Proceedings

Genetic imprinting analysis for alcoholism genes using variance components approach

Sanjay Shete* and Robert Yu

Address: Department of Epidemiology, Unit 1340, University of Texas M. D. Anderson Cancer Center, 1155 Hermann Pressler Boulevard, Houston, TX, 77030 USA

Email: Sanjay Shete* - sshete@mdanderson.org; Robert Yu - rkyu@mdanderson.org

* Corresponding author

Abstract
Genomic imprinting, which is also known as the parent-of-origin effect, is a mechanism that only expresses one copy of a gene pair depending upon the parental origin. Although many chromosomal regions in the human genome are likely to be imprinted, imprinting is not accounted for in the usual linkage analysis. In this study, using a variance-components approach with a quantitative phenotype ttth-FP1, we found significant evidence of imprinting at two loci, D7S1790 and D1S1631, on chromosome 1 and chromosome 7, respectively. Our results suggest that allowing for the possibility of imprinting can increase the power to detect linkage for localizing genes for alcoholism.

Background
Genomic imprinting (also known as the parent-of-origin effect) is a mechanism by which only one copy of a gene pair is expressed, and this expression is determined by the parental origin of the copy. The deregulation of imprinted genes has been implicated in a number of human diseases. Expression of imprinted genes is regulated by allele-specific epigenetic modifications of DNA and chromatin. These modifications affect central regulatory elements that control the allele-specific expression of neighboring genes. Although many chromosomal regions in the human genome are likely to be imprinted, particularly those involved in developmental disorders, imprinting is not accounted for in the usual linkage analysis [1-8]. In this summary, we analyzed the ttth-FP1 (far frontal left side channel), a quantitative measure of alcohol dependence, using the families provided by a multi-center consortium of the Collaborative Study on the Genetics of Alcoholism (COGA) [9,10].

Alcoholism is a complex disorder with involvement of genetic and environmental risk factors. Several studies have shown familial aggregation, segregation, and linkages to several regions [11]. Therefore, the purpose of our study was to evaluate the possibility of genomic imprinting in the regions that show some evidence of linkage using a recently developed method. Several regions on chromosomes 1 and 7 have been localized using parametric and nonparametric methods of linkage and association methods that do not allow for the possibility of genomic imprinting.

Methods
Variance components approach
Quantitative variation in a trait often occurs because of the underlying variation in genetic factors. We recently developed a method to analyze quantitative traits using the variance components approach and allowing for imprinting as described by Shete and Amos [3] and Shete et al. [4]. Let X_i be the phenotypic value for the ith individual in a pedigree:

$$X_i = \mu + g_i + G_i + \sum_{k=1}^{s} \beta_k Z_{i_k} + e_i$$

where μ is the overall mean, g_i is the major-gene effect, G_i is the polygenic effect, β_k values are covariate effects that are assumed to be uncorrelated with genetic and environmental factors, and e_i is the environmental effect. The major gene effect has a mean value of *a* when individual's genotype is BB, d_1 when the genotype is Bb, d_2 when the genotype bB, and *-a* when the genotype bb. Here, we assumed that the first allele is derived from the father and the second allele is derived from the mother. Let *d* be the dominance effect and *I* be the imprinting effect. Then, $d = (d_1 + d_2)/2$ and $I = (d_1 - d_2)/2$. When $d_1 = d_2$, there is no imprinting. Shete and Amos [3] decomposed genetic variance at this locus into three parts: an additive component due to the paternally derived allele, σ^2_{af}; an additive component due to the maternally derived allele, σ^2_{am}; and the usual dominance component, σ^2_d. These parent-specific additive components are:

$$\sigma^2_{a_f} = pq[(a+I) - d(p-q)]^2, \sigma^2_{a_m} = pq[(a-I) - d(p-q)]^2$$

where *p* and *q* are the frequencies of alleles *B* and *b*, respectively. Also, $\sigma^2_{af} + \sigma^2_{am} = \sigma^2_a$.

