

How variation in risk allele output and gene interactions shape the genetic architecture of schizophrenia

Merve Kasap

Vanderbilt University Medical Center

Donard S. Dwyer (✉ ddwyer@lsuhsc.edu)

LSU Health Shreveport <https://orcid.org/0000-0002-4201-0958>

Research Article

Keywords: epistasis, gene-gene interactions, risk allele frequency, schizophrenia, syntenic blocks

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1420609/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Schizophrenia is a highly heritable polygenic psychiatric disorder. Characterization of its genetic architecture may lead to a better understanding of the overall burden of risk variants and how they determine susceptibility to disease. A major goal of this project was to develop a modeling approach to compare and quantify the relative effects of single nucleotide polymorphisms (SNPs), copy number variants (CNVs) and other factors. We derived a mathematical model for the various genetic contributions based on the probability of expressing a combination of risk variants at a frequency that matched disease prevalence. The model included estimated risk variant allele outputs (VAOs) adjusted for population allele frequency. We hypothesized that schizophrenia risk genes would be more interactive than random genes and we confirmed this relationship. Gene-gene interactions may cause network ripple effects that spread and amplify small individual effects of risk variants. The modeling revealed that the number of risk alleles required to achieve threshold for susceptibility will be determined by the average functional locus output (FLO) associated with a risk allele, the risk allele frequency (RAF), the number of protective variants present and the extent of gene interactions within and between risk loci. The model can account for the quantitative impact of protective variants as well as CNVs on disease susceptibility. The fact that non-affected individuals must carry a non-trivial burden of risk alleles suggests that genetic susceptibility will inevitably reach threshold for schizophrenia at a recurring frequency in the population.

Introduction

Schizophrenia is a devastating psychiatric illness that afflicts approximately 0.7-1% of the population with no known cure and mainly symptomatic treatment (Ibrahim and Tamminga 2011; Jablensky et al. 1992; Tandon et al. 2008). It shows a high degree of heritability, but the genetic liability is complex with polygenic origins (Gejman et al. 2011; Gottesman and Shields 1967). Both common inherited gene variants of small genetic effect size (Lee et al. 2012; Stefansson et al. 2009; The International Schizophrenia Consortium 2009) and rarer *de novo* mutations with greater effect, such as copy number variants (CNVs) (Kirov et al. 2009; Stefansson et al. 2008; Xu et al. 2008), contribute to causation of this disease. Large genome-wide association studies (GWAS) in schizophrenia have identified over 100 candidate risk loci with several hundred associated genes (Arnedo et al. 2015; Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). The schizophrenia risk-gene candidates are highly enriched for essential and evolutionarily-conserved genes (Hussin et al. 2015; Kasap et al. 2018) and mutation-intolerant genes (Pardiñas et al. 2018). As the result of Hill-Robertson interference and balancing-selection mechanisms, this may lead to the persistence of risk variants in the genome despite liability for disease (Kasap et al. 2018; Pardiñas et al. 2018).

Based on an accumulation of evidence, genetic influences in schizophrenia appear to resemble gene specification of quantitative traits such as height or skin color rather than classical Mendelian patterns of inheritance (Wray et al. 2014). Consequently, the genetic architecture of schizophrenia is multifaceted (Mackay 2001; Wray and Visscher 2010) and its major features such as the number of risk alleles involved and their allele frequencies in the population are still unfolding. Mathematical modeling has

previously been used to characterize various aspects of complex human traits (Holland et al. 2017; Risch 1990; Zeng et al. 2018; Zhang et al. 2018).

There are divergent views on the main features of the architecture of schizophrenia with some groups proposing that a limited number of robust causal variants are adequate to contribute to disease (Mitchell and Porteous 2011), whereas others suggest compound models of strong non-synonymous mutations and rare CNVs acting in concert with common background alleles (Rodriguez-Murillo et al. 2012). More expansive models envision that susceptibility is determined by hundreds or even thousands of risk alleles with small individual effects (Boyle et al. 2017; The International Schizophrenia Consortium 2009; Wray and Visscher 2010). A major goal of this research project was to develop a modeling approach for comparing and quantifying the relative magnitude of effect of single nucleotide polymorphisms (SNPs), CNVs and other mutations.

The role of gene-gene interactions or epistasis in the genetic liability for psychiatric disorders is likewise controversial. Although epistasis is acknowledged as a theoretical contributor, the magnitude of its effects is viewed as ranging from negligible (Crow 2010; Hill et al. 2008) to highly significant (Carlborg and Haley 2004; Cheverud and Routman 1995; Jones et al. 2014; Phillips 2008). Gene interactions may help to explain the so-called missing heritability in psychiatric disorders (Woo et al. 2017) and genetic connectivity is extensive among the risk genes for bipolar disorder (Franklin and Dwyer 2020) and major depressive disorder (Sall et al. 2021). Davierwala et al. (2005) reported that essential genes are much more interactive than non-essential genes and schizophrenia risk genes are highly enriched for ones that are essential for life and broadly conserved during evolution (Kasap et al. 2018). Consequently, we hypothesized that risk genes for schizophrenia are similarly more interactive than random genes, which may amplify the effect sizes of interacting variants (Cheah et al. 2016). If so, the findings would highlight the importance of being able to account for gene-gene interactions in models of genetic liability for heritable psychiatric disorders.

Some of the genetic variants associated with schizophrenia may actually be protective rather than detrimental (Nishino et al. 2018), which complicates assessment of the total risk burden in an individual. Therefore, we explored the possibility that protective variants as well as risk variants (under negative selection) could be identified by analyzing allele frequencies and odds ratio (OR) scores in relation to case versus control status. Moreover, the studies reported here sought to answer two pertinent questions that turn out to be related to one another: how many risk genes (including both risk and protective) contribute to threshold liability for schizophrenia and what role do gene-gene interactions play in overall risk determination? As described here, we developed a novel mathematical model to address the first question and then explored the importance of gene interactions in shaping the risk for schizophrenia.

Methods

Theoretical Background and Equation for Risk-Gene Quantification

According to polygenic threshold models (Wray and Visscher 2010), multiple hits at different risk genes are required to achieve threshold risk for schizophrenia. Certain combinations of risk genes will significantly elevate liability for disease.

To estimate the number of genes required to reach threshold, we developed key equations as described in the Supplemental Methods section. Eq. (1) describes the functional output from a risk-associated locus:

$$(1) \text{ FLO} = (\text{RAF}(x) + (1 - \text{RAF})) ,$$

where x is the variant allele output (VAO or altered activity level of the gene affected by a polymorphism) and RAF is the risk allele frequency in the population. This term can be considered the average functional locus output (FLO) from two (or more) alleles at a genetic locus when one is affected by variation. It is important to add that a risk gene may harbor more than a single variant; several variants (e.g., SNPs) may collectively affect the expression/function of the gene. The FLO is intended to reflect the overall contributions of relevant variants of a particular risk gene. Thus, the total number of risk variants for a disorder that are present in the genome is different from, and greater than, the number of risk genes required to reach threshold for the disorder, which is the subject of this investigation.

