ARIC Manuscript Proposal #2553

PC Reviewed: 5/12/15	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Significance analysis of rare variant sets

b. Abbreviated Title (Length 26 characters):

2. Writing Group:
Writing group members:

Baolin Wu
Weihua Guan
James S. Pankow

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. __BW__ **[please confirm with your initials electronically or in writing]**

First author: **Baolin Wu**
Address: Division of Biostatistics
University of Minnesota
A460 Mayo Building, MMC 303
420 Delaware St SE
Minneapolis, MN 55455

Phone: 612-624-0647 Fax: 612-626-0660
E-mail: baolin@umn.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: **James S. Pankow**
Address: Division of Epidemiology, SPH
University Of Minnesota
Room 300 WBOB, 7525A
1300 S 2nd St
Minneapolis, MN 55455

Phone: 612-624-2883 Fax: 612-624-0315
E-mail: panko001@umn.edu

3. Timeline:
We anticipate around three manuscripts ready to submit for Publications Committee review in Fall, 2015.

4. Rationale:
Many methods have been developed to explore the influence of rare variants on the complex diseases (Lee *et al.*, 2014). A single rare variant association test often lacks power. Rare variant set based tests that aggregate information across rare variants within a gene have become attractive to enrich the rare variant association signal. Existing methods are broadly based on the burden test (Li and Leal, 2008; Madsen and Browning, 2009) and the sequence kernel association test (SKAT; Wu *et al.*, 2010, 2011). The burden test aggregates the variant scores within a set and performs well under strong assumptions of similar variant risks. The SKAT assumes unequal variant risks, and performs well for a mix of both protective and deleterious variants. To further improve the power, SKAT has proposed to weight rare variants inversely proportional to their minor allele frequencies (MAFs). Lee *et al.* (2012) proposed SKAT-O to adaptively weight the burden test and SKAT. SKAT-O numerically chooses an optimal weight to maximize the power, and is computationally very involved. SKAT based approaches have commonly used the score statistics, and there have not been any systematic studies of the impact of choice of score test statistics and the associated rare variant weights.

We propose to develop efficient algorithms to compute significance of adaptively weighted burden and SKAT statistics. We also propose and empirically evaluate alternative forms of score statistics and rare variant weights for SKAT. We further propose an adaptive burden test and study the power of adaptively combining SKAT and the proposed adaptive burden test.

5. Main Hypothesis/Study Questions:

We will develop statistical methods for an association test of rare variant sets. We will apply the proposed methods to the ARIC data to identify rare variant sets associated with type 2 diabetes and related phenotypes.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Consider a binary trait denoted as D. Denote X as the covariates to be adjusted for. For a given set of *m* rare variants, denote (G_1, …, G_m) as their copy of minor alleles. We model the binary trait with a logistic regression model, $\text{logit}[Pr(D=1|X,G_1,\ldots,G_m)] =\alpha_0+X\alpha_1+G_1\beta_1+\ldots+ G_m\beta_m$. The rare variant set association is testing the null hypothesis H_0: $\beta_1=\ldots=\beta_m=0$. Note that individual β_k can be estimated and tested using the score statistics U_k, which can be very efficiently computed under the null model. To reduce the number of degrees of freedom in the test, the burden test approach typically assumes $\beta_1=\ldots=\beta_m$, and thus consider a test statistic based on the sum, $U_1+\ldots+U_m$. To allow the

effect β_k vary across SNPs with different directions, SKAT considers sum of squared score statistics, $U_1^2+\ldots+U_m^2$. For both tests, we can incorporate a weight w_k for the *k*-th rare variant, which is typically computed based on the rare variant MAF.

We will evaluate several variations of the burden test and SKAT. First note that U_k is typically not standardized, and it is not intuitive to directly compare them across rare variants. We propose to use the standardized score statistics V_k in the burden test and SKAT, and comparatively study different forms of rare variant weights $\{w_k\}$. Secondly note that the burden test typically works well if all variants have similar effects, and has reduced performance with a mix of protective and deleterious variants. We propose the following adaptive burden test, $B=\max(\max(w_1V_1,0)+\ldots+\max(w_mV_m,0), \max(-w_1V_1,0)+\ldots+\max(-w_mV_m,0))$, which tries to adaptively choose the sum of positive or negative signals. Following Lin (2005), we can use Monte Carlo simulation to efficiently evaluate the significance of B since it is based on the score statistics. We further propose to use the minimum p-value of B and SKAT to adaptively test the significance of the rare variant set. For other type of traits, we will assume the exponential family distribution and generalized linear model. We will demonstrate the power of proposed methods using extensive simulations.

