1. Extended description of Quality Control methods and results

Genotyping Workflow

Concentration, fragmentation and response to PCR were determined. The concentrations of all samples were adjusted to 50 ng/μl and 15 μl of each sample robotically dispensed into barcoded 96-well plates. Samples from cases and controls were randomly distributed on plates.The plates were processed in a fully automated Illumina BeadLab equipped with liquid handling robots (Tecan Ltd, Dorset, UK), Illumina BeadArray readers and Illumina iScans (Illumina Ltd, CA, USA) . Genotyping was carried out using the Illumina HumanHap610 quad array according to manufacturer's recommendations. The raw data were analyzed using Beadstudio ((Illumina Ltd, CA, USA) and extracted for statistical analysis.

Confirmatory genotyping of SNPs showing suggestive evidence of association was performed to confirm the calling of alleles and to provide verification of the genotypes for imputed SNPs and was carried out at the MRC SGDP Centre using a Biomek FX liquid handling robot and a Equator Low Volume Pipetting System (Deerac Fluidics), PCR with predesigned Taqman® SNP assays was performed on MJ research 384 PCR machines and genotypes were called on an ABI 7900 with SDS v2.0 (PE Applied Biosystems, Foster City, CA, USA).

Quality control

Quality control procedures were performed using PLINK v1.04 (Purcell et al. 2007) using stringent thresholds of completeness, minor allele frequency (MAF) and Hardy-Weinberg equilibrium separately in case and control data sets, to ensure high quality results.

An initial analysis of individual-level data excluded samples with genotyping completeness of less 95% or outlier values of heterozygosity (<0.29 or >0.36), where gender assigned from genotypic data was inconclusive or inconsistent with phenotypic information (in cases only). Non-European ancestry was determined using EIGENSTRAT (see below). Related or duplicate individuals within and across case and control samples were identified through identity-by-state sharing analysis on an linkage disequilibrium-pruned set of SNPs (~ 18K SNPs); for each pair related up to second degree relationships, the individual with lower genotyping completeness was omitted.

To optimize data quality, all SNP QC thresholds were applied separately to cases and controls, then merged data sets were formed for analysis. SNP genotyping completeness was higher in cases than in controls, due primarily to DNA from 1289 controls (73.7%) being collected from cheek swabs. To ensure non-differential missing rates in cases and controls, genotyping completeness thresholds were applied separately in each group. SNPs with call rate < 99%; Hardy-Weinberg equilibrium P-values < 10^{-5} in cases or

controls, and minor allele frequency < 1% were removed. Finally, based on this set of high-quality SNPs, samples that had genotyping completeness < 99% were excluded.

Supplementary Table 1 shows the breakdown of SNP and sample filtering during the quality control procedures of the depression cases and controls. Analyses were performed using the PLINK v 1.04 and 1.05.

EIGENSTRAT analysis

All samples were tested for population stratification using the software EIGENSTRAT (Price et al. 2006), which performs a principal components analysis (PCA) on SNP data, in a two stage process. Analysis was performed on a subset of 80,304 SNPs selected from all genotyped SNPs, omitting regions of high LD.

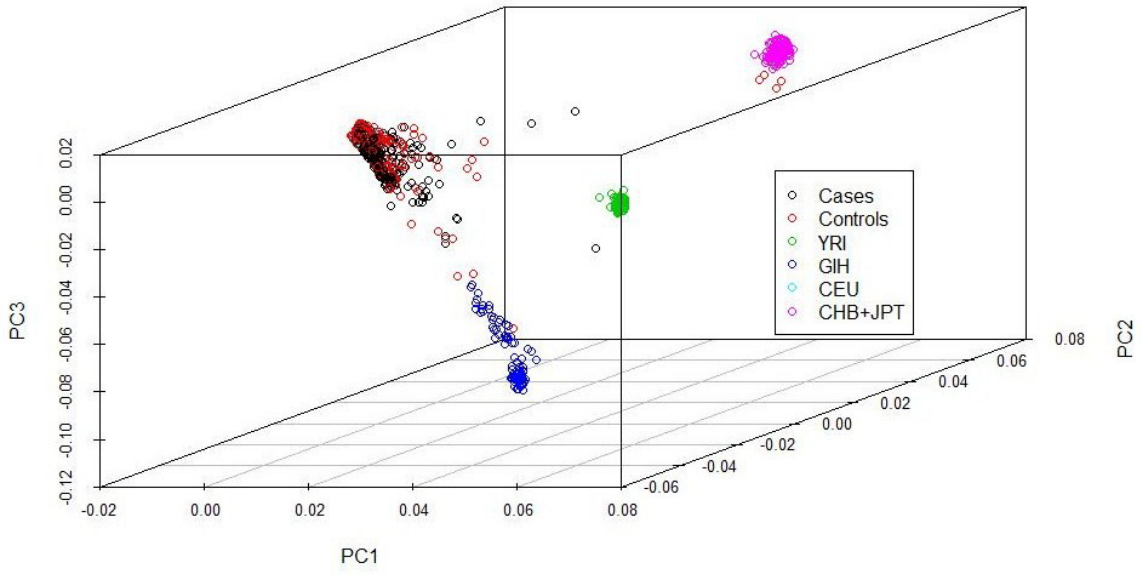
Individuals of non-European ancestry were identified by combining study genotypes with genotypes from HapMap data from the CEU, YRI, CBT, JPT and GIH populations (with GIH included since much of non-European ancestry in the UK is from the Indian sub-continent). In total, 9 cases and 12 controls were omitted (Supplementary Figure 1). The UK-ascertained cases provided a good match to control samples, and two principal components were necessary to correct for differences between cases and controls.