Based on the probability of obtaining a series of variants with different VAOs and RAFs over multiple potential risk loci, we devised Eq. (2):

$$(2) P_t = [\text{RAF1}(x_1) + (1 - \text{RAF1})]^{y_1} \times [\text{RAF2}(x_2) + (1 - \text{RAF2})]^{y_2} \times \dots n$$

where P_t is the probability of threshold risk-gene combination in the population, x and RAF correspond to above and y is the number of different risk genes associated with particular x and RAF values. We can substitute estimates of x and RAF and solve for y to calculate the number of risk genes needed to reach threshold for schizophrenia. We ran the model calculations until P_t converged on 0.01, the prevalence of schizophrenia in the population.

Risk allele frequency (RAF) of disease-associated variants

The PGC risk variant analysis included reference alleles that were both minor and major in terms of population expression; therefore, we have referred to their frequency in the genome as RAF to reflect this. We characterized individual RAFs of the risk variants from the PGC dataset to derive representative estimates for plugging into Eq. (2). To obtain OR scores from this dataset having the same directional relationship (> 1), frequencies below 1 were expressed as $1/\text{OR}$. From the original list of 128 variants, we excluded 14 indels from our analysis. For comparison, we generated a random list of genes with Molbiotools' Random Gene Set Generator

(<http://www.molbiotools.com/randomsequencegenerator.html>). We then randomly selected SNP variants from these genes to compile a similar-sized list of reference variants. The population allele frequencies of the randomly chosen reference variants were recorded.

In addition, we characterized the allele frequencies of a subset of risk variants that are associated with syntenic blocks of functionally-related genes as described previously (Kasap et al. 2018). The syntenic blocks examined here corresponded to the final ranking of variants assigned in the PGC (2014) study: 9, 22, 31, 38, 40, 47, 59, 84, and 85. We determined the RAFs and OR scores for this subset of alleles and compared these values with those from the complete list of risk variants.

Gene Interaction Analysis

Gene interactions are defined here as pairs of genes that when co-retained in cells affect cell survival as determined by Lin et al. (2010) with radiation hybrids. These genes comprise the Genetic Interactions data used by GeneMANIA (Zuberi et al. 2013).

To compare genetic interactions in the entire PGC gene dataset with interactions between genes randomly selected from the human genome, we used GeneMANIA and selected “Genetic interactions” with “Max resultant genes” (other genes) set to 0. For comparison, we generated 4 random gene lists (as above) of roughly the same size (333–346 genes) as the final PGC dataset (338) and analyzed these under the same conditions. We also compared a rheumatoid arthritis (RA) dataset of 326 genes derived by Okada et al. (2014).

We quantified genetic interactions as described previously (Franklin and Dwyer 2020). We then used the number of links per gene to compare connections involving the set of PGC risk genes with those of the randomly selected gene lists.

Statistical Methods

We compared the RAFs of the PGC SNP dataset to a set of randomly generated SNPs using a standard t-test.

To determine statistical significance for comparisons between the number of genetic interactions among genes in the different sets, we first calculated the mean and standard deviation from data derived from 4 random lists of genes expressed as the number of links per gene obtained from GeneMANIA. Based on the standard deviations from the random gene lists, we established confidence intervals (CI) for significant differences at the $p < 0.01$ or $p < 0.05$ levels. We then determined whether data from the PGC dataset surpassed the CI calculated from the four random gene sets.

Results

Distribution of RAFs

We sought to determine if there was evidence of negative and positive selection on the PGC risk variants and to derive an average allele frequency for entering into Eq. (2). Therefore, we characterized the allele frequencies of risk variants associated with schizophrenia and obtained profiles of cases vs. controls (Fig. 1) to obtain estimates of the average RAF. We compared RAFs from the PGC dataset to the

population allele frequencies of 114 randomly-generated SNPs and discovered no difference between the allele frequencies for these two groups of SNPs – both averaged 0.44. Thus, schizophrenia risk alleles showed the same overall frequency profile as a random set of alleles, which is similar to the data of Ohi et al. (2017) on an overlapping gene set (average RAF 0.46). This argues against strong selection of the risk variants, consistent with the findings of Liu et al. (2019) and Yao et al. (2020), and against these common variants belonging to a subset with unique characteristics.

Closer examination of the highest and lowest quartiles revealed additional interesting relationships. Regression analysis showed a stronger positive correlation between RAF and OR score for the highest quartile of variants occurring more often in Controls than Cases (Fig. 2), which suggests that they might represent protective variants. The opposite relationship was observed for the lowest quartile, perhaps indicating that adverse risk variants are under stronger negative selection. Positive and negative selection have both been observed previously for variants associated with susceptibility to schizophrenia. Overall, the data support the findings of Nishino et al. (2018) that protective variants co-occur in the genomes of individuals with schizophrenia and should be accounted for when estimating the burden of risk due to genetic factors.

In a previous study, we identified syntenic blocks of risk-gene candidates in schizophrenia that were collectively geared toward common purposes (Kasap et al. 2018). When we characterized the RAFs and OR scores for these 9 blocks, we found that the SNP variants did not differ significantly from the total SNP set in terms of RAF (0.47 ± 0.16 vs. 0.44 ± 0.26) or OR score (1.075 ± 0.009 vs. 1.086 ± 0.036) and they fell into the middle two quartiles of the overall data (Fig. 1A). This distribution suggests that the alleles associated with syntenic blocks of risk genes are not under strong selection. The lack of obvious selection on these functional blocks of genes suggests that they may reside in genomic regions such as recombination coldspots or regions that experience background selection with reduced mutability in order to preserve the local adaptive arrangement.

Risk Gene Quantification

To assess the underlying risk-gene architecture, we used Eq. (2). If the risk variants decrease output from the affected alleles to 80% of normal, P_t will reach the 1% criterion at 50 genes with an average RAF of 0.44 (Fig. 3A). By contrast, if the VAO is 95% of the non-risk allele and the average frequency of expression is low (RAF = 0.1), then it would require 900 risk alleles to reach the 0.01 threshold (Fig. 3B). Thus, the number of risk alleles needed to achieve threshold for disease decreases as the frequency of the risk allele in the population increases and in relation to the relative loss of functional output from the affected locus. This relationship is similar to the model calculations of Wray and Visscher (2010) and demonstrates the utility of simple models as starting points for analyzing the genetic architecture of schizophrenia.

Inclusion of Protective Variants

The work of Nishino et al. (2018) and logical considerations converge on the conclusion that individuals with schizophrenia will harbor some protective variants in their genome. In addition, Hess et al. (2021) recently identified resilience variants that modify the risk for schizophrenia. In comparison to protective variants, which are typically the alternate allele at a risk locus, resilience alleles moderate the adverse effects of risk variants. To account for the moderating effects of both types of protective alleles, we re-examined the model with inclusion of values for $x > 1$ to reflect the protective variants in Eq. (2). For simplicity, we chose a value for the FLO of 1.015 for the simulations because it is equal in magnitude but opposite in effect of the FLO used for adverse risk variants (0.985). Protective variants will actually have an array of effect sizes similar to the risk variants. Figure 3C reveals that the number of risk variants required to reach threshold increases as a function of the number of protective variants included in these simulations. Moreover, this modification of Eq. (2) allows us to quantitatively model the effects of protective variants on overall disease risk.

Genetic Interactions among Risk Variants

Based on previous investigation of risk genes for bipolar disorder (Franklin and Dwyer 2020), we hypothesized that the PGC risk genes will also show greater interaction with each other than would be observed in randomly-selected sets of genes. Gene interactions can potentially amplify the effects of risk variants and we refer to this phenomenon as network ripple effects.

