Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Title: Sensitivity Analysis of Self-Identified Race and Ethnicity (SIRE): Screening for Single Nucleotide Polymorphisms.

Authors:

Toni P. Miles

Stephen Hanson

LaCreis Kidd

Avonne Connor

Saeed A. Jortani

Affliation: University of Louisville, Louisville, KY.

Corresponding Author:

T. Miles, M.D., Ph.D.

Professor, Department of Family & Geriatric Medicine
501 E Broadway, Suite 240
School of Medicine
Louisville. KY 40202

Telephone: 502-852-5201
Email: tpmile01@louisville.edu
Fax: 502-852-0651

Word count: 3934

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Abstract – 150 words

Within the realm of medical care, Self-Identified Race and Ethnicity (SIRE) categories are promoted as an inexpensive tool to identify underlying genotypic diversity. Scientific opinion is divided about the adequacy of SIRE to serve this function. If genetic diversity can guide medical decisions, it is important to know the effectiveness of genetic screening *via* SIRE. Proper development of self-reported measures such as SIRE requires sensitivity and specificity studies. These types of formal evaluation are largely absent for SIRE. To begin this formal process, we estimate the sensitivity of SIRE in screening for variant Single Nucleotide Polymorphism (SNP). Our results indicate that the current use of SIRE is inadequate to screen for selected biotransformation related SNP in the N-Acetyl Transferase pathway. The widespread usage of SIRE to screen for genotypic diversity could promote erroneous assignment of patients to disease risk or therapeutic categories.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Introduction

Self-Identified Race and Ethnicity (SIRE) has been the topic of a number of conferences, editorial and opinion papers that suggest it is not an adequate pre-screen test for medically significant genotypic diversity (Love 2001, Collins FS 2004, Wolf SF 2005, Duster 2007). Despite this lack of consensus opinion, SIRE continues to be promoted for use as a tool to predict the presence of disease risk genes (Freeman et al 2006) or response to pharmacologic treatments (Wolf 2005). The scientific rationale for the use of SIRE as a genetic screening tool often rests on observed population-level differences in the proportion of variant alleles. In the development of tools for medical application, population-level proportions are not sufficient to justify this application. Development of a predictive marker to prevent disease occurrence or enhance treatment response requires the estimation of the risks associated with both accurate and inaccurate categorical assignment. In this paper, we have selected a SNP with the potential for clinical application. The following text provides prevalence rates of a SNP from four SIRE defined populations that were created to identify cancer related genes. To push this concept further, we then estimate the probability that SIRE *alone* predicts the presence of a variant SNP. Our results suggest that in a sample derived from populations across the globe, SIRE cannot satisfactorily identify the presence of variant SNP.

First do no harm is the guiding principle when contemplating the introduction of a new medical procedure. This directive commands us to develop a clear sense of the potential for harm that would accompany the use of SIRE as a genetic screener. Rather than a sole focus on the potential for positive effects, the translation of genetic-marker guided therapy into clinical practice also needs a clear assessment of the additional risks imposed with using SIRE to guide medical decisions. Such risks include the potential for harm associated with an incorrect

assignment of genetic constitution based on SIRE, the probability that a SIRE response will incorrectly assign disease risk, and the likelihood that a patient will be incorrectly assigned to a therapeutic regimen based on a SIRE response. To fully assess the utility of SIRE as a genetic marker, we need to compare the probability of benefit with the probability for causing harm.

SIRE is a self-reported measure and one obvious source of harm is the risk of misclassification. Although there have been formal discussions of the potential connection between human bio-geographic variation and categorization of race within medicine (Tishoff and Kidd 2004), a working definition for SIRE has never been subjected to formal evaluation. Scientific research involving the use of SIRE utilizes a categorization scheme that is largely defined by governmental agencies such as the U.S. Bureau of the Census (www.censusbureau.org). These categorizations may be of use in social science, but their validity in encompassing biological factors relevant for clinical evaluation is often unclear. To cite one example, The National Cancer Institute SNP500Cancer Database (http://www.nci.nih.gov) divides contributed samples into four SIRE categories – Caucasian, African / African American, Native American / Hispanic, and Pacific Rim. The website provides further detail on the country of origin for the samples but nothing concerning the logic used to assign individuals.

