Austin Biometrics and Biostatistics

Austin Publishing Group

Open Access Full Text Article

Research Article

A Likelihood Model for Linkage Analysis of Genetic Traits

Ao Yuan[1]*, Xiaogang Zhong[1] and George E Bonney[2]

[1]Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, USA

[2]National Human Genome Center, Howard University, USA

***Corresponding author:** Ao Yuan, Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington, DC 20057, USA

Received: August 18, 2014; **Accepted:** September 09, 2014; **Published:** September 22, 2014

Abstract

Linkage analysis is one of the major approaches for genetic studies of human diseases, for mapping putative genes or studying relationships between loci. Many of the existing methods use identity by descent data, or a particular familial structure, which may not be fully available in some practices. Here we propose a likelihood model for linkage analysis with pedigrees, along with segregation and regressive analysis. Without requiring identity by descent data, this model can be used for both quantitative and qualitative traits to study trait-trait linkage with/without observed genotypes, or trait-marker linkage with observed marker genotype, which include sib pair analysis as a special case. This model is applied to a real data example for illustration.

Keywords: Gamet; Gene loci; Linkage; Recombinant fraction

Introduction

The advances in biotechnology have led to the identification of more and more disease genes without the knowledge of the biochemical nature of the diseases. Linkage analysis is one of the most commonly used approaches for mapping human disease genes, which is often the first step to identify the chromosomal location of them, and may followed by various diagnosis and ultimately therapeutic treatment for these diseases. There are numerous methods, parametric, nonparametric and semi-parametric, for link- age/ association analysis [1-6]. Furthermore Kruglyak et al. [7] proposed a unified multipoint approach, Hor- vath et al. [8] considered family based approach for this problem, Sung et al. [9] suggested a multipoint analysis using Markov chain Monte Carlo algorithm. Many of them use the Identity by Descent (IBD) data, or require some particular familial structure such as infected relative pairs or extreme discordant sib pairs. But in practices IBD data cannot be uniquely determined or not fully available, and particular familial structure are difficult to collect, while marker genotyping data are commonly available. Many of these models are not for the study of trait-trait genetic relationship; some of them use only part of the data information, for example the squared trait value difference. Although robust, the nonparametric model-free methods may suffer potential loss of efficiency since they do not use knowledge of traits generating mechanism. In addition, complex traits are often affected by covariates such as sex, age, race and environmental factors. Here we consider a simple likelihood model for linkage analysis for pedigrees, along with segregation and covariates analysis based on the likelihood principle. This model can be used to study trait-trait linkage with/without observed genotypes, or trait-marker linkage with observed marker genotype, which include sib pair analysis as a special case. Using this model as an illustration, we analyze a set of nuclear family data to reveal the genetic connection of two traits which are known have close phenotypic relationships. Some possible extension of future work is discussed.

Methods

We describe the method for quantitative traits and nuclear family, the cases for qualitative traits or combined traits are similar, the general pedigree case can be analyzed by breaking it into nuclear families. Let y_f, y_m and y_o be d-dimensional observations of the father, mother and off spring respectively, where

$$y_f = \left(y_{f,1},..,y_{f,d}\right)^T,\quad y_m = \left(y_{m,1},..,y_{m,d}\right)^T,$$

$$y_o = \left(y_1,..,y_n\right)^T,\ y_j = \left(y_{j,1},..,y_{j,d}\right)^T,\ (j = 1,...,n),$$

and n is the number of sibs in the nuclear family. Denote $y = (y_f, y_m, y_o)^T$ and its underlying random variable by $Y = (Y_f, Y_m, Y_o)^T$. Let L_1 and L_2 be the two loci under consideration for linkage analysis, we assume there are two alleles at each locus, with $a_1|b_1$ for L_1 and $a_2|b_2$ for L_2. We code the genotype at each locus as 0, 1 and 2 for b|b, a|b (b|a) and a|a respectively, r be the recombinant fraction - the probability that a gamet is recombinant, n be the sib size for the family. Let g_{fi}, g_{mi} and g_{ji} be the genotypes of father mother and the j-th sib at locus i (i = 1, 2), p_{1i} and p_{2i} be the proportion of the corresponding genotype at locus i. Let p_{ij} be the proportion of the haplotype

$$\left(\frac{g_i}{g_j}\right),\ (i,j = 0,1,2),\ g_f = \left(\frac{g_{f1}}{g_{f2}}\right)$$

$$g_m = \left(\frac{g_{m1}}{g_{m2}}\right),\ g_j = \left(\frac{g_{j1}}{g_{j2}}\right),\ T\left(g_j \middle| g_f, g_m\right)$$

be the transmission probability of the sibs genotype given those of the parents. Note that there are 9 possible composite genotypes at the two loci for each individual. Consider the multivariate model and the notations as in Yuan and Bonney [10], assume unknown phase, the likelihood for a given nuclear family can be written as

$$L(y) = \sum_{g_f} P\left(g_f\right) f\left(y_f \middle| g_f\right) \sum_{g_m} P\left(g_m\right) f\left(y_m \middle| y_f, g_f, g_m\right) K\left(g_f, g_m\right)$$

$$\times \prod_{j=1}^{n} \sum_{g_j} T\left(g_j \middle| g_f, g_m\right) f\left(y_j \middle| y_f, g_f, y_m, g_m, g_j\right), \qquad (1)$$

where, each summation is over all the genotypes of that individual at the two loci, in its general form with un observed genotypes at both loci, and $T_{(gj\,|gf\,,\,gm)}$ is the transmission probability for the case of unknown phase. In model (1) the conditional densities $f(y_f|g_f)$, $f(y_m|y_f, g_f, g_m)$ and $f(y_j|y_f, g_f, y_m, g_m, g_j)$ can be any general densities. Latter on for easy of exposition and convenience of application, we will assume that $f_{(yf|gf)}$ is the d-dimensional normal density with mean