The results of GeneMANIA analysis are depicted in Fig. 4 and Supplemental Figs. S1 & S2. For the PGC risk-gene set, we detected a significant increase in the number of interactions per gene among members of this list compared to the random genes (Fig. 4). Gene interactions provide information about the network connectivity of risk genes and appear to be correlated with the degree of evolutionary conservation of the genes. The latter observation may stem from more extensive integration of older genes into functional networks. Therefore, gene interactions may be an important source of added liability in schizophrenia and may explain some of the missing heritability (Woo et al. 2017).

Effect of Gene Interactions and CNVs on Models of Genetic Risk

The extensive interactions among schizophrenia risk genes forces us to consider such interactions when estimating the number of risk alleles that predispose to disease. At a single SNP locus, multiple nearby risk-gene candidates may be present, which functionally interact either locally or with genes on different chromosomes. If the average VAO due to genetic variation is 0.965, it will result in a FLO of 0.985 when adjusted for average RAF (0.44). If this is a single independent site, then 0.985 will go into Eq. (2). Likewise, the value of 0.985 would represent the contributions of a second independent locus with similar output. However, when two risk genes interact it is essentially a non-random assorting of that gene combination. To mathematically represent these non-random or non-independent combinations we chose to assign a value of 0.985^2 (0.97) instead of 0.985 to each interacting locus. If a risk gene interacts with

more than one additional risk gene, then for calculation purposes the interacting gene FLO would be 0.985^{2+n} , where n in the exponent equals the number of additional interacting risk genes.

To model gene interactions, we used 30% as the fraction of interacting risk genes, which is close to the number derived from our previous work (Kasap et al. 2018). As can be seen in Fig. 5A, ~ 300 risk alleles are required to achieve the 0.01 threshold in the case of no genetic interactions, whereas this number falls as interactions increase. Providing this is a useful way to quantitate gene interactions, this means that gene-gene interactions enhance the liability of small effect sizes of individual risk variants.

CNVs represent an additional source of compound genetic interactions because they can potentially affect the expression of functional blocks of genes. Therefore, we sought to model estimates of the number of risk genes needed to achieve threshold risk for schizophrenia in the absence or presence of CNV FLOs (0.5) in Eq. (2). A single CNV replaces about 15% of the total risk genes needed to achieve threshold, whereas two CNVs reduce that number by 30% (Fig. 5B). Therefore, CNVs make significant contributions to disease liability; however, their effects must occur on a substantial background of risk variation. For example, in Fig. 5B a single CNV would replace the contributions of approximately 45 common risk alleles, but not obviate the need for many additional ones (250–260 in this case).

Taken together, we speculate that the total number of genes (risk + protective) that will achieve threshold for schizophrenia is in the range of 2500–2600. This estimate is based on solving Eq. (2) with the following parameters: an average VAO of 0.995 (a 0.5% reduction in gene expression/output due to the polymorphism), an average RAF of 0.44, which results in a FLO of 0.9978, 30% of the risk genes interact with one additional partner and 25% of the variation affects protective genes (FLO = 1.0022. This 'best guess' estimate of ~ 2500 risk **genes** is closer to the assessment of 8300 causal **variants** (note: each gene may have multiple variants associated with it) by Ripke et al. (2013) than to the 40,000 causal variants suggested by Nishino et al. (2018). By contrast, Zhang et al. (2018) suggested more than 10,000 susceptibility SNPs may be involved in schizophrenia based on the same gene dataset, whereas Frei et al. (2019) estimated 8500 variants that explained 90% SNP-heritability, which agrees with the assessment of Ripke et al. (2013) and our calculations. Holland et al. (2020) used mathematical modeling to arrive at 31,000 causal SNPs. Some of the tendency to overestimate the number of contributing risk genes may be due to differences in what is considered a risk variant. Moreover, the failure to incorporate gene-gene interactions in other estimates ignores the possibility that genetic interactions may amplify the effects of individual variants and decrease the genetic burden needed to reach threshold for the disorder.

Discussion

The major findings of this study include: 1) a simple equation relating altered genetic output from variant alleles and RAFs to the total number of risk genes required to reach threshold for schizophrenia, 2) support for the occurrence of protective variants undergoing positive selection and a means to deal with them computationally and 3) the discovery of extensive gene-gene interactions among schizophrenia risk genes and a strategy for including them in genetic liability calculations. Finally, we provide a quantitative

basis for harmonizing views about the relative contributions of SNPs and CNVs to the genetic architecture.

Wray and Visscher (2010) made impressive headway in characterizing the genetic architecture of schizophrenia and our analysis generally complements theirs. Specifically, we have developed a straightforward mathematical model for estimating risk gene burden with explicit terms for RAF, the effect size due to risk-gene variation (VAO), gene-gene interactions and inclusion of protective variants. The model provides a quick snapshot of the different factors that influence the genetic liability for schizophrenia. Parameters can be varied to take into account as much complexity as desired.

Analysis of the PGC data revealed that there was an equal number of schizophrenia risk variants with increased vs. decreased occurrence (negative vs. protective effects?) in the case population. Previous studies have revealed evidence for both positive and negative selection of variants that show differences in frequency comparing control subjects with those affected with schizophrenia (Liu et al. 2019; Polimanti and Gelernter 2017; Pardiñas et al. 2018; Yao et al. 2020). For an optimized phenotype, most mutations are expected to produce a negative effect because there is a statistically greater chance of an adverse outcome the closer the phenotype is to optimum. Since this trend was not observed in the PGC dataset, it suggests that most of the traits underlying schizophrenia are not optimized at this point or perhaps are not optimizable due to pleiotropy.

The work of Nishino et al. (2018) and Hess et al. (2021) suggest that many protective variants will be expressed in the genomes of the case population. These protective (resilience) variants counterbalance the adverse effects of risk variants (Hess et al. 2021). Therefore, the balance between adverse and protective alleles must tilt substantially toward the former to produce symptoms of schizophrenia. Taken together with the observation that risk variants for schizophrenia are enriched in essential genes (Kasap et al. 2018), these findings suggest that negative selection of adverse risk variants is minimal, perhaps because the associated genes have been optimized for critical pleiotropic purposes. Furthermore, SNPs associated with schizophrenia risk may affect the expression of multiple genes, sometimes with opposing functions (Peng et al. 2021). Therefore, the net result on relevant behavior may reflect the sum of effects on many genes (Peng et al. 2021), which can be accounted for with our concept of functional locus output (FLO).

A special subset of variants associated with syntenic blocks of genes exhibited intermediate RAFs and average OR scores under little apparent selective pressure. We previously speculated that the process of gene amalgamation to form these syntenic blocks may have been accompanied by creation of recombination coldspots, which would impede selection but also preserve weak risk variants in the DNA blocks (Kasap et al., 2018). The data presented here support this earlier suggestion.

Consistent with studies of gene interactions in bipolar disorder and depression (Franklin and Dwyer 2020; Sall et al. 2021), candidate risk genes for schizophrenia showed greater network interaction than random gene sets of the same size. These gene-gene interactions may cause network ripple effects among the connected genes that amplify the impact of small individual VAOs. Others have observed a similar

phenomenon among risk genes for schizophrenia (Cheah et al. 2016; Su et al. 2017). Furthermore, this notion is similar to that of Greenspan (2001) who described the functional connectivity of gene networks in terms of flexibility and pleiotropy. Despite some overlap with the omnigenic model of Boyle et al. (2017), there are also important distinctions. At some level, all genes are interconnected based on how the genome evolved (Dwyer, 2020). This common origin can obscure real differences between genes and disorders. For example, the risk genes for bipolar disorder and depression are much more interconnected (2-3-fold higher number of gene interactions) than the schizophrenia risk genes. If the various risk genes participated in omnigenic interactions, we would not expect to observe these large differences, especially in psychiatric disorders with overlapping symptoms and involving similar cell types.

