Development of Proteomic Prediction Models for Transition to Psychotic Disorder in the Clinical High-Risk State and Psychotic Experiences in Adolescence

David Mongan, MB BCh, BAO; Melanie Föcking, PhD; Colm Healy, MSc; Subash Raj Susai, MSc; Meike Heurich, PhD; Kieran Wynne, MSc; Barnaby Nelson, PhD; Patrick D. McGorry, PhD; G. Paul Amminger, PhD; Merete Nordinoft, PD; Marie-Odile Krebs, PhD; Anita Riecher-Rössler, PhD; Rodrigo A. Bressan, PhD; Neus Barrantes-Vidal, PhD; Stefan Borgwardt, MD; Stephan Ruhrmann, MD; Gabriele Sachs, PhD; Christos Pantelis, MD; Mark van der Gaag, PhD; Lieuwe de Haan, PhD; Lucia Valmaggia, PhD; Thomas A. Polliak, PhD; Matthew J. Kempton, PhD; Bart P. F. Rutten, PhD; Robert Whelan, PhD; Mary Cannon, PhD; Stan Zammit, PhD; Gerard Cagney, PhD; David R. Cotter, PhD; Philip McGuire, PhD; for the European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) High Risk Study Group

**IMPORTANCE** Biomarkers that are predictive of outcomes in individuals at risk of psychosis would facilitate individualized prognosis and stratification strategies.

**OBJECTIVE** To investigate whether proteomic biomarkers may aid prediction of transition to psychotic disorder in the clinical high-risk (CHR) state and adolescent psychotic experiences (PEs) in the general population.

**DESIGN, SETTING, AND PARTICIPANTS** This diagnostic study comprised 2 case-control studies nested within the European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) and the Avon Longitudinal Study of Parents and Children (ALSPAC). EU-GEI is an international multisite prospective study of participants at CHR referred from local mental health services. ALSPAC is a United Kingdom-based general population birth cohort. Included were EU-GEI participants who met CHR criteria at baseline and ALSPAC participants who did not report PEs at age 12 years. Data were analyzed from September 2018 to April 2020.

**MAIN OUTCOMES AND MEASURES** In EU-GEI, transition status was assessed by the Comprehensive Assessment of At-Risk Mental States or contact with clinical services. In ALSPAC, PEs at age 18 years were assessed using the Psychosis-Like Symptoms Interview. Proteomic data were obtained from mass spectrometry of baseline plasma samples in EU-GEI and plasma samples at age 12 years in ALSPAC. Support vector machine learning algorithms were used to develop predictive models.

**RESULTS** The EU-GEI subsample (133 participants at CHR; mean [SD] age, 22.6 [4.5] years; 68 [51.1%] male) comprised 49 (36.8%) who developed psychosis and 84 (63.2%) who did not. A model based on baseline clinical and proteomic data demonstrated excellent performance for prediction of transition outcome (area under the receiver operating characteristic curve [AUC], 0.95; positive predictive value [PPV], 75.0%; and negative predictive value [NPV], 98.6%). Functional analysis of differentially expressed proteins implicated the complement and coagulation cascade. A model based on the 10 most predictive proteins accurately predicted transition status in training (AUC, 0.99; PPV, 76.9%; and NPV, 100%) and test (AUC, 0.92; PPV, 81.8%; and NPV, 96.8%) data. The ALSPAC subsample (121 participants from the general population with plasma samples available at age 12 years (61 [50.4%] male) comprised 55 participants (45.5%) with PEs at age 18 years and 61 (50.4%) without PEs at age 18 years. A model using proteomic data at age 12 years predicted PEs at age 18 years, with an AUC of 0.74 (PPV, 67.8%; and NPV, 75.8%).

**CONCLUSIONS AND RELEVANCE** In individuals at risk of psychosis, proteomic biomarkers may contribute to individualized prognosis and stratification strategies. These findings implicate early dysregulation of the complement and coagulation cascade in the development of psychosis outcomes.

Published online August 26, 2020.
Early detection of psychosis may improve clinical outcomes. Clinical high-risk (CHR) criteria enable identification of vulnerable groups with 3-year transition rates to first-episode psychosis (FEP) of 16% to 35%. However, it is difficult to predict outcomes individually. Previous studies have also characterized an extended psychosis phenotype that includes individuals with psychotic experiences (PEs). These subthreshold symptoms are associated with an increased risk of psychotic and nonpsychotic disorders and reduced global functioning.

Biomarkers may augment prognosis and stratification strategies. We aimed to compare plasma protein expression in individuals at CHR who do and do not develop psychosis and to develop models incorporating proteomic data for individualized prediction of transition to FEP. This study also aimed to apply similar methods for prediction of PEs in a general population sample.

Methods

Ethical approval for this diagnostic study was granted by the Royal College of Surgeons in Ireland. Ethics committees of participating sites granted approval for the European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI). Approval was also obtained from the Avon Longitudinal Study of Parents and Children (ALSPAC) Ethics and Law Committee and local research ethics committees. Informed consent for collection of biological samples was obtained in accordance with the Human Tissue Act 2004. Informed consent for use of questionnaire and clinical data was obtained following recommendations of the ALSPAC Ethics and Law Committee at the time.

Study 1: CHR Sample

Participants and Study Design

EU-GEI study includes a prospective cohort of 344 participants at CHR recruited across 11 international sites. Individuals with CHR symptoms who were referred by local mental health services were eligible to participate if they met CHR criteria according to the Comprehensive Assessment of At-Risk Mental States (CAARMS) and provided written informed consent. Exclusion criteria were current or past psychotic disorder as assessed either by CAARMS interview or by contact with the clinical team or review of clinical records. Sixty-five of 344 participants at CHR (18.9%) developed psychosis on follow-up, 57 within 24 months and 8 after 24 months.

Baseline clinical measures were recorded. These included age, sex, body mass index (BMI; calculated as weight in kilograms divided by height in meters squared), years of education, General Assessment of Functioning (GAF) subscales for symptoms and disability, the Scale for the Assessment of Negative Symptoms (SANS), the Brief Psychiatric Rating Scale (BPRS), and the Montgomery-Åsberg Depression Rating Scale (MADRS).

Sample Preparation, Proteomics, Validation, and Replication

Laboratory procedures were conducted blind to case-control status. Protein depletion, digestion, and peptide purification resulted in 24 samples.

Figure 1. Derivation of Participants Included in the Initial EU-GEI Mass Spectrometry Experiment and Their Provision of Plasma Samples

344 Participants at CHR recruited

49 CHR-T analyzed

16 Excluded

14 Did not provide sample

2 Sample not available at time of experiment

45 CHR-T

279 CHR-NT

56 Excluded

56 Did not provide sample

139 Random subsample not selected

84 CHR-NT analyzed

CHR indicates clinical high risk; CHR-NT, participants at clinical high risk who did not transition to psychosis; CHR-T, participants at clinical high risk who transitioned to first-episode psychosis; and EU-GEI, European Network of National Schizophrenia Networks Studying Gene-Environment Interactions.
were performed using baseline plasma samples. Discovery-based proteomic methods were used. Briefly, 5 μL from each prepared sample was injected on a Q Exactive (Thermo Scientific) mass spectrometer operated in data-dependent acquisition mode for label-free liquid chromatography mass spectrometry (eMethods in Supplement 1 and eAppendix in Supplement 2).

