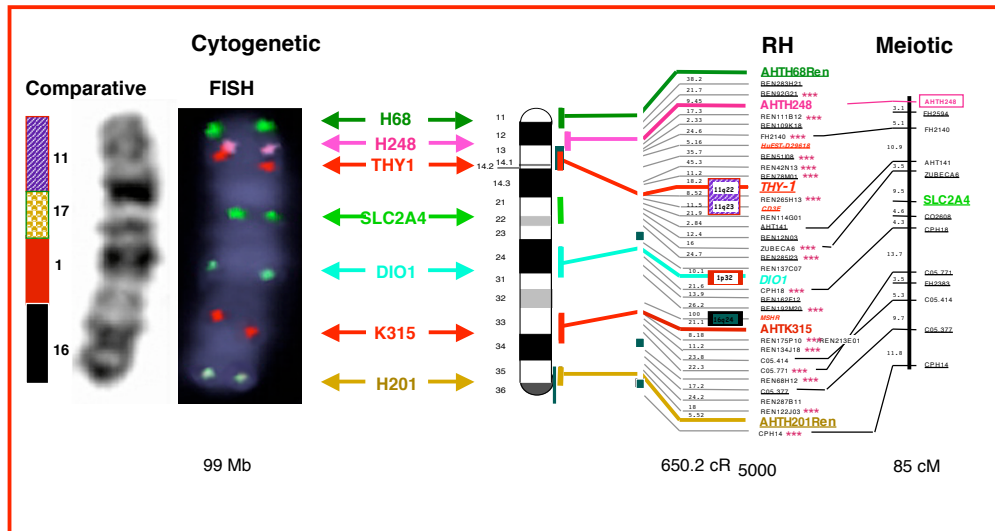# Everything you always wanted to know about genetic maps, markers, linkage analysis, Lod scores, and lots of other important stuff!

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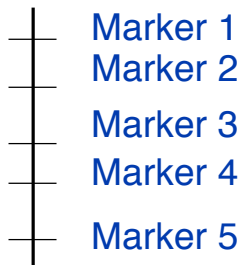Cancer Genetics Branch, NHGRI
March 2005

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

3

Example: RH, Meiotic Linkage and Cytogenetic Maps for Canine Chromosome 5

Cytogenetic   **RH**   **Meiotic**

Comparative   **FISH**

11

17

1

16

H68
H248
THY1

SLC2A4

DIO1

K315

H201

99 Mb

650.2 cR   85 cM

5000

4

Portion of Chromosome

Marker 1
Marker 2
Marker 3
Marker 4
Marker 5

Meiotic Linkage Maps

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

Polymorphism

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

Single Nucleotide Polymorphism-SNP

```
1        attctgggggaatctcttgcatatagtcc
         taagaccccttagagaacgtatatcagg

2        attctgggggaacctcttgcatatagtcc
         taagaccccttggagaacgtatatcagg
```

Kinds Of Genetic Markers

o 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
o 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
o 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

CACACACACACAC
GTGTGTGTCGTGT
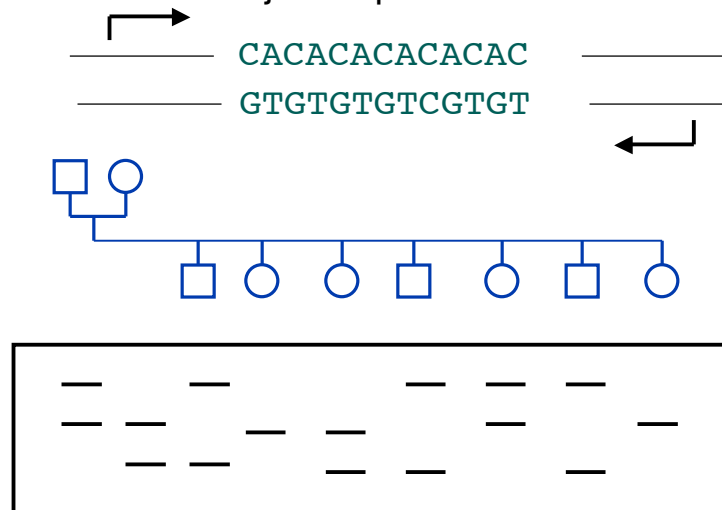
Two Common Measures of Polymorphism

o 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
o Polymorphic Information Content (PIC)
  – PIC - $\dfrac{(n-1)^2 (N +1)}{N^3}$
  – PIC 0 > P <1
  – Most "good" markers have PIC > 0.7
o "Polymorphic locus" is defined as one with Het or PIC >10%

Use Linkage Map to Localize Trait-Associated Locus

o 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

o Analysis of 250 families with about 7 sampled individuals per family will require close to 1 million genotypes
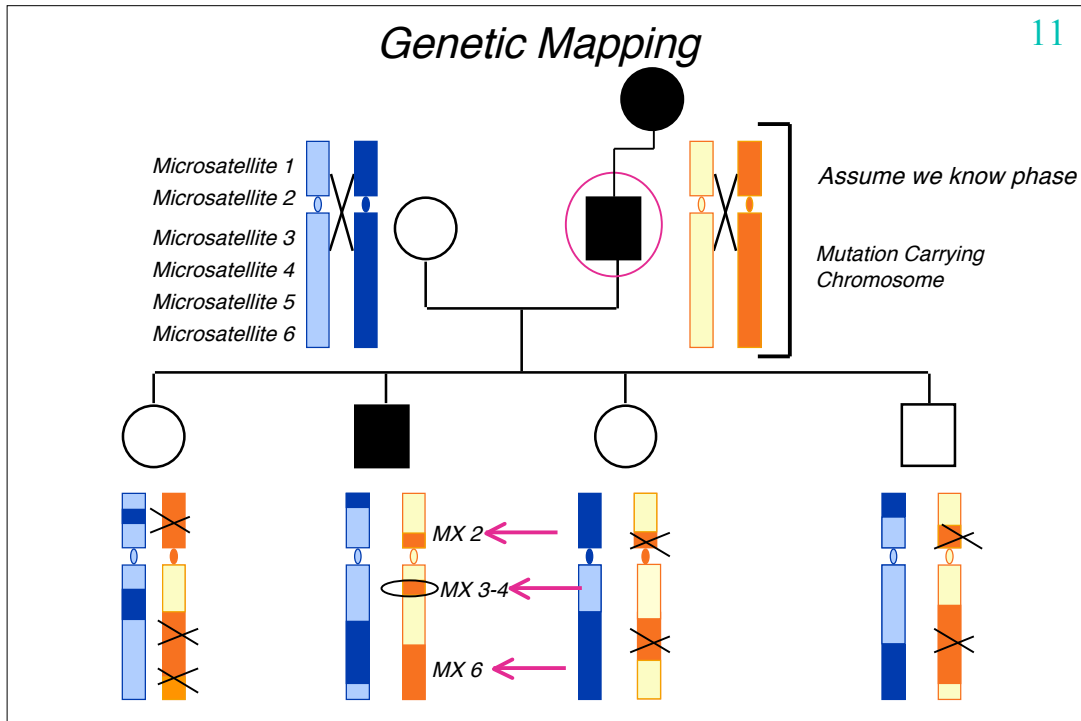
? ?

55 62 ?

