

ARTICLE

Symptom dimensions as alternative phenotypes to address genetic heterogeneity in schizophrenia and bipolar disorder

Aurélie Labbe^{1,2,3}, Alexandre Bureau^{4,5}, Isabel Moreau⁴, Marc-André Roy⁴, Yvon Chagnon⁴, Michel Maziade⁴ and Chantal Merette^{4,6}

This study introduces a novel way to use the lifetime ratings of symptoms of psychosis, mania and depression in genetic linkage analysis of schizophrenia (SZ) and bipolar disorder (BP). It suggests using a latent class model developed for family data to define more homogeneous symptom subtypes that are influenced by a smaller number of genes that will thus be more easily detectable. In a two-step approach, we proposed: (i) to form homogeneous clusters of subjects based on the symptom dimensions and (ii) to use the information from these homogeneous clusters in linkage analysis. This framework was applied to a unique SZ and BP sample composed of 1278 subjects from 48 large kindreds from the Eastern Quebec population. The results suggest that our strategy has the power to increase linkage signals previously obtained using the diagnosis as phenotype and allows for a better characterization of the linkage signals. This is the case for a linkage signal, which we formerly obtained in chromosome 13q and enhanced using the dimension mania. The analysis also suggests that the methods may detect new linkage signals not previously uncovered by using diagnosis alone, as in chromosomes 2q (delusion), 15q (bizarre behavior), 7p (anhedonia) and 9q (delusion). In the case of the 15q and 2q region, the results coincide with linkage signals detected in other studies. Our results support the view that dissecting phenotypic heterogeneity by modeling symptom dimensions may provide new insights into the genetics of SZ and BP.

European Journal of Human Genetics (2012) 20, 1182–1188; doi:10.1038/ejhg.2012.67; published online 25 April 2012

Keywords: phenotype refinement; linkage analysis; latent class analysis

INTRODUCTION

Over the past years, genome-wide association and linkage studies have made major contributions to advancing our understanding of many complex diseases, namely age-related macular degeneration, breast cancer, prostate cancer, obesity and particularly inflammatory bowel diseases and type-2 diabetes, for which the findings of GWA studies led to the discovery of > 10 loci or genes in each of these diseases.

In comparison, the uncovering of genotype–phenotype relationships in psychiatric disorders such as schizophrenia (SZ) and bipolar disorder (BP) is still at an early stage. In the past few years, significant efforts have been invested in bridging the gap between susceptible genotypes and diagnostic criteria by searching for alternative phenotypes, representing an aspect of the disorder believed to be influenced by fewer genes.¹ Given the well-known clinical heterogeneity of several psychiatric disorders, an informative phenotype may also result from the use of such alternative phenotypes to identify subgroups in which the relationship between genotype and diagnosis is more direct, simpler and stronger than on the total sample. The hypothesis that underlies such an approach is that genetic heterogeneity accounts largely for the variation seen in clinical expression. For SZ or BP, a new generation of studies have combined clustering techniques based either on endophenotypes or clinical

phenotypes with genetic data to seek more homogeneous subgroups of patients in order to increase power of linkage or association studies. An example for SZ is provided by Lin *et al*² who found a cluster of SZ families characterized by a deficit on the cognitive performance test within which family-based association in 6p24.3 was significant. Recently, Fanous *et al*³ used latent class analysis to divide psychotic subjects from Irish high-density families into six classes based on the operational criteria for psychotic illness. After assigning each subject to their most probable class, a linkage analysis was applied for each class after considering any individual belonging to that class as affected. Using this approach, four chromosomal regions were suggestively linked but provided little additional linkage evidence over that obtained with the original diagnostic phenotype.

Two points of great analytic importance for such studies are that: (1) the statistical analysis typically relies on traditional clustering algorithms, which ignore the dependency among family members and may therefore not be totally appropriate for analyzing family data and (2) the subjects are typically assigned to their most probable cluster without taking into account the level of uncertainty of cluster memberships. As we describe below, our study will address these particular drawbacks.

¹Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada; ²Department of Psychiatry, McGill University, Montréal, Québec, Canada; ³Douglas Mental Health University Institute, Verdun, Québec, Canada; ⁴Centre de Recherche Université Laval Robert Giffard, Québec, Canada; ⁵Département de Médecine Sociale et Préventive, Université Laval, Québec, Canada; ⁶Département de Psychiatrie, Université Laval, Québec, Canada

*Correspondence: Dr A Labbe, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue, Montréal, Québec H3A1A2, Canada. Tel: +514 398 5625; Fax: +514 398 4503; E-mail: Aurelie.labbe@mcgill.ca

Received 21 October 2011; revised 22 February 2012; accepted 8 March 2012; published online 25 April 2012