We will illustrate the developed methods using the genetic data in the Atherosclerosis Risk in Communities (ARIC) study. Specifically, we will apply our methods to re-examine the association of exome chip SNPs with type 2 diabetes and related traits, including fasting glucose, fasting insulin, 2-hour glucose after an oral glucose tolerance test in non-diabetic white samples in ARIC. In addition, we will compare results from the proposed methods to existing alternative methods. We will adjust for standard covariates (age, sex, center, and if necessary, principal components of ancestry).

7.a. Will the data be used for non-CVD analysis in this manuscript? ____ Yes __X__ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? ____ Yes ____ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript?
__X__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"?
__X__ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php

_____ Yes __X__ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

ARIC GWAS results for each of the glycemia traits have already been published as part of international meta-analyses conducted by the DIAGRAM Consortium (diabetes status in European populations) and MAGIC Consortium (diabetes-related continuous phenotypes in European populations). ARIC exome chip results for fasting glucose have been published recently.

Rasmussen-Torvik LJ, Alonso A, Li M, Kao W, Kottgen A, Yan Y, Couper D, Boerwinkle E, Bielinski SJ, Pankow JS. Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose. Genet Epidemiol 2010; 34: 665-673.

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL, Lindgren CM, Mägi R, Morris AP, Randall J, Johnson T, Elliott P, Rybin D, Thorleifsson G, Steinthorsdottir V, Henneman P, Grallert H, Dehghan A, Hottenga JJ, Franklin CS, Navarro P, Song K, Goel A, Perry JR, Egan JM, Lajunen T, Grarup N, Sparsø T, Doney A, Voight BF, Stringham HM, Li M, Kanoni S, Shrader P, Cavalcanti-Proença C, Kumari M, Qi L, Timpson NJ, Gieger C, Zabena C, Rocheleau G, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Payne F, Roccasecca RM, Pattou F, Sethupathy P, Ardlie K, Ariyurek Y, Balkau B, Barter P, Beilby JP, Ben-Shlomo Y, Benediktsson R, Bennett AJ, Bergmann S, Bochud M, Boerwinkle E, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Bumpstead SJ, Charpentier G, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Cornelis M, Crawford G, Crisponi L, Day IN, de Geus EJ, Delplanque J, Dina C, Erdos MR, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Fox CS, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Groves CJ, Grundy S, Gwilliam R, Gyllensten U, Hadjadj S, Hallmans G, Hammond N, Han X, Hartikainen AL, Hassanali N, Hayward C, Heath SC, Hercberg S, Herder C, Hicks AA, Hillman DR, Hingorani AD, Hofman A, Hui J, Hung J, Isomaa B, Johnson PR, Jørgensen T, Jula A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, Li Y, Lyssenko V, Mahley R, Mangino M, Manning AK, Martínez-Larrad MT, McAteer JB, McCulloch LJ, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Morken MA, Mukherjee S, Naitza S, Narisu N, Neville MJ, Oostra BA, Orrù M, Pakyz R, Palmer CN, Paolisso G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I, Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rathmann W, Rayner NW, Rice K, Ripatti S, Rivadeneira F, Roden M, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA, Scheet P, Scott LJ, Seedorf U, Sharp SJ, Shields B, Sigurethsson G, Sijbrands EJ, Silveira A, Simpson L, Singleton A, Smith NL, Sovio U, Swift A, Syddall H, Syvänen AC, Tanaka T, Thorand B, Tichet J, Tönjes A, Tuomi T, Uitterlinden AG, van Dijk KW, van Hoek M, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N, Waeber G, Wagner PJ, Walley A, Walters GB, Ward KL, Watkins H, Weedon MN, Wild SH, Willemsen G, Witteman JC, Yarnell JW, Zeggini E, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium, Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Hattersley AT, Silander K, Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR, Cooper C, Dedoussis GV, Serrano-Ríos M, Morris AD, Lind L, Palmer LJ, Hu FB, Franks PW, Ebrahim S, Marmot M, Kao WH, Pankow JS, Sampson MJ, Kuusisto J, Laakso M, Hansen T, Pedersen O, Pramstaller PP, Wichmann HE, Illig T, Rudan I, Wright AF, Stumvoll M, Campbell H, Wilson JF; Anders Hamsten on behalf of Procardis

