

# The Value of Regenotyping Older Linkage Data Sets with Denser Marker Panels

Veronica J. Vieland<sup>a</sup> Kimberly A. Walters<sup>a</sup> Marco Azaro<sup>b</sup>  
Linda M. Brzustowicz<sup>b</sup> Thomas Lehner<sup>c</sup>

<sup>a</sup>Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio, <sup>b</sup>Department of Genetics, Rutgers University, Piscataway, N.J., and <sup>c</sup>The National Institute of Mental Health, Bethesda, Md., USA

## Key Words

Linkage analysis · Data repositories · Microsatellites · SNPs · Information content · PPL

## Abstract

**Objectives:** Linkage analysis can help determine regions of interest in whole-genome sequence studies. However, many linkage studies rely on older microsatellite (MSAT) panels. We set out to determine whether results would change if we regenotyped families using a dense map of SNPs. **Methods:** We selected 47 Hispanic-American families from the NIMH Repository and Genomics Resource (NRGR) schizophrenia data repository. We regenotyped all individuals with DNA available from the NRGR on the Affymetrix Lat Array. After optimizing SNP selection for inclusion on the linkage map, we compared information content (IC) and linkage results using MSAT, SNP and MSAT+SNP maps. **Results:** As expected, SNP provided a higher average IC (0.78, SD 0.03) than MSAT (0.51, SD 0.10) in a direct 'apples-to-apples' comparison using only individuals genotyped on both platforms; while MSAT+SNP provided only a slightly higher IC (0.82, SD 0.03). However, when utilizing all available individuals, in-

cluding those who had genotypes available on only one platform, the IC was substantially increased using MSAT+SNP (0.76, SD 0.05) compared to SNP (0.61, SD 0.02). Linkage results changed appreciably between MSAT and MSAT+SNP in terms of magnitude, rank ordering and localization of peaks. **Conclusions:** Regenotyping older family data can substantially alter the conclusions of linkage analyses.

© 2014 S. Karger AG, Basel

## Introduction

Linkage analysis has fallen out of favor as a technique for gene discovery in psychiatric genetics and some other clinical areas. This is understandable given the paucity of definitive successes, despite considerable investment in psychiatric linkage studies over a number of years. One explanation for the lack of clear success is that the study design itself is flawed in application to psychiatric conditions; but another is that the successful execution of the design has hitherto been thwarted by limitations of available technologies. Here we investigate one aspect of this execution: reliance of the majority of psychiatric linkage

studies, which were carried out during the 1990s and early 2000s, on sparse microsatellite (MSAT) marker maps.

We consider a particular set of multiplex schizophrenia pedigrees obtained from the NIMH Repository and Genomics Resource (NRGR). The available NRGR data included genotypes for a sparse (approx. 9 cM) MSAT map, which was the state of the art at the time these studies were initially done. As part of the Combined Analysis of Psychiatric Studies (CAPS) project, a collaboration with the NRGR, we regenotyped all individuals from these families who had DNA available from the NRGR, in order to obtain dense SNP genotypes.

Our objectives were two-fold. First, we wanted to assess the extent to which information content (IC) would be augmented by regenotyping the older samples on an updated genotyping platform. Second, we wanted to know what the impact would be on the results of linkage analysis. As anticipated, the gain in IC was appreciable using SNPs, and it increased further when using both MSATs and SNPs simultaneously. In addition, there were notable changes in the linkage results when the high IC map was used relative to the results obtained based on the MSATs alone. This strongly suggests that linkage analysis results based on older studies were substantially hindered by the available marker maps at the time, and that an assessment of their results should therefore be made in light of this limitation.

## Methods

Because our genotyping budget was limited, we chose a single set of families for this study, in part based on a phenotype of particular interest. Specifically, we focus on the Hispanic-American multiplex schizophrenia sample originally collected by Escamilla and colleagues in 2003 and 2005 [1, 2], and the complete subset of families ( $n = 47$ ) from this data set that had at least one case of strictly defined multiplex schizophrenia as well as at least one case of schizophrenia with a strong affective component [for additional clinical details, see Vieland et al. 3].

