DOI:
10.1016/j.euroneuro.2018.07.101

Document Version
Other version

Supplementary Material

METHODS AND MATERIALS

For all studies, informed consent had been obtained and ethical approval for the study was granted by the local Research Ethics committees.

Assessment of self-reported smoking and definition of smoking categories in the four samples

NFBC1966: At age 14, participants reported if they have ever smoked and how much via a postal questionnaire using eight options: 1) 'Never'; 2) 'I have tried once'; 3) 'I have tried twice or more'; 4) 'I smoke occasionally'; 5) 'I smoke about twice a week'; 6) 'I smoke between 1–5 cigarettes a day'; 7) 'I smoke between 6–10 cigarettes a day' and 8) 'I smoke more than 10 cigarettes a day'.

The category *Ever-Tried* included groups 2-8; *Smokers* included groups 4-8; *Weekly-Smokers* included groups 5-8, and *Never-Tried* included group 1.

NFBC1986: At age 16, participants were asked by postal questionnaires: 'Have you ever smoked or used snuff in your life?' with four possible answers: 1) 'no',2) 'yes but only tried',3) 'yes, smoke' and4) 'yes, snuff'.

Participants in group 1 were classified as *Never-Tried*; those in groups 2-3 were classified as *Ever-Tried and* those in group 3 as *Smokers*. Participants in group 4 (N=36) were excluded. Cohort members were also asked how often they smoked and participants who were smoking every week or more were classified as *Weekly-Smokers*.

ALSPAC: At age 15, participants were invited to attend a face-to-face interview about tobacco use. Participants were asked multiple questions including: 'Have you ever tried a cigarette, even a puff?'. Participants who responded affirmatively to this question were classified as *Ever-Tried* and participants who answered 'no' were classified as *Never-Tried*. Participants who responded that they have smoked more than five cigarettes to the question 'How many cigarettes have you smoked in your lifetime' were classified as *Smokers*; those who responded that they smoked every week or more were classified as *Weekly-Smokers*.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (www.bris.ac.uk/alspac/researchers/data-access/data-dictionary). Ethical approval for the study was obtained from the ALSPAC ethics and Law Committee and the Local Research Ethics Committees.

IMAGEN: Smoking use was assessed using the European School Survey Project on Alcohol and Drugs questionnaires (1). Participants were asked 'On how many occasions during your lifetime have you smoked cigarettes?'. Seven options were given:
'0','1-2','3-5','6-9','10-19', '20-39' and '40 or more'.

Participants who responded zero were classified as *Never-Tried* and all the other as *Ever-Tried*. Participants who responded they have smoked more than five cigarettes were classified as *Smokers*. Participants were also asked how often they smoked and participants who were smoking every week or more frequently were classified as *Weekly-Smokers*. Although smoking behavior was assessed with different instruments in the four cohorts, the four categories were defined as similarly as possible.

Cotinine

Approximately 80% of nicotine is metabolized to cotinine in the liver by CYP2A6 enzyme (2). Cotinine is a relatively stable compound whose level reflects the cumulative intake of nicotine in the last week (3). Cotinine was measured only in the ALSPAC cohort from plasma samples collected at 15 years using the Cozart Cotinine Enzyme Immunoassay. Data was taken from the ALSPAC from EDTA blood plasma samples taken in a clinic assessment at 15 years. The plasma samples were stored at -80 °C and allowed to thaw at room temperature before use. Cotinine was measured using the Cozart Cotinine Enzyme Immunoassay (EIA) serum kit (M155B1). All samples were run in duplicates. Absorbance was measured spectrophotometrically at a wave length of 450 nm. Cotinine concentrations are expressed as ng/ml of blood. Cotinine level was quintile-transformed to reach normality. Quintile transformation has the advantage of including individuals with cotinine level equal to zero such as non-smokers in the analyses (4). Analyses using log-transformation (that exclude non-smokers) were also performed and showed substantially identical results (data not shown). Cotinine level was available in 2540 participants who also had genetic data available and were of European ancestry. Linear regression was performed to investigate the association between plasma continine level and smoking (see Supplementary-Table-3).