Nine proteins in plasma samples from the same participants at CHR described above (Figure 1) were assessed using enzyme-linked immunosorbent assay (ELISA). Details are available in eMethods in Supplement 1.

In an effort to reproduce our findings, we conducted a partial replication of the initial mass spectrometry experiment by analyzing baseline plasma samples from 49 CHR-T cases (2 of these cases were different from the initial experiment) and an entirely new group of 86 CHR-NT control cases. Details are available in eMethods in Supplement 1.

Study 2: General Population Sample
Participants and Study Design
The ALSPAC is a prospective birth cohort. Pregnant women in Avon, United Kingdom, with delivery dates between April 1, 1991, and December 31, 1992, were invited to participate, and 14,541 pregnancies were enrolled. When the oldest children were approximately age 7 years, an attempt was made to bolster the sample with children who did not join originally. The sample size for analyses using data from age 7 years is 15,454 pregnancies (14,901 children alive at 1 year).

Plasma samples obtained at age 12 years from ALSPAC participants who did or did not report PEs at age 18 years were previously investigated. In data-independent acquisition analyses focused on proteins of the complement pathway, several proteins were differentially expressed. Herein, we performed data-dependent acquisition analyses (rather than data-independent acquisition) in this sample to achieve broader proteome coverage.

Outcome, Sample Preparation, and Proteomics
Psychotic experiences were assessed in participants at age 12 years and age 18 years using the Psychosis-Like Symptoms Interview and were rated as not present, suspected, or definite. Of 4060 participants assessed at both time points, 190 (4.7%) had suspected or definite PEs at age 18 years but not at age 12 years. The present study was based on a subsample of case participants (who did not report PEs at age 12 years but reported at least 1 definite PE at age 18 years) and randomly selected control participants (who did not report PEs at either age 12 years or age 18 years).

Plasma samples at age 12 years were prepared as previously described. Data-dependent acquisition proteomic analyses were performed as for study 1.

Data and Statistical Analyses
Data were analyzed from September 2018 to April 2020. Clinical data were tested for differences using the 2-sided t test for continuous variables and χ² test for categorical variables in SPSS, version 25 (IBM). P values were corrected for multiple comparisons using the Benjamini-Hochberg procedure with a 5% false discovery rate (FDR). The threshold for statistical significance was FDR-corrected P < .05.

Label-free quantification was performed in MaxQuant, version 1.5.2.8 (Max Planck Institute of Biochemistry). Proteins identified with at least 2 peptides (1 uniquely assignable to the protein) and quantified in more than 80% of samples were taken forward for analysis and log₂ transformed. Missing values were imputed using imputeL.CMD (version 2.0) in RStudio. Label-free quantification values were converted to z scores and winsorized within ±3 z.

Analysis of covariance was performed in Stata, version 15 (StataCorp LLC), comparing the mean label-free quantification for each protein in cases and controls. Covariates included age, sex, BMI, and years of education in study 1 and sex, BMI at age 12 years, and maternal social class in study 2. P values were corrected for multiple comparisons with a 5% FDR.

Predictive Models
Neurominer, version 1.0, for MatLab 2018a (MathWorks Inc) was used to develop support vector machine (SVM) models (eMethods in Supplement 1). The development of each model is summarized in eTable 1 in Supplement 1.

Models 1a-c: Predicting Transition Using Clinical and Proteomic Data
First, we developed a model predicting transition using clinical and proteomic data together (model 1a). eTable 2 in Supplement 1 lists the included clinical features. Geographical generalizability was incorporated using leave-site-out cross-validation (eMethods in Supplement 1) as recommended for multisite consortia. To assess the relative contribution of clinical and proteomic data, we next developed models using the same cross-validation and training framework but based on clinical (model 1b) and proteomic (model 1c) features separately.

Model 2a and b: Parsimonious Model
We sought to generate a parsimonious model based on the 10 highest-weighted proteomic predictors and internally validate this model in unseen data (Figure 1 in Supplement 1). As the largest site, London, United Kingdom, was chosen as the test site, and data for these participants were held out.

To derive the 10 highest-weighted proteins, a model (model 2a) was generated using proteomic data from all sites except London (n = 30 for CHR-T and n = 50 for CHR-NT). A reduced model was then developed based solely on data for these 10 proteins in the non-London data set (model 2b) and then tested in the held-out London data (n = 19 for CHR-T and n = 34 for CHR-NT). Both models used leave-site-out cross-validation.

Model 3: Replication
Because of differences in protein identifications, it was not possible to apply models 1a-c and 2a-b to the replication data set. We instead sought to replicate our initial findings by performing a second discovery analysis, generating a new model (with leave-site-out cross-validation) predicting transition based on clinical and proteomic data in the replication data set.
Model 4: Predicting PEs Using Proteomic Data
We developed a model predicting PEs at age 18 years in the ALSPAC based on proteomic data at age 12 years. Repeated nested cross-validation with 5 inner folds and 5 outer folds was used.

Supplementary Analyses
Several supplementary analyses (eMethods in Supplement 1) were performed. These included the following: the development of a model predicting transition in EU-GEI based on ELISA data (model S1), the development of a model predicting functional outcome in EU-GEI (GAF disability subscale score ≤60 [poor functional outcome] vs >60 [good functional outcome] at 24 months) based on clinical and proteomic data (model S2), investigation of potential EU-GEI site associations for clinical and proteomic data, and the development of multivariate-corrected versions of SVM models whereby the variance associated with multiple covariates was extracted using principal components analysis.

Results
Study I: CHR Sample
Of 344 participants at CHR who were recruited, 152 (44.2%) attended face-to-face interviews at 12 months and 105 (30.5%) at 24 months. Baseline characteristics of participants who did or did not attend at least 1 follow-up interview are compared in eTable 3 in Supplement 1. After FDR correction, participants who attended interviews had a mean of 1 more year of education and a lower mean SANS total global score than those who did not attend interviews but were otherwise comparable.

The subsample for the initial experiment comprised 133 (49 CHR-T and 84 CHR-NT) participants with baseline plasma samples available, of whom 49 (36.8%) developed psychosis (Figure 1). The mean (SD) age of the participants was 22.6 (4.5) years; 68 participants (51.1%) were male. After FDR correction, participants included in the subsample had a higher mean SANS total composite, SANS total global, and BPRS total scores than nonincluded participants but were otherwise comparable on baseline characteristics (eTable 4 in Supplement 1).

Subsample characteristics are listed in Table 1. After FDR correction, there were no statistically significant group differences for CHR-T vs CHR-NT based on baseline characteristics. The median duration from baseline to transition was 219 days (interquartile range, 424 days). The CHR-T participants had lower mean functional outcome scores at 2 years compared with CHR-NT participants (mean GAF symptoms score at 2 years, 42.3 in CHR-T vs 62.2 in CHR-NT; FDR-corrected P < .007; mean GAF disability score at 2 years, 44.7 in CHR-T vs 64.5 in CHR-NT; FDR-corrected P < .007).

Differential Expression
Of 345 proteins identified, 166 were quantified in more than 80% of plasma samples. There was nominally statistically significant (P < .05) differential expression of 56 proteins in CHR-T vs CHR-NT, of which 35 remained statistically significant after FDR correction (eTables 5 and 6 in Supplement 1). eFigure 2 in Supplement 1 shows a functional association network for these proteins, and eTable 7 in Supplement 1 lists protein-protein interactions. On functional enrichment analysis, the topmost implicated pathway was the complement and coagulation cascade (eTable 8 in Supplement 1).

