Clinical characteristics indexing genetic differences in schizophrenia

Taylor, Jacob; de Vries, Ymkje Anna; van Loo, Hanna M.; Kendler, Kenneth S.

Published in:
Molecular Psychiatry

DOI:
10.1038/s41380-022-01850-x

Publication date:
2023

ARTICLE

Check for updates

Clinical characteristics indexing genetic differences in schizophrenia: a systematic review

Jacob Taylor[1,5], Ymkje Anna de Vries[2,5], Hanna M. van Loo[3] and Kenneth S. Kendler[4]✉

Genome-wide studies are among the best available tools for identifying etiologic processes underlying psychiatric disorders such as schizophrenia. However, it is widely recognized that disorder heterogeneity may limit genetic insights. Identifying phenotypes indexing genetic differences among patients with non-affective psychotic disorder will improve genome-wide studies of these disorders. The present study systematically reviews existing literature to identify phenotypes that index genetic differences among patients with schizophrenia and related disorders. We systematically reviewed family-based studies and genome-wide molecular-genetic studies investigating whether phenotypic variation in patients with non-affective psychotic disorders (according to DSM or equivalent systems) was associated with genome-wide genetic variation (PROSPERO number CRD42019136169). An electronic database search of PubMed, EMBASE, and PsycINFO from inception until 17 May 2019 resulted in 4347 published records. These records included a total of 813 relevant analyses from 264 articles. Two independent raters assessed the quality of all analyses based on methodologic rigor and power. We found moderate to strong evidence for a positive association between genetic/familial risk for non-affective psychosis and four phenotypes: early age of onset, negative/deficit symptoms, chronicity, and functional impairment. Female patients also tended to have more affected relatives. Severity of positive symptoms was *not* associated with genetic/familial risk for schizophrenia. We suggest that phenotypes with the most evidence for reflecting genetic difference in participating patients should be measured in future large-scale genetic studies of schizophrenia to improve power to discover causal variants and to facilitate discovery of modifying genetic variants.

Molecular Psychiatry; https://doi.org/10.1038/s41380-022-01850-x

INTRODUCTION

Genome-wide association studies (GWAS) and whole-exome or whole-genome sequencing studies are among the best available tools for identifying etiologic processes underlying human diseases. These studies have been particularly valuable for schizophrenia – a highly heritable condition [1] with poorly understood etiology. Recently, such studies have yielded unambiguous associations between schizophrenia and both common and rare genetic variants [2–4], some of which have led to promising etiologic hypotheses [5]. Increasing sample sizes is expected to continue to yield further genetic risk variants.

However, disorder heterogeneity may limit genetic insights. Since Kraepelin combined hebephrenia, dementia paranoides and catatonia into "dementia praecox," which is similar to our present day concept of schizophrenia, attempts to define this syndrome have grappled with its clinical heterogeneity [6]. Patients with schizophrenia can present quite differently from one another, vary in their premorbid characteristics and follow highly variable courses [7]. If this phenotypic heterogeneity is influenced by genetic factors, then measuring phenotypic traits that index genetic differences in patients will be valuable for several reasons. First, it will improve power to discover causal variants whose effects are most pronounced in patients with certain phenotypic features. For example, stratifying cases of major depression by exposure to adversity enabled discovery of novel genetic variants [8]. Second, it could facilitate discovery of modifying genetic variants – those that shape the form of a disorder without necessarily causing it [9]. Third, richer phenotypic data in large numbers of patients with available genome-wide data could enable molecular-genetic data to inform future changes to diagnostic criteria [10]. Finally, genetic associations in specific subgroups of patients could lead to identification of targets for pharmacologic intervention. Thus, "deeper phenotyping"–knowing more about the cases in genome-wide studies–could have an important impact on the design and utility of such investigations. However, given the size of these studies, and the time needed to collect detailed phenotypic data, it is important to determine which phenotypes are most likely to advance these goals

This study systematically reviews existing literature to identify phenotypes which index genetic differences among patients with non-affective psychotic disorders. We review both family-based

[1]Department of Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts, USA; Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Child and Adolescent Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [3]Department of Psychiatry and Interdisciplinary Center Psychopathology and Emotion regulation, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [4]Virginia Institute for Psychiatric and Behavioral Genetics and Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [4]These authors contributed equally: Jacob Taylor, Ymkje Anna de Vries. ✉email: Kenneth.Kendler@vcuhealth.org

Received: 9 April 2022 Revised: 11 October 2022 Accepted: 20 October 2022
Published online: 18 November 2022

and genome-wide molecular-genetic studies. We highlight those phenotypes with the strongest evidence as well as phenotypes that have been extensively investigated but for which no clear evidence exists. We conclude by arguing that future genetic sample collections of patients with schizophrenia and related conditions should measure those features which meaningfully index relevant genetic effects.

METHODS

Protocol registration

This systematic review's PROSPERO registration number is CRD42019136169.

