Optimum Designs for Next-Generation Sequencing to Discover Rare Variants for Common Complex Disease

Gang Shi[1]* and D.C. Rao[1,2]

[1]*Division of Biostatistics and Department of Genetics, Washington University in St. Louis, School of Medicine, Saint Louis, Missouri*
[2]*Departments of Psychiatry and Mathematics, Washington University in St. Louis, School of Medicine, Saint Louis, Missouri*

Recent advances in next-generation sequencing technologies make it affordable to search for rare and functional variants for common complex diseases systematically. We investigated strategies for enriching rare variants in the samples selected for sequencing so as to optimize the power for their discovery. In particular, we investigated the roles of alternative sources of enrichment in families through computer simulations. We showed that linkage information, extreme phenotype, and nonrandom ascertainment, such as multiply affected families, constitute different sources for enriching rare and functional variants in a sequencing study design. Linkage is well known to have limited power for detecting small genetic effects, and hence not considered to be a powerful tool for discovering variants for common complex diseases. However, those families with some degree of family-specific linkage evidence provide an effective sampling strategy to sub-select the most linkage-informative families for sequencing. Compared with selecting subjects with extreme phenotypes, linkage evidence performs better with larger families, while extreme-phenotype method is more efficient with smaller families. Families with multiple affected siblings were found to provide the largest enrichment of rare variants. Finally, we showed that combined strategies, such as selecting linkage-informative families from multiply affected families, provide much higher enrichment of rare functional variants than either strategy alone. *Genet. Epidemiol.* 2011. © 2011 Wiley-Liss, Inc.

Key words: next-generation sequencing; rare variants; enrichment; study design; complex diseases; linkage

INTRODUCTION

Genome-wide association studies (GWAS) employing dense single nucleotide polymorphisms (SNPs) have successfully identified a large number of genetic loci for common complex traits [Hindorff et al., 2009]. However, such approaches relying on common variants have, for the most part, only explained small proportions of the trait heritabilities. For example, the 13 common variants recently identified by two large consortia, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) [Levy et al., 2009] and Global Blood Pressure Genetics (Global BPgen) [Newton-Cheh et al., 2009], collectively explained less than 2% of blood pressure (BP) variance. More common variants are being identified through larger collaborative efforts in mega-consortia with gigantic sample sizes. A recent investigation by the International Consortium on BP identified 16 additional common loci for BP based on a total sample size of approximately 200,000. The 29 variants together explain less than 2.5% of the BP variance. The reliance on common variants is arguably becoming a barrier to further progress.

With the advance of high-throughput next-generation sequencing technologies [Metzker, 2010], discovering rare variants with much larger effect sizes on a massive scale and in a systematic way becomes more feasible and affordable. Genetic dissection for common complex diseases requires a paradigm shift from the "common disease, common variant" hypothesis to "common disease, rare variant" hypothesis [Schork et al., 2009]. Even though rare and functional variants may have very small minor allele frequencies (MAFs), they are likely to have much larger effects [Bodmer and Bonilla, 2008; Schork et al., 2009; Zhu et al., 2010]. Their discovery may therefore explain a potentially large part of the "missing heritability" [Manolio et al., 2009]. In order to identify a rare variant through sequencing, it is necessary to ensure that a certain number of copies of the minor allele is expected to exist in the sequencing sample. From a statistical point of view, the more the rare variants are enriched in the sequencing samples, the larger the statistical power for testing their associations with the diseases. This is true in general for any type of study design. For family studies, it is well known that if one parent has a copy of the minor allele, then half of the offspring are expected to carry it. Therefore, variants that are rare in the general population could be very common in certain families. Here we examine some strategies for selecting such families in which rare variants are enriched. We carried out a simulation study to evaluate the efficacy of enriching rare variants via family-specific

linkage information, extreme phenotype, and nonrandom ascertainments. For statistical analysis strategies involving rare variants, Bansal et al. [2010] recently presented a comprehensive review.

SIMULATION STUDY

We carried out extensive simulations to evaluate the three strategies for enriching rare variants using family data. A quantitative trait was simulated from joint effects of a "causal" SNP, a polygenic component, and a random component. The causal SNP has a MAF that varies from 0.1 to 5% in the general population. The minor allele of the SNP is assumed to increase the phenotype in an additive manner. Its effect size, measured as half the displacement between the two homozygous means, varied from 0.05 standard deviation (SD) units to 2 SD units of the simulated phenotype. A polygenic effect with 20% heritability was simulated to account for additional genetic effects other than the SNP. The rest of the phenotypic variance was generated from an independent and identically distributed normal random variable. A qualitative disease affection status was created for each subject using a liability threshold model. A subject was defined to be affected if the subject's phenotype was as least one SD larger than the phenotypic mean. This yielded a prevalence of 16% for the simulated disease in the general population. Effect sizes of the simulated SNP in terms of percent of phenotypic variance explained (R^2) in random populations are shown in Supplementary Table SI.