Model 1a: Predicting Transition Using Clinical and Proteomic Data
An SVM model predicted transition status based on clinical and proteomic features (model 1a), with excellent performance (area under the receiver operating characteristic curve [AUC], 0.95; [P < .001]; sensitivity, 98.0%; specificity, 81.0%; positive predictive value [PPV], 75.0%; and negative predictive value [NPV], 98.6%). Performance metrics are listed in Table 2. Figure 2A shows the mean algorithm scores and predicted outcomes stratified by site. The receiver operating characteristic curve is shown in Figure 2B. Table 3 lists the 10% highest-weighted features according to the mean feature weight. For example, the 5 highest-ranked predictive features were alpha-2-macroglobulin (A2M) (mean weight, −0.330), immunoglobulin heavy constant mu (IGHM) (mean weight, −0.256), C4b-binding protein alpha chain (C4BPA) (mean weight, −0.161), complement component 8 alpha chain (C8A) (mean weight, 0.158), and phospholipid transfer protein (PLTP) (mean weight, −0.146).

Model Ib and Ic: Clinical and Proteomic Data
The clinical model (model Ib) demonstrated poor predictive performance (AUC, 0.48; P = .63). These results are summarized in Table 2 and eFigure 3 in Supplement 1. For example, sensitivity was 46.9%, specificity was 53.6%, PPV was 37.1%, and NPV was 63.4%.

The proteomic model (model Ic) demonstrated excellent predictive performance (AUC, 0.96; P < .001). These results are summarized in Table 2 and eFigure 4 in Supplement 1. For example, sensitivity was 100%, specificity was 84.5%, PPV was 79.0%, and NPV was 100%.

Model 2a and b: Parsimonious Model
The AUC for the model based on proteomic data from all sites except London (model 2a) was 0.94 (P < .001) (Table 2 and eFigure 5 in Supplement 1). The 10 highest-weighted features were alpha-2-macroglobulin (A2M), immunoglobulin heavy constant mu (IGHM), C4b-binding protein alpha chain (C4BPA), vitamin K-dependent protein S, fibulin 1, transthyretin, N-acetylmuramoyl-l-alanine amidase, vitamin D-binding protein, clusterin, and complement component 6 (C6).

A reduced model based solely on these 10 most predictive proteins was developed using data from all sites except London (model 2b), with an AUC of 0.99 (P < .001), sensitivity of 100%, specificity of 82.0%, PPV of 76.9%, and NPV of 100% (Table 2 and eFigure 6 in Supplement 1). This model predicted transition status in the held-out London data, with an AUC of 0.92, sensitivity of 94.7%, specificity of 88.2%, PPV of 81.8%, and NPV of 96.8% (Table 2).

ELISA Validation
After FDR correction, 2 proteins assessed by ELISA showed statistically significant mean differences between CHRT and
### Table 1. Sample Characteristics for CHR-T and CHR-NT Groups in the Initial Experiment

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. (%)</th>
<th>CHR-T (n = 49)</th>
<th>CHR-NT (n = 84)</th>
<th>Test Statistic</th>
<th>P value</th>
<th>Corrected P value (FDR 5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline age, mean (SD), y</td>
<td>0</td>
<td>22.2 (5.0)</td>
<td>22.9 (4.2)</td>
<td>-0.824</td>
<td>.41</td>
<td>.78</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>26 (53.1)</td>
<td>42 (50.0)</td>
<td></td>
<td>0.116</td>
<td>.73</td>
<td>.91</td>
</tr>
<tr>
<td>Female</td>
<td>23 (46.9)</td>
<td>42 (50.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline body mass index, mean (SD)</td>
<td>20 (15.0)</td>
<td>24.5 (4.5)</td>
<td>24.4 (6.1)</td>
<td>0.116</td>
<td>.91</td>
<td>.91</td>
</tr>
<tr>
<td>Baseline years of education, mean (SD)</td>
<td>14 (10.5)</td>
<td>14.1 (3.4)</td>
<td>14.4 (3.0)</td>
<td>-0.625</td>
<td>.53</td>
<td>.79</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>33 (67.3)</td>
<td>58 (69.0)</td>
<td></td>
<td>2.370</td>
<td>.31</td>
<td>.65</td>
</tr>
<tr>
<td>Black</td>
<td>8 (16.3)</td>
<td>7 (8.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>8 (16.3)</td>
<td>19 (22.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever used cannabis</td>
<td>3 (2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>36 (73.5)</td>
<td>65 (77.4)</td>
<td></td>
<td>0.051</td>
<td>.82</td>
<td>.91</td>
</tr>
<tr>
<td>No</td>
<td>11 (22.4)</td>
<td>18 (21.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not known</td>
<td>2 (4.1)</td>
<td>1 (1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline cannabis use</td>
<td>29 (21.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>15 (30.6)</td>
<td>26 (31.0)</td>
<td></td>
<td>0.030</td>
<td>.86</td>
<td>.91</td>
</tr>
<tr>
<td>No</td>
<td>22 (44.9)</td>
<td>41 (48.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not known</td>
<td>12 (24.5)</td>
<td>17 (20.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline tobacco use**</td>
<td>14 (10.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>21 (42.9)</td>
<td>43 (51.2)</td>
<td></td>
<td>0.373</td>
<td>.54</td>
<td>.79</td>
</tr>
<tr>
<td>No</td>
<td>21 (42.9)</td>
<td>34 (40.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not known</td>
<td>7 (14.3)</td>
<td>7 (8.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline alcohol use***</td>
<td>3 (2.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>35 (71.4)</td>
<td>58 (69.0)</td>
<td></td>
<td>0.071</td>
<td>.79</td>
<td>.91</td>
</tr>
<tr>
<td>No</td>
<td>13 (26.5)</td>
<td>24 (28.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not known</td>
<td>1 (2.0)</td>
<td>2 (2.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline medication use</td>
<td>31 (23.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19 (38.8)</td>
<td>32 (38.1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antidepressant</td>
<td>13</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antipsychotic</td>
<td>9</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypnotic</td>
<td>2</td>
<td>6</td>
<td></td>
<td>0.042</td>
<td>.84</td>
<td>.91</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>20 (40.8)</td>
<td>31 (36.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not known</td>
<td>10 (20.4)</td>
<td>21 (25.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF symptoms score</td>
<td>12 (9.0)</td>
<td>52.4 (10.3)</td>
<td>56.0 (10.0)</td>
<td>-1.906</td>
<td>.06</td>
<td>.19</td>
</tr>
<tr>
<td>GAF disability score</td>
<td>5 (3.8)</td>
<td>52.3 (12.4)</td>
<td>54.8 (11.3)</td>
<td>-1.148</td>
<td>.25</td>
<td>.60</td>
</tr>
<tr>
<td>SANS total composite score</td>
<td>19 (14.3)</td>
<td>20.9 (14.0)</td>
<td>16.2 (11.6)</td>
<td>1.903</td>
<td>.06</td>
<td>.19</td>
</tr>
<tr>
<td>SANS total global score</td>
<td>11 (8.3)</td>
<td>6.6 (4.1)</td>
<td>5.8 (3.7)</td>
<td>1.158</td>
<td>.25</td>
<td>.60</td>
</tr>
<tr>
<td>BPRS total score</td>
<td>10 (7.5)</td>
<td>49.1 (11.5)</td>
<td>44.2 (10.2)</td>
<td>2.452</td>
<td>.02</td>
<td>.08</td>
</tr>
<tr>
<td>MADRS total score</td>
<td>7 (5.3)</td>
<td>20.3 (10.4)</td>
<td>19.2 (9.2)</td>
<td>0.657</td>
<td>.51</td>
<td>.79</td>
</tr>
<tr>
<td>GAF symptoms score at 2 y, mean (SD)</td>
<td>62 (46.6)</td>
<td>42.3 (13.2)</td>
<td>62.2 (10.3)</td>
<td>-7.125</td>
<td>&lt;.001</td>
<td>&lt;.007</td>
</tr>
<tr>
<td>GAF disability score at 2 y, mean (SD)**</td>
<td>54 (40.6)</td>
<td>44.7 (9.1)</td>
<td>64.5 (12.8)</td>
<td>-8.024</td>
<td>&lt;.001</td>
<td>&lt;.007</td>
</tr>
<tr>
<td>GAF disability score at 2 y, dichotomous outcome†</td>
<td>54 (40.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor functioning</td>
<td>29 (59.2)</td>
<td>18 (21.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good functioning</td>
<td>1 (2.0)</td>
<td>31 (36.9)</td>
<td></td>
<td>27.734</td>
<td>&lt;.001</td>
<td>&lt;.007</td>
</tr>
<tr>
<td>Not known</td>
<td>19 (38.8)</td>
<td>35 (41.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CHR-NT, participants at clinical high risk who did not transition to psychosis; CHR-T, participants at clinical high risk who transitioned to first-episode psychosis; EU-GEI, European Network of National Schizophrenia Networks Studying Gene-Environment Interactions; FDR, false discovery rate; GAF, General Assessment of Functioning; MADRS, Montgomery-Åsberg Depression Rating Scale (high score, greater number and severity of depressive symptoms; low score, lower number and severity of depressive symptoms).  
* Missing data were excluded in hypothesis tests.  
** Daily tobacco use for at least 1 month over the previous 12 months.  
*** At least 12 alcoholic beverages over the previous 12 months.  
† Data available for 71 of 133 participants (27 CHR-T and 44 CHR-NT).  
‡ Data available for 79 of 133 participants (30 CHR-T and 49 CHR-NT).  
§ A GAF disability subscale score of 60 or less indicates poor functioning, and a score greater than 60 indicates good functioning.
### Table 2. Performance Metrics for Unadjusted Support Vector Machine Models