In large multigenerational families, statistical methods to identify clusters of subjects based on their symptom patterns are usually not adapted to the design of the study. Because subjects within families are genetically dependent, a standard cluster analysis should not be carried out on all subjects. In a recent paper, we developed a latent class model for pedigree data that aimed to identify subgroups of subjects based on their symptom or endophenotype patterns.⁴ This method accounts for the dependence among subjects and addresses both genetic heterogeneity and phenotype definition within a unified framework. It assumes that clinical heterogeneity reflects etiological heterogeneity and that the subgroups of subjects identified should be closer to the underlying diseases, that is, groups of subjects with a distinct etiology. Our group proposed a strategy showing how to use these subtypes in linkage analysis, and the method has already successfully enhanced linkage results in autism.⁵ In the current paper, we apply this framework to two unique SZ and BP kindreds from the Eastern Quebec population that have been submitted to different genetic analyses.^{6–8} Our main objectives were to evaluate whether our proposed strategy would (i) lead to a better characterization of the signals than that obtained with DSM categories, (ii) allow the detection of new linkage signals and (iii) facilitate replication across two kindred samples when compared with the replication of linkage findings originally obtained with the BP or SZ diagnoses in one of the subsamples.

MATERIALS AND METHODS

Sample

A sample of 48 kindreds followed up longitudinally for >20 years has been collected from the Eastern Quebec population. A first wave of collection produced a first sample of 21 kindreds (sample 1) described in Maziade *et al.*⁷ and a second wave of data collection ended up with a second sample of 27 kindreds (sample 2) described in Merette *et al.*⁹ The combined sample of 48 multigenerational families comprises a total of 1278 individuals, 376 of whom were affected by a DSM-IV SZ or BP spectrum disorder. Lifetime DSM-III-R or DSM-IV diagnoses were made according to a stringent best-estimate lifetime procedure described in Maziade *et al.*¹⁰ and Roy *et al.*¹¹ Using these diagnoses, six phenotypes were derived: BP narrow (BP I only, $n = 121$), BP broad (BP I, $n = 121$; BP II, $n = 34$; and recurrent major depression, $n = 47$), SZ narrow (SZ only, $n = 125$), SZ broad (SZ, $n = 125$; schizophrenic form, $n = 7$; and schizotypal personality disorder, $n = 3$), common locus (CL) narrow (BP narrow, $n = 121$; SZ narrow, $n = 125$; and schizoaffective disorder, $n = 38$) and CL broad (BP broad, $n = 202$; SZ broad, $n = 135$; and schizoaffective disorder, $n = 38$).

Symptom definition

The lifetime presence and severity of psychotic, manic and depressive symptoms in a total of 455 subjects were evaluated on a six-point rating scale (each point corresponding to an operational definition of severity adapted for each symptom – 0: none, 1: questionable, 2: mild, 3: moderate, 4: marked and 5: severe) on each of the 82 items of the Comprehensive Assessment of Symptoms and History (CASH) instrument.¹² These 455 subjects contained the 376 subjects affected by SZ or BP spectrum disorder plus another set of 79 subjects who did not meet the DSM-IV criteria but featured some of the symptoms used in the DSM diagnosis. The procedure of assessment is described in detail in Maziade *et al.*⁶ Based on the information drawn during the review of the lifetime best-estimate diagnostic procedure, the CASH 82 items were assessed both in acute episodes and stabilized episodes in the 455 subjects above and covered the following 11 dimensions: delusion, hallucination, bizarre behavior, catatonia, thought disorder, alogia, anhedonia, apathy, affective blunting, mania and depression. The complete description of the dimensions is provided in Supplementary Table S1. In the current paper, these 11 – in acute episode – dimensions were used for subsequent analyses, except catatonia, which showed very limited variability across subjects.

Phenotype definition

For each of the 10 dimensions, we derived the average of the item ratings contained in a particular dimension for each subject. Note that a low level of this phenotype on a dimension composed of many items may indicate that the subject expresses symptoms only for a small number of items or expresses a low level on most items.

Genotyping

DNA extraction and genotyping procedures have been described in detail in previous papers.^{7,9} Microsatellite markers were selected from the NCBI database to produce a dense covering of the chromosomal region originally detected to be linked with SZ, BP or both in our cohort. In total, subjects were genotyped on 414 microsatellite markers on 22 chromosomes.

Statistical analysis

We repeated the following steps in the analysis of each of the 10 dimensions.

Step1 – Latent class model. We first identified dimension subtypes (ie, classes in the latent class model) corresponding to distinct levels of the phenotype by applying our latent class modeling approach for extended families¹³ in the combined sample of 48 kindreds, using our R package LCA extend available on the Comprehensive R Archive Network (www.r-project.org). Assuming that the K subtypes have been identified for a particular symptom dimension, our model is able to quantify the uncertainty of the phenotype's distribution among subjects by computing the probability of subtype membership for each subject. For example (hypothetical), if we suppose that a model with three subtypes was obtained using the dimension anhedonia and that subject no.1 had probabilities of 0.3, 0.6 and 0.1 to belong to the three subtypes, respectively, then this subject would highly contribute to subtype 2, subtly to subtype 1 and almost not to subtype 3. For the purpose of describing the resulting subtypes, these subtype membership probabilities are also used to form clusters of subjects who are most likely to belong to a given subtype. For instance, the above subject no. 1 would be assigned to the cluster corresponding to subtype 2.