We simulated sibships ascertained by three different methods: (1) Sibships randomly drawn from the general population; (2) Sibships recruited through one affected sib (i.e., a proband); and (3) Sibships with all-affected siblings. When simulating sibships ascertained through a proband, we did not restrict the other siblings to be unaffected; hence there could be more than one affected sibs in such sibships. To investigate the effect of family size, sibship size was varied from 2 to 5 while keeping the total sample size at 2,000 subjects, i.e., we simulated 1,000 two-sib families, 667 three-sib families, 500 four-sib families, and 400 five-sib families. Genotype and phenotype data for all possible ascertainment methods, sibship sizes, MAFs, and effect sizes of the SNP were simulated with 1,000 replications of each experimental condition. A microsatellite marker, used for linkage analysis, was simulated in linkage equilibrium with the SNP. Since the purpose of this exercise is to discover the causal SNP through appropriate sequencing, the SNP was treated as latent and therefore unavailable for linkage analysis. The microsatellite marker was assumed to be heterozygous "1/2" in fathers and heterozygous "3/4" in mothers, and zero recombination between the SNP and microsatellite marker was assumed within families. We performed variance component linkage analysis of the simulated family data using MERLIN [Abecasis et al., 2002], and top linkage-positive families were determined based on the family-specific logarithm of odds (LOD) scores.

RESULTS

ENRICHMENT THROUGH LINKAGE INFORMATION IN RANDOMLY ASCERTAINED FAMILIES

For a rare SNP, the minor allele exists only in a small percent of families in the general population. Therefore,

we evaluated the possibility of using linkage information to sub-select those few families that are most likely to segregate for the rare variants. We computed family-specific LOD scores and examined the MAFs in subsets of families with the highest family-specific LOD scores. We defined "enrichment ratio" as the ratio of the observed MAF in the selected sample to that in the general population.

We considered sampling 50, 100, and 200 subjects from families with the highest family-specific LOD scores as shown in Figure 1 for a SNP with MAF = 0.5% in the general population. Results for other MAF values are presented in supplementary materials. When examining Figure 1, it should be remembered that the MAF will not increase in a random sample of families regardless of how many and how large the families are (see panel A in Fig. 3). It is interesting to see that using linkage information helps to improve the concentration of the minor allele in the selected families. Selecting 200 subjects (10% of the total sample of 2,000 subjects) from linkage-positive families (panel A in Fig. 1) enriches the MAF modestly, with up to fivefold enrichment in the largest sibships assuming an effect size of 2 SD. Selecting 100 subjects (5% of the 2,000 subjects) in a like manner (panel B) enriches the MAF more than the larger sample of 200 subjects did. Although at first this may seem unreal, this is true because selecting 100 out of 2,000 is a more extreme sampling than selecting 200 out of the same 1,000, [e.g., see Li and Leal, 2009; Liu and Leal, 2009; Cirulli and Goldstein, 2010]. Finally, selecting only 50 subjects, which is only 2.5% of the total sample of 2,000, enriches the MAF the most, over 12-fold (see panel C). Minor alleles were found to be better enriched in larger families due to the fact that they are more informative for linkage. It is interesting and important to note that family-specific linkage evidence is useful for discriminating families in which the rare variant segregates from those where it does not segregate, even when the aggregate linkage evidence is not strong.

ENRICHMENT THROUGH SELECTING UNRELATED SUBJECTS WITH EXTREME PHENOTYPES IN RANDOMLY ASCERTAINED FAMILIES

Subjects taken from extremes of phenotypic distribution are known to have higher frequencies of rare causal variants compared with random samples from the general population [Li and Leal, 2009; Liu and Leal, 2009; Cirulli and Goldstein, 2010]. We examined this strategy by selecting subjects according to extreme phenotypes regardless of linkage information. We selected 50, 100, and 200 unrelated subjects with the largest phenotypic values from the same random families as discussed above (in the context of Fig. 1). The enrichment results are shown in Figure 2. Unlike when linkage information was used, families with smaller sizes showed better results compared to families with larger sizes. This is true because selecting unrelated subjects from larger number of small families constitutes more extreme selection than from smaller number of large families.

