REVIEW PAPER

The role of large pedigrees in an era of high-throughput sequencing

Ellen M. Wijsman

Abstract Rare variation is the current frontier in human genetics. The large pedigree design is practical, efficient, and well-suited for investigating rare variation. In large pedigrees, specific rare variants that co-segregate with a trait will occur in sufficient numbers so that effects can be measured, and evidence for association can be evaluated, by making use of methods that fully use the pedigree information. Evidence from linkage analysis can focus investigation, both reducing the multiple testing burden and expanding the variants that can be evaluated and followed up, as recent studies have shown. The large pedigree design requires only a small fraction of the sample size needed to identify rare variants of interest in population-based designs, and many highly suitable, well-understood, and available statistical and computational tools already exist. Samples consisting of large pedigrees with existing rich phenotype and genome scan data should be prime candidates for high-throughput sequencing in the search of the determinants of complex traits.

E. M. Wijsman
Department of Biostatistics, University of Washington, Seattle, WA, USA

E. M. Wijsman (✉)
Division of Medical Genetics, Department of Medicine, University of Washington, Box 357720, Seattle, WA 98195-7720, USA
e-mail: wijsman@u.washington.edu

E. M. Wijsman
Department of Genome Sciences, University of Washington, Seattle, WA, USA

Background

Major scientific successes typically build on previous discoveries. This is certainly true of human gene identification studies. Work in model organisms initially delineated the principles behind genetic mapping (Sturtevant 1913), although even in model organisms it took 50 years to generalize the principles to complex quantitative traits (Thoday 1961). Over the same period, statistical geneticists outlined specific approaches needed for gene mapping in human data, describing early ideas leading to allele-sharing statistics (Penrose 1935) and the classical lod score approach (Haldane and Smith 1947; Morton 1955). It then took another 20 years for development of the first practical computational algorithms based on pedigree peeling (Elston and Stewart 1971) with implementation in computer programs (Ott 1974). Another decade passed before an ample source of necessary markers in the form of DNA-based variation began to be available (Botstein et al. 1980). The principles elucidated by these fundamental ideas mean that all are still in use, although methodologies implementing the ideas, including high-throughput sequencing, are now vastly cheaper, faster and more efficient.

Once all the pieces were in place, progress toward human gene identification was rapid. The fundamental underlying principles coupled with an ample source of markers lead to early studies that proved feasibility in humans by leading to genes for both recessive (Riordan et al. 1989) and dominant (Huntington's Disease Collaborative Research Group 1993) genetic diseases. They also relatively quickly lead to early identification of genes affecting moderately complex traits, such as early-onset breast cancer (Miki et al. 1994) and Alzheimer's disease (Levy-Lahad et al. 1995; Sherrington et al. 1995). The overall strategy was described as positional cloning (Collins 1991), although more recent studies should more

properly be referred to as positional sequencing. These early studies illustrated the power of the pedigree-based design in leading to genes of interest with rare but highly influential disease alleles. With this approach, close to 4,500 genes relevant to human disease had been identified by the end of 2011 (Amberger et al. 2011). However, moving from a map location to gene identification was difficult because of the poor resolution of pedigree-based designs (Boehnke 1994), and most success stories represent Mendelian disorders or essentially Mendelian forms of more complex disorders.

Complex traits have yielded less frequently to this approach (McClellan and King 2010; Risch 2000). Struggles with the challenges of complex traits eventually lead to an alternative push toward use of population-based designs (Botstein and Risch 2003) under the assumption that much of complex trait genetics might be explained by common genetic variation (Collins et al. 1997) and detectable with association methods. With inexpensive genotyping of very dense SNP panels, genome-wide association studies in large population-based samples (GWAS) became common, with >1,200 publications since 2005 that report at least one association at a genome-wide significance level (http://www.genome.gov/gwastudies). There have, of course, been some clear success stories with identification of the underlying genes, such as complement H and macular degeneration (Klein et al. 2005). However, early optimistic predictions about the impact of these discoveries (Manolio et al. 2008) became more guarded for two reasons: (1) most GWAS studies have only lead to associations, with identification and verification of the actual risk variants a rare outcome (Hindorff et al. 2009); and (2) most estimated effects are small, thus explaining relatively little of the estimated genetic variance (Manolio et al. 2009). Given the small estimated effect sizes for most traits, sample sizes have now reached enormous sample sizes of up to hundreds of thousands (Speliotes et al. 2010) of subjects—several orders of magnitude more than that needed for pedigree-based studies. One current hypothesis is that rare variation is much more important than was earlier believed (Manolio et al. 2009). It is also probable that human variation includes all of the complex features found in model organisms: complex genetic architectures, existence of rare alleles, and the presence of many influential variants that are in noncoding regions (Flint and Mackay 2009), and therefore not easily amenable to functional prediction algorithms (Erlich et al. 2011).

Investigation of rare variants is the current frontier of human genetic disease. Ability to extensively survey existing DNA-based variation is increasingly feasible through use of rapidly improving high-throughput sequencing methods (Ng et al. 2010; Roach et al. 2010). The hypothesis that rare variants are important contributors to complex human traits is further supported by information from (1) early sequencing studies of subjects in the tails of the distribution of cardiovascular-related quantitative traits (Cohen et al. 2004), (2) indications that rare variants are more likely than common variants to have large effects (Bodmer and Bonilla 2008; Gorlov et al. 2011), (3) recent sequencing of regions surrounding seven genes associated with low-density lipoprotein levels (Sanna et al. 2011), and (4) large numbers of variants typically identified in genes affecting Mendelian disorders, such as familial hypercholesterolemia with >1,100 known variants (Leigh et al. 2008). Sample size requirements for GWAS of rare variants would need to increase substantially over current sample sizes. This leads to the inescapable conclusion that designs, other than simple population-based designs, may be critical, as high-throughput sequencing in the search for influential rare variation joins the research toolkit. Use of large pedigrees is one of several important designs in this context and the focus of this article.

