Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor

Matthew C. Keller[1,2]*, Matthew A. Simonson[1,2], Stephan Ripke[3,4,5], Ben M. Neale[3,4,5], Pablo V. Gejman[6,7], Daniel P. Howrigan[1,2], Sang Hong Lee[8], Todd Lencz[9,10,11], Douglas F. Levinson[12], Patrick F. Sullivan[13,14,15,16], The Schizophrenia Psychiatric Genome-Wide Association Study Consortium[¶]

1 Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, Colorado, United States of America, 2 Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, Colorado, United States of America, 3 Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 4 Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, 5 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, United States of America, 6 Department of Psychiatry and Behavioral Sciences, North Shore University Health System, Evanston, Illinois, United States of America, 7 Department of Psychiatry and Behavioral Sciences, University of Chicago, Chicago, Illinois, United States of America, 8 Queensland Institute of Medical Research, Brisbane, Australia, 9 Department of Psychiatry, Division of Research, The Zucker Hillside Hospital Division of the North Shore – Long Island Jewish Health System, Glen Oaks, New York, United States of America, 10 Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, New York, United States of America, 11 Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York, United States of America, 12 Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, United States of America, 13 Department Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden, 14 Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 15 Department of Psychiatry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 16 Department of Epidemiology, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

Abstract

Autozygosity occurs when two chromosomal segments that are identical from a common ancestor are inherited from each parent. This occurs at high rates in the offspring of mates who are closely related (inbreeding), but also occurs at lower levels among the offspring of distantly related mates. Here, we use runs of homozygosity in genome-wide SNP data to estimate the proportion of the autosome that exists in autozygous tracts in 9,388 cases with schizophrenia and 12,456 controls. We estimate that the odds of schizophrenia increase by ~17% for every 1% increase in genome-wide autozygosity. This association is not due to one or a few regions, but results from many autozygous segments spread throughout the genome, and is consistent with a role for multiple recessive or partially recessive alleles in the etiology of schizophrenia. Such a bias towards recessivity suggests that alleles that increase the risk of schizophrenia have been selected against over evolutionary time.

Introduction

Schizophrenia is a highly (.70–.80) heritable [1] neurodevelopmental disorder that has a lifetime prevalence of ~0.4% [2]. As with most complex disorders, the specific genetic variants that account for a majority of the heritability of schizophrenia remain to be discovered. Two primary factors may explain the difficulty in identifying risk variants. First, the results of genome-wide association studies (GWAS) make it clear that a very large number of genes contribute to schizophrenia risk, and the overall population risk attributable to any one risk variant must be small [3]. Second, although common causal variants almost certainly play an important role in the genetic etiology of schizophrenia [4,5], it is likely that the frequency distribution of schizophrenia risk alleles is biased towards the rare end of the spectrum [5]. Both

of these factors are consistent with selection keeping schizophrenia risk alleles with the largest effects rare, such that no single allele can contribute much to population risk.

If schizophrenia risk alleles have been selected against across evolutionary time (have been under "purifying" selection), another prediction is that schizophrenia risk alleles will be biased towards being recessive. This bias, called directional dominance, occurs in traits subject to purifying selection because selection more efficiently purges the additive and dominant alleles with the strongest effects, leaving the remaining pool of segregating alleles more recessive than otherwise expected [6]. Directional dominance has traditionally been inferred from observations of inbreeding depression, the tendency for offspring of close genetic relatives to have higher rates of congenital disorders and lower fitness [7]. Fitness traits such as survival, reproduction, resistance

Author Summary

Inbreeding occurs when genetic relatives have offspring. Because all humans are related to one another, even if very distantly, all people are inbred to various degrees. From a genetic standpoint, it is well known that inbreeding increases the risk that a child will have a rare recessive genetic disease, but there is also increasing interest in understanding whether inbreeding is a risk factor for more common, complex disorders such as schizophrenia. In this investigation, we used single-nucleotide polymorphism data to quantify the degree to which 9,388 schizophrenia cases and 12,456 controls were inbred, and we tested the hypothesis that people whose genome shows higher evidence of being inbred are at higher risk of having schizophrenia. We estimate that the odds of schizophrenia increase by ~17% for every 1% increase in inbreeding. This finding is consistent with a role for multiple recessive or partially recessive alleles in the etiology of schizophrenia, and it suggests that genetic variants that increase the risk of schizophrenia have been selected against over evolutionary time.

to disease, and predator avoidance tend to show more inbreeding depression than traits under less intense selection [8]. Interestingly, there are numerous reports of inbreeding effects on human complex traits such as heart disease [9], hypertension [10], osteoporosis [11], cancer [12], and IQ [13,14].

Studies that have investigated inbreeding effects on schizophrenia using pedigree data suggest that close (e.g., cousin-cousin) inbreeding is a risk factor [15,16,17,18,19,20], although three studies have failed to find the predicted effect [21,22,23]. However, close inbreeding cannot be a major contributor to schizophrenia risk in industrialized countries given its rarity (<1% of marriages) [24] and the modest increase in the odds of schizophrenia among highly inbred offspring (~2- to 5-fold) [15,16,17,18,19]. Nevertheless, inbreeding is a matter of degree; when distant relatives are considered, everyone is inbred to some degree. It is likely that the parents of the vast majority of people alive today share a common ancestor within ~15 generations [25]. Although such "distant" inbreeding would be prohibitively difficult to detect from pedigrees, it can leave signals in the genome that are detectable using genome-wide single nucleotide polymorphism (SNP) data.

The inbreeding coefficient of an individual, F, is defined as the probability that two randomly chosen alleles at a homologous locus within an individual are *identical by descent* (IBD, identical because they are inherited from a common ancestor) [26]. Homozygosity arising from the inheritance of two IBD genomic segments is termed autozygosity. Most estimates of F assume that marker data are independent, and provide an aggregate measure of homozygosity at measured variants across the genome [27]. Recently, however, several investigators have used *runs of homozygosity* (ROHs; long stretches of homozygous SNPs) to infer autozygosity, and have investigated whether the proportion of the genome that exists in such ROHs, F_{roh}, predicts complex traits [28,29,30,31,32,33,34,35].

Of several alternative estimates of F, including F estimated by treating markers independently and F estimated from pedigree information, Keller, Visscher, and Goddard [25] recently concluded that F_{roh} is optimal for inferring the degree of genome-wide autozygosity and for detecting inbreeding effects. However, given the small variation in genome-wide F_{roh} in unselected samples (e.g., SD ~.005), large sample sizes (e.g.,

>12,000) are necessary to detect inbreeding depression for likely effect sizes in samples not selected for recent inbreeding [25]. Studies investigating the effects of F_{roh} on human complex traits with samples sizes <3,000 and that failed to find significant inbreeding effects [28,33,34,35,36] are likely to have been underpowered. That said, the only study of F_{roh} in schizophrenia [29] found a very large inbreeding effect, but the effect was observed in a small sample ($n = 322$) and was significant only for ROHs caused by common haplotypes.

The present study uses imputed SNP data from 17 schizophrenia case-control datasets (total $N = 21,844$) that are part of the Psychiatric GWAS Consortium (PGC) [3,37] to investigate whether F_{roh} is associated with higher risk of schizophrenia. We also use an ROH mapping approach to investigate whether specific areas of the genome are predictive of case-control status when autozygous. This study represents the largest investigation to date on the potential consequences of autozygosity as estimated using F_{roh}, and may help elucidate the genetic architecture and natural history of schizophrenia.

