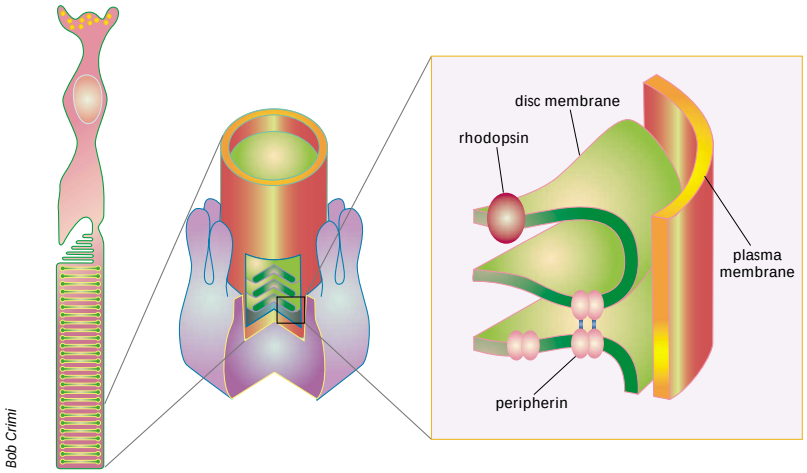

Fig. 2 The rod outer segment, which contains rhodopsin and the biochemical machinery to convert a photon to a chemical signal, is embedded into the retinal pigment epithelium. Peripherin is an adhesion molecule localized to the disc rim, where it assembles to maintain the structural integrity of the outer segment. (*This figure is drafted from one appearing in Trends In Genetics*[14] *with permission from Elsevier Science.)*

tional biochemical retinoid regeneration pathway is yet to be had. Prolonged exposure to light may effect degeneration of the treated cells if they cannot exchange retinoids with the RPE.

How does expression of the transgene activate the developmental program that restores the outer segment, both in terms of its architecture and its functional interactions with both the RPE and retinal circuitry? It might be argued that gene therapy works in the *rds* mouse because the outer segment continually regenerates, even in the adult animal. If this is the case, it will be key to introduce the transgene before apoptosis of the photoreceptor cell—the probable point of no return for therapeutic approaches that don't involve retinal transplantation[12].

How is gene therapy as applied to the *rds* mouse relevant to understanding retinal degenerative disorders in general? The study by Ali *et al.* indicates that replacement of the missing gene may restore function relatively late in the disease process, even in cases where the lack of a structural gene prevents complete rod cell development. This may contrast with some genetically dominant disorders in which the outer segment initially develops but then degenerates over time, after which the photoreceptor cell undergoes apoptosis. The question of whether a damaged outer segment can be restored to 'full health' remains open. Ribozyme-targeted degradation of mutant RNA may be one approach to treat dominant retinal degenerations[13].

The complex ontogeny, anatomy and physiology of the retina make it a common site of genetic disease, not to mention a fascinating structure. With the advent of gene therapy and appropriate models of retinal dystrophies come new ways to probe its development, form and function. Its tantalizing accessibility makes it even harder to resist as an object of wonder and awe. □

1. Ali, R.R. *et al. Nature Genet.* **25**, 306–310 (2000).
2. Jansen, H.G. & Sanyal, S. *J. Comp. Neurol.* **224**, 71–84 (1984).
3. Travis, G.H. *et al. Nature* **338**, 70–73 (1989).
4. Connell, G. *et al. Proc. Natl Acad. Sci. USA* **88**, 723–726 (1991).
5. Goldberg, A.F.X. & Molday, R.S. *Proc. Natl Acad. Sci. USA* **93**, 13726–13730 (1996).
6. Travis, G.H., Groshan, K.R., Lloyd, M. & Bok, D. *Neuron* **9**, 113–119 (1992).
7. Takahashi, M., Miyoshi, H., Verma, I.M. & Gage, F.H. *J. Virol.* **73**, 7812–7816 (1999).
8. Bennett, J. *et al. Nature Med.* **2**, 649–654 (1996).
9. Dowling, J.E. *Nature* **168**, 114–118 (1960).
10. Saari, J.C. *Invest. Ophthalmol. Vis. Sci.* **41**, 337–348 (2000).
11. Wright, A.F. *Nature Genet.* **17**, 132–134 (1997).
12. Gouras, P. *et al. Invest. Ophthalmol. Vis. Sci.* **35**, 3145–3153 (1994).
13. Lewin, A.S. *et al. Nature Med.* **4**, 967–971 (1998).
14. Gregory-Evans, I. & Bhattacharya, S.S. *Trends Genet.* **14**, 103–109 (1998).