Advantages of family-based designs

Harnessing segregation information

Large pedigrees provide a design that has considerable power in the search for rare trait variants. In this context, large pedigrees are those that are individually large enough to provide a statistically significant result, given good quality and quantity of data, and use of an efficient analysis approach. Such pedigrees will typically be at least 20–25 subjects, but will frequently be much larger. The trait loci may represent single genes with relatively high effects, or several closely linked loci, each with moderate effects that together have a large effect (Yazbek et al. 2011). Large pedigrees intrinsically have more power for detection of linkage or estimation of effects than do equivalent-sized samples of smaller families (Wijsman and Amos 1997) or unrelated subjects, particularly in the presence of rare variants such as those found in sequence data (Gagnon et al. 2011; Simpson et al. 2011; Wilson and Ziegler 2011). If pedigrees are sufficiently large, they can individually implicate genomic regions. For example, some current methods, which are modern versions of early allele-sharing methods, identify shared multilocus segments among small numbers of subjects. These methods rely on very large pedigrees with many meioses separating affected subjects within pedigrees (Leibon et al. 2008; Thomas et al. 2008). They do not require genotyping of intervening relatives, with the pedigree structure providing information about the expected distribution of the number and sizes of such segments. Large pedigrees, therefore, enable gene mapping and identification studies to be carried out with relatively small sample sizes (Wright et al. 1999). The large pedigree design is thus particularly useful in the presence of locus

heterogeneity among families. The fundamental disadvantage of pedigree designs is that genomic regions identified via linkage analysis tend to be relatively large because of the coarse nature of the meiotic process (Boehnke 1994). In earlier years, the cost and difficulty of evaluating all variation in an identified genomic region was a barrier to identifying the gene(s) driving the evidence for linkage, particularly in smaller sample sizes for which the implicated regions were particularly large. However, with the recent revolutionary changes provided by new high-throughput sequencing technologies, genotyping costs for follow-up studies in such a region are no longer a major limitation since the whole region can be evaluated.

Evidence for influence of rare variants often includes evidence for co-segregation with traits in pedigrees. This is sometimes carried out in an ad hoc manner without quantification of statistical support (Cruchaga et al. 2012). However, evaluation for co-segregation between the trait and variant can be obtained by carrying out a classical linkage analysis, using one of the many sophisticated and available approaches that are reviewed elsewhere (Bailey-Wilson and Wilson 2011). A linkage analysis provides a suitable and well-calibrated statistical framework for evaluation of the role of a candidate variant, as demonstrated in an analysis of lipoprotein lipase gene variants and LDL size (Hokanson et al. 1999), or a more recent evaluation of a variant identified through high-throughput sequencing that is implicated in Charcot-Marie-Tooth disease (Weedon et al. 2011).

Family-based designs enrich for variants of interest. For example, selection of subjects from families with multiple affected subjects, or selection of subjects from families with extreme values, can enrich for multiple copies of a rare variant (Gorlov et al. 2011; Ionita-Laza and Ottman 2011), providing improved ability to measure, and detect, the effects of such variants. A contrast of two recent evaluations of several candidate genes for late-onset Alzheimer's disease provides a useful illustration (Cruchaga et al. 2012; Gerrish et al. 2012). In both studies, genes identified through their role in early-onset AD and dementia were evaluated, with the same four genes evaluated in both studies. In the family-based study, investigators were able to use a highly efficient pooled-sequencing approach in 439 unrelated probands from late-onset families with multiple affected subjects/family (Cruchaga et al. 2012). This resulted in identification of a statistically significant excess of rare variants in these genes in the probands relative to unrelated controls, with follow-up co-segregation with the trait in the families used as part of the evaluation. In contrast, a population-based case–control sample was ∼40 times larger, consisting of 17,313 subjects (Gerrish et al. 2012), but with no preference for family history. Even a targeted analysis of the four candidate genes produced no evidence for association with SNPs in the genes tested.

Complexities handled

Large pedigrees provide some protection against the highly deleterious effects of heterogeneity. This is a key strategy to obtain interpretable results with relatively modest sample sizes, even though genetic heterogeneity is inevitable for complex traits (Sillanpaa and Auranen 2004). Large pedigrees can each be more homogeneous with respect to genetic variation than a combined sample of many smaller families, while providing enough information to detect linkage within individual large families. This strategy of using large families for initial screening has long been a highly successful strategy when there is high genetic heterogeneity, as illustrated by traits such as hereditary hearing loss (Varilo and Peltonen 2004) and dilated cardiomyopathy (Hershberger et al. 2010). The strategy has also been both proposed and used in the search for loci contributing to complex traits (Wright et al. 1999). This same strategy of using large pedigrees also worked well to identify loci for simulated trait data as part of Genetic Analysis Workshop 17 (Gagnon et al. 2011; Simpson et al. 2011; Wilson and Ziegler 2011) in which evidence for linkage was clearly identified in the large pedigrees, without a clear comparable signal in sequence data available for unrelated subjects.

