Yvon C. Chagnon[a], Michel Maziade[b, c,*], Thomas Paccalet[b,d], Jordie Croteau[b], Alain Fournier [b], Marc-André Roy[b, c], Alexandre Bureau[b,e].

A multimodal attempt to follow-up linkage regions using RNA expression, SNPs and CpG methylation in schizophrenia and bipolar disorder kindreds.

Running title: Multi-omics in schizophrenia and bipolar disorder

Supplementary material

Affiliations : [a] École de Psychologie, FESS, Université Laval, Québec, G1V 0A6 Canada; [b] Centre de recherche CERVO, Centre intégré universitaire de santé et des services sociaux de la Capitale-Nationale, Québec, G1J 2G3, Canada; [c] Université Laval, département de psychiatrie et neurosciences, Québec, G1V 0A6, Canada; [d] Université du Québec à Trois-Rivières, département de Sciences infirmières, Trois-Rivières, G9A 5H7, Canada and [e] Université Laval, département de médecine sociale et préventive, Québec, G1V 0A6, Canada.

Supplement 1

Linkage analysis. We performed a joint analysis of sample 1 on which we reported a genome-wide linkage scan (1) and sample 2 on which we reported replication of linkage signals in 16p with BP (2) and in 13q13-q14 with CL (3). Compared to these previous reports, 273 additional family members were recruited, among which 191 have been genotyped. The total sample now reaches 1547 subjects, including 1161 with genotype data. Diagnoses have been updated, including 20 incident cases of BP for a total of 225. The number of SZ cases remains 136, and the total number of affected subjects in the CL phenotype definition including SZ, BP and SAD reaches 404. We used an updated panel of microsatellite markers (N=730 in sample 1 and N=571 in sample 2) where mean spacing between markers was reduced between 5 to 10 megabases, and where all markers were genotyped in both samples. As in previous reports, parametric (or model-based) two-point linkage analyses were performed using both a dominant and a recessive mode of inheritance with the age-dependent penetrance values used in sample 1. The exact parameter values were provided in (4). Both models took into account the uncertainty of diagnosis by increasing the phenocopy rate in liability classes corresponding to "probable" or "possible" diagnoses (5). For each mode of inheritance, the analysis was first carried out with affected and unaffected subjects and then repeated by considering only the affected family members to allow for the presence of a few disease gene carriers who had not yet expressed the disorder (referred to as the "affected-only" affection-status type). Hence, for each phenotype definition, four models were tested resulting from the combination of the two modes of inheritance and affection-status types.

Gene expression measurement in immortalized lymphocytes. To evaluate RNA expression in immortalized lymphocytes (IML) in the 498 selected subjects we produced IML cultures that

were monitored to be harvested in exponential growth. Cells were grown in RPMI 1640 with glucose (Hyclone; Fisher) medium supplemented with (15%) fetal bovine serum (Wisent) at 37°C and 5% CO_2 (incubators Forma Scientific) in T25 Canted, Vented cap flask (Corning). Monitoring of the cell count was made with an hemacytometer. 10^7 cells were pelleted by centrifugation (316 g, 10 minutes, 4°C) and frozen at -80°C until use for RNA extraction that was made on affinity column (RNeasy Mini Kit; Qiagen) using a Qiacube robot (Qiagen). RNA quality (RNA integrity number or RIN) was evaluated on a Bioanalyser with RNA 6000 Nano Kit (Agilent). A RIN of 9 to 10 was usually observed and those RNA samples with a RIN below 7.0 were extracted again from a new cell culture. RNA quantification was made using Quant-iT™ RNA Assay Kit and a Qubit™ fluorometer (Life Technologies). RNA was kept at -80°C until used. About 1.5 ug of total RNA per sample at a concentration of 50 ng/ul was sent on dry ice (-80°C) to Genome Quebec (McGill, Montreal) and used to probe HumanHT-12 Expression Beachip v4 (47 K probes; Illumina).

Processing probe intensities. Background intensity level estimated using the background + signal model of Irizarry et al. (6) was subtracted from the intensity measurements and the resulting background-corrected intensities were quantile normalized across all chips using the Bioconductor MBCB package (www.bioconductor.org/packages/release/bioc/html/MBCB.html) with the "RMA" background correction and "quant" normalization options. This combination of methods was found to perform best among those implemented in the MBCB package by inspection of relative log expression (RLE) plots (7).