Our studies focused on gene-gene interaction networks, whereas others have explored pathway networks to gain insights into schizophrenia (Willsey et al. 2018). The networks have been derived from gene function analysis, protein-protein interactions, co-expression data and other sources (Gilman et al. 2012; Schwarz et al. 2016; Walker et al. 2019; Willsey et al. 2018). These studies have been very informative about possible mechanisms contributing to pathogenesis; however, they do not directly address the genetic architecture. Because the pathways, proteins and genes involved in schizophrenia are organized into tangible networks, genetic risk calculations need to reflect this inherent connectivity among networked genes. Here, we have attempted to quantitatively assess the impact of gene interactions on overall risk burden.

Multiple factors contribute to the total risk burden for schizophrenia: risk variants (SNPs, CNVs), protective variants, gene interactions, allele frequency and the overall effect size of the genetic mutation. We have represented these various factors in a simple equation that can reconcile relative contributions from common non-coding SNPs and CNVs. CNVs, null mutants and major functional mutations (loss or gain of function) will produce significant liability for schizophrenia; however, their effects manifest on a background of significant risk alleles (Bassett et al. 2017; Bergen et al. 2019; Tansey et al. 2016). Using estimated parameters, we calculated that a deletion CNV is roughly equivalent to around 15% of the total single-nucleotide variants required to reach threshold for disease. By itself, a single CNV in a risk gene would not be sufficient to cause schizophrenia without the contributions of many additional background risk variants.

Interpretation of the genetic architecture in schizophrenia must be considered cautiously due to various inherent limitations. GWAS and CNV studies have likely identified instances of false positive candidate genes, whereas some actual risk genes may have been missed. Rare MAFs have thus far been largely neglected; however, these rare alleles are likely to have large effects based on previous work (Bergen et al. 2019; Suárez-Rama et al. 2015; The International Schizophrenia Consortium 2008). Environmental factors and epigenetic changes will also complicate interpretation of genetic influences on disease susceptibility. We have limited our simulations to average values for parameters such as RAFs and VAO, so the real situation is much more complicated; however, the equation can manage greater complexity than presented here. Nevertheless, overall trends such as a requirement for a substantial number of risk variants to reach threshold and the potential significance of genetic interactions are based on solid

observations and logic. Finally, there may be alternative ways to handle gene-gene interactions mathematically; however, this work provides a useful conceptual framework to the problem.

According to the model, non-affected individuals must harbor a non-trivial complement of risk alleles that experience little selection. Furthermore, threshold combinations of risk alleles will be inherited at a set frequency in the population – a phenomenon previously described as inevitable bad luck (Kasap et al. 2018). Therefore, schizophrenia differs significantly from pathological genetic conditions such as inherited metabolic disorders or rare Mendelian diseases. Instead, schizophrenia risk variants via their different VAOs and RAFs determine whether certain quantitative traits fall in the normal range. In our distant ancestors, concurrent expression of various suboptimum traits may have carried little penalty for the individual. Without the complexity, artificiality and stress of modern society, someone showing a collection of traits that would be diagnosed today as schizophrenia may still have been largely functional when living a simpler existence in nature. This emerging view of schizophrenia has important implications for diagnosis and treatment.

Declarations

Funding

This work was supported by internal funding.

Author Statement

M. K. and D.D. both performed gene interaction analyses and jointly wrote the manuscript. D. D. derived the equation for risk gene analysis.

Declaration of Competing Interest

The authors have no competing or conflicts of interest to report.

Acknowledgments

The authors greatly appreciate ongoing support from the Department of Psychiatry and Behavioral Medicine at LSU Health Shreveport.

References

1. Arnedo J, Svrakic DM, Del Val C, Romero-Zaliz R, Hernandez-Cuervo H, Molecular Genetics of Schizophrenia Consortium, Fanous AH, Pato MT, Pato CN, de Erausquin GA, Cloninger CR, Zwir I (2015) Uncovering the hidden risk architecture of the schizophrenias: confirmation in three independent genome-wide association studies. *Am J Psychiatry* 172:139–153

2. Bassett AS, Lowther C, Merico D, Costain G, Chow EWC, van Amelsvoort T, McDonald-McGinn D, Gur RE, Swillen A, Van den Bree M, Murphy K, Gothelf D et al (2017) Rare genome-wide copy number variation and expression of schizophrenia in 22q11.2 deletion syndrome. *Am J Psychiatry* 174:1054–1063
3. Bergen SE, Pioner A, Howrigan D, CNV Analysis Group and the Schizophrenia Working Group of the Psychiatric Genomics Consortium, O'Donovan MC, Smoller JW, Sullivan PF, Sebat J, Neale B, Kendler KS (2019) Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. *Am J Psychiatry* 176:29–35
4. Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169:1177–1186
5. Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* 5:618–624
6. Cheah SY, Lurie JK, Lawford BR, Young RM, Morris CP, Voisey J (2016) Interaction of multiple gene variants and their effects on schizophrenia phenotypes. *Compr Psychiatry* 71:63–70
7. Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–1461
8. Crow JF (2010) On epistasis: why it is unimportant in polygenic directional selection. *Philos Trans R Soc London B Biol Sci* 365:1241–1244
9. Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, Mnaimneh S, Ding H, Zhu H, Chen Y, Cheng X, Brown GW et al (2005) The synthetic genetic interaction spectrum of essential genes. *Nat Genet* 37:1147–1152
10. Dwyer DS (2020) Genomic chaos begets psychiatric disorder. *Complex Psychiatry* 6:20–29
11. Franklin C, Dwyer DS (2021) Candidate risk genes for bipolar disorder are highly conserved during evolution and highly interconnected. *Bipolar Disord* 23:400–408
12. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, O'Connell KS, Wang Y, Djurovic S, Thompson WK, Andreassen OA, Dale AM (2019) Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat Commun* 10:2417
13. Gejman PV, Sanders AR, Kendler KS (2011) Genetics of schizophrenia: new findings and challenges. *Annu Rev Genomics Hum Genet* 12:121–144
14. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M, Vitkup D (2012) Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci* 15:1723–1728
15. Gottesman IJ, Shields A (1967) A polygenic theory of schizophrenia. *Proc Natl Acad Sci USA* 58:199–205
16. Greenspan RJ (2001) The flexible genome. *Nat Rev Genet* 2:383–387
17. Hess JL, Tylee DS, Mattheisen M, Schizophrenia Working Group of the Psychiatric Genomics Consortium; Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Borglum

