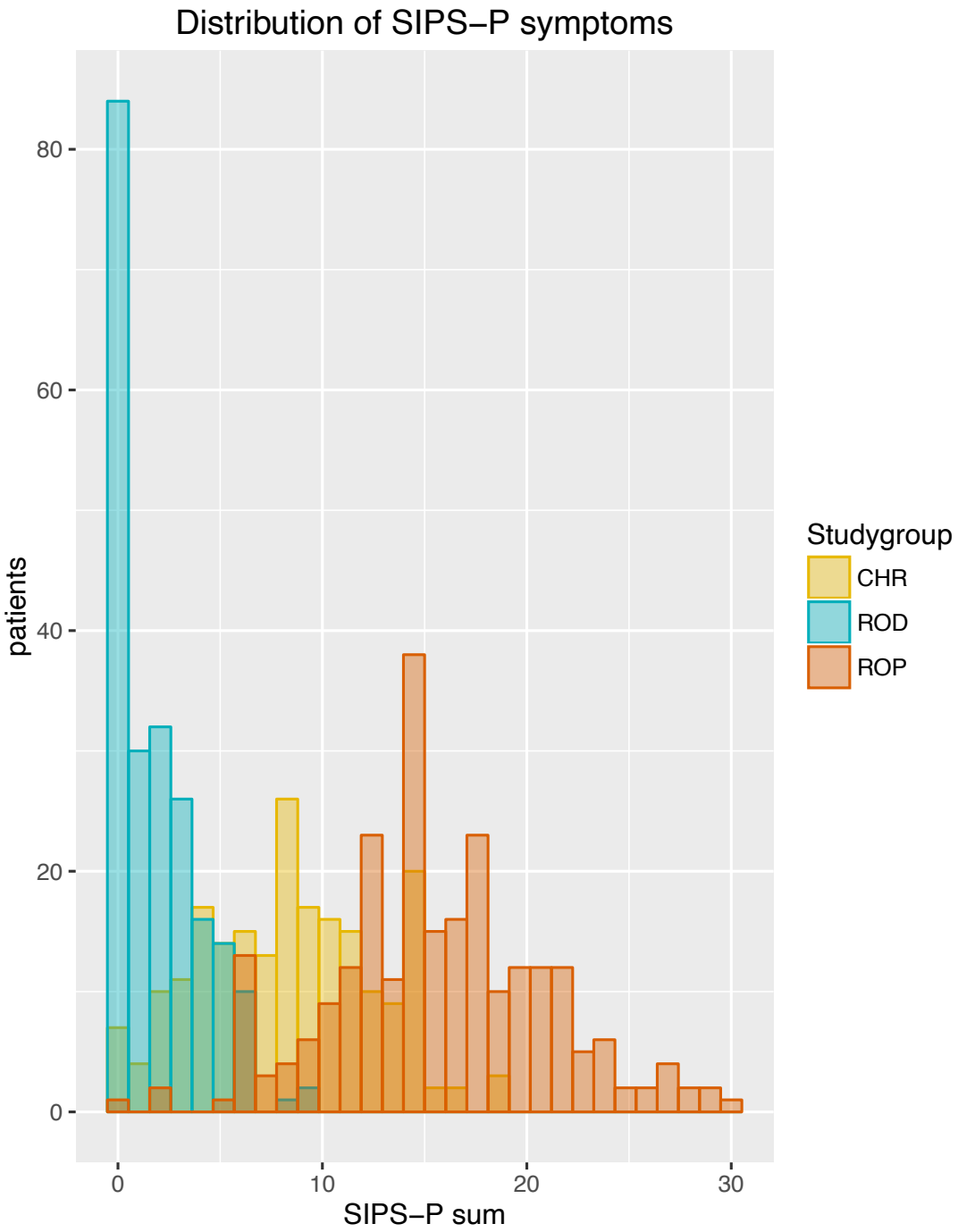
.



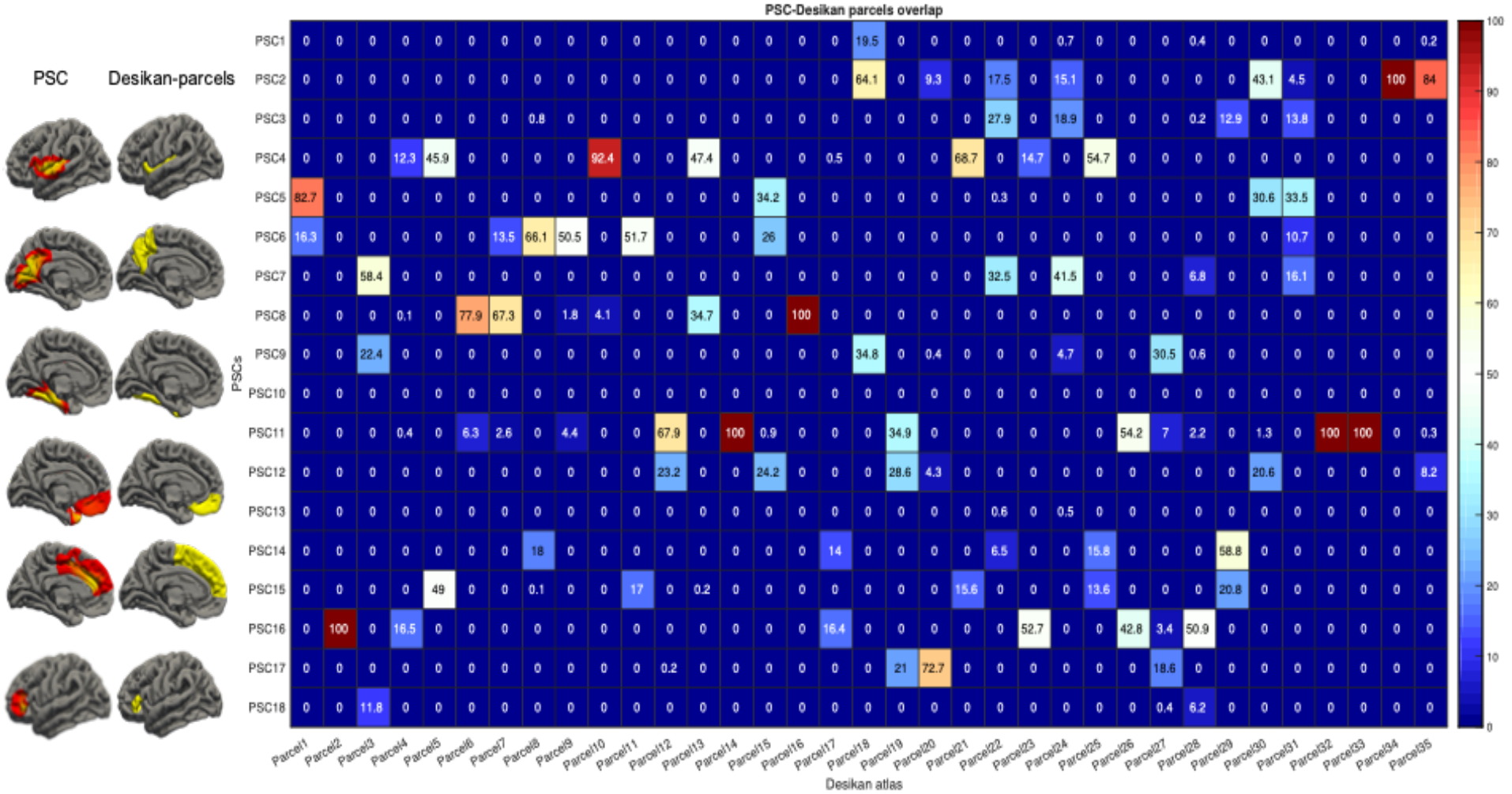
Supplementary Figure 1. Analyses outline. Abbreviations: NNMF: Non-Negative Matrix Factorization; PSC: Pattern of Structural Covariance; HC: Healthy Controls; PAT: patients.



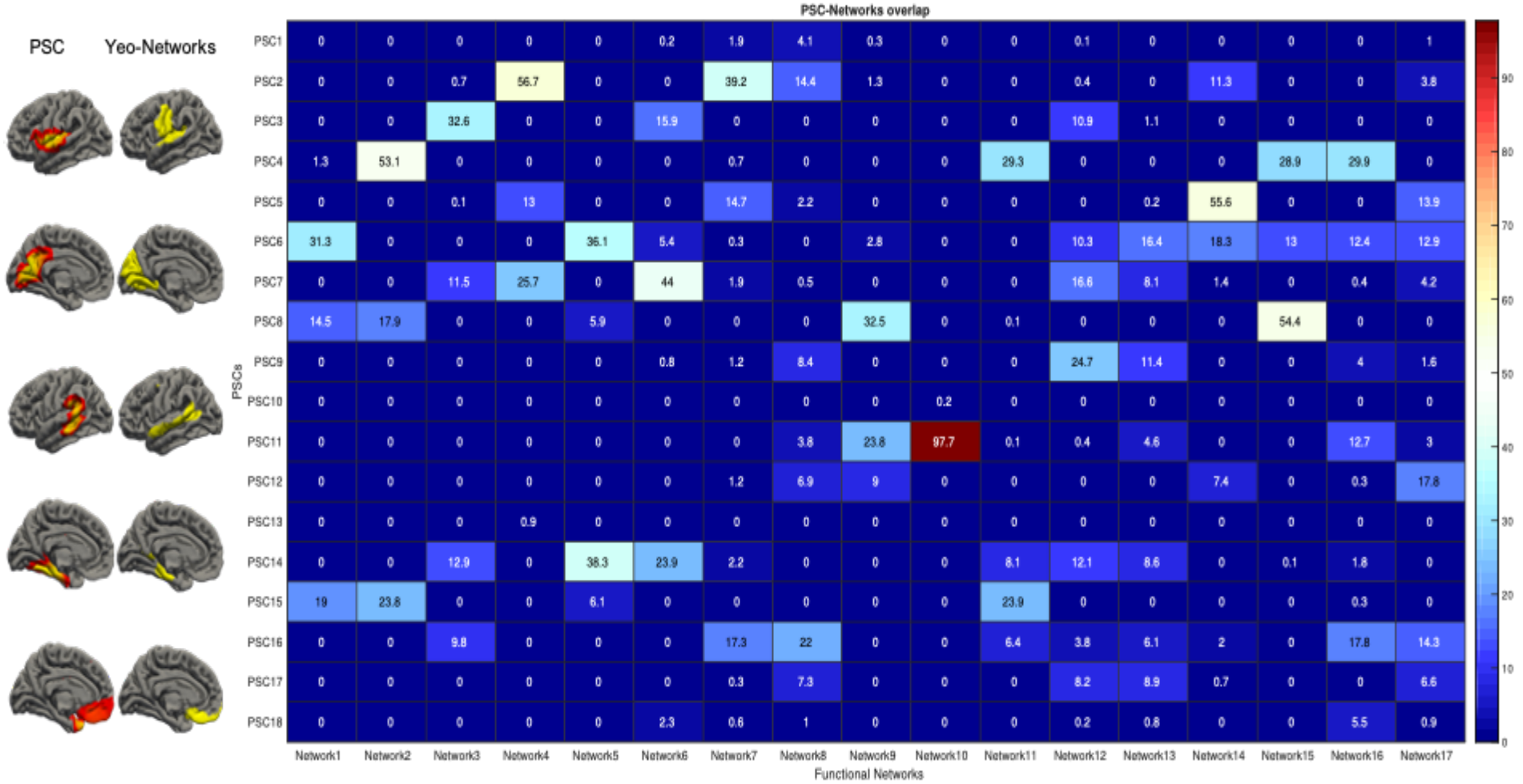
Supplementary Figure 2. Euler number distribution in the controls (HC, N=314) and patients (PAT, N=728) sample.



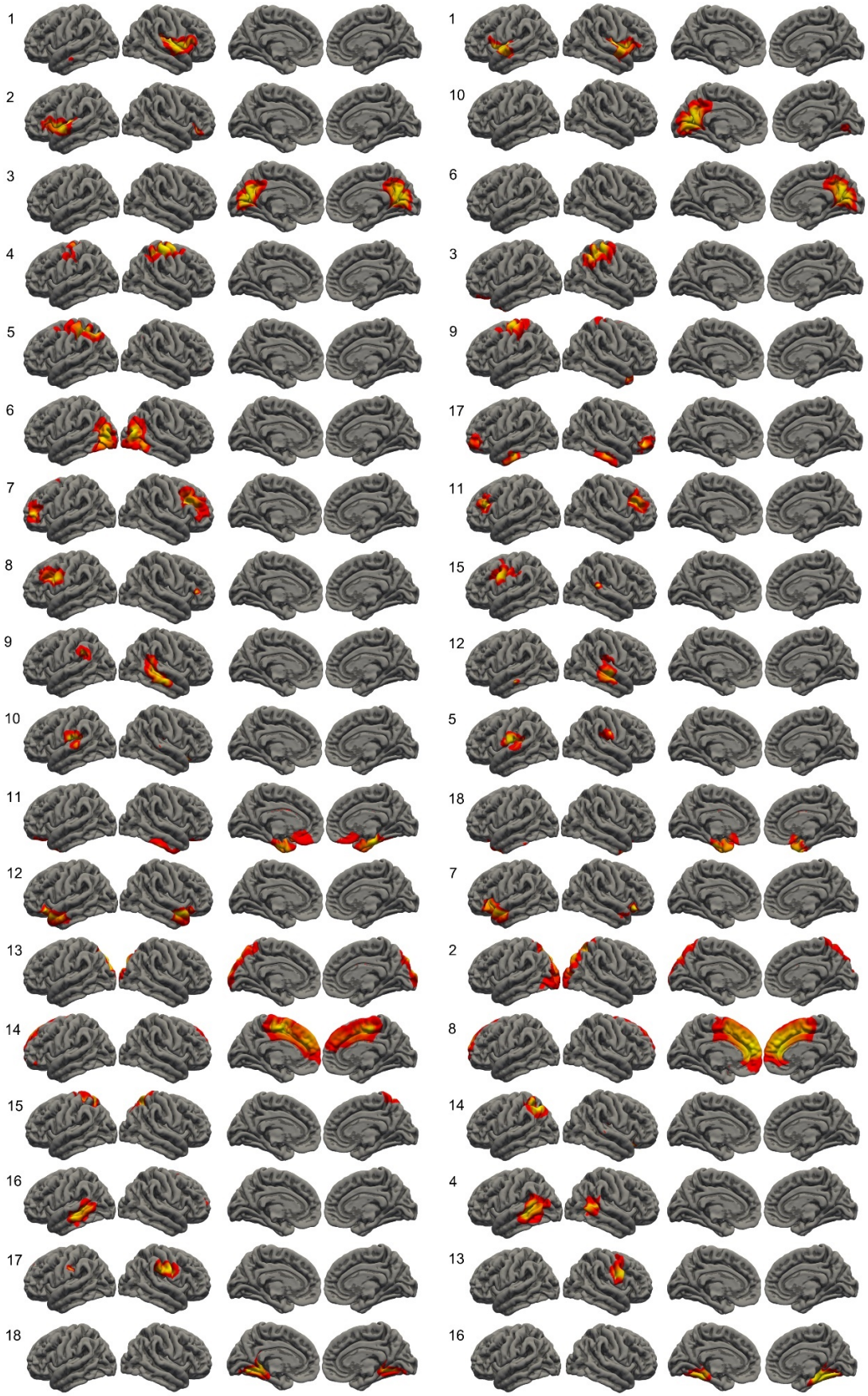
Supplementary Figure 3. Distribution of SIPS-P sum items in the patients' population. Abbreviations: SIPS-P: Structured Interview for the Prodromal Syndrom, sum of P-items; CHR: Clinical High Risk; ROD: Recent Onset Depression; ROP: Recent Onset Psychosis



