Analysis of Shared Haplotypes amongst Palauans Maps Loci for Psychotic Disorders to 4q28 and 5q23–q31

Corneliu A. Bodea[a]   Frank A. Middleton[b]   Nadine M. Melhem[c]   Lambertus Klei[d]
Youeun Song[c]   Josepha Tiobech[e]   Pearl Marumoto[e]   Victor Yano[e]
Stephen V. Faraone[b]   Kathryn Roeder[a]   Marina Myles-Worsley[f]   Bernie Devlin[c]
William Byerley[g]

[a]Department of Statistics, Carnegie Mellon University, Pittsburgh, Pa., [b]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, N.Y., [c]Department of Psychiatry, University of Pittsburgh School of Medicine, and [d]University of Pittsburgh Medical Center, Pittsburgh, Pa., USA; [e]Palau Ministry of Health, Koror, Republic of Palau; [f]University of Washington School of Public Health, Seattle, Wash., and [g]Department of Psychiatry, University of California San Francisco, San Francisco, Calif., USA

Abstract

To localize genetic variation affecting risk for psychotic disorders in the population of Palau, we genotyped DNA samples from 203 Palauan individuals diagnosed with psychotic disorders, broadly defined, and 125 control subjects using a genome-wide single nucleotide polymorphism array. Palau has unique features advantageous for this study: due to its population history, Palauans are substantially interrelated; affected individuals often, but not always, cluster in families; and we have essentially complete ascertainment of affected individuals. To localize risk variants to genomic regions, we evaluated long-shared haplotypes, ≥10 Mb, identifying clusters of affected individuals who share such haplotypes. This extensive sharing, typically identical by descent, was significantly greater in cases than population controls, even after controlling for relatedness. Several regions of the genome exhibited substantial excess of shared haplotypes for affected individuals, including 3p21, 3p12, 4q28, and 5q23–q31. Two of these regions, 4q28 and 5q23–q31, showed significant linkage by traditional LOD score analysis and could harbor variants of more sizeable risk for psychosis or a multiplicity of risk variants. The pattern of haplotype sharing in 4q28 highlights *PCDH10*, encoding a cadherin-related neuronal receptor, as possibly involved in risk.   © 2016 S. Karger AG, Basel

Introduction

Variation in DNA is a substantial source of risk for schizophrenia and related psychotic disorders [1, 2]. Genetic studies, therefore, can contribute to our understanding of their neurobiology. Specific discoveries have

Bernie Devlin
Department of Psychiatry
University of Pittsburgh School of Medicine
Pittsburgh, PA 15213 (USA)
E-Mail devlinbj@upmc.edu

William Byerley
Department of Psychiatry, University of California San Francisco
San Francisco, CA 94143 (USA)
E-Mail william.byerley@ucsf.edu

come slowly, apparently because of the complex genetic architecture of psychotic disorders. At the population level, risk arises from variation scattered throughout the genome that impacts either the function or expression of many genes or functional RNA and has a range of effect size and allele frequency.

Recent research demonstrates constraints on effect size as a function of allele frequency: common variants, for example, collectively play a large role in risk [2–4], while the effect of any one variant is small [5–7]. Individual rare variants surely have a larger range of effect on risk than common variants. Yet, even if a rare variant has a substantial impact, its role can been difficult to prove from genetic data alone due to its rareness [8, 9]. One class of rare variants, recurrent copy number variants (CNVs), has been implicated by their occurrence in cases at a significantly higher rate than that found in massive control databases [10–13]. More relevant to this study, some variants have been implicated due to their segregation with affection status in extended pedigrees. A classic example is a balanced chromosomal translocation, t(1;11)(q42.1;q14.3), disrupting *DISC1*, which generates a fusion product of DISC1 and Boymaw [14, 15]. This rare variant shows significant co-segregation with schizophrenia, bipolar disorder and depression in a large family [16]. Recent studies have also shown that variation is DISC1 is also associated with other psychiatric phenotypes such as cognitive dysfunction and drug abuse [17–20]. Findings of linkage from individual pedigrees are few. More commonly, signals arise from non-parametric linkage analyses of many small pedigrees [21–23], from which it can be challenging to identify risk loci.

At a fundamental level, linkage analysis relies on a mapping between identity by descent (IBD) segments of the genome and how they assort with affection status in the pedigree(s). For a single extended pedigree, a large score for linkage arises when a substantial number of cases are IBD for the same portion of the genome – they share an extended haplotype – while a far smaller fraction of unaffected family members share that haplotype, the expected number depending on the penetrance of the risk variant. This fundamental logic leads to a variation on linkage analysis, IBD mapping, which asks if there is a significant excess of IBD-sharing among cases, relative to controls, at a population level [24]. Here again, the data evaluated are shared haplotypes, which are inferred to be IBD. Like more standard linkage analysis, IBD mapping's success in discovering risk loci in human complex disorders has also been limited.

We hypothesize that a hybrid approach could be effective, combining more standard linkage with results from analysis of shared haplotypes, regardless of known pedigree relationships, especially when applied to a comprehensive sampling of a population such as we have obtained from the nation of Palau. Palauans originated from East Asians some 3,000 years ago [25, 26] and the current size of the population is roughly 20,000 individuals, of which two thirds are native Palauans. Within that population we have diagnosed 203 Palauan cases with schizophrenia and related psychotic disorders, representing almost complete ascertainment of affected individuals in the population [27, 28]. The rate of psychotic disorders in Palauan adults, about 2% in adults [27], is twofold higher than most populations [29]. Affected subjects are often related, and pedigrees of related, affected subjects can be constructed [27, 28]; yet, viewed from another perspective, due to Palauan population history, it is not unusual for two randomly drawn subjects to be 5th degree or closer relatives. Indeed, it is possible to represent the vast majority of affected individuals by a single extended pedigree [30], although the pedigree requires a multiplicity of founders. Therefore, to form smaller pedigrees in Palau requires some arbitrary partitions of related subjects, which can make interpretation of results from linkage mapping and such pedigrees challenging.

In the research reported here, after genotyping Palauan case and control subjects for a set of single nucleotide polymorphisms (SNPs) spanning the genome and phasing those genotypes into haplotypes genome-wide, we analyze haplotype sharing to identify genomic regions that could harbor variation affecting risk. To do so, we first explored patterns of haplotype sharing among nominally related individuals, such as first cousins, who are expected to share 1/8th of their genomes IBD. Dividing the genome into independent bins 5, 10, or 20 Mb in length, the total length of shared haplotypes for all known relative pairs was matched to their expected IBD. The total length of shared 10-Mb haplotypes matched best to the expected IBD pattern. Next, by evaluating sharing of these 10-Mb haplotypes among all Palauan subjects, we detected multiple regions of notable haplotype sharing among cases as compared to that among controls or between cases and controls. Two of the regions demonstrated significant linkage using traditional LOD score analysis. Comprehensive sequence analysis to characterize variation across these regions will be required before we can identify variation likely to cause risk for psychotic disorders.

