Published in final edited form as:

Genetica. 2009 June ; 136(2): 333–340. doi:10.1007/s10709-008-9305-3.

Human QTL Linkage Mapping

Laura Almasy and John Blangero

Department of Genetics, Southwest Foundation for Biomedical Research

Abstract

Human quantitative trait locus (QTL) linkage mapping, although based on classical statistical genetic methods that have been around for many years, has been employed for genome-wide screening for only the last 10-15 years. In this time, there have been many success stories, ranging from QTLs that have been replicated in independent studies to those for which one or more genes underlying the linkage peak have been identified to a few with specific functional variants that have been confirmed in *in vitro* laboratory assays. Despite these successes, there is a general perception that linkage approaches do not work for complex traits, possibly because many human QTL linkage studies have been limited in sample size and have not employed the family configurations that maximize the power to detect linkage. We predict that human QTL linkage studies will continue to be productive for the next several years, particularly in combination with RNA expression level traits that are showing evidence of regulatory QTLs of large effect sizes and in combination with high-density genome-wide SNP panels. These SNP panels are being used to identify QTLs previously localized by linkage and linkage results are being used to place informative priors on genome-wide association studies.

Keywords

genome screen; Haseman-Elston; IBD; quantitative traits; variance components

The methods are old the results are new

The basic statistical genetic methods that are used today for human quantitative trait locus (QTL) linkage mapping are extensions of classic models that were available twenty years ago at the second International Conference on Quantitative Genetics (ICQG2). The methods currently used for human QTL linkage mapping derive from classical variance component methods(1) and from the sibling pair regression methods proposed by Haseman and Elston (2). These methods are based on the simple intuition that relatives who are more alike phenotypically should be more likely to share alleles at the QTL (and at nearby linked markers) than relatives who are phenotypically discordant. Identity by descent (IBD) allele sharing is used to quantify the proportion of alleles a relative pair shares at a particular location in the genome that are copies of the same ancestral chromosome. In the simplest case of a sample of independent sibling pairs, the original Haseman-Elston linkage method proposed that the squared difference in the siblings' phenotypes can be regressed on IBD with evidence of linkage coming from a negative regression slope indicating that siblings who are relatively more discordant share fewer alleles at a marker than siblings who are relatively more phenotypically concordant. The variance component (1,3) and the newer revised Haseman-Elston methods (4,5) are somewhat more complex, allowing for the non-independence of multiple relative pairs from the same pedigree and incorporating polygenic background effects, effects of measured

environmental covariates, and gene-gene and gene-environment interactions. The variance component and revised Haseman-Elston linkage approaches have been shown to be asymptotically equivalent, with the maximum likelihood-based variance component approaches being slightly more powerful and the regression-based Haseman-Elston being more robust to non-normality in the trait distribution(4). Newer incarnations of the same basic model use Markov Chain Monte Carlo (MCMC) methods for parameter estimation and incorporate the number of QTLs as an additional parameter(6).

Despite the fact that the basic statistical genetic methods for human QTL linkage were in place at the time of ICQG2, linkage studies for loci influencing quantitative traits in humans were still in their infancy and were generally limited to a few markers, usually in candidate genes. Complete genome-wide screens for QTLs in humans did not begin to appear until the mid-1990s(7-9). The post-ICQG2 advances that have made whole genome QTL linkage mapping possible have been primarily technological. In 1987, a genetic map of human chromosome 7 was published with 63 two-allele restriction fragment length polymorphism (RFLP) markers spaced roughly every 3 cM and this was one of only a few chromosomes with complete coverage in a linkage map(10). Today, we have dense maps of highly polymorphic short tandem repeat (STR) markers and even more dense maps of less polymorphic single nucleotide polymorphisms (SNPs) with known order and map position. The UCSC Genome Browser (<http://genome.ucsc.edu>) currently shows > 2,200 known STRs on chromosome 7 and > 670,000 SNPs. Additionally, the methods for genotyping these markers have become increasingly efficient, both in terms of cost per marker and in terms of the time to genotype each marker, with high-throughput genotyping technologies contributing to both of these factors. Computing, a rate limiting resource in the past, has also become faster and cheaper.

