

Mapping Genes Related to Early Onset Major Depressive Disorder in Dagestan Genetic Isolates



Kazima BULAYEVA¹, Todd LENCZ², Stephen GLATT³, Toru TAKUMI⁴, Farida GURGENOVA⁵, Hideshi KAWAKAMI⁶, Oleg BULAYEV⁷

SUMMARY

Aim: The purpose of this study was to determine the molecular epidemiology of early onset major depressive disorder (MDD) in genetic isolates of the Caucasus Dagestan indigenous ethnic populations using molecular and statistical population-genetic approaches.

Methods: Two multigenerational pedigrees from two diverse remote highland isolates with aggregation of early onset MDD were ascertained within our long-term research program titled 'Dagestan Genetic Heritage, DGH'. The first isolate included 48 cases of MDD (19 living) with 11 suicides committed, and the second included 60 MDD cases (30 living) with 12 suicides committed. The phenotypes of the affected family members were determined using a database containing diagnoses from a regional psychiatric hospital and through our own clinical examinations, which were based on a Russian translation of DIGS software based on the DSM-IV criteria. A 10 cM genomic scan (Weber/CHLC 9.0 STRs) of the 64 affected and non-affected members of the pedigrees was performed and the data was used for multipoint parametric linkage analyses. Following this scan, selected cases were analyzed by Affymetrix 6.0 SNP arrays in order to refine the contribution of copy number variations (CNVs) to the genetic basis of MDD.

Results: We found a total of 18 genomic regions with nominal ($\text{LOD} > 1.3$) linkage to MDD across the two isolates. Three genomic regions had genome-wide significant ($\text{LOD} > 3$) linkages and were found at 2p13.2-p11.2, 14q31.12-q32.13 and 22q12.3. We also confirmed previous findings for MDD at 4q25, 11p15, 12q23-24, 13q31-32, 18q21-22 and 22q11-13. Six linkage regions were observed in both genetic isolates, while 12 other linkages demonstrated population-specific heterogeneity. We detected CNV rearrangements within 12 of the 18 linkage regions. Affected subjects had the highest rate of genomic instability within the linkage regions at 2p13.2-p11.2, 4q25-q28.2, 7p14.1, 8p23, 14q31.12-q32.13, 18q22.1 and 20p13.

Conclusion: The results obtained in this study suggest that mapping genes of complex diseases, including MDD, across genetically homogeneous isolates can help detect linkage signals and expedite the search for susceptibility genes when combined with methods that detect structural genomic variation in linkage regions.

Key words: Major depressive disorder, suicide, genetic linkage analysis, SNPs, DNA copy number variants

INTRODUCTION

Major depressive disorder (MDD) is a complex psychiatric disorder characterized by persistent sadness, extreme fatigue, feelings of guilt and hopelessness, appetite change, insomnia and suicidal thoughts (DSM-IV). The estimated risk of a diagnosis of major depression in suicide cases is almost 11 times higher

than in controls (Kim et al 2007). Although the heritability of MDD is lower than other psychiatric diseases, twin studies demonstrate that there is a significant genetic contribution of approximately 35-40% (Glowinski et al 2003; Sullivan et al 2000). Additionally, the morbid risk of MDD is consistently higher among the first-degree relatives of major depressive probands (~15 to 25%) than the first-degree relatives of unaf-

Received: 25.02.2011 - **Accepted:** 08.07.2011

Acknowledgement: The authors thank Dr. J. Weber and NHLBI Mammalian Genotyping Service for high-quality and efficient STR genotyping of the Dagestan genetic isolate pedigree samples (NIH/NHLBI Grant, 1998, Drs. H. Coon and K. Bulayeva) and Prof Takumi and Prof Kawakami (Hiroshima University, Japan) who genotyped AFFX SNP 6 in selected MDD cases.

¹Prof., ²MSc., Res. Assist. ³PhD, Senior Scientist, Dept. of Population Genetics, N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia, ⁴Prof., Division of Psychiatry Research, The Zucker Hillside Hospital, ⁵Prof., Department of Psychiatry and Behavioral Sciences, and Medical Genetics Research Center; SUNY Upstate Medical University; New York, USA, ⁶Prof., Laboratory of Integrative Bioscience Graduate School of Biomedical Sciences, ⁷Prof., Department of Epidemiology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan.

Kazima Bulayeva MD, e-mail: bulaeva@vigg.ru

affected controls (~5%). As with other forms of mental illness, the mode of genetic transmission is complex and poorly understood, and no susceptibility genes have been unequivocally identified to date (Kato 2007). Genome wide linkage studies have reported various peaks across multiple chromosomes, including 1q21-42, 4p16, 10q21-26, 11p15, 12q23-24, 13q11-32, 15q25-q26, 18p11, 18q21-22, 22q11-13, Xp11, and Xq24-28 (Abkevich et al 2003; Holmans et al 2004; Kendler et al 2001; Kia-Keating et al 2007; Zubenko et al 2002); however, most of these results have not been replicated (Levinson 2006). Notably, prior linkage and association studies have primarily focused on heterogeneous populations. Genetically homogeneous isolates are exceptional resources for the detection of susceptibility genes in complex human diseases because they contain less genetic and clinical heterogeneity.

This article expands our cross-isolates approach to mapping genes of complex diseases by utilizing a genome-wide multipoint linkage analysis in a clinical-epidemiological cohort of extended pedigrees who we previously reported on and who have high rates of psychiatric illness (Bulayeva et al 2002; Bulayeva et al 2007). In this report, we focus on two multigenerational pedigrees derived from two remote highland isolates that contain an aggregation of MDD and suicides. We recruited members of these pedigrees via the Dagestan Genetic Heritage research programme founded by K.B. Bulayeva in VIGG RAS in 1976. In addition to genome wide linkage analysis using short tandem repeat (STR) markers across both pedigrees, we also examined single nucleotide polymorphism (SNP) for copy number variation (CNV) using Affymetrix SNP 6.0 microarrays in selected affected cases from our isolates. CNVs and copy number polymorphisms (CNP) are novel tools that are important when searching for the susceptibility genes of complex diseases. CNV is a common type of genomic variability, with altered dosage of DNA fragments ranging in size 1 Kb to several Mb. CNVs (duplications or deletions) can influence gene expression by disrupting coding sequences or altering gene dosage in ways that can contribute to phenotypic variations or pathogenesis. Our goal was to collect preliminary data as to whether CNVs, which have been associated with other forms of psychiatric illness (Glessner et al 2009; Stefansson H et al 2008; Walsh et al 2008), could be detected in putative linkage regions for MDD. We report here the deletions and duplications found within the linked genomic regions associated with MDD.

