A weighted accumulation test for associating rare genetic variation with quantitative phenotypes

Chuanhua Xing, Duke University
Glen Satten, Emory University
Andrew S. Allen, Duke University

Journal Title: BMC Proceedings
Volume: Volume 5, Number SUPPL. 9
Publisher: BMC (part of Springer Nature) | 2011-12-01, Pages S6-S6
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1186/1753-6561-5-S9-S6
Permanent URL: https://pid.emory.edu/ark:/25593/v1pkb

Final published version: http://dx.doi.org/10.1186/1753-6561-5-S9-S6

Copyright information:
© 2011 Xing et al; licensee BioMed Central Ltd. This is an Open Access work distributed under the terms of the Creative Commons Attribution 2.0 Generic License (http://creativecommons.org/licenses/by/2.0/).

Accessed April 9, 2022 9:01 AM EDT
A weighted accumulation test for associating rare genetic variation with quantitative phenotypes

Chuanhua Xing1,2, Glen A Satten3, Andrew S Allen1,2*

From Genetic Analysis Workshop 17
Boston, MA, USA. 13-16 October 2010

Abstract
Currently there is a great deal of interest in developing methods for testing the role that rare variation plays in disease development. Here we propose a weighted association test that accumulates genetic variation across a signaling pathway. We evaluate our approach by analyzing simulated phenotype data from an exome sequencing study of 697 unrelated individuals from the Genetic Analysis Workshop 17 (GAW17) data set. Although our weighted approach identifies several interesting pathways associated with phenotype Q1, so does an alternative unweighted accumulation approach. Such a result is not unexpected because there is no systematic relationship between the allele frequency of a variant and its effect on phenotype in the GAW17 simulation model.

Background
Next-generation sequencing technology allows for the characterization of virtually all of an individual’s genetic variation. Genome-wide association studies (GWAS) have successfully detected hundreds of disease-susceptible loci that harbor common variants. However, the common variants identified so far have explained only a small portion of the genetic risk of most of the diseases studied. Some researchers have argued that this is due in part to rare variants having a larger role in disease etiology than previously suspected. Some recent studies support this reasoning [1-11].

Several approaches have been proposed to analyze rare variants for association with disease. The cohort allelic sums test (CAST) is a simple grouping method that compares the number of affected and unaffected individuals who have variants [4,12,13]. Li and Leal [14] introduced the combined multivariate and collapsing (CMC) method. In CMC, markers in a gene or other unit of analysis are collapsed into one or more indicator variables based on some criteria (e.g., the presence of at least one nonsynonymous mutation within a gene). Because many criteria could be used to define several such indicator variables, Li and Leal proposed a multivariate test using Hotelling’s T. Morris and Zeggini [15] proposed an accumulation approach that regresses phenotype on a genetic score, defined as the proportion of sites within the gene or pathway that harbor mutations. Price et al. [16] proposed a variable-threshold approach by finding the maximum z-score across all possible values for threshold $T$, assuming that the variants having minor allele frequency under this threshold are more likely to be functional. Madsen and Browning [12] proposed a weighted sum statistic. In this approach, single-nucleotide polymorphisms (SNPs), which are rare among the control subjects, are up-weighted with the goal of giving rare, highly penetrant mutations greater influence on the test statistic.

Here, we propose a weighted group-wise association test that accumulates genetic variation across a signaling pathway. We extend the basic idea behind the Madsen and Browning [12] weighting scheme to quantitative traits. Specifically, genetic markers that are rare among individuals in the center of the phenotypic distribution (those with nonextreme phenotypes) are up-weighted to reflect the assumption that rare genetic variation will tend to have a larger effect on a phenotype. Because the weight is a function of phenotype, as in Madsen and Browning’s study [12], permutation is used to assess statistical significance. In the next section, we detail our approach and

* Correspondence: asallen@duke.edu
1Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Road, Suite 1102 Hock Plaza, Box 2721, Durham, NC 27710, USA
Full list of author information is available at the end of the article

© 2011 Xing et al; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
highlight its application to phenotype Q1 of the Genetic Analysis Workshop 17 (GAW17) data.

Methods

Suppose that the number of SNPs in a genetic unit (a signaling pathway or a gene) is \( P \). Let \( Y_i \), \( i = 1, 2, \ldots, N \), be the phenotype for individual \( i \). We define \( I_{ij} \) for individual \( i \) to be the number of minor alleles at SNP \( j \). Let \( X_i \) be a genetic summary score, which we define as:

\[
X_i = \sum_{j=1}^{P} w_j I_{ij},
\]

where \( w_j \) is a weight that is applied to the \( j \)th SNP.

