



# Linkage disequilibrium measures for fine-scale mapping of disease loci are revisited

Carlos Zapata\*

Department of Genetics, University of Santiago de Compostela, Santiago de Compostela, Spain  
\*Correspondence: c.zapata@usc.es

**Edited by:**

Xuefeng Wang, Harvard University, USA

**Reviewed by:**

Alexandre Bureau, Université Laval, Canada

**Keywords:** genome-wide association studies, linkage disequilibrium and genetic recombination, linkage disequilibrium between diallelic loci, new disease mutations, 2 × 2 contingency tables

## A commentary on

### A comparison of linkage disequilibrium measures for fine-scale mapping

by Devlin, B., and Risch, N. (1995). *Genomics* 29, 311–322. doi: 10.1006/geno.1995.9003

Genome-wide association (GWA) studies for fine-scale mapping of disease loci are conceptually based on the occurrence and history of non-random association of alleles at different loci in populations (or linkage disequilibrium). Thus, a new disease mutation arises on a single ancestral haplotype of a given population in complete linkage disequilibrium with syntenic polymorphic markers. Owing to historical recombination events, markers close to the disease locus will tend to be in strongest disequilibrium than distant markers when the presence of evolutionary mechanisms generating non-uniform disequilibrium patterns is ignored. Therefore, the use of optimal measures to assess gradients of the strength of disequilibrium along chromosomes is crucial for fine mapping of disease loci. In practice, an ideal disequilibrium measure for fine-scale mapping should be a monotone function of the recombination fraction between the marker and the disease locus.

In an influential paper, Devlin and Risch (1995, henceforth referred to as “D&R”) compared the efficiency of five measures of linkage disequilibrium used to refine the location of disease loci: the robust formulation of the population attributable risk  $\delta$  (Lewin and Bertell, 1978), Lewontin’s  $D'$  (Lewontin, 1964), the correlation coefficient  $\Delta$  (Hill and Robertson, 1968; also described in the literature as  $r$ ), Yule’s  $Q$  (Nei and Li, 1980),

and Kaplan and Weir’s  $d$  (Kaplan and Weir, 1992). It was concluded that the measures  $\delta$  and  $D'$  outperform all other measures for fine-scale mapping. In particular, D&R established analytically that  $\delta$  is the ideal measure for fine mapping because changes of  $\delta$  over generations are directly related to the recombination fraction ( $\theta$ ) between the disease and the marker locus. In contrast, they found that  $D'$  not only depends on  $\theta$ , but also on the haplotype frequencies. This supposed advantage of  $\delta$  over  $D'$  for fine-scale mapping of disease loci has become a classical paradigm in the literature on disequilibrium. Nevertheless, D&R’s conclusion about the relative merits of  $\delta$  and  $D'$  for fine-scale mapping based on analytic results is erroneous. As is shown below, the source of the error lies in a misinterpretation of the formulas used by D&R to assess the relationship between the decay of disequilibrium by recombination and the measures  $\delta$  and  $D'$ . Despite the time elapsed, this misconception remains uncorrected.

Following D&R, let us consider that there are two alleles at each of two loci: a disease allele and a normal allele segregate at the first locus, and two marker alleles segregate at the other locus. The layout and notation for population haplotype, marker allele and disease allele frequencies are given in **Table 1**. The measures  $\delta$  and  $D'$  are defined as  $\delta = D/\pi_{+1}\pi_{22}$  and  $D' = D/D_{\max}$ , respectively; where  $D = \pi_{11}\pi_{22} - \pi_{12}\pi_{21}$  and  $D_{\max}$  is the lesser of  $\pi_{1+}\pi_{2+}$  and  $\pi_{+1}\pi_{2+}$  when  $D > 0$ , or the lesser of  $\pi_{1+}\pi_{+1}$  and  $\pi_{2+}\pi_{+2}$  when  $D < 0$ . Under a deterministic population model, D&R assumed initial complete linkage disequilibrium between disease and marker loci; no change of disease allele and marker allele frequencies over time;

**Table 1 | Layout and notation for population haplotypes, marker alleles, and disease allele frequencies in a 2 × 2 table.**

Marker	Disease allele	Normal allele	
$A_1$	$\pi_{11}$	$\pi_{12}$	$\pi_{1+}$
$A_2$	$\pi_{21}$	$\pi_{22}$	$\pi_{2+}$
	$\pi_{+1}$	$\pi_{+2}$	1

the disease allele arose in a haplotype carrying the allele  $A_1$  at the marker locus, so that  $\pi_{21} = 0$ ,  $\pi_{11} = \pi_{+1}$ ,  $\pi_{22} = \pi_{2+}$  and  $D > 0$  at generation 0; and  $\theta$  between the disease and the marker locus is constant along generations. After  $n$  generations, the decay of the initial disequilibrium ( $D_0$ ) can be obtained as a function of  $\theta$  by means of the expression  $D_n = (1 - \theta)^n D_0$ , where  $D_n$  is the value of disequilibrium in the  $n$ th generation (Hedrick, 2005). Expressing the measures  $\delta$  and  $D'$  in terms of  $(1 - \theta)^n$ , D&R obtained then the following mathematical relationship:

$$(1 - \theta)^n = D_n/D_0 = [\pi_{11}\pi_{22} - \pi_{12}\pi_{21}] / \pi_{+1}\pi_{22} = \delta \quad (1)$$

