Incorporating Linkage Information into a Common Disease/Rare Variant Framework

Anthony L Hinrichs[1] and **Brian K Suarez**[1,2]

[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO

[2]Department of Genetics, Washington University School of Medicine, St. Louis, MO

Abstract

Recent developments in sequencing technology have allowed the investigation of the common disease/rare variant hypothesis. In the Genetic Analysis Workshop 17 data set, we have sequence data on both unrelated individuals and eight large extended pedigrees with simulated quantitative and qualitative phenotypes. Group 11, whose focus was incorporating linkage information, considered several different ways to use the extended pedigrees to identify causal genes and variants. The first issue was the use of standard linkage or identity-by-descent information to identify regions containing causal rare variants. We found that rare variants of large effect segregating through pedigrees were precisely the bailiwick of linkage analysis. For a common disease, we anticipate many risk loci, so a heterogeneity linkage analysis or an analysis of a single pedigree at a time may be useful. The second issue was using pedigree data to identify individuals for sequencing. If one can identify linked regions and even carriers of risk haplotypes, the sequencing will be substantially more efficient. In fact, sequencing only 2.5% of the genome in carefully selected individuals can detect 52% of the risk variants that would be detected through whole-exome sequencing in a large number of unrelated individuals. Finally, we found that linkage information from pedigrees can provide weights for case-control association tests. We also found that pedigree-based association tests have the same issues of binning variants and variant counting as those in tests of unrelated individuals. Clearly, when pedigrees are available, they can provide great assistance in the search for rare variants that influence common disorders.

Keywords

Introduction

Can standard linkage analysis be used to help narrow the region to be searched in an association study? The short answer is, of course, yes. The larger question, however, is whether the advent of genome-wide association (GWA) studies sounds the death knell of linkage studies. This question formed part of the undercurrent explored by Group 11 of Genetic Analysis Workshop 17 (GAW17). We argue here that, because both linkage

Corresponding Author: Anthony L. Hinrichs Washington University School of Medicine 660 South Euclid Box 8134 St. Louis, MO 63110 Phone: 314-362-5370 tony@fire.wustl.edu.

analysis and GWA studies exploit the same physical phenomena, they are not in competition but are rather complementary methods and that the debate pitting the two approaches is a false dichotomy [Risch and Merikangas, 1996]. Each method has its strengths and weaknesses.

One hundred forty-six years ago an obscure Bohemian monk, Gregor Johann Mendel, announced the results of his experiments on the common garden pea (*Pisum sativum*). The full account of his seminal work was published the following year [Mendel, 1866]. However, for a variety of reasons, his findings were largely ignored by the scientific community of his day, only to be independently rediscovered in 1900 by three botanists (the Dutch biologist Hugo de Vries, the German botanist Carl Correns, and the Austrian Erich Tschermak). Mendel's observations led to the formulation of two important laws: the law of segregation and the law of independent random assortment. Ironically, it is the violation of Mendel's second law that underpins both linkage analysis and GWA studies. It is, as it were, the exception that proves the rule.

Mendel studied seven sets of alternative characters in a species that we now know has only seven pairs of chromosomes. Two of the characters he analyzed display phenotypes resulting from genes physically located on the same autosome. The map positions of these two loci, however, are sufficiently distant that their alleles segregate independently. They are syntenic but not linked.

Although chromosomes were first discovered in 1842 by the Swiss botanist and microscopist Karl Wilhelm von Nägeli, the hypothesis that chromosomes carry the genes in a more or less linear order was not confirmed until 1902 by Sutton [1902, 1903]. Indeed, it was not until 1956 that the diploid number of chromosomes for our species was finally determined to be 46 [Ford and Hamerton, 1956; Tjio and Levan, 1956]. Because multiple loci are packaged on the same chromosome, both linkage analysis and GWA studies are obvious approaches for locating those genes. In the not too distant past the challenge for either method was the dearth of markers. Before recombinant DNA technologies were available, the number of markers was restricted to a few dozen blood groups, serum proteins, and cytological markers. (Before use of DNA variants as markers for linkage studies, it was common practice to obtain blood samples from members of a family, centrifuge the samples, and discard the buffy coat containing the DNA!) What changed, beginning in the 1970s, was the revolutionary explosion of discoveries of DNA sequence variants (e.g., restriction fragment length polymorphism [RFLPs], variable number of tandem repeats [VNTRs], microsatellites, and single-nucleotide polymorphisms [SNPs]). It became obvious that the abundance of markers would allow the mapping of all simple Mendelian phenotypes given sufficient family data [Botstein et al., 1980].