61 56 56 51 46
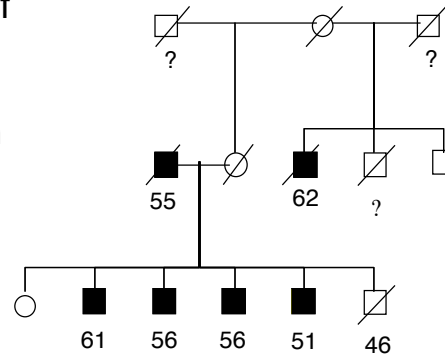
Genetic Linkage

o 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

o No Linkage
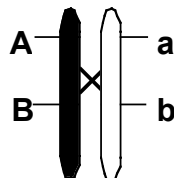– Two loci have a 50% chance of being co-inherited

Genetic Mapping

11

Microsatellite 1
Microsatellite 2
Microsatellite 3
Microsatellite 4
Microsatellite 5
Microsatellite 6

Assume we know phase

Mutation Carrying
Chromosome

MX 2
MX 3-4
MX 6

Genetic Distance

12

A — a
B — b

For a given area of the genome,
the recombination fraction, θ , 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

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

A "Linked" Marker

D = Disease
N = Normal

Putative linked allele

DN NN NN NN DN NN DN DN DD DN NN NN
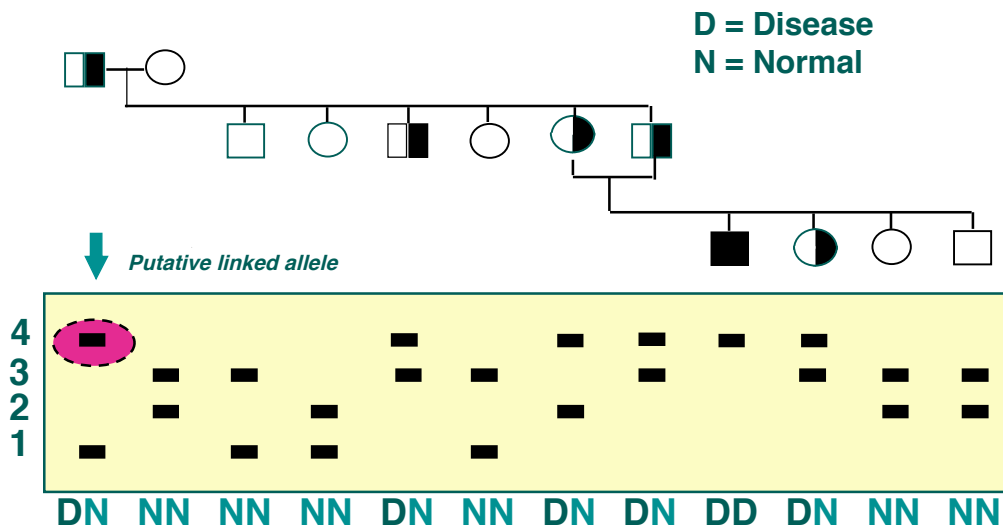
Terminology

₀Phenocopy--Individual with the same disease or trait, but it is associated with a different cause

₀Penetrance--How likely one is to get a disease (trait) if they carry a genetic variant

₀Sporadic--Disease (trait) due to factors other then those under consideration

   environmental factors

   weakly penetrant alleles

   stochastic events

Hypothetical Marker
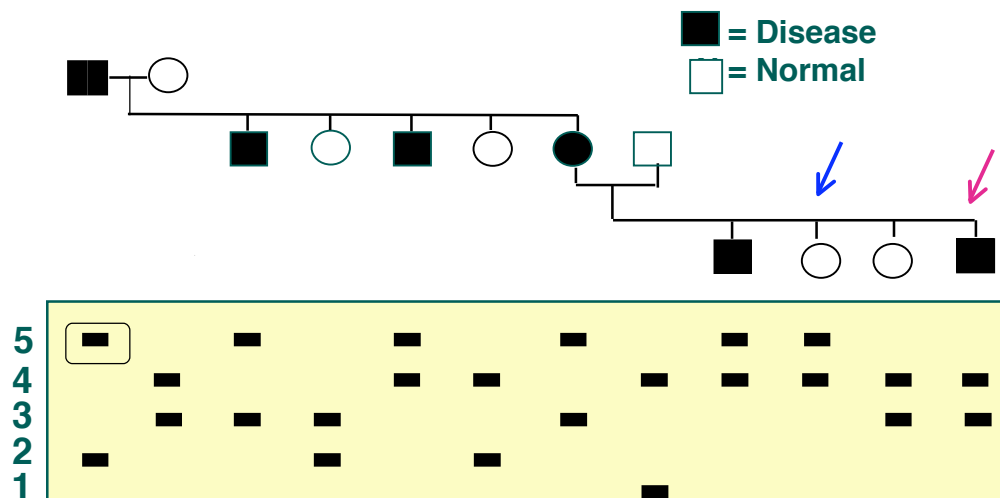
Is there evidence for linkage?

Three Possibilities (ok, Four) to Explain Data

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

Estimates of Linkage

o Genome-wide scan
  - Testing for linkage between markers and disease state
o 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 ≥ 3.3 significant
  - Indicate greater then 1000:1 odds in favor of linkage
o 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.
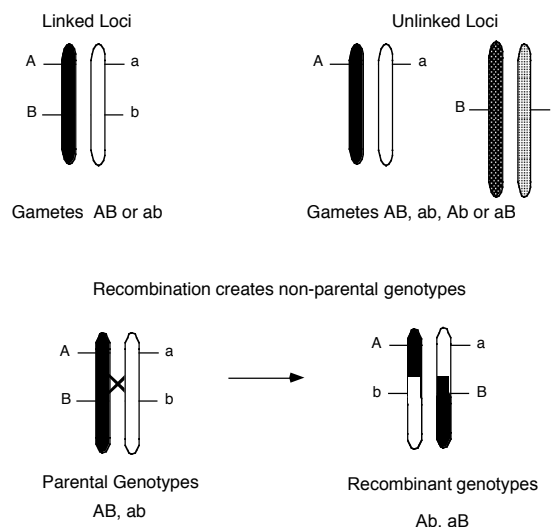
o While lod scores > 3.3 are generally accepted as evidence of linkage, Lod scores <-2.0 are accepted as evidence against linkage.
o Because they are calculated as Log of the Odds, lod scores may be summed across families.
o Adding families allows us to increase power.
o Lod scores are calculated for multiple values of θ.
o The value of θ at which the peak lod is obtained is taken as the maximum likelihood value of θ.

Genetic Recombination

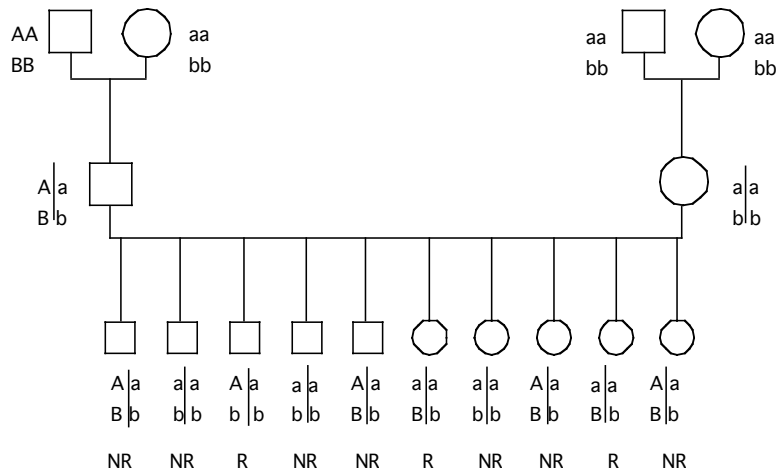n = total number of loci
k = number of recombinant alleles

Recombination fraction = θ = k/n

For linked loci θ <0.5
For unlinked loci θ = 0.5

Linked Loci

A ——— a
B ——— b

Gametes AB or ab

Unlinked Loci

A ——— a
B ——— b

Gametes AB, ab, Ab or aB

Recombination creates non-parental genotypes

A ——— a
B ——— b

A ——— a
b ——— B

Parental Genotypes
AB, ab

Recombinant genotypes
Ab, aB

AA □— ○ aa
BB   bb

aa □— ○ aa
bb   bb

A|a
B|b □ (son of AA/BB × aa/bb)

○ a|a
   b|b (daughter)

Offspring:
□ □ □ □ □ ○ ○ ○ ○ ○

A|a   a|a   A|a   a|a   A|a   a|a   a|a   A|a   a|a   A|a
B|b   b|b   b|b   b|b   B|b   B|b   b|b   B|b   B|b   B|b

NR   NR   R   NR   NR   R   NR   NR   R   NR

$\theta = 0.3$

$\theta < 0.5$ but is it significant?