MATERIALS AND METHODS

Description of the populations: The Caucasus region between the Black and Caspian Seas is characterized by extreme cultural and linguistic differentiation in a small geographic area that suggests a complex history. Twenty-six indigenous

Dagestan ethnic groups belong to four linguistic families: Dagestan-Nakh, South Caucasian, Indo-European, and Altaic. Our pioneering genetic-epidemiological study of the Dagestan ethnic populations, initiated in 1976, has demonstrated that there is the possibility of descent from a common ancestor population that existed several hundred generations ago (Bulayeva et al 2003). Subsequently, this common ancestral population has been fragmented into endogamous communities.

During annual expeditions among the Dagestan indigenous ethnics, we found several remote isolates aggregating certain complex diseases such as schizophrenia, mental retardation, depression, arterial hypertension, cancer etc., in communities where the neighboring isolates have no particular aggregation of any disease. In this report, we focus on two remote, highland villages of ethnic Laks (#6007 and #6008) with high aggregations of depression and 11-12 officially documented, completed suicides in each.

Clinical Assessment and Diagnosis: Most affected individuals from both isolates were previously diagnosed at Dagestan mental hospitals with affective or schizoaffective disorders. We re-diagnosed all affected probands using a structured psychiatric interview, the Diagnostic Interview for Genetic Studies (DIGS), based on DSM-IV diagnostic criteria (Nurnberger et al 1994). We had previously translated the DIGS into Russian for our research and adapted it, in both computerized and paper forms, to our multi-ethnic study. Diagnostic assessments were conducted by two Dagestan psychiatrists (R. Kurbanov and U. Guseinova) who had been trained in the DIGS and who were experienced with clinical evaluations from their long-term work at regional psychiatric hospitals. Individuals were placed in the category of 'probably' affected if they had no previous diagnoses at psychiatric hospitals but met the DSM-IV criteria for MDD based on our clinical interview. We also documented all additional diagnoses, including age at onset and severity in order to characterize any significant co-morbid conditions of the patients. The affected subjects who met the DSM-IV axis I diagnostic criteria for MDD were included as probands in our linkage study, whereas several probands with schizoaffective disorder symptoms were excluded from the analysis.

As part of the clinical analysis, we also interviewed three to four unaffected members of every family within each ascertained kindred using the Family Interview for Genetic Studies (FIGS), with the goal of obtaining corroborating diagnostic information and to detect MDD symptoms in unavailable pedigree members. Additional surveys were used for collecting demographic data about marriage structure, reproductive abilities, life span and morbidity of isolate and kindred members.

Written informed consent was obtained from each participant prior to clinical interviews and blood sample collections. This

study was approved by the Dagestan Center of the Russian Academy of Sciences IRB.

Pedigrees analysis: Using information about the family history, we reconstructed pedigrees for every affected and probably-affected MDD subject in both isolates. The final pedigrees reconstructed in the isolates contained 241 individuals in #6007 and 364 members in #6008 across 12-13 generations. It was possible to construct such deep, multi-generational extended pedigrees because these ethnic groups have a tradition in which fathers are responsible for transmitting information to younger generations about their direct ancestors extending back seven generations (Bulayeva et al 2005). Pedigree information was corroborated and verified by interviews taken from different family members within each isolate. Progeny (Desktop Version) software (Progeny Software LLC) was used to draw pedigree and manage data.

Genotyping: Blood samples were collected from three or four current generations of affected and healthy pedigree members. Genomic DNA was extracted from peripheral-blood leukocytes using standard protocols. Using standard methods, approximately 400 µg of DNA was isolated from each subject's blood sample. This DNA was sent to the high-throughput genotyping facilities at the Mammalian Genotyping Service of the Marshfield Medical Research Foundation. The Weber/CHLC map contained approximately 400 STR (mostly tri- and tetranucleotide repeat) markers, covering the genome at a 10 cM resolution.

As a pilot/exploratory analysis, a genome wide SNP genotyping (Affymetrix 6.0 microarrays) whole-genome scan was performed in eleven unrelated or distantly related affected MDD subjects from our isolates at labs in Hiroshima University (Profs. Takumi and Kawakami Labs). All selected cases were affected and all cases had a number of affected first- and second-degree genetic relatives. Two of the 11 samples failed Quality Control (QC) parameters and were excluded from our analysis. The average call rate of the remaining samples was $\geq 99\%$.

After obtaining linkage results for MDD and analyzing candidate genes, CNVs were identified within the linkage regions.

Statistical analysis: For the linkage analyses, we designated pedigree members with MDD as 'affected' and pedigree members who had no mental illness according to all available information were considered 'unaffected'. Other individuals, including pedigree members with unclear clinical symptoms, were considered 'unknown'. For the linkage study, we used Simwalk 2, based on the Markov chain/Monte Carlo (MCMC) algorithm (Sobel and Lange 1996), which is able to analyze large pedigrees because it considers the underlying configurations in proportion to their likelihood. Using multipoint parametric linkage analyses, we tested both dominant or recessive 90% penetrance model of inheritance, a disease allele frequency of 0.02, and an assumption of genetic

heterogeneity. Detailed description of the use of Simwalk2 for linkage analysis in Dagestan genetic isolates is described in our previous publications (Bulayeva et al 2005,2007).

CNVs were calculated with the Golden Helix SVS and Affymetrix GTC 4.0 software, which implements a Hidden Markov Model (HMM)-based algorithm to identify chromosomal gains and losses by comparing the signal intensity of each SNP probe set against a reference set. As described in SVS and GTC manuals, a CNV was identified when there were at least three consecutive SNPs that showed consistent deletion or duplication. A 5-state Hidden Markov Model (HMM) was used for smoothing and segmenting the CN data. The presence of a genome-wide technical artifact, called a spatial autocorrelation or 'wave', which occurs in a large dataset used to determine the location of CNV across the genome, may increase the number of false positive CNs detected. We used a method that removed this artifact so that a more biologically meaningful clustering of the data could be obtained and an increase in the number of CNVs could be identified (Marioni et al 2007). Because it is not possible to delineate the exact boundaries of each CNV with genome-wide SNP-genotyping arrays, we used SNP positions to determine approximate CNV boundaries.

RESULTS

A total of 64 individuals from two genetic isolates of ethnic Laks (#6007 & #6008) with predominant aggregation of early onset MDD were enrolled in this preliminary study. The demographic and clinical characteristics of these two isolates in comparison with other genetic isolates with an aggregation of schizophrenia and bipolar disorder as identified by our long-term study are presented in Table 1. The total number of living MDD, bipolar and schizophrenia-spectrum cases ranged from 19 to 20 in the isolates (Table 1). Many of the Dagestan neighboring highland villages of the same ethnic background and population size do not have any officially diagnosed cases of MDD, BP or schizophrenia, suggesting a possible founder effect. All affected MDD subjects within an isolate were found in one large, extended pedigree with a limited number of common ancestors.