Step2 – Linkage analysis. We performed a two-point parametric linkage analysis within each cluster of subjects while accounting for the uncertainty of the cluster assignment. The detailed method is described in Bureau *et al.*⁵ and the R function `latent.prob.covar` is available from the authors to insert the cluster assignment probabilities into linkage software input files. The analyses were performed using SUPERLINK¹⁴ under two transmission models (recessive and dominant), assuming a disease allele frequency of 0.1 for the recessive model and 0.01 for the dominant model. All these analyses were carried out using both affected and unaffected subjects and then using only affected subjects also. Note that an affected subject was defined as a subject with symptoms measured, even if the subject did not receive a diagnosis of SZ or BP. The parametric LOD score was then maximized in the combined sample over the recombination fraction and over the four possible combinations resulting from (i) two different transmission models (dominant vs recessive) and (ii) two different types of analyses (affected/unaffected analysis vs affected-only analysis) to produce a MOD score.

Step3 – Assessing significance of linkage findings. In the case of large pedigrees, a complex simulation process is required to assess the distribution of the MOD scores. Because of the strong dependence among the phenotypes and methods, a Bonferroni correction for multiple testing induces very conservative results and loss of power.¹⁵ Therefore, we provide here only MOD scores and not P -values corrected for the multiple tests. As a reference, note, however, that MOD scores of 2.7 and 4.18 are required for a genome-wide suggestive and significant linkage,¹⁶ respectively, using a conservative Bonferroni correction for each linkage analyses of a dimension with two subtypes (four linkage analyses per subtype). For three subtypes, the corresponding MOD scores are 2.94 (suggestive linkage) and 4.35 (significant linkage). In order to evaluate the contribution of each of the samples individually, we also report LOD scores on samples 1 and 2, corresponding to the model under which the MOD score was obtained in the combined sample.

RESULTS

Identification of dimension subtypes

We first applied the latent class model for extended families to each of the 10 phenotypes (corresponding to the 10 dimensions) individually. In general, the optimal number of subtypes for each dimension varied from two to three, with the majority of dimensions segregating two subtypes. The results are illustrated in Figure 1, which underscores the number of subtypes identified from the latent class model for each dimension, as well as the average severity level of the subjects classified in each subtype. For example, we can see that 8 out of the 10 dimensions led to two subtypes, characterized by a presence/absence of symptoms. Clustering subjects based on anhedonia or delusion, however, led to three subtypes corresponding to increasing severity levels. We also investigated the intersection between subtypes and diagnostic categories, as shown in Table 1. As we can see, none of the dimension subtypes exactly corresponded to SZ or BP, implying that the identified subtypes always represented new phenotype definitions.

In order to investigate the potential of the subtypes for genetic analysis, we measured the familial aggregation of each identified subtype by computing the number of families segregating a specific subtype (Table 2). As we can see, the identified subtypes show a strong familial aggregation, reflecting hopefully a relatively good genetic homogeneity of the corresponding clusters of subjects.

Genetic analysis of the dimension subtypes

As mentioned in the Materials and methods section, four linkage analyses were performed on each of the 10 subtypes of the phenotypes (22 subtypes in total). In this paper, we report four regions on chromosomes 2q34, 15q25, 7p14 and 9q33 with MOD scores > 3 , corresponding to signals undetected using the diagnosis as phenotype. We also report a signal on chromosome 13q14 that we classify as an enhancement of a signal previously detected with some of the DSM phenotypes on our samples. The MOD scores of these signals (with the corresponding model) for the combined sample, as well as the

corresponding LOD scores, for the individual samples are given in Table 3. Below, we detail and interpret each signal chromosome-by-chromosome.

Chromosome 13q14: enhancement of previously detected signals. Based on previous results, our data suggested a susceptibility locus in 13q13-q14 that is shared by SZ and BP. We previously reported a genome-wide suggestive linkage in this region with the CL phenotype in sample 1 that crossed the diagnosis boundaries by combining SZ, BP and schizoaffective disorders.⁷ This initial finding was also replicated in sample 2. In the combined sample, the linkage peak with CL at marker D13S1297 (42.1Mb) reached a parametric MOD score of 3.12 and an NPLpair $-\log_{10}$ *P*-value of 5.21, exceeding that obtained in each sample and indicating consistency across the two samples.⁸ Using our subtyping strategy, we found a significant linkage signal replicated across samples at marker D13S291 (44.82 Mb) reaching a MOD score of 5.20 on the second subtype identified from the mania dimension (Table 3), corresponding to a *P*-value of 1.0×10^{-6} or $-\log_{10}$ *P*-value of 6.0. As seen in Figure 1a (mania subtype 2), this subtype represents a cluster of 323 subjects expressing manic symptoms (ie, who had at least a questionable level of manic symptoms). Linkage analysis using this cluster as phenotype definition yielded a MOD score that exceeded the one based on a phenotype definition of narrow CL that encompasses five diagnoses and included 290 subjects. We must also emphasize that the difference between the MOD scores of mania subtype 2 and the narrow CL phenotype definition is not only because of the excess of 33 subjects but rather comes from the withdrawal of 52 subjects, mainly affected by SZ, who did not have any symptoms of mania combined with the addition of 85 subjects, not included in the narrow CL phenotype definition, who showed at least some questionable symptoms of mania. This reshuffle of the subjects used in linkage analysis suggests that a better genetic homogeneity was obtained and would lead to an optimal characterization of the signal on 13q14 by means of manic symptoms. As we can see in Figure 2a, the manic symptoms are well