Further analysis was performed on the cleaned data set with SNPs and individuals identified above omitted. Principal component loadings differed between cases and controls on the first two principal components, and these were therefore included as covariates in testing for association.

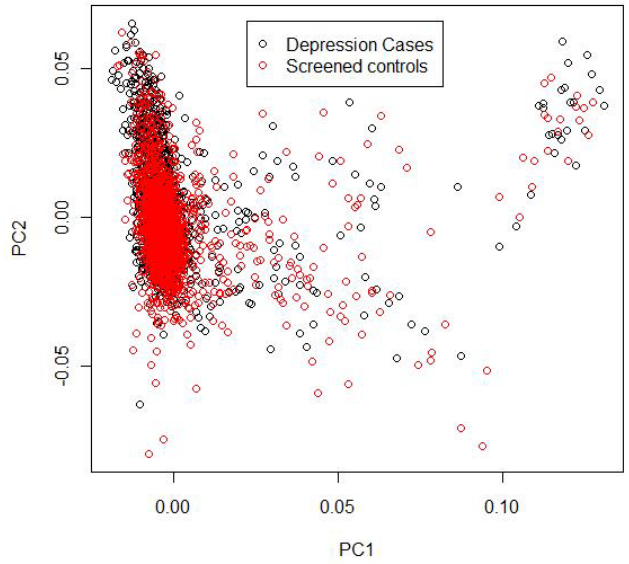
Supplementary Table 1; Breakdown of SNP and sample filtering during the quality control procedures of the depression case and control data. Association analysis is based on 471,747 SNPs that passed QC in both cohorts.

	Cases		**Controls**	
	SNPs	**Individuals**	**SNPs**	**Individuals**
No. before QC	582,574	1755	582,574	1,745
No. dropped at QC step:				
1. Genotyping completeness < 99%	-	59	-	117
2. Extreme heterozygosity	-	3	-	6
3. Non-European ancestry	-	9	-	12
4. Inconsistent genotype sex	-	16	-	4
5. Close relatives, duplicates	-	32	-	12
6. Call rate < 99%	18,527	-	82,875	-
7. Minor allele frequency < 0.01	31,136	-	25,827	-
8. HWE P-values < 10^{-5}	546	-	364	-
No. remaining after QC	532,365	1,636	473,508	1,594

Supplementary Figure 1:
HapMap CEU, YRI, CHB, JPT and GIH populations with cases and controls, to identify individuals study subjects of non-European origin



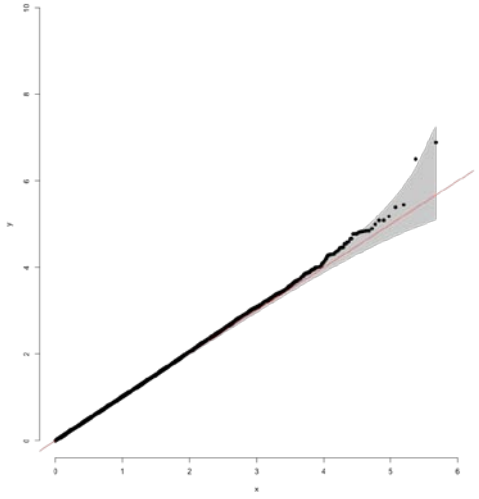
Supplementary Figure 2: Individual loadings on principal components 1 and 2 (PC1, PC2) for EIGENSTRAT analysis for depression cases v. controls