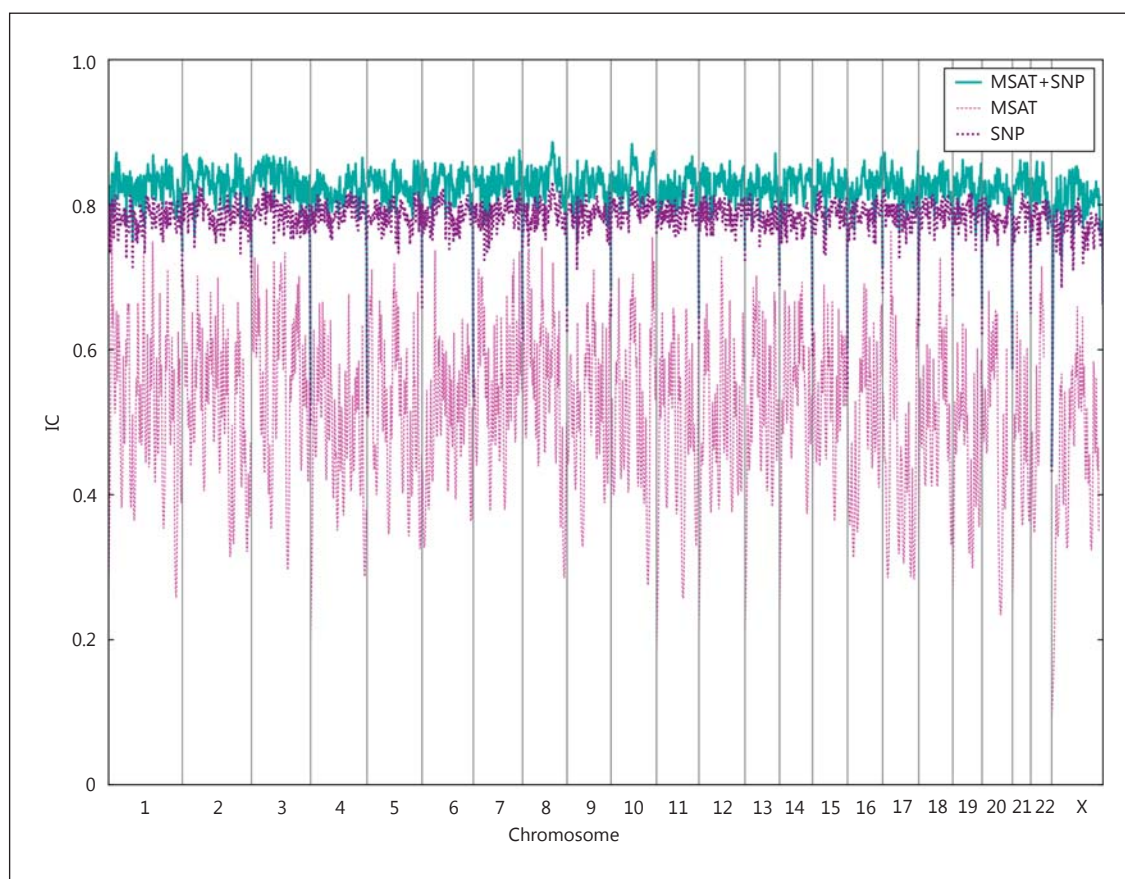
The available NRGR data included genotypes for 402 MSATs on 231 individuals in the 47 families (for descriptions of the pedigree structures, see below). We applied a data-cleaning protocol involving pedigree-structure verification (including heterozygosity checks on the X chromosome to confirm sex) as well as an evaluation of markers for excessive missingness and/or Mendelian errors, or violations of Hardy-Weinberg equilibrium with a  $p$  value  $< 0.01$  [for details, see 3]. Mendelian errors were zeroed out for the marker and family in which they occurred. Following data cleaning, 395 markers and 219 individuals remained in the analysis; two families were dropped due to unresolvable pedigree-structure problems (excessive non-Mendelizations). The average heterozygosity across the 395 markers was 0.23 (SD 0.07, range 0.09–0.51), and the average intermarker distance was 9.29 cM (SD 3.82, range 0.34–19.98).

The NRGR had DNA available for 238 individuals from these 47 families. The genotyping of 193 individuals (all individuals available from the 2005 data set and 17 of 62 individuals from the 2003 data set, see also below) on the Affymetrix Axiom Lat Chip was performed by RUCDR Infinite Biologics (formerly the Rutgers University Cell and DNA Repository) using standard methods. The data are available through the NRGR. (SNPs for 6 members of an uninformative branch of a large pedigree, ungenotyped with MSATs, were set aside, leaving a total of 187 individuals.) Repetition of the family-structure verification was done with the SNP data. This restored 3 families lost using MSATs alone: one that had previously had  $< 2$  genotyped cases and two with excess MSAT Mendelian errors, one of which required changing full siblings to half-siblings. Markers were removed if they were monomorphic or duplicative, if they had  $> 5\%$  missingness, and/or violated Hardy-Weinberg equilibrium with a  $p$  value  $< 0.01$ . Mendelian errors were zeroed out and missingness checks were rerun. 586,178 SNPs remained after these quality control measures (out of  $> 800,000$  probes on the chip).

For various reasons, there were some differences in the sets of individuals genotyped for the MSAT and SNP maps. In addition to the slight difference in available individuals (as noted above), 12 individuals with available DNA for SNP genotyping were missing MSAT genotypes after MSAT cleaning. Furthermore, 44 individuals from the 2003 sample included in the original (MSAT) study did not have SNP genotyping. The reason for this was budgetary: we originally budgeted for the 2005 sample only; however, this left 17 empty lanes. Since the 2003 data set included 15 eligible families, it was decided to additionally obtain SNPs for two affecteds from each of the first 2 of these families (numerically ordered) and for one affected/family from each of the remaining 13 families. While this detail relates to factors ancillary to scientific considerations, it is also typical of situations arising in actual studies, and it plays a role in the Results section. Finally, two of the pedigrees were too large for some of the analyses done here (see below), leaving a reduced set of 45 pedigrees for some analyses.