- AD, Als TD, Grove J, Werge T, Mortensen PB, Mors O, Nordentoft M et al (2010) A polygenic resilience score moderates the genetic risk for schizophrenia. *Mol Psychiatry* 26:800–815
18. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4:e1000008
 19. Holland D, Frei O, Desikan R, Fan CC, Shadrin AA, Smeland OB, Sundar VS, Thompson P, Andreassen OA, Dale AM (2020) Beyond SNP heritability: polygenicity and discoverability estimated for multiple phenotypes with a univariate Gaussian mixture model. *PLoS Genet* 16:e1008612
 20. Hussin JG, Hodgkinson A, Idaghdour Y, Grenier JC, Goulet JP, Gbeha E, Hip-Ki E, Awadalla P (2015) Recombination affects accumulation of damaging and disease-associated mutations in human populations. *Nat Genet* 47:400–404
 21. Ibrahim HM, Tamminga CA (2011) Schizophrenia: treatment targets beyond monoamine systems. *Annu Rev Pharmacol Toxicol* 51:189–209
 22. Jablensky A, Sartorius N, Emberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A (1992) Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychol Med Monogr Suppl* 20:1–97
 23. Jones AG, Bürger R, Arnold SJ (2013) Epistasis and natural selection shape the mutational architecture of complex traits. *Nat Commun* 5:3709
 24. Kang G, Yue W, Zhang J, Huebner M, Zhang H, Ruan Y, Lu T, Ling Y, Zuo Y, Zhang D (2008) Two-stage designs to identify the effects of SNP combinations on complex diseases. *J Hum Genet* 53:739–746
 25. Kasap M, Rajani V, Rajani J, Dwyer DS (2018) Surprising conservation of schizophrenia risk genes in lower organisms reflects their essential function and the evolution of genetic liability. *Schizophr Res* 202:120–128
 26. Kirov G, Grozeva D, Norton N, Ivanov D, Mantripragada KK, Holmans P, International Schizophrenia Consortium, Wellcome Trust Case Control Consortium, Craddock N, Owen MJ, O'Donovan MC (2009) Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum Mol Genet* 18:1497–1503
 27. Lee SH, DeCandia TR, Ripke S, Yang J (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 44:247–250
 28. Lin A, Wang RT, Ahn S, Park CC, Smith DJ (2010) A genome-wide map of human genetic interactions inferred from radiation hybrid genotypes. *Genome Res* 20:1122–1132
 29. Liu C, Everall I, Pantelis C, Bousman C (2019) Interrogating the evolutionary paradox of schizophrenia: a novel framework and evidence supporting recent negative selection of schizophrenia risk alleles. *Front Genet* 10:389
 30. Mackay TFC (2001) The genetic architecture of quantitative traits. *Annu Rev Genet* 35:303–339
 31. Mitchell KJ, Porteous DJ (2011) Rethinking the genetic architecture of schizophrenia. *Psychol Med* 41:19–32

32. Nishino J, Kochi Y, Shigemizu D, Kato M, Ikari K, Ochi H, Noma H, Matsui K, Morizono T, Boroevich KA, Tsunoda T, Matsui S (2018) Empirical Bayes estimation of semi-parametric hierarchical mixture models of unbiased characterization of polygenic disease architectures. *Front Genet* 9:115
33. Ohi K, Shimada T, Yasuyama T, Uehara T, Kawasaki Y (2017) Variability of 128 schizophrenia-associated gene variants across distinct ethnic populations. *Transl Psychiatry* 7:e988
34. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S, Graham RR, Manoharan A et al (2014) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506:376–381
35. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, Han J, Hubbard L et al (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 50:381–389
36. Peng X, Bader JS, Avramopoulos D (2021) Schizophrenia risk alleles often affect the expression of many genes and each gene may have a different effect on the risk: a mediation analysis. *Am J Med Genet* 186B:251–258
37. Phillips PC (2008) Epistasis – the essential role of gene interactions in the structural evolution of genetic systems. *Nat Rev Genet* 9:855–867
38. Polimanti R, Gelernter J (2017) Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. *PLoS Genet* 13:e1006618
39. Ripke S, O’Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH et al (2013) Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 45:1150–1158
40. Risch N (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet* 46:222–228
41. Rodriguez-Murillo L, Gogos JA, Karayiorgou M (2012) The genetic architecture of schizophrenia: new mutations and emerging paradigms. *Annu Rev Med* 63:63–80
42. Sall S, Thompson W, Santos A, Dwyer DS (2021) Analysis of major depression risk genes reveals evolutionary conservation, shared phenotypes and extensive genetic interactions. *Front Psychiatry* 12:698029
43. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511:421–427
44. Schwarz E, Izmailov R, Lio P, Meyer-Lindenberg A (2016) Protein interaction networks link schizophrenia risk loci to synaptic function. *Schizophr Bull* 42:1334–1342
45. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D, Werge T, Pietiläinen OP, Mors O, Mortensen PB, Sigurdsson E, Gustafsson O et al (2009) Common variants conferring risk of schizophrenia. *Nature* 460:744–747
46. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, Fossdal R, Sigurdsson E, Sigmundsson T, Buizer-Voskamp JE, Hansen T, Jakobsen KD et al (2008) Large recurrent

- microdeletions associated with schizophrenia. *Nature* 455:232–236
47. Su Y, Ding W, Xing M (2017) The interaction of TXNIP and AFq1 genes increases the susceptibility of schizophrenia. *Mol Neurobiol* 54:4806–4812
 48. Suárez-Rama JJ, Arrojo M, Sobrino B, Amigo J, Brenila J, Agra S, Paz E, Brión M, Carracedo A, Páramo M, Costas J (2015) Resequencing and association analysis of coding regions at twenty candidate genes suggest a role for rare risk variation at *AKAP9* and protective variation at *NRXN1* in schizophrenia susceptibility. *J Psychiatr Res* 2015 66–67:38–44
 49. Tandon R, Keshavan MS, Nasrallah HA (2008) Schizophrenia, “just the facts: what we know in 2008. 2. Epidemiology and etiology. *Schizophr Res* 102:1–18
 50. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL, MacCarroll SA, Holmans P, Kirov G, Walters J, Owen MJ, O’Donovan MC (2016) Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatry* 21:1085–1089
 51. The International Schizophrenia Consortium (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–240
 52. The International Schizophrenia Consortium (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748–752
 53. Walker RL, Ramaswami G, Hartl C, Mancuso N, de la Torre-Ubieta L, Pasaniuc B, Stein JL, Geschwind DH (2019) Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell* 179:750–771
 54. Willsey AJ, Morris MT, Wang S, Willsey HR, Sun N, Teerikorpi N, Baum TB, Cagney G, Bender KJ, Desai TA, Srivastava D, Davis GW et al (2018) The Psychiatric Cell Map Initiative: a convergent systems biological approach to illuminating key molecular pathways in neuropsychiatric disorders. *Cell* 174:505–520
 55. Woo HJ, Yu C, Kumar K, Reifman J (2017) Large-scale interaction effects reveal missing heritability in schizophrenia, bipolar disorder and posttraumatic stress disorder. *Transl Psychiatry* 7:e1089
 56. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorpe (2014) Research Review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry* 55:1068–1087
 57. Wray NR, Visscher PM (2010) Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr Bull* 36:14–23
 58. Xu B, Roos JL, Levy S et al (2008) Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885
 59. Yao Y, Yang J, Xie Y, Liao H, Yang B, Xu Q, Rao S (2020) No evidence for widespread positive selection signatures in common risk alleles associated with schizophrenia. *Schizophr Bull* 46:603–611
 60. Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, Yap CX, Xue A, Sidorenko J, McRae AF, Powell JE, Montgomery GW et al (2018) Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet* 50:746–753

61. Zhang Y, Qi G, Park JH, Chatterjee N (2018) Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat Genet* 50:1318–1326
62. Zuberi K, Franz M, Rodriguez H, Montojo J, Lopes CT, Bader GD, Morris Q (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res* 41:W115–W122