Results

SNP data from 9,388 schizophrenia cases and 12,456 controls were collected with institutional review board approval from 17 sites in 11 countries (Table 1). Due to the different SNP platforms used across datasets, the number of SNPs remaining after quality control and linkage-disequilibrium pruning procedures (see below) differed substantially between the datasets (column 6 of Table 1). This induced artifactual differences in ROH statistics across datasets and made it impossible to allelically match ROHs across datasets (see Methods). To circumvent these issues, our main analysis concentrated on ROH results from a common set of imputed SNPs, but we also report results from the raw (non-imputed) SNP data. We imputed 1,252,901 autosomal SNPs in each dataset using BEAGLE [38] and HapMap3 as the reference panel [3]. We used extremely stringent imputation QC thresholds that have been shown to achieve accuracy rates similar to those in genotyped SNPs [39], leaving 398,325 high-quality imputed SNPs. We then removed 303,513 SNPs that were in high linkage disequilibrium (LD) with other SNPs. We defined ROHs as being ≥65 consecutive homozygous SNPs in a row (~2.3 Mb) among the remaining 94,812 imputed SNPs [40]. We followed the same procedure for each dataset using the raw data, but defined ROHs as being ≥110 consecutive homozygous SNPs in a row (~1.7 to ~3.2 Mb, depending on the dataset). ROH thresholds were determined empirically (see *Methods*) so as to maximize the significance of the schizophrenia-F_{roh} relationship, but as shown below, results differed little for alternative thresholds. F_{roh} was defined as the proportion of an individual's genome that exists in ROHs. Descriptive statistics of ROHs and F_{roh} across individual and combined datasets are shown in Table 1, and distribution of ROH lengths and F_{roh} are shown in Figure 1 (Figure S1 shows the non-truncated distribution of F_{roh}).

ROH burden results

We regressed case-control status on F_{roh} separately in each of the 17 datasets using logistic regression, controlling for potential confounding factors such as population stratification and SNP quality metrics (see *Methods*). Figure 2 shows the estimated change in odds of schizophrenia for every 1% increase in F_{roh} and the 95% confidence intervals from these 17 logistic regression equations, and Figure S2 shows the same results from an analysis conducted on the raw (non-imputed) SNP data. It should be noted that confidence intervals are symmetric on the log odds scale but

Table 1. Descriptive statistics of SNPs and ROHs (derived from imputed data) across datasets.

Dataset	n	site	platform	SNPs passing QC	LD pruned SNPs passing QC	% imputed SNPs genotyped	avg Froh * 100	SD(Froh) *100	avg ROH length (kb)	SD(ROH length in kb)
AB	1418	UK	Affy 5	362461	180127	37%	0.13	0.21	2458	1897
BON	1778	Germany	Illum 550	494953	269497	21%	0.11	0.17	2424	2055
BULG	1135	Bulgaria	Affy 6	654278	288366	34%	0.21	0.45	3720	4014
CARWTC	3406	UK	Affy 500	365456	184440	37%	0.12	0.37	2563	2629
CAT2	606	USA	Affy 500	384538	234201	34%	0.18	1.05	3248	4107
DK	939	Denmark	Illum 650	533191	277949	20%	0.14	0.36	2544	2766
DUB	1130	Ireland	Affy 6	642723	271644	32%	0.16	0.21	2620	2237
EDI	650	UK	Affy 6	646310	275252	32%	0.11	0.21	2413	2272
MGS2	5163	USA/Australia	Affy 6	638937	280522	35%	0.11	0.21	2420	2021
MUC	785	Germany	Illum 317	295593	202543	20%	0.14	0.39	2901	3317
PORT	561	Portugal	Affy 5	333136	186277	39%	0.61	1.06	4662	4087
SW1	335	Sweden	Affy 5	330113	170520	35%	0.19	0.31	2844	2593
SW2	618	Sweden	Affy 6	661602	275690	31%	0.23	0.37	2911	2782
TOP3	598	Norway	Affy 6	630195	280545	32%	0.15	0.17	2375	1575
UCLA	1334	Netherlands	Illum 550	505922	270851	21%	0.15	0.37	2728	2681
UCL	1009	UK	Affy 5	277652	156506	36%	0.09	0.13	2212	1633
ZHH	379	USA	Affy 500	258470	159220	37%	0.36	0.99	3551	3385
Total	21844	N/A	Imputed	398325	94812	82%[†]	0.15	0.40	2788	2771

[†]82% of imputed SNPs used in overall ROH analysis were genotyped on at least one platform.
doi:10.1371/journal.pgen.1002656.t001

asymmetric on the odds ratio scale shown in Figure 2 and Figure S2. As indicated by the confidence intervals, there was a great deal of variability in the estimates of the *Froh*-schizophrenia association, and none of these 17 odds ratios significantly differed from one. Nevertheless, 13 of the odds ratios were greater than one (i.e.,

consistent with autozygosity being a schizophrenia risk factor) while 4 were less than one, a result inconsistent with chance (exact binomial test, $p = 0.025$). More formally, using a mixed linear effects logistic regression model that treated dataset as a random factor (which also controlled for SNP platform because dataset was
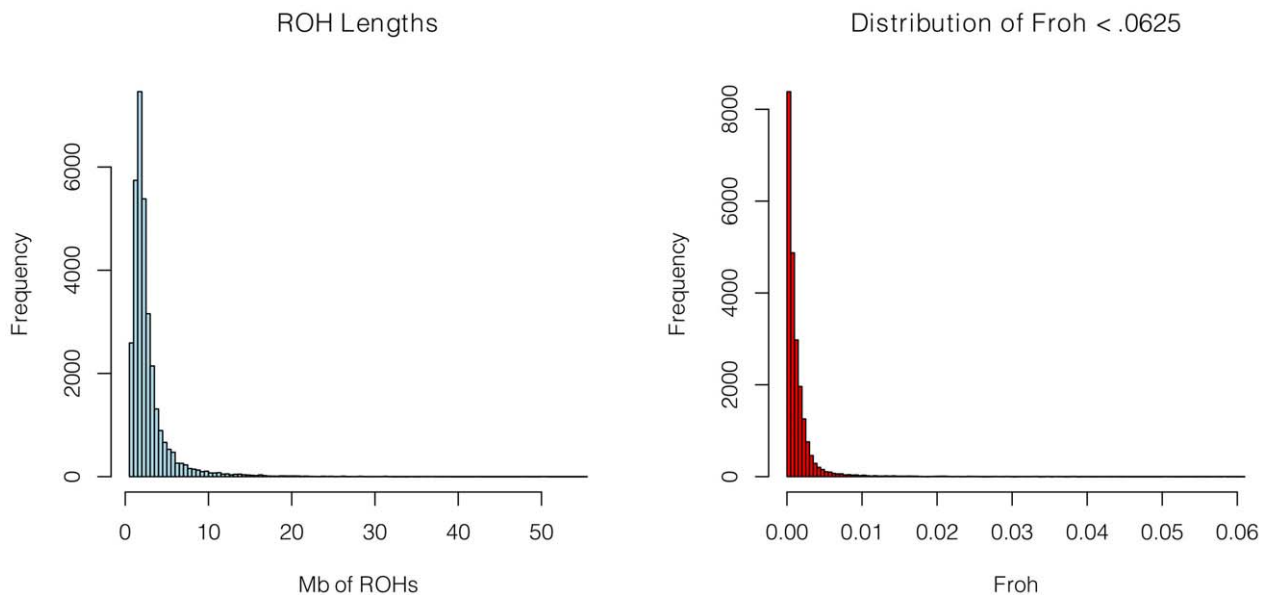