Isolate #6007 was smaller in terms of the total number (Nt=320) of residents in comparison with isolate #6008 (Nt=900). The pedigrees reconstructed for affected cases in each isolate also differed in size: isolate #6007 contained 241 members across 11 generations (with DNA available from 15 cases out of the total of 26 in the pedigree), while #6008 contained 327 members across 13 generations (with DNA available from 23 cases out of the total 38). The sex ratio in both pedigrees was similar and was close to 1:1. However, we found that the number of suicides committed in these isolates was 2.5-2.6 times higher in females than in males. The

Table 1. Description of Dagestan genetic isolates with aggregation of psychiatric diseases

Isolate ID	Ethnics	Nt	AFST*	Total # of affected alive	No of observed (affected)	No of pedigree members	No of suicides committed (female/male)†
6007	Laks	320	MDD	19	26(15)	241	11(8/3)
6008	Laks	900	MDD	30	38 (25)	327	12(8/4)
6010	Laks	440	SCZ	17	24(10)	214	0
6034	Tindals	1800	SCZ	42	86(39)	450	3 (2/1)
6022	Dargins	1340	SCZ	35	42(20)	310	1 (0/1)
6017	Dargins	700	SCZ	25	30(17)	257	3 (1/2)
6021	Kumyks	2000	SCZ	31	41(25)	342	2(1/1)
6009	Avars	1300	BP	32	42(23)	324	1(0/1)
6015	Kubachians	3000	SCZ	11	18(6)	157	0

Nt= total number of residents in the isolates;

* The mood affect with the highest prevalence in the isolate

† No of suicides committed in the pedigrees reconstructed in the last 3-4 generations

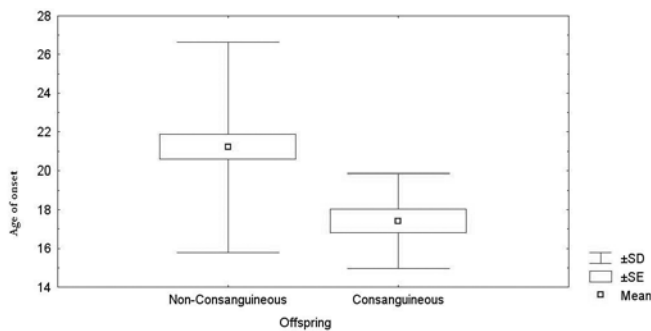


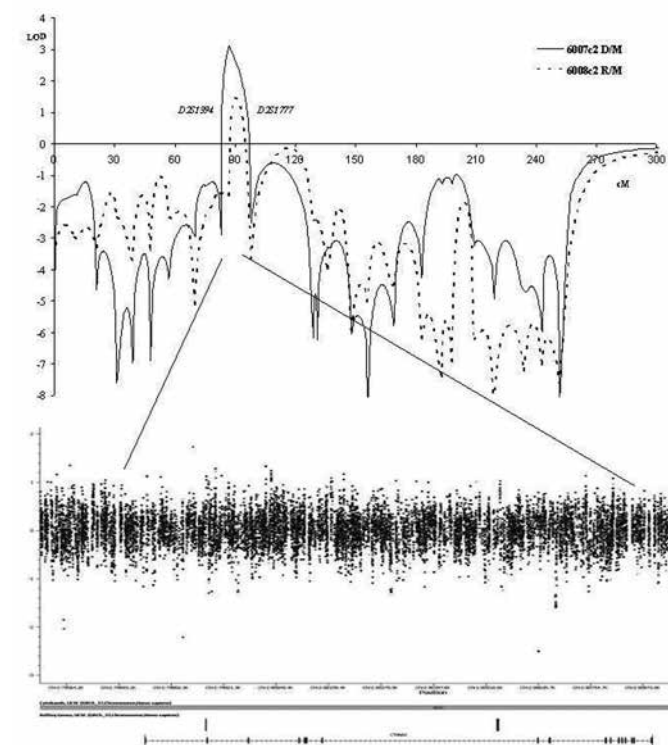
Figure 1. Age of onset for MDD among offspring of consanguineous and non-consanguineous marriages.

Mean=mean age of onset; SD =standard deviation; SE =standard error.

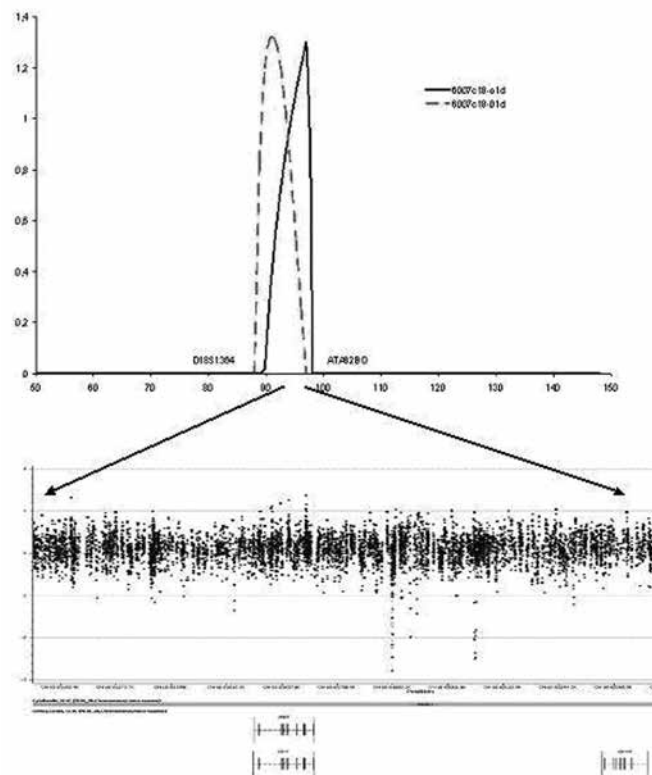
number of committed suicides, which were determined during DIGS/FIGS interviews of the probands' family members and by examining the family medical history documented at regional hospitals, is much higher (11-12 committed suicides) than in any other genetic isolates with aggregation of schizophrenia and bipolar disorder (0-3 suicides) (Table 1). Importantly, no suicide cases were reported among the elderly in the isolates. Most suicides were committed by young family members who were 16-22 years of age, and only three cases were committed between 27-35 years of age. Both genetic isolates and pedigrees had a high rate (85-90%) of traditional endogamy (within-villages marriages) and consanguineous marriages. The mean values of the inbreeding coefficients calculated in the pedigrees were 0.0375 for #6007 and 0.0348 for #6008. The age of onset for MDD was significantly earlier ($t=2.65$, $df=78$, $p=0.009$) among offspring of consanguineous marriages (17.4 ± 0.63) than among offspring of non-consanguineous marriages, i.e., without known consanguineous marriage within reconstructed generations (21.2 ± 0.67) (Figure 1). Although these two isolates belong to the same ethnicity (Laks), and were geographically close (~15km) to each other in the highland area, we found only two inter-village marriages across 11-13 generations.