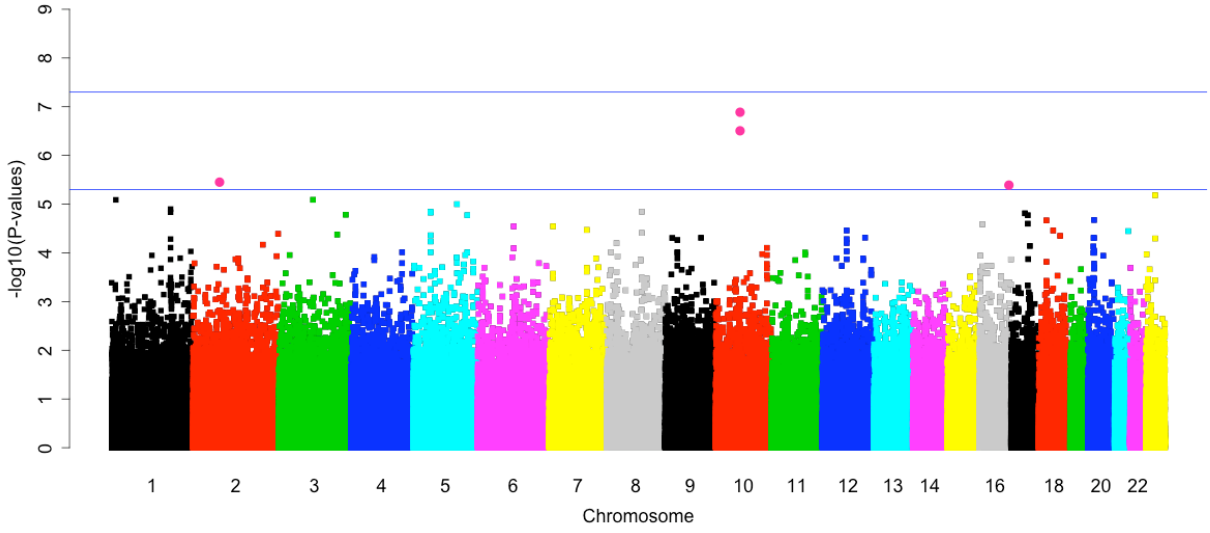
Association testing

The primary analysis was logistic regression for case control status on SNP genotypes, under an additive model for genotype (i.e. coded 0, 1, 2), with covariates comprising the eigenvector loadings for two principal components. Additional analysis was performed using logistic regression under a dominant and recessive inheritance models, and using a full genotypic model.

Supplementary Figure 3: Quantile-quantile plot of observed $-\log_{10}$(p-values) for 471,474 SNPs, analysed with logistic regression under an additive genetic model, with two ancestry principal components as covariates. Genomic control λ=1.02.



Supplementary Figure 4: Manhattan plot of genomic location against -1og$_{10}$(p-value) from logistic regression test statistic, showing 4 SNPs (on chromosomes 2, 10 and 16) exceeding the threshold shown for suggestive significance ($p=5 \times 10^{-6}$) but no SNPs reaching genome-wide significance ($p=5 \times 10^{-8}$).



References

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904-909.

Purcell, S., Cherny, S.S., and Sham, P.C. 2003. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19: 149-150.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.

Supplementary Table 2: Analysis of SNPs in candidate genes for association with depression (2654 SNPs in 84 genes). The SNP achieving the minimum p-value in each gene is shown, with genes ordered by Bonferroni-corrected p-values for each gene.