Genetic-association analyses within each sample

NFBC1966: Genome-wide genotyping was performed using the Illumina Infinium 370cnvDuo array. Quality control procedures used are described elsewhere (5). 33 directly genotyped SNPs covering the *TTC12-ANKK1-DRD2* were available.

NFBC1986: No genome-wide SNP data were available in this cohort. SNP rs2236709 was already genotyped as part of the custom Illumina Metabochip array (6). 13 additional SNPs were genotyped by Kbioscience (Hoddesdon, UK, http://www.kbioscience.co.uk/) using their

own system of fluorescence-based competitive allele-specific PCR (KASPar). SNPs were selected using a tagging approach to capture 75% of alleles at r^2>=0.8.

ALSPAC: Genome-wide genotyping was performed using the Illumina 550K array (Illumina, San Diego, CA, USA). We extracted the same SNPs available in NFBC1966.

IMAGEN: Participants were genotyped using the Illumina Quad 610 and 660 arrays. Among the SNPs that passed quality controls, 29 of the 33 SNPs were available.

Functional Magnetic Resonance Imaging (fMRI)

Monetary incentive delay (MID) task (7): Each trial involved an anticipation phase (4 sec), a response phase (2 sec), a feedback phase (2 sec), and a fixation period (4 sec). During the anticipation phase, cues indicating the amount of reward that could be won in a given trial (large, small, or none) were shown for 4 seconds. The participant could win large or small numbers of points (10 or 2) by responding as quickly as possible to a response cue. The points were converted to food snacks (small chocolate candies) following testing (5 points per candy). The participant completed 22 trials per condition, yielding 66 trials in total. One cue (circle with two lines) signaled that a large reward could be won, another (circle with one line) that a small reward could be won, and a third cue (triangle) that no reward could be won in the respective trial. Following a random time interval, a response cue was displayed, and the participant was instructed to respond as quickly as possible to this cue by means of a button press. The time window in which responses were counted as "wins" was adjusted dynamically during the course of the experiment according to the participant's performance, such that on average the participant won in 66% of all trials. The response and feedback phases had a total duration of 2 seconds. Four seconds of inter-trial fixation periods separated the trials.

Data Acquisition: Structural and functional Magnetic Resonance Imaging (fMRI) data were acquired by using 3-T MRI scanners from a range of manufacturers (Siemens, Philips, General Electric, Bruker), and the scanning variables were specifically chosen to be compatible with all scanners (8). For the present task, 300 volumes were acquired for each participant. Each volume consisted of 40 slices aligned to the line connecting the anterior-posterior commissure (2.4-mm thickness, 1-mm gap TR=2.20 s, TE=30 ms). The total scanning duration for the task took 11 min.

fMRI Preprocessing and Analysis

Structural and fMRI data were acquired using 3-T MRI scanners of different manufacturers (Siemens, GE, Philips). Pre-processing and data analyses were conducted using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) running in MATLAB® (9).

First, echo-planar images were co-registered with the T1 structural image. Functional images were realigned and resliced to the first volume. We created a custom template from the T_1 images of 552 adolescents by using the DARTEL toolbox (10) as implemented in SPM8. Single-subject contrast images were normalized to Montreal Neurological Institute (MNI) space by means of this custom template on the basis of the individual participants' DARTEL flow fields, and they were smoothed with a 5-mm Gaussian isotropic kernel. A first-level model was constructed on the unsmoothed single-participant data by using the following regressors: 1) anticipation of large reward, 2) anticipation of small reward, 3) anticipation of no reward, 4) feedback indicating large reward, 5) feedback indicating small reward, 6) feedback indicating no reward. Each regressor was included twice, once for successful trials, i.e., hits, and once for unsuccessful trials, i.e., misses. Trials in which participants failed to respond were modeled as separate error trials. Estimated movement

parameters were added to the design matrix in the form of 18 additional columns (3 translations, 3 rotations, 3 quadratic and 3 cubic translations, and each 3 translations with a shift of ±1 TR).

At the first level of analysis, changes in the blood-oxygen-level dependent (BOLD) response for each participant were assessed by linear combinations at the individual participant level, for each experimental condition; each trial (i.e. reward anticipation high reward) was convolved with the hemodynamic response function to form regressors that account for variance associated with the processing of reward anticipation and feedback. Contrast images of the parameter estimates were created for each participant.