Eligibility criteria

Studies were eligible for inclusion in the systematic review if they contained analyses that met inclusion criteria for the population of interest, phenotypic features of interest, and an investigation of the phenotypes' genetic relevance. Our population of interest consisted of samples in which at least 50% of individuals were over 18 years old and had a diagnosis of a non-affective psychotic disorder (schizophrenia, schizoaffective disorder, delusional disorder, or unspecified psychosis). We excluded studies where a majority of probands had ages of onset >45 years or had been ill for <6 months. We focused on phenotypic features that can be measured in a clinical or research interview, or with commonly used laboratory or psychometric assessments. Analyses involving phenotypic variation whose measurement required specialized technology (e.g., brain imaging, electrophysiologic measures) were excluded.

Relevant analyses investigated whether phenotypic variation measured in patients with non-affective psychotic disorders was associated with some form of genome-wide genetic variation. Genome-wide genetic variation could be measured directly using molecular methods (e.g. GWAS, whole-exome or whole-genome sequencing or genome-wide CNV microarrays) or indirectly through patterns of familial aggregation. We excluded candidate gene and linkage-based analyses given low reproducibility [11], as well as genome-wide studies that did not include an omnibus measure of genetic risk such as a polygenic risk score. We set a minimum sample size for inclusion at 50 probands for family studies and 100 probands for molecular-genetic studies. Included studies fell into 5 main categories:

1. *Polygenic risk score (PRS)-based studies* - The presence of a phenotype in a proband is associated with differences in PRS.
2. *Other molecular-genetic studies* - The presence of a phenotype in a proband is associated with other genome-wide differences e.g., copy number variant (CNVs) burden or number of rare single nucleotide polymorphisms (SNPs).
3. *Studies of familial aggregation* - The presence of a phenotype in a proband is associated with differences in familial risk for non-affective psychosis.
4. *Affected-relative pair studies* - A phenotype is correlated among two family members, both of whom have a non-affective psychotic disorder.
5. *Genetic modeling studies* (which examined heritability of a phenotype within multiplex families or concordant twin pairs). Such studies can determine whether a phenotype is heritable, as well as its relationship with genetic factors influencing psychosis liability.

Because our purpose is to guide phenotypic measurement among patients with non-affective psychosis in future genetic studies, we excluded studies in which phenotypic variation was measured primarily in individuals without psychotic disorders. This includes classic "endophenotype" studies [12] that explore group phenotypic differences between cases, family members of cases and unrelated controls. We also excluded studies that correlated a particular phenotype (e.g. intelligence) in affected probands and their unaffected relatives. We included only primary studies (i.e. no meta-analyses or reviews) published in peer-reviewed, Anglophonic journals.

Search strategy

We searched PubMed, EMBASE, and PsycINFO from inception until 17 May 2019. Our search strategy (see appendix for details) aimed to identify manuscripts meeting eligibility criteria described above. We also searched the reference lists of included articles.

Study selection and data extraction

After exclusion of duplicates, titles and abstracts were independently reviewed by two raters (JT, YV, or HL), with disagreements resolved by consensus. We obtained the full text of all papers that passed title-abstract review. Full text review and data extraction was performed by at least one rater (JT, YV, or HL) using a data extraction form specific to one of the main types of studies outlined above.

Quality rating

We devised quality ratings for molecular-genetic and family studies (see Supplementary Notes 1 and 2). Quality ratings were performed in duplicate, with disagreements resolved through consensus and discussion with a third rater if necessary. Power for each analysis was based on analysis type and whether the phenotype measure, and/or measure of genetic difference was continuous, categorical, or dichotomous (see Supplementary Tables 7 and 8). Based on methodologic rigor and power, we rated each analysis as "high", "moderate", "low" or "very low" in quality.

Data synthesis

We grouped similar phenotypes into larger categories and synthesized the results by category (and when feasible by specific phenotype), focusing on whether the evidence supported an effect in a uniform direction. We distinguished between analyses that found clearly statistically significant findings ($p < 0.01$), those that found marginally significant findings ($0.01 < p < 0.05$), those that found trend-level findings ($0.05 < p < 0.10$), and those that reported no evidence of association. If the same phenotype was analyzed multiple times in the same or in overlapping samples in different papers, we selected the analysis with the most general phenotype (e.g. aggregate severity of positive symptoms rather than severity of hallucinations). If multiple analyses in overlapping samples examined equally suitable phenotypes, we generally selected the largest available sample only. In rare cases, if sample sizes were similar and more than two analyses were performed in the same sample, we also considered consistency in results (e.g. if two analyses showed no association and one showed a significant association we would choose one of the former analyses). We examined the available evidence for each of the five types of studies separately. Where possible, for each phenotype category, we looked at the overall number of studies, and the number of high-quality studies, that reported an association in a particular direction vs. those that reported no association or an association in the opposite direction. Due to the heterogeneity in study methods and often limited reporting of results (e.g. missing effect sizes), we could not perform a meta-analysis.

RESULTS

Search results and study selection

Supplementary Fig. 1 depicts a flow chart of the study inclusion and selection. We identified 3942 published records through electronic searches and an additional 405 records through reference searches and other sources. We assessed the full texts of 584 articles. Of these, 320 articles were excluded as ineligible, leaving 264 articles for review. These included 193 articles containing family aggregation analyses, 37 with affected-pair analyses, 8 with genetic modeling analyses, 20 with PRS-based analyses, and 16 with other molecular-genetic analyses. Individual articles often contained multiple analyses. The number of each analysis type contained in each article and full references are provided in Supplemental Table 1.