Citation: Yuan A, Zhong X and Bonney GE. A Likelihood Model for Linkage Analysis of Genetic Traits. Austin Biom and Biostat. 2014;1(1): 7.

$$\mu_f = \sum_{i=1}^{9} \beta_i \chi\left(g_f = i\right) + \beta x_f$$

and variance matrix Σ_f, where the $\chi(g_f = i)$ denote the event that the father's composite genotype if of type *i*, β's are d-dimensional vector of parameters and x_f is the covariates matrix for the father; in the same manner, $f_{(ym|yf\,,\,gf\,,\,gm)}$ is the conditional normal density with mean

$$\mu_m + \Omega_p \sum\nolimits_f^{-1}\left(y_f - \mu_f\right) \quad ,$$

where

$$\mu_m = \sum_{i=1}^{9} \beta_i \chi\left(g_m = i\right) + \beta x_m$$

and variance matrix $\Sigma_m - \Omega_p \Sigma_f^{-1} \Omega_p$, and Σ_m is the variance matrix of mother alone and Ω_p is the between-parents correlation matrix. Furthermore, we take $K_{(gf,gm)}$ as the K-function as in Yuan and Bonney [10] which is an adjustment factor for the product of the penetrance of the sibs given the parents genotypes and $f_{(yj|yf,\ gf,\ ym,\ gm,\ gj)}$ is the conditional normal density function with mean

$$\mu_j + \Omega_{sp} \sum\nolimits_p^{-1}\left(y_p - \mu_p\right),$$

where $\Omega_{sp}=(\Omega_{sf},\Omega_{sm})$ is the sib-parents correlation matrix which is composed of the sib-father and sib-mother blocks of correlation matrices,

$$\sum\nolimits_p = \begin{pmatrix} \Sigma_f & \Omega_p \\ \Omega_p & \Sigma_m \end{pmatrix}, y_p = \begin{pmatrix} y_f \\ y_m \end{pmatrix}, \mu_p = \begin{pmatrix} \mu_f \\ \mu_m \end{pmatrix}$$

and

$$\mu_j = \sum_{i=1}^{9} \beta_i \chi\left(g_j = i\right) + \beta x_j$$

and variance matrix $\Sigma_s - \Omega_{sp} \Sigma_p^{-1} \Omega_{sp}$. Note that although we use the same coding for the two loci, but f_1=0 and f_2=0 do not mean the same gene at the two loci. The specification of the joint genotype proportion p_{ij}'s and the transmission probabilities $T_{(gj|gf,\ gm)}$ is put expression (10) latter, and its values are given in Table II.

Note in model (1), typically there are many zero components of the transmission probability $T_{(gj|gf,\ gm)}$, so that it will be more efficient to evaluate $T_{(gj|gf,\ gm)}$ first, if its non-zero then compute the penetrances for the family members, otherwise ignore the computation for that combination of genotypes. The $T_{(gj|gf,\ gm)}$'s are functions of the recombination fraction *r*. When the phase is known, (1) should be modified as

$$L(y) = \sum_{g_f} P(g_f)\, f(y_f \,|g_f) \sum_{g_m} P(g_m)\, f(y_m \,|y_f, g_f, g_m) K(g_f, g_m)$$

$$\times \prod_{j=1}^{n} \sum_{g_j} T_1(g_j \,|g_f, g_m; h(g_j, g_f, g_m))\, f(y_j \,|y_f, g_f, y_m, g_m, g_j), \qquad (2)$$

where $T_{1(gj|gf,gm;\ h(gj,gf,gm))}$ is the transmission probability for the give phase configuration $h_{(gj,\ gf,\ gm)}$ of *(gj|gf, gm)*. So (1) is can be rewritten as

$$L(y) = \sum_{g_f} P(g_f)\, f(y_f \,|g_f) \sum_{g_m} P(g_m)\, f(y_m \,|y_f, g_f, g_m) K(g_f, g_m)$$

$$\times \sum_{h(g_j, g_f, g_m)} P(h(g_j, g_f, g_m)) \prod_{j=1}^{n} \sum_{g_j} T_1(g_j \,|g_f, g_m, h(g_j, g_f, g_m))\, f(y_j \,|y_f, g_f, y_m, g_m, g_j),$$

where $\Sigma_{h(gj,gf,gm)}$ is summation across all different phase configurations $h_{(gj,\ gf,\ gm)}$s of *(gj, gf, gm)*, and $P(h_{(gj,\ gf,\ gm)})$ is the probability of configuration $h_{(gj,\ gf,\ gm)}$. The number of different phase configurations of *(gj, gf, gm)* depends on the number of heterozygote's in it. Note here we have two loci, each locus has two genotypes, and the genotypes of the parents are assumed independent, as common in the literature. If there are k ($0 \le k \le 6$) heterogygotes in *(gj, gf, gm)*, then there are 2k different phase configurations, and each has probability $P(h_{(gj,\ gf,\ gm)})$=1/2k. This method needs to enlist all the different phase configurations, since different triple *(gj, gf, gm)* may have different number of phase configurations, this method will not be easy in terms of programming. A more convenient way in programming is to treat each genotype as heterozygote, and sum over all the 26=64 phase configurations each with probability 1/64. Although this way will have some redundant computations, but is a general procedure, it does not require to enlist the phase configurations for each triple *(gj, gf, gm)*, and so is easy to programming. The values of $T_{(gj|gf,\ gm)}$ are given in Table II in the Appendix, for all possible composite genotypes of *(gj, gf, gm)*. This is a general procedure for programming without the knowledge of the phase configuration for each triple.

Linkage between trait loci

For simplicity, we only consider the case of two phenotypes controlled by their own loci with unobserved genotypes at both loci.