The normalized and smoothed single-subject contrast images were then taken to a second-level random effects analysis to identify brain regions activating by anticipation of reward. Gender, site of recruitment and handedness were used as covariates of no interest. For the present investigation, we focused on the analyses on the reward anticipation phase using the contrast 'anticipation of large reward > anticipation of no reward'.

Region of interest (ROI) analyses were run to investigate the association between activations in the main site of reward processing, i.e. the ventral striatum, and genotypes. The ROIs for the bilateral ventral striatum were extracted based on coordinates from previous findings (11) from the contrast 'anticipation of high reward > anticipation of no reward' as 9-mm sphere (x= ±15, y=9, z=-9) using MarsBaR version 0.42 (12), see Supplementary-Figure-2.

From all participants who provided fMRI data for the MID task (n=1668), 1610 and 1630 for the left and right ventral striatum, respectively, were left after outlier removal according to the following criterion: Mean +/- 2*SD. 1263 of these had genotypic data as well and these were included in the analyses between fMRI and genotypic data.

Genetic-association analyses within each sample:

To control for population stratification, principal components were computed from genome-wide data (NFBC1966, ALSPAC and IMAGEN) (13) or Multidimensional Scaling coordinates (14) computed using 130,000 SNPs from the Illumina metabochip array (NFBC1986) (6) and included as covariates into the analysis.

Statistical Analyses

Meta-analysis: The meta-analysis procedure in Plink v1.07 (15) requires input files with odds ratio, beta estimates, confidence interval, and p-value for each study. Meta-analysis outputs include a beta, odds ratios, and a p-value for fixed and random effect models. The fixed effects model assumes that the effect of the independent variable on the outcome is the same in each study. The random effect model allows the effect to vary between studies. The output gives two measures: Cochrane's Q statistic and I^2 heterogeneity index. The I^2 statistic lies between 0%-100% with I^2=0 indicating no heterogeneity (16).

TABLES

Supplementary Table 1. Minor allele frequencies (MAF) in the 4 different cohorts.

BP	Gene	SNP	Location	MAF					Ref allele/	N	
				Overall	Imagen	Alspac	1966	1986	Other allele	Samples	Individuals
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	0.3681	0.3586	0.4920	0.4944	C/T	4	8565
112691986	*TTC12*	rs2236709	Intron	0.2588	0.2533	0.2396	0.2961	0.3031	G/A	4	8866
112693407	*TTC12*	rs7927508	Intron	0.3667	0.3637	0.3578	0.4920	NA	G/A	3	6336
112694133	*TTC12*	rs2156486	Intron	0.1814	0.1817	0.2574	0.1865	0.1801	G/T	4	8569
112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	0.4955	0.4977	0.4260	*NA*	C/A	3	6344
112704356	*TTC12*	rs10502172	Intron	0.4596	0.4685	0.4851	0.4616	0.4526	T/C	4	8633
112705919	*TTC12*	rs2303380	Intron splice site	0.3819	0.3824	0.3664	0.2970	*NA*	G/A	3	6344
112716539	*TTC12*	rs2288159	Intron	0.1561	0.1480	0.1479	0.2417	*NA*	T/G	3	6343
112721133	*TTC12*	rs4987094	Intron	0.1004	0.0977	0.0995	0.1979	*NA*	A/G	3	6347
112735810	*TTC12*	rs2276070	Intron	0.1541	0.1481	0.1478	0.2410	*NA*	T/C	3	6345
112739889	*TTC12*	rs719802	Intron	0.3953	0.3915	0.3970	0.3242	*NA*	T/C	3	6344
112739985	*TTC12*	rs719804	Intron	0.249	0.2500	0.1941	0.1732	*NA*	G/A	3	6334

