Thumbs up

Limb mutants elucidating mechanisms of development.

Marijke Jeltje van Baren

Thumbs up

Limb mutants elucidating mechanisms of development.

Ledemaat mutanten verhelderen ontwikkelingsmechanismen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
Rector Magnificus

Prof. dr. ir. J. H. van Bemmel

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op

woensdag 6 maart 2002 om 13.45 uur

door

Marijke Jeltje van Baren

Geboren te Amersfoort

Promotiecommissie

Promotor   Prof. dr. B.A. Oostra

Overige leden   Prof. dr. Dzierzak
                            Prof. dr. Hovius
                            Prof. dr. R. Zeller.

Co-promotor   Dr. P. Heutink

Maar het meest bevredigende antwoord kan nimmer wedijveren

met de meest fascinerende vraag

Table of Contents

7

.

Chapter 1

Introduction into the genetics of limb formation

The genetics of limb formation

Limb development

The development of a single cell into a complete organism has intrigued biologists for many decades. How is the three-dimensional form of the organism established? How does a cell know if and when it should become a muscle cell or a skin cell, and how does it achieve that? We are only beginning to understand how development works, how tight timing and localisation of gene expression organises the embryo and how similar pathways and mechanisms are used over and over again to form structures as different as hands and kidneys.

One of the key components of development is the commitment of cells to different pathways. Initially, cells are 'totipotent': they are capable of differentiating into any cell type that makes up the organism (Driesch 1892). After the initial divisions, an embryo becomes 'patterned': it acquires a body plan in which each part is destined to become a specific part of the organism (Wolpert *et al*. 1969). To achieve this, cells choose a pathway that leads to differentiation, and in doing this they lose their potential to initiate another developmental pathway. Initially, the animal body plan is laid out in broad strokes, in which the body axes are established (Johnson and Tabin 1997). The anterioposterior axis runs from head to tail, the dorso-ventral axis from back to front, and the proximo-distal axis runs perpendicular to these two, from close to the trunk to far from it (fig. 1). In later stages, these regions are refined and some become semi-autonomous, after which a new process of pattern formation begins. One of these semi-autonomous regions is the developing vertebrate limb, which has been used extensively as a model for pattern formation (Johnson and Tabin 1997). The undifferentiated vertebrate limb is a self organising system: if it is transplanted to a favourable ectopic location, the limb bud is capable of developing into a morphologically normal limb (Harrison 1918), and the anterioposterior polarity of the transplanted limb is determined by the graft rather than the host environment. The limb is accessible and big and therefore easy to manipulate. Surgical manipulations of the chick limb have shown a great deal about patterning and
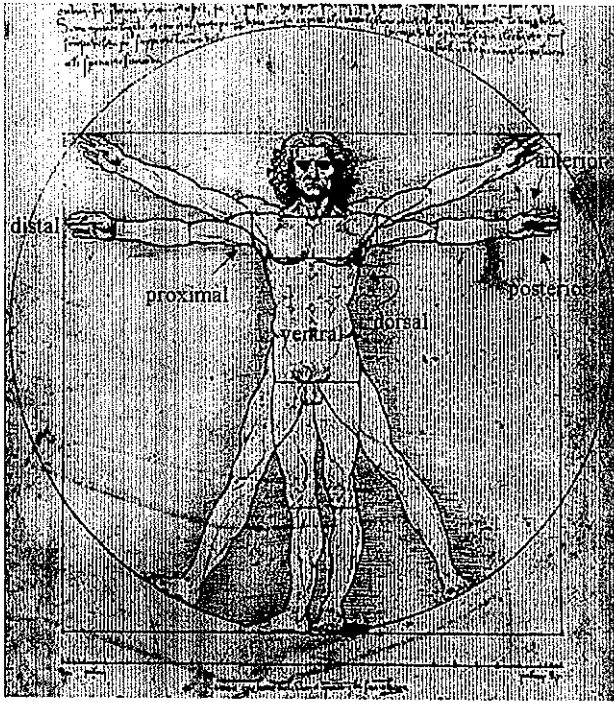
Figure 1. The body axes.

studies in mouse limb mutants have revealed many of the genetic components involved in limb formation (Johnson and Tabin 1997). These genes are also involved in the formation of other structures, such as the kidneys and gut, and in these processes, they often act in a very similar way (Burrow 2000). Many developmental genes continue to function in regulation of cell growth and differentiation after embryogenesis: for instance, mutations of some members of the hedgehog pathway are associated with human cancer as well as birth defects (Saldanha 2001).

Many of the genes involved in mammalian limb development have homologs in *Drosophila*. The development of the *Drosophila* wing has been studied extensively, especially in the wing disc, the primordium of the wing present in the larvae (for review see Gilbert 2000). The reverse is also true: many of the genes that function in invertebrate wing development have their counterparts in vertebrate limb development, showing that *'insects are not odd beasts with quirky developmental strategies of their own'* (Lawrence 1990). Indeed, the proteins involved in wing patterning often interact in a way that is

12

remarkably similar to the patterning of the vertebrate limb (Serrano and O'Farrell 1997). Because insights in many limb developmental pathways originate from studies in *Drosophila*, the development of the *Drosophila* wing will be discussed briefly later in this introduction.