Given the features of the data, we considered 3 separate versions of the data set: (i) the reduced set of 45 pedigrees including all and only individuals with available genotypes on both the MSAT and SNP maps (Baseline,  $n = 29$  families); (ii) the reduced set of pedigrees, including all pedigrees and individuals with genotypes on either the MSAT or SNP map (Baseline<sup>+</sup>,  $n = 45$  families), and (iii) the set of all families ( $n = 47$ ) including all individuals genotyped on either platform (Real Data). For Baseline, any founders or connecting relatives required for pedigree integrity who had genotypes on just one platform were left in the pedigree files with their genotypes set to 'missing'. The Real Data had an average of 7.1 individuals per pedigree (range 4–31); an average of 4.9 genotyped individuals per pedigree (range 2–19), and an average of 2.5 affecteds per pedigree (range 2–5). Virtually all affecteds were genotyped; however, only an average of 1.1 founders per pedigree were genotyped (range 0–5), which is typical of many of the older collections.

In preparation for linkage analysis, we tried a number of approaches aimed at maximizing IC of the SNP marker map while removing marker-to-marker linkage disequilibrium (LD). See online supplementary table 1 for details of the comparisons (for all online suppl. material, see [www.karger.com/doi/10.1159/000360003](http://www.karger.com/doi/10.1159/000360003)). In brief, the final algorithm selected for the construction of the SNP linkage map involved the following steps: (i) markers with minor



**Fig. 1.** Genome-wide plot of IC for each of the three marker maps, including exactly the same families and the same set of genotyped individuals within those families for all three maps. Two large pedigrees were omitted in order to use Merlin's IC calculation. The IC drop-off at telomeres is affected by the distance between the first (last) marker on the chromosome and the first (last) calculation

position. Here we calculated IC from 0 cM (start) to the last marker on the right + 2 cM (end). Across chromosomes, the average distance from start to the first marker is 5.76 cM for MSATs and 0.36 for SNPs, while the average distance from the last marker to end is 1.01 for MSATs and 0.77 for SNPs. This explains the differences in behavior at the telomeres.

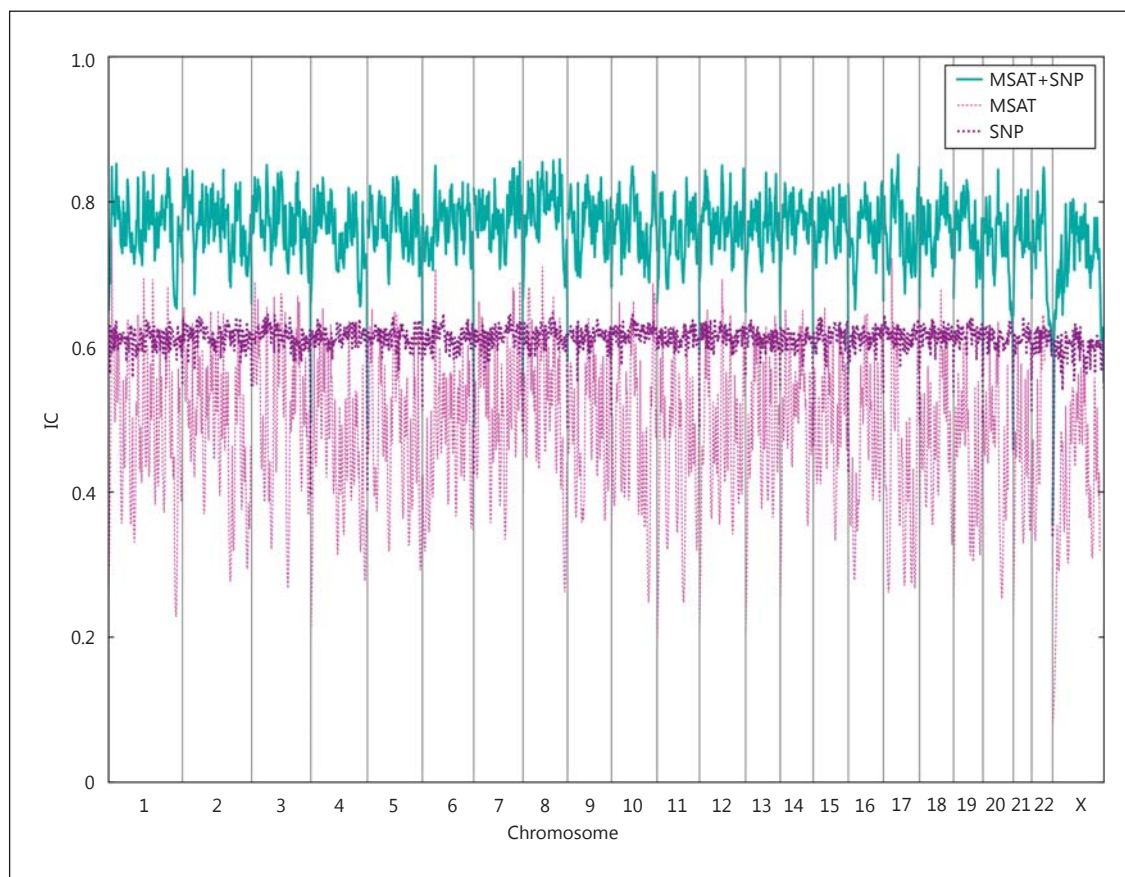
allele frequency (MAF) <0.20 were discarded; (ii) a base map was constructed using all markers with MAF >0.45; (iii) all gaps >0.3 cM in the base map were supplemented with an additional marker, based on the highest MAF marker available within the gap, and this process was repeated until all intermarker distances were <0.3 cM, and (iv) marker-to-marker LD was eliminated at an  $R^2$  threshold of 0.20 using an iterative algorithm described in detail in online supplementary table 1. For Baseline<sup>+</sup>, the resulting SNP map comprised 25,414 markers, with an average MAF of 0.46 (SD = 0.04, range 0.21–0.50) and an average intermarker distance of 0.15 cM (SD = 0.12, range 0.00–1.97). (For Baseline, there were 25,408 SNPs on the map and the remaining descriptive statistics were identical to two decimal places.) Note that these maps had identical IC to two decimal places when compared to the map with all SNPs, but they allow for far more efficient computation since they contain only about one quarter as many markers, even after pruning the 'all SNPs' set to remove marker-to-marker LD.