Materials and Methods

Samples and Genotyping

As described elsewhere [31], we genotyped DNA from blood samples of all known affected subjects (n = 203) and a set of control subjects (n = 125) who had no history of psychiatric treatment and were ≥3° relatives to any affected individual. Diagnoses were made from information obtained from direct interview, using the Schedule for Affective Disorders and Schizophrenia-Lifetime Version [32], and this information was supplemented by a detailed review of psychiatric medical records. Final best estimate diagnoses – assigned by 2 clinicians not involved in ascertainment – were based on DSM-IV criteria. Research protocols and procedures were approved by institutional review boards at each of the sites in the US and the Republic of Palau. All subjects provided written informed consent to participate after receiving a full explanation of the study in both English and Palauan.

Of the 203 affected individuals, 113 were diagnosed with schizophrenia and 6 with probable schizophrenia; 33 were schizoaffective, depressed (n = 3) or bipolar type (n = 19); 28 were diagnosed with psychosis not otherwise specified; 18 were diagnosed with bipolar disorder, 9 of whom had psychotic symptoms; and 5 had diagnoses aligned with schizophrenia or bipolar disorder.

DNA samples were genotyped using Affymetrix Genomewide Human SNP Array 5.0, which seeks to characterize genotypes for 500,568 SNPs. Quality control was conducted at the individual and SNP levels. Samples were retained for the study only if ≥ 98% of their SNPs produced genotypes. Retained SNPs met the following criteria: genotypes for ≥99% of samples; minor allele frequency ≥0.020 (estimated by count and assuming subjects were independent); Hardy-Weinberg statistic with p value >0.005. After quality assessment, 549 samples remained and were characterized by 267,754 SNPs.

Haplotype Sharing

To identify shared haplotypes, genotyped SNPs were jointly phased and imputed using Beagle [33]. The phasing ignored family relationships, yet we found phase to be accurate by evaluating segregation of haplotypes in families. The genome was scanned using moving bins of 5, 10, and 20 Mb, at 1-Mb intervals, analyzing the distribution of haplotypes formed by SNPs amongst affected and unaffected individuals, simultaneously, to identify regions of unusual sharing of haplotypes. In Results, we focus on 10-Mb haplotypes and loci in which at least 7 subjects shared the same haplotype exactly (as shown in Results, such sharing was rare). Once a region of interest was identified, the region of maximum sharing among the affected subjects is defined; i.e. ≥10 Mb, for a 10-Mb window.

Results from such sharing can be dependent, e.g. 11 subjects share a 10-Mb haplotype and 10 of them share a haplotype in an adjacent bin offset by 1 Mb. When important, we noted such clustering in the results, and we merged regions of sharing (online suppl. material; for all online suppl. material, see www.karger.com/doi/10.1159/000450726) for some analyses. This approach relies on inferred haplotypes; when phase is erroneous, it yields false negatives. To characterize segment sharing without declaring phase, we analyzed subsets of subjects using shared segment analysis (SSA) [34, 35]. For a given set, such as a pedigree, SSA searches the genome for regions of identity by state sharing among all or a subset of the subjects based on their genotypes, not haplotypes.

DASH

To search the genome without boundaries, we employed DASH [36], which builds on graph theory and phase information for subjects to infer clusters of individuals likely to share a long haplotype (see suppl. material for implementation). Analyses of these haplotype data using DASH did not reveal additional regions of the genome with unusual haplotype sharing (data not shown). Largely, the results for sharing using fixed boundaries were indistinguishable from those from DASH because once haplotype sharing region was detected within a 10-Mb bin, it was extended to its limit. Occasionally, DASH would highlight very long haplotype sharing among a subset of affected subjects and miss the somewhat smaller shared region for a larger group of subjects, an example being the 4q28.3 region.

Estimating Pairwise Relatedness of Subjects

Interpretation of haplotype sharing is enhanced by knowledge of relationships among subjects. While close relationships were typically known, more distant relationships often were not. For this reason, we estimated the kinship between pairs of individuals using the total length of all inferred IBD segments using IBD_{half} [37] (online suppl. material). In brief, pairwise relatedness was estimated from inferred IBD segments and then refined using a smoothing algorithm known as treelets [38, 39].

Estimating whether Shared 10-Mb Haplotypes Are IBD

To infer whether 10-Mb shared haplotypes are IBD, we analyzed data using IBDLD [40]. Using a hidden Markov model, IBDLD estimates the probability a pair of related individuals is IBD for a given locus from genotype data and pedigrees or estimated relationships amongst individuals. IBDLD estimates background linkage disequilibrium (LD) and adjusts for it in the model.

### Estimating Total Number of Meioses Separating Cases

A statistic of some utility for pedigree-based linkage analysis is the total number of meioses separating all affected subjects in the pedigree [see 34, 35]. Counting meioses in a pedigree is straightforward. The complex interrelationships among Palauans, however, made simple counting impossible in many cases, and thus this number was approximated (online suppl. material).

Results

Estimating Familial Relationships from Haplotypes

We first examined the pattern of sharing for haplotypes of lengths 5, 10, and 20 Mb, assessing the fraction of the genome shared by a pair of subjects of a given *nominal* relationship, as determined by established pedigrees [28]. The fraction of the genome shared by pairs of individuals as 5-Mb haplotypes typically exceeds their degree of relationship (fig. 1); that fraction is somewhat closer to the degree of relationship for total sharing of 10-Mb haplotypes, and it underestimates degree of relationship for 20-Mb haplotypes. That the total sharing of 5- or 10-Mb haplotypes typically exceeds the pedigree-based relationship