Comparing Figures 1 and 2, we can see that selecting unrelated subjects with extreme phenotypes (from random samples of families) seems to enrich slightly more than selecting the same number of subjects from linkage-positive
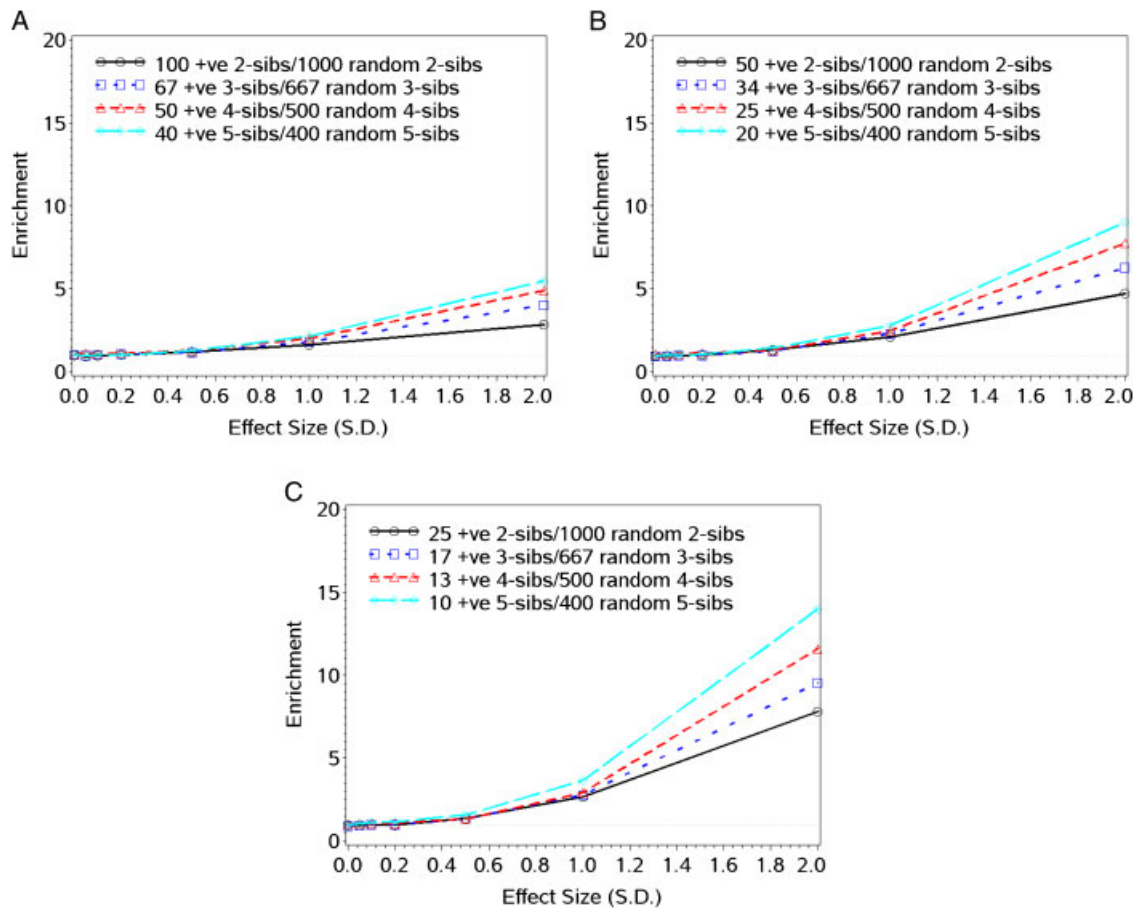
Fig. 1. Enrichment of rare variants through selection of linkage-positive families in random samples of sibships, MAF = 0.5%. (A) Selecting 200 subjects from top linkage positive families; (B) selecting 100 subjects from top linkage positive families; and (C) selecting 50 subjects from top linkage positive families.

families, at least for MAF = 0.5%. It is also clear that linkage information tends to perform better in larger families, while the extreme-phenotype method is more efficient in smaller families.

ENRICHMENT THROUGH ASCERTAINMENT

We evaluated enrichment of rare variants using three different ascertainment methods: randomly ascertained sibships, sibships ascertained through an affected sib (through a proband), and entire sibships with all-affected sibs. We computed MAFs of the simulated causal SNP observed in the sibships ascertained by the three methods and compared them with the MAFs in the general population. The enrichment results are shown in Figure 3 for various family sizes and ascertainment methods. Not surprisingly, as shown in panel A of Figure 3, there is no enrichment of rare variants in randomly sampled sibships: observed MAFs are identical to the values in the general population for all family sizes and effect sizes of the SNPs. It is worth mentioning that the results were averaged over 1,000 replications, and therefore represent the "expected" MAFs. For a single experiment, the observed MAF could drift from generation to generation and the deviation from the expected value could be large especially in small samples.

For sibships ascertained through one proband (panel B in Fig. 3), modest enrichment was observed, with a maximum enrichment of about threefold among the simulated scenarios. The enrichment ratio depends on the effect size of the SNP and the sibship size: the larger the effect size of the SNP and the smaller the sibship size, the greater the enrichment. In particular, the enrichment is inversely related to family size much as it was for extreme phenotype selection (Fig. 2). This is due to the fact that the prevalence of the disease is higher in sibpairs than in larger sibships. In other words, the enrichment gets diluted in larger sibships, which is true for this type of ascertainment method.

For sibships with all-affected siblings (panel C in Fig. 3), the enrichment is substantially larger than when sibships were recruited through proband-based ascertainment. In this case, larger sibships yield better enrichment that is expected since more affected sibs in a family represents more extreme sampling. In this case, rare variants (with MAF = 0.5%) can be enriched as much as 10- to 50-fold.