Although the basic statistical methods for assessing evidence of linkage in human pedigree data were in place 20 years ago, there have been some developments in statistical methods relating to estimation of IBD allele sharing using multiple markers in large pedigrees. The Elston-Stewart algorithm, used in the LINKAGE package, and the Lander-Green algorithm, used in programs such as Merlin, Allegro, and GeneHunter, each provide analytical formulae for estimating IBD allele sharing using multiple genotyped markers and properly weighting over all possible inheritance vectors. However, each of these algorithms has a limitation. The computing time for the Elston-Stewart approach is exponential with the number of markers but linear in the number of individuals in a pedigree, making it suitable for very large families but limiting the number of markers that can be used, an unsatisfactory state of affairs given the dense sets of STRs and SNPs available today. The Lander-Green algorithm, on the other hand, is linear in computing time with the number of markers but exponential with increasing pedigree size. Thus is suitable and widely used for studies with smaller families, but as larger pedigrees increase power to detect linkage, there was a need for programs that would facilitate the analysis of a dense map of markers in pedigrees of unlimited size and complexity.

Two main approaches have emerged for dealing with IBD estimation with an unlimited number of markers in pedigrees of unlimited size. The first of these involved a linear function of the IBDs at individual markers with the function depending on the probability of recombination between the markers, given their distance from each other and the type of relative pair involved. This was first proposed by Fulker et al for sibling pairs(11) and later expanded by Almasy and Blangero for larger pedigrees(12). Although this approach provided improved power and more accurate localization than sequential analyses of single markers, it was limited by the information content provided by each of the markers individually (i.e. better for relatively informative STRs than for less heterozygous markers such as SNPs) and has been for the most part abandoned in favor of more computer intensive, but more accurate, MCMC-based IBD estimation implemented in programs such as SimWalk2(13) and Loki(14). MCMC approaches to IBD estimation with multiple markers are based on sampling the space of possible

inheritance vectors that are consistent with the observed genotypes in a family and estimating IBD allele sharing for each relative pair by weighting the IBD value for each inheritance vector by the likelihood of that inheritance vector given the observed genotypes.

Human QTL linkage mapping works – success stories

There are numerous examples in recent years of successful human QTL linkage mapping studies. A number of these findings have been replicated in independent samples, a handful have been followed-up to the point of identifying a specific gene within the QTL linkage region, some of these have implicated specific variants within the gene as being potentially functional, and a few of those variants have been confirmed by *in vitro* functional assays.

One of the first widely replicated human QTL linkages is a peak for obesity-related traits on chromosome 2p. It was initially reported by Comuzzie et al. in 1997 as a LOD score of 4.95 for levels of leptin, an adipocyte-derived hormone, with weaker evidence for linkage of a QTL influencing fat mass to the same location in a sample of Mexican American families(15). Linkage of a QTL influencing leptin levels to chromosome 2p markers has since been replicated in French(16) and African American populations(17) and linkage to this region has also been observed for the related phenotype of body mass index(18,19). Despite the strong linkage signals and the replication of this peak in multiple independent populations, identification of the gene(s) underlying this QTL has proved elusive. An obvious positional candidate under the linkage peaks, the pro-opiomelanocortin (POMC) gene, was resequenced and two rare coding changes with minor allele frequencies of < 0.01 were identified. There was also good evidence for a third potentially functional variant, an insertion/deletion polymorphism in an intron. However, these variants only partially explain the observed linkage signals, suggesting that a second as yet unidentified gene may also contribute to this QTL(20,21).

Some replicated QTL linkages have been followed up to the point of identification of specific associated genes under the linkage peak, some of which are now being replicated in genome-wide association screens. Duggirala and colleagues reported linkage of a QTL influencing age at onset of type 2 diabetes and diabetes itself, analyzed using a variance component-based liability threshold model, to chromosome 10q(22). Another group found confirmatory linkage evidence in the same chromosomal region(23) and eventually identified the *TCF7L2* gene under their linkage peak as being strongly associated with type 2 diabetes(24). Association with *TCF7L2* has now been verified in the Mexican American sample in which the original linkage was reported(25). *TCF7L2* has since been picked up by multiple genome-wide association screens(26,27).

A number of human QTL linkages have not only genes, but also specific functional variants that have been identified and confirmed through *in vitro* functional studies. One of these is Factor VII (FVII), a protein involved in blood clotting that has been implicated in risk of venous and arterial sclerosis. A genome-wide linkage study identified a QTL for FVII on chromosome 13 with a LOD score of 3.18(28). The structural gene for FVII, *F7*, was directly under this linkage peak and was an obvious positional candidate. Association analyses of the 48 polymorphisms identified in the *F7* gene by sequencing in individuals from the linkage sample identified four variants with a very high posterior probability of effect and three additional variants with high support for functionality. Several of these variants were in putative promoter regions and their functionality has now been confirmed in *in vitro* expression assays(29).

