

1 **Title: Prefrontal co-expression of schizophrenia risk genes is associated with treatment response**
2 **in patients**

3
4 **One Sentence Summary:** Schizophrenia risk genes co-expressed in the dorsolateral prefrontal cortex
5 are associated with clinical outcome in patients with schizophrenia.

6
7 **Authors:** Giulio Pergola^{1,†}, Pasquale Di Carlo^{1,2,†}, Andrew E. Jaffe^{2,3,4,5}, Marco Papalino¹, Qiang
8 Chen², Thomas M. Hyde^{2,7,8}, Joel E. Kleinman^{2,7}, Joo Heon Shin², Antonio Rampino^{1,10}, Giuseppe
9 Blasi^{1,10}, Daniel R. Weinberger^{2,6,8,9}, Alessandro Bertolino^{1,10,*}

10

11 **Affiliations:**

12 ¹Group of Psychiatric Neuroscience, Department of Basic Medical Sciences, Neuroscience and Sense
13 Organs, University of Bari Aldo Moro, Bari, Italy

14 ²Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, Maryland,
15 USA

16 ³Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore,
17 Maryland, USA

18 ⁴Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland,
19 USA

20 ⁵Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, USA

21 ⁶Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

22 ⁷Department of Neurology, Johns Hopkins School of Medicine, Baltimore,

23 ⁸Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore,
24 Maryland, USA

25 ⁹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore,
26 Maryland, USA

27 ¹⁰Azienda Ospedaliero-Universitaria Consorziale Policlinico, Bari, Italy

28 † Indicates equal contribution as first authors

29 * Corresponding author

30 Alessandro Bertolino, MD, PhD
31 Piazza G Cesare, 11
32 70124 Bari, Italy
33 alessandro.bertolino@uniba.it
34

35

36 **Abstract:** Gene co-expression networks are relevant to functional and clinical translation of
37 schizophrenia (SCZ) risk genes. We hypothesized that SCZ risk genes may converge into co-
38 expression pathways which may be associated with gene regulation mechanisms and with response to
39 treatment in patients with SCZ. We identified gene co-expression networks in two prefrontal cortex
40 *post-mortem* RNA sequencing datasets (total N=688) and replicated them in four more datasets (total
41 N=227). We identified and replicated (all p-values<.001) a single module enriched for SCZ risk loci
42 (13 risk genes in 10 loci). *In silico* screening of potential regulators of the SCZ risk module via
43 bioinformatic analyses identified two transcription factors and three miRNAs associated with the risk
44 module. To translate *post-mortem* information into clinical phenotypes, we identified polymorphisms
45 predicting co-expression and combined them to obtain an index approximating module co-expression
46 (Polygenic Co-expression Index: PCI). The PCI-co-expression association was successfully replicated
47 in two independent brain transcriptome datasets (total N=131; all p-values<.05). Finally, we tested the
48 association between the PCI and short-term treatment response in two independent samples of patients
49 with SCZ treated with olanzapine (total N=167). The PCI was associated with treatment response in the
50 positive symptom domain in both clinical cohorts (all p-values<.05).

51 In summary, our findings in a large sample of human *post-mortem* prefrontal cortex show that co-
52 expression of a set of genes enriched for schizophrenia risk genes is relevant to treatment response.
53 This co-expression pathway may be co-regulated by transcription factors and miRNA associated with
54 it.

55

56 **KEYWORDS:** Gene co-expression networks, dorsolateral prefrontal cortex, olanzapine, RNA
57 sequencing, schizophrenia.

58 **[Main Text:]**

59 **Introduction**

60 Schizophrenia (SCZ) risk is highly related to genetic factors, and specific risk loci have recently
61 been identified by the Psychiatric Genomics Consortium (PGC) (1). The discovery that at least 108
62 genetic loci are associated with the disease suggests that multiple biological processes may be involved
63 in SCZ, perhaps converging into one or few common pathways (high coherence), or distributed across
64 many pathways of genetic risk (low coherence) (2). The question of genetic risk coherence is an
65 important issue in SCZ research because the functional and clinical translation of PGC SCZ risk
66 variants remains modest when they are considered on their own or additively cumulated. For example,
67 currently available cumulative scores do not explain a large fraction of the variance in treatment
68 response and treatment resistance (3, 4).

69 The challenge of translating genetic risk into common pathways associated with clinical
70 predictions is compounded by the fact that we know the risk *loci*, but in only a minority of cases do we
71 know which *genes* within them are causally implicated in the disorder. PGC risk loci include many
72 genes and are proximal to many more, such that risk variants in the loci may theoretically impact
73 hundreds of genes (5); additionally, the effect of genetic variants in the PGC loci is not necessarily
74 restricted to proximal genes (6). Understanding the relationship between risk variants and genes
75 involved in the disorder may require identification of *common pathways and biological processes*
76 *involving genes located in multiple loci* – rather than considering only the genetic variants associated
77 with GWAS hits. In turn, discovering biological pathways that bring together multiple SCZ risk loci
78 will contribute to identify molecular elements, such as transcription factors and miRNA, that may
79 represent *nodes of risk convergence* by regulating diverse gene functions.

80 A basic principle of biology is that the expression of individual genes is often coordinated by
81 regulatory molecules resulting in the co-expression of gene networks (7). Therefore, gene co-

82 expression is a biological process possibly relevant to the convergence of SCZ risk into common
83 pathways that are associated with clinical translation of PGC loci. At least some of the PGC risk
84 variants control gene expression (8, 9). Recent evidence suggests that genes in the PGC SCZ loci co-
85 segregate into co-expression pathways (8) and genetic variation in such pathways is relevant to SCZ
86 phenotypes (10).

87 We hypothesized that *genes located in PGC SCZ risk loci may converge into co-expression*
88 *pathways* which, in turn, may reveal molecular elements potentially contributing to orchestrate genetic
89 risk into common biological pathways and ultimately clinical outcome. The translational relevance of
90 such co-expression pathways to SCZ can be validated in terms of their association with clinical
91 phenotypes in patients, including treatment outcome. However, gene set clustering in co-expression
92 networks is variable and methodologically complex (11), and thus requires transcriptome-wide
93 replication to be considered reliable. Fromer and coworkers (8) previously used RNA sequencing in
94 *post-mortem* prefrontal cortex to identify gene expression patterns potentially relevant to SCZ risk.
95 Here, we used RNA sequencing data from the two largest collections of *post-mortem* prefrontal cortex
96 currently available: the Lieber Institute for Brain Development repository (LIBD) (12) and the
97 CommonMind Consortium collection (CMC) (8). We identified gene co-expression networks by means
98 of Weighted Gene Co-expression Network Analysis (WGCNA) (13). After assessing network
99 preservation of the LIBD co-expression network in CMC, we focused on one module showing
100 overrepresentation of genes located in the PGC SCZ loci. We aimed to identify potential genetic
101 regulators of the loci and to assess clinical translation. In order to translate *post-mortem* data mining
102 into clinical phenotypes, we used common genetic variation, i.e., we identified co-expression
103 quantitative trait loci (co-eQTLs) (10) and combined them to obtain a numeric index approximating
104 network co-expression (14). Genetic variation in this gene set was associated with short-term treatment
105 response to olanzapine in terms of positive symptoms in the largest double-blind clinical trial openly

106 available to date with genome-wide genotyping (CATIE; N = 121) (15). We replicated the clinical
107 results in an independent dataset of 46 patients with SCZ treated with olanzapine in Bari, Italy (16).
108 The current work complements further reports on partially overlapping datasets which focused on
109 network approaches to identify potential novel drug targets (17).

110

111 **Results**

112 *Co-expression Network of human prefrontal cortex.*

113 We selected frontal cortex samples from 343 LIBD subjects and 345 CMC subjects. The sample was
114 filtered based on RNA Integrity Number (≥ 7.0), age range (17-86 years), ethnicity (African-American
115 and Caucasians), and diagnosis (LIBD: patients with SCZ = 143, healthy controls [HCs] = 200; CMC:
116 patients with SCZ = 166; HCs = 179; demographics in Table 1). Transcripts available in both datasets
117 with Reads Per Kilobase per Million (RPKM) $> .1$ mapped to 20,993 genes. After preprocessing (11),
118 we computed WGCNA, separately for patients with SCZ and HCs within each of the two datasets, and
119 derived network preservation statistics (Fig. S1) (18, 19). We found that all co-expression modules
120 showed moderate to strong preservation between patients with SCZ and HCs both in the LIBD and in
121 CMC datasets (all Z-summary scores ≥ 2 , Fig. S1). Since all modules were thus relatively preserved
122 between patients with SCZ and HCs across both datasets at the selected threshold (results in
123 Supplementary Materials, SM, and Fig. S1), we pooled data from patients and controls and identified
124 one network for the LIBD and one for the CMC datasets. The results reported in the manuscript refer to
125 this WGCNA with pooled patients with SCZ and HCs. In this WGCNA, we selected the LIBD network
126 as the reference and tested its preservation in the CMC network, which was successful (Fig. 1A-B;
127 additional details in the SM; Table S1 and S2; Fig. S2). The whole network identified in the LIBD
128 dataset with gene modules and connectivity statistics is available in Data file S1.

129

130

TABLE 1 ABOUT HERE

131

132

133 *Prioritization of modules relevant for SCZ.*

134 We prioritized modules in terms of their relevance for SCZ. To assess their relevance to diagnosis, we
135 tested whether any module eigengenes (the first principal component of gene expression in the
136 modules, abbreviated as *ME* in the following) were associated with diagnosis. Six *ME* were
137 significantly different between patients with SCZ and HCs in the LIBD network (Bonferroni-corrected
138 p-value < .05, Table S3), suggesting potentially different co-expression. However, none of these
139 associations were replicated in CMC (all uncorrected p-values > .05, Table S3). Having found no
140 replicable co-expression signature of diagnosis, we asked whether any of the modules included more
141 risk genes for SCZ than expected by chance. This way we ***directly tested the hypothesis that risk***
142 ***converges into co-expression pathways***. To detect modules in which PGC SCZ risk genes (n = 310;
143 gene list in the Table S4) were overrepresented, we computed hypergeometric tests and corrected the
144 results for multiple comparisons (Bonferroni-corrected p-value < .05). We found that the *Darkgreen*
145 module in the LIBD network was the only module significantly enriched for genes in the PGC SCZ
146 loci (10 loci, 13 genes, p-value = 3.1×10^{-5} , see Table 2, see Table S5 for the full list of *Darkgreen*
147 genes). Notably, the enrichment remained significant when including both protein-coding and non-
148 protein-coding genes located in the PGC loci (p-value = 5.7×10^{-4}), further suggesting that *Darkgreen*
149 co-expressed genes co-localized with genetic risk for SCZ. Next, we asked whether this enrichment
150 was affected by genetic spatial proximity. Our hypothesis was that the overrepresentation of SCZ risk
151 genes should remain significant when expanding the boundaries of the loci within a genomic distance
152 compatible with an influence of sequence elements on gene expression (20, 21). The enrichment
153 survived permutation-based empirical p-value < .001 when loci were expanded up to 450 kbp (Fig. 1C;
154 see also Fig. S3, which includes protein-coding and non-protein-coding genes), indicating that many
155 genes in the same loci were co-expressed in *Darkgreen*. Additionally, gene set ‘competitive’
156 enrichment analysis with the software MAGMA (22) demonstrated that variants falling within

157 *Darkgreen* were associated with greater SCZ risk compared to the remaining sets (we excluded the
158 *Grey* module of non-clustered genes; p-value = .036, see Methods and Table S1 for details). Hence,
159 converging evidence from the gene list and the localization of genetic variants suggested that genetic
160 risk for SCZ converged into *Darkgreen*. Moreover, *ME Darkgreen* was not associated with possible
161 biological confounders such as smoking habit, nor with antipsychotic or antidepressant medications in
162 SCZ patients (we used a binary classification of whether or not patients used the substances;
163 uncorrected p-value > .1; Data file S2 reports uncorrected p-value for all modules in the LIBD
164 network).

165

166 *Functional significance of Darkgreen module.*

167 *Darkgreen* included 225 genes, of which 157 were protein coding (Table S5). We investigated the
168 functional significance of *Darkgreen* by means of gene ontology analyses. *Darkgreen* was functionally
169 enriched for gene products involved in homophilic cell adhesion via plasma membrane (Amigo2,
170 GO:0007156, 9 genes, fold-enrichment = 7.92, Bonferroni-corrected p-value = .022). Specific
171 Expression Analysis [<http://genetics.wustl.edu/jdlab/csea-tool-2/>] (23) revealed that *Darkgreen* was
172 enriched for genes preferentially expressed in the cortex during young adulthood (24) (Fig. S4).
173 Therefore, we asked whether *Darkgreen* genes were also co-expressed during neurodevelopment, given
174 the importance of developmental ages for SCZ liability (25). WGCNA on a sample of 93 LIBD
175 subjects from fetuses to 16 year old individuals (hereinafter, LIBD developmental series) non-
176 overlapping with the sample used in the main analysis revealed higher than chance topological
177 preservation (empirical p-value < .001 (26); Table S6), showing that *Darkgreen* gene-gene
178 relationships were significant also in independent subjects during developmental life stages. We also
179 explored further datasets to assess the robustness of the gene-gene relationships detected in this

180 module. *Darkgreen* was among 12 modules preserved in all of the three additional frontal cortex
181 microarray and RNA sequencing datasets we analyzed, showing that the gene-gene associations we
182 identified were robust (Fig. 1D; empirical p-value < .001 (26); Table S6). Fig.1E represents the hub
183 genes of *Darkgreen* and their relationship with PGC SCZ hits included in the module.

