

RESEARCH

Open Access



A significant risk locus on 19q13 for bipolar disorder identified using a combined genome-wide linkage and copy number variation analysis

Magnus Lekman¹, Robert Karlsson², Lisette Graae³, Ola Hössjer^{4*} and Ingrid Kockum¹

* Correspondence: ola@math.su.se

⁴Department of Mathematics,
Division of Mathematical Statistics,
Stockholm University, SE-106 91
Stockholm, Sweden

Full list of author information is
available at the end of the article

Abstract

Background: The genetic background to bipolar disorder (BPD) has been attributed to different genetic and genomic risk factors. In the present study we hypothesized that inherited copy number variations (CNVs) contribute to susceptibility of BPD. We screened 637 BP-pedigrees from the NIMH Genetic Initiative and gave priority to 46 pedigrees. In this subsample we performed parametric and non-parametric genome-wide linkage analyses using ~21,000 SNP-markers. We developed an algorithm to test for linkage restricted to regions with CNVs that are shared within and across families.

Results: For the combined CNV and linkage analysis, one region on 19q13 survived correction for multiple comparisons and replicates a previous BPD risk locus. The shared CNV map to the pregnancy-specific glycoprotein (*PSG*) gene, a gene-family not previously implicated in BPD etiology. Two SNPs in the shared CNV are likely transcription factor binding sites and are linked to expression of an F-box binding gene, a key regulator of neuronal pathways suggested to be involved in BPD etiology.

Conclusions: Our CNV-weighted linkage approach identifies a risk locus for BPD on 19q13 and forms a useful tool to future studies to unravel part of the genetic vulnerability to BPD.

Keywords: Bipolar disorder, Genome-wide, Linkage analysis, Copy number variation

Background

Bipolar disorder (BPD) is a burdensome [1] and common [2] spectrum of mental disorders [3]. The concordance rate which is up to 8.5 times higher for monozygotic than dizygotic twins for BPD shows that genetic factors contribute to susceptibility to BPD [4]. Non-genetic factors are however also of importance in the underlying etiology since the heritability rates are between 59 and 87 % [4]. Accumulating data demonstrates that BPD is a both clinically and genetically heterogeneous disorder with different risk factors in different subgroups and with a shared genetic overlap between different diagnoses of psychiatric disorders [5, 6]. The emerging picture further reveals that individual genetic risk loci contribute with relatively small effect. This complicated

picture has for a long time hampered the success to find robust genetic results in BPD. However, the availability to both larger sample sizes, more dense marker map [7, 8] and with application of new methodologies enforced by the venture from the Psychiatric Genomic Consortium (PGC) [9] have finally allowed for statistical robust signals from the *CACNA1C*, *CACNB2*, *ODZ4*, *SYNE1* and *NCAN* genes.

The genetic architecture of BPD is however more complicated than previously anticipated [10]. A variety of genomic polymorphisms, not only restricted to variation of single nucleotides may explain why it has been hard to identify BPD susceptibility genes [11–13]. This type of variations include larger genomic segments known as copy number variation (CNV) which, in comparison to a reference genome, are defined as gain or loss of genomic segments larger than 1 kb in size [13, 14]. It was previously thought that these forms of genetic polymorphisms are relatively rare and highly potent in conferring risk [12, 15]. Recent findings indicate that besides the highly penetrant rare risk variants, common variants also occur but with a more modest risk contribution than previously assumed [12, 16]. An emerging picture thus indicates that different forms of genomic variations may explain some of the expected genetic risk for a group of individuals [10, 11, 17].

In this study our hypothesis was that CNVs irrespective of their frequencies predispose to BPD. We further hypothesized that such structural variants are inherited. Under such assumptions, families ascertained for having high genetic liability of BPD constitutes an unprecedented opportunity to find such genomic variations in regions with evidence for linkage. To increase power to find families with linkage to BPD we screened 637 BP-pedigrees, provided by the NIMH Bipolar Genetic Initiative, and selected a subsample presumed to carry a genetic form of BPD. Pedigrees were selected based on family-wise genome-wide linkage analysis or by analyzing candidate genes for presence of stretches of deletions. We selected 46 BP-pedigrees for our present study and conducted two separate forms of analyses. First, we performed parametric and non-parametric genome-wide linkage analyses using a dense SNP-marker map with genetic data filtered for genotypes mapping to CNVs.

We next tested the same sample with parametric and non-parametric linkage analyses restricted only to regions with CNVs that are shared among at least two individuals within the same family. To do this we developed an algorithm to sum family-wise linkage scores in regions with CNVs that are shared within and across families.

Our results demonstrate that for the linkage part of this study, several signals surpassed threshold of suggestive linkage for both the non-parametric and parametric models and confirm several previously reported linkage regions. For the combined CNV and linkage analysis, one region on 19q13 survived correction for multiple comparisons and confirms a previously reported risk locus for BPD.

Several plausible candidate genes for BPD reside in 19q13. Moreover, two markers in the identified CNV have been reported as eQTLs for an F-box binding protein (*FBXO30*) with a suggestive role in BPD susceptibility.

Methods

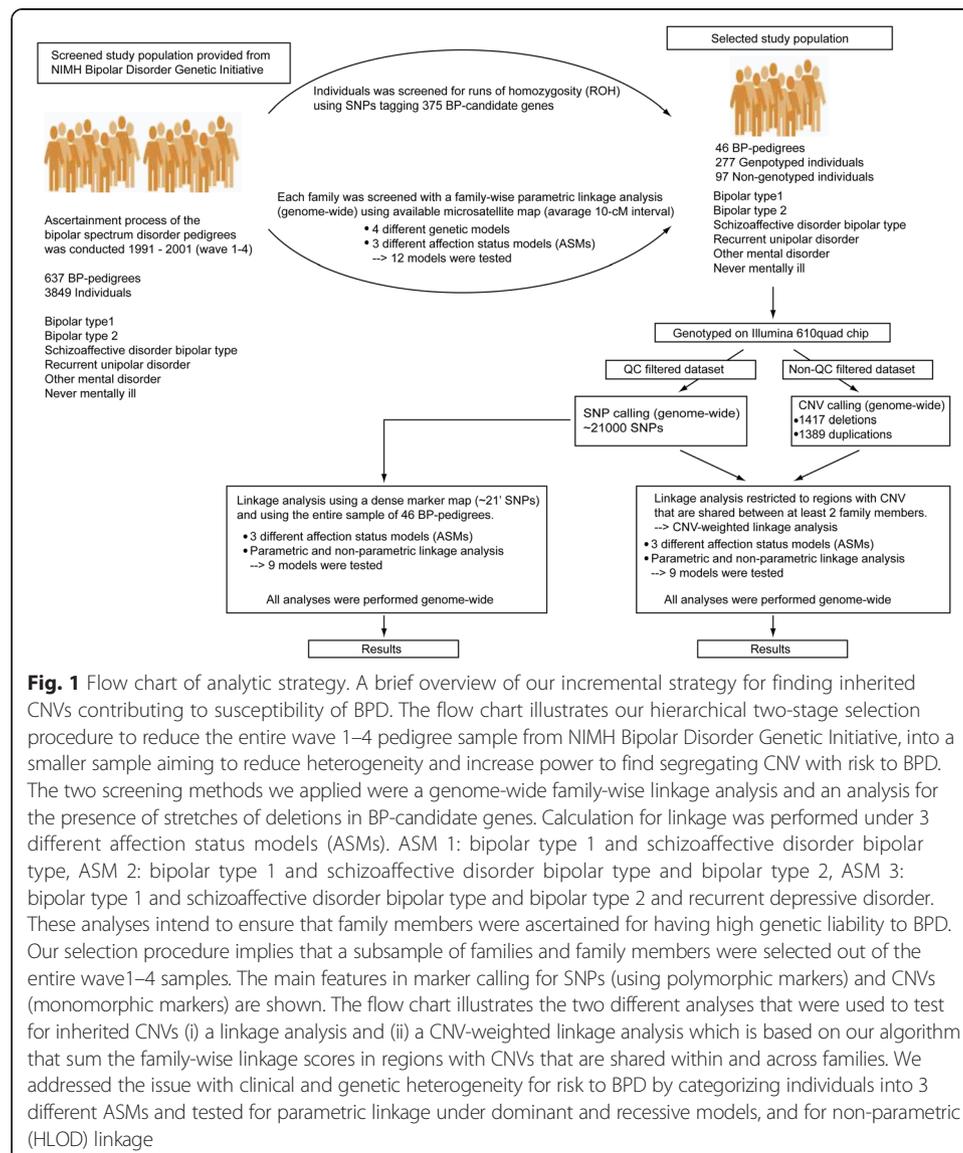
Study subjects and recruitment process

The BP-pedigrees and genotypic data were provided from the NIMH Bipolar Disorder Genetic Initiative [18] with ascertainment and diagnosis processes conducted during

1991 to 2001 (Wave 1–4) (detailed described in Additional file 1). This material has been analyzed for linkage to BPD in previous studies without yielding signals that meet the criterion of significant linkage (Additional file 1).

