

RESEARCH

Open Access

Genetic adaptation of the human circadian clock to day-length latitudinal variations and relevance for affective disorders

Diego Forni^{1†}, Uberto Pozzoli^{1†}, Rachele Cagliani¹, Claudia Tresoldi¹, Giorgia Menozzi¹, Stefania Riva¹, Franca R Guerini², Giacomo P Comi³, Elisabetta Bolognesi², Nereo Bresolin^{1,3}, Mario Clerici^{2,4} and Manuela Sironi^{1*}

Abstract

Background: The temporal coordination of biological processes into daily cycles is a common feature of most living organisms. In humans, disruption of circadian rhythms is commonly observed in psychiatric diseases, including schizophrenia, bipolar disorder, depression and autism. Light therapy is the most effective treatment for seasonal affective disorder and circadian-related treatments sustain antidepressant response in bipolar disorder patients. Day/night cycles represent a major circadian synchronizing signal and vary widely with latitude.

Results: We apply a geographically explicit model to show that out-of-Africa migration, which led humans to occupy a wide latitudinal area, affected the evolutionary history of circadian regulatory genes. The SNPs we identify using this model display consistent signals of natural selection using tests based on population genetic differentiation and haplotype homozygosity. Signals of natural selection driven by annual photoperiod variation are detected for schizophrenia, bipolar disorder, and restless leg syndrome risk variants, in line with the circadian component of these conditions.

Conclusions: Our results suggest that human populations adapted to life at different latitudes by tuning their circadian clock systems. This process also involves risk variants for neuropsychiatric conditions, suggesting possible genetic modulators for chronotherapies and candidates for interaction analysis with photoperiod-related environmental variables, such as season of birth, country of residence, shift-work or lifestyle habits.

Background

The temporal coordination of biological processes into daily cycles is a common feature of most living organisms. Although circadian cycles are determined by the presence of an internal cell-autonomous clock, they are synchronized (entrained) by environmental cues, most importantly visible light and external temperature [1].

In mammals, the suprachiasmatic nucleus (SCN) represents the central circadian pacemaker. SCN neurons sustain cell-autonomous cycles and, through the retinohypothalamic tract, receive inputs from melanopsin-expressing photosensitive retinal ganglion cells (RGC). Thus, although many peripheral tissues display autonomous circadian oscillations, the SCN hierarchically coordinates

internal rhythms by providing a link with the external environment [2].

At the molecular level, the mammalian core circadian circuit involves the CLOCK, ARNTL (also known as BMAL1) and NPAS2 transcription factors, which activate the transcription of cryptochrome (CRY1 and CRY2), period (PER1, PER2 and PER3) and other clock-controlled genes [3]. CRY/PER heterodimers translocate back into the nucleus and inhibit their own transcription by acting on the CLOCK/ARNTL complex. The degradation of CRY and PER relieves the inhibition and initiates a new cycle [3].

Detrimental effects for health and fitness accompany alterations of circadian rhythms, due either to genetic defects or to changes in external variables that function as entrainment cues [4,5]. Notably, disruption of circadian rhythms is a common feature of human psychiatric diseases including schizophrenia (SCZ), bipolar disorder (BPD) and autism, with seasonal affective disorder (SAD) being a

* Correspondence: manuela.sironi@bp.lnf.it

†Equal contributors

¹Scientific Institute IRCCS E. Medea, 23842 Bosisio Parini, LC, Italy
Full list of author information is available at the end of the article

common condition characterized by the occurrence of depressive symptoms during short winter days [2]. Light therapy is the most effective treatment for SAD [6] and circadian-related treatments (sleep deprivation and bright light) sustain antidepressant response in patients with BPD [7].

These observations, together with the notion that all species, from bacteria to plants and animals, have evolved circadian timing systems [1], suggest that genetic adaptations to seasonal variations in day length (photoperiod) are widespread. In animals, signatures of latitude-driven natural selection at circadian genes have been described for *Drosophila* [8], birds [9,10] and fishes [11].

Anatomically modern humans appeared in East Africa about 200,000 years ago, with archeological evidence placing the origin of our species in the Awash Valley of Ethiopia [12], a region immediately north of the equator. Equatorial regions are characterized by roughly 12 h/12 h day/night cycles; thus, the annual minimum and maximum photoperiod are almost identical (that is, day length tends to be constant throughout the year). Out-of-Africa migration led humans to occupy a wide latitudinal area where seasonal variation in photoperiod can be wide. We tested the hypothesis that seasonal photoperiod variation (that is, the occurrence of day/night cycles that deviate from the 12 h/12 h pattern) acted as a selective pressure. We suggest that this process influenced adaptive evolution at circadian regulatory loci and at risk variants for psychiatric and neurologic diseases.

Results

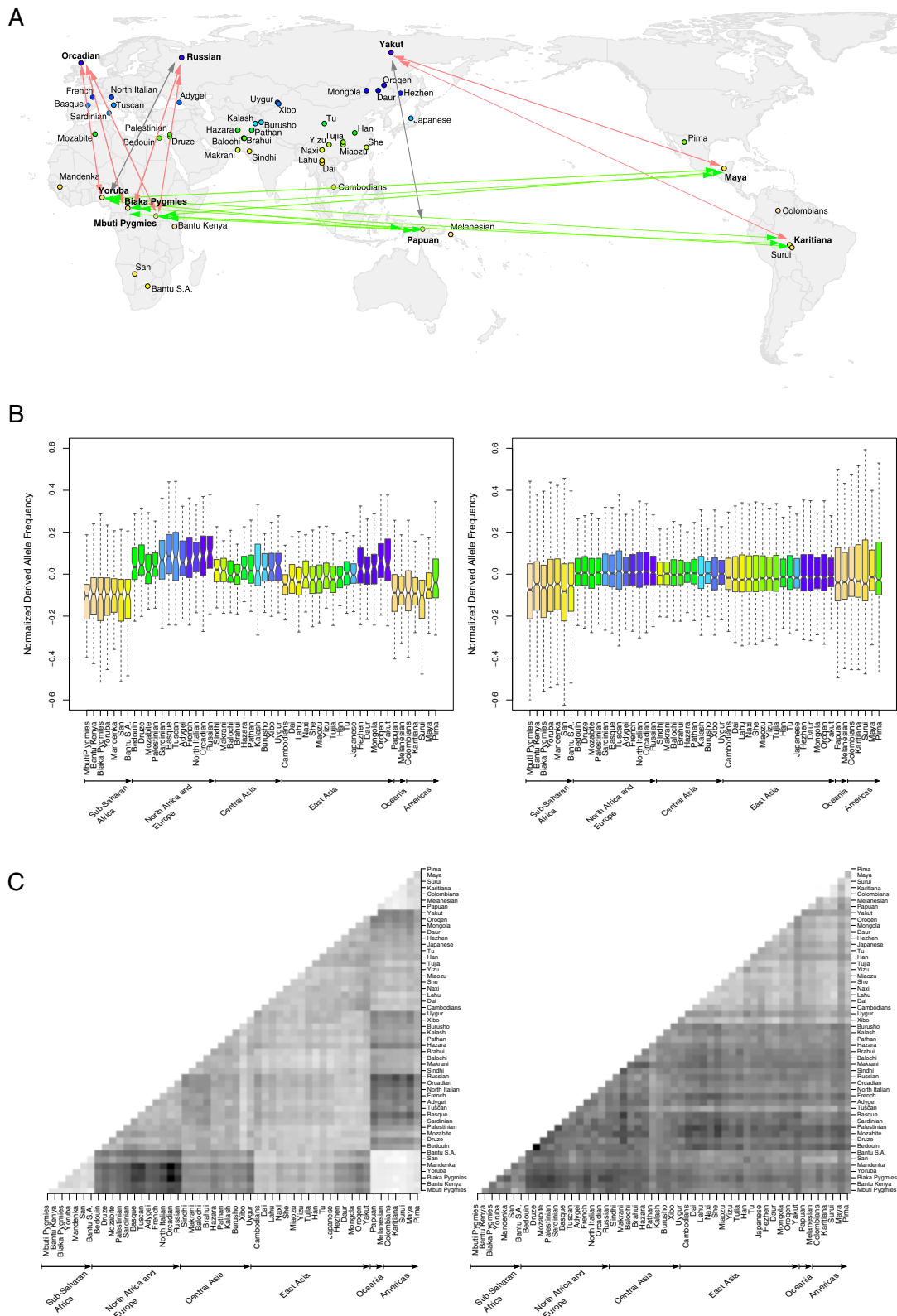
Day-length variation acts as a selective pressure on genes involved in circadian rhythms

Geographically explicit models are a powerful tool to study adaptation to environmental pressures [13-20]. Thus, we applied a previously developed approach that analyzes spatial correlations between genetic variation and environmental factors [15,16] to test the hypothesis that, during out-of Africa migration, seasonal variation of annual day length acted as a selective pressure on circadian clock genes. Briefly, we analyzed genotype data from 52 human populations distributed worldwide (Human Genome Diversity Project-Centre d'Étude du Polymorphisme Humain Human Genome Diversity Cell Line Panel (HGDP-CEPH panel) [21] and determined the annual minimum and maximum photoperiod for the geographic region where each population is located (Additional file 1, Figure 1A). The difference between the maximum and minimum photoperiod (hereafter referred to as Δ photoperiod) is a measure of the deviation from the almost constant annual day length observed at equatorial regions, and is hypothesized to represent a selective pressure. For all single nucleotide polymorphisms (SNPs) in the HGDP-CEPH panel ($n = 660,832$) we calculated Kendall's rank correlation

between allele frequencies and Δ photoperiod. Because genetic diversity in humans is affected by demographic factors [22,23], each SNP was assigned a percentile rank in the distribution of Kendall's correlation coefficients (τ) calculated for all SNPs with a similar (in the 1% range) minor allele frequency (MAF) calculated over all populations. This procedure was based on the assumption that demography affects the whole genome (and will be reflected in the distribution of τ values), whereas selection is a locus-specific force. Thus, SNPs that ranked high in the distribution of τ are more likely to have experienced a selective force (in addition to the demographic effect). The binning in MAF classes corrected for the fact that the power of the correlation tests was also affected by the overall SNP frequency in populations (because low MAF values result in several similar frequency values among populations - that is, several ties).

We first assessed whether 12 genes that compose the core circadian pacemaker (*CLOCK*, *PER1*, *PER2*, *PER3*, *NPAS2*, *ARNTL*, *CRY1*, *CRY2*, *CSNK1E*, *CSNK1D*, *NR1D1* and *ARNTL2*) [3] carry SNPs that correlate with Δ photoperiod (Additional file 2). Specifically, we considered a variant to significantly correlate with this parameter if it displayed a Kendall's correlation P value <0.05 after Bonferroni correction for the number of SNPs being analyzed (175 in the case of circadian core components) and a τ percentile rank >0.95 . Seven circadian genes carried at least one variant significantly correlated with Δ photoperiod (Table 1, Additional file 2). By performing 10,000 resamplings of 12 randomly selected human genes (see Methods for details), we verified that the empirical probability of obtaining seven or more genes with at least one significant SNP amounts to 0.031, indicating that core circadian genes are more likely than expected to carry variants correlated to Δ photoperiod. Because these loci represent a small set, we next analyzed 267 genes that have been identified in a large-scale RNA interference (RNAi) screen as modifiers of circadian rhythms in U2OS cells, a human osteosarcoma cell line (Additional file 2) [24]. After removing core circadian genes (to make the two sets independent) and genes that have no SNP genotyped in the HGDP-CEPH panel, 223 RNAi hits remained; of these, 40 carried variants that significantly correlated with Δ photoperiod (Table 1, Additional file 2). Using a resampling approach as described above, we calculated that the empirical probability of observing 40 or more significant genes is 0.043.