In order to align the markers on a genetic map and create a combined MSAT+SNP marker set, all markers were placed on

the v3 Rutgers combined linkage-physical map (hg19, Build 37.3, dbSNP 137) [4]. Markers from previous builds not present in v3 were located using either comparable v2 positions (this applied only to 20 markers at one telomere of chromosome 1) or assigned positions using interpolation (approx. 100 additional SNPs). In constructing the MSAT+SNP map, any SNP falling within 0.3 cM of an MSAT was also removed to avoid introducing any new LD between markers. This resulted in dropping 1,742, 1,738 and 1,743 additional SNPs for Baseline, Baseline<sup>+</sup> and Real Data analysis, respectively. The final MSAT+SNP map used for Real Data analysis contained 24,216 markers in total, with an average intermarker distance of 0.15 cM (SD 0.12, range 0.00–1.97).

IC was calculated in Merlin [5]. Two of the 47 pedigrees were too large for Merlin to handle, thus IC comparisons are shown for the subset of 45 pedigrees only. Multipoint linkage analysis was done using Kelvin [6]. Shown here are results for the posterior probability of linkage (PPL) statistic; for comparison purposes, nonparametric linkage (NPL) [7] results as calculated by Mer-





**Fig. 2.** Genome-wide plot of IC for each of the three marker maps, including all available families and individuals for all three maps. Two large pedigrees were omitted in order to use Merlin's IC calculation.

lin are shown in online supplementary figure S2, for the 45 pedigrees Merlin was able to process. [Kelvin performed exact (full) multipoint calculations for these 45 pedigrees and utilized a hybrid MCMC-exact likelihood for the remaining 2 pedigrees, 8]. Prior to analysis, marker allele frequencies were estimated using maximum likelihood. We also performed LD analysis using the full set of SNPs (see online suppl. fig. S3 for results).

The PPL is essentially a straightforward application of Bayes' theorem. Letting  $L$  represent 'linkage' to a given genomic position and  $D$  be the data, the PPL is calculated as