When the imprinting coefficient $I = 0$, σ^2_{af} and σ^2_{am} are equal to $\sigma^2_a/2$; and, when σ^2_{af} and σ^2_{am} are equal, $I = 0$. Hence, Shete and Amos [3] proposed that a test for the equality of these two parent-specific additive variances is a test for imprinting. In an extended pedigree, one must consider an allele that is shared IBD (identical by descent) by a pair of relatives in which one of the relatives received the copy from his/her father and the other received the copy from his/her mother. So, we define "parent-specific IBD sharing between a pair of relatives *i* and *j*" as follows:

$$\pi_{ff,ij} = \begin{cases} 1 = \text{if } i \text{ and } j \text{ share an allele IBD that is derived from their father} \\ 0 = \text{otherwise} \end{cases},$$

$$\pi_{fm,ij} = \begin{cases} 1 = \text{if } i \text{ and } j \text{ share an allele IBD that is derived father of } i \text{ and the mother of } j \\ 0 = \text{otherwise} \end{cases}.$$

We define $\pi_{mf,ij}$ and $\pi_{mm,ij}$ similarly. Then, the phenotypic covariance is given by [4]

$$Cov\left[x_i, x_j \middle| \left(\pi_{ffij}, \pi_{mmij}, \pi_{fmij}, \pi_{mfij}\right)\right] = \begin{cases} \sigma^2_{a_f} + \sigma^2_{a_m} + \sigma^2_d + \sigma^2_G + \sigma^2_e & \text{if } i = j \\ \pi_{ffij}\sigma^2_{a_f} + \pi_{mmij}\sigma^2_{a_m} + \left(\pi_{fmij} + \pi_{mfij}\right)\left(\sigma^2_a/2 - 2pqI^2\right) + \Delta_{ij}\sigma^2_d + \phi_{ij}\sigma^2_G & \text{if } i \neq j \end{cases}$$

From the above equation, it can be seen that the coefficients of $\pi_{ff,ij}$, $\pi_{mm,ij}$, and $(\pi_{fm,ij} + \pi_{mf,ij})$ are equal if and only if σ^2_{af} and σ^2_{am} are equal, and σ^2_{af} and σ^2_{am} are equal if and

only if the imprinting parameter $I = 0$ (i.e., there is no parental imprinting). Hence, the likelihood ratio test (LRT) for equality of these coefficients is a valid test for the null hypothesis of no imprinting. We do not estimate the parameters *p*, *q*, or *I* separately in the above equation, rather we estimate three parameters σ^2_{af}, σ^2_{am}, and $(\sigma^2_a/2 - 2pqI^2)$. Ordinarily, in a genome scan, one will test the joint null hypotheses of no linkage and no imprinting by testing $\sigma^2_{af} = \sigma^2_{am} = 0$.

Distribution of the LRT

The asymptotic distribution of the LRT is complex. For testing linkage without imprinting the LRT test is assumed to be a half-and-half mixture of χ^2 random variable with one and zero degrees of freedom. For joint testing of linkage and imprinting, we now have three parameters in the model. The two parameters σ^2_{af} and σ^2_{am} are independent; however, the third parameter $(\sigma^2_a/2 - 2pqI^2)$ is correlated with the first two parameters [4]. Because this parameter was used in our model, we used a mixture of χ^2 distribution with 0, 1, 2, and 3 degrees of freedom with mixing parameters of 1/8, 3/8, 3/8, and 1/8 for joint testing of linkage and imprinting following the same rationale as in the standard linkage analyses using the approach of Self and Liang [12]. Similarly, for testing the linkage model without imprinting against the linkage model allowing for imprinting we used a mixture of χ^2 distribution with 0, 1, and 2 degrees of freedom with mixing parameters of 4/8, 3/8, and 1/8. These asymptotic distributions can be used to obtain point-wise significance of the LRT test for testing linkage and/or imprinting.

Multipoint parent-specific IBD

Computation of multipoint parent-specific IBD is described by Shete et al. [4]. There are four alleles at a single locus for the relative pair *i* and *j*. The two alleles for individual *i* are denoted by a vector (i_m, i_f), where i_m and i_f are maternal and paternal alleles, respectively. Similarly, we define the vector (j_m, j_f) for individual *j*. There are 15 possible ordered states of IBD between these two individuals [13]. Of these 15 states, only 7 are essential for computation of IBD sharing in outbred populations. Using the notations of SIMWALK2 [14,15], we define probabilities of these states as $S_9 = (i_m j_m)(i_f j_f)$, $S_{10} = (i_m j_m)(i_f)(j_f)$, $S_{11} = (i_m)(i_f j_f)(j_m)$, $S_{12} = (i_m j_f)(i_f j_m)$, $S_{13} = (i_m j_f)(i_f)(j_m)$, $S_{14} = (i_m)(i_f j_m)(j_f)$, and $S_{15} = (i_m)(i_f)(j_m)(j_f)$. In these states, the pairs of alleles inside the parentheses are IBD. We used SIMWALK2 to obtain these 15 detailed states of identity sharing.