Figure 1. Distributions of ROH Lengths (left) and *Froh* (right) in the total sample. Distributions are based on ROHs from the imputed SNP data. For clarity, the distribution of *Froh* leaves omits 15 individuals who have *Froh*>.0625.
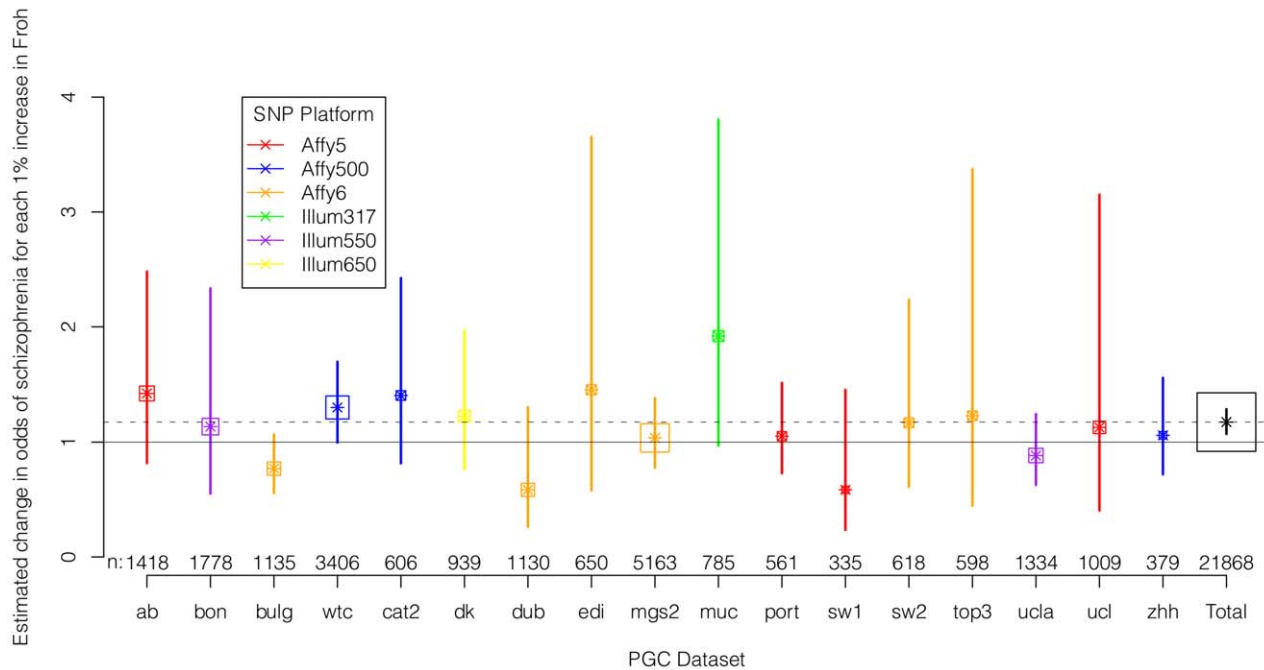doi:10.1371/journal.pgen.1002656.g001

Figure 2. Estimated changes in odds of schizophrenia for each 1% increase in *Froh* (odds ratios; asterisks) and their 95% confidence intervals (bars) across the 17 datasets (colored) and for the total sample (black) from the imputed SNP data. Boxes are proportional to the square root of sample sizes (also shown at the bottom). Dataset names are on the x-axis. Although none of the estimated odds ratios are significantly different from one individually, the overall effect (black) is highly significant.
doi:10.1371/journal.pgen.1002656.g002

nested within each platform), the overall association between schizophrenia and *Froh* in the combined sample was highly significant ($\beta = 16.1$, $z = 3.44$, $p = 0.0006$ in the imputed data, and $\beta = 17.98$, $z = 3.89$, $p = 0.0001$ in the raw data). A slope of *Froh* on schizophrenia of 16.1 is interpreted as saying that for every 0.01 increase in *Froh*, the odds of schizophrenia are multiplied by $e^{0.161} = 1.17$, or increased by 17%.

Several secondary analyses were undertaken to explore the robustness and generality of the *Froh*-schizophrenia association. There was no evidence that the *Froh*-schizophrenia association differed significantly between datasets ($\chi^2 = 0.253$, $p = 0.88$), and the association remained highly significant in 17 models that removed one dataset at a time. To understand if this association was sensitive to the covariates included in the model, we ran additional models that controlled for no covariates, various combinations of covariates, and dataset-by-covariate interactions. In all of these models, the association between *Froh* and schizophrenia remained significant. We also found that our conclusions were insensitive to the SNP threshold used to define ROHs; the association between *Froh* and schizophrenia remained relatively unchanged and significant for all SNP thresholds of ≥40 consecutive SNPs in both the imputed (Figure 3) and raw (Figure S3) data. Finally, both common ROHs ($\beta = 28.5$, $z = 2.51$, $p = .012$), which arose from haplotypes that were observed often in the data, and uncommon ROHs ($\beta = 20.4$, $z = 3.29$, $p = .001$) were predictive of case-control status (see *Methods*).

Autozygosity versus hemizygosity

Copy number variant deletions can create apparent ROHs in SNP data. We could not systematically catalog the overlap between deletions and ROHs in the full dataset because deletion information is not available on the entire sample. However, Levinson and colleagues [41] identified 501,890 deletions (using

their "broad" criteria) in the MGS2 dataset ($n = 5,163$), comprising about one-fourth of the total sample used here. The median length of a deletion in the MGS2 dataset was ~10 kb, whereas the median length of a ROH was ~2,000 kb, suggesting that very few deletions would be long enough to qualify as ROHs. Consistent with this expectation, we found that only 10 of 6,480 ROHs in the MGS2 dataset were possible deletions using the algorithm described by McQuillan et al. [31], which called a ROH a "possible deletion" if its total length was <500 kb after removing deletion regions from ROHs. The percentage of ROHs thus classified (0.15%) was similar to the percentage (0.30%) reported by McQuillan et al. [31]. This percentage is too small to have a meaningful impact on our results, because when we removed a larger percentage of ROHs that were identified as being the largest schizophrenia risk factors (see below), the *Froh*-schizophrenia association remained highly significant. We conclude that ROH results reported above are due to autozygosity rather than hemizygosity.

The effects of close versus distant inbreeding on schizophrenia

A reverse-causation explanation of the *Froh*-schizophrenia association is possible: people who have a higher "load" of schizophrenia risk alleles (and who transmit this risk to offspring) may be more likely to mate with a relative. This counter-explanation to the causal interpretation of the *Froh*-schizophrenia relationship is less likely if the relationship holds not only for close inbreeding, but also for autozygosity caused by distant and almost certainly unintended inbreeding (arising from common ancestors who lived many generations ago). One way to investigate this issue is to remove positive outliers on *Froh* and reassess the *Froh*-schizophrenia relationship. We reran models after dropping a) two individuals with *Froh*>0.125, the approximate equivalent of half-
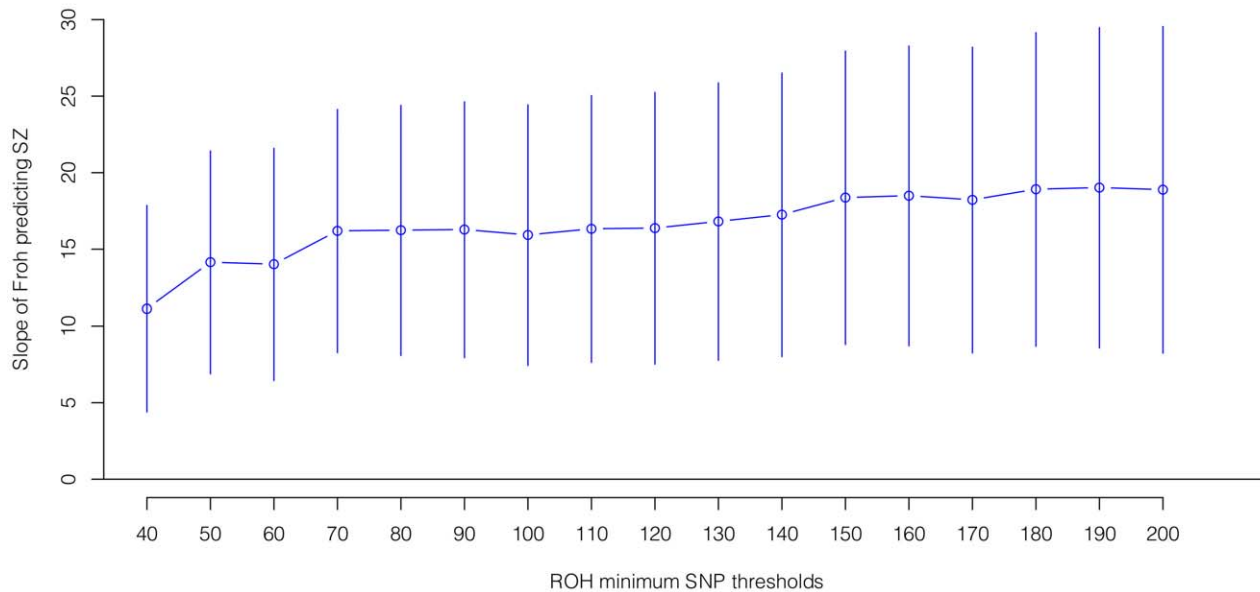
Figure 3. Slope estimates (the change in log odds for a 1% increase in *Froh*; points) and their 95% confidence intervals (bars) of *Froh* from imputed SNP data predicting schizophrenia for different SNP homozygosity thresholds of calling ROHs. Minimum SNP thresholds for full and reduced models are offset for clarity. All ROH thresholds were significant; the most significant result was for ROHs defined as being 65 or more homozygous SNPs in a row.
doi:10.1371/journal.pgen.1002656.g003