We evaluate two weighting schemes. In the first scheme, \( w_j \) is taken to be 1 for all \( j \). Thus rare and common SNPs are treated in the same way, and \( X_i \) is the simple sum of the number of minor alleles in the gene or pathway. This approach is similar to that of Morris and Zeggini’s method [15]; they defined a genetic score by the proportion of sites within the gene or pathway that harbored mutations. Because this scheme does not differentially weight SNPs, we refer to it as unweighted.

In the second scheme, we calculate the frequency of nonreference mutations among nonextreme individuals at position \( j \) as:

\[
\rho_j = \frac{\sum_{i=1}^{N} I_{ij} \delta(Y_i) + 1}{2 \sum_{i=1}^{N} \delta(Y_i) + 2},
\]

where \( \delta(Y_i) = 1 \) when \( Y_i \) is within one standard deviation of the mean and \( \delta(Y_i) = 0 \) otherwise. Adding a 1 to the numerator and a 2 to the denominator ensures that the frequency \( \rho_j \) is nonzero, so that the weight used in the second scheme,

\[
w_j = \frac{1}{\left[2N \rho_j (1 - \rho_j)\right]^{1/2}},
\]

remains finite [12]. Note that, with this weight, SNPs that are rare among those individuals whose phenotypes lie within the center of the phenotype distribution will be up-weighted and will have a larger role in the genetic summary score \( X_i \). We refer to this scheme simply as weighted.

For both approaches, once we have defined the genetic score \( X_i \), we assume that it is related to \( Y_i \) through the linear model:

\[
Y_i - \beta_0 + \beta_1 X_i + \epsilon,
\]

where \( \epsilon \) is an unknown error term. A Wald statistic, \( \hat{\beta}_1 / [\text{Var}(\hat{\beta}_1)]^{1/2} \), is computed, with the variance estimated using a sandwich estimator [17]. Because the weighted approach uses phenotypic information in defining the weight, we use permutation to assess statistical significance. We note that, in this case, the weight is recomputed for each permuted data set. We use 1 million permutations throughout.

We evaluate our approach using the simulated GAW17 data set. These data are described in detail elsewhere [18]. Although all 200 replicates are analyzed, for illustration purposes, we present results concerning replicate 1 in greater detail. Our analyses focus on one phenotype: quantitative trait Q1. Even though we had access to the answers for the underlying simulation model, our approach, including the characterization of signaling pathways, was developed without reference to these answers.

We characterize gene sets using information from two databases. The first, PharmGKB (https://www.pharmgkb.org/) [19], provides information on 1,400 signaling pathways. Unfortunately, the genes in the GAW17 data set are not well represented in PharmGKB, with only 713 out of 3,205 genes sequenced in the GAW17 data being included in 821 of these pathways. To compensate for this low coverage, we also classify genes by biological process from the Gene Ontology (GO) database (http://www.geneontology.org/) [19], provides information on 1,400 signaling pathways. Unfortunately, the genes in the GAW17 data set are not well represented in PharmGKB, with only 713 out of 3,205 genes sequenced in the GAW17 data being included in 821 of these pathways. To compensate for this low coverage, we also classify genes by biological process from the Gene Ontology (GO) database (http://www.geneontology.org/) [19]. The results of these analyses, applied to replicate 1, can be found in Tables 1 and 2. Pathways (processes) with a dagger are significant using the weighted approach, and pathways (processes) with an asterisk are significant using the unweighted approach. Table 1 presents those PharmGKB signaling pathways that were found to be significant by one of the two (weighted or unweighted) approaches. It is immediately apparent from Table 1 that trait Q1 seems to be related to vascular endothelial growth factor (VEGF). This, of course, is comforting, given that the simulation...
Table 1 Significant PharmGKB signaling pathways

