Everything you always wanted to know about genetic maps, markers, linkage analysis, Lod scores, and lots of other important stuff!

Elaine Ostrander, Ph.D.
Cancer Genetics Branch, NHGRI
March 2005

Types of Maps

- **Meiotic Linkage Map**
  Distance between markers/genes measured as a function of recombination

- **Radiation Hybrid Map**
  Distance is measured as a function of chromosome breakage

- **Cytogenetic Maps**
  Distance is measured by assignment of large clones or chromosomal paints to ordered cytogenetic bands

- **Comparative Maps**
  Chromosomal segments compared between species and segments containing genes of conserved order identified
Example: RH, Meiotic Linkage and Cytogenetic Maps for Canine Chromosome 5

Meiotic Linkage Maps

- Basic tool of genetics
- Sets of polymorphic markers that are used to follow inheritance of segments of a chromosome through generations within families
- Ideally, markers are evenly distributed and highly polymorphic in population under study
- Distance is measured indirectly as a function of recombination
- 1cM is equivalent to about 1 million bp in human genome
Polymorphism

- Positions in the genome where there is variation in DNA sequence
- Stable for tracking Mendelian inheritance
- Variants are termed alleles
- May be in either coding or non-coding regions
- Insertions, deletions, rearrangements, variable length repeats

Kinds Of Genetic Markers

- RFLP—Restriction Fragment Length Polymorphism (RFLP)
  - Addition, deletion, or change of base pair results in gain or loss of restriction enzyme site
    - Two allele systems
    - Detection by Southern blot analysis
- Microsatellites—Simple sequence repeats—small di, tri, or tetra nucleotide repeats that are reiterated in tandem—(CA)n, (GAAA)n, (CAG)n, etc.
  - Highly polymorphic—large numbers of alleles in population, n = 2-40
  - Stable within pedigrees (0.0004 mutations/gamete)
  - Common—50,000/genome
- Single Nucleotide Polymorphisms (SNPs)
  - Very frequent (1/kb)
  - Assays through sequencing or SNP chips
Genetic Mapping With Microsatellite Markers

PCR is used to follow inheritance of length alleles and adjacent portions of chromosome

---

Two Common Measures of Polymorphism

- Heterozygosity
  - For a locus with n alleles, $H = 1 - 1/n$
  - So a locus with three alleles has a maximum heterozygosity of $H = 1 - 1/3 = 2/3 = 0.67$

- Polymorphic Information Content (PIC)
  - $PIC = \frac{(n-1)^2}{N(N+1)} N^3$
  - $PIC > 0 < 1$
  - Most “good” markers have $PIC > 0.7$

- “Polymorphic locus” is defined as one with Het or PIC >10%
Use Linkage Map to Localize Trait-Associated Locus

- Take advantage of principles of linkage analysis:
  - Meiotic linkage map
  - Families with disease in question clearly phenotyped
  - Ability to accurately screen (genotype) families with markers placed at least every 5-10 cM
  - Analysis--error checking and linkage
- Analysis of 250 families with about 7 sampled individuals per family will require close to 1 million genotypes

Genetic Linkage

- Linkage
  - Two loci (e.g. marker and a putative disease gene) are physically close on a chromosome
  - Will be transmitted together from parent to child more often then expected by chance
- No Linkage
  - Two loci have a 50% chance of being co-inherited
For a given area of the genome, the recombination fraction, $\theta$, is proportional to the distance between the loci. Genetic distance is measured in Morgans (M) or cM.

If, for example, the recombination fraction between two loci is $\theta = 0.06$ the genetic distance between them is simply 0.06 Morgan (6cM).
Lots of Things Affect Recombination

Unit of map distance is a Morgan

- A Morgan is the length of chromosomal segment which, on average, experiences one exchange per strand
- For short intervals the frequency of recombination will be directly proportional to the map interval since the number of double crossovers will be negligible
- For longer intervals relationship is more complex
  - Occurrence of multiple crossover is no longer negligible
  - Phenomena of “interference”

A “Linked” Marker

\[ \begin{array}{c}
\text{D = Disease} \\
\text{N = Normal} \\
\end{array} \]
**Terminology**

- Phenocopy--Individual with the same disease or trait, but it is associated with a different cause
- Penetrance--How likely one is to get a disease (trait) if they carry a genetic variant
- Sporadic--Disease (trait) due to factors other than those under consideration
  - environmental factors
  - weakly penetrant alleles
  - stochastic events

**Hypothetical Marker**

Is there evidence for linkage?
Three Possibilities (ok, Four) to Explain Data

- Explain inconsistencies by:
  - Age dependent penetrance
  - Presence of sporadics and phenocopies
  - Recombination
  - Simply wrong!

Estimates of Linkage

- Genome-wide scan
  - Testing for linkage between markers and disease state
- LOD score - Log of Odds
  - Do number of recombinants between marker and putative disease locus differ significantly over chance?
  - Underlying model of inheritance
  - LOD score $\geq 3.3$ significant
  - Indicate greater then 1000:1 odds in favor of linkage
- NPL - Nonparametric Linkage Analysis
  - Significant allele sharing among affected individuals?
  - No model of inheritance
  - Assessed as $P$ value
Lod scores - Calculation to help assess if the association you observe differs significantly over what is expected by chance.