Large pedigree data sets frequently include extensive and unusually complete phenotype information. Such large families are typically collected by investigators with a deep interest in the phenotypes, and develop strong rapport with the subjects. As a result, new phenotypes may be collected on the same participants over time, and may include expensive or unusual phenotypes. Examples include the well-known Framingham Heart Study (Jaquish 2007), the Strong Heart Study (Lee et al. 1990), and large families with familial combined hyperlipidemia under study at several institutions (Jarvik et al. 1994; Pajukanta et al. 2003; Rosenthal et al. 2011). Such samples represent a rich phenotypic resource that can be mined for biological insight and phenotypic subgroups. A rich phenotypic data set also can provide the opportunity to identify pre-clinical phenotypes through evaluation of biomarkers in unaffected relatives of cases with known primary mutations, as is being attempted in both Crohn's disease (Hedin et al. 2012) and early onset Alzheimer's disease (Morris 2011). All of these sources of information can provide a unique and informative perspective on the genotype–phenotype relationship that is impossible to achieve in huge studies that depend on combining measures that are jointly collected in large numbers across different sites.

Pedigrees facilitate error detection. For example, most sample swaps are not only easy to identify in pedigree genome scan data, but also can be corrected in silico (Boehnke and Cox 1997). Similarly, genotype, and even phenotype, errors are detectable through identification of low probability situations in meiotic transmissions

(Buetow 1991; Ehm et al. 1996) with improved detection derived from using more subjects per pedigree (Mukhopadhyay et al. 2004). The higher rate of error associated with high-throughput sequencing means that error detection methods are likely to increase in importance. Similar to in silico correction of pedigree error, it is likely that discrepant sequencing calls can be corrected by capitalizing on information about genomic segments shared identical-by-descent within pedigrees. Software for error detection associated with high-throughput sequencing in large pedigrees has not yet been released. However, proof-of-concept studies show that modern computational tools enable extension of these ideas to error detection with dense data in large pedigrees (Cheung et al. 2011; Markus et al. 2011), so that it is only a matter of time before useable software is available.

Challenges

Use of large pedigrees also introduces certain complications. Some of these involve requirements for additional required information, including the complete pedigree structure, marker allele frequencies, and, for any kind of multipoint analysis, a meiotic map. A sequence-based map is not a particularly good proxy (Wijsman et al. 2007), so invariably some interpolation onto an existing high-quality meiotic map is necessary (Matise et al. 2007). There are also particular challenges with using very dense genotype data, since close markers typically are in linkage disequilibrium (LD). This affects the prior probability distribution of haplotype frequencies, which, in turn, strongly affects analysis results (Ott 1992; Sieh et al. 2007). Attempts to model the LD genomewide in pedigree analysis have so far been unsatisfactory (Thomas 2007). Alternatives are to treat blocks of markers as non-recombining multi-allelic segments with empirically determined haplotype frequencies (Abecasis and Wigginton 2005; Sieh et al. 2007) or to thin markers sufficiently to avoid the problem of LD. The former option has not yet been implemented for large pedigrees, although it works reasonably well for smaller pedigrees. The latter is an option that can be used in most settings, as discussed further below.

Computation

Computations on large pedigrees are no longer the major bottleneck of the past. While exact computation remains impractical on large pedigrees with many markers (Thompson 2011), there are now excellent alternatives. This makes it possible to include new types of data, such as genotypes from high-throughput sequencing. Practical computation on large pedigrees is based on Markov chain Monte Carlo (MCMC) methods, and implementations have steadily improved since their inception (Sobel and Lange 1993; Thompson 1994). Existing packages, including SIMWALK2 (Sobel et al. 2001), MORGAN (Thompson 2005, 2011; Tong and Thompson 2008), as well as other programs (Thomas et al. 2000), provide excellent results for analysis of multipoint marker data for pedigrees without a large number of loops. These MCMC methods sample from the possible underlying inheritance configurations, with the sampled configurations typically used under a general analytic framework that can incorporate many different linkage statistics (Sobel and Lange 1996). Other MCMC-based approaches allow analysis with complex trait models that otherwise are computationally intractable (Daw et al. 1999; Heath 1997), allowing use of models that may be more useful for analysis of complex traits.

The accuracy of results obtained with some MCMC-based approaches has been extensively tested for a variety of simulated pedigree structures and marker configurations (Wijsman et al. 2006). For initial computations on pedigrees, the coarse nature of meiosis means that thinning markers to 1 SNP marker per cM is both acceptable as measured by little loss of inheritance information despite reduction in the data used, and is computationally advantageous (Wijsman et al. 2006). This is consistent with analysis of real data carried out with a variety of methods (Wilcox et al. 2005) that showed that such marker thinning did not cause substantial loss of information, even though some data were then ignored. SIMWALK2 and MORGAN have comparable performance on sparser marker panels, such as panels of multiallelic markers. However, for denser markers, e.g., SNPs or sequence-based markers, the computation time for MORGAN used at the time of comparison was two orders of magnitude faster than that of SIMWALK2 (Wijsman et al. 2006). More recent improvements have yielded even better performance (Tong and Thompson 2008).

One current program that also is likely to be useful in the context of modern genotyping data is gl_auto from the MORGAN package. This is the first distributed program that outputs the sampled inheritance patterns so that they can be used for a variety of analyses without requiring the end user to understand and modify the more complex MCMC-based programs. For example, the investigator can take advantage of the factoring of computations associated with the markers versus the trait (Sobel and Lange 1996), or can carry out several different analyses that all depend on the same sample of inheritance patterns, such as analyses of different traits, or with different analysis methods. An additional computational cost saving comes from the program IBDgraph (Koepke and Thompson 2010), which

recognizes equivalent configurations in the inheritance patterns so that redundant computations can be avoided: recent application to a moderately large Alzheimer's disease family sped up computations by a factor of 10 (Marchani and Wijsman 2011).