Other well replicated human QTL linkages include loci influencing variation in triglyceride levels on chromosome 15 (30-33), body mass index on chromosome 3 (34-36), and reading disability on chromosome 6p (7,37-39) to name only a few. Another example of a replicated QTL linkage with a specific gene identified under the linkage peak by follow-up association studies is a chromosome 4 QTL influencing alcohol dependence and related endophenotypes

that show association with GABRA2(40-42). Further cases of specific functional variants that have been identified in genes originally localized through QTL linkages include three coding variants in a member of the TAS2R bitter taste receptor family that influence ability to taste phenylthiocarbamide (PTC)(43,44) and a promoter variant in the SEPS1 gene, in a region of chromosome 15 linked to a variety of inflammatory disorders, that affects inflammatory response(45).

Lessons from these studies

Although relatively few human QTL linkages have been followed up to the point of identifying the underlying gene(s) and functional variants, we may still be able to draw some conclusions about the spectrum of human QTLs from these examples. Given the sample size, these observations are necessarily anecdotal, but they provide some idea of what is possible within the range of human QTLs, if not a quantitative measure of how likely these possibilities are.

We may not be able to rely on picking an obvious positional candidate under a linkage peak. In the days of Mendelian penetrance model-based linkage, an initial linkage finding was generally followed up by fine mapping and genotyping of additional markers to more precisely define the recombinations within families and the minimal chromosomal area in which the gene of interest resided. Today, this fine mapping stage has largely been abandoned as the results of the Human Genome Project have made it easy to examine the catalog of genes in the region under a linkage peak to select one or a few obvious candidates to test as potential sources of a QTL linkage. Some of the known human QTL examples suggest that the obvious positional candidate gene may not be the correct one or that the obvious candidate may not be sufficient to explain the linkage in the case of multiple QTLs under a peak. As SNP genotyping becomes faster and cheaper, our reliance on choosing positional candidates is decreasing as it is more and more feasible to comprehensively screen all genes in the region under a linkage peak for association with the trait of interest. Currently, opinions differ on whether this association screening should be gene-centric, including known coding variants in the region, or should consist of markers of a uniform density selected to cover the whole region in a systematic fashion. This is likely to be a short-lived debate as sequencing technology becomes cheap enough to feasibly sequence the entire region under a linkage peak.

A linkage peak may represent the cumulative effects of more than one gene. For example, although there is strong evidence that the coding variants in the POMC gene affect leptin levels, these variants can not account for the linkage peak and there is also strong association with additional variants, not in linkage disequilibrium with the *POMC* markers, in another nearby gene.

There may be multiple functional variants, multiple quantitative trait nucleotides (QTNs), in each QTL, as was found with clotting Factor VII and the F7 gene, with ability to taste PTC and the TAS2R bitter taste receptor gene, and with leptin and the POMC gene. This allelic heterogeneity has implications for association studies. The power to find these genes by linkage involves the total variance in the trait explained by the QTL, which is the summed variance of the QTNs. The power to detect these genes by association, however, is a function of the individual variance attributable to a QTN and the strength of the linkage disequilibrium between that QTN and a genotyped marker. In the specific case of F7, the total effect size of the QTL was roughly 35%, but a single one of the QTNs accounted for roughly 14% of the trait variance and may have been easily detectable by association, particularly as this variant was common (minor allele frequency of 16.6%) and in strong linkage disequilibrium with other markers in the gene. However, in other cases, it may be easier to detect the QTL, the collective effect of multiple QTNs, by linkage than to detect any of the individual QTNs by association. For example, the insertion/deletion polymorphism in POMC had a minor allele frequency of

0.06 and both of the coding variants in this gene had frequencies of < 0.01 and low linkage disequilibrium with surrounding markers. However, collectively these three variants account for approximately 11% of the variation in leptin levels.

For the few human QTLs for which specific functional variants have been identified, it appears that the QTNs behind human QTLs consist of both rare and common variants and both coding and non-coding variants. In POMC and in F7 there were rare SNPs, some present in only a single family in the sample, that produced a large effect on each individual who carried them through changes in the amino acid structures of the proteins. In F7, there were also putative regulatory variants, some now confirmed through *in vitro* functional studies, with allele frequencies $> 20\%$.

Why is there a belief that linkage doesn't work for complex traits?