Controlling unwanted variation. Potential housekeeping genes were selected from a panel of 32 candidate genes (4398915C_7300_7500_Endogenous_Controls, Life Technologies). Among

the 63 probes in these potential housekeeping genes, we retained as housekeeping probes the 51 with little variation in expression across diagnosis groups (SZ, BP and NAR) as determined by a $p>0.10$ in an ANOVA of the log2-expression values of probes. All subsequent analyses were then conducted by including the first 10 principal components of the 51 housekeeping probes (following (7)), age at blood draw and sex as fixed covariates. In addition the Illumina slide and a polygenic effect were included as random effects in mixed models.

Supplementary results

Linkage analysis Table S1 shows the maximum LOD score in each region where a LOD score > 3.0 was reached, defining the boundaries of the linkage regions by the markers where the maximum LOD score had decreased by 2 units compared to the marker with the region-wide maximum LOD score. We also report a CL signal at the SCZD6 locus and an overlapping region linked to BP on 8p replicating previous linkage signals for SZ but also for BP (8) without reaching the 3.0 threshold. We also examined family-specific LOD scores to detect additional signals in regions where genetic heterogeneity reduces whole-sample LOD scores. Within a 5 Mb interval on 3q, we observed 2 families with a LOD score > 2.8 for CL, the only such interval over the genome. We defined the region where these family-specific LOD scores on 3q were > 2.0 as an additional linkage region. For our gene expression analysis we focused on these regions and the regions on 16p and 13q13-q14 defined by non-parametric linkage in previous reports ((2, 3) Table S1).

Supplementary tables and table legends

Table S1: Linkage analysis in the schizophrenia and bipolar disorder Eastern Quebec kindred sample

Chr	Pheno-type	Zmax[1]	Region start[2]		Maximal result		Region end[2]		Width
			Marker	Mb	Marker	Mb	Marker	Mb	Mb
3	CL	2.9[3],	D3S1616	107.6	D3S1267	138.4,	D3S1764	139.2	31.6
4	BP	3.5	D4S2431	174.8	D4S2417	180.2	D4S2951	183.2	8.3
6	SZ	**4.1**	D6S470	10.0	D6S334	16.1	D6S1959	19.9	9.9
8	BP	2.8	D8S515	39.6	D8S509	55.6	D8S166	57.4	17.8
8	CL	2.4	D8S136	22.4	D8S278	32.7	D8S1115	42.5	
13	CL	5.2[4]	D13S127	35.9	D13S1297	43.2	D13S119	58.5	22.6
13	SZ	3.3	D13S1253	40.1	D13S788	51.9	D13S1325	52.5	
15	BP	3.1	D15S122	26.0	D15S113	26.2	D15S165	31.3	5.3
16	BP	3.9[5]	D16S2619	13.6	D16S3060	15.8	D16S403	23.0	9.4
18	BP	**4.6**	D18S978	38.3	D18S1145	42.2	D18S851	50.1	11.8
							Total lenght		105.8

[1] Maximum two-point parametric LOD score. Scores > 4 are bolded.

[2] The beginning and end of the regions are defined by the marker where the maximum LOD score had decreased by 2 units compared to the marker with the region-wide maximum LOD score. Genomic positions are from GRC37.3.

[3] Zmax for the family yielding the highest signal instead of the whole kindred sample.

[4] $-\log_{10}$ p-value of the nonparametric linkage score reported in Maziade et al. (3) on smaller sample instead of Zmax.

[5] $-\log_{10}$ p-value of the nonparametric linkage score reported in Mérette et al. (2) on smaller sample instead of Zmax.

Table S2. Number of genes per linked regions according to Build GRCh38.p10, and expressed in immortalized lymphocytes (IMLs)

Linked regions	3q21	4p34	6p22	8p21	8p11	13q11	15q13	16p12	18q21
# genes 2051	539	63	149	319	213	346	101	199	122
#genes expressed 559	160	12	38	102	42	83	24	62	36
#probes expressed 686*	201	13	48	57*	128*	113	27	76	44
*: 21 probes overlap both regions									

Table S3 Results of the eQTN analysis for the common locus phenotype. SNP: Single Nucleotide Polymorphism reference sequence number; chr: chromosome; bp_build37: genomic position from GRC37.3; MAF: Minor Allele Frequency; gene: RefSeq gene or genes located within 50kb of the SNP; probes: gene expression probes for the gene or genes; b: regression coefficient estimate of the transcript model; b.var: estimated variance of the regression coefficient estimate of the transcript model; X2: Wald chi-square statistic (b^2/b.var); p: p-value of the Wald test; OR: odds ratio of the phenotype vs. NARs between the minor allele homozygous and major allele homozygous genotypes.

