Effect of Linkage between vWA and D12S391 in Kinship Analysis

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Email:
kristen.oconnor@nist.gov

Kristen L. O'Connor[1], Andreas O. Tillmar[2], and John M. Butler[1]

[1] U.S. National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899-8314
[2] National Board of Forensic Medicine, Department of Forensic Genetics and Forensic Toxicology, Linköping, Sweden

ISFG

P-181

The expanded European Standard Set of loci includes D12S391, which is found on the short arm of chromosome 12 and is only 6.3 megabases (Mb) from the established vWA locus that is widely used in Europe and the U.S. Ideally for use in forensic analyses, genetic markers on the same chromosome should be more than 50 Mb in physical distance in order to ensure full recombination and thus independent inheritance. Recent studies have shown no significant linkage disequilibrium between vWA and D12S391 in U.S. and worldwide populations [1-3], although genetic linkage has been found (estimated recombination fraction of 0.108 using 28 meioses) [2]. We evaluated genetic linkage in six multi-generation families (109 meioses) using the program LINKAGE [4]. The recombination fraction was estimated at 0.089 (95% CI 0.044-0.158). The effect of linkage is an increased tendency for alleles at physically close loci to be transmitted together during meiosis. For linked loci, the use of haplotype frequencies and the recombination fraction has been recommended in likelihood calculations for kinship testing [5]. Moreover, ignoring linkage in likelihood calculations may lead to incorrect conclusions of kinship in some instances [6]. Using pedigree simulations, we evaluated the impact of including or ignoring linkage between vWA and D12S391 on likelihood ratio values. We demonstrate the case-specific impact on likelihood ratio values when the "incorrect" model is used.

Estimation of Linkage between D12S391 and vWA

Methods

Logarithm of odds (LOD score) analysis [7] is used to estimate the most probable value for the recombination rate between two loci using family genotype data. The LOD score (often called Z) is the ratio of Likelihood(rec)/Likelihood(rec=0.5). In other words, the LOD score determines how much more likely it is that two markers are linked with a certain recombination frequency of "rec" compared to being unlinked (where rec is 0.5). In linkage analysis, the value of "rec" can be altered from 0 to 0.5. The value that maximizes the LOD score is the most probable recombination frequency.

Linkage analysis was performed using the program LINKAGE [4]. Six multi-generation Caucasian families (109 meioses total) were genotyped at D12S391 and vWA. For some meioses, the phase was unknown (leading to non-integer values for k and n). U.S. Caucasian allele frequencies [8] were used to estimate the recombination rate for non-informative meioses (e.g., homozygous parent or both parents had same genotype). The support interval was computed based on the 1-lod method [9] and a 95% confidence interval was calculated with the BINOM program [7].

Results

LOD score, Z_{Max}[a]	18.7
Z_{Max}-1	17.7
Recombination fraction at Z_{Max}[a]	0.089
# recombinants (k)[b]	9.75
# meioses (n)[b]	109.5
Support interval (estimated from Z_{Max}-1)	0.05-0.15
Confidence interval (95 %)[c]	0.044-0.158

[a] Estimated using *LINKAGE*
[b] Estimated using *lods*
[c] Estimated using *binom*

Conclusions

A linkage study by Budowle et al. tested 28 meioses and estimated the recombination frequency between D12S391 and vWA as 10.8 % [2].

- No confidence interval was reported

Using 109 meioses, we refined the estimated recombination frequency as 8.9 %.

- 95 % confidence interval is 4.4 – 15.8 %

Large-scale recombination patterns are conserved across human populations [10].