Linkage between trait and marker loci

Suppose the data y is controlled by one locus with unobserved genotype, and we have the genotype g_2 of *y* at the marker locus, a common assumption is that, g_2 has no epistatic interaction with *y*, i.e. g and y has no direct connection, but g_2 has relationship with the unobserved genotype of *y*, and phase unknown. In this case (1) becomes

$$L(y) = \sum_{g_{f1}} P(g_f)\, f(y_f \,|g_f) \sum_{g_{m1}} P(g_m)\, f(y_m \,|y_f, g_f, g_m) K(g_{f1}, g_{m1})$$

$$\times \prod_{j=1}^{n} \sum_{g_{f1}} T(g_j \,|g_f, g_m)\, f(y_j \,|y_f, g_f, y_m, g_m, g_j), \qquad (3)$$

here the summation is only for all the genotypes at the trait locus.

Point analysis

One way of multi-point linkage analysis is to perform 3-point analysis step by step across the segment span the multipoint. Here we use our model to address the 3-point analysis. In this problem, we have two markers and an unknown disease locus, which may lie between the two markers or outside the interval between them. We assume that the case is unknown, while the model is similar when the phase is known. Again, we only need to specify the likelihood for one family. The composite genotypes are $g_f = (g_{f1}, g_{f2}, g_{f3})$ for the father, $g_m = (g_{m1}, g_{m2}, g_{m3})$ for the mother, and $g_j = (g_{j1}, g_{j2}, g_{j3})$ for the j-th sib. We assume the first and second genotypes in the composite genotype of each individual are the observed genotypes at markers 1 and 2, the third marker g_{j3} is the unobserved disease genotype, assuming marker g_{j1} is located at the left side of marker g_{j2} on the chromosome. Since we have three loci, there are three recombination fractions for the three pair wise loci. Denote r_1 as the recombination fraction between marker 1 and the disease marker, r_2 as that between marker 2 and the disease, r_3 as that between the first two markers, and $T(g_{j1}, g_{j2}, g_{j3}|g_{f1}, g_{f2}, g_{f3}); (g_{m1}, g_{m2}, g_{m3}))$ the 3-point transmission probability, which is a function of (r_1, r_2, r_3). In this case, (3) is rewritten as

Appendix A:

No.	g_f $g_m \setminus g_s$	$\binom{0}{0}$	$\binom{0}{1}$	$\binom{0}{2}$	$\binom{1}{0}$	$\binom{1}{1}$	$\binom{1}{2}$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$
1	$\binom{0}{0}\binom{0}{0}$	1								
2	$\binom{0}{0}\binom{0}{1}$	$\frac{1}{2}$	$\frac{1}{2}$							
3	$\binom{0}{0}\binom{1}{0}$	$\frac{1}{2}$			$\frac{1}{2}$					
4	$\binom{0}{1}\binom{0}{0}$	$\frac{1}{2}$	$\frac{1}{2}$							
5	$\binom{1}{0}\binom{0}{0}$	$\frac{1}{2}$			$\frac{1}{2}$					
6	$\binom{0}{0}\binom{0}{2}$		1							
7	$\binom{0}{0}\binom{2}{0}$				1					
8	$\binom{0}{2}\binom{0}{0}$		1							
9	$\binom{2}{0}\binom{0}{0}$				1					
10	$\binom{1}{0}\binom{0}{0}$	$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$				
11	$\binom{1}{0}\binom{1}{0}$	$\frac{1}{4}$			$\frac{1}{2}$			$\frac{1}{4}$		
12	$\binom{1}{0}\binom{0}{1}$	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$				
13	$\binom{0}{0}\binom{1}{1}$	$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$				
14	$\binom{0}{1}\binom{0}{1}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$						
15	$\binom{0}{1}\binom{1}{0}$	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$				
16	$\binom{2}{1}\binom{0}{0}$				$\frac{1}{2}$	$\frac{1}{2}$				
17	$\binom{2}{0}\binom{1}{0}$				$\frac{1}{2}$			$\frac{1}{2}$		
18	$\binom{2}{0}\binom{0}{1}$				$\frac{1}{2}$	$\frac{1}{2}$				
19	$\binom{1}{2}\binom{0}{0}$		$\frac{1}{2}$			$\frac{1}{2}$				
20	$\binom{0}{2}\binom{1}{0}$		$\frac{1}{2}$			$\frac{1}{2}$				

No.	g_f $g_m \setminus g_s$	$\binom{0}{0}$	$\binom{0}{1}$	$\binom{0}{2}$	$\binom{1}{0}$	$\binom{1}{1}$	$\binom{1}{2}$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$
41	$\binom{1}{1}\binom{0}{2}$		$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$			
42	$\binom{2}{1}\binom{0}{1}$				$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$			
43	$\binom{1}{2}\binom{0}{1}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$			
44	$\binom{0}{2}\binom{1}{1}$		$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$			
45	$\binom{0}{1}\binom{2}{1}$				$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$			
46	$\binom{0}{1}\binom{1}{2}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$			
47	$\binom{2}{0}\binom{1}{1}$				$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$	
48	$\binom{1}{0}\binom{2}{1}$				$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	
49	$\binom{1}{0}\binom{1}{2}$		$\frac{1}{4}$			$\frac{1}{2}$			$\frac{1}{4}$	
50	$\binom{0}{1}\binom{2}{2}$					$\frac{1}{2}$	$\frac{1}{2}$			
51	$\binom{0}{2}\binom{1}{2}$			$\frac{1}{2}$			$\frac{1}{2}$			
52	$\binom{0}{2}\binom{2}{1}$					$\frac{1}{2}$	$\frac{1}{2}$			
53	$\binom{1}{0}\binom{2}{2}$					$\frac{1}{2}$			$\frac{1}{2}$	
54	$\binom{2}{0}\binom{1}{2}$					$\frac{1}{2}$			$\frac{1}{2}$	
55	$\binom{2}{0}\binom{2}{1}$							$\frac{1}{2}$	$\frac{1}{2}$	
56	$\binom{2}{2}\binom{0}{1}$					$\frac{1}{2}$	$\frac{1}{2}$			
57	$\binom{1}{2}\binom{0}{2}$			$\frac{1}{2}$			$\frac{1}{2}$			
58	$\binom{2}{1}\binom{0}{2}$					$\frac{1}{2}$	$\frac{1}{2}$			
59	$\binom{2}{1}\binom{2}{0}$							$\frac{1}{2}$	$\frac{1}{2}$	
60	$\binom{2}{2}\binom{1}{0}$					$\frac{1}{2}$			$\frac{1}{2}$	

