Published in final edited form as:

Prev Sci. 2018 January ; 19(1): 49–57. doi:10.1007/s11121-016-0730-8.

Informing Prevention and Intervention Policy Using Genetic Studies of Resistance

Brion S. Maher¹, Shawn Latendresse², and Michael M. Vanyukov³

¹Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

²Department of Psychology and Neuroscience, Baylor University, Waco, TX

³Departments of Pharmaceutical Sciences, Psychiatry, and Human Genetics, University of Pittsburgh, Pittsburgh, PA

Abstract

The common paradigm for conceptualizing the influence of genetic and environmental factors on a particular disease relies on the concept of risk. Consequently, the bulk of etiologic, including genetic, work focuses on “risk” factors. These factors are aggregated at the high end of the distribution of liability to disease, the latent variable underlying the distribution of probability and severity of a disorder. However, liability has a symmetric but distinct aspect to risk, resistance to disorder. Resistance factors, aggregated at the low end of the liability distribution and supporting health and recovery appear to be more promising for effective prevention and intervention. Herein, we discuss existing work on resistance factors, highlighting those with known genetic influences. We examine the utility of incorporating resistance genetics in prevention and intervention trials and compare the statistical power of a series of ascertainment schemes to develop a general framework for examining resistance outcomes in genetically informative designs. We find that an approach that samples individuals discordant on measured liability, a low-risk design, is the most feasible design and yields power equivalent to or higher than commonly used designs for detecting resistance genetic and environmental effects.

Keywords

Genetic Resistance; Resilience; Prevention; High-risk design

The bulk of research into the human condition has focused on disease, disorder, or infection. Accordingly, guided by the medical priority of dealing with disease, it is the factors that increase the probability and severity of the disorder, i.e., “risk” factors, that are pursued in biomedical studies. These factors include environmental (e.g., infectious agents) and genetic (mutations and risk-associated alleles of genetic polymorphisms). Given the obvious success in identifying risk/causal factors for “simple” diseases (e.g., infections; Mendelian or single-

Corresponding author: Brion Maher, PhD, 624 N. Broadway Ave., Baltimore, MD 21205, brion@jhu.edu.

Compliance with Ethical Standards

Ethical approval. For this type of study ethical approval is not required.

Formal consent: For this type of study formal consent is not required.

Claims of potential conflict of interest. Drs. Latendresse, Vanyukov and Maher have no potential conflicts of interest to report.

gene disorders), researchers have consequently carried these approaches to the study of complex (multifactorial) disorders. It should be noted, however, that success in *translation* of findings even in Mendelian disorders has been low, as the strong genetic effects are difficult to compensate for or correct. For much more common complex disorders, the detection of single genetic effects is difficult because of their small impact: many genes contribute to individual differences in the risk, and that number is further augmented by the genetic and phenotypic (observed trait) heterogeneity of the disorder. As a rule, that implies the necessity of extremely large sample for such effects' detection, while compensating for them may be both difficult and ineffective.

Importantly, risk for a disorder is one aspect of liability, a latent trait comprising effects of all factors influencing an individual's probability of developing a disorder and its severity (Falconer, 1965). Hence, variation in liability to a complex disorder is due to individual differences in a large number of genetic and environmental influences. According to the liability-threshold model, whereby affected phenotypes are generally above a certain threshold point on the liability scale, higher values on the liability continuum correspond to higher risk for disease and increased severity (REICH, James, & Morris, 1972).

Vanyukov et al (Vanyukov et al., 2016) present the concept of *resistance* as the aspect of liability that is symmetric to risk (Figure 1). Accordingly, resistance factors are defined as those decreasing the probability and severity of a disorder. Work on resistance to disease has until now largely been limited to one of its facets, resilience. Resilience is conceptualized as "successful" adaptation after exposure to adversity (Luthar, Cicchetti, & Becker, 2000; Masten, 2001) and is thus contingent on the latter. Reviewed elsewhere (Cicchetti, 2010), adoption of a resilience perspective has allowed investigators to recast research on outcomes of high-risk populations, from the deterministic impact of the environment to individual differences in response to the environment. While this is valuable, resilience captures only the organism's dynamic response, or adaptation, to a stressor rather than the complete nature of disease risk, including that outside of the context of a particular exposure or exposure response. Resilience is also commonly inferred post-hoc upon observation of "successful" adaptation to a presumed stressor ("social disadvantage or highly adverse conditions" (Windle, 2002)), rather than being a disorder-specific measurable construct. Unsurprisingly, the genetic studies of resilience (Feder, Nestler, & Charney, 2009) are in effect studies of stress response mechanisms rather than resilience as a trait. The examples of findings reviewed by Feder and colleagues (Feder et al., 2009) are largely if not exclusively related to risk and disease (psychopathology) rather than lack thereof, which would be indicative of resistance (and potentially resilience). Adversity causing stress is also only one characteristic of the environment that potentially contributes to disorder liabilities.

