Supplemental Data
Genome-wide Association Analysis Reveals Putative Alzheimer's Disease Susceptibility Loci in Addition to *APOE*

Lars Bertram, Christoph Lange, Kristina Mullin, Michele Parkinson, Monica Hsiao, Meghan F. Hogan, Brit M. M. Schjeide, Basavaraj Hooli, Jason DiVito, Iuliana Ionita, Hongyu Jiang, Nan Laird, Thomas Moscarillo, Kari L. Ohlsen, Kathryn Elliott, Xin Wang, Diane Hu-Lince, Marie Ryder, Amy Murphy, Steven L. Wagner, Deborah Blacker, K. David Becker, and Rudolph E. Tanzi

Table S1. Demographic characteristics of screening sample and follow-up datasets.

Sample	No. families (subjects)	No. women (%)	No. affecteds (AAO+SD [range])	No. unaffecteds (AAE+SD [range])
*NIMH	410 (1,376)	930 (68%)	941 (72.3+7.7 [50-97])	404 (70.0+10.7 [31-93])
NIA	329 (1,040)	639 (61%)	748 (74.2+7.1 [52-98])	282 (73.4+9.6 [36-94])
NCRAD	331 (1,108)	706 (64%)	799 (71.3+7.6 [50-98])	293 (70.6+8.1 [39-93])
CAG	215 (483)	294 (61%)	220 (69.3+9.0 [50-89])	263 (73.3+8.6 [50-92)

*Sample used for initial GWA analysis; other samples used for follow-up analyses. Numbers missing to total subjects when adding affecteds and unaffecteds = phenotype unknown.

Table S2. Accuracy of genotype calls using the DM or BRLMM genotype calling algorithms.

	Chip Call Rates				DM Algorithm		BRLMM Algorithm	
Chip Type	Patient 4	Patient 2	Patient 1	Average Call rate	Number of Inheritance Errors	Accuracy	Number of Inheritance Errors	Accuracy
StyI	95.91%	95.86%	93.29%	95.0%	1269	99.47%	631	99.74%
StyI	92.59%	90.78%	91.35%	91.6%	1866	99.22%	1192	99.50%
StyI	88.52%	82.19%	87.28%	86.0%	3559	98.51%	2046	99.14%
NspI	95.20%	97.94%	98.04%	97.1%	707	99.74%	272	99.90%
NspI	91.86%	88.83%	91.72%	90.8%	2655	99.01%	1377	99.49%
NspI	85.70%	84.80%	86.24%	85.6%	4011	98.51%	2427	99.10%
NspI	82.26%	80.62%	81.11%	81.3%	4173	98.44%	1817	99.32%
NspI	77.23%	76.94%	74.69%	76.3%	4923	98.17%	2300	99.14%
NspI	70.85%	73.19%	70.03%	71.4%	5204	98.06%	2107	99.21%

Accuracy was assessed by determining the number of inheritance errors for a family trio (mother, father, child). Inheritance errors were identified on replicate data with varying call rates collected for each member of the trio using PedCheck[1]. The number of inheritance errors consistently decreased with increasing initial DM call rate, indicating that higher call rates correlate with better accuracy in our dataset. BRLMM analysis of the same raw data CEL files resulted in even fewer inheritance errors as compared to the DM genotypes, indicating that the genotypes generated by BRLMM are highly accurate. Similar to the DM data, the number of inheritance errors found in genotypes called by BRLMM decreased with increasing initial DM chip call rate.

Reference to Supplementary Table 2:

1. O'Connell, J.R. and Weeks, D.E. (1998). PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet *63*:259-66.

Legends to Supplementary Figures

Figure S1. An increased number of heterozygous genotypes called by BRLMM as compared to DM was observed. This suggested that the accuracy for calling heterozygous genotypes is improved using BRLMM. For a representative subset of our data, we observed a 2.7% increase in the number of heterozygotes called across both GeneChip® arrays.