The French flag model

An important model on patterning was proposed by Lewis Wolpert (Wolpert *et al.* 1969). During embryonic development, cells must generate a complex pattern from a field of equivalent precursor cells. To do this, they must 'know where they are' and then interpret this information in their development. To simplify this problem, Wolpert formulated the 'French flag problem': what mechanism would ensure that a line of totipotent cells, no matter how long, would always have a French flag pattern: one third blue, one third white and one third red? In order to develop in the proper way, cells must receive a signal which specifies their position within the system (fig. 2). One way this can be established is when a chemical is produced on one side of the field, and a concentration gradient through the field is set up. Cells would need a system to detect the concentration of the chemical, or 'morphogen', and differentiate accordingly: for instance, 'blue' at low concentrations, 'white' at intermediate, and 'red' at high concentrations of the chemical. In this system, cells must recognise threshold values of the morphogen and it has been shown that some
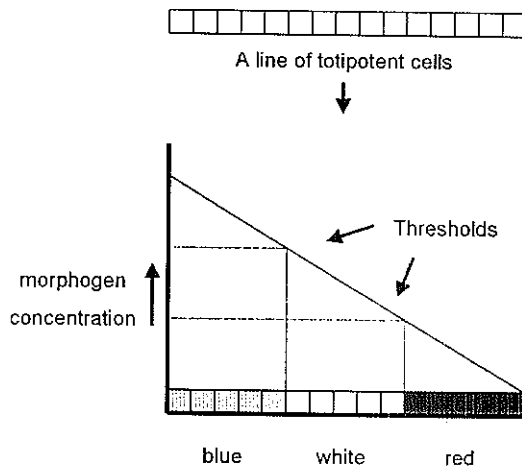
Figure 2. Wolpert's French Flag model. Cells detect the concentration of a morphogen and behave accordingly.

cells are capable of doing this. For example, in the *Drosophila* wing, the morphogen Decapentaplegic (DPP) is secreted by a stripe of cells in the middle of the developing wing. It induces transcription of genes such as *spalt* and *optomotor-blind* (*omb*) in neighbouring cells (de Celis *et al.* 1996, Grimm and Pflugfelder 1996). *Omb* is expressed both in the vicinity of the DPP source and further away, while *spalt* is expressed close to the DPP source only, indicating that *omb* has a lower threshold for DPP (Nellen *et al.* 1996).

The morphogen was long thought to be a diffusible component (Wolpert 1996). It could be taken up by the cells or destroyed outside the cells, which would create the gradient. However, such a linear decrease of morphogen concentration is not the only way in which positional information can be generated. Another means of recording positional information is by measuring the time spent in a region. Dividing cells push themselves away from the signalling source, as is the case in the tip of the developing limb (see below).

Cells can only interpret the positional information according to their own developmental program. Cells that have committed themselves to become part of a leg will interpret signals as if they are part of a leg, even when they are moved to a different location. In one experiment, mesoderm from the developing chick leg was grafted to the tip of a wing bud. The grafted tissue was derived from the future thigh, but after the grafting it received signals for distal positioning. In response, toes developed, not wing digits. This shows that the cells committed to leg development interpret information according to a 'leg development' program (Krabbenhoft and Fallon 1989)

Drosophila wing development

The *Drosophila* wing arises from a layer of epidermis on the side of the embryo (Gilbert 2000). This layer evaginates to become a disc, which later grows out in a telescopic way, so that the dorsal and ventral sides of the wing disc become the dorsal and ventral sides of the wing (fig. 3). The wing disc is divided into an anterior (A) and a posterior compartment (P) which are separated by the compartment boundary. This boundary is not visible as a structure, but it can be visualised by adding a tracer to cells from one compartment and following this through the divisions: cells derived from progenitors in

14
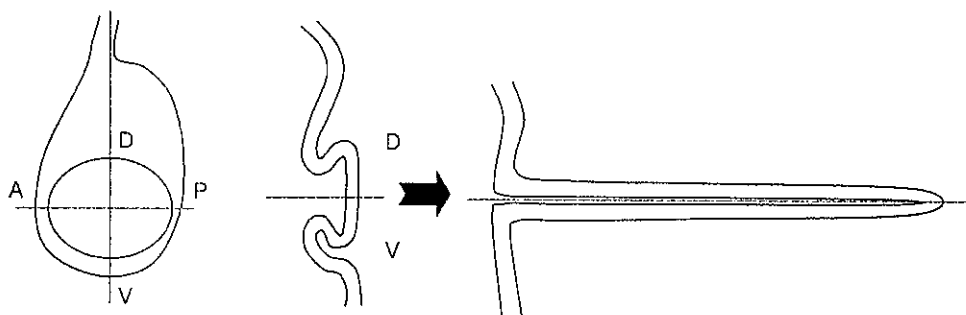
Figure 3. The developing wing of Drosophila. A: anterior, P: posterior, D: dorsal, V: ventral. Antero-posterior and dorso-ventral boundaries are indicated by dotted lines.

either compartment never cross the boundary (Garcia-Bellido *et al.* 1973). The wing disc is highly patterned, especially by genes that are expressed in one compartment, but exert their effects in the other.