No.	g_f $g_m \setminus g_s$	$\binom{0}{0}$	$\binom{0}{1}$	$\binom{0}{2}$	$\binom{1}{0}$	$\binom{1}{1}$	$\binom{1}{2}$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$
21	$\binom{0}{2}\binom{0}{1}$		$\frac{1}{2}$	$\frac{1}{2}$						
22	$\binom{1}{0}\binom{2}{0}$				$\frac{1}{2}$			$\frac{1}{2}$		
23	$\binom{0}{1}\binom{2}{0}$				$\frac{1}{2}$	$\frac{1}{2}$				
24	$\binom{0}{0}\binom{2}{1}$				$\frac{1}{2}$	$\frac{1}{2}$				
25	$\binom{0}{0}\binom{1}{2}$		$\frac{1}{2}$			$\frac{1}{2}$				
26	$\binom{1}{0}\binom{0}{2}$		$\frac{1}{2}$			$\frac{1}{2}$				
27	$\binom{0}{1}\binom{0}{2}$		$\frac{1}{2}$	$\frac{1}{2}$						
28	$\binom{2}{2}\binom{0}{0}$					1				
29	$\binom{2}{0}\binom{2}{0}$							1		
30	$\binom{2}{0}\binom{0}{2}$					1				
31	$\binom{0}{2}\binom{2}{0}$					1				
32	$\binom{0}{2}\binom{0}{2}$			1						
33	$\binom{0}{0}\binom{2}{2}$					1				
34	$\binom{1}{1}\binom{1}{0}$	$\frac{1-r}{4}$	$\frac{r}{4}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{r}{4}$	$\frac{1-r}{4}$	
35	$\binom{1}{1}\binom{0}{1}$	$\frac{1-r}{4}$	$\frac{1}{4}$	$\frac{r}{4}$	$\frac{r}{4}$	$\frac{1}{4}$	$\frac{1-r}{4}$			
36	$\binom{0}{1}\binom{1}{1}$	$\frac{1-r}{4}$	$\frac{1}{4}$	$\frac{r}{4}$	$\frac{r}{4}$	$\frac{1}{4}$	$\frac{1-r}{4}$			
37	$\binom{1}{0}\binom{1}{1}$	$\frac{1-r}{4}$	$\frac{r}{4}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{r}{4}$	$\frac{1-r}{4}$	
38	$\binom{1}{1}\binom{2}{0}$				$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$	
39	$\binom{1}{2}\binom{1}{0}$		$\frac{1}{4}$			$\frac{1}{2}$			$\frac{1}{4}$	
40	$\binom{2}{1}\binom{1}{0}$				$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$	

No.	g_f $g_m \setminus g_s$	$\binom{0}{0}$	$\binom{0}{1}$	$\binom{0}{2}$	$\binom{1}{0}$	$\binom{1}{1}$	$\binom{1}{2}$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$
61	$\binom{1}{2}\binom{2}{0}$					$\frac{1}{2}$			$\frac{1}{2}$	
62	$\binom{0}{2}\binom{2}{2}$						1			
63	$\binom{2}{0}\binom{2}{2}$								1	
64	$\binom{2}{2}\binom{0}{2}$						1			
65	$\binom{2}{2}\binom{2}{0}$								1	
66	$\binom{1}{1}\binom{1}{1}$	$\frac{(1-r)^2}{4}$	$\frac{r(1-r)}{2}$	$\frac{r^2}{4}$	$\frac{r(1-r)}{2}$	$\frac{1-2r+2r^2}{2}$	$\frac{r(1-r)}{2}$	$\frac{r^2}{4}$	$\frac{r(1-r)}{2}$	$\frac{(1-r)^2}{4}$
67	$\binom{2}{1}\binom{1}{1}$				$\frac{1-r}{4}$	$\frac{1}{4}$	$\frac{r}{4}$	$\frac{r}{4}$	$\frac{1}{4}$	$\frac{1-r}{4}$
68	$\binom{1}{2}\binom{1}{1}$		$\frac{1-r}{4}$	$\frac{r}{4}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{r}{4}$	$\frac{1-r}{4}$
69	$\binom{1}{1}\binom{2}{1}$				$\frac{1-r}{4}$	$\frac{1}{4}$	$\frac{r}{4}$	$\frac{r}{4}$	$\frac{1}{4}$	$\frac{1-r}{4}$
70	$\binom{1}{1}\binom{1}{2}$		$\frac{1-r}{4}$	$\frac{r}{4}$		$\frac{1}{4}$	$\frac{1}{4}$		$\frac{r}{4}$	$\frac{1-r}{4}$
71	$\binom{2}{2}\binom{1}{1}$					$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$
72	$\binom{2}{1}\binom{2}{1}$							$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
73	$\binom{2}{1}\binom{1}{2}$					$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$
74	$\binom{1}{2}\binom{2}{1}$					$\frac{1}{4}$	$\frac{1}{4}$		$\frac{1}{4}$	$\frac{1}{4}$
75	$\binom{1}{2}\binom{1}{2}$			$\frac{1}{4}$			$\frac{1}{2}$			$\frac{1}{4}$
76	$\binom{1}{1}\binom{2}{2}$					$\frac{1-r}{2}$	$\frac{r}{2}$		$\frac{r}{2}$	$\frac{1-r}{2}$
77	$\binom{1}{2}\binom{2}{2}$						$\frac{1}{2}$			$\frac{1}{2}$
78	$\binom{2}{1}\binom{2}{2}$								$\frac{1}{2}$	$\frac{1}{2}$
79	$\binom{2}{2}\binom{1}{2}$						$\frac{1}{2}$			$\frac{1}{2}$
80	$\binom{2}{2}\binom{2}{1}$								$\frac{1}{2}$	$\frac{1}{2}$
81	$\binom{2}{2}\binom{2}{2}$									1