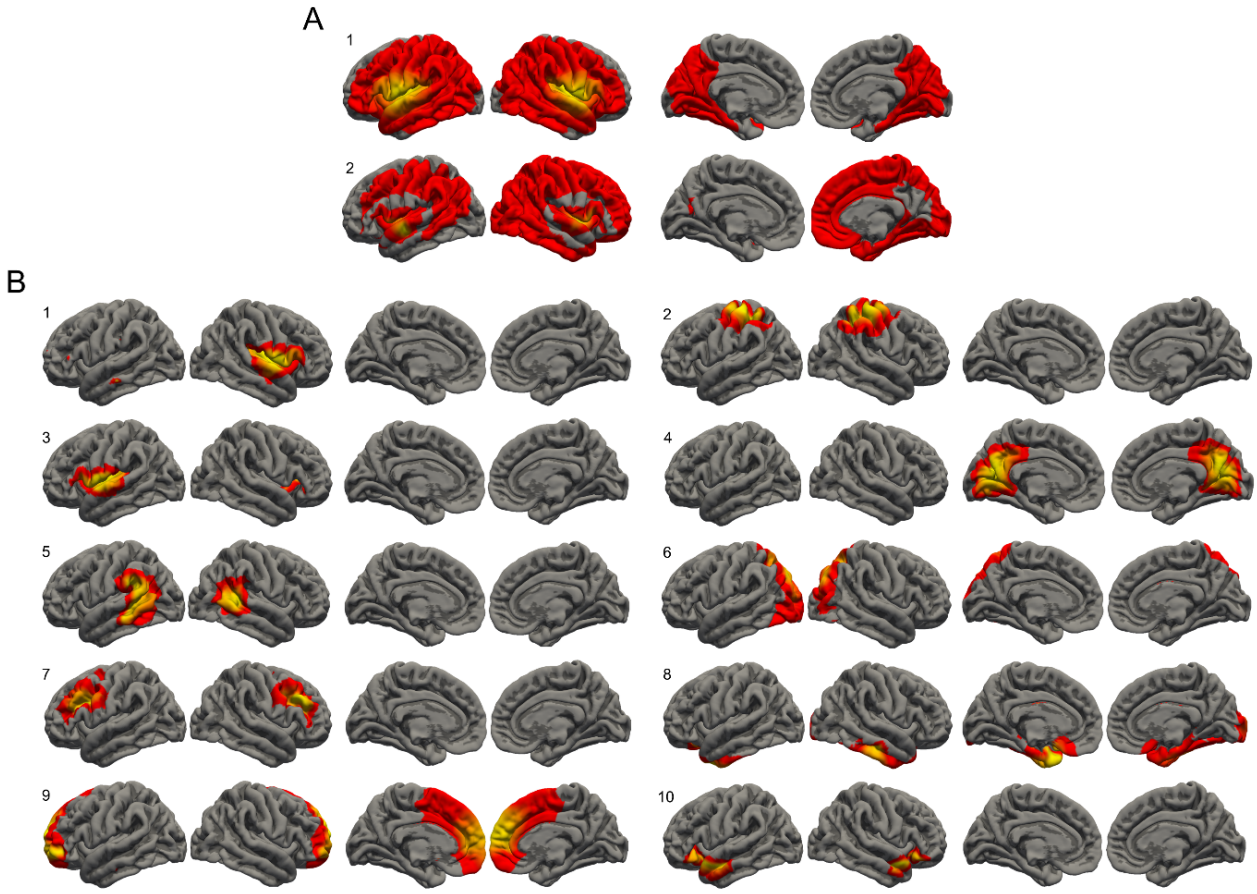
Supplementary Figure 4. Spatial overlap between the 18 PSCs and the 35 parcels from the structural Desikan-Killiany atlas. Displayed is the overlap between the 18 PSC (Y-axis) and the 35 parcels from the Desikan-Killiany structural atlas (X-axis) for the left hemisphere. The tables' values represent percentages of vertices overlap, where warmer colors mean higher overlap (see figure legend for percentages). Abbreviations: PSC: Patterns of Structural Covariance.



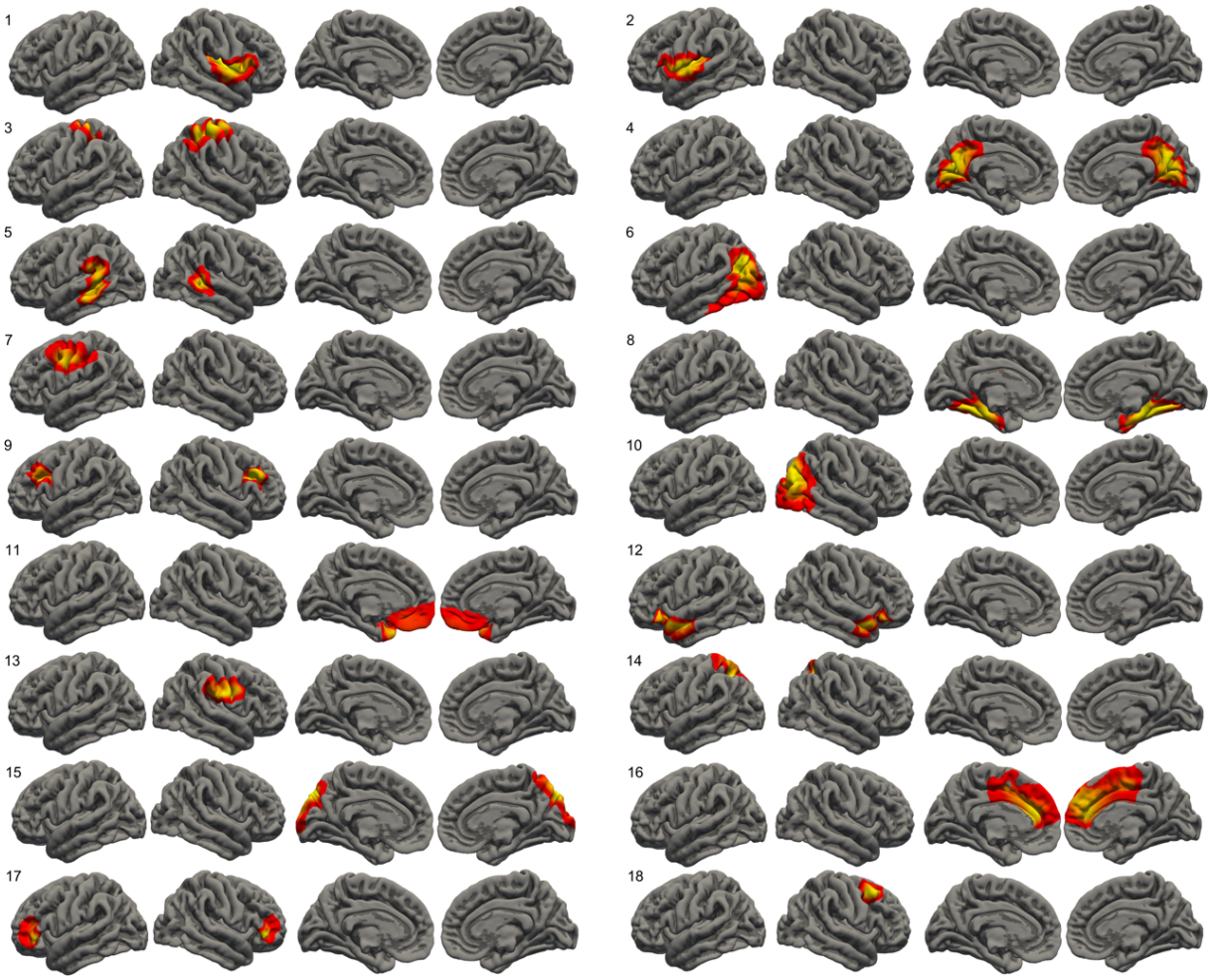
Supplementary Figure 5. Spatial overlap between the 18 PSCs and the 17 networks from the Yeo-functional connectivity atlas. Displayed is the overlap between the 18 PSC (Y-axis) and the 17 Yeo-functional networks atlas (X-axis) for the left hemisphere. The tables' values represent percentages of vertices overlap, where warmer colors mean higher overlap (see figure legend for percentages). Abbreviations: PSC: Patterns of Structural Covariance.



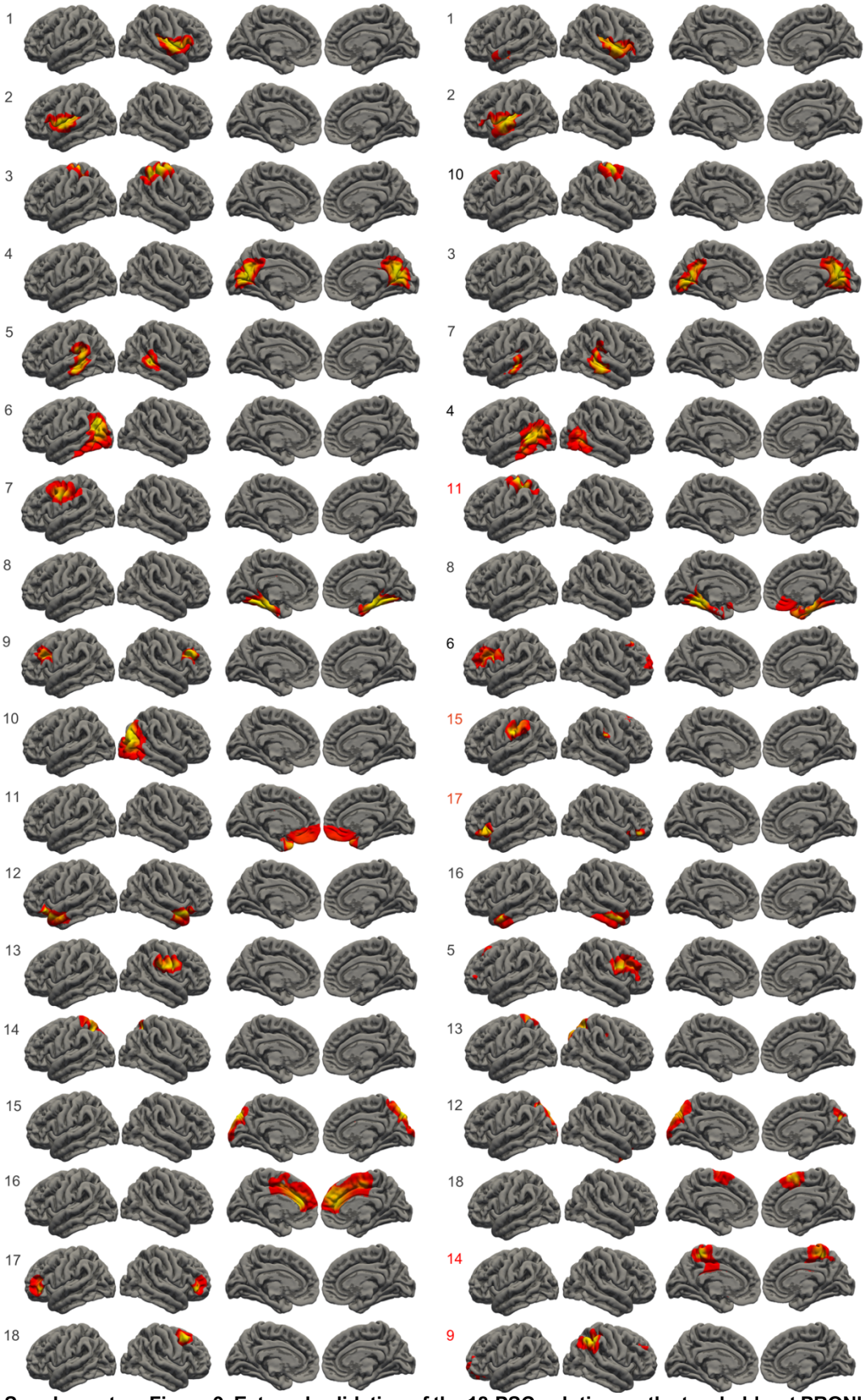
Supplementary Figure 6. 18 Patterns of Structural Covariance (PSC) in the randomly split HC sample. Split1 and 2 are depicted on the left and right side, respectively. PSCs are projected on a template brain (pial surface) and warmer colors represent higher covariance. Left side: lateral left and right view, right side: medial left and right view. The correspondent PSC number is indicated in the upper left corner.