Table 2 presents the results of the genome-wide multipoint linkage scan. All loci with LOD scores ≥ 1.2 (nominal $p < .05$) in either pedigree, under either the dominant model (D/M) or recessive model (R/M) are presented along with the flanking markers and peak location in cM. Most of the identified linkage peaks are isolate-specific, reflective of genetic heterogeneity. However, six linkage regions (1p36.1-p35.2, 2p13.2-p11.2, 13q31.1-q32.1, 17q25.3, 18q22.1 and 22q12.3) with LODs=1.2-3.4 demonstrated homogeneity in both pedigrees, although some inter-isolate differences observed at these loci were also noted in the genetic model (dominant or recessive) (Table 2). As seen in the Table 2, we identified 11 nominally significant linkage signals in #6007 and 13 such signals in #6008. The strongest genetic homogeneity, both in the location of the linked genomic region and the mode of transmission, was obtained at 22q12.3, with significant linkage (LOD=3.44) observed in pedigree #6007, and strongly suggestive linkage (LOD=2.80) in #6008 (Table 2). Although the strongest results in this region were obtained under the dominant model of inheritance, it should be noted that 5-6 affected subjects demonstrated extended homozygotic STR genotypes across the linkage region. In addition to the region at 22q12.3, two genome-wide significant (LOD>3) linkage peaks were observed in #6007: one at 14q31.12-q32.13 (LOD=3.417, recessive model), and another at 2p13.2-p11.2 (LOD=3.103, dominant model) (Table 2). No genome wide significant LOD scores ≥ 3 we observed in #6008. Suggestive level (LOD ≥ 2) linkages were found at 9q33.3-q34.2 (#6008, R/M), 13q31.1-q31.2 (#6007, R/M), 11p15 (#6008, R/M), 17q25.3 (#6007, R/M), 19q13.31-q13.33 (#6008, D/M) and 22q12.3 (#6008, D/M).

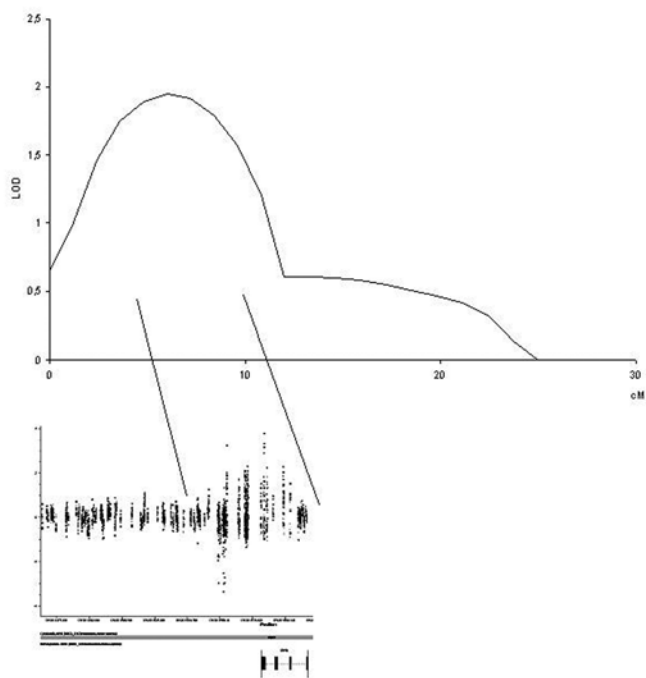
After a specific chromosomal region has been identified by linkage analysis, additional work is required to refine the location of the candidate susceptibility gene(s). We used exploratory examination of CNV, based on Affymetrix 6.0 microarray data, to detect structural variations within the MDD linkage regions that may be involved in the disease pathogen-



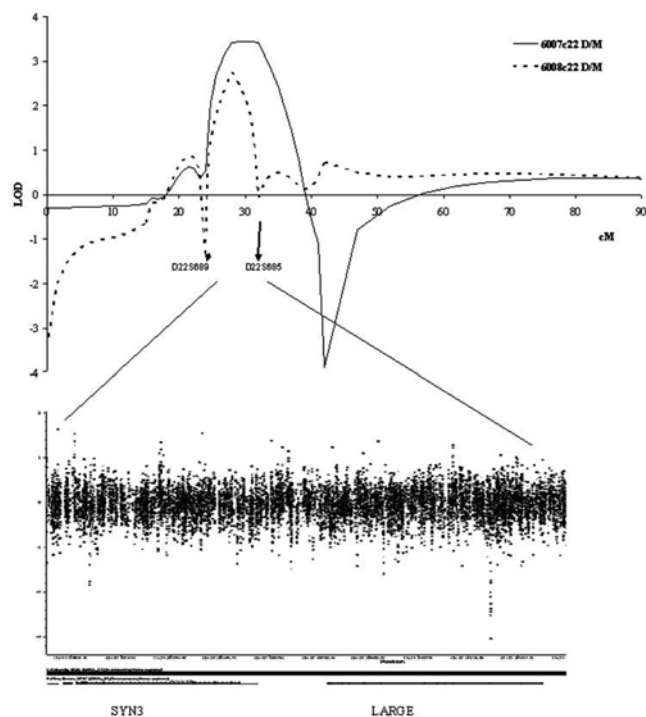
A



B



C



D

Figure 2. Linkage scan in Chromosomes 2 (A), 18 (B), 20(C) and 22 (D) for MDD in pedigrees coming from two Dagestan genetic isolates and the CNVs found in the linkage regions. CNVs detected:

- A- In genes CTNNA2 and LRRTM1 located within the linked 2p13.2-p11.2 region;
- B- In genes CDH7 and CDH19 (and intergenic) located within the linked 18q22.1 region.
- C- In gene SIRPB1 located within the linked 20p13 region
- D- In genes SYN3 and LARGE located within the linked 22q12.3 region

Table 2. Genome –Wide Multipoint Linkage Scan for MDD in Two Dagestan Genetic Isolates with a High Prevalence of Early Onset Major Depressive Disorder

#6007			#6008		
MAP	LOD, R/M-D/M	Flanking loci, Peak/cM	MAP	LOD, R/M-D/M	Flanking loci, Peak/cM
1p36.1-p35.2	1.3, D/M	D1S552-D1S1622, 52.4	1p36.21	1.43, R/M	D1S1356-D1S1352, 62.2
2p13.2-p11.2	3.103, D/M	D2S1394-D2S1777, 87	2p12-p11.2	1.4, R/M	D2S1777-D2S1790, 90
4q25-q28.2	1.873, R/M	D4S2623-D4S2394, 114			
			5q14.1-q14.3	1.9, D/M	D5S1501-D5S1725, 94.5
			7p12.3-p14.1	1.64, R/M	D7S2846-D7S1818, 60.2
			8p23.1-p23.3	1.56, R/M	D8S264-D8S277, 3
9q21.33-q22.33	1.2, D/M	D9S257-D9S910, 84.8	9q33.3-q34.2	2.1, R/M	D9S1825-D9S157, 129.5
10p13-p12.33	1.7, R/M	D10S1430-D10S1423, 33.5			
			11p15.4-p15.5	2.1, R/M	D11S1984-D11S2362, 3
			12q24.1-q24.3	1.56, D/M	D12S2078-D12S1045, 149.4
13q31.1-q32.1	2.31, R/M	D13S317-D13S793, 53	13q31.1-q32.1	1.3, D/M	D13S317-D13S793, 51
14q31.12-q32.13	3.417, R/M	D14S617-D14S1434, 101			
17q25.3	2.48, R/M	D17S784-D17S928, 123	17q25.3	1.3, D/M	D17S784-D17S928, 123
18q22.1	1.3 D/M	D18S1364- ATA82BO2, 97	18q22.1	1.32, D/M	D18S1364- ATA82BO2, 90.7
			19q13.31-q13.33	2.7, D/M	D19S178-D19S246, 66.9
20p13	1.95, D/M	D20S103-D20S482, 5			
22q12.3	3.44, D/M	D22S685-D22S683, 29.9	22q12.3	2.80, D/M	D22S685-D22S683, 28