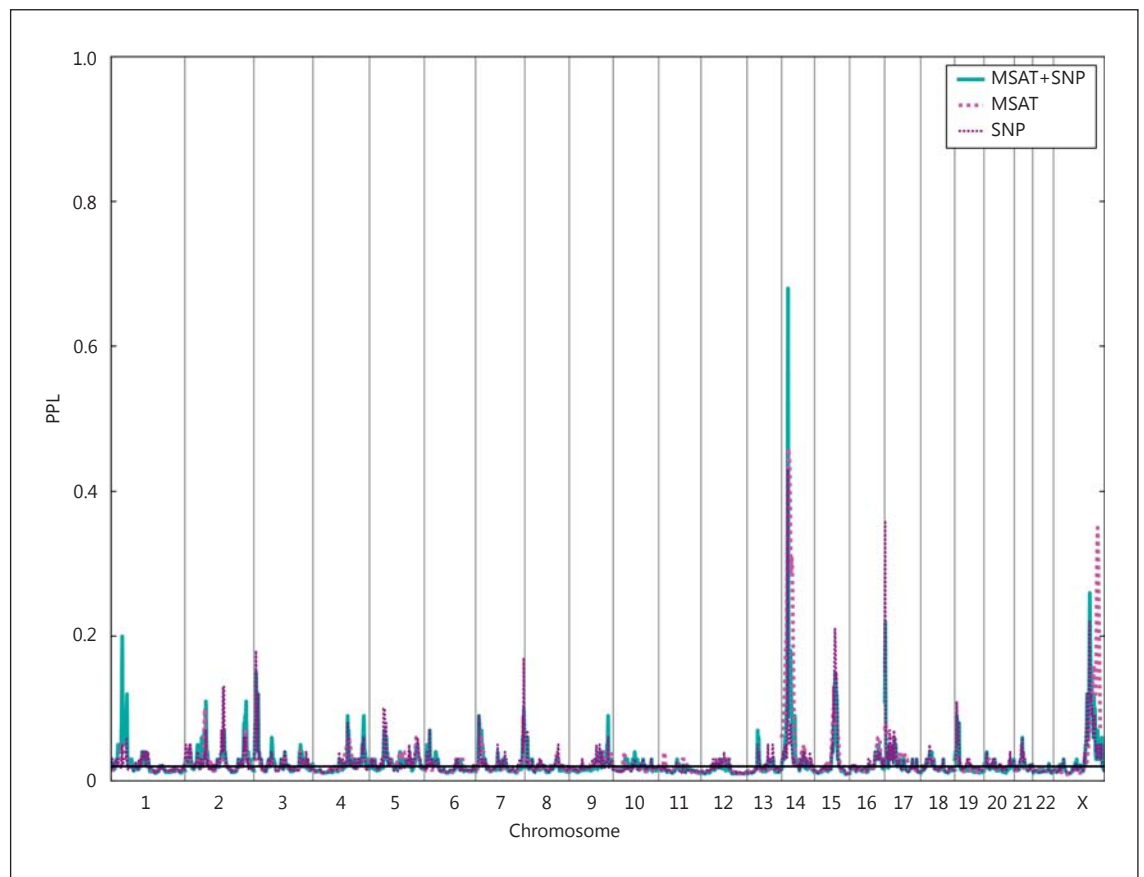
$$PPL = P(L|D) = \frac{P(D|L)P(L)}{P(D|L)P(L) + P(D|no L)P(no L)}.$$

The prior probability of linkage  $P(L)$  is set to 2% based on empirical calculations. The two likelihoods appearing in this equation,  $P(D|L)$  and  $P(D|no L)$ , are the numerator and denominator of the (exponentiated) ordinary LOD score, respectively. They are functions of the parameters of the unknown trait model; Kelvin handles these parameters via (numerical) integration. The PPL is on the probability scale  $[0, \dots, 1]$  and, as applied here, it represents the probability of a schizophrenia disorder gene at each location, given the available data.

## Results

We first compared IC for the Baseline data subset (fig. 1). This comparison establishes the relative IC of the three different maps in a direct apples-to-apples comparison. Not surprisingly, the SNPs alone yielded a higher average IC across the genome compared to MSAT data alone (SNP mean IC = 0.78; MSAT mean IC = 0.51). However, the MSAT IC had far greater variance (SNP IC SD = 0.03, range 0.43–0.83; MSAT IC SD = 0.10, range 0.09–0.80), and the SNP map had some gaps, with the result that the IC was actually higher for the MSATs than SNPs at certain locations. The IC was only modestly increased when comparing the MSAT+SNP map to the SNP map itself (MSAT+SNP mean IC = 0.82, SD = 0.03, range 0.44–0.89).

While yielding a direct assessment of additional IC on the SNP map relative to the MSAT map, figure 1 tends to understate the true value of adding dense SNP information on top of existing MSATs in typical research situa-