sibling inbreeding ($\beta = 15.57$, 95% CI(β) = [25.0, 6.14], $z = 3.24$, $p = 0.001$); b) 15 individuals with *Froh*>0.0625, the approximate equivalent of cousin-cousin inbreeding ($\beta = 15.13$, 95% CI(β) = [26.1, 4.25], $z = 2.73$, $p = 0.006$); c) 56 individuals with *Froh*>0.03125, the approximate equivalent of half-cousin inbreeding ($\beta = 8.43$, 95% CI(β) = [21.43, −4.55], $z = 1.27$, $p = 0.20$); d) 942 individuals with *Froh*>.005, consistent with elevated levels of distant inbreeding ($\beta = 5.17$, 95% CI(β) = [34.84, −24.50], $z = 0.34$, $p = .73$); and e) 6,101 individuals with *Froh* scores above the mean level of *Froh* ($\beta = 66.91$, 95% CI(β) = [139.2, −5.4], $z = 1.81$, $p = .07$). To test whether the change in significance after dropping outliers was due to the *Froh*-schizophrenia association being stronger for individuals with high levels of autozygosity, we included a quadratic term (*Froh*^{2}) in the regression model. In contrast to the highly significant linear term of *Froh*, the quadratic term of *Froh* was non-significant ($p = .09$), suggesting that the effect of autozygosity is linear across the range of *Froh* observed here.

The simple approach—dropping outliers—to distinguishing the effects of distant versus close inbreeding is problematic for two reasons. First, *Froh* is naturally extremely right-skewed (Figure 1 and Figure S1), even in large, simulated populations where close inbreeding is disallowed [25], and so dropping even a small number of outliers greatly reduces the variation in *Froh*, decreases the statistical power to detect an association, and degrades the precision of point estimates. Indeed, there is no evidence that the schizophrenia-*Froh* association changes as outliers are removed, because the original point estimate ($\beta = 16.1$) is contained within every confidence interval above. Thus, the results from dropping outliers demonstrate that the *Froh*-schizophrenia relationship is not driven by a few highly inbred individuals, but do not allow us to distinguish the effects of distant vs. close inbreeding. Second, individuals with high *Froh* can arise by chance from the accumulation of many paths of distant inbreeding [25], and are not necessarily the products of close inbreeding. For example, the distribution of lengths of observed ROHs among individuals with *Froh*>0.0625 is more consistent with inbreeding from common

ancestors living ~6 generations ago than with first cousin inbreeding (Figure 4).

An alternative and more robust approach for assessing the relative importance of distant versus close inbreeding is to compare the effects of short versus long ROHs. We defined *Froh*<5 Mb as the proportion of the autosome in ROHs of length <5 Mb and *Froh*>5 Mb as the converse, with 5 Mb chosen as the threshold because the variances of *Froh*<5 Mb and *Froh*>5 Mb were equal. An autozygous segment spanning <5 Mb should originate from a common ancestor ≥ 10 generations ago on average [41]. The effect of *Froh*<5 Mb ($\beta = 27.6$, $z = 2.23$, $p = 0.026$) was similar to the effect of *Froh*>5 Mb ($\beta = 24.3$, $z = 2.01$, $p = 0.044$), consistent with the hypothesis that autozygosity arising from distant inbreeding is about as much of a schizophrenia risk factor as autozygosity arising from more recent common ancestors.

ROH mapping analysis

The top of Figure 5 shows the $-\log_{10}$ p-values for the 5,742 logistic regressions predicting case-control status from ROHs at each 500 kb bin along the autosome. No regions reached genome-wide significance although two (1p13.2 and 3p24.1) exceeded the "suggestive significance" threshold. Table 2 shows the twelve genes located in these two regions along with their potential functional significances. Neither region has been previously implicated in linkage analyses [42], copy number variant analyses [43], or GWAS meta-analyses [3] of schizophrenia. After recalculating *Froh* with the two suggestively significant regions removed, results of the burden analysis remained essentially unchanged, showing that these regions have only a minor influence on the overall *Froh*-schizophrenia association and suggesting that the effect of autozygosity is diffused across the genome.

The bottom of Figure 5 shows the frequencies of ROHs occurring at each 500 kb bin across the autosome. With one exception, less than 1.5% of the sample had an ROH at each
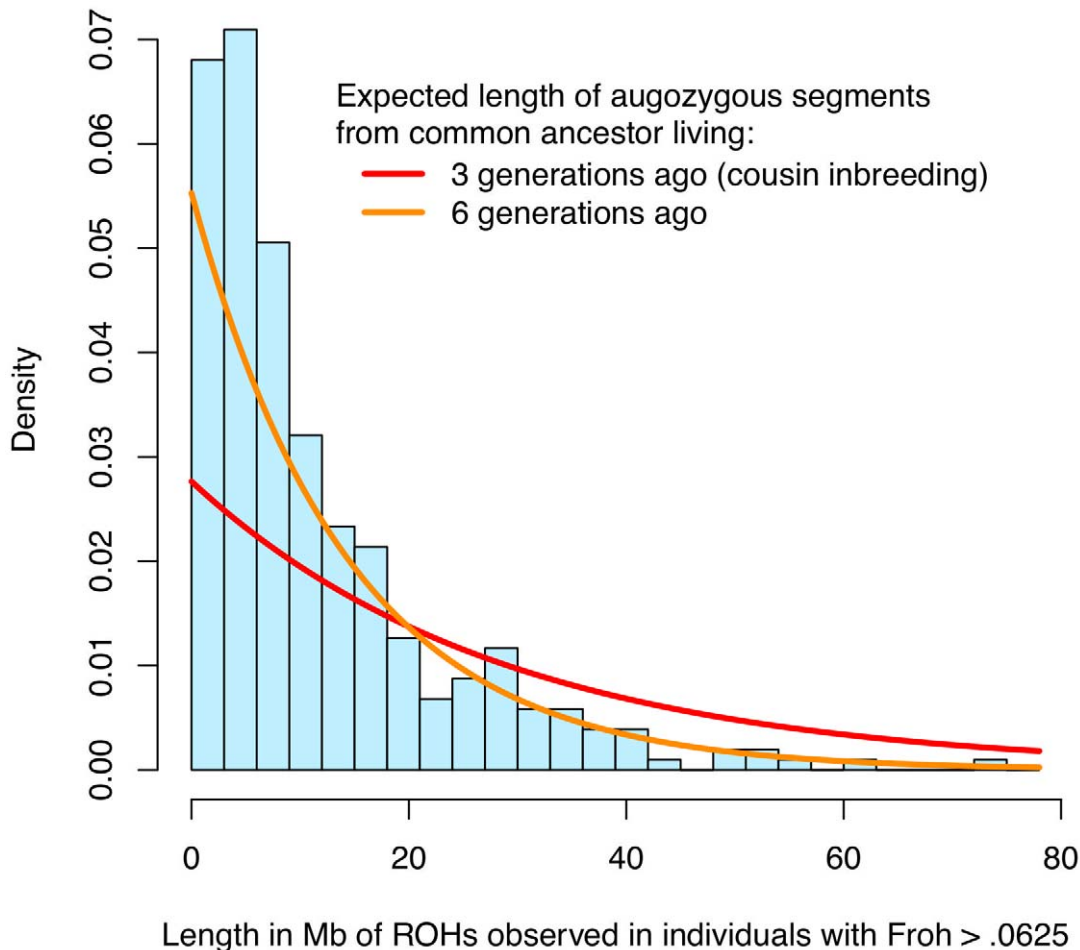
Figure 4. Distribution of ROH lengths for the 15 individuals with *Froh*>.0625 in the sample (blue) and the expected lengths of autozygous segments for different levels of inbreeding (red and orange). Nearby ROHs that were broken up by a possible heterozygous SNP miscall were joined together. Assuming Haldane's recombination model, the length of an autozygous segment should follow an exponential distribution with mean equal to 1/(2×number of generations since the common ancestor) in Morgans. The figure shows that the distribution of ROH lengths among individuals with *Froh*>.0625 is most consistent with autozygosity caused by common ancestors between parents who lived ~6 generations ago.
doi:10.1371/journal.pgen.1002656.g004