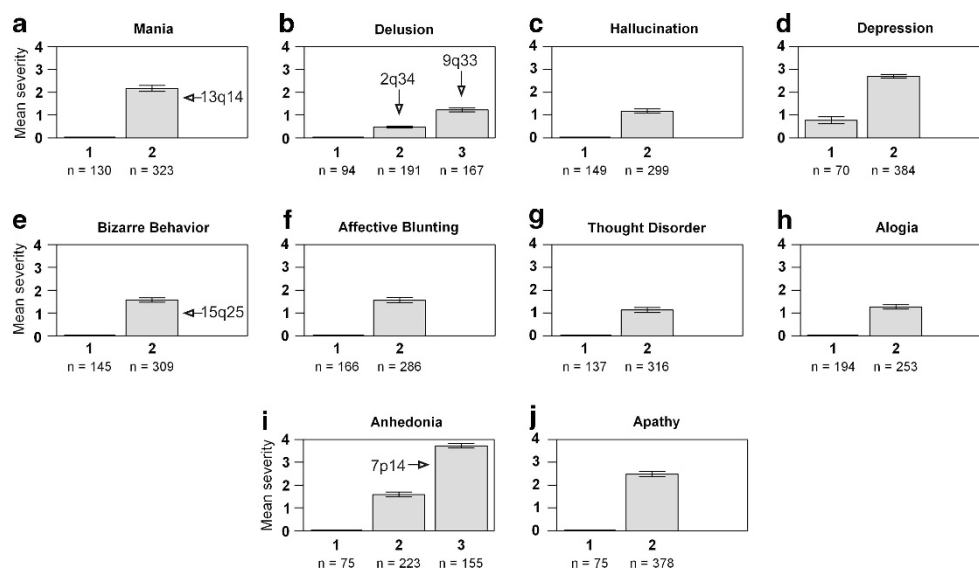


Figure 1 Each figure panel (a–j) represents the summary of the mean level severity of a given dimension within subtypes. For each subtype and each dimension, the figures show the mean severity level and its 95% confidence interval, and the number of subjects classified in each subtype (according to their most likely subtype). On each barplot, the arrow identifies the subtype yielding an enhanced or a new linkage result on the genome region identified above the arrow.

Table 1 Distribution of diagnostic phenotype in the combined sample in each of the subtypes identified by the latent class analysis

Dimension	Subtype	Signal	Number of subjects per diagnostic phenotype (proportion)				Total n = 455 ^a
			BP	SZ	SZ	No	
			Broad n = 202	Broad n = 135	Affective n = 38	Diagnosis n = 80	
Mania	1		34 (0.26)	43 (0.33)	1 (0.01)	52 (0.40)	130 (1)
	2	13q14	168 (0.52)	91 (0.28)	37 (0.11)	27 (0.08)	323 (1)
Delusion	1		52 (0.55)	2 (0.02)	0 (0)	40 (0.43)	94 (1)
	2	2q34	112 (0.59)	37 (0.19)	13 (0.07)	29 (0.15)	191 (1)
	3	9q33	38 (0.23)	94 (0.56)	25 (0.15)	10 (0.06)	167 (1)
Hallucination	1		88 (0.59)	6 (0.04)	0 (0)	55 (0.37)	149 (1)
	2		113 (0.38)	124 (0.41)	38 (0.13)	24 (0.08)	299 (1)
Depression	1		12 (0.17)	42 (0.60)	3 (0.04)	13 (0.19)	70 (1)
	2		190 (0.49)	93 (0.24)	35 (0.09)	66 (0.17)	384 (1)
Bizarre behavior	1		63 (0.43)	18 (0.12)	3 (0.02)	61 (0.42)	145 (1)
	2	15q25	139 (0.45)	117 (0.38)	35 (0.11)	18 (0.06)	309 (1)
Affective blunting	1		98 (0.59)	6 (0.04)	4 (0.02)	58 (0.35)	166 (1)
	2		103 (0.36)	128 (0.45)	34 (0.12)	21 (0.07)	286 (1)
Thought disorder	1		56 (0.41)	18 (0.13)	3 (0.02)	60 (0.44)	137 (1)
	2		146 (0.46)	116 (0.37)	35 (0.11)	19 (0.06)	316 (1)
Alogia	1		111 (0.57)	20 (0.10)	5 (0.03)	58 (0.30)	194 (1)
	2		89 (0.35)	112 (0.44)	33 (0.13)	19 (0.08)	253 (1)
Anhedonia	1		44 (0.59)	4 (0.05)	1 (0.01)	26 (0.35)	75 (1)
	2		116 (0.52)	44 (0.20)	23 (0.10)	40 (0.18)	223 (1)
	3	7p14	41 (0.26)	87 (0.56)	14 (0.09)	13 (0.08)	155 (1)
Apathy	1		40 (0.53)	4 (0.05)	3 (0.04)	28 (0.37)	75 (1)
	2		161 (0.43)	131 (0.35)	35 (0.09)	51 (0.13)	378 (1)