Supplementary Figure 7. 2-PSC and 10-PSC solutions. A: 2-PSC solution, B: 10-PSC solution. PSCs are projected on a template brain (pial surface) and warmer colors represent higher covariance. Left side: lateral left and right view, right side: medial left and right view. The correspondent PSC number is indicated in the upper left corner.

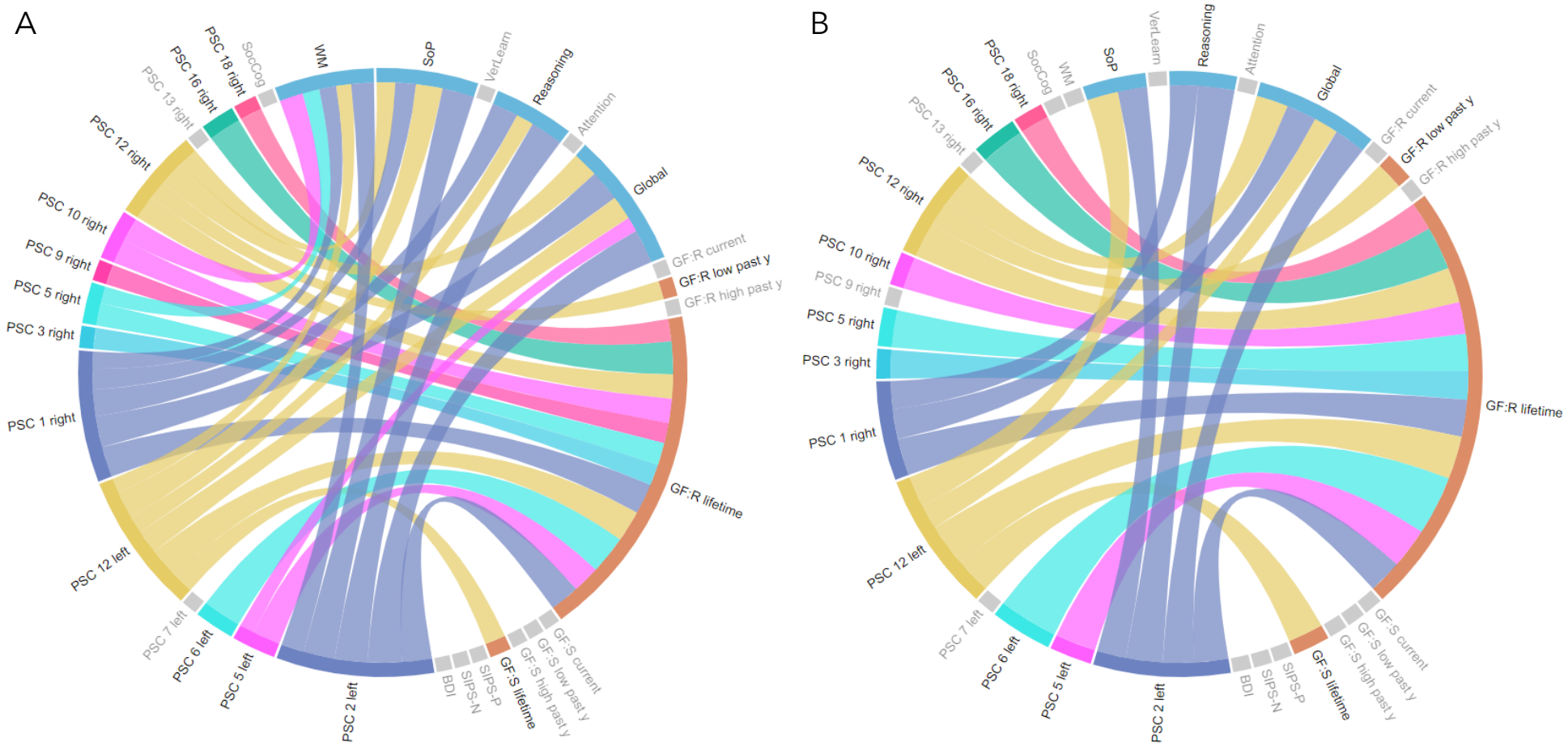


Supplementary Figure 8. 18 gyrification-based Patterns of Structural Covariance. PSCs are projected on a template brain (pial surface) and warmer colors represent higher gyrification structural covariance between individuals. Left side: lateral left and right view, right side: medial left and right view. The correspondent PSC number is indicated in the upper left corner.
PSC1 and 2: right and left superior temporal gyrus, pars orbitalis and triangularis and insula. PSC3: part of the superior parietal lobule and the intraparietal sulcus bilaterally. PSC4: bilateral posterior cingulate, precuneus, occipital visual areas. PSC5: bilateral angular gyrus and the medial temporal sulcus and gyrus, mainly on the left hemisphere. PSC6 and 10: asymmetrical occipital cortex, angular and inferior temporal gyrus. PSC7 and 13: asymmetrical area crossing horizontally the left and right inferior precentral gyrus, the inferior frontal sulcus and the inferior frontal gyrus. PSC8: bilateral medial temporal lobe, perirhinal and enthorinal cortex, parahippocampal and fusiform gyrus. PSC9: bilateral dorsolateral prefrontal cortex. PSC11: bilateral ventromedial prefrontal cortex. PSC12: bilateral medial and superior temporal gyrus. PSC14: mainly left frontal and parietal lobe, part of the primary somatosensory areas and the superior parietal lobule. PSC15: bilateral occipital pole and the parietoccipital sulcus. PSC16: bilateral medial prefrontal cortex, motor and premotor cortex, cingulate. PSC17: bilateral lateral frontopolar cortex. PSC18: right posterior part of the middle frontal gyrus.



Supplementary Figure 9. External validation of the 18-PSC solution on the two held-out PRONIA sites. The original pooled HC sample and the validation sample (Milan and Birmingham, N=84) are depicted on the left and right side, respectively. PSCs are projected on a template brain (pial surface) and warmer colors represent higher covariance. Left side: lateral left and right view, right side: medial left and right view. The correspondent PSC number is indicated in the upper left corner. Red numbers represent non-matching PSCs.

Supplementary Figure 10. Inner product distribution. The graph plots the maximum inner product values (X-axis) for each of the 18 PSCs (Y-axis) and represents the distribution of spatial overlap between the original Pattern of Structural Covariance (PSC) extracted from the discovery sample and the PSCs calculated in the external validation sample. The inner product ranges from 0 (low overlap) to 1 (perfect overlap).



Supplementary Figure 11. Comparison of neurocognitive and clinical results with and without correction for IQ. The circular plots represent associations between components and neurocognitive (in blue) and functioning (in orange) domains, with thicker lines reflecting higher effect sizes/lower p-values. Grey domains are those not significant in the analyses. **A:** associations resulting from linear models for effect of PSCs on cognition/functioning with age and sex (FDR correction $p<0.01$). **B:** associations resulting from linear models for effect of PSCs on cognition/functioning with age, sex and IQ (FDR correction $p<0.05$). For a visual display of significant components overlaid on a common cortical surface please refer to Figure 4. Same colors correspond to symmetrical PSCs. Abbreviations: PSC: Pattern of Structural Covariance; SocCog: social cognition; WM: working memory; SoP: speed of processing; VerLearn: verbal learning; Global: global cognition; GF:R/S: global functioning: Role/Social; lifetime: the highest functioning lifetime; low: lowest; high: highest; y: year; SIPS-P/N: Structured Interview for the Prodromal Syndrome, Positive/Negative symptoms; BDI: Beck Depression Inventory.

Supplementary Table 1. Sociodemographic information for the controls' and patients' sample, subdivided by site.

		Pooled sample	Munich	Basel	Cologne	Turku	Udine	X^2/F	p
N (%)	HC	308	63 (20.5)	59 (19.2)	70 (22.7)	47 (15.3)	69 (22.4)		
	PAT	713	357 (50.1)	68 (9.5)	139 (19.5)	90 (12.6)	59 (8.3)		
Mean age [y] (SD)	HC	25.7 (6.1)	26.6 (6.9)	25.3 (5.6)	24.6 (5.5)	26.7 (5.4)	25.5 (6.6)	1.27	0.28
	PAT	25.2 (5.9)	25.0 (6.0)	24.4 (5.2)	25.0 (5.5)	26.0 (5.7)	26.0 (6.7)	1.20	0.31
Sex, female (%)	HC	183 (59.4)	41 (22.4)	34 (18.6)	34 (18.6)	31 (16.9)	43 (23.5)	5.41	0.25
	PAT	326 (45.7)	153 (46.9)	26 (8.0)	67 (20.6)	50 (15.3)	30 (9.2)	7.19	0.13
Handedness mean (SD)[a]	HC	76.2 (44.5)	74.0 (47.8)	77.7 (41.4)	78.7 (39.3)	79.8 (36.4)	71.7 (53.4)	0.32	0.86
	PAT	68.8 (54.0)	68.4 (56.8)	57.8 (63.9)	74.3 (42.9)	66.7 (51.3)	73.8 (50.7)	1.06	0.38
Education years (SD)	HC	15.5 (3.3)	15.3 (3.8)	15.0 (3.1)	15.8 (3.3)	15.5 (2.3)	16.0 (3.7)	0.78	0.54
	PAT	14.0 (5.4)	14.2 (7.0)	12.4 (3.3)	15.0 (3.2)	13.5 (2.5)	13.9 (2.8)	2.96	0.02

Abbreviations: y=years; SD: standard deviation; HC: healthy controls; PAT: patients. Notes: [a]: Edinburgh Handedness Score.

Supplementary Table 2. MRI Acquisition parameters in the PRONIA consortium.

Site	Model	Field strength	Coil channels	Flip angle	TR (ms)	TE (ms)	Voxel size (mm)	FOV	Slice num.
Munich	Philips Ingenia	3T	32	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Milan	Philips Achieva Intera	1.5T	8	12	8.1	3.7	0.93 x 0.93 x 1.0	240 x 240	170
Basel	SIEMENS Verio	3T	12	8	2000	3.4	1.0 x 1.0 x 1.0	256 x 256	176
Cologne	Philips Achieva	3T	8	8	9.5	5.5	0.97 x 0.97 x 1.0	250 x 250	190
Birmingham	Philips Achieva	3T	32	8	8.4	3.8	1.0 x 1.0 x 1.0	288 x 288	175
Turku	Philips Ingenuity	3T	32	7	8.1	3.7	1.0 x 1.0 x 1.0	256 x 256	176
Udine	Philips Achieva	3T	8	12	8.1	3.7	0.93 x 0.93 x 1.0	240 x 240	170

Supplementary Table 3. Correlations between the 18 loading coefficients-PSC and the extracted gyrification-PSCs.

PSCs	gyrification-PSCs - left		gyrification-PSCs - right	
	r	p	r	p
PSC 1	0.818	2.52E-75	1.000	0
PSC 2	1.000	0	0.770	1.31E-61
PSC 3	0.916	1.61E-123	0.981	3.33E-221
PSC 4	0.949	4.45E-156	0.959	7.31E-170
PSC 5	0.961	3.19E-173	0.890	3.21E-106
PSC 6	0.990	2.08E-264	0.728	3.63E-52
PSC 7	0.996	0	0.837	4.04E-82
PSC 8	0.929	2.47E-134	0.922	1.72E-128
PSC 9	0.935	9.29E-140	0.894	5.31E-109
PSC 10	0.218	1.12E-04	0.993	4.72E-283
PSC 11	0.913	2.58E-121	0.947	2.29E-152
PSC 12	0.950	1.55E-156	0.885	1.86E-103
PSC 13	0.700	1.18E-46	0.989	1.26E-257
PSC 14	0.983	1.67E-225	0.802	2.25E-70
PSC 15	0.905	9.99E-116	0.918	3.06E-125
PSC 16	0.938	9.30E-143	0.982	9.91E-223
PSC 17	0.942	1.69E-147	0.867	1.57E-94
PSC 18	0.728	5.20E-52	0.993	1.34E-286

All p-values are FDR ($p<0.01$) corrected for multiple comparisons. Abbreviations: PSC, Pattern of Structural Covariance

Supplementary Table 4. Effects of age, sex and site on loading coefficients-PSC in HC.

PSCs	Age		Sex		Site	
	r	p	t	p	F	p
PSC 1	-0.342	7.00E-10	6.381	6.52E-10	0.556	0.695
PSC 2	-0.363	5.17E-11	7.042	1.25E-11	0.706	0.588
PSC 3	-0.561	5.73E-27	4.530	8.45E-06	1.033	0.390
PSC 4	-0.227	5.93E-05	6.117	2.91E-09	0.412	0.800
PSC 5	-0.436	1.08E-15	4.648	4.99E-06	0.412	0.800
PSC 6	-0.463	9.51E-18	3.831	1.55E-04	0.547	0.702
PSC 7	-0.550	8.48E-26	4.188	3.69E-05	0.815	0.516
PSC 8	-0.270	1.57E-06	7.282	2.79E-12	0.771	0.545
PSC 9	-0.476	7.45E-19	4.344	1.90E-05	0.628	0.643
PSC 10	-0.456	3.23E-17	5.070	6.91E-07	0.983	0.417
PSC 11	-0.466	5.49E-18	2.417	0.016	0.355	0.841
PSC 12	-0.371	1.79E-11	5.224	3.24E-07	0.538	0.708
PSC 13	-0.434	1.51E-15	3.262	0.001	0.764	0.549
PSC 14	-0.459	1.80E-17	4.406	1.46E-05	0.692	0.598
PSC 15	-0.374	1.10E-11	4.637	5.25E-06	0.577	0.679
PSC 16	-0.461	1.22E-17	4.197	3.55E-05	0.926	0.449
PSC 17	-0.481	2.87E-19	3.891	1.23E-04	0.958	0.431
PSC 18	-0.511	6.55E-22	4.743	3.24E-06	1.502	0.202

Abbreviations: PSC, Pattern of Structural Covariance

Supplementary Table 5. Effects of age, sex and site on gyrification-PSCs in HC.