There are research methods that can be used to assess accuracy, validity, and consistency of SIRE as a surrogate for genetic constitution. These methods have been used to evaluate similar data. Family history, like SIRE, is a self-reported measure used as a surrogate for shared genetic constitution (http://www.cdc.gov). Family members share their genes, as well as their environment, lifestyles and habits. The clinical validity of self-reported *family history for cancer*

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

has been evaluated using an evidence-based review of precision and accuracy (Murff, Spigel, and Syngal 2004).

Precision reflects the reproducibility of the report. There are multiple factors that decrease the precision of a family history report, including poor patient-physician communication, increasing distance of relation to the affected relative, and lack of disease in the reporting patient. *Accuracy* captures the how well a test actually measures the value it intends to measure. A similar set of factors diminish the accuracy of self-reported cancer family history. The best accuracy is obtained for patients with active colon or breast cancer who identify first-degree relatives who have had colon or breast cancer (Murff, Spigel, and Syngal 2004). All other site-specific family cancer history reports lack precision and accuracy. The results of self-reported family cancer history studies indicate that the precision and accuracy of SIRE reports should be thoroughly evaluated prior to any implementation as a genetic screening tool.

Translating genetic knowledge from population-level statistics to individual-level risk assessment is another potential contributor to SIRE-induced clinical harm. The argument used to justify the use of SIRE as a genetic screening tool is based on population differences in the distribution of variant alleles. In medical epidemiologic research, it has been long understood that population distribution cannot be used as the indicator of individual risk. Within the setting of medical care, the important goal is to identify the probability that an individual patient has the genes of interest. One seeks to know, e.g., the likelihood that an individual man has a gene that substantially increases his risk of prostate cancer, or the chance that an individual patient has the pharmacogenetics constitution to derive benefit from a specific medication. The application of a screening tool is the appropriate mechanism to determine these odds but the underlying mechanics of this application have not been explored .

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Screening is a logical first-step process to distinguish between categories of genetically determined response. SIRE is being promoted as an appropriate screening tool within the realm of medical care. Clinically useful screening tests need to be robust enough to identify high risk individuals embedded in a non-selected sample. SIRE is conceptually designed to create a few broadly defined categories. This concept is in direct opposition to sampling strategies employed in most gene identity studies, which have always capitalized on the power of a highly select sample of related individuals. Lack of precision in the definition of categories, then, is a significant source of variation in estimating the sensitivity of SIRE as a screening tool to identify underlying SNP.

Screening has a recognized place in the ascertainment of medical history, assignment of clinical risk, and determination of treatment appropriateness. The practice of evidence-based medicine now *requires* a process of formal evaluation for all screening tools. When SIRE is associated with SNP, then the distinction is being made between persons having a wild-type SNP and those who have a variant SNP. In this report, we present an approach to evaluate the sensitivity of SIRE as a screener for the presence of selected variant SNP. To conduct this technical evaluation of SIRE, we use the nomenclature present on the publicly available National Cancer Institute SNP500Cancer Database. (http://www.nci.nih.gov.) The SNP500 Cancer database was not created to support definitive evaluation of SIRE as a screening tool, but it is ideal for this conceptual evaluation for a number of reasons. It is one of the publicly available datasets where SIRE and SNP data are available for this type of analyses. The database contains samples from persons across the globe. This sample composition more closely mimics the clinical environment for the application of SIRE guided medical treatment. The stability of genotyping technology has been assured for all samples. This technical report is designed to

present a strategy that can be used to design future formal testing of SIRE as a SNP screening tool.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Results

Although screening is a conceptually simple idea, the algebra underlying translation from biomarker to clinical tool is not. In this report, we compare the information content of three statistical approaches widely used in genetics, medicine and social science to detect the presence of population-level differences - Prevalence; Screening measures, and Linear Regression. Prevalence is a simple population-based calculation of percentage. Screening tests measurement evaluates performance and quantifies the probability of misclassification errors. Linear regression (LR) is also a prediction statistic that further measures the association between a marker and a related trait. The widespread usage of LR by the scientific community makes it an intuitively useful way of illustrating the association between SIRE and SNPs. Each measure is mathematically consistent, but what do they tell us about the use of SIRE to guided SNP allele detection?

[Table 1 about here.]

Prevalence is a straight forward method for identifying population differences in the proportion of individuals with a specific SNP as shown in Table 1. Each row contains percent distribution of a single allele in the three NAT genes for a SIRE-defined group. Each NAT gene has three allelic forms determined by the combination of wt for wild type and var for variant. For the NAT1 rs1057126, the AA allele is least common. The percent of persons in each SIRE category with this allele ranges from 8.2% (Caucasians) to 25.0% (African/African American). There is no known risk associated with any of these genes. They were selected solely for calculations in this report.

