1	Pathway-based polygenic score analysis identifies the NRF2-
2	KEAP1 and mRNA splicing - minor pathways as enriched in
3	shared genetic variation between chronotype and bipolar disorder.
4	
5	
6	Laura Fahey ¹ , Lorna M. Lopez ¹
7	
8	
9	1 Department of Biology, Maynooth University, Maynooth, Co. Kildare, Ireland
10	

Author Summary 12

13 This study investigates shared genetic influences that may contribute to the circadian rhythm disruption 14 and sleep issues in neurodevelopmental and neuropsychiatric conditions. Using polygenic score 15 analysis on large-scale genetic studies of autism, attention-deficit/hyperactivity disorder, schizophrenia, and bipolar disorder, we tested their ability to predict chronotype and insomnia status of participants in 16 17 the UK Biobank. Our findings reveal that polygenic scores for autism, schizophrenia, and bipolar disorder are associated with an evening chronotype, while polygenic scores for attention-18 19 deficit/hyperactivity disorder, autism, schizophrenia, and bipolar disorder are associated with insomnia. 20 Pathway analysis identified the enrichment of shared genetic variation between chronotype and bipolar 21 disorder in the NRF2-KEAP1 and mRNA splicing minor pathways. Previous studies have linked the 22 NRF2-KEAP1 pathway to the pathology of bipolar disorder and schizophrenia. NRF2 and splicing 23 components have been previously reported to be rhythmically regulated by circadian clock genes. These 24 results suggest a potential role for the NRF2-KEAP1 and mRNA splicing minor pathways in mediating 25 circadian rhythm disturbances in bipolar disorder, providing insights into the genetic basis of sleep 26 issues in neuropsychiatric conditions.

27

11

²⁹ NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

30 Abstract

It has been postulated that circadian dysfunction may contribute to the sleep problems prevalent in neurodevelopmental and neuropsychiatric conditions. Genetic correlations between numerous pairs of neurodevelopmental or neuropsychiatric and sleep phenotypes have been identified. We hypothesize that this overlapping genetic variation is enriched in certain biological pathways.

35 We used genome-wide polygenic score analysis to confirm previously reported genetic correlations and

- 36 pathway-based polygenic score analysis to identify enriched pathways. We created polygenic scores
- 37 using summary statistics from the largest genome-wide association studies (GWAS) of autism (AUT),
- 38 attention-deficit/hyperactivity disorder (ADHD), schizophrenia (SCZ) and bipolar disorder (BP). We
- 39 tested the performance of these polygenic scores in predicting chronotype and insomnia status of UK
- 40 Biobank participants. For the pathway-based polygenic scores, we restricted genetic variation to SNPs
- 41 that mapped to genes within 451 pathways from the Reactome database.
- 42 Genome-wide polygenic scores for AUT, SCZ and BP were found to be associated with an evening 43 chronotype, and polygenic scores for ADHD, AUT, SCZ and BP were found to be associated with 44 insomnia status. Pathway-based polygenic score analysis identified the NRF2 KEAP1 and mRNA 45 splicing minor pathways as being enriched for genetic variation overlapping between chronotype and BP. For the NRF2 KEAP1 pathway, the signal is enriched in the subset of genes that function with 46 47 KEAP1 to regulate NRF2 expression. Examination of eQTL data pointed to BP associated SNPs within 48 these gene-sets being associated with expression changes of many genes to which they map. A number 49 of these eQTL SNPs were reported to be genome-wide significant for SCZ in previous studies.

These results demonstrate that the overlapping genetic variation between chronotype and BP is enriched in genes involved in the *NRF2-KEAP1* and *mRNA splicing minor* pathways. Animal model and human cell line studies have previously linked the NRF2 pathway to the pathology of BP and SCZ. Additionally, NRF2 and splicing components have been reported to be rhythmically regulated by circadian clock genes. Our results suggest that these pathways could be involved in mediating the disrupted circadian rhythm phenotype of BP.

- 56
- 57
- 59

60 Introduction

61 Neurodevelopmental and neuropsychiatric conditions (NDPCs) encompass conditions that affect brain function, cognitive-processes, emotions and behaviour. Neurodevelopmental conditions manifest in 62 childhood, with the most prevalent ones including autism spectrum disorder (AUT) and attention deficit 63 64 hyperactivity disorder (ADHD) (1). Neuropsychiatric conditions typically have onset in adulthood and 65 include schizophrenia (SCZ) and bipolar disorder (BP). NDPCs are characterised by their 66 symptomatology, with many of them displaying overlapping symptoms. One example of such overlapping symptoms are sleep problems, considered a core symptom in the diagnostic criteria for BP 67 68 (DSM-5; American Psychiatric Association, 2013), present in up to 80% of individuals with SCZ (3) 69 and AUT (4,5) and observed in up to 50% of those with ADHD (6).

Sleep is regulated by sleep/wake homeostasis, which balances the urge to sleep and the requirement for wakefulness. As wakefulness persists, the homeostatic sleep drive intensifies, promoting sleep. Sleep is also regulated by the circadian clock, a biological timing mechanism that is controlled by 24-hour transcription-translational feedback loops involving several core clock genes. The central circadian pacemaker in the suprachiasmatic nucleus (SCN) of the hypothalamus responds to external cues, such

as light signals, and synchronizes circadian rhythms across all cells in the body (7).

76 Both sleep/wake homeostasis and circadian rhythm have been shown to be disrupted in NDPCs. 77 Disruption of homeostatic sleep drive, as indicated by reduced slow-wave activity, has been reported in 78 SCZ and early-stage psychosis (8,9). Melatonin and cortisol are two hormones involved in circadian 79 regulation. Melatonin is produced at night, while cortisol is primarily produced in the morning (10). 80 Dysregulation of both hormones has been observed in NDPCs. For example, AUT has being associated 81 with lower levels of melatonin at night (10,11) and ADHD has been associated with lower levels of 82 cortisol after wakening (12). Significant reductions in melatonin levels have also been reported in the 83 cerebrospinal fluid of BP patients (13). Conversely, activities that disrupt circadian rhythms, such as 84 shift work and frequent transatlantic flights, have been associated with adverse mental health (14). 85 Hence, a bidirectional relationship appears to exist, where environmental disruptions to circadian 86 rhythm exacerbate NDPC symptoms, while inherent circadian rhythm disturbances potentially 87 contribute to the development of NDPCs or share underlying mechanisms. This supports the hypothesis 88 that overlapping genetic variation contributes to both circadian disruption and the development of 89 NDPCs. Indeed, mouse models with circadian clock gene mutations display symptoms reminiscent of 90 human NDPCs, with different gene or protein domain mutations resulting in varying sets of NDPC 91 symptoms (15).

92 Genome-wide association studies (GWASs) have identified hundreds of genome-wide single nucleotide 93 polymorphisms (SNPs) associated with different NDPCs and sleep related phenotypes. Chronotype 94 refers to individual's preference for wakening and preforming activity earlier in the day (morning

95 chronotype) or later (evening chronotype). It is a behavioural manifestation of an individual's circadian 96 clock and is often used as a proxy phenotype for circadian timing (16). In modern society, having a 97 morning chronotype is considered advantageous due to its alignment with early school and work start 98 times. Consequently, individuals with an evening chronotype may experience reduced sleep on 99 work/school days and compensate for it on free days - this is known as 'social jetlag' (17). Evening chronotypes have being associated with numerous adverse psychological outcomes (18). The most 100 101 recent GWAS of a morning chronotype reported enrichment of chronotype associated loci in the clock 102 genes PER1, PER2, PER3, CRY1 and BMAL1. Furthermore, the top enriched biological pathways relate 103 to the circadian clock (19).