Finally, exact enrichment results when selecting 100 subjects using each of the strategies are presented in Table I for the SNP with MAF = 0.5% in the general population. Results for additional MAF values are shown in Supplementary Tables SII–SIV.
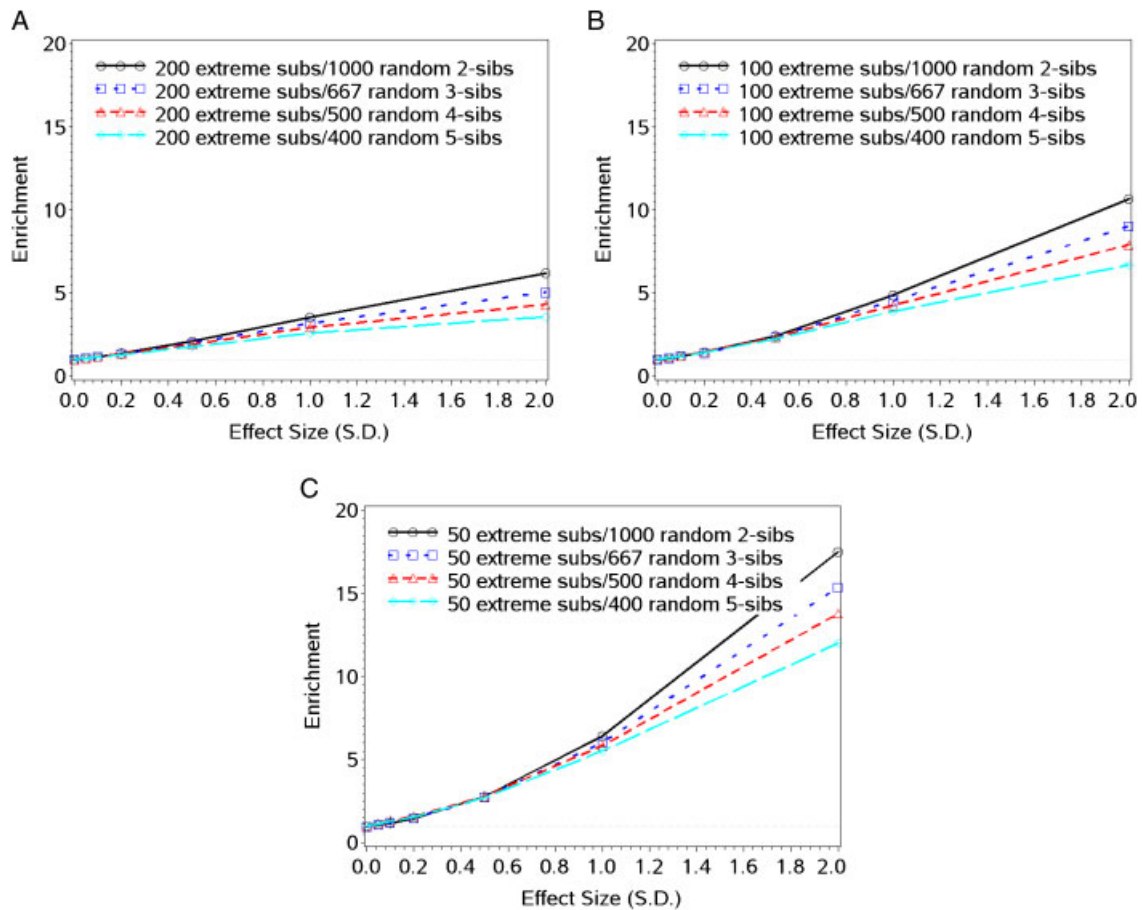
Fig. 2. Enrichment of rare variants through selection of unrelated subjects with extreme phenotypes in random samples of sibships, MAF = 0.5%. (A) Selecting 200 unrelated subjects with extreme phenotypes; (B) selecting 100 unrelated subjects with extreme phenotypes; and (C) selecting 50 unrelated subjects with extreme phenotypes.

OPTIMUM ENRICHMENT THROUGH SELECTION OF LINKAGE-POSITIVE FAMILIES FROM ALL-AFFECTED SIBSHIPS

As shown previously, sequencing samples selected based on family-specific linkage information, extreme phenotype, and nonrandom ascertainment all enrich rare variants. A combination of some of these strategies could provide compound enrichment. Subject to realistic design constraints, e.g., sample size, number of subjects selected for sequencing, etc., it will be useful to evaluate possible combinations of multiple enrichment strategies. Here we consider a particular combination to illustrate the added gain, namely, selecting linkage-positive families from all-affected sibships. Enrichment due to selecting linkage-positive families with 50, 100, and 200 subjects from multiply affected (all-affected) sibships are shown in Figure 4. Compared with results in Figures 1 and 3, such a combination provides much better enrichment of rare variants than either of the individual strategies. For example, when selecting 50 subjects (panel C), the enrichment varies between approximately 50- and 78-fold depending on the sibship sizes assuming that the rare variant in the population with MAF = 0.5% has an effect size of 2 SD.