Figures

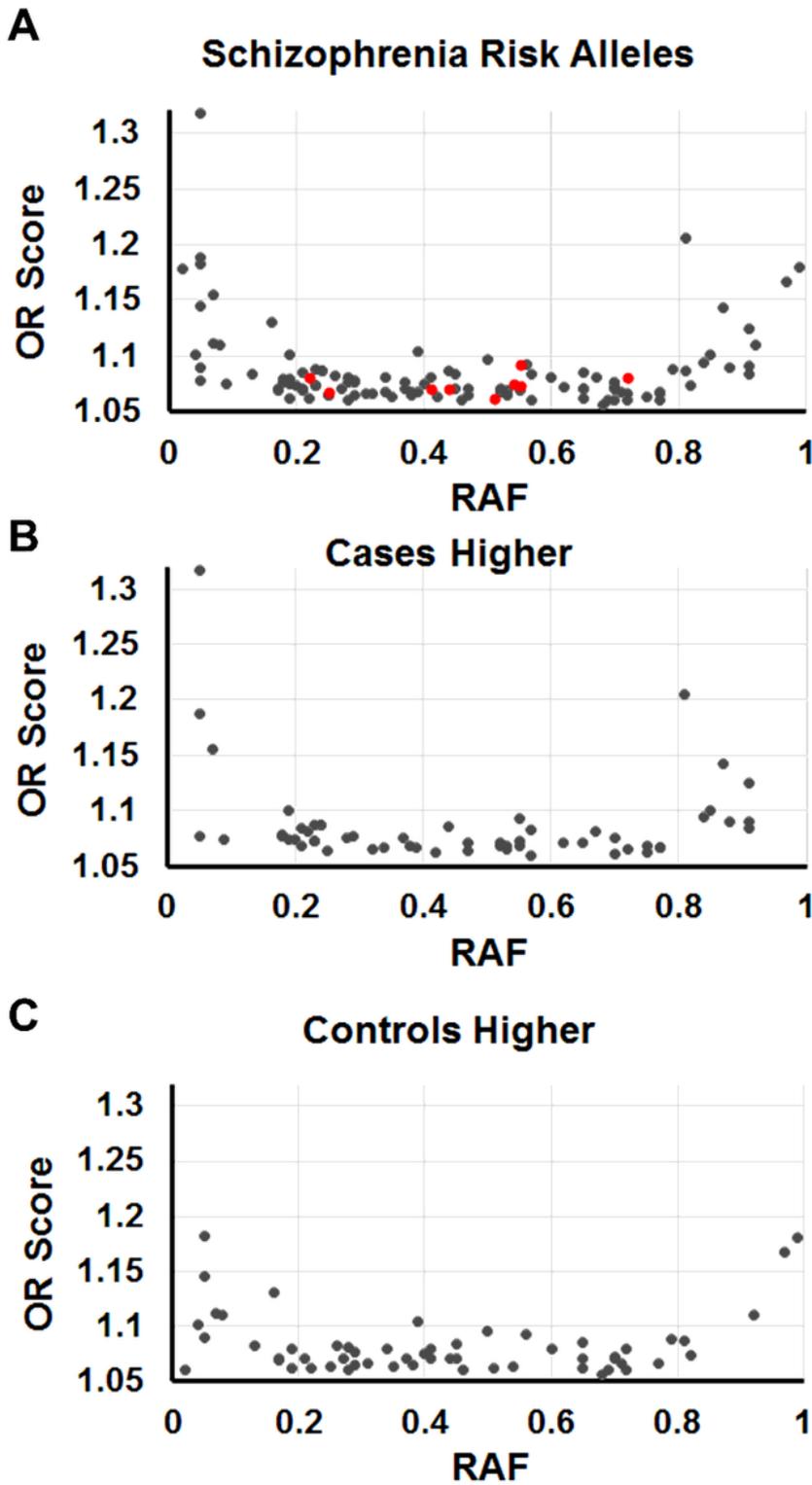


Figure 1

Distribution of OR scores as a function of RAF for the PGC risk alleles. (A) Data from 114 risk alleles are plotted. The red circles represent SNPs associated with functional syntenic blocks as described in the text. (B) Plots of 57 SNPs that occurred more frequently in cases of schizophrenia than controls and (C) 57 SNPs that were higher in controls than cases. It is worth noting that the case and control distributions are superficially very similar to each other.