104 Surprisingly, no core clock genes have been linked with any NDPC through large-scale GWAS. 105 However, genome-wide correlations in GWAS identified SNP effects have been reported between many sleep traits and NDPCs using the linkage disequilibrium score regression tool (Figure 1). This 106 107 relationship is negative for the advantageous morning chronotype and positive for insomnia and 108 daytime sleepiness, which are both debilitating. Polygenic score (PGS) analysis can also be used to 109 assess genetic correlations between phenotypes. If a PGS created using effect sizes from a GWAS of 110 one phenotype can significantly predict the phenotype of samples in an independent cohort, it indicates 111 shared genetic variation between the two phenotypes in question (20). PGSs for ADHD and depression 112 have been reported to positively correlate with initiating and maintaining sleep and excessive somnolence in children. Additionally, a PGS for anxiety disorder correlated with nightmares in children 113 (21). These genetic correlation findings suggest that the shared genetic background between NDPCs 114 and circadian rhythms likely extends genome-wide, beyond the core clock genes. 115

Exploring the genes, biological pathways and tissue types impacted by the shared genetic variation 116 117 between sleep and NDPCs has been a relatively unexplored area. A gene-based cross-trait meta-analysis identified forty-four genes common to both insomnia and ADHD, and found them to be enriched for 118 genes within synaptic-related pathways (22). Another study identified 149 loci shared between 119 120 psychiatric disorders (SCZ, BP and depression) and sleep traits (chronotype and sleep duration). Forty-121 nine lead SNPs at these loci were found to act as eQTLs for 115 genes across various brain tissues 122 including the basal ganglia, cortex, hippocampus and cerebellum (23).

123 The aims of this study were two-fold. First, we investigated if genome-wide polygenic scores for 124 ADHD, AUT, SCZ and BP could predict insomnia and chronotype in UKB. Secondly, using pathway-125 based polygenic score analysis, we investigated the biological pathways implicated by shared genetic 126 variation between these NDPC and sleep phenotypes. We report that polygenic scores for AUT, SCZ and BP were associated with chronotype status and polygenic scores for all four NDPC phenotypes 127 128 were associated with insomnia. BP polygenic scores, restricted to genetic variation in the KEAP1-NRF2 129 and *mRNA splicing-minor* biological pathways were both significant in predicting chronotype, over and

- 130 above the background signal. To gain further biological insights, we reviewed expression quantitative
- 131 trait loci data, and found many BP associated SNPs within these gene sets to be associated with
- 132 expression changes of the genes to which they map.
- 133



134

Figure 1. Published genetic correlations (rg) between pairs of neuropsychiatric or 135 136 neurodevelopmental and sleep phenotypes. Heatmap cells marked with an asterisk (*) indicate that 137 they were reported as statistically significant in the original publications, after multiple testing 138 correction. The original publications include the latest and largest genome-wide association studies of 139 Autism (AUT; Grove et al., 2019), morning chronotype (Jones et al., 2019), attention-deficit hyperactivity disorder (ADHD; Demontis et al., 2023), bipolar disorder (BP; Mullins et al., 2021), major 140 141 depressive disorder (MDD; Howard et al., 2019), sleep duration (Dashti et al., 2019) and daytime sleepiness (Wang et al., 2019). Rows and columns are clustered based on similarity. 142

Results 143

Genome-wide polygenic score analysis 144

145 To test for indication of shared genetic variation, we investigated if genome-wide polygenic scores for AUT, ADHD, SCZ and BP could predict chronotype and/or insomnia status of UKB participants. We 146 147 implemented this using the PGS tool, SBayesRC, which re-estimates GWAS SNP effect sizes using a 148 model that incorporates external linkage disequilibrium (LD) and biological annotation data (24). As 149 discovery data for PGS analysis, we used publicly available GWAS summary statistics for the NDPC

150 phenotypes, which were based on studies containing between 46,350 (AUT) and 225,534 (ADHD) 151 samples. As target data for PGS analysis, we used UKB genotype and questionnaire data related to 152 chronotype and insomnia. After quality control, we had 409,630 samples with chronotype data and 153 239,918 samples with insomnia data (see methods).

Polygenic scores for AUT, SCZ and BP, were statistically significant for association with chronotype 154 status (AUT: $r^2 = 3.57 \times 10^{-4}$ [95% CI = 2.51×10⁻⁴ - 4.82×10⁻⁴], p = <2.20×10⁻¹⁶; SCZ: $r^2 = 5.13\times 10^{-4}$ [95% 155 $CI = 3.84 \times 10^{-4} - 6.61 \times 10^{-4}$], $p = \langle 2.20 \times 10^{-16}$; BP: $r^2 = 2.97 \times 10^{-5}$ [95% $CI = 5.71 \times 10^{-6} - 7.25 \times 10^{-5}$], $p = \langle 2.20 \times 10^{-16} \rangle$ 156 4.84x10⁻⁴) (Figure 2, Supplementary Table 1). Polygenic scores for all four GWAS phenotypes were 157 statistically significant for association with insomnia, the strongest association being between the 158 ADHD PGS and insomnia (ADHD: $r^2 = 2.36 \times 10^{-3}$ [95% CI = 1.99x10⁻³ - 2.76x10⁻³], p = <2.20x10⁻¹⁶; 159 AUT: $r^2 = 3.69 \times 10^{-5}$ [95% CI = 4.29×10⁻⁶ - 1.02×10⁻⁴], p = 2.93×10⁻³; BP: $r^2 = 1.10 \times 10^{-4}$ [95% CI = 160 $4.19 \times 10^{-5} - 2.09 \times 10^{-4}$], p = 2.90×10^{-7} ; SCZ: r² = 3.32×10^{-5} [95% CI = $3.07 \times 10^{-6} - 9.52 \times 10^{-6}$], p = 4.79×10^{-5} 161 ³; Figure 2; Supplementary Table 1). To enable direct comparison to the pathway PGS analysis in the 162 next section, we also report genome-wide PGSs calculated using the tool, PRSice-2 (Choi & O'Reilly, 163 164 2019), with a P-value threshold of 0.5 (Supplementary Table 1). These results are very similar to those 165 obtained with SBayesRC, with the exception that SBayesRC performed better in the prediction of an 166 evening chronotype based on all four NDPC PGSs, and also performed better in the prediction of 167 insomnia based on the ADHD PGS. This improvement is to be expected given that the calculation of PGSs using LD clumping and p-value thresholding, as in PRSice-2, is not the optimal method for 168 genome-wide PGS analysis (26). 169

170 Pathway-based polygenic score analysis

171 Pathway-based polygenic score analysis was performed to test if shared genetic variation exists between 172 AUT, ADHD, SCZ or BP and chronotype or insomnia status at the biological pathway level. Here, all eight pairs of phenotypes were tested, regardless of whether statistically significant results were 173 174 observed at the genome-wide level, because it is possible that shared genetic effects are present in 175 certain biological pathways, but the effects are too small to be observable at a genome-wide level. We tested all Reatome pathways that contained between 50 and 500 genes, resulting in 451 pathways to 176 177 test. Pathway-based polygenic scores were derived using SNPs that map by position to each of the 451 178 pathways.

- Two pathways were statistically significant for the prediction of chronotype status based on BP genetic 179
- variation. These pathways encompass the mRNA splicing minor pathway ($r^2 = 7.5 \times 10^{-5}$ [95% CI = 180
- $3.13 \times 10^{-5} 1.26 \times 10^{-4}$], P = 3.18×10^{-8} , competitive P = 1.33×10^{-5}) and the KEAP1 NRF2 pathway (r² = 181
- 7.18x10⁻⁵ [95% CI = $2.92x10^{-5}$ - $1.33x10^{-3}$], P = $6.26x10^{-8}$, competitive P = $1.33x10^{-5}$; Figure 3, 182
- Supplementary Table 2). Note that the competitive P value achieved for each of these significant gene-183
- 184 sets was the lowest possible based on the 75,000 permutations performed. Interestingly, the mRNA

185 splicing - minor pathway also ranked the highest, in terms of competitive P value, for the prediction of 186 chronotype based on the SCZ PGS, but this did not surpass the multiple testing corrected threshold (r^2 = 5.62×10^{-5} [95% CI = 1.96×10^{-5} - 1.12×10^{-4}], P = 1.65×10^{-6} , competitive P = 1.0×10^{-3} ; Supplementary 187 188 Table 2).

189 No pathway PGS achieved a competitive P value that passed multiple testing correction for the prediction of insomnia. The pathway Nuclear Events Mediated by NRF2, which is a subpathway of 190 KEAP1 NRF2, ranked as the second and third highest performing pathways, out of all 451 tested, for 191 the prediction of insomnia based on SCZ and BP PGSs, respectively (Figure 3, Supplementary Table 192 193 3). Competitive P values for the top four pathway PGSs for each of the eight phenotype pairs are 194 visualised as a heatmap in Figure 3.