The first gene that is thought to be expressed asymmetrically is *engrailed* (*en*) (Kornberg 1981). It is expressed in the P compartment, and this expression pattern is already present before the wing disc invaginates. *En* promotes the expression of the secreted signalling protein Hedgehog (HH) in the P compartment (Heemskerk and DiNardo 1994). At the same time, the HH receptor Patched (PTC) is expressed in the A compartment, where it interacts with another transmembrane protein, Smoothened (SMO) (Alcedo *et al.* 1996, Phillips *et al.* 1990). A third gene expressed in the A compartment is *cubitus interruptus* (*ci*), which encodes a transcription factor (Alexandre *et al.* 1996, Dominguez *et al.* 1996). The inactive CI is bound to microtubuli (Robbins *et al.* 1997). When no HH protein is bound to PTC, cytoplasmic CI protein is cleaved, and the short form of CI transports to the nucleus where it represses transcription of HH target genes (fig. 4a) (Aza-Blanc *et al.* 1997). When HH from the P compartment reaches the PTC receptors in the A compartment, SMO is released and starts a signal transduction cascade which ultimately releases CI from the microtubuli and prevents its proteolysis. Instead, CI is modified into a short lived transcriptional activator of HH target genes, one of which is *ptc* (Aza-Blanc and Kornberg 1999) (fig. 4b). This means that HH promotes the expression of its own receptor. It is thought that this enhanced expression restricts the range of HH activity, because most of the protein will be bound by nearby receptors, leaving only a fraction to move to cells further away (Chen and Struhl 1996).

15
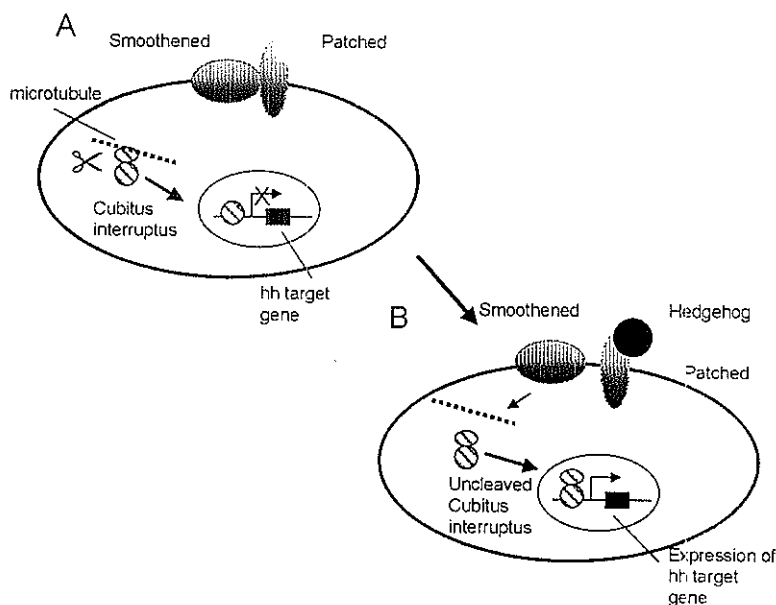
Figure 4. Hh signalling in the P compartment of the wing disc. **A.** In the absence of hh, Ci is cleaved and the short form transduces to the nucleus, where it represses hh target genes. **B.** When hh is bound, Smoothened is released from the Patched receptor, full length Ci is released from the microtubuli and functions as a transcriptional activator.

Another gene that is transcribed in the boundary region in response to HH signalling is *decapentaplegic* (*dpp*) (Basler and Struhl 1994). DPP is a good candidate for the morphogen that specifies anterioposterior patterning: a gradient of DPP forms throughout the wing with the highest concentration along the A/P boundary (Lecuit *et al.* 1996). This means that the concentration gradient in both the A and the P compartments is similar, but it is interpreted differently by the cells in each compartment. Interestingly, *dpp* is expressed at uniform intensity within its domain (St Johnston and Gelbart 1987), therefore the activity gradient must be formed by posttranscriptional modulation of DPP distribution or signalling capability (Ferguson and Anderson 1992).

The A/P boundary is not the only compartment boundary specified in the developing wing. There is also a boundary between the dorsal and ventral compartments, which is established in a slightly different way. This D/V boundary is set up with the Notch (N) receptor, which is expressed in the wing disc (Shellenbarger and Mohler 1978). The Notch receptor has two ligands, Delta and Serrate (Fehon *et al.* 1990, Rebay *et al.*

16

1991). Cells in the ventral ectoderm express *Delta (Dl)*, while cells in the dorsal ectoderm express *Serrate (Ser)* and *fringe (fng)* (Doherty *et al.* 1996, Irvine and Wieschaus 1994, Speicher *et al.* 1994). FNG modulates N in such a way that it becomes sensitive to DL and refractory to SER (Fleming *et al.* 1997, Johnston *et al.* 1997, Klein and Arias 1998, Panin *et al.* 1997). The result is that SER signals to ventral cells and DL signals to dorsal cells, so that N is activated symmetrically. Upon activation, the intracellular domain of N is released from the membrane and translocates to the nucleus where it regulates the transcription of downstream target genes (Blaumueller *et al.* 1997, Struhl and Adachi 1998). One of these is *wingless (wg)*, which is secreted and patterns the D and V compartments (Rulifson and Blair 1995).