Figure S2. (A) Genotype calls made by DM and BRLMM were in very close agreement (>99.2% concordant). Concordance in genotype calls increased as DM call rates increased, suggesting that data generated from arrays with high call rates may be more accurate. (B) Though BRLMM was able to make calls on a significant number of SNPs previously not called with DM, a much lower number of SNPs were observed for which DM made a genotype call and BRLMM did not ("Lost" genotypes).

Figure S3. BRLMM genotype calling experiments were carried out with batch sizes of 50 and 100 chip data CEL files. Initial genotype call rates were determined using the DM algorithm (0.33 threshold). We tested the genotype call rate outcome for samples with *moderate* (93%), *good* (95%) and *excellent* (98%) chip call rates when processed in varying batch environments. Eighteen test samples were analyzed by BRLMM in different batch environments. Raw data from one test sample was combined with either 49 or 99 other samples for batch analysis by BRLMM. Six test samples had approximately 93% DM call rates; six samples had approximately 95% call rates, and six samples had greater than 98% call rates. Three batch environments contained either 49 or 99 samples, and were defined based upon their initial DM call rate. The moderate group call rates ranged from 93 to 94%; the mixed group call rates ranged from 93 to 99%; the excellent group call rates ranged from 98 to 99%. "Like" and "unlike" refer to the similarity of the test sample call rate compared to the call rate for the other 49 or 99 samples used in the cluster. (A) Data from Nsp chips and (B) data from Sty chips from the same individual were analyzed separately. The majority of the cases tested did show an increase in call rates where samples were analyzed in "like" vs "unlike" environments with a few exceptions (C). In addition, "Like" outperformed "mixed" batches in the majority of cases as well. To ensure that accuracy was maintained, inheritance errors were measured for

trios called in these three different environments (like, mixed, unlike) and in all cases, Mendelian errors fell below the 0.5% threshold.

Figure S4. BRLMM-derived allele signals of a SNP transformed into Cluster-Center-Stretch space (http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf) Data for a single SNP in multiple samples has been plotted based upon the signal intensity versus allele contrast (signal strength on the A versus B allele probes on the chip). (A) Data from a moderate call rate batch of samples (green dots) is compare to data from an excellent call rate batch (pink triangles). The pink triangles form distinct clusters for the three possible genotypes (BB – left, AB – center, AA – right). This example illustrates that both the allele contrast and the signal strength can shift markedly with different input data sets. (B) Excellent call rate data for which the “call zone” (shown in blue) is shown for the BB genotype cluster. The “call zone” is calculated based upon the variance in the distance from the center of the cluster for each data point, as well as the distance between cluster centers. (C) The “call zone” for the BB cluster is shown for the moderate data. In both B and C a single data point (blue diamond) derived from a moderate call rate chip, is either excluded from the “call zone” and therefore not called (B), or is included in the “call zone” (C) and is given the correct genotype.

Figure S5. (A) Distribution of GeneChip® call rates. The average chip call rates for DM =0.33 were 96.45%. Application of the BRLMM algorithm further improved call rates, increasing the average call rate to 98.95% improving the average chip call rates by approximately 2.3% across the entire sample set. (B) Overall SNP call rate performance of DM versus BRLMM. The percentage of SNPs amenable for genetic analysis (i.e. those having greater than 90% of the samples with genotypes called) increased from 88.3% to 98.8% across both arrays when the BRLMM algorithm was applied. This resulted in a gain of 52,500 SNPs (441,500 SNPs versus. 494,000 SNPs) available for genetic analysis.

Figure S6. Distribution of FBAT-GEE P-values for all 404,604 SNP on the 500K array with ≥10 informative families as histogram depicting the frequency of the range of observed P-values (from 0 to 1; bin size = 20 P-values). The P-value distribution is in line with what would be expected under the null hypothesis.

Figure S1.