PSCs	Age		Sex		Site	
	r	p	t	p	F	p
PSC 2 L	-0.356	1.21E-10	7.115	8.00E-12	0.722	0.577
PSC 3 L	-0.514	3.91E-22	4.244	2.92E-05	0.746	0.561
PSC 4 L	-0.220	1.00E-04	5.953	7.19E-09	0.342	0.849
PSC 5 L	-0.381	4.32E-12	4.176	3.88E-05	0.297	0.880
PSC 6 L	-0.428	3.66E-15	3.549	4.47E-04	0.552	0.698
PSC 7 L	-0.541	7.62E-25	4.126	4.76E-05	0.836	0.503
PSC 8 L	-0.206	2.72E-04	6.985	1.78E-11	0.458	0.766
PSC 9 L	-0.436	9.49E-16	4.019	7.37E-05	0.632	0.640
PSC 11 L	-0.384	3.06E-12	2.017	0.045	0.343	0.849
PSC 12 L	-0.329	3.48E-09	4.847	1.99E-06	0.554	0.697
PSC 14 L	-0.421	1.18E-14	3.753	2.09E-04	0.651	0.627
PSC 15 L	-0.281	5.48E-07	3.903	1.17E-04	0.522	0.720
PSC 16 L	-0.355	1.46E-10	2.962	0.003	0.927	0.448
PSC 17 L	-0.415	3.15E-14	3.111	2.04E-03	0.846	0.497
PSC 1 R	-0.333	2.13E-09	6.420	5.18E-10	0.559	0.693
PSC 3 R	-0.545	3.28E-25	4.411	1.43E-05	1.058	0.377
PSC 4 R	-0.182	0.001	5.417	1.23E-07	0.461	0.764
PSC 5 R	-0.403	1.92E-13	4.190	3.66E-05	0.597	0.665
PSC 8 R	-0.211	1.89E-04	6.825	4.73E-11	0.859	0.489
PSC 9 R	-0.390	1.33E-12	3.325	9.93E-04	0.356	0.840
PSC 10 R	-0.431	2.32E-15	4.973	1.10E-06	0.974	0.422
PSC 11 R	-0.424	6.72E-15	1.936	**0.054**	0.140	0.967
PSC 12 R	-0.295	1.35E-07	4.418	1.38E-05	0.362	0.836
PSC 13 R	-0.395	5.91E-13	2.553	0.011	0.686	0.602
PSC 14 R	-0.350	2.74E-10	3.935	1.03E-04	0.404	0.806
PSC 15 R	-0.315	1.59E-08	4.229	3.10E-05	0.492	0.741
PSC 16 R	-0.472	1.64E-18	4.305	2.25E-05	0.805	0.523
PSC 17 R	-0.393	7.83E-13	3.378	8.25E-04	0.824	0.511
PSC 18 R	-0.481	3.18E-19	4.595	6.32E-06	1.459	0.215

Bold values do not survive multiple comparisons' correction. Abbreviations: PSC, Pattern of Structural Covariance; L, left hemisphere; R, right hemisphere.

Supplementary Table 6. Effects of age, sex and site on gyrification-PSCs in PAT.

PSCs	Age		Sex		Site	
	r	p	t	p	F	p
PSC 2 L	-0.275	8.41E-14	9.152	5.81E-19	0.631	**0.640**
PSC 3 L	-0.447	3.08E-36	5.635	2.52E-08	2.454	**0.045**
PSC 4 L	-0.139	1.89E-04	8.896	4.74E-18	2.680	**0.031**
PSC 5 L	-0.287	5.45E-15	5.596	3.13E-08	3.295	**0.011**
PSC 6 L	-0.324	7.36E-19	5.912	5.23E-09	4.508	0.001
PSC 7 L	-0.442	1.85E-35	4.878	1.33E-06	1.127	**0.342**
PSC 8 L	-0.180	1.32E-06	10.701	7.03E-25	2.497	**0.042**
PSC 9 L	-0.379	8.59E-26	2.869	0.004	0.237	**0.917**
PSC 11 L	-0.334	4.39E-20	3.558	3.99E-04	2.345	**0.053**
PSC 12 L	-0.234	2.41E-10	8.919	3.93E-18	3.253	**0.012**
PSC 14 L	-0.366	5.00E-24	3.256	0.001	1.533	**0.191**
PSC 15 L	-0.184	7.46E-07	2.098	0.036	1.195	**0.312**
PSC 16 L	-0.324	6.19E-19	4.853	1.49E-06	2.360	**0.052**
PSC 17 L	-0.354	1.71E-22	4.314	1.83E-05	5.083	0.000
PSC 1 R	-0.286	6.17E-15	8.532	8.61E-17	1.049	**0.381**
PSC 3 R	-0.477	8.20E-42	5.807	9.57E-09	0.686	**0.602**
PSC 4 R	-0.161	1.50E-05	7.480	2.20E-13	2.645	**0.033**
PSC 5 R	-0.245	3.08E-11	5.362	1.11E-07	2.929	**0.020**
PSC 8 R	-0.185	6.75E-07	9.667	7.48E-21	2.922	**0.020**
PSC 9 R	-0.371	1.14E-24	2.780	0.006	0.716	**0.581**
PSC 10 R	-0.329	2.00E-19	5.735	1.44E-08	1.182	**0.317**
PSC 11 R	-0.332	8.12E-20	3.991	7.25E-05	1.819	**0.123**
PSC 12 R	-0.226	1.06E-09	8.929	3.63E-18	2.860	**0.023**
PSC 13 R	-0.330	1.49E-19	2.852	0.004	2.004	**0.092**
PSC 14 R	-0.347	1.32E-21	2.373	0.018	0.926	**0.448**
PSC 15 R	-0.194	1.80E-07	2.140	0.033	2.045	**0.086**
PSC 16 R	-0.395	4.36E-28	5.586	3.32E-08	1.754	**0.136**
PSC 17 R	-0.296	7.12E-16	3.070	0.002	4.079	0.003
PSC 18 R	-0.451	5.11E-37	5.172	3.02E-07	1.192	**0.313**

Bold values do not survive multiple comparisons' correction. Abbreviations: PSC, Pattern of Structural Covariance; L, left hemisphere; R, right hemisphere.

Supplementary Table 7. Correlations between the PSCs, Euler number and age in controls and patients.

PSCs	Euler number – HC		Euler number – PAT	
	r	p	r	p
PSC 2 L	-0.3489	**3.03E-10**	-0.2642	**7.49E-13**
PSC 3 L	-0.4134	**3.83E-14**	-0.3405	**8.24E-21**
PSC 4 L	-0.3058	**4.33E-08**	-0.1772	**1.92E-06**
PSC 5 L	-0.3349	**1.66E-09**	-0.3305	**1.24E-19**
PSC 6 L	-0.4251	**6.03E-15**	-0.3811	**4.67E-26**
PSC 7 L	-0.4366	**9.18E-16**	-0.3669	**3.82E-24**
PSC 8 L	-0.3428	**6.41E-10**	-0.2785	**3.64E-14**
PSC 9 L	-0.3640	**4.43E-11**	-0.2732	**1.13E-13**
PSC 11 L	-0.3377	**1.19E-09**	-0.2812	**2.00E-14**
PSC 12 L	-0.3129	**2.01E-08**	-0.2674	**3.88E-13**
PSC 14 L	-0.3704	**1.89E-11**	-0.2880	**4.36E-15**
PSC 15 L	-0.3649	**3.92E-11**	-0.2183	**3.86E-09**
PSC 16 L	-0.3126	**2.08E-08**	-0.2840	**1.07E-14**
PSC 17 L	-0.2983	**9.47E-08**	-0.2485	**1.72E-11**
PSC 1 R	-0.3003	**7.73E-08**	-0.2835	**1.21E-14**
PSC 3 R	-0.3589	**8.58E-11**	-0.3366	**2.39E-20**
PSC 4 R	-0.2085	**2.29E-04**	-0.1779	**1.75E-06**
PSC 5 R	-0.3581	**9.48E-11**	-0.2685	**3.06E-13**
PSC 8 R	-0.2833	**4.28E-07**	-0.2922	**1.69E-15**
PSC 9 R	-0.2401	**2.06E-05**	-0.2802	**2.51E-14**
PSC 10 R	-0.3454	**4.67E-10**	-0.2810	**2.09E-14**
PSC 11 R	-0.3201	**9.14E-09**	-0.2919	**1.80E-15**
PSC 12 R	-0.2422	**1.73E-05**	-0.3041	**1.02E-16**
PSC 13 R	-0.3465	**4.09E-10**	-0.2633	**9.09E-13**
PSC 14 R	-0.2558	**5.43E-06**	-0.2139	**8.03E-09**
PSC 15 R	-0.2738	**1.07E-06**	-0.1826	**9.21E-07**
PSC 16 R	-0.3242	**5.74E-09**	-0.3243	**6.36E-19**
PSC 17 R	-0.2823	**4.74E-07**	-0.2633	**9.08E-13**
PSC 18 R	-0.3469	**3.87E-10**	-0.3648	**7.20E-24**
age	0.3227	**6.78E-09**	0.1654	**9.01E-06**

Bold values represent significant correlations (after FDR correction p<0.01).
Abbreviations: PSC: Patterns of Structural Covariance; HC: healthy controls; PAT: patients.

Supplementary Table 8. Information on neuropsychological tests used in the PRONIA study.

Test name	Cognitive domain	Administration
Rey-Osterrieth Complex Figure (ROCF)	Visual-spatial construction; visual-spatial memory (both short- and long-term)	Paper-pencil format with tablet support
Diagnostic Analysis of Non-Verbal Accuracy (DANVA-2-AF) [a]	Social cognition	Tablet-based
Auditory Digit Span, Forward & Backward trials (ADS-F&B) [a]	Verbal short-term memory and verbal working memory	Auditory presentation of numbers by recorded (male) voice
Verbal Fluency, Phonemic & Semantic trials (VF-P/S) [a]	Verbal fluency in a phonemic ('S'-words) and a semantic ('Animals') condition	Named words were recorded and written down by examiner
Rey Auditory Verbal Learning Test (RAVLT) [a, b]	Short- and long-term verbal memory	Auditory presentation of word list by recorded (male) voice
Trail Making Task, A and B trials (TMT-A/B) [a]	Processing speed, sequencing, graphical-motor capacity, visual attention and search ability, flexibility	Paper- pencil format
Continuous Performance Test, Identical Pairs version (CPT-IP) [a]	Selective and sustained visual attention	Tablet-based
Self-Ordered Pointing Test (SOPT)	Short-term visual-spatial memory and working memory	Tablet-based
Digit Symbol Substitution Test (DSST) [a]	Sustained attention, working memory and processing speed	Paper- pencil format
Salience Attribution Task (SAT-SV)	Explicit and implicit, adaptive and aberrant salience	Tablet-based
Wechsler Adult Intelligence Scale (WAIS-III) [a]		
Vocabulary	Premorbid verbal intelligence	Paper-pencil format
Matrices	Visual processing and abstract reasoning	Paper-pencil format

Tests are presented in order of administration. Notes: [a] test used to construct the cognitive scores, [b] test substituted by the Hopkins Verbal Learning Test-Revised for Turku.

Supplementary Table 9. Linear models for effect of group (controls vs. patients) with PSC, age and sex.

PSCs	Est.	SE	t	r - group	p	Adj. R^2	F	model's p
PSC 2 L	-0.100	0.022	-4.486	-0.139	8.10E-06	0.198	85.198	3.52E-49
PSC 5 L	-0.055	0.015	-3.688	-0.115	2.38E-04	0.142	57.129	3.93E-34
PSC 6 L	-0.031	0.009	-3.564	-0.111	3.83E-04	0.167	69.011	1.26E-40
PSC 7 L	-0.030	0.010	-2.983	-0.093	0.003	0.255	117.320	3.02E-65
PSC 12 L	-0.070	0.019	-3.759	-0.117	1.80E-04	0.155	63.150	1.91E-37
PSC 1 R	-0.100	0.021	-4.729	-0.147	2.58E-06	0.186	78.828	7.50E-46
PSC 3 R	-0.038	0.010	-3.784	-0.118	1.63E-04	0.289	139.132	1.57E-75
PSC 5 R	-0.045	0.015	-3.053	-0.095	0.002	0.123	48.696	2.07E-29
PSC 9 R	-0.053	0.015	-3.474	-0.108	5.33E-04	0.156	63.917	7.29E-38
PSC 10 R	-0.026	0.009	-3.080	-0.096	0.002	0.174	72.723	1.29E-42
PSC 12 R	-0.086	0.017	-5.075	-0.157	4.60E-07	0.149	60.429	5.93E-36
PSC 13 R	-0.048	0.014	-3.305	-0.103	9.82E-04	0.135	54.107	1.89E-32
PSC 16 R	-0.020	0.006	-3.325	-0.104	9.14E-04	0.215	93.887	1.22E-53
PSC 18 R	-0.038	0.011	-3.574	-0.111	3.68E-04	0.248	113.200	3.00E-63

Displayed are the 14 significant PSCs. Abbreviations: PSC: Patterns of Structural Covariance, L: left, R: right Est.: estimate, SE: standard error, t: t-stat, r - group: partial correlation coefficient for the group effect; p: p-value for the specific contribution of PSC in the model, Adj. R^2: model's adjusted R^2, F: F-test, model's p: p-value for the whole linear model with age, sex and PSC.

Supplementary Table 10. P-values for linear models for effect of group (i.e., HC vs PAT) with age, sex and scan quality measures.