where  $\pi_{+1}\pi_{22}$  was considered to be the best estimate of  $D_0$ , the initial amount of linkage disequilibrium, given that  $\pi_{21} = 0$  at generation 0 and hence  $\pi_{11} = \pi_{+1}$ . In contrast:

$$(1 - \theta)^n = D'[1 + (\pi_{21}/\pi_{22})] \quad (2)$$

D&R concluded, therefore, that the measure  $\delta$  is a function of  $\theta$  only, whereas the relationship between  $D'$  and  $\theta$  depends on haplotype frequencies. Nevertheless, the formulas (1) and (2) are conceptually erroneous. Any measure of the strength of

disequilibrium is computed from haplotype frequencies at a given generation. Note, however, that the numerator and the denominator in the formula (1) refer to haplotype frequencies at generations  $n$  and  $0$ , respectively. As initial complete linkage disequilibrium between disease locus and marker locus decays from  $D > 0$  to  $D = 0$ , coupling ( $\pi_{11}$  and  $\pi_{22}$ ) and repulsion ( $\pi_{12}$  and  $\pi_{21}$ ) haplotype frequencies decrease and increase, respectively, by a proportion  $\theta$  each generation. Accordingly, haplotype frequencies at generations  $0$  and  $n$  are distinct and they should be denoted differently. Distinguishing haplotype frequencies at generations  $0$  and  $n$  by  $\pi_{ij}(0)$  and  $\pi_{ij}(n)$ , respectively, where  $i, j = \{1, 2\}$ ; the formula (1) can be then rewritten as follows:

$$\begin{aligned}(1 - \theta)^n &= D_n/D_0 \\ &= [\pi_{11}(n)\pi_{22}(n) - \pi_{12}(n)\pi_{21}(n)] / \\ &\quad \pi_{+1}\pi_{22}(0) = \delta[\pi_{22}(n)/\pi_{22}(0)]\end{aligned}$$

Therefore, the relationship between  $\delta$  and  $\theta$  depends on the frequencies of the haplotype  $\pi_{22}$  at generations  $n$  and  $0$ . In terms of  $D'$ , the formula (1) can be rewritten as:

$$\begin{aligned}(1 - \theta)^n &= D_n/D_0 \\ &= D_n/\pi_{+1}\pi_{22}(0) \\ &= D_n/\pi_{+1}\pi_{2+} \\ &= D_n/D_{\max} = D'_n\end{aligned}$$

given that  $\pi_{22}(0) = \pi_{2+}$  and  $D_{\max} = \pi_{+1}\pi_{2+}$  when  $D > 0$  and  $\pi_{+1}\pi_{2+} < \pi_{1+}\pi_{+2}$ . It is shown, therefore, that  $D'$  is directly related to the recombination fraction. Likewise, the formula (2) can be rewritten as:

$$(1 - \theta)^n = D'_n \{1 + [\pi_{21}(0)/\pi_{22}(0)]\} = D'_n$$

given that  $\pi_{21}(0) = 0$ . D&R studied the relationship between  $D'$  and  $\theta$  when  $\pi_{1+} < \pi_{+1}$ . Note, however, that  $\pi_{1+} = [\pi_{11}(0) + \pi_{12}(0)]$  cannot be lower than  $\pi_{+1} = [\pi_{11}(0) + \pi_{21}(0)] = \pi_{11}(0)$ .

Our reanalysis on the relationship of the recombination fraction with the measures of disequilibrium  $\delta$  and  $D'$  thus, contradicts the contention by D&R that  $\delta$  outperforms  $D'$  for fine-scale mapping because changes of  $\delta$  over generations are directly related to the recombination fraction. In fact, we reached the opposite conclusion to that of D&R. This finding reinforces the view that  $D'$  exhibits better statistical properties as a general measure of linkage disequilibrium than other commonly used measures (Zapata, 2011). In particular,  $D'$  seems to be an optimal measure for mapping of marker association and localization of disease loci sensu D&R.

## REFERENCES

Devlin, B., and Risch, N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 29, 311–322. doi: 10.1006/geno.1995.9003

Hedrick, P. W. (2005). *Genetics of Populations*. Boston: Jones and Bartlett Publishers.

Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622

Kaplan, N., and Weir, B. S. (1992). Expected behavior of conditional linkage disequilibrium. *Am. J. Hum. Genet.* 51, 333–343.

Lewin, M. L., and Bertell, R. (1978). Re: "Simple estimation of population attributable risk from case-control studies." *Am. J. Epidemiol.* 108, 78–79.

Lewontin, R. C. (1964). The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 49, 49–67.

Nei, M., and Li, W.-H. (1980). Non-random associations between electromorphs and inversion chromosomes in finite populations. *Genet. Res.* 35, 65–83. doi: 10.1017/S001667230001394X

Zapata, C. (2011). On the uses and applications of the most commonly used measures of linkage disequilibrium from the comparative analysis of their statistical properties. *Hum. Hered.* 71, 186–195. doi: 10.1159/000327732

Received: 30 July 2013; accepted: 17 October 2013; published online: 05 November 2013.

Citation: Zapata C (2013) Linkage disequilibrium measures for fine-scale mapping of disease loci are revisited. *Front. Genet.* 4:228. doi: 10.3389/fgene.2013.00228

This article was submitted to *Statistical Genetics and Methodology*, a section of the journal *Frontiers in Genetics*.

Copyright © 2013 Zapata. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.