While resilience has garnered a fair amount of attention, the broader concept of *resistance* has only recently been discussed (Vanyukov et al., 2016) as the preferred and necessary alternative to the "risk" perspective. Resistance factors are not limited to the individual's ability to rebound from or adapt to stressors (corresponding to resilience) and to the opposites of respective risk factors (usually described as "protective factors"), but also include factors that limit exposure to risk increasing environments and response to stressors or pathogens (e.g., drugs), increase exposure and sensitivity to environments which decrease

risk, as well as enable absence of relapse upon recovery. Whereas risk factors are expected to be aggregated among the affected and other “high-risk” individuals, resistance factors are likely to be aggregated among high-resistance individuals, representing the opposite end of the liability distribution (Figure 1).

Nevertheless, research has been focused on risk factors and high-risk sampling, which largely relies on oversampling affected cases or individuals expected to be at a high risk for a particular disorder. One reason for that is the belief that, akin to some infectious diseases, like tuberculosis, once the risk/etiological factor is identified, it can be removed. Another possible reason is the difficulty in selecting a priori low-risk samples: there are usually no overt face-valid indicators of high resistance that are comparable to the common indicators of high risk, the symptoms and the categorical diagnosis of the disorder. Whereas the low liability to a Mendelian disorder can be relatively readily ascertained (virtually anybody from the general population, if unaffected, is likely to have no/low risk), the same approach cannot be applied to multifactorial liabilities. When an allele is defined as “protective” in genetic association studies, it is usually by virtue of being alternative to the “risk” allele. However, such alternatives of risk factors, genetic and non-genetic, derived from the control samples representing ~90–95% of the population – are likely to be of an average/neutral rather than a high resistance effect (Vanyukov et al., 2016). The rare exceptions resulting in such an effect (e.g., the low activity ALDH2 allele for alcoholism, and smallpox vaccine) have historically been discovered due to the inadvertent selection of a high-resistance sample.

Genetics presents a telling example of the asymmetry of the “risk” approach. One of the once most popular association tests, the transmission-disequilibrium test (TDT), deals with alleles that are “overtransmitted” from heterozygous parents to the affected offspring. Not only are the unaffected offspring not studied, but the “undertransmitted” alleles are disregarded – obviously because they are not “risk” alleles. Regrettably, such family-based studies were even proposed to be the only type of genetic association research acceptable for publication consideration, with case-control research rejected in its entirety (Paterson, 1997), preceding the ascendance of whole genome scans, which are largely case-control studies ill-powered to detect resistance factors. As currently construed, however, neither genetic case-control nor family-based studies target resistance factors and enable their efficient detection. Below, we discuss application of the case-control design to introduce changes needed to reorient research from risk to resistance. We focus initially on genetic resistance factors in addiction, where significant successes have been realized, and extend that to other cases.

Addiction

Investigations of the impact of measured genetic (i.e., DNA polymorphisms) influences on variation in the liability to drug addiction, in common with other complex disorders, have had limited success in explaining its total heritability, the proportion of the trait’s variance that is due to genetic factors. Replicable findings of a meaningful effect of genetic variants on disease risk or trajectory in drug addiction have been few (Bierut, 2011; Hartz & Bierut, 2010; Maher et al., 2011; Treutlein & Rietschel, 2011), both for specific drugs (Tobacco and

Genetics Consortium, 2010; Xie et al., 2011) as well as for general liability to addiction (Maher et al., 2011; Schwantes-An et al., 2015).

As noted above, case-control designs focus on identifying factors for increased disease risk and increase power to identify risk factors by oversampling cases and consequently enriching the sample for ‘risk’ factors. Early genetic studies of addiction frequently contrasted drug-dependent cases and age/sex matched unaffected population controls. However, pathogenesis of substance use disorder (SUD) is a multi-stage process, progressing first through drug use, then abuse, and ultimately psychological and physiological dependence and compulsive pursuit of drugs (addiction), with varying influence of genes and environment on individual variation by stage. Including controls who have not ever used drugs hinders focusing on a specific stage of dependence. Study designs that use controls who report drug exposure but do not satisfy criteria for SUD (addiction) likely deliver greater power to detect allelic associations pertaining to risk for dependence, since it is unclear if controls who have never used would develop dependence if exposure occurred. Consistent with this observation, the recent successes in addiction genetics have been in studies focusing on variation within drug-exposed populations or included controls with significant exposure (Cornelis et al., 2011; Sulem et al., 2011; Tobacco and Genetics Consortium, 2010). In genetic studies of tobacco addiction (e.g., (S. F. Saccone et al., 2007; Tobacco and Genetics Consortium, 2010), the importance of using screened controls is highlighted. Using a very large sample (N=143,023), Tobacco and Genetics Consortium researchers performed a genome-wide association scan, or tests of association between approximately one million SNPs genome-wide, and phenotypes of smoking including age of initiation, use and cessation, with minimal success. In a smaller sample restricted to smokers (N=73,853) for smoking quantity, a highly significant association of characteristics of smoking with the nicotinic acetylcholine receptor gene cluster (CHRNA5-A3-B4) on chromosome 15 was replicated. In prior studies, similar results had been found using controls who were limited to smokers who had not become dependent (N. L. Saccone et al., 2010). It is believed that specific variants in these loci confer elevated risk to nicotine dependence after commencing smoking via differences in corticostriatal response to nicotine (Janes et al., 2012).