We next analyzed genes that, when mutated or over-expressed, affect circadian rhythmicity *in vivo*. To this aim, we searched the Mouse Genome Informatics (MGI) resource for mutant strains showing circadian disturbance. Because regulation of sleep/wake cycles is a major physiological output of the circadian clock, genes that determine abnormal sleep patterns in mice were also included. After excluding core circadian genes and RNAi



(See figure on previous page.)

Figure 1 F_{ST} (Fixation index) and allele frequency analysis. (A) HGDP-CEPH populations are shown on a world map and color coded according to their Δ photoperiod (in topological color scale, minimum = orange, maximum = dark blue). Arrows denote populations included in F_{ST} pairwise comparisons, with red and green indicating a significant excess of variants with an F_{ST} rank >0.95 or <0.05 , respectively (among the 84 SNPs in circadian rhythm genes, see text). Gray arrows denote pairwise comparisons that showed non-significant excess of high F_{ST} variants. Fisher's exact test P values (see text) for F_{ST} comparisons are as follows: Orcadian-Biaka Pygmy, 0.02; Orcadian-Mbuti Pygmy, 0.05; Orcadian-Yoruba, 0.05; Russian-Biaka Pygmy, 0.03; Russian-Mbuti Pygmy, 0.03; Russian-Yoruba, 0.08; Papuan-Biaka Pygmy, <0.01 ; Papuan-Mbuti Pygmy, <0.01 ; Papuan-Yoruba, <0.01 ; Papuan-Yakut, 0.12; Yakut-Maya, 0.03; Yakut-Karitiana, 0.02; Maya-Biaka Pygmy, <0.01 ; Maya-Mbuti Pygmy, <0.01 ; Maya-Yoruba, <0.01 ; Karitiana-Biaka Pygmy, <0.01 ; Karitiana-Mbuti Pygmy, <0.01 ; Karitiana-Yoruba, <0.01 . **(B)** Box-plot of normalized derived allele frequencies for the 84 SNPs significantly correlated with Δ photoperiod (left) and for 840 minor allele frequency (MAF)-matched control variants (right). Populations are color coded as in panel (A) and are grouped in six broad geographic areas; within these areas populations are ordered according to increasing Δ photoperiod. **(C)** Pairwise F_{ST} for all HGDP-CEPH populations calculated for the 84 Δ photoperiod-selected variants (left) and for 840 MAF-matched control SNPs (right). Pairwise F_{ST} percentile ranks are reported in gray scale (F_{ST} rank increasing with gray shading). Populations are ordered as in **(A)**.

hits, we identified 107 loci (Additional file 2). Eighty-two of these had been included in the HGDP-CEPH panel and 23 carried variants that correlate with Δ photoperiod (Table 1). Again, these genes are more like to carry significant SNPs than expected if randomness alone were responsible (empirical $P = 0.025$).

Likewise, we analyzed the Online Mendelian Inheritance in Man (OMIM) and PhenomicDB databases to compile a list of genes that have been associated with sleep pattern disturbance in humans. Specifically, we included genes only if the phenotype (either syndromic or not) resulted from mutation and could be ascribed to a single gene. Although sleep disturbances have been observed in patients with neurodegenerative diseases (for example, Parkinson's and Huntington's disease), the causal genes were not considered because sleep problems in these conditions might be secondary to the general disruption of neural circuitry [2]. Also, we excluded *CSNK1D* and *PER2* (circadian core set), whose mutations cause familial advanced sleep phase syndrome, as well as *RAI1*, *BHLHE41*, *UBE3A*, and *HCRT* (responsible for Smith-Magenis/Potocki-Lupski syndromes, short sleep phenotype, Angelman syndrome and narcolepsy, respectively), which had already been included in the mouse strain gene set. Thus, we obtained a list of 11 genes, seven of which carry SNPs that correlated with Δ photoperiod (Table 1; Additional file 2). As above, this number is significantly higher than expected from random sampling ($P = 0.0177$).

Finally, we analyzed a small set of genes ($n = 13$) that transduce photic inputs in melanopsin-expressing RGCs; these cells sustain non-image-forming responses to light, including circadian entrainment and light-induced melatonin suppression. Again, these genes show more SNPs that correlated with Δ photoperiod than expected by chance (seven significant genes, empirical $P = 0.0476$) (Table 1, Additional file 2). We next merged together all gene sets mentioned above (core circadian, RNAi hit, mouse circadian disturbance, human sleep disturbance and melanopsin-signaling genes). Analysis of these 341 genes confirmed that they are significantly more likely than

expected to carry variants that correlated with Δ photoperiod (empirical $P = 0.0335$). To confirm that the results we obtained were not influenced by the SNP content of the gene sets (for example, large genes with many SNPs may be more likely to carry at least one variant that correlates with Δ photoperiod), we performed resampling analyses using SNP-matched random gene sets (Additional files 3, 4 and 5). This approach largely confirmed the results detailed above (Additional file 4).

As a further control, the same analyses were performed using environmental factors different from photoperiod variation, namely annual minimum and maximum temperature, Δ temperature (the difference between the two previous measures), and annual short-wave (UV) radiation flux. These variables are strongly correlated with Δ photoperiod, as they all depend on latitude (Kendall's correlation coefficients with Δ photoperiod for annual minimum temperature = 0.75, maximum temperature = -0.35, Δ temperature = 0.68 and radiation flux = -0.50; all P values <0.0001) and may contribute to circadian entrainment [1]. In a total of 20 resampling analyses (five genes sets and four variables), only the core circadian gene set was found to display more variants than expected that correlate with Δ temperature (Additional file 6), suggesting that Δ photoperiod exerted the strongest pressure on the analyzed genes.

Thus, we analyzed five independent sets of genes involved in circadian regulation or in sleep homeostasis and observed that they are more likely than expected by chance to carry variants showing signals of Δ photoperiod-driven selection. Notably, genes included in those five sets do not represent a full inventory of genes with a known role in circadian rhythm regulation; rather, they were selected because they could be ascribed to specific categories through dedicated searches.

In total we analyzed 406 genes in the five independent sets; 341 of these had at least one SNP in the HGDP-CEPH panel and 84 displayed signals of Δ photoperiod-driven selection. The majority ($n = 273$) of the full gene set ($n = 406$) could be included in a protein-protein interaction network (Figure 2) that comprises all core circadian

Table 1 Single nucleotide polymorphisms that correlate with Δ photoperiod

SNP	Gene	Kendall's correlation P value ^a	τ rank ^b	F _{ST} rank ^c		LnRsb rank ^d		Functional annotation ^e
				Biaka-Orcadian	Biaka-Maya	CEPH	CHB + JPT	
Core circadian								
rs11048980	<i>ARNTL2</i>	7.167×10^{-8}	1.000	0.506	0.000	0.79	0.89	intronic
rs6811520	<i>CLOCK</i>	1.444×10^{-5}	0.978	0.831	0.466	0.84	0.87	intronic
rs11038695	<i>CRY2</i>	1.626×10^{-4}	0.957	0.624	0.247	0.64	0.33	intronic; in LD ($r^2 = 0.93$ in CEPH) with rs11038697 (within ARNTL and CLOCK binding sites)
rs4789846	<i>CSNK1D</i>	5.546×10^{-5}	0.970	0.601	0.000	0.69	0.84	intronic
rs5995572	<i>CSNK1E</i>	1.053×10^{-5}	0.984	0.973	0.515	0.05	0.14	intronic
rs17654772	<i>NPAS2</i>	2.549×10^{-7}	0.998	0.615	0.377	0.84	0.43	intronic
rs4663868	<i>PER2</i>	9.462×10^{-5}	0.962	0.286	0.219	0.97	0.90	intronic
Circadian hits (RNAi screen)								
rs2583836	<i>ABL1</i>	3.071×10^{-7}	0.995	0.936	0.311	0.97	0.63	intronic
rs12713670	<i>ANTXR1</i>	4.331×10^{-9}	1.000	0.108	0.468	0.57	0.58	intronic
rs12193789	<i>ASCC3</i>	9.351×10^{-6}	0.994	0.371	0.166	0.94	0.96	intronic
rs11581556	<i>ATF6</i>	7.326×10^{-7}	0.995	0.845	0.615	0.96	0.90	intronic
rs5996513	<i>BCR</i>	8.368×10^{-6}	0.996	0.932	0.626	0.11	0.21	intronic
rs8	<i>CDK6</i>	9.581×10^{-7}	0.998	0.800	0.299	0.82	0.85	intronic
rs2304593	<i>CMTM7</i>	5.239×10^{-6}	0.993	0.370	0.000	0.16	0.34	intronic
rs2279103	<i>CTDP1</i>	6.403×10^{-8}	0.999	0.745	0.276	0.70	0.50	missense T340M
rs657801	<i>DENND2D</i>	9.280×10^{-6}	0.983	0.797	0.352	0.94	0.94	intronic
rs2354420	<i>EMP2</i>	6.654×10^{-6}	0.988	0.657	0.167	0.08	0.18	intronic
rs11792480	<i>ENG</i>	1.281×10^{-5}	0.989	0.954	0.041	0.83	0.68	intronic
rs550897	<i>FAM55D</i>	3.670×10^{-8}	0.998	0.848	0.308	0.99	0.99	missense Y398H
rs1011814	<i>FGF10</i>	6.501×10^{-7}	0.994	0.926	0.513	0.99	1.00	intronic
rs17679400	<i>FHIT</i>	1.3645×10^{-7}	0.998	0.742	0.255	0.26	0.20	intronic
rs1416995	<i>GPR158</i>	2.921×10^{-7}	0.997	0.995	0.973	0.04	0.29	intronic
rs12787863	<i>GRM5</i>	3.597×10^{-7}	0.996	0.860	0.088	0.94	0.81	intronic
rs17422	<i>HCFC1</i>	1.623×10^{-7}	0.997	0.936	0.004	0.84	0.81	intronic
rs11586100	<i>HNRNPR</i>	4.271×10^{-6}	0.984	0.968	0.248	1.00	1.00	intronic
rs1122821	<i>HOMER3</i>	8.660×10^{-6}	0.980	0.479	0.013	0.46	0.61	intronic
rs2158622	<i>JAZF1</i>	5.810×10^{-6}	0.991	0.716	0.000	0.04	0.49	intronic
rs9952025	<i>LIPG</i>	1.269×10^{-5}	0.994	0.764	0.787	0.18	0.11	intronic
rs955816	<i>MARCH4</i>	1.735×10^{-6}	0.991	0.910	0.186	0.85	0.37	intronic; in LD ($r^2 = 0.80$ in CHB + JPT) with rs11691655 (within ARNTL and CLOCK binding sites)
rs566125	<i>MMP3</i>	1.247×10^{-6}	0.995	0.716	0.254	0.98	0.96	intronic
rs710080	<i>MPG</i>	1.641×10^{-7}	0.998	0.989	0.840	0.44	0.54	5' UTR variant
rs1053000	<i>PFKP</i>	9.234×10^{-6}	0.988	0.971	0.152	1.00	1.00	3' UTR variant
rs578096	<i>PTGER3</i>	1.272×10^{-5}	0.985	0.733	0.000	0.77	0.51	intronic
rs2020945	<i>PWP2</i>	5.640×10^{-8}	0.998	0.965	0.780	0.62	0.50	missense D25N
rs11617401	<i>RAB20</i>	5.288×10^{-7}	0.997	0.791	0.000	0.89	0.47	intronic
rs1204897	<i>RCC2</i>	7.331×10^{-6}	0.987	0.540	0.496	0.37	0.46	intronic
rs17682132	<i>RCVRN</i>	3.072×10^{-7}	0.998	0.943	0.000	0.96	0.33	intronic
rs492786	<i>SCARA3</i>	7.034×10^{-6}	0.982	0.197	0.417	0.82	0.68	intronic