| PharmGKB ID   | Pathway name                                      | Permutation p-value (weighted approach) | Permutation p-value (unweighted approach) |
|---------------|--------------------------------------------------|-----------------------------------------|------------------------------------------|
| PA1647135821* | Actions of nitric oxide in the heart             | 6.0 × 10^{-6}                          | <1.0 × 10^{-6}                           |
| PA164713652*  | VEGF hypoxia and angiogenesis                    | <1.0 × 10^{-6}                         | 6.0 × 10^{-6}                           |
| PA164728105*  | Signaling events mediated by VEGFR1 and VEGFR2  | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| PA164728138†  | S1P3 pathway                                     | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| PA164713890†  | Neurophil interactions with VEGF and VEGFR       | <1.0 × 10^{-6}                         | 1.7 × 10^{-5}                           |
| PA164714260†  | VEGF binds to VEGFR leading to receptor dimerization | <1.0 × 10^{-6}                         | 3.3 × 10^{-5}                           |
| PA164728144†  | VEGFR1-specific signals                          | <1.0 × 10^{-6}                         | 1.70 × 10^{-4}                          |
| PA164728199†  | Integrins in angiogenesis                        | 9.0 × 10^{-6}                          | (a)                                     |
| PA164728205*  | S1P1 pathway                                     | 8.0 × 10^{-6}                          | 1.1 × 10^{-5}                           |
| PA164728223†  | HIF-1-alpha transcription factor network         | <1.0 × 10^{-6}                         | 1.92 × 10^{-4}                          |
| PA164728227†  | Glypican 1 network                               | 1.0 × 10^{-6}                          | <1.0 × 10^{-6}                           |
| PA2032†       | VEGF pathway                                     | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |

† Pathways (processes) that are significant using the weighted approach.
* Pathways (processes) that are significant using the unweighted approach.

Table 2 Significant GO processes

| GO term               | Description                                      | Permutation p-value (weighted approach) | Permutation p-value (unweighted approach) |
|-----------------------|--------------------------------------------------|-----------------------------------------|------------------------------------------|
| GO:0000186†           | Activation of MAP kinase activity, especially during sporulation | <1.0 × 10^{-6}                          | 4.0 × 10^{-7}                           |
| GO:0001569*           | Branching involved in blood vessel morphogenesis | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0001666†           | Response to lowered oxygen tension               | <1.0 × 10^{-6}                         | 1.0 × 10^{-6}                           |
| GO:0001938†           | Up-regulation of endothelial cell proliferation  | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0002040†           | Sprouting angiogenesis                           | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0006355†           | Regulation of cellular transcription, DNA-dependent | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0006916*           | Anti-apoptosis                                   | 7.5 × 10^{-5}                          | 1.0 × 10^{-6}                           |
| GO:0006940†           | Any process that modulates the frequency, rate, or extent of smooth muscle contraction | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0006952†           | Defense/immunity protein activity                | 8.0 × 10^{-6}                          | 1.9 × 10^{-5}                           |
| GO:0007169†           | Transmembrane receptor protein tyrosine kinase signaling pathway | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0008152†           | Metabolic process                                | 3.0 × 10^{-6}                          | 2.2 × 10^{-5}                           |
| GO:0010595†           | Up-regulation of endothelial cell migration      | <1.0 × 10^{-6}                         | 7.0 × 10^{-6}                           |
| GO:0030097†           | Blood cell formation                             | <1.0 × 10^{-6}                         | (a)                                     |
| *GO:0030522†          | Intracellular receptor-mediated signaling pathway | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0030499†           | Up-regulation of vascular endothelial growth factor (VEGF) receptor signaling pathway | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0043129†           | Surfactant homeostasis                           | <1.0 × 10^{-6}                         | 1.1 × 10^{-5}                           |
| GO:0045446†           | Endothelial cell differentiation                 | <1.0 × 10^{-6}                         | (a)                                     |
| GO:0045745*           | Positive regulation of G-protein coupled receptor protein signaling pathway | 6.425 × 10^{-2}                       | 1.2 × 10^{-5}                           |
| GO:0048286†           | Lung alveolus development                        | <1.0 × 10^{-6}                         | 5.0 × 10^{-6}                           |
| GO:0048661†           | Up-regulation of smooth muscle cell proliferation | <1.0 × 10^{-6}                         | <1.0 × 10^{-6}                           |
| GO:0050927†           | Up-regulation of positive chemotaxis             | <1.0 × 10^{-6}                         | (a)                                     |
| GO:0051894†           | Up-regulation of focal adhesion formation        | <1.0 × 10^{-6}                         | (a)                                     |
| GO:0055074†           | Regulation of calcium ion concentration          | <1.0 × 10^{-6}                         | (a)                                     |

† Pathways (processes) that are significant using the weighted approach.
* Pathways (processes) that are significant using the unweighted approach.
* Once 100 permuted data sets were found to have a larger Wald statistic for a given process than that observed in the original (unpermuted) data set, that process was deemed nonsignificant and further permutations were not performed.
was designed so that genes affecting Q1 came primarily from this pathway.