It is noteworthy that it is risk for, rather than resistance to, the development of dependence that has been targeted in most research. Resistance-enhancing effects, if any reported, are either defined by an allele’s virtue of being the alternative to the “risk” allele, or noted inadvertently, as, for instance, the detection of the “protective” effect of the minor allele of the rs16969968 SNP of the CHRNA5 gene (a “risk” allele for nicotine dependence) on the risk for cocaine dependence (Gruza et al., 2008). The largest addiction-related genetic effect, however, has been established when a phenotypic resistance factor was specifically targeted. A higher sensitivity to ethanol (“flushing” upon exposure) among East Asians (particularly, the Japanese) was established by comparison with Caucasians (Wolff, 1972) and subsequently experimentally related to a higher level of acetaldehyde upon exposure to ethanol (Wolff, 1973). Flushing among the Japanese was then related to the “atypical” form of acetaldehyde dehydrogenase deficiency (Harada, Agarwal, & Goedde, 1981) and, finally, to the ALDH2*2 allele of the gene coding for the β 2 subunit of the enzyme (Goedde, Agarwal, & Harada, 1983). Variants in genes involved in alcohol metabolism leading to

increases in acetaldehyde either through the increased oxidation of alcohol to acetaldehyde (ADH1B) or through slowing the rate of conversion of acetaldehyde to acetate (ALDH2) have been associated with quantitative differences in alcohol consumption (Crabb, Edenberg, Bosron, & Li, 1989; Luczak, Glatt, & Wall, 2006; Peng & Yin, 2009). Obviously, these resistance genetic effects are discoverable only in samples examining use variation in those who have tried drinking. Consequently, the use of exposed controls has become a more common approach in substance abuse genetics. It should be noted, however, that while exposed controls facilitate the detection of factors related to elevated risk for physical dependence, the factors – including genetic ones – that lower the probability of drug exposure and thus enhance resistance are undetectable with this sampling scheme. To a large degree, the same pertains to the factors that influence the magnitude of exposure (usually limited to a substantial but relatively low number of drug experiences).

Other diseases

The recent explosion of genomic data has allowed systematic searches for “fairy godmother” mutations or genetic variants that confer disease resistance (Williams, 2016). For example, recent work screened genomic sequence data from >589,000 subjects to identify 13 subjects with Mendelian disease causing mutations but without apparent disease. Other work has concluded that roughly 5% of healthy adults carry mutations predicted to “cause” disease (Johnston et al., 2015) but that screening of cohorts is vastly underpowered to detect specific protective factors, whether genetic or environmental. Hence, investigators focused on specific diseases or mechanisms have relied on several different, and frequently ad hoc, approaches to identify resistance variants.

The usual approach to discovering protective variants has been to focus on those thought to be at risk but without disease. This approach has been frequently applied to infectious disease, where exposure risk is known. Possibly most notable, a specific variant in CCR5, the gene encoding for an HIV co-receptor for binding and entry into CD4+ T cells called the CC-chemokine receptor 5, confers resistance to HIV infection. The 32 base pair deletion, CCRdelta32, results in a nonfunctioning CCR5 and confers immunity to HIV to those homozygous for the mutation (Dean, Carrington, Winkler, & Huttley, 1996; Liu et al., 1996). Importantly, identifying these variants relied on the study of highly exposed controls. Utilizing a novel study design, focusing on HIV-seropositive subjects with and without tuberculosis from high TB exposure regions, Sobota and colleagues (Sobota et al., 2016) successfully identified a TB-resistance variant in a histone mark on Chromosome 5 near IL12B. Other work using exposed controls has identified, or is aiming to identify, resistance variants for other infectious diseases, including influenza (Hsu & Spindler, 2012), malaria (Shelton et al., 2015), Kuru (Mead et al., 2009) and other prion diseases (Asante et al., 2015).

The concept of biological resistance exists and has relevance beyond infectious disease and addiction, however identification of those “at risk” may often be less discrete than having been exposed to an infectious pathogen or addictive substance and instead rely on ad hoc definitions of liability. Freudenberger-Hua (Freudenberger-Hua et al., 2014) focused on whole genome sequencing of 44 centenarians with known common disease risk factors (e.g.,

obesity, smoking) but who were healthy in a quest to identify disease-protective variants. The study of overweight individuals without Type 2 Diabetes successfully identified resistance variants in SLC30A8 (Flannick et al., 2014). In a study of 500 Ashkenazi Jews over 95 years old, adipopectin resistance mutations, protective against arterial inflammation, and resistance mutations in the cholesterylester transfer protein gene, protective against high cholesterol and consequent health outcomes, were identified (Barzilai & Gabriely, 2010; Schechter, Barzilai, Crandall, & Atzmon, 2010). Not all investigations of genetic resistance in subjects at-risk for but without disease have yielded similar successes. For example, recently a genome sequencing study of 454 “welllderly” subjects over 80 with no chronic disease, failed to identify protective variants (Erikson et al., 2016). Thus, a framework for investigating genetic resistance is needed to avoid further reliance on serendipitous discovery or secondary analyses in samples not ascertained to study resistance. The need for innovative study design is more pronounced in prevention trials, where the use of random samples of hundreds of thousands of subjects to identify tens of resistant subjects is infeasible.