Given the complex nature of BPD we applied a screening process intended to select families presumed to carry a relatively strong genetic influence of risk to BPD (Fig. 1). In this process genotypic data from 3849 individuals in 637 nuclear and extended BP-pedigrees was downloaded from the NIMH Genetic Initiatives' website [19] and screened with a genome-wide family-wise (defined as the type 1 error for one single family) parametric linkage scan (microsatellite map, average 10-cM interval) under different genetic models (dominant, recessive, hyper dominant or affected only) using GENEHUNTER [20] (allele frequencies and penetrance vectors described under linkage analysis).

We followed a common practice for finding linked loci of a complex disorder with an unclear classification between the different subtypes of the disorder and considered three affection status models (ASMs) [18].



The criteria we used for selecting BP-pedigrees were a family-wise logarithm of odds (LOD) score > 1.1 , or if several families were found to have overlapping family-wise LOD scores > 1.0 in the same genomic region. Pedigrees were also screened for stretches of homozygous genotypes (ROH), possibly indicating deletions, as this type of genetic polymorphism has been shown to have a suggestive role in the etiology of BPD [21]. To do this, individuals in Wave1-4 BP-pedigrees were screened for deletions in SNPs tagging 357 candidate genes for BPD, by scanning for runs of homozygosity and Mendelian error analysis using the PLINK software [22]. Our ROH-based inclusion criterion was thus to find larger stretches of deletions in single individuals, and include the corresponding family regardless of whether deletions occurred in other family members.

In total we selected 46 families for our analyses consisting of 277 individuals with DNA and 97 individuals for whom DNA was not available (Additional file 2). We applied a pre-agreed analysis plan not to report a final result in the present study for a region that formed inclusion criterion for any family. In absence of interaction, this eliminates any systematic bias between the removed and retained regions. Table 1 summarizes the pedigree structure for the selected sample (Additional file 3 provides full details of the 46 pedigrees).

Table 1 Pedigree summary statistics and hierarchical affection status models

Pedigree structure			
Families	46		
Founders	147		
Non-founders	219		
Total number of individuals	366		
Number of affected individuals	176		
Number of BP-I individuals	168		
Number of BP-II individuals	9		
Number of SABP individuals	9		
Number of RUDD individuals	12		
Family size (average)	7.96		
Generations (average)	2.74		
Number of affected relative pairs			
1 degree	224	183	200
2 degree	106	99	91
3 degree	33	27	34
4 degree	6	8	10
5 degree	4	3	3
6 degree	2	2	2
Hierarchical affection status models (ASM)			
Very narrow affection status model (ASM1)	BP-I and SABP		
Narrow affection status model (ASM2)	BP-I and SABP and BP-II		
Broad affection status model (ASM3)	BP-I and SABP and BP-II and RUDD		

The table displays the summary statistics for the 46 pedigrees under the three different affection status models (ASMs). BP-I: Bipolar type 1, BP-II: Bipolar type 2, SABP: Schizoaffective bipolar type, RUDD: Recurrent unipolar depressive disorder. Individuals with a diagnosis of bipolar spectrum disorders that only apply to a certain ASM were coded as "unknown" under the other ASMs

Given the complex background of BPD, with clinical and genetic heterogeneities, and the additional uncertainty in diagnosis classification, we prioritized a smaller sample rather than a large collection of extended BP-pedigrees in the NIMH Bipolar Disorder Genetic Initiative wave 1–4. The complex pattern of inheritance in extended pedigrees complicates detection of segregating risk loci with linkage analysis. In the present linkage study, we therefore aimed to increase power using one family based criterion that tends to select pedigrees that are genetically homogeneous, and one ROH criterion which allows several smaller pedigrees to be chosen.

Genotyping and quality controls

For the final selected individuals DNA samples (lymphoblastoid cells) were obtained from Rutgers University and Cell Repository (New Jersey, USA) and were genotyped using the Illumina Human 610quad chip at Uppsala University, Sweden. We applied quality control (QC) analysis of individuals, SNP and CNV data consisting of an ordered series of steps to prevent spurious signals that may otherwise mislead statistical inference. Of notice, in order to reduce the presence of erroneous genotypes, SNPs located within CNV-regions were removed. Methods of genotyping and QC analysis are described in Additional file 1.

After final QC filter 269 individuals with DNA from 46 pedigrees (44 pedigrees with Caucasian ancestry and 2 pedigrees with African American ancestry) and 20,714 SNP-markers were ready for the linkage analysis. CNV calling with PennCNV and QC analysis, as described in the Additional file 1, identified 2806 CNVs (1417 deletions and 1389 duplications). The mean CNV length was 110,455 bp with a maximum and minimum length of 4,580,011 bp and 10,046 bp respectively.

All of our CNVs overlapped with variations reported in three publicly available databases in August 2013; Database of Genomic Variants [23], Wellcome Trust Sanger Institute [24] and finally from Children's Hospital of Philadelphia [25].

Linkage analysis

To unravel a segregating risk locus to BPD in our sample the different levels of heterogeneity, both clinical and genetic, prompted us to categorize individuals in different affection status models (ASMs) and to calculate linkage using multipoint parametric dominant and recessive models that take into account inter-familial heterogeneity using heterogeneity LOD scores (HLOD) and multipoint non-parametric models in MERLIN (v.1.1.2) [26].

For the non-parametric analysis we used the linear model [27] with an 'all affected relative pairs' (NPL_{ALL}) statistics, in order to identify linkage. The Z scores were converted to LOD scores and *P* values according to Kong and Cox 1997 [20]. For the parametric dominant model we assumed a risk allele frequency of 0.0045 with penetrance vector 0.001, 0.50 and 0.75 for the three different genotypes, and for the recessive model we assumed a risk allele frequency of 0.065 with a penetrance vector of 0.0015, 0.0015 and 0.50 [28]. The GENEHUNTER software [20] was used to generate phased haplotypes and positions of recombinants.

Empirical significance levels for linkage analysis

For the linkage part of this study we defined the suggestive linkage level to be that which on average would be exceeded by one linkage peak by chance in a genome-wide

scan, and significant linkage to be that which would be expected to exceed once per 20 genome-wide scans as proposed by Lander and Kruglyak 1995 [29]. To define these thresholds we simulated 1000 datasets using the MERLIN software in which phenotypic status and pedigree structure were retained while simulating random multilocus genotypes. More specifically, these simulations are based on gene dropping in all pedigrees under the null hypothesis of no linkage. For each pedigree, marker alleles are first simulated for founders (according to their allele frequencies), then haplotypes are propagated to all non-founders based on Mendelian segregation and recombinations.

The threshold for the empirical P value corresponding to suggestive linkage was then calculated based on a family-wise error rate (FWER) of $1 - e^{-1} = 0.632$, this is the probability that a Poisson distributed random variable with an expected value of 1, is positive, and it approximates the probability of at least one significant linkage peak [29].

As the models are nested, correction for multiple comparisons across the different models would have been too conservative, we corrected only within each model. Additional file 4 reaffirms that the linkage peaks of all nine models are strongly correlated.

Algorithm of CNV-weighted linkage scores

In order to find inherited genomic regions conferring risk for BPD, the sum of average family-wise parametric LOD scores or non-parametric Z scores were calculated over regions and families sharing overlapping CNVs. For families with at least two members with overlapping CNVs, the average linkage score in the region was calculated and added to those observed in the same region in other families. Thus, the algorithm generates a CNV-weighted linkage scores for genomic segments representing regions with CNVs that are shared within and across families (detailed described in Additional file 1 and Additional file 5).

Empirical significance levels for CNV-weighted linkage scores

Empirical P values for the CNV-weighted linkage score analyses were also derived by gene dropping from simulated multilocus genotypes in MERLIN, using the same marker allele frequencies of the founders as in the unweighted linkage analysis simulations. For these analyses the threshold for significance was defined based on a FWER of $\alpha = 0.05$. The nested models made us to correct within each model and not across the different models. Figure 2 reaffirms that the CNV-weighted linkage peaks of all nine models are strongly correlated.