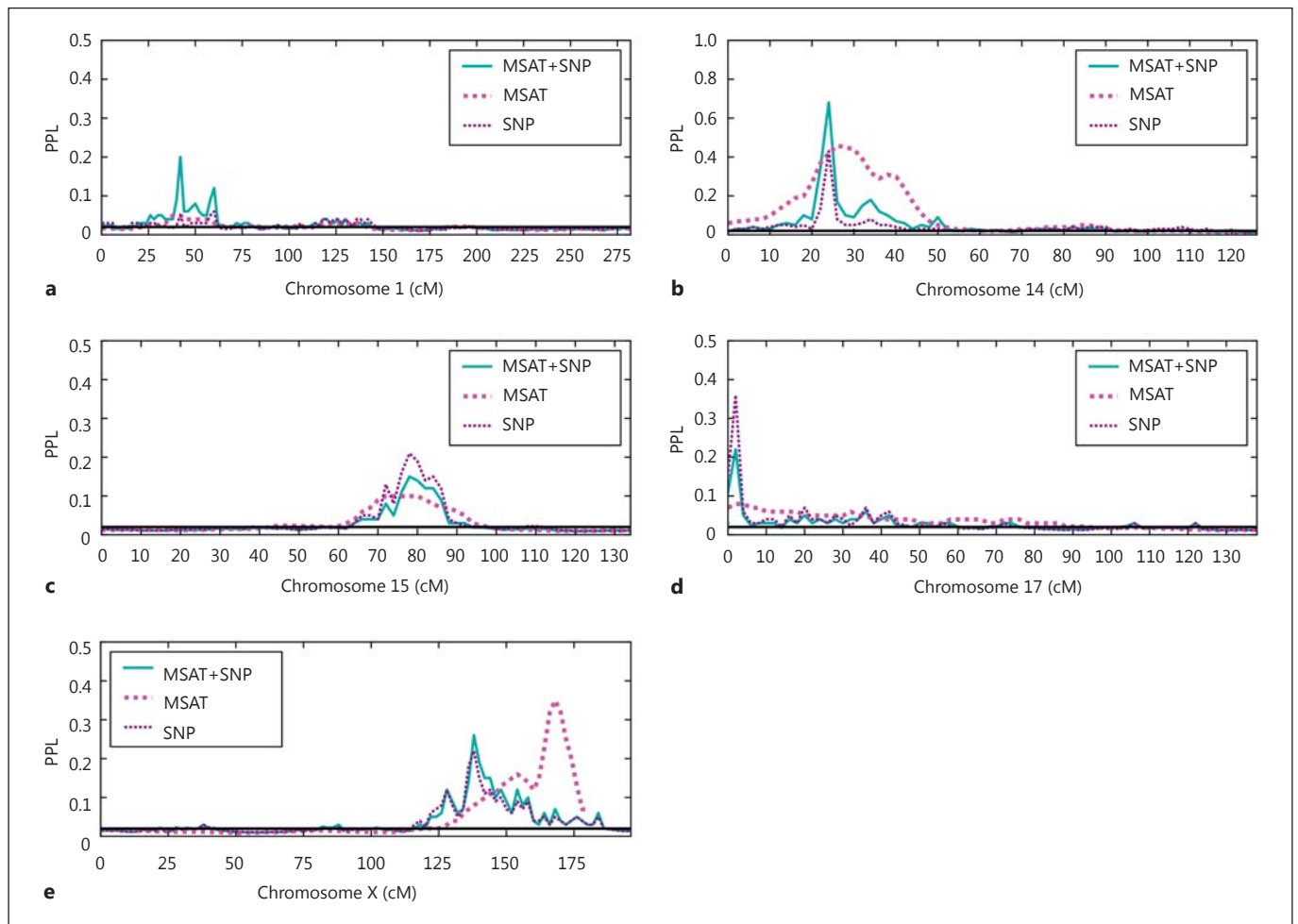
**Fig. 3.** Genome-wide linkage results (PPL) for each of the three marker maps, including all available families and individuals for all three maps. The horizontal line at 0.02 represents the prior probability of linkage; values below this line indicate evidence against linkage.

tions. Figure 2 shows the same comparison but based on the Baseline<sup>+</sup> data subset. Here it is clear that consideration of both MSATs and SNPs together yields substantial increases in IC relative to SNPs alone: SNP mean IC = 0.61, SD = 0.02, range 0.34–0.65; MSAT+SNP mean IC = 0.76, SD = 0.05, range 0.36–0.86. (For comparison, the MSAT statistics for the Baseline<sup>+</sup> group were MSAT mean IC = 0.48, SD = 0.10, range 0.08–0.75.) Note that the IC levels are lower overall for the MSAT and SNP maps comparing Baseline<sup>+</sup> to Baseline, precisely because Baseline<sup>+</sup> contains individuals with missing genotypes on one or the other platform, while Baseline does not.

Finally, in order to assess the impact of the level of increased IC on linkage results, we carried out linkage analysis for each of the three maps (fig. 3) using the full Real Data set. While the results based on the three maps are correlated, there are also salient differences. These include differences in the magnitudes of peaks (e.g. on

chromosome 14) as well as changes in the rank ordering of peaks from highest to lowest (e.g. comparing peaks on chromosomes 17 and X between the SNP and MSAT+SNP maps). They also include dramatic differences in localization, particularly on chromosomes 14 and X, in both cases corresponding to map regions where the MSAT IC is quite low. Thus a low IC can result not only in under- or overestimates of linkage signals, but also in a substantial displacement of signals along the chromosome. Figure 4 shows sample comparisons for selected chromosomes.