$$L(y)=\sum_{g_{f3}}P(g_f)f(y_f|g_f)\sum_{g_{m3}}P(g_m)f(y_m|y_f,g_f,g_m)K(g_{f_3},g_{m_3})$$
$$\times\prod_{j=1}^{n}\sum_{g_{j3}}T(g_{j_1},g_{j_2},g_{j_3}|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})f(y_j|y_f,g_f,y_m,g_m,g_j) \quad (4)$$

Note in this model, although $f(\cdot|\cdot)$ has the same form as in model (1), but the mean μ's has 27 coefficients for all the possible different 3-point composite genotypes, instead of 9. In equation (4) the key is the specification of the 3-point transmission probability. Note that the three recombination fractions are not independent. During meiosis, when there is a cross-over at marker 1, and no cross-over at the other two loci, then there is a recombination event between marker 1 and the disease marker, it is also a recombination event between marker 1 and marker 2; however if there is also a cross-over at marker 2, then there is no recombination event between marker 1 and marker 2. If we consider all the possibilities of crossovers at the 3 markers, the relationships of r_1, r_2 and r_3, and the combinatory outcomes of the 3-point gametes can be complicated. In this case a complete Table of all the 3-poin transmission probabilities as in Table II will have 729×27 entries. So it is impractical to list all such probabilities. One may use Haldane's model (Lange 1997, p110) for the specification, but this model is not easy to implement into software. Observe that

$$T(g_{j_1},g_{j_2},g_{j_3}|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})=\frac{P(g_{j_1},g_{j_2},g_{j_3};g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})}{P(g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})}$$

$$=\frac{P(g_{j_1},g_{j_2}|g_{j_3};g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})P(g_{j_3};|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})}{P(g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})}$$

$$=P(g_{j_1},g_{j_2}|g_{j_3};g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})P(g_{j_3};|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})$$
$$=P(g_{j_1},g_{j_2}|g_{f_1},g_{f_2};g_{m_1},g_{m_2})P(g_{j_3}|g_{f_3};g_{m_3})=T_3(g_{j_1},g_{j_2}|g_{f_1},g_{f_2};g_{m_1},g_{m_2})P(g_{j_3}|g_{f_3},g_{m_3})$$

Similarly,

$$T(g_{j_1},g_{j_2},g_{j_3}|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})=T_2(g_{j_2},g_{j_3}|g_{f_2},g_{f_3};g_{m_2},g_{m_3})P(g_{j_1}|g_{f_1},g_{m_1})$$

and

$$T(g_{j_1},g_{j_2},g_{j_3}|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})=T_1(g_{j_1},g_{j_3}|g_{f_1},g_{f_3};g_{m_1},g_{m_3})P(g_{j_2}|g_{f_2},g_{m_2})$$

Where $Ti(\cdot|\cdot)$ is a function of r_i and its values are given in Table II, just replace r there by r_i (i = 1, 2, 3). The values of $P(g_j|g_f, g_m)$ are given in Table III for convenience. So we specify the 3-point transmission probability as

$$T(g_{j_1},g_{j_2},g_{j_3}|g_{f_1},g_{f_2},g_{f_3};g_{m_1},g_{m_2},g_{m_3})=\frac{1}{3}\big(T_1(g_{j_1},g_{j_3}|g_{f_1},g_{f_3};g_{m_1},g_{m_3})P(g_{j_2}|g_{f_2},g_{m_2})$$
$$+T_2(g_{j_2},g_{j_3}|g_{f_2},g_{f_3};g_{m_2},g_{m_3})P(g_{j_1}|g_{f_1},g_{m_1})+T_3(g_{j_1},g_{j_2}|g_{f_1},g_{f_2};g_{m_1},g_{m_2})P(g_{j_3}|g_{f_3},g_{m_3})\big)$$

Finally, the MLE ($\hat{r}_1$, $\hat{r}_2$, $\hat{r}_3$) of (r_1, r_2, r_3) is computed. If $\hat{r}_1$ = max {$\hat{r}_1$, $\hat{r}_2$, $\hat{r}_3$}, the disease locus is more likely lies on the right side of marker 2; if $\hat{r}_2$ = max {$\hat{r}_1$, $\hat{r}_2$, $\hat{r}_3$}, the disease locus is more likely lies on the left side of marker 1; If $\hat{r}_3$ = max {$\hat{r}_1$, $\hat{r}_2$, $\hat{r}_3$}, the disease locus is more likely lies between markers 1 and 2.