What kinds of questions are we asking when comparisons are made? With the percentages in the bottom row, we can ask – 'Does any one group have a greater than expected

percent of persons with the variant SNP?' This implies a comparison between percents observed in the combined sample (bottom row) and a single SIRE category. Statistical comparison of Caucasians and African/African American categories yield a result indicates statistical differences. A similar pattern of allelic difference can be observed when comparing the Pacific Rim category (4.4%) with Caucasians (19.8%) for RS1801280. Again, there is a statistically significant between group differences in prevalence of the NAT1 rs1801280 CC allele. This type of statistical difference appears to suggest the utility of SIRE as a screening tool for at least two NAT1 alleles.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

----Table 2 about here----

Does a statistically significant difference in population prevalence translate into a clinically useful screening question? To be clinically useful, a screening test must maximize identification of persons have an elevated risk ('true positive') and minimize misclassification of who do not have an elevated risk ('false positive'). In addition, medical treatment or lack thereof is not without its own inherent risks. Therefore, it is also important to have a screening test that accurately identifies persons without risk ('true negative'). Finally, it is important that the test results do not erroneously deny treatment to a person who might benefit ('false negative'). In Table 2, we present sensitivity estimates (i.e. Is the SNP present?) and specificity (i.e. Is the SNP absent?).

The performance of self-reported racial identity as a screening tool for the rs1057126 AA allele and the rs1801280 CC allele appears to be much less reliable than prevalence statistics would suggest. When a person self identifies as African / African American, there is only a 37% probability that the AA allele is present. Similarly, there is only a 34% probability that the CC

allele is present when a person self-identifies as a Caucasian. Each of these is a low probability events are well below the 85 percent standard used to judge clinical utility for screening tests (www.cochrane.org; (http://www.healthit.ahrq.gov). Clearly in the case of these two alleles, a positive response to an individual SIRE category is not a clinically useful screening tool for the presence of this SNP.

Alternatively, what if the critical decision rested on the probability that a trait was absent? Specificity statistics are higher for both responses (76%, respectively). If the medical decision involved application of a treatment with potential toxicity, then increasing the certainty associated with a negative test result would be important and prevent unnecessary exposure to a toxic treatment. For each gene, not being a member of the African / African-American category (AA allele) or the Caucasian category (CC allele) would only be accurate 76% of the time.

---Table 3 here---

Does SIRE predict the presence of a single variant SNP? A non-clinical approach to this question could also involve the use of a regression model. Regression models are a statistical tool that is most familiar to social scientists evaluating multivariable effects. Table 3 shows results from a modeling process. There are SIRE group separate models shown in each row, i.e. each separate response is evaluated for its likelihood to predict the presence of the variant allele. The models were computed using either the *sensitivity* estimate (results shown) or the *prevalence* (models not shown). None of the SIRE variables achieved statistical significance when prevalence was used. Is there added information for study design that can be gained from this statistical approach? In each of the models the t-statistic results suggests a large amount of unexplained variance in the association between SIRE and the presence of the AA (Table 3A results) or CC (Table 3B results). Biopsychosocial models of health and disease incorporate

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

concepts that combine allelic data with other factors to compare and contrast biological and sociological influences on disease risk. More work is needed to develop appropriate statistical tools. These statistical models suggest the size of the gap between SNP and the ability of SIRE to predict the presence of variant SNP.

Discussion

This is a technical report describing the application of sensitivity estimation to SIRE when it is used as a screening tool for the presence of variant SNP. SIRE has been promoted as a screening tool for the assignment of future disease risk or the selection of medications in clinical practice. An assumption underlying this practice is that SIRE accurately indicates the presence of clinically important SNP. By formulating SIRE as a screening tool, we have been able to begin the process of articulating a formal test of SIRE. It is our hope that with this approach, the discussion of race and medicine can be taken out of the realm of opinion. Treatment decision-making has significant consequences for patients. Physicians are more comfortable using tools that are evidence-based. Social scientific research has provided evidence that SIRE is a complex measure capturing genetic and environmental factors (Hayward et al 2000).