Results

Linkage analysis

The multipoint non-parametric (NPL_{ALL}) linkage analysis met suggestive level on 3p14.1. The three affection status models generated comparable results with ASM3 exhibiting the strongest peak, $NPL Z = 3.56$ [20] (Tables 2 and 3 and Additional file 4). With the parametric dominant model suggestive linkage, $HLOD = 2.41$, was reached for the same region (3p14.1) as in the non-parametric model (Table 2 and 3 and Additional file 4). For the recessive model, the most notable evidence of linkage occurred at 6p12.3 ($HLOD = 2.64$). Other regions reaching suggestive linkage were 1q23.3 ($HLOD = 2.04$) and 10q26.2 ($HLOD = 2.25$), all with ASM3.

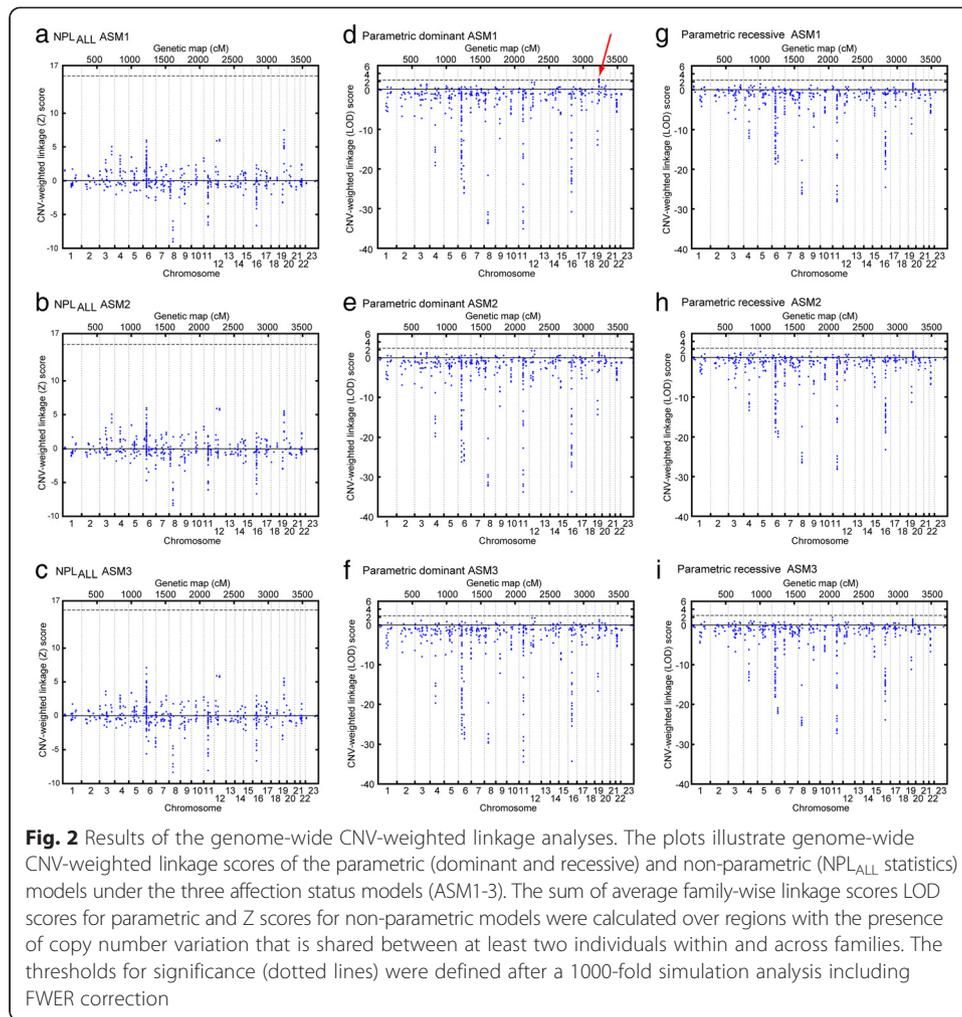


Fig. 2 Results of the genome-wide CNV-weighted linkage analyses. The plots illustrate genome-wide CNV-weighted linkage scores of the parametric (dominant and recessive) and non-parametric (NPL_{ALL} statistics) models under the three affection status models (ASM1-3). The sum of average family-wise linkage scores LOD scores for parametric and Z scores for non-parametric models were calculated over regions of copy number variation that is shared between at least two individuals within and across families. The thresholds for significance (dotted lines) were defined after a 1000-fold simulation analysis including FWER correction

Table 2 Summary table of the suggestive linkage results

Maximum linkage scores of parametric linkage analysis									
Chr	95 % C.I. (cM)	Peak marker	Locus	HLOD (threshold)	Alpha	IC	Diagnostic model	Genetic model	Empirical P value (threshold)
1	6.2	rs10919096	1q23.3	2.04 (1.97)	0.20	0.95	ASM3	Recessive	0.57 (0.632)
3	4.9	rs1001763	3p14.1	2.41 (1.97)	0.29	0.91	ASM3	Dominant	0.32 (0.632)
6	5.2	rs9381631	6p12.3	2.64 (1.97)	0.28	0.91	ASM3	Recessive	0.19 (0.632)
10	4.5	rs638395	10q26.2	2.25 (1.97)	0.28	0.93	ASM3	Recessive	0.40 (0.632)

The table displays maximum linkage scores for the suggestive linkage results of the non-parametric and parametric linkage analyses

A 1000-fold simulation analysis generated genome-wide thresholds for suggestive linkage Z and parametric HLOD scores (shown in parentheses). The approximate 95 % confidence intervals for the highest linkage scores are defined as a LOD-drop, or Z-drop, corresponding to a unit of 1.0. The mean information content (IC) estimate across all chromosomes was 0.92 (range 0.96–0.70), indicating that most of the genetic information was successfully captured using our high density mapping approach. The fraction of linked families, alpha, was less than 0.29 for regions showing suggestive linkage, indicating that there is evidence for a marked heterogeneity with respect to linked loci in these 46 pedigrees

Abbreviations: NPL Z the linkage statistics for the estimation of *identity-by-descent* (IBD) allele sharing. LOD is the LOD calculated from NPL Z scores according to Kong and Cox (1997). IC the information content and is a measure of the probability that the IBD status at a certain locus can be determined for a given pairs of relatives. Alpha a measure of the locus heterogeneity that indicates the proportion of families with alleles linked to disease at a certain locus.

HLOD estimate of the heterogeneity LOD score. Delta measure of allele sharing among affected individuals within pedigrees. Empirical P value estimated probability of having a score by chance that is at least as large as the observed one, after correction for multiple comparisons

Table 3 Summary table of the suggestive linkage results

Maximum linkage scores of non-parametric linkage analysis									
Chr	95 % C.I. (cM)	Peak marker	Locus	NPL Z (threshold)	LOD	IC	Delta	Diagnostic model	Empirical <i>P</i> value (threshold)
3	18.5	rs4855407	3p14.1	3.56 (3.34)	2.34	0.91	0.50	ASM3	0.46 (0.63)
3	3.49	rs4855407	3p14.1	3.49 (3.32)	2.32	0.91	0.49	ASM2	0.48 (0.63)
3	18.5	rs4855407	3p14.1	3.56 (3.30)	2.34	0.91	0.43	ASM1	0.41 (0.63)

The table displays maximum linkage scores for the suggestive linkage results of the non-parametric and parametric linkage analyses

A 1000-fold simulation analysis generated genome-wide thresholds for suggestive linkage Z and parametric HLOD scores (shown in parentheses). The approximate 95 % confidence intervals for the highest linkage scores are defined as a LOD-drop, or Z-drop, corresponding to a unit of 1.0. The mean information content (IC) estimate across all chromosomes was 0.92 (range 0.96–0.70), indicating that most of the genetic information was successfully captured using our high density mapping approach. The fraction of linked families, alpha, was less than 0.29 for regions showing suggestive linkage, indicating that there is evidence for a marked heterogeneity with respect to linked loci in these 46 pedigrees

Abbreviations: NPL Z the linkage statistics for the estimation of *identity-by-descent* (IBD) allele sharing. LOD is the LOD calculated from NPL Z scores according to Kong and Cox (1997). IC the information content and is a measure of the probability that the IBD status at a certain locus can be determined for a given pairs of relatives. Alpha a measure of the locus heterogeneity that indicates the proportion of families with alleles linked to disease at a certain locus.