Of course, there is no way to know the true locations of relevant genes. However, it is clear that the ‘take home’ messages of a study based on the MSAT+SNP map would have been quite different from the original results based on the MSAT map alone. Because going from MSAT to SNP to MSAT+SNP involves increasing IC, we can treat the results based on the latter as yielding ‘best estimate peak’ (BEP) findings. Using the SNP map compared to the MSAT



**Fig. 4.** Linkage results on selected chromosomes. Note the different y-axis scales (in **a**, **c–e** it is 0, ..., 0.5, while in **b** it is 0, ..., 1.0). The horizontal line at 0.02 represents the prior probability of linkage; values below this line indicate evidence against linkage.

map, the peak on chromosome 14 is somewhat higher but considerably narrower and somewhat displaced, while the BEP shows a similar localization to SNPs alone but is considerably higher. On chromosome 17, the MSAT map misses the peak altogether relative to the BEP, while the SNPs alone seem to overestimate the size of the peak. On chromosome X, the MSATs place the peak quite far from the BEP, while the SNPs perform similarly to the BEP.

## Discussion

We have demonstrated a substantial increase in IC based on a combined MSAT+SNP map, compared to either the MSAT map alone or the SNP map alone, in real-

istic settings (Baseline<sup>+</sup> comparisons). These changes lead to substantially different linkage findings comparing either the SNP or MSAT+SNP results to the MSAT results, which are typical of older linkage studies. While it is impossible to know whether the modest peaks obtained in the new analyses are ‘real’ or not, they are certainly more accurate insofar as they are based on substantially higher IC than the MSAT-only analyses. This exercise makes clear that an earlier generation of linkage studies might have failed to live up to expectations in part simply because the markers available at the time when many of these studies were done were not sufficiently informative.

But these results also illustrate that the older MSAT data sets still have value, particularly insofar as not all samples originally genotyped will be available for regeno-

typing with a new panel. Of course all results shown here are a function in part of the particulars of this data set, but the data considered here are typical in many respects of the majority of psychiatric linkage studies in the NRGR and many other older linkage data sets as well. Thus these older genotype sets are still highly valuable and should not be ignored even when new SNP data are available. Of note, too, is that the lower IC inherent in the MSAT map is not due to an intrinsic problem with MSATs *per se*, but rather to the fact that the MSAT maps tended to be quite sparse.

We conclude from this exercise that the failure of linkage studies to yield genes for many psychiatric data sets is likely due, at least in part, to a low efficiency in finding peaks and a poor localization as a result of reliance on marker maps with inadequate IC. Of course, many of these studies may also have been underpowered due to a small sample size, leading to an interest in using meta-analysis to combine results across studies. However, the deleterious effects of low IC would need to be removed prior to meta-analysis, which would not be able to adjust for failures to correctly estimate or localize peaks within individual studies due to inadequate marker coverage in the primary data.

Moreover, while the focus of this paper is on IC rather than the genetics of schizophrenia *per se*, even in this small sample we have some new and potentially interesting findings, particularly under the linkage peak on chromosome 14. The MSAT peak is quite broad, on the order of 30 Mb, and encompasses nearly 200 genes. By contrast, the MSAT+SNP peak is less than 6 Mb and contains fewer than 20 genes. This smaller interval contains three GWAS signals of interest. Two SNPs (rs17111920 and rs915071) have previously been identified in a study of schizophrenia and bipolar disorder [9]. (Recall that our sample includes only families in which at least one individual has schizophrenia with a strong affective component.) While rs915071 is more than 100 kb from the nearest gene, rs17111920 is approximately 29 kb from miR-4307. This miRNA has been identified in a study of human neural precursors [10], and so could be of importance in early neural development, but it is located very near the edge of the linkage peak. Perhaps of greatest interest is the GWAS signal at rs2038256, located more centrally under the linkage peak, which has been associated with brain glutamate concentration in an imaging study conducted in patients with multiple sclerosis [11]; alterations in brain glutamate have also been observed in imaging studies of individuals with schizophrenia [reviewed in 12]. This signal is within 10 kb of *FOXG1*,

the gene for a member of the forked-head transcription factor family that plays an important role in the neurogenesis of glutamatergic neurons [13] and has been identified as the causative factor in some variants of Rett syndrome and related neurodevelopmental disorders [reviewed in 14].

Additionally, while our sample is extremely small for purposes of detecting allelic association, of the 9 SNPs (out of 559,855) that showed PPL scores  $\geq 5\%$  (online suppl. table S3), rs5957600 is notable because it occurs under the second largest linkage peak in this sample on Xq24. This SNP is in *GLUD2*, which plays a role in the recycling of glutamate during neurotransmission and has been associated with Parkinson disease and amyotrophic lateral sclerosis [reviewed in 15]. The remaining genes in this table have no known relationships to the brain, neurological processes, or neuropsychiatric disorders, with the exception of *SLIT3* which is important in axonal guidance and has been implicated in schizophrenia [16], major depression [17], and autism [18].