Consortium; MAGIC investigators, Bergman RN, Buchanan TA, Collins FS, Mohlke KL, Tuomilehto J, Valle TT, Altshuler D, Rotter JI, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector TD, Frayling TM, Ferrucci L, Kong A, Thorsteinsdottir U, Stefansson K, van Duijn CM, Aulchenko YS, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A, Jarvelin MR, Waterworth DM, Vollenweider P, Peltonen L, Mooser V, Abecasis GR, Wareham NJ, Sladek R, Froguel P, Watanabe RM, Meigs JB, Groop L, Boehnke M, McCarthy MI, Florez JC, Barroso I. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature Genetics 2010; 42: 105-116. PMCID: PMC3018764

Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Mägi R, Strawbridge RJ, Rehnberg E, Gustafsson S, Kanoni S, Rasmussen-Torvik LJ, Yengo L, Lecoeur C, Shungin D, Sanna S, Sidore C, Johnson PC, Jukema JW, Johnson T, Mahajan A, Verweij N, Thorleifsson G, Hottenga JJ, Shah S, Smith AV, Sennblad B, Gieger C, Salo P, Perola M, Timpson NJ, Evans DM, Pourcain BS, Wu Y, Andrews JS, Hui J, Bielak LF, Zhao W, Horikoshi M, Navarro P, Isaacs A, O'Connell JR, Stirrups K, Vitart V, Hayward C, Esko T, Mihailov E, Fraser RM, Fall T, Voight BF, Raychaudhuri S, Chen H, Lindgren CM, Morris AP, Rayner NW, Robertson N, Rybin D, Liu CT, Beckmann JS, Willems SM, Chines PS, Jackson AU, Kang HM, Stringham HM, Song K, Tanaka T, Peden JF, Goel A, Hicks AA, An P, Müller-Nurasyid M, Franco-Cereceda A, Folkersen L, Marullo L, Jansen H, Oldehinkel AJ, Bruinenberg M, Pankow JS, North KE, Forouhi NG, Loos RJ, Edkins S, Varga TV, Hallmans G, Oksa H, Antonella M, Nagaraja R, Trompet S, Ford I, Bakker SJ, Kong A, Kumari M, Gigante B, Herder C, Munroe PB, Caulfield M, Antti J, Mangino M, Small K, Miljkovic I, Liu Y, Atalay M, Kiess W, James AL, Rivadeneira F, Uitterlinden AG, Palmer CN, Doney AS, Willemsen G, Smit JH, Campbell S, Polasek O, Bonnycastle LL, Hercberg S, Dimitriou M, Bolton JL, Fowkes GR, Kovacs P, Lindström J, Zemunik T, Bandinelli S, Wild SH, Basart HV, Rathmann W, Grallert H; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, Maerz W, Kleber ME, Boehm BO, Peters A, Pramstaller PP, Province MA, Borecki IB, Hastie ND, Rudan I, Campbell H, Watkins H, Farrall M, Stumvoll M, Ferrucci L, Waterworth DM, Bergman RN, Collins FS, Tuomilehto J, Watanabe RM, de Geus EJ, Penninx BW, Hofman A, Oostra BA, Psaty BM, Vollenweider P, Wilson JF, Wright AF, Hovingh GK, Metspalu A, Uusitupa M, Magnusson PK, Kyvik KO, Kaprio J, Price JF, Dedoussis GV, Deloukas P, Meneton P, Lind L, Boehnke M, Shuldiner AR, van Duijn CM, Morris AD, Toenjes A, Peyser PA, Beilby JP, Körner A, Kuusisto J, Laakso M, Bornstein SR, Schwarz PE, Lakka TA, Rauramaa R, Adair LS, Smith GD, Spector TD, Illig T, de Faire U, Hamsten A, Gudnason V, Kivimaki M, Hingorani A, Keinanen-Kiukaanniemi SM, Saaristo TE, Boomsma DI, Stefansson K, van der Harst P, Dupuis J, Pedersen NL, Sattar N, Harris TB, Cucca F, Ripatti S, Salomaa V, Mohlke KL, Balkau B, Froguel P, Pouta A, Jarvelin MR, Wareham NJ, Bouatia-Naji N, McCarthy MI, Franks PW, Meigs JB, Teslovich TM, Florez JC, Langenberg C, Ingelsson E, Prokopenko I, Barroso I. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nature Genetics 2012; 44: 991-1005. PMCID: PMC3433394