Post-hoc analyses 195

196 Interrogations of significant pathway-based polygenic score analysis results

197 The mRNA Splicing – Minor Pathway encompasses 50 genes, to which 518 SNPs map, while the KEAP1 NRF2 pathway comprises 107 genes, to which 1,356 SNPs map. To explore potential 198 199 enrichment of the observed signal within specific subsets of these pathways, we investigated Reactome 200 subpathways, some of which were initially omitted due to our criteria of excluding pathways with fewer than 50 genes. Given the relatively small size of the mRNA splicing – minor pathway, it does not contain 201 any Reactome subpathways. In contrast, the KEAP1 NRF2 pathway could be broken down into two 202 203 Reactome subpathways, and an additional pathway that we established comprising of all genes that do 204 not fall into these two subpathways, which we referred to as Regulation of NRF2 by KEAP1. Among the KEAP1 NRF2 sub-pathways, the gene-set that explained the most variance was Regulation of NRF2 205 *by KEAP1* ($R^2 = 5.0x10^{-5}$ [95% CI = 1.6x10⁻⁵-1.03x10⁻⁴], P = 6.25x10⁻⁶, competitive P = 1.87x10⁻⁴; 206 Figure 4a; Supplementary Table 4). To account for the size of the pathways, we plotted the average 207 208 variance in chronotype explained per SNP for these pathway PGSs, in comparison to the average 209 variance explained per SNP by a PGS constructed using SNPs that map to all HGNC gene symbols 210 (Figure 4b; Supplementary Tables 2 and 4).

211 eQTL Analysis Results

- 212 To gain a better understanding of the effect of BP genetic variation on the expression of genes within
- 213 the mRNA Splicing – Minor Pathway and KEAP1 NRF2 pathways, we reviewed expression quantitative
- 214 trait loci (eQTLs) data in the database, QTLBase (27). For the KEAP1 NRF2 pathway, five genes
- 215 satisfied the criteria of having gene P values $< 1x10^{-2}$ for BP and $< 1x10^{-3}$ for chronotype based on gene-
- 216 based MAGMA analysis. Of these genes, three contained BP SNPs associated with their expression
- change: cullin-3 (CUL3), proteasome 26S subunit ubiquitin receptor, non-ATPase 2 (PSMD2) and 217
- 218 proteasome 20S subunit beta 10 (PSMB10).

medRxiv preprint doi: https://doi.org/10.1101/2024.02.16.24302920; this version posted February 18, 2024. The copyright holder for this preprint doi: https://doi.org/10.1101/2024.02.10.24002020, tills voision posted 1 doi and you and you

- 219 CUL3 is a core component of the CUL3-dependent E3 ubiquitin ligase complex which mediates 220 ubiquitination and subsequent degradation of target proteins. This complex associates with KEAP1 to 221 ubiquitinate NRF2, maintaining its abundance at a low level (28) At CUL3, the BP allele of the most
- 222 significant BP SNP (rs6748341) is associated with decreased expression of CUL3 in the brain cortex (P
- = 4.91×10^{-6}). Interestingly, a SNP in very high (LD) with this SNP (rs11685299; LD R² = 0.99) is 223 genome-wide significant for association with SCZ (29). The BP allele of a SNP in high LD with both
- 224 225 of these SNPs (rs11681451) is also associated with decreased expression of CUL3, this time in three
- brain regions (Cortex $P = 1.81 \times 10^{-6}$, Hippocampus $P = 1.95 \times 10^{-5}$, Caudate $P = 2.33 \times 10^{-5}$; Supplementary 226
- 227 Table 5).
- The gene PSMD2 encodes a subunit of the 26S proteasome, which degrades NRF2 (28). At PSMD2, 228
- 229 the BP allele of rs843355 is associated with decreased expression of PSMD2 in various immune cells
- in the blood, including Blood-Neutrophils CD16+ ($P = 1.38 \times 10^{-6}$) and Blood-Monocytes ($P = 1.79 \times 10^{-6}$) 230
- 231 ⁶) (Supplementary Table 5).
- 232 Similarly to PSMD2, the PSMB10 gene also encodes a proteasomal subunit involved in NRF2
- 233 degradation (28). At PSMB10, the BP allele of rs5923 is associated with both increased and decreased
- 234 expression of PSMB10 in various blood immune cells (Increased: Blood P = 4.2×10^{-9} , Blood-Monocytes
- $CD14+P = 5.11x10^{-9}$, Blood-Neutrophils $CD16+P = 2.94x10^{-7}$; Decreased: Blood-Monocytes CD14+235
- $P = 9.56 \times 10^{-7}$, Blood-Neutrophils CD16+ $P = 8.21 \times 10^{-6}$). 236
- For the mRNA Splicing Minor Pathway, six genes satisfied the criteria of having gene P values $< 1 \times 10^{-1}$ 237
- ² for BP and $< 1x10^{-3}$ for chronotype: splicing factor 3b subunit 1 (SF3B1), splicing factor 3b subunit 1 238
- (SF3B2), RNA polymerase II, I And III subunit H (POLR2H), small nuclear ribonucleoprotein U5 239
- Subunit 200 (SNRNP200) and pre-mRNA processing factor 6 (PRPF6). 240
- SF3B1 encodes subunit 1 of the splicing factor 3b protein complex (SF3b) which is a component of the 241 242 U12 small nuclear ribonucleoprotein (snRNP) of the minor spliceosome and a component of the U2 243 snRNP of the major spliceosome. At SF3B1, the BP allele of the SNP most associated with BP 244 (rs55775495) is associated with increased expression of SF3B in the central nervous system (P =245 2.68x10-15), peripheral nervous system ($P = 1.13x10^{-6}$), brain cerebellum ($P = 1.30x10^{-6}$) and brain caudate ($P = 2.19 \times 10^{-6}$). Interestingly, the SNP rs6434928, which is in very high LD ($R^2 = 0.88$) with 246 this SNP is genome-wide significant for association with SCZ (30) and has almost the exact same eOTL 247
- 248 results. SF3B2 encodes subunit 2 of SF3b. The BP allele of rs2270448 is associated with decreased
- expression of SF3B2 in the brain cerebellum ($P = 1.28 \times 10^{-5}$). 249
- 250 POLR2H encodes a subunit shared by the three eukaryotic polymerases. The BP allele of rs62287507
- is associated with increased expression of POLR2H in blood ($P = 2.33 \times 10^{-6}$). SNRNP200 encodes an 251
- RNA helicase that comprises one of the U5 snRNP-specific proteins. At SNRNP200, the BP allele is 252
- associated with increased expression in blood neutrophils CD16+ ($P = 2.02 \times 10^{-11}$). *PRPF6* encodes a 253

254 protein that acts as a bridging factor between U5 and U4/U6 snRNPs in formation of the spliceosome.

- 255 The encoded protein can also bind androgen receptor, providing a link between transcriptional
- 256 activation and splicing. The BP allele of the most significant SNP at this loci, rs3829704, is associated
- 257 with decreased expression of PRPF6 in blood cells (Lymphocyte $P = 1.29 \times 10^{-31}$, Blood $P = 3.95 \times 10^{-11}$,
- Blood-T cell CD4+ activated $P = 1.48 \times 10^{-7}$ and decreased expression of PRPF6 in the brain (Cortex P 258
- 259 $= 1.2 \times 10^{-6}$; Caudate, P = 6.0x10^{-6}).