Abbreviations: BP, bipolar disorder; SZ, schizophrenia. ^aSome subjects had missing values for some of the dimensions. Therefore, the total number of subjects may slightly vary from one dimension to another (from 447 to 455).

Table 2: Familial aggregation of the subtypes identified by the latent class analysis (based on 48 families)

Dimension	Number of families with >80% members classified in the same subtype ^a (proportion of families)		
	Subtypes		
	1	2	3
Mania	1 (0.02)	21 (0.44)	—
Delusion	0 (0)	7 (0.15)	11 (0.23)
Hallucination	0 (0)	23 (0.48)	—
Depression	0 (0)	34 (0.71)	—
Bizarre behavior	0 (0)	20 (0.42)	—
Affective blunting	3 (0.06)	20 (0.42)	—
Thought disorder	0 (0)	21 (0.44)	—
Alogia	4 (0.06)	16 (0.33)	—
Anhedonia	0 (0)	6 (0.13)	8 (0.17)
Apathy	0 (0)	36 (0.75)	—

^aEach subject was classified in its most probable subtype, according to the subtype probabilities computed in the latent class analysis.

distributed among the eight related items, with may be an under-representation of the distractibility (mandist) symptoms.

Chromosome 9q: new signal. We obtained a new linkage signal on 9q33 (D9S299, 108.69 Mb) using a subtype 3 derived from delusion. Both samples contribute to this linkage, with LOD scores of 1.26 (sample 1) and 2.28 (sample 2). This subtype contains subjects with

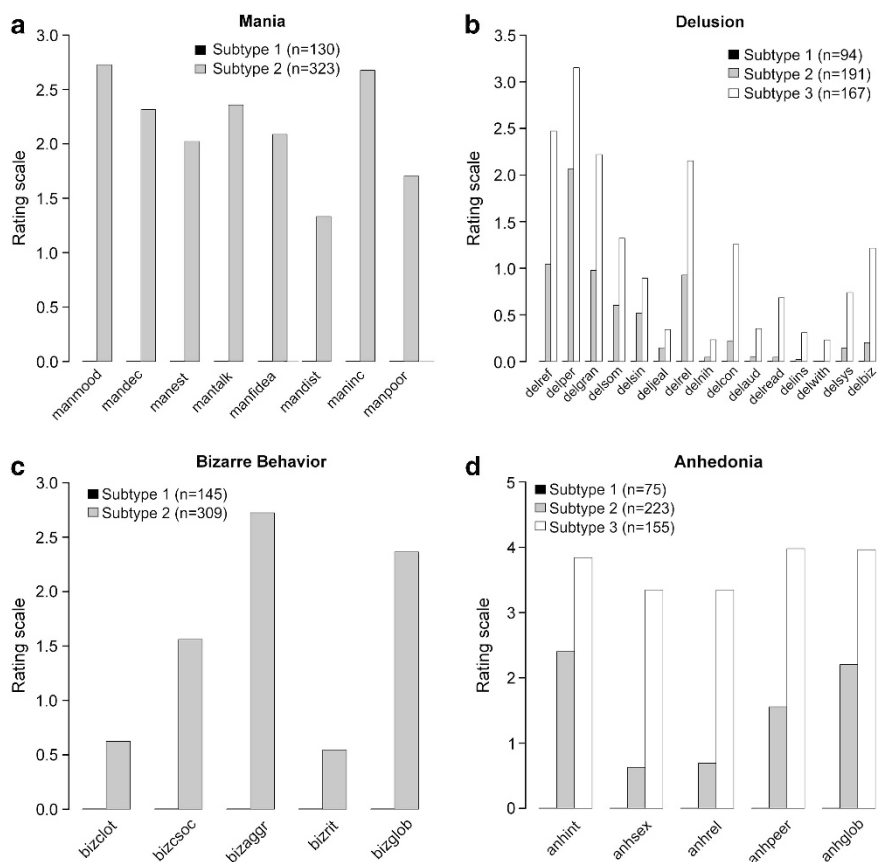
the most severe average level of delusion in our sample. Most of them are affected – but not exclusively – by SZ (Table 1) and express symptoms (mild level or more) for four, five or six items of this dimension, especially on persecutory delusion (*delper*), delusion of reference (*delref*), grandiose delusions (*delgran*) and religious delusion (*delrel*) (see Figure 2b).