Data

To identify imprinted genes that affect the risk for alcoholism, we proposed to study 143 families consisting of a total of 1,614 individuals. We used the quantitative trait denoted by ttth1-FP1, far frontal left side channel, a quan-
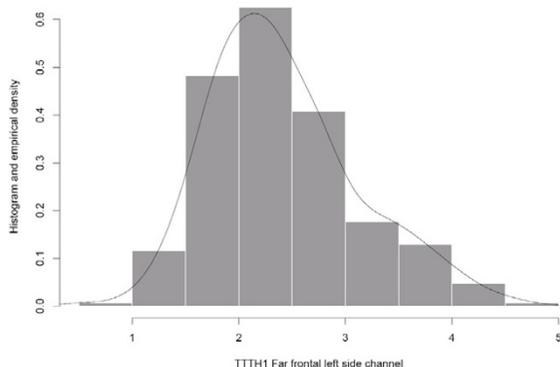
Figure 1
Empirical distribution of ttth1-FP1, far frontal left side channel.

titative measure of alcoholism. The empirical distribution of this quantitative trait is shown in the Figure 1. Because the normality of the trait distribution is underlying assumption in our variance component approach, we performed standardization of the data by subtracting mean and dividing by standard deviation followed by winsorization at a cut-off value of 1% from both side of the empirical distribution. Data winsorization has been shown to reduce the false-positive rate using variance components approach [16,17]. Finally, we used the residuals obtained from a polygenic model with sex as a covariate as our quantitative trait. From previous Genetic Analysis Workshop 11, two regions on chromosome 1 and a region chromosome 7 were selected to illustrate the utility of the method for detecting imprinting. These regions showed some evidence of linkage using standard model with at least one of the COGA defined phenotypes [11]. The markers we considered were D1S532, and D1S1631 on chromosome 1, and D7S1790 on chromosome 7 from Genetic Analysis Workshop 14 data.

Results and Discussion
For each of the markers listed, we calculated multipoint parent-specific IBDs using the methods described. On chromosome 7, for the marker D7S1790, we obtained a negative log likelihood value of 140.35 for the model in which the major gene variance component was fixed to zero, the model without linkage. The same value under the linkage without imprinting model was found to be 135.84. Using the LRT discussed above, we obtained a suggestive significant *p*-value of 0.00129 for linkage. Furthermore, we obtained a negative log likelihood value of 129.95 using the variance components approach that tests for linkage allowing for imprinting. Using the LRT, the significance for joint testing of linkage and imprinting at this marker is 0.00003. When we compared the log likelihoods with linkage but no imprinting model with joint

linkage and imprinting model, we obtain a *p*-value of 0.00057, which is significant evidence of imprinting. The evidence of imprinting is also evident from the lower *p*-value obtained using the imprinting model.

On chromosome 1, for the marker D1S1631, we obtained a negative log likelihood value of 136.90 for the model with linkage without imprinting. This gives a significance value of 0.00417 for linkage. A negative log likelihood value of 131.90 was obtained for the linked model with imprinting. This leads to a significant *p*-value of 0.00018 for joint testing of linkage and imprinting. When we compared the log likelihoods with linkage but no imprinting model with joint linkage and imprinting model, we obtain a *p*-value of 0.00143, which is a suggestive significant evidence for imprinting. Evidence of imprinting for marker D1S532 was not very significant, as shown in the Table 1. These results showed that a genome scan approach using a linkage model that allows for imprinting is important for alcoholism. There may be other regions on the genome that did not show evidence for linkage using standard models but could show significance if imprinting is allowed. Therefore, we recommend a genome scan with an imprinting model as oppose to just testing for the regions showing linkage using the standard models. Shete and Amos [3] note that it is important to allow for the sex-specific recombination fraction while testing for linkage allowing imprinting to reduce the false-positive rate.

Conclusion
Imprinting is not accounted for traditional linkage analyses. We found evidence of imprinting, even allowing for the multiple testing, on two loci. It may also be important to allow for other covariates of environmental exposures, such as smoking, in the model. In addition, the asymptotic distribution that we used may not be very accurate and recommend simulation-based *p*-values at the significant loci to confirm evidence of linkage. In conclusion, our results suggest that allowing for imprinting in the linkage analyses can increase the power to detect genes responsible for the alcoholism.