Information

The large amount of underlying variation creates challenges for use of data generated by high-throughput sequencing. The extent of this rare variation was first revealed by early applications of high-throughput sequencing (Ng et al. 2010; Roach et al. 2010). While samples of unrelated subjects are either severely underpowered or exceedingly expensive because of large required sample sizes, large pedigrees are uniquely suited for study of rare variants. Large pedigrees also allow efficient use of resources: in order to extract maximal information, sequence data on only a small number of subjects needs to be added to an existing genome scan marker data set, which typically either already exists or can be generated relatively cheaply. The two sources of data can easily be combined and interrogated with available statistical genetic analysis methods. The linkage analysis results can also be used as filters to effectively focus investigations on small segments of the genome (Musunuru et al. 2010; Smith et al. 2011; Wang et al. 2010), or as weights to incorporate results from different sources of information (Roeder et al. 2006). Use of large pedigrees, therefore, can prevent the twin problems of high cost and low power that are intrinsic to use of population-based studies of rare variants. For example, two recent studies were able to identify causal variants in pedigrees of sizes 24 and 42 by sequencing only 2–4 subjects per pedigree (Musunuru et al. 2010; Wang et al. 2010), and by combining this information with existing linkage panel markers.

Most early successful applications of high-throughput sequencing methods have demonstrated usefulness of high-throughput sequencing through application to rare Mendelian disorders. The earliest such applications focused on a handful of mostly unrelated subjects, but were feasible only because they combined a focus on exquisitely rare Mendelian recessive traits with a series of ad hoc but reasonable filters to nominate genes and variants based on predicted function and frequency (Ng et al. 2009, 2010; Roach et al. 2010). This strategy fails in situations that are typical of complex traits where (1) there is genetic and etiological heterogeneity among subjects, (2) the traits are continuous, (3) the frequency of relevant variants may be greater than vanishingly rare, and (4) the effects of influential variants are unlikely to be easy to predict based on sequence information. In these situations, other analysis approaches and/or filtering strategies are necessary. These are found in the repertoire of existing analytical approaches for pedigree data.

Use of linkage analysis is an effective filtering strategy. Relative to older technologies, positional sequencing with current technologies is much faster, provides much more extensive data on a region, and provides a more comprehensive evaluation of regional variation. As a result, the whole region implicated by a positive linkage signal is now easily sequenced, and it is no longer necessary to accrue a large sample just for the purpose of narrowing a region of interest (Boehnke 1994). Linkage analysis results can therefore be used as a filter to focus on a comprehensive analysis of all or most variants in the region with evidence of linkage (Smith et al. 2011). For traits with a clear mode of inheritance, the region of interest would presumably be that bounded by obligate recombinants. For more complex traits, other criteria may need to be determined. This presumably would include greater weight given to variants in regions with the strongest evidence of linkage than in more peripheral regions, although functional annotation or other biologically based information could also be used, in principle, to prioritize variants. With significant evidence of linkage, it is also only necessary to sequence a small number of carefully chosen subjects per pedigree, preferably with information from the linkage analysis results guiding the choice of subjects (Marchani and Wijsman 2011). This strategy has proven to be efficient for identifying causal mutations for a rapidly increasing number of Mendelian genetic diseases and traits, including some that are found in relatively large pedigrees (Southgate et al. 2011; Weedon et al. 2011), and some of which represent exceedingly rare diseases in large pedigrees (Raskind et al. 2009, 2011). Also, by focusing analysis on a limited number of variants defined by a particular region, this strategy does not require the use of imperfect functional prediction tools that may miss key variants (Erlich et al. 2011), and allows interrogation of all or most variants in a region.

The same general strategy of sequencing small numbers of individuals in key pedigrees also yields genes that affect complex traits. Several studies have reported gene identification in the context of Mendelian traits that represent very high levels of genetic heterogeneity. Each study focused on regions with significant linkage evidence in single large pedigrees, including ataxia (Wang et al. 2010), familial dilated cardiomyopathy (Norton et al. 2011), and thoracic aortic aneurism (Regalado et al. 2011). Some of the earliest successful studies of more complex, multilocus traits also started with one or two large pedigrees, with subsequent focus on genes in region(s) with significant evidence of linkage in those pedigrees. These studies include: (1) a variant in affecting plasma adiponectin levels

identified as a candidate in a region with linkage evidence (Bowden et al. 2010), (2) two studies that also used quantitative trait linkage analysis and followed this with a comprehensive measured genotype approach of all variants in the region identified from sequencing (Musunuru et al. 2010; Rosenthal et al. 2011), and (3) one study with a focus on a candidate gene, but with comprehensive sequencing and a measured genotype approach to evaluate all variants in the gene (Calafell et al. 2010). In these latter studies, the measured genotype approach (Almasy and Blangero 2004) was feasible because of the restricted region of interest: in this context, it is computationally feasible to every variant in a region for its ability to explain the segregating variance. Also, it is important to note that none of these studies required sequencing of large number of subjects.