PSC	IQR	CNR	SNR
PSC 2 L	**5.16E-06**	**4.81E-04**	**6.31E-05**
PSC 5 L	**1.81E-04**	0.004	**0.002**
PSC 6 L	**2.61E-04**	0.004	**0.003**
PSC 7 L	**0.002**	0.086	0.025
PSC 12 L	**1.45E-04**	**2.06E-04**	**1.54E-04**
PSC 1 R	**2.02E-06**	**9.29E-06**	**8.32E-07**
PSC 3 R	**1.25E-04**	0.005	**8.73E-04**
PSC 5 R	**0.002**	0.004	**0.002**
PSC 9 R	**4.46E-04**	0.018	**0.001**
PSC 10 R	**0.001**	0.012	**0.002**
PSC 12 R	**3.93E-07**	**1.00E-05**	**1.12E-06**
PSC 13 R	**8.34E-04**	0.014	**0.002**
PSC 16 R	**6.70E-04**	0.008	**0.003**
PSC 18 R	**2.24E-04**	0.007	**0.001**

Displayed are the significant PSCs for linear models for effect of group (HC vs patients). Bold values represent significant p-values in the models augmented with each scan quality measure (FDR-correction $p<0.01$). Abbreviations: PSC: Patterns of Structural Covariance; L: left, R: right; IQR: weighted average Image Quality Rating; CNR: Contrast to Noise Ratio; SNR: Signal to Noise Ratio.

Supplementary Table 11. Partial correlations for effect of group (controls vs. patients) with PSC, age and sex in the external validation sample.

PSC	r	p
PSC2 L	-0.185	**0.005**
PSC3 L	-0.160	0.015
PSC4 L	-0.017	0.802
PSC5 L	-0.227	**0.001**
PSC6 L	-0.212	**0.001**
PSC7 L	-0.230	**<0.001**
PSC8 L	-0.119	0.073
PSC9 L	-0.151	0.022
PSC11 L	-0.294	<0.001
PSC12 L	-0.204	**0.002**
PSC14 L	-0.200	0.002
PSC15 L	-0.154	0.020
PSC16 L	-0.211	0.001
PSC17 L	-0.135	0.041
PSC1 R	-0.211	**0.001**
PSC3 R	-0.202	**0.002**
PSC4 R	-0.100	0.132
PSC5 R	-0.151	0.022
PSC8 R	-0.160	0.015
PSC9 R	-0.200	**0.002**
PSC10 R	-0.209	**0.001**
PSC11 R	-0.188	0.004
PSC12 R	-0.251	**<0.001**
PSC13 R	-0.218	**0.001**
PSC14 R	-0.154	0.020
PSC15 R	-0.166	0.012
PSC16 R	-0.229	**<0.001**
PSC17 R	-0.169	0.010
PSC18 R	-0.166	0.012

Displayed are the partial correlations (correction for age and sex) for effect of group (controls vs. patients) on PSC components in the external validation sample. Bold values represent significant p-values for group effect in the models, which overlap with original results in the discovery sample, while blue values indicate terms, which do not overlap with the previous analyses (FDR-correction p<0.01). Abbreviations: r: correlation coefficient; p: p-value; PSC: Patterns of Structural Covariance, L: left, R: right.

Supplementary Table 12. Linear models to predict cognitive domains with PSCs, age and sex.

Working Memory								
PSCs	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 2 L	0.445	0.116	3.837	0.148	1.37E-04	0.027	7.070	1.11E-04
PSC 12 L	0.422	0.141	2.996	0.116	0.003	0.019	5.137	0.002
PSC 1 R	0.403	0.124	3.250	0.126	0.001	0.021	5.670	7.75E-04
PSC 5 R	0.557	0.173	3.217	0.125	0.001	0.021	5.599	8.55E-04
PSC 10 R	1.176	0.299	3.930	0.152	9.41E-05	0.028	7.313	7.92E-05
Speed of Processing								
PSCs	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 2 L	0.422	0.091	4.655	0.182	3.95E-06	0.045	10.983	4.89E-07
PSC 12 L	0.392	0.139	2.826	0.166	0.005	0.025	6.343	0.000
PSC 1 R	0.342	0.247	1.385	0.143	0.167	0.015	4.285	0.005
PSC 12 R	0.226	0.211	1.072	0.135	0.284	0.014	4.023	0.008
Reasoning								
PSCs	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 2 L	0.5544	0.1170	4.7394	0.186	2.66E-06	0.0383	9.3571	4.67E-06
PSC 12 L	0.4774	0.1447	3.3000	0.131	0.0010	0.0209	5.4665	0.0010
PSC 1 R	0.5296	0.1259	4.2070	0.166	2.97E-05	0.0312	7.7557	4.31E-05
Global Cognition								
PSCs	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 2 L	2.5224	0.5176	4.8729	0.197	1.41E-06	0.0386	8.9479	8.36E-06
PSC 5 L	2.3782	0.7941	2.9947	0.122	0.0029	0.0149	3.9973	0.0078
PSC 12 L	2.5234	0.6367	3.9634	0.161	8.29E-05	0.0259	6.2553	0.0003
PSC 1 R	2.5730	0.5590	4.6030	0.186	5.10E-06	0.0346	8.0909	2.74E-05
PSC 12 R	2.9770	0.7165	4.1552	0.168	3.73E-05	0.0283	6.7770	0.0002

Displayed are the significant PSCs for each linear model with cognitive domains. Abbreviations: PSC: Patterns of Structural Covariance, L: left, R: right Est.: estimate, SE: standard error, t: t-stat, r-PSC: partial correlation coefficient for the effect of PSC; p: p-value for the specific contribution of PSC in the model, Adj. R^2: model's adjusted R^2, F: F-test. model's p: p-value for the whole linear model with age, sex and PSC.

Supplementary Table 13. P-values for linear models for effect of PSC on cognitive domains with age, sex and quality measures.

PSC	IQR						
	SocCog	WM	SoP	VerLearn	Reas	Attention	Global
PSC 2 L	0.146	**1.1E-04**	**5.0E-06**	0.017	**2.1E-06**	0.005	**1.1E-06**
PSC 5 L	0.830	0.006	0.006	0.019	0.062	0.004	**0.002**
PSC 12 L	0.439	**0.002**	**3.4E-05**	0.400	**8.4E-04**	0.007	**6.5E-05**
PSC 1 R	0.092	**0.001**	**3.1E-04**	0.054	**3.3E-05**	0.002	**5.6E-06**
PSC 5 R	0.513	**0.001**	0.035	0.048	0.022	0.089	0.008
PSC 10 R	0.736	**7.3E-05**	0.134	0.020	0.079	0.318	0.023
PSC 12 R	0.531	0.011	**1.0E-03**	0.074	0.013	0.001	**2.1E-05**
PSC	CNR						
	SocCog	WM	SoP	VerLearn	Reas	Attention	Global
PSC 2 L	0.217	**2.3E-04**	**3.0E-06**	0.013	**2.2E-06**	0.005	**1.7E-06**
PSC 5 L	0.959	0.013	0.004	0.015	0.075	0.004	**0.003**
PSC 12 L	0.542	***0.004***	**3.1E-05**	0.429	**0.002**	0.016	**2.2E-04**
PSC 1 R	0.125	**0.002**	**2.0E-04**	0.042	**3.6E-05**	0.001	**9.1E-06**
PSC 5 R	0.468	**0.002**	0.029	0.023	0.027	0.093	0.011
PSC 10 R	0.637	**1.6E-04**	0.092	0.013	0.076	0.339	0.027
PSC 12 R	0.706	0.019	**7.5E-04**	0.079	0.037	0.002	**1.1E-04**
PSC	SNR						
	SocCog	WM	SoP	VerLearn	Reas	Attention	Global
PSC 2 L	0.241	**1.4E-04**	**2.5E-06**	0.010	**1.7E-06**	0.003	**9.8E-07**
PSC 5 L	0.960	0.009	0.004	0.012	0.058	0.003	**0.002**
PSC 12 L	0.543	**0.003**	**3.0E-05**	0.385	**0.001**	0.012	**1.2E-04**
PSC 1 R	0.144	**0.001**	**1.6E-04**	0.032	**3.3E-05**	***9.7E-04***	**5.7E-06**
PSC 5 R	0.461	**0.001**	0.027	0.019	0.020	0.075	0.007
PSC 10 R	0.662	**9.4E-05**	0.084	0.011	0.067	0.293	0.021
PSC 12 R	0.743	0.010	**6.8E-04**	0.056	0.026	***0.001***	**4.6E-05**

Displayed are the significant PSCs for each linear model with cognitive domains. Bold values represent significant p-values in the models augmented with quality measures, which overlap with original results. Grey italic values indicate terms, which do not overlap with the previous analyses without quality measures. Blue italic values indicate originally significant terms, which are no longer significant (FDR-correction p<0.01). Abbreviations: PSC: Patterns of Structural Covariance, L: left, R: right; IQR: weighted average Image Quality Rating; CNR: Contrast to Noise Ratio; SNR: Signal to Noise Ratio; SocCog: social cognition, WM: working memory, SoP: speed of processing, VerLearn: verbal learning, Reas: reasoning; Global: Global Cognition.

Supplementary Table 14. P-values for linear models for effect of PSC on cognitive domains with age, sex and education.

PSCs	SocCog	WM	SoP	VerLearn	Reas	Attention	Global
PSC 2 L	0.105	**5.50E-05**	**3.80E-06**	0.018	**3.19E-06**	0.003	**4.46E-07**
PSC 5 L	0.705	0.007	0.004	0.030	0.049	0.003	**0.002**
PSC 12 L	0.280	**8.04E-04**	**2.45E-05**	0.265	**6.51E-04**	0.006	**1.71E-05**
PSC 1 R	0.098	**0.003**	**4.72E-04**	0.082	**8.77E-05**	0.002	**1.22E-05**
PSC 5 R	0.655	**9.01E-04**	0.025	0.076	0.023	0.079	0.009
PSC 10 R	0.614	**2.30E-04**	0.072	0.028	0.064	0.346	0.028
PSC 12 R	0.469	***0.003***	**5.74E-04**	0.019	0.014	***6.11E-04***	**4.76E-06**

Displayed are the significant PSCs for each linear model with cognitive domains. Bold values represent significant p-values in the models augmented with education, which overlap with original results, while grey italic values indicate significant terms not detected in the previous analyses without education (FDR-correction p<0.01). Abbreviations: PSC: Patterns of Structural Covariance, L: left, R: right; SocCog: social cognition, WM: working memory, SoP: speed of processing, VerLearn: verbal learning, Reas: reasoning; Global: Global Cognition.

Supplementary Table 15. P-values for linear models for effect of PSC on cognitive domains with age, sex and IQ.

PSCs	SocCog	WM	SoP	VerLearn	Reas	Attention	Global
PSC 2 L	0.464	*0.051*	**0.004**	0.487	**8.40E-04**	0.184	**0.002**
PSC 5 L	0.804	0.136	0.052	0.109	0.339	0.047	*0.030*
PSC 12 L	0.806	*0.091*	**0.002**	0.594	*0.040*	0.107	**0.009**
PSC 1 R	0.327	*0.144*	*0.054*	0.614	**0.005**	0.058	**0.003**
PSC 5 R	0.511	*0.026*	0.324	0.200	0.142	0.436	0.103
PSC 10 R	0.966	*0.006*	0.932	0.109	0.337	0.986	0.182
PSC 12 R	0.714	0.132	*0.012*	0.268	0.111	0.022	**0,002**

Displayed are the significant PSCs for each linear model with cognitive domains. Bold values represent significant p-values in the models augmented with IQ, which overlap with original results when applying FDR-correction p<0.05. Blue italic values indicate originally significant terms, which are no longer significant. Abbreviations: PSC: Patterns of Structural Covariance, L: left, R: right; SocCog: social cognition, WM: working memory, SoP: speed of processing, VerLearn: verbal learning, Reas: reasoning; Global: Global Cognition.

Supplementary Table 16. Linear models to predict Global Functioning (GF) domains with PSCs, age and sex.

GF:R lowest past year								
PSC	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 12 R	0.9151	0.2678	3.4167	0.130	0.0007	0.0240	6.5570	0.0002
GF:S highest lifetime								
PSC	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 12 L	0.4478	0.1233	3.6309	0.138	0.0003	0.0473	12.2313	8.42E-08
GF:R highest lifetime								
PSCs	Est.	SE	t	r - PSC	p	Adj. R^2	F	p
PSC 2 L	0.446	0.100	4.456	0.169	9.78E-06	0.063	16.260	3.25E-10
PSC 5 L	0.650	0.152	4.279	0.162	2.15E-05	0.061	15.723	6.80E-10
PSC 6 L	1.361	0.263	5.168	0.195	3.12E-07	0.072	18.640	1.26E-11
PSC 12 L	0.562	0.122	4.626	0.175	4.47E-06	0.065	16.795	1.56E-10
PSC 1 R	0.483	0.106	4.546	0.172	6.47E-06	0.064	16.542	2.21E-10
PSC 3 R	0.828	0.226	3.667	0.140	2.65E-04	0.054	14.035	6.94E-09
PSC 5 R	0.569	0.149	3.814	0.145	1.49E-04	0.056	14.418	4.10E-09
PSC 9 R	0.533	0.149	3.583	0.137	3.64E-04	0.054	13.824	9.30E-09
PSC 10 R	1.036	0.260	3.989	0.152	7.36E-05	0.058	14.891	2.14E-09
PSC 12 R	0.551	0.138	4.002	0.152	6.98E-05	0.058	14.926	2.03E-09
PSC 16 R	1.768	0.374	4.721	0.179	2.85E-06	0.066	17.105	1.02E-10
PSC 18 R	0.826	0.217	3.803	0.145	1.56E-04	0.056	14.388	4.27E-09

Displayed are the significant PSCs for each linear model with functioning domains. Abbreviations: GF:R: GF role, GF:S: GF social, PSC: Patterns of Structural Covariance, L: left, R: right, Est.: estimate, SE: standard error, t: t-stat, r-PSC: partial correlation coefficient for the effect of PSC; p: p-value for the specific contribution of PSC in the model, Adj. R^2: model's adjusted R^2, F: F-test; model's p: p-value for the whole linear model with age, sex and PSC.