Fig. 1. Pedigree relatedness versus genetically inferred relatedness. Genetic relatedness is computed as the proportion of the genome shared through IBD segments of lengths greater than 5, 10, and 20 Mb, respectively. We expect genetically inferred relatedness to be greater than the pedigree level because the small effective population size of Palau, probably throughout its history, has induced substantial cryptic relatedness.
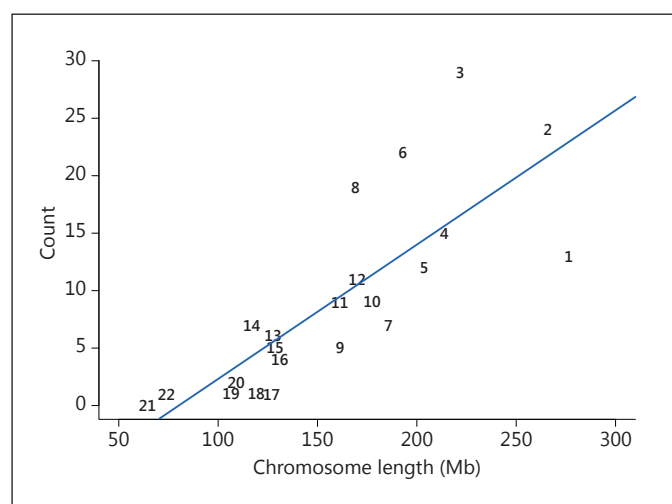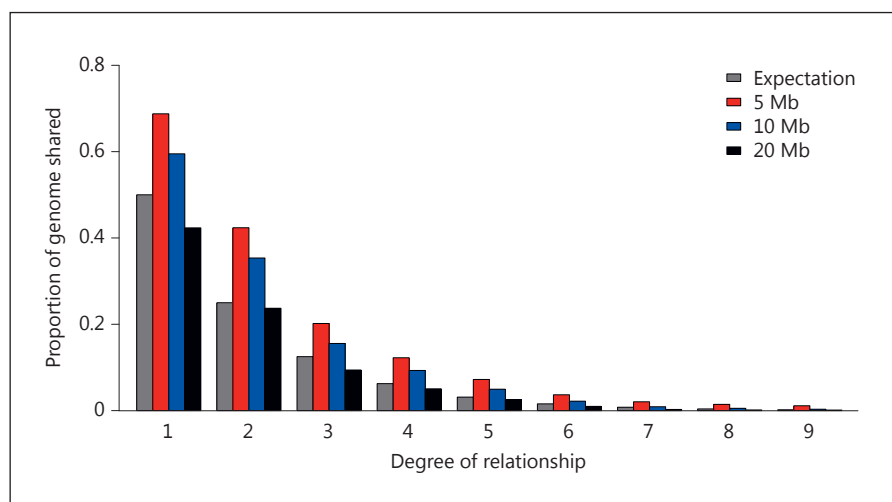
Fig. 2. Count of unique clusters of large haplotypes found in ≥7 cases versus the length of the chromosome.

of a pair of individuals is consistent with population history of Palau, which was established by a relatively small number of founders and experienced several subsequent bottlenecks. This history increases cryptic relatedness within the population and makes Palauans more closely related than their immediate pedigree relationships suggest. These results are also consistent with our earlier study that finds substantial runs of homozygosity in Palauan genomes [31]. For this reason, and because the relationships of many subjects was unknown, we estimated relationships amongst all pairs of individuals based on their total length of shared haplotypes and used these es-

timates for further analyses, except for linkage analyses. When we have evaluated how the total sharing of 5-, 10- and 20-Mb haplotypes relates to a smoothed estimate of relatedness, which is not based on haplotype cutoffs, the haplotype sharing slightly underestimated relationships and was most disparate for 20-Mb haplotypes, as expected (data not shown).

Pattern of Haplotype Sharing

We next evaluated the pattern of haplotype sharing among case and control subjects to identify regions with an unusual pattern of sharing. For this and all subsequent analyses, we focus on 10-Mb haplotypes. Chromosome X showed the greatest abundance of matching, far more than the autosomes, consistent with its smaller effective population size. Because of this excess, which occurs throughout the chromosome, we excluded chromosome X from further analyses. For the autosomes, chromosomes 3, 6, and 8 showed some evidence for excess haplotype sharing as a function of their length, which could reflect reduced recombination and extended LD, while chromosome 1 showed a dearth (fig. 2). There are 12 regions where 15 or more subjects shared a 10-Mb haplotype (i.e. ≥4.6% of 328 subjects), 10 involving either chromosome 3 or 6. As noted in Materials and Methods, these regions were not necessarily independent, and indeed we concatenated two regions on chromosome 6 containing the MHC genes, jointly spanning ≈25–36 Mb, and another two spanning ≈71.5–82.5 Mb. Likewise, on chromosome 3, three regions were combined to span 43–57 Mb and three combined to span from ≈78 to 90 Mb.
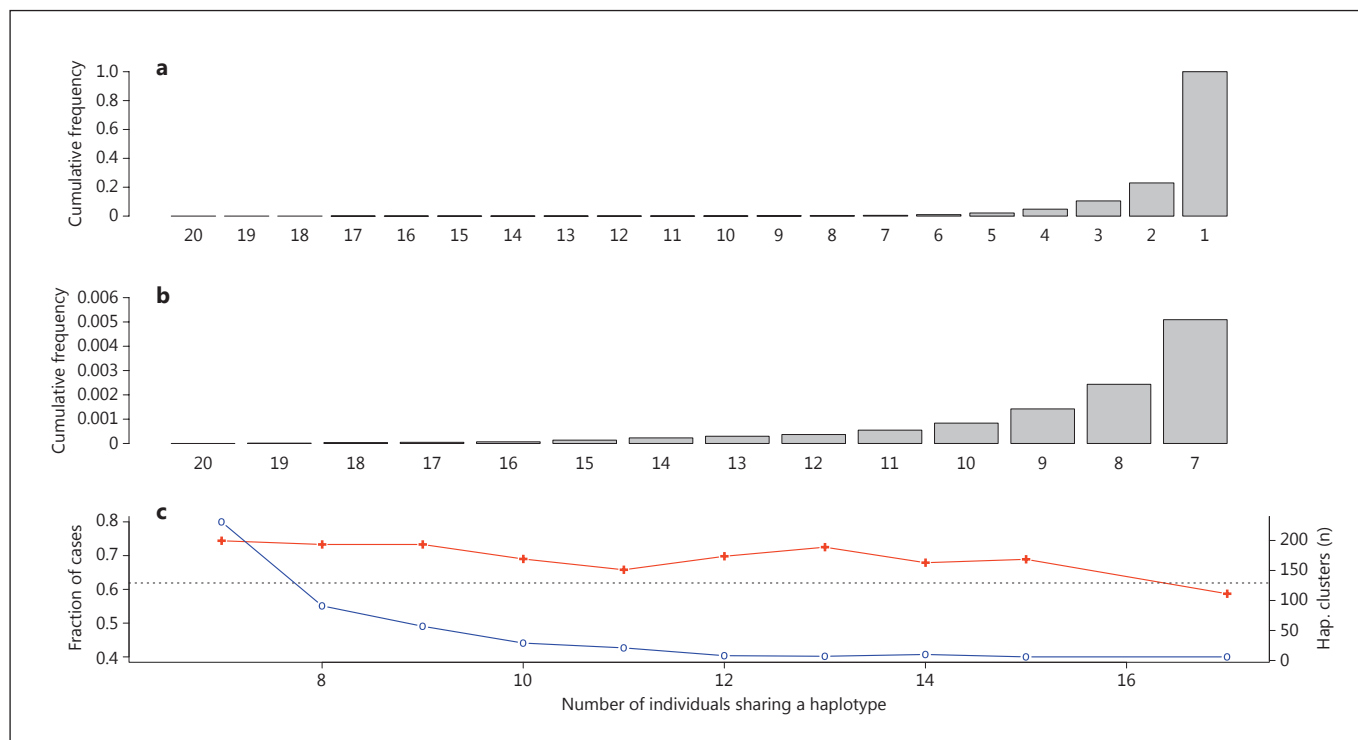
Fig. 3. Cumulative distribution of subjects sharing large haplotypes in the sample (**a**, **b**), and distribution of subjects sharing large haplotypes versus fraction of those subjects who have a psychotic disorder (**c**). Reference line is the expected fraction, based on the sample of 203 affected individuals and 125 controls (i.e. 203/328).