- While lod scores > 3.3 are generally accepted as evidence of linkage, Lod scores < -2.0 are accepted as evidence against linkage.
- Because they are calculated as Log of the Odds, lod scores may be summed across families.
- Adding families allows us to increase power.
- Lod scores are calculated for multiple values of $\theta$.
- The value of $\theta$ at which the peak lod is obtained is taken as the maximum likelihood value of $\theta$.

\[ n = \text{total number of loci} \]
\[ k = \text{number of recombinant alleles} \]

Recombination fraction = $\theta = k/n$

For linked loci $\theta < 0.5$
For unlinked loci $\theta = 0.5$
Dreaded Linkage Summary

- Estimate of recombination frequency is only valid if number of offspring is sufficient to be certain the observed ratio of nonrecombinants/recombinants is statistically different from the 50:50 ratio expected by chance.

- To evaluate this--Calculate a series of likelihood ratios (relative odds) at various values of $\theta$ ranging from $\theta$ (no recombination) to $\theta = 0.5$ (random assortment).

- Thus, the likelihood at a given value of $\theta = \frac{\text{Likelihood of data if loci unlinked at } \theta}{\text{Likelihood of data if loci unlinked}}$

- The computed likelihood are usually expressed as the $\log_{10}$ of this ratio and called a “lod score” (Z) for “logarithm of the odds.”

- Value at which Z is greatest is accepted as the best estimate of the recombination fraction and called the maximum likelihood estimate.
**Linkage Application**

To calculate whether two loci are linked or unlinked calculate:

- The Likelihood ($L$) for linkage ($\theta < 0.5$)

- Versus $L$ (non-linkage) ($\theta > 0.5$)

The Lod score method (the likelihood odds ratio):

$$Z(\theta) = \log_{10} \frac{L(\theta)}{L(0.5)}$$

Example:

- $n = 10$  
- $k = 3$  
- $n-k = 7$  
- $\theta = 0.3$

$$Z(\theta) = \log_{10} \left( \frac{0.3^{3} \times 0.7^{7}}{0.5} \right)$$

$$Z \theta \approx n \log(2) + k \log(\theta) + (n-k) \log(1-\theta) \text{ if } \theta > 0$$

$$10 \log(2) + 3 \log(0.3) + 7 \log(0.7) = 3.01 + (-1.6) + (-1) = 0.357$$

**Do we have linkage?**

Generally, Lod scores $> 3.3$ are accepted as evidence for linkage.

---

**Types of Analyses**

- **Parametric**
  - Derive “transmission model” where assumptions are made regarding:
    - Age-dependent penetrance
    - Frequency of mutant allele in population
    - Mode of transmission

- **Non parametric**
  - “model free”
    - Association between disease state and chromosome sharing
    - Omit data from unaffected individuals
    - When is this a problem?
Kidney cancer maps to chromosome 5 with \( \text{Lod} = 16.7 \ (\theta = 0.016) \)

Jonnasdottir et al., (2003) PNAS 100: 5296-5301

Linkage Analysis Using Founder Effects: Two Strategies

- One large pedigree with disease derived from single founder
  - Leon et al., (1992) PNAS 89: 5281-5284
- Many smaller pedigrees with similar features that share common heritage
Linkage Disequilibrium

Linkage Disequilibrium Around Variant

- Ancestral Chromosome
- Present Day Chromosomes

**Isolated Population Model**
- Initial Population
- Bottleneck

**Founder Effect Model**
- Long Bottleneck
- Small number of founders
- Expansion
- 10-100 generations

**Founder Effect Model**
- Small number of founders
- Expansion
- 10-100 generations
Leon et al., 1992

- See Table 1—What is it telling you about?
  - Information content and polymorphism
- See Table 2. How do you explain the lod scores going up and down?
  - Hint: map order and/or marker informativeness
- See Figure 1
  - What markers is the gene between?
  - Right side of pedigree allows us to narrow region of linkage to between what markers?
  - How do you explains persons D and E?

![Table 1. Polymorphisms on chromosome 5q linked to deafness in the Monge Linded](image)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>p</th>
<th>Primer or probe/enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>FMR1</td>
<td>8</td>
<td>0.12</td>
<td>AAG GTG TTC TTT GTA TGC TCA GC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GTA ATG TGT TGT TAG TGC AAG GC</td>
</tr>
<tr>
<td>LCH (5E9)</td>
<td>6</td>
<td>0.07</td>
<td>AGG TCC AGG GTA GCT CAT GCT CAT CTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTA ATG CAG GGC TTT AGG GC</td>
</tr>
<tr>
<td>GRL</td>
<td>2</td>
<td>0.00</td>
<td>ORT/Ref 1</td>
</tr>
<tr>
<td>DES170</td>
<td>2</td>
<td>0.00</td>
<td>TPSE/Tag 1</td>
</tr>
<tr>
<td>DES575 (6/6132)</td>
<td>5</td>
<td>0.07</td>
<td>AGG CAG AGT CTA GAA GGA GC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTT TAA CAT GCT TTA AGA GC</td>
</tr>
<tr>
<td>DES190 (6/592)</td>
<td>3</td>
<td>0.00</td>
<td>TGG GAA GGG TTA GAA GAT GC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AGG AGT TCT AGT TGC GAT ACG G</td>
</tr>
<tr>
<td>DES115 (6/596)</td>
<td>4</td>
<td>0.14</td>
<td>TCC TAG CTT AAT TCC TCC CCT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GCA GCT TGT TCA ATT TCG GC</td>
</tr>
<tr>
<td>DES200 (6/6115)</td>
<td>8</td>
<td>0.00</td>
<td>CTG CAG TAG CAA CGG AGA AA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TGG AGG AGG AAA CAG UAA GC</td>
</tr>
<tr>
<td>DES522</td>
<td>3</td>
<td>0.11</td>
<td>JG90F1-01/C/Map 1</td>
</tr>
</tbody>
</table>