Summary of the evidence

We grouped all extracted phenotypes into 15 categories (see Table 1 for details). The largest number of analyses were available for the phenotypes disease onset, negative and positive symptoms ($N = 112$, 107 and 106 respectively). Below we describe the strongest evidence that emerged from reviewing each study type. Additionally, a quantitative summary of the findings from familial aggregation and affected-relative pairs studies is provided in

Table 1. Number of each analysis type by phenotype category.

Phenotype category	Number of analyses (number of papers/number of independent samples)				
	PRS	Other molecular genetic	Familial aggregation	Affected pair	Genetic modeling
Negative symptoms	10 (8/2)	5 (4/3)	60 (38/31)	35 (20/16)	2 (1/1)
Onset (age, mode)	5 (4/3)	6 (6/5)	69 (56/42)	27 (17/18)	0
Positive symptoms	9 (9/3)	6 (5/4)	45 (31/25)	44 (18/14)	2 (2/1)
Mood (e.g., mania, depression)	12 (7/2)	4 (3/2)	34 (19/17)	25 (13/13)	0
Course/severity	9 (6/8)	5 (4/3)	43 (37/35)	15 (13/11)	0
Cognition/education	18 (7/6)	8 (6/5)	36 (23/19)	2	7 (4/2)
Diagnosis/subtypes (e.g., schizophrenia vs. other conditions)	0	1	58 (51/32)	10	1
Functioning	4	2	35 (27/23)	20 (6/7)	1
Putative environmental risk factors	0	5 (3/3)	31 (27/23)	2	0
Gender	0	3	30 (30/21)	6 (6/5)	0
Other	0	4 (3 /2)	20 (16/16)	8 (5/5)	1
Substance use	1	6 (3/3)	6 (6/5)	1	0
Medication tolerance	0	1	11 (11/9)	0	0
Family History of other conditions	3 (3/1)	3	3 (3/2)	0	0
Physical exam findings (anomalies, soft neurological signs)	0	0	8 (7/6)	0	0
Total number of analyses	71	58	489	195	14

Number of papers refers to number of separate articles that contain relevant analyses. Number of independent samples is often smaller than number of analyses because similar analyses are often performed in the same sample and reported in either the same or separate articles.

Tables 2 and 3. Detailed summaries of all findings are provided in Supplementary Note 3 and all underlying data is provided in Supplementary Tables 2–6.

Evidence from PRS-based studies

We examined 71 PRS-based analyses, most of which explored associations between PRS for schizophrenia and patient phenotypic variation. The strongest evidence for phenotype-PRS associations came from one study of over 8,000 patients [13]. Patients were assigned a score for four symptomatic factors: manic, depressive, positive and negative. Associations between these factors and PRSs for schizophrenia and bipolar disorder were examined. This study found strong statistical associations between SCZ PRS and the severity or presence of negative symptoms and between BP PRS and manic symptoms. There was no evidence for associations between either PRS and severity of positive or depressive symptoms.

Several other PRS-based studies looked at similar phenotypes in sub-populations from this study. Other reasonably high-quality studies performed analyses in different but substantially smaller samples ($N < 1000$). Even where statistically significant findings were made in these small, high-quality studies, replication across samples was poor. We could not draw conclusions with confidence about the genetic relevance of other phenotypes from these studies.

Evidence from other molecular-genetic studies

In total, there were 59 analyses using other molecular-genetic techniques. The strongest available evidence, based on findings from two high-quality studies based on largely overlapping samples, suggested that rare variant burden is associated with moderate to severe intellectual disabilities in people with schizophrenia [14, 15]. The methods for measuring CNV and other rare variant burden in other studies were too varied to permit definitive conclusions about any other phenotype.

Evidence from family aggregation studies

A total of 489 family aggregation analyses compared proband phenotypic variation to risk of schizophrenia or related disorders in family members (Table 2). Although there were many null findings for all phenotypes -i.e. no statistically significant association between the phenotypes and familial risk of psychosis-, for certain phenotypes, a majority of non-null findings were in the same direction. The strongest evidence for a positive association with familial risk of psychosis was available for lower age of onset and negative symptoms (Fig. 1a, c). The strongest evidence against a relationship between familial risk of psychosis and phenotypic variation was for severity of positive symptoms (Fig. 1b). Moderate evidence also suggested that functional impairment and chronicity are associated with familial aggregation of psychotic disorders. In addition, female probands tended to have more affected relatives compared with male probands. For most phenotype categories the evidence was equivocal. For example, while four high-quality studies suggested that probands with schizophrenia may have more affected family members than probands with other non-affective psychotic disorders, one high-quality study found evidence to the contrary and eight found no evidence either way (Fig. 1d). Evidence regarding other phenotype categories was inconsistent or too limited to permit conclusions about either the presence or absence of a relationship between the phenotype and familial aggregation (see Supplementary Note 3 for further information about phenotypes with fewer available analyses).

Evidence from affected-relative pair studies

A total of 195 affected-relative pair analyses examined the phenotypic correlation among related pairs of patients with non-affective psychosis. Most of these analyses involved siblings with schizophrenia. Table 3 shows the number of findings demonstrating significant positive correlation of phenotypes among affected relative pairs as well as the number of null findings; as expected, no study showed

Table 2. Findings from family aggregation studies.