The measured genotype approach is particularly useful for evaluating effects of variants in large pedigrees, and essentially combines information from both association and linkage. The approach efficiently extracts information about sequence-based variants through two outcomes. First, there is an inherent increase in information obtained by having multiple genotyped copies of a single rare variant in a pedigree. This occurs naturally when a risk variant drives a linkage signal because of co-segregation with the trait. Second, by combining sequence data with existing marker data in large pedigrees, it is possible to impute variants into other individuals who have phenotype, but at most partial genotype data, further increasing the information available for evaluating the effects of variants. Such imputation can be carried out jointly with the linkage analysis (Heath 1997; Rosenthal et al. 2011; Wijsman et al. 2010), but this can be computationally intensive. Alternatively, it is possible to use the existing data on the pedigree to impute variants with a genotype calling algorithm, similar to approaches used for genotype imputation in unrelated subjects (Li et al. 2009). In the context of pedigrees, the shared inheritance of chromosomal segments within a pedigree provides the correlation needed for imputation. Methods that provide imputed genotypes in pedigrees already exist for pedigrees that are small enough for exact computation (Burdick et al. 2006). Similar methods that accommodate much larger pedigrees are also now becoming available (Cheung et al. 2010). These methods sample inheritance vectors (IVs) within pedigrees (Thompson 2011) obtained from analysis with a framework panel of markers. The IVs are obtained with either exact computation or Markov chain Monte Carlo (MCMC) by, e.g., gl_auto in MORGAN. Deterministic (Wijsman 1987) followed by probabilistic inference then allows computationally rapid imputation of the dense sequence-based variants in the remaining subjects, with a pre-defined, and tunable, error rate. Although it would be inappropriate to use them for linkage analysis, the imputed genotypes may

be used for, e.g., a computationally rapid analysis approach, such as a variance components analysis (Almasy and Blangero 2004) to determine which variant(s) explain the segregating variance, followed by targeted genotyping of key variants in more subjects.

Conclusions

We have argued here that large pedigrees should be one of the designs for which high-throughput sequence data should be generated, particularly when rare variants are suspected to play a role in disease risk. Large pedigrees are uniquely suited for the study of rare variants, and can provide statistically significant evidence of both co-segregation with the disease or trait and of existence of genotype–phenotype association. Also, a small amount of sequencing added to an otherwise well-characterized large pedigree can yield considerable information even on subjects with no sequence data, thus making this approach highly cost effective. Although the goal here was not to provide a comprehensive evaluation of existing software, appropriate computational tools and statistical methods already exist for analysis of large pedigrees, so there is no major practical barrier to data analysis. In the search for the genetic basis of complex traits, it is important to use the best available tools, and not to eschew well-established designs and analysis methods for lack of novelty. This will enable the information in the sequence data to be efficiently obtained and used by drawing upon existing methods with well-understood properties. As a result, the quality of the resulting scientific inference obtained from the use of novel genotyping technologies will be maximized.

References

Abecasis GR, Wigginton JE (2005) Handling marker–marker linkage disequilibrium: pedigree analysis with clustered markers. Am J Hum Genet 77:754–767

Almasy L, Blangero J (2004) Exploring positional candidate genes: linkage conditional on measured genotype. Behav Genet 34:173–177

Amberger J, Bocchini C, Hamosh A (2011) A new face and new challenges for Online Mendelian Inheritance in man (OMIM (R)). Hum Mutat 32:564–567

Bailey-Wilson JE, Wilson AF (2011) Linkage analysis in the next-generation sequencing era. Hum Hered 72:228–236

Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40:695–701

Boehnke M (1994) Limits of resolution of genetic linkage studies: implications for positional cloning of human disease genes. Am J Hum Genet 55:379–390

Boehnke M, Cox N (1997) Accurate inference of relationships in sib-pair linkage studies. Am J Hum Genet 61:423–429

Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33:228–237

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

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 identify a rare genetic variant in the ADIPOQ gene in the IRAS family study. Hum Mol Genet 19:4112–4120

Buetow K (1991) Influence of aberrant observations on high-resolution linkage analysis outcomes. Am J Hum Genet 49:985–994

Burdick JT, Chen WM, Abecasis GR, Cheung VG (2006) In silico method for inferring genotypes in pedigrees. Nat Genet 38:1002–1004

Calafell F, Almasy L, Sabater-Lleal M, Buil A, Mordillo C, Ramirez-Soriano A, Sikora M, Souto JC, Blangero J, Fontcuberta J, Soria JM (2010) Sequence variation and genetic evolution at the human F12 locus: mapping quantitative trait nucleotides that influence FXII plasma levels. Hum Mol Genet 19:517–525

Cheung CYK, Thompson EA, Wijsman EM (2010) In silico genotype imputation on large pedigrees. Genet Epidemiol 34:919

Cheung CYK, Wijsman EM, Thompson EA (2011) Detection of genotype errors in dense markers on large pedigrees. International Congress of Human Genetics, Montreal

Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305:869–872

Collins FS (1991) Of needles and haystacks—finding human-disease genes by positional cloning. Clin Res 39:615–623

Collins FS, Guyer MS, Chakravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581

Cruchaga C, Chakraverty S, Mayo K, Vallania FL, Mitra RD, Faber K, Williamson J, Bird T, Diaz-Arrastia R, Foroud TM, Boeve BF, Graff-Radford NR, St Jean P, Lawson M, Ehm MG, Mayeux R, Goate AM (2012) Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. PLoS One 7:e31039

Daw EW, Heath SC, Wijsman EM (1999) Multipoint oligogenic analysis of age-at-onset data with applications to Alzheimer's disease pedigrees. Am J Hum Genet 64:839–851

Ehm MG, Kimmel M, Cottingham RW (1996) Error detection for genetic data, using likelihood methods. Am J Hum Genet 58:225–234

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Erlich Y, Edvardson S, Hodges E, Zenvirt S, Thekkat P, Shaag A, Dor T, Hannon GJ, Elpeleg O (2011) Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. Genome Res 21:658–664

Flint J, Mackay TFC (2009) Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res 19:723–733

Gagnon F, Roslin NM, Lemire M (2011) Successful identification of rare variants using oligogenic segregation analysis as a prioritizing tool for whole-exome sequencing studies. BMC Proc 5:S11