Chromosome 2q: new signal. We also obtained a new suggestive linkage signal (MOD of 3.4) in our samples on 2q34 using a subtype derived from the delusion dimension on marker D2S322 (210.8 Mb) (Table 3). Both samples contribute to this linkage, with LOD scores of 1.58 (sample 1) and 2.05 (sample 2). As can be seen in Figure 1b (delusion subtype 2), this signal was derived from a cluster of subjects with a low level of delusion, in general. Note that delusion is a dimension composed of 15 items, and a low level of delusion does not always indicate that a subject is mildly affected by delusion (a subject severely affected for a single type of delusion would be rated 5 for one item and 0 for the others, resulting in a very low average). As seen in Figure 2b, the second subtype represents subjects who express on average questionable levels of delusion of reference (*delref*), grandiose delusion (*delgran*) and religious delusions (*delrel*). These subjects have also on average a mild level of persecutory delusions (*delper*) and suffer from fewer symptoms in general, compared with the third subtype in which another linkage signal was found on chromosome 9q. Indeed, 47 subjects in subtype 2 have at least four symptoms on which they were rated moderate or more, vs 128 subjects in the third subtype (results not shown). As shown in Table 1, subjects in subtype 2 are mostly (but not exclusively) affected by BP, as expected, given the low level of delusion usually observed in BP patients.

Table 3 Summary of the linkage signals detected in the combined sample and in samples 1 and 2 separately by our latent class modeling

Region	Marker (Mb)	Dimension	Model	MOD score ^a		LOD score ^b		Subtype	Diagnosis	Description
				Sample 1 and 2	1	2				
<i>Enhancement of previously detected signals</i>										
13q14	D13S291 (44.82)	Mania	D-AO	5.20	1.68	3.64		Subtype 2 (Figure 1a)	Across all diagnoses	Presence of manic symptoms
<i>New signals</i>										
2q34	D2S322 (210.80)	Delusion	R-AO	3.40	1.58	2.05		Subtype 2 (Figure 1b)	Mostly BP (but not exclusively)	Questionable levels of <i>delref</i> , <i>delgran</i> and <i>delrel</i> . Mild level of <i>delper</i> .
15q25	D15S211 (81.19)	Bizarre behavior	D-AU	3.81	1.81	2.03		Subtype 2 (Figure 1e)	Across all diagnoses	Mild or moderate levels of bizarre behavior, especially <i>bizaggr</i> and <i>bizglob</i> .
7p14	D7S671 (41.96)	Anhedonia	D-AO	3.48	3.88	0.62		Subtype 3 (Figure 1i)	Across all diagnoses	Marked level of anhedonia, across all items.
9q33	D9S299 (108.69)	Delusion	R-AO	3.48	1.26	2.28		Subtype 3 (Figure 1a)	Mostly SZ (but not exclusively)	Mild or moderate levels of <i>delper</i> , <i>delref</i> , <i>delgran</i> and <i>delrel</i> .

The map used for the genomic locations was the Ensembl map (http://uswest.ensembl.org/Homo_sapiens/Info/Index). ^aThe MOD score was obtained in the combined sample by maximizing the LOD score over the mode of inheritance (D, dominant vs R, recessive) and the type of analysis (AO, affected-only vs AU, affected/unaffected). ^bA LOD score was obtained in samples 1 and 2 by fixing the mode of inheritance and the type of analysis to the one previously obtained in the combined sample. The sum of the LOD scores of samples 1 and 2 may not exactly add up to the MOD score obtained in the combined samples (1 and 2) because of the possibility that the LOD scores were maximized for different values of the recombination fraction. MOD scores or LOD scores showing suggestive or significant linkage are indicated in bold.

**Figure 2** Summary of the mean level severity per symptoms for the four dimensions (panels a-b-c-d, respectively) for which we report a linkage signal.

Chromosome 15q: new signal. We obtained a linkage signal on 15q25 (D15S211, 81.19 Mb) using a class derived from the bizarre behavior dimension. Again, both samples contribute to the signal obtained

(Table 3), leading to a combined MOD score of 3.81. The use of this dimension in latent class analysis led to two subtypes, defined only by the presence/absence of bizarre behavior. The class showing a

linkage signal contains subjects showing a presence of bizarre behavior (see Figure 1e – subtype 2). This is a very large subtype containing 309 affected subjects in the combined sample, heterogeneous in terms of diagnosis (grouping together subjects across all diagnostic spectrum, as shown in Table 1) but homogeneous in terms of bizarre behavior. As seen in Figure 2c, this cluster of subjects express mild or moderate aggressive and agitated behavior (*bizag*) as well as global rating of bizarre behavior (*bizaglob*).

