*Proc. Natl. Acad. Sci. USA*
Vol. 87, pp. 701–704, January 1990
Medical Sciences

# Localizing multiple X chromosome-linked retinitis pigmentosa loci using multilocus homogeneity tests

**(genetic linkage/human gene mapping)**

JURG OTT[a,b], SHOMI BHATTACHARYA[c], J. D. CHEN[d], MICHAEL J. DENTON[d], JENNY DONALD[d], CHRISTOPHER DUBAY[e], G. JANE FARRAR[f], GERALD A. FISHMAN[g], DANIEL FREY[h], ANDREAS GAL[i], PETE HUMPHRIES[f], B. JAY[j], MARCELLE JAY[j], MICHAEL LITT[e], MARCO MÄCHLER[h], MARIA MUSARELLA[k], MEINHARD NEUGEBAUER[l], ROBERT L. NUSSBAUM[m], JOSEPH D. TERWILLIGER[a], RICHARD G. WELEBER[e], BRUNHILDE WIRTH[i], FULTON WONG[g], RONALD G. WORTON[k], AND ALAN F. WRIGHT[n]

[a]Columbia University, Department of Genetics and Development, and [b]New York State Psychiatric Institute, New York, NY 10032; [c]University of Newcastle upon Tyne, Department of Human Genetics, Newcastle, NE2 4AA United Kingdom; [d]Prince of Wales Hospital, Randwick, NSW 2031, Australia; [e]Oregon Health Sciences University, Department of Biochemistry, Portland, OR 97201; [f]Trinity College, Department of Genetics, Dublin 2, Ireland; [g]University of Illinois, Department of Ophthalmology, Chicago, IL 60612; [h]Institute of Medical Genetics, Zurich, Switzerland; [i]Institute of Human Genetics, Bonn, Federal Republic of Germany; [j]University of London, Department of Ophthalmology, London EC1V 2PD, United Kingdom; [k]Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada; [l]Institute of Medical Statistics, Bonn, Federal Republic of Germany; [m]Howard Hughes Medical Institute, University of Pennsylvania, Department of Human Genetics, Philadelphia, PA 19104; and [n]Medical Research Council Human Genetics Unit, Edinburgh EH4 2XU, United Kingdom

*Communicated by Richard L. Sidman, November 6, 1989 (received for review June 2, 1989)*

**ABSTRACT** **Multilocus linkage analysis of 62 family pedigrees with X chromosome-linked retinitis pigmentosa (XLRP) was undertaken to determine the presence of possible multiple disease loci and to reliably estimate their map location. Multilocus homogeneity tests furnished convincing evidence for the presence of two XLRP loci, the likelihood ratio being 6.4 × $10^9$:1 in favor of two versus a single XLRP locus and gave accurate estimates for their map location. In 60–75% of the families, location of an XLRP gene was estimated at 1 centimorgan distal to OTC, and in 25–40% of the families, an XLRP locus was located halfway between *DXS14* (p58-1) and *DXZ1* (Xcen), with an estimated recombination fraction of 25% between the two XLRP loci. There is also good evidence for a third XLRP locus, midway between *DXS28* (C7) and *DXS164* (pERT87), supported by a likelihood ratio of 293:1 for three versus two XLRP loci.**

Retinitis pigmentosa (RP) is a group of hereditary progressive disorders of the retina characterized initially by night blindness, often within the first two decades of life, reduction of peripheral or side vision, eventual decrease in central vision to variable degrees, in many cases leading to total blindness due to degeneration of the retina (1). There are several subtypes of RP, including autosomal recessive, autosomal dominant, and X chromosome-linked forms (XLRP).

Linkage analyses in families with XLRP have shown conflicting results regarding the location of the disease locus. Consequently, the presence of two XLRP loci was hypothesized (2–5). One subtype of XLRP, referred to as RP2, was linked to locus *DXS7* (L1.28) (6, 7); another subtype, RP3, was linked to locus *OTC* (ornithine carbamoyltransferase) (8–10). Further evidence for location of an XLRP locus distal to *OTC* came from a patient with a deletion starting between *OTC* and *DXS84* (754) and extending toward the telomere, in whom several X-chromosome linked diseases occurred including RP (11). Friedrich *et al.* (12) found another locus responsible for XLRP to be closely linked to *DXS7*, between Xcen and *DXS7* (L1.28).