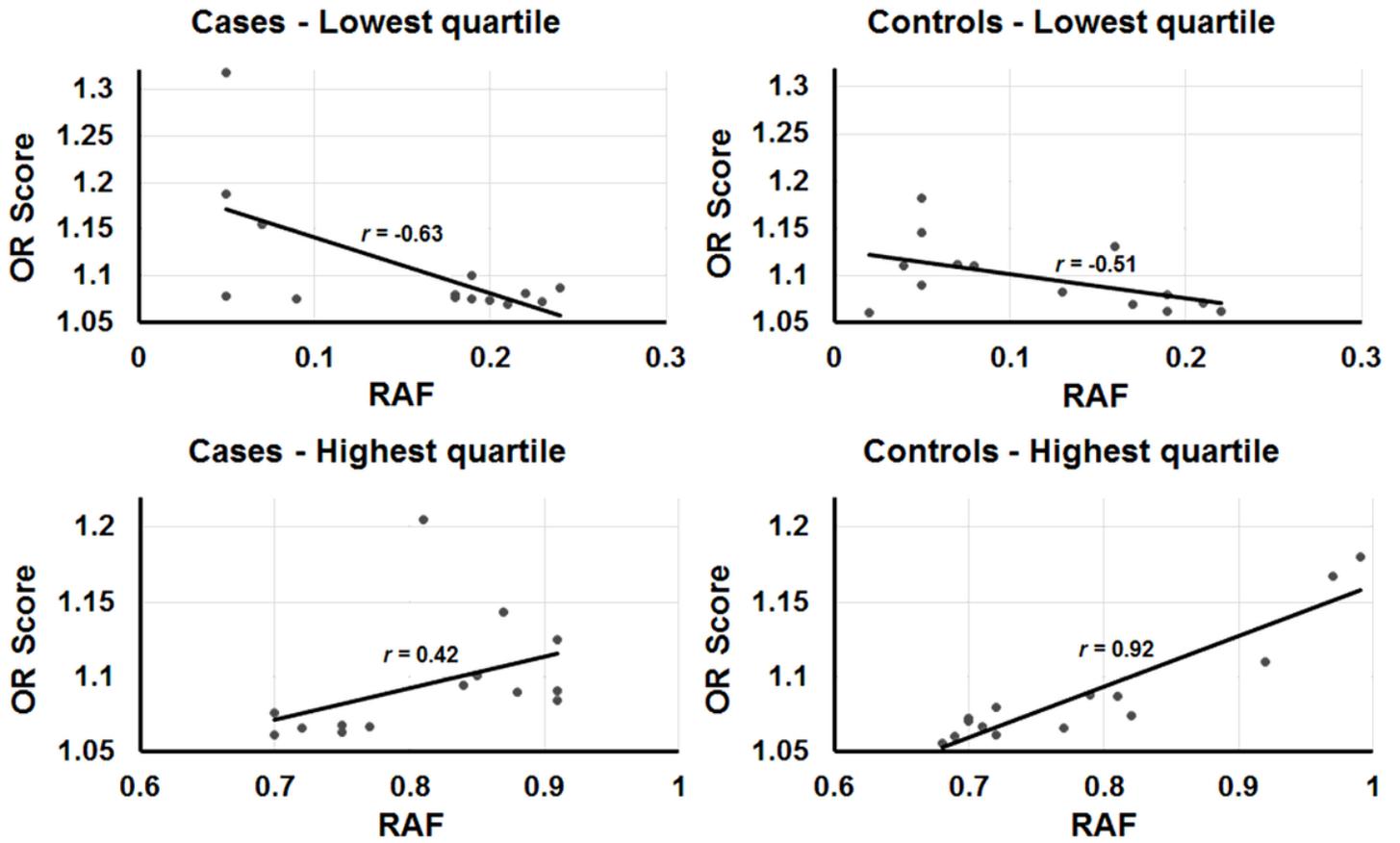


Figure 2

Regression analyses comparing RAF with OR score. Data from the highest and lowest quartiles were evaluated by standard regression analysis for relationships between their allele frequencies and adjusted OR scores. The plots show the lowest and highest quartile of data points from SNPs that occurred more frequently in Cases than Controls and vice versa (as labeled). The trend lines are included along with the correlation coefficient, r .

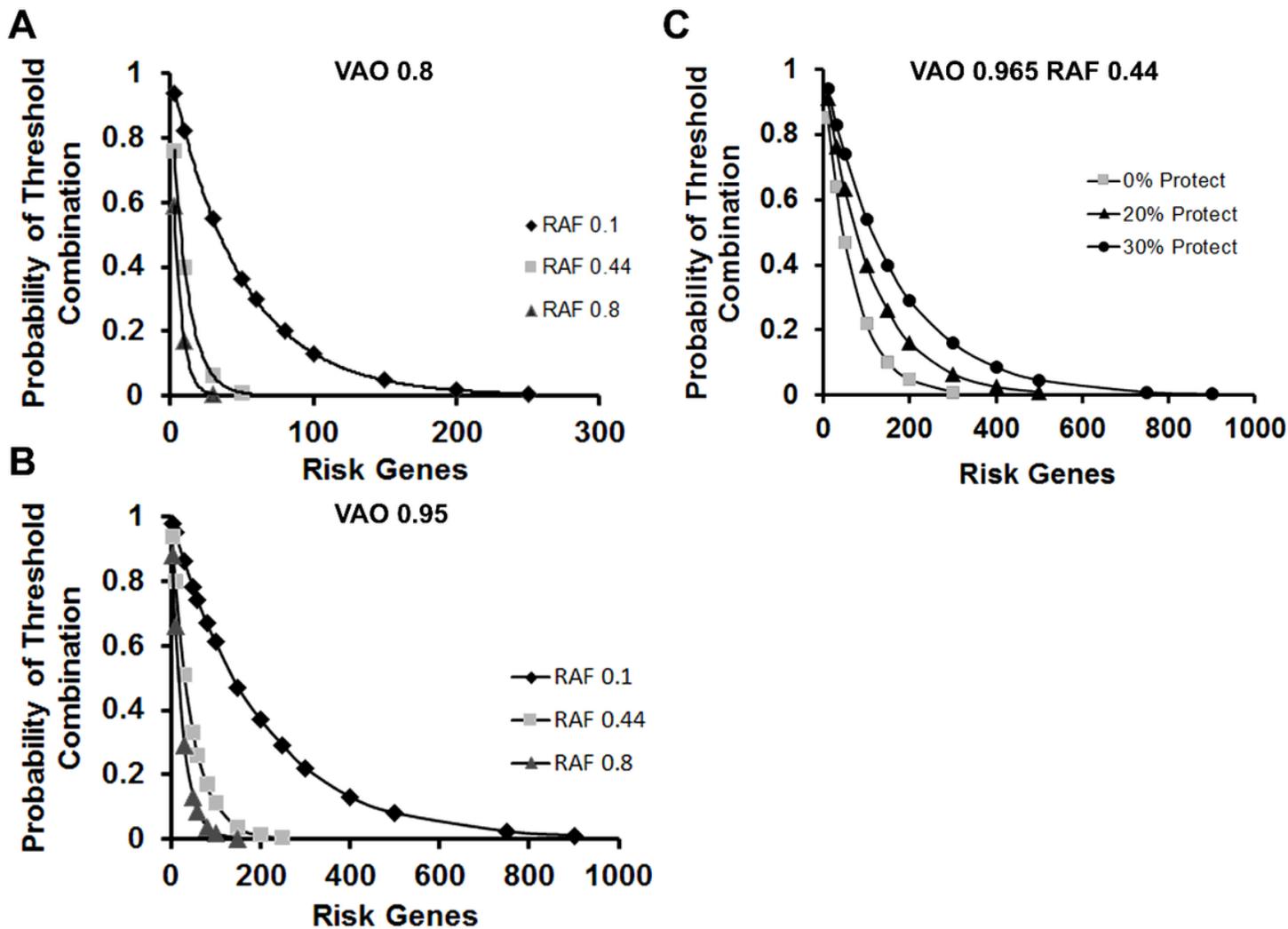


Figure 3

Risk gene estimation. Graphs were rendered for the number of risk genes required to reach the 0.01 probability of threshold risk-gene combination in the population (P_T) from pairs of alleles associated with risk polymorphisms. The VAOs are labeled: 0.8 (A) and 0.95 (B). The average RAFs used in equation (2) are indicated in the inset legends. (C) Estimation of risk alleles needed to reach threshold with protective variants included, either 20% or 30% of the total (as labeled). We assumed that the protective variants had a VAO of 1.035 with a resulting FLO of 1.015, which was equal in magnitude, but opposite in effect from the VAO of 0.965 (FLO = 0.985) for adverse risk variants.