We next asked if sharing a haplotype 10 Mb or more was unusual in Palau either in controls or cases and how this varied by degree of relationship. After sampling 100 bins, out of the 2,351 10-Mb bins across the autosomes, a randomly selected bin yielded 350.1 unique haplotypes in cases, on average, or 1.72 ± 0.11 (mean ± SD) unique haplotypes per case subject. Control samples yielded a somewhat larger number of unique haplotypes per subject, 1.91 ± 0.06. These results imply that there are over a million 10-Mb haplotypes distributed amongst cases and controls in the 2,351 bins.

The difference between cases and controls in the count of unique haplotypes probably arose because a pair of case samples was, on average, more closely related than a pair of control samples: for case pairs, average relationship was 0.020; for control pairs, it was 0.016 (1° = 0.5, 2° = 0.25, … , 5° = 0.03125, 6° = 0.015625, …). To evaluate this conjecture, we randomly sampled 250,000 pairs of subjects, then sampled a 10-Mb bin and determined if the pair of subjects shared a haplotype: 595 out of 250,000 (0.0024) draws yielded a pair of identical haplotypes. Expressed as the fraction of total case-case, control-control

and case-control pairs, the rate of matching was 0.0036 (366/96,053), 0.0015 (54/36,215), and 0.0017 (195/117,732), respectively. To determine if sharing was a function of relatedness alone, or if it was also a function of case status, we then fitted a logit model to these data: the outcome was yes/no for matching, and it had two predictors, estimated relatedness of pair and their case status. The degree of relatedness was the best predictor of matching (z = 41.61; p < 10−16) and it accounted for 19% of the variability in matching (pseudo-R^2). Even after accounting for relatedness, however, case-case pairs were 1.64 times more likely to share a haplotype than control-control pairs (z = 3.16; p = 0.0016), whereas case-control pairs were not more likely to share 10-Mb haplotypes than control-control pairs (z = 1.87; p = 0.062). Still, the case status accounted for only 0.2% of the variance in matching.

Regions of Unusual Haplotype Sharing amongst Affected Subjects

As illustrated in figure 3, it was unusual for 7 or more subjects to share a 10-Mb haplotype (henceforth a 'haplotype cluster'; fig. 3a, b). Yet, such sharing did not neces-

Fig. 4. Average estimated relatedness of affected individuals sharing a long haplotype versus the estimated total number of meioses separating the affected subjects. The fitted line explains 70% of the variance in estimated average relatedness. Numbered points fall outside of the 95% confidence band of the fitted line; these clusters of subjects show greater average relatedness than expected, which is consistent with their shared large haplotype conferring greater risk for psychotic disorders than other shared haplotypes depicted by circles.
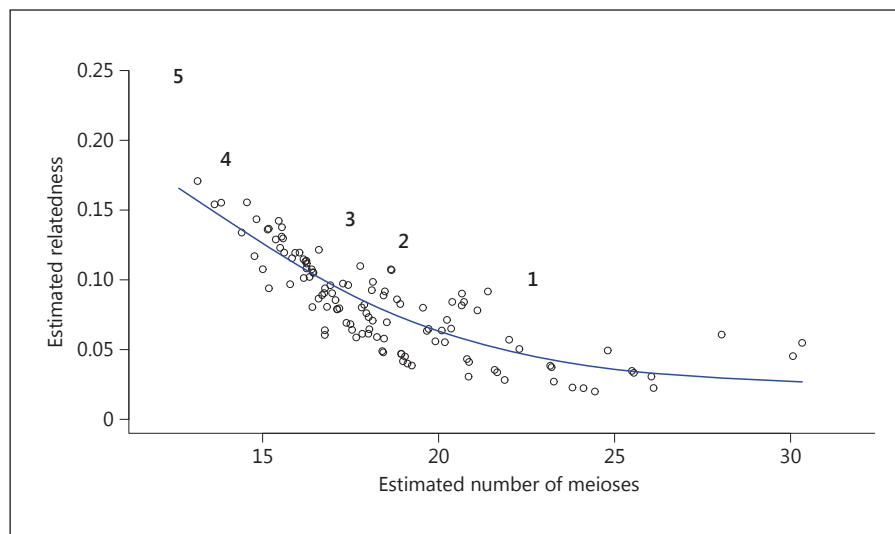
sarily indicate a region affecting risk for psychotic disorders; there were haplotype clusters that were not biased toward case-case sharing: for example, there were 6 haplotype clusters in which 16 or more subjects shared a 10-Mb haplotype, and each was characterized by a larger fraction of controls than would be expected based on case/control representation (fig. 3c). Case-biased haplotype clusters were common only when 15 or fewer subjects carried the haplotype (fig. 3c). When 15 subjects shared a haplotype, at least 3 controls also always carried it, making these haplotypes also unlikely to bear risk alleles. At 14, however, there were two haplotype clusters shared by 13 cases and 1 control, one occurring on chromosome 3 at 48–58 Mb and the other also on chromosome 3 at 76–86 Mb. Note that earlier we identified both overlap regions as having an excess of such sharing. Nonetheless, such an event was rare in the data. Enumerating the outcomes in terms of haplotypes shared by two or more subjects, 14 or more subjects shared a 10-Mb haplotype in only 0.023% of the outcomes (fig. 3a, b), regardless of case-control status.

Following the logic (and statistical basis) for genetic linkage analysis, we assume here that genomic regions shared by the greatest number of cases, relative to controls, would have the highest likelihood of harboring risk variation. To highlight regions that could be involved in risk, we describe in some detail haplotype clusters represented by 7 or more affected subjects and no more than 2 control subjects (online suppl. table 1), which occurs 0.16% of the time. Of these, again following the logic of linkage analysis, clusters represented by 9–13 cases and 2

or fewer controls would be the most likely to harbor risk variants because such clusters are quite rare (occurring 0.03% of the time) and more strongly biased for case representation.

If a haplotype were to bear one or more variants affecting risk, the probability it will be transmitted to and shared by closely related, affected subjects would be proportional to its penetrance. Therefore, variants having a substantial impact on risk should be found in pedigrees multiplex for affected individuals, an expectation born out repeatedly for autosomal dominant Mendelian disorders. For this reason, we asked if haplotype clusters in online supplementary table 1, specifically those with 7–13 affected subjects per cluster, were associated with unusually dense family structure. We first estimated the average relationship among affected individuals in each haplotype cluster and the total number of meioses separating the affected subjects. When the relationship between these two variables was fitted by a cubic spline (fig. 4), the model explained a substantial amount of the variance ($R^2 = 0.70$). For most haplotype clusters, the average relationship amongst affected subjects declined steadily as the total number of meioses separating them increased (fig. 4). Thus, most haplotype clusters were homogeneous in their familial properties. Five observations, however, fell outside of the 95% confidence envelope for the curve, and each of these outliers was consistent with closer familial relationships among affected individuals (fig. 4, numbered points).