Gerrish A, Russo G, Richards A, Moskvina V, Ivanov D, Harold D, Sims R, Abraham R, Hollingworth PCJ, Hamshere M, Pahwa JS, Dowzell K, Williams A, Jones N, Thomas C, Stretton A, Morgan AR, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Morgan K, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Johnston JA, Holmes C, Mann D, Smith AD, Love S, Kehoe PG, Hardy J, Mead S, Fox N, Rossor M, Collinge J, Maier W, Jessen F, Kölsch H, Heun R, Schürmann B, Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Hüll M, Rujescu D, Goate AM, Kauwe JS, Cruchaga C, Nowotny P, Morris JC, Mayo K, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Davies G, Harris SE, Starr JM, Deary IJ, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel KH, Klopp N, Wichmann HE, Carrasquillo MM, Pankratz VS, Younkin SG, Jones L, Holmans PA, O'Donovan MC, Owen MJ, Williams J (2012) The role of variation at AβPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease. J Alzheimers Dis 28:377–387

Gorlov IP, Gorlova OY, Frazier ML, Spitz MR, Amos CI (2011) Evolutionary evidence of the effect of rare variants on disease etiology. Clin Genet 79:199–206

Haldane JBS, Smith CAB (1947) A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann Eugen 14:10–31

Heath SC (1997) Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61:748–760

Hedin CR, Stagg AJ, Whelan K, Lindsay JO (2012) Family studies in Crohn's disease: new horizons in understanding disease pathogenesis, risk and prevention. Gut 61:311–318

Hershberger RE, Morales A, Siegfried JD (2010) Clinical and genetic issues in dilated cardiomyopathy: a review for genetics professionals. Genet Med 12:655–667

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 Nat Acad Sci USA 106:9362–9367

Hokanson JE, Brunzell JD, Jarvik GP, Wijsman EM, Austin MA (1999) Linkage of low-density lipoprotein size to the lipoprotein lipase gene in heterozygous lipoprotein lipase deficiency. Am J Hum Genet 64:608–618

Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Ionita-Laza I, Ottman R (2011) Study designs for identification of rare disease variants in complex diseases: the utility of family-based designs. Genetics 189:1061–U500

Jaquish CE (2007) The Framingham Heart Study, on its way to becoming the gold standard for cardiovascular genetic epidemiology? BMC Med Genet 8:63

Jarvik GP, Brunzell JD, Austin MA, Krauss RM, Motulsky AG, Wijsman EM (1994) Genetic predictors of FCHL in four large pedigrees: influence of ApoB level major locus predicted genotype and LDL subclass phenotype. Arterioscler Thromb Vasc Biol 14:1687–1694

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389

Koepke H, Thompson EA (2010) Efficient testing operations on dynamic graph structures using strong hash functions. Department of Statistics, technical reports. University of Washington, Seattle

Lee ET, Welty TK, Fabsitz R, Cowan LD, Le NA, Oopik AJ, Cucchiara AJ, Savage PJ, Howard BV (1990) The Strong Heart Study—a study of cardiovascular-disease in American–Indians—design and methods. Am J Epidemiol 132:1141–1155

Leibon G, Rockmore DN, Pollak MR (2008) A SNP streak model for the identification of genetic regions identical-by-descent. Stat Appl Genet Mol Biol 7:16

Leigh SEA, Foster AH, Whittall RA, Hubbart CS, Humphries SE (2008) Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Ann Hum Genet 72:485–498

Levy-Lahad E, Wasco W, Poorkaj P, Romano DM, Oshima J, Pettingell WH, Yu CE, Jondro PD, Schmidt SD, Wang K, Crowley AC, Fu YH, Guenette SY, Galas D, Nemens E, Wijsman EM, Bird TD, Schellenberg GD, Tanzi RE (1995) Candidate gene for the chromosome 1 familial Alzheimer's disease locus. Science 269:973–977

Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10:387–406

Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118:1590–1605

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 TFC, McCarroll SA, Visscher PM (2009) Finding the missing heritability of complex diseases. Nature 461:747–753

Marchani EE, Wijsman EM (2011) Estimation and visualization of identity-by-descent within pedigrees simplifies interpretation of complex trait analysis. Hum Hered 72:289–297

Markus B, Birk OS, Geiger D (2011) Integration of SNP genotyping confidence scores in IBD inference. Bioinformatics 27:2880–2887

Matise TC, Chen F, Chen WW, De la Vega FM, Hansen M, He CS, Hyland FCL, Kennedy GC, Kong XY, Murray SS, Ziegle JS, Stewart WCL, Buyske S (2007) A second-generation combined linkage-physical map of the human genome. Genome Res 17:1783–1786

McClellan J, King MC (2010) Genetic heterogeneity in human disease. Cell 141:210–217

Miki Y, Swensen J, Shattuckeidens D, Futreal PA, Harshman K, Tavtigian S, Liu QY, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugenstrano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering L, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannonalbright L, Goldgar D, Wiseman R, Kamb A, Skolnick MH (1994) A strong candidate for the breast and ovarian-cancer susceptibility gene BRCA1. Science 266:66–71

Morris JC (2011) Dominantly Inherited Alzheimer Network (DIAN): registry characteristics and biomarker findings. Neurology 76:A416–A416

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318

Mukhopadhyay N, Buxbaum SG, Weeks DE (2004) Comparative study of multipoint methods for genotype error detection. Hum Hered 58:175–189

Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ, Fennell T, Banks E, Ambrogio L, Cibulskis K, Kernytsky A, Gonzalez E, Rudzicz N, Engert JC, DePristo MA, Daly MJ, Cohen JC, Hobbs HH, Altshuler D, Schonfeld G, Gabriel SB, Yue P, Kathiresan S (2010) Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. N Engl J Med 363:2220–2227