The goal of this work is to explore the power of a range of study designs to discover novel genetic and environmental resistance factors. Those study designs include case-control, representative sampling, a low risk-design, and the high-risk design. We also examine modifications of the high-risk design, an approach where subjects are selected based on the presence of a phenotype of interest in a parent. The intent of the high-risk design is to enrich for “risk” factors, both genetic and family-environment, to increase the probability of disease outcomes in the offspring. We treat the case-control design as the reference approach, since it is the most widely used approach to over-ascertainment of cases relative to population prevalence in order to gain power. For resistance studies, we explore whether sampling based on environmental exposures and parental phenotype impacts the power to detect specific genetic and environmental resistance factors. We use a simulation-based approach and assume an underlying continuous disease liability to compare the power to detect novel resistance factors under each study design.

Materials and Methods

In this section we describe (1) the study designs aimed at detecting resistance factors; (2) the multifactorial liability threshold model used to generate disease status; and (3) the approach to data simulation we used.

Study Designs

We compared the power of six different study designs to detect resistance effects for a dichotomous outcome by simulation (Figure 2). Our intent is to mimic sampling strategies that could be used in a typical intervention or prevention trial, with subjects ascertained prior to onset of the outcome and assume a sampling unit of a parent-child trio. We treated a case-control sampling strategy as the reference group (Figure 2a) with a sample of 500 trios with an affected offspring and 500 control trios, irrespective of parental affection status. We examined random sampling of 1000 trios from the population (Figure 2b), where trios were simulated as randomly ascertained from the population with respect to all parameters. We

included a low-risk (high resistance) design (Figure 2c). Under the low-risk design, a sample of 1,000 offspring from a larger total population is selected from below the 10th percentile and above the 60th percentile respectively on a known liability distribution. We examined the power of the high-risk design, where subjects were ascertained through the father, with 500 case fathers and 500 control fathers (Figure 2d). We also examined two modifications of the high risk design, both of which included sampling conditional on a single measured continuous environmental factor which accounted for 25% of the total variance in the underlying liability. Under the low environmental risk sampling approach (Figure 2e), 500 high-risk (case father) and 500 low-risk (control father) trios were selected from the bottom half of the measured environmental risk scale. Under the extreme environmental sampling approach, 500 high-risk (affected father) trios were selected with values on the environmental risk below the mean and 500 low-risk (control father) trios were selected with environmental risk above the mean (Figure 2f).

Disease Model

We assumed a liability threshold model, where an underlying continuous risk trait captures all genetic and environmental contributions to risk and resistance. Under this model and for this work, this is operationalized as:

$$Z_i = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$$

where Z_i is the liability trait, β_1 is the correlation between polygenic risk X_1 and outcome (assuming standardized coefficients), β_2 is the correlation between family environment X_2 and outcome, β_3 is the correlation between a specific SNP genotype X_3 and outcome, β_4 is the correlation between a specific protective environment X_4 and outcome, and ε is the residual variance. Outcomes were dichotomized to a disease outcome (Y_i) by:

$$\text{Prob}(Y_i) = \frac{e^{Z_i}}{1 + e^{Z_i}}$$

Simulation

Samples of trios were directly simulated via Monte Carlo simulation. First mother and father polygenic risk, family environmental risk, and individual protective environment were simulated from a standard normal distribution. A single resistance SNP was simulated in parents. Genetic variables were simulated in the offspring conditional on parental SNPs using gene-dropping simulation and on parental polygenic risk as the mean of parental polygenic risk. Given a set of correlations and residual variance, a random normal (0,1) underlying risk was directly simulated per individual. Risk was dichotomized to disease status via the logit model with a constant added to simulate a specified population prevalence of disorder. In all simulations, we assumed a population prevalence of 5%. For the case-control and high-risk design samples we over-simulated to generate 500 case or high-risk families from the stated environmental quantiles, respectively. In the low-risk design, we over-simulated to generate 500 subjects with liability at or below the 10th percentile and 500 subjects above the 60th percentile. In the random design, we simulated

1000 trios irrespective of outcome and environmental stratum. We selected parameter values for genetic and environmental effects similar to those seen for common behavioral outcomes. However, since the goal of this work is to examine the comparative power of these approaches, the values of the parameters is not of critical importance since the relative power of the approaches is unlikely to change under different effect sizes. We assumed a minor allele frequency of the protective allele of 0.3, an overall trait heritability of 0.50, the variance accounted for by the family environment of 0.25. We varied the impact of the protective SNP and individual environment. For each of 1000 iterations under each study design, we tested the association between the offspring protective SNP and protective individual environment with outcome via logistic regression. Power was calculated at $\alpha = .05$ or the proportion of Monte Carlo replicates that yielded a p-value less than .05.