260 **Differential Expression Analysis**

261 We next investigated if genes within the mRNA Splicing – Minor Pathway and KEAP1 NRF2 pathway 262 are differentially expressed between BP cases and controls. We used PsychEncode differential 263 expression data based on postmortem dorsolateral prefrontal cortex samples from 144 BP cases and 899 264 healthy controls. Two genes within the mRNA Splicing – Minor Pathway were found to be differentially expressed. These are Serine and Arginine Rich Splicing Factor 6 (SRSF6; log2FC = -0.07; P = 2.0x10⁻ 265 ⁴; FDR = 0.01) and Small Nuclear Ribonucleoprotein Polypeptides B and B1 (SNRPB; log2FC = 0.08; 266 P 4.4x10⁻⁴ = FDR = 0.02). Four differentially expressed genes were observed in the *KEAP1 NRF2* 267 pathway. These are Sestrin 1 (SESN1; log2FC = 0.07; $P = 2.8 \times 10^{-4}$; FDR = 0.02), cyclin dependent 268 kinase inhibitor 1A (CDKN1A log2FC = 0.32; P = 2.51×10^{-4} ; FDR = 0.02), AKT serine/threonine kinase 269 1 (AKT1; $\log 2FC = 0.04$; P = 6.8x10⁻⁴; FDR = 0.03) and MAF BZIP Transcription Factor G (MAFG; 270 log2FC = -0.05; P = 4.9x10⁻⁴; FDR = 0.02). Interestingly, three of these genes are present in *Regulation* 271 of NRF2 by KEAP1, which had a higher average R² per SNP than the KEAP1 NRF2 pathway itself 272 (Figure 4b). However, none of these gene-sets had a statistically significant enrichment of differentially 273 expressed genes (Regulation of NRF2 by KEAP1 P= 0.09; KEAP1 NRF2 P = 0.5; mRNA Splicing -274 275 *Minor Pathway* P =0.5).

276

Discussion 277

Using genome-wide polygenic score analysis, we have found that polygenic scores for AUT, SCZ and 278 BP could predict an evening chronotype and polygenic scores for all four NDPC phenotypes could 279 280 predict insomnia status in UKB. Pathway-based polygenic score analysis identified the NRF2 KEAP1 and mRNA splicing minor pathways as being enriched for genetic variation overlapping between 281 282 chronotype and BP. For the NRF2 KEAP1 pathway, the signal is enriched in the subset of genes that 283 function with KEAP1 to regulate NRF2 expression. Examination of eQTL data pointed to BP associated 284 SNPs being associated with expression changes of many genes within the highlighted pathways to 285 which they map. Interestingly, three of the ten eQTL SNPs are genome-wide significant for SCZ 286 (Supplementary Table 5).

287 The genome-wide polygenic score results we report support previously published genetic correlation 288 results. AUT and SCZ have previously been reported to exhibit a negative genetic correlation with a 289 morning chronotype, as demonstrated by linkage disequilibrium score regression (31,32). Using a 290 different method and data, our study provides further evidence for these findings, as well as 291 demonstrating a positive association between the BP PGS and an evening chronotype. Additionally, our 292 results confirm that insomnia is genetically correlated with ADHD, SCZ and BP (33–35), the strongest 293 relationship being between ADHD and insomnia. We also report a negative genetic correlation between 294 AUT and insomnia, whereas, in previous studies this relationship was not statistically significant (31).

295 BP polygenic scores created using SNPs that map to the KEAP1 NRF2 and mRNA splicing-minor Reactome pathways were associated with chronotype, over and above the background signal. The 296

297 majority of the variance explained by the KEAP1 NRF2 gene-set is concentrated within genes belonging

298 to the subpathway, Regulation of NRF2 by KEAP1. It must be noted that the R² values of these pathway-

299 based PGSs are low. This is partly to be expected, given the small numbers of genes in these pathways.

300 Clear enrichment of explained variance by the mRNA splicing – minor and KEAP1 NRF2 pathways are

301 observed when plotting the average variance explained per SNP for these pathways in comparison to

302 the average variance explained per SNP by PGSs created using all genes based on BP genetic variation

303 (Figure 4b).

304 A sub-pathway of KEAP1-NRF2, Nuclear Events Mediated by NRF2, ranked very highly in predicting insomnia status, based on BP and SCZ genetic variation. This did not reach statistical significance, 305 306 possibly due there being less insomnia samples (N=239,918) than chronotype samples (N=409,630), and therefore reduced power. NRF2 is a transcription factor that, under normal conditions, is bound by 307 308 KEAP1 in the cytoplasm as part of the CUL3-KEAP1 E3 ubiquitin ligase complex, which facilitates 309 the ubiquitination of NRF2, marking it for proteasomal degradation. When cells are exposed to 310 oxidative stress, toxins, bacteria, viruses or other forms of cellular damage, KEAP1 cysteine residues are modified, disrupting its binding affinity for NRF2. As a result, NRF2 stabilises and translocates to 311 312 the nucleus where it binds antioxidant response elements (AREs) in the promoter regions of genes 313 encoding antioxidant enzymes and detoxification proteins (36). Of particular interest regarding sleep 314 and circadian rhythms, NRF2 has also been observed to undergo rhythmic regulation by the two 315 circadian clock genes, BMAL1 and CLOCK, in myeloid cells. These clock genes promote NRF2 316 expression by binding to the NRF2 E-box domain (37). Through these pathways, NRF2 mediates 317 coupling between oscillations in redox balance and circadian timekeeping mechanisms, with this loop 318 serving to optimise the timing of circadian and seasonal timing of antioxidant responses (38,39).

319 Elevated oxidative stress has been consistently documented in numerous NDPCs (including BP), and 320 this contributes to the concurrent neuroinflammation observed in these conditions (40,41). The 321 increased expression of NRF2 target genes help cells mitigate oxidative stress, improve mitochondrial

322 function and reduce inflammation, all of which are involved in the pathology of NDPCs (36,42). NRF2 323 also acts as a negative regulator of the NF κ B signalling pathway, while the NF κ B pathway reciprocally 324 negatively regulates NRF2 signalling. This interaction helps maintain control over basal inflammation 325 and oxidative stress levels (43). When the NRF2 pathway becomes saturated or is deficient, activation

326 of the NFkB pathway is heightened, which leads to an increased production of pro-inflammatory cytokines (36). Given this relationship, it is therefore interesting that the KEAP1 NRF2 and Nuclear 327 328 Events Mediated by NRF2 pathways clustered together with the FCERI mediated NF-kB activation

329 pathway competitive P value heatmap of the top four pathways across all phenotype pairs (Figure 3).

330 Studies in a seasonal animal model (Meduka fish) that exhibits a constitutive depressive winter phenotype have shown that NRF2 signalling pathways were inactivated in winter photoperiods, and 331 332 that this was associated with a depressive phenotype that could be reversed by treatment with NRF2 activators (44). These functions of NRF2 signalling in regulating seasonal changes in behaviour are 333 334 significant since seasonality is such an important clinical feature of human NDPCs. For example, the 335 prevalence of major depressive disorder and seasonal affective disorder show seasonal variation and 336 the oscillations in mood and metabolic function that characterise BP recur with a seasonal rhythm (45). 337 Furthermore, people with BP have higher self-reported seasonality scores (46) and altered sensitivity to 338 changes in day light (47). Chronotype also links seasonality and BP, as evening chronotype is associated 339 with increased self-reported seasonality (48) and with the presence and severity of signs of BD (49).

340 Given the beneficial antioxidant and anti-inflammatory roles of NRF2, it may be hypothesized that 341 NRF2 is downregulated in NDPCs. This was observed in animal models of depression, where lower levels of NRF2 and KEAP1 were reported in the prefrontal cortex and the dentate gyrus and Cornu 342 Ammonis 3 (CA3) regions of the hippocampus, along with increased blood levels of inflammatory 343 344 cytokines (50). Similar results were seen using human postmortem samples, where protein levels of KEAP1 and NRF2 in the parietal cortex of major depressive disorder, schizophrenia, and bipolar 345 disorder samples were significantly lower than controls. Indeed, there is evidence supporting the 346 347 potential of natural and synthetic molecules, including melatonin, that elevate NRF2 levels to be used 348 as treatments for psychiatric conditions (36,42).

349 However, there is some disagreement in the literature, with some studies reporting that NRF2 signalling 350 is upregulated in NDPCs. For example, Lukic et al. (2014) reported increased abundance of NRF2,

KEAP1, and NF-κB within the cytoplasm of peripheral blood mononuclear cells of MDD patients (51). 351

352 This lack of consistent direction, in terms of up- or down- regulation of NRF signalling, is reflected in

353 our eQTL results, where BP genetic variation is associated with inconsistent changes in expression of

genes within the NRF2 KEAP gene-set. For example, the CUL3 gene encodes the cullin 3 protein, 354

355 which is a core component of ubiquitin ligase complex which ubiquitinates NRF2 and the PSMD2 gene

356 encodes a subunit of a proteasome that degrades NRF2. Therefore, CUL3 and PSMD2 function together

to degrade NRF2, yet BP eQTLs within these genes are associated with expression changes that are in opposite directions. Additionally, the four genes within the NRF2 KEAP1 gene-set that are differentially expressed between BP cases and controls, based on PsychEncode differential expression data, are upregulated.