The current study was initiated as a collaborative effort (*i*) to obtain evidence for XLRP heterogeneity if at all present, and, if so, (*ii*) to localize the disease loci in a comprehensive linkage and heterogeneity analysis.

**Family Data.** A total of 62 families were available for this analysis, most but not all of which have been published previously. For calculation efficiency, some pedigrees had earlier been broken down into smaller families and analyzed separately; here, they were analyzed undivided. Disease status for both affected males and carrier females was determined by the investigators and was incorporated unaltered in this analysis.

In many cases, heterozygous women also have symptoms that are, however, generally much milder than in men. Where detectable, such symptoms have been used for carrier status determination.

**Linkage Analysis.** The linkage analysis was carried out with the LINKAGE programs version 4.7 (13). Map distances between markers whose physical order is largely known were predominantly determined from the marker data in the present 62 families, except that the distance between *DXS85* (782) and *DXS41* (99–6) was obtained from H. Willard (personal communication).

For given marker distances, each interval between adjacent markers was divided in 5 subintervals whose end points represented possible locations of the disease locus. This way, $13 \times 5 + 1 = 66$ possible disease locations, $x_i$, were defined. For each family, at each possible disease location, the multipoint lod score (logarithm of odds), $\log_{10}[L(x)/L(x_\infty)]$, was calculated, where $L(x)$ is the likelihood with respect to position $x$ of the disease locus, and $x_\infty$ denotes an infinite map distance (disease locus not on map). In several families, not all markers present could be accommodated in the Mlink program, in which case the largest number possible, 4 or 5, around each possible disease location was used.

All multipoint likelihood calculations assumed absence of interference, and the resulting recombination fraction estimates were converted to map distances by the Kosambi mapping function. The values, $x$, designate map distances proximal from the *XG* blood group locus in the region pter–p22.3 on the short arm of the X chromosome where *XG* is taken to be at map location $x = 0$.

**Tests for Heterogeneity.** Heterogeneity was allowed for under a model of a mixture of two family types, type 1 with disease location at $x_1$, and type 2 with disease location at $x_2$,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RP, retinitis pigmentosa; XLRP, X chromosome-linked RP; LR, likelihood ratio; cM, centimorgan(s).

Table 1. Marker loci used

| Gene symbol | Probe name | Θ | Z | x |
|---|---|---|---|---|
| *XG* | — | — | — | 0 |
| *DXS85* | 782 | 0.196 | 1.24 | 21 |
| *DXS41* | 99-6 | 0.300 | * | 55 |
| *DXS28* | C7 | 0.134 | 2.10 | 69 |
| *DXS164* | pERT87 | 0.120 | 4.76 | 81 |
| *DXS206* | XJ1.1/XJ5.1 | 0.000 | 14.07 | 81 |
| *DXS84* | 754 | 0.032 | 22.35 | 85 |
| *OTC* | — | 0.068 | 27.22 | 91 |
| *DXS7* | L1.28 | 0.084 | 30.52 | 100 |
| *OATL1* | — | 0.074 | 18.17 | 107 |
| *DXS14* | p58-1 | 0.078 | 16.04 | 115 |
| *DXZ1*/Xcen | pBamX7 | 0.050 | * | 120 |
| *DXS1* | p8 | 0.100 | 3.30 | 130 |
| *DXYS1* | pDP34 | 0.100 | * | 140 |

Θ = estimate of recombination fraction with locus listed on line above, $x$ = Kosambi map location in cM, $Z$ = maximum lod score obtained in present data.
*Θ is based on published values, and the map distance estimate between *DXS14* (p58-1) and Xcen is from H. Willard (personal communication).

the proportion of families of type 1 being denoted by $\alpha_1$. Based on the multipoint lod scores for each family at each point, the parameters $\alpha_1$ ($\alpha_2 = 1 - \alpha_1$), $x_1$, and $x_2$ were estimated by evaluating the likelihood at points of $\alpha_1$ from 0 through 1 in steps of 0.05, and at all possible points of $x_1$ and $x_2$, $x_1 < x_2$. This approach represents a previously proposed multilocus extension (14) of Smith's homogeneity test. The HOMOG2 computer program, an adaptation of the HOMOG program (15), was used for these calculations.