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461:272–U153

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–U41

Norton N, Li DX, Rieder MJ, Siegfried JD, Rampersaud E, Zuchner S, Mangos S, Gonzalez-Quintana J, Wang LB, McGee S, Reiser J, Martin E, Nickerson DA, Hershberger RE (2011) Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. Am J Hum Genet 88:273–282

Ott J (1974) Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet 26:588–597

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283–290

Pajukanta P, Allayee H, Krass KL, Kuraishy A, Soro A, Lilja HE, Mar R, Taskinen MR, Nuotio I, Laakso M, Rotter JI, de Bruin TWA, Cantor RM, Lusis AJ, Peltonen L (2003) Combined analysis of genome scans of Dutch and Finnish families reveals a susceptibility locus for high-density lipoprotein cholesterol on chromosome 16q. Am J Hum Genet 72:903–917

Penrose LS (1935) The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. Ann Eugen 6:133–138

Raskind WH, Matsushita M, Peter B, Biberston J, Wolff J, Lipe H, Burbank R, Bird TD (2009) Familial dyskinesia and facial myokymia (FDFM): follow-up of a large family and linkage to chromosome 3p21–3q21. Am J Med Genet Part B Neuropsychiat Genet 150B:570–574

Raskind WH, Chen YZ, Matsushita MM, Robertson PD, Rieder M, Girirajan S, Lipe H, Eichler EE, Nickerson DA, Bird TD (2011) Linkage and single exome analyses identify ADCY5 as the gene for familial dyskinesia with facial myokymia. Int Congr Hum Genet, Montreal

Regalado ES, Guo DC, Villamizar C, Avidan N, Gilchrist D, McGillivray B, Clarke L, Bernier F, Santos-Cortez RL, Leal SM, Bertoli-Avella AM, Shendure J, Rieder MJ, Nickerson DA, Milewicz DM (2011) Exome sequencing identifies SMAD3 mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. Circ Res 109:680–U220

Riordan JR, Rommens JM, Kerem B-S, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou J-L, Drumm ML, Iannuzzi MC, Collins FS, Tsui L-C (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245:1066–1073

Risch N (2000) Searching for genetic determinants in the new millennium. Nature 405:847–856

Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ (2010) Analysis of genetic inheritance in a family Quartet by whole-genome sequencing. Science 328:636–639

Roeder K, Bacanu SA, Wasserman L, Devlin B (2006) Using linkage genome scans to improve power of association in genome scans. Am J Hum Genet 78:243–252

Rosenthal EA, Ronald J, Rothstein J, Rajagopalan R, Ranchalis J, Wolfbauer G, Albers JJ, Brunzell JD, Motulsky AG, Rieder MJ, Nickerson DA, Wijsman EM, Jarvik GP (2011) Linkage and association of phospholipid transfer protein activity to LASS4. J Lipid Res 52:1837–1846

Sanna S, Li BS, Mulas A, Sidore C, Kang HM, Jackson AU, Piras MG, Usala G, Maninchedda G, Sassu A, Serra F, Palmas MA, Wood WH, Njolstad I, Laakso M, Hveem K, Tuomilehto J, Lakka TA, Rauramaa R, Boehnke M, Cucca F, Uda M, Schlessinger D, Nagaraja R, Abecasis GR (2011) Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. Plos Genet 7:e1002198

Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, Chi H, Lin C, Li G, Holman K, Tsuda T, Mar L, Foncin JF,

Bruni AC, Montesi MP, Sorbi S, Rainero I, Pinessi L, Nee L, Chumakov I, Pollen D, Brookes A, Sanseau P, Polinsky RJ, Wasco W, Dasilva HAR, Haines JL, Pericak-Vance MA, Tanzi RE, Roses AD, Fraser PE, Rommens JM, St George-Hyslop PH (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature 375:754–760

Sieh W, Yu C-E, Bird TD, Schellenberg GD, Wijsman EM (2007) Accounting for linkage disequilibrium among markers in linkage analysis: impact of haplotype frequency estimation and molecular haplotypes for a gene in a candidate region for Alzheimer's disease. Hum Hered 63:26–34

Sillanpaa MJ, Auranen K (2004) Replication in genetic studies of complex traits. Ann Hum Genet 68:646–657

Simpson CL, Justice CM, Krishnan M, Wojciechowski R, Sung H, Cai J, Green T, Lewis D, Behneman D, Wilson AF, Bailey-Wilson JE (2011) Old lessons learned anew: family-based methods for detecting genes responsible for quantitative and qualitative traits in the Genetic Analysis Workshop 17 mini-exome sequence data. BMC Proc 5:S83

Smith KR, Bromhead CJ, Hildebrand MS, Shearer AE, Lockhart PJ, Najmabadi H, Leventer RJ, McGillivray G, Amor DJ, Smith RJ, Bahlo M (2011) Reducing the exome search space for Mendelian diseases using genetic linkage analysis of exome genotypes. Genome Biol 12:R85

Sobel E, Lange K (1993) Metropolis sampling in pedigree analysis. Stat Methods Med Res 2:263–282

Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58:1323–1337

Sobel E, Sengul H, Weeks DE (2001) Multipoint estimation of identity-by-descent probabilities at arbitrary positions among marker loci on general pedigrees. Hum Hered 52:121–131

Southgate L, Machado RD, Snape KM, Primeau M, Dafou D, Ruddy DM, Branney PA, Fisher M, Lee GJ, Simpson MA, He Y, Bradshaw TY, Blaumeiser B, Winship WS, Reardon W, Maher ER, FitzPatrick DR, Wuyts W, Zenker M, Lamarche-Vane N, Trembath RC (2011) Gain-of-function mutations of ARHGAP31, a Cdc42/Rac1 GTPase regulator, cause syndromic cutis aplasia and limb anomalies. Am J Hum Genet 88:574–585