Is there a place for SIRE-guided medical treatment? This analysis would suggest that SIRE cannot be used as it is currently defined to screen for variant SNP in the N-Acetyl Transferase system. These findings need to be replicated in other treatment decisions and for other SNP. Our analysis indicates that SIRE performs well below the recommended 85 percent minimum sensitivity standard when applied as a screening test to indicate an underlying variant SNP. In a white paper (2003), The British National Health Service evaluated the future application of genetic testing in clinical medicine but did not recommend use of SIRE as a screening tool for clinically important SNP. SIRE lacks adequate sensitivity as a screening tool

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

and its application in clinical practice has the potential for erroneously assuming the presence of a variant SNP when it is not present.

Statistical Sample and Methods

SIRE definition and Sampling Frame

SNP500CANCER Dataset

The Cancer Genome Anatomy Project: SNP500Cancer Database was created to re-sequence 102 anonymous reference samples as well as the Human Diversity Panel, consisting of 280 persons from a variety of populations world-wide from the Coriell Biorepository (Camden, NJ) (Cann et al 2002, Packer et al 2006). The Project seeks to validate known or newly discovered genetic variants of potential importance to studies of cancer and other diseases. Initial inclusion of a SNP in the database is based upon review of the literature with input from intramural and extramural investigators. There is currently heavy weighting towards non-synonymous SNP (i.e., those that result in an amino acid substitution). For purposes of this sensitivity analysis, the least common form is designated as a *variant*. For the NAT1 rs1057126 SNP, the AA genotype was designated as the variant and is present among 18.0 % of the combined sample (Case Plus Human Diversity Panel). A similar decision process was applied to each of the other SNP categories. For rs 15561 the AA genotype was designated as variant (22.2%). For the rs 18011280 SNP, the TT genotype was designated as variant (9.4 %). The SIRE sensitivity calculation, then estimates the probability that a patient report of specific category within SIRE indicates the presence of a low frequency *variant* SNP.

Sensitivity is a ratio comparing the performance of a screening test with a different so-called *gold standard* indicator (Loong 2003). To develop the analysis shown in this report, we started with the standard 2x2 Table.

[Insert Table 4 about here]

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

The sensitivity of SIRE is compared with the gold standard – Molecular confirmation of a SNP in Table 4. **A** represents all persons who report group membership and have the SNP of interest. **A** divided by **A + C** indicates the sensitivity of a self-report. There are other measures of test performance that can be obtained from this approach. *False positive* (**A / A+B)** directly calculates the probability that a self-report is incorrect. *Likelihood ratio* (LR) is another evidence-based ratio that is used to compare test sensitivity and false positive rates. The interpretation of LR balances the tension between a test that is highly sensitive but also generates large rates of false positives. In the assignment of treatment, each individual circumstance requires determining whether greater harm comes from missing a person with the trait of interest or misclassifying a person who does not have the trait.

The SNP500Cancer Database is a broad sampling of SNP standardized for future cancer risk research. It was not designed for screening test development. The rationale for using this database for sensitivity analysis is based on the idea that it contains consensus-derived genetic markers derived from geographically dispersed populations that have been assigned to broad categories of SIRE. One future technical issue to resolve is the development of a more appropriate dataset.

Table 1 shows the percent or *prevalence* of variant SNP for each SIRE group across the three SNP categories used in this analysis. The analysis shown in Table 2 could have been accomplished using a variety of statistical methods. We chose this approach because it is the preferred application for medical decision purposes (Loong 2003). Clinical decision-making requires that we are able to address the likelihood that our test leads to the false assumption that a person has a SNP when, in fact, it is not present. *Sensitivity* and *Specificity* are companion statistics. In this technical report, we are solely focused on examining the technical issues

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

surrounding the ability of SIRE to indicate the presence of a variant SNP. In a future studies advocating clinical application for SIRE as a screening tool, both its sensitivity and specificity should be presented.

Sources of variation in Sensitivity

The calculation of sensitivity is generally a population-based statistic which capitalizes on the concepts underlying the classical Chi-X^2 statistic which is a ratio comparing the observed frequency of an event with the expected frequency based on random occurrence. The proportion of individuals in the specific sample with the gold standard marker influences the estimated sensitivity of a screening test. For example, if a particular variant SNP is present in 100 percent of members of one SIRE category, then SIRE would accurately reflect the underlying presence of the SNP. Screening tests are known to have falsely elevated sensitivity rates when the prevalence of a characteristic is low in the population. The estimated test sensitivity can be corrected for chance agreement using a method developed by Coughlin and Pickle (1992). Creating an experimental sample for the assessment of test performance is another approach to distinguishing a main effect of SIRE and interaction between SIRE and SNP variants (Goldman and Flanders 2007).