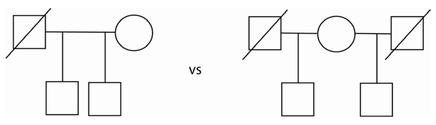
- Might expect the recombination frequency between D12S391 and vWA to be similar across major worldwide populations

Impact of Linkage and Linkage Disequilibrium Assumptions on Computed Likelihood Ratios for Different Case Scenarios

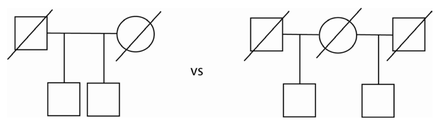
Methods

In order to assess the effects of linkage and linkage disequilibrium (LD) between D12S391 and vWA on likelihood ratio (LR) values, we performed a simulation study with four different case scenarios:

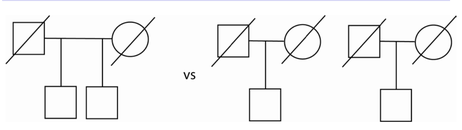
Pedigree 1. Full siblings with mother (H1) vs. Half siblings with shared mother (H2)



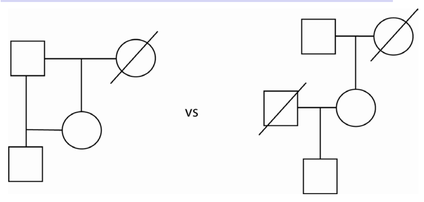
Pedigree 2. Full siblings (H1) vs. Half siblings (H2)



Pedigree 3. Full siblings (H1) vs. Unrelated (H2)



Pedigree 4. Incest (H1) vs. No incest (H2)



Dataset

For each case scenario, genotypes of the involved individuals were simulated 10,000 times under each hypothesis.

Founder DNA profiles were generated with haplotype frequencies taken from O'Connor et al. [1] and 0.089 was used as the recombination frequency (θ) between D12S391 and vWA.

Algorithms

Two likelihood ratio comparisons were carried out to measure the impact on computed LR values by (1) assuming linkage equilibrium (LE) versus LD, or (2) ignoring linkage versus accounting for linkage.

1. **LE vs. LD Comparison**
   LRs calculated using **allele frequencies** and θ = 0.089 ("rough LD") with LRs calculated using **haplotype frequencies** and θ = 0.089 ("exact LD").

   If there is no influence of LD between D12S391 and vWA, then we expect the LR(rough LD) to equal the LR(exact LD).

2. ***Unlinked vs. Linked Comparison***
   Compared LRs calculated with $\boldsymbol{\theta = 0.5}$ and allele frequencies ("rough linked") with LRs calculated with $\boldsymbol{\theta = 0.089}$ and allele frequencies ("exact linked").

Haplotype frequencies were estimated from Tillmar et al. [11] using observed haplotypes from O'Connor et al. [1] and allele frequencies from Hill et al. [8].

Correlation between exact and rough estimates for the different case scenarios was tested with Pearson's correlation coefficient (r).

References

[1] K.L. O'Connor, et al., Corrigendum to Linkage disequilibrium analysis of D12S391 and vWA in U.S. population and paternity samples. Forensic Sci. Int. Genet. (in press), doi:10.1016/j.fsigen.2010.09.003.

[2] B. Budowle, et al., Population genetic analyses of the NGM STR loci. Int. J. Legal Med. 2011;125:101-109.

[3] C. Phillips, et al., Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. Forensic Sci. Int. Genet. 2011;5(3):155-169.

[4] LINKAGE Program: http://linkage.rockefeller.edu/soft/linkage/

[5] M. Krawczak, Kinship testing with X-chromosomal markers: mathematical and statistical issues. Forensic Sci. Int. Genet. 2007;1(2):111-114.

[6] M. Nothnagel et al., Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. Int. J. Legal Med. 2010;124(3):205-215.

[7] J. Ott, Analysis of Human Genetic Linkage, third ed., The Johns Hopkins University Press, Baltimore, 1999.

[8] C.R. Hill, et al., Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. Forensic Sci. Int. Genet. 5(4) (2011) 269-75.

[9] P.M. Conneally, et al., Report of the Committee on Methods of Linkage Analysis and Reporting. Cytogenet. Cell Genet. 40(1-4) (1985) 356-9.

[10] D. Serre, et al., Large-scale recombination rate patterns are conserved among human populations. Genome Res. 15(11) (2005) 1547-52.