Both linkage analysis and GWA techniques aim to locate the chromosomal position of genes that give rise to a measurable phenotype. Historically, the older of these two techniques is linkage analysis, which proved to be an especially useful tool for mapping genes in organisms for which controlled experimental crosses could be performed. In addition to the scarcity of markers, linkage analysis in humans was much more difficult because of the relatively small size of families and the cumbersome statistical methods. Ott's [1974]

implementation of Elston and Stewart's [1971] algorithm in the computer program LIPED revolutionized human linkage analysis. Association studies did enjoy a brief period of popularity in the 1950s, but like linkage analysis, these early case-control studies suffered from two problems: a paucity of genetic markers, so they were largely restricted to searching for an association between various diseases and blood groups; and sample sizes that were in retrospect woefully underpowered [Suarez and Hampe, 1996].

To be sure, there are significant differences between linkage analysis and GWA studies. Whereas most linkage analyses are carried out in families ranging in size from sib pairs to large extended pedigrees (the exception is somatic cell hybrid studies that can locate loci on a particular chromosome.), GWA studies are primarily carried out on unrelated samples of case and control subjects drawn from the same breeding population. Both methods exploit the fact that genes that are located close enough together do not display independent random assortment. In families, the propinquity between the causal gene and the marker need only be close enough that the recombination fraction between the two loci is significantly less than 50%. The causal gene and the marker can be dozens of megabases apart and still provide mapping information. On the other hand, notwithstanding a marker that *is* the culprit responsible for the phenotype under study, a successful GWA study requires that the marker be in linkage disequilibrium with alleles at the causal gene. In an idealized closed random mating population, not subject to mutation, selection, admixture, or drift, the expectation is that, given enough time, alleles at any two genes (or a gene and a marker) will not display linkage disequilibrium. However, the approach to equilibrium may take many generations [Suarez and Hampe, 1996] and, moreover, no human population is, or has ever been, free of the evolutionary forces supposed for the idealized population.

All the contributors to Group 11, which focused on incorporating linkage information, carried out some sort of linkage or family-based analysis. Because a main focus of GAW17 dealt with the analysis of rare variants, standard linkage analysis with the 24,487 SNPs in 3,205 genes would likely have been underpowered because so many SNPs had a minor allele frequency less than 1% (and would not meet the standard definition of a polymorphism [Ford, 1940]). In fact, in the sample of 697 unrelated individuals, 9,433 of the 24,487 SNPs (38.5%) had only a single occurrence. The abundance of these private polymorphisms (in the sense used by Race and Sanger [1950]) would yield a low information content for linkage analysis. Fortunately, the data providers simulated completely informative markers for each of the 3,205 genes and provided the participants with identity-by-descent (IBD) matrices for all pairs of pedigree members. Without these matrices, standard linkage analysis would have had low power to detect any of the causal genes without adding more polymorphic markers with higher information content.

Methods

There were three questions addressed by Group 11. First, we tried to understand how well IBD information and linkage analyses could be used to identify regions containing causal rare variants. Second, we considered the costs and power of using limited sequencing based on the family data. Finally, we investigated the role of rare variants for association tests with family data.

IBD Information and Linkage Analyses

All of the work groups investigating IBD information and linkage analyses used the fully informative IBD matrices. Analyses included all three quantitative traits as well as the affection status. The methods of linkage analysis included affected relative pairs implemented in LODPAL (from the SAGE package [SAGE Project, 2009]), variance components implemented in SOLAR [Almasy and Blangero, 1998], the Elston-Stewart algorithm implemented in FASTLINK [Cottingham et al., 1993; Schäffer et al., 1994], the modified Haseman-Elston algorithm [Elston et al., 2000], the Lander-Green algorithm and variance components implemented in Merlin [Abecasis et al., 2002], and a Markov chain Monte Carlo algorithm implemented in MCLINK [Thomas et al., 2000].