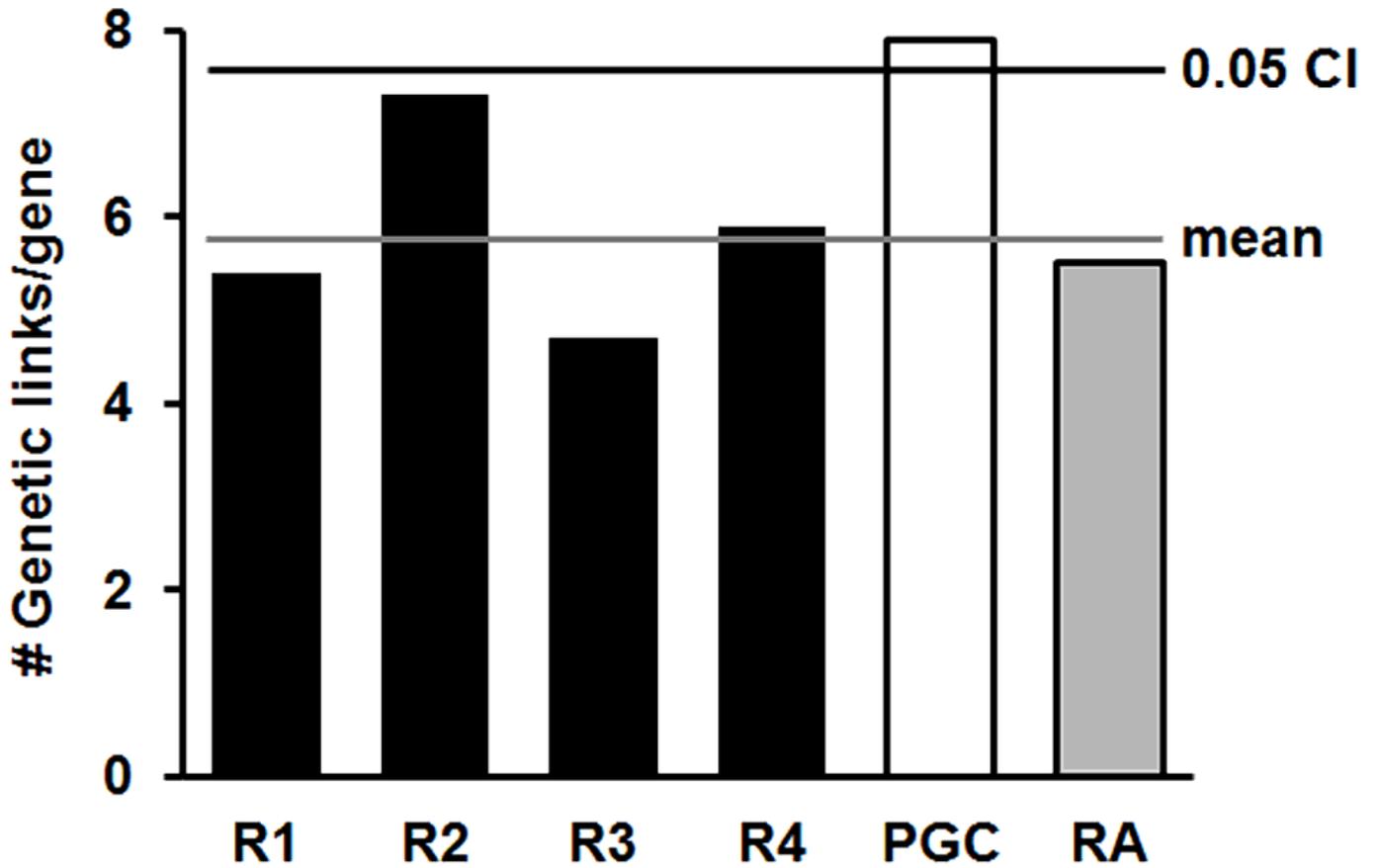


Figure 4

Genetic interactions among PGC risk genes. (A) The degree of genetic interaction among genes from 4 random gene lists (R1-R4) was analyzed with GeneMANIA and compared with the degree of interaction among risk-gene candidates from the PGC dataset and a rheumatoid arthritis (RA) dataset. The data are plotted as the number of genetic links per gene in the input lists that showed interactions in the Lin *et al.* (2010) dataset. The average of the 4 random lists is indicated with a gray line and the CI ($p < 0.05$) with a black line.

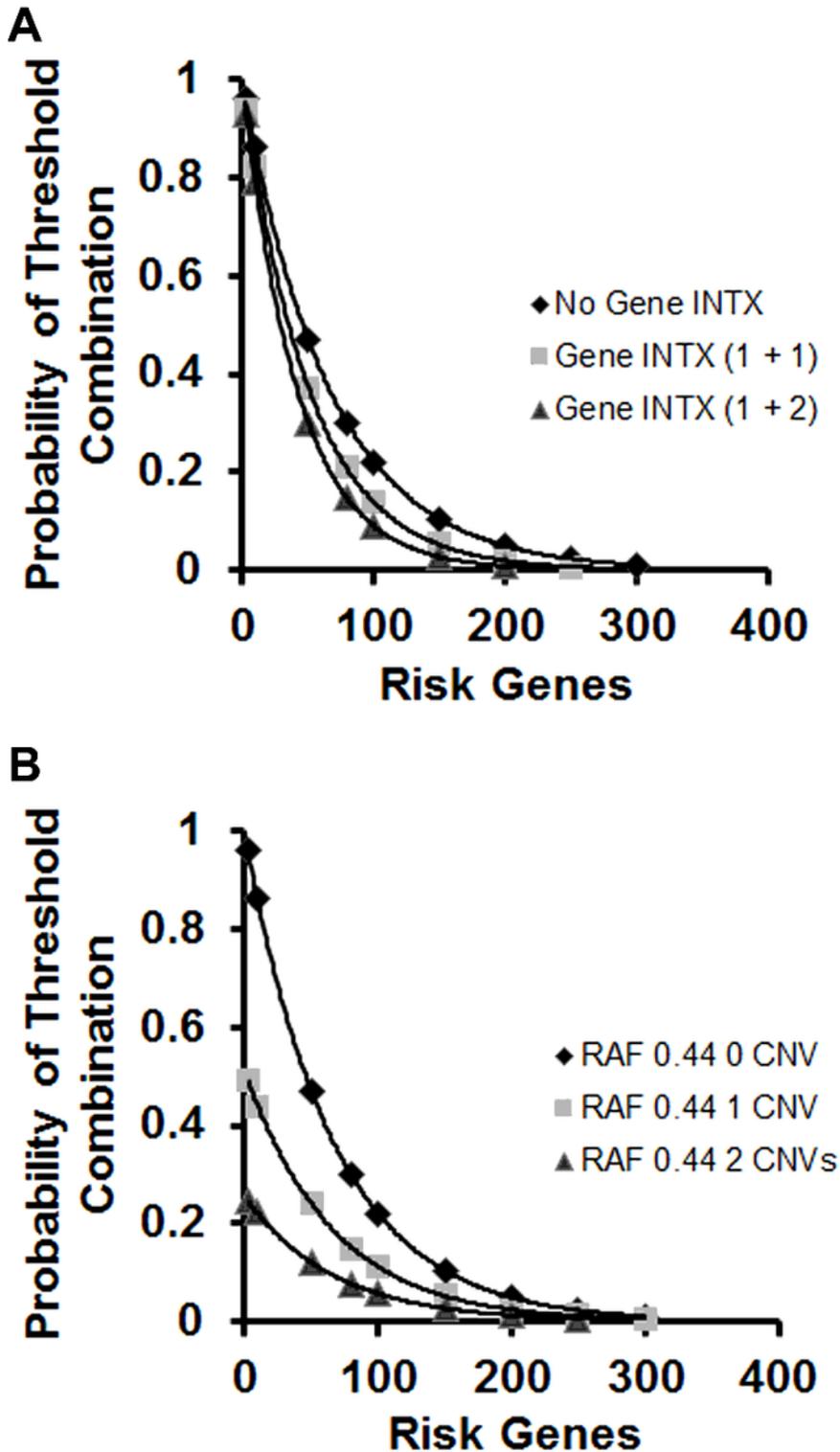


Figure 5

Estimation of risk gene threshold with inclusion of gene-gene interactions and CNVs. (A) Data were plotted for the number of risk genes required to reach the 0.01 probability of threshold combination when gene-gene interactions were included in the calculation. For this model, the VAO was 0.965 and the FLO 0.985 with RAF set at 0.44. Of the total risk genes, 30% were modeled as interacting with 1 (1 + 1) or 2 (1 + 2) other genes, which required substitution of either 0.97 (0.985^2) or 0.956 (0.985^3) into equation (2) for

genes with 1 or 2 partners, respectively. For comparison, calculations with no gene interactions (INTX) were also shown. (B) With the same settings as (A), the model included 0, 1 or 2 CNVs, which were represented with FLOs of 0.5.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation.docx](#)