HLOD estimate of the heterogeneity LOD score. Delta measure of allele sharing among affected individuals within pedigrees. Empirical *P* value estimated probability of having a score by chance that is at least as large as the observed one, after correction for multiple comparisons

CNV-weighted linkage analysis

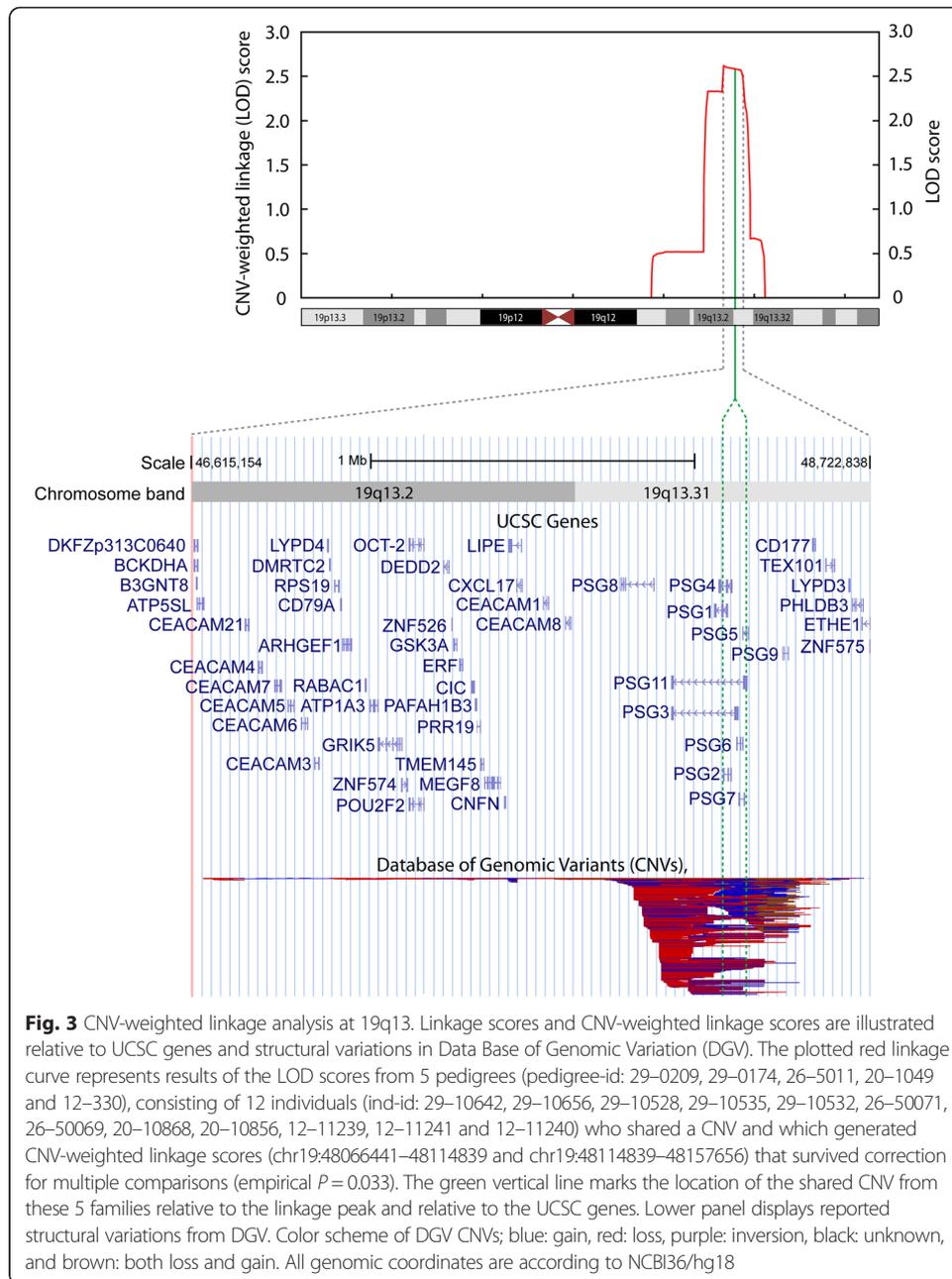
We identified 2806 CNVs in our families all of which overlapped with previously reported CNVs in three publicly available databases (see Materials and Methods).

Of the nine different linkage and affection status models that were tested only the parametric dominant model, under a narrow diagnose classification (ASM1), exhibited a significant CNV-weighted linkage score after a 1000-fold simulation analysis including correction for multiple comparisons, with empirical $P = 0.033$ (Fig. 2). The significant signal that exceeded the genome-wide threshold (dashed line of Panel d in Fig. 2) resides on 19q13 and represents a region of CNVs that due to our algorithm design was divided into two separate segments (chr19:48066441–48114839 and chr19:48114839–48157656) (Fig. 3) and was generated from 5 families (Fig. 4a). The positions of each CNV are provided in Additional file 6. CNVs were present among 12 individuals, 9 of which were affected under the narrow affection status model (ASM1) in the 5 families contributing to the CNV weighted linkage score (illustrated in Fig. 3).

The 19q13 region exhibits frequent CNVs in the general population and stretches over a region that harbors a family of genes encoding the pregnancy-specific glycoproteins (PSGs) (Fig. 3).

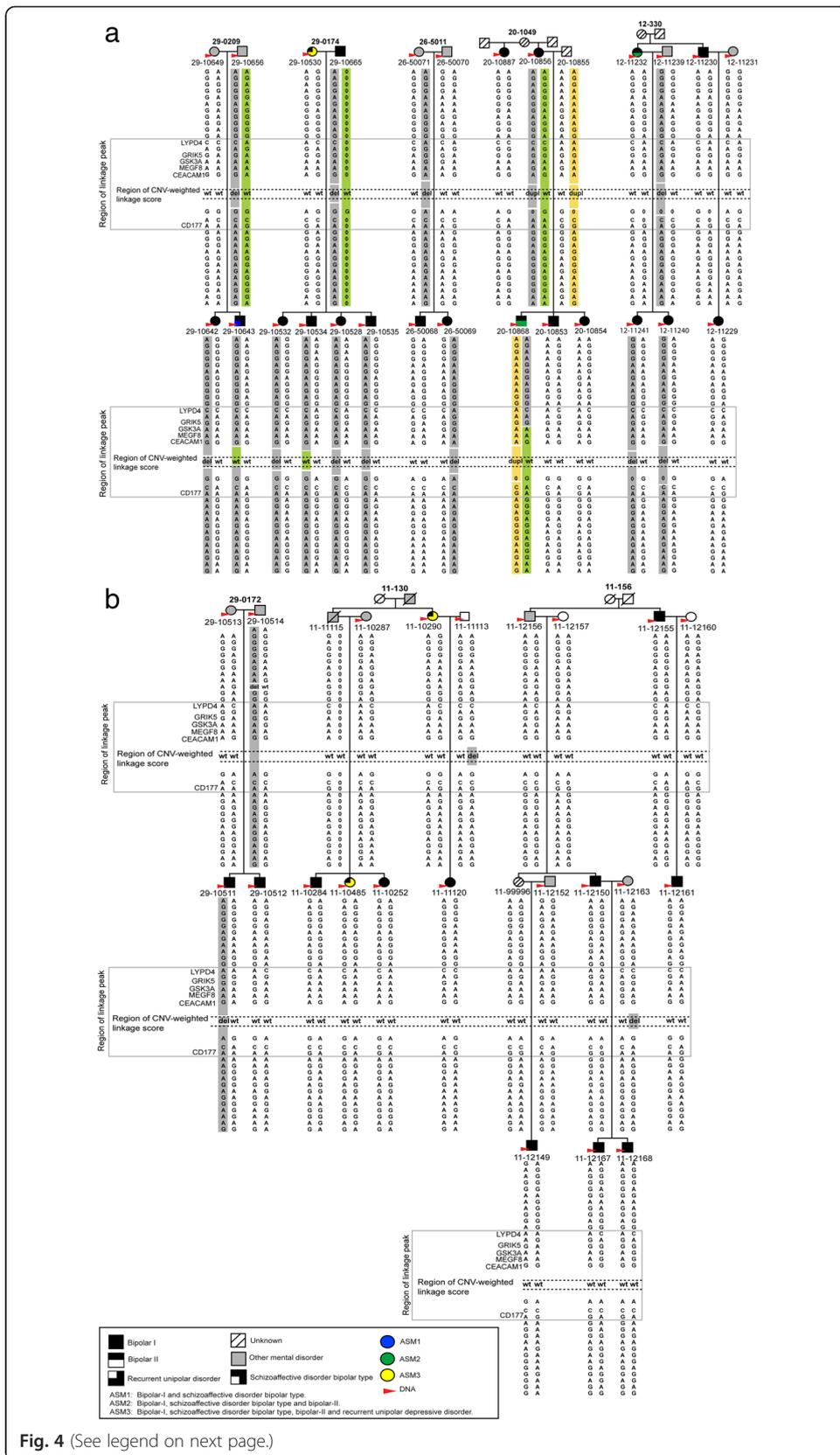
Since the CNV-weighted linkage score analysis did not require a positive CNV status only among those affected, according to the different affection status models, we next attempted to find the implications of structural variations in the PSG genes for risk to BPD. In particular, we wanted to found if the CNVs are *de novo* or inherited by performing a haplotype analysis in the families that contributed to the CNV-weighted linkage peak (Fig. 4a).