184

185 *Genetic regulation potentially implicated in Darkgreen module co-expression.*

186 We hypothesized that co-expressed genes may be co-regulated by elements such as transcription
187 factors (TFs) and miRNA. We tested this hypothesis by investigating transcription factors targeting
188 *Darkgreen* genes. Using the software Pscan (<http://159.149.160.88/pscan/>)(27) we identified two TFs
189 (NRF1, KLF14) whose binding motif was overrepresented in the promoter regions of our co-expressed
190 genes (Bonferroni-corrected p-value < .05). Interestingly, seven out of 13 SCZ risk genes showed an
191 association with NRF1 (*GIGYF2*, *NDUFA6*, *SCAF1*, *CACNA1C*, *IGSF9B*, *TMX2*, *ANKRD44*); also
192 KLF14 had seven PGC risk gene targets (*SCAF1*, *ANKRD44*, *GIGYF2*, *TMX2*, *CACNA1C*, *IGSF9B*,
193 *AKT3*). However, the identified TF were related with several other modules (corrected p-value < .05;
194 NRF1 to 24 modules; KLF14 to 18 modules; Fig. S5), hindering conclusions about their specificity. It
195 is also possible that some TFs may exert their effects on multiple modules because of tissue expression
196 specificity or biological coherence of the identified modules.

197 Micro-RNAs (miRNAs) are also regulators of gene co-expression (28). Hauberg and coworkers (29)
198 have shown that the targetome of 10 miRNAs is enriched for SCZ risk variants. Here, we assessed the
199 overrepresentation of the targetome of each of these miRNAs in *Darkgreen*. We found that the targets
200 of three SCZ-related miRNA (miR-101, miR-374, miR-28) were overrepresented in *Darkgreen*
201 (Bonferroni-corrected p-value < .05; see Table S7 for further details), suggesting that these miRNAs

202 may plausibly promote the correlated expression of *Darkgreen* genes. Both miR-374 and miR-28
203 targets shared the same seven SCZ risk genes (*AKT3*, *ANKRD44*, *CACNA1C*, *PCDHA3*, *PCDHA4*,
204 *PCDHA5*, *PCDHA6*), while among miR-101 targets we found two risk genes (*ANKRD44* and *AKT3*).
205 To assess the specificity of these findings, we computed the enrichment for miRNA targetomes in all
206 other modules and reported uncorrected p-values in Data file S3. The identified miRNAs overlapped
207 with only few modules (miR-101/miR-374/miR-28 = 8/6/10 modules, corrected p-value < .05, Fig. S6),
208 suggesting some degree of specificity. Overall, these results are consistent with the idea that genetic
209 risk convergence in *Darkgreen* may be mediated by TFs and miRNAs.

210

211 *Overlap with genes regulated by antipsychotics*

212 We defined our network based on data from patients with SCZ and HCs. Since SCZ patients are
213 usually treated with antipsychotics, it can be hypothesized that drugs contributed to the aggregation of
214 genes into modules. In a recent study, Kim and coworkers (30) identified genes differentially expressed
215 in the striatum and in the whole brain of mice exposed to haloperidol vs. not exposed mice. We
216 computed for all modules the enrichment for the differentially expressed genes and found a single
217 module (*Brown*) enriched (11 genes, 16% of total differentially expressed genes, Bonferroni-corrected
218 p-value = .00447, Data file S4). Specifically, *Brown* was enriched for down-regulated genes (9 genes,
219 26.5% of down-regulated genes, Bonferroni-corrected p-value = .00411; Fig. S7). *Darkgreen* did not
220 show any significant overlap with haloperidol target genes, suggesting that its relevance for SCZ risk
221 genes was not a by-product of medication, at least to the extent that haloperidol is a representative
222 antipsychotic.

223

FIGURE 1 ABOUT HERE

224 TABLE 2 ABOUT HERE

225

226 *Polygenic Co-expression Index.*

227 To translate *Darkgreen* co-expression into clinical phenotypes, we generated an index predicting
228 *Darkgreen* co-expression based on the genetic background of each individual. We first identified single
229 nucleotide polymorphisms (SNPs) predicting co-expression (co-eQTLs) of the whole module and
230 generated a Polygenic Co-expression Index (PCI (10, 14)). We used a Robust Linear Model to assess
231 allelic dose effects on *Darkgreen ME* (which explained 28% of the variance in the LIBD dataset). The
232 linear model was adjusted for diagnosis, age, sex, RNA integrity (RIN), total RPKM mapped, total
233 RPKM mapped to mitochondrial DNA, and 10 genomic principal components accounting for
234 population stratification. With the aim of increasing our statistical power, we computed a meta-analytic
235 p-value for each SNP based on the effect size in the LIBD and CMC datasets (meta-analytic dataset;
236 overall, 688 subjects). We ranked SNPs based on their meta-analytic p-value and computed several
237 PCIs by adding one SNP at a time (SNPs weights are available in Table S8). Our purpose was to
238 identify an ensemble of SNPs affording prediction of co-expression (correlation between *Darkgreen*
239 *ME* and PCIs), rather than identifying single genetic variants associated with co-expression per se
240 (although it is noteworthy that the first ranked SNP, rs9836592, would survive Bonferroni correction
241 for multiple comparisons). To determine how many variants should be included in the PCI, we
242 replicated the association between *Darkgreen ME* and PCIs in two additional transcriptomic and
243 genomic datasets (BRAINEAC samples with RIN > 5.5, N = 38; LIBD developmental series samples
244 with RIN ≥ 7.0, N = 93) (31, 32)). The test sets did not affect the model at any stage, because both the
245 *ME* and the weights of the SNPs in the PCI were derived from the training sets. We found that all PCIs
246 including between 6 and 32 SNPs afforded significant predictive capacity in both datasets with an

247 effect size comparable between discovery and replication sets (BRAINEAC: p-value < .05, Fig. 2A-B;
248 LIBD developmental series: p-value < .05; Fig 2A and Fig. S8). Table 3 includes annotations of the
249 first 32 SNPs.

250 To study translational phenotypes in a clinical population, we performed a meta-analysis of the
251 BRAINEAC and the LIBD developmental series - both test datasets independent of the training sets -
252 to select the most reliable predictors of co-expression. Prediction strength reached a plateau between 14
253 and 17 SNPs, with no further improvement when more SNPs were added (Fig. 2C). Based on these
254 results, we used the PCIs including 14 to 17 SNPs as predictors of symptom improvement (positive,
255 negative, and general PANSS) in the CATIE clinical trial of antipsychotic efficacy.

256 FIGURE 2 ABOUT HERE

257 TABLE 3 ABOUT HERE

258

259 *Clinical study.*

260 We focused on patients treated with olanzapine because it showed the best response in the study (33)
261 and because we had a replication sample available undergoing the same treatment. The outcome
262 variable was percent change of symptom severity from baseline to one-month follow-up both in CATIE
263 and in UNIBA datasets. We computed a Robust Multiple Regression to assess the association with the
264 PCIs, controlling for age, gender, education level and ancestry (indexed using the first ten genomic
265 principal components). We corrected statistics for multiple comparisons using pACT (34). Table 4
266 illustrates the results. This correction procedure accounts for the high correlation between the
267 predictors and between the dependent variables. We found the most significant relationship between
268 the PCI-16 and positive PANSS improvement (corrected p-value = .033, partial- η^2 = .061; Fig. 3A),

269 which replicated in the UNIBA independent clinical sample (one-tailed $p = .0475$, $\text{partial-}\eta^2 = .067$;
270 Fig. 3B; Table 4). We assessed the biological significance of this set of 16 SNPs by interrogating
271 Haploreg v. 4.1. Haploreg tests the presence of genetic regulatory elements in a given SNP list (35, 36).
272 Our SNP list was specifically enriched for H3K27ac-H3K9ac marks in the dorsolateral prefrontal
273 cortex including Brodmann Areas (BA) 46 and 9 (Bonferroni-corrected p -value = .029). It is worth
274 mentioning that the LIBD RNA sequencing was obtained on BA 46 cortical tissue.

275

276

277 **Discussion**

278 We investigated the convergence of SCZ PGC loci into co-expression networks with the aim of
279 identifying a biological pathway of SCZ risk and regulatory elements associated with gene co-
280 expression that could be translated to the clinic. We identified a gene co-expression module enriched
281 for genes located in risk loci for SCZ. This finding was reproducible, as demonstrated by network
282 preservation and replicated topological overlap in four independent brain gene expression datasets.
283 Module genes were associated with potential gene expression regulation elements. Co-eQTLs
284 identified in 688 subjects were associated with short-term treatment response to olanzapine – a first line
285 antipsychotic - in patients with SCZ. These findings suggest a significant degree of coherence of SCZ
286 risk genes and co-expression partners that might be translated to the clinic.

287

288 Gene co-expression in schizophrenia

289 In the context of noncoding variation, which characterizes most GWAS significant SNPs and common
290 disorders, gene expression is likely the phenotype closest to DNA in which inter-individual differences
291 can be directly associated with genetic variation. The multifold preservation of the network is
292 important because one may expect that gene co-expression in patients with SCZ may be confounded by
293 state-related factors such as pharmacological treatment; instead, our results demonstrate that such state-
294 related factors did not dominate the topology of the network, which was replicated in three independent
295 datasets of non-psychiatric individuals of various ages totaling 227 subjects. Therefore, it is unlikely
296 that our results are biased because of the use of data from patients. Moreover, we failed to associate the
297 gene-gene relationships within *Darkgreen* with smoking or antipsychotic medication (90 patients were
298 treated and 50 showed no evidence of treatment with antipsychotics), though these phenomenological

299 factors are poorly quantified in *post-mortem* tissue. Notably, it is difficult to conclusively rule out the
300 effect of antipsychotic medication because the possible confounding effects of medication may depend
301 on the specific antipsychotic administered and on the dosage.

302 Jaffe and coworkers (12) have suggested that preprocessing RNA data controlling for hidden RNA
303 quality is a key factor affecting the inferences drawn from transcriptome studies and the topology of the
304 network we report here holds also when preprocessing data with the conservative approach they
305 described (Fig. S11). In summary, the network we identified and validated in the largest sample tested
306 to date (including data from overall 915 *post-mortem* samples) is robust in terms of reproducibility and
307 highlights gene-gene relationships revealing non-random clustering of SCZ risk genes.

308

309 The schizophrenia risk co-expression module

310 Gene ontology analysis revealed involvement of *Darkgreen* genes in cell-cell adhesion, a biological
311 process previously associated with risk for SCZ and bipolar disorder (37, 38). It should be noted that
312 we selected genes expressed in the brain, whereas ontologies were not filtered in the same way. This
313 implies that our approach was conservative and more biological functions than currently detected may
314 be shared by these genes. Interestingly, the same gene ontology characterized differentially expressed
315 genes in induced pluripotent stem cell-derived differentiated neurons, in a recent study comparing
316 populations of monozygotic twins with discordant response to clozapine in treatment-resistant SCZ
317 (39). Taken together, both findings highlight the potential importance of the genes co-expressed in
318 *Darkgreen* for the physiology of olanzapine and clozapine, two atypical antipsychotics. *Darkgreen*
319 included also genes coding for proteins involved in synaptic transmission mediated by serotonin,
320 glutamate and GABA (*HTR1F*, *GRM5*, *GABRB1*, *GABRG3*), or involved in neural excitability

321 (*KCNH1*, *KCNA3*, *KCNH7*, *KCNH5*), along with *CACNA1C*, a risk gene for SCZ and bipolar disorder
322 supported by multiple lines of evidence (40-44). The functions of the genes in *Darkgreen* are consistent
323 with previous pathway analyses of SCZ risk (45) and enhance the biological plausibility that the co-
324 regulation of this module has functional relevance.

325 Although co-expression does not necessarily imply gene co-regulation, it is noteworthy that the 13
326 PGC hits of *Darkgreen* are distributed across 10 different loci, rather than encompassing a single locus
327 that is co-transcribed because of genetic proximity (20, 21). This finding suggests that there may be
328 co-regulators of these 10 loci, which we attempted to identify via bioinformatics analyses. The findings
329 that promoter sequences of *Darkgreen* genes were enriched for the target sequence of two transcription
330 factors (NRF1 and KLF14) and for the targetome of three miRNAs previously associated with SCZ
331 (miR-101, miR-374, miR-28) represent a potentially relevant clue about regulatory elements and target
332 sequence patterns potentially implicated in the co-regulation of SCZ risk genes. However, the
333 association of TFs and, to a lesser extent, of miRNAs to other modules hinder conclusions about
334 specificity. The targetome of genetic regulatory elements is generally larger than the size of a single co-
335 expression module and may therefore be associated with multiple gene sets. Furthermore, the role of
336 these regulatory elements in neurodevelopment (46-48) and the significant preservation of *Darkgreen*
337 topology in very young subjects is consistent with the hypothesis that SCZ risk genes are coordinated
338 by processes relevant to neurodevelopmental trajectories.