DISCUSSION

Affected sibpair design [Suarez et al., 1978] has long been popular for linkage study of common complex diseases. Zhu et al. [2010] showed that affected sibpairs enriched rare risk haplotypes better than unrelated "cases" in GWAS. A similar design with multiple affected subjects in families was recently discussed in the context of sequencing studies [Cirulli and Goldstein, 2010]. Common complex diseases usually have high prevalences and therefore affected subjects may not represent very "extremes" in populations. However, multiply affected siblings represent much more "extreme" when the family is considered as the sampling unit. Assuming bivariate normality of quantitative phenotypes in sibpairs, affected sibpairs represent 3.1% of random sibpairs assuming a heritability of 0.2, and 5.4% assuming a heritability of 0.8, compared with 16% for unrelated affected subjects in random populations. Therefore, families with multiple affected subjects are appealing for sequencing studies. For existing family studies where subjects were recruited previously, ascertainment is not a design parameter. However, sibships with more affected sibs should be chosen over those with fewer affected sibs when possible.
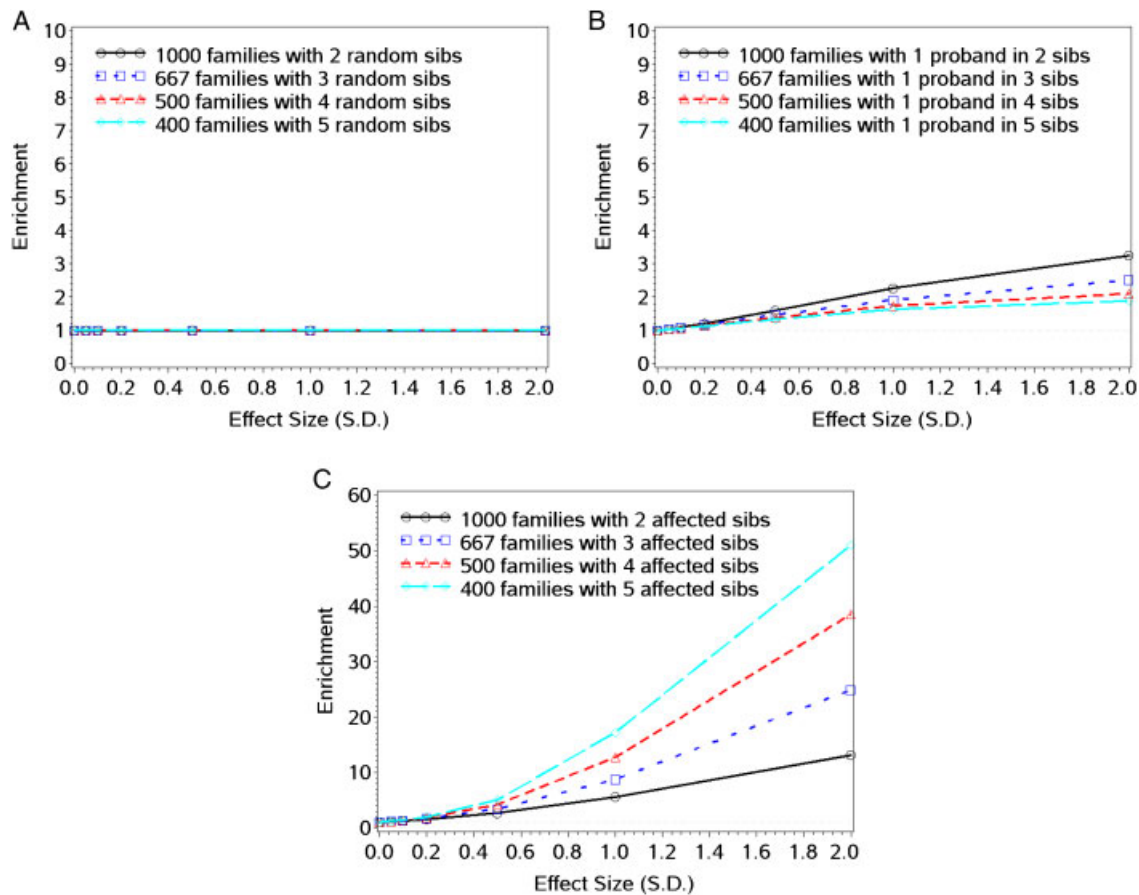
Fig. 3. Enrichment of rare variants through three different ascertainment approaches: random ascertainment, through an affected member (proband), and nonrandom ascertainment of multiply affected sibships, MAF = 0.5%. (A) Sibships recruited randomly; (B) sibships recruited through one affected proband; and (C) multiply affected sibships.