361 The other pathway highlighted by our study is the *mRNA-splicing minor pathway*. Splicing of mRNA 362 refers to the removal of introns from pre-mRNA transcripts, a process performed by the spliceosome. The spliceosome comprises distinct complexes responsible for various mRNA splicing steps, composed 363 364 of both proteins and small nuclear RNAs. While the majority of splicing events are carried out by the major spliceosome, a subset of introns, known as minor introns, necessitates the involvement of the 365 minor spliceosome for their excision (52). Minor splicing is less efficient than major splicing, this 366 367 reduced efficiency has been proposed to play a role in the regulation of genes containing minor introns, as minor intron containing transcripts are retained in the nucleus until the minor spliceosome 368 369 components become available. Of relevance to NDPCs, the cerebral cortex is one of the tissues with 370 the highest expression of genes containing minor introns (53).

371 The importance of minor intron splicing is evident by the fact that minor-intron containing genes are 372 highly conserved and thirteen diseases have been linked to mutations in minor splicing components, 373 including the neurodegenerative diseases, amyotrophic lateral sclerosis, and spinal muscular atrophy 374 (52,54,55). A study by Buel and colleagues found oscillations in many components of the major spliceosome pathway mirror that of the clock and bmall transcripts (56), indicating that the major 375 376 spliceosomal pathway is at least partly regulated by the circadian clock. This is interesting in relation to the minor spliceosome too, given the overlap in these pathways. Interestingly, the biologically related 377 378 pathway, Processing of Capped Intron-Containing Pre-mRNA, ranked as the second highest performing 379 pathway PGS in the prediction of insomnia based on ADHD genetic variation. This pathway clustered

together with *mRNA-splicing minor pathway* in the heatmap in Figure 3.

381 Neither the NRF2 KEAP1 or the mRNA-splicing minor pathways were found to be enriched in genes 382 differentially expressed between BP cases and controls in the post-mortem dorsolateral prefrontal 383 cortex. It could be that these genes are differentially expressed in a different tissue type to the one we 384 investigated or that the shared genetic variation we have identified only applies to a subset of 385 individuals, and differential expression of these genes would only occur for this subset. A future analysis could select BP cases with high NRF2 KEAP1 / mRNA-splicing minor PGSs and restrict the differential 386 387 analysis to just these cases (this would require genotype and RNA-sequencing data on BP cases and 388 controls).

We implemented MAGMA gene-set analysis to investigate if the *NRF2 KEAP1* and *mRNA splicing minor* pathways could be identified using standard single phenotype pathway enrichment methods.
Neither the *NRF2 KEAP1* and *mRNA splicing minor pathways* were enriched for genetic variation

individually associated with BP, chronotype or SCZ (Supplementary Table 6). This indicates that these
pathways are enriched for the genetic variation shared between these two phenotypes and are not
enriched for the genetic variation associated with each individual phenotype.

395 There are some limitations to the pathway-based PGS method we have implemented. Firstly, we 396 mapped SNPs to genes solely based on the physical distance of SNPs to genes. While, it has been 397 reported that the majority (76%) of GWAS SNPs affect the closest gene (57), SNPs can be located in 398 regulatory elements up to hundreds of megabases away from the causal gene. Such distal causal genes 399 were not included in our analysis. Another limitation is that polygenic score analysis cannot distinguish between horizontal versus vertical pleiotropy i.e. whether genetic variation within the NRF2 KEAP1 400 401 and mRNA splicing minor pathways independently contribute to both BP and an evening chronotype, 402 or whether genetic variation within these pathways contribute to an evening chronotype, which in turn, causes BP, or vice versa. 403

This is the first study to investigate the biological pathways enriched in genetic variation overlapping between NDPC and sleep phenotypes, using pathway-based PGS analysis in a global, rather than hypothesis-based design. We used the largest available datasets from GWASs of ADHD, AUT, BP as our discovery datasets, and our target dataset consisted of up to 409,630 samples from UKB.

408 Our findings of shared genetic variation between chronotype and BP in a pathway that is known to 409 regulate circadian rhythms and seasonality in animals (*NRF2 KEAP1*) in addition to the strongly 410 seasonal nature of the clinical features and prevalence of BP, are further evidence to support the 411 hypothesis that disrupted circadian rhythm and seasonal adaptive responses underlie the pathology of 412 BP (58). Furthermore, our findings highlight the *mRNA splicing-minor* pathway, which has also been 413 shown to be at least partly regulated by the circadian clock. Thus, this study provides further evidence 414 to support the circadian clock hypothesis of BP disorder.

415

416 Methods

417 Target Data for Polygenic Score Analysis – UK Biobank

418 **Quality Control (QC)**

Genotype data was collected, processed and imputed by UKB (59). We accessed and analysed the data through the UKB Research Analysis Platform (Application ID: 71469). UKB research participants were excluded based on the following criteria: not being of Caucasian ancestry (data-field: 22006), containing a sex chromosomal aneuploidy (data-field: 22019), having a high SNP missingness and/or having unusually high or low heterozygosity (data-field: 22027), having ten or more third-degree relatives (data-field: 22021) and performing shift work (data-field: 826). Further samples were removed

425 based on relatedness using the UKB supplied relatedness file, which lists pairs of individuals related up 426 to a third degree. One individual from each pair was removed without removing samples from pairs 427 where a sample had already been removed. A total of 487,409 samples passed these quality control measures. To minimize any potential overlap between UKB and PGC SCZ and BP cohorts, participants 428 429 who had received SCZ or BP diagnoses were excluded from the construction of the SCZ PGS or BP PGS, respectively. Such samples were identified through self-reported diagnoses (data-field: 20544) 430 and linked health records reporting international classification of disease (ICD10) codes, F200-F209 431 for SCZ and F310-F319 for BP (data-field: 41270). This resulted in a further removal of 1,223 SCZ 432 433 samples and 1,866 BP samples, in the construction of the SCZ and BP PGSs, respectively.

SNPs were excluded based on the following criteria: imputation quality score (INFO) < 0.7, proportion 434

435 of missing genotypes > 0.02, minor allele frequency (MAF) < 0.005, Hardy–Weinberg equilibrium

 $(HWE) < 1 \times 10^{-6}$. Duplicate SNPs were removed using Plink2 with the flag "--rm-dup force-first". 436

- 437 Plink2 binary file-sets were created using the "--make-bed" flag and chromosome specific files were
- merged using the Plink2 "--merge-list" flag. Finally, ambiguous SNPs and SNPs not present in the 438
- 439 corresponding neuropsychiatric GWAS were removed.

440 **Phenotype Generation**

- 441 Two UKB phenotypes were used for this study: chronotype ("Morning/evening person (chronotype)";
- data-field 1180) and insomnia ("sleeplessness/insomnia"; data-field 1200). These data points were 442
- gathered as part of a questionnaire participants completed on touchscreen computers at UKB assessment 443
- 444 centres.
- 445 For chronotype, participants were asked the question, "do you consider yourself to be"?, and they had 446 six optional answers, which we encoded as a quantitative eveningness phenotype: "Definitely an 447 'evening' person" was encoded as 4, "More an 'evening' than a 'morning' person" was encoded as 3, "More a 'morning' than an 'evening' person" was encoded as 2, "Definitely a 'morning' person" was 448 encoded as 1, "Prefer not to answer" and "Do not know" were both encoded as missing. This resulted 449 450 in 409,630 post-QC samples with chronotype data.
- For insomnia, participants were asked the question, "do you consider yourself to be"? and had four 451 452 optional answers, which we encoded in a binary fashion: "never/rarely" was encoded as 0, "usually" was encoded as 1 and "sometimes" and "Prefer not to answer" were both encoded as missing. This 453
- 454 resulted in post-QC 239,918 samples with insomnia data.