112749387	*TTC12*	rs2282511	flanking 3'UTR	0.3455	*NA*	0.3274	0.2957	0.2959	A/C	3	7145
112754346	*TTC12*	rs754672	flanking 3'UTR	0.4894	0.4831	0.4621	0.4629	0.4513	T/C	4	8552
112761718	*ANKK1*	rs877138	flanking 5'UTR	0.3431	0.3456	0.3336	0.3003	*NA*	G/A	3	6347
112768580	*ANKK1*	rs4590907	Intron	0.1436	0.1397	0.1362	0.2342	0.2487	G/T	4	8578
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	0.1870	0.1869	0.1565	*NA*	A/G	3	6339
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.44529	*NA*	0.3059	0.3775	0.0000	G/C	3	7239
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	0.4976	0.4925	0.4624	0.4750	G/A	4	8575
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys,TaqI A	0.2026	0.2015	0.1976	0.1703	0.1806	A/G	4	8561
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	*NA*	0.4505	0.4658	0.4537	A/G	3	7249
112801119	*DRD2*	rs1076563	Intron	0.4127	0.4041	0.3924	0.4972	*NA*	C/A	3	6346
112815891	*DRD2*	rs7125415	Intron	0.09914	0.0944	0.0959	0.1896	*NA*	T/C	3	6346
112818599	*DRD2*	rs4648318	Intron	0.2538	0.2473	0.2439	0.3407	*NA*	C/T	3	6346
112824662	*DRD2*	rs4274224	Intron	0.4823	0.4814	0.4917	0.2391	*NA*	G/A	3	6342
112829684	*DRD2*	rs4581480	Intron	0.09919	0.0954	0.1027	0.0683	*NA*	C/T	3	6344
112834984	*DRD2*	rs7131056	Intron	0.4205	0.4202	0.4264	0.4942	*NA*	C/A	3	6341
112846601	*DRD2*	rs4938019	Intron	0.1447	0.1441	0.1401	0.2335	0.2392	C/T	4	6334

112852165	*DRD2*	rs12364283	flanking 5'UTR	0.07563	0.0779	0.0775	0.0787	*NA*	G/A	3	8583
112857971	*DRD2*	rs10891556	intergenic	0.1758	0.1751	0.1734	0.2380	*NA*	T/G	3	6339
112860946	*DRD2*	rs6589377	intergenic	0.3776	*NA*	0.3820	0.1689	*NA*	G/A	4	6345
112863421	*DRD2*	rs4482060	intergenic	0.30032	*NA*	0.4330	0.2589	0.2545	T/A	3	8576

Supplementary Table 2. Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and self-reported smoking behavior for *Smokers* vs *Never-Tried*.

Notes: *Smokers* (N=2819) are compared *Never-Tried* (N=6049) (binary logistic regression). Results from NFBC1966, NFBC1986, IMAGEN and ALSPAC are combined using meta-analysis. FEM=Fixed effect model, REM=Random effect model, *Q*=Cochrane's *Q* statistic, I^2=heterogeneity index (0-100). Significant p values are in bold, of these those that remain significant after Bonferroni correction for multiple testing are indicated by an asterisk. Thick lines indicate LD blocks.

BP	Gene	SNP	Location	MAF	Ref allele/ Other allele	N		P		OR		Q	I^2
						Samples	Individuals	(FEM)	(REM)	(FEM)	(REM)		
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	C/T	4	8565	0.01		1.09		0.26	25
112691986	*TTC12*	rs2236709	Intron	0.2588	G/A	4	8866	0.001*		1.13		0.46	0
112693407	*TTC12*	rs7927508	Intron	0.3667	G/A	3	6336	0.008		1.12		0.26	25.84
112694133	*TTC12*	rs2156486	Intron	0.1814	G/T	4	8569	0.022		0.9		0.46	0
112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	C/A	3	6344	0.3		0.96		0.39	0
112704356	*TTC12*	rs10502172	Intron	0.4596	T/C	4	8633		0.14		0.9	0.01	76.24
112705919	*TTC12*	rs2303380	Intron splice site	0.3819	G/A	3	6344		0.83		1.02	0.001	84.44
112716539	*TTC12*	rs2288159	Intron	0.1561	T/G	3	6343	0.009		1.14		0.86	0
112721133	*TTC12*	rs4987094	Intron	0.1004	A/G	3	6347	0.013		1.16		0.94	0