Results

The results are presented in Table 1 corresponding directly to the models in Figure 2. The impact of the protective environment was varied at R^2 of .0225, .04, and .0625. The effect of the protective allele was varied at effect sizes, indexed as deviation from the heterozygote mean of $-.15$, $-.20$ and $-.25$, yielding an additive genetic variance of .009, .017 and .026 respectively at a minor (protective) allele frequency of .3. Power is presented at an alpha of .05. It is important to note that relative, and not absolute, power is of interest in comparative analyses such as this. It is expected that any factors which would increase or decrease power, such as change in effect size or sample size, would similarly affect the power of all approaches. As expected, case-control sampling yields substantial improvements over random sampling for detecting protective SNP and environment effects across all effect sizes. The low risk design, where subjects are ascertained at a low liability, yields power equivalent to a case-control design to detect SNP effects, and much higher power to detect the protective environmental effect. The high-risk design yields power increases similar to case-control sampling with the exception of at smaller SNP effect sizes, where the high-risk design performs better. This is most likely due to an enrichment in SNP effects in high risk families with control offspring. Sampling with a high-risk design from an environmental risk below the mean yields a power profile similar to a case-control design. Sampling high-risk families from below the population mean in environmental risk and low-risk families from above the population mean yields dramatic increases in power across the range of effects to detect both protective SNP and environmental effects.

Discussion

Since the advent of the genome-wide era, the approach has been plagued by the multiple testing problem, something that was actually recognized before most genome-wide analyses were performed (Zaykin & Zhivotovsky, 2005). Realistic effect sizes may not stand out among a sea of 1 million tests. The problem of separating the genomic wheat from the chaff has led many to question the value of large scale testing. Prevention trials are faced with an additional difficulty in that sampling occurs prior to the onset of disorder or the transition of interest. We performed simulation under a range of effect sizes. While the selection of sample and effect sizes to represent real world settings is important, we present power of the approaches for comparative purposes. That is, while the absolute power of a given sample

size at a given effect is interesting, we are interested here only in the relative power of the designs and expect that the relative power of the designs will remain constant across sample size and effect size.

We demonstrated that sampling approaches that increase the proportion of resistant cases versus random sampling or case-control sampling lead to increases in the power to detect resistance effects. The high-risk design, and modified approaches, yield power similar to, or much greater than, case-control sampling by increasing the presence and variability of factors that impact the liability distribution. Notably, a sampling approach that ascertains case and control fathers from below and above the mean on an environmental risk factor respectively yields dramatic increases in power to detect resistance factors. However, this approach, which is contingent both on parental phenotype and measured parental environment would rely on a much larger sample from which sub-sampling would occur, thus substantially limiting its feasibility. The obvious exception to this is extant cohorts with parental data. More importantly, the low-risk design which ascertains based solely on measured liability irrespective of parental phenotype, and is hence much more feasible, yields increases in power similar to the high-risk designs with the previously mentioned exception of extreme environmental risk sampling. We contend that the low-risk design is substantially more feasible and similar to sampling approaches that have been used with success in complex disease genetics for more than a decade (N. Risch & Zhang, 1995; N. J. Risch & Zhang, 1996).

Successful implementation of the low risk design, the most powerful approach that is not contingent upon parental phenotypes or the presence/absence of a particular outcome (the usual scenario in prevention trials), presents several barriers which are easy to overcome. The first is the establishment of common measures of resistance. An important way of defining a high resistance phenotype is via developing a continuous index of liability to the disorder and selecting individuals from the low end of the distribution. Such an index, for instance, has been developed for measurement liability to drug addiction in children, before any exposure to drugs (Vanyukov, Tarter et al., 2003; Vanyukov, Kirisci et al., 2003; Vanyukov et al., 2009; Vanyukov et al., 2016). A second related barrier is the development and use of items, scales and constructs that measure not just the “bad” but also the “good”. A vestige of the disease era of research is the widespread existence of items, test and scales designed to measure only a small part of the liability distribution, usually most informative around the threshold between “normal” and “disease”. The recent movement to researching “domains” is an important step in the direction of observing the full range of liability, at one end capturing those at high risk for affection with disease or in the disease spectrum, and at the other, those with high resistance phenotypes. Similarly, polygenic approaches, which index tens, hundreds, or thousands of SNPs to create composite indices of genetic risk, have been developed (Maher, 2015) and it is worthwhile to consider those as part of multifactorial resistance in the same way that polygenic risk scores are used in the context of multifactorial risk. To fully inform research on resistance, an integration of biology, context and their interplay is essential. As described by Falconer, liability is a latent trait comprising effects of all factors, genetic and environmental, influencing an individual’s probability of developing a disorder (Falconer, 1965). Hence, scales developed to measure liability for sampling into

prevention trials should comprise not only measures of phenotypic precursors but also genetic and environmental factors known to impact liability variation for studied outcomes.