Table 2 presents GO biological processes that were found to be significant by one of the two (weighted or unweighted) approaches. Although VEGF is clearly implicated through one of the significant GO processes, the overall importance of VEGF is far less clear. This does not suggest that the information encoded in the GO database is somehow inferior to that represented by PharmGKB; it suggests only that the organization of PharmGKB makes the involvement of VEGF more transparent in this particular analysis. Results from the analysis of the other replicates are entirely similar. Almost all of the 200 replicates clearly implicate the VEGF pathway as influencing trait Q1. For example, using the weighted approach, PharmGKB pathway PA2032 was found to be significant in all 200 replicates, whereas GO process GO:0030949 was significant in 195 of 200 replicates. The unweighted approach performed even better, with all 200 replicates finding both PA2032 and GO:0030949 significant.

We informally explored which genes were important in these significant pathways and processes by enumerating the number of times a given gene was present in a significant pathway or process. The genes occurring in the biological processes from the GO data set are compared with the ones in the signaling pathways from the PharmGKB data set. Table 3 presents the top 10 genes with the most representation in the list of significant pathways or processes. From this table we can see that, although the GO and PharmGKB approaches may appear different at the pathway or process level, they seem to identify similar structure at the gene level. The genes VEGFA, FLT1, KDR, HIF1A, and ARNT are consistently represented both in the significant PharmGKB pathways and the significant GO processes. A comparison of the results in Table 3 with those in Table 4 suggests that the unweighted analysis also gives similar results.

**Discussion**

We presented two tests (weighted and unweighted) that accumulate genetic variation across a signaling pathway or biological process. In the analyses presented here, we found that the unweighted approach worked as well as, or better than, the weighted approach. We believe that this is strictly due to the structure of this particular simulation, in which the effect sizes of causal SNPs show no trend with the frequency of the causal variant. In situations where rarer SNPs are, in fact, more highly penetrant, we would expect a weighted approach to be more powerful.

In the analyses presented here, an accumulation approach seemed to work well. However, we must offer two important caveats. First, when moving from a gene-based to a pathway-based approach, the power of the approach becomes increasingly dependent on the state of existing biological knowledge and its representation in databases such as PharmGKB and GO. Even though we constructed pathways and biological processes without considering the true simulation model, our results are bound to be an overly optimistic representation of the power of a pathway-based approach. After all, the GAW17 simulation was constructed by accessing the same biological knowledge (although perhaps not the same databases) that we used to construct our pathways. Second, we computed a genetic score by simply summing the number of mutations in a gene or pathway and ignoring the directionality of the effect. This approach will be powerful when mutations lead to a shift in the phenotype in only one direction (as in the GAW17 simulation). However, it is likely that some mutations could lead to higher values of a phenotype and that other mutations could lead to lower values. This is possible even within a gene and becomes even more likely when considering a collection of genes, such as in a signaling pathway.

### Table 3

**Ten most frequent genes in significant PharmGKB pathways and GO processes using the weighted approach**

| PharmGKB | GO | Number of nonsynonomous SNPs | Number of nonsynonomous SNPs |
|----------|----|-----------------------------|-----------------------------|
| VEGFA    | KDR | 2                           | 11                          |
| FLT1     | SRC | 20                          | 1                           |
| HSP90AA1 | FLT1| 9                           | 2                           |
| SRC      | VEGFA| 1                          | 20                          |
| KDR      | SRC | 11                          | 1                           |
| SRC      | HSP90AA1| 2                         | 9                           |
| PRKCA    | KDR | 6                           | 9                           |
| HIF1A    | SRC | 5                           | 2                           |
| HIF1A    | SRC | 4                           | 6                           |
| SRC      | SRC | 3                           | 4                           |
| ARNT     | SRC | 1                           | 1                           |

Bold denotes genes that were found using both databases.
Table 4 Ten most frequent genes in significant PharmGKB pathways and GO processes using the unweighted approach

| PharmGKB | Number of nonsynonomous SNPs | GO | Number of nonsynonomous SNPs |
|----------|-----------------------------|----|-----------------------------|
| VEGFA    | 2                           | FLT1| 20                          |
| FLT1     | 20                          | VEGFA| 2                          |
| SRC      | 1                           | KDR | 11                          |
| HSP90AA1 | 9                           | KIT | 5                           |
| HIF1A    | 6                           | ARNT| 9                           |
| PRKCA    | 2                           | PTK2| 4                           |
| KDR      | 11                          | PDGFB| 3                          |
| SHC1     | 3                           | SHH | 4                           |
| PTK2     | 5                           | EGF R| 4                          |
| ARNT     | 9                           | ROR2| 2                           |

Bold denotes genes that were found using both databases.