The (one-sided) test of homogeneity ($H_0$, $\alpha_1 = 1$) against the hypothesis of heterogeneity ($H_1$, $\alpha_1 < 1$) was carried out by calculating the maximum likelihood under each of the two hypotheses and reporting the likelihood ratio, LR, in favor of heterogeneity over homogeneity. As the distributional properties of the LR in this situation are unknown, one of the problems being that multipoint likelihoods are multimodal, no error probabilities are attached to observed likelihood ratios. Conventionally, likelihood ratios exceeding 50–100 are considered meaningful.

Support regions for parameters were calculated under the assumption that a set of parameter values ($\alpha_1, x_1, x_2$) belongs to the support region when its $\log_{10}$ likelihood is within 1.7 units of the maximum $\log_{10}$ likelihood. Parameter values within such a support region have an associated likelihood ratio of at most 50:1 against the maximum likelihood estimates. For each parameter, its smallest and largest value in the support region are taken to constitute the end points of its support interval. Note that this conservative definition of support intervals will yield relatively wide error margins, as compared to conventional support regions based on a difference of 1 unit of $\log_{10}$ likelihood, which is equivalent to a likelihood ratio of 10:1.

Particularly in small families, the possibility of an autosomal mode of inheritance cannot usually be ruled out with certainty. Therefore, the heterogeneity analysis was carried out with different selections of families, the selection criteria being the maximum multipoint lod score, $Z$, achieved in a family. To minimize the chance of admixture with autosomal forms of RP, most analyses were carried out using only those 40 families with at least a multipoint lod score of 0.5, which eliminates from the analysis many small and marginally informative families. On the other hand, such a selection reduces the chance of finding XLRP loci outside of the map spanned by the markers if such a locus exists in the families investigated.

Calculations under a model allowing for three or four disease loci were carried out with the HOMOG3 and HOMOG4 programs, which are extensions of the HOMOG2 program. HOMOG4 evaluated the proportions, $\alpha_i$, in steps of 0.20.

**Results Assuming Two XLRP Loci.** Table 1 shows the marker loci used and the map locations estimated or assumed for them. The maximum lod scores, $Z$, shown were obtained by analyzing the 62 XLRP families. Table 2 presents the results of the analysis for heterogeneity. The evidence for two XLRP loci is convincing, even when only those families with a maximum lod score of at least ½ were analyzed. While the map positions of the two XLRP loci were estimated quite accurately, the proportion of type 1 families has a rather wide support interval given in parentheses below. Based on the 40 families with $Z > ½$, 75% (45–90%) of them had the *RP* locus at map position 90 (87–90) centimorgans (cM), 1 cM distal to *OTC*, and in 25% (10–55%) of the families the *RP* locus was estimated at map location 117 (104–119) cM, halfway between *DXS14* (p58-1) and Xcen. When based on the 56 families with positive lod scores, the latter position is obtained at 118 cM with a much shorter support interval (116–119 cM). The estimated distance between the two XLRP loci is, thus, 27–28 cM, corresponding to a recombination fraction Θ = 0.25. For all analyses summarized below, only those families with a multipoint lod score, $Z > ½$, were used.

The analysis of heterogeneity presented in Table 2 assumes the same proportions of family types and locus positions in all sources of families (homogeneity among investigators). Variation of these parameters between investigators may be allowed for in an extended heterogeneity analysis in the following way. Consider the three hypotheses: Homogeneity within and between investigators ($H_1$, $\alpha_1 = 1$; one map location estimated), heterogeneity within but homogeneity between investigators ($H_2$, $\alpha_1 < 1$, the same for each investigator; one proportion and two map locations estimated), and heterogeneity both within and between investigators ($H_3$, $\alpha_1 < 1$, but allowing for $\alpha_1$ varying among investigators; three parameters estimated per investigator with at least three families, one parameter estimated per investigator with less than three families). Testing $H_2$ versus $H_1$ corresponds to the test for heterogeneity already carried out (Table 2); the maximum $\log_{10}$ likelihoods obtained were 70.63 under $H_1$ and 80.44 under $H_2$. An analysis under $H_3$ leads to a maximum $\log_{10}$ likelihood of 88.31 shown in Table 3, which presents good evidence for heterogeneity from three investigators and little or no heterogeneity from the other investigators, suggesting that the proportion of families with one or the other XLRP locus may vary in different research centers. These