While risk and resistance are symmetric liability aspects, resistance factors are not necessarily the opposites or alternative variants of the risk factors. This differentiates resistance from a more narrow concept of protective factors, which are “opposite ends of the same continuum [as risk factors]” or, alternatively, moderate, mediate or buffer the effects of risk factors (Hawkins, Catalano, & Miller, 1992). Aggregation, and thus the probability of detection, of risk and resistance factors differ in different portions of the liability distribution (Fig. 1). These sets of factors are likely to partially overlap. An example could be a SNP allele conferring elevated risk and the alternative allele that may confer enhanced resistance. The alternative allele, however, may be neutral, i.e., of no discernible phenotypic effect on the background of all other factors, rather than protective. Analogously, a resistance-enhancing variant may exist and be detectable within high-resistance populations, while the alternative variant(s) is neutral: e.g., the ALDH2*2 allele (Goedde et al., 1983; Harada et al., 1981). The rest of ALDH2 variation only weakly, if at all, influences alcoholism risk (Macgregor et al., 2009). The distinction between the risk and resistance perspectives is nontrivial also because it entails differences in approaches to the identification of respective factors. Research that is focused on risk factors alone is likely to detect those with the strongest risk-increasing effects, aggregated, e.g., among the affected individuals, representing the high end of the liability distribution (5–10% of the population). In contrast, phenotypes carrying strongest resistance-increasing effects – genetic or environmental – are diluted by the heterogeneity of the typical unaffected control, sampled from 90–95% of the population. Symmetric to high-risk groups, factors conferring high resistance are aggregated among those with phenotypes at the low end of the liability scale, requiring high-resistance sampling for their detection.

Factors enhancing resistance may have the highest translational impact. The smallpox vaccine is an example. It was developed before the main risk factor, the variola virus, was discovered, and vaccine production does not even involve this virus. The immune milkmaids, inoculated with cowpox, were the high-resistance population in the discovery of the smallpox vaccine. This example shows that resistance factors are not merely opposite to risk factors, and that extending the action of a resistance factor to the rest of the population may be more practicable than the removal of a risk factor. The most effective measures for HIV/AIDS prevention – condom use and abstention from risky sex or, for injection drug users, from needle-sharing – are also not specific to HIV but increase resistance to infection. Dealing directly with the “pathogen” is appropriate in some cases (e.g., antiretroviral therapy for HIV; antibiotics for bacteria). Such an approach may be infeasible or ineffective in many other cases, while raising resistance to the pathogen may be more promising.

Researchers can use our findings in two clear ways. First, to guide etiologic research to enhance the likelihood of discovering resistance factors. Second, our results can guide the modification of existing approaches to prevention and intervention research, to enhance our ability to identify factors that increase resistance to disease among apparently “at-risk” subjects. Importantly, inclusion of resistant subjects in a randomized trial will limit the

apparent effectiveness of the intervention unless those resistance factors are known and properly accounted for in the design.

Acknowledgments

Funding. This work was supported by National Institute on Drug Abuse (NIDA) Grants R01DA036525 and R01DA039408, and National Institute on Alcoholism and Alcohol Abuse Grant K01AA020333.

References

- Asante EA, Smidak M, Grimshaw A, Houghton R, Tomlinson A, Jeelani A, Linehan JM. A naturally occurring variant of the human prion protein completely prevents prion disease. *Nature*. 2015
- Barzilai N, Gabriely I. Genetic studies reveal the role of the endocrine and metabolic systems in aging. *The Journal of Clinical Endocrinology & Metabolism*. 2010; 95(10):4493–4500. [PubMed: 20926537]
- Bierut LJ. Genetic vulnerability and susceptibility to substance dependence. *Neuron*. 2011; 69(4):618–627. [PubMed: 21338875]
- Cicchetti D. Resilience under conditions of extreme stress: A multilevel perspective. *World Psychiatry*. 2010; 9(3):145–154. [PubMed: 20975856]
- Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, Bennett SN, Caporaso NE. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genetics*. 2011; 7(4):e1002033. [PubMed: 21490707]
- Crabb DW, Edenberg HJ, Bosron WF, Li TK. Genotypes for aldehyde dehydrogenase deficiency and alcohol sensitivity. the inactive ALDH2(2) allele is dominant. *The Journal of Clinical Investigation*. 1989; 83(1):314–316. [PubMed: 2562960]
- Dean M, Carrington M, Winkler C, Huttley GA. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. *Science*. 1996; 273(5283):1856. [PubMed: 8791590]
- Erikson GA, Bodian DL, Rueda M, Molparia B, Scott ER, Scott-Van Zeeland AA, Topol EJ. Whole-genome sequencing of a healthy aging cohort. *Cell*. 2016; 165(4):1002–1011. [PubMed: 27114037]
- Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics*. 1965; 29(1):51–76.
- Feder A, Nestler EJ, Charney DS. Psychobiology and molecular genetics of resilience. *Nature Reviews Neuroscience*. 2009; 10(6):446–457. [PubMed: 19455174]
- Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, Benediktsson R. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nature Genetics*. 2014; 46(4):357–363. [PubMed: 24584071]
- Freudenberg-Hua Y, Freudenberg J, Vacic V, Abhyankar A, Emde A, Ben-Avraham D, Koppel J. Disease variants in genomes of 44 centenarians. *Molecular Genetics & Genomic Medicine*. 2014; 2(5):438–450. [PubMed: 25333069]
- Goedde HW, Agarwal DP, Harada S. The role of alcohol dehydrogenase and aldehyde dehydrogenase isozymes in alcohol metabolism, alcohol sensitivity, and alcoholism. *Isozymes*. 1983; 8:175–193. [PubMed: 6354997]
- Gruzza RA, Wang JC, Stitzel JA, Hinrichs AL, Saccone SF, Saccone NL, Budde JP. A risk allele for nicotine dependence in CHRNA5 is a protective allele for cocaine dependence. *Biological Psychiatry*. 2008; 64(11):922–929. [PubMed: 18519132]
- Harada S, Agarwal D, Goedde H. Aldehyde dehydrogenase deficiency as cause of facial flushing reaction to alcohol in Japanese. *The Lancet*. 1981; 318(8253):982.
- Hartz SM, Bierut LJ. Genetics of addictions. *The Psychiatric Clinics of North America*. 2010; 33(1):107–124. [PubMed: 20159342]
- Hawkins JD, Catalano RF, Miller JY. Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin*. 1992; 112(1):64. [PubMed: 1529040]