[11] A.O. Tillmar, et al., Using X-chromosomal markers in relationship testing: Calculation of likelihood ratios taking both linkage and linkage disequilibrium into account. Forensic Sci. Int. Genet. (in press), doi:10.1016/j.fsigen.2010.11.004.

Funding
Support for K. O'Connor provided by the National Research Council Research Associateship Program.

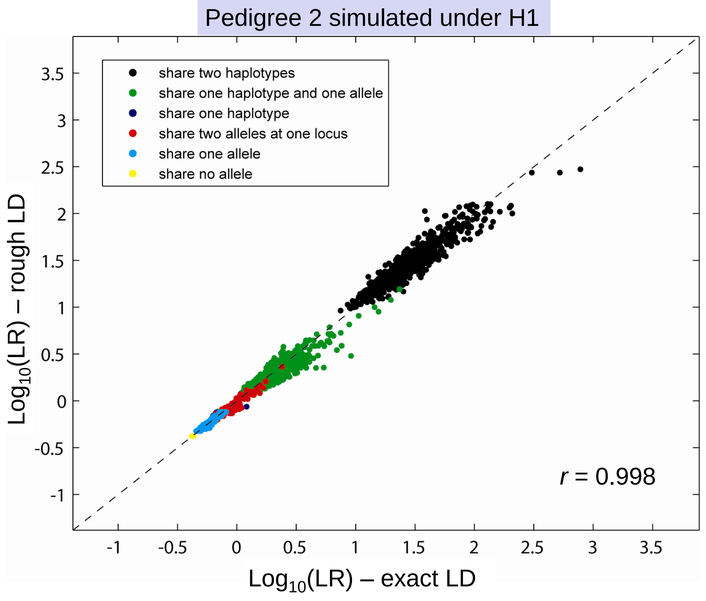
Poster available for download from STRBase
http://www.cstl.nist.gov/strbase/pub_pres/OConnorISFG2011poster.pdf

Results and Conclusions

Measuring the Impact of Assuming LE or LD

Results for LE vs. LD comparisons were similar for all case scenarios and hypotheses. All *r*-values were greater than 0.96. Thus, one representative figure is provided below.

Pedigree 2 simulated under H1



If there is a random association between alleles at D12S391 and vWA, then we expect the LR(rough LD) value to equal the LR(exact LD) value.

- When plotted against each other, these values would fall along the y=x-axis (dashed line in plot).

Assuming LE or LD has only a subtle effect on the LR values when linkage is taken into account.

Thus, allele or haplotype frequencies can be used for LR calculations involving D12S391 and vWA.

- If one assumes LE between D12S391 and vWA, it is easier to infer haplotype frequencies from allele frequencies (particularly for unobserved haplotypes in population samples).

Measuring the Impact of Ignoring Linkage

Table 1. Difference (computed as a ratio) between LR values calculated assuming unlinked markers (θ = 0.5) or assuming linkage (θ = 0.089) between D12S391 and vWA. Difference ratio = LR(unlinked)/LR(linked).

Case scenario	Simulated under H1			Simulated under H2		
	median	minimum	maximum	median	minimum	maximum
Pedigree 1	0.72	0.23	3.14	0.60	0.42	4.18
Pedigree 2	0.86	0.60	2.82	0.86	0.60	3.03
Pedigree 3	0.84	0.36	7.31	0.84	0.36	6.76
Pedigree 4	0.63	0.14	25.00	1.00	0.24	31.44

If one assumes that linkage between D12S391 and vWA has no effect on the computed LR value, then the difference ratio would equal one.