Table 1 Single nucleotide polymorphisms that correlate with Δ photoperiod (Continued)

rs9812406	<i>SCHIP1</i>	1.487×10^{-7}	0.998	0.989	0.577	0.60	0.68	intronic
rs6695715	<i>SEC16B</i>	1.113×10^{-5}	0.982	0.247	0.123	0.44	0.78	intronic
rs2224957	<i>SH3GL2</i>	2.399×10^{-7}	0.996	0.896	0.180	0.01	0.04	intronic
rs6081636	<i>SLC24A3</i>	5.980×10^{-6}	0.988	0.612	0.246	0.16	0.47	intronic
rs830142	<i>SLC8A2</i>	7.264×10^{-6}	0.993	0.325	0.000	0.67	0.80	intronic
rs732611	<i>TBC1D9</i>	4.321×10^{-8}	0.999	0.943	0.000	0.78	0.87	intronic
rs1514685	<i>TPO</i>	5.156×10^{-8}	0.998	0.720	0.199	0.23	0.17	intronic
rs11760463	<i>WDR86</i>	1.279×10^{-5}	0.975	0.854	0.378	1.00	0.97	intronic
rs10227271	<i>WNT2</i>	6.406×10^{-6}	0.983	0.971	0.173	0.07	0.70	intronic
Mouse circadian/sleep disturbance								
rs3753472	<i>ADORA1</i>	1.155×10^{-5}	0.981	0.941	0.445	0.80	0.21	intronic
rs2830044	<i>APP</i>	5.206×10^{-6}	0.985	0.833	0.311	0.53	0.71	intronic
rs228188	<i>BTBD9</i>	3.149×10^{-6}	0.993	0.895	0.155	1.00	0.72	intronic
rs7801807	<i>CADPS2</i>	1.465×10^{-5}	0.993	0.891	0.000	0.90	0.90	intronic
rs17818083	<i>EBF2</i>	2.924×10^{-6}	0.992	0.939	0.000	0.11	0.24	intronic
rs809192	<i>FYN</i>	1.270×10^{-5}	0.987	0.804	0.061	0.82	0.63	intronic
rs2134294	<i>HCRTR2</i>	5.923×10^{-6}	0.986	0.741	0.405	1.00	0.96	intronic
rs17187747	<i>KCNB2</i>	6.700×10^{-7}	0.994	0.652	0.349	0.33	0.54	intronic
rs12762512	<i>KCNMA1</i>	4.857×10^{-8}	0.999	0.307	0.494	0.29	0.33	intronic
rs1943620	<i>NCAM1</i>	9.774×10^{-9}	0.999	0.838	0.410	0.18	0.19	intronic
rs10774910	<i>NOS1</i>	1.693×10^{-6}	0.990	0.916	0.118	1.00	1.00	intronic
rs2294678	<i>NOX3</i>	1.530×10^{-5}	0.980	0.950	0.172	0.18	0.87	intronic
rs12043436	<i>OMA1</i>	2.192×10^{-6}	0.991	0.840	0.129	0.99	0.95	intronic
rs595146	<i>PHLPP1</i>	8.637×10^{-6}	0.999	0.877	0.000	0.34	0.29	intronic
rs10508958	<i>PRKG1</i>	4.490×10^{-7}	0.996	0.873	0.193	0.99	0.69	intronic
rs3733553	<i>PRKG2</i>	1.212×10^{-5}	0.980	0.868	0.823	0.96	0.99	intronic
rs752579	<i>RAI1</i>	1.023×10^{-5}	0.981	0.954	0.473	1.00	0.96	intronic
rs2271733	<i>RAX</i>	1.464×10^{-5}	0.984	0.688	0.443	0.89	0.63	missense D44E
rs10519052	<i>RORA</i>	3.667×10^{-6}	0.991	0.448	0.234	0.68	0.68	intronic
rs968357	<i>RORB</i>	2.655×10^{-6}	0.992	0.755	0.560	0.63	0.83	intronic
rs2236409	<i>TNC</i>	9.344×10^{-6}	0.985	0.455	0.063	0.40	0.96	intronic
rs7300641	<i>TPH2</i>	1.676×10^{-5}	0.975	0.827	0.400	0.01	0.02	intronic
rs4445877	<i>UBE3A</i>	1.849×10^{-5}	0.974	0.939	0.607	0.90	0.99	intronic
Mendelian diseases causing sleep disturbance in humans								
rs5909187	<i>CDKL5</i>	6.275×10^{-6}	0.987	0.550	0.017	0.27	0.55	intronic
rs11603330	<i>DHCR7</i>	2.848×10^{-7}	0.995	0.806	0.479	0.75	0.98	intronic
rs13414769	<i>HDAC4</i>	3.134×10^{-5}	0.969	0.573	0.264	0.94	0.98	intronic
rs2239464	<i>MECP2</i>	3.189×10^{-7}	0.996	0.663	0.517	0.48	0.28	intronic
rs858953	<i>NRXN1</i>	1.226×10^{-9}	1.000	0.780	0.254	0.83	0.49	intronic
rs8137951	<i>SHANK3</i>	3.148×10^{-6}	0.989	0.517	0.127	0.59	0.44	intronic
rs12406072	<i>SLC2A1</i>	2.430×10^{-5}	0.980	0.952	0.518	0.56	0.52	intronic
Melanopsin signaling								
rs308039	<i>GNA11</i>	2.387×10^{-5}	0.972	0.568	0.079	0.06	0.29	intronic
rs4745672	<i>GNAQ</i>	4.391×10^{-6}	0.986	0.790	0.610	1.00	0.92	intronic

Table 1 Single nucleotide polymorphisms that correlate with Δ photoperiod (Continued)

rs2476197	<i>INADL</i>	1.834×10^{-7}	0.998	0.943	0.252	0.92	0.96	intronic
rs2224361	<i>PLCB4</i>	4.222×10^{-5}	0.991	0.426	0.000	0.32	0.37	intronic
rs2138004	<i>PRKCA</i>	3.025×10^{-5}	0.971	0.974	0.708	0.24	0.33	intronic
rs10910030	<i>PRKCZ</i>	1.588×10^{-8}	0.999	0.771	0.234	0.18	0.27	intronic
rs1392171	<i>TRPC7</i>	2.123×10^{-6}	0.993	0.371	0.000	0.48	0.70	intronic

^aKendall's correlation P values for the correlation between allele frequency and Δ photoperiod; ^bpercentile rank in the distribution of Kendall's correlation coefficients (τ) calculated for minor allele frequency (MAF)-matched SNPs; ^c F_{ST} percentile rank in the distribution of SNPs showing a similar average MAF (calculated over all populations); ^dlnRsb percentile rank; ^einformation on ARNTL, CLOCK and CRY1 binding sites was obtained from [80]; significant values are shown in bold. CEPH, 1000 Genomes Phase I data for Utah Residents with Northern and Western European ancestry; CHB + JPT, 1000 Genomes Phase I data for Han Chinese in Beijing plus Japanese in Tokyo; RNAi, RNA interference; SNP, single nucleotide polymorphism, UTR, untranslated region.

and melanopsin-signaling components plus 56% of RNAi hits, 85% of circadian disturbance genes identified in mice, and 83% of human circadian phenotype loci. Two clusters of highly connected nodes are evident in the network, one that mainly includes components of the core circadian pacemaker and the other comprising neuropeptides such as hypocretin (*HCTR*), neuromedin (*NMU*), oxytocin (*OTX*), pro-melanin-concentrating hormone (*PMCH*) and their receptors. As expected, some hub nodes in the network are represented by genes involved in multiple cellular processes such as *FOS*, *APP*, *MTOR*, *PPARG* and *PTEN*.

For each of the 84 genes, we selected the SNP showing the strongest association with Δ photoperiod (based on the percentile rank of τ), obtaining a list of 84 independent variants. Four of these (4.7%) were missense substitutions (Table 1). An overview of SNP location relative to ENCODE functional elements is available as Additional file 7 for core circadian genes. Analysis of copy number variant genotypes in HGDP-CEPH and 1000 Genomes Phase I subjects [25,26] revealed no linkage disequilibrium (LD) with variants reported in Table 1.