Acknowledgments

CX and ASA acknowledge support from the National Institutes of Health (NIH) through National Institute of Mental Health (NIMH) grant R01 MH084680. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The Genetic Analysis Workshop is supported by NIH grant R01 GM031575. This article has been published as part of BMC Proceedings Volume 5 Supplement 9, 2011: Genetic Analysis Workshop 17. The full contents of the supplement are available online at http://www.biomedcentral.com/1753-6561/5?issue=S9.

Author details
1Department of Biostatistics and Bioinformatics, Duke University, 2424 Erwin Road, Suite 1102 Hock Plaza, Box 2721, Durham, NC 27710, USA. 2Duke Clinical Research Institute, Duke University, 2400 Pratt Street, Durham, NC, 27705-3976, USA. 3Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA. 4Road, Suite 1102 Hock Plaza, Box 2721, Durham, NC 27710, USA. 2Duke Clinical Research Institute, Duke University, 2400 Pratt Street, Durham, NC, 27705-3976, USA. 5Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA. 6Road, Atlanta, GA 30333, USA.

Authors’ contributions
CX, GAS, and ASA designed the study. CX implemented the methods and conducted the analyses. CX, GAS, and ASA wrote the manuscript.

Competing interests
The authors declare that there are no competing interests.

Published: 29 November 2011

References
1. Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, Doelle H, Escoy B, Kryukov G, Schmidt S: Medical sequencing at the extremes of human body mass. Am J Hum Genet 2007, 80:779-791.
2. Azzopardi D, Dalloso AR, Blasón K, Hendrickson BC, Jones N, Rawstorne E, Colley J, Moskvin A, Frye C, Sampson JR. Multiple rare nonsynonymous variants in the adenosomatous polyposis coli gene predispose to colorectal adenomas. Cancer Res 2008, 68:358-363.
3. Brunham LR, Singaraja RR, Hayden MR. Variations on a gene: rare and common variants in ABCA1 and their impact on HDL cholesterol levels and atherosclerosis. Annu Rev Nutr 2006, 26:105-129.
4. Cohen JC, Kiss RS, Pertsemidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 2004, 305:869-872.
5. Cohen JC, Pertsemidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci USA 2006, 103:1810-1815.
6. Ji W, Foo JN, O’Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP: Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet 2008, 40:592-599.
7. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009, 324:387-389.
8. Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet 2007, 39:513-516.
9. Romeo S, Yin W, Kazitina J, Pennacchio LA, Boerwinkle E, Hobbs HH, Cohen JC. Rare loss-of-function mutations in ANGPTL4 family members contribute to plasma triglyceride levels in humans. J Clin Invest 2009, 119:70-79.
10. Slatter TL, Jones GT, Williams MJ, van Rij AM, McCormick SP. Novel rare mutations and promoter haplotypes in ABCA1 contribute to low-HDL-C levels. Clin Genet 2008, 73:179-184.
11. Walsh T, McLellan J, McCarthy SE, Addington AM, Pierce SB, Cooper GM, Nord AS, Kusenda M, Malhotra D, Bhandani A. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008, 320:539-543.
12. Madsen BE, Browning SR. A group-wise association test for rare mutations using a weighted sum statistic. PLOS Genet 2009, 5:e1000384.
13. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). Mutat Res 2007, 615:28-56.
14. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Gen 2008, 83:311-321.
15. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol 2010, 34:188-193.
16. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ, Sunyaev SR. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 2010, 86:832-838.
17. Stephens LA, Boos DD. The calculus of M-estimation. Am Stat 2002, 56:29-38.
18. Almasy LA, Dyer TD, Peralta JM, Kent JW Jr, Charlesworth JC, Curran JE, Blangero J. Genetic Analysis Workshop 17 mini-exome simulation. BMC Proc 2011, 5(suppl 9):S2.
19. Klein TE, Chang JT, Cho MK, Easton KL, Ferguson R, Hewett M, Lin Z, Liu Y, Liu S, Oliver D, et al. Integrating genotype and phenotype information: an overview of the PharmGKB Project. Pharmacogenomics J 2001, 1:167-170.

doi:10.1186/1753-6561-5-S9-S6

Cite this article as: Xing et al.: A weighted accumulation test for associating rare genetic variation with quantitative phenotypes. BMC Proceedings 2011 5(Suppl 9):S6.