(8. Approximate distance between adjacent markers.)
Leon et al., 1992

Table 2. Maximum two-point lod scores ($Z$), maximum likelihood recombination fractions ($\theta$), and 95% confidence intervals for $\theta$ for linkage of deafness to chromosome 5 markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>$Z$</th>
<th>$\theta$</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIB5</td>
<td>6.63</td>
<td>0.06</td>
<td>(0.01, 0.12)</td>
</tr>
<tr>
<td>IL9</td>
<td>13.55</td>
<td>0.06</td>
<td>(0.01, 0.12)</td>
</tr>
<tr>
<td>GRL</td>
<td>2.65</td>
<td>0.10</td>
<td>(0.02, 0.30)</td>
</tr>
<tr>
<td>D5S210/D5S207</td>
<td>7.50</td>
<td>0.13</td>
<td>(0.05, 0.31)</td>
</tr>
</tbody>
</table>

Leon et al., 1992

[Diagram with genetic data]
What are problems associated with mapping deafness genes?
- Number of forms of disease
- Mating not random
- Association with other congenital problems
- Incomplete penetrance

What other genes would you imagine would be this hard to map?
- Cancer, epilepsy, behavioral (alcoholism), mental illness (schizophrenia)

Where would you find other families like this to map other diseases?
- Iceland, Finland, Bedouins, etc.
What specific advantages did this pedigree (Leon et al., 1992) offer?

- Autosomal dominant
- Highly penetrant
- Single founder (Felix Monge born in 1754)
- Low mobility of family members (Cartago, Costa Rica)
- Societal acceptance so families opt for multiple children
- Statistical power--150 people available-99 in this report

Ruiz-Perez et al., 2000

- What is the disease?
  - Recessive form of dwarfism found in Old Order Amish
- What features of Amish populations make them ideal for studies like this one?
  - Descended from small number of founders
  - Strict endogamy
  - Centrifugal gene flow
  - Genealogical records
  - Large families
- In this case, how far back could lineage be traced?
  - All 50 cases trace to single couple who immigrated in in 1744 to Lancaster county
  - Amish are distributed into 3 consanguineous groups (demes) that live in Pennsylvania, Ohio, Northern Indiana
What is the Primary Result?

Ellis-van Creveld syndrome is an autosomal recessive skeletal dysplasia characterized by short limbs, short ribs, postaxial polydactyly and dysplastic nails and teeth.

Mapped to chromosome 4p16 in nine Amish sub pedigrees and pedigrees from Mexico, Ecuador and Brazil.

Weyers acrodental dysostosis is an autosomal dominant disorder with a similar but milder phenotype, has been mapped in a single pedigree to an area including the EvC critical region.

Researchers identified a new gene (EVC), encoding a 992-amino-acid protein, mutated in EvC patients.

Observed a splice-donor change in an Amish pedigree and six truncating mutations and a single amino acid deletion in seven pedigrees. The heterozygous carriers of these mutations did not manifest features of EvC.
What is the problem?

However, they also found two heterozygous missense mutations associated with Weyers acrodental dysostosis in a single individual and a father/daughter pair.

How do they know the missense are not neutral polymorphisms?
- Not observed in 200 chromosomes
- In critical region of protein

How do you explain these results?

Summary

- There are several kinds of maps. Those composed of large numbers of well spaced polymorphic markers are used for genetic mapping studies.
- Pedigree analysis and accurate phenotypes are key for successful linkage studies.
- Multiple kinds of analyses can be done. In general, however, one is looking to find markers whose alleles segregate faithfully with phenotypes in families.
- The LOD method allows you to group data from similar families together, thus increasing power to accurately map loci genes or exclude regions of the genome.
- Making use of founder effects is a way to increase power and reduce locus heterogeneity problems in linkage mapping.
The gene for an inherited form of deafness maps to chromosome 5q31

PEDRO E. LEON*, HENRIETTE RAVENTOS*, ERIC LYNCH†, JAN MORROW†, and MARY-CLAIRE KING†

*Center for Research in Molecular and Cell Biology, University of Costa Rica, San Jose, Costa Rica; and †School of Public Health and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Communicated by Victor A. McKusick, February 13, 1992

ABSTRACT Primary—i.e., nonsyndromal—postlingual deafness is inherited as an autosomal dominant phenotype in a large kindred in Costa Rica. Genetically susceptible individuals begin to lose hearing at low frequencies at about age 10 years, after language and speaking are learned. Deafness inevitably progresses by age 30 years to bilateral hearing loss of all frequencies. Intelligence, fertility, and life expectancy are normal. The family traces its ancestry to an affected founder born in Costa Rica in 1754. We have mapped the gene for deafness in this kindred to chromosome 5q31, between the markers IL2 and GRL, by linkage analysis involving 99 informative relatives.