Song et al. [2011] developed a novel method to analyze affection status. Each affected individual was assigned a propensity score [Doan et al., 2006]; this score is the predicted probability of being affected based on a logistic regression using the significant covariates, calculated in each replicate separately. Linkage analysis was performed using LODPAL, and SNPs were placed into the model in turn. Each SNP was assigned the value of the LOD score with the SNP included after subtracting the LOD score of the base model (without any SNP genotypes). This was called the LodDiff.

Shi et al. [2011] used SOLAR on the two quantitative traits with simulated risk loci, Q1 and Q2. They examined the role of family-specific linkage analysis and power across 200 replicates. Hinrichs et al. [2011] also used SOLAR on the quantitative traits but generated a novel phenotype by considering, first, 50 replicates as the repeated measures of a single genotyped individual; they derived a quantitative liability score. For comparison, they also considered the combination of 50 replicates analyzed using Fisher's method [Province, 2001]. Simpson et al. [2011] analyzed both the quantitative traits and the affection status using the modified Haseman-Elston algorithm, the Lander-Green algorithm, and variance components. They also analyzed affection status using the Elston-Stewart algorithm.

Akula et al. [2011] performed a novel variation on the affected relative pairs linkage method. They examined only affected pairs of fourth- and fifth-degree relatives to identify loci with higher sharing than expected. This was repeated across all replicates.

Identifying Sequencing Candidates

The second question addressed by Group 11 was a cost-benefit analysis of performing limited resequencing to identify rare variants using family data. In the GAW17 data set, we were provided with sequence data for all exons in all 3,205 genes in all individuals generated by gene dropping from subjects in the 1000 Genomes Project. In any other sample, this much sequencing would be prohibitively expensive (although prices are dropping rapidly). We considered the advantages and disadvantages of restricting attention to only genes underneath established linkage peaks and selecting individuals likely to be carriers of rare variants with large effect size.

Allen-Brady et al. [2011] used MCLINK for the affection status across 23 unilineal pedigrees in each of 10 replicates. They identified regions with a heterogeneity LOD [Smith, 1961] score of 3.3 or greater (significant) or less than 3.3 but greater than 1.86 (suggestive).

They considered a number of different methods to select individuals for these gene regions (such as the youngest affected subject or low-covariate-risk subjects) and tested to see how frequently a rare causal variant would have been identified. Similarly, Choi et al. [2011] considered linkage analysis of the quantitative phenotypes and how frequently a causal rare variant would have been identified through exonic sequencing of the linked family.

Association Testing

The final topic addressed by Group 11 was to examine the role of rare variants for association tests with family data. Feng et al. [2011] developed a novel statistical method in which data from affected sibling pairs within arbitrary pedigrees was used to weight variants; they then ran association tests based on those weights. In particular, to test a set of variants belonging to a group (gene, pathway, genomic region, etc.), they generated a weight for each variant on the basis of affected siblings carrying the minor allele and unaffected siblings carrying the major allele. For a sample of unrelated case and control subjects, each individual was given a genetic score based on the weights for each variant. This approach is similar to a method developed by Madsen and Browning [2009], where weights were created using the inverse of the variance of the minor allele frequency in control subjects. Feng and colleagues first tested a single replicate and then compiled replicates to generate 400 affected sib pairs and 2,000 case subjects and 2,000 control subjects.

Almeida et al. [2011] compared a variety of association methods and tested sensitivity, specificity, and positive and negative predictive values. These methods included looking for reduced heritability estimates from SOLAR by including polymorphisms as covariates and p-values from QTDT [Abecasis et al., 2000]. They considered several different methods for binning rare alleles. They also examined the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to look for interconnected pathways with influence on the traits.