Supplementary Table 17. Linear models to predict Global Functioning (GF) domains with PSCs, age and sex and education.

PSCs	GF:R curr	GF:R L past y	GF:R H past y	GF:R lifetime	GF:S curr	GF:S L past y	GF:S H past y	GF:S lifetime
PSC 2 L	0.057	0.025	0.021	**2.1E-05**	0.478	0.391	0.041	0.020
PSC 5 L	0.056	0.098	0.027	**3.5E-05**	0.921	0.504	0.459	0.113
PSC 6 L	0.083	0.193	0.002	**6.0E-07**	0.436	0.655	0.160	0.005
PSC 12 L	0.102	0.063	0.015	**2.1E-06**	0.347	0.402	0.013	**2.0E-04**
PSC 1 R	0.183	0.024	0.071	**4.2E-05**	0.098	0.054	0.097	0.090
PSC 3 R	0.270	0.368	0.150	**6.6E-04**	0.374	0.210	0.884	0.412
PSC 5 R	0.385	0.288	0.016	**3.8E-04**	0.425	0.243	0.448	0.595
PSC 9 R	0.147	0.136	0.033	***0.002***	0.499	0.341	0.053	0.042
PSC 10 R	0.920	0.804	0.186	**1.8E-04**	0.749	0.917	0.752	0.457
PSC 12 R	0.007	**7.4E-04**	0.022	**6.5E-05**	0.121	0.006	0.026	0.028
PSC 16 R	0.071	0.122	0.011	**7.7E-06**	0.616	0.667	0.044	0.030
PSC 18 R	0.095	0.084	0.056	**4.8E-04**	0.166	0.217	0.098	0.078

Displayed are the significant PSCs for each linear model with functioning domains. Bold values represent significant p-values in the models augmented with education, which overlap with original results, while grey italic values indicate not significant terms, detected in the previous analyses without education (FDR-correction p<0.01). Abbreviations: GF:R: GF role, GF:S: GF social, PSC: Patterns of Structural Covariance, L: left, R: right, L: lowest; H: highest; y: year.

Supplementary Table 18. Linear models to predict Global Functioning (GF) domains with PSCs, age and sex and IQ.

PSCs	GF:R curr	GF:R L past y	GF:R H past y	GF:R lifetime	GF:S curr	GF:S L past y	GF:S H past y	GF:S lifetime
PSC 2 L	0.106	0.053	0.133	**0.002**	0.684	0.544	0.120	0.014
PSC 5 L	0.038	0.064	0.054	**3.98E-04**	0.856	0.437	0.603	0.080
PSC 6 L	0.121	0.226	0.012	**1.08E-05**	0.658	0.990	0.393	0.011
PSC 12 L	0.278	0.154	0.099	**2.37E-04**	0.802	0.829	0.072	**0.001**
PSC 1 R	0.310	0.044	0.170	**7.85E-04**	0.211	0.115	0.201	0.053
PSC 3 R	0.431	0.523	0.405	**0.004**	0.832	0.498	0.642	0.492
PSC 5 R	0.375	0.240	0.016	**7.89E-04**	0.612	0.398	0.415	0.479
PSC 9 R	0.557	0.487	0.166	0.008	0.755	0.774	0.254	0.034
PSC 10 R	0.841	0.954	0.380	**0.002**	0.893	0.694	0.877	0.728
PSC 12 R	0.040	**0.005**	0.047	**0.001**	0.393	0.045	0.064	0.084
PSC 16 R	0.165	0.191	0.038	**2.30E-04**	0.973	0.920	0.067	0.017
PSC 18 R	0.258	0.248	0.194	**0.004**	0.435	0.409	0.289	0.074

Displayed are the significant PSCs for each linear model with functioning domains. Bold values represent significant p-values in the models augmented with IQ, which overlap with original results when applying FDR-correction p<0.05. Grey bold values indicate terms which additionally survive FDR-correction p<0.01. Abbreviations: GF:R: GF role, GF:S: GF social, PSC: Patterns of Structural Covariance, L: left, R: right, L: lowest; H: highest; y: year.

Supplementary Table 19. Linear models to predict Global Functioning (GF) domains with PSCs, age and sex and scan quality measures.

PSC	IQR							
	GF:R curr	GF:R L past y	GF:R H past y	GF:R lifetime	GF:S curr	GF:S L past y	GF:S H past y	GF:S lifetime
PSC 2 L	0.052	0.021	0.010	**1.0E-05**	0.453	0.331	0.038	0.010
PSC 5 L	0.046	0.068	0.014	**2.2E-05**	0.883	0.690	0.414	0.068
PSC 6 L	0.086	0.183	0.002	**3.3E-07**	0.347	0.462	0.137	0.003
PSC 12 L	0.126	0.058	0.007	**4.8E-06**	0.459	0.405	0.017	**3.9E-04**
PSC 1 R	0.132	0.015	0.025	**6.7E-06**	0.069	0.039	0.068	0.048
PSC 3 R	0.230	0.285	0.086	**2.8E-04**	0.373	0.142	0.878	0.428
PSC 5 R	0.278	0.193	0.006	**1.6E-04**	0.332	0.176	0.325	0.501
PSC 9 R	0.130	0.122	0.020	**3.8E-04**	0.401	0.223	0.034	0.018
PSC 10 R	0.800	0.656	0.127	**7.8E-05**	0.640	0.763	0.694	0.416
PSC 12 R	0.011	**6.9E-04**	0.008	**7.4E-05**	0.173	0.007	0.024	0.055
PSC 16 R	0.059	0.077	0.004	**3.0E-06**	0.581	0.453	0.026	0.014
PSC 18 R	0.067	0.066	0.036	**1.6E-04**	0.120	0.149	0.087	0.057
PSC	CNR							
	GF:R curr	GF:R L past y	GF:R H past y	GF:R lifetime	GF:S curr	GF:S L past y	GF:S H past y	GF:S lifetime
PSC 2 L	0.117	0.059	0.009	**6.5E-06**	0.676	0.606	0.087	0.019
PSC 5 L	0.075	0.122	0.014	**3.2E-05**	0.981	0.486	0.556	0.104
PSC 6 L	0.102	0.238	*8.7E-04*	**6.0E-07**	0.360	0.532	0.142	0.008
PSC 12 L	0.192	0.103	0.006	**6.8E-06**	0.581	0.551	0.026	**0.001**
PSC 1 R	0.310	0.055	0.036	**1.2E-05**	0.143	0.105	0.169	0.086
PSC 3 R	0.456	0.659	0.151	**5.9E-04**	0.534	0.350	0.880	0.433
PSC 5 R	0.410	0.293	0.009	**2.2E-04**	0.369	0.271	0.530	0.601
PSC 9 R	0.280	0.306	0.039	**5.6E-04**	0.699	0.597	0.113	0.030
PSC 10 R	0.932	0.980	0.195	**2.3E-04**	0.821	0.882	0.964	0.572
PSC 12 R	0.033	*0.004*	0.011	**1.2E-04**	0.346	0.031	0.067	0.091
PSC 16 R	0.105	0.156	0.009	**7.7E-06**	0.755	0.735	0.059	0.022
PSC 18 R	0.142	0.173	0.047	**1.8E-04**	0.206	0.352	0.187	0.067
PSC	SNR							
	GF:R curr	GF:R L past y	GF:R H past y	GF:R lifetime	GF:S curr	GF:S L past y	GF:S H past y	GF:S lifetime
PSC 2 L	0.072	0.031	0.007	**3.8E-06**	0.565	0.471	0.059	0.015
PSC 5 L	0.064	0.096	0.013	**2.4E-05**	0.913	0.578	0.474	0.097
PSC 6 L	0.096	0.209	*8.4E-04*	**4.2E-07**	0.322	0.435	0.106	0.007
PSC 12 L	0.168	0.080	0.006	**4.7E-06**	0.517	0.445	0.017	**0.001**
PSC 1 R	0.195	0.026	0.026	**6.6E-06**	0.103	0.068	0.120	0.070
PSC 3 R	0.322	0.460	0.122	**3.7E-04**	0.432	0.251	0.961	0.380
PSC 5 R	0.340	0.220	0.008	**1.5E-04**	0.313	0.205	0.432	0.564
PSC 9 R	0.147	0.153	0.025	**3.2E-04**	0.556	0.458	0.079	0.021
PSC 10 R	0.904	0.790	0.165	**1.5E-04**	0.721	0.985	0.917	0.521
PSC 12 R	0.017	*0.001*	0.008	**6.5E-05**	0.247	0.014	0.035	0.074
PSC 16 R	0.068	0.099	0.007	**5.1E-06**	0.666	0.634	0.045	0.018
PSC 18 R	0.075	0.081	0.034	**9.4E-05**	0.139	0.218	0.112	0.053

Displayed are the significant PSCs for each linear model with functioning domains. Bold values represent significant p-values in the models augmented with quality measures (i.e., IQR, CNR and SNR), which overlap with original results, while grey italic values indicate terms, which do not overlap with the previous analyses without quality measures. Blue italic values indicate originally significant terms, which are no longer significant (FDR-correction p<0.01). Abbreviations: IQR: weighted average Image Quality Rating; CNR: Contrast to Noise Ratio; SNR: Signal to Noise Ratio; GF:R: GF role, GF:S: GF social, PSC: Patterns of Structural Covariance, L: left, R: right, L: lowest; H: highest; y: year.

Supplementary Table 20. Associations between clinical variables and PSCs.

	SIPS-P		SIPS-N		SIPS-D		SIPS-G		COGDIS		BDI	
	r	p	r	p	r	p	r	p	r	p	r	p
PSC 2 L	-0.080	0.039	-0.025	0.521	0.002	0.950	-0.079	0.264	-0.111	0.039	-0.018	0.661
PSC 5 L	-0.016	0.671	0.001	0.974	0.064	0.100	-0.022	0.095	-0.049	0.561	0.023	0.568
PSC 6 L	-0.016	0.682	-0.035	0.364	0.068	0.082	-0.003	0.367	-0.022	0.938	0.008	0.847
PSC 7 L	-0.030	0.431	0.026	0.498	0.037	0.340	-0.025	0.475	-0.039	0.519	0.027	0.509
PSC 12 L	-0.089	0.022	-0.019	0.624	0.012	0.755	-0.072	0.732	-0.105	0.060	-0.017	0.674
PSC 1 R	-0.077	0.046	-0.028	0.471	0.024	0.530	-0.053	0.429	-0.054	0.167	-0.032	0.436
PSC 3 R	-0.054	0.163	-0.013	0.731	0.009	0.810	-0.050	0.548	-0.066	0.197	-0.028	0.486
PSC 5 R	0.021	0.590	-0.015	0.704	0.058	0.132	-0.011	0.378	-0.008	0.771	0.021	0.609
PSC 9 R	0.014	0.727	-0.044	0.258	0.056	0.153	-0.013	0.992	-0.006	0.735	-0.068	0.096
PSC 10 R	-0.030	0.443	0.014	0.712	0.023	0.548	-0.019	0.075	-0.036	0.622	0.039	0.337
PSC 12 R	-0.110	0.004	-0.037	0.335	-0.017	0.663	-0.115	0.768	-0.118	0.003	-0.037	0.360
PSC 13 R	-0.041	0.291	-0.025	0.523	-0.016	0.680	-0.015	0.212	-0.026	0.691	-0.010	0.799
PSC 16 R	-0.026	0.508	-0.032	0.412	0.066	0.091	-0.047	0.212	-0.059	0.225	-0.013	0.755
PSC 18 R	-0.024	0.534	0.016	0.676	0.044	0.258	-0.027	0.827	-0.017	0.488	-0.033	0.417

Depicted are partial correlations (correction for age and sex) between the 14 significant PSC components and the clinical variables. Abbreviations: r: correlation coefficient; p: p-value; PSC: Pattern of Structural Covariance; L: left; R: right; SIPS: Structured Interview for the Psychosis-Risk Syndrome, P, N, D, G: positive, negative, disorganized and general symptoms, respectively; COGDIS: Cognitive Disturbances from the Schizophrenia Proneness Instrument-Adult version; BDI: Beck's Depression Inventory.

Supplementary Table 21. P-values for linear models for effect of study group (i.e., HC vs ROD, CHR and ROP) with age and sex.

PSC	ROD		CHR		ROP	
	r	p	r	p	r	p
PSC 2 L	-0.076	*0.015*	-0.134	*1.80E-05*	-0.121	*1.09E-04*
PSC 5 L	-0.081	*0.010*	-0.110	*4.54E-04*	-0.084	*0.008*
PSC 6 L	-0.088	*0.005*	-0.075	*0.017*	-0.100	*0.001*
PSC 7 L	-0.071	*0.023*	-0.063	0.045	-0.086	*0.006*
PSC 12 L	-0.076	*0.015*	-0.095	*0.003*	-0.106	*6.89E-04*
PSC 1 R	-0.092	*0.003*	-0.114	*2.61E-04*	-0.141	*6.39E-06*
PSC 3 R	-0.078	*0.013*	-0.068	*0.030*	-0.131	*2.87E-05*
PSC 5 R	-0.095	*0.002*	-0.073	*0.020*	-0.060	0.056
PSC 9 R	-0.093	*0.003*	-0.086	*0.006*	-0.080	*0.011*
PSC 10 R	-0.085	*0.007*	-0.048	0.127	-0.094	*0.003*
PSC 12 R	-0.105	*0.001*	-0.119	*1.35E-04*	-0.148	*2.10E-06*
PSC 13 R	-0.076	*0.015*	-0.055	0.078	-0.111	*3.77E-04*
PSC 16 R	-0.082	*0.009*	-0.093	*0.003*	-0.073	*0.021*
PSC 18 R	-0.078	*0.013*	-0.059	0.059	-0.125	*6.84E-05*

Depicted are partial correlations (corrected for age, sex and study-group) between the 14 significant PSC components and group (HC vs. PAT) and p-values deriving from linear models for group effect (HC vs. PAT) with age, sex and study-group. Blue italic values represent significant associations at FDR-correction threshold p<0.05. Abbreviations: ROD: recent onset depression, CHR: clinical high risk, ROP: recent onset psychosis, PSC: pattern of structural covariance, L: left hemisphere, R: right hemisphere, r: partial correlation coefficient, p: p-value.

