Supplementary Methods

GSMA

Since the GSMA operates on linkage analysis results, its ability to localize signals is limited based on marker density. While others have proposed an approach that shifts large bins to refine the regional signals[1], we divided the genome into 5, 10, 15 and 20 cM bins. Each bin was assigned a p-value based on the maximal significance level reached in that bin. In applying GSMA, bins within a scan are ranked according to this test statistic value. Since the individual studies varied substantially by sample size, and simulation studies showed that weighting increases the power of the GSMA[2], we also performed a series of analyses where we weighted the rank order statistics by sample size. For ANX, a dichotomous trait, we used the square root of the number of genotyped affected individuals to weight individual studies. For N, a quantitative trait, we used the square root of the total sample size as the weighting factor, since all genotyped subjects provide power to such an analysis.

GSMA calculates the average rank (R_{avg}) for a given genomic bin across the combined studies. The procedure yields two measures of significance [2]. The first, termed P_{SR}, refers to the significance of the R_{Avg} statistic based solely on its rank. For unweighted analyses the R_{Avg} test statistic follows a known distribution for rank order statistics and the probability of a given average rank order can be computed directly. In weighted analyses, the distribution must be determined empirically by random rank ordering studies for combination to determine a probability distribution. We used 10,000 iterations to determine empirical significance for weighted runs. The second statistic, termed P_{OR}, is the significance of R_{Avg} given its overall rank. Significance is determined empirically. For example, the actual value of R_{Avg} for the first place bin is compared to an empirical distribution of R_{Avg} values of all first place finishers from

randomly ranked GSMA procedures to determine P_{OR} while P_{SR} is calculated by comparison to the empiric distribution constructed from all bins. Importantly, neither of these statistics is ideally informative alone. The primary test is the P_{SR}, but a bin's P_{SR} must be considered in the light of its P_{OR}. Since the GSMA procedure will always yield "highest ranking" bins, it is important to assess by P_{OR} whether that particular bin is significant for a bin of its rank. This was illustratively described by Levinson et al.[2] as determining whether a particular scan is a "fast" race or a "slow" race. Previous work has suggested Bonferroni correction of .05/(# bins) for the conclusion of genome-wide significance using GSMA. Given the experience of the field, under this stringent approach we do not anticipate genome-wide significant signals using linkage analysis for complex disease. Instead, our approach aims to highlight unique and overlapping interesting regions using all available linkage datasets for NEU and ANX.

We also assessed whether any individual study was particularly influential in our GSMA procedure by using a drop-one-out sensitivity analysis. In this procedure, we remove one study at a time and rerun the GSMA analysis. Each result is then compared to the original scan across the bins using Spearman's rank correlation to assess the stability of the result.

FDR rational

A q-value is an estimate of the proportion of false discoveries among all significant bins when the corresponding p-value is used as the threshold for declaring significance[3, 4]. This FDR-based approach (1) provides a good balance between the competing goals of true positive findings versus false discoveries, (2) allows the use of more similar standards in terms of the proportion of false discoveries produced across studies because it is much less dependent on the arbitrary number or sets of statistical tests that are performed, (3) is relatively robust against the effects of

correlated tests, and (4) provides a more subtle picture about the possible relevance of the tested bins rather than an all-or-nothing conclusion about whether a study produces significant results[3, 5-12].

Reference List

1. Cooper,M.E., Goldstein,T.H., Maher,B.S., & Marazita,M.L. Identifying genomic regions for fine-mapping using genome scan meta-analysis (GSMA) to identify the minimum regions of maximum significance (MRMS) across populations. *BMC. Genet.* **6 Suppl 1:S42.**, S42 (2005).

2. Levinson,D.F., Levinson,M.D., Segurado,R., & Lewis,C.M. Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: methods and power analysis. *Am. J. Hum. Genet.* **73**, 17-33 (2003).

3. Storey,J.D. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Stat.*2013-2035 (2003).

4. Storey,J.D. & Tibshirani,R. Statistical significance for genome-wide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440-9445 (2003).

5. Benjamini,Y. & Hochberg,Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**, 289-300 (1995).

6. Brown,B.W. & Russell,K. Methods correcting for multiple testing: Operating characteristics. *Statistics in Medicine* **16**, 2511-2528 (1997).

7. Sabatti,C., Service,S., & Freimer,N. False discovery rate in linkage and association genome screens for complex disorders. *Genetics* **164**, 829-833 (2003).