Human deafness is a major medical and public health concern. A generation ago, most congenital deafness was attributable to viral diseases during pregnancy. Now that maternal viral diseases are far less common, perhaps 50% of deafness in newborns has a genetic basis (1, 2). Probably an even higher proportion of deafness among older children and young adults is genetically influenced. However, the genetics of human deafness is complex and heterogeneous. There are at least 32 genetic forms of primary human deafness (i.e., deafness not secondary to some other disease), some of which may themselves be heterogeneous (3). In addition to genetic heterogeneity across affected families, deafness among affected children in the same family may be due to different genes, because deaf persons often marry each other. Furthermore, some congenital forms of deafness are associated with severe abnormalities, sometimes leading to early death and often to limited family size, so for these conditions few families with more than one affected child are to be found. Finally, many genetic forms of deafness are variable in expression with incomplete penetrance. It is probably not surprising, therefore, that heretofore no genes have been mapped for primary human deafness.

Inherited low-frequency hearing loss (LHFL1; no. 124900 in ref. 3) is an autosomal dominant, fully penetrant, sensory-neural deafness originally described in an extended Costa Rican kindred (4). In this kindred, low-tone deafness begins at about age 10 and progresses by age 30 in males and females to profound, irreversible, bilateral deafness involving all frequencies. Impedance tests suggest normal stapedial reflexes and no mechanical damage (5). Tone decay and other audiometric studies indicate normal retrocochlear function. Speech development before onset of deafness is normal, as are intelligence, fertility, and life expectancy.

Deafness in the Costa Rican kindred has been traced back eight generations to Felix Monge, who was born in Costa Rica in 1754. Testaments by Felix Monge and two of his brothers indicate they were deaf but were born hearing. Felix Monge's sibship was the seventh generation of their family in Costa Rica. Their ancestors migrated to Costa Rica from Jerez de la Frontera, Spain, about 1600. Most of the descendants of Felix Monge still live near Cartago, Costa Rica. At least 150 living adults from the family are informative for linkage analysis. For this study, relatives were considered informative if they either were diagnosed as deaf or were older than age 25 years with no symptoms of hearing loss. Ninety-nine of these informative relatives are included in this report.

MATERIALS AND METHODS

Clinical evaluation of deafness was carried out by audiometric testing at the Centeno Guell Auditory Testing Facility in San Jose, Costa Rica, as described (4, 5). Audiologic testing was conducted in an Industrial Acoustics Corporation sound chamber (model 400-SER). Interior noise levels did not exceed 40 decibels, scale A. A clinical audiometer (Maico, model MA-22) and a portable audiometer (Maico, model MA-16) were used for pure tone and bone conduction testing at 250 Hz through 8000 Hz. An impedance bridge (Teledyne, model TA-3D) was employed for all tympanometry and acoustic reflex testing. All audiometric equipment was calibrated before each period of testing. Before audiometric evaluation, subjects were given otoscopic examinations to determine the condition of their ear canals and then given the opportunity to practice the required tasks. Pure tone and bone conduction thresholds were measured using the modified Hughson–Westlake technique (6). Frequencies tested by air conduction were 250, 500, 1000, 2000, 3000, 4000, 6000, and 8000 Hz. Frequencies tested by bone conduction were 250, 500, 1000, 2000, 3000, and 4000 Hz.

Affected relatives older than age 25 in the kindred had pure tone air conduction hearing threshold levels above 50 decibels for all frequencies. Affected relatives younger than age 20 had mild to moderate hearing losses through 1000 Hz but retained normal or near-normal thresholds for higher frequencies. By age 30, thresholds were at severe levels across the entire frequency range, leading to flat profound hearing loss by age 40. Audiometric configurations for right and left ears were similar, with all deaf persons affected bilaterally. A minimal criterion for deafness in this kindred was a hearing threshold greater than 50 decibels at 250 Hz and 500 Hz. Relatives younger than age 25 with no apparent symptoms were not included in the analysis.

Whole blood for 99 informative relatives was drawn into acid citrate dextrose, lymphoblastoid cell lines were prepared, and DNA was extracted using methods previously described (7, 8). Southern hybridizations were carried out according to standard procedures (9–11). Amplification and electrophoresis of sequences containing microsatellite polymorphisms were carried out by methods described for these loci (12–15). In Costa Rica, where 32P was not available, sequences containing microsatellite polymorphisms were amplified without labeled nucleotides, then electrophoresed.

Abbreviations: ln, logarithm of odds; cM, centimorgan(s).

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.
FIG. 1. Linkage of deafness in the Monge kindred to markers on chromosome 5q31. Dark symbols indicate deaf persons; symbols with diagonal slashes represent deceased persons. The position of the “S.M.” branch in the kindred is not certain. Genotypes of some deceased persons are suggested on the pedigree in brackets, but these inferred genotypes were not included in the statistical analysis. Boxes indicate the haplotypes apparently linked to deafness in each branch of the kindred. By multipoint analysis, odds in favor of linkage of deafness to the region between IL9 and DSS210/DSS207 are >10^{11}.1. Recombination events in persons A, C, E, and F indicate that the deafness gene lies above GRL; recombination events in persons B, D, and G indicate that the deafness gene lies below IL9. The distance between GRL and IL9 is ≈7 centimorgans (cM).
Table 1. Polymorphisms on chromosome 5q linked to deafness in the Monge kindred

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele</th>
<th>θ</th>
<th>Primer or probe/enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIB5</td>
<td>8</td>
<td>0.12</td>
<td>AAG GTG TTC TTT GGA TGT TCA CC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GTA ATG TGT ATC TAG AGA G</td>
</tr>
<tr>
<td>IL9 (SE9)</td>
<td>6</td>
<td>0.07</td>
<td>AGG TTC AGG CTA GCT CAT GT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTA ATG CAG AGA TTT AGG GC</td>
</tr>
<tr>
<td>GRL</td>
<td>2</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>DSS70</td>
<td>2</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSS210 (Mfd122)</td>
<td>5</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>DSS207 (Mfd42)</td>
<td>3</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>DSS119 (Mfd6)</td>
<td>4</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>DSS209 (Mfd116)</td>
<td>8</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>DSS22</td>
<td>3</td>
<td>0.11</td>
<td></td>
</tr>
</tbody>
</table>

θ, Approximate distance between adjacent markers.

in 18–20% acrylamide gels, and stained with silver nitrate or ethidium bromide.