All the proteins mentioned here have one or several counterparts in vertebrates (table 1). In most cases, a single gene in *Drosophila* corresponds to a family of related homologs in vertebrates, and many of these interact in a similar way. However, the *Drosophila* wing disc arises from ectoderm only, and signals do not need to travel from one cell lineage to another. The vertebrate limb has a more complex organisation.

Hox genes

How does a cell respond to the positional information? Candidates for recording these positional values are the Hox genes, which are expressed in overlapping patterns in many developing tissues. Mutations in these genes lead to 'homeotic transformations' in which one structure replaces another. This was first reported in 1915, by Calvin Bridges, who studied a fly that had undergone a mutation to produce an extra set of wings in place of a pair of balancers (small, winglike appendages). Bridges called the mutation 'homeotic' because it changed one body part into another. This transformation- along with another homeotic mutation in a gene called *Antennapedia* – that caused legs to sprout from a fly's head where antennae would normally be – was interesting because it suggested the existence of control genes responsible for directing the development of large parts of the body (reviewed in (Lewis 1995).

In vertebrates, four *Hox* gene clusters are found, named *Hoxa*, *Hoxb*, *Hoxc* and *Hoxd*. The order of these genes in the genome corresponds to their timing and location of expression: more 3' located genes are expressed earlier in development than more 5'

Table 1. Mammalian homologs of *Drosophila* genes.

Drosophila gene	Mammalian homolog(s)
engrailed (en)	Engrailed homolog 1 (En1)
	Engrailed homolog 2 (En-2)
hedgehog (hh)	Sonic hedgehog (Shh)
	Indian hedgehog (Ihh)
	Desert hedgehog (Dhh)
patched (ptc)	Patched homolog (ptc)
	Patched homolog 2 (ptc2)
smoothened (smo)	Smoothened homolog (Smoh)
cubitus interruptus (ci)	GLI-Kruppel family member GLI1 (Gli1)
	GLI-Kruppel family member GLI2 (Gli2)
	GLI-Kruppel family member GLI3 (Gli3)
decapentaplegic (dpp)	Bone morphogenetic protein 2 (Bmp2)
	Bone morphogenetic protein 4 (Bmp4)
notch	Notch homolog 1 (Notch1)
	Notch homolog 2 (Notch2)
	Notch homolog 3 (Notch3)
	Notch homolog 4 (Notch4)
delta	Delta-like 1 homolog (Dll1)
	Delta-like 2 homolog (Dll2)
	Delta-like 3 homolog (Dll3)
	Delta-like 4 homolog (Dll4)
serrate (ser)	Jagged 1 (Jag1)
	Jagged 2 (Jag2)
fringe (fng)	Manic fringe (Mfng)
	Lunatic fringe (Lfng)
	Radical fringe (Rfng)
wingless (wg)	Wingless-type MMTV integration site family, members (currently 20) (Wnt1, Wnt2, Wnt 2b etc.)

located genes. This phenomenon is called temporal colinearity. There are 13 homology groups of *Hox* genes, but not every cluster contains a gene from every group. For example, *Hoxa11* is homologous to *Hoxc11* and *Hoxd11*, but there is no *Hoxb11*. In total, 39 Hox genes are known.

In the developing limb, *Hoxa* and *Hoxd* genes are expressed (Dolle *et al.* 1991, Dolle *et al.* 1989, Haack and Gruss 1993, Izpisua-Belmonte *et al.* 1991). The *Hoxd* cluster is expressed in overlapping patterns that run from anterior to posterior (fig. 5a). The transcript domain of a gene is contained within the domain of the gene located 3', in such a way that *Hoxd9* is expressed almost throughout the limb bud while *Hoxd13* is expressed only in the posterior part. The *Hoxa* cluster is expressed in a proximo-distal way, with *Hoxa9* expression throughout the limb and *Hoxa13* in the distal part only (fig. 5b).