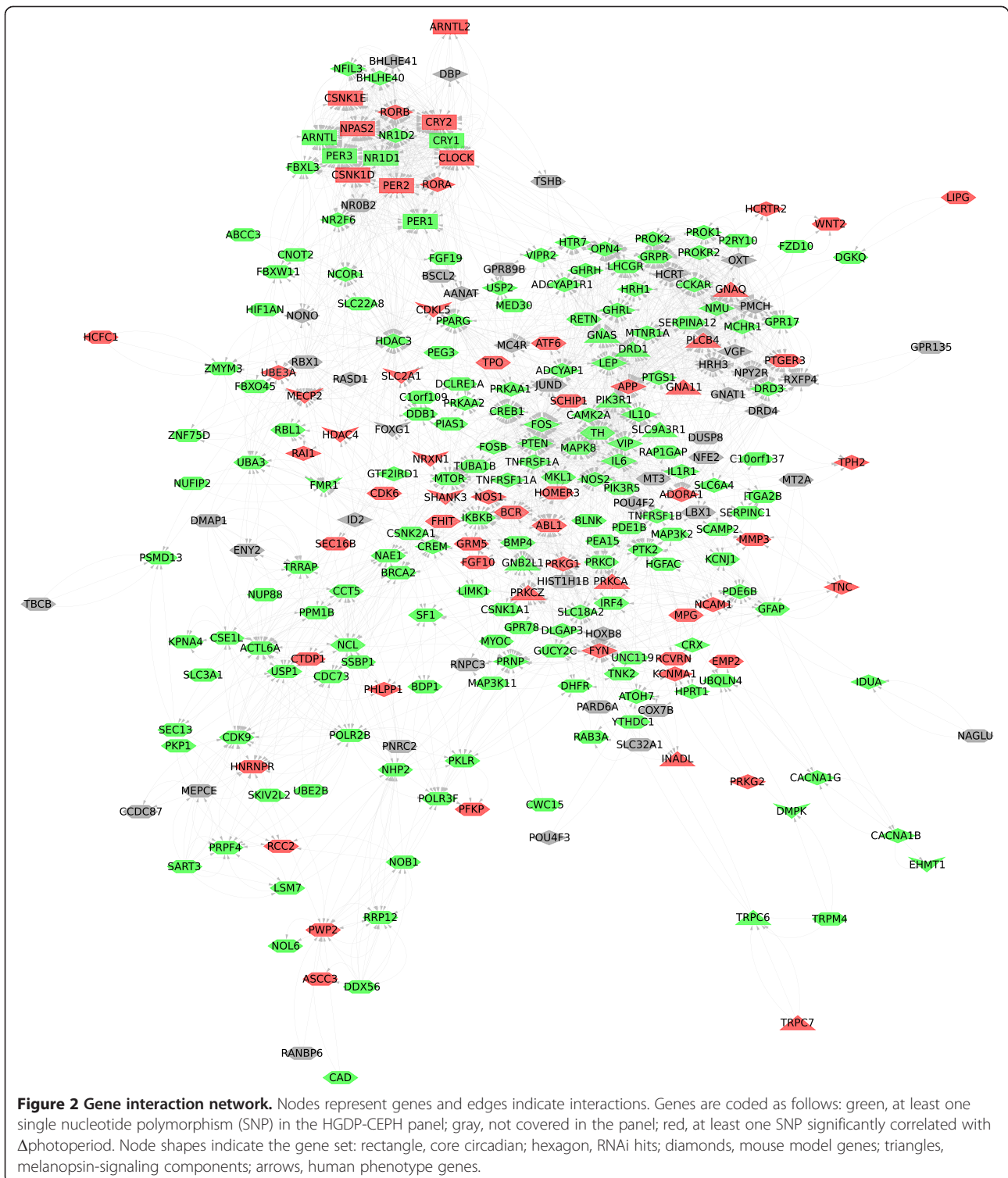
Natural selection signals at circadian rhythm regulatory genes

To validate the results reported above, and to gain further insight into the evolutionary history of the 84 Δ photoperiod-selected variants, we analyzed their normalized derived allele frequency (DAF): a progressive increase with Δ photoperiod was observed in most geographic areas, suggesting that selection has operated on these variants throughout human migration (Figure 1B). As expected, the normalized allele frequencies of 840 SNPs randomly selected from the HGDP-CEPH panel for having the same average MAF as selected variants (control SNPs) showed no such Δ photoperiod-dependent increase (Figure 1B).

We next reasoned that signatures of Δ photoperiod-driven selection should translate into strong genetic differentiation (fixation index, F_{ST}) between population pairs that live in regions with very different Δ photoperiod (for example, Biaka Pygmies and Orcadians, Figure 1A);

conversely, F_{ST} is expected to be low for geographically distant populations if they live in areas where annual variation in day length is similar (for example, Biaka Pygmies and Maya, Figure 1A). To test this hypothesis we calculated F_{ST} for the 84 variants in several pairwise comparisons (Figure 1A). To account for demographic effects, the same F_{ST} comparisons were calculated for all HGDP-CEPH SNPs and each of the 84 variants was ascribed a percentile rank in the distribution of SNPs showing a similar average MAF (calculated over all populations). In the Biaka Pygmy-Orcadian comparison, 13 SNPs showed an F_{ST} percentile rank higher than 0.95 (Table 1); this represents a three-fold enrichment over the expected (expected = 4, Fisher's exact test, $P = 0.019$). Conversely, in the Biaka Pygmy-Maya comparison, 18 variants showed an F_{ST} below the 0.05 percentile, representing a strong enrichment over expectation (Fisher's exact test, $P = 0.0011$, Table 1). Similar results were obtained for most additional pairwise comparisons (Figure 1A), with population pairs living in areas with very different Δ photoperiod showing an excess of SNPs with high F_{ST} values and distant populations sharing similar Δ photoperiod displaying many low- F_{ST} variants. Overall, these results confirm strong spatial signatures of natural selection at the 84 selected variants. Thus, calculation of the percentile rank of F_{ST} comparisons between all population pairs in the HGDP-CEPH panel revealed a clear signal for the 84 variants, with population differentiation increasing with Δ photoperiod across most geographic areas and low F_{ST} observed for population pairs living at similar latitudes (Figure 1C).

Although different selective regimes may underlie the spatial distribution of Δ photoperiod-selected variants, it is conceivable that at least some of the signals we detected at the 84 variants are determined by selective sweeps that ensued after out-of-Africa expansion. Therefore, we tested the hypothesis that these variants are preferential targets of positive selection in non-African populations by two methods, the Derived Intra-allelic Nucleotide Diversity (DIND) [27] and lnRsb [28] tests, computed using the 1000 Genomes Phase I data for Yoruba (YRI), Utah Residents with Northern and Western European ancestry (CEPH), and Han Chinese in Beijing plus Japanese in



Tokyo (CHB + JPT). DIND is based on the ratio of intra-allelic diversity associated with the ancestral and derived alleles (π_A/π_D) analyzed against the frequency of the derived allele [27] (Figure 3). The InRsb test evaluates the ratio of extended haplotype homozygosity between

two populations (in this case CEPH versus YRI and CHB + JPT versus YRI) [28] (Figure 3). For both tests the significance cut-off was set at $P < 0.05$. DIND was calculated for CEPH and CHB + JPT for all variants with DAF > 0.12 ($n = 66$) and > 0.08 ($n = 68$) among the 84 we

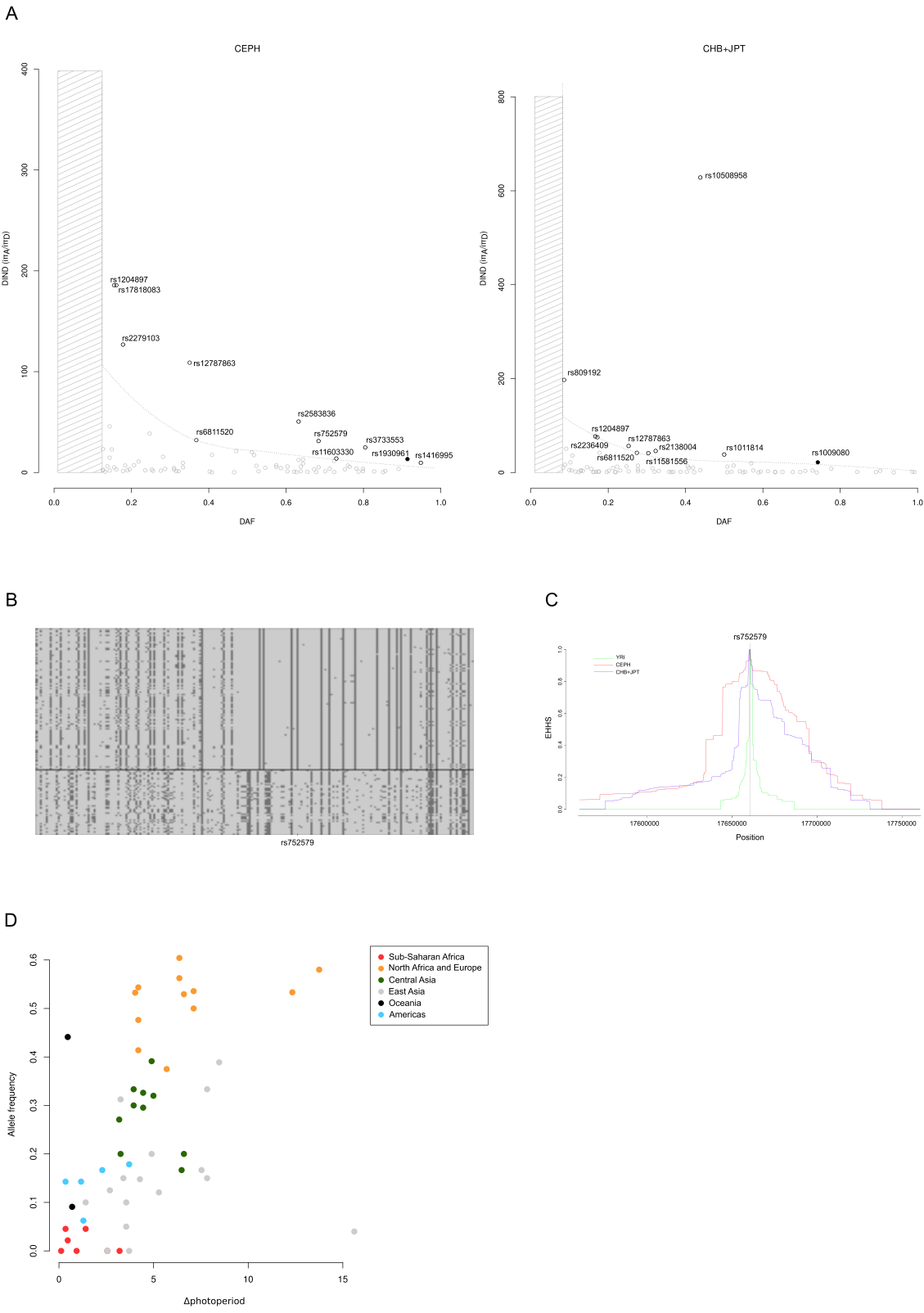


Figure 3 (See legend on next page.)

(See figure on previous page.)

Figure 3 Positive selection signatures for variants that correlate with Δ photoperiod. (A) DIND test results: the dashed line represents the 95th percentile of a distribution of 20,000 randomly selected HGDP-CEPH single nucleotide polymorphisms (SNPs). The gray shaded areas indicate frequency ranges where the ratio could not be calculated (see Methods). SNPs above the 95th percentile are shown as black circles (open: top variants from the five gene sets; solid: disease-associated SNPs). (B) Schematic representation of CEPH haplotypes in the region surrounding rs752579 (in *RAI1*), selected as an example. Each line represents a haplotype, columns indicate polymorphic positions. Dark gray, derived alleles; light gray, ancestral alleles. (C) Extended haplotype homozygosity decay plot for rs752579. (D) Correlation between Δ photoperiod and allele frequency for rs752579. Populations from different broad geographic areas are coded by different colors. CEPH, 1000 Genomes Phase I data for Utah Residents with Northern and Western European ancestry; CHB + JPT, 1000 Genomes Phase I data for Han Chinese in Beijing plus Japanese in Tokyo; DIND, Derived Intra-allelic Nucleotide Diversity; EHHS, extended haplotype homozygosity of an individual SNP site.

identified (see Methods). Ten SNPs had a significant DIND test in CEPH, and nine in CHB + JPT (Figure 3). In both populations the number of DIND-significant SNPs was higher than expected (Fisher's exact test $P = 0.038$ for CEPH and $P = 0.06$ for CHB + JPT); although statistical significance is borderline in CHB + JPT, it should be noted that a three-fold excess is observed and that the sample size for Fisher's exact test was small. Likewise, results from the InRsb tests (calculated for all 84 variants) indicated that there are 18 significant SNPs in CEPH and 17 in CHB + JPT (Table 1), representing a significant enrichment (Fisher's exact test, $P = 0.0008$ for CEPH and $P = 0.0015$ for CHB + JPT). Signals detected using the DIND and InRsb tests only partially overlap: this is expected as the two methods have different power to detect selection depending on the frequency of the target variant [27,28] and because the DIND can only capture selection at the derived allele.