<table>
<thead>
<tr>
<th>Model description</th>
<th>Transition, No./total No. (%)</th>
<th>Nontransition, No./total No. (%)</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>Balanced accuracy, %</th>
<th>AUC (95% CI)</th>
<th>PPV, %</th>
<th>NPV, %</th>
<th>Positive likelihood ratio</th>
<th>Negative likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1a: EU-GEI clinical and proteomic data*</td>
<td>48/49 (98.0)</td>
<td>1/49 (2.0)</td>
<td>98.0</td>
<td>81.0</td>
<td>89.5</td>
<td>0.95 (0.91-0.99)</td>
<td>75.0</td>
<td>98.6</td>
<td>5.1</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites</td>
<td>Features: 69 clinical and 166 proteomic</td>
<td>Target: transition status</td>
<td>N: 49 transition, 84 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1b: EU-GEI clinical data*</td>
<td>23/49 (46.9)</td>
<td>26/49 (53.1)</td>
<td>46.9</td>
<td>53.6</td>
<td>50.3</td>
<td>0.48 (0.38-0.58)</td>
<td>37.1</td>
<td>63.4</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites</td>
<td>Features: 69 clinical</td>
<td>Target: transition status</td>
<td>N: 49 transition, 84 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1c: EU-GEI proteomic data*</td>
<td>49/49 (100)</td>
<td>0/49 (0)</td>
<td>100</td>
<td>84.5</td>
<td>92.3</td>
<td>0.96 (0.92-1.00)</td>
<td>79.0</td>
<td>100</td>
<td>6.5</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites</td>
<td>Features: 166 proteomic</td>
<td>Target: transition status</td>
<td>N: 49 transition, 84 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2a: EU-GEI proteomic data (non-London)*</td>
<td>28/30 (93.3)</td>
<td>2/30 (6.7)</td>
<td>93.3</td>
<td>80.0</td>
<td>86.7</td>
<td>0.94 (0.88-1.00)</td>
<td>73.7</td>
<td>95.2</td>
<td>4.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites except London</td>
<td>Features: 166 proteomic</td>
<td>Target: transition status</td>
<td>N: 30 transition, 50 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2b: top 10, training data*</td>
<td>30/30 (100)</td>
<td>0/30</td>
<td>100</td>
<td>82.0</td>
<td>91.0</td>
<td>0.99 (0.96-1.00)</td>
<td>76.9</td>
<td>100</td>
<td>5.6</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites except London</td>
<td>Features: 10 proteomic</td>
<td>Target: transition status</td>
<td>N: 30 transition, 50 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
## Table 2. Performance Metrics for Unadjusted Support Vector Machine Models (continued)

<table>
<thead>
<tr>
<th>Model description</th>
<th>Transition, No./total No. (%)</th>
<th>Nontransition, No./total No. (%)</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>Balanced accuracy, %</th>
<th>AUC (95% CI)</th>
<th>PPV, %</th>
<th>NPV, %</th>
<th>Positive likelihood ratio</th>
<th>Negative likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2b: top 10, test data&lt;sup&gt;a&lt;/sup&gt;</td>
<td>18/19 (94.7)</td>
<td>1/19 (5.3)</td>
<td>30/34 (88.2)</td>
<td>4/34 (11.8)</td>
<td>94.7</td>
<td>88.2</td>
<td>91.5</td>
<td>0.92 (0.83-1.00)</td>
<td>81.8</td>
<td>96.8</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, London site</td>
<td>Features: 10 proteomic</td>
<td>Target: transition status</td>
<td>N: 19 transition, 34 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 3: EU-GEI clinical and proteomic data&lt;sup&gt;a&lt;/sup&gt;</td>
<td>48/49 (98.0)</td>
<td>1/49 (2.0)</td>
<td>77/86 (89.5)</td>
<td>9/86 (10.5)</td>
<td>98.0</td>
<td>89.5</td>
<td>93.7</td>
<td>0.98 (0.95-1.00)</td>
<td>84.2</td>
<td>98.7</td>
</tr>
<tr>
<td>Data set: EU-GEI replication experiment, all sites</td>
<td>Features: 69 clinical and 119 proteomic</td>
<td>Target: transition status</td>
<td>N: 49 transition, 86 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 4: ALSPAC proteomic data&lt;sup&gt;a&lt;/sup&gt;</td>
<td>40/55 (72.7)</td>
<td>15/55 (27.3)</td>
<td>47/66 (71.2)</td>
<td>19/66 (28.8)</td>
<td>72.7</td>
<td>71.2</td>
<td>72.0</td>
<td>0.74 (0.65-0.83)</td>
<td>67.8</td>
<td>75.8</td>
</tr>
<tr>
<td>Data set: ALSPAC</td>
<td>Features: 265 proteomic</td>
<td>Target: PEs at age 18 y</td>
<td>N: 55 PEs, 66 no PE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model S1: ELISA</td>
<td>33/44 (75.0)</td>
<td>11/44 (25.0)</td>
<td>51/82 (62.2)</td>
<td>31/82 (37.8)</td>
<td>75.0</td>
<td>62.2</td>
<td>68.6</td>
<td>0.76 (0.67-0.85)</td>
<td>51.6</td>
<td>82.3</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites</td>
<td>Features: 9 ELISA</td>
<td>Target: transition status</td>
<td>N: 44 transition, 82 nontransition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model S2: functional outcome</td>
<td>27/47 (57.4)</td>
<td>20/47 (42.6)</td>
<td>22/32 (68.8)</td>
<td>10/32 (31.3)</td>
<td>57.4</td>
<td>68.8</td>
<td>63.1</td>
<td>0.74 (0.63-0.85)</td>
<td>73.0</td>
<td>52.4</td>
</tr>
<tr>
<td>Data set: EU-GEI initial experiment, all sites</td>
<td>Features: 69 clinical and 166 proteomic</td>
<td>Target: functional outcome</td>
<td>N: 47 poor functioning (GAF &lt;60); 32 good functioning (GAF &gt;60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ALSPAC, Avon Longitudinal Study of Parents and Children; AUC, area under the receiver operating characteristic curve; ELISA, enzyme-linked immunosorbent assay; EU-GEI, European Network of National Schizophrenia Networks Studying Gene-Environment Interactions; NPV, negative predictive value; PE, psychotic experience; PPV, positive predictive value.