Hox genes encode transcription factors. The structure of the genes differs considerably, but all Hox proteins contain a conserved motif, the homeodomain, which is a stretch of 60 amino acids that binds to specific DNA motifs in promoters of target genes. It is thought that target genes can be activated by several *Hox* genes, but that the more 5' *Hox* gene exerts a dominant effect (Duboule 1995). The homeodomain is also found in many other genes which are not located in the *Hox* clusters.
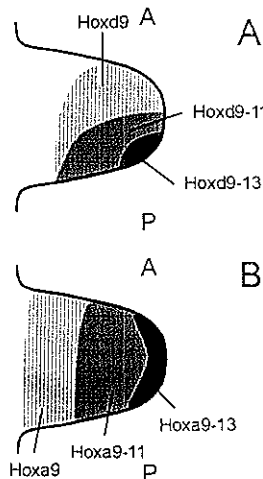
Figure 5. Hox clusters are expressed in overlapping patterns. A is anterior, P is posterior. **A.** The Hoxd cluster overlaps in an antero-posterior pattern. **B.** The Hoxa cluster overlaps in a proximo-distal pattern.
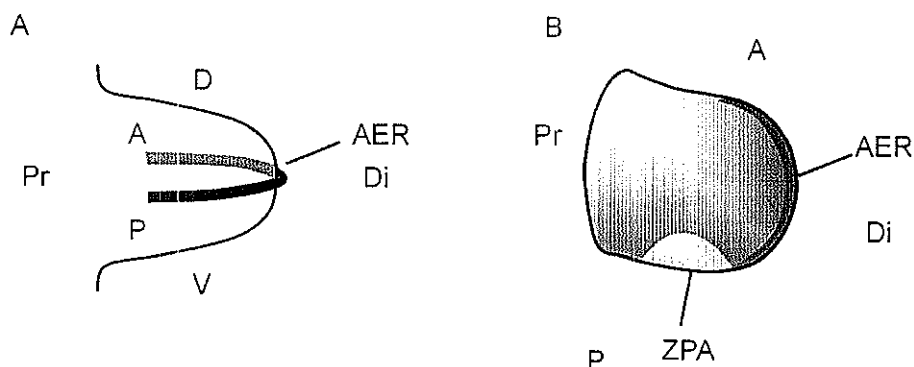
The developing limb

Formation of the limb starts with proliferation of cells of the lateral plate mesoderm at specific places along the axis of the embryo (Searls and Janners 1971). This proliferation occurs at the end of a phase of rapid proliferation of the rest of the mesoderm, and soon the proliferating cells form a bulge under the ectoderm, the limb bud. Later, blood vessels penetrate the limb bud and neural crest cells and cells from the somitic mesoderm migrate into the limb bud to form nerves and muscles (Gilbert 2000). The skeleton is derived from lateral plate mesoderm, whereas the ectoderm develops into skin (Hall and Miyake 2000). Throughout limb development, cells from these different layers signal to each other, and this communication is necessary for proper limb outgrowth. This interaction is mediated by signalling centres.

Signalling centres: the proximo-distal axis

Cells at the tip of the limb bud form the apical ectodermal ridge (AER), a thin, tightly packed band of epithelial cells which signals to the underlying mesoderm to promote outgrowth of the limb (Saunders 1948). The AER runs along the boundary between the dorsal and ventral side of the limb (fig. 6a) and it specifies the proximo-distal outgrowth. When the AER is removed from chick limbs, outgrowth stops, and the result is a truncated limb (Saunders 1948). Removing the AER early during development leads to severe truncation: in the chick wing only the humerus is formed. If the AER is removed later, the radius and ulna do form, but digits are absent (Summerbell 1974). The AER keeps the underlying mesenchymal cells in an undifferentiated and proliferative state, the progress zone. As outgrowth continues, cells at the proximal side leave this zone and differentiate. By this time, their fate is specified: cells that leave the zone first, condense to form the humerus, whereas cells that leave the zone later, become the radius and ulna, and lastly the digits (Summerbell *et al.* 1973).

In the AER, several *Fibroblast Growth Factors* (*Fgfs*) are expressed. *Fgf-8* and *Fgf-2* are expressed throughout the AER and *Fgf-4* is expressed in the posterior two thirds of the ridge (Dono and Zeller 1994, Niswander and Martin 1992, Savage *et al.* 1993, Suzuki *et al.* 1992, Vogel *et al.* 1996). *Fgf-8* is also expressed prior to the formation of the AER, in a broad stripe of cells along the distal end of the limb. When the AER has

formed, *Fgf-8* expression is restricted to the ridge (Crossley *et al.* 1996, Heikinheimo *et al.* 1994, Ohuchi *et al.* 1994, Vogt and Leder 1996). When removal of the AER is followed by application of any of these *Fgfs*, outgrowth is restored (Crossley *et al.* 1996, Fallon *et al.* 1994, Niswander *et al.* 1993). Thus, *Fgfs* are sufficient for limb outgrowth.

Figure 6. The developing limb bud. **A** .The AER runs from anterior (A) to posterior (P) on the boundary between the dorsal (D) and ventral (V) side of the limb bud. Pr is proximal, Di is distal. **B.** The zone of polarising activity (ZPA) is located in the mesenchyme at the posterior side of the limb bud.