Correcting for Confounders 455

456 The UKB phenotypes were corrected for the effects of confounders by taking the residuals of regression 457 models with the UKB phenotype as the dependent variable and the confounders as the independent

458 variables. For the binary insomnia phenotype, a logistic regression model was used, while a linear

459 regression model was used for the quantitative chronotype phenotype. The residuals of these models,

- 460 representing the variation in phenotype left unexplained by the effects of confounders, were assigned
- 461 as the corrected phenotypes. The fourteen confounders corrected for were: age, sex, UKB assessment
- centre, genotyping batch and genetic principal components (PCs) 1-10. 462

Discovery Data for Polygenic Score Analysis – GWAS Summary Statistics 463

Discovery data for polygenic score analysis included summary statistics from the latest and largest 464 published GWAS of AUT (18,381 cases and 27,969 controls from a Danish population (31)), ADHD 465 (38,691 cases and 186,843 controls from European, Danish and Icelandic populations (33)), SCZ 466 467 (74, 776)101,023 controls from cases and European, East Asian, 468 African American and Latino populations (30)) and BP (41,917 cases and 31,358 controls from a 469 European population (35)).

Genome-Wide Polygenic Score Analysis 470

471 Effect sizes for all imputed SNPs were estimated using SBayesRC based on GWAS summary statistics. 472 SBayesRC models SNP effects using a mixture of normal distributions that differ in their variances and 473 uses SNP annotation information to select which of these distributions best model SNP effect sizes. 474 This allows for SNP effect probability distributions to differ across annotation groups and has been 475 shown to perform better than other popular PGS methods (24). Inputs to SBayesRC include GWAS 476 summary statistics, LD European reference data and data containing 96 functional annotations for 8 477 million imputed **SNPs** (the latter two data files were downloaded from 478 github.com/zhilizheng/SBayesRC)

479 The first step involves converting GWAS summary statistics to COJO format, and converting odds 480 ratios to beta values using log(OR) when beta values were not provided. Subsequently, the SBayesRC 481 tidy function was run to filter SNPs based on inconsistent alleles between the GWAS summary statistic 482 and LD data and per SNP sample sizes less than three standard deviations of the mean. The SBayesRC impute function was implemented to impute the GWAS summary data based on LD. This resulted in 483 7,356,519 SNPs for each phenotype. The main SBayesRC function was then run to estimate SNP 484 485 effects. These SNP effects were provided to the --score command in Plink1.9 to calculate PGSs for all 486 individuals in the UKB test dataset (60). Performance metrics were calculated in R by running linear regressions with UKB chronotype or insomnia status as the dependent variable and the SCZ, BP, AUT 487 or ADHD PGS as the independent variables. P values and variance explained (R²) were obtained using 488 489 R's summary function. Confidence intervals were created using the ci rsquared function as part of the 490 confintr R package (https://github.com/mayer79/confintr).

491 Pathway-Based Polygenic Score Analysis

492 Pathway data

493 Pathway data for pathway based PGS analysis was sourced from Reactome, an open-source, manually 494 curated and peer-reviewed pathway database (61). We used all *Homo sapien* pathways that contained 495 greater than 50 genes, to remove overly specific pathways, and less than 500 genes, to remove overly 496 broad pathways and limit the computational intensity of pathway-based PGS analysis. This resulted in

497 451 pathways to test.

498 Pathway-based polygenic score analysis using PRSet

- 499 Polygenic score analysis was performed using the tool, PRSet (62). Inputs to PRSet included Reactome 500 pathway data, published GWAS summary statistics to be used as discovery data, UKB files in PLINK 501 binary format (.bed, .bim and .fam) to be used as target data and a GTF file for the genome build, 502 GRCh37. SNPs with a reported P > 0.5 from the corresponding GWAS were removed. This resulted in 503 final SNP numbers of N = 2,374,077 for the AUT PGS, N = 2,187,591 for the ADHD PGS, N =
- 504 2,525,546 for the SCZ PGS and N = 2,416,496 for the BP PGS.
- Pathway-based PGS analysis in PRSet involves three steps. The first step is to map SNPs to genes by 505 506 genomic position using GRCh37 start-stop coordinates extended 35kb upstream and 10kb downstream 507 to include potential regulatory elements. The second step is to perform pathway specific LD clumping 508 (parameters used: --clump-kb 250kb --clump-p 1 --clump-r2 0.1). Clumping is performed for each 509 pathway independently to retain the genetic signal for each pathway i.e to prevent a SNP outside of the pathway being assigned as the lead clumped SNP. The third step involves calculating pathway specific 510 PGSs for each individual, by summing up, for each SNP, the number of minor alleles at that SNP 511 multiplied by the GWAS effect size. Finally, performance metrics (R^2 and P value) are calculated for 512 513 each pathway. Competitive analysis was performed by creating 5,000 background polygenic scores, each constructed by randomly sampling the same number of post LD clumped SNPs as contained within 514 gene-set being tested and conservatively including the gene-set being tested as one background set. A 515 516 competitive P value is created using a formula that counts the number of times the P value for a background gene-set is less than the P value for the gene-set being tested and dividing by the number 517 518 of permutations (62). A disadvantage of this method is that the lowest possible P value is 1 / (number 519 of permutations + 1). Therefore, gene-sets that reached the lowest possible competitive P value at 5,000 520 permutations were re-analysed with a higher number of permutations. Confidence intervals were 521 calculated for the top performing pathways in R using the same method as for the genome-wide PGSs.
- 522 Multiple testing correction was performed for pathway-based PGS analysis by calculating a Bonferroni
- 523 corrected P value threshold, where 0.05 was divided by the number of tests performed (four GWAS
- 524 phenotypes x 451 pathways x 2 UKB phenotypes = 3,608).

525

526 **Post-hoc analysis**

527 MAGMA gene-set analysis

Gene-sets that were statistically significant in the PGS analysis, and the genes within them, were tested 528 for enrichment of common genetic variation using region-based multi-marker analysis of genomic 529 530 annotation (MAGMA; http://ctg.cncr.nl/software/magma; (63) and summary statistics from published 531 GWAS on chronotype (69,369 cases and 236,642 controls of British ancestry; 3) and BP. An analysis involves three steps. First, in the annotation step, SNPs with available GWAS results are mapped on to 532 533 genes (GRCh37/hg19 start-stop coordinates +35kb/-10kb). Second, in the gene analysis step, gene P 534 values are computed for each GWAS dataset. This gene analysis is based on a multiple linear principal 535 components regression model that accounts for LD between SNPs in each gene, number of SNPs in 536 each gene, inverse of the mean minor allele count of variants in each gene and the GWAS sample size. 537 The European panel of the 1000 Genomes data was used as a reference panel for LD. Third, a 538 competitive GSA based on the gene P values, also using a regression structure, was used to test if the 539 genes in each gene-set were more strongly associated with either phenotype than other genes in the 540 genome.

541 eQTL Analysis

542 Expression quantitative trait loci (eQTL) data was examined using QTLbase2 (27), a database that 543 aggregates published genome-wide QTL summary statistics across more than 95 tissue and cell types. 544 We reviewed eQTL data related to genes found within the gene-sets prioritised by pathway-based PGS analysis. Specifically, we considered genes that achieved significance levels of $P < 1x10^{-2}$ for BP and P 545 $< 1 \times 10^{-3}$ for chronotype in the gene-based MAGMA analysis. This was to identify genes individually 546 547 enriched for BP and chronotype genetic variation, as these are the genes most likely contributing to the significant pathway-based PGS results and therefore of interest. Since these genes were already 548 549 highlighted as being of interest through pathway-based PGS analysis, it was not necessary that they 550 were individually statistically significant in the MAGMA analysis.

The most statistically significant SNP at each gene locus was searched in QTLbase to identify if it had been associated with expression changes of the mapped gene in biologically relevant tissues or cell types, including those related to the brain, central nervous system, peripheral nervous system, blood, or immune systems. If the top SNP at a locus did not show any eQTL associations, we investigated other SNPs with similarly low P values at that locus. It was assumed that the SNPs demonstrating the highest statistical significance for association with BP would also contribute to chronotype prediction, given that BP effect sizes were used to compute the pathway-based PGSs.