112735810	*TTC12*	rs2276070	Intron	0.1541	T/C	3	6345	0.009		1.16		0.81	0
112739889	*TTC12*	rs719802	Intron	0.3953	T/C	3	6344	0.21	0.66	1.05	1.04	0.011	77.98
112739985	*TTC12*	rs719804	Intron	0.249	G/A	3	6334	0.09		1.09		0.19	39.05
112749387	*TTC12*	rs2282511	flanking 3'UTR	0.3455	A/C	3	7145		0.19		1.11	0.01	77.25
112754346	*TTC12*	rs754672	flanking 3'UTR	0.4894	T/C	4	8552		0.07		0.89	0.02	68.26
112761718	*ANKK1*	rs877138	flanking 5'UTR	0.3431	G/A	3	6347		0.65		1.05	0.004	81.87
112768580	*ANKK1*	rs4590907	Intron	0.1436	G/T	4	8578	0.05		1.08		0.43	0
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	A/G	3	6339	0.78		0.99		0.58	0
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.44529	G/C	3	7239		0.06		1.12	0.08	61.34
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	G/A	4	8575		0.09		0.9	0.03	66.72
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys,TaqI	0.2026	A/G	4	8561	0.85		1.01		0.72	0

			A										
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	A/G	3	7249		0.03		0.89	0.09	59.03
112801119	*DRD2*	rs1076563	Intron	0.4127	C/A	3	6346	0.003		0.89		0.14	48.44
112815891	*DRD2*	rs7125415	Intron	0.09914	T/C	3	6346	0.01		1.16		0.77	0
112818599	*DRD2*	rs4648318	Intron	0.2538	C/T	3	6346	0.003		1.14		0.18	41.98
112824662	*DRD2*	rs4274224	Intron	0.4823	G/A	3	6342	0.93		1		0.14	49.82
112829684	*DRD2*	rs4581480	Intron	0.09919	C/T	3	6344		0.53		1.07	0.1	56.29
112834984	*DRD2*	rs7131056	Intron	0.4205	C/A	3	6341	0.22		1.05		0.31	15.05
112846601	*DRD2*	rs4938019	Intron	0.1447	C/T	4	6334	0.09		1.08		0.52	0
112852165	*DRD2*	rs12364283	flanking 5'UTR	0.07563	G/A	3	8583	0.88		1.01		0.88	0
112857971	*DRD2*	rs10891556	intergenic	0.1758	T/G	3	6339	0.51		1.03		0.38	0
112860946	*DRD2*	rs6589377	intergenic	0.3776	G/A	4	6345		0.86		1.01	0.06	59.84
112863421	*DRD2*	rs4482060	intergenic	0.30032	T/A	3	8576	0.68		1.02		0.14	48.38

Supplementary Table 3. Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and self-reported smoking behavior for *Weekly-Smokers* vs *Never-Tried*.

Notes: *Weekly-Smokers* (N=1624) are compared to *Never-Tried* (N=6049) (binary logistic regression). Results from NFBC1966, ALSPAC, NFBC1986, and IMAGEN are combined using meta-analysis. FEM=Fixed effect model, REM=Random effect model, *Q*=Cochrane's *Q* statistic, I^2=heterogeneity index (0-100). Significant p values are in bold (uncorrected, Bonferroni corrected threshold: 0.0015 are indicated by *). Thick lines indicate LD blocks.

BP	**Gene**	**SNP**	**Location**	**MAF**	**Ref/**	**N**		***P***		***OR***		***Q***	**I^2**
					Other allele	**Samples**	**Individuals**	**(FEM)**	**(REM)**	**(FEM)**	**(REM)**		
112685412	*TTC12*	rs4517559	flanking 5'UTR	0.3736	C/T	4	7385	**0.04**		1.09		0.72	0
112691986	*TTC12*	rs2236709	Intron	0.2588	G/A	4	7672	**0.001***		1.16		0.94	0
112693407	*TTC12*	rs7927508	Intron	0.3667	G/A	3	5281	**0.05**		1.12		0.62	0
112694133	*TTC12*	rs2156486	Intron	0.1814	G/T	4	7390	**0.04**		0.89		0.61	0