Bodea et al.
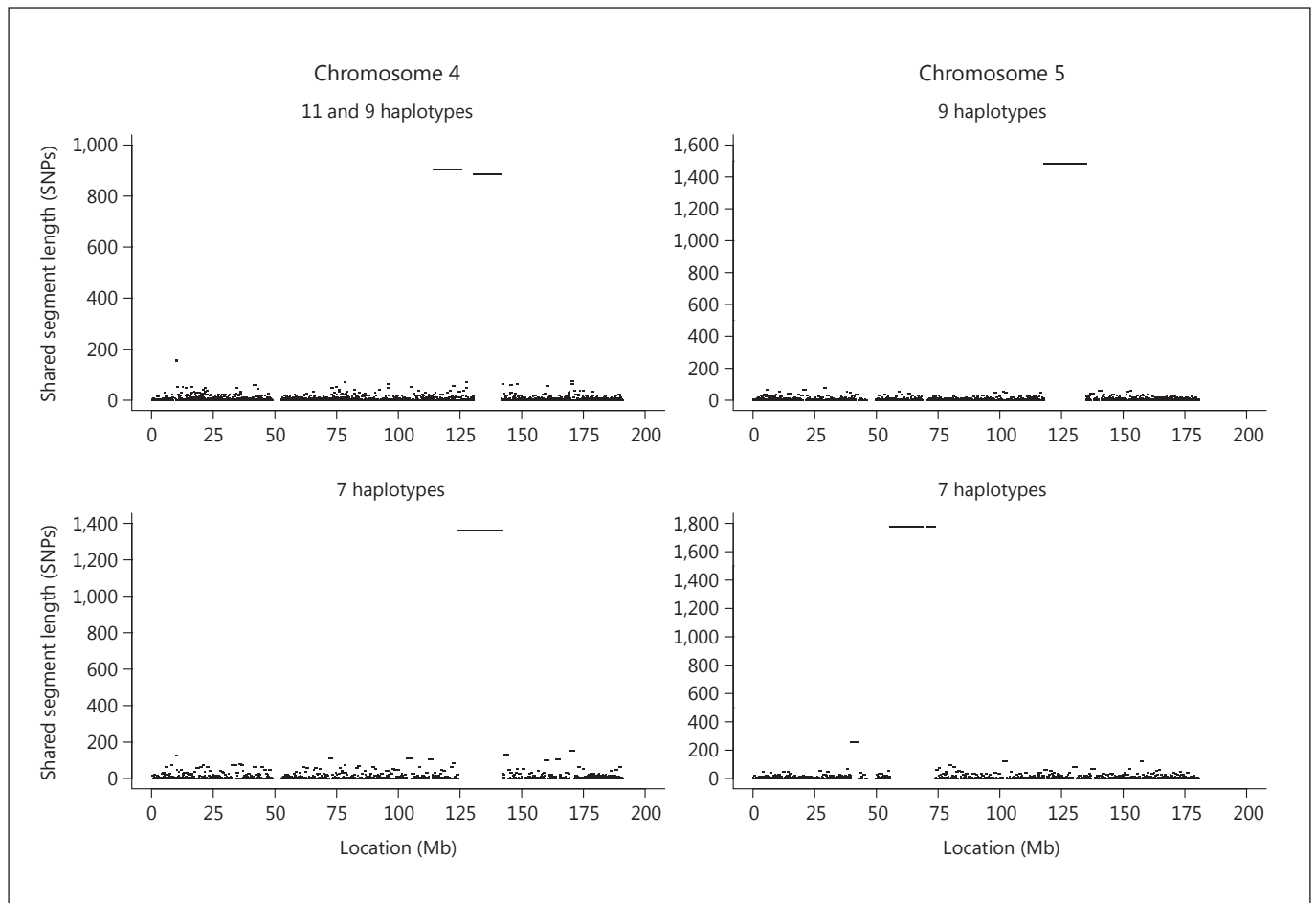
Fig. 5. Haplotype regions shared by subjects highlighted in figure 3 as assessed by identity by state segment sharing. Analysis over all autosomes is presented in supplementary figure 1.

Properties of Outliers and Linkage Analysis of Shared Haplotypes

Of the outliers in figure 4, two (No. 1 and 3) were from essentially the same pedigree and almost overlap in terms of genomic location. Cluster 1 had 11 affected individuals who shared a haplotype from chromosome 4 spanning 131–141 Mb (online suppl. table 1). Two control subjects also shared this haplotype. Cluster 3, spanning 115–125 Mb on chromosome 4, held 9 affected subjects who were a perfect subset of cluster 1; no controls shared this haplotype. Of note, the 9 affected subjects shared by cluster 1 and 3 fell in a multiplex pedigree (K1584) previously identified by epidemiological studies in Palau [27]. SSA [34, 35] was used to determine if 4q28.3 is the only region of the genome thus shared among these 11 and 9 affected subjects. For these affected subjects, only the same 4q28 region was identified (fig. 5; online suppl. fig. 1).

Remarkably, another of the outlier haplotype clusters, cluster 5 (fig. 4), also fell in 4q28 and its localization strongly overlapped cluster 1. This haplotype covered 125–139 Mb and involved 7 cases and no controls. The haplotypes for clusters 1 and 5 were quite different, with the exception of a 0.5-Mb region, chr4:133,731,053–134,238,544, where they were identical (fig. 6). The shared 0.5-Mb subhaplotype was common in Palau, 11.6% in controls, yet the odds ratio for carrying (vs. not) this subhaplotype in cases versus controls was 1.94 (20% in cases vs. 11.6% in controls; MQLS statistic = 13.3, p value = 0.00027), and the difference in haplotype frequency was not lowered substantially after removing the sets of subjects in clusters 1 and 5 (17% in cases, 11.6% in controls). Indeed when we inspected the 4q28 region for other groups of subjects who share a 10-Mb haplotype, there were two others (online suppl. table 1), and both shared
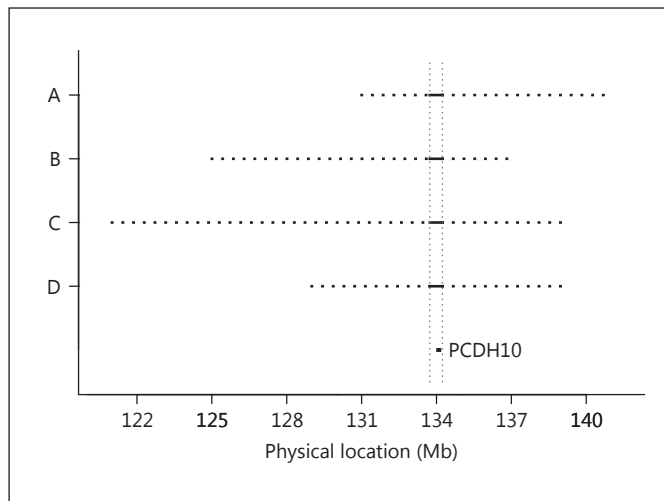
Fig. 6. Pattern of haplotype sharing at 4q28.3 for four clusters of affected subjects. The only gene found in the shared 0.5 Mb region is *PCDH10*.

this subhaplotype, but were distinct for 10-Mb haplotypes (fig. 6). Both haplotype clusters had 9 cases and 0 controls, the first shared haplotype spanning 121–137 Mb and the other spanning 128–138 Mb.