Candidate Gene	Chr.	SNP	Position (base pairs)	Alleles	No. SNPs in gene	P-value	Odds Ratio	Frequency in Cases	Frequency in Controls
PDE9A	21	rs13050655	43045971	A/G	53	3.58E-05	0.60	0.03	0.06
SLC6A1	3	rs2697153	10998672	G/A	42	0.0002623	0.83	0.37	0.41
NR3C1	5	rs10482682	142659590	T/C	20	0.001022	1.19	0.40	0.36
TBX21	17	rs11650354	43177091	T/C	5	0.007115	1.19	0.19	0.16
DRD5	4	rs1878943	9375986	T/C	4	0.01851	1.16	0.20	0.17
FGFR3	4	rs743682	1767650	A/G	2	0.04806	0.84	0.08	0.09
HTR2C	23	rs2428700	113916920	A/G	11	0.008948	0.82	0.15	0.18
HTR4	5	rs17108435	147933440	C/T	38	0.002617	0.80	0.12	0.15
AVPR1A	12	rs7308008	61844229	A/G	9	0.0118	0.84	0.14	0.16
SLC6A2	16	rs3785152	54274051	T/C	36	0.003313	1.28	0.12	0.09
AVPR1B	1	rs28419084	204413276	T/C	10	0.01521	1.24	0.10	0.09
CACNA1C	12	rs7960297	2416780	T/C	173	0.0009678	0.74	0.08	0.10
GSK3B	3	rs16830594	121088768	G/A	23	0.007536	0.81	0.11	0.13
FGFR1	8	rs13279569	38460773	A/G	12	0.01877	0.80	0.07	0.08
GRIN1	9	rs10870198	139162141	G/A	3	0.08201	0.91	0.38	0.41
SLC6A4	17	rs1487971	25596879	T/C	13	0.02009	0.89	0.37	0.40
GRIN3A	9	rs1983812	103372864	T/C	53	0.004983	0.87	0.48	0.51
PDE11A	2	rs9288008	178254180	C/T	96	0.003993	0.76	0.06	0.09
PSMD9	12	rs895959	120804735	C/T	5	0.08273	1.14	0.13	0.12
ZNF804A	2	rs1583048	185491386	C/T	30	0.01382	0.85	0.19	0.21
GRIN2D	19	rs1799286	53624149	T/C	8	0.05202	0.78	0.04	0.05
GRIK2	6	rs609531	102161603	C/A	134	0.003194	1.18	0.31	0.28
GRIK1	21	rs466612	30186260	T/G	109	0.00415	1.45	0.05	0.03
GRIN2A	16	rs17571080	10199561	A/G	137	0.003359	1.22	0.18	0.15
PDE1A	2	rs1430153	183109918	C/T	69	0.006749	1.17	0.26	0.24
HTR7	10	rs7916720	92616547	C/T	20	0.02339	0.86	0.19	0.21
LEP	7	rs791608	127660228	T/C	10	0.04766	1.22	0.07	0.06
TPH1	11	rs172424	17979526	G/T	4	0.1207	0.92	0.44	0.46
GRIK4	11	rs1944522	120370132	T/C	99	0.005053	1.15	0.51	0.48
HTR2A	13	rs1328674	46339708	T/C	46	0.01099	0.73	0.04	0.05
DRD2	11	rs4274224	112824662	G/A	26	0.02111	0.89	0.47	0.50
MAOA	23	rs1137070	43488335	T/C	9	0.06552	1.12	0.30	0.28
PER2	2	rs7582286	238880426	C/A	10	0.06014	0.90	0.31	0.33
GSK3A	19	rs11878620	47439069	A/G	1	0.618	1.05	0.07	0.07
FGFR4	5	rs451643	176444038	T/G	4	0.1616	0.92	0.29	0.31
GRIN2C	17	rs1568447	70348607	C/T	5	0.1307	1.08	0.37	0.35
LEPR	1	rs6588147	65708082	G/A	40	0.01686	0.88	0.33	0.35
GRIA3	23	rs5911634	122461630	A/C	45	0.01499	1.16	0.30	0.26
OLIG1	21	rs2834078	33345769	G/T	6	0.1183	0.83	0.04	0.05
ADRA2C	4	rs6822427	3749870	T/C	4	0.1845	0.94	0.48	0.50