339

340 Genetic variants associated with co-expression of schizophrenia risk genes

341 Since it is not possible to directly assess gene expression in the living human brain, it is of interest to
342 translate models of gene co-expression into genetic variants (co-eQTLs) which index co-expression in

343 living individuals. The co-eQTLs detected here merit further investigation as potential indicators of loci
344 affected by genetic regulatory elements associated with positive symptoms and their clinical course.
345 For example, the first ranked SNP, rs9836592, has been associated with risk for bipolar disorder (49),
346 another disorder frequently treated with antipsychotic drugs such as olanzapine. Furthermore, this SNP
347 has been already associated with the regulation of gene expression (49) and is an eQTL for *CACNA1D*
348 (9). Moreover, the entire set of 16 SNPs was enriched for histone acetylation marks. Previous evidence
349 supports the relevance of histone modification pathways to SCZ (45) and the specific role of H3K27ac
350 markers in autism (50) a neurodevelopmental disorder sharing some genetic risk with SCZ (51). The
351 clinical evidence obtained in two independent samples supports the functional role of these SNPs in the
352 clinical treatment of SCZ.

353

354 Clinical translation of transcriptome data mining

355 We found that the PCI computed using the genetic variants above described was reproducibly
356 associated with treatment response to olanzapine. On the one hand, this finding suggests that the 13
357 PGC hit genes co-expressed in *Darkgreen* are candidates within their loci for mechanistic
358 interpretations of response to treatment. On the other hand, the PCI indexes a wider group of genes,
359 going beyond the 13 PGC hits, suggesting a broader transcriptomic landscape of risk and more relevant
360 here, of the biology of treatment response. Such landscape stratifies patients with SCZ in terms of
361 treatment response even though *Darkgreen* co-expression is not reproducibly associated with diagnosis
362 and the PCI variants per se are not associated with diagnosis. Another implication of the present
363 findings is that antipsychotic efficacy may involve many more genes than those coding for the
364 traditional targets, e.g., dopamine and serotonin antagonism, and may depend on the convergence in
365 terms of genetic regulation of multiple neural transmission systems, including glutamate and GABA

366 receptors, as well as calcium and potassium channels. This possibility is implicit in the fact that
367 dopamine and serotonin are engaged in tuning glutamate and GABA neuronal activity in cortex (52,
368 53).

369 Our findings further suggest a link of SCZ risk loci and their molecular interactors with inter-individual
370 variation in response to treatment with olanzapine selectively in terms of positive symptoms domain,
371 despite the differences between the clinical datasets we used. The current evidence is limited by the
372 relatively restricted sample size in the clinical groups (total N=167) and by the modest size of the
373 clinical effects. Therefore, this evidence awaits further independent replications in larger clinical
374 samples. However, this clinical translation is promising with respect to the feasibility of patient
375 stratification based on biological measures, in line with dimensional views of the diagnosis of SCZ (6,
376 54-57).

377 This study demonstrates the potential for co-expression genetic studies to be translated in the clinic.
378 However, several limitations suggest caution. First, while WGCNA is a flexible and extensively used
379 tool, gene co-expression network analyses can be implemented with different methodological
380 nuancing. For example, reproducible gene-gene relationships can be reflected in different gene
381 clustering across datasets and studies. Second, a large portion of the variance in treatment response
382 remains unexplained (> 90%), suggesting the potential role of other factors not assessed here. Large
383 datasets including longitudinal clinical information, genome-wide genotyping, along with brain
384 imaging data and environmental variables, may bring us closer to the clinical utility of this work (58).
385 Third, the role of potential regulators of gene co-expression requires biological evidence to offer
386 mechanistic explanation of how their targets are related with response to olanzapine. Addressing these
387 limitations will be necessary steps to more routinely apply genetic screening in the clinic.

388 Nevertheless, this work demonstrates that a proportion of SCZ risk genes converge into gene co-

389 expression networks and provides information on potentially relevant molecules implicated in this
390 process. The findings offer a proof of concept that translation of genetic risk into clinical information
391 requires the study of multiple levels of biological organization, starting from the very beginning of the
392 information flow from DNA to phenotypes, i.e., gene expression.

393

394 **Material and Methods**

395 *Study design*

396 Table 1 summarizes the demographic data and relative statistics for the subjects included in all
397 experiments. For the co-expression network study, we used RNA sequencing data from the LIBD (12)
398 and from the CMC (8) *post-mortem* series for a transcriptome-wide WGCNA (13). Both datasets
399 included *post-mortem* mRNA expression levels of HCs and patients with SCZ in the human prefrontal
400 cortex, whereas the three additional datasets used for replication included only non-psychiatric
401 individuals (31, 59). Additionally, the LIBD dataset included toxicological tests performed on frozen
402 *post-mortem* tissue long after death. Smoking habit was assessed based on nicotine and cotinine
403 quantification, as well as on reports from familiars. Drug consumption assessment, particularly
404 regarding antipsychotics and antidepressants, has been recorded as a yes/no variable. Permission to use
405 *post-mortem* brain materials was obtained by the next of kin (see the original reports for further
406 information). We selected subsets of individuals in the LIBD and CMC datasets to match possible
407 confounding variables across the datasets as closely as possible. Therefore, we included samples with
408 age ≥ 17 years of Caucasian or African American ancestry, RNA integrity number (RIN) ≥ 7.0 . We
409 used χ^2 tests to assess the effects of gender, ethnicity and diagnosis between datasets and a two-sample
410 t-test to assess the effect of age.

411 In the clinical studies, all participants provided written informed consent following the guidelines of
412 the Declaration of Helsinki after receiving a complete description of the study. Protocols and
413 procedures were approved by the ethics committee of the University of Bari (UNIBA) and by the
414 institutional review board of each clinical site involved in the CATIE program. Diagnosis of
415 Schizophrenia was established via Structured Clinical Interview for DSM-IV-TR (SCID). Symptom
416 severity was assessed with Positive and Negative Syndrome Scale (PANSS) (60) at study entry and at

417 several follow-up visits. The first clinical cohort included patients recruited in CATIE study by the
418 NIMH and treated with olanzapine (N = 121) (33). The second cohort included 46 patients recruited
419 from the region of Apulia, Italy, also treated with olanzapine in monotherapy (16). Study protocols and
420 exclusion criteria are available in SM Materials and Methods.

421

422 *Co-expression Network of human prefrontal cortex*

423 We processed RNA sequencing raw data as previously described (12) (SM Materials and Methods).
424 We selected 20,993 Ensembl ID transcripts with median Reads Per Kilobase per Million mapped reads
425 (RPKMs) > 0.1 in both the LIBD and CMC datasets. We log₂-transformed RPKMs values with an
426 offset of 1, e.g., log₂(RPKM+1). RNA expression data are affected by systematic noise, e.g., as a
427 consequence of batch effects. We used the Remove Unwanted Variation (RUV) tools (RUVcorr R
428 Bioconductor package) developed by Freytag et al. (11) to model systematic but latent sources of noise,
429 without explicitly modeling nuisance covariates (61-63). RUV capitalizes on the putatively low
430 physiological variation of housekeeping genes (HK). Therefore, variation in HK expression may reflect
431 more closely systematic noise than inter-individual variability. This version of RUV was specifically
432 designed to correct the signal prior to WGCNA (SM Material and Methods).

433 WGCNA (13, 64) uses gene-gene Pearson's correlation indices as a continuous, i.e., weighted, measure
434 of gene-gene relationships. We computed unsigned networks, i.e., negatively correlated genes are
435 considered connected rather than non-connected (65). The correlation matrix was transformed into an
436 adjacency matrix by raising Pearson's coefficients to a positive exponent, β , which is chosen to meet
437 the "scale-free" power law connectivity distribution. Scale invariance is widely considered a common
438 organization feature of cellular functions (66). A hierarchical clustering method was then used to group

439 genes into clusters, called “modules” (SM Materials and Methods). Colors were used to arbitrarily label
440 co-expression modules, with the “grey” module representing genes that did not cluster into any
441 particular module. Co-expression was summarized by the *ME*, the first principal component of the
442 expression of genes in any given module. A unique *ME* was computed for each module.

443 First, we computed separate co-expression networks for patients with SCZ and HCs within each of the
444 two datasets. To identify possible differences in network topology between patients and controls, we
445 employed the same β value for all datasets, because this parameter affects mean network connectivity
446 ($\beta = 6$ was the minimum value that satisfied the scale invariance criterion for all datasets, which fits
447 well with the authors’ suggestions for unsigned networks; for signed networks higher exponents are
448 generally needed, e.g., $\beta = 12$). We used the methods described by Langfelder et al.(18) and by Johnson
449 et al.(26) to compare graph properties using permutation approaches. These procedures are
450 complementary because the first relies on evaluation of graph parameters, while the second entails an
451 empirical, parameter-free test. The preservation technique published by Langfelder and coworkers (18)
452 uses connectivity and density to derive a summary score that characterizes optimal preservation with Z
453 ≥ 10 , partial preservation with $2 \leq Z < 10$, and no preservation with $Z < 2$ (1,000 permutations). The
454 technique developed by Johnson et al. (26), instead, assesses whether the topological relationships
455 between genes in the second dataset mirror those of the first dataset at a level greater than chance.
456 Therefore, for each module we computed the median of its topological overlap matrix and compared
457 this value against the null distribution of medians computed on random modules of identical size. We
458 used 10,000 re-samplings and a threshold for replication significance of empirical p-value $< .001$.

459 Network statistics showed strong preservation between HCs and patients with SCZ within each dataset.
460 Moreover, we used Wilcoxon signed rank test to demonstrate that preservation statistics (Z-values)
461 were greater between groups within the same dataset than between the same group across the two

462 datasets (SM Results, Fig. S1). Based on these results, we adopted an alternative approach. We pooled
463 data from patients with SCZ and HCs and identified one network for the LIBD and one for the CMC
464 datasets, allowing greater statistical power for the next steps of the analysis. All the following analyses
465 used the LIBD network with pooled patients with SCZ and HCs as the reference set. The minimum
466 value of β that satisfied scale invariance criterion both in the LIBD and in the CMC datasets was 5.
467 This network was comprised of 43 modules, with 6,706 transcripts falling in the grey module, i.e., not
468 clustered (Data file S1). These modules were strongly preserved in CMC (Fig. 1a-b, Table S1).
469 Importantly, the “gold” module, i.e., a random module whose size was defined as equal to the median
470 of the sizes of all modules, showed the lowest Z preservation statistic (Fig 1a). We cross-checked
471 preservation using CMC as the reference (57 modules, grey: 7,228 transcripts; also in this case, all
472 modules had $Z \geq 2$; Table S2).

473 We assessed the association of the LIBD Module Eigengenes (ME_{LIBD}) with case-control status (i.e.,
474 HCs vs. patients with SCZ) with a Robust Linear Model with the *lmRob* function of the *robust* R
475 package. We used Bonferroni correction for multiple comparisons (corrected p-value < .05). We
476 introduced observed demographics, RNA quality feature and RNA sequencing coverage as covariates
477 since they may potentially affect gene expression measures. The model accounted for age, gender, RIN,
478 total reads mapped, total reads mapped at mitochondrial DNA and 10 genomic ancestries as covariates
479 to account for potential genetic stratification (SM Materials and Methods). Then, we replicated the
480 findings in the CMC dataset. In order to obtain factor scores in the CMC network (ME_{SCMC}) for each
481 corresponding LIBD module, we computed the factor loadings for each ME_{LIBD} . Factor loadings
482 express the weighted contribution of each gene in the module to the ME . Then, we projected factor
483 loadings into the corresponding CMC gene expression data to obtain projected- ME_{SCMC} . Thus, we
484 evaluated the replication of the association between co-expression ($projected-ME_{SCMC}$) and case-control
485 status in the CMC dataset (p-value < .05).

486 In order to address the effect of potential confounders on the identified network, we employed Robust
487 Linear Models to assess the association between each ME_{LIBD} and nicotine, cotinine and smoking status
488 separately. Moreover, we evaluated the association between each ME_{LIBD} with antipsychotics and
489 antidepressants in the SCZ group ($\alpha = .1$; Data file S2).

490

491 *Co-expression Network replication*

492 We used several datasets of *post-mortem* brain samples to validate the LIBD modules through the
493 above-mentioned permutation procedure (26).

494 i) The LIBD Developmental Series (spanning ages from fetal to adolescent);

495 ii) The CMC dataset (as already described);

496 iii) The BRAINEAC Frontal Cortex dataset (31);

497 iv) The GTEx Brain Cortex dataset (59, 67);

498 v) The GTEx Frontal Cortex Brodmann Area 9 dataset (59, 67).

499 Datasets i-ii) were pre-processed as described in the pre-processing section. Dataset iii) is publicly
500 available at <http://www.braineac.org/>. Microarray expression data were downloaded and pre-processed
501 through RUV tools, selecting the same parameters used for the LIBD and CMC datasets ($k = 5$).
502 Datasets iv-v) are available at <https://www.gtexportal.org/home/>. RNA sequencing data have been
503 downloaded in the already pre-processed release format (GTEx Analysis V6p), with the aim to test for
504 module replication regardless of the pre-processing pipeline.

505

506 *Prioritization of modules relevant for SCZ*

507 The enrichment analysis is used to characterize the functional profile of gene sets identified *a priori*. It
508 consists in identifying over-represented gene classes within another gene set.

509 (1) We investigated the overlap between the LIBD modules and genetic association with SCZ (1). We
510 referred to genes identified by the PGC study (n = 310 genes were included in the network based on
511 transcript expression levels, Table S4) and used a hypergeometric test to assess the significance of the
512 over-representation in each module. We selected the modules surviving Bonferroni correction for
513 multiple comparisons (number of modules = 43, p-value = $.05/43 = .00116$). Moreover, we conducted
514 hypergeometric tests at multiple levels of PGC loci expansion (from ± 50 kbp to ± 10 Mbp) to investigate
515 the range of the gene-gene interactions potentially involved in the convergence of SCZ risk genes. We
516 used the biomaRt R package (68) to select protein coding genes located within the expanded PGC loci.
517 Finally, we derived an empirical p-value through a permutation approach ($p < .001$, SM Materials and
518 Methods). Since restricting the analysis to protein coding genes may bias results because network
519 analysis encompasses different gene biotypes (protein-coding and non-protein-coding), we repeated the
520 same analysis also including all the genes located in the expanded PGC loci, regardless of gene biotype
521 (Fig. S3).