Despite the success of many genome-wide linkage scans for quantitative traits in humans, there is a common perception that although linkage analyses were excellent for localizing genes influencing simple Mendelian traits, they do not work for complex traits, a category that is usually defined as traits influenced by multiple genes and their interactions with each other and with the environment. And indeed, there are many examples of unsuccessful genome-wide quantitative trait linkage screens for a variety of complex phenotypes in humans.

The power of quantitative trait linkage methods depends on the proportion of variance due to the QTL being sought, the total heritability of the trait, the density and heterozygosity of the genotyped markers, the sample size, and the configuration of these individuals in families (46). The informativeness of the genotyped markers is generally a minor issue as there are well-established linkage maps in humans with a high density of microsatellites with very high heterozygosity or an even higher density of SNPs providing equivalent or better information about segregation within families. Additionally, the multipoint IBD methods discussed above have made it possible to estimate IBD allele sharing with a high degree of accuracy. Of the other factors relating to power, it is possible, or even likely, that some heritable quantitative traits of interest in humans represent the aggregate effects of many QTLs, none of which has an individual effect size large enough to be detected in linkage studies. However, for many human linkage studies, the limiting factor has been sample size or the family configurations available for study.

For a fixed sample size, the power to detect linkage for a quantitative trait is maximized when those individuals are concentrated into as few families as possible so that the number of relative pairs in the sample is maximized. For example, if I can afford a sample size of six individuals in my study, I could collect three families with two siblings each for a total of three relative pairs, two families with three siblings each for a total of six relative pairs, or one family of six siblings for a total of 15 relative pairs. Concentrating the individuals into fewer families dramatically increases the number of comparisons and contrasts that can be made among relatives in the study. Collecting large families can be challenging and the goal of six siblings per family may not be possible in many western populations. However, large sibship sizes are still the norm in some population groups. Alternatively, the same number of relative pairs can be obtained by collecting our six individuals in three sibling pairs, provided these three sibling pairs are cousins of each other. In this case, the result is still a total of 15 relative pairs in a single family.

Blangero et al(47) expressed this more formally, comparing the relative efficiency of quantitative trait linkage studies in families of different configurations. A relative efficiency of 1 was assigned to a linkage study of a population isolate in which all of the individuals studied coalesce into a single pedigree of 2000 individuals, an idealized situation but an attainable one given that it was based on the actual pedigree structure of an existing study.

Compared to this ideal, the relative power per person studied is only 0.04 for sibling pairs, 0.11 for nuclear families with three siblings examined, or 0.17 for nuclear families with four siblings examined. Nuclear families, even large nuclear families, provide relatively poor power for quantitative trait linkage. Studies of American or European families of moderate size, an average of 19 or 31 individuals per multigenerational pedigree, provided much better power, with relative efficiencies in the range of 0.35 – 0.59.

The smaller the family configuration, the larger the sample size will need to be to maintain an equivalent level of power in a human QTL linkage study. Williams and Blangero(46) provide analytical formulae for estimating the power to detect linkage for a quantitative trait in a variance component-based analysis as a function of the QTL-specific heritability, the total additive genetic heritability, the sample size, and the family configuration. For a QTL-specific heritability of 0.20, to achieve 80% power to detect linkage at a LOD score of 3 requires approximately 1000 individuals in extended pedigrees of size 48, > 2000 individuals in sibships of size 4, or > 6000 individuals in sibling pairs. Alternatively, for a sample size of 1000 – 2000 individuals, one has 80% power to detect linkage at a LOD score of 3 for QTL-specific heritabilities in the range of 0.14 – 0.20 if the sample is composed of pedigrees of 48 individuals each or 0.35 – 0.5 if the sample consists of sibling pairs.

A QTL-specific heritability of 0.14 is still quite large, even if we consider that it may be spread over multiple QTNs within the QTL, and it is likely that many traits of interest will not have any single QTL that contributes so much to the trait variance. However, the examples above show that some traits will have QTLs with effect sizes in this range and that linkage will work in some cases. It is impossible to know beforehand which traits these will be, just as it is impossible to know which traits will be influenced by common variants likely to be captured by the SNP panels used in genome-wide association studies. Given which, it seems likely that the complementary approaches of genome-wide linkage, optimized to find QTLs of large effect that may represent the joint effects of multiple QTNs in a region, and genome-wide association, optimized to find common QTNs of small effect, will both be useful for gene localization for complex traits, providing potentially different and complimentary results.