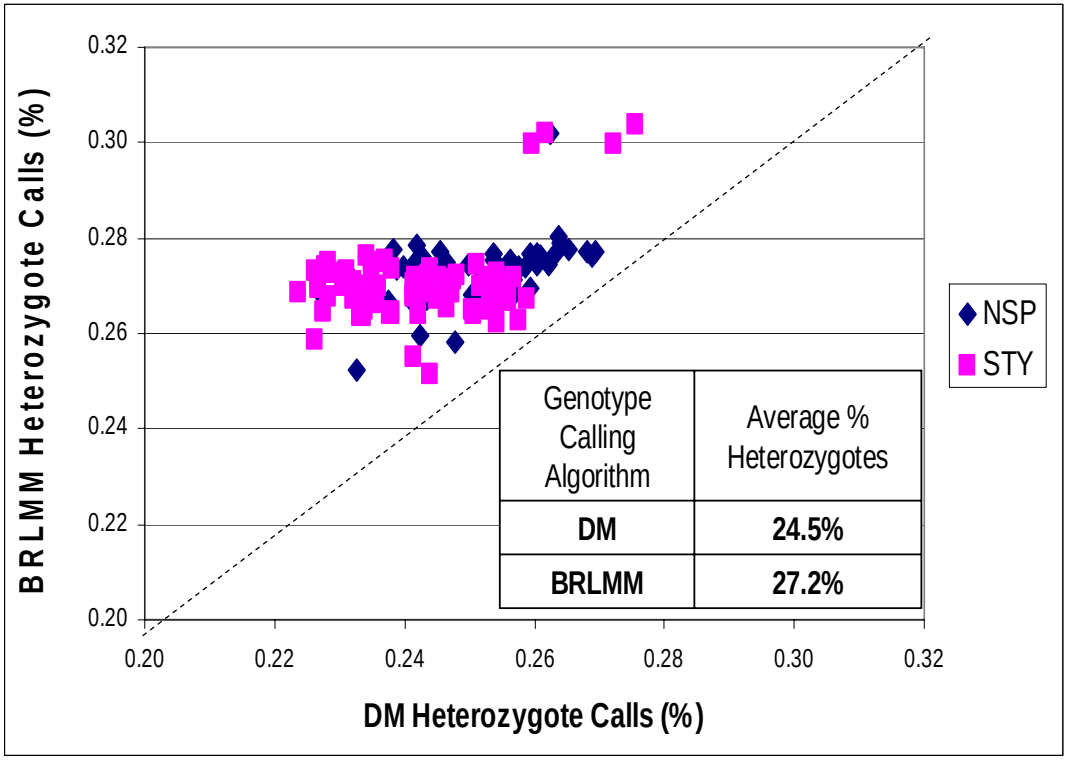

BRLMM Heterozygote Calls (%)
0.32
0.30
0.28
0.26
0.24
0.22
0.20
0.20
0.22
0.24
0.26
0.28
0.30
0.32
DM Heterozygote Calls (%)
NSP
STY
Genotype Calling Algorithm
Average % Heterozygotes
DM
24.5%
BRLMM
27.2%

Figure S2A.