522 (2) In addition, we explored the enrichment for common SCZ variants. We used summary statistics of
523 9.4 million SNPs from the largest GWAS in SCZ by PGC (1) publically available
524 (<http://www.med.unc.edu/pgc/results-and-downloads>) and excluded the MHC region on chromosome 6
525 because high LD in this locus could bias gene set enrichment statistics as already suggested by other
526 authors (45). We used Multi-marker Analysis of GenoMic Annotation (MAGMA)(22) to perform a
527 gene-set competitive enrichment analysis adjusted for confounding variables (SI Materials and

528 Methods). The significance threshold was set at the nominal p-value < .05 because we were only
529 interested in modules that already showed a significant overrepresentation of SCZ risk genes.

530

531 *Functional enrichment analyses*

532 We used Amigo2 (<http://amigo2.geneontology.org/amigo>, Gene Ontology database released 2017-06-
533 09) online available tools to perform functional enrichment analyses of the *Darkgreen* module, which
534 was selected based on the overrepresentation of SCZ risk genes. We listed *Darkgreen* protein coding
535 genes and performed online automatic searches in the Gene Ontology Database Released on 2017-06-
536 29 with the PANTHER Overrepresentation Test (release 2017-04-13). Furthermore, we used Specific
537 Expression Analysis (SEA) software [<http://genetics.wustl.edu/jdlab/csea-tool-2/>] (23) to track cell-
538 and tissue-specific expression pattern during neurodevelopment (Fig. S4).

539

540 *Enrichment analysis of Transcription Factor Binding Sites*

541 We used Pscan, a freeware web interface (<http://159.149.160.88/pscan/>)(27) to scan promoter regions
542 of our co-expressed genes looking for binding specificity of known Transcription Factors (TF). We
543 referred to the JASPAR 2016 (69) database of TF binding profiles and defined the promoter regions
544 spanning 1,000 bp upstream the transcription starting site by selecting these options from the web
545 interface. We scanned 472 different TF binding domains. We considered statistically significant TFs
546 surviving Bonferroni correction for multiple comparisons (corrected p-value < .05). Then, we explored
547 the contribution of single genes to the selected TFs and reported SCZ risk genes contained in
548 *Darkgreen* related with the TF more strongly than genome-wide average for the same TF (27)). Finally,

549 to explore the specificity of our findings, we evaluated the enrichment of all the other modules and
550 reported corrected p-values (Fig. S5).

551

552 *Micro-RNA target prediction*

553 We investigated the overlap between SCZ related miRNA targetomes (29) and *Darkgreen*, with the
554 purpose to identify specific regulatory elements of co-expressed genes. We used four miRNA target
555 repositories to obtain different lists of targets for each miRNA family (i. TargetScan v7.1,
556 http://www.targetscan.org/vert_71/ (70); ii) MirTarget, <http://www.mirdb.org/>(71); TargetMiner,
557 http://www.isical.ac.in/~bioinfo_miu/targetminer20.htm(72) and TarBase V7.0(73)). Then, we
558 performed a hypergeometric test for over-representation of miRNA targets in *Darkgreen* and combined
559 p-values with sum-log Fisher's method across different lists for each miRNA family. The corrected
560 significance threshold for the combined p-values was set to p-value = .00125, after having applied
561 Bonferroni correction (10 miRNA families tested times 4 tools used). We inspected targetomes
562 overlapping with *Darkgreen* and reported SCZ risk genes available in at least one gene list (Table S7).
563 Finally, we investigated the specificity of these enrichments by extending the same analysis to all the
564 LIBD modules and miRNA families (Data file S3). We expected that each miRNA would be associated
565 with only few modules. We showed results corrected for multiple comparisons (Bonferroni rule,
566 number of modules = 43; Fig. S6).

567

568 *Overlap with genes regulated by haloperidol*

569 We investigated the overlap between putative antipsychotics target genes and *Darkgreen* as well as all
570 other modules. We used lists of genes differentially expressed (DEG) between haloperidol-treated mice

571 and the control mice (30). We used lists of DEG at q-value < .05 in the striatum and in the whole brain
572 of mice (30). Moreover, we separately tested up- and down- regulated genes. We employed biomaRt R
573 package (68) to convert mouse genes into human orthologs. We performed hypergeometric test for
574 over-representation of haloperidol targets in network modules and used Bonferroni correction for
575 multiple comparisons (number of modules = 43; Fig. S7).

576

577 *Meta-analysis of co-expression quantitative trait loci*

578 SNP genotyping procedures and genotype imputation have been described previously for LIBD (9),
579 CMC (8), BRAINEAC (31), CATIE (33) and UNIBA (10) subjects (also see SM Materials and
580 Methods). We selected SNPs in the genes encompassed in *Darkgreen*, expanded by 100 kbp up- and
581 down-stream gene start and end, consistent with previous studies (10, 14). We employed a relatively
582 conservative extension of the genes because with larger flanks, e.g., 500 kbp to 1 mbp, the SNP sets
583 would largely overlap between modules. We selected SNPs with $MAF \geq 0.1$ because the sample size
584 was too limited to investigate uncommon variants and pooled minor allele carriers when $MAF \leq 0.15$
585 to avoid biasing estimations of population variance with small genotypic groups. These filters resulted
586 in 52,198 SNPs available in both the LIBD and the CMC datasets that we selected for further analyses.

587 We aimed to identify an ensemble of SNPs that, together, could predict gene co-expression (co-
588 eQTLs). The biological plausibility of co-eQTLs is supported by findings that only 30% of mRNA
589 expression heritability is associated with *cis*-active elements (74), suggesting a role of distant
590 regulatory elements and possibly *trans*-elements in heritable mRNA expression. With this purpose, we
591 investigated the association between the *Darkgreen* co-expression module summarized by the *ME*-
592 *Darkgreen* (see section 2.3) and SNP allelic dosage. We used a Robust Linear Model to estimate the

593 effect of the SNP allelic dosage separately in the LIBD and the CMC datasets with the *lmRob* function
594 of the *robust* R package. We included diagnosis (HCs vs. patients with SCZ), age, gender, RIN, total
595 count of mapped reads, total count of mitochondrial mapped reads and 10 ancestries as covariates.
596 Notably, co-varying for diagnosis allowed us to detect markers of co-expression valid both in patients
597 and controls rather than risk markers for SCZ. Finally, we performed a fixed-effect meta-analysis over
598 the two datasets with the *rma.uni* function of the *metaphor* R package, using partial correlation
599 coefficients of allelic dosage as an estimate of effect size. Then, we ranked SNPs according to their
600 meta-analytic p-value.

601 Following previous work on polygenic summaries of additive genetic effects (*1, 10*), we restricted the
602 analysis to independent SNPs. In this perspective, we evaluated pair-wise R^2 between SNPs within 250
603 kbp. We considered two SNPs independent when $R^2 < 0.1$ (*1*). We then performed a priority LD
604 pruning by iteratively discarding the SNP with the weaker association. We used this procedure to
605 enrich our selection for relevant variants (for further applications of a similar procedure see
606 <http://prioritypruner.sourceforge.net/documentation.html>). The final selection included 2,266 tagging
607 SNPs with negligible residual interdependence. We used the top 100 ranked co-eQTLs for the PCI
608 computation.

609

610 *Polygenic Co-expression Index*

611 We employed a previously published procedure based on Signal Detection Theory to assign weights
612 (A') to each SNP genotype (*10, 14*) (SM Materials and Methods). For each genotypic population of
613 each of the 100 top-ranked SNPs, we computed the A' weights separately in the LIBD and the CMC
614 datasets. Then, we averaged the weights across the two datasets (Table S8). We defined the PCI as the

615 average of A' values corresponding to all the genotypes of each subject. In this way, the PCI could be
616 interpreted as the genetically indexed inter-individual variability associated with gene co-expression
617 measured by Darkgreen Module Eigengene (*ME-Darkgreen*). The PCI is positively correlated with
618 *ME-Darkgreen* and is not confounded by ethnicity (SM Materials and Methods and Fig. S9).

619 A relevant issue is how many SNPs need to be included in the PCI. Including too few SNPs may not
620 afford sufficient predictive power, while too many SNPs may yield overfitting effects on the positive
621 correlation between the PCI and the *ME-Darkgreen*. To identify a SNP ensemble with significant
622 predictive power, we computed 100 different PCIs with an increasing number of SNPs (the first PCI
623 included just the first ranked co-eQTL, the second PCI included the first and the second co-eQTL, and
624 so on up to the 100th co-eQTL) and assessed the Pearson's correlations between the PCIs and the *ME-*
625 *Darkgreen* both in the LIBD and the CMC datasets. In case of overfitting, the effect size of the
626 correlation PCIs-*ME* in the discovery sets should monotonically increase when more SNPs are added,
627 whereas the effect size in the replication datasets should reach a plateau and then decrease (Fig. 2A).
628 We assessed the statistical significance of the PCI-*ME* correlation in the two independent replication
629 sets also via a permutation approach (p-value < .05, SM Materials and Methods).

630 Then, we performed a fixed-effect meta-analysis separately on the discovery and replication datasets.
631 In this way, we estimated global replication effect sizes using PCI-*ME* correlation coefficients (Fig.
632 2C). In order to identify the best set of predictors for the clinical study, we selected PCIs based on the
633 largest replication effect size. We started to include PCIs at the beginning of the plateau and stopped
634 when the effect size reached the absolute maximum and then started to decline (a possible effect of
635 overfitting; Fig. 2C).

636

637 *Clinical study*

638 We used two samples of patients with SCZ treated with olanzapine to assess the association between
639 the PCIs and the clinical outcome measured with the PANSS. Clinical outcome was defined as the
640 difference between baseline and early clinical response (one month) relative to baseline symptoms in
641 PANSS sub-scales and total scores. Patients were genome-wide genotyped (SM Materials and
642 Methods) and SNP genotypes were used to compute PCIs for each patient. We tested the association
643 between clinical outcome and PCIs through a Robust Linear Model using age, gender, education level
644 and ten genomic PCs as nuisance covariates. The CATIE cohort was used as discovery sample and
645 results were corrected for multiple comparisons, i.e. the multiple clinical subscales and PCIs we tested
646 (corrected p-value < .05). Due to the high correlation among the set of predictors – the PCIs – and
647 among the set of outcomes, we used an appropriate procedure for p-values adjustment of multiple
648 correlated tests (34). We selected the best model and replicated the association in the UNIBA cohort
649 (one-tailed p-value < .05). We reported the effect size as partial η^2 .

650 Finally, to assess the biological significance of the SNPs encompassed in the PCI, we submitted the list
651 and the selected variants in full linkage disequilibrium with them to HaploReg 4.1(36) selecting
652 American ancestry and all four epigenome sources. Haploreg is a repository of genetic regulatory
653 elements across multiple tissues according to previous genomic studies (35, 36). Since the reference
654 network was identified in dorsolateral prefrontal cortex, we specifically interrogated this brain region
655 including BA46 and 9. Finally, we computed the statistics for overrepresentation of regulatory
656 elements (Bonferroni-corrected p-value < .05).

657

658

659 **Supplementary Materials**

660 Material and Methods

661 Results

662 Fig. S1. Results of intra-dataset preservations (Langfelder method).

663 Fig. S2. Preservation of CMC network in the LIBD dataset (Langfelder method).

664 Fig. S3. *Darkgreen* module enrichment for schizophrenia risk genes (all gene biotypes).

665 Fig. S4. Specific Expression Analysis (SEA).

666 Fig. S5. Transcription factor binding sites enrichment analysis.

667 Fig. S6. MicroRNA targets prediction.

668 Fig. S7. Overlap of LIBD modules with haloperidol targets.

669 Fig. S8. PCIs replication in LIBD developmental ages set.

670 Fig. S9. Association between PCIs and *Darkgreen* ME separately in Caucasian and African-American
671 subjects.

672 Fig. S10. RUV pre-processing.

673 Fig. S11. Preservation of the LIBD network in LIBD dataset with an alternative preprocess pipeline
674 (Langfelder method).

675 Table S1. The LIBD network: module replication in the CMC dataset and enrichment statistics for
676 schizophrenia risk.

677 Table S2. The CMC network: module replication in the LIBD dataset.

678 Table S3. The LIBD network: association between module eigengenes (MEs) and diagnosis (HC vs.
679 SCZ).

680 Table S4. Genes in the PGC list included in the network.

681 Table S5. Chart of *Darkgreen* genes and connectivity statistics.

682 Table S6. The LIBD network: module replication (empirical p-values).

683 Table S7. Overrepresentation of miRNA targetomes in *Darkgreen* module.

684 Table S8. SNP annotations: A' weights of SNP genotypes used for PCI computation.

685 Data file S1. (Microsoft Excel format). Chart of LIBD network genes, module labels assignments and
686 connectivity statistics.

687 Data file S2. (Microsoft Excel format). Chart of p-values: association between LIBD module
688 eigengenes and biological confounders.

689 Data file S3. (Microsoft Excel format). Chart of combined p-values: enrichment of LIBD modules for
690 miRNA targetomes.