Dreaded Linkage Summary

o 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.

o To evaluate this--Calculate a series of likelihood ratios (relative odds) at various values of θ ranging from 0 (no recombination) to $\theta = 0.5$ (random assortment).

o Thus, the likelihood at a given value of θ =

   Likelihood of data if loci unlinked at θ /likehood of data if loci unlinked

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

o 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 $\theta = 0.3$

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

$Z\theta = n \log(2) + k\log(\theta) + (n-k)\log(1-\theta)$ if $\theta > 0$

10 log (2) + 3 (log) 0.3 + (7) log (0.7) =

3.01 + (-1.6) + (-1) =

lod score = 0.357   **Do we have linkage?**

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

---

Types of Analyses

o Parametric

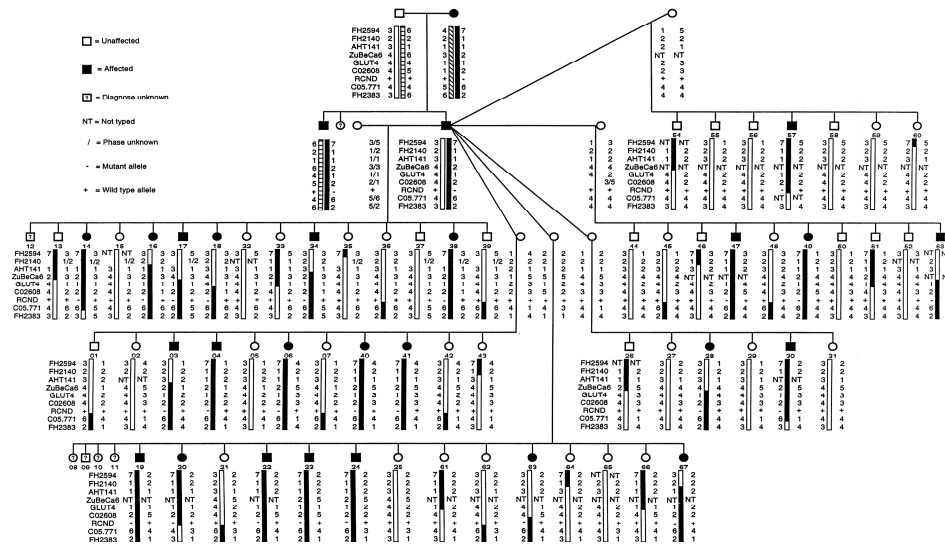   Derive "transmission model" where assumptions are made regarding:
   - Age-dependent penetrance
   - Frequency of mutant allele in population
   - Mode of transmission

o 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 Lod = 16.7 (θ = 0.016)

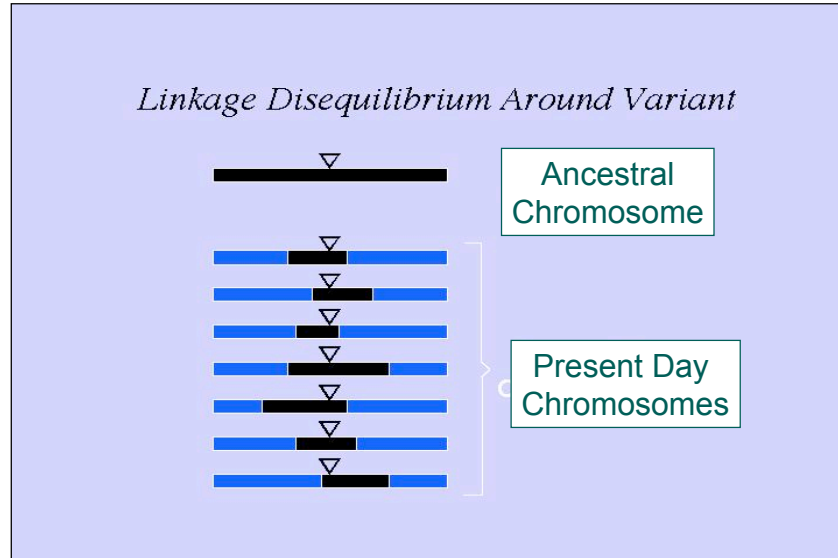Jonnasdottir et al., (2003) PNAS 100: 5296-5301

Linkage Analysis Using Founder Effects: Two Strategies

o One large pedigree with disease derived from single founder
  - Leon et al., (1992) PNAS 89: 5281-5284
o Many smaller pedigrees with similar features that share common heritage
  - Ruiz-Perez et al., (2000) Nature Genetics 24: 283-286.

Linkage Disequilibrium

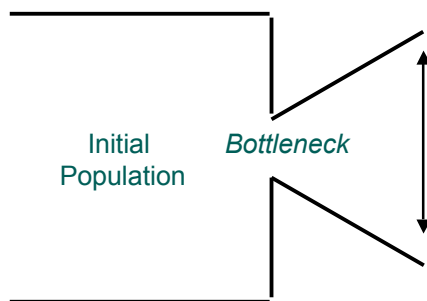### Linkage Disequilibrium Around Variant

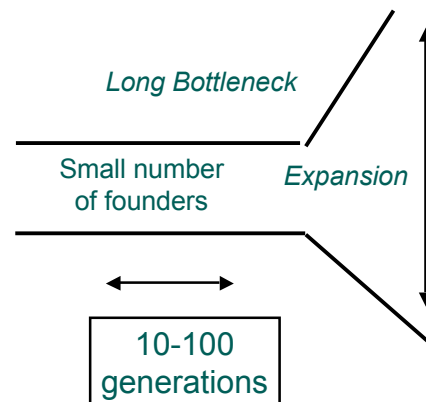Ancestral Chromosome

Present Day Chromosomes

Isolated Population Model

Initial Population

Bottleneck

Founder Effect Model

Long Bottleneck

Small number of founders

Expansion

10-100 generations

Leon et al., 1992

- o See Table 1--What is it telling you about?
  - o Information content and polymorphism
- o See Table 2. How do you explain the lod scores going up and down?
  - • Hint: map order and/or marker informativeness
- o 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?

Leon et al., 1992

Table 1. Polymorphisms on chromosome 5q linked to deafness in the Monge kindred

Locus	Allele	θ	Primer or probe/enzyme
FIB5	8		AAG GTG TTC TTT GCA TGT TCA CC GTA ATG TGT TCT ATC TAG TTC AAC G
IL9 (SE9)	6	0.12	AGG TCC AGG CTA GCT CAT GT CTA ATG CAG AGA TTT AGG GC
GRL	2	0.07	OB7/*Bcl* I
D5S70	2	0.00	TP5E/*Taq* I
D5S210 (Mfd122)	5	0.07	ATG CAG AAT CTA CAA GGA CC CTT TAA CAT CCT TTA ACA GC
D5S207 (Mfd42)	3	0.00	TTG GAA GCC TTA GGA AGT GC AAG AAT TCT AGT TTC AAT ACC G
D5S119 (Mfd6)	4	0.14	TCC TAC CTT AAT TTC TGC CT GCA GGT TGT TTA ATT TCG GC
D5S209 (Mfd116)	8	0.00	CTG CAG TAG AAA GGC AGA GT TGC AGC ACC AAA CAC CAA GT
D5S22	3	0.11	JO205H-C/*Msp* I

θ, Approximate distance between adjacent markers.

Leon et al., 1992

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

Marker	Z	θ	95% confidence interval
FIB5	6.63	0.06	(0.01, 0.12)
IL9	13.55	0.06	(0.01, 0.12)
GRL	2.65	0.10	(0.02, 0.30)
D5D210/D5S207	7.50	0.13	(0.05, 0.31)