region. The exception occurs in the Major Histocompatibility Complex region in 6p21.3, where 15.5% of the sample had an ROH. This high number of ROHs is explained by the low recombination and long, common, geographically-specific haplotypes that occur here [44,45].

Discussion

These results suggest that the odds of schizophrenia increase by ~17% for every 0.01 increase in the proportion of estimated autosomal autozygosity (*Froh*). Given the standard deviation of *Froh* (0.004), this effect is modest, explaining <0.1% of the risk of schizophrenia in outbred populations (Nagelkerke $r = 0.026$). Nevertheless, this effect implies that close inbreeding is a significant risk factor for schizophrenia. Cousin-cousin inbreeding is predicted to increase the odds of schizophrenia 2.74-fold (by 174%) and second-cousin inbreeding is expected to increase the odds of schizophrenia 1.29-fold (by 29%). These estimates are roughly in line with previous reports on schizophrenia from samples selected based on pedigree inbreeding [15,16,17,18,19,20] and similar to the increased risk of major birth defects following close inbreeding [46]. Given that second cousin or closer

inbreeding occurs frequently in several world cultures, and that progeny from such unions account for about 10% of the world's population [47], autozygosity may be an important risk factor for schizophrenia worldwide.

The apparent effect of autozygosity on schizophrenia suggests that risk alleles that are more dominant have disappeared over evolutionary time at a faster rate than risk alleles that are more recessive. This is consistent with the hypothesis that alleles that increase the risk of schizophrenia have been under purifying (negative) selection [48].

There are three main limitations to the current study. The most important is that this was a mega-analysis of SNP data collected at 17 different sites using six different platforms. The collection and handling of samples, the distribution of samples on plates, and the calling of SNPs differed between and within sites in ways that were impossible to quantify in the analysis. This certainly added noise to the results, reducing the apparent effect size, but also may have introduced subtle biases. We have tried to statistically control for as many of these as possible, but the possibility remains that uncorrected biases made these results appear stronger or weaker than they actually are.
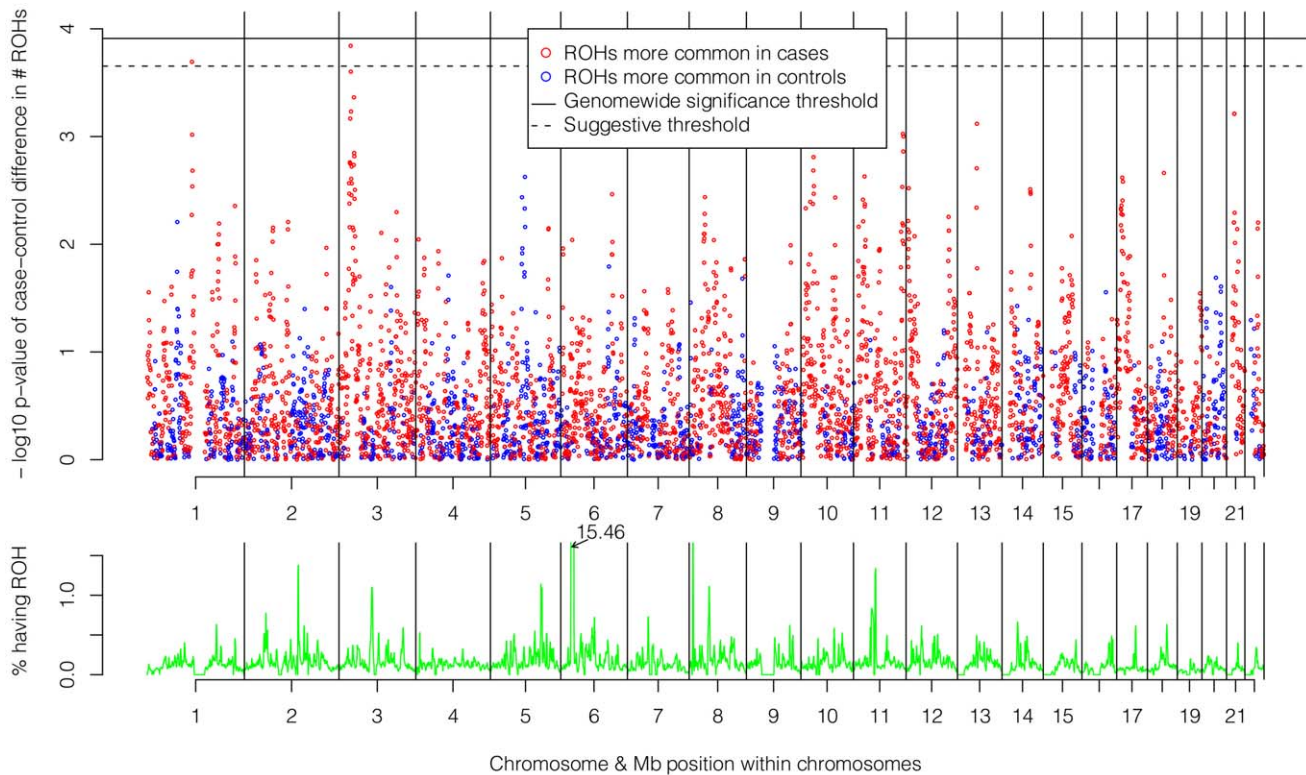
Figure 5. Risk and protective effects of ROHs on schizophrenia risk and frequencies of ROHs across the autosome. Top panel: $-\log_{10}$ p-values for the risk (red) and protective (blue) effects of ROHs on schizophrenia risk at each 500 kb region along the autosome. Bottom panel: frequencies of ROHs across the autosome.
doi:10.1371/journal.pgen.1002656.g005

Second, while our results clearly support the hypothesis that autozygosity is a risk factor for schizophrenia, they are less clear about how confidently we can differentiate the roles of distant versus close inbreeding. On one hand, when enough outliers on *Froh* values are excluded, the case-control difference is no longer significant. On the other hand, there are good statistical reasons to consider the analysis of short versus long ROHs more valid than the analyses that exclude individuals with the highest *Froh* values. Thus, the authors favor the conclusion that both distant and close

inbreeding are risk factors for schizophrenia. A more definitive answer to this question would either require a substantially larger sample size or a sample of similar size to the current one but drawn from a population with greater variation in levels of distant inbreeding.

A final limitation has to do with the correlational nature of these findings. We argue that the *Froh*-schizophrenia association is likely to be causal because the association is consistent with a known genetic mechanism, directional dominance, and because the

Table 2. Genes within the two 500 kb bins that reached "suggestive" significance in the ROH mapping analysis.