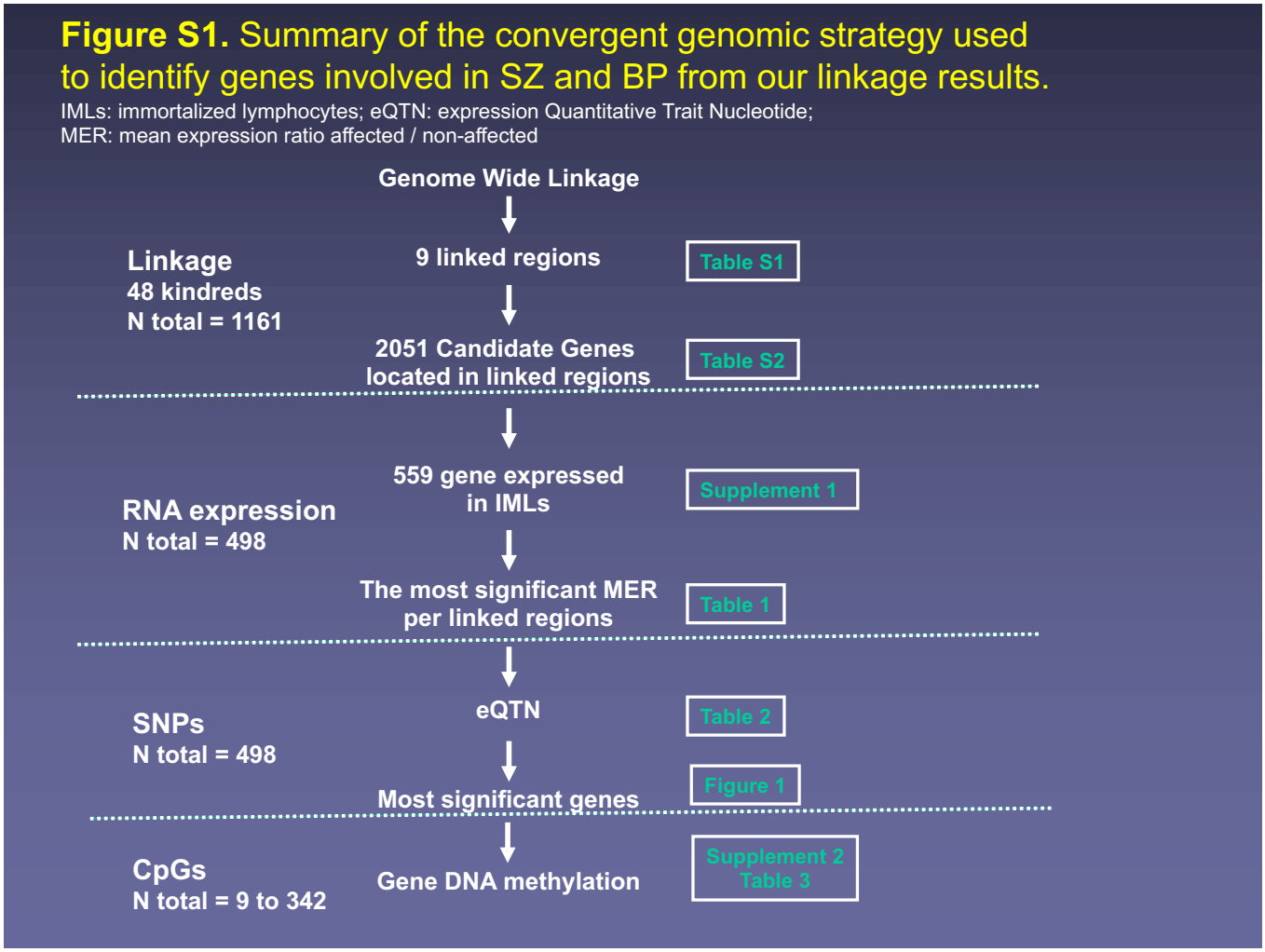
Table S4 Results of the eQTN analysis for the bipolar disorder phenotype. See legend of Table S3 for definition of the headers.

Table S5 Results of the eQTN analysis for the schizophrenia phenotype. See legend of Table S3 for definition of the headers.