The future – RNA expression, whole genome SNP data, and sequencing

One way to try to increase the chances that a trait is influenced by fewer QTLs of larger single effects may be to try to select phenotypes that are likely to be less complex and closer to the level of gene action. This is the rationale behind many studies of quantitative traits as risk factors for complex diseases - that intermediate traits or endophenotypes may be less complex and closer to the level of gene action than disease endpoints(48-51). A new class of quantitative traits, RNA expression levels, is showing great promise in this respect(52,53). These traits are measurements of the amount of transcribed RNA for a particular gene, reflecting levels of gene expression. It is possible to assay expression levels for tens of thousands of human genes in a single experiment using chips that have been designed for this purpose. Göring et al.(53) reported that of the 20,413 transcripts with detectable expression in lymphocytes, approximately 85% had significant heritabilities and over a thousand of these traits had QTLs that were genome-wide significant at an experiment-wide false discovery rate of 0.05. For QTL linkages at the position of the structural gene coding for the RNA whose transcript levels were measured, the median QTL-specific heritability for the QTLs that met that 0.05 false discovery rate cut-off was 24.6%. Although these QTL-specific effect sizes are likely to be overestimates (54,55), they still suggest that there are many human QTLs with substantial effect sizes that can be detected by linkage. Göring et al. used these linkages of RNA expression traits to their structural loci as a starting point for identifying QTNs influencing traits of biomedical interest. They identified transcript levels with linkages to their own structural gene that were correlated with HDL cholesterol levels, sequenced the promoter regions of these genes to identify the

QTNs responsible for the variation in gene expression level, and tested whether these QTNs were also associated with variation in HDL cholesterol levels. Essentially, they were able to identify QTLs with relatively modest effects on HDL cholesterol levels by first localizing them through linkage analyses of correlated traits that were closer to gene action, i.e. RNA expression levels.

Other recent advances in molecular technology are also changing the uses and trajectory of human QTL linkage studies. It is now technologically and economically feasible, if not quite routine, to genotype a million or more SNPs genome-wide in a large sample of individuals. Initially the markers in these SNP panels were selected to provide good coverage of common variation in genome-wide association studies. That is, the panels were composed of markers with high allele frequencies chosen because they represented a group of markers that were in linkage disequilibrium with each other and with the 'tag' SNP selected for inclusion in the panel. Newer panels also include rarer variants, particularly known coding SNPs that change the amino acid sequence of a protein, and markers selected to identify polymorphic copy number variations in humans. Whole genome association studies with these panels are now rivaling linkage as a method for gene localization.

As discussed above, although they are often viewed as being in competition, linkage and association methods are complementary and are optimized to find different sorts of QTLs. Linkage has an advantage when the minor allele of a QTL is rare in the population but has a large effect in each individual who carries the variant allele or when there are multiple QTNs in a region that add up to a substantial total QTL effect size. Association has the advantage when a QTL is due to one or a few QTNs that are among the genotyped markers on a SNP panel or are common but of modest effect in each individual who carries the variant allele. Additionally, the two approaches can be used jointly to localize and identify QTLs. For a large study with dozens of quantitative phenotypes, association analyses with markers from a genome-wide SNP panel may be an economical way to follow up the multiple linkage signals in multiple chromosomal regions for the different phenotypes, helping to move from a general region of linkage to a specific gene or genes under the linkage peak. Conversely, linkage results, even ones that are not significant on a genome-wide level, may help to inform genome-wide association studies. Roeder et al. proposed that linkage results could be used to adjust p-values in genome-wide association studies, up-weighting the association results in regions with prior evidence for linkage with the weighting factor depending on the strength of the linkage signal (56). Sun et al. proposed a more categorical approach to the same issue, suggesting that false discovery rate could be controlled in a stratified manner, with SNPs being grouped by factors such as whether they were in a region with prior evidence of linkage(57).

The next logical step in the progression of genotyping technology for human gene mapping is to move to complete resequencing for purposes of genome scanning. At present, this is infeasible due to the cost and speed of available sequencing platforms. However, the X Prize Foundation has issued a challenge with a reward of 10 million dollars to the team that can sequence the complete genomes of 100 individuals in no more than 10 days at a cost of no more than \$10,000 per sample(58). If this challenge is met, it is a short step from that point to a situation where it becomes economically feasible to resequence 1000 or more individuals for a QTL localization study. When we have complete sequence data, we will be able to guarantee that any functional variants, any QTNs, that are present in a sample are among the genotyped markers and we will be able to test each variant in a fixed effects model that will provide more power than either association studies that depend on linkage disequilibrium between markers or linkage studies using random effects models. This is the juncture at which we believe linkage studies will become obsolete and a whole new class of quantitative genetic methods will be required to sift through the vast amounts of sequence data in ways that maintain power and control false positives in the face of massive multiple testing, take into account the structure

within the genome and the functional grouping of DNA segments into genes and regulatory elements, and take advantage of biological knowledge regarding related genes, traits, and pathways.

