Genes Involved in Brain Development Influence Crying Habits – A Genome Wide Association Study

Chao Tian[1], Cory Y. McLean[1], Eric Y. Durand[1], Nicholas Eriksson[1], Joyce Y. Tung[1], David A. Hinds[1]
[1]23andMe, Inc., Mountain View, CA;

Introduction

Crying, for emotional reasons, is considered unique to humans. Crying habits vary greatly--some crying easily and others rarely. Thus far it is not clear why, though emotional stability has been shown to play a role and excessive emotionality appears to have a familial tendency. In this genome-wide association study, we searched for possible genetic associations with crying easily using a large sample of unrelated customers of 23andMe, Inc., with European ancestry.

Methods

We conducted genome-wide association studies (GWAs) on 'Cry easily' phenotype. All participants were drawn from the customer base of 23andMe, Inc., a personal genetics company. All participants were of primarily European ancestry, and we excluded relatives who were first cousins or closer. All analyses controlled for sex and five principal components of genetic ancestry. P-values were calculated using likelihood ratio test.

Phenotypic Data

Phenotype Statistics					Age Distribution			
Phenotype	Group	Total	M	F	(0,30]	(30,45]	(45,60]	(60,Inf]
Cry Easily	Case	72841	25835	47006	11644	21834	20500	18863
	Control	112368	71139	41229	16354	34677	29678	31659

Table 1 Our discovery cohorts

Our discovery cohorts are drawn from the more than 650,000 genotyped customers who reported via a web-based questionnaire whether they cry easily. Sex and ancestry were determined based on genetic data; for the GWAS, participants were of European ancestry.

Genotyping

Participants were genotyped for 946,181 SNPs across four versions of Illumina–based beadchips. An additional 12,817,048 imputed SNPs were included in the analysis. For imputation, we first used Beagle[1] (version 3.31) to phase batches of 8000-9000 individuals across chromosomal segments of no more than 10,000 genotyped SNPs, with overlaps of 200 SNPs. We excluded SNPs with Hardy-Weinberg equilibrium $P<10^{-20}$, call rate < 95%, or with large allele frequency discrepancies compared to European 1000 Genomes reference data. We imputed each phased segment against all-ethnicity 1000 Genomes haplotypes (excluding monomorphic and singleton sites) using a high-performance version of Minimac[2], using 5 rounds and 200 states for parameter estimation.

Genetic Association Tests

We performed logistic regression assuming an additive model for allelic effects, using the model:
Cry_easily ~ age + sex + pc.0 + pc.1 + pc.2 + pc.3 + pc.4 + genotype
The results are adjusted for a genomic control inflation factor lambda=1.143.

Results

72841 cases self reported that they cry easily and 112368 controls said that they did not cry easily. Females are three times more likely to self-identify as crying easily than males (OR=3.14, $P < 2.2\text{x}10^{-16}$).
We identified genome-wide significant associations with 12 loci (Figure 1), including rs62335062 ($P=2.4\text{x}10^{-36}$, OR=1.107) upstream of *IRX2*, rs7196282 ($P=8.3\text{x}10^{-16}$, OR=0.931) in an intron of *ZNF423*, rs876714 ($P=1.3\text{x}10^{-13}$, OR=1.058) in an intron of *BIN3* and close to *EGR3*, rs10838125 ($P=5.8\text{x}10^{-13}$, OR=1.059) between *TTC17* and *HSD17B12*, rs2206271 ($P=2.0\text{x}10^{-12}$, OR=0.944) upstream of *TFAP2B*, rs62170343 ($P=8.7\text{x}10^{-12}$, OR=1.06) in an intron of *LRRTM4*, rs4961492 ($P=1.8\text{x}10^{-10}$, OR=1.052) in the intron of *BNC2*, rs62509237 ($P=2.1\text{x}10^{-10}$, OR=1.054) upstream of *LRP12*, rs17133429 ($P=5.810\text{x}^{-10}$, OR=1.096) in an intron of *SDK1*, rs5869436 ($P=2.6\text{x}10^{-9}$, OR=0.955) in *microRNA9-2* (*MIR9-2*), rs4365937 ($P=2.9\text{x}10^{-8}$, OR=1.047) near *MIR2113* and rs4655870 ($P=4.8\text{x}10^{-8}$, OR=1.042).