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Conclusion

SIRE appears to be a poor screening tool for variant N-acetyl transferase SNPs, well below the recommended level of 85% specificity. Physicians generally ought not recommend screening tests where the risk of a false positive is high and potential harms could come from an erroneous screening result. Nor should SIRE be utilized as a screening test for clinically relevant variant SNPs without good evidence of significant sensitivity, since significant harms can come from

treating a patient based on an erroneous assumption of a variant SNP. Further research should be done to more carefully test this specificity for other relevant SNPs, but our analysis suggests that the current evidence gives little reason, in most cases, to correlate SIRE with any given SNP.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

References

Burke W, Atkins D, Gwinn M, Guttmacher A, Haddow J, Lau J, Palomaki G, Press N, Richards S, Wideroff L, Weisner G. Genetic test evaluation: information needs of clinicians, policy makers, and the public. 2002 Am J Epidemiology 156 (4): 311 – 318.

Cann, H et al. A Human Genome Diversity Cell Line Panel. Science. 2002. 296(5566): 261-262.

Collins FS. 2004 What we do and don't know about race….[Commentary] Nature Genetics 36(11): S13 – S15.

Coughlin S, Pickle L. Sensitivity and specificity-like measures of the validity of a diagnostic test that are corrected for chance agreement. Epidemiology 1992. 3: 178 – 181.

Duster T. Medicalisation of race. [Commentary] Lancet 2007. 369: 702 – 704

Freeman ML, Haiman CA, Patterson N et al. Admixture mapping identifies 8q24 as a prostate cancer risk in African-American men. PNAS 2006. 103(38): 14068 – 14073.

Goldman M, Flanders WD. Study design options in evaluating gene-environment interaction: practical considerations for a planned case-control study of pediatric leukemia. Pediatric Blood Cancer 2007. 48: 375 – 379.

Goldstein DB, Tate SK, Sisodiya SM. 2003. Pharmacogenetics goes genomic. Nature Genetics: Reviews 4: 937 – 947.

Hayward MD, Crimmons EM, Miles TP**,** Yang Y. The significance of socioeconomic status in explaining the race gap in chronic health conditions. American Sociologic Reviews. 65:910-30, 2000.

Kim H, Neubert J, Miguel A, Xu K, Krishnaraju R, Iadavola M, Goldman D, Dionne R. Genetic influence on variability in human acute experimental pain sensitivity associated with gender, ethnicity, and psychological temperment. Pain 109:488 – 496.

Loong, T. 2003. Understanding sensitivity and specificity with the right side of the brain. British Medical Journal 327: 716-719.

Our inheritance, our future. Realising the potential genetics in the National Health Service. Presentation to Parliament by the Secretary of State for Health. June 2003. Published by The Stationary Office. Available at www.tso.co.uk/bookshop.

Packer BR, Yeager M, Burdett L et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. Nucleic Acids Res. 2006. 34: D617-D621.

Wolf SF. 2005 Are we making progress in the debate over racial and ethnic categories in biomedical research? [Meeting report] Nature Genetics 37(8): 789 – 790.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Table 1: Prevalence Variant N-Acetyl Transferase (NAT) SNP allele within Self-Identified Racial and Ethnic (SIRE) Category: Percent allele type, total number within SIRE category.

	SIRE Category				
NAT SNP	**African / African American N (%)**	**Caucasian N (%)**	**Hispanic / Native American N (%)**	**Pacific Rim N (%)**	**Combined Sample N (%)**
Genotype / Allelic Distribution					
rs1057126					
- wt / wt	25 (25.0)	63 (64.9)	19 (26.8)	42 (37.2)	149 (39.1)
- wt / var	52 (52.0)	26 (26.8)	37 (52.1)	47 (41.6)	162 (42.5)
- var / var	23 (23.0)	8 (08.2)	15 (21.1)	24 (21.2)	70 (18.4)
- Total column	100	97	71	113	381
- % SIRE	26.1	25.4	18.6	29.7%	
rs15561					
- wt / wt	21 (21.4)	52 (55.9)	18 (25.3)	29 (25.9)	120 (32.1)
- wt / var	49 (50.0)	29 (31.2)	37 (52.1)	54 (48.2)	169 (45.2)
- var / var	28 (28.6)	12 (12.9)	16 (22.5)	29 (25.9)	85 (22.7)
- Total column	98	93	71	112	374
- % SIRE	26.2	24.9	19.0	29.9	
rs1801280					
- wt / wt	44 (50.0)	30 (31.2)	31 (43.1)	78 (70.3)	183 (49.9)
- wt / var	38 (43.2)	47 (49.0)	34 (47.2)	29 (26.1)	148 (40.3)
- var / var	6 (06.8)	19 (19.8)	7 (09.7)	4 (03.6)	36 (09.8)
- Total column	88	96	72	111	367
- % SIRE	24.0	26.2	19.6	30.2	