In pedigree 29–0174 we observed a deletion in three affected children, but not in the fourth affected child (29–10534) which otherwise has inherited the same haplotype as his siblings from the father for whom we did not have any DNA sample. This indicates either that there is a double crossover in this region or that we missed the detection of



a CNV in the child or that the father has two identical haplotypes with the exception of the presence of the CNV.

Similarly, in pedigree 29-0209 the son (29-10643) has inherited the same haplotype as his sister from the father, but no CNV was detected despite being present in both the father and the sister. This could be explained either by a double crossover or by us failing to detect the CNV. In family 26-5011 the haplotype with the CNV was only transmitted to one of the affected children. In family 20-1049 the CNV was transmitted from the untyped father (20-10855) to a child with diagnosis under ASM2 but not to the sibling with diagnosis under ASM1.



(See figure on previous page.)

Fig. 4 Haplotype analysis of families with CNVs in 19q13. Phasing analysis of genotypes to generate the most likely haplotype in pedigrees with CNVs on chromosome 19q13, was carried out with the GENEHUNTER software. Forty-seven markers representing all available markers in this region, spanning a region of 17.04 Mb, were included for the haplotype analysis. To simplify illustration of results, flanking markers were removed and only genotypes for 33 markers most proximal to the CNV are depicted, mapping a 6.05 Mb region. The linkage peak region is marked with a gray window and spans 1.5 Mb. The region with the two adjacent significant CNV-weighted linkage scores (91,215 bp in size) is illustrated with a gray dashed line. CNVs of duplication are denoted 'dupl', deletions are denoted 'del' and the normal state (wild type) are denoted 'wt'. Haplotypes are displayed in colors (only for relevant chromosomes) to illustrate inheritance of gain/loss of genomic segments. The relative genomic region of each CNV is illustrated by separate colored segments. Of note, CNV calling was made based on a complete set of non-QC filtered sample of both monomorphic and polymorphic probes whereas analysis of the haplotypes was made using QC filtered polymorphic probes only. In order to retrieve recombinant mapping of high resolution, all available SNP-markers located within the linkage peak region were included. Representative gene-id's are displayed. All genomic coordinates are according to NCBI36/hg18.

a Results of the initial analysis of CNV-weighted linkage scores with 5 pedigrees consisting of 12 individuals with a shared CNV. In pedigree 29–0174 no DNA was available for individual 29–10665 who was therefore excluded from the initial CNV-weighted analysis. The CNV status for this individual was revealed in the subsequent phased haplotype analysis. Moreover, in pedigree 20–1049 the CNV-carriership in 20–10855 was detected using the subsequent phased haplotype analysis. **b** Results of the extended analysis to find undetected CNVs. In our first attempt to identify undetected CNVs in this region we manually checked the CNV calling and identified 3 individuals with deletions, 11–11113, 11–112163 and 29–10511, and individual 29–10514 with a deletion in the adjacent region. Finally, a phased haplotype analysis indicates that the CNV in 29–10511 is a *de novo* event and that no transmission of CNVs occurs in the pedigrees 11–130 and 11–156. This analysis further indicates deletions in 29–10665 and 20–10855

This prompted us to identify plausible undetected CNVs, by investigating non-QC data in the entire sample in this region. We observed deletions for three individuals (11–11113, 11–12163 and 29–10511) in three additional pedigrees (11–130, 11–156 and 29–0172) (Fig. 4b). We further observed that individual 29–10514 has a deletion adjacent to the region of interest. A haplotype analysis of these three pedigrees revealed a *de novo* CNV in pedigree 20–0172 for individual 29–10511 and deletions in families 11–130 and 11–156 that were not transmitted (Fig. 4b) and therefore not included in our CNV-weighted linkage analysis. From the haplotype analysis it seems unlikely that we have missed to detect CNVs in this region in other family members of these three families.

Discussion

Linkage analysis

Our genome-scan using 20,123 markers resulted in a high resolution mapping to identify narrow regions linked to BPD. We performed QC analyses aiming to reduce influences from markers located in CNV regions and selected informative markers using an independent study population. We estimated thresholds for suggestive linkage through simulation according to a well established criteria of Lander and Krygluak 1995 [29] to obtain robust linkage data. Four loci reached suggestive evidence of linkage to BPD.

The most notable linkage was to chromosome 3p14.1. This signal reached suggestive linkage in both the non-parametric and parametric dominant models under ASM3, and is consistent with several previous reports of linkage to BPD [30–32]. Several candidate genes related with synaptic and other functions of relevance to BPD susceptibility reside in this region [32]. Of note, one of the suggested genes, *MAGII*, has recently been reported [33] to harbor large structural variation (CNV) in the same BP-pedigree sample

as was used in the present study. The *MAGII* gene polymorphism was further validated in an independent sample of unrelated bipolar, schizophrenia and schizoaffective disorder cases, and thus, represents an interesting candidate for future studies. The synaptic function gene, *MAGII*, has also been proposed as a putative candidate in a linkage and expression analysis by Lopez de Lara 2010 [32]. However, it is unlikely that the *MAGII* CNV is the major reason for the observed linkage at 3p14 in our sample since the CNV reported by Karlsson et al. [33] was only observed in two (11–158 and 11–130) of the 46 BP-pedigrees used in both studies. This argument is supported by the notion that under the parametric linkage 29 % of the pedigrees are estimated to be linked to this region (see Table 2). The non-parametric linkage obtained without pedigree 11–158 and 11–130 was as high as 1.56, while it was 1.16 in these two families alone.

Our suggestive linkage to 1q23 in the parametric recessive analysis is consistent with the model-free linkage analyses reported previously [34, 35]. This locus has also been linked to schizophrenia [36] which further supports a shared genetic vulnerability between schizophrenia (SZ) and BPD. There have been several reports of linkage to 10q26.2 [31, 37, 38] where we detected suggestive linkage for the recessive model.

Although chromosome 6 has been a focus for a BPD risk locus [39, 40], none of these regions overlap with 6q16 reported in our study.

The present study was aimed for a CNV-weighted analysis and apparently several pedigrees in our sample were not optimal for a linkage analysis. Since the number and sizes of pedigrees were small, the number of meioses was limited, leading both to low power and reducing the ability to pinpoint linkage to a small region. Of note, the same BP-pedigrees have been included in previous linkage scans of both parametric and non-parametric models [41] without yielding suggestive linkage in regions overlapping with those identified in the present study. This notion underlines the profoundly heterogenous background of BPD and motivates methods attempting to search for shared segregating risk loci in a more homogenous sample than in the present study.

CNV-weighted linkage analysis

Based on 2806 CNV segments, which were all found in the general population, the parametric dominant ASM1 model identified two CNV-weighted linkage scores on 19q13 that remained significant after a 1000-fold simulation including a FWER correction (empirical $P = 0.033$). We did not adjust thresholds for significance for the CNV-weighted linkage analyses according to all tested models. Our CNV-weighted linkage score on 19q13 would not be significant if all nine models were accounted for according to Bonferroni's approach. However, a Bonferroni correction would be too conservative as the models are correlated (see Fig. 2). Our algorithm identified 12 CNV-carriers in 5 families that contributed to this signal (Fig. 4a). The CNV polymorphism is located in a region with frequent structural variations and harbors a gene family encoding the pregnancy-specific glycoprotein (*PSG*) genes (Fig. 3).

In a series of subsequent analyses we aimed to find possible implications of structural variation in the *PSG* gene in vulnerability to BPD as well as to unravel if the CNVs are inherited or *de novo*. To do this we analyzed phased haplotypes in the 19q13 region and manually checked the CNV calling prior to QC filtering. We found that in the 5 BP-pedigrees contributing to the CNV-weighted linkage score on 19q13 under the

ASM1 model, 9 of 17 were affected (BP-I) but also that 5 of 15 unaffected (i.e. not classified BP-I) individuals were positive for the CNV, that is, they had a deletion (del) or duplication (dup) of a chromosomal region that included the CNV-weighted linkage peak. This indicates that if the CNV is functional in causing risk for BPD it still has incomplete penetrance, whereas some individuals got BPD due to other reasons.