Table S6 Overview of findings for highlighted genes

Gene	RNA differential expression	SNP associated with diagnostic through RNA expression	Differential methylation of CpG sites
FZD3	Best in region	Nearly gene-wide significant	Negative
ITGB5	Region-wide significant	Negative	Negative
SPCS3	Best in region	Gene-wide significant	Negative

Figure S1. Summary of the multi-modal genomic strategy used to identify genes involved in SZ and BP from our linkage results. IMLs: immortalized lymphocytes; eQTN: expression Quantitative Trait Nucleotide; MER: mean expression ratio affected / non-affected



Supplement 2: Forward, reverse and sequencing primers used and targetted sequences analyzed in DNA methylation experiments. C/T are the methylation variable CpG sites. A biotin was added 5' of the reverse primers.

ITGB5 island 1. Forward: 5'GGTTGTTTTGGTTGGT, reverse: 5'ACTTCCACCTAAACCC, sequencing: 5'GGGTAGGGGAGGGGTTGGTTT, with target: C/TGTTC/TGGC/TGGTC/TGC/TGATTTGGGTTTC/TGAGAC/TGC/TGTTTAGC/TGTC/TG

ITGB5 island 2. Forward: 5'TGAGATTGTGGATAATAGTAG, reverse: 5'AAACTACAAATACCTACCAC, sequencing : 5'AATTTTAGTATTTTGGGAGG, with target: TC/TGAGGTAGGC/TGGATTAC/TGAC/TGTTAGGAGATC/TGAG.

ITGB5 rs10934702. Forward: 5' TGAGGTTGGTGGTATATGAG, reverse: 5' CCTCTTCAAAAATTCTCAAAACTACAAC, sequencing : 5' GTTAATTTTTTGATATTTTAAATGTTTTTAAG, with target: TT/CGTTA.

SPCS3 island 1. Forward: 5'TTTAAAGATAGGAGCG, reverse: 5'ATATTACTCCCTCATTAACC, sequencing: 5'GGATTATGTTGTGAGTGAGG, with target: TC/TGGGTC/TGGC/TGGTGTAGGAC/TGTC/TGGGATC/TGGGTTGGGGC/TGGAAGTC/TGAAGTC/TGGGTC/TGTC/TGC/TGTC/TG

SPCS3 island 2. Forward: 5' TGAGTAGTTGGGATTATAGG, reverse: 5' TCACCATACAACTCTCTAATTACAAC, sequencing: AGTGTTAGGATTATAGGTATAAG, with target: TTATC/TGC/TGTTC/TGGTC/TGTTG

SPCS3 island 3. Forward: 5' TAGGTTAGAGTGAAGTGG, reverse: 5' ACCAACCTAACCAACATAAC, sequencing: 5'AGTAGTTGGGATTATAGG, with target: C/TGTTC/TGTTATTAC/TGTTC/TGGTT

SPCS3 rs7694145. Forward: 5' ATGGTTTTTATTATTTTGTGGTATG, reverse: 5' ACTTCTAATAAAATCCAAC, sequencing : 5' AGTTTTTTGAGAGATTTTGTTAGG, with target: AAAC/TGTTG.

FZD3. Forward: 5'GATGTGAATGGATTATTAG, reverse: 5' ACCTATTCACACAATTCACCTTC, sequencing: 5' TGTGTGTTTTATAATTTTTTTTAGAAAGG, with target : 5'C/TGTTTGATAC/TGTGTTAAGTAAATGTTGAC/TGAG/ATGTT where G/A is the SNP rs1946583.

Références.

1. Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N, et al. Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in Eastern Quebec families. Molecular psychiatry. 2005;10(5):486-99.

2. Merette C, Roy MA, Bureau A, Fournier A, Emond C, Cliche D, et al. Replication of linkage with bipolar disorder on chromosome 16p in the Eastern Quebec population. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2008;147B(6):737-44.

3. 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. European journal of human genetics : EJHG. 2009;17(8):1034-42.

4. Maziade M, Fournier A, Phaneuf D, Cliche D, Fournier JP, Roy MA, et al. Chromosome 1q12-q22 linkage results in eastern Quebec families affected by schizophrenia. American journal of medical genetics. 2002;114(1):51-5.

5. Ott J, Bhat A. Linkage analysis in heterogeneous and complex traits. European child & adolescent psychiatry. 1999;8 Suppl 3:43-6.

6. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249-64.

7. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. Biostatistics. 2012;13(3):539-52.

8. Tabares-Seisdedos R, Rubenstein JL. Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer. Molecular psychiatry. 2009;14(6):563-89.