Gene	Location	Functional Significance
MRP63P1	1p13.2	mitochondrial ribosomal protein 63 pseudogene
LOC100421116	1p13.2	involved with trafficking protein, a kinesin binding 2 pseudogene
TRIM33	1p13.2	thought to be a transcriptional co-repressor
RPL26P10	1p13.2	ribosomal protein pseudogene
EIF2S2P5	1p13.2	subunit 2 beta pseudogene
LOC643586	1p13.2	pyruvate kinase, muscle pseudogene
BCAS2	1p13.2	believed to play a role in increasing estrogen receptor function
DENND2C	1p13.2	gene of unknown function; highly conserved across mammals
AMPD1	1p13.2	catalyzes deamination of AMP to IMP in skeletal muscle; role in the purine nucleotide cycle
RBMS3	3p24.1	encodes an RNA-binding protein; implicated in smoking cessation, diabetes, and bone mass
RPS12P5	3p24.1	pseudogene of unknown function
LOC100129900	3p24.1	pseudogene of unknown function

doi:10.1371/journal.pgen.1002656.t002

association appears to be as robust for short ROHs as long ROHs. Short ROHs are likely to represent autozygosity caused by distant inbreeding, and therefore seem less likely to differ between parents as a function of their load of schizophrenia risk alleles. Nevertheless, we cannot eliminate the possibility that parents of offspring who have schizophrenia differ in ways that make distant inbreeding more likely, such as an increased propensity to mate with individuals who have culturally, geographically, or ethnically similar backgrounds.

Conclusion

Inbreeding has had a central place in population genetics since its inception, but until recently, the effects of inbreeding could only be investigated from careful analysis of pedigrees and only for close inbreeding. SNP data allows investigation into the effects of potentially very distant inbreeding in non-selected samples, and allows insight into where the signal comes from in the genome. However, unless samples are specifically selected based on inbreeding, very large samples are required to reliably detect effects of autozygosity due to the low variation between individuals in their levels of autozygosity. The present investigation used SNP data from a large sample to conclude that autozygosity is a risk factor for schizophrenia. If the relationship between F_{roh} and schizophrenia is due to directional dominance, such that schizophrenia risk alleles are more recessive than otherwise expected, this suggests that alleles that increase the risk of schizophrenia have been under negative selection ancestrally.

Methods

Psychiatric GWAS consortium data

Full methods are given elsewhere [3]. Briefly, 9,388 schizophrenia cases and 12,456 controls were collected with institutional review board approval from 17 sites in 11 countries (Australia, Bulgaria, Denmark, Germany, Ireland, Netherlands, Norway, Portugal, Sweden, United Kingdom, and Unites States of America; see Table 1). As is typical in the field, individuals with schizophrenia or schizoaffective disorder were included as cases [49] [50]. The quality of phenotypic data was verified by a systematic review of data collection methods to ensure consistency between sites.

Quality control (QC) procedures for raw SNP data

The initial set of samples and SNPs passed common GWAS QC procedures [3]. In particular, we removed a) one individual from any pair of individuals who were related with pi-hat >0.2, b) individuals with non-European ancestry as determined by principal components analysis; c) samples with SNP missingness >0.02; or d) samples with genome-wide heterozygosities >6 standard deviations above the mean. SNPs were excluded if they a) deviated from Hardy-Weinberg equilibrium at $p<1\times10^{-6}$; b) had missingness >0.02; c) showed a minor allele frequency difference to HapMap CEU>0.15; or d) had a missingness difference between cases and controls >0.02. On average the QC processes excluded 15 individuals (0–100) and 38K SNPs (5K–160K) per dataset. The number of SNPs per dataset after QC varied between 250K and 680K (Table 1).

Imputation and QC procedures for imputed SNP data

Six different SNP platforms (Affymetrix 500K, 5.0, and 6.0 chips along with the Illumina 317K, 550K, and 650K chips; Table 1) were used across the 17 datasets. Differences across platforms in SNP densities, frequency distributions, LD patterns, and missingness led to variation in ROH statistics across datasets.

For example, the DK dataset contains 280K SNPs after LD pruning (1 SNP per 11 kb) whereas the UCL datset contains 156K SNPs after LD pruning (1 SNP per 21 kb). ROHs therefore would have to be about twice as long in the UCL dataset to qualify, which induces artifactual noise in ROH statistics due to platform effects. This issue is not circumvented by using an ROH threshold based on length rather than number of SNPs; in this case, half as many homozygous SNPs in a row would be required to call an ROH in the less dense dataset. In both cases, the type-I and type-II error rates of autozygosity detection differ systematically between datasets.

To overcome these issues, we imputed dosages for 1,252,901 autosomal SNPs in each dataset using BEAGLE [38] and HapMap3 as the reference panel [3]. We converted imputation dosages to best-guess (highest posterior probability) SNP calls because ROH detection algorithms require discrete SNP calls. Because typical imputation QC thresholds can lead to a high number of missed ROHs, we used extremely stringent imputation QC thresholds that have been shown to achieve accuracy rates similar to those in genotyped SNPs [39]. In particular, we removed 854,566 imputed SNPs with dosage $r^2<0.90$ in any dataset (the dosage r^2 is equivalent to *MACH's* r^2 measure described in [51]), that had a dosage $r^2<0.98$ or >1.02 in the overall sample, or that had MAF<0.05, leaving 398,325 high-quality imputed SNPs. Because only ~100K SNPs are use to make ROH calls (see below), we could afford to lose a large number of imputed SNPs from QC procedures.

ROHs called from imputed data were less variable across platform and across datasets in terms of basic descriptive statistics, in the effects of potential artifacts (e.g., SNP missingness rates and excess heterozygosity on F_{roh}), and in their associations with schizophrenia. We therefore report results on ROHs called from imputed data. However, results for the ROHs called from raw data were similar, and are shown in Figures S2 and S3.

ROH calling procedures

Of three programs investigated (PLINK, GERMLINE, and BEAGLE), a recent investigation by three of the authors of the current report [40] concluded that PLINK (using the –homozyg commands) optimally detected autozygous stretches and maximized power to detect an effect of autozygosity on a phenotype. In particular, the authors recommended: a) pruning for strong LD (removing any SNPs having a multiple $R^2>0.90$ with all other SNPs in a 50 SNP window), which reduced false autozygosity calls by removing redundant markers in SNP-dense regions and by making SNP coverage more uniform; and b) defining ROHs as being ≥65 consecutive homozygous SNPs with no heterozygote calls allowed [40]. We used these recommendations to detect ROHs in all analyses, although to ensure that we did not miss potential effects of autozygosity, we report on results from the specific ROH threshold (number of homozygous SNPs in a row) that minimized the p-value of the F_{roh}-schizophrenia association (see Figure 3 and Figure S3). This threshold was 65 SNPs-in-a-row (spanning ~2.3 Mb) in the imputed SNP data and 110 SNPs-in-a-row (spanning ~1.7 Mb to ~3.2 Mb depending on the dataset) in the raw data. It should be noted that results were relatively insensitive to the specific threshold chosen (Figure 3 and Figure S3). Finally, to ensure that no ROH crossed a region of low SNP density (e.g., a centromere), we also required that ROHs have a density greater than 1 SNP per 200 kb, and we broke an ROH in two if a gap >500 kb existed between adjacent homozygous SNPs.

ROHs can also be categorized by their frequency (how often a particular haplotype creates ROHs at a given location). We used PLINK's –homozyg-group and –homozyg-match arguments to

understand whether uncommon ROHs or common ROHs were particularly predictive of case-control status, defining ROHs in a given region as "uncommon" when they allelically matched with 16 (the median) or fewer other ROHs in the combined data; all other ROHs were defined as "common."