A look at linkage disequilibrium

Michael Boehnke

Department of Biostatistics, University of Michigan School of Public Health, 1420 Washington Heights, Ann Arbor, Michigan 48109-2029, USA.
e-mail: boehnke@umich.edu

The extent of linkage disequilibrium critically determines the efficiency of strategies to identify genetic variants that predispose to human disease. New data indicate that the extent of disequilibrium is highly variable across the genome, and that differences in disequilibrium levels between isolated and mixed populations are modest.

Identifying the genetic variants that predispose to common human diseases such as heart disease, diabetes, hypertension and depression is a major goal of human geneticists. The primary strategy to identify these variants is to test for linkage in families with 300 to 400 multiallelic short tandem repeat (STR) markers and to follow up regions of suggestive linkage with a dense set of markers to better localize the variants. The development of a dense map of biallelic single-nucleotide polymorphisms (SNPs) will facilitate this process, and support the alternative strategy of a genome-wide search for linkage disequilibrium using simpler families or cases and controls[1]. For these reasons, a map of approximately 100,000 SNPs was identified as a goal of the current five-year plan of the Human Genome Project[2]. The plan also calls for research to estimate the number of SNPs required to map traits in different populations. Papers presented on pages 324 and 320 shed light on this issue.

Patricia Taillon-Miller *et al.*[3] typed 39 SNPs in the Xq25–q28 region of the X chromosome in three samples: 100 Finns, 150 Sardinians, and a mixed sample of 92 individuals comprised primarily of North Americans of European descent and French

people. In each sample, they identified two extended regions of strong disequilibrium (over 1 Mb at Xq25 and 340 kb at Xq28), bracketed by regions of weak disequilibrium. Even within the regions of strong disequilibrium, they observed substantial variability in disequilibrium as a function of physical distance.

Iain Eaves *et al.*[4] typed 21 STRs across 6.5 cM of chromosome 18 on samples of 800 chromosomes from Sardinia, Finland, the United Kingdom and the United States. The relationship between degree of disequilibrium and genetic distance was similar in the four samples, and in each sample there was substantial variability about the mean. These observations held for the actual STR markers and for pseudo-biallelic markers that resulted from dichotomizing the STR allele size distributions. They concluded that genetic isolates are not significantly more valuable than mixed populations for disequilibrium mapping of common variants underlying common diseases.

In general, the studies by Eaves *et al.* and Taillon-Miller *et al.* tell a consistent story. Both find substantial variability in disequilibrium levels beyond that which can be explained by differences in distances or allele frequencies, consistent with findings for smaller regions around individual genes[5,6]. Both studies also indicate that disequilibrium levels in the relatively isolated Finnish and Sardinian populations are not substantially greater than those in more mixed populations like those of the US and the UK.

Which populations?

Still, care needs to be taken not to overgeneralize these results. These studies are limited to a few European populations or populations of European descent, and to regions of only two chromosomes. Taillon-Miller *et al.* considered only SNPs with allele frequencies more than or equal to 0.2; such SNPs are likely to be relatively old, with disequilibrium patterns more likely to be similar across populations than random SNPs. Further, they considered the X chromosome, for which results might well be different than for the autosomes. Eaves *et al.* do not consider SNPs at all, and the relevance of findings from STRs to patterns of disequilibrium for SNPs is unclear. At a technical level, some of the observed variability in disequilibrium as a function of distance must be due to the uncertainty of estimates of physical and genetic distance.