Phenotype category	Positive association				No association	Negative association (total)
	Significant	Marginal	Trend	Total		
Age of onset	11 (8)	6 (1)	5 (3)	**22 (12)**	**19 (8)**	**0 (0)**
Course/severity	5 (2)	5 (2)	1 (0)	**11 (4)**	**22 (5)**	**2 (0)**
Recovery	2 (0)	3 (0)	1 (0)	**6 (0)**	**5 (1)**	**0 (0)**
Severity	0 (0)	1 (1)	0 (0)	**1 (1)**	**9 (3)**	**0 (0)**
Negative symptoms	5 (1)	8 (3)	0 (0)	**13 (4)**	**18 (7)**	**0 (0)**
Negative/deficit syndrome	4 (1)	8 (3)	0 (0)	**12 (4)**	**17 (7)**	**0 (0)**
Disorganized symptoms	1 (1)	2 (1)	0 (0)	**3 (2)**	**7 (2)**	**0 (0)**
Positive symptoms	0 (0)	2 (1)	0 (0)	**2 (1)**	**20 (9)**	**3 (1)**
Functioning	2 (1)	4 (3)	2 (0)	**8 (4)**	**15 (5)**	**0 (0)**
Global functioning	1 (0)	2 (1)	0 (0)	**3 (1)**	**7 (3)**	**0 (0)**
Putative environmental risk factors	2 (0)	3 (1)	1 (0)	**6 (1)**	**11 (4)**	**6 (1)**
Season of birth (winter vs. other seasons)	1 (0)	3 (1)	1 (0)	**5 (1)**	**8 (4)**	**4 (0)**
Female gender	4 (3)	3 (2)	0 (0)	**7 (5)**	**14 (3)**	**0 (0)**
Diagnosis/subtype (SCZ vs. other diagnoses)	0 (0)	3 (2)	3 (2)	**6 (4)**	**13 (9)**	**2 (1)**
Diagnosis/subtype (SCZ vs. SZA)	0 (0)	1 (0)	2 (1)	**3 (1)**	**10 (8)**	**2 (1)**
Diagnosis/subtype (hebephrenic vs. paranoid)	0 (0)	3 (0)	3 (0)	**6 (0)**	**9 (4)**	**0 (0)**
Cognition	1 (0)	3 (0)	0 (0)	**4 (0)**	**14 (5)**	**1 (0)**
Mood symptoms	0 (0)	4 (3)	1 (0)	**5 (3)**	**11 (5)**	**1 (0)**
Depression	0 (0)	3 (2)	1 (0)	**4 (2)**	**9 (4)**	**1 (0)**
Medication tolerance	1 (0)	0 (0)	0 (0)	**1 (0)**	**7 (2)**	**1 (0)**
Anomalies	1 (0)	0 (0)	1 (0)	**2 (0)**	**3 (0)**	**1 (0)**
Substance use	0 (0)	0 (0)	0 (0)	**0 (0)**	**5 (1)**	**0 (0)**

For a summary of main results, see the bolded final three columns.
Numbers indicate the number of analyses in a specific category; numbers in brackets indicate the number of high-quality analyses in a specific category. Significant findings reflect P values < 0.01, marginal findings reflect P values that are between 0.01 and 0.05 and trend findings reflect P values that are between 0.05 and 0.1. For negative associations, significant, marginal, and trend-level findings were combined because of the low number of negative associations. All broad phenotype categories, and all individual phenotypes with ≥10 independent analyses, are shown. In general, positive associations reflect that worse outcomes on a phenotype (e.g. more symptoms, decreased functioning) are associated with increased familial risk of schizophrenia. For the following phenotypes, the association was considered positive if increased familial risk was associated with: lower age of onset, birth in winter, female gender, schizophrenia diagnosis, and hebephrenic subtype.

significant negative correlations. The strongest evidence was available for age of onset, disorganized symptoms, and course/severity. Mood symptoms and severity of positive symptoms were also frequently correlated in affected pairs with schizophrenia, which means that these phenotypes may index genetic factors that modify the clinical presentation of the disease.

Evidence from genetic modeling studies

Only eight studies containing 14 analyses used these methods and they produced no consistent findings. For details see Supplementary Note 3.

Overall summary of the evidence

Overall, we found moderate to strong evidence for a positive association between genetic/familial risk for non-affective psychosis and four phenotypes: early age of onset, negative/deficit symptoms, chronicity, and functional impairment (Table 4). Female patients also tended to have more affected relatives. While it is likely that virtually all patients in included studies have or have had some positive symptoms, aggregate results from both PRS-based and familial aggregation studies showed that severity of positive symptoms is *not* associated with genetic/familial risk for schizophrenia.

Age of onset, disorganized symptoms, and chronicity were highly correlated among pairs of affected family members, with somewhat less robust evidence for positive and mood symptoms. This suggests that these phenotypes may be influenced or modified by genetic factors that do not necessarily reflect genetic risk for schizophrenia. Finally, there is moderate evidence that rare genetic variants also affect the risk of intellectual disability among patients with these disorders.