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705 **References and notes**

- 706 1. C. Schizophrenia Working Group of the Psychiatric Genomics, Biological insights from 108
707 schizophrenia-associated genetic loci. *Nature* **511**, 421-427 (2014).
- 708 2. K. S. Kendler, What psychiatric genetics has taught us about the nature of psychiatric illness and what is
709 left to learn. *Molecular psychiatry* **18**, 1058-1066 (2013).
- 710 3. N. C. Hettige, C. B. Cole, S. Khalid, V. De Luca, Polygenic risk score prediction of antipsychotic dosage in
711 schizophrenia. *Schizophrenia research* **170**, 265-270 (2016).
- 712 4. T. Wimberley, C. Gasse, S. M. Meier, E. Agerbo, J. H. MacCabe, H. T. Horsdal, Polygenic Risk Score for
713 Schizophrenia and Treatment-Resistant Schizophrenia. *Schizophrenia bulletin*, (2017).
- 714 5. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic.
715 *Cell* **169**, 1177-1186 (2017).
- 716 6. P. J. Harrison, D. R. Weinberger, Schizophrenia genes, gene expression, and neuropathology: on the
717 matter of their convergence. *Molecular psychiatry* **10**, 40-68; image 45 (2005).
- 718 7. C. Gaiteri, Y. Ding, B. French, G. C. Tseng, E. Sibille, Beyond modules and hubs: the potential of gene
719 coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes,
720 brain, and behavior* **13**, 13-24 (2014).
- 721 8. M. Fromer, P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C.
722 Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gumus, A. E. Cicek, K. K. Dang, A. Browne,
723 C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y. C. Wang, M. C.
724 Mahajan, J. M. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De
725 Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H.
726 Toyoshiba, R. E. Gur, C. G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum,
727 E. E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin, P. Sklar, Gene
728 expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience* **19**,
729 1442-1453 (2016).
- 730 9. A. E. Jaffe, R. E. Straub, J. H. Shin, R. Tao, Y. Gao, L. Collado Torres, T. Kam-Thong, H. S. Xi, J. Quan, Q.
731 Chen, C. Colantuoni, W. S. Ulrich, B. J. Maher, A. Deep-Soboslay, T. B. Consortium, A. Cross, N. J.
732 Braindon, J. T. Leek, T. M. Hyde, J. E. Kleinman, D. R. Weinberger, Developmental And Genetic
733 Regulation Of The Human Cortex Transcriptome In Schizophrenia. *bioRxiv*, (2017).
- 734 10. G. Pergola, P. Di Carlo, E. D'Ambrosio, B. Gelao, L. Fazio, M. Papalino, A. Monda, G. Scozia, B.
735 Pietrangelo, M. Attrotto, J. A. Apud, Q. Chen, V. S. Mattay, A. Rampino, G. Caforio, D. R. Weinberger, G.
736 Blasi, A. Bertolino, DRD2 co-expression network and a related polygenic index predict imaging,
737 behavioral and clinical phenotypes linked to schizophrenia. *Translational psychiatry* **7**, e1006 (2017).
- 738 11. S. Freytag, J. Gagnon-Bartsch, T. P. Speed, M. Bahlo, Systematic noise degrades gene co-expression
739 signals but can be corrected. *BMC bioinformatics* **16**, 309 (2015).
- 740 12. A. E. Jaffe, R. Tao, A. L. Norris, M. Kealhofer, A. Nellore, J. H. Shin, D. Kim, Y. Jia, T. M. Hyde, J. E.
741 Kleinman, R. E. Straub, J. T. Leek, D. R. Weinberger, qSVA framework for RNA quality correction in
742 differential expression analysis. *Proceedings of the National Academy of Sciences of the United States
743 of America* **114**, 7130-7135 (2017).
- 744 13. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis.
745 *Statistical applications in genetics and molecular biology* **4**, Article17 (2005).
- 746 14. G. Pergola, P. Di Carlo, I. Andriola, B. Gelao, S. Torretta, M. T. Attrotto, L. Fazio, A. Raio, D. Albergò, R.
747 Masellis, A. Rampino, G. Blasi, A. Bertolino, Combined effect of genetic variants in the GluN2B coding
748 gene (GRIN2B) on prefrontal function during working memory performance. *Psychological medicine* **46**,
749 1135-1150 (2016).
- 750 15. R. Rosenheck, J. Doyle, D. Leslie, A. Fontana, Changing environments and alternative perspectives in
751 evaluating the cost-effectiveness of new antipsychotic drugs. *Schizophrenia bulletin* **29**, 81-93 (2003).

- 752 16. A. Bertolino, G. Caforio, G. Blasi, M. De Candia, V. Latorre, V. Petruzzella, M. Altamura, G. Nappi, S.
753 Papa, J. H. Callicott, V. S. Mattay, A. Bellomo, T. Scarabino, D. R. Weinberger, M. Nardini, Interaction of
754 COMT (Val(108/158)Met) genotype and olanzapine treatment on prefrontal cortical function in
755 patients with schizophrenia. *The American journal of psychiatry* **161**, 1798-1805 (2004).
- 756 17. E. Radulescu, A. E. Jaffe, R. E. Straub, Q. Chen, J. H. Shin, T. M. Hyde, J. E. Kleinman, D. R. Weinberger,
757 Identification and prioritization of gene sets associated with schizophrenia risk by co-expression
758 network analysis in human brain. *bioRxiv*, (2018).
- 759 18. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible?
760 *PLoS computational biology* **7**, e1001057 (2011).
- 761 19. M. R. Johnson, J. Behmoaras, L. Bottolo, M. L. Krishnan, K. Pernhorst, P. L. M. Santoscoy, T. Rossetti, D.
762 Speed, P. K. Srivastava, M. Chadeau-Hyam, N. Hajji, A. Dabrowska, M. Rotival, B. Razzaghi, S. Kovac, K.
763 Wanisch, F. W. Grillo, A. Slaviero, S. R. Langley, K. Shkura, P. Roncon, T. De, M. Mattheisen, P.
764 Niehusmann, T. J. O'Brien, S. Petrovski, M. von Lehe, P. Hoffmann, J. Eriksson, A. J. Coffey, S. Cichon, M.
765 Walker, M. Simonato, B. Danis, M. Mazzuferi, P. Foerch, S. Schoch, V. De Paola, R. M. Kaminski, V. T.
766 Cunliffe, A. J. Becker, E. Petretto, Systems genetics identifies Sestrin 3 as a regulator of a proconvulsant
767 gene network in human epileptic hippocampus. *Nature communications* **6**, 6031 (2015).
- 768 20. P. Michalak, Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic
769 genomes. *Genomics* **91**, 243-248 (2008).
- 770 21. G. Kustatscher, P. Grabowski, J. Rappsilber, Pervasive coexpression of spatially proximal genes is
771 buffered at the protein level. *Molecular systems biology* **13**, 937 (2017).
- 772 22. C. A. de Leeuw, J. M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS
773 data. *PLoS computational biology* **11**, e1004219 (2015).
- 774 23. X. Xu, A. B. Wells, D. R. O'Brien, A. Nehorai, J. D. Dougherty, Cell type-specific expression analysis to
775 identify putative cellular mechanisms for neurogenetic disorders. *The Journal of neuroscience : the*
776 *official journal of the Society for Neuroscience* **34**, 1420-1431 (2014).
- 777 24. K. Ohi, T. Shimada, Y. Nitta, H. Kihara, H. Okubo, T. Uehara, Y. Kawasaki, Specific gene expression
778 patterns of 108 schizophrenia-associated loci in cortex. *Schizophrenia research* **174**, 35-38 (2016).
- 779 25. D. R. Weinberger, From neuropathology to neurodevelopment. *Lancet* **346**, 552-557 (1995).
- 780 26. M. R. Johnson, K. Shkura, S. R. Langley, A. Delahaye-Duriez, P. Srivastava, W. D. Hill, O. J. Rackham, G.
781 Davies, S. E. Harris, A. Moreno-Moral, M. Rotival, D. Speed, S. Petrovski, A. Katz, C. Hayward, D. J.
782 Porteous, B. H. Smith, S. Padmanabhan, L. J. Hocking, J. M. Starr, D. C. Liewald, A. Visconti, M. Falchi, L.
783 Bottolo, T. Rossetti, B. Danis, M. Mazzuferi, P. Foerch, A. Grote, C. Helmstaedter, A. J. Becker, R. M.
784 Kaminski, I. J. Deary, E. Petretto, Systems genetics identifies a convergent gene network for cognition
785 and neurodevelopmental disease. *Nature neuroscience* **19**, 223-232 (2016).
- 786 27. F. Zambelli, G. Pesole, G. Pavesi, Pscan: finding over-represented transcription factor binding site
787 motifs in sequences from co-regulated or co-expressed genes. *Nucleic acids research* **37**, W247-252
788 (2009).
- 789 28. A. Ultsch, J. Lotsch, What do all the (human) micro-RNAs do? *BMC genomics* **15**, 976 (2014).
- 790 29. M. E. Hauberg, M. H. Holm-Nielsen, M. Mattheisen, A. L. Askou, J. Grove, A. D. Borglum, T. J. Corydon,
791 Schizophrenia risk variants affecting microRNA function and site-specific regulation of NT5C2 by miR-
792 206. *European neuropsychopharmacology : the journal of the European College of*
793 *Neuropsychopharmacology* **26**, 1522-1526 (2016).
- 794 30. Y. Kim, P. Giusti-Rodriguez, J. J. Crowley, J. Bryois, R. J. Nonneman, A. K. Ryan, C. R. Quackenbush, M. D.
795 Iglesias-Ussel, P. H. Lee, W. Sun, F. P. de Villena, P. F. Sullivan, Comparative genomic evidence for the
796 involvement of schizophrenia risk genes in antipsychotic effects. *Molecular psychiatry* **23**, 708-712
797 (2018).
- 798 31. D. Trabzuni, M. Ryten, R. Walker, C. Smith, S. Imran, A. Ramasamy, M. E. Weale, J. Hardy, Quality
799 control parameters on a large dataset of regionally dissected human control brains for whole genome
800 expression studies. *Journal of neurochemistry* **119**, 275-282 (2011).

- 801 32. A. Schroeder, O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M.
802 Granzow, T. Ragg, The RIN: an RNA integrity number for assigning integrity values to RNA
803 measurements. *BMC Mol Biol* **7**, 3 (2006).
- 804 33. T. S. Stroup, J. P. McEvoy, M. S. Swartz, M. J. Byerly, I. D. Glick, J. M. Canive, M. F. McGee, G. M.
805 Simpson, M. C. Stevens, J. A. Lieberman, The National Institute of Mental Health Clinical Antipsychotic
806 Trials of Intervention Effectiveness (CATIE) project: schizophrenia trial design and protocol
807 development. *Schizophrenia bulletin* **29**, 15-31 (2003).
- 808 34. K. N. Conneely, M. Boehnke, So many correlated tests, so little time! Rapid adjustment of P values for
809 multiple correlated tests. *American journal of human genetics* **81**, 1158-1168 (2007).
- 810 35. L. D. Ward, M. Kellis, HaploReg: a resource for exploring chromatin states, conservation, and regulatory
811 motif alterations within sets of genetically linked variants. *Nucleic acids research* **40**, D930-934 (2012).
- 812 36. L. D. Ward, M. Kellis, HaploReg v4: systematic mining of putative causal variants, cell types, regulators
813 and target genes for human complex traits and disease. *Nucleic acids research* **44**, D877-881 (2016).
- 814 37. Z. Zhang, H. Yu, S. Jiang, J. Liao, T. Lu, L. Wang, D. Zhang, W. Yue, Evidence for Association of Cell
815 Adhesion Molecules Pathway and NLGN1 Polymorphisms with Schizophrenia in Chinese Han
816 Population. *PLoS one* **10**, e0144719 (2015).
- 817 38. N. The, C. Pathway Analysis Subgroup of the Psychiatric Genomics, Psychiatric genome-wide
818 association study analyses implicate neuronal, immune and histone pathways. *Nature neuroscience* **18**,
819 199-209 (2015).
- 820 39. T. Nakazawa, M. Kikuchi, M. Ishikawa, H. Yamamori, K. Nagayasu, T. Matsumoto, M. Fujimoto, Y.
821 Yasuda, M. Fujiwara, S. Okada, K. Matsumura, A. Kasai, A. Hayata-Takano, N. Shintani, S. Numata, K.
822 Takuma, W. Akamatsu, H. Okano, A. Nakaya, H. Hashimoto, R. Hashimoto, Differential gene expression
823 profiles in neurons generated from lymphoblastoid B-cell line-derived iPSCs from monozygotic twin
824 cases with treatment-resistant schizophrenia and discordant responses to clozapine. *Schizophr Res* **181**,
825 75-82 (2017).
- 826 40. Y. Kim, P. Giusti-Rodriguez, J. J. Crowley, J. Bryois, R. J. Nonneman, A. K. Ryan, C. R. Quackenbush, M. D.
827 Iglesias-Ussel, P. H. Lee, W. Sun, F. P. de Villena, P. F. Sullivan, Comparative genomic evidence for the
828 involvement of schizophrenia risk genes in antipsychotic effects. *Molecular psychiatry*, (2017).
- 829 41. S. Erk, A. Meyer-Lindenberg, P. Schmierer, S. Mohnke, O. Grimm, M. Garbusow, L. Haddad, L. Poehland,
830 T. W. Muhleisen, S. H. Witt, H. Tost, P. Kirsch, N. Romanczuk-Seiferth, B. H. Schott, S. Cichon, M. M.
831 Nothen, M. Rietschel, A. Heinz, H. Walter, Hippocampal and frontolimbic function as intermediate
832 phenotype for psychosis: evidence from healthy relatives and a common risk variant in CACNA1C.
833 *Biological psychiatry* **76**, 466-475 (2014).
- 834 42. A. Devor, O. A. Andreassen, Y. Wang, T. Maki-Marttunen, O. B. Smeland, C. C. Fan, A. J. Schork, D.
835 Holland, W. K. Thompson, A. Witoelar, C. H. Chen, R. S. Desikan, L. K. McEvoy, S. Djurovic, P. Greengard,
836 P. Svenningsson, G. T. Einevoll, A. M. Dale, Genetic evidence for role of integration of fast and slow
837 neurotransmission in schizophrenia. *Molecular psychiatry* **22**, 792-801 (2017).
- 838 43. Q. Zhang, Q. Shen, Z. Xu, M. Chen, L. Cheng, J. Zhai, H. Gu, X. Bao, X. Chen, K. Wang, X. Deng, F. Ji, C.
839 Liu, J. Li, Q. Dong, C. Chen, The effects of CACNA1C gene polymorphism on spatial working memory in
840 both healthy controls and patients with schizophrenia or bipolar disorder. *Neuropsychopharmacology :
841 official publication of the American College of Neuropsychopharmacology* **37**, 677-684 (2012).
- 842 44. B. Dietsche, H. Backes, D. Laneri, T. Weikert, S. H. Witt, M. Rietschel, J. Sommer, T. Kircher, A. Krug, The
843 impact of a CACNA1C gene polymorphism on learning and hippocampal formation in healthy
844 individuals: a diffusion tensor imaging study. *NeuroImage* **89**, 256-261 (2014).
- 845 45. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone
846 pathways. *Nature neuroscience* **18**, 199-209 (2015).
- 847 46. G. Lippi, C. C. Fernandes, L. A. Ewell, D. John, B. Romoli, G. Curia, S. R. Taylor, E. P. Frady, A. B. Jensen, J.
848 C. Liu, M. M. Chaabane, C. Belal, J. L. Nathanson, M. Zoli, J. K. Leutgeb, G. Biagini, G. W. Yeo, D. K. Berg,