Autosomal dominant transmission of deafness in the Monge kindred was confirmed previously by segregation analysis (4). Linkage was evaluated using LINKAGE (16), postulating a rare autosomal dominant allele (q = 0.01) with complete penetrance and no sporadic cases. Multipoint analysis was carried out using LINKAGE (17). Only genotypes that were obtained explicitly for each person were included in the linkage analysis: no inferred genotypes nor any haplotypes were included in the statistical analysis.

RESULTS

Primer sequences or probe/enzyme combinations for markers on chromosome 5q are listed in Table 1 (see refs. 12–15 for allele frequencies). Order of markers and approximate distances between adjacent markers are based on data from CEPH (ref. 13; unpublished data) and from the Monge kindred. Relative orders of DSS210 vs. DSS207, of DSS119 vs. DSS209, and of GRL vs. DSS70 cannot be determined from our data so far.

Coinheritance of deafness with chromosome 5q markers in the Monge kindred is shown in Fig. 1. Haplotypes and genotypes for deceased persons are indicated in Fig. 1 but were not included in the statistical analysis. Two-point logarithm of odds (lod) scores for linkage of deafness to markers in this region are indicated in Table 2. Multipoint analysis of the 14-cM interval defined by IL9 and DSS210/DSS207 strongly suggests that the LFH1 gene is between these markers. The maximum lod score for a gene proximal to IL9 is 9.95; the maximum lod score for a gene within the interval is 12.42; the maximum lod score for a gene distal to DSS207/DSS210 is 7.46. The maximum lod score within the interval occurs ≈5 cM distal to IL9, which corresponds to a locale slightly proximal to GRL.

Table 2. Maximum two-point lod scores (Z), maximum likelihood recombination fractions (θ), and 95% confidence intervals for θ for linkage of deafness to chromosome 5 markers

<table>
<thead>
<tr>
<th>Marker</th>
<th>Z</th>
<th>θ</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIB5</td>
<td>6.63</td>
<td>0.06</td>
<td>(0.01, 0.12)</td>
</tr>
<tr>
<td>IL9</td>
<td>13.55</td>
<td>0.06</td>
<td>(0.01, 0.12)</td>
</tr>
<tr>
<td>GRL</td>
<td>2.65</td>
<td>0.10</td>
<td>(0.02, 0.30)</td>
</tr>
<tr>
<td>DSS210/DSS207</td>
<td>7.50</td>
<td>0.13</td>
<td>(0.05, 0.31)</td>
</tr>
</tbody>
</table>

Seven persons with informative recombination events place the deafness gene LFH1 between IL9 and GRL. In Fig. 1, persons A, C, E, and F indicate that LFH1 is proximal to GRL; B, D, and G indicate that LFH1 is distal to IL9.

DISCUSSION

Deafness in this kindred appears to be determined by a gene between IL9 and GRL, an interval of ≈7 cM. Because LFH1 is primary deafness, with no associated abnormalities, the normal function of the LFH1 gene may be specific to hearing. Other occurrences of sensorineural deafness, whether inherited or apparently sporadic, may be due to other mutations at the LFH1 locus (3, 18). Linkage information from the Monge kindred will permit this hypothesis to be tested in other families. The LFH1 gene appears not to be identical to the gene for Treacher Collins syndrome, a complex mandibulofacial disorder sometimes involving deafness, which maps distal to GRL and, hence, distal to the LFH1 gene (19, 20).

Physical maps of human chromosome 5q indicate that IL9 and GRL map to 5q31 (21, 22). Genes coding for early growth response (EGRI), antigen CD14, and endothelial growth factor (FGF-A) are located between IL9 and GRL (21). This region of human chromosome 5q is homologous to a portion of mouse chromosome 11. The mouse shaker-2 locus maps to this region, but outside the interval defined by Sparc and the interleukin loci (23). If genes in this region are in the same order in mouse and humans, the shaker-2 locus would be outside the critical region of linkage for LFH1 (21).

The LFH1 gene can be further localized by mapping deafness in the Monge kindred relative to additional markers in this region of chromosome 5q. Given that ≈150 informative relatives have agreed to participate (of whom 99 have been sampled thus far), this large kindred offers the potential of mapping this gene to a resolution of ≈1 cM. Coding sequences in close physical proximity to markers showing no recombination with LFH1 can then be investigated.

We thank J. R. Sanchez, F. Mengel, and C. Vargas for assistance with audiometric testing and sampling; E. Fournier for help with the Monge genealogy; J. L. Weber; J. A. Warrington, J. J. Wasmuth, E. W. Jabs, M. A. Vance, L. Yamaoka, and M. C. Speer for sharing data on chromosome 5; M. K. Lee for statistical analysis; R. Giles for technical assistance; and K. Shaver-Arns for helpful discussion. This work was supported in part by National Institutes of Health Grant RO1 DC01076-01.
Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis

Victor L. Ruiz-Perez1, Susan E. Ide2,5, Tim M. Strom3, Bettina Lorenz3, David Wilson1, Kathryn Woods1, Lynn King4, Clair Francomano6, Peter Freisinger3, Stephanie Spranger7, Bruno Marino8, Bruno Dallapiccola9, Michael Wright1, Thomas Meitinger3, Mihael H. Polymeropoulos2 & Judith Goodship1

*These authors contributed equally to this work.