DISCUSSION

We systematically reviewed, for the first time to our knowledge, evidence for how genetic heterogeneity among patients with schizophrenia and other non-affective psychotic disorders is related to phenotypic variation. The 264 articles included in this review indicated several traits with moderate or strong evidence that they index a higher genetic loading for schizophrenia, including early age of onset, high level of negative symptoms, functional impairment, chronicity, and female gender. The finding that earlier age of onset is associated with greater risk of psychosis in family members of probands is among the most highly replicated of our findings and is also consistent with results across a range of other human disorders [16–21]. Several of the other traits associated with increased genetic loading for schizophrenia

Table 3. Findings from affected pair studies.

Findings from affected pair studies: total (high quality)					
Phenotype category	**Positive association**				**No association**
	Significant	**Marginal**	**Trend**	**Total**	
Age of onset	8 (3)	6 (1)	0 (0)	**14 (4)**	**4 (1)**
Negative symptoms	5 (1)	4 (1)	0 (0)	**9 (2)**	**7 (2)**
Negative/deficit syndrome	4 (1)	3 (0)	0 (0)	**7 (1)**	**7 (3)**
Disorganized symptoms	3 (1)	5 (3)	0 (0)	**8 (4)**	**2 (0)**
Positive symptoms	5 (1)	2 (1)	0 (0)	**7 (2)**	**7 (2)**
Hallucinations	4 (0)	0 (0)	0 (0)	**4 (0)**	**7 (4)**
Mood symptoms	6 (1)	0 (0)	1 (1)	**7 (2)**	**6 (1)**
Depression	4 (1)	0 (0)	0 (0)	**4 (1)**	**7 (1)**
Course/severity	6 (1)	2 (1)	0 (0)	**8 (2)**	**3 (1)**
Functioning	4 (1)	0 (0)	0 (0)	**4 (1)**	**3 (0)**
Diagnosis/subtype (hebephrenic vs. paranoid)	3 (0)	0 (0)	0 (0)	**3 (0)**	**3 (1)**
Gender	0 (0)	2 (1)	0 (0)	**2 (1)**	**3 (1)**

For a summary of main results, see the bolded final three columns.
Numbers indicate the number of analyses in a specific category; numbers in brackets indicate the number of high-quality analyses in a specific category. Significant findings reflect *P* values < 0.01, marginal findings reflect P values that are between 0.01 and 0.05 and trend findings reflect *P* values that are between 0.05 and 0.1. No studies found negative associations. All broad phenotype categories with ≥5 independent analyses, and all individual phenotypes with ≥10 independent analyses, are shown.

–more severe negative symptoms, greater functional impairment, and greater overall severity/chronicity– are also associated with earlier age of onset [22]. Thus, this set of replicated findings is mutually consistent. Indeed, this set of traits have been recognized as characterizing a distinct subgroup of schizophrenia patients since Kraepelin. The diagnostic category he described as hebephrenia corresponds to patients who become ill early in life, follow a deteriorating course and have prominent disorganized symptoms and more prominent negative symptoms compared with positive symptoms [23].

One possible way to account for the well-replicated finding that female probands have more relatives with psychotic disorders than male probands is a female protective effect [24]. The incidence of schizophrenia is higher in men than in women [25]. Thus, it may take higher genetic liability for a woman to be diagnosed with schizophrenia compared to a man. As a result, women with schizophrenia may have greater average genetic risk compared with men. Another possible explanation is that female patients tend to become ill later in life and are less likely to be hospitalized compared with male patients [26]. Thus, efforts to match male and female patients by age, or any aspect of ascertainment that make it more likely for patients who have been hospitalized to end up in a study, may result in female patients having higher genetic loadings than male patients. This is a pattern that has been fairly well-established in autism, which is substantially more common in males than females, though is not always seen in other heritable disorders with sex differences in incidence [27, 28].

Many studies explored the hypothesis that severity of positive symptoms (hallucinations and delusions) reflect a higher degree of genetic loading for psychotic disorders. A large majority of such studies found no association between severity of positive symptoms and familiality. While some studies did find associations between familial risk and the presence of particularly severe forms of positive symptoms, such as bizarre delusions [29], these specific symptoms were not measured in enough independent samples to include in our results. However, findings from affected-relative pair studies suggest that the severity and expression of positive symptoms may nonetheless be influenced by familial/genetic factors ('modifier genes'). Similarly, our review suggests that mood symptoms (depressive, manic) do not index genetic vulnerability for schizophrenia, but the expression of these mood symptoms in patients with schizophrenia may be under genetic influence. The hypothesis that manic symptoms in particular are under genetic influence, but that the genetic factors that influence them do not themselves confer risk for developing schizophrenia is supported by the finding that manic symptoms in schizophrenia are associated with polygenic risk for bipolar disorder, but not for schizophrenia [14]. Thus, positive symptom severity and mood symptoms in non-affective psychotic disorders seem to be influenced by *modifying* genetic factors – those that shape the course or expression of psychotic disorders without influencing risk of developing a psychotic illness itself [9]. Because modifying genetic factors may not differ between cases and controls, genome-wide studies with minimal phenotyping will be unable to uncover them. Direct assessment of positive psychotic symptoms and mood symptoms in patients may therefore be necessary to identify genetic variants that influence these symptoms. Furthermore, because positive symptoms and mood symptoms are also the manifestations of severe psychiatric illness that respond best to existing medications, it is not currently clear whether these modifying genetic influences directly shape illness presentation versus affecting medication adherence or response.