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Allen HL, Lindgren CM, Luan J, Magi R, Randall JC, Vedantam S, Winkler TW, Qi L, Workalemahu T, Heid IM, Steinthorsdottir V, Stringham HM, Weedon MN, Wheeler E, Wood AR, Ferreira T, Weyant RJ, Segre AV, Estrada K, Liang LM, Nemesh J, Park JH, Gustafsson S, Kilpelanen TO, Yang JA, Bouatia-Naji N, Esko T, Feitosa MF, Kutalik Z, Mangino M, Raychaudhuri S, Scherag A, Smith AV, Welch R, Zhao JH, Aben KK, Absher DM, Amin N, Dixon AL, Fisher E, Glazer NL, Goddard ME, Heard-Costa NL, Hoesel V, Hottenga JJ, Johansson A, Johnson T, Ketkar S, Lamina C, Li SX, Moffatt MF, Myers RH, Narisu N, Perry JRB, Peters MJ, Preuss M, Ripatti S, Rivadeneira F, Sandholt C, Scott LJ, Timpson NJ, Tyrer JP, van Wingerden S, Watanabe RM, White CC, Wiklund F, Barlassina C, Chasman DI, Cooper MN, Jansson JO, Lawrence RW, Pellikka N, Prokopenko I, Shi JX, Thiering E, Alavere H, Alibrandi MTS, Almgren P, Arnold AM, Aspelund T, Atwood LD, Balkau B, Balmforth AJ, Bennett AJ, Ben-Shlomo Y, Bergman RN, Bergmann S, Biebermann H, Blakemore AIF, Boes T, Bonnycastle LL, Bornstein SR, Brown MJ, Buchanan TA et al (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42:937–U53

Sturtevant AH (1913) The linear arrangement of six sex-linked factors in Drosophila as shown by their mode of association. J Exp Zool 14:43–59

Thoday JM (1961) Location of polygenes. Nature 191:368–370

Thomas A (2007) Towards linkage analysis with markers in linkage disequilibrium by graphical modelling. Hum Hered 64:16–26

Thomas A, Gutin A, Abkevich V, Bansal A (2000) Multilocus linkage analysis by blocked Gibbs sampling. Stat Comput 10:259–269

Thomas A, Camp NJ, Farnham JM, Allen-Brady K, Cannon-Albright LA (2008) Shared genomic segment analysis. Mapping disease predisposition genes in extended pedigrees using SNP genotype assays. Ann Hum Genet 72:279–287

Thompson EA (1994) Monte Carlo likelihood in the genetic mapping of complex traits. Philos Trans R Soc B 344:345–351

Thompson EA (2005) MCMC in the analysis of genetic data on pedigrees. In: Kendall WS, Wang JS, Liang F (eds) Markov chain Monte Carlo: innovations and applications. World Scientific Publishing Company, Singapore

Thompson EA (2011) The structure of genetic linkage data: from LIPED to 1M SNPs. Hum Hered 71:86–96

Tong LP, Thompson E (2008) Multilocus lod scores in large pedigrees: combination of exact and approximate calculations. Hum Hered 65:142–153

Varilo T, Peltonen L (2004) Isolates and their potential use in complex gene mapping efforts—commentary. Curr Opin Genet Dev 14:316–323

Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS (2010) TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 133:3510–3518

Weedon MN, Hastings R, Caswell R, Xie WJ, Paszkiewicz K, Antoniadi T, Williams M, King C, Greenhalgh L, Newbury-Ecob R, Ellard S (2011) Exome sequencing identifies a DYNC1H1 mutation in a large pedigree with dominant axonal Charcot-Marie-Tooth disease. Am J Hum Genet 89:308–312

Wijsman EM (1987) A deductive method of haplotype analysis in pedigrees. Am J Hum Genet 41:356–373

Wijsman EM, Amos CI (1997) Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. Genet Epidemiol 14:719–735

Wijsman EM, Rothstein JH, Thompson EA (2006) Multipoint linkage analysis with many multiallelic or dense diallelic markers: MCMC provides practical approaches for genome scans on general pedigrees. Am J Hum Genet 79:846–858

Wijsman EM, Sung YJ, Buil A (2007) Summary of GAW15: group 9 linkage analysis of the CEPH expression data. Genet Epidemiol 31:S75–S85

Wijsman EM, Rothstein JH, Igo RP, Brunzell JD, Motulsky AG, Jarvik GP (2010) Linkage and association analyses identify a candidate region for apoB level on chromosome 4q32.3 in FCHL families. Hum Genet 127:705–719

Wilcox MA, Pugh EW, Zhang H, Zhong X, Levinson DF, Kennedy GC, Wijsman EM (2005) Comparison of single-nucleotide polymorphisms and microsatellite markers for linkage analysis in the COGA and simulated data sets for Genetic Analysis Workshop 14: presentation groups 1, 2, and 3. Genet Epidemiol 29(Suppl 1):S7–S28

Wilson AF, Ziegler A (2011) Lessons learned from Genetic Analysis Workshop 17: transitioning from genome-wide association studies to whole-genome statistical genetic analysis. Genet Epidemiol 35:S107–S114

Wright AF, Carothers AD, Pirastu M (1999) Population choice in mapping genes for complex diseases. Nat Genet 23:397–404

Yazbek SN, Buchner DA, Geisinger JM, Burrage LC, Spiezio SH, Zentner GE, Hsieh CW, Scacheri PC, Croniger CM, Nadeau JH (2011) Deep congenic analysis identifies many strong, context-dependent QTLs, one of which, Slc35b4, regulates obesity and glucose homeostasis. Genome Res 21:1065–1073