1Human Genetics Unit, School of Biochemistry and Genetics, Newcastle University, Newcastle upon Tyne, UK. 2Novartis Pharmaceuticals Corporation, Pharmacogenetics Division, Gaithersburg, Maryland, USA. 3Abteilung Medizinische Genetik, Klinikum Innenstadt, Ludwig-Maximilians-Universität, München, Germany. 4Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA. 5George Washington University, Graduate Genetics Program GWIBS, Washington DC, USA. 6Kinderklinik der Technischen Universität, München, Germany. 7Centre for Human Genetics, University of Bremen, Bremen, Germany. 8Ospedale Pediatrico Bambino Gesù, Rome, Italy. 9La Sapienza University, CSS-Mendel Institute, Rome, Italy. Correspondence should be addressed to J.G. (e-mail: j.a.goodship@ucl.ac.uk).

Ellis-van Creveld syndrome (EvC, MIM 225500) is an autosomal recessive skeletal dysplasia characterized by short limbs, short ribs, postaxial polydactyly and dysplastic nails and teeth.1,2 Congenital cardiac defects, most commonly a defect of primary atrial septation producing a common atrium, occur in 60% of affected individuals. The disease was mapped to chromosome 4p16 in nine Amish subpedigrees and single pedigrees from Mexico, Ecuador and Brazil.3 Weyers acrodental dysostosis (MIM 193530), an autosomal dominant disorder with a similar but milder phenotype, has been mapped in a single pedigree to an area including the EvC critical region.4 We have identified a new gene (EVC), encoding a 992–amino-acid protein, that is mutated in individuals with EvC. We identified a splice-donor change in an Amish pedigree and six truncating mutations and a single amino acid deletion in seven pedigrees. The heterozygous carriers of these mutations did not manifest features of EvC. We found two heterozygous missense mutations associated with a phenotype, one in a man with Weyers acrodental dysostosis and another in a father and his daughter, who both have the heart defect characteristic of EvC and polydactyly, but not short stature. We suggest that EvC and Weyers acrodental dysostosis are allelic conditions. The EvC locus was localized between D4S2957 and D4S827 in Amish and Brazilian pedigrees.5,6 We constructed a physical map of this region by searching the MIT database (http://www.genome.wi.mit.edu) for YAC clones and the Stanford Human Genome Center (SHGC) database for BAC clones (Fig. 1a, http://www-shgc.stanford.edu). We analysed the available genomic sequence from these BACs (http://www.shgc.stanford.edu) for microsatellite repeats and identified two polymorphic markers, 500H20P5 and 164F16P2. We analysed microsatellites in the Amish and Brazilian pedigrees in the region initially defined as segregating with the disease.3 Recombinations in the Amish and Brazilian pedigrees placed the disease centromeric to D4S2375 and a recombination in the Brazilian pedigree placed the disease telomeric to 164F16P2 (Fig. 2), refining the EvC region to an interval of less than 1 Mb (Fig. 1a).
We analysed unfinished genomic sequence from BACs 135O5, 69D13, 164F16 and 568K5 (http://www-shgc.stanford.edu) using exon-prediction programmes. This revealed two previously identified genes, aCI and CRMP1 (encoding collapsin response mediator protein), neither of which were mutated in our EvC families. Sequence analysis also predicted a gene encoding a protein with homology to serine-threonine kinases, which was confirmed by RACE and RT–PCR (Fig. 1a). We also screened orphan predicted exons in the region, and identified a homozygous nonsense mutation in an affected individual (NCL12) which led us to focus our attention on BAC 69D13. There were two groups of predicted exons in the sequence available from BAC 69D13 (Fig. 1b, labelled exons 8–11 and 14–20). We linked these by RT–PCR and used the resulting product, RT1, to probe northern blots of fetal and adult tissue to estimate the size of the gene. A 7-kb transcript was seen in fetal kidney and a fainter 7-kb band in fetal lung. A faint transcript of the same size was seen in adult kidney following a one-week exposure.

Sequence analysis identified two ESTs (IMAGE clones 2168182 and 365528) downstream of exons 8–20. Northern-blot analysis of clone 2168182 showed that it had the same tissue distribution and transcript size as RT1, indicating they are part of the same gene. The IMAGE clones 2168182 and 365528, derived from adult brain and 19-week fetal heart, respectively, are identical at the 5′ end, but their 3′ ends differ (Fig. 1b). Clone 2168182 aligned directly with the genomic sequence. Alignment of clone 365528 with the genomic sequence indicated the presence of introns within the 3′ UTR. Moreover, sequence from the centromeric end of 365528 was complementary to the coding sequence of CRMP1 exon 12. RT–PCR between exon 18 and unique sequence from each IMAGE clone confirmed the alternate 3′ ends of this gene (Fig. 1b), identified the stop codon and also revealed an alternative splice acceptor at the beginning of exon 21, leading to the presence or absence of a serine residue. We obtained the 5′ sequence of the gene by 5′-RACE and confirmed it by RT–PCR using primers upstream of the initiation codon. We identified an ATG within a KOZAK sequence with loss of the ORF 12 bases upstream of the ATG. The length of the cDNA is 6.43 kb or 7.06 kb depending on 3′ end usage, consistent with the transcript seen on northern-blot analysis.