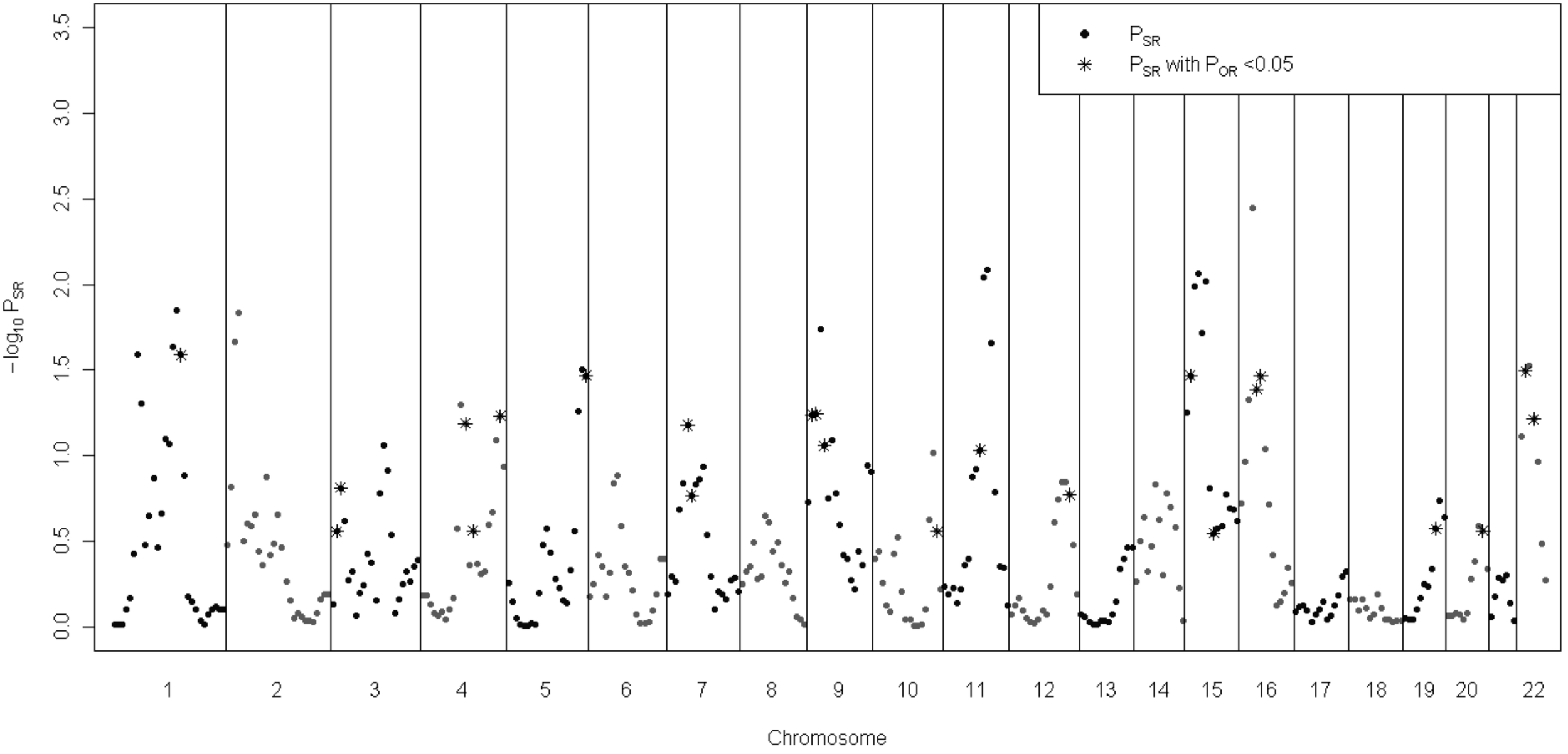
8. Tsai,C.A., Hsueh,H.M., & Chen,J.J. Estimation of false discovery rates in multiple testing: application to gene microarray data. *Biometrics.* **59**, 1071-1081 (2003).