The two remaining outlier haplotype clusters, clusters 2 and 4, involved the q arm of chromosome 5. The haplotype for cluster 2 covered 119–134 Mb and involved 9 cases and no controls. The haplotype for cluster 4 covered 56–73 Mb and involved 7 cases and no controls. There was no overlap amongst the affected subjects involving clusters 2, 4, or 5 and no overlap of subjects from those three clusters with subjects in clusters 1 and 3. For clusters 2 and 4, SSA analysis showed that the haplotype clusters were unique for each set of subjects (fig. 5; online suppl. fig. 1). For cluster 5, however, the same subjects also shared a 13-Mb region of chromosome 9 IBS, extending from roughly 2 to 15 Mb (fig. 5; online suppl. fig. 1).

Analysis of IBD

To infer whether 10-Mb shared haplotypes are IBD for the 5 outlier clusters, we analyzed data using IBDLD [40]. IBDLD inferred IBD for a given locus from genotype data and estimated relationships amongst pairs of individuals. The shared 10-Mb haplotypes were inferred to be IBD for a majority of pairs of subjects (online suppl. fig. 2). Because IBDLD estimates IBD status for pairs of subjects, however, as opposed to the entire cluster of subjects, its analysis apparently was not yet fully efficient; i.e. the IBD relationship is not always transitive (A and B share a seg-

ment, and B and C share the same segment, but A and C do not share the segment); if we forced the equality, it was apparent all subjects were IBD for a shared haplotype (online suppl. fig. 2). We next randomly selected 5 other clusters for which 7 affected subjects shared a 10-Mb haplotype. Results from IBDLD again show that the shared haplotype for these clusters is IBD for most pairs of subjects.

Linkage Analysis

Pedigrees for subjects from each of these five clusters were derived from a unified extended pedigree of Palau subjects described elsewhere [30]. This multifounder pedigree captured the relationships amongst most affected subjects in Palau and all of the subjects in each of the clusters; thus, it was straightforward to form a subpedigree for each cluster. Included in each pedigree were all affected subjects in the cluster, their parents, offspring, siblings and any intervening relatives required to connect affected individuals in a complete pedigree. Haplotype information was included when known, as was affection status. For the calculation of LOD score for linkage, we treated the shared ($\geq$10 Mb) haplotype as an allele and pooled all other haplotypes as the alternative allele; the risk haplotype frequency was set to its relative frequency in control subjects or $1/(260 + 2)$, whichever was larger; and we assumed a parametric model in which the risk allele acted dominantly and had a penetrance of 50%, and the probability of psychosis while not carrying the haplotype was 0.02. Using these assumptions, the LOD scores for linkage for clusters 1–5, respectively, were 3.32 (p $\approx$ 4.6 $\times$ 10^{-5}), 3.59 (p $\approx$ 2.4 $\times$ 10^{-5}), 2.64 (p $\approx$ 2.4 $\times$ 10^{-4}), 1.37 (p $\approx$ 6.0 $\times$ 10^{-3}), and 2.07 (p $\approx$ 1.0 $\times$ 10^{-3}). The calculation for cluster 1 ignored an affected individual in K1584 who carried 9.5 Mb of the shared haplotype, which no other controls shared, and who was therefore IBD with her other relatives; accounting for this subject raised the LOD score at 4q28.3 to 3.85 (p $\approx$ 1.3 $\times$ 10^{-5}). Thus, two regions exhibited significant linkage, given that 3.3 is a conventional threshold for significance of linkage in extended pedigrees [41].

Discussion

Most populations studied genetically for psychotic disorders are of European descent. The population of Palau is quite different in this regard. Palau was populated only 3,000 years ago [25, 26] and the population has experienced several bottlenecks even in historical times. The genomes of its peoples reflect that history, being

characterized by extensive LD, which in some instances can be unusually large for markers separated by many megabases [42], and by extensive runs of homozygosity (ROH) [31]. The symptoms of psychotic disorders in Palauans, however, are indistinguishable from those found the world over [32], and indeed we have previously reported evidence that some features of the genetics of risk are similar too. For example, CNVs shown to be risk factors in European populations [10–13] are risk factors in Palau as well [43, 44], including 1q21.1, 15q11.2 (BP1-BP2), and SLC1A1 [12]. Furthermore, subjects with psychotic disorders tend to have longer and more ROH than do controls from the same population [31], a feature seen in various populations [45–50]. Still, recessive inheritance does not seem to explain a substantial portion of risk [31]. To augment our knowledge of the genetics of psychotic disorders in Palau, the work reported here evaluated sharing of haplotypes to localize variants affecting risk in a dominant or additive manner.

Quite long segments of chromosomes were sometimes shared among Palauan subjects who were nominally unrelated. We targeted sharing of very long haplotypes, ≥10 Mb, which analyses suggest are almost always identical by descent. The best predictor for a pair of subjects sharing a haplotype was their degree of relatedness, with close relatives being much more likely to share a long haplotype. Even after taking relatedness into account, 2 subjects who were diagnosed with a psychotic disorder were significantly more likely to share such a haplotype than were case-control or control-control pairs, consistent with the genetic basis for these disorders in Palau and supporting the efficacy of haplotype sharing for genetic analysis.

We found numerous regions of the genome that exhibited substantial haplotype sharing amongst cases (online suppl. table 1). By the metric of largest number of cases sharing the same haplotype, relative to controls, two independent regions on chromosome 3 emerged as notable, both having 13 cases and one control sharing the same 10-Mb haplotype (48–58 and 76–86 Mb). These regions harbored many genes (online suppl. table 1), and it is impossible without sequence data to identify which genes – if any – contain variants affecting risk. Complicating interpretation of these results is that they fell on chromosome 3 in regions that, by our analyses (fig. 2), showed reduced diversity of haplotypes, presumably due to recombination suppression or possibly natural selection.

A result apparent from mathematical analysis [51], and one observed empirically from a host of pedigrees segregating Mendelian risk variants, is that a variant with large impact on risk should, on average, segregate in a pedigree multiplex for affected individuals than a variant of modest effect on risk. Using this fact, we searched for clusters of affected subjects who shared the same haplotype and also demonstrated unusually close relationships among them (fig. 5). Five such clusters were identified, four were independent. The most noteworthy was a cluster of 11 affected subjects and 2 controls who share a 10-Mb haplotype at 4q28.3, with a twelfth affected subject sharing 9.5 Mb of the haplotype. Parametric linkage analysis of 10 of these affected individuals, who fell in a single pedigree, yielded an LOD score of 3.85, a significant result. Another of these clusters yields significant linkage to 5q23–q31, based on 9 subjects and no controls sharing a 15-Mb haplotype (LOD = 3.59).