Multi-point analysis

In genome wide linkage analysis, there are often hundreds of markers to be considered. Instead one marker at a time, it is known that analyzing all the makers together will enhance the power. Let k be the total number of markers under consideration, there are k (k – 1)/2 pair wise recombination's fractions r_{ij}s. It is difficult to estimated all the recombination's in a model, and it is unnecessary, but usually the map distances of the markers are known, so the recombination fractions among the markers can be estimated automatically using map functions, for example the Hadane function or Kosamby function. If we know actually all the particular marker positions on the chromosome, their recombination's fractions then can be determined. So if we let r_{0j} be the recombination fraction between the disease locus and the locus of marker j, which are the only unknown recombination fractions to be estimated, we assume that the other r_{ij}s ($i, j \neq 0$) are known. As far as we know, usually the markers are from haplotype blocks, different blocks are weakly dependent, and the markers within the same block are strongly dependent, but not perfectly dependent. For some blocks, only one marker is typed, while in some other blocks there are more than one marker. Then a likelihood using all the traits as in equation (4) will be impractical as it will involve too many parameters. Instead, we may consider the likelihood only use the observed marker composite genotypes. Let $r = (r_{01},...,r_{0k})$, (g_{f0}, g_{m0}, g_{j0}) be the unobserved genotypes of (father, mother, sib), $g_f = (g_{f0}, g_{f1},...,g_{fk})$ be the composite genotype of the father, $g_m = (g_{m0}, g_{m1}, ...,g_{mk})$ be that of the mother, and $g_j = (g_{j0}, g_{j1},...,g_{jk})$ be that of the j-th sib. The likelihood for one family is

Table 1: Linkage Results on chromosome 4 for ntth1.

SNP	Marker Name	Map Distance(cM)	Allele	Lodscore
84	tsc1276837	34.24	2	1.75
146	tsc0526379	52.99	1	1.92
148	tsc0045058	53.26	1	3.35
159	tsc0527513	55.57	2	1.93
295	tsc1213381	85.42	2	2.22
319	tsc0055068	89.31	2	3.03
714	tsc0051777	172.86	2	2.57

$$L(r)=\sum_{g_{f0}}P(g_f)\sum_{g_{m0}}P(g_m)\prod_{j=1}^{n}\sum_{g_{j0}}T(g_j|g_f,g_m) \quad (5)$$

Then the problems are how to specify $P(g_f)$ and how to specify $T(g_j|g_f, g_m)$? For the transmission probability, Haldane's model (Lange [11]) is not easy to use, since it requires the recombination status among the markers, which are always unknown with the phases. Let $T_{rs} = T_{rs}(g_{jr}, g_{js}|g_{fr}, g_{fs}; g_{mr}, g_{ms})$ be the transmission probability at marker loci (r, s), we can specify the transmission as in the three point case, as

$$T(g_j|g_f,g_m)=\frac{1}{k}\begin{cases}T_{0,1}T_{2,3}T_{4,5}...T_{k-1,k}+T_{0,2}T_{1,3}T_{4,5}...T_{k-1,k}+...+T_{0,k}T_{1,2}T_{3,4}...T_{k-2,k-1};\\ T_{0,1}T_{2,3}T_{4,5}...T_{k-2,k-1}P(g_{j_k}|g_{f_k},g_{m_k})+T_{0,2}T_{1,3}T_{4,5}...T_{k-2,k-1}P(g_{j_k}|g_{f_k},g_{m_k})\end{cases}$$
$$\begin{cases}\quad if \qquad k=2l-1;\\ +...+T_{0,k}T_{1,2}T_{3,4}...T_{k-3,k-2}P(g_{j_{k-1}}|g_{f_{k-1}},g_{m_{k-1}});\quad if \quad k=2l.\end{cases} \quad (6)$$

For $r \neq 0$, the $T_{r,s}$'s are given in Table II, with r replaced by r_{rs}; $P(g_j|g_f, g_m)$ is the corresponding one-locus transmission probability at the unpaired left-over locus. Once $P(g_f)$ (and so $P(g_f)$) is specified, $T(g_j|g_f, g_m)$ in equation (6) is a quadratic function of r. It can be applied to any nuclear family design.

The method in Liang et al.[12] is also simple, but it requires to known the trans- mitted allele status of father and mother at each loci, which are sometimes uncertain, or can only be inferred with 1/2

probability. Also this method applies to only to the case-parent trio design.

Specification of the haplotype and the transmission probabilities

Specification of the haplotype probability: A simple way is to assume linkage equilibrium between the two loci and set

$$P(g) = P(g1)P(g2) \quad (7)$$

However this assumption is inappropriate with the presence of linkage [13-15]. When dealing with Linkage Disequilibrium (LD), we usually need to consider all possible gametic disequilibrium within the haplotype [16], which will be very complicated for three or more alleles. With model (1) and (2), we can define the LD parameter as

$$\delta = P(g) - P(g1)P(g2) \quad (8)$$

In case of Hardy-Weinberg Disequilibrium (HWD), let f be the common HWD parameter [17,18] at the two loci, at each locus

$$p_{kk} := P(A_k A_k) = p_k^2 + p_k(1-p_k)f,\ p_{kl} := P(A_k A_l) = p_k p_l(1-f),\ k \neq l,$$

we have

$$P(g) = p_{ij}^1 p_{kl}^2 + \delta \quad (9)$$

Where $p_{ij}^1 = P(a_i a_j)$ is the genotype probability at the trait locus and $p_{kl}^2 = P(A_k A_l)$ is the probability at the marker, and both of them satisfy the above HWD specification.

Specification of the transmission probability: Let r be recombinant fraction – the probability that a given sib's genotype is a recombinant of those of his/her parents'. $0 \le r \le 1/2$, $r = 0$ corresponds to complete linkage of the two loci, $r = 1/2$ corresponds to no linkage (Sham [19]). For the two-allele two loci case, there are $3^6 = 729$ possible values of $T(g_s|g_f, g_m)$, but only a few different ones and many zeros. If the genotypes are ordered, let $g_s = g_{sf}||g_{sm}$, where g_{sf} be the paternal gamete and g_{sm} the maternal gamete, we have (Lange 1997).