Chromosome 7p: new signal

We obtained a linkage signal on 7p14 (D7S671, 41.96 Mb) using a subtype derived from anhedonia. This subtype contains subjects with a marked average level of anhedonia (level 4). Anhedonia is a dimension composed of five items. Such a high level for the average phenotype indicates that subjects show a strong pattern of anhedonia across all symptoms (see Figure 2d) and diagnoses (see Table 1). Note that the signal is mainly in sample 1 (LOD score of 3.88, using 69 affected subjects – results not shown) but remains present in the combined sample (MOD score of 3.48).

DISCUSSION

Refining the phenotype of psychiatric disorders in genetic studies is probably one of the key elements that will help researchers to uncover the genetic basis of such complex diseases. We introduced in this paper, a novel way to use the lifetime ratings of symptoms of psychosis, mania and depression in genetic linkage analysis of SZ and BP. Our approach was developed under a solid statistical framework and, as of today, we are not aware of any other analytical framework that simultaneously allows (i) to form homogeneous groups of subjects based on their symptom patterns while accounting for the dependency of individuals in families and (ii) to use the information from these homogeneous groups in linkage analysis, accounting for the uncertainty of the subjects' group assignment. Moreover, to our knowledge, this study is the first attempt to integrate clinical symptoms of both BP and SZ in a single linkage analysis – a strategy supported by increasing evidence for commonalities between these disorders.¹⁷ Our approach both enhanced previously obtained linkage signals and facilitated the detection of new signals not previously obtained with diagnosis alone.

We note that the multivariate quantitative trait methods are not suitable for the situation where multivariate symptoms are measured only in affected individuals. This implies that all family members not meeting the DSM diagnosis criteria may be coded as having missing phenotypes, which can lead to serious biases in the quantitative trait linkage results with current methods. We show in this paper that our new strategy has the capability to enhance linkage signals obtained with a diagnosis only as phenotype, as well as potentially uncover new signals. In particular, we show that going beyond the diagnosis definition and studying a broader population in terms of affection status may provide a new insight into the genetics of SZ and BP. In fact, all the classes of symptoms from which we obtained a linkage signal contained a significant proportion of symptomatic subjects who did not meet a DSM diagnosis, but whose symptoms had been evaluated. The three signals found on chromosomes 13q, 15q and 7p illustrate perfectly a situation where subjects are extremely heterogeneous in terms of diagnosis, but share instead heritable clinical features such as manic symptoms (chr 13q), mild levels of bizarre behavior (chr 15q) or marked levels of anhedonia (chr 7p). Even in signals obtained from subjects sharing a higher degree of similarity in the diagnosis, as in 2q (mostly BP) and 9q (mostly SZ), adding subjects sharing similar clinical features across diagnoses turned out

to be the key element in uncovering a linkage signal that could not be captured using the diagnosis alone. Indeed, using the BP diagnosis as a phenotype on 2q or the SZ diagnosis on 9q does not give a linkage signal, despite having a larger number of subjects for the genetic analysis.

Some of the results obtained coincide with linkage signals previously obtained in our sample⁸ and also detected in other studies such as the signal on chromosome 13q that has been widely reported in the literature (see a review of 25 studies in Detera-Wadleigh and McMahon¹⁸), indicating two main zones of linkage with SZ or BP: one in 13q13-q14 (≈ 40 Mb) and another in 13q21-q33 (≈ 95 Mb). Our data suggest that the manic symptom dimension as a phenotype may help to reconcile the findings overlapping SZ and BP from different studies or to better understand the probable heterogeneity at this locus. Another example is the signal we found on 15q (D15S211) using the dimension based on bizarre behavior. The literature shows that this locus may be shared by other diseases, and marker D15S211 is the exact same one on which a linkage with a syndrome of severe mental retardation, spasticity and tapetoretinal degeneration was reported.¹⁹ This marker is also close to D15S652 (92.52 Mb), on which major depression disorder was mapped.²⁰ Finally, the signal on chromosome 2q34 (210.8 Mb) is close to the SNP rs17662626 on 2q32.3 (193.7 Mb) recently published by the Schizophrenia Psychiatric GWAS Consortium and the Psychiatric GWAS Consortium Bipolar Disorder Working Group^{21,22} as being associated with both SZ and BP disorders. Note that another close signal reported in the literature was found on 2q37.1 using the SZ phenotype.²³ Finally, although several diseases, such as deafness, epilepsy, colorectal cancer, Usher's syndrome or macular degeneration, have been mapped in the 9q33 region, we are not aware of any psychiatric disorder showing linkage to markers nearby. Similarly, we are not aware of any other studies showing a signal in the 7p region.

It is important to note that the power of our approach for uncovering new signals may have advantages over the traditional diagnostic approach only if affected individuals are genetically heterogeneous with respect to the 'disease' locus. Otherwise, use of diagnoses is still recommended, as it has the advantage of using all the affected individuals in the study. This is probably one of the reasons that would explain why we did not enhance linkage in well-known chromosomal regions for SZ and BP such as 15q11.1 or 18q.⁷ Furthermore, non-parametric linkage analyses did not produce any significant signals using our classes of symptoms on the six regions presented in this paper. Because non-parametric analyses have greater power than parametric analyses in the presence of heterogeneity or in the case of incomplete penetrance, the presently obtained stronger parametric results suggest that the subtypes obtained from psychotic, manic and depressive symptoms may be more genetically homogeneous when compared with the original DSM diagnosis. Note that heterogeneity within symptoms may still be detected, as shown by the two signals detected from the two subtypes representing different levels of delusion (2q-mild level and 9q-moderate level).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by grants NPE-80219 and MOP-93723 from the Canadian Institutes for Health Research. A Bureau and A Labbe are supported by scientist awards from the Fonds de la recherche en santé du Québec.