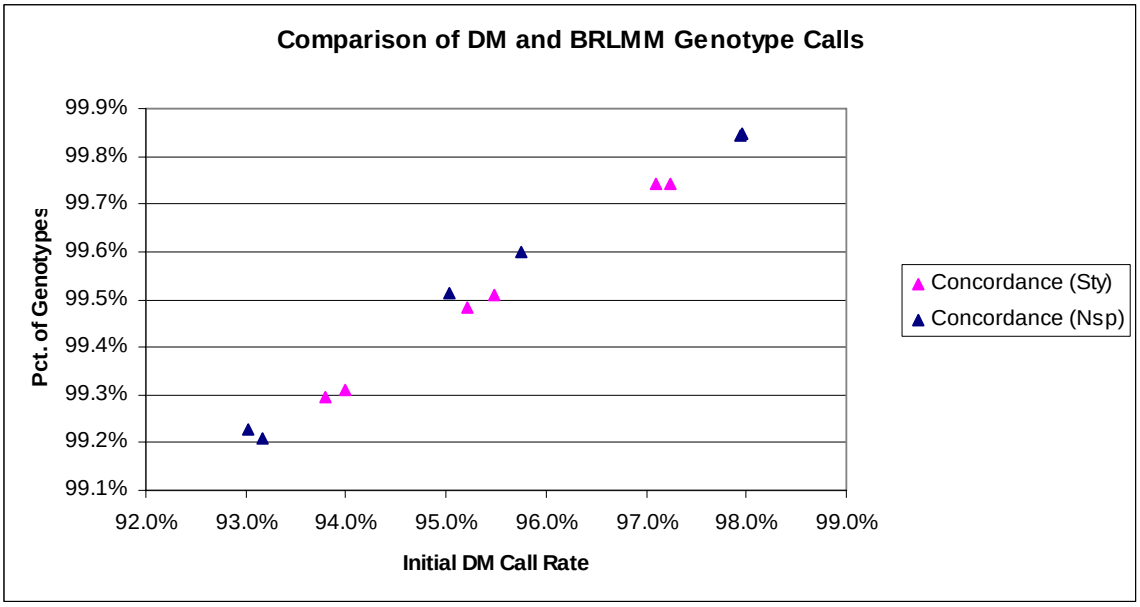


Figure S2B.

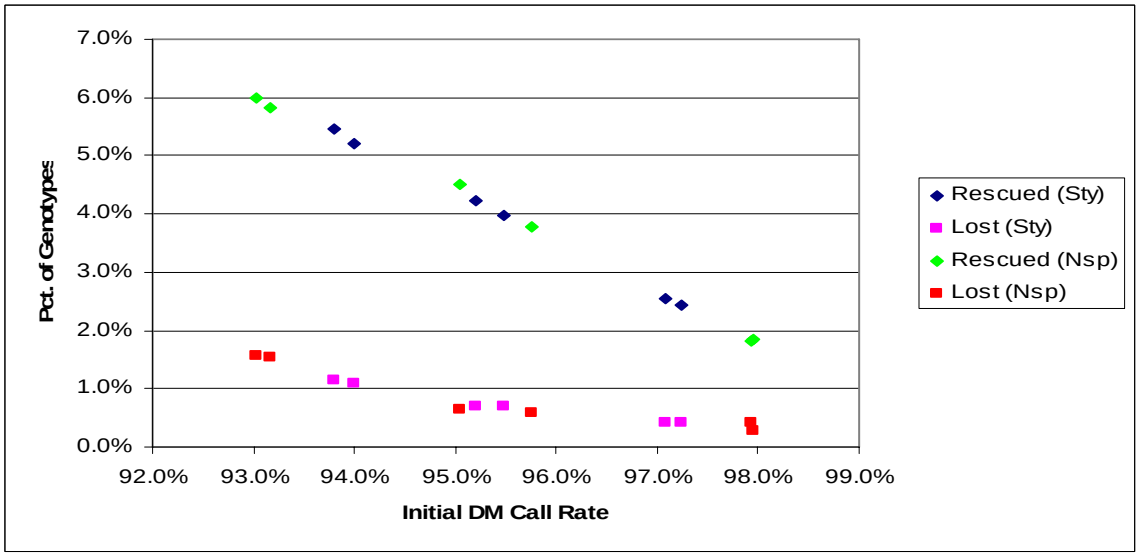


Figure S3A.