We identified a mouse EST (AI615515) by ESTblast searching that was 82% identical to EVC. We completed the mouse Evc cDNA by 5′- and 3′-RACE. There is 79% similarity and 66.8% identity between mouse and human EVC at the amino acid level. As seen in the human cDNA, there is alternate splicing of the mouse transcript at the position corresponding to the beginning of human exon 21. This leads to the presence of 13 additional amino acids in some mouse transcripts (Fig. 3). The intron between exons 20 and 21 has been sequenced in human and there is no sequence that corresponds with these amino acids.

Translation of the human coding sequence gives a 992-amino-acid protein. Motif searching identified a leucine zipper, three putative nuclear localization signals and a putative transmembrane domain (Fig. 3).

We designed intronic primers to amplify the coding exons (1–21) and screened samples by SSCP, sequencing the products with altered migration patterns. We found four homozygous mutations in the coding sequence: Q879X in exon 18, a 1-nt insertion in exon 7 (nt910-911 insA), deletion of 3 nt from exon 7 (ΔK302; Fig. 4) and deletion of exons 12–21 (Fig. 5). We identified two truncating mutations, R340X and 734delT, in patient 16086, whose parents were not available for study. An affected child, whose father has Weyers acrodental dysostosis, was a compound heterozygote with a missense mutation (S507P) inherited from her father and a 1-nt deletion (2456delG) on the maternal allele. We tested a panel of 100 normal control chromosomes for each of the mutations. We identified another heterozygous missense change, R443Q, in a father and his daughter, who both have postaxial polydactyly of hands and feet, partial atrioventricular canal with common atrium and agenesis of the upper lateral incisors bilaterally with enamel abnormalities. We did not find the R443Q change in a larger panel of 194 normal chromosomes.

We sequenced the entire coding sequence in affected individuals from the Amish and Brazilian pedigrees. We observed two variants in the Amish, a missense change and a splice-donor-site change, R760Q and a G→T substitution in intron IVS13+5, respectively. Both variants segregate with the disease in all nine branches of the family. We found the R760Q change in 1 British control and 1 CEPH control of 97 normal chromosomes (194 chromosomes) tested. As the gene is not transcribed in lymphocytes, we have not yet
Carnegie stages 19 and 21, we detected low levels of nucleotide insertion. NCL 16 has a single-codon deletion. Patient NCL12 has a sequence in NCL14 and NCL16 are from the reverse strand. NCL 14 has a single-

Fig. 4 has been shown for association with disease in a number of genes and alternate splicing. A change at this position has been reported in asso-

Fig. 3 Comparison of human and mouse protein sequence. The leucine zipper region is shaded, the putative transmembrane domain is boxed and the putative nuclear localization signals are underlined. Serine residue 965 is not present in all human transcripts. The 13 aa (TYSSASPRPRHVS) in the mouse protein sequence are not present in all transcripts.

been able to demonstrate that the intronic change leads to alternate splicing. A change at this position has been reported in asso-

Fig. 4 DNA sequence showing homozygous mutations in three patients. The sequence in NCL14 and NCL16 are from the reverse strand. NCL 14 has a single-nucleotide insertion. NCL 16 has a single-codon deletion. Patient NCL12 has a C→T transversion, leading to Q879X.

Methods

Patients. We obtained samples from the Amish and Brazilian pedigrees, a child with EvC whose father had features of Weyers acrodental dysostosis, an individual from the island of Barra, and a father and daughter with an EvC-like phenotype. We obtained samples from a further six EvC families, five of whom were known to be consanguineous.

Physical map. We confirmed the presence or absence of chromosome 4p16.1 STSs as indicated in the SHGC map by PCR (data not shown). We identified a gap between BACs 135OS and 69D13 in the Stanford map. We designed STSs by directly sequencing the ends of clones 135O5 and 565K5, which has now been integrated into the Stanford contig. We iso-

Microsatellite analysis. We carried out saturation genotyping using repeat-con-

EVC cDNA cloning. We performed 5´ RACE using marathon ready human brain cDNA (Clontech) or human fetal kidney poly(A) RNA, 16–32 weeks (Clontech), in conjunction with the 5´ RACE kit (Gibco-BRL). To achieve 5´ RACE or RT–PCR products containing the GC-rich exon 1, PCR mixtures were supplemented with betaine (2 M). All other RT–PCR reactions were carried out by standard methods.
Polymeropoulos, M.H. et al. Dwarfism in the Amish I. 14 from patient 16218. The 7-kb fragment corresponding to exon 11 is present
exon 12 and the absence of a 12-kb fragment corresponding to exons 13 and
Autoradiography revealed the absence of a 1-kb fragment corresponding to
Northern-blot analysis. We amplified RT–PCR product comprising exons 11–14.
We used the integrated NIX programme (http://
www.hgmp.mrc.ac.uk) or the integrated RUMMAGE programme (http://
We performed all hybridizations following the manufacturer’s
tools and checked lane loading in the northern blot by hybridizing with
Tissue in situ hybridization. We obtained ethical permission for the
use and collection of human embryos. We collected the embryos following either
surgical or medically induced termination of pregnancy and determined their
developmental stage by stereomicroscopy14. They were then fixed in
tissue  in  situ  by  standard techniques
1. Ellis, R.W.B. & van Creveld, S. A syndrome characterised by ectodermal dysplasia,
polydactyly, chondrodysplasia and congenital morbus gorois: report of three
9. Ishida, Y. et al. Isolation and characterisation of 21 novel expressed DNA
10. Wilson, G.M. et al. Regulation of AUF1 expression via conserved alternatively
splice variants in the 3’ untranslated region. Mol. Cell. Biol. 6, 4056–4064
12. Ubeda, M., Schmitt-Nei, M., Ferrer, J. & Hubener, J.F. CHOPIGADD153 and the
methionyl-tRNA synthetase (MetRS) genes overlap in a conserved region that
13. Hill, R.D. Two cases of Ellis-van Creveld syndrome in a small island population.
guide. in Molecular Genetics of Early Human Development (eds Strachan, T.
15. Hanley, N.A. et al. Expression of steroidogenic factor 1 and Wilm’s tumour 1
during early human gonadal development and sex determination. Mech. Dev. 87,