- Hsu T, Spindler KR. Identifying host factors that regulate viral infection. *PLoS Pathogens*. 2012; 8(7)
- Janes AC, Smoller JW, David SP, Haddad S, Basu A, Fava M, Kaufman MJ. Association between CHRNA5 genetic variation at rs16969968 and brain reactivity to smoking images in nicotine dependent women. *Drug and Alcohol Dependence*. 2012; 120(1):7–13. [PubMed: 21764527]
- Johnston JJ, Lewis KL, Ng D, Singh LN, Wynter J, Brewer C, Gonsalves SG. Individualized iterative phenotyping for genome-wide analysis of loss-of-function mutations. *The American Journal of Human Genetics*. 2015; 96(6):913–925. [PubMed: 26046366]
- Liu R, Paxton WA, Choe S, Ceradini D, Martin SR, Horuk R, Landau NR. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell*. 1996; 86(3):367–377. [PubMed: 8756719]
- Luczak SE, Glatt SJ, Wall TJ. Meta-analyses of ALDH2 and ADH1B with alcohol dependence in asians. *Psychological Bulletin*. 2006; 132(4):607. [PubMed: 16822169]
- Luthar SS, Cicchetti D, Becker B. The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*. 2000; 71(3):543–562. [PubMed: 10953923]
- Macgregor S, Lind PA, Bucholz KK, Hansell NK, Madden PA, Richter MM, Whitfield JB. Associations of ADH and ALDH2 gene variation with self report alcohol reactions, consumption and dependence: An integrated analysis. *Human Molecular Genetics*. 2009; 18(3):580–593. [PubMed: 18996923]
- Maher BS. Polygenic scores in epidemiology: Risk prediction, etiology, and clinical utility. *Current Epidemiology Reports*. 2015; 2(4):239–244. [PubMed: 26664818]
- Maher BS, Vladimirov VI, Latendresse SJ, Thiselton DL, McNamee R, Kang M, Vanyukov MM. The AVPR1A gene and substance use disorders: Association, replication, and functional evidence. *Biological Psychiatry*. 2011; 70(6):519–527. [PubMed: 21514569]
- Masten AS. Ordinary magic: Resilience processes in development. *American Psychologist*. 2001; 56(3):227. [PubMed: 11315249]
- Mead S, Whitfield J, Poulter M, Shah P, Uphill J, Campbell T, Mein CA. A novel protective prion protein variant that colocalizes with kuru exposure. *New England Journal of Medicine*. 2009; 361(21):2056–2065. [PubMed: 19923577]
- Paterson AD. Case-control association studies in complex traits--the end of an era? *Molecular Psychiatry*. 1997; 2:277–278. [PubMed: 9246662]
- Peng G, Yin S. Effect of the allelic variants of aldehyde dehydrogenase ALDH2* 2 and alcohol dehydrogenase ADH1B* 2 on blood acetaldehyde concentrations. *Human Genomics*. 2009; 3(2):1.
- REICH T, James J, Morris C. The use of multiple thresholds in determining the mode of transmission of semi-continuous traits*. *Annals of Human Genetics*. 1972; 36(2):163–184. [PubMed: 4676360]
- Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science*. 1995; 268(5217):1584. [PubMed: 7777857]
- Risch NJ, Zhang H. Mapping quantitative trait loci with extreme discordant sib pairs: Sampling considerations. *American Journal of Human Genetics*. 1996; 58(4):836–843. [PubMed: 8644748]
- Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, Bierut LJ. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: A meta-analysis and comparison with lung cancer and COPD. *PLoS Genetics*. 2010; 6(8) 10.1371/journal.pgen.1001053.
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, Bierut LJ. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Human Molecular Genetics*. 2007; 16(1):36–49. doi:ddl438 [pii]. [PubMed: 17135278]
- Schechter CB, Barzilai N, Crandall JP, Atzmon G. Cholesteryl ester transfer protein (CETP) genotype and reduced CETP levels associated with decreased prevalence of hypertension. *Mayo Clinic Proceedings*. 2010; 85(6):522–526. [PubMed: 20511482]
- Schwantes-An T, Zhang J, Chen L, Hartz SM, Culverhouse RC, Chen X, Konte B. Association of the OPRM1 variant rs1799971 (A118G) with non-specific liability to substance dependence in a collaborative de novo meta-analysis of european-ancestry cohorts. *Behavior Genetics*. 2015:1–19. [PubMed: 25502189]