- 849 MicroRNA-101 Regulates Multiple Developmental Programs to Constrain Excitation in Adult Neural
850 Networks. *Neuron* **92**, 1337-1351 (2016).
- 851 47. A. Jauhari, T. Singh, A. Pandey, P. Singh, N. Singh, A. K. Srivastava, A. B. Pant, D. Parmar, S. Yadav,
852 Differentiation Induces Dramatic Changes in miRNA Profile, Where Loss of Dicer Diverts Differentiating
853 SH-SY5Y Cells Toward Senescence. *Molecular neurobiology* **54**, 4986-4995 (2017).
- 854 48. M. C. Chiang, Y. C. Cheng, H. M. Chen, Y. J. Liang, C. H. Yen, Rosiglitazone promotes neurite outgrowth
855 and mitochondrial function in N2A cells via PPARgamma pathway. *Mitochondrion* **14**, 7-17 (2014).
- 856 49. H. Chang, L. Li, T. Peng, M. Grigoriou-Serbanescu, S. E. Bergen, M. Landen, C. M. Hultman, A. J.
857 Forstner, J. Strohmaier, J. Hecker, T. G. Schulze, B. Muller-Myhsok, A. Reif, P. B. Mitchell, N. G. Martin,
858 S. Cichon, M. M. Nothen, S. Jamain, M. Leboyer, F. Bellivier, B. Etain, J. P. Kahn, C. Henry, M. Rietschel,
859 G. Swedish Bipolar Study, D. S. C. Moo, X. Xiao, M. Li, Identification of a Bipolar Disorder Vulnerable
860 Gene CHDH at 3p21.1. *Molecular neurobiology*, (2016).
- 861 50. W. Sun, J. Poschmann, R. Cruz-Herrera Del Rosario, N. N. Parikshak, H. S. Hajan, V. Kumar, R.
862 Ramasamy, T. G. Belgard, B. Elanggovan, C. C. Wong, J. Mill, D. H. Geschwind, S. Prabhakar, Histone
863 Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* **167**, 1385-1397 e1311 (2016).
- 864 51. M. C. O'Donovan, M. J. Owen, The implications of the shared genetics of psychiatric disorders. *Nature*
865 *medicine* **22**, 1214-1219 (2016).
- 866 52. P. Celada, M. V. Puig, F. Artigas, Serotonin modulation of cortical neurons and networks. *Frontiers in*
867 *integrative neuroscience* **7**, 25 (2013).
- 868 53. R. Brisch, A. Saniotis, R. Wolf, H. Bielau, H. G. Bernstein, J. Steiner, B. Bogerts, K. Braun, Z. Jankowski, J.
869 Kumaratilake, M. Henneberg, T. Gos, The role of dopamine in schizophrenia from a neurobiological and
870 evolutionary perspective: old fashioned, but still in vogue. *Frontiers in psychiatry* **5**, 47 (2014).
- 871 54. R. Birnbaum, D. R. Weinberger, Functional neuroimaging and schizophrenia: a view towards effective
872 connectivity modeling and polygenic risk. *Dialogues in clinical neuroscience* **15**, 279-289 (2013).
- 873 55. J. E. Kleinman, A. J. Law, B. K. Lipska, T. M. Hyde, J. K. Ellis, P. J. Harrison, D. R. Weinberger, Genetic
874 neuropathology of schizophrenia: new approaches to an old question and new uses for postmortem
875 human brains. *Biological psychiatry* **69**, 140-145 (2011).
- 876 56. A. Meyer-Lindenberg, D. R. Weinberger, Intermediate phenotypes and genetic mechanisms of
877 psychiatric disorders. *Nature reviews. Neuroscience* **7**, 818-827 (2006).
- 878 57. T. R. Insel, The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *The*
879 *American journal of psychiatry* **171**, 395-397 (2014).
- 880 58. T. Moberget, N. T. Doan, D. Alnaes, T. Kaufmann, A. Cordova-Palomera, T. V. Lagerberg, J. Diedrichsen,
881 E. Schwarz, M. Zink, S. Eisenacher, P. Kirsch, E. G. Jonsson, H. Fatouros-Bergman, L. Flyckt, KaSp, G.
882 Pergola, T. Quarto, A. Bertolino, D. Barch, A. Meyer-Lindenberg, I. Agartz, O. A. Andreassen, L. T.
883 Westlye, Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite
884 mega-analysis of 983 patients and 1349 healthy controls. *Molecular psychiatry*, (2017).
- 885 59. G. T. Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue
886 gene regulation in humans. *Science* **348**, 648-660 (2015).
- 887 60. S. R. Kay, A. Fiszbein, L. A. Opler, The positive and negative syndrome scale (PANSS) for schizophrenia.
888 *Schizophrenia bulletin* **13**, 261-276 (1987).
- 889 61. L. Jacob, J. A. Gagnon-Bartsch, T. P. Speed, Correcting gene expression data when neither the
890 unwanted variation nor the factor of interest are observed. *Biostatistics* **17**, 16-28 (2016).
- 891 62. J. A. Gagnon-Bartsch, T. P. Speed, Using control genes to correct for unwanted variation in microarray
892 data. *Biostatistics* **13**, 539-552 (2012).
- 893 63. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control
894 genes or samples. *Nature biotechnology* **32**, 896-902 (2014).
- 895 64. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC*
896 *bioinformatics* **9**, 559 (2008).

- 897 65. P. Roussos, B. Guennewig, D. C. Kaczorowski, G. Barry, K. J. Brennand, Activity-Dependent Changes in
898 Gene Expression in Schizophrenia Human-Induced Pluripotent Stem Cell Neurons. *JAMA psychiatry* **73**,
899 1180-1188 (2016).
- 900 66. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A. L. Barabasi, Hierarchical organization of
901 modularity in metabolic networks. *Science* **297**, 1551-1555 (2002).
- 902 67. L. J. Carithers, K. Ardlie, M. Barcus, P. A. Branton, A. Britton, S. A. Buia, C. C. Compton, D. S. DeLuca, J.
903 Peter-Demchok, E. T. Gelfand, P. Guan, G. E. Korzeniewski, N. C. Lockhart, C. A. Rabiner, A. K. Rao, K. L.
904 Robinson, N. V. Roche, S. J. Sawyer, A. V. Segre, C. E. Shive, A. M. Smith, L. H. Sobin, A. H. Undale, K. M.
905 Valentino, J. Vaught, T. R. Young, H. M. Moore, A Novel Approach to High-Quality Postmortem Tissue
906 Procurement: The GTEx Project. *Biopreservation and biobanking* **13**, 311-319 (2015).
- 907 68. D. Smedley, S. Haider, S. Durinck, L. Pandini, P. Provero, J. Allen, O. Arnaiz, M. H. Awedh, R. Baldock, G.
908 Barbiera, P. Bardou, T. Beck, A. Blake, M. Bonierbale, A. J. Brookes, G. Bucci, I. Buetti, S. Burge, C.
909 Cabau, J. W. Carlson, C. Chelala, C. Chrysostomou, D. Cittaro, O. Collin, R. Cordova, R. J. Cutts, E. Dassi,
910 A. Di Genova, A. Djari, A. Esposito, H. Estrella, E. Eyra, J. Fernandez-Banet, S. Forbes, R. C. Free, T.
911 Fujisawa, E. Gadaleta, J. M. Garcia-Manteiga, D. Goodstein, K. Gray, J. A. Guerra-Assuncao, B. Haggarty,
912 D. J. Han, B. W. Han, T. Harris, J. Harshbarger, R. K. Hastings, R. D. Hayes, C. Hoede, S. Hu, Z. L. Hu, L.
913 Hutchins, Z. Kan, H. Kawaji, A. Keliet, A. Kerhornou, S. Kim, R. Kinsella, C. Klopp, L. Kong, D. Lawson, D.
914 Lazarevic, J. H. Lee, T. Letellier, C. Y. Li, P. Lio, C. J. Liu, J. Luo, A. Maass, J. Mariette, T. Maurel, S.
915 Merella, A. M. Mohamed, F. Moreews, I. Nabihoudine, N. Ndegwa, C. Noirot, C. Perez-Llamas, M.
916 Primig, A. Quattrone, H. Quesneville, D. Rambaldi, J. Reecy, M. Riba, S. Rosanoff, A. A. Saddiq, E. Salas,
917 O. Sallou, R. Shepherd, R. Simon, L. Sperling, W. Spooner, D. M. Staines, D. Steinbach, K. Stone, E.
918 Stupka, J. W. Teague, A. Z. Dayem Ullah, J. Wang, D. Ware, M. Wong-Erasmus, K. Youens-Clark, A.
919 Zadissa, S. J. Zhang, A. Kasprzyk, The BioMart community portal: an innovative alternative to large,
920 centralized data repositories. *Nucleic acids research* **43**, W589-598 (2015).
- 921 69. A. Mathelier, O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-
922 Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W. W. Wasserman, JASPAR 2016: a major
923 expansion and update of the open-access database of transcription factor binding profiles. *Nucleic*
924 *acids research* **44**, D110-115 (2016).
- 925 70. V. Agarwal, G. W. Bell, J. W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian
926 mRNAs. *eLife* **4**, (2015).
- 927 71. N. Wong, X. Wang, miRDB: an online resource for microRNA target prediction and functional
928 annotations. *Nucleic acids research* **43**, D146-152 (2015).
- 929 72. S. Bandyopadhyay, R. Mitra, TargetMiner: microRNA target prediction with systematic identification of
930 tissue-specific negative examples. *Bioinformatics* **25**, 2625-2631 (2009).
- 931 73. I. S. Vlachos, M. D. Paraskevopoulou, D. Karagkouni, G. Georgakilas, T. Vergoulis, I. Kanellos, I. L.
932 Anastasopoulos, S. Maniou, K. Karathanou, D. Kalfakakou, A. Fevgas, T. Dalamagas, A. G. Hatzigeorgiou,
933 DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA
934 interactions. *Nucleic acids research* **43**, D153-159 (2015).
- 935 74. A. L. Price, A. Helgason, G. Thorleifsson, S. A. McCarroll, A. Kong, K. Stefansson, Single-tissue and cross-
936 tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS*
937 *genetics* **7**, e1001317 (2011).
- 938 75. R. B. Scharpf, R. A. Irizarry, M. E. Ritchie, B. Carvalho, I. Ruczinski, Using the R Package crlmm for
939 Genotyping and Copy Number Estimation. *Journal of statistical software* **40**, 1-32 (2011).
- 940 76. B. N. Howie, P. Donnelly, J. Marchini, A flexible and accurate genotype imputation method for the next
941 generation of genome-wide association studies. *PLoS genetics* **5**, e1000529 (2009).
- 942 77. O. Delaneau, C. Coulonges, J. F. Zagury, Shape-IT: new rapid and accurate algorithm for haplotype
943 inference. *BMC bioinformatics* **9**, 540 (2008).