<sup>a</sup> Models 1a-c, 2, and 3 are adjusted for age, sex, body mass index, and years of education, and model 4 is additionally adjusted for race/ethnicity and tobacco use.
CHR-NT. These were A2M and complement component 1r (C1r) (eTables 9 and 10 in Supplement 1). The A2M mean in CHR-T was 1173.1 μg/mL vs 1150.1 μg/mL in CHR-T (FDR-corrected \( P = .02 \)), and the C1r mean in CHR-T was 52803.9 μg/mL vs 52803.9 μg/mL in CHR-T (FDR-corrected \( P = .04 \)).

Model 3: Replication
Replication subsample characteristics are listed in eTables 11 and 12 in Supplement 1. Of 485 proteins identified, 119 were quantified in more than 80% of plasma samples. There was nominally statistically significant (\( P < .05 \)) differential expression of 82 proteins, of which 78 remained statistically significant after FDR correction (eTable 13 in Supplement 1).

Model 3 demonstrated excellent performance for prediction of transition in the replication data set (AUC, 0.98 \( [ P < .001 \); sensitivity, 98.0%; specificity, 89.5%; PPV, 84.2%; and NPV, 98.7%) (Table 2 and eFigure 7 in Supplement 1). The highest-weighted 10% of features are listed in Table 3. For example, the 5 highest-ranked predictive features were A2M (mean weight, −0.286), carboxypeptidase N subunit 2 (mean weight, 0.210), IGHM (mean weight, −0.193), complement C1s subcomponent (mean weight, −0.181), and alpha-1-antichymotrypsin (mean weight, 0.168). Proteins among the highest-weighted 10% of features in both model 1a and model 3 (and weighted in similar directions) included A2M, IGHM, C4BPA, plasminogen, and C6.

Study 2: General Population Sample
The initial subsample was composed of plasma samples from 132 participants (65 case and 67 control samples). Eleven plasma samples were excluded because of poor protein identification profiles, resulting in 55 case and 66 control samples from 121 participants (61 [50.4%] male). Case samples were more likely to be from female participants. There was no evidence for differences in BMI, race/ethnicity, or maternal social class (eTable 14 in Supplement 1).

Differential Expression
Of 506 proteins identified, 265 were quantified in more than 80% of samples. There was nominally statistically significant (\( P < .05 \)) differential expression of 40 proteins at age 12 years (eTable 15 in Supplement 1), of which the following 5 remained statistically significant after FDR correction: C4BPA (ratio of means in PE vs no PE, 0.77), serum paraoxonase/arylesterase 1 (ratio of means, 0.80), IGHM (ratio of means, 0.78), inhibin beta chain (ratio of means, 1.31), and clusterin (ratio of means, 0.92).

Model 4: Predicting PEs Using Proteomic Data
An SVM model using 265 proteomic features from plasma samples obtained at age 12 years predicted PEs at age 18 years, with an AUC of 0.74 \( [ P < .001 \); sensitivity of 72.7%, specificity of 71.2%, PPV of 67.8%, and NPV of 75.8% (Table 2 and eFigure 8 in Supplement 1). For example, the 5 highest-ranked predictive features were C4BPA (mean weight, −0.227), serum paraoxonase/arylesterase 1 (mean weight, −0.218), complement factor H–related protein 1 (mean weight, −0.181), vitamin K–dependent protein S (mean weight, −0.130), and lysozyme C (mean weight, −0.145) (Table 3).
Supplementary Analyses

Model S1 used ELISA data to predict transition status in EU-GEI, with an AUC of 0.76 (P < .001). These results are summarized in Table 2 and eFigure 9 in Supplement 1.

Model S2 used clinical and proteomic data to predict poor (GAF disability subscale score ≤60) vs good (>60) functional outcome at 2 years in EU-GEI, with an AUC of 0.74 (P = .003) (Table 2 and eFigure 10 in Supplement 1). The 10% highest-weighted features are listed in Table 16 in Supplement 1.

There was evidence of differences for the clinical data between the London and the Netherlands sites compared with others (eTable 17, eFigure 11, and eFigure 23 in Supplement 1), likely because of group differences in age, years in education, and BPRS score (eMethods and eFigures 13-22 in Supplement 1). There was no strong evidence of systematic site associations for the proteomic data (eTable 18, eFigure 12, and eFigure 24 in Supplement 1).

Performance metrics of multivariate-corrected SVM models are listed in eTable 19 in Supplement 1. There were generally slight reductions in AUCs of the corrected models compared with their uncorrected counterparts (median change in AUC, 0.04; range, 0.01-0.10), although in all cases the 95% CIs overlapped.