Acknowledgments

This work was supported in part by NIH grants R01 MH59490, R01 GM31575, U10 AA08403, R01 HL45522, R01 MH78111, and R01 HL070751. Thanks also to the NIH and NSF for travel support to attend ICQG3.

References

1. Hopper JL, Mathews JD. Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet* 1982;46(4):373–83. [PubMed: 6961886]
2. Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 1972;2(1):3–19. [PubMed: 4157472]
3. Blangero J, Williams JT, Almasy L. Variance component methods for detecting complex trait loci. *Adv Genet* 2001;42:151–81. [PubMed: 11037320]
4. Chen WM, Broman KW, Liang KY. Quantitative trait linkage analysis by generalized estimating equations: unification of variance components and Haseman-Elston regression. *Genet Epidemiol* 2004;26(4):265–72. [PubMed: 15095386]
5. Wang T, Elston RC. Two-level Haseman-Elston regression for general pedigree data analysis. *Genet Epidemiol* 2005;29(1):12–22. [PubMed: 15838848]
6. Daw EW, Wijsman EM, Thompson EA. A score for Bayesian genome screening. *Genet Epidemiol* 2003;24(3):181–90. [PubMed: 12652522]
7. Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. Quantitative trait locus for reading disability on chromosome 6. *Science* 1994;266(5183):276–9. [PubMed: 7939663]
8. Duggirala R, Stern MP, Mitchell BD, Reinhart LJ, Shipman PA, Uresandi OC, Chung WK, Leibel RL, Hales CN, O'Connell P, Blangero J. Quantitative variation in obesity-related traits and insulin precursors linked to the OB gene region on human chromosome 7. *Am J Hum Genet* 1996;59(3):694–703. [PubMed: 8751871]
9. Daniels SE, Bhattacharya S, James A, Leaves NI, Young A, Hill MR, Faux JA, Ryan GF, le Souef PN, Lathrop GM, Musk AW, Cookson WO. A genome-wide search for quantitative trait loci underlying asthma. *Nature* 1996;383(6597):247–50. [PubMed: 8805698]
10. Barker D, Green P, Knowlton R, Schumm J, Lander E, Oliphant A, Willard H, Akots G, Brown V, Gravius T, et al. Genetic linkage map of human chromosome 7 with 63 DNA markers. *Proc Natl Acad Sci U S A* 1987;84(22):8006–10. [PubMed: 2891136]
11. Fulker DW, Cherny SS, Cardon LR. Multipoint interval mapping of quantitative trait loci, using sib pairs. *Am J Hum Genet* 1995;56(5):1224–33. [PubMed: 7726180]
12. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62(5):1198–211. [PubMed: 9545414]
13. Sobel E, Sengul H, Weeks DE. Multipoint estimation of identity-by-descent probabilities at arbitrary positions among marker loci on general pedigrees. *Hum Hered* 2001;52(3):121–31. [PubMed: 11588394]
14. Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 1997;61(3):748–60. [PubMed: 9326339]
15. Comuzzie AG, Hixson JE, Almasy L, Mitchell BD, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J. A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nat Genet* 1997;15(3):273–6. [PubMed: 9054940]
16. Hager J, Dina C, Francke S, Dubois S, Houari M, Vatin V, Vaillant E, Lorentz N, Basdevant A, Clement K, Guy-Grand B, Froguel P. A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nat Genet* 1998;20(3):304–8. [PubMed: 9806554]
17. Rotimi CN, Comuzzie AG, Lowe WL, Luke A, Blangero J, Cooper RS. The quantitative trait locus on chromosome 2 for serum leptin levels is confirmed in African-Americans. *Diabetes* 1999;48(3):643–4. [PubMed: 10078570]