Overall, these data suggest that variants correlated to Δ photoperiod in circadian rhythm regulatory genes represented preferential targets of positive selection during the out-of-Africa expansion of human populations.

Genetic adaptation to photoperiod and disease susceptibility

As mentioned above, disruption of circadian rhythms is a common feature of affective disorders such as SCZ, SAD, BPD, major depressive disorder (MDD) and autism spectrum disorder (ASD). Indeed, mutations in several genes that cause sleep/wake disturbances in humans or mice (Table 1) determine a phenotype that includes ASD features (for example, *MECP2*, *SHANK3*, *NRXN1*, *RAI1*, *UBE3A*). Likewise, mice with mutations in *Clock* or *Fbxl3* (included in the RNAi hit set but showing no signatures of natural selection) display both circadian disturbances and a behavioral profile reminiscent of BPD [4]; these animals are considered models for mania [4]. Notably, genes identified in the *in vitro* RNAi screen have been implicated in susceptibility to SCZ (*SEC16B* and *CAECAM21*), ASD (*FHIT*) and MDD (*SHIP1* and *FHIT*) by genome-wide association studies (GWAS) in humans; the same holds for *ARNTL* and *RORA*, which are associated with SCZ.

Thus, we explored the possibility that risk variants for these psychiatric conditions represent targets of Δ photoperiod-driven selection. To this aim, we retrieved all GWAS SNPs associated with any trait or disease from the National Human Genome Research Institute (NHGRI) Catalog of Published Genome-Wide Association Studies [29]. We retained only variants that have been genotyped in the HGDP-CEPH panel and collapsed SNPs in LD ($r^2 > 0.4$) into single loci, remaining with a list of 4,192 variants. Because more than 100 SNPs have been identified in meta-analyses of BPD and SCZ, the two diseases were considered together. Overall, our set included 128 variants associated with SCZ and/or BPD, 14 (11%) of these showing significant correlation with Δ photoperiod. By performing 10,000 resamplings we calculated that this proportion is significantly higher than that obtained for randomly selected GWAS SNPs (empirical $P = 0.0112$) and MAF-matched HGDP-CEPH variants (empirical $P = 0.003$). Thus, SCZ and/or BPD risk variants are preferential targets of natural selection driven by annual photoperiod variation. This was not the case for MDD or depression, as only two out of 50 associated SNPs significantly correlated with Δ photoperiod (Table 2). Finally, in the case of ASD, few SNPs have been associated with this condition in GWAS; thus, the LD-pruned set only contained 10 ASD variants, one of these significantly correlated with Δ photoperiod (Table 2). As for SAD, no GWAS has been performed; a variable number tandem repeat (VNTR) in the serotonin transporter gene (*SLC6A4*) promoter, which affects expression, is the only variant reproducibly associated with SAD and seasonality [30-32]. We used data from a recent analysis [33] and obtained a significant correlation between the frequency of the 14-repeat (short) allele, which predisposes to SAD [30-32], and Δ photoperiod ($\tau = 0.331$, Kendall's correlation P value = 0.00115).

Although not a psychiatric disease, restless leg syndrome (RLS) is a human condition with a strong circadian component. RLS symptoms worsen in the evening or night, and include sleep fragmentation. GWAS have identified seven unlinked SNPs, most of these mapping to four distinct genes (*MEIS1*, *BTBD9*, *PTPRD* and *MAP2K5/SKOR1*). Four of these variants have been included in the HGDP-CEPH panel and one of them significantly correlated

Table 2 Genome-wide association study single nucleotide polymorphisms that correlate with Δ photoperiod

SNP	Gene(s) ^a	Disease	Risk allele	Risk allele frequency decreases with Δ photoperiod	Kendall's correlation <i>P</i> value ^b	τ rank ^c	F _{ST} rank ^d	lnRsb rank ^e		Gene functional connection to circadian rhythm
								Biaka-Orcadian	Biaka-Maya	
rs1930961	<i>LRP5L, ADRBK2</i>	BPD	C	Yes	1.182×10^{-6}	0.993	0.996	0.942	0.530 0.700	ADRBK2 (aka GRK3) phosphorylates melanopsin [34]
rs6746896	<i>LMAN2L, CNNM4</i>	BPD	A	Yes	8.421×10^{-7}	0.995	0.550	0.350	0.780 0.790	<i>CNNM4</i> is an RNAi hit (the gene does not appear in Table 1 because the SNP is non-genic)
rs8099939	<i>GRIK5</i>	BPD	T	Yes	6.538×10^{-5}	0.956	0.968	0.615	0.960 0.980	-
rs2286492	<i>FAM126A</i>	BPD	G	No	8.521×10^{-5}	0.966	0.811	0.000	0.160 0.260	-
rs7570682	<i>TMEM18, POU3F3</i>	BPD	A	Yes	2.320×10^{-4}	0.950	0.455	0.004	0.160 0.090	-
rs4075511	<i>KCNS3, RDH14</i>	BPD	A	No	7.345×10^{-9}	0.999	0.797	0.156	0.720 0.730	-
rs7319311	<i>COL4A2</i>	BPD and SCZ	A	No	1.083×10^{-4}	0.952	0.474	0.319	0.400 0.470	<i>Clock</i> -regulated in mouse cardiomyocytes [35]
rs1124376	<i>KAT2B</i>	BPD and SCZ	G	Yes	3.852×10^{-4}	0.950	0.860	0.076	0.880 0.150	Directly interacts with CLOCK and NPAS2 [36]
rs1605834	<i>APOB, KLHL29</i>	BPD and SCZ	C	Yes	2.857×10^{-4}	0.950	0.966	0.877	0.350 0.240	-
rs17002034	<i>MKL1</i>	BPD and SCZ	G	Yes	2.322×10^{-5}	0.978	0.689	0.000	0.990 0.990	RNAi hit (gene does not appear among selected hits because of heavier Bonferroni correction)
rs589249	<i>CSF3R, GRIK3</i>	SCZ	G	Yes	4.565×10^{-5}	0.964	0.996	0.624	1.000 1.000	<i>Grik3</i> expression is regulated by Clock in the mouse ventral tegmental area [37]
rs7004633	<i>MMP16, RIPK2</i>	SCZ	A	No	1.433×10^{-6}	0.992	0.925	0.538	0.260 0.330	-
rs1009080	<i>PTPRU, MATN1</i>	SCZ	G	Yes	2.419×10^{-6}	0.988	0.987	0.708	0.900 0.690	-
rs1572299	<i>TLR4, DBC1</i>	SCZ	A	Yes	3.901×10^{-4}	0.951	0.950	0.321	1.000 0.200	DBC1 regulates NR1D1 stability [38]
rs892055	<i>RASGRP4</i>	Asperger	nr	na	4.268×10^{-5}	0.964	0.603	0.585	0.610 0.420	-
rs9601248	<i>NDFIP2, SPRY2</i>	MDD	C	No	7.198×10^{-5}	0.958	0.896	0.085	0.390 0.330	<i>SPRY2</i> is a direct target of ARNTL in mouse liver [39]
rs8020095	<i>GPHN</i>	Depression	A	Yes	1.273×10^{-4}	0.950	0.944	0.504	0.620 0.980	<i>GPHN</i> shows rhythmic expression in mouse SCN [40]
rs12593813	<i>MAP2K5</i>	RLS	A	No	2.066×10^{-5}	0.971	0.498	0.175	1.000 0.830	-

^aOne gene is reported if the variant is genic, for intergenic SNPs the two flanking protein-coding genes are listed; ^bKendall's *P* values for the correlation between allele frequency and Δ photoperiod; ^cpercentile rank in the distribution of Kendall's correlation coefficients (τ) calculated for minor allele frequency (MAF)-matched SNPs; ^dF_{ST} percentile rank in the distribution of SNPs showing a similar average MAF (calculated over all populations); ^elnRsb percentile rank; significant values are shown in bold. BPD, bipolar disorder; CEPH, 1000 Genomes Phase I data for Utah Residents with Northern and Western European ancestry; CHB + JPT, 1000 Genomes Phase I data for Han Chinese in Beijing plus Japanese in Tokyo; MDD, major depressive disorder; na, not available; nr, not reported; RLS, restless leg syndrome; SCZ, schizophrenia; SNP, single nucleotide polymorphism.

with Δ photoperiod (rs12593813, at the *MAP2K5/SKOR1* locus) (Table 2).

As above, for all GWAS variants showing signals of Δ photoperiod-driven selection we performed the DIND and lnRsb tests. Natural selection signatures at several of these variants were detected (Table 2, Figure 3). Notably, many risk variants showing signals of Δ photoperiod-driven selection map within or close to genes with strong evidence of involvement in circadian rhythm regulation (Table 2). For example, *ADRB2* phosphorylates melatonin [34], *KAT2B* directly interacts with *CLOCK* and *NPAS2* [36], *CNNM4* and *MKL1* have been identified in the circadian RNAi screen, *DBC1* regulates *NR1D1* stability [38], and *GPHN* shows circadian oscillations in the mouse SCN [40].

Discussion

We analyzed adaptation of human populations to seasonal variations in day length determined by latitude using a method based on the spatial correlation between allele frequencies and environmental variables. Our working hypothesis builds on the previously reported suitability of geographic-explicit models to study human evolution [13-17,20]; the simple conundrum whereby genes that play a role in circadian regulation should represent preferential targets of Δ photoperiod-driven selection; and the expectation that a significant fraction of variants identified using models of spatial variation in allele frequencies should display natural selection signatures detectable with different approaches. Thus, we analyzed five independent gene sets selected on the basis of distinct evidence of involvement in circadian regulation and show that they are significantly more likely to carry SNPs that correlate with Δ photoperiod than expected. Photoperiod is a function of latitude, as is the case for other environmental variables such as temperature and UV radiation flux. These other climatic factors might contribute to the evolution of the circadian genes we analyzed. Nonetheless, resampling analysis using temperature and UV radiation revealed no excess of correlated variants, with the only exception of circadian core genes with Δ temperature. This might either depend on the difficulty of disentangling the effects of correlated environmental variables or on the fact that external temperature also acts as a circadian entrainment cue [1]. Nonetheless, these analyses indicate that Δ photoperiod exerted a stronger selective pressure on these genes than temperature or UV radiation flux.