AKT1	14	rs2498804	104304140	A/C	1	0.7907	0.99	0.31	0.31
BDNF	11	rs908867	27702340	T/C	14	0.06222	0.85	0.09	0.10
DRD4	11	rs4963126	646845	A/G	3	0.3119	0.95	0.44	0.45
PSMB4	1	rs4603	149640649	C/T	5	0.1965	1.09	0.18	0.17
PER1	17	rs2304911	7991704	G/A	5	0.2044	0.86	0.04	0.05
SLC6A3	5	rs2292023	1514389	A/C	23	0.04633	0.86	0.12	0.13
CRHR1	17	rs17763104	41261576	A/G	15	0.0769	0.87	0.12	0.14
GRIK5	19	rs8099939	47212948	T/G	5	0.2369	1.06	0.44	0.42
ABCB1	7	rs9282564	87067376	C/T	56	0.02335	0.83	0.10	0.11
PER3	1	rs10462018	7802214	T/C	17	0.07858	1.13	0.17	0.15
IL6	7	rs2069837	22734552	G/A	17	0.08029	1.19	0.08	0.07
ADRA1B	5	rs17455628	159281476	G/T	19	0.07303	1.21	0.06	0.05
OLIG2	21	rs762236	33301494	A/G	11	0.1301	0.93	0.47	0.49
CD3E	11	rs12576947	117663265	C/T	5	0.2869	0.94	0.32	0.34
HTR1A	5	rs1364043	63286607	G/T	2	0.7879	1.02	0.24	0.24
GRIA4	11	rs597395	105226679	A/G	48	0.03455	0.78	0.04	0.05
GNB3	12	rs4963511	6813631	A/G	13	0.1288	1.11	0.16	0.14
ANK3	10	rs2393596	61513573	A/C	114	0.015	1.17	0.20	0.18
HTR2B	2	rs17586405	231687502	G/A	2	0.8881	1.02	0.03	0.03
HTR1B	6	rs2000292	78223664	A/G	13	0.1396	0.92	0.25	0.26
HTR3B	11	rs2011249	113273848	A/G	13	0.1424	1.10	0.20	0.18
STAT3	17	rs9252	37808375	A/G	10	0.1967	0.89	0.08	0.09
PCLO	7	rs7807790	82621782	C/T	68	0.0293	1.13	0.29	0.26
NTRK2	9	rs7875184	86690918	T/C	74	0.02723	0.85	0.12	0.14
CREB1	2	rs2709373	208094269	C/T	7	0.2989	1.07	0.19	0.18
HTR5A	7	rs2919435	154489085	A/G	17	0.1256	0.93	0.39	0.41
HTR3A	11	rs11214800	113368140	C/A	19	0.1233	0.93	0.47	0.49
OLIG3	6	rs13200244	137867324	T/G	14	0.1743	0.92	0.22	0.23
TPH2	12	rs12231341	70695703	T/C	30	0.08216	0.81	0.04	0.05
COMT	22	rs737866	18310109	C/T	22	0.1159	1.10	0.29	0.27
FKBP5	6	rs9380529	35783674	G/A	18	0.1647	0.93	0.48	0.50
PRKCH	14	rs959729	61003953	T/C	74	0.0421	0.80	0.06	0.07
GRIA2	4	rs4691394	158468483	A/G	5	0.6575	0.97	0.14	0.14
P2RX7	12	rs10849851	120081027	G/A	23	0.1595	1.20	0.05	0.04
DRD1	5	rs265995	174782552	T/C	18	0.2053	0.91	0.13	0.14
DRD3	3	rs11706283	115367825	T/C	22	0.1834	1.12	0.10	0.09
ADRA2B	2	rs4426564	96144713	C/T	5	0.8428	1.01	0.33	0.33
ADRA1A	8	rs4236679	26682474	A/G	64	0.07791	0.83	0.05	0.06
ADRA2A	10	rs11195419	112829358	A/C	10	0.5259	0.95	0.12	0.12
CLOCK	4	rs3792603	55996815	G/A	10	0.5303	0.96	0.20	0.20
GRIA1	5	rs11741924	153103056	A/G	91	0.0656	0.91	0.32	0.35
GRIN2B	12	rs7970177	13630255	T/C	164	0.03872	1.20	0.09	0.08
GRIK3	1	rs1027599	37164007	G/A	29	0.2482	0.90	0.09	0.10
FGFR2	10	rs3135761	123266081	A/G	35	0.2361	1.08	0.17	0.16

Supplementary Table 3: Association analysis of UK depression males cases (n=477) against all controls (n=1594), using logistic regression with two covariates of ancestry principal components. Genomic control λ=1.02