Abbreviations
COGA: Collaborative Study on the Genetics of Alcoholism

LRT: Likelihood ratio test

IBD: Identity by descent

Authors' contributions
SS conceived the study, participated in the statistical analyses, and drafted the manuscript. RY participated in the statistical analyses and helped draft the manuscript.

Table 1: Significance of likelihood ratio test for linkage and imprinting.

Marker	L(unlinked)	L(linked)	L(linkage and imprinting)	*p*-Value (linkage)	*p*-Value (linkage and imprinting)	*p*-Value (imprinting)
D7S1790	140.38	135.84	129.95	0.00129	0.00003	0.00057
D1S1631	140.38	136.90	131.90	0.00417	0.00018	0.00143
D1S532	140.38	137.35	134.66	0.00700	0.00270	0.01612

References

1. Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP: **Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization.** *Am J Hum Genet* 2000, **66**:1945-1957.
2. Hanson RL, Kobes S, Lindsay RS, Knowler WC: **Assessment of parent-of-origin effects in linkage analysis of quantitative traits.** *Am J Hum Genet* 2001, **68**:951-962.
3. Shete S, Amos CI: **Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting.** *Am J Hum Genet* 2002, **70**:751-757.
4. Shete S, Zhou X, Amos CI: **Genomic imprinting and linkage test for quantitative-trait loci in extended pedigrees.** *Am J Hum Genet* 2003, **73**:933-938.
5. Wu CC, Shete S, Amos CI: **Linkage analysis of affected sib pairs allowing for parent-of-origin effects.** *Ann Hum Genet* 2005, **69**:113-126.
6. Knapp M, Strauch K: **Affected-sib-pair test for linkage based on constraints for identical-by-descent distributions corresponding to disease models with imprinting.** *Genet Epidemiol* 2004, **26**:273-285.
7. Mukhopadhyay N, Weeks DE: **Linkage analysis of adult height with parent-of-origin effects in the Framingham Heart Study.** *BMC Genet* 2003, **4(Suppl 1)**:S76.
8. Shete S, Zhou X: **Parametric approach to genomic imprinting analysis with applications to Angelman's syndrome.** *Hum Hered* 2005, **59**:26-33.
9. Reich T: **A genomic survey of alcohol dependence and related phenotypes: results from the Collaborative Study on the Genetics of Alcoholism (COGA).** *Alcohol Clin Exp Res* 1996, **20**:133A-137A.
10. Begleiter H, Reich T, Nurnberger J Jr, Li TK, Conneally PM, Edenberg H, Crowe R, Kuperman S, Schuckit M, Bloom F, Hesselbrock V, Porjesz B, Cloninger CR, Rice J, Goate A: **Description of the Genetic Analysis Workshop 11 Collaborative Study on the Genetics of Alcoholism.** *Genet Epidemiol* 1999, **17(Suppl 1)**:S25-S30.
11. Almasy L, Borecki IB: **Exploring genetic analysis of complex traits through the paradigm of alcohol dependence: summary of GAW11 contributions.** *Genet Epidemiol* 1999, **17(Suppl 1)**:S1-24.
12. Self SG, Liang K-Y: **Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions.** *J Am Stat Assoc* 1987, **82**:605-610.
13. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits Sunderland: Sinauer Associates*; 1997.
14. Weeks DE, Sobel E, O'Connell JR, Lange K: **Computer programs for multilocus haplotyping of general pedigrees.** *Am J Hum Genet* 1995, **56**:1506-1507.
15. Sobel E, Lange K: **Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics.** *Am J Hum Genet* 1996, **58**:1323-1337.
16. Shete S, Beasley TM, Etzel CJ, Fernandez JR, Chen J, Allison DB, Amos CI: **Effect of winsorization on power and type 1 error of variance components and related methods of QTL detection.** *Behav Genet* 2004, **34**:153-159.
17. Etzel CJ, Shete S, Beasley TM, Fernandez JR, Allison DB, Amos CI: **Effect of Box-Cox transformation on power of Haseman-Elston and maximum-likelihood variance components tests to detect quantitative trait loci.** *Hum Hered* 2003, **55**:108-116.