Supplementary Table 22. Associations between PSCs and cognitive domains in the healthy control sample.

	SocCog		WM		SoP		VerLearn		Reas		Attention		Global	
	r	p	r	p	r	p	r	p	r	p	r	p	r	p
PSC 2 L	0.084	0.145	0.069	0.236	0.104	0.076	0.027	0.645	0.081	0.176	0.090	0.121	*0.119*	*0.015*
PSC 5 L	0.007	0.904	0.035	0.546	0.013	0.822	0.085	0.143	0.051	0.394	0.040	0.493	0.013	0.404
PSC 6 L	0.018	0.762	0.027	0.644	0.041	0.488	0.074	0.205	0.016	0.784	0.018	0.751	0.026	0.764
PSC 7 L	0.008	0.895	0.040	0.496	0.048	0.409	*0.109*	*0.059*	0.021	0.728	0.032	0.578	0.007	0.963
PSC 12 L	0.035	0.546	0.094	0.104	0.068	0.248	0.046	0.429	0.008	0.896	0.043	0.462	0.071	0.108
PSC 1 R	0.086	0.137	0.078	0.180	0.083	0.158	0.006	0.919	0.049	0.416	*0.160*	*0.006*	*0.137*	*0.014*
PSC 3 R	0.008	0.884	0.010	0.869	0.016	0.782	*0.113*	*0.050*	0.014	0.813	0.037	0.520	0.009	0.946
PSC 5 R	0.010	0.862	0.071	0.224	0.008	0.894	0.017	0.765	0.045	0.450	*0.116*	*0.045*	0.082	0.110
PSC 9 R	*0.155*	*0.007*	0.058	0.322	*0.108*	*0.065*	0.058	0.321	0.067	0.264	0.069	0.236	*0.133*	*0.032*
PSC10 R	0.021	0.718	0.031	0.589	0.033	0.574	0.013	0.827	0.005	0.927	0.052	0.373	0.028	0.433
PSC12 R	0.022	0.704	*0.116*	*0.045*	0.011	0.857	0.006	0.923	0.066	0.272	0.089	0.124	0.083	0.074
PSC13 R	0.017	0.768	0.042	0.472	0.007	0.911	0.068	0.245	0.083	0.167	*0.107*	*0.066*	0.064	0.166
PSC16 R	0.011	0.853	0.067	0.250	0.008	0.897	0.019	0.749	0.093	0.119	0.067	0.247	0.072	0.249
PSC18 R	0.095	0.102	0.040	0.492	0.030	0.613	0.052	0.369	0.090	0.133	0.067	0.251	0.063	0.166

Displayed are the partial correlations (correction for age and sex) between the 14 significant PSC components and the cognitive domains for the healthy controls sample. No associations were significant. Blue values indicate terms showing trends to significance. Abbreviations: r: correlation coefficient; p: p-value; SocCog: Social cognition; WM: working memory; SoP: Speed of Processing; VerLearn: verbal learning; Reas: reasoning; Global: global cognition; PSC: Patterns of Structural Covariance, L: left, R: right.

SUPPLEMENTAL REFERENCES

Beck AT, Steer RA. 1984. Internal consistencies of the original and revised beck depression inventory. J Clin Psychol.

Benedict RHB, Schretlen D, Groninger L, Brandt J. 2003. Hopkins Verbal Learning Test – Revised: Normative Data and Analysis of Inter-Form and Test-Retest Reliability. Clin Neuropsychol.

Cornblatt BA, Risch NJ, Faris G, Friedman D, Erlenmeyer-Kimling L. 1988. The continuous performance test, identical pairs version (CPT-IP): I. new findings about sustained attention in normal families. Psychiatry Res.

Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, Adams P, Cooper C, Fava M, McGrath PJ, McInnis M, Phillips ML, Trivedi MH, Weissman MM, Shinohara RT. 2018. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage.

Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, Roalf DR, Satterthwaite TD, Gur RC, Gur RE, Schultz RT, Verma R, Shinohara RT. 2017. Harmonization of multi-site diffusion tensor imaging data. Neuroimage. 161:149–170.

Harrison JE, Buxton P, Husain M, Wise R. 2000. Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. Br J Clin Psychol.

Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 8:118–127.

Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, Rosen M, Ruef A, Dwyer DB, Paolini M, Chisholm K, Kambeitz J, Haidl T, et al. 2018. Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or with Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis. JAMA Psychiatry. 75:1156–1172.

Mathias SR, Knowles EEM, Mollon J, Rodrigue A, Koenis MMC, Alexander-Bloch AF, Winkler AM, Olvera RL, Duggirala R, Göring HHH, et al. 2020. Minimal Relationship between Local Gyrification and General Cognitive Ability in Humans. Cereb Cortex. 30:3439–3450.

McMinn MR, Wiens AN, Crossen JR. 1988. Rey auditory-verbal learning test: Development of norms for healthy young adults. Clin Neuropsychol. 2:67–87.

Miller TJ, Mcqlashan TH, Rosen JL, Cadenhead K, Ventura J, Mcfarlane W, Perkins DO, Pearlson QD, Woods SW. 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.

Nowicki S, Duke MP. 1994. Individual differences in the nonverbal communication of affect: The diagnostic analysis of nonverbal accuracy scale. J Nonverbal Behav.

Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, Essock S, Fenton WS, Frese FJ, Gold JM, et al. 2008. The MATRICS consensus cognitive battery, part 1: Test selection, reliability, and validity. Am J Psychiatry. 165:203–213.

Osterrieth PA. 1944. Le test de copie d'une figure complexe. Arch Psychol.

Roiser JP, Stephan KE, Ouden HEM Den, Friston KJ, Joyce EM. 2010. Adaptive and aberrant reward prediction signals in the human brain. Neuroimage. 50:657–664.

Rosen AFG, Roalf DR, Ruparel K, Blake J, Seelaus K, Villa LP, Ciric R, Cook PA, Davatzikos C, Elliott MA, et al. 2018. Quantitative assessment of structural image quality. Neuroimage. 169:407–418.

Ross TP, Hanouskova E, Giarla K, Calhoun E, Tucker M. 2007. The reliability and validity of the self-ordered pointing task. Arch Clin Neuropsychol.

Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, Rodríguez-Sánchez JM, Ríos-Lago M, Tirapu J, Barceló F. 2009. Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. J Int Neuropsychol Soc.

Schaer M, Cuadra MB, Schmansky N, Fischl B, Thiran J-P, Eliez S. 2012. How to Measure Cortical Folding from MR Images: a Step-by-Step Tutorial to Compute Local Gyrification Index. J Vis Exp. 1–8.

Schultze-Lutter F, Addington J, Ruhrmann S, Klosterkötter J. 2007. Schizophrenia Proneness Instrument. Adult Version Roma Giovanni Fioriti Ed.

Schultze-Lutter F, Ruhrmann S, Fusar-Poli P, Bechdolf A, G. Schimmelmann B, Klosterkotter J. 2012. Basic Symptoms and the Prediction of First-Episode Psychosis. Curr Pharm Des.

Schultze-Lutter F, Schimmelmann BG, Ruhrmann S, Michel C. 2013. "A rose is a rose is a rose", but at-risk criteria differ. Psychopathology.

Tierney MC, Nores A, Snow WG, Fisher RH, Zorzitto ML, Reid DW. 1994. Use of the Rey Auditory Verbal Learning Test in Differentiating Normal Aging From Alzheimer's and Parkinson's Dementia. Psychol Assess.

Wechsler D, Coalson DL, Raiford SE. 2008. Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). San Antonio.

Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, Charman T, Tillmann J, Banaschewski T, Dumas G, et al. 2019. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. Biol Psychiatry Cogn Neurosci Neuroimaging. 4:567–578.

Zgaljardic DJ, Temple RO. 2010. Reliability and Validity of the Neuropsychological Assessment Battery-Screening Module (NAB-SM) in a Sample of Patients with Moderate-to-Severe Acquired Brain Injury. Appl Neuropsychol. 17:27–36.

7. References

1. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:1000097.
2. Sotiras A, Resnick SM, Davatzikos C. Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *Neuroimage*. 2015;108:1-16.
3. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. *The Lancet Psychiatry*. 2016;3(2):171-178.
4. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The Science of Prognosis in Psychiatry: A Review. *JAMA Psychiatry*. 2018;75(12):1280-1288.
5. Fusar-Poli P, Borgwardt S, Bechdolf A, et al. The psychosis at risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry*. 2013;70(1):107-120.
6. Fusar-Poli P, Bonoldi I, Yung AR, et al. Predicting Psychosis. *Arch Gen Psychiatry*. 2012;69(3):220-229.
7. Riecher-Rössler A, Studerus E. Prediction of conversion to psychosis in individuals with an at-risk mental state: A brief update on recent developments. *Curr Opin Psychiatry*. 2017;30(3):209-219.
8. Fusar-Poli P, Tantardini M, De Simone S, et al. Deconstructing vulnerability for psychosis: Meta-analysis of environmental risk factors for psychosis in subjects at ultra high-risk. *Eur Psychiatry*. 2017;40:65-75.
9. Perkins DO, Jeffries CD, Addington J, et al. Towards a psychosis risk blood diagnostic for persons experiencing high-risk symptoms: preliminary results from the NAPLS project. *Schizophr Bull*. 2015;41(2):419-428.
10. Gifford G, Crossley N, Fusar-Poli P, et al. Using neuroimaging to help predict the onset of psychosis. *Neuroimage*. 2017;145:209-217.
11. Perez VB, Woods SW, Roach BJ, et al. Automatic auditory processing deficits in schizophrenia and clinical high-risk patients: Forecasting psychosis risk with mismatch negativity. *Biol Psychiatry*. 2014;75(6):459-469.
12. Andreou C, Borgwardt S. Structural and functional imaging markers for susceptibility to psychosis. *Mol Psychiatry*. 2020;25(11):2773-2785.
13. Dwyer DB, Falkai P, Koutsouleris N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol*. 2018;14(1):1-28.
14. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction Models of Functional Outcomes for Individuals in the Clinical High-Risk State for Psychosis or with Recent-Onset Depression: A Multimodal, Multisite Machine Learning Analysis. *JAMA Psychiatry*. 2018;75(11):1156-1172.
15. Koutsouleris N, Dwyer DB, Degenhardt F, et al. Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*. 2020;1-16.
16. Koutsouleris N, Worthington M, Dwyer DB, et al. Towards generalizable and transdiagnostic tools for psychosis prediction:An independent validation and improvement of the NAPLS-2 risk calculator in the multi-site PRONIA cohort. *Biol Psychiatry*. 2021; https://doi.org/10.1016/j.biopsych.2021.06.023.
17. Schmidt A, Cappucciati M, Radua J, et al. Improving Prognostic Accuracy in Subjects at Clinical High Risk for Psychosis: Systematic Review of Predictive Models and Meta-analytical Sequential Testing Simulation. *Schizophr Bull*. 2017;43(2):375-388.
18. Smieskova R, Fusar-Poli P, Allen P, et al. Neuroimaging predictors of transition to psychosis—A systematic review and meta-analysis. *Neurosci Biobehav Rev*. 2010;34(8):1207-1222.

19. Dazzan P. Neuroimaging biomarkers to predict treatment response in schizophrenia: The end of 30 years of solitude? *Dialogues Clin Neurosci*. 2014;16(4):491-503.

20. Llinares-Benadero C, Borrell V. Deconstructing cortical folding: genetic, cellular and mechanical determinants. *Nat Rev Neurosci*. 2019;20(3):161-176.

21. Armstrong E, Schleicher A, Omran H, Curtis M, Zilles K. The ontogeny of human gyrification. *Cereb Cortex*. 1995;5(1):56-63.

22. Rogers J, Kochunov P, Zilles K, et al. On the genetic architecture of cortical folding and brain volume in primates. *Neuroimage*. 2010;53(3):1103-1108.

23. Rash BG, Duque A, Morozov YM, Arellano JI, Micali N, Rakic P. Gliogenesis in the outer subventricular zone promotes enlargement and gyrification of the primate cerebrum. *Proc Natl Acad Sci U S A*. 2019;116(14):7089-7094.

24. Borrell V. How cells fold the cerebral cortex. *J Neurosci*. 2018;38(4):776-783.

25. Sun T, Hevner RF. Growth and folding of the mammalian cerebral cortex: from molecules to malformations. *Nat Rev Neurosci*. 2014; 15(4):217-32.

26. Palaniyappan L, Liddle PF. Aberrant cortical gyrification in schizophrenia: A surface-based morphometry study. *J Psychiatry Neurosci*. 2012;37(6):399-406.

27. Wallace GL, Robustelli B, Dankner N, Kenworthy L, Giedd JN, Martin A. Increased gyrification, but comparable surface area in adolescents with autism spectrum disorders. *Brain*. 2013;136(6):1956-1967.

28. Kippenhan JS. Genetic Contributions to Human Gyrification: Sulcal Morphometry in Williams Syndrome. *J Neurosci*. 2005;25:7840-7846.

29. Pillay P, Manger PR. Order-specific quantitative patterns of cortical gyrification. *Eur J Neurosci*. 2007;25(9):2705-2712.

30. Gautam P, Anstey KJ, Wen W, Sachdev PS, Cherbuin N. Cortical gyrification and its relationships with cortical volume, cortical thickness, and cognitive performance in healthy mid-life adults. *Behav Brain Res*. 2015;287:331-339.

31. Gregory MD, Kippenhan JS, Dickinson D, et al. Regional variations in brain gyrification are associated with general cognitive ability in humans. *Curr Biol*. 2016;26(10):1301-1305.

32. Nanda P, Tandon N, Mathew IT, et al. Local gyrification index in Probands with psychotic disorders and their first-degree relatives. *Biol Psychiatry*. 2014;76(6):447-455.