Simpson et al. [2011] compared power and type I error rates of two family-based association tests as applied to the quantitative traits: regression on mid-parent (ROMP) [Pugh et al., 2001; Roy-Gagnon et al., 2008] and ASSOC from the SAGE package [SAGE Project, 2009].

Results

First and foremost, we note that the contributors chose a variety of subsets of the data for testing: The number of replicates used varied between 1 and all 200; in some cases only a few of the large extended pedigrees were used; and several methods decomposed the large pedigrees into nuclear subsets. The results presented here are based on the internal testing that each work group performed, and the results may differ on different data sets. In the Discussion section, we present results that were observed by several teams.

IBD Information and Linkage Analyses

Song et al. [2011] looked for an increase in LOD score (the LodDiff) when a causal SNP was included in the trait model. They found that the LodDiff was highly skewed toward true positives. However, their method still requires work into the statistical properties of the

LodDiff. In particular, the theoretical and empirical distributions of the LodDiff need to be developed.

Shi et al. [2011], Hinrichs et al. [2011], and Simpson et al. [2011] all revealed the ability of linkage analysis to localize candidate genes with true causal variants. Many of the significant findings occurred in only a single family or in a small number of families where a causal variant happened to segregate throughout a pedigree (Table I). In fact, a number of causal variants that occurred only a few times in the unrelated individuals and thus could not have been detected in that sample showed significant linkage. These three work groups also noticed odd patterns of false positives that may have been due to identical genotype data across replicates. Hinrichs and colleagues found that the repeated-measures framework, with multiple phenotypic observations for the same genotype, might be useful for longitudinal data in human or animal genetics.

Simpson et al. [2011] showed that methods using extended pedigrees (such as variance components analysis or Elston-Stewart model-based linkage in extended pedigrees) had substantially greater power than sibling pair and nuclear pedigree-based linkage analysis. However, the Elston-Stewart algorithm is severely limited in the number of markers it can analyze at one time in multipoint linkage, which reduces its utility in modern sequencing studies. Finally, following up linkage signals that are suggestive rather than genome-wide significant does not adequately control for type I error.

Akula et al. [2011] found that looking at sharing for distantly affected pairs of relatives resulted in a low true-positive rate (4.6%). However, this method may be more useful in a different context, because there are lots of variants that increase risk in these data. The technique may be most useful for Mendelian traits, rare traits for which there are unlikely to be multiple risk variants within a pedigree, or inbred pedigrees for which distantly related individuals have IBD = 2. Because the GAW17 data set has no inbreeding and simulates a common disease with many risk variants, further investigation of the method in other data sets is necessary.

Identifying Sequencing Candidates

Allen-Brady et al. [2011] found that it is far less efficient to sequence unrelated case subjects than to use linkage information to identify linked pedigrees and to select outlying case subjects from the pedigrees, such as the youngest affected individual or affected individuals with low covariate risk. Identifying haplotype carriers through examination of segregation patterns, when possible, is also extremely useful. Similarly, Choi et al. [2011] found that by sequencing only 2.5% of the exome (the areas under significant linkage peaks in the families), they were able to identify 52% of the loci that would have been identified by whole-exome sequencing in the unrelated individuals.

Association Testing

Feng et al. [2011] found that the weights generated from family data substantially improved on the Madsen-Browning weighting scheme. Although neither method had power in a single replicate, when considering a larger data set, their novel method found more significance for

a true positive than the Madsen-Browning method in one case and identified a true positive missed by the Madsen-Browning method in another case.

Almeida et al. [2011] showed that several different ways of dealing with the information provided by rare and common variants showed low sensitivity and low positive predictive value to detect causal genes. The polygenic model using the information provided by common variants that altered trait heritability in at least one family presented the highest level of sensitivity but a lower positive predictive value. However, these analyses were based on only a single replicate. Almeida and colleagues also noted that only a third of genes present in the exome had entries in the KEGG database, and consequently any kind of annotation would neglect a significant proportion of genetic information

Simpson et al. [2011] found that the nuclear family-based tests of association (ROMP and ASSOC) were more powerful than the linkage methods in detecting variants with causal effects on the quantitative traits Q1 and Q2 but that they also lacked stringent control of type I error.