Several variants identified using geographic-explicit model display signatures of natural selection when tests based on population genetic differentiation or haplotype homozygosity are applied, supporting their adaptive role. Although these tests (in particular the DIND and lnRsb) likely detected signals that derive from so-called hard sweeps (that is, sweeps of newly arisen mutations), adaptation to

seasonal variation in photoperiod possibly also occurred through balancing selection or selection on standing variation (soft sweeps), as previously suggested for other environmental variables [41-43]. The signatures left by soft sweeps are known to be more difficult to identify and typically do not originate an extended LD pattern [44,45].

Finally, we should add that variants genotyped in the HGDP-CEPH panel suffer from a moderate ascertainment bias [21], which might potentially affect our results. Nonetheless, the ascertainment scheme is common to all variants and we based most of our results on the drawing of empirical distributions, which, at least partially, correct for panel-wide biases. Also, we validated our results by the application of haplotype-based tests with the use of unbiased genotype data. Overall, the results reported here indicate that, during out-of-Africa migration, human populations adapted to life at different latitudes by tuning their circadian clock systems. In line with this observation, the length of the endogenous circadian period differs in people depending on ethnicity, and Africans show shorter periods than people with European or Asian ancestry [46].

The selective effect of photoperiod differences across geographic areas has previously been demonstrated for vertebrates and invertebrates, by studying latitudinal clines in circadian gene polymorphisms [8-11], and by dissecting rhythmic functions in animals living at extreme latitudes, where loss of ultradian patterns may be observed during periods of continuous light or darkness [1]. Indeed, the synchronization of internal clocks with environmental cues is thought to be essential for health and fitness [5]. As an example, *Clock*-mutant mice display abnormal rhythmicity of several physiological functions (for example, sleep/wake cycles, food intake and basal metabolism), show behavioral disturbances, are susceptible to diet-induced obesity, and present frequent pregnancy failures [3]. Likewise, wild-type rodents exposed to experimental jet lag show accelerated tumor growth, spatial cognitive deficits, cardiovascular problems, hastened mortality upon aging and increased body mass [5]. Similar health problems are reported to occur with high prevalence in humans who experience frequent changes in daily rhythms as a result of rotating shift work [5].

A large body of evidence suggests that circadian rhythm abnormalities determine or exacerbate mood-related disorders [4]. Alterations in sleep/wake cycles, activity, body temperature and hormonal secretions have been described in BPD, SCZ, MDD and ASD [4]. Reduced amplitude and lower expression of core circadian components have been described in fibroblasts from patients with BPD [47]. In parallel, studies in diurnal rodent species have demonstrated that prolonged exposure to a very short photoperiod results in a depression-like behavior [48,49], whereas in rats prolonged dark phase conditions administered post-natally result in increased anxiety and decreased social interaction

in adult life [50]. These observations indicate a causal link between psychiatric diseases and circadian disruption and suggest that mood disorders in modern human populations might at least partially result from adaptation or mis-adaptation to latitudinal photoperiod variations. We thus tested whether risk variants for psychiatric disorders represent preferential targets of Δ photoperiod-driven selection. Results indicated that the proportion of risk variants targeted by Δ photoperiod-driven selection is significantly higher than expected for BPD and SCZ, but not for MDD, although sleep and circadian rhythm disruption represent a common feature in depressive disorders. Nonetheless, MDD differs from SCZ and BPD by showing higher prevalence and substantially lower heritability [51,52]. Fewer risk variants with smaller effect have been identified for MDD compared with SCZ and BPD. This is reflected in our study, whereby the sample size of MDD SNPs is less than half that of SCZ or BPD variants.

Additionally, recent works have suggested that the consequences of MDD on reproductive fitness might be much less severe than those of other psychiatric disorders. Large-scale studies based on historical registries have indicated that patients with SCZ have strongly reduced fecundity [53-55], whereas their siblings may have increased or decreased reproductive success depending on gender [53,54]. The overall effect calculated for affected patients and their siblings indicated that these families contribute fewer descendants to the new generation compared with families with no members affected by SCZ [53,54]. Similar results have been reported for BPD, although the decrease in fecundity seems to be less marked than that observed for SCZ [53,55]. Contrasting evidences have been reported about the fertility effect of MDD; some authors [55] indicated lower reproductive rates for patients with MDD compared with controls (however reproductive success was higher than in SCZ and BPD). Conversely, a recent analysis revealed that no reduction in fecundity is observed in patients with MDD [53]. Interestingly, siblings of patients with MDD were shown to have significantly more children.

These observations lead to a long-debated hypothesis on the adaptive role of depression [56], including SAD [57], and suggest that MDD variants are subject to selective regimes different from other psychiatric disease. As mentioned above, SAD is the best example of a mood disorder directly triggered by photoperiod changes. The observation whereby the short VNTR allele at *SCL6A4* increases with Δ photoperiod, especially across Africa and Europe [33], may not be directly supportive of the adaptive view, but at least indicates that a SAD-predisposing variant is not counter-selected at high latitudes. As for the psychiatric disease variants we identified as targets of Δ photoperiod-driven selection, the two for MDD show different patterns: the risk allele frequency increases with

Δ photoperiod for *GPHN* and decreases for *NDFIP2/SPRY2*. Conversely, for the majority of SCZ or BPD risk alleles (10 in 14), a lower frequency is observed in populations showing greater Δ photoperiod. Thus, as populations expanded at higher latitudes, non-risk variants for SCZ and/or BPD might have been favored by selection to counterbalance the effects of photoperiod length variations, these latter possibly representing an environmental stress for people at risk of mental illness.

Conclusions

Signatures of latitude-driven natural selection at circadian genes have been described for vertebrates and invertebrates [8-11]. Herein we address this same issue in humans by showing that humans adapted to seasonal variations in photoperiod.

Also, we provide a link among circadian rhythm regulators, human adaptation to latitude, and susceptibility to affective disorders. Whereas the evolutionary scenarios underlying the maintenance of affective disorder risk variants will require further exploration, the Δ photoperiod-selected risk variants we identified represent excellent candidates for interaction analysis with photoperiod-related environmental variables (for example, season of birth, country of residence, shift-work or lifestyle habits). Also, these SNPs might modulate the effect of chronotherapy, which is gaining increasing interest *per se* or in supplementation to pharmacological approaches.

Methods

Environmental variables

Geographic coordinates for HGDP-CEPH populations were obtained by the HGDP-CEPH website [58]. When population locations were provided as latitude ranges, the midpoint was used. For each population, the annual maximum and minimum day lengths (photoperiods) were calculated using the C++ code available as a part of the Google Wide Open Smart Home framework [59]. The annual minimum and maximum photoperiod, as well as Δ photoperiod, are monotone functions of latitude (absolute value). Thus, photoperiod variation is consistent over latitude ranges. Because we applied a rank-based correlation test, the results are robust to minor misspecification of latitudes.

The annual minimum and maximum temperatures as well as short-wave radiation flux were retrieved from the National Centers for Environmental Prediction/National Center for Atmospheric Research database (Legates and Willmott Average, re-gridded dataset).

HGDP-CEPH data and statistical analysis

Genotype data for the HGDP-CEPH panel derive from a previous work [21]. Eighty-six SNPs were removed as they were invariant (MAF = 0) or because they failed in

all populations. Atypical or duplicated samples and pairs of close relatives were removed [60]. An SNP was ascribed to a specific gene if it was located within the transcribed region or no farther than 500 base pairs upstream of the transcription start site. MAF for any single SNP was calculated as the average over all populations. When necessary, the derived or ancestral state was determined based on the comparison with the chimpanzee, orangutan and macaque reference genomes.

Genes that determine circadian or sleep pattern disturbances in mouse were identified through the MGI website [61]. Specifically, we searched the Mammalian Phenotype Ontology using the term “circadian” and the MGI phenotypes using “circadian” and “abnormal sleep pattern” as keywords. Entries were manually inspected so that only known protein-coding genes were included (the MGI search also retrieved “heritable phenotypic markers”, with undefined genomic location); and phenotypes resulting from double gene knock-outs were not considered (as the effect cannot be unequivocally attributed to one gene). The MGI search method was preferred over conventional literature analysis to obtain an unbiased set of genes involved in mouse circadian phenotypes. Genes responsible for sleep disturbance phenotypes in humans were identified through search of the OMIM [62] and PhenomicDB [63] databases using “circadian” or “sleep disturbances” as keywords. Entries were manually inspected and genes were included only if the phenotype can be ascribed to the mutation of a single gene (for example, contiguous gene syndromes were not included); and the phenotype results from mutation. Specifically, we decided not to include polymorphism-phenotype associations, because they have been shown to often lack robustness and fail to replicate in independent analyses [64]; as an example, variants in the *MAOA* gene have been associated with poor sleep quality and insomnia in two studies that analyzed few subjects and obtained partially contrasting results [65,66]). Thus, *MAOA* was not included in the study gene set.

Genes involved in melanopsin signaling were derived from a previous work [67], with the inclusion of *TRPC6*, *TRPC7* and *PLCB4* from Xue *et al.* [68].

Correlations between allele frequencies and environmental variables were calculated by Kendall's rank correlation coefficient (τ), a non-parametric statistic used to measure the degree of correspondence between two rankings. Specifically, because all SNPs in the HGDP-CEPH panel are bi-allelic, the frequency of one allele (randomly selected) was used in the correlation test (using the other allele would simply change the sign of the correlation coefficient). To account for demographic events, each SNP is then assigned a percentile rank in the distribution of τ absolute values calculated for all SNPs having a MAF (averaged over all populations) similar (in the 1% range) to that of the SNP being analyzed [15,16].

To estimate the probability of obtaining n genes carrying at least one significant SNP out of a group of m genes, we applied a resampling approach: samples of m genes were randomly extracted from a list of all genes covered by at least one SNP in the HGDP-CEPH panel (number of genes = 15,280) and for each sample the number of genes with at least one significant SNP was counted (Bonferroni correction was applied to each sampled SNP set). The empirical probability of obtaining n genes was then calculated from the distribution of counts deriving from 10,000 random samples of m genes. A similar resampling strategy was used for SNP sets, as mentioned in the text.

Pairwise F_{ST} was calculated using the R package *hierfstat* [69]. Because F_{ST} values are not independent from allele frequencies, we binned variants based on their MAF (50 classes) and calculated the 95th percentile for each MAF class, as detailed in the text. The populations used for the pairwise comparisons were selected on the basis of their Δ photoperiod and for having a sufficiently large sample of individuals genotyped in the HGDP-CEPH panel (at least 13, Additional file 1). The Yakut-Orcadian, Yakut-Russian and Russian-Orcadian comparisons are not reported because F_{ST} is generally low between these populations and the 0.05 percentile could not be unequivocally determined.

1000 Genomes data and positive selection tests

1000 Genomes Project Phase I data were retrieved from the dedicated website [26,70]. Data refer to the following populations: CEPH (Utah residents with Northern and Western European ancestry), population code = CEU, number of individuals = 85; YRI (Yoruba in Ibadan, Nigeria), population code = YRI, number of individuals = 88; CHB (Han Chinese in Beijing, China), population code = CHB, number of individuals = 97; JPT (Japanese in Tokyo, Japan), population code = JPT, number of individuals = 89. Data for CHB and JPT were combined.