The 4q28.3 linkage region is even more unusual because three other haplotype clusters fell therein (fig. 6), a result not seen elsewhere in the genome. The full haplotypes for these 4 clusters of subjects were independent, but all shared a 0.5-Mb subhaplotype covering the gene PCDH10. This shared subhaplotype is relatively common in Palau, almost 12% of control subjects carry a copy, but its presence is associated with an almost twofold greater risk for psychotic disorders (p value = 0.00027). It is thus possible a variant on the subhaplotype conveys risk, but it is impossible to rule out other models of risk as well, including additional or independent risk factors.

PCDH10 is a gene of some interest regarding psychiatric disorders. Disruption of this gene has been associated with risk for autism [52], and it incurs a de novo missense mutation detected by a recent exome sequencing study of schizophrenia [9]. The protein encoded by this gene, a calcium-dependent adhesion molecule, has myriad functions including a role in the neuronal circuit formation during development. Indeed a recent study has shown that the interaction of myocyte enhancer factor 2, fragile X mental retardation protein, and Pcdh10 contributes to the process of excitatory synapse elimination in mouse neurons [53].

A substantial number of other plausible candidate genes reside in these two linkage regions (online suppl. table 2). There are only 14 named genes falling in 4q28.3, 131–141 Mb, yet it would be no simple task to say which, if any, affect the risk for psychotic disorders.

Of the 76 genes falling in 5q23–q31, 119–134 Mb, several are of particular interest. One such gene, MEGF10, falls in the interval and it encodes a protein playing a critical role in synapse remodeling and therefore neural circuit refinement [54]. There is some evidence for its involvement in the risk for schizophrenia [55].

While rare risk variants could drive patterns of haplotype sharing, especially when many affected subjects share the same haplotype, it is also possible that sharing can be explained by one or more common variants that confer risk. While the Palau sample and population are not powered to detect common variants of subtle impact on risk, such as those reported in the largest genome-wide association study [7], we did compare the two regions of significant linkage from Palau with those GWAS results. Of the 108 significant loci identified by the Schizophrenia Working Group of the Psychiatric Genomics Consortium, none mapped to the 131- to 141-Mb interval on chromosome 4 or to the 119- to 134-Mb interval of chromosome 5. Still, it is possible that a substantial proportion of the regions highlighted in online supplementary table 1 owe their excess sharing of haplotypes among cases to the impact of common variants on risk.

Given the two significant findings of linkage, we wondered why we missed these linkages in earlier linkage studies using short tandem repeat (STR) loci. A more complete diagnostic inventory was a major contributing factor, especially given the fact that some affected individuals were not yet diagnosed or ascertained in our earlier studies, while the uneven distribution of STRs was a minor contributor. The 5q31 linkage region was detected, although with diminished signal, in a previous multipoint linkage analysis [56].

A weakness of our approach is that errors of phase or genotyping could disrupt true patterns of haplotype sharing and produce false negatives. An example is a recent finding from our group about a deletion CNV that disrupts *SLC1A1* in 7 individuals with psychosis [44]. Looking specifically at this region and set of subjects, the deletion altered the pattern of haplotype sharing for 13 SNPs (16 prior to quality control). Moreover, apparently because of the deletion, the genotype caller labeled some SNPs heterozygous when they were hemizygous; the true status was confirmed both by our inspection of the genotype clusters and by molecular analyses reported previously [44]. For other genomic regions, it is also possible that CNVs could affect apparent haplotype sharing estimates, usually reducing estimates but occasionally increasing them. While we systematically screened for CNVs in the same subjects analyzed in the present study, we cannot rule out the possibility that rare events could exert influence on some of the findings (online suppl. table 1).

The obvious next step is whole-genome sequencing of 2 or 3 subjects having selected haplotypes. This will refine our understanding of what variants could be relevant to risk, discriminating between the impact of rare mutations versus common variants, both intra- and intergenic. It seems reasonable to predict, based on past experience, that sorting out these effects will be just as challenging as interpreting association and linkage signals generated by earlier genetic technology.

Statement of Ethics

The authors have no ethical conflicts to disclose.

References

1 Wray NR, Gottesman II: Using summary data from the Danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front Genet 2012;3:118.

2 Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al: Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet 2013;45:984–994.

3 International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748–752.

4 Lee SH, DeCandia TR, Ripke S, Yang J; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC), et al: Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 2012;44:247–250.

5 Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, et al: Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013;45:1150–1159.

6 Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium: Genome-wide association study identifies five new schizophrenia loci. Nat Genet 2011;43:969–976.

7 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014;511:421–427.

8 Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, O'Dushlaine C, et al: A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506:185–190.

9 Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al: De novo mutations in schizophrenia implicate synaptic networks. Nature 2014;506:179–184.

10 International Schizophrenia Consortium: Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008;455:237–241.

11 Stefansson H, Rujescu D, Cichon S, Pietiläi-nen OP, Ingason A, Steinberg S, et al: Large recurrent microdeletions associated with schizophrenia. Nature 2008;455:232–236.

12 Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al: CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 2014;505:361–366.

13 Rees E, Walters JT, Chambert KD, O'Dushlaine C, Szatkiewicz J, Richards AL, et al: CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and *SLC1A1* and duplications at 1p36.33 and *CGNL1*. Hum Mol Genet 2014;23:1669–1676.

14 Zhou X, Chen Q, Schaukowitch K, Kelsoe JR, Geyer MA: Insoluble DISC1-Boymaw fusion proteins generated by DISC1 translocation. Mol Psychiatry 2010;15:669–672.

15 Eykelenboom JE, Briggs GJ, Bradshaw NJ, Soares DC, Ogawa F, Christie S, et al: A t(1;11) translocation linked to schizophrenia and affective disorders gives rise to aberrant chimeric *DISC1* transcripts that encode structurally altered, deleterious mitochondrial proteins. Hum Mol Genet 2012;21:3374–3386.

16 Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ: Schizophrenia and affective disorders – cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. Am J Hum Genet 2001;69:428–433.

17 Rampino A, Walker RM, Torrance HS, Anderson SM, Fazio L, Di Giorgio A, Taurisano P, Gelao B, Romano R, Masellis R, Ursini G, Caforio G, Blasi G, Millar JK, Porteous DJ, Thomson PA, Bertolino A, Evans KL: Expression of DISC1-interactome members correlates with cognitive phenotypes related to schizophrenia. PLoS One 2014;9:e99892.

18 Carless MA, Glahn DC, Johnson MP, Curran JE, Bozaoglu K, Dyer TD, Winkler AM, Cole SA, Almasy L, MacCluer JW, Duggirala R, Moses EK, Göring HH, Blangero J: Impact of *DISC1* variation on neuroanatomical and neurocognitive phenotypes. Mol Psychiatry 2011;16:1096–1104.