Finally, rare genetic variants, including large copy number variants and single nucleotide variants predicted to disrupt the functioning of proteins, are increasingly recognized as risk factors for schizophrenia [3, 14, 30–32]. We found moderate evidence that severe intellectual disability is associated with rare protein disrupting mutations in individuals with schizophrenia. Of note, this type of variation is also increased in individuals with intellectual disability without psychotic illnesses. This finding suggests that schizophrenia accompanied by moderate to severe premorbid intellectual disability may be associated with distinct genetic risk factors.

Our results should be interpreted in the context of several potential methodological limitations. First, although we included a very large number of studies, there are many potentially relevant phenotypes, and most of these phenotypes have not been studied sufficiently to draw conclusions as to whether or not they index genetic risk. Second, the nature of the data available to us made

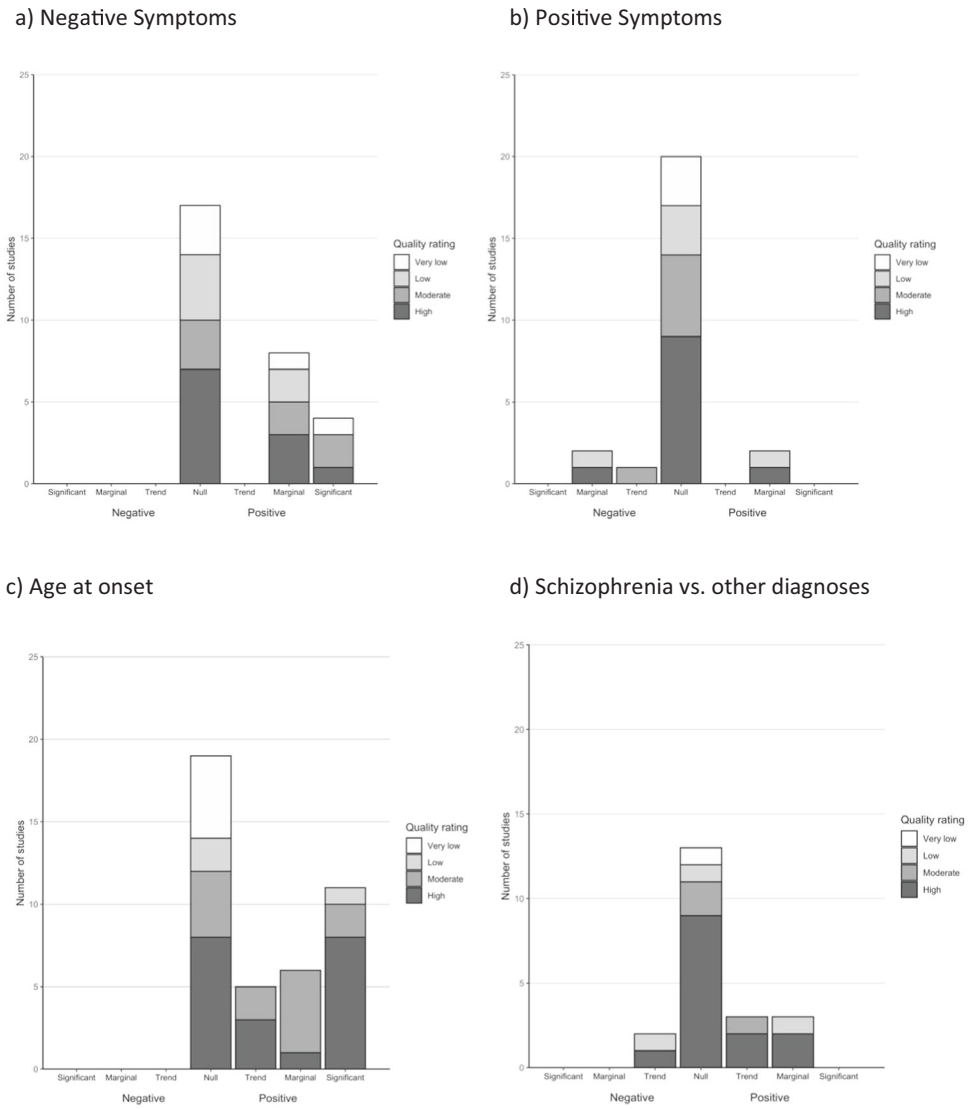

Fig. 1 Evidence from independent familial aggregation analyses relating features of schizophrenia with increased familial risk. Evidence from independent family aggregation analyses that increased familial risk of non-affective psychotic disorders is associated with **a** greater burden of negative symptoms, **b** greater burden of positive symptoms, **c** lower age of onset and **d** a diagnosis of schizophrenia rather than a different non-affective psychotic disorder. The Y-axis indicates the number of studies in each category. The X-axis divides studies into those which present significant, marginal or trend negative results, null results or trend, marginal and significant positive results.

formal meta-analyses infeasible, so we were unable to compare systematically the magnitudes of effect sizes for different phenotypes against each other. Third, in the familial aggregation studies, the vast majority of studies indicated no difference in familial risk for almost all of the investigated phenotypes. This is likely due to insufficient power because of small sample sizes, which may have resulted in false negative findings.