- 944 78. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de
945 Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based
946 linkage analyses. *American journal of human genetics* **81**, 559-575 (2007).
- 947 79. J. O'Connell, D. Gurdasani, O. Delaneau, N. Pirastu, S. Ulivi, M. Cocca, M. Traglia, J. Huang, J. E.
948 Huffman, I. Rudan, R. McQuillan, R. M. Fraser, H. Campbell, O. Polasek, G. Asiki, K. Ekoru, C. Hayward,
949 A. F. Wright, V. Vitart, P. Navarro, J. F. Zagury, J. F. Wilson, D. Toniolo, P. Gasparini, N. Soranzo, M. S.
950 Sandhu, J. Marchini, A general approach for haplotype phasing across the full spectrum of relatedness.
951 *PLoS genetics* **10**, e1004234 (2014).
- 952 80. B. Howie, C. Fuchsberger, M. Stephens, J. Marchini, G. R. Abecasis, Fast and accurate genotype
953 imputation in genome-wide association studies through pre-phasing. *Nature genetics* **44**, 955-959
954 (2012).
- 955 81. C. Genomes Project, G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E.
956 Handsaker, H. M. Kang, G. T. Marth, G. A. McVean, An integrated map of genetic variation from 1,092
957 human genomes. *Nature* **491**, 56-65 (2012).
- 958 82. B. Howie, J. Marchini, M. Stephens, Genotype imputation with thousands of genomes. *G3* **1**, 457-470
959 (2011).
- 960 83. O. Delaneau, J. Marchini, J. F. Zagury, A linear complexity phasing method for thousands of genomes.
961 *Nature methods* **9**, 179-181 (2012).
- 962 84. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: accurate alignment of
963 transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36
964 (2013).
- 965 85. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning
966 sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
- 967 86. E. Eisenberg, E. Y. Levanon, Human housekeeping genes, revisited. *Trends in genetics : TIG* **29**, 569-574
968 (2013).
- 969 87. L. Peixoto, D. Risso, S. G. Poplawski, M. E. Wimmer, T. P. Speed, M. A. Wood, T. Abel, How data analysis
970 affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic
971 acids research* **43**, 7664-7674 (2015).
- 972 88. Network, C. Pathway Analysis Subgroup of the Psychiatric Genomics, Corrigendum: Psychiatric
973 genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature
974 neuroscience* **18**, 1861 (2015).

975

976

977 **Acknowledgments:** This article was based on results from the Clinical Antipsychotic Trials of
978 Intervention Effectiveness (CATIE) project supported with Federal funds from the National Institute of
979 Mental Health (NIMH) under contract NO1 MH90001. The project was carried out by principal
980 investigators from the University of North Carolina, Duke University, the University of Southern
981 California, the University of Rochester, and Yale University in association with Quintiles, Inc., and the
982 program staff of the Division of Interventions and Services Research of the NIMH and investigators

983 from 84 sites in the United States. AstraZeneca Pharmaceuticals LP, Bristol-Myers Squibb Company,
984 Forest Pharmaceuticals, Inc., Janssen Pharmaceutica Products, L.P., Eli Lilly and Company, Otsuka
985 Pharmaceutical Co., Ltd., Pfizer Inc., and Zenith Goldline Pharmaceuticals, Inc., provided medications
986 for the studies. CMC data were generously provided to GP by the NIMH and CommonMind
987 Consortium. We gratefully acknowledge the work by Prof. Roberto Bellotti, Dr. Alfonso Monaco
988 (Department of Physics – University of Bari Aldo Moro), Marco Zezza, Leonardo Sportelli and
989 Elisabetta Volpe (Department of Basic Medical Science, Neuroscience, and Sense Organs – University
990 of Bari Aldo Moro), who contributed to data analysis. We are also in debt to Dr. Gianluca Ursini, Dr.
991 Richard Straub, and Dr. Venkata S. Mattay (Lieber Institute for Brain Development) for insightful
992 discussions on the procedures employed. **Funding:** This work was supported by a “Capitale Umano ad
993 Alta Qualificazione” grant by Fondazione Con Il Sud, by the NARSAD grant (number: 28935), and by
994 the “Ricerca Finalizzata” (grant number: PE-2011-02347951) awarded to Alessandro Bertolino; by the
995 Lieber Institute for Brain Development; and by a Hoffmann-La Roche Collaboration Grant awarded to
996 Giulio Pergola. This project has received funding from the European Union Seventh Framework
997 Programme for research, technological development and demonstration under grant agreement no.
998 602450 (IMAGEMEND). This paper reflects only the author's views and the European Union is not
999 liable for any use that may be made of the information contained therein. **Author contributions:** GP,
1000 PDC, DRW, and AB designed the study; TMH, JEK, GB, DRW, AR, and AB were involved in data
1001 collection; PDC, AEJ, MP, QC analyzed the data; GP, PDC, AEJ, DRW, and AB interpreted the data;
1002 GP, PDC, and AB wrote the first draft of the manuscript; all authors revised and approved the
1003 manuscript. **Competing interests:** Alessandro Bertolino is a stockholder of Hoffmann-La Roche Ltd.
1004 He has also received consulting fees from Biogen and lecture fees from Otsuka, Janssen, Lundbeck,
1005 and consultant fees from Biogen. Giulio Pergola has been the academic supervisor of a Roche
1006 collaboration grant (years 2015-16) that funds his and Antonio Rampino's salary. Antonio Rampino

1007 has received travel fees from Lundbeck. All other authors have no biomedical financial interests and no
1008 potential conflicts of interest.

1009

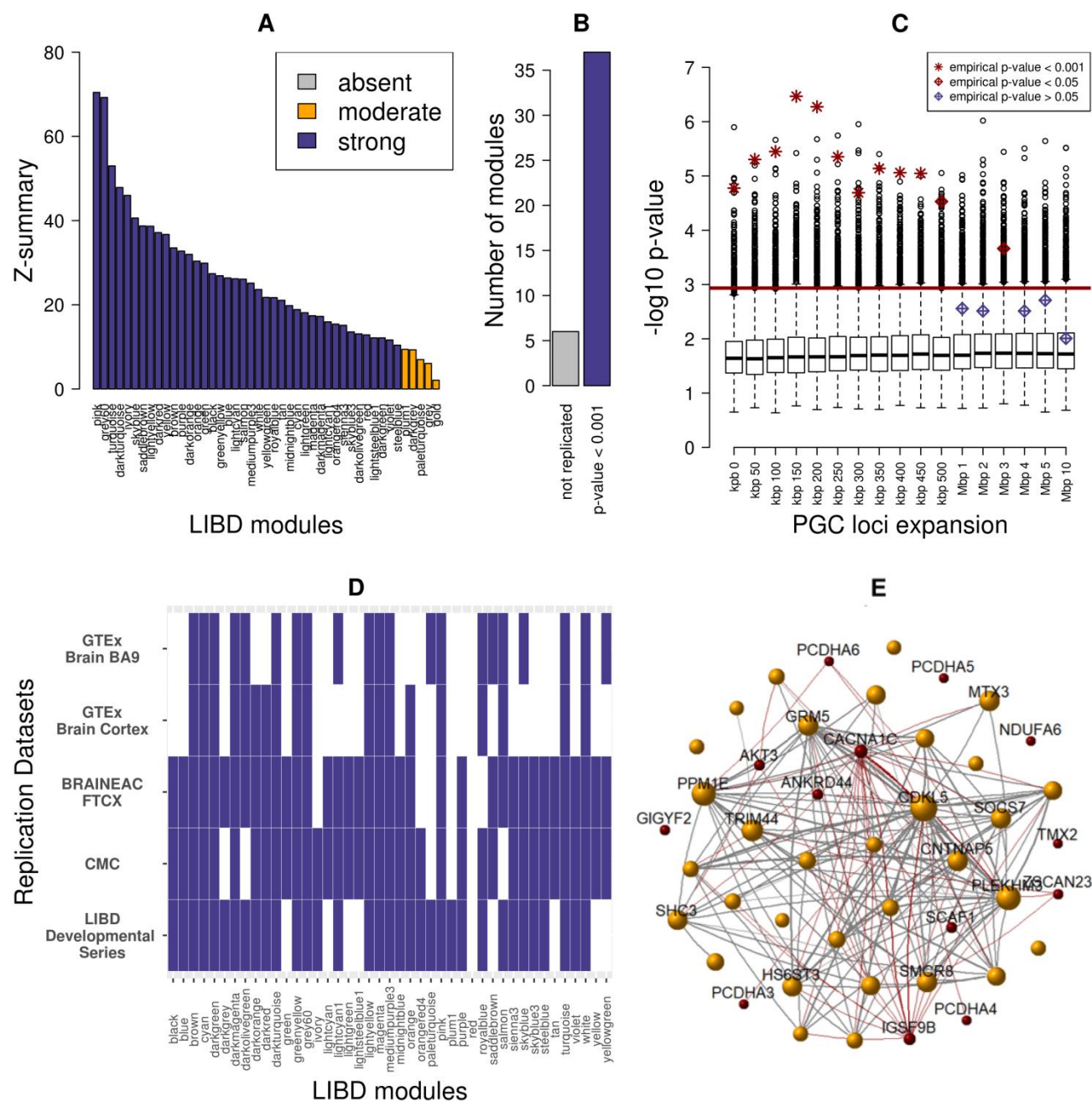
1010

1011

1012

1013

1014 **Figures:**



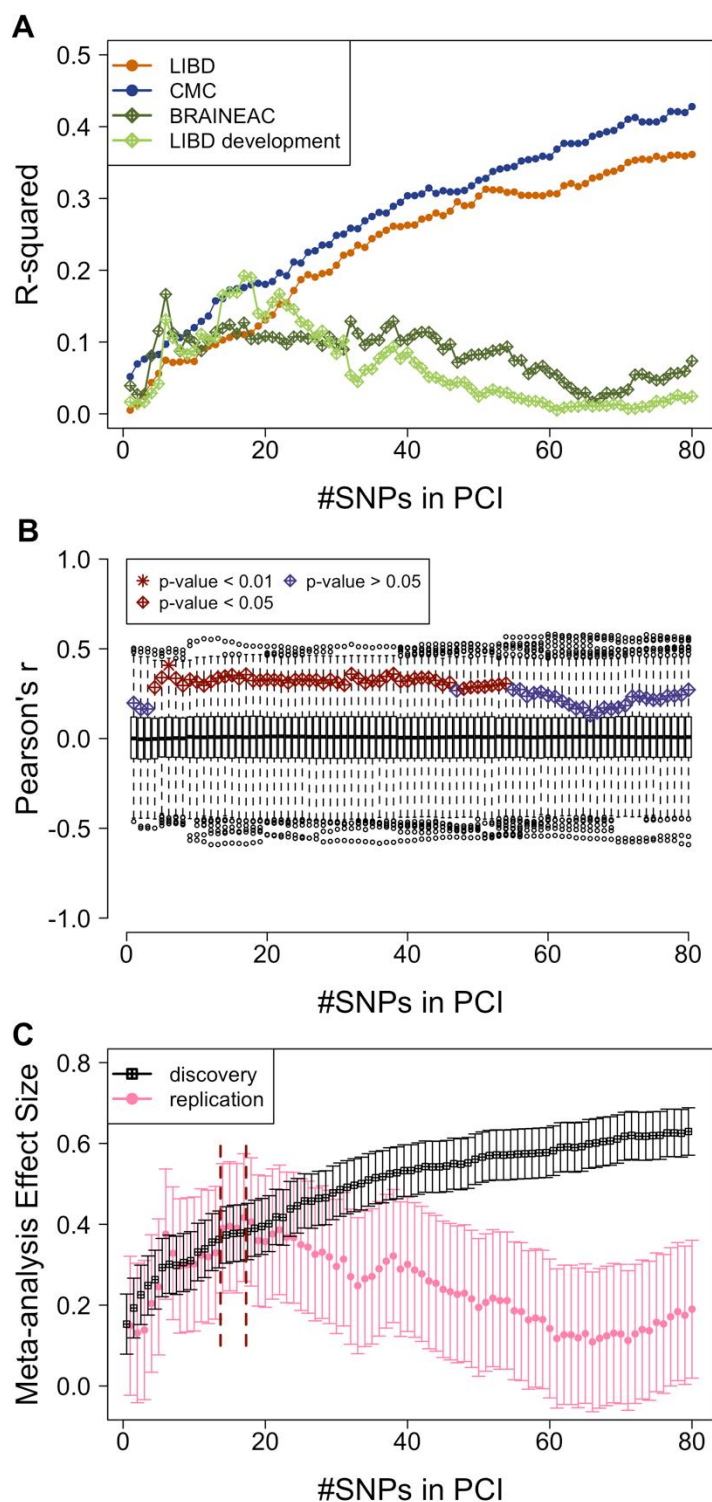
1015

1016 **Fig. 1. Co-expression Network.** (A) Preservation of the LIBD network in the CMC dataset
 1017 (Langfelder method). The LIBD modules are shown on the x-axis ranked by Z-summary preservation
 1018 score (y-axis). $Z \geq 10$ denotes strong preservation, $2 \leq Z < 10$ moderate, and $Z < 2$ absent (18). (B)
 1019 Replication of the LIBD modules topology in the CMC dataset (Johnson method). Bars indicate the

1020 number of replicated modules at empirical p-value $< .001$ vs. not replicated modules (10,000
1021 permutations). **(C)** *Darkgreen* module enrichment for schizophrenia risk genes. Enrichment
1022 significance is shown over increasing expansion of schizophrenia risk loci boundaries. The x-axis
1023 reports the size of expansion in kilo-base pairs (kbp). The y-axis indicates the $-\log_{10}$ p-value of the
1024 hypergeometric test for overrepresentation of schizophrenia risk loci in *Darkgreen*. Boxplots show the
1025 null distribution of the lowest enrichment p-value over all network modules obtained after network
1026 labels permutation (n=10,000). The red horizontal line shows the Bonferroni threshold selected
1027 (number of modules = 43, $\alpha = .0012$). Stars and diamonds denote *Darkgreen* exact enrichment p-value.
1028 **(D)** Replication of the LIBD modules in several different datasets (Johnson method). Slate-blue fields
1029 denote modules (x-axis) replicated at empirical p-value $< .001$ (over 10,000 permutations). **(E)**
1030 *Darkgreen* graph. The nodes of the graphs (spheres) are genes and schizophrenia risk genes are colored
1031 in dark red. Gold spheres represent a selection of the most connected genes in the module (scaled intra-
1032 modular connectivity ≥ 0.3) and have a diameter proportional to intra-modular connectivity. i.e., larger
1033 spheres denote genes harboring more connections within *Darkgreen*. Lines denote gene-gene
1034 relationships and their width is proportional to connection strength.