18. Palmer LJ, Buxbaum SG, Larkin E, Patel SR, Elston RC, Tishler PV, Redline S. A whole-genome scan for obstructive sleep apnea and obesity. *Am J Hum Genet* 2003;72(2):340–50. [PubMed: 12506338]
19. Moslehi R, Goldstein AM, Beerman M, Goldin L, Bergen AW. A genome-wide linkage scan for body mass index on Framingham Heart Study families. *BMC Genet* 2003;4:S97. [PubMed: 14975165]
20. Cole SA, Comuzzie AG, Dyer T, Britten M, Bounpheng M, Rines D, Almasy L, Williams JT, Blangero J. Genetic variation at the POMC locus partially defines the chromosome 2p human obesity quantitative trait locus. *Am J Hum Genet* 2003;73(Suppl):195.
21. Hixson JE, Almasy L, Cole S, Birnbaum S, Mitchell BD, Mahaney MC, Stern MP, MacCluer JW, Blangero J, Comuzzie AG. Normal variation in leptin levels is associated with polymorphisms in the proopiomelanocortin gene, POMC. *J Clin Endocrinol Metab* 1999;84(9):3187–91. [PubMed: 10487685]
22. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am J Hum Genet* 1999;64(4):1127–40. [PubMed: 10090898]
23. Reynisdottir I, Thorleifsson G, Benediktsson R, Sigurdsson G, Emilsson V, Einarsdottir AS, Hjorleifsdottir EE, Orlygsdottir GT, Bjornsdottir GT, Saemundsdottir J, Halldorsson S, Hrafnkelsdottir S, Sigurjonsdottir SB, Steinsdottir S, Martin M, Kochan JP, Rhees BK, Grant SF, Frigge ML, Kong A, Gudnason V, Stefansson K, Gulcher JR. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34–q35.2. *Am J Hum Genet* 2003;73(2):323–35. [PubMed: 12851856]
24. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, Styrkarsdottir U, Magnusson KP, Walters GB, Palsdottir E, Jonsdottir T, Gudmundsdottir T, Gylfason A, Saemundsdottir J, Wilensky RL, Reilly MP, Rader DJ, Bagger Y, Christiansen C, Gudnason V, Sigurdsson G, Thorsteinsdottir U, Gulcher JR, Kong A, Stefansson K. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006;38(3):320–3. [PubMed: 16415884]
25. Lehman DM, Hunt KJ, Leach RJ, Hamlington J, Arya R, Abboud HE, Duggirala R, Blangero J, Goring HH, Stern MP. Haplotypes of transcription factor 7-like 2 (TCF7L2) gene and its upstream region are associated with type 2 diabetes and age of onset in Mexican Americans. *Diabetes* 2007;56(2):389–93. [PubMed: 17259383]
26. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445(7130):881–5. [PubMed: 17293876]
27. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007;316(5829):1341–5. [PubMed: 17463248]
28. Soria JM, Almasy L, Souto JC, Sabater-Lleal M, Fontcuberta J, Blangero J. The F7 gene and clotting factor VII levels: dissection of a human quantitative trait locus. *Hum Biol* 2005;77(5):561–75. [PubMed: 16596941]
29. Sabater-Lleal M, Chillon M, Howard TE, Gil E, Almasy L, Blangero J, Fontcuberta J, Soria JM. Functional analysis of the genetic variability in the F7 gene promoter. *Atherosclerosis*. 2007
30. Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, O'Connell P, Stern MP. A major susceptibility locus influencing plasma triglyceride concentrations is located on chromosome 15q in Mexican Americans. *Am J Hum Genet* 2000;66(4):1237–45. [PubMed: 10729112]
31. Coon H, Leppert MF, Eckfeldt JH, Oberman A, Myers RH, Peacock JM, Province MA, Hopkins PN, Heiss G. Genome-wide linkage analysis of lipids in the Hypertension Genetic Epidemiology Network (HyperGEN) Blood Pressure Study. *Arterioscler Thromb Vasc Biol* 2001;21(12):1969–76. [PubMed: 11742872]