TABLE I. Exact enrichment results when 100 subjects are selected for sequencing based on linkage information, extreme phenotype, ascertaining sibships through one affected proband, and ascertaining multiply affected sibships for MAF = 0.5%

	Enrichment method			
Number and size of families	Linkage	Extreme phenotype	One proband	Multiply affected
1,000 two-sibs	2.1/4.7	4.9/10.7	2.3/3.3	5.6/13.1
667 three-sibs	2.3/6.3	4.5/9.0	1.9/2.5	8.7/24.9
500 four-sibs	2.5/7.8	4.3/7.9	1.7/2.1	12.7/38.5
400 five-sibs	2.8/9.0	3.9/6.7	1.6/1.9	17.2/51.0

First number: enrichment ratio assuming the effect size of the SNP is 1 SD; second number: enrichment ratio assuming the effect size of the SNP is 2 SD.

Many complex diseases are defined based on quantitative disease traits. Our second strategy that selects subjects with extreme phenotype may sound similar to the third one that selects multiply affected sibships. However, the two methods are different parting important ways: First, most current enrichment methods based on extreme phenotype select unrelated subjects typically from the extreme tails of the trait distribution. The third strategy, on the other hand, focuses on selecting sibships with several affected sibs. Second, as another possible combined strategy, selecting subjects with more extreme phenotypes in multiply affected sibships could provide further

enrichment of rare causal variants (results not shown). Third, for affected subjects with medication usage, only their disease status can be identified validly while their (disease) quantitative risk factor values may be confounded by medication effects. In this paper, we used the qualitative affection status and quantitative disease trait for different purposes, the first for sampling sibships and the second for sampling unrelated subjects.

When searching for rare functional variants in random families, overall linkage evidence is likely driven by a few families segregating the rare causal variant. Family-specific linkage evidence provides an effective way to
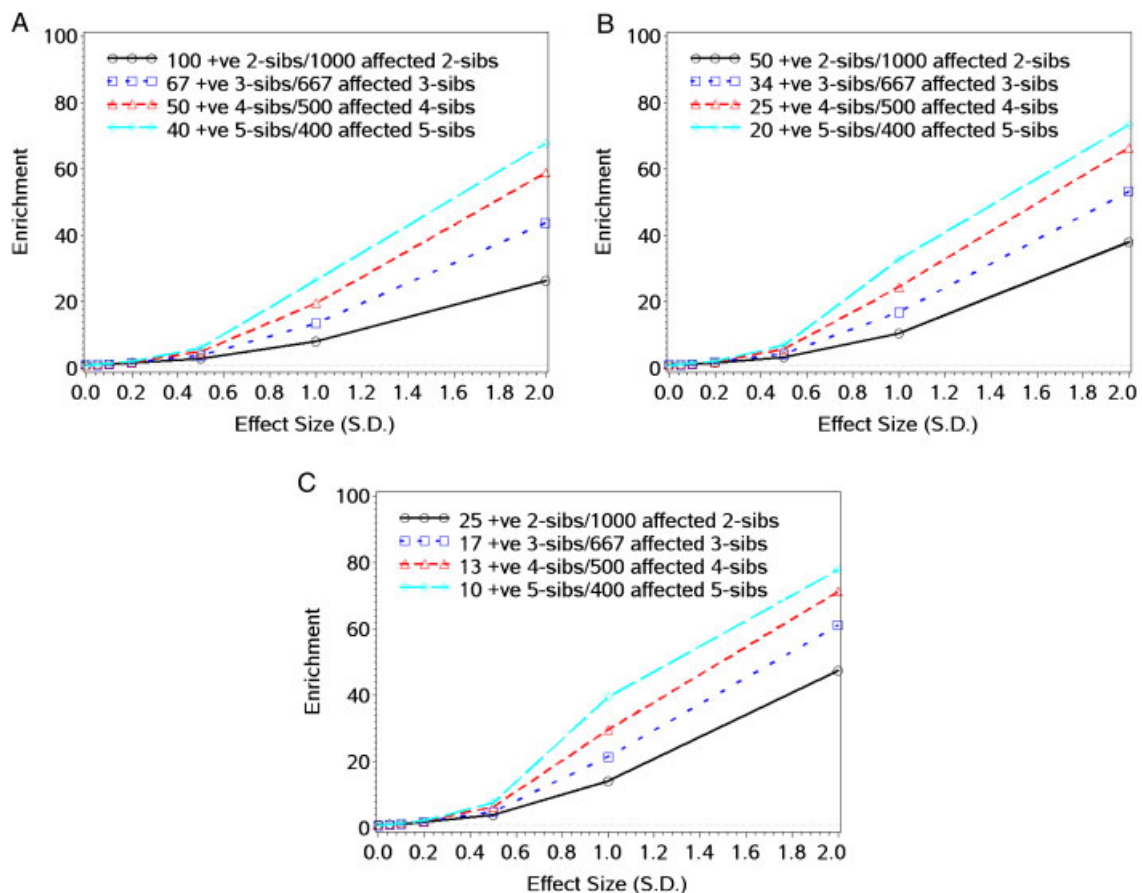
Fig. 4. Enrichment of rare variants through selection of linkage-positive families in multiply affected sibships, MAF = 0.5%. (A) Selecting 200 subjects from top linkage positive families; (B) selecting 100 subjects from top linkage positive families; and (C) selecting 50 subjects from top linkage positive families.

sub-select certain families where the rare variant is much more prevalent (increased MAF). Bowden et al. [2010] were able to enrich a causal variant MAF from 1.1 to 18% by selecting the top two linkage-informative families out of a total of 80 families. The variant explained 63% of the phenotypic variance in the two carriers' families. By selecting linkage-informative families, investigators have discovered two causal variants for an ocular disorder [Nikopoulos et al., 2010] using next-generation sequencing technologies.