Saxena R, Hivert MF, Langenberg C, Tanaka T, Pankow JS, Vollenweider P, Lyssenko V, Bouatia-Naji N, Dupuis J, Jackson AU, Kao WH, Li M, Glazer NL, Manning AK, Luan J, Stringham HM, Prokopenko I, Johnson T, Grarup N, Boesgaard TW, Lecoeur C, Shrader P, O'Connell J, Ingelsson E, Couper DJ, Rice K, Song K, Andreasen CH, Dina C, Köttgen A, Le Bacquer O, Pattou F, Taneera J, Steinthorsdottir V, Rybin D, Ardlie K, Sampson M, Qi L, van Hoek M, Weedon MN, Aulchenko YS, Voight BF, Grallert H, Balkau B, Bergman RN, Bielinski SJ, Bonnefond A, Bonnycastle LL, Borch-Johnsen K, Böttcher Y, Brunner E, Buchanan TA, Bumpstead SJ, Cavalcanti-Proença C, Charpentier G, Chen YD, Chines PS, Collins FS, Cornelis M, J Crawford G, Delplanque J, Doney A, Egan JM, Erdos MR, Firmann M, Forouhi NG, Fox CS, Goodarzi MO, Graessler J, Hingorani A, Isomaa B, Jørgensen T, Kivimaki M, Kovacs P, Krohn K, Kumari M, Lauritzen T, Lévy-Marchal C, Mayor V, McAteer JB, Meyre D, Mitchell BD, Mohlke KL, Morken MA, Narisu N, Palmer CN, Pakyz R, Pascoe L, Payne F, Pearson D, Rathmann W, Sandbaek A, Sayer AA, Scott LJ, Sharp SJ, Sijbrands E, Singleton A, Siscovick DS, Smith NL, Sparsø T, Swift AJ, Syddall H, Thorleifsson G, Tönjes A, Tuomi T, Tuomilehto J, Valle TT, Waeber G, Walley A, Waterworth DM, Zeggini E, Zhao JH; GIANT consortium; MAGIC investigators, Illig T, Wichmann HE, Wilson JF, van Duijn C, Hu FB, Morris AD, Frayling TM, Hattersley AT, Thorsteinsdottir U, Stefansson K, Nilsson P, Syvänen AC, Shuldiner AR, Walker M, Bornstein SR, Schwarz P, Williams GH, Nathan DM, Kuusisto J, Laakso M, Cooper C, Marmot M, Ferrucci L, Mooser V, Stumvoll M, Loos RJ, Altshuler D, Psaty BM, Rotter JI, Boerwinkle E, Hansen T, Pedersen O, Florez JC, McCarthy MI, Boehnke M, Barroso I, Sladek R,

Froguel P, Meigs JB, Groop L, Wareham NJ, Watanabe RM. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nature Genetics 2010; 42: 142-148. PMCID: PMC2922003