558 Differential Expression Analysis

- 559 Differential expression summary data was downloaded as Supplementary Table 1 from Gandal et al.
- 560 (2018). This is a PsychENCODE Consortium study that performed RNA-sequencing, followed by
- 561 differential expression analysis, of samples isolated form the postmortem dorsolateral prefrontal cortex
- samples for 144 BP cases and 899 controls. Specific details of the analysis performed are provided in
- 563 the paper (64)
- 564 More specific methodology details and code for all analyses are provided here:
- 565 https://github.com/laurafahey02/ Polygenic_Score_Analyses.

566 Acknowledgements

567 This project has received funding from the European Research Council (ERC) under the European 568 Union's Horizon 2020 research and innovation programme (grant agreement No 950010). We would 569 like to thank Dr Cathy Wyse for discussions related to chronobiology and for reviewing a draft of the 570 manuscript.

571

572 Bibliography

573	1.	Scandurra V, Emberti Gialloreti L, Barbanera F, Scordo MR, Pierini A, Canitano R.
574		Neurodevelopmental Disorders and Adaptive Functions: A Study of Children With
575		Autism Spectrum Disorders (ASD) and/or Attention Deficit and Hyperactivity Disorder
576		(ADHD). Front Psychiatry. 2019;10:673.
577	2.	American Psychiatric Association. Diagnostic and Statistical Manual of Mental
578		Disorders. American Psychiatric Association; 2013.
579	3.	Cohrs S. Sleep Disturbances in Patients with Schizophrenia. CNS Drugs.
580		2008;22(11):939–62.
581	4.	Cortesi F, Giannotti F, Ivanenko A, Johnson K. Sleep in children with autistic spectrum
582		disorder. Sleep Med. 2010 Aug;11(7):659–64.
583	5.	Galli J, Loi E, Visconti LM, Mattei P, Eusebi A, Calza S, et al. Sleep Disturbances in
584		Children Affected by Autism Spectrum Disorder. Front Psychiatry. 2022 Feb 17;13.
585	6.	Wajszilber D, Santisteban JA, Gruber R. Sleep disorders in patients with ADHD: impact
586		and management challenges. Nat Sci Sleep. 2018 Dec;Volume 10:453-80.

587 588 589	7.	Deboer T. Sleep homeostasis and the circadian clock: Do the circadian pacemaker and the sleep homeostat influence each other's functioning? Neurobiol Sleep Circadian Rhythms. 2018 Jun;5:68–77.
590 591 592	8.	Zhang Y, Quiñones GM, Ferrarelli F. Sleep spindle and slow wave abnormalities in schizophrenia and other psychotic disorders: Recent findings and future directions. Schizophr Res. 2020 Jul;221:29–36.
593 594 595	9.	D'Agostino A, Castelnovo A, Cavallotti S, Casetta C, Marcatili M, Gambini O, et al. Sleep endophenotypes of schizophrenia: slow waves and sleep spindles in unaffected first-degree relatives. NPJ Schizophr. 2018 Feb 9;4(1):2.
596 597 598	10.	Alachkar A, Lee J, Asthana K, Vakil Monfared R, Chen J, Alhassen S, et al. The hidden link between circadian entropy and mental health disorders. Transl Psychiatry. 2022 Jul 14;12(1):281.
599 600	11.	Logan RW, McClung CA. Rhythms of life: circadian disruption and brain disorders across the lifespan. Nat Rev Neurosci. 2019 Jan;20(1):49–65.
601 602 603	12.	Coogan AN, McGowan NM. A systematic review of circadian function, chronotype and chronotherapy in attention deficit hyperactivity disorder. ADHD Attention Deficit and Hyperactivity Disorders. 2017 Sep 7;9(3):129–47.
604 605 606	13.	Bumb JM, Enning F, Mueller JK, van der List T, Rohleder C, Findeisen P, et al. Differential melatonin alterations in cerebrospinal fluid and serum of patients with major depressive disorder and bipolar disorder. Compr Psychiatry. 2016 Jul;68:34–9.
607 608 609	14.	Brown JP, Martin D, Nagaria Z, Verceles AC, Jobe SL, Wickwire EM. Mental Health Consequences of Shift Work: An Updated Review. Curr Psychiatry Rep. 2020 Feb 18;22(2):7.
610 611 612	15.	von Schantz M, Leocadio-Miguel MA, McCarthy MJ, Papiol S, Landgraf D. Genomic perspectives on the circadian clock hypothesis of psychiatric disorders. In 2021. p. 153–91.
613	16.	von Schantz M. Natural Variation in Human Clocks. In 2017. p. 73–96.
614 615	17.	Roenneberg T. How can social jetlag affect health? Nat Rev Endocrinol. 2023 Jul 23;19(7):383–4.
616 617	18.	Zou H, Zhou H, Yan R, Yao Z, Lu Q. Chronotype, circadian rhythm, and psychiatric disorders: Recent evidence and potential mechanisms. Front Neurosci. 2022;16:811771.

618 619 620	19.	Jones SE, Lane JM, Wood AR, van Hees VT, Tyrrell J, Beaumont RN, et al. Genome- wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nat Commun. 2019 Dec 29;10(1):343.
621 622 623	20.	Mistry S, Harrison JR, Smith DJ, Escott-Price V, Zammit S. The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: Systematic review. Schizophr Res. 2018 Jul;197:2–8.
624 625 626 627	21.	Ohi K, Ochi R, Noda Y, Wada M, Sugiyama S, Nishi A, et al. Polygenic risk scores for major psychiatric and neurodevelopmental disorders contribute to sleep disturbance in childhood: Adolescent Brain Cognitive Development (ABCD) Study. Transl Psychiatry. 2021;11(1):187.
628 629 630 631	22.	Carpena MX, Bonilla C, Matijasevich A, Martins-Silva T, Genro JP, Hutz MH, et al. Sleep-related traits and attention-deficit/hyperactivity disorder comorbidity: Shared genetic risk factors, molecular mechanisms, and causal effects. The World Journal of Biological Psychiatry. 2021 Nov 26;22(10):778–91.
632 633 634	23.	O'Connell KS, Frei O, Bahrami S, Smeland OB, Bettella F, Cheng W, et al. Characterizing the Genetic Overlap Between Psychiatric Disorders and Sleep-Related Phenotypes. Biol Psychiatry. 2021 Nov;90(9):621–31.
635 636 637 638	24.	Zheng Z, Liu S, Sidorenko J, Yengo L, Turley P, Ani A, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. bioRxiv [Internet]. 2022 Jan 1;2022.10.12.510418. Available from: http://biorxiv.org/content/early/2022/10/14/2022.10.12.510418.abstract
639 640	25.	Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience. 2019 Jul 1;8(7).
641 642 643	26.	Ni G, Zeng J, Revez JA, Wang Y, Zheng Z, Ge T, et al. A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. Biol Psychiatry. 2021 Nov 1;90(9):611–20.
644 645 646	27.	Zheng Z, Huang D, Wang J, Zhao K, Zhou Y, Guo Z, et al. QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. Nucleic Acids Res. 2020 Jan 8;48(D1):D983–91.
647 648 649	28.	Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016 Jun 20;54:1.30.1-1.30.33.

650 651 652	29.	Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022 Apr 21;604(7906):502–8.
653 654 655	30.	Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature. 2022 Apr;604(7906):502–8.
656 657 658	31.	Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019 Mar 25;51(3):431–44.
659 660 661	32.	Jones SE, Lane JM, Wood AR, van Hees VT, Tyrrell J, Beaumont RN, et al. Genome- wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. Nat Commun. 2019 Jan 29;10(1):343.
662 663 664	33.	Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019 Jan 26;51(1):63–75.
665 666 667	34.	Watanabe K, Jansen PR, Savage JE, Nandakumar P, Wang X, Agee M, et al. Genome- wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. Nat Genet. 2022 Aug 14;54(8):1125–32.
668 669 670	35.	Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021 Jun 17;53(6):817–29.
671 672 673	36.	Bhandari R, Kaur J, Kaur S, Kuhad A. The Nrf2 pathway in psychiatric disorders: pathophysiological role and potential targeting. Expert Opin Ther Targets. 2021 Feb 1;25(2):115–39.
674 675 676	37.	Early JO, Menon D, Wyse CA, Cervantes-Silva MP, Zaslona Z, Carroll RG, et al. Circadian clock protein BMAL1 regulates IL-1β in macrophages via NRF2. Proc Natl Acad Sci U S A. 2018 Sep 4;115(36):E8460–8.
677 678 679	38.	Wible RS, Ramanathan C, Sutter CH, Olesen KM, Kensler TW, Liu AC, et al. NRF2 regulates core and stabilizing circadian clock loops, coupling redox and timekeeping in Mus musculus. Elife. 2018 Feb 26;7.