Received 21 October 1999; accepted 7 February 2000.

Table 1 • EVC mutations and polymorphisms

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Sequence change</th>
<th>Exon</th>
<th>Protein effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>16086 (compound heterozygous)</td>
<td>734delT</td>
<td>10</td>
<td>R340X</td>
</tr>
<tr>
<td>ncl-16 (homozygous)</td>
<td>904-906delAAG</td>
<td>7</td>
<td>ΔK302v</td>
</tr>
<tr>
<td>ncl-14 (homozygous)</td>
<td>901-911insA</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>15219 (compound heterozygous)</td>
<td>919T→C</td>
<td>7</td>
<td>S307P</td>
</tr>
<tr>
<td>I-1 (heterozygous)</td>
<td>1328G→A</td>
<td>10</td>
<td>R443Q</td>
</tr>
<tr>
<td>16218 (homozygous)</td>
<td>deletion</td>
<td>12-21</td>
<td></td>
</tr>
<tr>
<td>Amish (homozygous)</td>
<td>IVS13+5G→T</td>
<td>intron13</td>
<td></td>
</tr>
<tr>
<td>ncl-12 (homozygous)</td>
<td>2635C→T</td>
<td>18</td>
<td>Q879X</td>
</tr>
</tbody>
</table>

polymorphisms

Mutations in patients where the mutation on each allele has been identified and a list of
polymorphisms that lead to amino acid changes. The R760Q change, observed in 2 of 194 nor-
mal chromosomes, segregated with the disease allele in the Amish pedigree.

Mutation analysis. We designed intronic primers to amplify 20 of the
coding exons by aligning the cDNA sequence with genomic sequence (http://www-shgc.stanford.edu). Genomic sequence corresponding to exon 13 was not in the database and we obtained it by direct BAC sequencing using exon13–specific primers by direct BAC sequencing in order to
design intronic primers. We extracted genomic DNA from
peripheral blood leucocytes or fetal tissue and undertook mutation
screening by SSCP. We sequenced PCR fragments showing aberrant
migration patterns. We tested the changes identified in a panel of 100
normal control chromosomes by sequencing, SSCP or on the basis of
altered restriction sites. We sequenced exons 1–21 in affected individuals
from the Amish and Brazilian pedigrees. We performed Southern-blot
analysis by standard methods.

GenBank accession numbers. EVC, AF216168, AF216185; Evc, AJ250841;
human serine/threonine protein kinase, AJ250839; Mus musculus
serine/threonine protein kinase, AJ250840; microsatellites, AF195414,
AF195415, AF195416.

Acknowledgements

We thank A. Verloes, A. Nerlich and J. Young for sending samples from
patients, and I. Burn and T. Strachan for encouragement. This work was
supported by the British Heart Foundation, Newcastle Hospital Special
Trustees, the Knott Trust, the Borwick Trust and the Deutsche
Forschungsgemeinschaft.

1. Ellis, R.W.B. & van Creveld, S. A syndrome characterised by ectodermal dysplasia,
polydactyly, chondrodysplasia and congenital morbus gorois: report of three
3. Polymeropoulos, M.H. et al. The gene for the Ellis-van Creveld syndrome is
and dental abnormalities map to chromosome 4p16, in the region containing the
5. Ide, S.E., Ortiz de Luna, R.I., Francamano, C.A. & Polymeropoulos, M.H. Exclusion of
the MSX1 homeobox gene as the gene for the Ellis van Creveld syndrome in the
7. Digilio, M.C., Marino, B., Giannotti, A. & Dallapiccola, B. Single aturium,
atrioventricular canal/postaxial hexadactyly indicating Ellis-van Creveld syndrome.
9. Ishida, Y. et al. Isolation and characterisation of 21 novel expressed DNA
10. Wilson, G.M. et al. Regulation of AUF1 expression via conserved alternatively
spliced elements in the 3’ untranslated region. Mol. Cell. Biol. 6, 4056–4064
(1999).
12. Ubeda, M., Schmitt-Nei, M., Ferrer, J. & Hubener, J.F. CHOPIGADD153 and the
methionyl-tRNA synthetase (MetRS) genes overlap in a conserved region that
13. Hill, R.D. Two cases of Ellis-van Creveld syndrome in a small island population.
guide. in Molecular Genetics of Early Human Development (eds Strachan, T.
15. Hanley, N.A. et al. Expression of steroidogenic factor 1 and Wilm’s tumour 1
during early human gonadal development and sex determination. Mech. Dev. 87,