$$T(g_s|g_f, g_m) = T(g_{sf}|g_f)T(g_{sm}|g_m).$$

Given, $g_f = \left(\frac{a\|b}{A\|B}\right)$ we list all the non-zero values of $T(\cdot|\left(\frac{a\|b}{A\|B}\right))$, with various settings of g_{sf}, as the following

$$T(\cdot|\left(\frac{a\|b}{A\|B}\right)) = \begin{cases} \left(\frac{a}{A}\right), & a=b\ \&\ A=B; \\ 1 & \\ \left(\frac{a}{A}\right)\ \left(\frac{b}{A}\right), & a\neq b\ \&\ A=B; \\ \frac{1}{2}\ \ \frac{1}{2} & \\ \left(\frac{a}{A}\right)\ \left(\frac{a}{B}\right), & a=b\ \&\ A\neq B; \\ \frac{1}{2}\ \ \frac{1}{2} & \\ \left(\frac{a}{A}\right)\ \left(\frac{b}{B}\right)\ \left(\frac{a}{B}\right)\ \left(\frac{b}{A}\right), & a\neq b\ \&\ A\neq B; \\ \frac{1-r}{2}\ \ \frac{1-r}{2}\ \ \frac{r}{2}\ \ \frac{r}{2} & \end{cases} \quad (10)$$

the values of $T(g_{sm}|g_m)$ are the same.

Using (10) and the product $T(g_{sf}|g_f)\ T(g_{sm}|g_m)$ we can get all the transmission probabilities in allelic representation for each sib genotype $g_{sf}|g_{sm}$.

We list all the non-zero values of the transmission probabilities in numerical notation in Table II in Appendix A. All the 81 combinations of parent's genotypes are given in the second column, all those 9 for the sib given in the first row. An illustration of the computation of the entries in the table using (10) or directly by hand is given in the Appendix B.

Table 2: *(f_m, leptin) = (X_g, sex, age).*

θ	$\hat{\theta}_0$	$\hat{\theta}$
$\mu_{0,1}$	11.682(2.855)	11.659(2.904)
$\mu_{0,2}$	-2.146(0.464)	-2.292(0.464)
$\alpha_{1,1}$	2.527(0.574)	2.591(0.480)
$\alpha_{2,1}$	1.430(0.670)	1.639(0.550)
$\beta_{2,1}$	12.929(1.024)	12.913(1.024)
$\beta_{2,2}$	19.558(1.192)	19.531(1.192)
$\beta_{3,1}$	0.211(0.032)	0.210(0.032)
$\beta_{3,2}$	0.202(0.037)	0.202(0.037)
σ_1^2	212.630(12.225)	212.575(12.245)
σ_2^2	289.787(16.109)	289.785(16.132)
ρ_w	0.636(0.023)	0.636(0.023)
$\rho_{b[1,1]}$	0.262(0.047)	0.262(0.047)
$\rho_{b[1,2]}$	0.242(0.038)	0.242(0.038)
$P_{b[2,2]}$	0.264(0.042)	0.264(0.042)
qA_1	0.736(0.313)	0.738(0.325)
qA_2	0.757(0.609)	0.753(0.610)
r	0.500	0.000
loglike	-4505.583996	-4505.559749

Application

We analyze the data set released by the Gegetic Analysis Workshop14 using the pro- posed method. Recently, evidence has been found to relate alcoholism to genetic factors [20-22]. The Collaborative Study on the Genetics of Alcoholism (COGA) is a program to study this phenomenon extensively. The data set contains multiple phenotypes and genome wide scans from 229 families and 1490 individuals, in which 720 of them have incomplete/missing observations. Each individual has 20 records, among which the first 5 are i.d. or categorical, most of the other variables are continuous traits, including fat mass (fm) and leptin. We break the data into nuclear families. Sibs with missing response(s)/covariate(s) are deleted from the data, parents with missing response(s)/covariate(s) are kept in order for tracking down the family structure.

We first study the genetic association of electrophysiological measures related to alcoholism focusing on the NTTH phenotypes and the 786 Affymetrix SNPs on chromo- some 4. This chromosome has been shown to be involved in NTTH phenotypes in some previous studies.

There are four NTTH quantitative phenotypes: ntth1, ntth2, ntth3 and ntth4. Typically for this problem, one may perform a linkage analysis to pinpoint the highly spurious region, but this is computationally intensive and time consuming. In this dataset, the number of SNPs is large. For chromosome 4 alone there are 786 SNPs. We did a two- stage analysis. The first stage is an association

analysis, in which we regressed the trait on age, sex, and the SNPs, one at a time, across all the 786 SNPs on chromosome 4. This will analyze the statistical association between the phenotypes and the SNPs, which will provide us the phenotypes/SNPs with significant association for the next stage analysis. In the second stage, a formal linkage analysis was performed using model (1) on the SNPs selected from the first stage. After the two-stage analysis, we found ntth1 has strong linkage to some SNPs, while the trait linkage for ntth2-httn4 is not significant. The results on ntth1 for those SNPs with significant linkage are presented in Table I.

From this table, we find strong linkage in four regions: SNPs tsc0045058, tsc1213381, tsc0055068 and tsc0051777 at chromosome positions 148, 295, 319 and 714, with map distances 53.26, 85.42, 89.31 and 172.86cM; and moderate linkage in three regions: tsc1276837, tsc0526379 and tsc0527513 at positions 84, 146 and 159, with map distances 34.24, 52.99 and 55.57 cM.