Leon et al., 1992
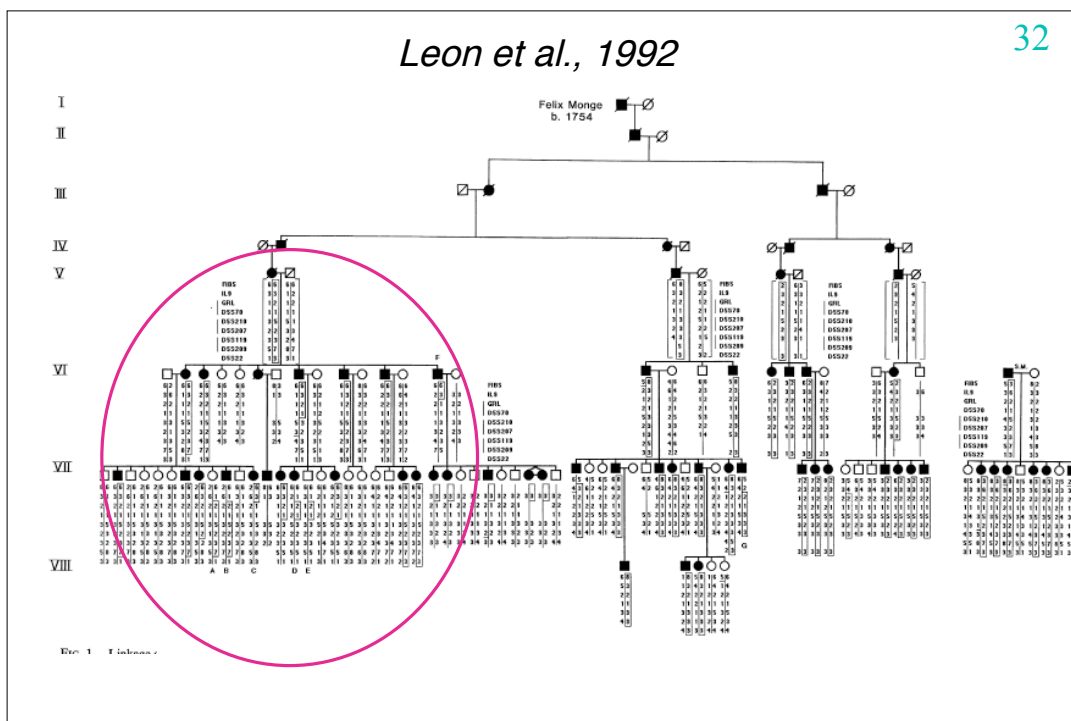
Fig. 1. Linkage

Leon et al., 1992 Pedigree
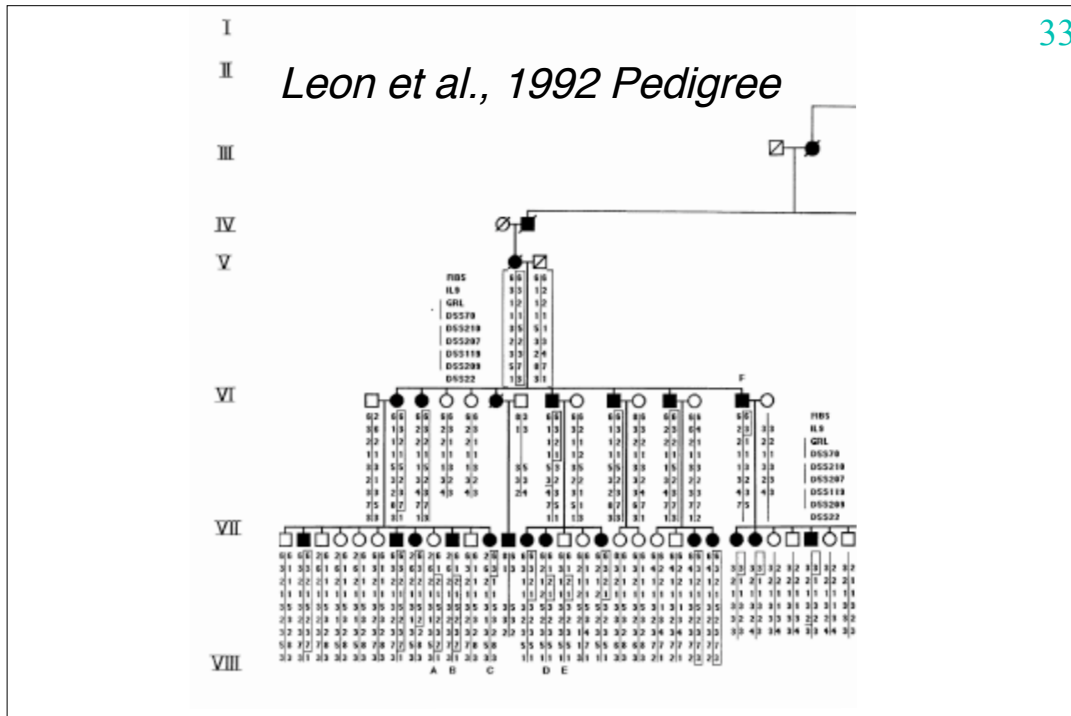
Leon et al.,1992

o What are problems associated with mapping deafness genes?
  - Number of forms of disease
  - Mating not random
  - Association with other congenital problems
  - Incomplete penetrance

o What other genes would you imagine would be this hard to map?

- Cancer, epilepsy, behavioral (alcoholism), mental illness (schizophrenia)

o 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?

o Autosomal dominant

o Highly penetrant

o Single founder (Felix Monge born in 1754)

o Low mobility of family members (Cartago, Costa Rica)

o Societal acceptance so families opt for multiple children

o Statistical power--150 people available-99 in this report

Ruiz-Perez et al., 2000

o What is the disease?
  - Recessive form of dwarfism found in Old Order Amish
o What features of Amish populations make them ideal for studies like this one?
  - Descended from small number of founders
  - Strict endogamy
  - Centrifigal gene flow
  - Genealogical records
  - Large families
o 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
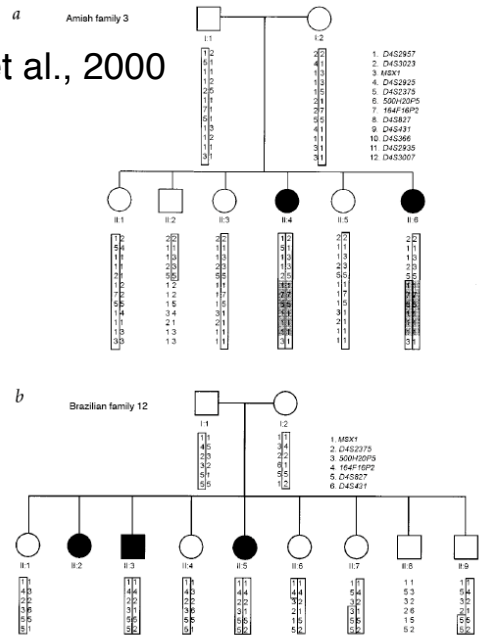
Ruiz-Perez et al., 2000

Fig. 2 Haplotype analysis in Amish family 3 and Brazilian family 12. The markers are ordered from telomere to centromere. The disease haplotypes are boxed and regions of homozygosity are shaded.

What is the Primary Result?

Ellis-van Creveld syndrome is an autosomal <u>recessive</u> 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 <u>dominant</u> 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.

39

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?

40

Summary

o There are several kinds of maps. Those composed of large numbers of well spaced polymorphic markers are used for genetic mapping studies.

o Pedigree analysis and accurate phenotypes are key for successful linkage studies.

o Multiple kinds of analyses can be done. In general, however, one is looking to find markers whose alleles segregate faithfully with phenotypes in families.

o 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.

o 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

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 *IL9* 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 (*LFHL1*; no. 124900 in ref. 3) is an autosomal dominant, fully penetrant, sensorineural 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, 100, 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 ^{32}P was not available, sequences containing microsatellite polymorphisms were amplified without labeled nucleotides, then electrophoresed