Table 2. Results of test for heterogeneity (two XLRP loci versus one XLRP locus)

| | | | Estimates (support intervals) | | |
|---|---|---|---|---|---|
| Z | n | LR | $\alpha_1$ | $x_1$ | $x_2$ |
| 0 | 62 | $2.2 \times 10^{16}$ | 0.60 (0.35–0.80) | 90 (87–90) | 118 (116–119) |
| >0 | 56 | $1.5 \times 10^{15}$ | 0.60 (0.40–0.80) | 90 (87–90) | 118 (116–119) |
| >½ | 40 | $6.3 \times 10^{9}$ | 0.75 (0.45–0.90) | 90 (87–90) | 117 (103–119) |

$Z$ = maximum lod score required for inclusion of family, $n$ = number of families satisfying inclusion criterion, LR = likelihood ratio in favor of two versus one XLRP loci.

Table 3. Heterogeneity analysis for each investigator (families with $Z > ½$)

| Investigator(s) | $n$ | $\log_{10}L$ | LR | Estimates $\alpha_1$ | $x_1$ | $x_2$ |
|---|---|---|---|---|---|---|
| S.B. and A.F.W. | 16 | 16.18 | 75,849 | 0.30 | 74 | 107 |
| M.J.D. | 3 | 22.03 | 1 | 1.00 | 91 | — |
| D.F. | 2 | 1.72 | — | 1.00 | 85 | — |
| P.H. | 3 | 3.82 | 1.73 | 0.65 | 85 | 128 |
| M. Musarella | 5 | 14.47 | 345 | 0.50 | 76 | 91 |
| F.W. | 5 | 6.37 | 1 | 1.00 | 100 | — |
| M.L. | 1 | 3.56 | — | 1.00 | 120 | — |
| A.G. | 4 | 15.73 | 364,762 | 0.75 | 89 | 117 |
| R.L.N. | 1 | 3.80 | — | 1.00 | 91 | — |
| | Total | 88.31 | | | | |

$n$ = number of families, $\log_{10}L$ = maximum log likelihood under heterogeneity, LR = likelihood ratio in favor of two versus one XLRP loci.

differences between investigators ($H_3$ versus $H_2$) are associated with a LR of antilog (88.31 − 80.44) = $7.4 \times 10^7$, which appears large. However, under $H_3$, 18 additional parameters are estimated over those under $H_2$ so that the increase in LR amounts to a (geometric) mean of LR = 2.7 per additional parameter. Therefore, significant variation between investigators cannot clearly be established.

For most investigators, the estimated location of the two XLRP loci agrees fairly well with the overall estimates of 90 and 118 cM (Table 2). For at least two investigators, however, a third location appears at a map location well below 90 cM, which points to the possibility of a third XLRP locus. This question is pursued further in the next section.

**More Than Two XLRP Loci.** Analyzing all 40 families with $Z > ½$ for the presence of up to four XLRP loci showed the results given in Table 4. Extending the two-locus to a three-locus mixture of families by estimating two additional parameters (one proportion and one map location) increased the maximum $\log_{10}$ likelihood by 2.47, which corresponds to a likelihood ratio of 293. There is, therefore, good evidence for a third XLRP locus at map position 76 cM, midway between *DXS28* (C7) and *DXS164* (pERT87), although the presence of this XLRP locus is by far not as well supported by the data as the other two XLRP loci. In an analysis of the 56 families with positive lod scores, evidence for the third XLRP locus is even stronger (results not shown).

As Table 4 also shows, allowing for a fourth XLRP locus results only in a small increase in the log likelihood. There is, thus, no evidence furnished by these families for more than three XLRP loci in the portion of the X chromosome investigated here.