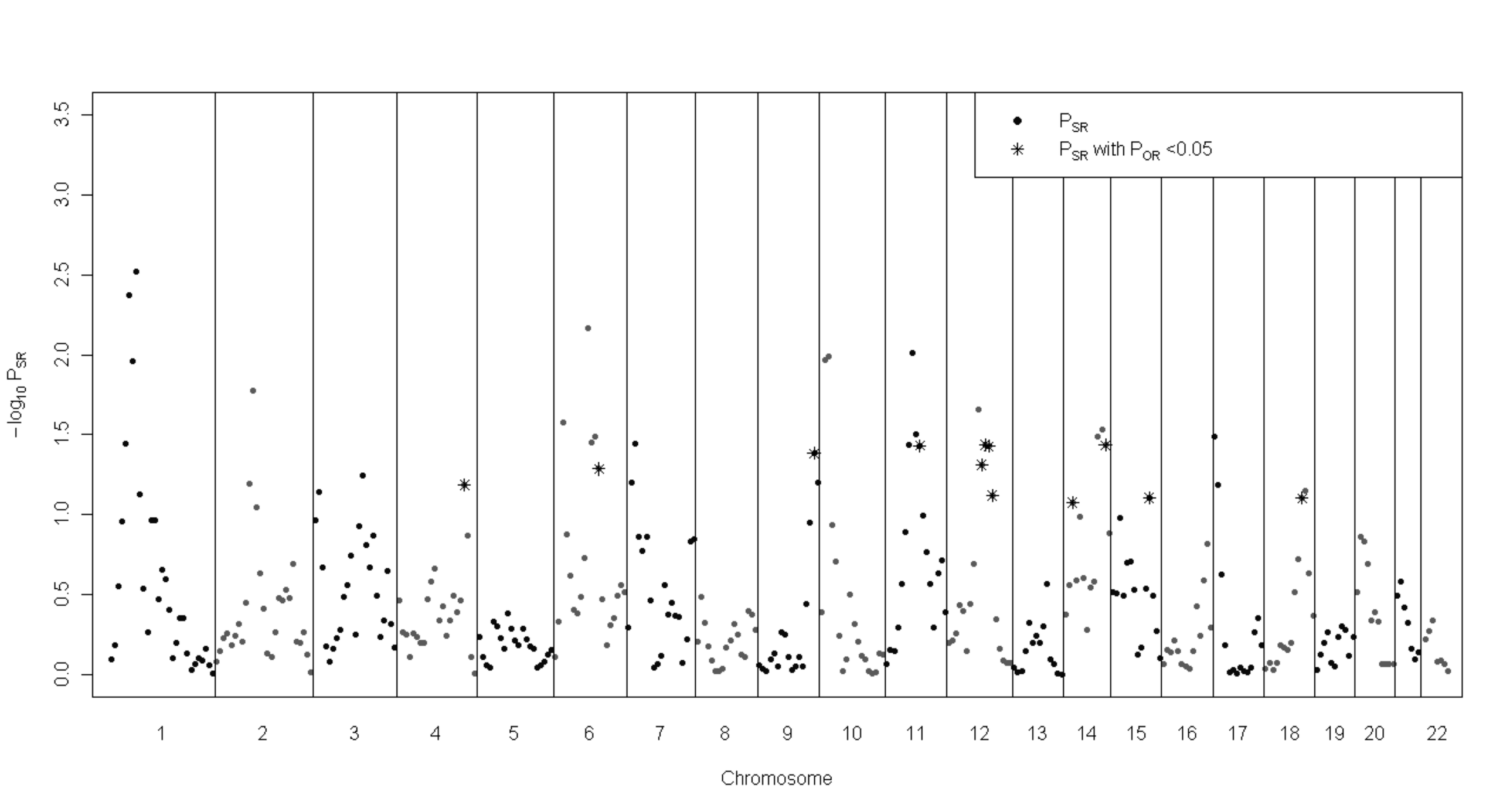
9. van den Oord,E.J.C.G. & Sullivan,P.F. A framework for controlling false discovery rates and minimizing the amount of genotyping in the search for disease mutations. *Human Heredity* **56**, 188-189 (2003).

10. Fernando,R.L. *et al.* Controlling the proportion of false positives in multiple dependent tests. *Genetics* **166**, 611-619 (2004).

11. Korn,E.L., Troendle,J.F., McShane,L.M., & Simon,R. Controlling the number of false discoveries: application to high-dimensional genomic data. *Journal of Statistical Planning and Inference* **124**, 379-398 (2004).

12. van den Oord,E.J. Controlling false discoveries in candidate gene studies. *Mol. Psychiatry.* **10**, 230-231 (2005).


P_{SR}
P_{SR} with P_{OR} <0.05
$-\log_{10} P_{SR}$
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
Chromosome


P_{SR}
P_{SR} with P_{OR} <0.05
$-\log_{10} P_{SR}$
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
Chromosome

Sensitivity: Spearman Rank Correlations - Study Dropped vs All

Neuroticism		Study										
GSMA method	grid	ATR	DENT	GENESiS	IASPSAD	UKSP	NTR	NZ-ND	BATS	min	max	mean
unweighted	5cM	0.922	0.933	0.928	0.932	0.920	0.926	0.926	0.923	0.920	0.933	0.926
unweighted	10cM	0.925	0.937	0.930	0.934	0.927	0.930	0.926	0.927	0.925	0.937	0.930
unweighted	15cM	0.929	0.936	0.934	0.938	0.929	0.930	0.930	0.932	0.929	0.938	0.932
unweighted	20cM	0.937	0.940	0.937	0.943	0.939	0.937	0.932	0.937	0.932	0.943	0.938
weighted	5cM	0.717	0.954	0.975	0.966	0.959	0.990	0.931	0.953	0.717	0.990	0.931
weighted	10cM	0.727	0.956	0.976	0.968	0.961	0.991	0.932	0.956	0.727	0.991	0.934
weighted	15cM	0.744	0.953	0.976	0.968	0.962	0.990	0.932	0.958	0.744	0.990	0.935
weighted	20cM	0.764	0.956	0.976	0.969	0.966	0.990	0.937	0.958	0.764	0.990	0.939