This observation can also be construed as the CNV having no involvement in the etiology of BPD. Although this cannot be entirely ruled out, the counter-argument that nonetheless underscores the involvement of the CNV polymorphism in 19q13 in the etiology of BPD concerns issues of disease classification. In spite of the fact that this polymorphism occurs in individuals without disease classification according to ASM1, they are not classified as never mentally ill. The fact that the CNV deletions and duplications are not entirely overlapping are most likely due our CNV calling algorithm based on SNPs, as described in the Additional file 1. Of importance, none of the suggestive linkage signals or significant CNV-weighted linkage scores occurred in regions in which families were selected (Additional file 2).

Possible candidate for BPD in 19q13?

Of interest, our result with a significant signal on 19q13 for BPD susceptibility agrees with previous reports.

First, without implicating a specific gene, Francks et al. [42] detected linkage of 19q13 to both SZ and BPD. The 19q13 locus harbors several putative candidate risk genes for BPD, e.g. the glycogen synthase kinase 3- α (*GSK3A*) gene and the glutamate receptor, ionotropic kainate 5 (*GRIK5*) gene. The *GSK3A* protein is homologous to *GSK3B*, a target molecule for lithium treatment [43], which has regulatory functions on proteins with a reported role in BPD susceptibility [44]. Several genes pertaining to the glutamate system have consistently been associated with BPD [39]. Taken together, there is a strong support for both the glutamate and the cell growth-maintenance related genes in BPD etiology.

Secondly, Alkelai and collaborators [45] found the *CEACAM21* gene in the 19q13 region to associate with SZ. The *CEACAM* genes, or the carcinoembryonic antigen-related cell adhesion (*CEA*) gene family, have several structural and functional similarities to the *PSG* genes (OMIM: 109770). Recent results show that the *CEA* genes are both brain and immune cell expressed (<http://genome.ucsc.edu/>). Although the exact molecular function of these gene products remains elusive, some studies reveal function related to cell-cell adhesion, innate immune system and signal modulation in various tissues [46–48]. These lines of results, in parallel to the enrichment of immune system genes among those associated or linked to BPD [49, 50] suggest a possible role for the *PSG* gene in BPD etiology.

Other candidates in this linkage peak region include genes related to neurotrophin [39] and immune systems [51] (Fig. 3), which is interesting, given that such mechanisms have been proposed to be involved in the etiology of BPD and other psychiatric disorders.

The mechanisms of the clinical manifestations and phenotypic effects of CNVs are well documented [52] and include alteration of gene dosage, truncated protein or positional effects. The latter include a transcription that may be directly controlled by promoters in the CNV or by alteration of chromatin structure [53]. Of note, regulatory elements have been identified as far away as 2 Mb from the transcription unit [52, 53].

Based on these observations we made a bioinformatic search (<http://regulome.stanford.edu/>) and screened all available markers in the CNV region for being transcription factor binding sites. Two markers (rs4802370 and rs7252967) are likely transcription factor binding sites and linked to expression of the F-box binding gene (*FBXO30*). The F-box protein functions as an ubiquitin-ligase and targets the transcription factor *NF- κ B* [54]. Of interest, the *NF- κ B* pathway has been shown to be a key regulator of neuroplasticity, neuronal survival and pro-inflammatory status and thus serves as a one of many etiology correlates to BPD [55, 56].

In a previous study, Karlsson et al. 2012 [33], identified a rare and highly penetrant CNV which map in the *MAGI1* gene. The Karlsson et al. study analyzed the identical dataset as was used in the present study and reported the *MAGI1*-CNV in the pedigrees 11–158 and 11–130. Although the polymorphism in the *MAGI1* gene in pedigrees 11–158 and 11–130, reported by Karlsson et al. 2012 [33], was ranked among the 25 strongest candidates in the CNV-weighted linkage analysis it was not the strongest observation (Additional file 7). This is explained by the non-overlapping CNV segments between these two families.

Limitations

Although our approach was successful in reducing genetic heterogeneity and evaluating linkage restricted to regions with shared CNVs there are several limitations in this study.

Firstly, CNVs were called if they were longer than 10 kb. Several lines of evidence suggest that CNVs shorter than 100 kb are less consistent using SNP-array CNV calling [57]. Of note, shorter structural variations have gained a great deal of attention for their role in complex disorders [13, 58]. Thus, for the purpose of not rejecting *a priori* true positive CNV segments with a putative role for BPD-vulnerability, shorter segments were allowed in this study. The CNVs in 19q13 were all longer than 100 kb, except for one CNV of length 91,215 bp.

Secondly, properties of the algorithm and definition of shared CNVs have consequences for the final results. It can be argued that sharing of CNVs between more than two individuals in the same family should be used for identification of CNV with major implications on susceptibility for BPD. The limited pedigree size prompted us to set this threshold at two individuals.

Thirdly, the CNV-weighted LOD score method has the potential to highlight regions where presence of a segregating CNV in pedigrees correlates with a higher family-wise LOD score. Since this approach simply sums LOD scores across families with at least two CNVs, but otherwise do not weight the family LOD scores by the CNV frequencies, CNVs which are strongly correlated with high LOD scores in a small fraction of pedigrees would rank lower in comparison to what would be seen for less strongly correlated CNVs found in a large proportion of pedigrees. It would therefore be of interest in future studies to weight the family linkage scores in different ways in terms of their family members CNV. In fact, CNV-weighted linkage analysis is an instance of combined association and linkage analysis, for which different methods and score functions have been proposed, see for instance [59–61] and references herein.

Fourthly, in the present study we set the threshold for defining a shared CNV as being strict overlapping. It is unclear whether this categorization is necessary for it to

contribute to increased risk for disease [12, 62, 63]. Thus, our study design may possibly have led to rejection of putative disease causing CNVs.

Lastly, although the PennCNV algorithm reports CNVs with a high power and at a low false positive rate [64], we cannot exclude that we missed to detect CNVs in our dataset.

In summary, our study provides statistical evidence that a region on 19q13 could be tied to BPD, which raises the possibility that this region confers risk to BPD for a subsample of individuals. Still, these results are as yet inconclusive for a specific candidate gene. Further studies are needed in independent samples in order to confirm the involvement of CNVs at 19q13 in BPD susceptibility and to understand the molecular consequences of such a CNV. Nonetheless, we conclude that our CNV-weighted linkage approach is a useful tool for future studies, attempting to address the role of larger structural variants in multifactorial diseases such as BPD.

Additional files

Additional file 1: Supplementary materials. (DOC 43 kb)

Additional file 2: Selection criteria for inclusion of BP-pedigrees. Data obtained from a large number of pedigrees from NIMH Genetic Initiative Wave 1–4 was screened to generate informative pedigrees and to reduce any sporadic and environmental form of BPD. Two different analyses were used to select pedigrees. Test for runs of homozygosity (ROH) and analyzing regions with increased parametric family LOD scores under different assumed modes of inheritance. (DOC 62 kb)

Additional file 3: Pedigrees included in linkage and CNV weighted analysis for BPD. To retrieve informative pedigrees from the NIMH Genetic Initiative Wave 1–4 sample, 46 pedigrees consisting of 269 genotyped individuals and 97 individuals with no available DNA were selected. Three affection status models were considered (ASM1-3) based on the different bipolar affective disorder sub-types, described in the figure-box. Individuals with a diagnosis of bipolar spectrum disorders which only apply to a certain ASM were coded as “unknown” under the other ASMs. In order to illustrate full details of pedigree composition different subtypes of bipolar spectrum disorders are illustrated. For linkage analyses with different ASMs only genotyped individuals were coded as ASM1-3. All other individuals, irrespective of diagnosis, were coded as ‘unknown’. (PDF 364 kb)

Additional file 4: Results of the parametric and non-parametric genome-wide linkage analyses under affection status models ASM1-3. Simulation based thresholds for genome-wide significant and suggestive linkage levels are illustrated in red and black dashed lines respectively. A-C: Non-parametric linkage analyses. NPL Z scores are illustrated. D-F: Parametric dominant linkage analyses. Heterogeneity LOD scores (HLOD) are illustrated. G-I: Parametric recessive linkage analyses. Heterogeneity LOD scores (HLOD) are illustrated. (PDF 367 kb)