112699378	*TTC12*	rs723077	Coding, Met73Leu	0.4894	C/A	3	5287	0.45		0.96		0.43	0
112704356	*TTC12*	rs10502172	Intron	0.4596	T/C	4	7456		0.08		0.89	0.08	56.11
112705919	*TTC12*	rs2303380	Intron splice site	0.3819	G/A	3	5288		0.63		1.06	0.03	72.38
112716539	*TTC12*	rs2288159	Intron	0.1561	T/G	3	5286	0.11		1.12		0.96	0
112721133	*TTC12*	rs4987094	Intron	0.1004	A/G	3	5290	0.09		1.15		0.8	0
112735810	*TTC12*	rs2276070	Intron	0.1541	T/C	3	5289	0.11		1.12		0.97	0
112739889	*TTC12*	rs719802	Intron	0.3953	T/C	3	5289		0.43		1.07	0.12	52.08
112739985	*TTC12*	rs719804	Intron	0.249	G/A	3	5278		0.22		1.14	0.12	53.62
112749387	*TTC12*	rs2282511	flanking 3'UTR	0.3455	A/C	3	6077		0.07		1.14	0.12	52.19
112754346	*TTC12*	rs754672	flanking 3'UTR	0.4894	T/C	4	7377	**0.008**		0.89		0.13	46.17
112761718	*ANKK1*	rs877138	flanking 5'UTR	0.3431	G/A	3	5290		0.46		1.08	0.08	60.56

112768580	*ANKK1*	rs4590907	Intron	0.1436	G/T	4	7398	0.41		1.04		0.79	0
112772031	*ANKK1*	rs7118900	Coding, Ala239Thr	0.1896	A/G	3	5283	0.84		0.99		0.7	0
112775225	*ANKK1*	rs4938016	Coding, Gly442Arg	0.44529	G/C	3	6153	**0.01**		1.12		0.14	48.79
112775370	*ANKK1*	rs2734849	Coding, His490Arg	0.4886	G/A	4	7397	**0.005**		0.89		0.34	10.17
112776038	*ANKK1*	rs1800497	Coding, Glu713Lys	0.2026	A/G	4	7386	0.65		1.02		0.82	0
112788669	*DRD2*	rs6277	Coding, Pro219Pro	0.12201	A/G	3	6161	**0.005**		0.88		0.39	0
112801119	*DRD2*	rs1076563	Intron	0.4127	C/A	3	5289	**0.03**		0.88		0.55	0
112803549	*DRD2*	rs2471857	Intron	0.1577	T/C	3	5289	0.91		1.01		0.54	0
112815891	*DRD2*	rs7125415	Intron	0.09914	T/C	3	5289	0.24		1.1		0.91	0
112818599	*DRD2*	rs4648318	Intron	0.2538	C/T	3	5286	**0.02**		1.16		0.32	11.96
112824662	*DRD2*	rs4274224	Intron	0.4823	G/A	3	5287	0.85		0.99		0.45	0

112829684	*DRD2*	rs4581480	Intron	0.09919	C/T	3	5284		0.87		1.03	0.09	57.52
112834984	*DRD2*	rs7131056	Intron	0.4205	C/A	3	5280	0.09		1.1		0.61	0
112846601	*DRD2*	rs4938019	Intron	0.1447	C/T	4	7404	0.06		1.1		0.99	0
112852165	*DRD2*	rs12364283	flanking 5'UTR	0.07563	G/A	3	5283	0.63		1.05		0.22	34.83
112857971	*DRD2*	rs10891556	intergenic	0.1758	T/G	3	5289	0.26		1.09		0.83	0
112860946	*DRD2*	rs6589377	intergenic	0.3776	G/A	4	7396	0.58		1.03		0.51	0
112863421	*DRD2*	rs4482060	intergenic	0.30032	T/A	3	6148	0.18		1.07		0.466	0

Supplementary Table 4. Association between SNPs spanning the *TTC12-ANKK1-DRD2* gene-cluster and cotinine level (quantile-transformed in the ALSPAC cohort (N=2,540) (linear regression). Significant p values are in bold (uncorrected, Bonferroni corrected threshold: 0.0015 are indicated by *). Thick lines indicate LD blocks.