Note: Identification based on categories in the SNP500Cancer Database. Denominator = Cancer Control Sample plus Population Diversity Panel. See http://www.nci.nih.gov. wt / wt = Wild type allele; wt / var = Heterozygous allele; var/var = Homozygous allele. Prevalence = [Number with designated allele divided by Number with other alleles]. These are calculated within each SIRE category. Denominator is also shown for each SIRE category. Percents within columns do not add to 100 due to rounding.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Table 2: Screening test performance: SIRE performance as a indictor of variant NAT allele (var).

	Variant Present *			Variant Absent *		
	Not Caucasian (NC)#	**Caucasian (C) @**	**Sensitivity (%)**	**Caucasian (C) @**	**Not Caucasian (NC)#**	**Specificity (%)**
Genotype						
rs1057126 - % SIRE	69.7	35.0		64.9	30.3	
- Test Parameter			85.3			42.3
rs15561 - % SIRE	74.2	44.1		55.9	26.8	
- Test Parameter			70.5			17.0
rs1801280 - % SIRE	64.1	68.8		31.2	35.9	
- Test Parameter			47.2			16.3

Note: Identification based on categories in the SNP500Cancer Database. See http://www.nci.nhi.gov. @Taken from Table 1. # calculated by combining groups. Presence of variant determined by laboratory testing.

- **Sensitivity of SIRE response** = [$VAR^{+}NC$ / ($VAR^{+}NC$ plus $VAR^{+}C$)];
- **Specificity of SIRE response** = [VAR^{-} C / (VAR^{-} C plus VAR^{-} NC)].

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Table 3: Does SIRE predict the presence of a single variant SNP? Trial with 2 NAT1 SNPs.

A. Dependent Variable: NAT1 Sensitivity 1057126

Model		Standardized Coefficients B	t	Sig.	95% Confidence Interval for B	
1	(Constant)		8.668	.013	20.146	59.854
	Caucasians / Others	-.854	-2.319	.146	-61.109	18.309
2	(Constant)		3.999	.057	-2.679	73.212
	African / Others	-.098	-.140	.902	-78.358	73.425
3	(Constant)		3.847	.061	-3.946	70.613
	Hispanic / Others	.210	.304	.790	-69.293	79.826
4	(Constant)		4.894	.039	2.853	44.347
	Pacific Rim / Others	.088	.124	.912	-40.293	42.693

B. Dependent Variable: NAT1 Sensitivity 1801280

Model		Standardized Coefficients B	t	Sig.	95% Confidence Interval for B	
1	(Constant)		5.271	.034	3.227	31.907
	Caucasians / Others	.966	5.286	.034	6.553	63.913
2	(Constant)		2.179	.161	-27.066	82.599
	African / Others	-.153	-.218	.847	-115.232	104.099
3	(Constant)		2.302	.148	-24.949	82.349
	Hispanic / Others	-.255	-.373	.745	-116.597	97.997
4	(Constant)		2.942	.099	-14.560	77.493

Pacific Rim / Others	-.558	-.952	.442	-112.420	71.686

Note: Each row represents a separate model with SIRE recoded as a dummy variable. To create the variable, the group of interest is coded as 1 and all others are coded as 0. Using a mixed group sample, the model reflects the probability that a specific SIRE reflects the presence of a single SNP. Models use SNP *sensitivity* estimate from Table 2.

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Nature Precedings : hdl:10101/npre.2008.1524.1 : Posted 17 Jan 2008

Table 4: Sample 2x2 table for comparison of SNP data with SIRE

	Molecular Confirmation of SNP	
SIRE Membership	Yes	No
Yes	A	B
No	C	D
	A + C	B + D