Discussion

Group 11 set out to examine how linkage information could be incorporated into a common disease/rare variant framework using the GAW17 data. Three questions were considered: Can linkage analyses be used to detect rare variants? Can we select sequencing candidates from family data? Can we perform association with rare variants in family data? However, we note that the term *rare variant* may be misleading. In particular, a variant that appears in only a small percentage of unrelated individuals may be enriched within a family and actually be quite common. Similarly, a variant that is common in unrelated individuals but that happens to not occur in the founders of a pedigree will appear to be monomorphic. All the methods described here depend to some extent on the frequency of the risk variants, but they tend to rely more on the segregation of those variants conditional on their presence in founders.

We found that linkage analyses work well in the common disease/rare variant framework. In particular, when a rare variant of large effect segregates through a pedigree, this perfectly fits a linkage model for quantitative or qualitative traits. After examining the families contributing to the true-positive signals, our group found that often a single family or a small number of families were responsible for the linkage signal. For example, in Table I we see that the single causal variant in *VEGFC* occurs only once in the unrelated sample but is observed in 31 individuals in Family 7. On the other hand, any risk variants that do not appear in the relatively small number of founders cannot be detected through any family-based analysis.

Our second issue, the selection of sequencing candidates, investigates the circumstance in which sequencing resources are limited and available for only a small number of individuals. Our results indicate that appropriate selection of individuals for sequencing can greatly reduce sequencing costs while maintaining reasonable power. In particular, when family-specific analysis shows strong linkage to a particular region, the affected or

quantitatively extreme individuals, when sequenced, are much more likely to identify rare variants than randomly selected case subjects. If a risk haplotype segregating through the pedigree can be identified, then the chance of finding risk variants will be further increased.

For the final issue, association tests, we find that families also have advantages. We observed that family linkage information can provide weights for an analysis of unrelated case and control subjects. This technique outperforms other methods based only on weight or on minor allele frequencies. We also found that when directly testing association in pedigrees, we encounter the same issues as with unrelated case and control subjects, namely, that when multiple rare variants occur within a gene, we need to consider binning or allele counts rather than simply performing separate tests on each variant.

Clearly then, in a common disease/rare variant framework, extended pedigrees have tremendous promise. Using linkage analysis to identify candidate genes, identifying individuals for sequencing, and performing association tests within pedigrees can all have better power than simply looking at unrelated case and control subjects. On the other hand, these methods require the use of informative markers outside what would be typed during exonic sequencing in case and control groups. This suggests that the many studies that have collected large pedigrees and typed them with microsatellites or SNPs for linkage would be well served by reexamining them to identify loci where rare variants of large effect may be segregating.

Acknowledgments

References

Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. Am J Hum Genet. 2000; 66:279–92. [PubMed: 10631157]

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin: rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet. 2002; 30:97–101. [PubMed: 11731797]

Akula N, Detera-Wadleigh S, Shugart YY, Nalls M, Steele J, McMahon FJ. Identity-by-descent filtering as a tool for the identification of disease alleles in exome sequence data from distant relatives. BMC Proc. 2011; 5(suppl 9):S76. [PubMed: 22373213]

Allen-Brady K, Farnham J, Cannon-Albright L. Strategies for selection of subjects for sequencing after detection of a linkage peak. BMC Proc. 2011; 5(suppl 9):S77. [PubMed: 22373232]

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62:1198–211. [PubMed: 9545414]

Almeida MAA, Horimoto A, Oliveira P, Krieger J, Pereira A. Different approaches for dealing with rare variants in family-based genetic studies: application of a Genetic Analysis Workshop 17 problem. BMC Proc. 2011; 5(suppl 9):S78. [PubMed: 22373261]

Botstein D, White R, Skolnick M, Davis R. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet. 1980; 32:314–31. [PubMed: 6247908]

Choi S, Liu C, Dupuis J, Logue M, Jun G. Using linkage analysis of large pedigrees to guide association analyses. BMC Proc. 2011; 5(suppl 9):S79. [PubMed: 22373287]