Additional file 5: Principles for region for which one CNV-weighted linkage score is calculated. LOD scores from two families (family ID: 20–1049 and 12–330) for a certain chromosome are depicted. The occurrence of CNVs for individuals within these families is also illustrated. The average linkage score (parametric LOD score or non-parametric Z-score) is calculated over the region with overlapping CNV from at least two individuals in the same family. Note, for those individuals with the presence of a CNV were SNPs zeroed out. In the non-CNV carriers were genotypes intact. The average linkage score from all families with overlapping CNV in at least two individuals is added. A CNV-weighted linkage score is thus generated for a defined region that share overlapping CNVs for more than 1 individual per family. To illustrate the calculation; an example is given with three markers that are located in a region with four overlapping CNVs, at the position 60–75 cM. For these 3 markers, the LOD scores for the two families are: Fam-ID 12–330: 0.9, 0.9 and 0.9, Fam-ID: 20–1049: 0.2, 0.4 and 0.1. The CNV-weighted linkage score (the sum of average linkage score) is then calculated as: Fam-ID 12–330: $(0.9 + 0.9 + 0.9)/3 = 0.9$, Fam-ID: 20–1049: $(0.2 + 0.4 + 0.1)/3 = 0.23$. Then, a total CNV-weighted linkage score = $0.9 + 0.23 = 1.13$. (PDF 179 kb)

Additional file 6: Genomic positions of CNVs identified in 19q13. The table shows the positions of all CNVs that were identified in 19q13 (chr19:48066441–48157656) and that harbors a significant CNV-weighted linkage score. All genomic coordinates are according to NCBI36/hg18. (DOC 38 kb)

Additional file 7: Table of the 25 strongest CNV-weighted linkage scores of the parametric dominant ASM1 model. The two strongest signals withstood test for significance using a 1000-fold simulation analysis including a FWER correction. Asterisk (*) denotes a significant value. Start and end positions are shown for the genomic regions that were identified to contain a shared CNV within (and across) family members. The full length of the CNV is also presented for each individual. Affection status is given for each individual, with BP-I = bipolar type I, BP-II = bipolar type II and OMD = other mental disorder. Of note, different ASMs were assumed, and affection status under a certain model implies a non-diseased status under the remaining ones. The copy number variation state is deletion or gain, where CN = 1 refers to 1 copy deletion and CN = 3 refers to 1 copy gain. All genomic coordinates are according to NCBI36/hg18. (DOC 326 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ML, OH and IK are responsible for (i) analysis plan and study design (ii) calculations and statistical analyses (iii) development of logistic regression model (iv) interpretation and conclusion of results and (v) writing of manuscript. RK, ML and LG are responsible for selection of the study sample. RK is responsible for development of algorithm of the CNV-weighted linkage scores. None of the authors have taken part of the sampling procedures of the NIMH Genetic Initiative BP-families, which is easily explained by the fact that this study is based on a biobank material. These samples are available upon request and after qualification for being authorized investigators. DNA samples were obtained from the cell and DNA repository at Rutgers University (Camden, New Jersey). Genotyping was performed using the Illumina Human 610quad chip at the SNP Technology Platform, Uppsala University, Sweden. All authors read and approved the final manuscript.

Acknowledgements

The authors gratefully acknowledge Silvia Paddock for the planning and design of this study. We also thank Silvia for all critical discussions throughout this project and her comments on this manuscript. The authors gratefully acknowledge Dr. Kai Wang (University of Southern California) for the inspirational and critical discussions in analysis of CNV. Ola Hössjer was financially supported by the Gustafsson Foundation for Research in Natural Sciences and Medicine.

This work was supported by the Swedish Research Council (Reference Number 2005–6929 and L2010-61x-20377-04-3), the Swedish Foundation for Strategic Research (Reference Number A3 05:207b), National Alliance for Research on Schizophrenia and Depression (SP), Karolinska Institutet, and the Center for Gender Medicine at Karolinska Institutet. The Single Nucleotide Polymorphism Technology Platform, Uppsala, Sweden is supported by Uppsala University, Uppsala University hospital, the Swedish Council for Research Infrastructure, and the Knut and Alice Wallenberg Foundation.

Data and biomaterials were collected in four projects that participated in the National Institute of Mental Health (NIMH) Bipolar Disorder Genetics Initiative. From 1991 to 1998, the Principal Investigators and Co-Investigators were Indiana University, Indianapolis, Indiana, U01 MH46282, John Nurnberger, M.D., Ph.D., Marvin Miller, M.D., and Elizabeth Bowman, M.D.; Washington University, St. Louis, Missouri, U01 MH46280, Theodore Reich, M.D., Allison Goate, Ph.D., and John Rice, Ph.D.; Johns Hopkins University, Baltimore, Maryland, U01 MH46274, J. Raymond DePaulo, Jr., M.D., Sylvia Simpson, M.D., M.P.H., and Colin Stine, Ph.D.; NIMH Intramural Research Program, Clinical Neurogenetics Branch, Bethesda, Maryland, Elliot Gershon, M.D., Diane Kazuba, B.A., and Elizabeth Maxwell, M.S.W. Data and biomaterials were collected as part of 10 projects that participated in the NIMH Bipolar Disorder Genetics Initiative. From 1999 to 2003, the Principal Investigators and Co-Investigators were Indiana University, Indianapolis, Indiana, R01 MH59545, John Nurnberger, M.D., Ph.D., Marvin J. Miller, M.D., Elizabeth S. Bowman, M.D., N. Leela Rau, M.D., P. Ryan Moe, M.D., Nalini Samavedy, M.D., Rif El-Mallakh, M.D. (at University of Louisville), Hussein Manji, M.D. (at Wayne State University), Debra A. Glitz, M.D. (at Wayne State University), Eric T. Meyer, M.S., Carrie Smiley, R.N., Tatiana Foroud, Ph.D., Leah Flury, M.S., Danielle M. Dick, Ph.D., and Howard Edenberg, Ph.D.; Washington University, St. Louis, Missouri, R01 MH059534, John Rice, Ph.D., Theodore Reich, M.D., Allison Goate, Ph.D., and Laura Bierut, M.D.; Johns Hopkins University, Baltimore, Maryland, R01 MH59533, Melvin McInnis, M.D., J. Raymond DePaulo, Jr., M.D., Dean F. MacKinnon, M.D., Francis M. Mondimore, M.D., James B. Potash, M.D., Peter P. Zandi, Ph.D., Dimitrios Avramopoulos, and Jennifer Payne; University of Pennsylvania, Philadelphia, Pennsylvania, R01 MH59553, Wade Berrettini, M.D., Ph.D.; University of California at Irvine, Irvine, California, R01 MH60068, William Byerley, M.D., and Mark Vawter, M.D.; University of Iowa, Iowa City, Iowa, R01 MH059548, William Coryell, M.D., and Raymond Crowe, M.D.; University of Chicago, Chicago, Illinois, R01 MH59535, Elliot Gershon, M.D., Judith Badner, Ph.D., Francis McMahon, M.D., Chunyu Liu, Ph.D., Alan Sanders, M.D., Maria Caserta, Steven Dinwiddie, M.D., Tu Nguyen, and Donna Harakal; University of California at San Diego, San Diego, California, R01 MH59567, John Kelsoe, M.D., and Rebecca McKinney, B.A.; Rush University, Chicago, Illinois, R01 MH059556, William Scheftner, M.D., Howard M. Kravitz, D.O., M.P.H., Diana Marta, B.S., Annette Vaughn-Brown, M.S.N., R.N., and Laurie Bederow, M.A.; NIMH Intramural Research Program, Bethesda, Maryland, 1Z01MH002810-01, Francis J. McMahon, M.D., Layla Kassem, PhD, Sevilla Detera-Wadleigh, Ph.D, Lisa Austin, Ph.D, and Dennis L. Murphy, M.D. Genotyping services were provided by the Center for Inherited Disease Research. The Center for Inherited Disease Research is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, Contract Number N01-HG-65403.

Author details

¹Neuroimmunology Unit, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. ²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ³Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden. ⁴Department of Mathematics, Division of Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden.

Received: 25 August 2015 Accepted: 7 December 2015

Published online: 18 December 2015

References

- Hirschfeld RM, Vornik LA. Bipolar disorder—costs and comorbidity. *Am J Manag Care*. 2005;11:S85–90.
- Merikangas KR, Jin R, He JP, Kessler RC, Lee S, Sampson NA, et al. Prevalence and correlates of bipolar spectrum disorder in the world mental health survey initiative. *Arch Gen Psychiatry*. 2011;68:241–51.
- Mameros A, Angst J. *Bipolar Disorders, 100 years after manic-depressive insanity*. New York: Kluwer Academic Publishers; 2002.
- Smoller JW, Finn CT. Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet*. 2003;123C:48–58.
- Schulze TG, Akula N, Breuer R, Steele J, Nalls MA, Singleton AB, et al. Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *World J Biol Psychiatry*. 2012;15:200–8.
- Craddock N, Sklar P. Genetics of bipolar disorder. *Lancet*. 2013;381:1654–62.