D/M, R/M – dominant or recessive model of pathogenic locus transmission

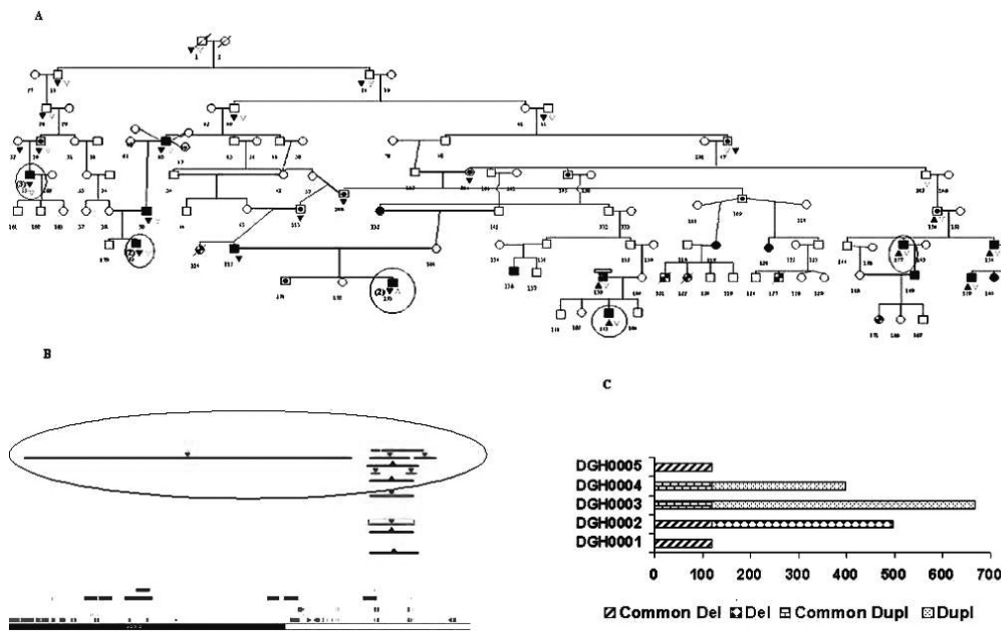
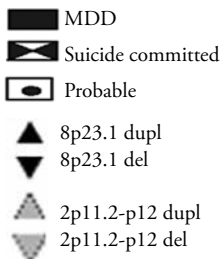


Figure 3. Recurrent CNVs found in 5 MDD cases within linkage regions at 8p23 and 2p12.

A. MDD pedigree branch from isolate 6008 with 5 distantly related cases observed (marked by circle) with common ancestors.



B. Segmental deletions (3 cases) and duplications (2 cases) at 8p23.1-p23.2 observed among pedigree members. Probable mutation carriers among ancestors up to the common ancestor are marked by points.

C. There are individual differences between the 5 affected MDD cases in the sizes (kb) of CNV, however, the genomic segment (120 kb) is common in all cases and was probably received from a common ancestor.

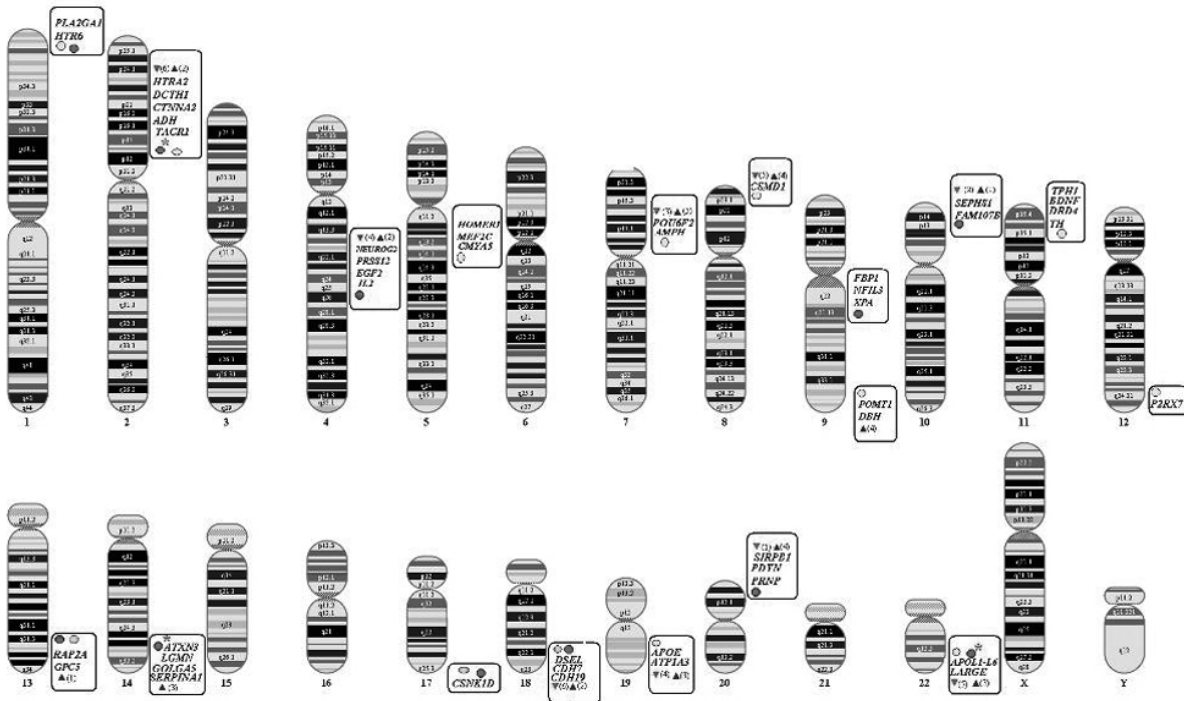


Figure 4. Karyoview-summarized results of a genome-wide linkage scan in two genetic isolate pedigrees with an aggregation of MDD, and the CNVs found in the linkage regions with the probable candidate genes identified (number of affected cases with CNVs).