### Table 3. Ten Percent Highest-Weighted Features for Model 1a, Model 3, and Model 4*

<table>
<thead>
<tr>
<th>Model/Feature</th>
<th>Mean weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1a: EU-GEI clinical and proteomic data, initial experiment, all sites</td>
<td></td>
</tr>
<tr>
<td>P01023 Alpha-2-macroglobulin</td>
<td>−0.330</td>
</tr>
<tr>
<td>P01871 Immunoglobulin heavy constant mu</td>
<td>−0.256</td>
</tr>
<tr>
<td>P04003 C4b-binding protein alpha chain</td>
<td>−0.161</td>
</tr>
<tr>
<td>P07357 Complement component 8 alpha chain</td>
<td>0.158</td>
</tr>
<tr>
<td>P50558 Phospholipid transfer protein</td>
<td>−0.146</td>
</tr>
<tr>
<td>O75636 Ficolin 3</td>
<td>−0.145</td>
</tr>
<tr>
<td>P02774 Vitamin D-binding protein</td>
<td>0.135</td>
</tr>
<tr>
<td>P07225 Vitamin K-dependent protein 5</td>
<td>−0.132</td>
</tr>
<tr>
<td>P43320 Beta-crystallin B2</td>
<td>0.132</td>
</tr>
<tr>
<td>P02766 Transthyretin</td>
<td>−0.130</td>
</tr>
<tr>
<td>P23142 Fibulin 1</td>
<td>0.125</td>
</tr>
<tr>
<td>P10909 Clusterin</td>
<td>0.121</td>
</tr>
<tr>
<td>P05155 Plasma protease C1 inhibitor</td>
<td>−0.114</td>
</tr>
<tr>
<td>Sex</td>
<td>−0.111</td>
</tr>
<tr>
<td>P00747 Plasminogen</td>
<td>0.111</td>
</tr>
<tr>
<td>P13671 Complement component 6</td>
<td>0.111</td>
</tr>
<tr>
<td>P02747 Complement C1q subcomponent subunit C</td>
<td>0.109</td>
</tr>
<tr>
<td>P02753 Retinol-binding protein 4</td>
<td>0.109</td>
</tr>
<tr>
<td>Q76LX8 A disintegrin and metalloproteinase with thrombospondin motifs 13</td>
<td>−0.108</td>
</tr>
<tr>
<td>P08697 Alpha-2-antiplasmin</td>
<td>−0.106</td>
</tr>
<tr>
<td>P19627 Inter-alpha-trypsin inhibitor heavy chain H1</td>
<td>0.105</td>
</tr>
<tr>
<td>MADRS: concentration difficulties</td>
<td>−0.104</td>
</tr>
<tr>
<td>P02489 Alpha-crystallin A chain</td>
<td>0.101</td>
</tr>
</tbody>
</table>

Model 3: EU-GEI clinical and proteomic data, replication experiment, all sites

<table>
<thead>
<tr>
<th>Model/Feature</th>
<th>Mean weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01023 Alpha-2-macroglobulin</td>
<td>−0.286</td>
</tr>
<tr>
<td>P22792 Carboxypeptidase N subunit 2</td>
<td>0.210</td>
</tr>
<tr>
<td>P01871 Immunoglobulin heavy constant mu</td>
<td>−0.193</td>
</tr>
<tr>
<td>P09671 Complement C1s subcomponent</td>
<td>−0.181</td>
</tr>
<tr>
<td>P10101 Alpha-1-antichymotrypsin</td>
<td>0.168</td>
</tr>
<tr>
<td>P00747 Plasminogen</td>
<td>0.163</td>
</tr>
<tr>
<td>P08571 Monocyte differentiation antigen CD14</td>
<td>0.161</td>
</tr>
<tr>
<td>P10909 Clusterin</td>
<td>0.158</td>
</tr>
<tr>
<td>Q16610 Extracellular matrix protein 1</td>
<td>0.157</td>
</tr>
<tr>
<td>G3XAM2 Complement factor I</td>
<td>0.140</td>
</tr>
<tr>
<td>P04003 C4b-binding protein alpha chain</td>
<td>−0.140</td>
</tr>
<tr>
<td>P13671 Complement component 6</td>
<td>0.132</td>
</tr>
<tr>
<td>P25311 Zinc alpha-2-glycoprotein</td>
<td>−0.131</td>
</tr>
<tr>
<td>P07359 Platelet glycoprotein Ib alpha chain</td>
<td>0.126</td>
</tr>
<tr>
<td>P01031 Complement C5</td>
<td>0.125</td>
</tr>
<tr>
<td>Q75882 Attractin</td>
<td>0.123</td>
</tr>
<tr>
<td>P00569 Immunoglobulin lambda constant</td>
<td>−0.120</td>
</tr>
<tr>
<td>P15169 Carboxypeptidase N catalytic chain (CPN)</td>
<td>0.115</td>
</tr>
</tbody>
</table>

Model 4: ALSPAC proteomic data

<table>
<thead>
<tr>
<th>Model/Feature</th>
<th>Mean weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>P04003 C4b-binding protein alpha chain</td>
<td>−0.227</td>
</tr>
<tr>
<td>P27169 Serum paraoxonase/arylesterase 1</td>
<td>−0.180</td>
</tr>
<tr>
<td>Q03591 Complement factor H-related protein 1</td>
<td>−0.152</td>
</tr>
<tr>
<td>P07225 Vitamin K-dependent protein 5</td>
<td>−0.145</td>
</tr>
<tr>
<td>P61626 Lysozyme C</td>
<td>−0.142</td>
</tr>
</tbody>
</table>

(continued)
Discussion

We described evidence of differential baseline plasma protein expression in individuals at CHR who developed psychosis compared with those who did not. Machine learning algorithms that incorporated clinical and proteomic data were used to predict transition outcome (AUC, 0.95). Proteomic features were of greater predictive value than clinical features. A parsimonious model based on 10 highly predictive proteins showed excellent performance in training (AUC, 0.99) and test (AUC, 0.92) data. Furthermore, a predictive model was developed using proteomic data at age 12 years for PEs at age 18 years in a general population sample (AUC, 0.74).

Although only 16% to 35% of individuals at CHR transition to FEP, the CHR state remains a strong risk factor. Clinical data have previously shown value for prediction of transition, and the poor performance of the clinical features in our study does not imply that clinical data in general are of little prognostic use. Previous studies have attempted to augment accuracy using neuroimaging and neurocognitive data, but blood-based tests have the advantage of greater accessibility. Perkins et al derived a panel of 15 proteins using immunoassays that distinguished between CHR-T and CHR-NT, with an AUC of 0.88. Chan et al used 22 blood-based biomarkers to predict schizophrenia onset, with an AUC of 0.82 that increased to 0.90 with incorporation of the CAARMS positive symptoms subscale. Our parsimonious model used data for 10 proteins, and, with further validation, may contribute to individualized prognosis and treatment stratification strategies.

eTable 20 in Supplement 1 summarizes our findings of differential expression in CHR-T vs CHR-NT and the predicted functional implications (modeled in eFigure 25 in Supplement 1). We found particularly strong evidence for dysregulation of the complement and coagulation cascade, previously implicated in schizophrenia. Similar processes have been previously implicated in proteomic studies of the development of PEs in the general population. Changes in the present CHR study that were consistent with results from these previous PE studies include increases in plasminogen, Clr, clusterin, and complement factor H and decreases in A2M and IGHM. The primary causes of these changes remain unknown but are consistent with evidence of enhanced inflammatory tone preceding psychosis and other mental disorders and schizophrenia risk associated with genetic variation of complement C4.

Several complement proteins emerged as important predictors of transition, including C4BPA, Clr of the antibody-antigen complex mediated pathway, key regulatory protease complement factor I, and terminal pathway components C6 and C8A. These arise from common pathways or functionally interact with coagulation proteins plasminogen and vitamin K-dependent protein S, supporting hypotheses of coagulation activation in psychosis. In both the initial and replication experiments, the most highly weighted predictor of transition was A2M (decreased in CHR-T vs CHR-NT), a protease inhibitor with diverse functions, including inhibition of proinflammatory cytokines such as interleukin 1β (consistently elevated in FEP). A2M is a key coagulation inhibitor and thus links functionally to our observations of elevated plasminogen in CHR-T. This finding is intriguing given the evidence that blood-derived plasminogen is associated with brain inflammation and complement activation. In models of multiple sclerosis, blood-brain barrier disruption facilitates transfer of fibrinogen into the brain, where it is deposited as fibrin, causing local inflammation. Given evidence for blood-brain barrier disruption in psychosis, fibrin may be associated with etiopathogenic mechanisms providing novel therapeutic avenues, but this hypothesis requires further investigation.