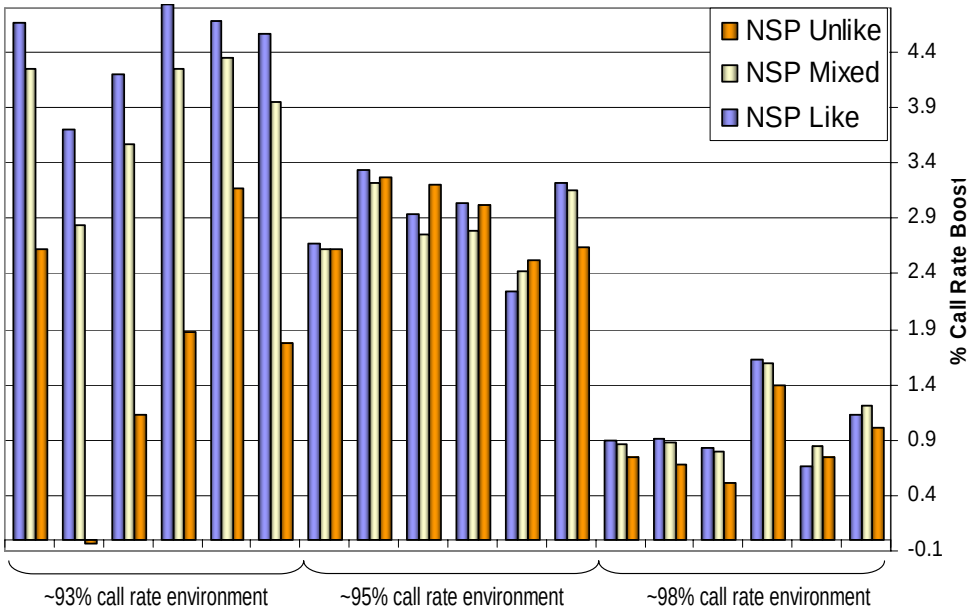


Figure S3B.

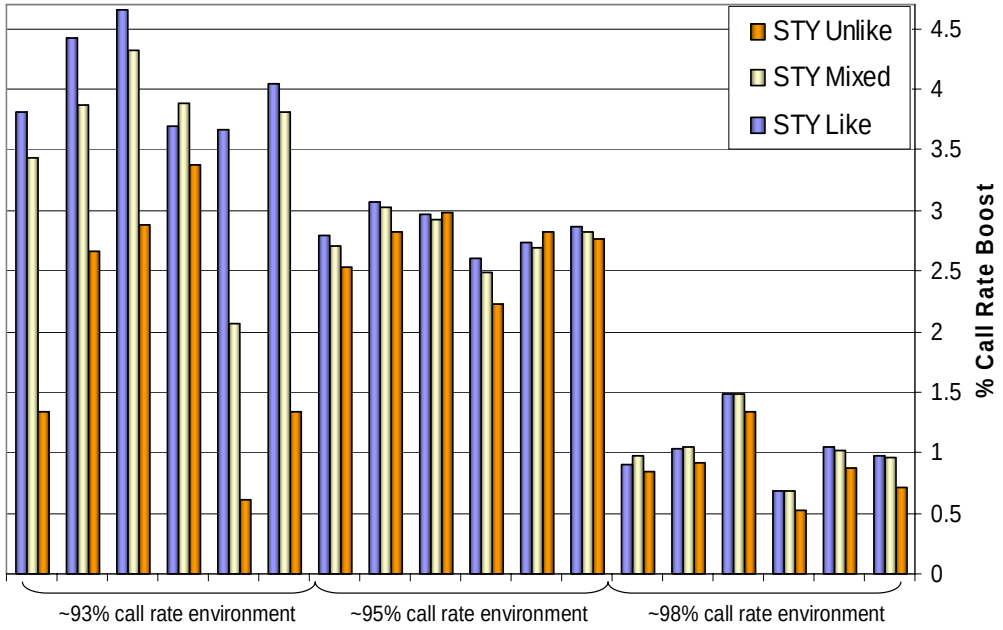


Figure S3C.

	LIKE > UNLIKE	LIKE > MIXED
Moderate CR (93%)	12/12	11/12
Good CR (95%)	8/12	11/12
Excellent CR (98%)	11/12	9/12

Figure S4A.

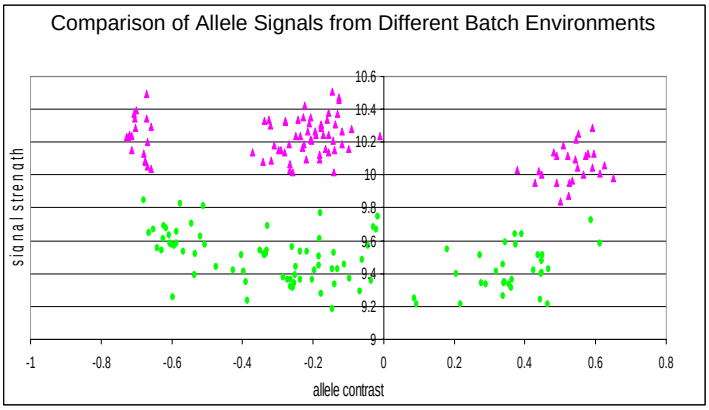


Figure S4B.