Pink stars indicate regions with LODs ≥ 3 .

- ▲ Gain
- ▼ Loss
- Linkage analysis results in isolate 6007;
- Linkage analysis results in isolate 6008.

esis. In total, genome-wide scans of nine MDD cases were generated with 133 CN segments (116 in autosomes and 15 in X and 2 in Y chromosomes) with sizes of greater than 100bp, containing 5 or more markers. Of these 133 detected segments, 87 were gains (9 homozygotic) and 46 were losses (5 homozygotic). The CN segment sizes varied from 101 to 4040 kb, with a mean size of 308 kb, a median size of 178 kb and a SD of 416.2 kb. The results obtained in this study indicated that there are CNV structural variations in 12 of the 18 regions linked with MDD. The highest rate of genomic instability was found within six of the regions linked with MDD: 2p13.2-p11.2 (6 losses and 2 gains), 4q25-q28.2 (4 losses and 2 gains), 7p14.1 (3 losses and 5 gains), 8p23.2-p23.1 (5 losses, 4 gains), 18q22.1 (6 losses and 2 gains) and 20p13 (1 loss and 4 gains). Linkage scan results for Chromosomes 2 (A), 18 (B), 20(C) and 22 (D), and the CNVs found in the linkage regions, are presented in Figure 2.

A large deleted segment (4040 kb) was seen in one affected case at 8p23.2 (Figure 3). The 8p23 linkage region contains 17 genes, including ARHGEF10, DLGAP2, CSMD1 and MCPH1, which were previously reported to be associated with ASD, schizophrenia, BP and mental retardation. The CNV analyses demonstrated a 'hot spot' of genomic

instability at 8p23.1-p23.2 and identified two more MDD candidate genes - CSMD1 and MCPH1. We analyzed the locations and sizes of the CN at the 8p23 linkage region in 5 of the 9 observed MDD cases and found that three cases had deletions and two had duplications in the same genomic region (Figure 3B). All these cases are from isolate 6008 and are distantly related to each other, with a common ancestor within 4-7 generations (Figure 3A). The branch of the pedigree contained 15 diagnosed MDD cases and 5 suicides. Our analysis showed that all 5 MDD cases had a common CNV segment at 8p23 that was 120 kb in size. The common segment can be explained in terms of identity by descent (IBD) (Fig 3A-C). In the three cases with deletions, the mutation was most likely derived from their common ancestors 1 or 2 (Fig 3A), whereas the common ancestor for the other two MDD cases with duplications is unclear (Fig 3A). CNVs were examined at 2p12 and were found to be located within the CTNNA2 gene in the same 5 MDD cases. The CNs detected in regions linked with MDD are summarized in Figure 4. In the cases examined in this study, CNs were not found in the 6 genomic regions linked with MDD: 1p35-p36, 5q14.1-q14.3, 9q21.33-q22.33, 11p15.4-p15.5, 12q24 and 17q25.3 (Figure 4).

DISCUSSION

In the long-term research conducted by the Dagestan Genetic Heritage programme among indigenous ethnics, it has been shown that prolonged reproductive isolation in severe highland environments during hundreds of generations favored high genetic diversity between isolated populations and low heterogeneity within them (Bulayeva et al 2003). Also, family constellations in these genetic isolates are large and multigenerational, with several generations living together due to the long lifespan of the highlanders and traditionally un-regulated reproductive practices. This family environment allows for identification of a sufficient number of affected cases and collection of quality clinical and pedigree data from these genetic isolates.

The genetic isolates described in this report, with an aggregation of MDD cases, showed a significant number of committed suicides (Table 1). In contrast to other data suggesting that suicidal behavior is a separate domain from psychosis and is connected with gender differences in socio-cultural roles (Oquendo et al 2001), the results of this study suggest that suicides in these isolates were mostly connected with MDD due to the fact that all MDD affected cases are close genetic relatives of the people who completed suicide. All the Dagestan remote highland villages have had the same (insufficient) medical services, as well as very similar socio-cultural traditions and socioeconomic circumstances. This suggests that the high aggregation of MDD and suicides in these Dagestan pedigrees is not connected with particular social and economic conditions that are unique only to these two highly isolated cohorts.

In both isolates, we found that the most affected MDD subjects, as well subjects who committed suicides, were offspring of close consanguineous marriages. To identify genes involved in moderately heritable diseases such as MDD, it is useful to study a subset of pedigrees with a high rate of consanguineous marriages because these pedigrees have maximal genetic homogeneity which can enhance the genetic "signal:noise" ratio. This is an especially important strategy for complex disorders with many small or weak genetic signals. Previous linkage and association studies have shown some candidate gene locations for MDD, most of which have been replicated only in two or fewer studies (Abkevich et al 2003; Holmans et al 2004; Kendler et al 2001; Kia-Keating et al 2007; Levinson 2006). Most association studies have focused on neurotransmitter systems when attempting to identify candidate genes, including the serotonin transporter, serotonin receptors, dopamine receptors, tyrosine hydroxylase, MAO-A, COMT, and tryptophan hydroxylase (Kato 2007; Kia-Keating et al 2007; Levinson 2006). Although the role of a TPH2 mutation (12q21), as well as 5-HTTLPR (17q11.2) and BDNF (11p14.1) promoter polymorphisms have drawn attention, these associations have been found to be more complex than previously thought (Kato 2007; Levinson 2006). Meta-

analyses suggest that there are small positive associations between polymorphisms in the serotonin transporter promoter region (5-HTTLPR) and bipolar disorder, suicidal behavior, and depression-related personality traits but no association has been seen with MDD yet (Levinson 2006).

Our genome-wide linkage scan in these two genetic isolates replicated previous findings for MDD at 4q25, 11p15, 12q23-24, 13q31-32, 18q21-22 and 22q11-13 (Table 2), whereas other linkage signals found at the 1p36, 5q14, 7p, 9q, 10p and 17q25 regions had not been previously reported as linked with MDD but contain genes which have evidence for association with other psychiatric diseases (e.g., bipolar disorder, anxiety, schizoaffective disease, alcohol abuse, autism etc). We found two new genomic regions that were significantly linked with MDD in #6007 with LODs =3.1-3.4 at 2p13.2-p11.2 (a somewhat similar signal was seen in the same regions in cases from #6008) and in 14q31.12-q32.13 (Table 2).