Abbreviations: lod, logarithm of odds; cM, centimorgan(s).
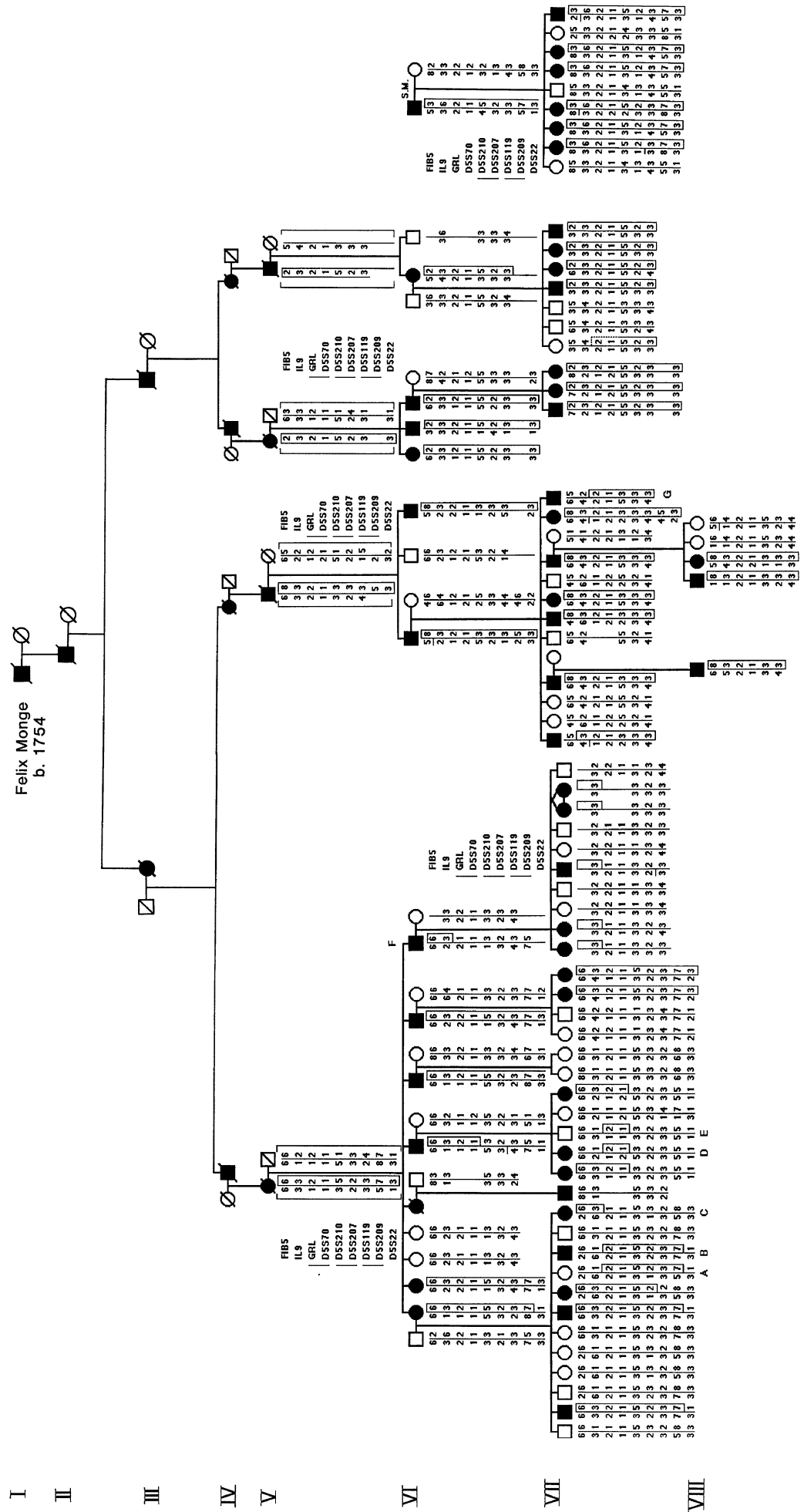
FIG. 1. Linkage of deafness in the Monge kindred to markers on chromosome 5q31. Dark symbols indicate deaf persons; symbols with diagonal slashes represent deceased persons. The position of the "S.M." branch in the kindred is not certain. Genotypes of some deceased persons are suggested on the pedigree 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 *D5S210/D5S207* are >10¹²:1. 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

Locus	Allele	θ	Primer or probe/enzyme
FIB5	8		AAG GTG TTC TTT GCA TGT TCA CC GTA ATG TGT TCT ATC TAG TTC AAC G
IL9 (SE9)	6	0.12	AGG TCC AGG CTA GCT CAT GT CTA ATG CAG AGA TTT AGG GC
GRL	2	0.07	OB7/*Bcl* I
D5S70	2	0.00	TP5E/*Taq* I
D5S210 (Mfd122)	5	0.07	ATG CAG AAT CTA CAA GGA CC CTT TAA CAT CCT TTA ACA GC
D5S207 (Mfd42)	3	0.00	TTG GAA GCC TTA GGA AGT GC AAG AAT TCT AGT TTC AAT ACC G
D5S119 (Mfd6)	4	0.14	TCC TAC CTT AAT TTC TGC CT GCA GGT TGT TTA ATT TCG GC
D5S209 (Mfd116)	8	0.00	CTG CAG TAG AAA GGC AGA GT TGC AGC ACC AAA CAC CAA GT
D5S22	3	0.11	JO205H-C/*Msp* I

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 LIPED (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 *D5S210* vs. *D5S207*, of *D5S119* vs. *D5S209*, and of *GRL* vs. *D5S70* 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 *D5S210*/*D5S207* strongly suggests that the *LFHL1* 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 *D5S207*/*D5S210* 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

Marker	Z	θ	95% confidence interval
FIB5	6.63	0.06	(0.01, 0.12)
IL9	13.55	0.06	(0.01, 0.12)
GRL	2.65	0.10	(0.02, 0.30)
D5D210/*D5S207*	7.50	0.13	(0.05, 0.31)

Seven persons with informative recombination events place the deafness gene *LFHL1* between *IL9* and *GRL*. In Fig. 1, persons A, C, E, and F indicate that *LFHL1* is proximal to *GRL*; B, D, and G indicate that *LFHL1* 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 *LFHL1* is primary deafness, with no associated abnormalities, the normal function of the *LFHL1* gene may be specific to hearing. Other occurrences of sensorineural deafness, whether inherited or apparently sporadic, may be due to other mutations at the *LFHL1* locus (3, 18). Linkage information from the Monge kindred will permit this hypothesis to be tested in other families. The *LFHL1* 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 *LFHL1* 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 (*EGR1*), antigen CD14, and endothelial growth factor (*FGFA*) 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 *LFHL1* (21).

The *LFHL1* 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 *LFHL1* can then be investigated.

1. Konigsmark, B. W. & Gorlin, R. J. (1976) *Genetic and Metabolic Deafness* (Saunders, Philadelphia).
2. Fraser, G. R. (1976) *The Causes of Profound Deafness in Childhood* (Johns Hopkins Univ. Press, Baltimore).
3. McKusick, V. A. (1990) *Mendelian Inheritance in Man* (Johns Hopkins Press, Baltimore), 9th Ed.
4. Leon, P. E., Bonilla, J. A., Sanchez, J. R., Vanegas, R., Villalobos, M., Torres, L., Leon, F., Howell, A. L. & Rodriguez, J. A. (1981) *Am. J. Hum. Genet.* **33,** 209–214.
5. Moulton, C. (1983) Thesis (Portland State Univ., Portland, OR).
6. Carhart, R. & Jerger, J. (1959) *J. Speech Hear. Disord.* **24,** 330–345.
7. Hall, J. M., Zuppan, P., Anderson, L. A., Huey, B., Carter, C. & King, M.-C. (1989) *Am. J. Hum. Genet.* **44,** 577–584.
8. Louie, L. & King, M.-C. (1991) *Am. J. Hum. Genet.* **48,** 637–638.
9. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
10. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
11. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
12. Weber, J. L. & May, P. E. (1989) *Am. J. Hum. Genet.* **44,** 388–396.
13. Weber, J. L., Polymeropoulos, M. H., May, P. E., Kwitek, A. E., Xiao, H., McPherson, J. D. & Wasmuth, J. J. (1991) *Genomics* **11,** 695–700.
14. Polymeropoulos, M. H., Xiao, H., Rath, D. S. & Merril, C. R. (1991) *Nucleic Acids Res.* **19,** 688.
15. Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattei, M.-G., Sarfarazi, M., Tsipouras, P., Ramirez, F. & Hollister, D. W. (1991) *Nature (London)* **352,** 330–334.
16. Ott, J. (1985) *Analysis of Human Genetic Linkage* (Johns Hopkins Univ. Press, Baltimore).
17. Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. (1985) *Am. J. Hum. Genet.* **37,** 482–498.
18. Ives, E. & Collis, E. (1991) *Am. J. Hum. Genet.* **49,** Suppl., 142.
19. Jabs, E. W., Li, X., Coss, C. A., Taylor, E. W., Meyers, D. A. & Weber, J. L. (1991) *Genomics* **11,** 193–198.
20. Dixon, M. J., Read, A. P., Donnai, D., Colley, A., Dixon, J. & Williamson, R. (1991) *Am. J. Hum. Genet.* **49,** 17–22.
21. Warrington, J. A., Hall, L. V., Hinton, L. M., Miller, J. N., Wasmuth, J. J. & Lovett, M. (1991) *Genomics* **11,** 701–708.
22. Westbrook, C. A., Neuman, W. L., Hewitt, J., Kidd, K. K., LeBeau, M. M. & Williamson, R. (1991) *Genomics* **10,** 1105–1109.
23. Davisson, M. T., Doolittle, D. P. & Roderick, T. H. (1991) *GBASE: The Genomic Database of the Mouse* (The Jackson Laboratory, Bar Harbor, ME).

letter

# Mutations in a new gene in Ellis-van Creveld syndrome and Weyers acrodental dysostosis

Victor L. Ruiz-Perez[1]*, Susan E. Ide[2,5]*, Tim M. Strom[3]*, Bettina Lorenz[3], David Wilson[1], Kathryn Woods[1], Lynn King[4], Clair Francomano[4], Peter Freisinger[6], Stephanie Spranger[7], Bruno Marino[8], Bruno Dallapiccola[9], Michael Wright[1], Thomas Meitinger[3], Mihael H. Polymeropoulos[2] & Judith Goodship[1]

*These authors contributed equally to this work.