ROH burden analysis

For each individual, we summed the total length of all their ROHs in the autosome and divided by the total SNP-mappable autosomal distance (2.77×10^9 bases) to derive F_{roh}, the proportion (0 to 1) of the autosome in ROHs. F_{roh} was used as the predictor of case-control status in ROH burden analyses. F_{roh} can be influenced by confounding factors like population stratification (e.g., if background levels of heterozygosity or autozygosity differed by ancestry), low quality DNA leading to incorrect SNP calls, and heterozygosity levels that vary across plates, DNA sources, etc. To control for the effects of stratification, we included the first 20 principal components based on ~30K SNPs genotyped in all datasets. We also controlled for the percentage of missing calls in the raw SNP data and excess heterozygosity as these track the quality of SNP calls [52]. Using simulations, Keller et al. [25] showed that the ability of F_{roh} to accurately estimate autozygosity is negligibly affected by statistically controlling for excess heterozygosity, and therefore doing so should have minimal effect on results when genotyping error rates are low, but may help elucidate effects of ROHs when such errors are present.

We regressed case-control status on F_{roh} separately in each of the 17 datasets using logistic regression, controlling for the potential confounders discussed above. We then employed a mixed linear effects logistic regression model (using the lme4 package in R version 2.11) to estimate the overall effect of F_{roh} across datasets, treating dataset as a random factor. This also controlled for SNP platform because dataset was nested within each platform (controlling for platform was statistically redundant in a model also controlling for dataset).

ROH mapping analysis

To understand whether any genomic area was predictive of case-control status, we divided the autosome into 5,742 segments of length 500 kb each. At each segment, an individual was scored as either having a ROH that partially or completely overlapped the segment or not. We performed 5,742 logistic regressions, regressing case-control status on whether or not individuals had an ROH in each segment, controlling for covariates described above. To derive a genome-wide significance threshold corrected for multiple testing, we permuted case-control status within the 17 datasets and reran the 5,742 logistic regressions, preserving the most significant result of each permutation. We repeated this permutation 1,000 times. The 50th most significant p-value was the genome-wide significance threshold and the 100th most significant p-value was the "suggestive" genome-wide significance threshold.

Supporting Information

Figure S1 Distributions of ROH Lengths (left) and F_{roh} (right) in the total sample, including individuals with $F_{roh} > .0625$. Distributions are based on ROHs from the imputed SNP data.
(TIF)

Figure S2 Estimated changes in odds of schizophrenia for each 1% increase in F_{roh} (odds ratios; asterisks) and their 95% confidence intervals (bars) across the 17 datasets (colored) and for the total sample (black) from the raw SNP data. Boxes are proportional to the square root of sample sizes (also shown at the

bottom). Dataset names are on the x-axis. Although none of the estimated odds ratios are significantly different from one individually, the overall effect (black) is highly significant.
(TIF)

Figure S3 Slope estimates (the change in log odds for a 1% increase in F_{roh}; points) and their 95% confidence intervals (bars) of F_{roh} from raw SNP data predicting schizophrenia for different SNP homozygosity thresholds of calling ROHs. Minimum SNP thresholds for full and reduced models are offset for clarity. All ROH thresholds were significant; the most significant result was for ROHs defined as being 110 or more homozygous SNPs in a row.
(TIF)

Glasgow, UK), Pat Malloy (University of Edinburgh, Edinburgh, UK), Alan W. Maclean (University of Edinburgh, Edinburgh, UK), Andrew McIntosh (University of Edinburgh, Edinburgh, UK). **Molecular Genetics of Schizophrenia (MGS2):** Pablo V. Gejman (NorthShore University HealthSystem, Evanston, USA, and University of Chicago, Chicago, USA), Alan R. Sanders (NorthShore University HealthSystem, Evanston, USA, and University of Chicago, Chicago, USA), Jubao Duan (NorthShore University HealthSystem, Evanston, USA, and University of Chicago, Chicago, USA), Douglas F. Levinson (Stanford University, Stanford, USA), Jianxin Shi (National Cancer Institute, Bethesda, USA), Nancy G. Buccola (Louisiana State University, New Orleans, USA), Bryan J. Mowry (Queensland Brain Institute, Brisbane, Australia), Robert Freedman (University of Colorado Denver, Aurora, USA), Farooq Amin (Emory University, Atlanta, USA, and Atlanta Veterans Affairs Medical Center, Atlanta, USA), Donald W. Black (University of Iowa, Iowa City, USA), Jeremy M. Silverman (Mount Sinai School of Medicine, New York, USA, and Veterans Affairs Medical Center, New York, USA), William F. Byerley (University of California at San Francisco, San Francisco, USA, and Northern California Institute for Research And Education, San Francisco, USA), C. Robert Cloninger (Washington University, St. Louis, USA). **Munich (MUC):** Ina Giegling (Ludwig-Maximilians University, Munich, Germany), Annette M. Hartmann (Ludwig-Maximilians University, Munich, Germany), Heike Konnerth (Ludwig-Maximilians University, Munich, Germany), Marion Friedl (Ludwig-Maximilians University, Munich, Germany), Bettina Konte (Ludwig-Maximilians University, Munich, Germany), Pierandrea Muglia (University of Toronto, Toronto, Canada, and NeuroSearch A/S, Ballerup, Denmark), Dan Rujescu (Ludwig-Maximilians University, Munich, Germany). **Portugal (PORT):** Michele T. Pato (University of Southern California, Los Angeles, USA), Carlos N. Pato (University of Southern California, Los Angeles, USA), Ayman Fanous (Washington VA Medical Center, Washington, DC, USA, Georgetown University School of Medicine, Washington, DC, USA, and Virginia Commonwealth University School of Medicine, Richmond, USA). **SW1 & SW2:** Christina M. Hultman (Karolinska Institutet, Stockholm, Sweden), Paul Lichtenstein (Karolinska Institutet, Stockholm, Sweden), Sarah E. Bergen (Massachusetts General Hospital, Boston, USA), Shaun Purcell (Broad Institute, Cambridge, USA), Edward Scolnick (Broad Institute, Cambridge, USA), Pamela Sklar (Massachusetts General Hospital, Boston, USA, and Mount Sinai School of Medicine, New York, USA), Patrick F. Sullivan (Karolinska Institutet, Stockholm, Sweden, and The University of North Carolina at Chapel Hill, Chapel Hill, USA). **TOP3:** Srdjan Djurovic (University of Oslo, Oslo, Norway, and Oslo University Hospital, Oslo, Norway), Morten Mattingsdal (University of Oslo, Oslo, Norway, and Sorlandet Hospital, Kristiansand, Norway), Ingrid Agartz (University of Oslo, Oslo, Norway, and Diakonhjemmet Hospital, Oslo, Norway), Ingrid Melle (University of Oslo, Oslo, Norway, and Oslo University Hospital, Oslo, Norway), Ole A. Andreassen (University of Oslo, Oslo, Norway, and Oslo University Hospital, Oslo,

Norway). **UCLA:** Roel A. Ophoff (University Medical Center Utrecht, Utrecht, The Netherlands, and University of California at Los Angeles, Los Angeles, USA), Rita M. Cantor (University of California at Los Angeles, Los Angeles, USA), Nelson B. Freimer (University of California at Los Angeles, Los Angeles, USA), René S. Kahn (University Medical Center Utrecht, Utrecht, The Netherlands), Don H. Linszen (University of Amsterdam, Amsterdam, The Netherlands), Jim van Os (Maastricht University Medical Centre, Maastricht, The Netherlands), Durk Wiersma (University of Groningen, Groningen, The Netherlands), Richard Bruggeman (University of Groningen, Groningen, The Netherlands), Wiepke Cahn (University Medical Center Utrecht, Utrecht, The Netherlands), Lieuwe de Haan (Academic Medical Centre University of Amsterdam, Amsterdam, The Netherlands), Lydia Krabbendam (Maastricht University Medical Centre, Maastricht, The Netherlands), Inez Myin-Germeys (Maastricht University Medical Centre, Maastricht, The Netherlands), Eric Strengman (University Medical Center Utrecht, Utrecht, The Netherlands). **London (UCL):** Andrew McQuillin (University College London Medical School, London, UK), Khalid Choudhury (University College London Medical School, London, UK), Susmita Datta (University College London Medical School, London, UK), Jonathan Pimm (University College London Medical School, London, UK), Srinivasa Thirumalai (West Berkshire NHS Trust, Reading, UK), Vinay Puri (University College London Medical School, London, UK), Robert Krasucki (University College LondonMedical School, London, UK), Jacob Lawrence (University College London Medical School, London, UK), Digby Quested (University of Oxford, Oxford, UK), Nicholas Bass (University College London Medical School, London, UK), Hugh Gurling (University College London Medical School, London, UK). **Zucker Hillside:** Anil K. Malhotra (The Zucker Hillside Hospital Division of the North Shore, Glen Oaks, USA, The Feinstein Institute for Medical Research, Manhasset, USA, and Albert Einstein College of Medicine of Yeshiva University, Bronx, New York, USA), Todd Lencz (The Zucker Hillside Hospital Division of the North Shore, Glen Oaks, USA, The Feinstein Institute for Medical Research, Manhasset, USA, and Albert Einstein College of Medicine of Yeshiva University, Bronx, New York, USA).