19 Gancarz A, Jouroukhin Y, Saito A, Shevelkin A, Mueller LE, Kamiya A, Dietz DM, Pletnikov MV: DISC1 signaling in cocaine addiction: towards molecular mechanisms of comorbidity. Neurosci Res 2016;105:70–74.

20 Xie P, Kranzler HR, Krystal JH, Farrer LA, Zhao H, Gelernter J: Deep resequencing of 17 glutamate system genes identifies rare variants in *DISC1* and *GRIN2B* affecting risk of opioid dependence. Addict Biol 2014;19:955–964.

21 Levinson DF, Holmans P, Straub RE, Owen MJ, Wildenauer DB, Gejman PV, et al: Multicenter linkage study of schizophrenia candidate regions on chromosomes 5q, 6q, 10p, and 13q: schizophrenia linkage collaborative group III. Am J Hum Genet 2000;67:652–663.

22 Holmans PA, Riley B, Pulver AE, Owen MJ, Wildenauer DB, Gejman PV, et al: Genome-wide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. Mol Psychiatry 2009;14:786–795.

23 Ng MY, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T, et al: Meta-analysis of 32 genome-wide linkage studies of schizophrenia. Mol Psychiatry 2009;14:774–785.

24 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;15:319–325.

25 Carson MT: First Settlement of Remote Oceania: Earliest Sites in the Mariana Islands. Berlin, Springer, Springer Briefs in Archaeology, 2014.

26 Liston J: An assessment of radiocarbon dates from Palau, Western Micronesia. Radiocarbon 2005;47:295–354.

27 Myles-Worsley M, Coon H, Tiobech J, Collier J, Dale P, Wender P, Reimherr F, Polloi A, Byerley W: A genetic epidemiological study of schizophrenia in Palau, Micronesia. I. Prevalence and familiality. Am J Med Genet B Neuropsychiatr Genet 1999;88:4–10.

28 Myles-Worsley M, Tiobech J, Blailes F, Middleton FA, Vinogradov S, Byerley W, et al: Familial transmission of schizophrenia in Palau: a 20-year genetic epidemiological study in three generations. Am J Med Genet B Neuropsychiatr Genet 2011;156B:247–254.

29 Jablensky A: The 100-year epidemiology of schizophrenia. Schizophr Res 1997;28:111–125.

30 Klei L, Bacanu SA, Myles-Worsley M, Galke B, Xie W, Tiobech J, et al: Linkage analysis of a completely ascertained sample of familial schizophrenics and bipolars from Palau, Micronesia. Hum Genet 2005;117:349–356.

31 Melhem NM, Lu C, Dresbold C, Middleton FA, Klei L, Wood S, Faraone SV, Vinogradov S, Tiobech J, Yano V, Roeder K, Byerley W, Myles-Worsley M, Devlin B: Characterizing runs of homozygosity and their impact on risk for psychosis in a population isolate. Am J Med Genet B Neuropsychiatr Genet 2014; 165B:521–530.

32 Endicott J, Spitzer RL: A diagnostic interview: the schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 1978;35:837–844.

33 Browning SR, Browning BL: High-resolution detection of identity by descent in unrelated individuals. Am J Hum Genet 2010;86:526–539.

34 Thomas A, Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA: Shared genomic segment analysis. Mapping disease predisposition genes in extended pedigrees using SNP genotype assays. Ann Hum Genet 2008;72:279–287.

35 Knight S, Abo RP, Abel HJ, Neklason DW, Tuohy TM, Burt RW, et al: Shared genomic segment analysis: the power to find rare disease variants. Ann Hum Genet 2012;76:500–509.

36 Gusev A, Kenny EE, Lowe JK, Salit J, Saxena R, Kathiresan S, et al: DASH: a method for identical-by-descent haplotype mapping uncovers association with recent variation. Am J Hum Genet 2011;88:706–717.

37 Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, Mountain JL: Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. PLoS One 2012;7:e34267.

38 Crossett A, Lee AB, Klei L, Devlin B, Roeder K: Refining genetically inferred relationships using treelet covariance smoothing. Ann Appl Stat 2013;7:669–690.

39 Lee AB, Nadler B, Wasserman L: Treelets: an adaptive multi-scale basis for sparse unordered data. Ann Appl Stat 2008;2:435–471.

40 Han L, Abney M: Identity by descent estimation with dense genome-wide genotype data. Genet Epidemiol 2011;35:557–567.

41 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995; 11:241–247.

42 Devlin B, Roeder K, Otto C, Tiobech S, Byerley W: Genome-wide distribution of linkage disequilibrium in the population of Palau and its implications for gene flow in Remote Oceania. Hum Genet 2001;108:521–528.

43 Melhem N, Middleton F, McFadden K, Klei L, Faraone SV, Vinogradov S, et al: Copy number variants for schizophrenia and related psychotic disorders in Oceanic Palau: risk and transmission in extended pedigrees. Biol Psychiatry 2011;70:1115–1121.

44 Myles-Worsley M, Tiobech J, Browning SR, Korn J, Goodman S, Gentile K, et al: Deletion at the SLC1A1 glutamate transporter gene cosegregates with schizophrenia and bipolar schizoaffective disorder in a 5-generation family. Am J Med Genet B Neuropsychiatr Genet 2013;162B:87–95.

45 Gibson J, Morton NE, Collins A: Extended tracts of homozygosity in outbred human populations. Hum Mol Genet 2006;15:789–795.

46 Li LH, Ho SF, Chen CH, Wei CY, Wong WC, Li LY, et al: Long contiguous stretches of homozygosity in the human genome. Hum Mutat 2006;27:1115–1121.

47 Simon-Sanchez J, Scholz S, Fung HC, Matarin M, Hernandez D, Gibbs JR, et al: Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 2007;16:1–14.

48 McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al: Runs of homozygosity in European populations. Am J Hum Genet 2008;83:359–372.

49 Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF: Genomic runs of homozygosity record population history and consanguinity. PLoS One 2010;5:e13996.

50 Wang LS, Hranilovic D, Wang K, Lindquist IE, Yurcaba L, Petkovic ZB, et al: Population-based study of genetic variation in individuals with autism spectrum disorders from Croatia. BMC Med Genet 2010;11:134.

51 He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, et al: Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS Genet 2013;9:e1003671.

52 Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al: Identifying autism loci and genes by tracing recent shared ancestry. Science 2008;321:218–223.

53 Tsai NP, Wilkerson JR, Guo W, Maksimova MA, DeMartino GN, Cowan CW, et al: Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95. Cell 2012;151: 1581–1594.

54 Chung WS, Clarke LE, Wang GX, Stafford BK, Sher A, Chakraborty C, et al: Astrocytes mediate synapse elimination through MEGF10 and MERTK pathways. Nature 2013;504:394–340.

55 Chen X, Sun C, Chen Q, O'Neill FA, Walsh D, Fanous AH, et al: Apoptotic engulfment pathway and schizophrenia. PLoS One 2009; 4:e6875.

56 Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B, et al: Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. Mol Psychiatry 2002;7:689–694.