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[3,5]. 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).
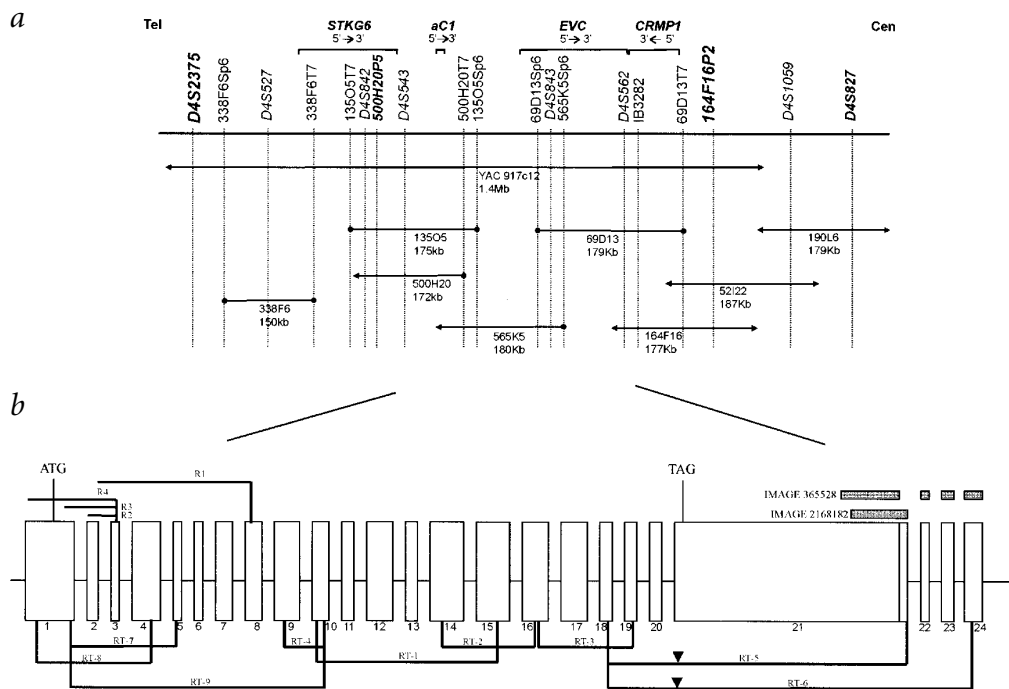
Fig. 1 Physical map of the region and genomic organization of *EVC*. **a**, Physical map of the region showing STSs, polymorphic markers (larger type) and gene transcripts. Clone ends are denoted by circles where there is a corresponding end clone and arrows where the end clones have not been identified. **b**, Genomic organization of *EVC* and cDNA contig assembly. RACE products are shown above the exons and RT–PCR products beneath. The arrow indicates the alternative splicing at the start of exon 21. ATG, translation initiation codon; TAG, stop codon. The two IMAGE clones are shown.