In the 2p12 region we identified as being linked with MDD, other researchers have recently reported a linkage with suicidality in recurrent MDD with a LOD=1.2, $P<0.0087$ (at marker D2S428) (Butler et al 2010). Additionally, there is evidence of linkage to the BP phenotype and suicidality at 2p12 with a LOD score = 4.2 and a LOD=3.82 (Hesselbrock et al 2004; Willour et al 2007). In another study, no evidence of linkage to the MDD phenotype and suicide attempts was found at the 2p12 region (Cheng et al 2006; Zubenko et al 2004). In this study, we have found evidence of 2p12-p11 linkage with early onset recurrent MDD and suicidality among MDD cases in two diverse genetic isolates. There are several genes of interest located in this region, including DCTN1, HTRA2, CTNNA2 and in the flanking region, the RNF103 gene (Fig 2A). All of these genes have been associated with multiple sclerosis, Parkinson's, bipolar, ADHD or Alzheimer disease in at least one study (Anney et al 2008; Scott et al 2009). The results obtained from the CNV analysis in the 2p12 region indicated gains in two of the affected MDD subjects and losses in five of the affected cases within the CTNNA2 (alpha catenin) and LRRTM1 (leucine-rich repeat transmembrane neuronal protein 1) genes, and there is a segment (between SNPs rs7591785 and rs7572995) that is significantly enriched in MDD subjects (Fig 2A). It is known that members of the catenin family of proteins play a major role in the folding and lamination of the cerebral cortex. CTNNA2 functions are related to regulation of the stability of synaptic contacts across the central nervous system (Hirano and Takeichi 1994). It was also shown that the LRRTM1 gene is associated with left-handedness and that patients with autism, schizophrenia and other psychiatric disorders have a greater tendency to be left-handed (Francks et al 2007).

The linkage region at 14q31-32 contains the ATXN3 gene, which has been associated with schizophrenia, as well as the LGMN gene associated with multiple sclerosis, and the

GOLGA5 gene related with thyroid system pathology. Any of these genes may also be involved in the MDD region that had a high LOD result in this study (Table 2). Four of the nine MDD cases have losses and two have gains at chromosome 14q32.2 within the SERPINA1 (serpin peptidase inhibitor, clade A, member 1) gene, whereas other structural variations within this linked region were intergenic. The SERPINA1 gene is the template for the protein alpha-1 antitrypsin, which helps to control several types of chemical reactions by inhibiting certain enzymes. The role of the CNV found in this gene in MDD is currently unclear.

The highest LOD score that emerged from this linkage analysis, LOD=3.44, occurred in the smaller genetic isolate pedigree #6007 at genomic region 22q12.3 (peak at 30 cM) and a LOD=2.8 was found in an adjacent genomic region for the second isolate pedigree #6008 (peak at 28 cM) (Fig 2D). In this linkage region, there are 19 genes, including the APOL3-6, TOM1, SYN3, RBM9, and RASD2 genes, which have been previously reported to be associated with unipolar or bipolar affective disorders and with schizophrenia (Liu et al 2008; Potash et al 2008). Our results support that some of these genes may be associated with MDD. There were three MDD cases with duplications and five with deletions within the SYN3 and LARGE genes that were previously reported to be associated with unipolar and bipolar disorders and with schizophrenia (Liu et al 2008; Potash et al 2008) (Fig 2D).

Several other interesting genes are found in regions with possible MDD linkage. The linkage region at 1p35-p36 contains the serotonin receptor genes HTR6, HTR1D and other genes related to neurodevelopmental or neurodegenerative diseases. The linkage region at 4q25-q26 contains PRSS12, which has been connected with non-syndromic mental retardation autosomal recessive type 1 (MRT1) [MIM:249500], as well as the genes EGF2 and IL2 that were previously reported to be associated with schizophrenia (Bulayev et al 2009). Additionally, it has been previously reported that there is a suggestive level of linkage on 4q with bipolar disorder in Old Order Amish, which is consistent with our results (Pauls et al 1995). However, the two gains and four losses found in the 4q25-q26 linkage region are completely intergenic, overlapping no genes.

In total, we detected CNV segmental rearrangements in 12 of the 18 linked regions, with 29 losses and 25 gains. However, we also found smaller sized deletions and duplications within certain genes located in linkage regions. These recurrent genomic features may be connected with specific MDD susceptibility gene(s) variations transmitted to the diverse isolates from a common ancestral meta-population. While several CNVs within the MDD linkage regions were found to be common amongst affected subjects (Table 2, Figs 3, 4), some represent moderately common CNVs observed in healthy control cohorts according to the Database of Genomic Variation (DGV). On the other hand, some of the detected

CNVs are novel. For instance, five affected MDD cases had novel CNVs (three deletions and two duplications) within the POU6F2 gene that was earlier reported to be a candidate gene for autism spectrum disorders (Glessner et al 2009) and is located within the linkage region at 7p14 identified in our MDD study. Multiple rare coding variants were identified in the ABCA13 gene also located at 7p14. An earlier study of the same location had identified one nonsense and nine missense mutations and compound heterozygosity/homozygosity in six cases of schizophrenia, bipolar disorder, and depression (Knight et al 2009). Our study supports this previous finding and our data shows that there are microdeletions in the ABCA13 gene in 5 out of 9 MDD cases.

We also detected homozygous CNVs in three of the affected MDD cases and heterozygous CNVs in two of the affected cases, within the SIRPB1 gene in the 20p13 region linked with MDD (Fig 2C). This finding may have significance in the pathogenesis of MDD because in a recent study, it was found that SIRP-beta1 was up-regulated and acted as a phagocyte receptor on microglia in the amyloid precursor protein J20 (APP/J20) in transgenic mice and in Alzheimer's disease (AD) patients (Gaikwad et al 2009).

The largest deletion, spanning more than 4 MB was seen in the 8p23.2 region in one of the affected subjects (Fig 3A-C). Another eight subjects demonstrated segmental deletions and duplications ranging from 116-667 kb in size in the region flanking 8p23.1 that contains the ARHGEF10, CSMD1 and MCPH1 genes associated with schizophrenia, bipolar disorder and microcephaly. Currently, microcephalin (MCPH1) has been identified as a plausible candidate gene for autism. The CSMD1 gene (CUB and Sushi multiple domains) is a distinguishing factor between dependent versus nondependent individuals abusing methamphetamine, alcohol, nicotine and other substances (Uhl et al 2008). Recent research demonstrates that duplication of 8p23.1-8p23.2 was observed in a child with speech delay and a diagnosis of ICD-10 autism (Glancy et al 2009) as well as in his mother, who had epilepsy and learning problems. It was also shown that the distal breakpoint of the duplicated region interrupts the CSMD1 gene in 8p23.2 and that the medial breakpoint lies between the MSRA and RP1L1 genes.

We believe that combining genome-wide STRs linkage scans with deeper microarray based CNVs analyses in linkage regions can be an effective way to examine the role that rare, common, and novel CNVs play in the pathogenesis of complex diseases. Using pedigrees found in genetically homogeneous isolates that contain an aggregation of a particular complex disease increases the ability to identify disease susceptibility genes. However, the SNP-array-based CNV results presented here must be considered preliminary and merely suggestive, and further investigations of the entire cohort, including unaffected members of the pedigrees, are needed.