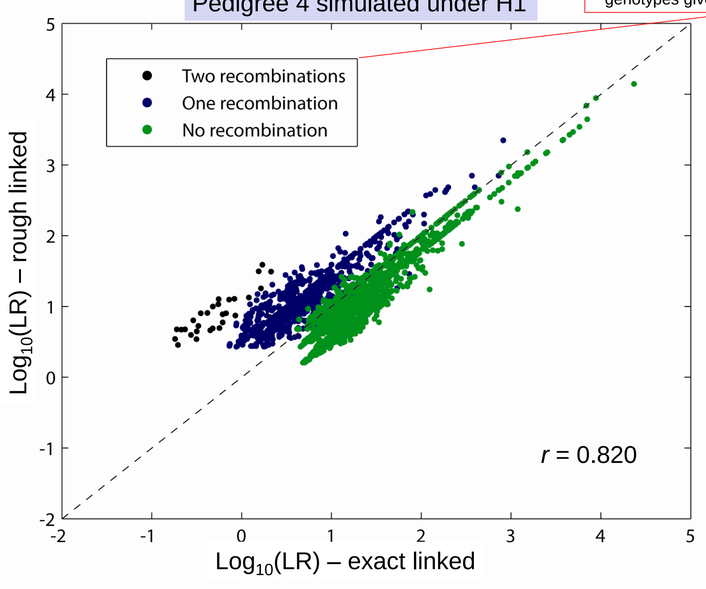
If linkage is ignored between D12S391 and vWA, then difference ratio values greater than one indicate overestimates of the real LR value, while difference ratio values less than one indicate underestimates of the real LR value.

On average, for all pedigree scenarios, LR values ignoring linkage were slightly underestimated than when linkage was considered. However, in extreme cases LR values could be overestimated up to 3-7 times if linkage was ignored for Pedigrees 1-3.

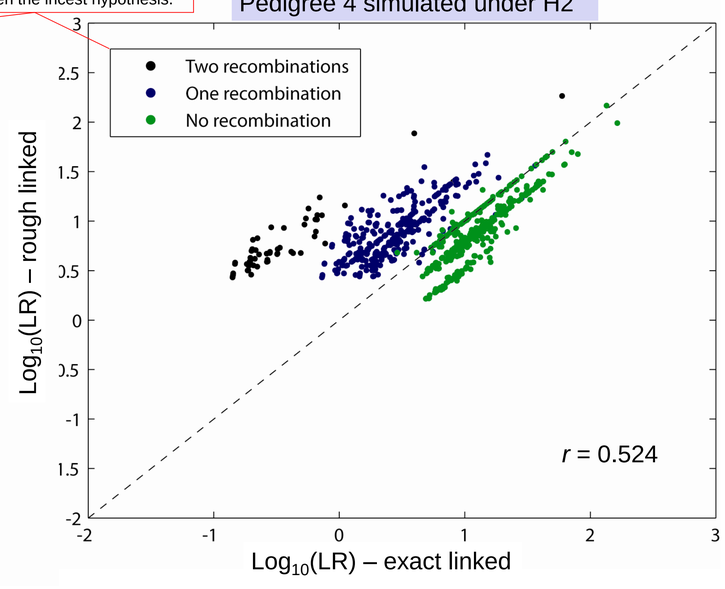
The incest case (Pedigree 4) displays the most extreme differences in LR values when linkage is ignored. The figures below illustrate the effect that recombination patterns have on LR values when accounting for linkage or not.

Refers to the minimum number of recombination(s) necessary to explain the genotypes given the incest hypothesis.

Pedigree 4 simulated under H1



Pedigree 4 simulated under H2



Approximately 90% of cases excluded the grandfather as the father of the child (LR = 0)

If linkage is ignored in a true incest case when:

No recombination is necessary to explain the genotype data, then the strength of the relationship may be underestimated by 1/7.

Recombination(s) have occurred, then the strength of the relationship may be overestimated up to 25 times.

If linkage is ignored in a false incest case when:

No recombination is necessary to explain the genotype data, then the strength of the relationship may be underestimated by 1/4.

Recombination(s) have occurred, then the strength of the relationship may be overestimated up to 30 times.

Overall, ignoring linkage between D12S391 and vWA has a greater impact on LR values for incest cases than for other more common kinship scenarios, particularly when recombination(s) have occurred between the markers.