Chr.	SNP	Position in base pairs	Alleles	P-value	Odds Ratio	Frequency in Cases	Frequency in Screened controls	Closest gene
8	rs6989226	15710484	C/G	1.81E-06	1.6	0.1897	0.129	*TUSC3*
17	rs16957797	8839862	G/T	6.70E-06	0.5771	0.09644	0.1518	*LOC728685*
14	rs4982207	20102464	T/C	8.21E-06	1.418	0.4653	0.3858	*RNASE9*
5	rs253890	116246177	C/T	8.57E-06	1.799	0.1038	0.06152	*SEMA6A*
12	rs7306356	124342083	T/C	1.07E-05	0.6874	0.2379	0.3128	*TMEM132B*
2	rs12471464	67058146	C/T	1.74E-05	1.475	0.2243	0.1647	*LOC644838*
2	rs17035931	47035536	C/A	2.21E-05	0.6651	0.1866	0.2487	*TTC7A*
4	rs12643648	79543504	T/G	2.23E-05	0.6751	0.1981	0.2604	*FRAS1*
22	rs9306489	43904148	G/A	2.34E-05	1.392	0.3742	0.3008	*NUP50*
2	rs17415659	1476810	T/C	2.48E-05	0.7149	0.3008	0.3739	*TPO*
2	rs4663163	235290899	A/G	2.76E-05	1.377	0.4088	0.3369	*ARL4C*
13	rs9549067	39600657	A/G	2.99E-05	1.415	0.3155	0.2459	*LOC646953*
4	rs13116982	189273316	G/A	3.13E-05	1.589	0.1405	0.09379	*TRIML2*
2	rs4927594	1504565	C/T	3.32E-05	1.363	0.5335	0.4601	*TPO*
8	rs2049833	15711171	G/A	3.42E-05	1.581	0.1447	0.09849	*TUSC3*
20	rs2210455	9867383	T/G	3.70E-05	0.6735	0.1783	0.2391	*ANKRD5*
3	rs10513580	163422601	C/T	4.38E-05	1.537	0.1604	0.1094	*LOC131149*
5	rs10043664	54464067	A/C	4.59E-05	0.6212	0.1059	0.1603	*CDC20B*
1	rs12143304	86278343	G/A	4.78E-05	1.487	0.196	0.1402	*COL24A1*
16	rs1947261	58964075	G/A	5.02E-05	1.353	0.5189	0.4412	*LOC729159*

Supplementary Table 4: Association analysis of UK depression females cases (n=1152) against all controls (n=1594), using logistic regression with two covariates of ancestry principal components. Genomic control λ=1.02

Chr.	SNP	Position in base pairs	Alleles	P-value	Odds Ratio	Frequency in Cases	Frequency in Screened controls	Closest gene
10	rs9416742	60212700	A/G	1.82E-08	0.6729	0.1762	0.2346	*BICC1*
10	rs999845	60205926	T/C	3.80E-08	0.6801	0.1785	0.2358	*BICC1*
8	rs987390	86734603	G/A	1.58E-06	0.7656	0.4362	0.5016	*REXO1L1*
17	rs8067196	49822630	C/A	2.26E-06	1.6	0.1033	0.06838	*TOM1L1*
8	rs2930553	86918262	G/T	2.48E-06	0.7694	0.4609	0.5254	*REXO1L1*
3	rs13079811	82697593	C/A	2.69E-06	0.7572	0.3082	0.3714	*GBE1*
17	rs8066010	49808584	A/G	2.77E-06	1.594	0.1033	0.06864	*TOM1L1*
1	rs606149	192188171	T/C	5.05E-06	1.288	0.4952	0.4307	*LOC647167*
1	rs10921464	192156156	T/C	6.68E-06	0.7767	0.3901	0.4551	*LOC647167*
3	rs7433760	187284322	G/A	8.12E-06	0.7606	0.2695	0.3262	*ETV5*
20	rs2423618	11757442	C/T	1.59E-05	1.295	0.3455	0.2889	*LOC728450*
5	rs2359879	76841025	A/C	2.12E-05	1.511	0.1076	0.0756	*WDR41*
3	rs2731943	21355681	T/C	2.63E-05	1.307	T/C	2.63E-05	*VENTXP7*
17	rs1547966	51779870	C/T	2.70E-05	1.301	C/T	2.70E-05	*ANKFN1*
6	rs10806179	80584274	C/A	2.75E-05	0.7659	C/A	2.75E-05	*ELOVL4*
16	rs4257224	7335293	G/T	2.76E-05	1.46	G/T	2.76E-05	*A2BP1*
10	rs10763492	58952986	G/A	2.86E-05	0.7456	G/A	2.86E-05	*IPMK*
17	rs9944407	40017486	C/T	2.89E-05	0.7863	C/T	2.89E-05	*FZD2*
18	rs4891441	62222556	C/A	2.94E-05	0.7485	C/A	2.94E-05	*CDH19*
5	rs10512971	51897194	G/A	3.34E-05	0.76	G/A	3.34E-05	*PELO*