REFERENCES

- Abkevich V, Camp N, Hensel C, et al (2003): Predisposition Locus for Major Depression at Chromosome 12q22-12q23.2. *Am J Hum Genet* 73:1271-81.
- Anney RJ, Lasky-Su J, O'Dushlaine C, et al (2008): Conduct disorder and ADHD: evaluation of conduct problems as a categorical and quantitative trait in the international multicentre ADHD genetics study. *Am J Med Genet B Neuropsychiatr Genet* 147B:1369-78.
- Bulayev O, Spitzin V, Bulayeva K (2009): Population approach to mapping genes of complex diseases. *J MedGenetics (Russia)* 4:3-17.
- Bulayeva K, Jorde L, Ostler C, et al (2003): Genetics and population history of Caucasus populations. *Hum Biol* 75:837-53.
- Bulayeva K, Pavlova T, Kurbanov R, et al. (2002): [Mapping genes of complex diseases in genetic isolates of Dagestan] Kartirovanie genov kompleksnykh zabolovaniy v geneticheskikh izolyatakh Dagestana. *Genetika* 38:1539-48.
- Bulayeva KB, Glatt SJ, Bulayev OA, et al (2007): Genome-wide linkage scan of schizophrenia: a cross-isolate study. *Genomics* 89:167-77.
- Bulayeva KB, Leal SM, Pavlova TA, et al (2005): Mapping genes of complex psychiatric diseases in Daghestan genetic isolates. *Am J Med Genet* 132B:Pages 76 - 84.
- Butler AW, Breen G, Tozzi F, et al (2010): A genomewide linkage study on suicidality in major depressive disorder confirms evidence for linkage to 2p12. *Am J Med Genet B Neuropsychiatr Genet* 153B:1465-73.
- Cheng R, Juo SH, Loth JE, et al (2006): Genome-wide linkage scan in a large bipolar disorder sample from the National Institute of Mental Health genetics initiative suggests putative loci for bipolar disorder, psychosis, suicide, and panic disorder. *Mol Psychiatry* 11:252-60.
- Francks C, Maegawa S, Lauren J, et al (2007): LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Mol Psychiatry* 12:1129-1139, 1057.
- Gaikwad S, Larionov S, Wang Y, et al (2009): Signal regulatory protein-beta1: a microglial modulator of phagocytosis in Alzheimer's disease. *Am J Pathol* 175:2528-39.
- Glancy M, Barnicoat A, Vijeratnam R, et al (2009): Transmitted duplication of 8p23.1-8p23.2 associated with speech delay, autism and learning difficulties. *Eur J Hum Genet* 17:37-43.
- Glessner J, Wang K, Cai G, et al (2009): Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459:569-73.
- Glowinski A, Madden P, Bucholz K, et al. (2003): Genetic epidemiology of self-reported lifetime DSM-IV major depressive disorder in a population-based twin sample of female adolescents. *J Child Psychol Psychiatry* 44:988-96.
- Hesselbrock V, Dick D, Hesselbrock M, et al (2004): The search for genetic risk factors associated with suicidal behavior. *Alcohol Clin Exp Res* 28:70S-76S.
- Hirano S, Takeichi M (1994): Differential expression of alpha N-catenin and N-cadherin during early development of chicken embryos. *Int J Dev Biol* 38:379-84.
- Holmans P, Zubenko G, Crowe R, et al (2004): Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *Am J Hum Genet* 74:1154-1167.
- Kato T (2007): Molecular genetics of bipolar disorder and depression. *Psychiat Clin Neuros* 61:1-17.
- Kendler KS, Thornton LM, Gardner CO (2001): Genetic risk, number of previous depressive episodes, and stressful life events in predicting onset of major depression. *Am J Psychiat* 158:582-6.
- Kia-Keating B, Glatt S, Tsuang M (2007): Meta-Analyses Suggest Association Between COMT, but Not HTR1B, Alleles, and Suicidal Behavior. *Am J Med Genet B* 144:1048-53.
- Kim S, Choi K, Baykiz A, et al. (2007): Suicide candidate genes associated with bipolar disorder and schizophrenia: an exploratory gene expression profiling analysis of post-mortem prefrontal cortex. *BMC Genomics* 12:413.
- Knight HM, Pickard BS, Maclean A, et al (2009): A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. *Am J Hum Genet* 85:833-46.
- Levinson D (2006): The genetics of depression. *Biol Psychiat* 60:84-92.
- Liu Y, Fann C, Liu C, et al (2008): RASD2, MYH9, and CACNG2 genes at chromosome 22q12 associated with the subgroup of schizophrenia with non-deficit in sustained attention and executive function. *Blol Psychiat* 64:789-96.
- Marioni JC, Thorne NP, Valsesia A, et al (2007): Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol* 8:R228.
- Nurnberger J, Blehar M, Kaufman C, et al (1994): Diagnostic interview for genetic studies: rationale, unique features and training. *Arch Gen Psychiatry* 51:849-62.
- Oquendo MA, Ellis SE, Greenwald BS, et al. (2001): Ethnic and Sex Differences in Suicide Rates Relative to Major Depression in the United States. *Am J Psychiatry* 158:1652-8.
- Pauls DL, Ott J, Paul SM, et al (1995): Linkage analyses of chromosome 18 markers do not identify a major susceptibility locus for bipolar affective disorder in the Old Order Amish. *Am J Hum Genet* 57:636-43.
- Potash J, Buervenich S, Cox N, et al (2008): Gene-based SNP mapping of a psychotic bipolar affective disorder linkage region on 22q12.3: association with HMG2L1 and TOM1. *Am J Med Genet B Neuropsychiatr Genet* 147B:59-67.
- Scott LJ, Muglia P, Kong XQ, et al (2009): Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *P Natl Acad Sci USA* 106:7501-06.
- Sobel E, Lange K (1996): Descent Graphs in Pedigree Analysis: Applications to Haplotyping, Location Scores, and Marker Sharing Statistics. *Am J Hum Genet* 58:1323-37.
- Stefansson H, Rujescu D, Cichon S, et al (2008): Large recurrent microdeletions associated with schizophrenia. *Nature* 455:232-6.
- Sullivan P, Neale M, Kendler K (2000): Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 157:1552-62.
- Uhl GR, Drton T, Liu QR, et al (2008): Genome-wide association for methamphetamine dependence: convergent results from 2 samples. *Arch Gen Psychiat* 65:345-55.
- Walsh T, McClellan J, McCarthy S, et al (2008): Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science (New York, NY)* 320:539-43.
- Willour VL, Zandi PP, Badner JA, et al (2007): Attempted suicide in bipolar disorder pedigrees: evidence for linkage to 2p12. *Biol Psychiat* 61:725-7.
- Zubenko G, Hughes H, Stiffler J, et al. (2002): Genome survey for susceptibility loci for recurrent, early-onset major depression: results at 10 cM resolution. *Am J Med Genet* 114:413-22.
- Zubenko GS, Maher BS, Hughes HB, et al (2004): Genome-wide linkage survey for genetic loci that affect the risk of suicide attempts in families with recurrent, early-onset, major depression. *Am J Med Genet B* 129B:47-54.