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

283

We analysed unfinished genomic sequence from BACs 135O5, 69D13, 164F16 and 565K5 (http://www-shgc.stanford.edu) using exon-prediction programmes. This revealed two previously identified genes, *aC1* 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. 1*a*). 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. 1*b*, 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. 1*b*). 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. 1*b*), 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[6], was a compound heterozygote with a missense mutation (S307P) 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[7]. 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 controls (194 chromosomes) tested. As the gene is not transcribed in lymphocytes, we have not yet
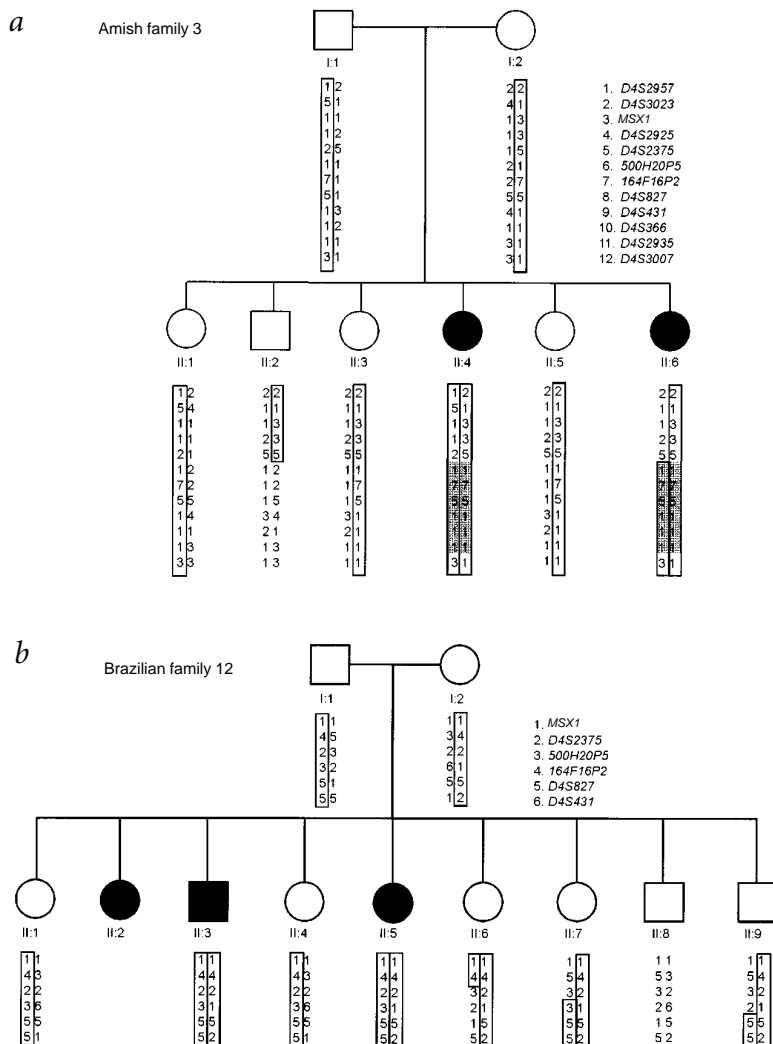
Fig. 2 Haplotype analysis in Amish family 3 and Brazilian family 12. The markers are ordered from telomere to centromere. The disease haplotypes are boxed and regions of homozygosity are shaded.

```
EVC    1  MARGGAACKSDARLLLGRDALRPAPA LLAPAVLLGAALGLGLGLWLG CRAGRQRTRHQKDDTQNLLKNLESNAQTPSETGS PSRRRKREV
          :   .  . :::: :::.:::.  :: . : .::::::..:::::::::::::.. :. ::: : .: .  .::  .:::.
Evc    1  MT-----CTKDARLQLGREALQAAPT LLVPAVLLGGVLGLGLGLWLG CRASHLRARLQKDDRKRLLGSSEPPAQSLRDTGSQAKARPRQR

EVC   91  QMSKDKEAVDECEPPSNSNITAFALKAKVIYPINQKFRPLADGSSNPSLHENLKQAV-----LPHQPVEASPSSSLGSLSQGEKDDCSSS
          . ..:.:. : :::  .:::::::::.:::::::::::::::::::::::.:::::::  ::   :::::.:::::::::. :.: :::
Evc   86  ETTRDEDAPEVCEPSLSGNITAFALKARVVYPINQKFRPLADGSSHPSLHENLTQAAAILPHLPHQPAEASPASSLGSLSQAGKEDGSSS

EVC  176  SSVHSATSDDRFLSRTFLRVNAFPEVLACESVDVDLCIYSLHLKDLLHLDTALRQEKHMMFIQIFKMCLLDLL PKKKSDDD ELYQKILSKQ
          ::..:. ::::..: .:::::..:::::::::::.:::  ::::::::.:::::::::.:::::.: :::: ::::: :::: :::.:
Evc  176  SSMRSTYSDDRILQCAFLRVGSFPEILACESVDIDLCVCSLHLKDLLQVDTALRQEKHLMFIQILKACLLDFF PKKKPDD LCQKVLSKQ

EVC  266  EKDLEELEKGLQVKLSNTEMSGAGDSEYITLADVEKKEREYSEQLIDNMEAFWKQMANIQHFLVDQFKCSSSKARQLMMTLTERMIAAEG
          :.:::::::.:..::.::::. ::. : .:::::::: ::::::: ::::::: ::: .  .: :.:::::::::..::::: ::::.:::
Evc  266  EHDLEELEKGLQARLANTEMLGTGDSGYVSLADVERKERELSEQLIDNMGAFWKQMESIQPTLMDQFKCSSSKARQFMMTLTGRMIVAEG

EVC  356  LLCDSQELQALDALERTMGRAHMAKVIEFLKLQVQEETRCRLAAISHGLELLAGEGKLSGRQKEELLTQQHKAFWQEAERFSREFVQRGK
          :: :::.:::.:.::::::::::: .:::.:..::. :.::::.:::: :::::. :::::::::::::::::: :::: :::::::: :::
Evc  356  LLHDSQDLHVLDTLERTMGRAHLARMVEFLRTQIQEETKCRLAAISRGLELLTVQGQLSGRQKEELLTQQHKAFWEEAERFGREFMQRGK

EVC  446  DLVTASLAHQVEGTAKLTLAQEEEQRSFLAEAQPTADPEKFLEAFHEVLERQRLMQCDLEEEENVRATEAVVALCQELYFSTVDTFQKFV
          :::  :::: :  : :..  :::.: :::::::::..:: :::::::::::: :::::::::  :::.::::::: :::::: .:::::::
Evc  446  DLVQASQARQAEAAAELTQTQEEERRSLLADSQLTSDPGEFLKAFHEVLERQRLTRSDQEGDEDTRITEAMAALCQELYCSTMGTFQKFV

EVC  536  DALFLQTLPGMTGLPPEECDYLRQEVQENAAWQLGKSNRFRRQQWKLFQELLEQDQQVWMEECALSSVLQTHLREDHEGTIRGVLGRLGG
          :.::::: .:.:  : :: :  :::.::: ::::::::::::::..:::.:::::::::.::.. .:::.:::::::::.::::::..::.
Evc  536  DSLFLKTLPEVTSLPVAECETLRQQVQEQAARQLGQADRFFRRQWGLLCDLLEQDKRVWLEEGTLSTVLQRQLRDHHESTIHGVLSRFSG

EVC  626  LTEESTRCV LQGHDLLLRSALRRLALRGNAL ATLTQMRLSGKKHLLQELREQRALEQGSSQCLDEHQWQLLRALEARVLEEASRLEEEAQ
          :.::::. : :::::.::: ::::::::::: :::::::::.:::::::..:: ::::::.::: :::.::::::::.:::::::.::::::
Evc  626  LSEESSRGI LQGHELLLCSALRRLALRGTTI TALAQMRLSGKKRLLQELHEQLALEQGVSPCLEEHQWQLLRALEARIQEEAARLEDEAQ

EVC  716  QTRLQLQQRLLAEAQEVGQLLQQHMECAIGQALLVHARNAATKSRAKDRDDFKRTLMEAAVESVYVTSAGVSRLVQAYYQQIGRIMEDHE
          :: :::.:::::::::::::::::::..:::::::::::::::::: :::::::::::::::::::: ::::::::::::::::::::::::
Evc  716  QTGLRLQQQLLAEAQEAGRLLQLHMERVIGQALLVHARNVASKGRTREKEDFKRTLVETVVESVYVTSTSVNRLVQAHYQAVGRKLLQAHE

EVC  806  ERKLQHLKTLQGERMENYKLRKKQELSNPSSGSRTAGGAHETSQAVHQRMLSQQKRFLAQFPVHQQMRLHAQQQQAGVMDLLEAQLETQL
          . ::::.:::: :::: :. :: :::::::::.:.::::.: :::::::::::::::.:::.  :::::::.::  :::.:::::::::::
Evc  806  EQLLQRLKTLQGERINAYKLWKKQEFSDPSLESQTADGTHGASQGVQQRMLSQQKRLLDQFTKHQQGRLNSQRQKAQELDQLQAQLETQL

EVC  896  QEAEQNFISELAALARVPLAESKLLPA KRGLLEKPLRTKRKKPL QERGDLGVPNNEDLASGDQTSGSLS-------------SKRLSQQ
          ::::::::::.::::::.:::::::: :   :::.: :.:::::: :::::: ::::: .:: ::: ::  . ::..   . ::: .:::.
Evc  896  QEAEQTFISELSTLARVPLPENKPFSNNRGLPEK PVRTKRKKPP PREREDLGTPNDDHLALADHTTGPLSTTYSASPPIRVHSGGRLDQQ

EVC  973  ESEAGDSGNSKKMLKRRSNL
          .:::::  ...:.:. :::
Evc  986  DSEAGDGESTSKILQKGSNL
```

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 (TTYSASPPIRVHS) 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 association with disease in a number of genes and alternate splicing has been shown for *CAT* (intron 4), *APC* (intron 9), *COL1A1* (introns 8, 14) and *COL2A1* (intron 20; ref. 8; http://www.uwcm.ac.ukuwcm/mg/hgmd0.html). We have not found a mutation in the Brazilian pedigree or in two consanguineous families in which the affected individuals were homozygous across the region. We found many polymorphisms, and those that change the amino acid sequence are listed (Table 1).

We have undertaken a limited study to determine *EVC* expression in human embryonic tissue using *in situ* hybridization. At Carnegie stages 19 and 21, we detected low levels of *EVC* expres-

sion in developing bone, heart, kidney and lung. In bone, *EVC* was expressed in the developing vertebral bodies, ribs and both upper and lower limbs. The expression was higher in the distal limb compared with the proximal limb. In addition, we detected *EVC* expression in the branching epithelium and surrounding mesenchyme of the lung, metanephros and atrial and ventricular myocardium, including both atrial and interventricular septa.

We have identified a new gene that is mutated in patients with EvC syndrome. The predicted protein has no homology to known proteins other than in a short region that may be a leucine zipper. We have seen alternate splicing in the 3′ UTR and overlap between *EVC* and *CRMP1*. In addition, a 2.6-kb fetal brain cDNA has been described, one end of which is identical to *EVC* exon 21 and the other, complementary to the *CRMP1* 3′ UTR (ref. 9). From the size of this clone we deduce that there is another 3′ splice variant. Alternate splicing in the 3′ UTR has been reported in other genes and may lead to differences in mRNA stability, thus affecting expression levels[10]. There are also examples of other genes transcribed from opposite DNA strands overlapping at the 3′ end[11], and *in vivo* interaction between two mRNAs may influence their stability and thus expression[12]. In this instance, the overlap is not confined to the 3′ UTR but includes *CRMP1* coding sequence. As *CRMP1* is expressed in the developing limb and ectoderm, it is possible that this 3′ complementarity has functional significance.

Methods

Patients. We obtained samples from the Amish and Brazilian pedigrees[3], a child with EvC whose father had features of Weyers acrodental dysostosis[6], an individual from the island of Barra[13], and a father and daughter with an EvC-like phenotype[7]. 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 135O5 and 69D13 in the Stanford map. We designed STSs by directly sequencing the ends of clones 135O5 and 69D13 using Sp6 and T7 primers and used them to identify BAC clone 565K5, which has now been integrated into the Stanford contig. We isolated BAC 338F6 by screening the Research Genetics library using *D4S527*. We developed the sequence-based map using the CrossMatch sequence alignment program (http://www.phrap.org).

Microsatellite analysis. We carried out saturation genotyping using repeat-containing microsatellite markers located in the EvC region. The position of these microsatellites is shown on the Stanford chromosome 4 YAC map (http://www-shgc.stanford.edu/Mapping/phys_map/Chr4YAC.html). We isolated markers 500H20P5 and 164F16P2 by analysing the BAC sequence for microsatellites.