33. Palaniyappan L, Mallikarjun P, Joseph V, White TP, Liddle PF. Folding of the prefrontal cortex in schizophrenia: Regional differences in gyrification. *Biol Psychiatry*. 2011;69(10):974-979.

34. Depping MS, Thomann PA, Wolf ND, et al. Common and distinct patterns of abnormal cortical gyrification in major depression and borderline personality disorder. *Eur Neuropsychopharmacol*. 2018;28(10):1115-1125.

35. Sasabayashi D, Takayanagi Y, Takahashi T, et al. Increased Occipital Gyrification and Development of Psychotic Disorders in Individuals With an At-Risk Mental State: A Multicenter Study. *Biol Psychiatry*. 2017;82(10):737-745.

36. Matsuda Y, Ohi K. Cortical gyrification in schizophrenia: Current perspectives. *Neuropsychiatr Dis Treat*. 2018;14:1861-1869.

37. Hauke DJ, Schmidt A, Studerus E, et al. Multimodal prognosis of negative symptom severity in individuals at increased risk of developing psychosis. *Transl Psychiatry*. 2021;11(312), doi.org/10.1038/s41398-021-01409-4.

38. Das T, Borgwardt S, Hauke DJ, et al. Disorganized gyrification network properties during the transition to psychosis. *JAMA Psychiatry*. 2018;75(6):613-622.

39. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31(3):968-980.

40. Destrieux C, Fischl B, Dale AM, Halgren E. Automatic parcellation of human cortical gyrii and sulci using standard anatomical nomenclature. *Neuroimage*. 2010;34(1):1-15.

41. Evans AC. Networks of anatomical covariance. *Neuroimage*. 2013;80:489-504.

42. Alexander-Bloch A, Giedd JN, Bullmore E. Imaging structural co-variance between human brain regions. *Nat Rev Neurosci*. 2013;14(5):322-336.

43. Bassett DS, Bullmore E, Verchinski BA, Mattay VS, Weinberger DR, Meyer-Lindenberg A. Hierarchical organization of human cortical networks in health and Schizophrenia. *J Neurosci*. 2008;28(37):9239-9248.

44. Van Den Heuvel MP, Sporns O, Collin G, et al. Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry*. 2013;70(8):783-792.

45. Wang T, Wang K, Qu H, et al. Disorganized cortical thickness covariance network in major depressive disorder implicated by aberrant hubs in large-scale networks. *Sci Rep*. 2016;6(June):1-12.

46. Yeh PH, Zhu H, Nicoletti MA, Hatch JP, Brambilla P, Soares JC. Structural equation modeling and principal component analysis of gray matter volumes in major depressive and bipolar disorders: Differences in latent volumetric structure. *Psychiatry Res - Neuroimaging*. 2010;184(3):177-185.

47. Palaniyappan L, Park B, Balain V, Dangi R, Liddle P. Abnormalities in structural covariance of cortical gyrification in schizophrenia. *Brain Struct Funct*. 2015;220(4):2059-2071.

48. Palaniyappan L, Marques TR, Taylor H, et al. Globally Efficient Brain Organization and Treatment Response in Psychosis: A Connectomic Study of Gyrification. *Schizophr Bull*. 2016;42(6):1446-1456.

49. Sotiras A, Toledo JB, Gur RE, Gur RC, Satterthwaite TD, Davatzikos C. Patterns of coordinated cortical remodeling during adolescence and their associations with functional specialization and evolutionary expansion. *Proc Natl Acad Sci*. 2017;114(13):3527-3532.

50. Yang Z, Oja E. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Trans Neural Networks*. 2010;21(5):734-749.

51. Fusar-Poli P, Solmi M, Brondino N, et al. Transdiagnostic psychiatry: a systematic review. *World Psychiatry*. 2019;18(2):192-207.

52. Patel Y, Parker N, Shin J, et al. Virtual Histology of Cortical Thickness and Shared Neurobiology in 6 Psychiatric Disorders. *JAMA Psychiatry*. 2020;78(1):47-63.

53. Musliner KL, Mortensen PB, McGrath JJ, et al. Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia with Risk for Depression in the Danish Population. *JAMA Psychiatry*. 2019;76:516-525.

54. Goodkind M, Eickhoff SB, Oathes DJ, et al. Identification of a common neurobiological substrate for mental Illness. *JAMA Psychiatry*. 2015;72(4):305-315.

55. McGorry P, Nelson B. Why we need a transdiagnostic staging approach to emerging psychopathology, early diagnosis, and treatment. *JAMA Psychiatry*. 2016;73(3):191-192.

56. Sasabayashi D, Takahashi T, Takayanagi Y, Suzuki M. Anomalous brain gyrification patterns in major psychiatric disorders: a systematic review and transdiagnostic integration. *Transl Psychiatry*. 2021;11(176),https://doi.org/10.1038/s41398-021-01297-8.

57. Owen MJ, O'Donovan MC, Thapar A, Craddock N. Neurodevelopmental hypothesis of schizophrenia. *Br J Psychiatry*. 2011;198(3):173-175.

58. Addington J, Farris M, Stowkowy J, Santesteban-Echarri O, Metzak P, Kalathil MS. Predictors of Transition to Psychosis in Individuals at Clinical High Risk. *Curr Psychiatry Rep*. 2019;21(6):39.

59. Schaer M, Cuadra MB, Schmansky N, Fischl B, Thiran J-P, Eliez S. How to Measure Cortical Folding from MR Images: a Step-by-Step Tutorial to Compute Local Gyrification Index. *J Vis Exp*. 2012;(59):1-8.

60. Cornblatt BA, Auther AM, Niendam T, et al. Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophr Bull*. 2007;33(3):688-702.

Appendix: Gyrification-based predictive models

Methods

A total of 158 CHR individuals (mean age: 23.87, SD: 5.43; 49.4% females) were recruited as part of the PRONIA consortium (see Supplementary Material of Paper II for further information). Twenty-three patients transitioned to psychosis after one year (Table 1).

Table 1. Basic demographic information of the study sample.

N	158
Age [years, mean] (SD)	23.87 (5.43)
Sex [females, N] (%)	78 (49.4)
Psychosis Transition [N] (%)	23 (14.6)
GF-R Baseline [mean] (SD)	6.0 (1.59)
1year Follow-up	6.55 (1.74)
GF-S Baseline [mean] (SD)	6.34 (1.44)
1year Follow-up	6.89 (1.44)

Abbreviations: GF: Global Functioning Role (R) or Social (S), [0:10],higher scores indicate better functioning.

Cortical surfaces were reconstructed from structural MRI images using the FreeSurfer software package (v. 6.0.0, https://surfer.nmr.mgh.harvard.edu/). Local gyrification Index (LGI)[59] was calculated across the whole cortical mesh (Supplementary Material of Paper II).

We built supervised machine learning models based on a Support Vector Machine algorithm using the in-house software NeuroMiner (http://proniapredictors.eu/neurominer/), which ensures a strict validation procedure through a nested cross-validation design[15]. Models were trained and tested to I) predict transition at follow-up (i.e., one year after study inclusion), and II) predict role and social functioning outcome at follow-up using the Global Functioning scale at a cut-off of 7 (GF:R and GF:S[60])[14]. As part of the machine learning pipeline, patients' gyrification meshes underwent preprocessing steps as follows: 1) age and sex correction using partial correlation analysis, 2) site correction thresholding for between-scanner voxel reliability[14] at the 25%, 50%

and 75% percentile, 3) dimensionality reduction using Principal Component Analysis (PCA; 10, 20, 40, 80 and 160 eigenvariates) and 4) standardization using median and winsorization. We used a linear, non-kernelized L2-regularized, L1-loss SVM algorithm and employed wrapper-based feature selection strategies to extract the most predictive features among the large gyrification mesh. Model optimization included hyperparameter combination (i.e., map percentile thresholds, PCA dimensions and SVM's *C* regularization parameter range of $2^{[-4 \underset{\in \mathbb{Z}}{\rightarrow} +4]}$) across all *k*=50 (=5 repetitions × 10 folds) available models in the training partition. The wrapper-based feature selection strategy was based on a greedy sequential forward search (SFS) at each SVM *C* regularization parameter, stopping when 80% of the features had been selected. Models' effectiveness was calculated based on the Balanced Accuracy (BAC) resulting from the test partition. Machine learning pipeline is represented in Figure 1.

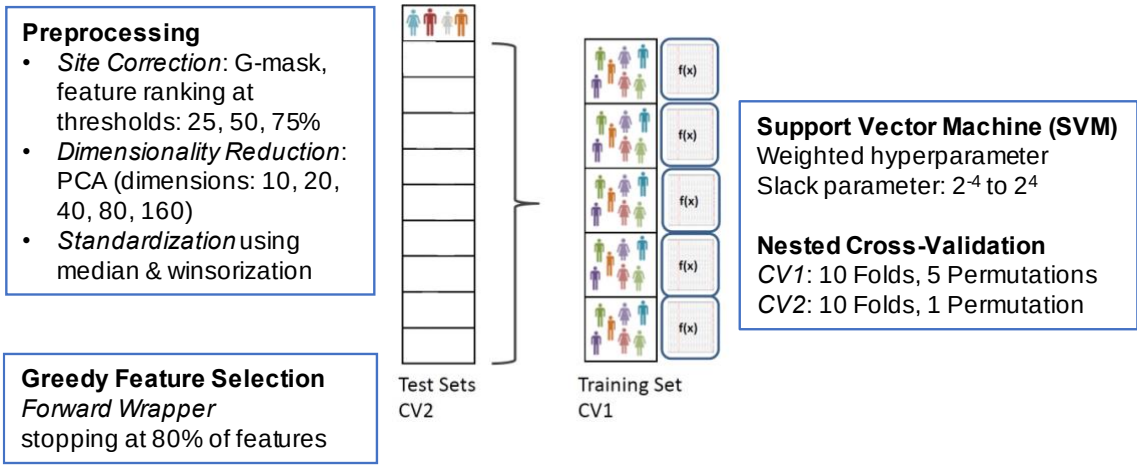


Figure 1. Machine learning pipeline in NeuroMiner.

Results

The SVM algorithm could not predict a transition to psychosis above chance based on gyrification patterns (BAC: 45.9%, sensitivity: 38.5% specificity: 53.3%) and the classification didn't reach statistical significance (Wilcoxon test Z=-1.60, p=.11). Similar results were found for prediction of functional outcome at 1 year follow-up, both for the social subscale (BAC: 50.0%, Z=-0.3, p=0.76) and the role subscale (BAC: 53.6, Z=0.3, p=0.73). Results are summarized in Table 2:

Table 2. Machine learning results

	Transition	GF:S follow-up	GF:R follow-up
Balanced accuracy [%]	45.9	50.0	53.6
Sensitivity [%]	38.5	50.0	55.6
Specificity [%]	53.3	50.0	51.7
Positive predictive value [%]	6.8	34.7	63.3
Negative predictive value [%]	90.6	65.3	43.7

Abbreviations: GF: Global Functioning Role (R) or Social (S), [0:10],higher scores indicate better functioning. Positive and negative predictive value were calculated in NeuroMiner from the initial true/false positive and true/false negative matrix.

Conclusion

Gyrification patterns in CHR individuals were not informative of a future transition to the overt disease, nor they could predict functional outcome after one year. These negative findings suggest that gyrification might be influenced both by early neurodevelopmental factors and by re-wiring processes during adolescence, which might be detectable throughout several psychiatric diseases, rather than in samples of at-risk subjects wherein the etiology and ultimate prognosis is unknown.

In order to better investigate differences in cortical folding and address the role of gyrification as neuroanatomical biomarker for psychosis, future research should focus further on transdiagnostic psychiatric populations in the early stages of disease. The high complexity of this cortical measure also calls for more advanced multivariate statistical approaches, which might be able to better capture subtler morphological patterns. We tackled this challenge by using cutting-edge methods to extract structural covariance at the neuroanatomical level, as well as by investigating larger and more heterogenous psychiatric samples at early disease stages (Paper II).

Acknowledgements

A mia nonna Mina.

Grazie per avermi insegnato a non arrendermi mai

e a combattere sempre con il sorriso.

This doctoral thesis is the result of years of research, commitment, pitfalls, dedication, but also luck and privilege. None of this could have been possible without me being a healthy, European woman who has a wealthy supportive family and had access to education. I hope society will give in the future the same opportunities I had also to those less privileged than me, because good research and discoveries are only to be found with diversity and inclusiveness.

I would like to thank my supervisor, Prof. Koutsouleris, for his enduring support and mentorship. Nikos, you have been an amazing teacher and have inspired me with your vision and talent. Thank you for welcoming me in your group with kindness and accompanying me throughout my personal and career growth.

Thank you, Dom, my supervisor and mentor. It has been a wonderful experience to learn from and with you during this journey. You have shown me how to really push forward and I will always be among grateful for your supervision. Thank you for believing in me.

In the PRONIA group and Neurodiagnostic Application lab I found some of my very best friends in Munich. Thank you for showing me what outstanding things a group can achieve. You are all incredible people and I wish you all the joy and success life can give you. Johanna, Linda, Mafe, Shalaila: I would not have been able to succeed in this challenge without having you by my side. Thank you for being such amazing women, I will always cherish and love you.

Nora, meine Ci. Danke für alles, was wir schon zusammen erlebt haben, und auch schon für was noch kommt. *Immer werden wir so bleiben, lachen über schlechte Zeiten. Deine Schmerzen sind auch meine, Jahr für Jahr.*

Basti, mein sicherer Hafen. Danke für die Abendessen, die du vorbereitet hast, während ich lange gearbeitet habe, für das stressabbauende Lachen, für deine Geduld und dass du immer mein größter Fan bist. Mit dir habe ich Ruhe und Glück in meinem Leben gefunden.

Mamma, papà e Giacomo. Grazie per il vostro supporto e l'amore incondizionato che mi date ogni giorno. Questi anni di studio e di ricerca separata da voi mi hanno mostrato quanto voi siate la cosa a me più cara al mondo.