Table 5 presents an analysis of a mixture of three XLRP loci individually for those investigators with three or more families with $Z > ½$ that had shown at least a hint for the presence of a second locus. There is, however, not much evidence for a third XLRP locus for any single investigator; in fact, only the families investigated by S.B. and A.F.W. show a slightly increased likelihood for three versus two XLRP loci. The cautious interpretation of these findings is

Table 4. Evidence for a varying number, $n$, of XLRP loci

| $n$ | LR | Parameter estimates $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|---|---|---|---|
| 1 | — | 1 | — | — | 95 | — | — | — |
| 2 | $6.4 \times 10^9$ | 0.75 | 0.25 | — | 90 | 117 | — | — |
| 3 | 293 | 0.20 | 0.60 | 0.20 | 76 | 91 | 117 | — |
| 4 | 1.6 | 0.20 | 0.40 | 0.20 | 76 | 91 | 98 | 117 |

LR = likelihood ratio for $n$ versus $n - 1$ XLRP loci (families with $Z > ½$).

Table 5. Heterogeneity analysis for each investigator, allowing for three XLRP loci (families with $Z > ½$)

| Investigator(s) | $n$ | LR | Estimates $\alpha_1$ | $\alpha_2$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|---|---|
| S.B. and A.F.W. | 16 | 3.41 | 0.2 | 0.3 | 82 | 98 | 107 |
| P.H. | 3 | 1 | 0.0 | 0.7 | — | 85 | 128 |
| M. Musarella | 5 | 1 | 0.5 | 0.0 | 76 | 91 | — |
| A.G. | 4 | 1.19 | 0.3 | 0.4 | 85 | 91 | 117 |

$n$ = number of families, LR = likelihood ratio in favor of three over two XLRP loci.

that a third XLRP locus exists but that only up to two XLRP loci are present in families studied by any one investigator, which explains the variation found above between investigators in the analysis of two XLRP loci.

The statistical evidence for a third XLRP locus appears to be contradicted by the existence of males carrying deletions of part or all of the region between the *DXS28* (C7) and *DXS164* (pERT87) loci, none of whom have been reported to have RP. For example, the two males, JH (16) and JO (17), both have deletions including the region between *DXS28* and *DXS164* but do not have RP. These observations render the existence of an XLRP locus between *DXS28* and *DXS164* implausible. An explanation for this contradiction might be that the estimate of the map position of the third XLRP locus is not very good. In fact, based on the families with $Z > ½$, its approximate support interval extends from *DXS41* (99–6) at 55 cM to 90 cM so that perhaps the third XLRP locus is distal to *DXS28* (some of the deletions without associated RP extended proximal to *DXS164* but none involves large segments distal to *DXS28*).

**Discussion.** To gauge the possible effect of mutation at the XLRP loci, various two-point linkage analyses were carried out with mutation either present ($\mu = 10^{-6}$) or absent ($\mu = 0$). The resulting lod scores differed at most in the sixth decimal place so that it appeared safe to assume absence of mutation in the linkage analysis. In addition, owing to the low mutation rate at XLRP loci, germinal mosaicism (18) was assumed absent.

As mentioned above, the two XLRP loci postulated in the literature are *RP2* linked to *DXS7* (L1.28) and *RP3* linked to *OTC*. In this analysis, the XLRP locus estimated at 90 cM, 1 cM distal to *OTC*, must correspond to the *RP3* locus. For the other XLRP locus (at map location 117 cM), the distal end of the support interval extends to the map location 104 cM, which is halfway between *DXS7* and *OATL1*. Therefore, this latter XLRP locus may be taken to correspond to the *RP2* locus.

The presence of heterogeneity has implications for the calculation of genetic risks. If it is known which of the XLRP loci is segregating in a particular family, genetic risks may be calculated in the usual manner—for example, by using one of the standard options in a linkage program. Such calculations will also be appropriate when the risk is more or less the same no matter which XLRP locus is assumed to segregate. Under heterogeneity, the genetic risk for a particular individual in a family may be represented as a weighted average over the conditional risks given each of the XLRP loci in turn. Detailed formulations of such calculations are proposed elsewhere (19).

Generous support by the National Retinitis Pigmentosa Foundation of America and the W. M. Keck Foundation is gratefully acknowledged.