Cottingham RW, Idury RM, Schäffer AA. Faster sequential genetic linkage computations. Am J Hum Genet. 1993; 53:252–63. [PubMed: 8317490]

Doan BQ, Sorant AJM, Frangakis CE, Bailey-Wilson JE, Shugart YY. Covariate-based linkage analysis: application of a propensity score as the single covariate consistently improves power to detect linkage. Eur J Hum Genet. 2006; 14:1018–26. [PubMed: 16736037]

Elston RC, Buxbaum S, Jacobs KB, Olson JM. Haseman and Elston revisited. Genet Epidemiol. 2000; 19:1–17. [PubMed: 10861893]

Elston R, Stewart J. A general model for the genetic analysis of pedigree data. Hum Hered. 1971; 21:523–42. [PubMed: 5149961]

Feng T, Elston RC, Zhu X. A novel method to detect rare variants using both family and unrelated case-control data. BMC Proc. 2011; 5(suppl 9):S80. [PubMed: 22373319]

Ford C, Hamerton J. The chromosomes of man. Nature. 1956; 178:1020–3. [PubMed: 13378517]

Ford, E. Polymorphisms and taxonomy. In: Huxley, J., editor. The New Systematics. Clarendon Press; Oxford: 1940. p. 407-10.

Hinrichs AL, Culverhouse RC, Suarez BK. Linkage analysis merging replicate phenotypes: an application to three quantitative phenotypes in two African samples. BMC Proc. 2011; 5(suppl 9):S81. [PubMed: 22373343]

Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet. 2009; 5:e1000384. [PubMed: 19214210]

Mendel G. Versuche über Pflanzenhybriden. Verhdlg Naturf Verein Brünn. 1866; 4:3–47.

Ott J. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet. 1974; 26:588–97. [PubMed: 4422075]

Province MA. The significance of not finding a gene. Am J Hum Genet. 2001; 69:660–3. [PubMed: 11481587]

Pugh EW, Papanicolaou GJ, Justice CM, Roy-Gagnon MH, Sorant AJ, Kingman A, Wilson AF. Comparison of variance components, ANOVA, and regression of offspring on midparent (ROMP) methods for SNP markers. Genet Epidemiol. 2001; 21(suppl 1):S794–9. [PubMed: 11793780]

Race, R.; Sanger, R. Blood Groups in Man. FA Davis; Philadelphia: 1950.

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273(5281):1516–7. [PubMed: 8801636]

Roy-Gagnon MH, Mathias RA, Fallin MD, Jee SH, Broman KW, Wilson AF. An extension of the regression of offspring on mid-parent to test for association and estimate locus-specific heritability: the revised ROMP method. Ann Hum Genet. 2008; 72:115–25. [PubMed: 18042270]

SAGE Project. SAGE: Statistical Analysis for Genetic Epidemiology. release 6.0.12009. http://darwin.cwru.edu/sage

Schäffer AA, Gupta SK, Shriram K, Cottingham RW. Avoiding recomputation in linkage analysis. Hum Hered. 1994; 44:225–37. [PubMed: 8056435]

Shi G, Simino J, Rao D. Enriching rare variants using family-specific linkage information. BMC Proc. 2011; 5(suppl 9):S82. [PubMed: 22373363]

Simpson CL, Justice CM, Krishnan M, Wojciechowski R, Sung H, Cai J, Green T, Lewis D, Behneman D, Wilson AF, et al. Old lessons learned anew: family-based methods for detecting genes responsible for quantitative and qualitative trait variation in the Genetic Analysis Workshop 17 mini-exome sequence data. BMC Proc. 2011; 5(suppl 9):S83. [PubMed: 22373393]

Smith C. Homogeneity test for linkage data. Proc Sect Intl Congr Hum Genet. 1961; 1:212–3.

Song YE, Namkung J, Shields RW, Baechle DJ, Song S, Elston RC. A method to detect single-nucleotide polymorphisms accounting for a linkage signal using covariate-based affected relative pair linkage analysis. BMC Proc. 2011; 5(suppl 9):S84. [PubMed: 22373405]

Suarez B, Hampe C. Linkage and association. Am J Hum Genet. 1996; 54:554–9. [PubMed: 8116626]

Sutton WS. On the morphology of the chromosome group in *Brachystola magna*. Biol Bull. 1902; 4:24–39.