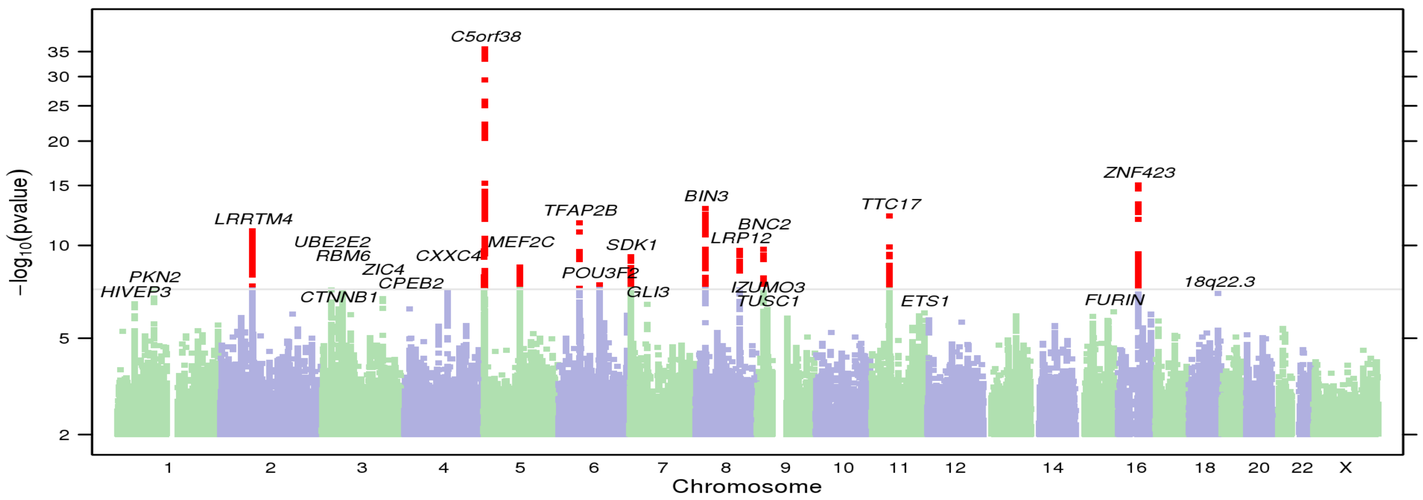

Figure 1: Manhattan Plot of association with 'cry_easily'. –log10 P-values across all SNPs tested. SNPs shown in red are genome-wide significant ($P< 5\text{x}10^{-8}$). Regions are named with the postulated candidate gene.

Cytoband	Name	Scaffold	Position	Alleles	Pvalue	OR	95% CI	Gene.context
5p15.33	rs62335062	chr5	2793679	A/G	$2.4\text{x}10^{-36}$	1.107	[1.090,1.125]	C5orf38--[]---IRX1
16q12.1	rs7196282	chr16	49728195	G/T	$8.3\text{x}10^{-16}$	0.931	[0.915,0.947]	[ZNF423]
8p21.3	rs876714	chr8	22482799	A/G	$1.3\text{x}10^{-13}$	1.058	[1.042,1.074]	[BIN3]
11p11.2	rs10838125	chr11	43565476	C/T	$5.8\text{x}10^{-13}$	1.059	[1.043,1.076]	TTC17--[]---HSD17B12
6p12.3	rs2206271	chr6	50786008	A/T	$2.0\text{x}10^{-12}$	0.944	[0.929,0.959]	TFAP2D--[]TFAP2B
2p12	rs62170343	chr2	77186361	A/G	$8.7\text{x}10^{-12}$	1.06	[1.043,1.078]	[LRRTM4]
9p22.3	rs4961492	chr9	16520958	A/G	$1.8\text{x}10^{-10}$	1.052	[1.036,1.069]	[BNC2]
8q22.3	rs62509237	chr8	105826782	C/T	$2.1\text{x}10^{-10}$	1.054	[1.037,1.071]	LRP12---[]---ZFPM2
7p22.2	rs17133429	chr7	3530038	C/T	$5.8\text{x}10^{-10}$	1.096	[1.065,1.128]	[SDK1]
5q14.3	rs5869436	chr5	87992715	D/I	$2.6\text{x}10^{-9}$	0.955	[0.941,0.970]	TMEM161B---[]--MEF2C
6q16.1	rs4365937	chr6	98595522	C/T	$2.9\text{x}10^{-8}$	1.047	[1.030,1.064]	MMS22L---[]---POU3F2
1p22.2	rs4655870	chr1	88684629	C/G	$4.8\text{x}10^{-8}$	1.042	[1.027,1.058]	LMO4---[]---PKN2

Table 2. Index SNPs for regions under $P = 5 \times 10^{-8}$. The index SNP is defined as the SNP with the smallest P-value within a region. The listed gene is our postulated candidate gene near the SNP.