Wessel, J., Chu, A.Y., Willems, S.M., Wang, S., Yaghootkar, H., Brody, J.A., Dauriz, M., Hivert, M.-F., Raghavan, S., Lipovich, L., Hidalgo, B., Fox, K., Huffman, J.E., An, P., Lu, Y., Rasmussen-Torvik, L.J., Grarup, N., Ehm, M.G., Li, L., Baldridge, A.S., Stančáková, A., Abrol, R., Besse, C., Boland, A., Bork-Jensen, J., Fornage, M., Freitag, D.F., Garcia, M.E., Guo, X., Hara, K., Isaacs, A., Jakobsdottir, J., Lange, L.A., Layton, J.C., Li, M., Hua Zhao, J., Meidtner, K., Morrison, A.C., Nalls, M.A., Peters, M.J., Sabater-Lleal, M., Schurmann, C., Silveira, A., Smith, A.V., Southam, L., Stoiber, M.H., Strawbridge, R.J., Taylor, K.D., Varga, T.V., Allin, K.H., Amin, N., Aponte, J.L., Aung, T., Barbieri, C., Bihlmeyer, N.A., Boehnke, M., Bombieri, C., Bowden, D.W., Burns, S.M., Chen, Y., Chen, Y.-D., Cheng, C.-Y., Correa, A., Czajkowski, J., Dehghan, A., Ehret, G.B., Eiriksdottir, G., Escher, S.A., Farmaki, A.-E., Frånberg, M., Gambaro, G., Giulianini, F., Goddard, W.A., Goel, A., Gottesman, O., Grove, M.L., Gustafsson, S., Hai, Y., Hallmans, G., Heo, J., Hoffmann, P., Ikram, M.K., Jensen, R.A., Jørgensen, M.E., Jørgensen, T., Karaleftheri, M., Khor, C.C., Kirkpatrick, A., Kraja, A.T., Kuusisto, J., Lange, E.M., Lee, I.T., Lee, W.-J., Leong, A., Liao, J., Liu, C., Liu, Y., Lindgren, C.M., Linneberg, A., Malerba, G., Mamakou, V., Marouli, E., Maruthur, N.M., Matchan, A., McKean-Cowdin, R., McLeod, O., Metcalf, G.A., Mohlke, K.L., Muzny, D.M., Ntalla, I., Palmer, N.D., Pasko, D., Peter, A., Rayner, N.W., Renström, F., Rice, K., Sala, C.F., Sennblad, B., Serafetinidis, I., Smith, J.A., Soranzo, N., Speliotes, E.K., Stahl, E.A., Stirrups, K., Tentolouris, N., Thanopoulou, A., Torres, M., Traglia, M., Tsafantakis, E., Javad, S., Yanek, L.R., Zengini, E., Becker, D.M., Bis, J.C., Brown, J.B., Adrienne Cupples, L., Hansen, T., Ingelsson, E., Karter, A.J., Lorenzo, C., Mathias, R.A., Norris, J.M., Peloso, G.M., Sheu, W.H.-H., Toniolo, D., Vaidya, D., Varma, R., Wagenknecht, L.E., Boeing, H., Bottinger, E.P., Dedoussis, G., Deloukas, P., Ferrannini, E., Franco, O.H., Franks, P.W., Gibbs, R.A., Gudnason, V., Hamsten, A., Harris, T.B., Hattersley, A.T., Hayward, C., Hofman, A., Jansson, J.-H., Langenberg, C., Launer, L.J., Levy, D., Oostra, B.A., O'Donnell, C.J., O'Rahilly, S., Padmanabhan, S., Pankow, J.S., Polasek, O., Province, M.A., Rich, S.S., Ridker, P.M., Rudan, I., Schulze, M.B., Smith, B.H., Uitterlinden, A.G., Walker, M., Watkins, H., Wong, T.Y., Zeggini, E., The EPIC-InterAct Consortium, Laakso, M., Borecki, I.B., Chasman, D.I., Pedersen, O., Psaty, B.M., Shyong Tai, E., van Duijn, C.M., Wareham, N.J., Waterworth, D.M., Boerwinkle, E., Linda Kao, W.H., Florez, J.C., Loos, R.J.F., Wilson, J.G., Frayling, T.M., Siscovick, D.S., Dupuis, J., Rotter, J.I., Meigs, J.B., Scott, R.A., Goodarzi, M.O., 2015. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat Commun 6. doi:10.1038/ncomms6897

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ____ Yes __X__ No

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* _____)

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ _____ _____)

*ancillary studies are listed by number at http://www.cscc.unc.edu/aric/forms/

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PUBMED Central** whenever the journal does

not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to Pubmed central.

References:

Lee, S., Abecasis, G.R., Boehnke, M., Lin, X., 2014. Rare-Variant Association Analysis: Study Designs and Statistical Tests. The American Journal of Human Genetics 95, 5–23.

Lee,S., Wu,M.C. and Lin,X. (2012) Optimal tests for rare variant effects in sequencing association studies. Biostatistics, 13 (4), 762–775.

Li,B. and Leal,S.M. (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. The American Journal of Human Genetics, 83 (3), 311–321.

Lin, D.Y., 2005. On Rapid Simulation of P Values in Association Studies. The American Journal of Human Genetics 77, 513–514.

Madsen,B.E. and Browning,S.R. (2009) A groupwise association test for rare mutations using a weighted sum statistic. PLOS Genetics, 5 (2), e1000384.

Wu,M., Lee,S., Cai,T., Li,Y., Boehnke,M. and Lin,X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. The American Journal of Human Genetics, 89 (1), 82–93.

Wu,M.C., Kraft,P., Epstein,M.P., Taylor,D.M., Chanock,S.J., Hunter,D.J. and Lin,X. (2010) Powerful SNP-Set analysis for case-control genome-wide association studies. The American Journal of Human Genetics, 86 (6), 929–942.