Sutton WS. The chromosomes in heredity. Biol Bull. 1903; 4:231–51.

Thomas A, Gutin A, Abkevich V, Bansal A. Multilocus linkage analysis by blocked Gibbs sampling. Stat Comput. 2000; 10:259–69.

Tjio J, Levan A. The chromosome number of man. Hereditas. 1956; 42:1–6.

Table I

Distribution of causal SNPs by sample

Gene	Trait	Chromo some	Number of SNPs[a]	CC[b]	F1[c]	F2	F3	F4	F5	F6	F7	F8
AKT3	DX[d]	1	1	1	0	0	0	0	0	0	0	2
ARNT	Q1	1	5	23	3	0	0	0	6	0	0	0
BCHE	Q2	3	13	19	4	2	0	0	1	0	2	2
BCL2L11	DX	2	3	6	0	0	0	0	0	2	0	0
ELAVL4	Q1, DX	1	2	2	0	0	1	0	0	0	0	0
FLT1	Q1	13	11	170	0	0	5	0	28	16	13	12
FLT4	Q1	5	2	3	0	0	0	0	0	0	0	0
GCKR	Q2	2	1	17	2	0	0	0	0	0	0	0
HIF1A	Q1	14	4	22	0	0	0	0	1	0	0	0
HIF3A	Q1	19	3	3	0	0	0	0	3	0	0	0
HSP90AA1	DX	14	4	367	61	78	74	64	59	65	115	60
INSIG1	Q2	7	3	3	0	0	0	0	0	0	0	0
KDR	Q1	4	10	271	14	37	43	31	6	0	93	0
LPL	Q2	8	3	25	7	20	2	1	0	1	23	7
NRAS	DX	1	1	1	0	0	0	0	0	0	0	0
PDGFD	Q2	11	4	16	0	0	0	0	0	0	0	0
PIK3C2B	DX	1	24	62	3	4	8	0	4	6	2	0
PIK3C3	DX	18	2	25	0	0	0	0	2	0	5	0
PIK3R3	DX	1	1	1	0	0	0	0	0	0	0	0
PLAT	Q2	8	8	23	0	1	2	0	0	3	0	0
PRKCA	DX	17	2	233	53	57	70	51	29	47	48	60
PRKCB1	DX	16	1	1	0	0	0	0	0	0	0	0
PTK2	DX	8	2	2	0	0	0	0	0	0	0	0
PTK2B	DX	8	3	5	0	0	0	0	0	0	0	0
RARB	Q2	3	2	8	0	0	0	0	0	0	0	0
RRAS	DX	19	2	4	0	0	0	0	0	2	1	0
SHC1	DX	1	1	9	0	0	0	0	0	0	0	0
SIRT1	Q2	10	8	12	0	0	3	0	0	4	21	0

Gene	Trait	Chromo some	Number of SNPs[a]	CC[b]	F1[c]	F2	F3	F4	F5	F6	F7	F8
SOS2	DX	14	2	6	0	0	0	0	0	0	0	0
SREBF1	Q2	17	10	31	16	21	0	5	9	12	8	0
VEGFA	Q1	6	1	3	4	20	0	0	0	0	22	0
VEGFC	Q1	4	1	1	0	0	0	0	0	0	31	0
VLDLR	Q2	9	8	14	2	0	0	0	11	0	6	0
VNN1	Q2	6	2	246	35	60	27	15	24	17	77	0
VNN3	Q2	6	7	201	24	17	27	0	21	0	42	42
VWF	Q2	12	2	9	0	0	0	2	0	0	0	0

[a] Number of variants within the gene that influence the trait.

[b] Number of causal variants observed in the unrelated sample.

[c] F1 to F8 are the number of causal variants observed in each family (Family 1 through Family 8).

[d] DX is diagnosis, that is, presence or absence of a discrete phenotype.