We validated differential expression of A2M and Clr using ELISA. The ELISA-based model (model S1) demonstrated fair, although reduced, predictive accuracy. This finding may reflect reduced sensitivity of ELISA and the inability to accurately quantify specific protein isoforms. Several proteins in the highest-weighted 10% of features for transition in study 1 were similarly highly weighted for PEs in study 2, including C4BPA, vitamin K-dependent protein S, A2M, and IGHM (eTable 21 in Supplement 1 summarizes the directionality of association of the 10% highest predictors in model 1a, model 3, and model 4). This observation may suggest a degree of similarity in proteomic changes between young people in the general population who develop PEs and help-seeking individuals at CHR who develop psychosis, but this hypothesis requires confirmation.

Outside of psychosis outcomes, several proteomic features contributed to prediction of functional outcome (model S2). A2M, IGHM, phospholipid transfer protein, and clusterin were among the 10% highest-weighted predictors. The results of the present study are also in keeping with studies in bipolar disorder and depression reporting decreased A2M, IgM, and C4BPA. At least some of these proteomic changes may be common to multiple clinical phenotypes, including neurodegenerative disorders, such as Alzheimer disease. Rather than considering such changes as biomarkers of individual disorders, phenotypic manifestations may be clinical markers of a variety of overlapping neuroimmune abnormalities that have their origin in combined genetic and environmental factors.

Limitations

This study has some limitations. First, these models require validation in independent cohorts to assess generalizability and real-world applicability. Second, differences in protein identifications precluded application of models between studies. However, there are valid reasons not to do so, including differences in outcome (psychotic disorder vs PEs) and age (postpubertal vs peripubertal). Third, data on duration of follow-up and reasons for dropout were not systematically collected in EU-GEI, and we were unable to fully assess the potential implications of these factors. Fourth, the replication experiment was partial because only 2 CHR-T cases were different from the initial experiment. Although our findings were generally replicated, no statement can be made regarding generalizability of model sensitivity. Fifth, participants were nonfasting, and there were no restrictions on time of sample collection. Sixth, other factors, such as childhood adversity, may have contributed to the proteomic changes that we observed, but these factors require further study.
Conclusions

We developed models incorporating proteomic data predicting transition to psychotic disorder in the CHR state. In a general population sample, several of the same proteins contributed to prediction of PEs. Further studies are required to validate these findings, evaluate their causes, and elucidate tractable targets for prediction and prevention of psychosis.

ARTICLE INFORMATION

Accepted for Publication: June 15, 2020.
Published Online: August 26, 2020.

Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2020 Mongan D et al. JAMA Psychiatry.

Author Affiliations: Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin, Ireland (Mongan, Föcking, Healy, Susai, Cannon, Cotter); School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff, United Kingdom (Heurich); School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Dublin, Ireland (Wynne, Cagney); Centre for Youth Mental Health, University of Melbourne, Parkville, Victoria, Australia (Nelson, McGorry, Amminger); Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark (Nordentoft); University Paris Descartes, Groupe Hospitalier Universitaire (GHU) Paris-Sainte Anne, Evaluation Centre for Young Adults and Adolescents (CJAAAD), Service Hospitalo-Universitaire, Institut National de la Santé et de la Recherche Medicale (INSERM) U1266, Institut de Psychiatrie (Centre National de la Recherche Scientifique [CNRS] 3557), Paris, France (Krebs); Department of Psychiatry, Medical Faculty, University of Basel, Basel, Switzerland (Riecher-Rössler, Borgwardt); LINC-Lab Interdisciplinary Neurocognitives Clinics, Dept Psiquiatría, Escola Paulista de Medicina, Universidade Federal de São Paulo (UNIFESP), São Paulo, Brazil (Bressan); Department of Psycologia Clinica i de la Salut (Universitat Autònoma de Barcelona), Fundació Sanitaria Sant Pere Claver (Spain), Spanish Mental Health Research Network (Centro de Investigación Biomédica en Red de Salud Mental [CIBERSAM]), Barcelona, Spain (Barrantes-Vidal); Department of Psychiatry and Psychotherapy, Translational Psychiatry Unit, University zu Lübeck, Lübeck, Germany (Borgwardt); Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany (Ruhmann); Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria (Sachs); Melbourne Neuropsychiatry Centre, Department of Psychiatry, University of Melbourne and Melbourne Health, Carlton South, Victoria, Australia (Pantelis); Faculty of Behavioural and Movement Sciences, Department of Clinical and EMGO+ Institute for Health and Care Research, Vrije Universiteit (VU) University, Amsterdam, the Netherlands (van der Gaag); Department of Psychosis Research, Parnassia Psychiatric Institute, The Hague, the Netherlands (van der Gaag); Academic Medical Centre (AMC), Academic Psychiatric Centre, Department Early Psychosis, Amsterdam, the Netherlands (de Haan); Institute of Psychiatry, Psychology & Neuroscience, Department of Psychology, King's College London, London, United Kingdom (Valmaggia); Institute of Psychiatry, Psychology & Neuroscience, Department of Psychiatry Studies, King's College London, London, United Kingdom (Pollak, Kempton, McGuire); Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, the Netherlands (Rutten); Trinity Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland (Whelan); Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, United Kingdom (Zammit); Bristol Medical School, University of Bristol, Bristol, United Kingdom (Zammit).

Author Contributions: Drs Cotter and McGuire are co-senior authors. Drs Mongan and Cotter had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Healy, Amminger, Krebs, Bressan, Borgwardt, Ruhmann, de Haan, Pollak, Rutten, Cannon, Cagney, Cotter, McGuire.


Drafting of the manuscript: Mongan, Föcking, Healy, Wynne, Cannon, Cotter, McGuire.


Statistical analysis: Mongan, Föcking, Healy, Cagney, Cotter.


Administrative, technical, or material support: Susai, Wynne, Riecher-Rössler, Bressan, Barrantes-Vidal, Borgwardt, Pantels, van der Gaag, Valmaggia, Pollak, Kempton, Cagney, Cotter, McGuire.

Supervision: Amminger, Nordentoft, Bressan, Borgwardt, Ruhmann, Sachs, de Haan, Cannon, Cagney, Cotter, McGuire.

Conflict of Interest Disclosures: A United Kingdom (UK) patent application has been filed in relation to the development of a prognostic test derived from this work (UK patent application 1919155.0). Dr Mongan received grants from the Wellcome Trust and Health Research Board Ireland and having UK patent application 1919155.0 pending. Mr Healy reported receiving grants from the European Research Council. Dr Krebs reported receiving grants from the French Ministry Programme Hospitalier de Recherche Clinique AOM07-118 and Elsevier, receiving grants and personal fees from Otsuka-Lundbeck and Janssen, and having a pending patent. Dr Borgwardt reported receiving grants from the European Community's Seventh Framework Programme under grant agreement HEALTH-F2-2010-241909 (Project EU-GEI). Dr Ruhrmann received grants from the European Commission and receiving nonfinancial support from Boehringer Ingeheim. Dr Sachs reported receiving honoraria for consulting and lectures on the topic of schizophrenia. Dr Pantelis reported receiving grants from the Australian National Health and Medical Research Council (NHMRC) and The Lundbeck Foundation and receiving personal fees from Lundbeck Australia Pty Ltd. Dr Kempton reported receiving grants from the European Commission and the Medical Research Council. Dr Cagney reported receiving grants from Health Research Board Ireland and having a patent for a biomarker panel pending. Dr Cotter reported receiving grants from Health Research Board Ireland and having UK patent 1919155.0 pending. No other disclosures were reported.