Author Contributions

Conceived and designed the experiments: MCK MAS SR BMN PVG DPH TL DFL. Analyzed the data: MCK MAS SR SHL. Wrote the paper: MCK PFS DPH DFL BMN. Collected the data: SPGWASC.

References

1. Sullivan PF, Kendler KS, Neale MC (2003) Schizophrenia as a complex trait: Evidence from a meta-analysis of twin studies. Arch Gen Psychiat 60: 1187–1192.
2. Saha S, Chant D, Welham J, McGrath J (2005) A systematic review of the prevalence of schizophrenia. PLoS Med 2: e141. doi:10.1371/journal.pmed.0020141.
3. Psychiatric GWAS Consortium (2011) Genome-wide association study identifies five novel schizophrenia loci. Nat Genin press.
4. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460: 748–752.
5. Lee SH, De Candia T, Ripke S, Yang J, PGC-SCZ, et al. (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 44: 247–250.
6. Charlesworth D, Willis JH (2009) The genetics of inbreeding depression. Nat Rev Genet 10: 783–796.
7. Roff DA (1997) Evolutionary quantitative genetics. New York, NY: Chapman & Hall.
8. DeRose MA, Roff DA (1999) A comparison of inbreeding depression in life-history and morphological traits in animals. Evolution 53: 1288–1292.
9. Shami SA, Qaisar R, Bittles AH (1991) Consanguinity and adult morbidity in Pakistan. Lancet 338: 954.
10. Rudan I, Smolej-Narancic N, Campbell H, Carothers A, Wright A, et al. (2003) Inbreeding and the genetic complexity of human hypertension. Genetics 163: 1011–1021.

11. Rudan I, Skaric-Juric T, Smolej-Narancic N, Janicijevic B, Rudan D, et al. (2004) Inbreeding and susceptibility to osteoporosis in Croatian island isolates. Coll Antropol 28: 585–601.
12. Lebel RR, Gallagher WB (1989) Wisconsin consanguinity studies. II: Familial adenocarcinomatosis. Am J Med Genet 33: 1–6.
13. Afzal M (1988) Consquences of consanguinity on cognitive behavior. Beh Genet 18: 583–594.
14. Morton NE (1979) Effect of inbreeding on IQ and mental retardation. Proc Nat Acad Sci 75: 3906–3908.
15. Abaskuliev AA, Skoblo GV (1975) Inbreeding, endogamy and exogamy among relatives of schizophrenia patients. Genetika 11: 145–148.
16. Bulayev OA, Pavlova TA, Bulayeva KB (2009) The effects of inbreeding on aggregation of complex diseases in genetic isolates. Russ J Genet 45: 961–968.
17. Nimgaonkar VL, Mansour H, Fathi W, Klei L, Wood J, et al. (2010) Consanguinity and increased risk for schizophrenia in Egypt. Schizophrenia Res 120: 108–112.
18. Chaleby K (1987) Cousin Marriages and Schizophrenia in Saudi-Arabia. Brit J Psychiat 150: 547–549.
19. Gindilis VM, Gainullin RG, Shmaonova LM (1989) Genetico-demographic patterns of the prevalence of various forms of endogenous psychoses. Genetika 25: 734–743.
20. Rudan I, Rudan D, Campbell H, Carothers A, Wright A, et al. (2003) Inbreeding and risk of late onset complex disease. J Med Genet 40: 925–932.
21. Ahmed AH (1979) Consanguinity and Schizophrenia in Sudan. Brit J Psychiat 134: 635–636.

22. Chaleby K, Tuma TA (1986) Cousin marriages and schizophrenia in Saudi Arabia. Brit J Psychiat 150: 547–549.

23. Saugstad L, Ødegard Ø (1986) Inbreeding and schizophrenia. Clin Genet 30: 261–275.

24. Bittles AH, Neel JV (1994) The costs of human inbreeding and their implications for variations at the DNA level. Nat Genet 8: 117–121.

25. Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense SNP data. Genetics 189: 237–249.

26. Wright S (1922) Coefficients of inbreeding and relationships. Am Nat 56: 330–339.

27. Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. Mol Ecol 11: 2467–2474.

28. Vine AE, McQuillin A, Bass NJ, Pereira A, Kandaswamy R, et al. (2009) No evidence for excess runs of homozygosity in bipolar disorder. Psychiatr Genet 19: 165–170.

29. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, et al. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci U S A 104: 19942–19947.

30. Ku CS, Naidoo N, Teo SM, Pawitan Y (2011) Regions of homozygosity and their impact on complex diseases and traits. Hum Genet 129: 1–15.

31. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, et al. (2008) Runs of homozygosity in European populations. Am J Hum Genet 83: 359–372.

32. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, et al. (2010) Genomic runs of homozygosity record population history and consanguinity. PLoS ONE 5: e13996. doi:10.1371/journal.pone.0013996.

33. Spain SL, Cazier JB, Houlston R, Carvajal-Carmona L, Tomlinson I (2009) Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. Cancer Res 69: 7422–7429.

34. Enciso-Mora V, Hosking FJ, Houlston RS (2010) Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. Eur J Hum Genet 18: 909–914.

35. Hosking FJ, Papaemmanuil E, Sheridan E, Kinsey SE, Lightfoot T, et al. (2010) Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. Blood 115: 4472–4477.

36. Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, et al. (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenet 10: 183–190.

37. Sullivan PF (2010) The psychiatric GWAS consortium: big science comes to psychiatry. Neuron 68: 182–186.

38. Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84: 210–223.

39. Hao K, Chudin E, McElwee J, Schadt EE (2009) Accuracy of genome-wide imputation of untyped markers and impacts on statistical power for association studies. BMC Genet 10: 27.

40. Howrigan DP, Simonson MA, Keller MC (2011) Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genom 12: 460–475.

41. Fisher RA (1954) A fuller theory of "junctions" in inbreeding. Heredity 8: 187–197.

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

43. Levinson DF, Duan J, Oh S, Wang K, Sanders AR, et al. (2011) Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiat 168: 302–316.

44. Alper CA, Larsen CE, Dubey DP, Awdeh ZL, Fici DA, et al. (2006) The haplotype structure of the human major histocompatibility complex. Hum Immunol 67: 73–84.

45. Traherne JA (2008) Human MHC architecture and evolution: implications for disease association studies. Int J Immunogenet 35: 179–192.

46. Chakraborty R, Chakravarti A (1977) On consanguineous marriages and the genetic load. Hum Genet 36: 47–54.

47. Bittles AH, Black ML (2010) Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci U S A 107 Suppl 1: 1779–1786.

48. Keller MC, Miller GF (2006) Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? Behav Brain Sci 29: 385–452.

49. Kendler KS, McGuire MT, Gruenberg AM, O'Hare A, Spellman M, et al. (1993) The Roscommon family study. Arch Gen Psychiat 50: 527–540.

50. Faraone SV, Blehar M, Pepple J, Moldin SO, Norton J, et al. (1996) Diagnostic accuracy and confusability analyses: an application to the Diagnostic Interview for Genetic Studies. Psychol Med 26: 401–410.

51. Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11: 499–511.

52. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, et al. (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol 34: 591–602.