The antero-posterior axis

The antero-posterior (A/P) axis in the developing limb is established by another signalling centre, the zone of polarising activity (ZPA), located at the posterior side of the limb bud (fig. 6b). When cells from this side of a chick limb were transplanted to the opposite side of another bud, an entire set of additional digits occurred, in a mirror image of the normal set (Saunders and Gasseling 1968). This outcome is expected when a morphogen is involved in setting up the digit pattern, a hypothesis supported by the finding that when the ZPA was separated from the rest of the limb by an impenetrable barrier, the anterior structures did not form (Summerbell 1979). Subsequently it was found that application of retinoic acid could substitute for the ZPA, inducing mirror image duplications when applied to the anterior side of the limb bud (Tickle *et al.* 1982). However, the ZPA does not contain retinoic acid in concentrations high enough to activate the patterning genes (Noji *et al.* 1991). ZPA cells express *Sonic hedgehog (Shh)*, one of three vertebrate

21

homologs of *Drosophila hedgehog* (Riddle *et al.* 1993). Application of *Shh* to the anterior margin of the limb bud, either by grafting of *Shh* expressing cells or by implantation of beads soaked in Shh protein, led to mirror image duplications (Chang *et al.* 1994, Lopez-Martinez *et al.* 1995, Riddle *et al.* 1993). This showed that Shh has polarising activity and that it can take over the ZPA function. Notably, high concentrations of retinoic acid induce *Shh* expression, which explains the polarising results of retinoic acid (Johnson *et al.* 1994). Mutations that cause loss of Shh function result in severe truncations of the limb in mice, further suggesting that Shh is crucial for limb outgrowth (Chiang *et al.* 1996).

It has been proposed that *Shh* expression is maintained by a positive feedback loop with *Fgf-4* in the AER (Laufer *et al.* 1994, Niswander *et al.* 1994). This postulation was made after the finding that ectopic expression of *Shh* in the chick leads to induction and maintenance of *Fgf-4* expression, and application of Fgf-4 after removal of the AER maintains *Shh* expression. Furthermore, the mouse *limb deformity* mutant, which fails to maintain *Shh* expression in the limb, does not express *Fgf-4* in the AER (Haramis *et al.* 1995, Niswander *et al.* 1994). A second positive feedback loop was proposed between *Fgf-8* in the AER and *Fgf-10* in the limb mesenchyme. *Fgf-10* is expressed in the distal limb mesenchyme of the progress zone and is essential for limb outgrowth (Sekine *et al.* 1999). Removal of the AER leads to downregulation of *Fgf-10* expression, and subsequent addition of *Fgf-8* induces *Fgf-10*. In addition, *Fgf-10* deficient mice do not express *Fgf-8* in the ectoderm and do not form an AER (Ohuchi *et al.* 1997). However, recent evidence suggests that both these feedback loops are not necessary for limb formation: in the *dominant hemimelia* mouse mutant, *Fgf-8* and *Fgf-4* expression are shifted anteriorly, so that no expression of *Fgf-4* is present close to *Shh*, and *Fgf-8* expression is lower throughout the AER. Neither *Shh* nor *Fgf-10* expression is reduced in this mutant, indicating that the feedback loops are not necessary for limb outgrowth (Lettice *et al.* 1999).

The dorso-ventral axis
As described above, the AER forms at the border between the dorsal and ventral ectoderm. On the dorsal side of this border, *Radical fringe* (*Rfng*) is expressed, a secreted protein which is homologous to *Drosophila fringe* (Johnston *et al.* 1997). On the ventral

side, the homeobox containing transcription factor *Engrailed 1* (*En1*) is expressed (Davis *et al.* 1991, Davis and Joyner 1988, Gardner and Barald 1992). *En1* expression restricts *Rfng* expression, so that a sharp boundary is maintained (Logan *et al.* 1997). The AER forms at this interface between *Rfng* and *En1* expressing cells, and ectopic expression of *Rfng* leads to the formation of new AERs at the boundaries of *En1* and *Rfng* expressing cells (Laufer *et al.* 1997, Rodriguez-Esteban *et al.* 1997). En1 also restricts expression of *Wnt7a*, a *wg* homolog that is expressed in the dorsal ectoderm (Dealy *et al.* 1993, Parr *et al.* 1993). Wnt7a induces the LIM-homeodomain containing transcription factor *Lmx1* in the dorsal mesoderm (Riddle *et al.* 1995, Vogel *et al.* 1995). Expression of *Wnt7a* or *Lmx1* on the ventral side of the developing limb bud leads to dorsalisation of the mesoderm, and loss of *Wnt7a* results in limb buds with a double ventral phenotype (Parr and McMahon 1995, Riddle *et al.* 1995). In these mutants, AERs do form at the proper position, indicating that *Wnt7a* is not necessary for the formation of the AER.

A recently published model suggests that the AER is formed by recruitment of cells from the ventral ectoderm (Kimmel *et al.* 2000) (fig. 7). These pre-AER cells assemble along a fixed dorsal border, which functions as a reference point. En1 then regulates the formation of a dorso-ventral (D/V) border, parallel to the dorsal border, and cells on both sides of this D/V border are pulled towards it, so that the AER rises up and forms a triangular ridge. If the D/V border is lost, or *En1* is misexpressed on the dorsal side of the AER, the AER flattens. Furthermore, mice that lack *En1* have a flattened AER that spreads into ectopic ventral locations (Loomis *et al.* 1996). After formation of the AER, a ventral border is also established, separating the AER tissue from the ventral
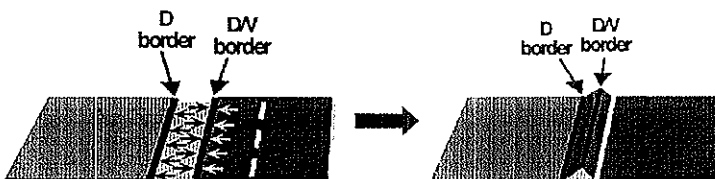
Figure 7. A model for AER formation. At first, cells from the ventral side of the limb bud assemble along a dorsal border (D). Later, cells are pulled towards a dorso-ventral border (D/V) so that a triangular ridge forms (Adapted from Kimmel *et al.* 2000).

ectoderm. When *En1* is ectopically expressed, a dorsally shifted AER is induced adjacent to the new *En1* boundary (Kimmel *et al*. 2000).