1035

1036



1037

1038 **Fig. 2. Polygenic Co-expression Index.** (A) The plot illustrates the variation of the effect size of the
1039 correlation between the PCI and the *Darkgreen* module eigengene (Y-axis) for a series of PCIs with

1040 incrementally added SNPs. The discovery (LIBD, CMC) and replication datasets (BRAINEAC, LIBD
1041 development) are represented with different colors. For increasing number of SNPs included in the PCI
1042 (x-axis), the effect size in the discovery sets increases monotonically because of overfitting, while it
1043 remains stable and then drops in the replication set, suggesting an optimal signal-to-noise ratio in the
1044 replication set between 6 and about 40 SNPs. **(B)** PCI replication. Empirical significance of the
1045 correlations between PCIs and Module Eigengene (ME) in the replication set (BRAINEAC). Stars and
1046 diamonds display on the y-axis the significance of each *ME*-PCI correlation over an increasing number
1047 of SNPs (x-axis). Box plots show the corresponding null distribution of the correlation coefficients
1048 when genotypes are permuted (2,000 permutations). Color and shape key in the panel highlight
1049 different empirical significance cut-offs. **(C)** Meta-analysis of the effect sizes in the discovery and
1050 replication datasets. Dark red vertical dashed lines delimit a plateau in the replication effect sizes
1051 between 14 and 17 SNPs. Note that the effect size never increases above the level observed at the 17th
1052 SNP.

1053

1054

1055

1056

1057

1058

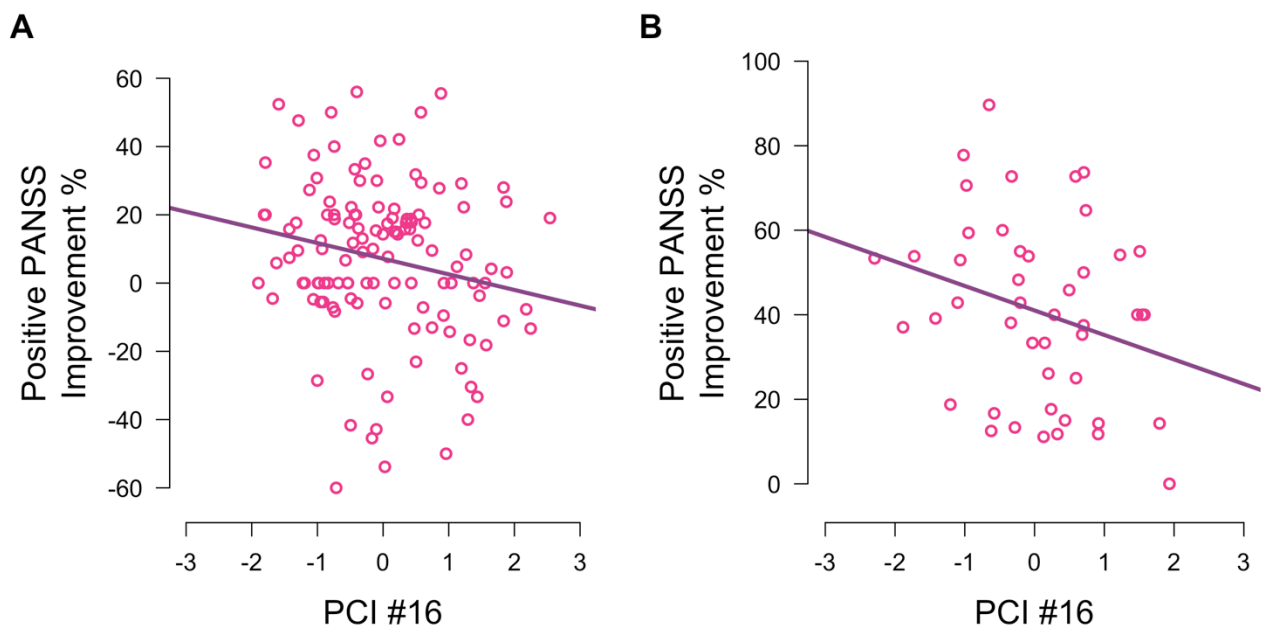
1059

1060

1061

1062

1063



1064

1065 **Fig. 3. Association between the PCI and clinical outcome.** Negative correlation between the PCI
1066 with 16 SNPs and symptom improvement in the positive domain of the PANSS (difference between
1067 endpoint and baseline relative to baseline, shown on the Y-axis) in the (A) CATIE and (B) UNIBA
1068 cohorts.

1069

1070

1071

1072

Tables

Table 1. Demographics.

	LIBD	CMC	Stats LIBD vs. CMC		BRAINEAC	LIBD Developmental Ages	GTEx: Brain Cortex	GTEx: Brain BA9	CATIE	UNIBA
Sample size	343	345			38	93	96	92	121	46
Female (male) [ratio]	105 (238) [0.44]	114 (231) [0.49]	$\chi^2 = 0.4$	$p = 0.55$	9 (29) [0.31]	33 (60) [0.55]	#	#	27 (94) [0.28]	9 (37) [0.24]
Age mean \pm s.d. (years)	45.2 \pm 14.8	60.8 \pm 17.4	$t = -12.7$	$p < 2.2 \times 10^{-16}$	56.6 \pm 19.1	3.5 \pm 6.0	#	#	41.5 \pm 10.9	27.7 \pm 6.8
Age range (years)	17-85	17-86			20-89	0-16	#	#	19-65	16-42
Diagnosis SCZ^a (HC^b)	143 (200)	166 (179)	$\chi^2 = 2.6$	$p = 0.11$	0 (38)	0 (93)	#	#	121 (0)	46 (0)
Ethnicity CAUC^c (AA^d)	166 (177)	283 (62)	$\chi^2 = 84.3$	$p < 2.2 \times 10^{-16}$	38 (0)	40 (53)	#	#	77 (44)	46 (0)

^a patients with schizophrenia; ^b healthy controls; ^c Caucasian; ^d African-American; # data unavailable as per ref. 49.

Table 2. PGC loci and genes overlapping with the module *Darkgreen*.

Ensembl gene ID	HGNC ^a Symbol	Gene name	PGC loci rank	PGC loci index SNP	PGC loci position (hg19 ^b)	PGC index SNP p-value
ENSG00000187987	ZSCAN23	zinc finger and SCAN domain containing 23	1	rs115329265	chr6:28303247-28712247	3.48×10^{-31}
ENSG00000151067	CACNA1C	calcium channel, voltage-dependent, L type, alpha 1C subunit	4	rs2007044, rs2239063	chr12:2321860-2523731	3.22×10^{-18}
ENSG00000204120	GIGYF2	GRB10 interacting GYF protein 2	22	rs6704768	chr2:233559301-233753501	2.32×10^{-12}
ENSG00000065413	ANKRD44	ankyrin repeat domain 44	31	rs6434928	chr2:198148577-198835577	2.06×10^{-11}
ENSG00000080854	IGSF9B	immunoglobulin superfamily, member 9B	36	rs75059851	chr11:133808069-133852969	3.87×10^{-11}
ENSG00000184983	NDUFA6	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 6, 14kDa	57	rs1023500, rs6002655	chr22:42315744-42689414	1.71×10^{-9}
ENSG00000213593	TMX2	thioredoxin-related transmembrane protein 2	59	rs9420	chr11:57386294-57682294	2.24×10^{-9}
ENSG00000117020	AKT3	v-akt murine thymoma viral oncogene homolog 3	64	rs10803138, rs77149735, rs14403, chr1_243881945_I	chr1:243503719-244002945	3.73×10^{-9}
ENSG00000126461	SCAF1	SR-related CTD-associated factor 1	106	rs56873913	chr19:50067499-50135399	4.69×10^{-8}
ENSG00000255408	PCDHA3	protocadherin alpha 3	108	chr5_140143664_I	chr5:140023664-140222664	4.85×10^{-8}
ENSG00000204967	PCDHA4	protocadherin alpha 4				
ENSG00000204965	PCDHA5	protocadherin alpha 5				
ENSG00000081842	PCDHA6	protocadherin alpha 6				

^aHUGO Gene Nomenclature Committee ID. ^bHuman Genome version 19.

14	rs260098	chr15:99641932	A	G	0.46	0.0010960	1	intergenic	SYNM	SYNM	ENSG00000259475, ENSG00000182253	antisense, protein coding
15	rs7115028	chr11:88483559	A	C	0.24	0.0014171	1	intron	GRM5	GRM5	ENSG00000168959	protein coding
16	rs2429175	chr12:2045085	A	G	0.34	0.0017808	1	intron	LINC00940	CACNA1C, DCP1B, LRTM2	ENSG00000151067, ENSG00000151065, ENSG00000166159	protein coding, protein coding, protein coding
17	rs2015586	chr10:119021737	C	T	0.43	0.0018226	1	intron	SLC18A2	SLC18A2	ENSG00000165646	protein coding
18	rs73055782	chr19:46500197	C	T	0.48	0.0018739	1	intron	CCDC61	CCDC61	ENSG00000104983	protein coding
19	rs12650211	chr4:91578496	T	C	0.17	0.0020413	1	intron	CCSER1	CCSER1	ENSG00000184305	processed transcript
20	rs2909160	chr7:102840922	T	C	0.12	0.0020747	1	intron	DPY19L2P2	DPY19L2P2	ENSG00000170629	processed transcript
21	rs2839149	chr21:47632580	T	C	0.26	0.0025836	1	intron	LSS	LSS	ENSG00000160285	protein coding
22	rs1510173	chr16:54253227	T	C	0.46	0.0028605	1	intergenic		FTO	ENSG00000140718	protein coding
23	rs6828754	chr4:6877902	T	C	0.40	0.0029061	1	intron	KIAA0232	KIAA0232	ENSG00000170871	protein coding
24	rs55844460	chr8:6648691	G	T	0.45	0.0029581	1	intergenic		AGPAT5	ENSG00000249898, ENSG00000155189	antisense, protein coding
25	rs7756776	chr6:13051652	G	A	0.36	0.0029726	1	intron	PHACTR1	PHACTR1	ENSG00000112137	protein coding
26	rs7964786	chr12:121132100	C	T	0.44	0.0030576	1	intron	MLEC	MLEC	ENSG00000110917	protein coding
27	rs56140649	chr2:56577297	C	T	0.19	0.0034265	1	intron	CCDC85A	CCDC85A	ENSG00000055813	protein coding
28	rs7688945	chr4:141642216	C	A	0.26	0.0034656	1	intron	TBC1D9	TBC1D9	ENSG00000109436	protein coding
29	rs35310447	chr17:7018444	T	G	0.32	0.0034804	1	promoter	ASGR2	ASGR1, BCL6B, DLG4, MIR497HG, RPL7AP64, SLC16A11	ENSG00000141505, ENSG00000161940, ENSG00000132535, ENSG00000267532, ENSG00000213876, ENSG00000174326	protein coding, protein coding, protein coding, antisense, pseudogene, protein coding
30	rs1858719	chr8:17977212	T	G	0.50	0.0035667	1	intergenic		ASAH1, PCM1	ENSG00000245281, ENSG00000253384, ENSG00000104763, ENSG00000078674	antisense, pseudogene, protein coding, protein coding
31	rs16965349	chr17:36614524	A	G	0.22	0.0036403	1	intron	ARHGAP23	ARHGAP23	ENSG00000225485	protein coding
32	rs5999223	chr22:34549440	A	G	0.34	0.0038297	1	intergenic			ENSG00000224404	lincRNA

^aReference Allele; ^bMinor Allele Frequency; ^cHUGO Gene Nomenclature Committee ID; ^dHGNC symbol of the closest gene (± 500 kbp) encompassed in *Darkgreen* Module;

^eEnsembl ID of the closest gene (± 500 kbp) encompassed in *Darkgreen* Module.

Table 4. Association between PCIs and positive PANSS early treatment response.

PANSS subscales	# SNPs in the PCI	CATIE			UNIBA	
		t-value	p-value	Corrected p-value	t-value	One-sided p-value
Positive	PCI #14	-2.11	.03681	.134	-1.93	.0306
	PCI #15	-2.74	.00714	.033	-1.79	.0403
	PCI #16	-2.74	.00708	.033	-1.71	.0475
	PCI #17	-2.58	.01119	.049	-1.60	.0553
Negative	PCI #14	0.59	.5575	1	1.14	.1311
	PCI #15	0.29	.7750	1	1.18	.1218
	PCI #16	0.28	.7816	1	1.35	.0925
	PCI #17	0.22	.8198	1	1.42	.0817
General	PCI #14	-1.52	.1301	.277	-0.28	.3889
	PCI #15	-1.81	.0721	.191	-0.19	.4234
	PCI #16	-1.74	.0837	.208	-0.24	.4075
	PCI #17	-1.78	.0771	.198	-0.23	.4106