SNP genotypes were organized in a MySQL database. A set of programs was developed to retrieve genotypes from the database and to analyze them according to selected regions or populations. These programs were developed in C++ using the *GeCo++* [71] and the *libsequence* [72] libraries.

The DIND test [27,73] was calculated for all 84 SNPs in Table 1 and for disease risk variants (Table 2); the statistical significance was derived from an empirical distribution of DIND-DAF value pairs for 20,000 SNPs randomly selected among those genotyped in the HGDP-CEPH panel. Specifically, DIND values were calculated for all SNPs using a constant number of 40 flanking variants (20 up- and downstream). The distributions of DIND-DAF pairs for CEPH and CHB + JPT were binned in DAF intervals (100 classes) and for each class the 95th percentile

was calculated [74]. As suggested previously [27], for values of $\pi_D = 0$ we set the DIND value to the maximum obtained over the whole dataset plus 20. For low DAF values most π_D were equal to 0 (that is, the 95th percentile could not be calculated); thus, we could not calculate the DIND test for SNPs with DAF <0.12 in CEPH and 0.08 in CHB + JPT.

The InRsb tests and the extended haplotype homozygosity of an individual SNP site (EHHS) were calculated as previously described [28] for the 84 SNPs in Table 1 and for disease risk variant in Table 2. Specifically, InRsb was calculated for all SNPs surrounding the selected variant in a 200 kilobase region; the statistical significance was derived from an empirical distribution of values from all SNPs in these regions. The EHHS test was calculated for the selected SNP as the marker and until it reached 0.05 in the surrounding region. The analysis was performed using the rehh R package [75].

Analysis of disease-associated variants

A list of GWAS SNPs was obtained from the NHGRI Catalog of Published Genome-Wide Association Studies [29,76]. Among these variants we retained those that had been genotyped in the HGDP-CEPH; these were collapsed on the basis of LD, originating a set of 4,192 variants. In particular, LD between SNP pairs was calculated using SNP Annotation and Proxy Search [77] with data for Europeans (CEU); SNPs showing $r^2 \geq 0.4$ were collapsed in a single locus by randomly selecting one of the two polymorphisms. Thus, all analyzed variants, including those associated with affective disorders and RLS, show $r^2 < 0.4$.

Frequency data for the *SLC4A6* VNTR were kindly provided by Dr J Murdoch as per a recent worldwide analysis [33]. Data for African Americans and European Americans were not included in the analysis as these groups cannot be located in a specific geographic area (or are admixed).

Gene interaction network and functional annotation

To visualize gene interactions, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) utility [78]. STRING queries different interaction databases and we limited our search to the human species. The resulting interaction file was used as an input for Cytoscape 2.8.3 [79].

Information on ARNTL, CLOCK and CRY1 binding sites was obtained from previous work [80]. Data derive from chromatin immunoprecipitation sequencing experiments in U2OS cells.

Data accession

All data used in this manuscript are publicly accessible through the HGDP-CEPH panel database (HGDP-CEPH

Genome Diversity Panel Database Version 3.0, <http://www.cephb.fr/hgdp/index.php>) and the 1000 Genomes Project website (<http://www.1000genomes.org/>; 1000 Genomes Project Phase I data).

Additional files

Additional file 1: Populations in the HGDP-CEPH panel with Δ photoperiod.

Additional file 2: List of analyzed genes, significant SNPs and phenotype descriptions.

Additional file 3: Resampling analysis with SNP number-matched gene sets.

Additional file 4: Resampling results with SNP-matched random gene sets.

Additional file 5: Distribution of SNP content.

Additional file 6: Resampling analysis with different environmental variables.

Additional file 7: Functional annotation for core circadian genes.

Abbreviations

ASD: autism spectrum disorder; BPD: bipolar disorder; CEPH: Utah Residents with Northern and Western European ancestry; CHB + JPT: Han Chinese in Beijing plus Japanese in Tokyo; DIND: Derived Intra-allelic Nucleotide Diversity; EHHS: extended haplotype homozygosity of an individual SNP site; GWAS: genome-wide association studies; LD: linkage disequilibrium; MAF: minor allele frequency; MDD: major depressive disorder; MGI: Mouse Genome Informatics; NHGRI: National Human Genome Research Institute; OMIM: Online Mendelian Inheritance in Man; RGC: retinal ganglion cells; RLS: restless leg syndrome; RNAi: RNA interference; SAD: seasonal affective disorder; SCN: suprachiasmatic nucleus; SCZ: schizophrenia; SNP: single nucleotide polymorphism; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins; YRI: Yoruba.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MS, UP, DF and MC designed the study. RC, CT, GM, SR, FRG, GPC, EB and NB performed research and/or analyzed data. MS, MC and UP wrote the article. All authors read and approved the final manuscript.

Authors' information

MC is Chair of Immunology at the Department of Physiopathology and Transplantation, University of Milan, Italy.

Acknowledgments

DF is supported by a fellowship of the Doctorate School of Molecular Medicine, University of Milan, Italy. The Funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author details

¹Scientific Institute IRCCS E. Medea, 23842 Bosisio Parini, LC, Italy. ²Don C. Gnocchi Foundation ONLUS, IRCCS, 20148 Milan, Italy. ³Dino Ferrari Centre, Department of Physiopathology and Transplantation, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, University of Milan, 20122 Milan, Italy. ⁴Department of Physiopathology and Transplantation, University of Milan, 20090 Milan, Italy.

Received: 23 April 2014 Accepted: 15 October 2014

Published online: 30 October 2014

References

- Hut RA, Beersma DG: **Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod.** *Philos Trans R Soc Lond B Biol Sci* 2011, **366**:2141–2154.
- Barnard AR, Nolan PM: **When clocks go bad: neurobehavioural consequences of disrupted circadian timing.** *PLoS Genet* 2008, **4**:e1000040.
- Zhang EE, Kay SA: **Clocks not winding down: unravelling circadian networks.** *Nat Rev Mol Cell Biol* 2010, **11**:764–776.
- McClung CA: **How might circadian rhythms control mood? Let me count the ways.** *Biol Psychiatry* 2013, **74**:242–249.
- Harrington M: **Location, location, location: important for jet-lagged circadian loops.** *J Clin Invest* 2010, **120**:2265–2267.
- Golden RN, Gaynes BN, Ekstrom RD, Hamer RM, Jacobsen FM, Suppes T, Wisner KL, Nemeroff CB: **The efficacy of light therapy in the treatment of mood disorders: a review and meta-analysis of the evidence.** *Am J Psychiatry* 2005, **162**:656–662.
- Wu JC, Kelson JR, Schachat C, Bunney BG, DeModena A, Golshan S, Gillin JC, Potkin SG, Bunney WE: **Rapid and sustained antidepressant response with sleep deprivation and chronotherapy in bipolar disorder.** *Biol Psychiatry* 2009, **66**:298–301.
- Kyriacou CP, Peixoto AA, Sandrelli F, Costa R, Tauber E: **Clines in clock genes: fine-tuning circadian rhythms to the environment.** *Trends Genet* 2008, **24**:124–132.
- Johnsen A, Fidler AE, Kuhn S, Carter KL, Hoffmann A, Barr IR, Biard C, Charmanier A, Eens M, Korsten P, Siitari H, Tomiuk J, Kempnaers B: **Avian Clock gene polymorphism: evidence for a latitudinal cline in allele frequencies.** *Mol Ecol* 2007, **16**:4867–4880.
- Liedvogel M, Szulkin M, Knowles SC, Wood MJ, Sheldon BC: **Phenotypic correlates of Clock gene variation in a wild blue tit population: evidence for a role in seasonal timing of reproduction.** *Mol Ecol* 2009, **18**:2444–2456.
- O'Malley KG, Banks MA: **A latitudinal cline in the Chinook salmon (*Oncorhynchus tshawytscha*) Clock gene: evidence for selection on PolyQ length variants.** *Proc Biol Sci* 2008, **275**:2813–2821.
- White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, Howell FC: **Pleistocene *Homo sapiens* from Middle Awash, Ethiopia.** *Nature* 2003, **423**:742–747.
- Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A: **Adaptations to climate in candidate genes for common metabolic disorders.** *PLoS Genet* 2008, **4**:e32.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK: **Using environmental correlations to identify loci underlying local adaptation.** *Genetics* 2010, **185**:1411–1423.
- Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M: **Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach.** *PLoS Genet* 2010, **6**:e1000849.
- Pozzoli U, Fumagalli M, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M: **The role of protozoa-driven selection in shaping human genetic variability.** *Trends Genet* 2010, **26**:95–99.
- Hancock AM, Witonsky DB, Ehler E, Alkorta-Aranburu G, Beall C, Gebremedhin A, Sukernik R, Utermann G, Pritchard J, Coop G, Di Rienzo A: **Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency.** *Proc Natl Acad Sci U S A* 2010, **107**:8924–8930.
- Casto AM, Feldman MW: **Genome-wide association study SNPs in the human genome diversity project populations: does selection affect unlinked SNPs with shared trait associations?** *PLoS Genet* 2011, **7**:e1001266.
- Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R: **Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution.** *PLoS Genet* 2011, **7**:e1002355.
- Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, Sukernik R, Utermann G, Pritchard JK, Coop G, Di Rienzo A: **Adaptations to climate-mediated selective pressures in humans.** *PLoS Genet* 2011, **7**:e1001375.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM: **Worldwide human relationships inferred from genome-wide patterns of variation.** *Science* 2008, **319**:1100–1104.
- Handley LJ, Manica A, Goudet J, Balloux F: **Going the distance: human population genetics in a clinal world.** *Trends Genet* 2007, **23**:432–439.
- Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, Myers RM, Cavalli-Sforza LL, Feldman MW, Pritchard JK: **The role of geography in human adaptation.** *PLoS Genet* 2009, **5**:e1000500.
- Zhang EE, Liu AC, Hirota T, Miraglia LJ, Welch G, Pongsawakul PY, Liu X, Atwood A, Huss JW 3rd, Janes J, Su AI, Hogenesch JB, Kay SA: **A genome-wide RNAi screen for modifiers of the circadian clock in human cells.** *Cell* 2009, **139**:199–210.
- Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI, Mefford H, Ying P, Nickerson DA, Eichler EE: **Population analysis of large copy number variants and hotspots of human genetic disease.** *Am J Hum Genet* 2009, **84**:148–161.
- 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: **An integrated map of genetic variation from 1,092 human genomes.** *Nature* 2012, **491**:56–65.
- Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, Kidd JR, Kidd KK, Alcais A, Ragimbeau J, Pellegrini S, Abel L, Casanova JL, Quintana-Murci L: **Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense.** *PLoS Genet* 2009, **5**:e1000562.
- Tang K, Thornton KR, Stoneking M: **A new approach for using genome scans to detect recent positive selection in the human genome.** *PLoS Biol* 2007, **5**:e171.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci U S A* 2009, **106**:9362–9367.
- Willeit M, Praschak-Rieder N, Neumeister A, Zill P, Leisch F, Stastny J, Hilger E, Thierry N, Konstantinidis A, Winkler D, Fuchs K, Sieghart W, Aschauer H, Ackenheil M, Bondy B, Kasper S: **A polymorphism (5-HTTLPR) in the serotonin transporter promoter gene is associated with DSM-IV depression subtypes in seasonal affective disorder.** *Mol Psychiatry* 2003, **8**:942–946.
- Sher L, Hardin TA, Greenberg BD, Murphy DL, Li Q, Rosenthal NE: **Seasonality associated with the serotonin transporter promoter repeat length polymorphism.** *Am J Psychiatry* 1999, **156**:1837.
- Rosenthal NE, Mazzanti CM, Barnett RL, Hardin TA, Turner EH, Lam GK, Ozaki N, Goldman D: **Role of serotonin transporter promoter repeat length polymorphism (5-HTTLPR) in seasonality and seasonal affective disorder.** *Mol Psychiatry* 1998, **3**:175–177.
- Murdoch JD, Speed WC, Pakstis AJ, Heffelfinger CE, Kidd KK: **Worldwide population variation and haplotype analysis at the serotonin transporter gene SLC6A4 and implications for association studies.** *Biol Psychiatry* 2013, **74**:879–889.
- Blasic JR Jr, Lane Brown R, Robinson PR: **Light-dependent phosphorylation of the carboxy tail of mouse melanopsin.** *Cell Mol Life Sci* 2012, **69**:1551–1562.
- Bray MS, Shaw CA, Moore MW, Garcia RA, Zanutta MM, Durgan DJ, Jeong WJ, Tsai JY, Bugger H, Zhang D, Rohrwasser A, Rennison JH, Dyck JR, Litwin SE, Hardin PE, Chow CW, Chandler MP, Abel ED, Young ME: **Disruption of the circadian clock within the cardiomyocyte influences myocardial contractile function, metabolism, and gene expression.** *Am J Physiol Heart Circ Physiol* 2008, **294**:H1036–H1047.
- Curtis AM, Seo SB, Westgate EJ, Rudic RD, Smyth EM, Chakravarti D, FitzGerald GA, McNamara P: **Histone acetyltransferase-dependent chromatin remodeling and the vascular clock.** *J Biol Chem* 2004, **279**:7091–7097.
- Mukherjee S, Coque L, Cao JL, Kumar J, Chakravarty S, Asaithamby A, Graham A, Gordon E, Enwright JF 3rd, DiLeone RJ, Birnbaum SG, Cooper DC, McClung CA: **Knockdown of Clock in the ventral tegmental area through RNA interference results in a mixed state of mania and depression-like behavior.** *Biol Psychiatry* 2010, **68**:503–511.
- Chini CC, Escande C, Nin V, Chini EN: **DBC1 (Deleted in Breast Cancer 1) modulates the stability and function of the nuclear receptor Rev-erbalpha.** *Biochem J* 2013, **451**:453–461.
- Hatanaka F, Matsubara C, Myung J, Yoritaka T, Kamimura N, Tsutsumi S, Kanai A, Suzuki Y, Sassone-Corsi P, Aburatani H, Sugano S, Takumi T: **Genome-wide profiling of the core clock protein BMAL1 targets reveals a strict relationship with metabolism.** *Mol Cell Biol* 2010, **30**:5636–5648.
- Shapiro-Reznik M, Jilg A, Lerner H, Earnest DJ, Zisapel N: **Diurnal rhythms in neurexins transcripts and inhibitory/excitatory synapse scaffold proteins in the biological clock.** *PLoS One* 2012, **7**:e37894.