7. Psychiatric GCBDWG. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet.* 2011;43:977–83.
8. Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X, Strohmaier J, et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet.* 2011;88:372–81.
9. Psychiatric Genomic Consortium [<http://www.med.unc.edu/pgc/>]
10. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nat Rev Genet.* 2012;13:537–51.
11. Alaerts M, Del-Favero J. Searching genetic risk factors for schizophrenia and bipolar disorder: Learn from the past and back to the future. *Hum Mutat.* 2009;30:1139–52.
12. Cook Jr EH, Scherer SW. Copy-number variations associated with neuropsychiatric conditions. *Nature.* 2008;455:919–23.
13. Lee JA, Lupski JR. Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron.* 2006;52:103–21.
14. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature.* 2006;444:444–54.
15. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell.* 2012;148:1223–41.
16. Hiroi N, Takahashi T, Hishimoto A, Izumi T, Boku S, Hiramoto T. Copy number variation at 22q11.2: from rare variants to common mechanisms of developmental neuropsychiatric disorders. *Mol Psychiatry.* 2013;18:1153–65.
17. Barnett JH, Smoller JW. The genetics of bipolar disorder. *Neuroscience.* 2009;164:331–43.
18. Nurnberger Jr, DePaulo JR, Gershon ES, Reich T, Blehar MC, Edenberg HJ, et al. Genomic survey of bipolar illness in the NIMH genetics initiative pedigrees: A preliminary report. *Am J Med Genet.* 1997;74:227–37.
19. NIMH Genetic Initiative [<http://www.nimhgenetics.org>]
20. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet.* 1997;61:1179–88.
21. Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T, et al. Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol Psychiatry.* 2009;14:376–80.
22. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
23. Database of Genomic Variants [<http://dgv.tcag.ca/dgv/app/home;build36/hg18>]
24. Welcome Trust Sanger Institute [http://sanger.ac.uk/research/areas/humangenetics/cnv/highres_discovery.html;NCBI36/hg18]
25. Children's Hospital of Philadelphia [<http://cnv.chop.edu;hg18/build36.1>]
26. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002;30:97–101.
27. Whittemore AS, Halpern J. A class of tests for linkage using affected pedigree members. *Biometrics.* 1994;50:118–27.
28. Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL, et al. Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. *Am J Hum Genet.* 2004;74:403–17.
29. Lander E, Kruglyak L. Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nat Genet.* 1995;11:241–7.
30. Etain B, Mathieu F, Rietschel M, Maier W, Albus M, McKeon P, et al. Genome-wide scan for genes involved in bipolar affective disorder in 70 European families ascertained through a bipolar type I early-onset proband: Supportive evidence for linkage at 3p14. *Mol Psychiatry.* 2006;11:685–94.
31. Cichon S, Schumacher J, Muller DJ, Hurter M, Windemuth C, Strauch K, et al. A genome screen for genes predisposing to bipolar affective disorder detects a new susceptibility locus on 8q. *Hum Mol Genet.* 2001;10:2933–44.
32. Lopez de Lara C, Jaitovich-Groisman I, Cruceanu C, Mamdani F, Lebel V, Yerko V, et al. Implication of synapse-related genes in bipolar disorder by linkage and gene expression analyses. *Int J Neuropsychopharmacol.* 2010;13:1397–410.
33. Karlsson R, Graae L, Lekman M, Wang D, Favis R, Axelsson T, et al. MAG11 copy number variation in bipolar affective disorder and schizophrenia. *Biol Psychiatry.* 2012;71:922–30.
34. Fallin MD, Lasseter VK, Wolyniec PS, McGrath JA, Nestadt G, Valle D, et al. Genomewide linkage scan for bipolar-disorder susceptibility loci among Ashkenazi Jewish families. *Am J Hum Genet.* 2004;75:204–19.
35. Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, et al. A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc Natl Acad Sci U S A.* 1999;96:5604–9.
36. Kohn Y, Lerer B. Genetics of schizophrenia: A review of linkage findings. *Isr J Psychiatry Relat Sci.* 2002;39:340–51.
37. Ewald H, Flint T, Kruse TA, Mors O. A genome-wide scan shows significant linkage between bipolar disorder and chromosome 12q24.3 and suggestive linkage to chromosomes 1p22-21, 4p16, 6q14-22, 10q26 and 16p13.3. *Mol Psychiatry.* 2002;7:734–44.
38. Kelsoe JR, Spence MA, Loetscher E, Foguet M, Sadovnick AD, Remick RA, et al. A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. *Proc Natl Acad Sci U S A.* 2001;98:585–90.
39. Serretti A, Mandelli L. The genetics of bipolar disorder: genome 'hot regions', genes, new potential candidates and future directions. *Mol Psychiatry.* 2008;13:742–71.
40. Baron M. Manic-depression genes and the new millennium: Poised for discovery. *Mol Psychiatry.* 2002;7:342–58.
41. Ross J, Berrettini W, Coryell W, Gershon ES, Badner JA, Kelsoe JR, et al. Genome-wide parametric linkage analyses of 644 bipolar pedigrees suggest susceptibility loci at chromosomes 16 and 20. *Psychiatr Genet.* 2008;18:191–8.
42. Francks C, Tozzi F, Farmer A, Vincent JB, Rujescu D, St Clair D, et al. Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13. *Mol Psychiatry.* 2010;15:319–25.
43. O'Brien WT, Klein PS. Validating GSK3 as an in vivo target of lithium action. *Biochem Soc Trans.* 2009;37:1133–8.
44. Li X, Jope RS. Is glycogen synthase kinase-3 a central modulator in mood regulation? *Neuropsychopharmacology.* 2010;35:2143–54.

45. Alkelai A, Lupoli S, Greenbaum L, Kohn Y, Kanyas-Sarner K, Ben-Asher E, et al. DOCK4 and CEACAM21 as novel schizophrenia candidate genes in the Jewish population. *Int J Neuropsychopharmacol*. 2012;15:459–69.
46. Pils S, Gerrard DT, Meyer A, Hauck CR. CEACAM3: An innate immune receptor directed against human-restricted bacterial pathogens. *Int J Med Microbiol*. 2008;298:553–60.
47. Thompson JA, Grunert F, Zimmermann W. Carcinoembryonic antigen gene family: Molecular biology and clinical perspectives. *J Clin Lab Anal*. 1991;5:344–66.
48. Streydio C, Swillens S, Georges M, Szpirer C, Vassart G. Structure, evolution and chromosomal localization of the human pregnancy-specific beta 1 glycoprotein gene family. *Genomics*. 1990;6:579–92.
49. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry*. 2011;16:286–92.
50. Jones KA, Thomsen C. The role of the innate immune system in psychiatric disorders. *Mol Cell Neurosci*. 2013;53:52–62.
51. Eaton WW, Pedersen MG, Nielsen PR, Mortensen PB. Autoimmune diseases, bipolar disorder, and non-affective psychosis. *Bipolar Disord*. 2010;12:638–46.
52. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437–55.
53. Kleinjan DA, van Heyningen V. Long-range control of gene expression: Emerging mechanisms and disruption in disease. *Am J Hum Genet*. 2005;76:8–32.
54. Arabi A, Ullah K, Branca RM, Johansson J, Bandarra D, Haneklaus M, et al. Proteomic screen reveals Fbw7 as a modulator of the NF-kappaB pathway. *Nat Commun*. 2012;3:976.
55. Brietzke E, Kapczinski F. TNF-alpha as a molecular target in bipolar disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. 2008;32:1355–61.
56. Barbosa IG, Nogueira CR, Rocha NP, Queiroz AL, Vago JP, Tavares LP, et al. Altered intracellular signaling cascades in peripheral blood mononuclear cells from BD patients. *J Psychiatr Res*. 2013;47:1949–54.
57. Zhang D, Qian Y, Akula N, Alliey-Rodriguez N, Tang J, Bipolar Genome S, et al. Accuracy of CNV Detection from GWAS Data. *PLoS One*. 2011;6:e14511.
58. Lachman HM, Pedrosa E, Petruolo OA, Cockerham M, Papolos A, Novak T, et al. Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B:259–65.
59. Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design. *Hum Biol*. 2000;72:63–132.
60. Perez-Enciso M. Fine mapping of complex trait genes combining pedigree and linkage disequilibrium information: a Bayesian unified framework. *Genetics*. 2003;163:1497–510.
61. Hossjer O. Combined association and linkage analysis for general pedigrees and genetic models. *Stat Appl Genet Mol Biol*. 2005;4:Article11.
62. Lupski JR. Genomic rearrangements and sporadic disease. *Nat Genet*. 2007;39:543–47.
63. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315:848–53.
64. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ. Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res*. 2010;38:e105.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