680	39.	Tamaru T, Hattori M, Ninomiya Y, Kawamura G, Varès G, Honda K, et al. ROS Stress
681		Resets Circadian Clocks to Coordinate Pro-Survival Signals. PLoS One. 2013 Dec
682		2;8(12):e82006.
683	40.	Teleanu DM, Niculescu AG, Lungu II, Radu CI, Vladâcenco O, Roza E, et al. An
684		Overview of Oxidative Stress, Neuroinflammation, and Neurodegenerative Diseases. Int
685		J Mol Sci. 2022 May 25;23(11).
686	41.	Salim S. Oxidative stress and psychological disorders. Curr Neuropharmacol. 2014
687		Mar;12(2):140–7.
688	42.	Morris G, Walker AJ, Walder K, Berk M, Marx W, Carvalho AF, et al. Increasing Nrf2
689		Activity as a Treatment Approach in Neuropsychiatry. Mol Neurobiol. 2021
690		May;58(5):2158–82.
691	43.	Gao W, Guo L, Yang Y, Wang Y, Xia S, Gong H, et al. Dissecting the Crosstalk
692		Between Nrf2 and NF-KB Response Pathways in Drug-Induced Toxicity. Front Cell Dev
693		Biol. 2022 Feb 2;9.
694	44.	Nakayama T, Okimura K, Shen J, Guh YJ, Tamai TK, Shimada A, et al. Seasonal
695		changes in NRF2 antioxidant pathway regulates winter depression-like behavior.
696		Proceedings of the National Academy of Sciences. 2020 Apr 28;117(17):9594-603.
697	45.	Geoffroy PA, Bellivier F, Scott J, Etain B. Seasonality and bipolar disorder: A systematic
698		review, from admission rates to seasonality of symptoms. J Affect Disord. 2014
699		Oct;168:210–23.
700	46.	Reynaud E, Berna F, Haffen E, Weiner L, Maruani J, Lejoyeux M, et al. Validity and
701		Usage of the Seasonal Pattern Assessment Questionnaire (SPAQ) in a French Population
702		of Patients with Depression, Bipolar Disorders and Controls. J Clin Med. 2021 Apr
703		27;10(9):1897.
704	47.	Roguski A, Ritter P, Smith DJ. Sensitivity to light in bipolar disorder: implications for
705		research and clinical practice. The British Journal of Psychiatry. 2024 Jan 4;1–4.
706	48.	Bjorvatn B, Saxvig IW, Waage S, Pallesen S. Self-reported seasonality is strongly
707		associated with chronotype and weakly associated with latitude. Chronobiol Int. 2021
708		Feb 1;38(2):278–85.
709	49.	Menculini G, Steardo LJr, Verdolini N, D'Angelo M, Chipi E, Cirimbilli F, et al.
710		Chronotype is associated with affective temperaments, clinical severity and worse

711 712		treatment outcomes in bipolar disorders: results from a two-center, cross-sectional study. Int J Psychiatry Clin Pract. 2023 Sep 1;27(3):248–56.
713 714	50.	Hashimoto K. Essential Role of Keap1-Nrf2 Signaling in Mood Disorders: Overview and Future Perspective. Front Pharmacol. 2018;9:1182.
715 716 717	51.	Lukic I, Mitic M, Djordjevic J, Tatalovic N, Bozovic N, Soldatovic I, et al. Lymphocyte levels of redox-sensitive transcription factors and antioxidative enzymes as indicators of pro-oxidative state in depressive patients. Neuropsychobiology. 2014;70(1):1–9.
718 719	52.	El Marabti E, Malek J, Younis I. Minor Intron Splicing from Basic Science to Disease. Int J Mol Sci. 2021 Jun 4;22(11):6062.
720 721 722	53.	Olthof AM, Hyatt KC, Kanadia RN. Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. BMC Genomics. 2019 Dec 30;20(1):686.
723 724 725	54.	Reber S, Stettler J, Filosa G, Colombo M, Jutzi D, Lenzken SC, et al. Minor intron splicing is regulated by <scp>FUS</scp> and affected by <scp>ALS</scp> -associated <scp>FUS</scp> mutants. EMBO J. 2016 Jul 15;35(14):1504–21.
726 727 728	55.	Jangi M, Fleet C, Cullen P, Gupta S V., Mekhoubad S, Chiao E, et al. SMN deficiency in severe models of spinal muscular atrophy causes widespread intron retention and DNA damage. Proceedings of the National Academy of Sciences. 2017 Mar 21;114(12).
729 730 731 732	56.	Buel SM, Debopadhaya S, De los Santos H, Edwards KM, David AM, Dao UH, et al. The PAICE suite reveals circadian posttranscriptional timing of noncoding RNAs and spliceosome components in <i>Mus musculus</i> macrophages. G3 Genes Genomes Genetics. 2022 Aug 25;12(9).
733 734 735	57.	Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, et al. Genome-wide enhancer maps link risk variants to disease genes. Nature. 2021 May;593(7858):238–43.
736 737 738 739	58.	McCarthy MJ, Gottlieb JF, Gonzalez R, McClung CA, Alloy LB, Cain S, et al. Neurobiological and behavioral mechanisms of circadian rhythm disruption in bipolar disorder: A critical multi-disciplinary literature review and agenda for future research from the ISBD task force on chronobiology. Bipolar Disord. 2022 May 10;24(3):232–63.
740 741	59.	Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018 Oct;562(7726):203–9.

742	60.	Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK:
743		A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.
744		The American Journal of Human Genetics. 2007 Sep;81(3):559–75.
745	61.	Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The
746		Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018 Jan 4;46(D1):D649-55.
747	62.	Choi SW, García-González J, Ruan Y, Wu HM, Porras C, Johnson J, et al. PRSet:
748		Pathway-based polygenic risk score analyses and software. PLoS Genet. 2023
749		Feb;19(2):e1010624.
750	63.	de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set
751		analysis of GWAS data. PLoS Comput Biol. 2015 Apr;11(4):e1004219.
752	64.	Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S, et al. Transcriptome-
753		wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science.
754		2018 Dec 14;362(6420).



756

757

Figure 2. Genome-wide polygenic score analysis results. Linear regressions were run with UK Biobank chronotype or insomnia status as the dependent variable and SCZ, BP, AUT or ADHD polygenic score as the independent variables. UK Biobank phenotypes were corrected for sex, genotyping batch, UK Biobank assessment centre and the top ten ancestry-informative genetic principal components prior to running this regression. Horizontal bars show the variance explained (R^2), error bars indicate 95% confidence intervals. Bars marked with an asterisk (*) indicate that they surpass the Bonferronicorrected *p* value threshold of $p \le 0.0001$.



764

Figure 3. Top performing pathways from pathway-based polygenic score analysis. Performance metrics for pathway-based polygenic score analysis were calculated by regressing insomnia or chronotype case status on each of the pathway-based polygenic scores. Competitive P values were calculated by creating at least 5,000 background polygenic scores, consisting of random sets of post-clump SNPs, the same size as the pathway under test (in this way, pathway size is controlled for). The competitive P value takes into account the number of times a background gene-set obtains a lower null P value than the pathway being tested. The four pathways that resulted in the lowest competitive P values for each of the eight phenotype pairs were identified, and their competitive P values across each of the eight phenotype pairs were plotted as a heatmap. The white stars represent pathway based polygenic scores statistically significant after multiple testing correction ($P < 1.39x10^{-5}$). Rows are clustered based on similarity.



Figure 4. Variance explained (\mathbb{R}^2) by the statistically significant pathway-based BP PGSs, and their subpathways, in predicting chronotype. Bars marked with an asterisk (*) indicate that they surpass the Bonferroni-corrected P value threshold of $P \le 1.39 \times 10^{-5}$. A) Total variance explained per pathway. Horizontal bars show the variance explained (\mathbb{R}^2), error bars indicate 95% confidence intervals. B) Average variance explained per SNP. This was obtained by dividing the total variance explained per pathway by the number of SNPs in that pathway.