In aggregate, our results are relevant as the scientific field returns to a focus on ‘co-segregation’ (linkage) within families as a way to cope with the volume of data generated by high-throughput sequencing experiments. Restricting our attention to sequence variation under linkage peaks makes sense only if linkage analysis is able to correctly rank-order loci by relative strength of evidence for or against linkage at any given peak, in addition to providing an accurate localization. Given the large investments already made in pedigree collection and, in many cases, extensive and very careful phenotyping, surely there is sufficient value to be added to these older collections by regenotyping available samples on more modern genotyping platforms, or even (as costs continue to fall) using whole-genome sequencing technologies.

### Acknowledgements

This work was supported by National Institutes of Health grants R01 MH086117 and U24 MH068457 and Ohio Supercomputer Center grant PCCR0002. We also wish to thank the investigators who contributed these data sets to the NRGR and the families who participated in these studies.



## References

- Escamilla MA, Ontiveros A, Nicolini H, Raventos H, Mendoza R, Medina R, Munoz R, Levinson D, Peralta JM, Dassori A, Almasy L: A genome-wide scan for schizophrenia and psychosis susceptibility loci in families of Mexican and Central American ancestry. *Am J Med Genet B Neuropsychiatr Genet* 2007; 144B:193–199.
- Escamilla M, Hare E, Dassori AM, Peralta JM, Ontiveros A, Nicolini H, Raventos H, Medina R, Mendoza R, Jerez A, Muñoz R, Almasy L: A schizophrenia gene locus on chromosome 17q21 in a new set of families of Mexican and Central American ancestry: evidence from the NIMH genetics of schizophrenia in Latino populations study. *Am J Psychiatry* 2009; 166: 442–449.
- Vieland VJ, Walters KA, Lehner T, Azaro M, Tobin K, Huang Y, Brzustowicz LM: Revisiting schizophrenia linkage data in the NIMH repository: reanalysis of regularized data across multiple studies. *Am J Psychiatry* 2014; 171:350–359.
- Matisse TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FC, Kennedy GC, Kong X, Murray SS, Ziegler JS, Stewart WC, Buyske S: A second-generation combined linkage physical map of the human genome. *Genome Res* 2007; 17:1783–1786.
- 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.
- Vieland VJ, Seok SC, Burian J, Catalyurek U, O'Connell J, Segre A, Valentine-Cooper W: KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. *Hum Hered* 2011; 72:276–288.
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; 58:1347–1363.
- Huang Y, Thomas A, Vieland VJ: Employing MCMC under the PPL framework to analyze sequence data in large pedigrees. *Front Genet* 2013; 4:59.
- Wang KS, Liu XF, Aragam N: A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res* 2010; 124:192–199.
- Goff LA, Davila J, Swerdel MR, Moore JC, Cohen RI, Wu H, Sun YE, Hart RP: Ago2 immunoprecipitation identifies predicted micro-RNAs in human embryonic stem cells and neural precursors. *PLoS One* 2009; 4:e7192.
- Baranzini SE, Srinivasan R, Khankhanian P, Okuda DT, Nelson SJ, Matthews PM, Hauser SL, Oksenberg JR, Pelletier D: Genetic variation influences glutamate concentrations in brains of patients with multiple sclerosis. *Brain* 2010; 133:2603–2611.
- Poels EM, Kegeles LS, Kantrowitz JT, Javitt DC, Lieberman JA, Abi-Dargham A, Girgis RR: Glutamatergic abnormalities in schizophrenia: a review of proton MRS findings. *Schizophr Res* 2014; 152:325–332.
- Kumamoto T, Toma K, Gunadi, McKenna WL, Kasukawa T, Katzman S, Chen B, Hanashima C: Foxg1 coordinates the switch from nonradially to radially migrating glutamatergic subtypes in the neocortex through spatiotemporal repression. *Cell Rep* 2013; 3: 931–945.
- Guerrini R, Parrini E: Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. *Epilepsia* 2012; 53:2067–2078.
- Shashidharan P, Plaitakis A: The discovery of human of GLUD2 glutamate dehydrogenase and its implications for cell function in health and disease. *Neurochem Res* 2014; 39:460–470.
- Shi Y, Zhao X, Yu L, Tao R, Tang J, La Y, Duan Y, Gao B, Gu N, Xu Y, Feng G, Zhu S, Liu H, Salter H, He L: Genetic structure adds power to detect schizophrenia susceptibility at SLIT3 in the Chinese Han population. *Genome Res* 2004; 14:1345–1349.
- Glessner JT, Wang K, Sleiman PM, Zhang H, Kim CE, Flory JH, Bradfield JP, Imielinski M, Frackelton EC, Qiu H, Mentch F, Grant SF, Hakonarson H: Duplication of the SLIT3 locus on 5q35.1 predisposes to major depressive disorder. *PLoS One* 2010; 5:e15463.
- Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, Leyva N, Konidari I, Gentry RC, Hulme WF, Booven DV, Mayo V, Hofmann NK, Schmidt MA, Martin ER, Haines JL, Cuccaro ML, Gilbert JR, Pericak-Vance MA: Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* 2014; 5:1.