SNP	**Gene**	**BP**	**Location**	**MAF**	***Ref allele/ Other allele***	**BETA**	**STAT**	***p***
rs4517559	*TTC12*	1.1E+08	flanking 5'UTR	0.3736	C/T	0.09	3.18	**0.001***
rs2236709	*TTC12*	1.1E+08	Intron	0.2588	G/A	0.11	3.49	**0.0005***
rs7927508	*TTC12*	1.1E+08	Intron	0.3667	G/A	0.09	3.16	**0.002**
rs2156486	*TTC12*	1.1E+08	Intron	0.1814	G/T	0.03	0.82	0.41
rs723077	*TTC12*	1.1E+08	Coding, Met73Leu	0.4894	C/A	-0.03	-1.2	0.23
rs10502172	*TTC12*	1.1E+08	Intron	0.4596	T/C	-0.01	-0.36	0.72
rs2303380	*TTC12*	1.1E+08	Intron splice site	0.3819	G/A	-0.02	-0.77	0.49
rs2288159	*TTC12*	1.1E+08	Intron	0.1561	T/G	0.06	1.58	0.11
rs4987094	*TTC12*	1.1E+08	Intron	0.1004	A/G	0.04	0.82	0.41
rs2276070	*TTC12*	1.1E+08	Intron	0.1541	T/C	0.06	1.57	0.12
rs719802	*TTC12*	1.1E+08	Intron	0.3953	T/C	-0.01	-0.29	0.78
rs719804	*TTC12*	1.1E+08	Intron	0.249	G/A	-0.01	-0.4	0.69

rs2282511	*TTC12*	1.1E+08	flanking 3'UTR	0.3455	A/C	0	0.15	0.88
rs754672	*TTC12*	1.1E+08	flanking 3'UTR	0.4894	T/C	-0.03	-1.19	0.23
rs877138	*ANKK1*	1.1E+08	flanking 5'UTR	0.3431	G/A	0.01	0.35	0.73
rs4590907	*ANKK1*	1.1E+08	Intron	0.1436	G/T	0.07	1.7	0.09
rs7118900	*ANKK1*	1.1E+08	Coding, Ala239Thr	0.1896	A/G	-0.05	-1.36	0.18
rs4938016	*ANKK1*	1.1E+08	Coding, Gly442Arg	0.44529	G/C	0.07	2.34	**0.02**
rs2734849	*ANKK1*	1.1E+08	Coding, His490Arg	0.4886	G/A	-0.03	-1.17	0.24
rs1800497	*ANKK1*	1.1E+08	Coding, Glu713Lys, TaqI A	0.2026	A/G	-0.04	-0.98	0.33
rs6277	*DRD2*	1.1E+08	Coding, Pro219Pro	0.12201	A/G	-0.05	-1.81	0.07
rs1076563	*DRD2*	1.1E+08	Intron	0.4127	C/A	-0.05	-1.64	0.1
rs2471857	*DRD2*	1.1E+08	Intron	0.1577	T/C	-0.02	-0.42	0.67
rs7125415	*DRD2*	1.1E+08	Intron	0.09914	T/C	0.07	1.45	0.15
rs4648318	*DRD2*	1.1E+08	Intron	0.2538	C/T	0.06	1.97	**0.05**
rs4274224	*DRD2*	1.1E+08	Intron	0.4823	G/A	-0.03	-1.03	0.3
rs4581480	*DRD2*	1.1E+08	Intron	0.09919	C/T	-0.01	-0.2	0.84

rs7131056	*DRD2*	1.1E+08	Intron	0.4205	C/A	0	-0.07	0.94
rs4938019	*DRD2*	1.1E+08	Intron	0.1447	C/T	0.05	1.36	0.17
rs12364283	*DRD2*	1.1E+08	flanking 5'UTR	0.07563	G/A	0.05	0.87	0.38
rs10891556	*DRD2*	1.1E+08	intergenic	0.1758	T/G	0.04	1.06	0.29
rs6589377	*DRD2*	1.1E+08	intergenic	0.3776	G/A	-0.04	-1.24	0.22
rs4482060	*DRD2*	1.1E+08	intergenic	0.30032	T/A	-0.03	-1.01	0.31