- 1 Szatmari P, Maziade M, Zwaigenbaum L *et al*: Informative phenotypes for genetic studies of psychiatric disorders. *Am J Med Genet* 2007; **114B**: 581–588.
- 2 Lin SH, Liu CM, Liu YL *et al*: Clustering by neurocognition for fine mapping of the schizophrenia susceptibility loci on chromosome 6p. *Genes Brain Behav* 2009; **8**: 785–794.
- 3 Fanous AH, Neale MC, Webb BT *et al*: Novel linkage to chromosome 20p using latent classes of psychotic illness in 270 Irish high-density families. *Biol Psychiatry* 2008; **64**: 121–127.
- 4 Labbe A, Bureau A, Merette C: Integration of genetic familial structure in latent class models. *Int J Biostat* 2009; **5**: 6.
- 5 Bureau A, Labbe A, Croteau J, Merette C: Using disease symptoms to improve detection of linkage under genetic heterogeneity. *Genet Epidemiol* 2008; **32**: 476–486.
- 6 Maziade M, Roy MA, Martinez M *et al*: Negative, psychoticism, and disorganized dimensions in patients with familial schizophrenia or bipolar disorder: continuity and discontinuity between the major psychoses. *Am J Psychiatry* 1995; **152**: 1458–1463.
- 7 Maziade M, Roy MA, Chagnon YC *et al*: Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in Eastern Quebec families. *Mol Psychiatry* 2005; **10**: 486–499.
- 8 Maziade M, Chagnon YC, Roy MA, Bureau A, Fournier A, Merette C: Chromosome 13q13–q14 locus overlaps mood and psychotic disorders: the relevance for redefining phenotype. *Eur J Hum Genet* 2009; **17**: 1034–1042.
- 9 Merette C, Roy MA, Bureau A *et al*: Replication of linkage with bipolar disorder on chromosome 16p in the Eastern Quebec population. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 737–744.
- 10 Maziade M, Raymond V, Cliche D *et al*: Linkage results on 11Q21–22 in Eastern Quebec pedigrees densely affected by schizophrenia. *Am J Med Genet* 1995; **60**: 522–528.
- 11 Roy MA, rette C, Cliche D *et al*: Clinical and methodological factors related to reliability of the best-estimate diagnostic procedure. *Am J Psychiatry* 1997; **154**: 1726–1733.
- 12 Andreasen NC, Flaum M, Arndt S: The comprehensive assessment of symptoms and history (CASH) an instrument for assessing diagnosis and psychopathology. *Arch Gen Psychiatry* 1992; **49**: 615–623.
- 13 Tayeb A, Labbe A, Bureau A, Merette C: Solving genetic heterogeneity in extended families by identifying sub-types of complex diseases. *Comput Stat* 2011; **26**: 539–560.
- 14 Fishelson M, Geiger D: Exact genetic linkage computations for general pedigrees. *Bioinformatics* 2002; **18**: S189–S198.
- 15 Romanos M, Freitag C, Jacob C *et al*: Genome-wide linkage analysis of ADHD using high-density SNP arrays: novel loci at 5q13.1 and 14q12. *Mol Psychiatry* 2008; **13**: 522–530.
- 16 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- 17 Van Snellenberg JX, de Candia T: Meta-analytic evidence for familial coaggregation of schizophrenia and bipolar disorder. *Arch Gen Psychiatry* 2009; **66**: 748–755.
- 18 Detera-Wadleigh SD, McMahon FJ: G72/G30 in schizophrenia and bipolar disorder: review and meta-analysis. *Biol Psychiatry* 2006; **60**: 106–114.
- 19 Mitchell SJ, McHale DP, Campbell DA *et al*: A syndrome of severe mental retardation, spasticity, and tapetoretinal degeneration linked to chromosome 15q24. *Am J Hum Genet* 1998; **62**: 1070–1076.
- 20 Holmans P, Zubenko GS, Crowe RR *et al*: Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *Am J Hum Genet* 2004; **74**: 1154–1167.
- 21 Ripke S, Sanders AR, Kendler KS *et al*: Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
- 22 Sklar P, Ripke S, Scott LJ *et al*: Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43**: 977–983.
- 23 Paunio T, Ekelund J, Varilo T *et al*: Genome-wide scan in a nationwide study sample of schizophrenia families in Finland reveals susceptibility loci on chromosomes 2q and 5q. *Hum Mol Genet* 2001; **10**: 3037–3048.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)