32. Austin MA, Edwards KL, Monks SA, Koprowicz KM, Brunzell JD, Motulsky AG, Mahaney MC, Hixson JE. Genome-wide scan for quantitative trait loci influencing LDL size and plasma triglyceride in familial hypertriglyceridemia. *J Lipid Res* 2003;44(11):2161–8. [PubMed: 12923221]
33. Arnett DK, Miller MB, Coon H, Ellison RC, North KE, Province M, Leppert M, Eckfeldt JH. Genome-wide linkage analysis replicates susceptibility locus for fasting plasma triglycerides: NHLBI Family Heart Study. *Hum Genet* 2004;115(6):468–74. [PubMed: 15375693]
34. Kissebah AH, Sonnenberg GE, Myklebust J, Goldstein M, Broman K, James RG, Marks JA, Krakower GR, Jacob HJ, Weber J, Martin L, Blangero J, Comuzzie AG. Quantitative trait loci on chromosomes 3 and 17 influence phenotypes of the metabolic syndrome. *Proc Natl Acad Sci U S A* 2000;97(26):14478–83. [PubMed: 11121050]
35. Wu X, Cooper RS, Borecki I, Hanis C, Bray M, Lewis CE, Zhu X, Kan D, Luke A, Curb D. A combined analysis of genomewide linkage scans for body mass index from the National Heart, Lung, and Blood Institute Family Blood Pressure Program. *Am J Hum Genet* 2002;70(5):1247–56. [PubMed: 11923912]
36. Luke A, Wu X, Zhu X, Kan D, Su Y, Cooper R. Linkage for BMI at 3q27 region confirmed in an African-American population. *Diabetes* 2003;52(5):1284–7. [PubMed: 12716766]
37. Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL. Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. *Am J Hum Genet* 1997;60(1):27–39. [PubMed: 8981944]
38. Grigorenko EL, Wood FB, Meyer MS, Pauls DL. Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. *Am J Hum Genet* 2000;66(2):715–23. [PubMed: 10677331]
39. Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, Richardson AJ, Weeks DE, Stein JF, Monaco AP. A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am J Hum Genet* 1999;64(1):146–56. [PubMed: 9915953]
40. Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, Bennett PH, Goldman D. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosomal-wide scan in an American Indian population. *Am J Med Genet* 1998;81(3):216–21. [PubMed: 9603607]
41. Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, Crowe RR, Goate A, Hesselbrock V, Jones K, Kwon J, Li TK, Nurnberger JI Jr, O'Connor SJ, Reich T, Rice J, Schuckit MA, Porjesz B, Foroud T, Begleiter H. Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. *Am J Hum Genet* 2004;74(4):705–14. [PubMed: 15024690]
42. Covault J, Gelernter J, Hesselbrock V, Nellisery M, Kranzler HR. Allelic and haplotypic association of GABRA2 with alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet* 2004;129(1):104–9. [PubMed: 15274050]
43. Drayna D, Coon H, Kim UK, Elsner T, Cromer K, Otterud B, Baird L, Peiffer AP, Leppert M. Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. *Hum Genet* 2003;112(56):567–72. [PubMed: 12624758]
44. Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* 2003;299(5610):1221–5. [PubMed: 12595690]
45. Curran JE, Jowett JB, Elliott KS, Gao Y, Gluschenko K, Wang J, Abel Azim DM, Cai G, Mahaney MC, Comuzzie AG, Dyer TD, Walder KR, Zimmet P, MacCluer JW, Collier GR, Kissebah AH, Blangero J. Genetic variation in selenoprotein S influences inflammatory response. *Nat Genet* 2005;37(11):1234–41. [PubMed: 16227999]
46. Williams JT, Blangero J. Power of variance component linkage analysis to detect quantitative trait loci. *Ann Hum Genet* 1999;63(Pt 6):545–63. [PubMed: 11246457]
47. Blangero J, Williams JT, Almasy L. Novel family-based approaches to genetic risk in thrombosis. *J Thromb Haemost* 2003;1(7):1391–7. [PubMed: 12871272]
48. Almasy L. Quantitative risk factors as indices of alcoholism susceptibility. *Ann Med* 2003;35(5):337–43. [PubMed: 12952020]

49. Almasy L, Blangero J. Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *Am J Med Genet* 2001;105(1):42–4. [PubMed: 11424994]
50. Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 2003;160(4):636–45. [PubMed: 12668349]
51. Ghosh S, Collins FS. The geneticist's approach to complex disease. *Annu Rev Med* 1996;47:333–53. [PubMed: 8712786]
52. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004;430(7001):743–7. [PubMed: 15269782]
53. Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, Maccluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet.* 2007
54. Beavis, WD. QTL analyses: power, precision, and accuracy. In: Paterson, AH., editor. *Molecular Dissection of Complex Traits*. CRC Press; New York: 1998. p. 145-162.
55. Goring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 2001;69(6):1357–69. [PubMed: 11593451]
56. Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve power of association in genome scans. *Am J Hum Genet* 2006;78(2):243–52. [PubMed: 16400608]
57. Sun L, Craiu RV, Paterson AD, Bull SB. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genet Epidemiol* 2006;30(6):519–30. [PubMed: 16800000]
58. Pennisi E. Genomics. On your mark. Get set. Sequence! *Science* 2006;314(5797):232. [PubMed: 17038592]