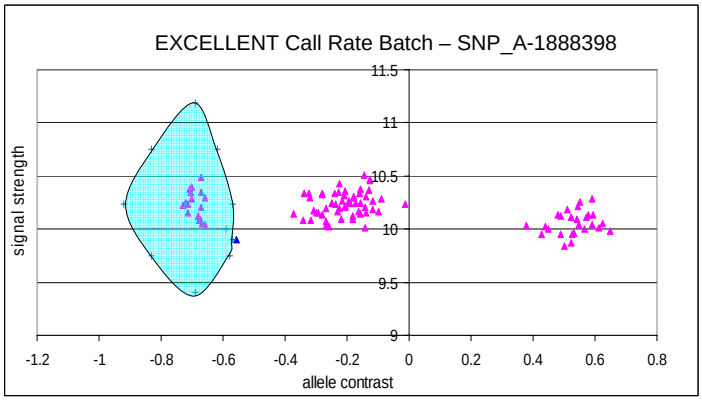


Figure S4C.

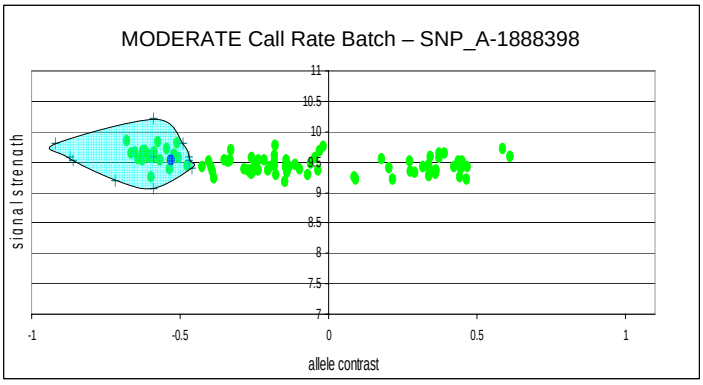


Figure S5A.