FIGURE LEGENDS:

Supplementary-Figure 1: Linkage disequilibrium (LD) of the chromosome 11q23 region in the NFBC1966 **(A)**, ALSPAC **(B)**, IMAGEN **(C)** and NFBC1986 **(D)** cohorts. Three haplotype blocks are indicated with borders. LD is coded according to the following color scheme: LOD < 2 and D' < 1: white; LOD < 2 and D' = 1: blue; LOD ≥ 2 and D' < 1: shades of red; LOD ≥ 2 and D' = 1: bright red. The *TTC12-ANKK1-DRD2* gene-cluster covers a large genomic area (chromosome 11 *TTC12*: 112,690,539-112,749,226, 58,688 bp; *ANKK1*: 112,763,723-112,776,350, 12,628 bp; *DRD2* short isoform: 112,785,528-112,851,103, 12,628 bp and *DRD2* long isoform: 112,785,528-112,851,103, 65,576 bp).

Supplementary-Figure 2: (A) Box plot of blood-oxygen level-dependent (BOLD) response in left ventral striatum stratified by rs2236709 genotype (note that G-carriers are pooled together). The x-axis represents the number of minor G-alleles (N_{AA}=706, $N_{AG/GG}$=557).

Mean BOLD response within each group was as follows: M_{AA}=0.365, SD_{AA}=0.370 and $M_{AG/GG}$=0.417, $SD_{AG/GG}$=0.391. (**B)** Region of interest as extracted from the contrast 'anticipation of high reward > anticipation for no reward' from the modified Monetary Incentive Delay (MID) task in the IMAGEN sample (N=1263) indicating left and right ventral striatum centred at Montreal Neurological Institute (MNI): x=±15, y=9, z=-9.

REFERENCES

1. Hibell B AB, Bjarnason T, Ahlstro¨m S, Balakireva O,, A K. The ESPAD Report: Alcohol and
Other Drug Use Among Students in 35 European Countries. Modintryckoffset AB: Stockholm, 2003.

2. Hukkanen J, Jacob P, 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. Pharmacol Rev. 2005; 57(1): 79-115.

3. Benowitz NL, Kuyt F, Jacob P, 3rd, Jones RT, Osman AL. Cotinine disposition and effects. Clinical Pharmacology & Therapeutics. 1983; 34(5): 604-11.

4. Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D, et al. Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. Proc Natl Acad Sci U S A. 2011; 108(17): 7119-24.

5. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature genetics. 2009; 41(1): 35-46.

6. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature genetics. 2012.

7. Peters J, Bromberg U, Schneider S, Brassen S, Menz M, Banaschewski T, et al. Lower ventral striatal activation during reward anticipation in adolescent smokers. The American journal of psychiatry. 2011; 168(5): 540-9.

8. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Buchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Molecular psychiatry. 2010; 15(12): 1128-39.

9. The MathWorks Inc. MATLAB. The MathWorks Inc., 2011.

10. Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage. 2007; 38(1): 95-113.

11. Yacubian J, Glaescher J, Schroeder K, Sommer T, Braus DF, Buechel C. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. Journal of Neuroscience. 2006; 26(37): 9530-7.

12. Brett M, Anton J-L, Valabregue R, Poline J-B. Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage. 2002; 16: S497.

13. Ducci F, Enoch MA, Yuan Q, Shen PH, White KV, Hodgkinson C, et al. HTR3B is associated with alcoholism with antisocial behavior and alpha EEG power--an intermediate phenotype for alcoholism and co-morbid behaviors. Alcohol. 2009; 43(1): 73-84.

14. Wang D, Sun Y, Stang P, Berlin JA, Wilcox MA, Li Q. Comparison of methods for correcting population stratification in a genome-wide association study of rheumatoid arthritis: principal-component analysis versus multidimensional scaling. BMC Proc. 2009; 3 Suppl 7: S109.

15. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3): 559-75.

16. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. British Medical Journal. 2003; 327(7414): 557-60.