Specification of limb identity

Forelimbs and hindlimbs are serially homologous structures that use the same set of signalling molecules to control their outgrowth and patterning. But how are the differences between both structures generated? As described above, *Fgf-10* is expressed in the mesenchyme in a feedback loop with *Fgf-8*. Both genes are expressed in the embryonic flank before initiation of limb outgrowth. It is thought that *Fgf-8* expression in the intermediate mesoderm restricts expression of *Fgf-10* to the lateral plate mesoderm of the presumptive limb regions, after which *Fgf-10* induces *Fgf-8* expression in the surface ectoderm, and limb outgrowth begins (Kawakami *et al*. 2001). Recently it was found that two members of the Wnt family, *Wnt-2B* and *Wnt-8C* contribute to the restriction and/or maintenance of *Fgf-10* expression. Interestingly, *Wnt-2B* is expressed in the forelimb area, whereas *Wnt-8C* is expressed in the hindlimb, indicating that the differences between fore- and hindlimbs are (partially) specified by these genes.

Another gene family with differential expression in fore- and hindlimbs is the *T-box* family of transcription factors. These genes are related to *Drosophila omb*, and four of the genes, *Tbx-2, 3, 4* and *5* are expressed in developing mouse limbs (Chapman *et al*. 1996). *Tbx-4* and *–5* are expressed in developing hindlimbs and forelimbs, respectively. In an experiment where ectopic limbs were induced by implantation of Fgf-2 containing beads, overlapping expression patterns of *Tbx-4* and *–5* were found in the limb buds (Gibson-Brown *et al*. 1998). The *Tbx-4* expression pattern was always more posterior than the *Tbx-5* expression pattern, and the amount of expression of each gene corresponded to the position of the limb: ectopic limbs induced close to the normal hindlimb expressed more *Tbx-4* than *Tbx-5*, and the inverse ratio was visible in limbs induced near forelimbs. Limbs that expressed mostly *Tbx-4* subsequently developed into limbs with mostly structural characteristics of hindlimbs, and limbs expressing mostly *Tbx-5* developed into limbs that resembled forelimbs. It was suggested that *Tbx* genes are induced by the *Hox* genes that have overlapping expression patterns along the body axis of the developing

embryo, and that *Fgf-10* could be a downstream target of the *Tbx* genes (Gibson-Brown *et al.* 1998).

Apoptosis and further outgrowth

An important process in development and maintenance of an organism is programmed cell death, or apoptosis: in an adult human, millions of cells die every minute. Apoptosis is a well organised process, in which cells activate an intracellular death programme: DNA, RNA and protein synthesis decrease, the cells shrink and are rapidly taken up by neighbouring cells (Steller 1995). During development, apoptosis helps in sculpting structures, such as the developing limb: digits are formed by condensation of mesenchyme, which then differentiates into cartilage. The interdigital mesenchyme and its overlying ectoderm undergo apoptosis, and this process is mediated by the mesenchyme. This was proven by Saunders and Fallon, who replaced the interdigital mesenchyme of chick and duck developing limbs (Saunders and Fallon 1967). Replacing chick mesenchyme with duck mesenchyme resulted in a webbed foot, because in duck limbs apoptosis is reduced. Transplantation must be performed before cells are destined to die: after a certain stage, interdigital chick mesenchyme cannot be rescued by a 'duck' environment because the cells, although still alive, have started their apoptosis programme.

In the developing vertebrate limb, zones of apoptotic activity are also found at the anterior and posterior margins of the limb bud and between the radius and ulna (Hurle *et al.* 1996). Apoptosis in the limb is mediated by Bone Morphogenic Proteins (Bmp) -2, -4 and -7 (Ganan *et al.* 1996, Macias *et al.* 1997). Beads soaked in these Bmps accelerate apoptosis when implanted in the interdigital region, and, when implanted in the tip of a growing digit, lead to the formation of an ectopic area of apoptosis. Furthermore, when the function of the Bmp receptor is blocked in the developing chick leg, the digits are webbed (Yokouchi *et al.* 1996).