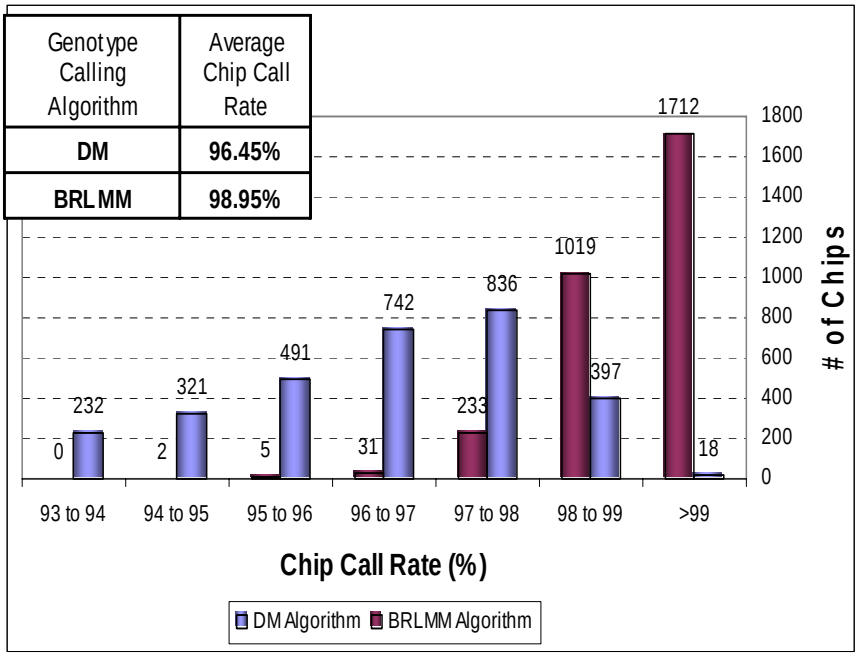
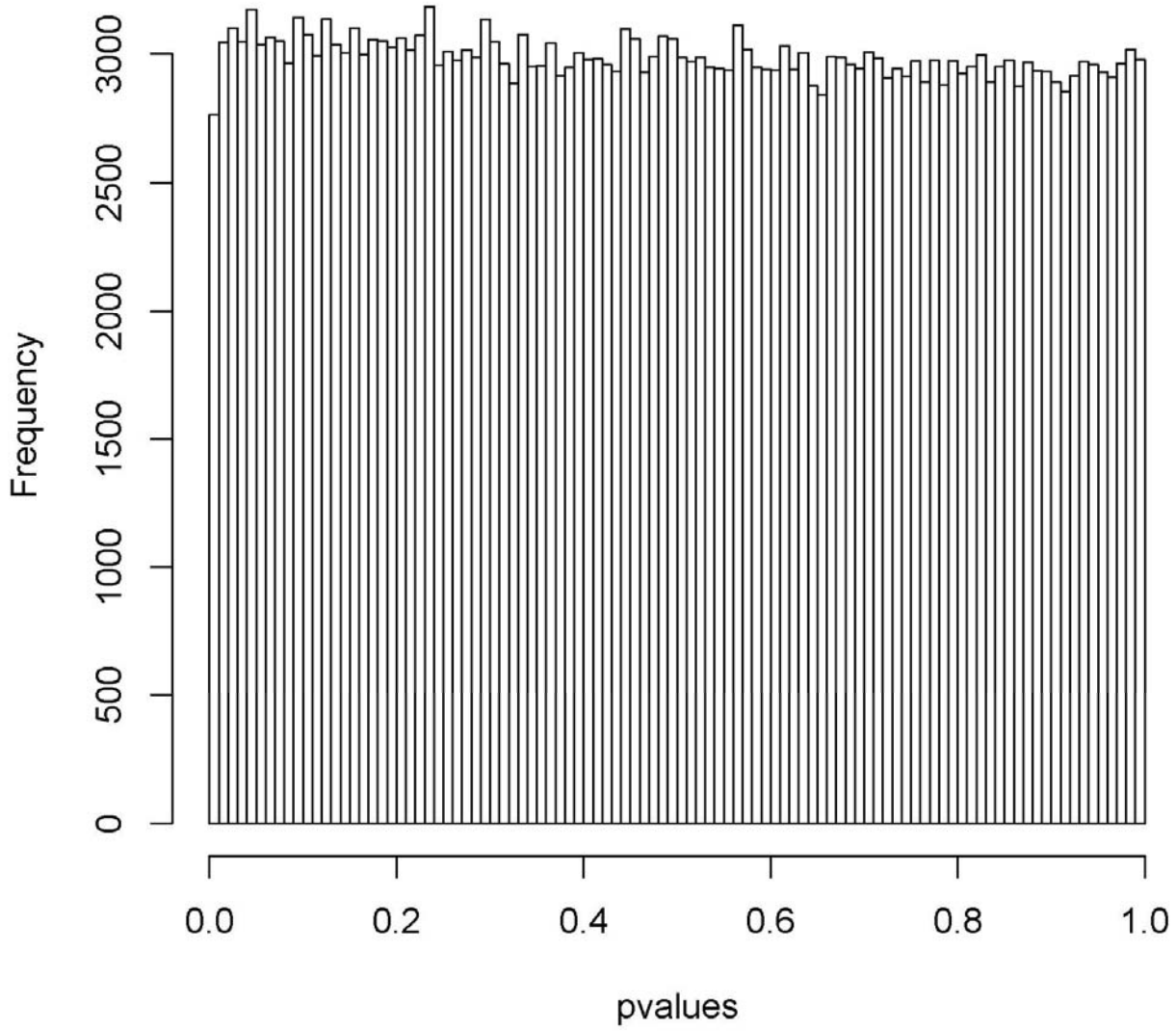


Figure S5B.

NSP I Chips		STY I Chips		BOTH Chips	
DM	BRLMM	DM	BRLMM	DM	BRLMM
89.14%	99.20%	87.40%	98.50%	88.30%	98.80%

Figure S6.


Frequency
0
500
1000
1500
2000
2500
3000
0.0
0.2
0.4
0.6
0.8
1.0
pvalues