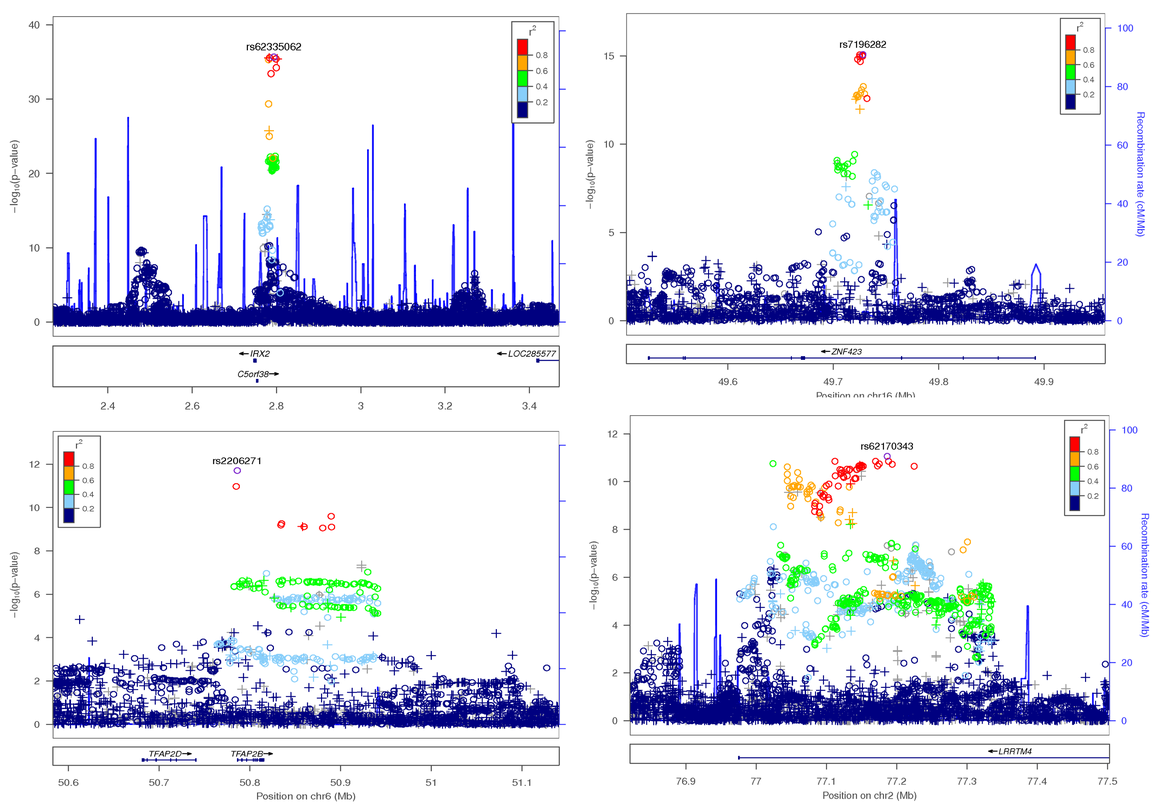

Figure 2. Associations with 'cry_easily' in four regions with genome-wide significant SNPs. Colors depict the squared correlation (r^2) of each SNP with the most associated SNP (which is shown in purple). Gray indicates SNPs for which r^2 information was missing.

Discussion

The top two associated genes, *IRX* and *ZNF423*, have been shown to be involved in cerebellum development and brain regionalization. The cerebellum is an important brain region that has been linked to crying. *LRRTM4* gene product is known to trigger the formation of excitatory synapses and shows highly selective expression in the brain. Mutations in *LRRTM* family genes have been associated with human handedness and schizophrenia. *TFAP2B* product, an important factor in the development of ectodermal and neural tissues, is involved in monoaminergic regulation and has been associated with neonatal temperament, alcohol addiction, adolescent depressive symptoms, and attention deficit hyperactivity disorder. The expression of *EGR3* is rapidly regulated by neural synaptic activity in the brain cortex, which may be important in defining neuroplastic responses following stimulus. *MIR9-2* has been associated with neuron-specific expression and neuronal differentiation during brain development. The finding in the 6q16.1 region was previously shown to be a genome-wide significant association with bipolar. *POU3F2* has been found to contribute to neocortex development in mice. This study highlights some important genes that potentially impact the neural basis of crying in response to emotional stimulus, which may also shed light on the disease etiology of more serious affective disorders.

Acknowledgments

We thank the customers of 23andMe for participating in this research and the employees of 23andMe for their contributions to this work.

References

1. Browning, S.R. & Browning, B.L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
2. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955-959 (2012).