Fourth, due to the heterogeneity of the study types and analyses performed, it is likely that our search was not complete despite our best efforts. Indeed, in several cases, papers identified through cross-references contained relevant analyses not mentioned in their abstracts. Fifth, the heavy reliance on family studies means that most of the analyses reviewed in this paper cannot definitively distinguish between genetic effects and non-genetic factors. While twin studies have been able to provide estimates of the heritability of schizophrenia itself, existing twin samples are too small (or not phenotyped deeply enough) to provide stable estimates for the heritability of phenotypic variation within patients with schizophrenia However, the findings that have emerged consistently from family aggregation and affected pair studies are very likely to be under substantial genetic influence, since the role of shared family environment has been found to be modest for schizophrenia and other psychiatric disorders [1].

The findings of this review are important for several reasons. Because of the very large number of participants necessary to draw inferences from genetic studies, consensus that a core group of phenotypes index clinical heterogeneity and should be measured could improve the efficiency of psychiatric genetic research. In addition to impacting future psychiatric genetic research, these findings help us understand how clinical differences among patients with the same diagnosis may reflect etiologic differences. As with other psychiatric diagnoses, the definitions of schizophrenia and schizoaffective disorder emphasize diagnostic reliability over biological coherence. Limiting phenotypic data in genetic studies to case/control status using these definitions will therefore likely lead us to miss important biology. Furthermore, genetic data has historically been important in refining diagnoses [33]. For molecular-genetic data to be used to modify current nosology, it is necessary to know more about patients than just their diagnosis.

Table 4. Summary of evidence.

	Phenotype	Level of evidence	Type of studies
Phenotypes that index genetic risk	Lower age of onset	Strong	FA
	Increased negative/deficit symptoms	Strong	FA (moderate) + PRS (strong)
	Female gender	Moderate	FA
	More chronic, severe course	Moderate	FA (moderate) + PRS (moderate)
	Functioning	Moderate	FA (moderate) + PRS (weak)
Phenotypes that do NOT index genetic risk	Positive symptoms	Strong	FA + PRS
Phenotypes that reflect modifying genetic influences	Age of onset	Strong	AP
	Disorganized symptoms	Strong	AP
	More chronic, severe course	Strong	AP
	Manic symptoms	Moderate	AP + PRS
	Positive symptoms	Moderate	AP
Phenotypes that reflect rare disruptive variant burden	Intellectual disability	Moderate	Other molecular-genetic

FA family aggregation, *PRS* polygenic risk score, *AP* affected pair.
Strong or moderate evidence from FA and AP studies generally required at least 10 independent analyses, and at least 5 high-quality analyses. For FA studies, strong evidence required that positive findings clearly outnumbered negative findings, and that the proportion of null findings was not too large. This was operationalized as the total positive findings minus total negative findings divided by total analyses. We required this quantity to be ≥0.4 among both high-quality analyses and among all analyses for strong evidence and ≥0.2 for moderate evidence. For AP studies, we considered the evidence to be strong if ≥75% of all analyses and of all high-quality analyses had positive findings, and as moderate if ≥50% of all analyses and of all high-quality analyses did. For both FA and AP studies, we still rated a phenotype as having moderate or strong evidence if there were slightly too few studies, but the evidence bar would still be met even with additional negative findings (for FA) or null findings (for AP) (e.g., if only 4 high-quality studies were available, all of which had positive findings, we still rated a phenotype as having strong evidence). We also took other factors in consideration for "borderline" cases according to the above operationalization, e.g., sample sizes in studies with negative or null vs. positive findings. The PRS study [14] and rare variant studies [15, 16] that allow for meaningful inferences are discussed in the main text.

CONCLUSIONS

Future genetic studies of schizophrenia should measure at least those phenotypes with the most evidence for reflecting genetic difference in participating patients: negative symptoms, age of onset, severity/chronicity, and functional impairment. The best way to assess these constructs may differ across studies and will depend on available time, expertise and sources of information such as medical records and collateral informants. However, efforts to collect measures related to each of these phenotypes in many different contexts (along with a willingness to share phenotypic data across studies) will facilitate analyses at the scale necessary to identify causal and modifying genetic factors that might otherwise be missed. This effort toward deep phenotyping will contribute toward making genome-wide association studies more effective tools for understanding the underlying pathogenic basis of psychiatric disease and informing the future of psychiatric nosology.