41. Novembre J, Di Rienzo A: **Spatial patterns of variation due to natural selection in humans.** *Nat Rev Genet* 2009, **10**:745–755.
42. Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, Bresolin N, Sironi M: **Widespread balancing selection and pathogen-driven selection at blood group antigen genes.** *Genome Res* 2009, **19**:199–212.
43. Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A: **Adaptations to new environments in humans: the role of subtle allele frequency shifts.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365**:2459–2468.
44. Barrett RD, Schluter D: **Adaptation from standing genetic variation.** *Trends Ecol Evol* 2008, **23**:38–44.
45. Newburg DS, Ruiz-Palacios GM, Morrow AL: **Human milk glycans protect infants against enteric pathogens.** *Annu Rev Nutr* 2005, **25**:37–58.
46. Eastman CI, Molina TA, Dziepak ME, Smith MR: **Blacks (African Americans) have shorter free-running circadian periods than whites (Caucasian Americans).** *Chronobiol Int* 2012, **29**:1072–1077.
47. Yang S, Van Dongen HP, Wang K, Berrettini W, Bucan M: **Assessment of circadian function in fibroblasts of patients with bipolar disorder.** *Mol Psychiatry* 2009, **14**:143–155.
48. Ashkenazy-Frolinger T, Kronfeld-Schor N, Juetten J, Einat H: **It is darkness and not light: depression-like behaviors of diurnal unstriped Nile grass rats maintained under a short photoperiod schedule.** *J Neurosci Methods* 2010, **186**:165–170.
49. Einat H, Kronfeld-Schor N, Eilam D: **Sand rats see the light: short photoperiod induces a depression-like response in a diurnal rodent.** *Behav Brain Res* 2006, **173**:153–157.
50. Toki S, Morinobu S, Imanaka A, Yamamoto S, Yamawaki S, Honma K: **Importance of early lighting conditions in maternal care by dam as well as anxiety and memory later in life of offspring.** *Eur J Neurosci* 2007, **25**:815–829.
51. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, Ripke S, MacIntyre DJ, McGhee KA, Maclean AW, Smit JH, Hottenga JJ, Willemsen G, Middeldorp CM, de Geus EJ, Lewis CM, McGuffin P, Hickie IB, van den Oord EJ, Liu JZ, Macgregor S, McEvoy BP, Byrne EM, Medland SE, Statham DJ, Henders AK, Heath AC, Montgomery GW, Martin NG, Boomsma DI, et al: **Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned.** *Mol Psychiatry* 2012, **17**:36–48.
52. Wray NR, Gottesman II: **Using summary data from the Danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder.** *Front Genet* 2012, **3**:118.
53. Power RA, Kyaga S, Uher R, MacCabe JH, Langstrom N, Landen M, McGuffin P, Lewis CM, Lichtenstein P, Svensson AC: **Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings.** *JAMA Psychiatry* 2013, **70**:22–30.
54. Haukka J, Suvisaari J, Lonnqvist J: **Fertility of patients with schizophrenia, their siblings, and the general population: a cohort study from 1950 to 1959 in Finland.** *Am J Psychiatry* 2003, **160**:460–463.
55. Laursen TM, Munk-Olsen T: **Reproductive patterns in psychotic patients.** *Schizophr Res* 2010, **121**:234–240.
56. Nesse RM: **Is depression an adaptation?** *Arch Gen Psychiatry* 2000, **57**:14–20.
57. Davis C, Levitan RD: **Seasonality and seasonal affective disorder (SAD): an evolutionary viewpoint tied to energy conservation and reproductive cycles.** *J Affect Disord* 2005, **87**:3–10.
58. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, et al: **A human genome diversity cell line panel.** *Science* 2002, **296**:261–262.
59. **Google WOSH Code for Photoperiod Calculation.** [http://code.google.com/p/wosh/source/browse/trunk/src/bundles/services/Weather/3rdparty/sunstuff.cpp]
60. Rosenberg NA: **Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives.** *Ann Hum Genet* 2006, **70**:841–847.
61. **Mouse Genome Informatics (MGI) Database.** [http://www.informatics.jax.org/]
62. **Online Mendelian Inheritance in Man (OMIM).** [http://www.omim.org/]
63. **PhenomicDB Database.** [http://www.phenomicdb.de/]
64. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: **Replication validity of genetic association studies.** *Nat Genet* 2001, **29**:306–309.
65. Du L, Bakish D, Ravindran A, Hrdina PD: **MAO-A gene polymorphisms are associated with major depression and sleep disturbance in males.** *Neuroreport* 2004, **15**:2097–2101.
66. Brummett BH, Krystal AD, Siegler IC, Kuhn C, Surwit RS, Zuchner S, Ashley-Koch A, Barefoot JC, Williams RB: **Associations of a regulatory polymorphism of monoamine oxidase-A gene promoter (MAOA-uVNTR) with symptoms of depression and sleep quality.** *Psychosom Med* 2007, **69**:396–401.
67. Peirson SN, Oster H, Jones SL, Leitges M, Hankins MW, Foster RG: **Microarray analysis and functional genomics identify novel components of melanopsin signaling.** *Curr Biol* 2007, **17**:1363–1372.
68. Xue T, Do MT, Riccio A, Jiang Z, Hsieh J, Wang HC, Merbs SL, Welsbie DS, Yoshioka T, Weissgerber P, Stolz S, Flockecki V, Freichel M, Simon MI, Clapham DE, Yau KW: **Melanopsin signalling in mammalian iris and retina.** *Nature* 2011, **479**:67–73.
69. Goulet J: **Hierfstat, a package for R to compute and test hierarchical F-statistics.** *Mol Ecol Notes* 2005, **5**:184–186.
70. **1000 Genomes Project.** [http://www.1000genomes.org/]
71. Cereda M, Sironi M, Cavalleri M, Pozzoli U: **GeCo++: a C++ library for genomic features computation and annotation in the presence of variants.** *Bioinformatics* 2011, **27**:1313–1315.
72. Thornton K: **Libsequence: a C++ class library for evolutionary genetic analysis.** *Bioinformatics* 2003, **19**:2325–2327.
73. Fagny M, Patin E, Enard D, Barreiro LB, Quintana-Murci L, Laval G: **Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets.** *Mol Biol Evol* 2014, **31**:1850–1868.
74. Forni D, Cagliani R, Pozzoli U, Colleoni M, Riva S, Biasin M, Filippi G, De Gioia L, Gnudi F, Comi GP, Bresolin N, Clerici M, Sironi M: **A 175 million year history of T cell regulatory molecules reveals widespread selection, with adaptive evolution of disease alleles.** *Immunity* 2013, **38**:1129–1141.
75. Gautier M, Vitalis R: **Rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure.** *Bioinformatics* 2012, **28**:1176–1177.
76. **A Catalog of Published Genome-Wide Association Studies.** [http://www.genome.gov/gwastudies]
77. **SNP Annotation and Proxy Search (SNAP).** [http://www.broadinstitute.org/mpg/snap/ldsearchpw.php]
78. **STRING database.** [http://string-db.org/]
79. **Cytoscape.** [http://www.cytoscape.org/]
80. Hoffmann J, Symul L, Shostak A, Fischer T, Naef F, Brunner M: **Non-circadian expression masking clock-driven weak transcription rhythms in U2OS cells.** *PLoS One* 2014, **9**:e102238.

doi:10.1186/s13059-014-0499-7

Cite this article as: Forni et al.: Genetic adaptation of the human circadian clock to day-length latitudinal variations and relevance for affective disorders. *Genome Biology* 2014 **15**:499.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