EVC cDNA cloning. We performed 5′-RACE using marathon ready human brain cDNA (Clontech) or human fetal kidney poly(A)+ RNA, age 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.
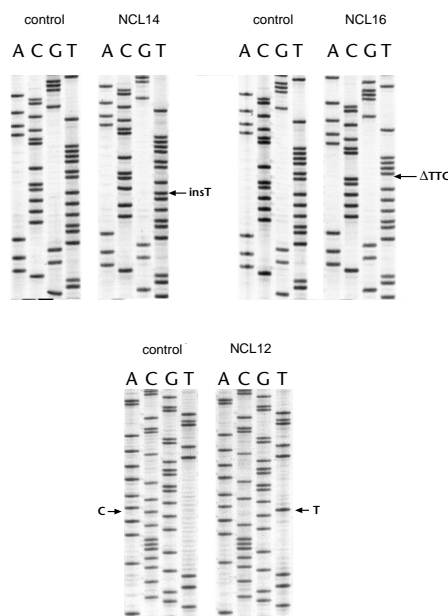
Fig. 4 DNA sequence showing homozygous mutations in three patients. The sequence in NCL14 and NCL16 are from the reverse strand. NCL 14 has a single-nucleotide insertion. NCL 16 has a single-codon deletion. Patient NCL12 has a C→T transversion, leading to Q879X.
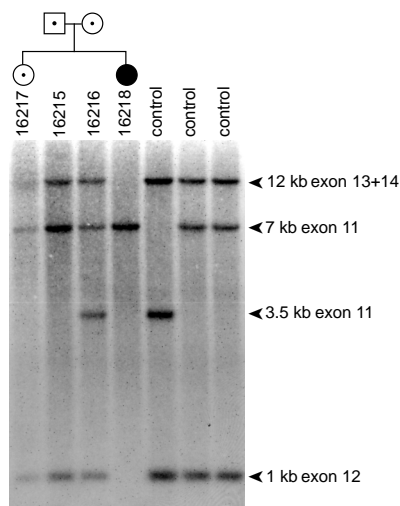
Fig. 5 Deletion analysis of family EVC015. A dosage blot of *Hind*III digests of genomic DNA was hybridized with an RT–PCR product comprising exons 11–14. Autoradiography revealed the absence of a 1-kb fragment corresponding to exon 12 and the absence of a 12-kb fragment corresponding to exons 13 and 14 from patient 16218. The 7-kb fragment corresponding to exon 11 is present in this patient. Individual 16216 is heterozygous for an exon 11 RFLP (3.5 kb/7 kb). Carrier status of individuals 16215, 16216 and 16217 is demonstrated by half dosage of both 1-kb and 12-kb fragments compared with controls.

Computer analysis. We used the integrated NIX programme (http://www.hgmp.mrc.ac.uk) or the integrated RUMMAGE programme (http://gen100.imb-jena.de/rummage) for DNA analysis and PSORTII for protein sequence analysis.

Northern-blot analysis. We amplified RT–PCR product RT1 from human brain cDNA using primers 5′–TCCAGCGAGGCAAAGACC–3′ and 5′–GCCGTGTCTGCTGTGCTTCCTC–3′, and TA subcloned RT1 in plasmid pCR 2.1 (Invitrogen). We purified RT1 insert from low-melting-point agarose gel, random primer labelled with [^{32}P] α–dCTP and used it as a probe on a human fetal multiple-tissue northern blot (Clontech, MTN Fetal II) as well as on a human adult northern blot (Clontech MTN-I). After removal of the previous probe, we probed the human fetal II northern blot again with the *Not*I-*Eco*RI 609-bp insert of the IMAGE clone 2168182. We performed all hybridizations following the manufacturer's protocols and checked lane loading in the northern blot by hybridizing with the human β-actin control probe supplied by the manufacturer.

Tissue *in situ* hybridization. We obtained ethical permission for the collection and use of human embryos. We collected the embryos following either surgical or medically induced termination of pregnancy and determined their developmental stage by stereomicroscopy[14]. They were then fixed in paraformaldehyde and embedded in paraffin. We cloned the RT–PCR product RT-1 in both orientations in pCR2.1 (Invitrogen) and linearized it with *Bam*HI to provide sense and antisense riboprobes. We carried out RNA tissue *in situ* using a standard technique[15].

Table 1 • *EVC* mutations and polymorphisms

Patient number	Sequence change	Exon	Protein effect
mutations			
16086 (compound heterozygous)	734delT	6	
	1018C→T	8	R340X
ncl-16 (homozygous)	904–906delAAG	7	ΔK302v
ncl-14 (homozygous)	910–911InsA	7	
15219 (compound heterozygous)	919T→C	7	S307P
	2456delG	17	
I-1 (heterozygous)	1328G→A	10	R443Q
16218 (homozygous)	deletion	12-21	
Amish (homozygous)	IVS13+5G→T	intron13	
ncl-12 (homozygous)	2635C→T	18	Q879X
polymorphisms			
	221A→C	2	Q74P
	772T→C	6	Y258H
	1207G→A	9	G403S
	1346C→A	10	T449K
	1727G→A	12	R576Q
	2279G→A	15	R760Q
	2858A→G	20	D953G

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 normal 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 exon-13–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*, AF216184, AF216185; *Evc*, AJ250841; human serine/threonine protein kinase, AJ250839; *Mus musculus* serine/threonine protein kinase, AJ250840; microsatellites, AF195414, AF195415, AF195416.

1. Ellis, R.W.B. & van Creveld, S. A syndrome characterised by ectodermal dysplasia, polydactyly, chondrodysplasia and congenital morbus cordis: report of three cases. *Arch. Dis. Child.* **15**, 65–84 (1940).
2. McKusick, V.A., Egeland, J.A., Eldridge, R. & Krusen, D.E. Dwarfism in the Amish I. The Ellis-van Creveld Syndrome. *Bull. Johns Hopkins Hosp.* **115**, 306–336 (1964).
3. Polymeropoulos, M.H. *et al.* The gene for the Ellis-van Creveld syndrome is located on chromosome 4p16. *Genomics* **35**, 1–5 (1996).
4. Howard, T.D. *et al.* Autosomal dominant postaxial polydactyly, nail dystrophy, and dental abnormalities map to chromosome 4p16, in the region containing the Ellis-van Creveld syndrome locus. *Am. J. Hum. Genet.* **61**, 1405–1412 (1997).
5. Ide, S.E., Ortiz de Luna, R.I., Francomano, C.A. & Polymeropoulos, M.H. Exclusion of the MSX1 homeobox gene as the gene for the Ellis van Creveld syndrome in the Amish. *Hum. Genet.* **98**, 572–575 (1996).
6. Spranger, S. & Tariverdian, G. Symptomatic heterozygosity in the Ellis-van Creveld syndrome? *Clin. Genet.* **47**, 217–220 (1995).
7. Digilio, M.C., Marino, B., Giannotti, A. & Dallapiccola, B. Single atrium, atrioventricular canal/postaxial hexodactyly indicating Ellis-van Creveld syndrome. *Hum. Genet.* **96**, 251–253 (1996).
8. Krawczak, M. & Cooper, D.N. The human gene mutation database. *Trends Genet.* **13**, 121–122 (1997).
9. Ishida, Y. *et al.* Isolation and characterisation of 21 novel expressed DNA sequences from the distal region of human chromosome 4p. *Genomics* **22**, 302–312 (1994).
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).
11. Petrukhin, K. *et al.* Identification of the gene responsible for Best macular dystrophy. *Nature Genet.* **3**, 241–247 (1998).
12. Ubeda, M., Schmitt-Ney, M., Ferrer, J. & Habener, J.F. CHOP/GADD153 and the methionyl-tRNA synthetase (MetRS) genes overlap in a conserved region that controls mRNA stability. *Biochem. Biophys. Res. Commun.* **19**, 31–38 (1999).
13. Hill, R.D. Two cases of Ellis-van Creveld syndrome in a small island population. *J. Med. Genet.* **14**, 33–36 (1977).
14. Bullen, P. & Wilson, D.I. The Carnegie staging of human embryos: a practical guide. in *Molecular Genetics of Early Human Development* (eds Strachan, T., Lindsay, S. & Wilson, D.I.) 27–35 (Bios Scientific, Oxford, 1997).
15. Hanley, N.A. *et al.* Expression of steroidogenic factor 1 and Wilms' tumour 1 during early human gonadal development and sex determination. *Mech. Dev.* **87**, 175–180 (1999).