REFERENCES

1. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60:1187–92.
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
3. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet. 2017;49:27–35.
4. Howrigan DP, Rose SA, Samocha KE, Fromer M, Cerrato F, Chen WJ, et al. Exome sequencing in schizophrenia-affected parent–offspring trios reveals risk conferred by protein-coding de novo mutations. Nat Neurosci. 2020;23:185–93.
5. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, et al. Schizophrenia risk from complex variation of complement component 4. Nature. 2016;530:177–83.
6. Jablensky A. Subtyping schizophrenia: implications for genetic research. Mol Psychiatry. 2006;11:815–36.
7. Carpenter WT Jr, Kirkpatrick B. The heterogeneity of the long-term course of schizophrenia. Schizophr Bull. 1988;14:645–52.
8. Peterson RE, Cai N, Dahl AW, Bigdeli TB, Edwards AC, Webb BT, et al. Molecular genetic analysis subdivided by adversity exposure suggests etiologic heterogeneity in major depression. Am J Psychiatry. 2018;175:545–54.
9. Fanous AH, Kendler KS. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. Mol Psychiatry. 2005;10:6–13.
10. Smoller JW, Andreassen OA, Edenberg HJ, Faraone SV, Glatt SJ, Kendler KS. Psychiatric genetics and the structure of psychopathology. Mol Psychiatry. 2019;24:409–20.
11. Duncan LE, Ostacher M, Ballon J. How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. Neuropsychopharmacology. 2019;44:1518–23.
12. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry. 2003;160:636–45.
13. Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address: douglas.ruderfer@vanderbilt.edu, Bipolar Disorder and Schizophrenia Working Group of the Psychiatric Genomics Consortium. Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell. 2018;173:1705–1715.e16.
14. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506:185–90.
15. Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, Torniainen M, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat Genet. 2017;49:1167–73.
16. Escott-Price V, Consortium IPDG, Nalls MA, Morris HR, Lubbe S, Brice A, et al. Polygenic risk of P arkinson disease is correlated with disease age at onset. Ann Neurol. 2015;77:582–91.
17. Baldessarini RJ, Tondo L, Vazquez GH, Undurraga J, Bolzani L, Yildiz A, et al. Age at onset versus family history and clinical outcomes in 1,665 international bipolar-I disorder patients. World Psychiatry. 2012;11:40–46.
18. Klein DN, Schatzberg AF, McCullough JP, Dowling F, Goodman D, Howland RH, et al. Age of onset in chronic major depression: relation to demographic and clinical variables, family history, and treatment response. J Affect Disord. 1999;55:149–57.

19. Grant BF. The impact of a family history of alcoholism on the relationship between age at onset of alcohol use and DSM-IV alcohol dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. Alcohol Health Res World. 1998;22:144–7.
20. Monsén U, Broström O, Nordenvall B, Sörstad J, Hellers G. Prevalence of inflammatory bowel disease among relatives of patients with ulcerative colitis. Scand J Gastroenterol. 1987;22:214–8.
21. Scheuner MT, Whitworth WC, McGruder H, Yoon PW, Khoury MJ. Expanding the definition of a positive family history for early-onset coronary heart disease. Genet Med. 2006;8:491–501.
22. Immonen J, Jääskeläinen E, Korpela H, Miettunen J. Age at onset and the outcomes of schizophrenia: a systematic review and meta-analysis. Early Inter Psychiatry. 2017;11:453–60.
23. Kendler KS. The development of Kraepelin's mature diagnostic concept of hebephrenia: a close reading of relevant texts of Hecker, Daraszkiewicz, and Kraepelin. Mol Psychiatry. 2020;25:180–93.
24. Robinson EB, Lichtenstein P, Anckarsäter H, Happé F, Ronald A. Examining and interpreting the female protective effect against autistic behavior. Proc Natl Acad Sci USA. 2013;110:5258–62.
25. McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. BMC Med. 2004;2:13.
26. Leung MD DA, Chue MRC Psych DP. Sex differences in schizophrenia, a review of the literature. Acta Psychiatr Scand. 2000;101:3–38.
27. Salvatore JE, Cho SB, Dick DM. Genes, environments, and sex differences in alcohol research. J Stud Alcohol Drugs. 2017;78:494–501.
28. Merikangas AK, Almasy L. Using the tools of genetic epidemiology to understand sex differences in neuropsychiatric disorders. Genes Brain Behav. 2020;19:e12660.
29. Peralta V, Cuesta MJ. Familial liability and schizophrenia phenotypes: a polydiagnostic approach. Schizophr Res. 2007;96:125–34.
30. Bergen SE, Ploner A, Howrigan D, CNV Analysis Group and the Schizophrenia Working Group of the Psychiatric Genomics Consortium, O'Donovan MC, Smoller JW, et al. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. Am J Psychiatry. 2019;176:29–35.
31. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landén M, et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci. 2016;19:1433–41.
32. Halvorsen M, Huh R, Oskolkov N, Wen J, Netotea S, Giusti-Rodriguez P, et al. Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. Nat Commun. 2020;11:1842.
33. Robins E, Guze SB. Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. Am J Psychiatry. 1970;126:983–7.

AUTHOR CONTRIBUTIONS

JT, YADV, HVL and KSK conceived and designed the study. JT, YADV, and HVL performed the literature review and data coding. KSK oversaw the project. JT drafted the first version of the manuscript and YADV, HVL and KSK revised it. All authors reviewed and approved the final version.

FUNDING

This study is supported in part by the Stanley Center for Psychiatric Research. HVL was supported by a VENI grant from the Talent Programme of the Netherlands Organization of Scientific Research (NWO-ZonMW 09150161810021).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-022-01850-x.

Correspondence and requests for materials should be addressed to Kenneth S. Kendler.

Reprints and permission information is available at http://www.nature.com/reprints

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