- Shelton JM, Corran P, Risley P, Silva N, Hubbard C, Jeffreys A, Hensmann M. Genetic determinants of anti-malarial acquired immunity in a large multi-centre study. *Malaria Journal*. 2015; 14(1):1–18. [PubMed: 25557741]
- Sobota RS, Stein CM, Kodaman N, Scheinfeldt LB, Maro I, Wieland-Alter W, Chervenak K. A locus at 5q33. 3 confers resistance to tuberculosis in highly susceptible individuals. *The American Journal of Human Genetics*. 2016; 98(3):514–524. [PubMed: 26942285]
- Sulem P, Gudbjartsson DF, Geller F, Prokopenko I, Feenstra B, Aben KK, Stefansson K. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. *Human Molecular Genetics*. 2011; 20(10):2071–2077. [PubMed: 21357676]
- Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics*. 2010; 42(5):441–447. [PubMed: 20418890]
- Treutlein J, Rietschel M. Genome-wide association studies of alcohol dependence and substance use disorders. *Current Psychiatry Reports*. 2011; 13(2):147–155. [PubMed: 21253885]
- Vanyukov MM, Kirisci L, Moss L, Tarter RE, Reynolds MD, Maher BS, Clark DB. Measurement of the risk for substance use disorders: Phenotypic and genetic analysis of an index of common liability. *Behavior Genetics*. 2009; 39(3):233–244. [PubMed: 19377872]
- Vanyukov MM, Kirisci L, Tarter RE, Simkevitz HF, Kirillova GP, Maher BS, Clark DB. Liability to substance use disorders: 2. A measurement approach. *Neuroscience & Biobehavioral Reviews*. 2003; 27(6):517–526. [PubMed: 14599433]
- Vanyukov MM, Tarter RE, Conway KP, Kirillova GP, Chandler RK, Daley DC. Risk and resistance perspectives in translation-oriented etiology research. *Translational Behavioral Medicine*. 2016; 1–11. [PubMed: 27012248]
- Vanyukov MM, Tarter RE, Kirisci L, Kirillova GP, Maher BS, Clark DB. Liability to substance use disorders: 1. common mechanisms and manifestations. *Neuroscience & Biobehavioral Reviews*. 2003; 27(6):507–515. [PubMed: 14599432]
- Williams SC. News feature: Genetic mutations you want. *Proceedings of the National Academy of Sciences of the United States of America*. 2016; 113(10):2554–2557. [PubMed: 26957571]
- Windle M. Resilience and development. Springer; 2002. Critical conceptual and measurement issues in the study of resilience; p. 161–176.
- Wolff PH. Ethnic differences in alcohol sensitivity. *Science (New York, N.Y.)*. 1972; 175(4020):449–450.
- Wolff PH. Vasomotor sensitivity to alcohol in diverse mongoloid populations. *American Journal of Human Genetics*. 1973; 25(2):193–199. [PubMed: 4689039]
- Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF, Gelernter J. Rare nonsynonymous variants in alpha-4 nicotinic acetylcholine receptor gene protect against nicotine dependence. *Biological Psychiatry*. 2011; 70(6):528–536. [PubMed: 21683344]
- Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. *Genetics*. 2005; 171(0016-6731; 2):813–823. [PubMed: 16020784]

Figure 1.
Liability, Risk and Resistance

Figure 2.
Ascertainment Schemes for Resistance Studies

Table 1
Power to Detect A Protective SNP or Resistance Environment Under Several Ascertainment Schemes

SNP Additive Variance	Environment R^2	Case Control		Random		Low Risk		High Risk		Low Environmental Risk		Extreme Environmental Risk	
		SNP	E	SNP	E	SNP	E	SNP	E	SNP	E	SNP	E
0.009	0.0225	0.263	0.536	0.181	0.332	0.267	0.64	0.334	0.545	0.276	0.51	0.859	0.994
0.017		0.442	0.507	0.285	0.317	0.401	0.607	0.482	0.521	0.449	0.487	0.979	0.998
0.026		0.561	0.510	0.415	0.295	0.553	0.627	0.627	0.576	0.598	0.514	0.999	0.992
0.009	0.04	0.253	0.768	0.199	0.562	0.278	0.874	0.329	0.760	0.255	0.747	0.866	1.000
0.017		0.412	0.764	0.278	0.534	0.385	0.869	0.448	0.760	0.435	0.774	0.987	1.000
0.026		0.597	0.778	0.412	0.534	0.564	0.857	0.643	0.759	0.63	0.76	0.999	1.000
0.009	0.0625	0.279	0.937	0.166	0.725	0.250	0.972	0.347	0.913	0.268	0.939	0.877	1.000
0.017		0.438	0.924	0.289	0.762	0.424	0.973	0.500	0.907	0.455	0.904	0.981	1.000
0.026		0.573	0.939	0.418	0.723	0.552	0.966	0.610	0.894	0.589	0.919	1.000	1.000

-resistance allele frequency = .3, heritability of outcome = .5, common environmental variance = .25