As part of the Genetic Analysis Workshop 17, enriching rare variants using family-specific linkage information was evaluated using the publicly available simulated data sets. Results based on large extended pedigrees [Shi et al., 2011] showed that linkage information works best for selecting families in which founders carried minor alleles and the alleles were also transmitted to multiple offspring. In our simulation study, although only siblings were simulated, enrichment through linkage information agrees with the results based on extended pedigrees. It is expected that enrichment of rare variants through linkage evidence would perform even better when the effect size of the SNP and/or the family size were larger. We wish to note that computation of family-specific LOD scores serves as an effective sampling strategy for selecting a small subset of families which have higher odds of segregating (rare) causal variants. In particular, this does not entail any hypothesis testing.

It is important to note that even though we simulated several combinations of MAFs and effect sizes, not all of them are biologically meaningful. Current GWAS findings [Hindorff et al., 2009] suggest that while common variants have small effect sizes, rarer functional variants are likely to have much larger effects [Bodmer and Bonilla, 2008; Schork et al., 2009; Zhu et al., 2010]. For the recently discovered common variants associated with BP and hypertension [Levy et al., 2009; Newton-Cheh et al., 2009], the average allelic effect is around 0.05 SD unit. The lower end of the effect size used in our simulation represents a reasonable effect for common variants but is an excessively conservative value for rare variants. On the other hand, a SNP with MAF = 5% and an effect size of 2 SD would explain 38% of the phenotypic variance in populations, which is equally unlikely for common complex diseases. As shown by Bowden et al. [2010], however, the same effect size would be realistic for rare variants.

It is known that multiple rare functional variants may coexist in a gene. In terms of statistical treatment, multiple rare variants are usually combined first through a collapsing method [Li and Leal, 2008], scored by the number of minor alleles [Morris and Zeggini, 2010], or weighted by allele frequencies [Madsen and Browning, 2009]. The first method implicitly pools all rare functional variants into a single ''mega'' variant whose MAF is the sum of MAFs of all variants. The second approach, where

MAFs are used as weights, assumes that the effect sizes of all rare variants are the same and that they collectively affect the phenotype in an additive manner. For both of them, the overall effect of multiple variants is much larger than the effect size of each individual variant. As a result, even though each variant may be rare in the general population, collective frequency of all rare functional variants in a gene could be more common. Although we simulated only a single SNP in this study, the same conclusions can be applied to the cases with multiple rare variants in one genetic unit. MAFs and effect sizes of the single SNP would represent the overall MAF and effect size for the mega variant (multiple rare variants). When multiple rare variants contribute to the disease through different genes and pathways, strategies based on ascertainment and extreme phenotype could enrich them simultaneously, hence are desirable for variant discovery. On the other hand, linkage information could help to alleviate such locus heterogeneity by enriching variants at targeted loci. In this work, we did not simulate variants in multiple causal genes. Therefore, we could not evaluate the level of enrichment for multiple variants. Enriching multiple rare variants through selecting subjects with extreme phenotype in random populations was recently investigated [Li and Leal, 2009].

Although rare variant enrichment strategies are phenotype-dependent, they allow for cost-effective designs for studying particular diseases. In this work, we used a fixed sample size (number of subjects), fixed size of polygenic effect, and prevalence of disease. Without loss of generality, the conclusions apply to other scenarios. When designing sequencing studies, it would be desirable to conduct simulations with parameters applicable to the particular studies and traits at hand. In real studies, deviation of the quantitative trait from normal distribution will affect the predicted enrichment results for all the three strategies. Measured phenotypes could be influenced by medication use, which will affect the exact enrichment levels of rare variants through extreme phenotype as well as linkage information. Therefore, actual enrichment ratios in real studies may not be exactly as predicted from simulations. From a study design point of view, longitudinal measurements can be used to leverage linkage information as well as account for the so-called regress-to-mean effect when selecting subjects with extreme phenotypes. Finally, using families allows for identifying and confirming private (and functional) mutations, which could lead to identifying genes in the pathophysiology of complex diseases.

ACKNOWLEDGMENTS

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101.

Bansal V, Libiger O, Torkamani A, Schork NJ. 2010. Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 11:773–785.

Bodmer W, Bonilla C. 2008. Common and rare variants in multi-factorial susceptibility to common diseases. Nat Genet 40:695–701.

Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, Hawkins GA, Guo X, Rotter JI, Chen YDI, Wagenknecht LE, Langefeld CD. 2010. Molecular basis of a linkage peak: Exome sequencing and family-based analysis identifies a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. Hum Mol Genet 19:4112–4120.

Cirulli ET, Goldstein DB. 2010. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11:415–425.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 106:9362–9367.

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Köttgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, van Duijn CM. 2009. Genome-wide association study of blood pressure and hypertension. Nat Genet 41:677–687.

Li B, Leal SM. 2009. Discovery of rare variants via sequencing: implications for the design of complex trait association studies. PLoS Genet 5:e1000481.

Liu DJ, Leal SM. 2009. A unified mixed effects likelihood framework for detecting associations with rare variants using sib and unrelated individuals with extreme quantitative phenotypes: application to the next generation sequencing data. Am J Hum Genet S82:51.

Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 5:e1000384.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753.

Metzker ML. 2010. Sequencing technologies—the next generation. Nat Rev Genet 11:31–46.

Morris AP, Zeggini E. 2010. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol 34:188–193.

Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, Najjar SS, Zhao JH, Heath SC, Eyheramendy S, Papadakis K, Voight BF, Scott LJ, Zhang F, Farrall M, Tanaka T, Wallace C, Chambers JC, Khaw KT, Nilsson P, van der Harst P, Polidoro S, Grobbee DE, Onland-Moret NC, Bots ML, Wain LV, Elliott KS, Teumer A, Luan J, Lucas G, Kuusisto J, Burton PR, Hadley D, McArdle WL; Wellcome Trust Case Control Consortium, Brown M, Dominiczak A, Newhouse SJ, Samani NJ, Webster J, Zeggini E, Beckmann JS, Bergmann S, Lim N, Song K, Vollenweider P, Waeber G, Waterworth DM, Yuan X, Groop L, Orho-Melander M, Allione A, Di Gregorio A, Guarrera S, Panico S, Ricceri F, Romanazzi V, Sacerdote C, Vineis P, Barroso I, Sandhu MS, Luben RN, Crawford GJ, Jousilahti P, Perola M, Boehnke M, Bonnycastle LL, Collins FS, Jackson AU, Mohlke KL, Stringham HM, Valle TT, Willer CJ, Bergman RN, Morken MA, Döring A, Gieger C, Illig T, Meitinger T, Org E, Pfeufer A, Wichmann HE, Kathiresan S, Marrugat J, O'Donnell CJ, Schwartz SM, Siscovick DS, Subirana I, Freimer NB, Hartikainen AL, McCarthy MI, O'Reilly PF, Peltonen L, Pouta A, de Jong PE, Snieder H, van Gilst WH, Clarke R, Goel A, Hamsten A, Peden JF, Seedorf U, Syvänen AC, Tognoni G,

Lakatta EG, Sanna S, Scheet P, Schlessinger D, Scuteri A, Dörr M, Ernst F, Felix SB, Homuth G, Lorbeer R, Reffelmann T, Rettig R, Völker U, Galan P, Gut IG, Hercberg S, Lathrop GM, Zelenika D, Deloukas P, Soranzo N, Williams FM, Zhai G, Salomaa V, Laakso M, Elosua R, Forouhi NG, Völzke H, Uiterwaal CS, van der Schouw YT, Numans ME, Matullo G, Navis G, Berglund G, Bingham SA, Kooner JS, Connell JM, Bandinelli S, Ferrucci L, Watkins H, Spector TD, Tuomilehto J, Altshuler D, Strachan DP, Laan M, Meneton P, Wareham NJ, Uda M, Jarvelin MR, Mooser V, Melander O, Loos RJ, Elliott P, Abecasis GR, Caulfield M, Munroe PB. 2009. Genome-wide association study identifies eight loci associated with blood pressure. Nat Genet 41:666–676.

Nikopoulos K, Gilissen C, Hoischen A, van Nouhuys CE, Boonstra FN, Blokland EA, Arts P, Wieskamp N, Strom TM, Ayuso C, Tilanus MA, Bouwhuis S, Mukhopadhyay A, Scheffer H, Hoefsloot LH,

Veltman JA, Cremers FP, Collin RW. 2010. Next-generation sequencing of a 40 Mb linkage interval reveals TSPAN12 mutations in patients with familial exudative vitreoretinopathy. Am J Hum Genet 86:240–247.

Schork NJ, Murray SS, Frazer KA, Topol EJ. 2009. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212–219.

Shi G, Simino J, Rao DC. 2011. Enriching rare variants using family-specific linkage information. BMC Proc 2011 (in press).

Suarez BK, Rice J, Reich T. 1978. The generalized sib pair IBD distribution: its use in the detection of linkage. Ann Hum Genet 42:87–94.

Zhu X, Feng T, Li Y, Lu Q, Elston RC. 2010. Detecting rare variants for complex traits using family and unrelated data. Genet Epidemiol 34:171–187.