Funding/Support: EU-GEI was funded by a Framework 7 Grant (HEALTH-F2-2010-241909) for the European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) study and by Health Research Board Ireland through a Clinician Scientist Award to Dr Cotter. Additional support was provided by a Medical Research Council Fellowship to Dr Kempton (grant MR/J008915/1) and by the Ministerio de Ciencia, Innovación e Universidades (grant PS2017-87512-21-R) and Generalitat de Catalunya (grant 2017SGR1612 and Catalan Institution for Research and Advanced Studies [ICREA] Academia award) to Dr Barrantes-Vidal. The UK Medical Research Council and the Wellcome Trust (grant 102215/2/13/2 and the University of Bristol provide core support for the Avon Longitudinal Study of Parents and Children (ALSPAC). A comprehensive list of grant funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). The outcomes data collected in the ALSPAC study that were used in the present study were specifically funded by the Medical Research Council (grant G0701503/85179). Dr Zammit is supported by the Bristol National Institute for Health Research Biomedical Research Centre. Dr Mongan is a fellow of the Irish Clinical Academic Training (ICAT) Programme, which is supported by the Wellcome Trust and Health Research Board Ireland (grant 203930/B/16/2), the Health Service Executive National Doctors Training and Planning, and the Health and Social Care Research and Development Division, Northern Ireland.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Group Information: EU-GEI High Risk Study Group members are Philip McGuire, Department of Psychiatry, Psychology & Neuroscience, Department of Psychiatry Studies, King's College London, London, United Kingdom (Pollak); Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, the Netherlands (Rutten); Trinity Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland (Whelan); Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, United Kingdom (Zammit); Bristol Medical School, University of Bristol, Bristol, United Kingdom (Zammit).

jmapsychnology.com
Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Lucia Valmaggia, Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Matthew J. Kempleton, Department of Psychiatry Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Thomas A. Pollak, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Conrad jeghe, Department of Psychiatry Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Stefania Tognin, Department of Psychiatry Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Adonis Mendos, Department of Psychiatry Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London; Lieuwe de Haan, Academic Medical Centre (AMC), Academic Psychiatric Centre, Department Early Psychosis; Mark van der Gaag, VU University, Faculty of Science, Department of Clinical Psychology and EMGO+ Institute for Health and Care Research, and Parnassia Psychiatric Institute, Department of Psychosis Research; Eva Velthorst, AMC, Academic Psychiatric Centre, Department Early Psychosis, and Department of Medicine at Mount Sinai, Department of Psychiatry; Tamar C. Kraan, AMC, Academic Psychiatric Centre, Department Early Psychosis; Daniella S. van Dam, AMC, Academic Psychiatric Centre, Department Early Psychosis; Nadine Burger, Parnassia Psychiatric Institute, Department of Psychosis Research, Banaby Nelson, Centre for Youth Mental Health, University of Melbourne; Patrick D. McGorry, Centre for Youth Mental Health, University of Melbourne; G. Paul Amminger, Centre for Youth Mental Health, University of Melbourne; Christos Pantelis, Center for Neuropsychiatric Schizophrenia Research (CNSR) and Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research (CINS), University of Copenhagen, and University of Copenhagen, Faculty of Health and Medical Sciences, Department of Clinical Medicine, Athena Politis, Centre for Youth Mental Health, University of Melbourne; Neive Goodall, Centre for Youth Mental Health, University of Melbourne; Anita Riecher-Rössler, University of Basel Psychiatric Hospital; Stefan Borgwardt, University of Basel Psychiatric Hospital; Charlotte Rapp, University of Basel Psychiatric Hospital; Sarah Ittig, University of Basel Psychiatric Hospital; Enrich Stedrus, University of Basel Psychiatric Hospital; Renata Smieszkova, University of Basel Psychiatric Hospital; Rodriguez A. Bressan, LINC-Lab interdisciplinary Neuroradiology Clinics, Dept of Psychiatry, Escola Paulista de Medicina, Universidade Federal de São Paulo; Liu J., University of Connecticut, Department of Psychology, Psychology & Neuroscience, King’s College London; Anja F. Vreugdenhil, Department of Neurology, Università degli Studi di Milano; Maria Gabriela Benitez, Department of Psychiatry, Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; Anna Racopi, Department of Psychobiology and Psychology, Universidad Autónoma de Barcelona; Thomas R. Kwapil, Department of Psychology, University of Illinois at Urbana-Champaign; Manel Monsonet, Department of Psicología Clínica i de la Salut (Universitat Autònoma de Barcelona); Araceli Rosa, Departamento de Biología Evolutiva, Ecología y Ciencias Ambientales (Universitat de Barcelona); Oussama Keber, University Paris Descartes, Hôpital Sainte-Anne, C’JAAAD, Service Hospitalo-Universitaire, INSERM U894, Institut de Psychiatrie (CNRS 3557); Claire Daban, University of Paris, GHU-Paris, Sainte-Anne, C’JAAAD, Hospitalo-Universitaire Department SHU; Julie Bourgin, University of Paris, GHU-Paris, Sainte-Anne, C’JAAAD, Hospitalo–Universitaire Department SHU; Boris Chaumette, University Paris Descartes, Hôpital Sainte-Anne, C’JAAAD, Service Hospitalo-Universitaire, INSERM U894, Institut de Psychiatrie (CNRS 3557); Jacky van der Weele, University of Amsterdam, Department of Psychology and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre; Birte Glenthøj, CNSRandCINS, University of Copenhagen; Merete Nordentoft, Mental Health Center Copenhagen and CINS, Mental Health Center Glostrup, Mental Health Services in the Capital Region of Copenhagen, University of Copenhagen; Louise Birkedal Genthøj, Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, and CINS and CINS, University of Copenhagen; Birte Genthøj, CNS and CINS, University of Copenhagen; Dominika Gebhard, Department of Psychiatry, Psychology & Psychotherapy, University of Cologne; Gabriele Sachs, Department of Psychiatry and Psychotherapy, Medical University of Vienna; Iris Lasser, Department of Psychiatry and Psychotherapy, Medical University of Vienna; Bernadette Winklbaur, Department of Psychiatry and Psychology, Medical University of Vienna; Philippe A. Delepaul, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, and Mondriaan Mental Health Trust; Bart P. F. Rutten, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, and Jim van Os, Department of Psychiatry Studies, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, and Department of Psychiatry, UMC Utrecht Brain Center, Utrecht University Medical Centre; Additional Contributions: We thank Magda Hryniewiec, MSc (Royal College of Surgeons in Ireland), for technical assistance and contributions in preparing samples for mass spectrometry. She was not compensated for her contributions. We also thank the Mass Spectrometry Core Facility at the Conway Institute, University College Dublin, for support in the development of our proteomics workflows. We are extremely grateful to all the participants, clinical teams, and research staff who contributed to the EU-GEI project. We are extremely grateful to all the families who took part in the ALSPAC, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

REFERENCES

Research Original Investigation Development of Proteomic Prediction Models for Transition to Psychotic Disorder
Development of Proteomic Prediction Models for Transition to Psychotic Disorder

Original Investigation Research

p.p.b.-range mass accuracies and proteome-wide peptide identification rates, individualized functioning.