1. Marmor, M. F., Aguirre, G., Arden, G., Berson, E., Birch, D. G., Boughman, J. A., Carr, R., Chatrian, G. E., Del Monte, M., Dowling, J., Enoch, J., Fishman, G. A., Fulton, A. B., Garcia, C. A., Gouras, P., Heckenlively, J., Hu, D.-N., Lewis,

R. A., Niemeyer, G., Parker, J. A., Perlman, I., Ripps, H., Sandberg, M. A., Siegel, I., Weleber, R. G., Wolf, M. L., Wu, L. & Young, R. S. L. (1983) *Ophthalmology* **90,** 126–131.
2. Nussbaum, R. L., Lewis, R. A., Lebo, R. G. & Ferrell, R. (1985) *Hum. Genet.* **70,** 45–50.
3. Buetow, K. H., Chakravarti, A., Nussbaum, R. L. & Ferrell, R. E. (1985) *Cytogenet. Cell Genet.* **40,** 595 (abstr.).
4. McKusick, V. A. (1985) *Mendelian Inheritance of Man* (Johns Hopkins Press, Baltimore).
5. Farrar, G. J., Geraghty, M. T., Moloney, J. M. B., McConnell, D. J. & Humphries, P. (1988) *J. Med. Genet.* **25,** 222–226.
6. Bhattacharya, S. S., Wright, A. F., Clayton, J. F., Price, W. H., Phillips, C. F., McKeown, C. M. E., Jay, M., Bird, A. C., Pearson, P. L., Southern, E. M. & Evans, H. J. (1984) *Nature (London)* **309,** 253–255.
7. Clayton, J. F., Wright, A. F., Jay, M., McKeown, C. M. E., Dempster, M., Jay, B. S., Bird, A. C. & Bhattacharya, S. S. (1986) *Hum. Genet.* **74,** 168–171.
8. Musarella, M. A., Burghes, A., Anson-Cartwright, L., Mahtai, M. M., Argonza, R., Tsui, L. C. & Worton, R. (1988) *Am. J. Hum. Genet.* **43,** 484–494.
9. Denton, M. J., Chen, J. D., Serraville, S., Colley, P., Halliday, F. B. & Donald, J. (1988) *Hum. Genet.* **78,** 60–64.
10. Wirth, B., Denton, M. J., Chen, J. D., Neugebauer, M., Halliday, F. B., van Schooneveld, M., Donald, J., Bleeker-Wagemakers, E. M., Pearson, P. L. & Gal, A. (1988) *Genomics* **2,** 263–266.
11. Francke, U., Ochs, H. P., de Martinville, B., Giacalone, J., Lindgren, V., Disteche, C., Pagon, R. A., Hofker, M. H., van Omenn, G. J. B., Pearson, P. L. & Wedgwood, R. J. (1985) *Am. J. Hum. Genet.* **37,** 250–267.
12. Friedrich, U., Warburg, M., Wieacker, P., Wienker, T. F., Gal, A. & Ropers, H. H. (1985) *Hum. Genet.* **71,** 93–99.
13. Lathrop, G. M. & Lalouel, J.-M. (1988) *Am. J. Hum. Genet.* **42,** 498–505.
14. Ott, J. (1986) in *A Short Guide to Linkage Analysis*, ed. Davies, K. E. (IRL, Arlington, VA), pp. 19–32.
15. Ott, J. (1985) *Analysis of Human Genetic Linkage* (Johns Hopkins Press, Baltimore).
16. Clarke, A., Roberts, S. H., Thomas, N. S. T., Whitfield, A., Williams, J. & Harper, P. S. (1986) *J. Med. Genet.* **23,** 501–508.
17. Van Ommen, G. J. B., Verkerk, J. M. H., Hofker, M. H., Monaco, A. P., Kunkel, L. M., Ray, P., Worton, R., Wieringa, B., Bakker, E. & Pearson, P. L. (1986) *Cell* **47,** 499–504.
18. Edwards, J. H. (1986) *J. Med. Genet.* **23,** 521–530.
19. Weeks, D. E. & Ott, J. (1989) *Am. J. Hum. Genet.* **45,** 819–821.