Next we test the hypothesis of no linkage between the loci and the trait. It is known that fat mass and leptin are closely related phenotypically, we are interested in the genotypic relationship between them, and assume they are controlled by their own gene loci, with sex and age as covariates. Without genotype data at both loci, the parameters of interest include the effects of the covariates, the unobserved genotypes, of their allele proportions and the recombination fraction between the two loci. Let θ be the vector of all the parameters in the model, $\theta\hat{}$ be its M.L.E. from the mixture model. Consider the application of model (1) with the haplotype probability given by (5). Let H_0 be the hypothesis that there is no linkage between the two trait loci, i.e. H_0: $r = 0.5$. The results are shown in Table 2 below, where $\theta\hat{}$ and $\theta\hat{}0$ are the m.l.e. of θ under the full model and H_0 respectively (in brackets are the estimated standard deviations). In this case the hull hypothesis is $r = 0.5$ lies on the boundary of the parameter, instead of the standard likelihood ratio test, the 2 times log-likelihood ratio statistic in this case is asymptotically a 0.5:0.5 mixture of χ_0^2 and χ_1^2

[Self and Liang, 1987], which in our case is 0.485 with an approximate P-value of 0.2. Thus the hypothesis H_0 of no genetic linkage between fat mass and leptin is rejected at a high significance level (Table 2).

-2 log-likelihood ratio = 0.48494, with a P-value ≈ 0.2 under the 0.5:0.5 mixture of χ_0^2 and χ_1^2.

Discussion

We have considered a simple likelihood model to study linkage between traits and between trait and marker loci, without requiring IBD data as most linkage studies do, thus makes it easy to use for both the quantitative and qualitative traits. The hypothesis of no linkage can be tested by the standard likelihood ratio under this model. The model is applicable to pedigree, nuclear family or sib pairs, along with segregation and regressive analysis. Using this model to the GWA14 data, we find strong genetic linkage between ntth1 at some SNP loci.

The usual linkage analysis is based on the assumption of linkage equilibrium between loci, which is inappropriate. Some other approaches with combined linkage and linkage disequilibrium [4,15,23], this will yield more information. The disequilibrium may be specified as in (7) or some other measures as reviewed by Devlin and Risch [24]. But LD may be affected by many factors, such as mutation, drift, selection, population stratification or admixture, etc., which create difficulties in LD analysis. Our method can also be extended to this case along with Hardy-Weinberg disequilibrium, as in Wright [17] and Cockerham [18] and to different likelihood formulations, or even to semi-parametric and nonparametric models. We can also incorporate the multipoint marker information into the model to increase its power. Although we only presented the model for the two-allele case at each locus, this model cab be extended to multiple allele case, while the corresponding transmission matrix to that in Appendix A will be a real challenge. For two loci with k_1 and k_2 alleles each, one needs to compute a $(n_1n_2)^2 \times n_1n_2$ transmission matrix, where $n_i = k_i (k_i + 1)/2$ is the number of genotypes at the i-th locus. This can be partially resolved by a stepwise procedure; we can cut the loci to two alleles at each step, and then select the sections with stronger linkage for next step.

There is a trade-off between the effectiveness and robustness of methods. The non- parametric and semi-parametric models are in general robust since they require no or little model assumptions, but they may suffer from potential loss of efficiency by the same reason.

References

1. Risch N, Lange K. Application of a recombination model in calculating the variance of sib pair genetic identity. Ann Hum Genet. 1979; 43: 177-186.
2. Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. Science. 1995; 268: 1584-1589.
3. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet. 1994; 54: 535-543.
4. Schaid DJ, Rowland C. Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. Am J Hum Genet. 1998; 63: 1492-1506.
5. Zhao LP, Aragaki C, Hsu L, Quiaoit F. Mapping of complex traits by single-nucleotide polymorphisms. Am J Hum Genet. 1998; 63: 225-240.
6. Dudoit S, Speed TP. A score test for the linkage analysis of qualitative and quantitative traits based on identity by descent data from sib-pairs. Biostatistics. 2000; 1: 1-26.
7. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996; 58: 1347-1363.
8. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol. 2004; 26: 61-69.
9. Sung YJ, Thompson EA, Wijsman EM. MCMC-based linkage analysis for complex traits on general pedigrees: multipoint analysis with a two-locus model and a polygenic component. Genet Epidemiol. 2007; 31: 103-114.
10. Yuan A, Bonney GE. Two new recursive likelihood calculation methods for genetic analysis. Hum Hered. 2006; 54: 82-98.
11. Lange K. Statistics for Biology and Health. Springer-Verlag: New York, Inc. 1997.
12. Liang KY, Hsu FC, Beaty TH, Barnes KC. Multipoint linkage-disequilibrium-mapping approach based on the case-parent trio design. Am J Hum Genet. 2001; 68: 937-950.
13. Tienari PJ, Wikström J, Sajantila A, Palo J, Peltonen L. Genetic susceptibility to multiple sclerosis linked to myelin basic protein gene. Lancet. 1992; 340: 987-991.
14. Terwilliger JD, Ott J. Handbook of Human Genetic Linkage. The Johns Hopkins University Press. 1993.
15. Xiong M, Jin L. Combined linkage and linkage disequilibrium mapping for

genome screens. Genet Epidemiol. 2000; 19: 211-234.

16. Weir BS. Genetic Data Analysis II. Sinauer Associates, Inc., Publishers. 1996.
17. WRIGHT S. The genetical structure of populations. Ann Eugen. 1951; 15: 323-354.
18. Cockerham CC. Variance of gene frequencies. Evolution. 1969; 23: 72-84.
19. Sham P. Statistics in Human Genetics. Arnold. 1998.
20. Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, et al. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an auto some-wide scan in an American Indian population. Am J Med Genet. 1998; 81: 216-221.
21. Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, et al. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet. 1998; 81: 207-215.
22. Zinn-Justin A, Abel L. Genome search for alcohol dependence using the weighted pair wise correlation linkage method: interesting findings on chromosome 4. Genet Epidemiol. 1999; 17: 421-426.
23. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet. 1993; 52: 506-516.
24. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 1995; 29: 311-322.

Austin Biom and Biostat - Volume 1 Issue 1 - 2014
Submit your Manuscript | www.austinpublishinggroup.com

Citation: Yuan A, Zhong X and Bonney GE. A Likelihood Model for Linkage Analysis of Genetic Traits. Austin Biom and Biostat. 2014;1(1): 7.