Isolated populations established by a limited number of founders have proven extremely useful for mapping genes for rare monogenic disorders, and many assumed that the same advantages would hold for common diseases. The papers in this issue suggest that the value of isolates for common diseases may have been overrated, consistent with a study[7] based on simulation by Leonid Kruglyak. Still, one should beware of going too far in the opposite direction. Linkage disequilibrium levels, although similar across populations in the two studies, seem modestly stronger in the Finns and Sardinians than in the samples from populations of perceived heterogeneity. Eaves *et al.* found that, for distances of less than 1 cM, disequilibrium estimates (d^2) for dichotomized STR markers were 25–50% larger in the Finnish than in the North American sample. This suggests a sample 20–33% smaller would be sufficient to detect disequilibrium, or would grant greater tolerance for a wider spacing of SNPs. Similarly, Taillon-Miller *et al.* identified 43 pairs of markers with significant disequilibrium in the mixed sample ($P<0.005$), and 59 significant pairs in the Finns.

These differences are not overwhelming, but if they are generally true, should not be ignored. More data will help obtain a more accurate gauge of linkage disequilibria. Even if patterns are similar across populations, isolates may still provide the advantages of less variable environment, ease of study and, in the case of Finland, a population strongly supportive of biomedical research. Smaller isolates or sub-isolates also may prove helpful.

How dense a map?

The critical issue is map density. Whereas the Human Genome Project initially envisaged 1 SNP for every 30 kb, Kruglyak[7] suggested that useful levels of disequilibrium were unlikely to extend beyond 3 kb in mixed populations (requiring a SNP every 6 kb), and that most isolated populations would be similar. The former prediction dictates 100,000 SNPs, and the latter, 500,000. Neither of the current papers is ideally suited to test these predictions, as one describes results for STRs, and the other specifically over-samples for markers in regions of apparent disequilibrium. Still, both provide some basis for optimism. Taillon-Miller *et al.* detect significant disequilibrium between SNPs over large regions of the X chromosome, and Eaves *et al.* observe useful levels of disequilibrium in about half of pseudo-biallelic marker pairs for distances up to 0.1–0.2 cM. Whereas they are by no means definitive, these data indicate that Kruglyak's predictions may be too pessimistic and that the initial estimate of 100,000 well spaced, informative SNPs could be reasonable. This discrepancy in findings may be due to the particular chromosomes and populations considered, to differences in mutation patterns for STRs and SNPs, or to the simplifying assumptions of Kruglyak's analysis, including no population bottlenecks and no selection. Either number will require identification of a much larger pool of SNPs, as spacing will not be optimal and not all SNPs will be informative in all populations. Other studies demonstrate the possibility of the near absence of disequilibrium over distances much shorter than 3 kb (ref. 5).

Even if disequilibrium extends across distances of 30 to 50 kb, obtaining a very dense panel of SNPs will be of great value. Finding evidence of association between a SNP and a phenotype of interest will immediately lead to a desire to test other SNPs in the region. And so the efforts of the SNP Consortium[8] to develop a catalogue of a million or more SNPs are well justified, regardless of disequilibrium levels such as those observed in the present studies.

There is clear need for a whole-genome linkage disequilibrium map[9], and these papers provide a small initial instalment. At least in its early stages, this map should be undertaken for several populations, to assess whether the observation of a common pattern across populations is confirmed for relatively common variants, the sort of allele frequencies at which differences between populations start to become significant, and to continue to assess the number of SNP markers that will be required and how that number may vary between populations. Ideally this should be done with an eye to assessing not just disequilibrium between SNPs, but also, SNP-phenotype association; after all, identifying such associations is a primary inspiration for this undertaking. □

1. Risch, N. & Merikangas, K. *Science* **273**, 1516–1517 (1997).
2. Collins, F.S. *et al. Science* **282**, 682–689 (1998).
3. Taillon-Miller, P. *et al. Nature Genet.* **24**, 324–328 (2000).
4. Eaves, I.A. *et al. Nature Genet.* **24**, 320–323 (2000).
5. Nickerson, D.A. *et al. Nature Genet.* **19**, 233–240 (1998).
6. Martin, E.R. *et al. Genomics* **63**, 7–12 (2000).
7. Kruglyak, L. *Nature Genet.* **22**, 139–144 (1999).
8. http://snp.cshl.org/
9. Kruglyak, L. *Proc. Natl Acad. Sci. USA* **96**, 1170–1172 (1999).