Digit identity is specified by the interdigital mesenchyme before its regression. This was shown in a series of experiments whereby interdigital mesenchyme was removed or foil barriers prevented signalling from mesenchyme to digit (Dahn and Fallon 2000). Digits develop according to the most posterior signals they receive: when the

mesenchyme between digits 2 and 3 is removed, digit 2 develops into a digit similar to digit 1. Interestingly, this is independent of *Hox* gene expression, and instead appears to be mediated by Bmps. This function of digit specification is in sharp contrast with that of regulating apoptosis. It was suggested that these different functions of Bmps, digit specification and induction of apoptosis, are dependent on the formation of heterodimers between Bmp-2, -4 and –7 (Dahn and Fallon 2000, Drossopoulou *et al*. 2000).

Glycosylation and pattern formation

Regulation of signalling pathways is extremely important, both in development and maintenance of an organism. One major means of regulation is the availability of signalling proteins, the level of which can be changed by altering transcription rates, mRNA or protein stability, or by active degradation. Another mechanism is the sequestering of proteins by binding to inhibitors, thereby preventing their action. Some proteins need to be activated before they can perform their functions, and one means of activation is that produced by glycosylation of proteins. Glycosylation is the addition of sugar polymers to specific serine residues in a protein by glycosyltransferases, a process that takes place in the Golgi apparatus (Stryer 1995). All glycosylated proteins, or proteoglycans, function in the extracellular matrix or at the cell surface. Some are transmembrane proteins, others are secreted and attached in the outer layer of the membrane by attachment of their sugar chains to glycosylphosphatidylinositol (GPI).

Several proteoglycans have been implicated in growth factor signalling. *Drosophila dally*, for instance, encodes a GPI linked protein which is required for morphogenesis of many tissues including the eye, wing and genitalia (Lin *et al*. 1999). Mutations in this gene lead to delay of cell division in the developing eye (dally stands for division abnormally delayed). The exact function of *dally* is not yet known, but it has been shown that it influences *dpp* signalling in several tissues (Jackson *et al*. 1997). In contrast, in the developing wing and the epidermis, *dally* enhances *wingless* signalling and it can rescue *wg* loss of function mutants (Lin *et al*. 1999, Tsuda *et al*. 1999). Although these functions are contrasting, they are tissue specific, and it has been postulated that the sugar residues attached to the Dally protein differ per tissue (Tsuda *et al*. 1999). A human homolog of *dally*, *Glypican 3*, was found to be mutated in an overgrowth and tumour-

susceptibility syndrome named SGBS (for Simpson-Golabi-Behmel syndrome) (MIM: 312870) (Pilia *et al.* 1996).

A few *Drosophila* genes involved in glycosylation have been found. One of these is *tout-velu* (*ttv*), which means 'all hairy' (Bellaiche *et al.* 1998). The gene is homologous to vertebrate *Ext* genes, which are associated with cartilaginous tumours of the growth plate, called exostoses (MIM: 133700 and 133701). Several lines of evidence suggest that the *Ext* genes are involved in the synthesis of heparan sulphate, a disaccharide repeat that is often found in the sugar chains (Lind *et al.* 1998, McCormick *et al.* 1998, The *et al.* 1999, Toyoda *et al.* 2000). *Ttv* mutants have reduced levels of heparan sulphate. Interestingly, loss of *ttv* alters the distribution of Hedgehog: in wildtype cells, HH disperses from the HH synthesising cells across 8-10 cells that respond to HH. In *ttv* mutants, HH stays attached to the HH synthesising cells. It was suggested that this attachment was mediated by a GPI linked protein, since these have been shown to move from one cell membrane to the next. GPI is attached to the outer layer of the cell membrane only, which makes it relatively easy to 'flip'. Loss of heparan sulphate would impair this mechanism (Bellaiche *et al.* 1998). However, later it was shown that HH is covalently attached to cholesterol (Porter *et al.* 1996 and see below) and until now, no GPI-linked Hedgehog has been found.

Recently, it was shown that glycosylation is also important in the Notch pathway. The Notch receptor contains many Epidermal Growth Factor receptor (EGF) motifs, cysteine rich domains that are involved in the proper folding of the protein. Some of these motifs contain a consensus sequence for O-glycosylation and it was indeed shown that mammalian Notch1 is glycosylated by O-linked fucose (Moloney *et al.* 2000b). Interestingly, Fringe, the protein that modulates Notch sensitivity, is a glycosyltransferase, which can add GlcNAc (N-acetylglucosamine) to O-fucose (Moloney *et al.* 2000a). It was postulated that Fringe mediated modification of the O-linked fucose saccharides attached to Notch increases Notch's sensitivity to Delta-like ligands (Blair 2000).

Glycosylation can be a means of regulation of proteins and in turn, the sugar residues can be modified as well. The sugar polymers are often modified by sulphation (adding sulphate), deacetylation or epimerisation (a change in chirality) (Selleck 2000).

These processes act on specific residues, leaving others untouched, and in this way distinct structural motifs can be established in the sugar chains.

Sulphation is needed for interaction of heparan sulphate with growth factors such as Fgf-2 (Olwin and Rapraeger 1992, Rapraeger *et al.* 1991). Heparan sulphate proteoglycans are expressed on the surface of most animal cells, and their structure varies extensively between tissues during development (Bernfield *et al.* 1992). In mice, mutations in Heparan sulphate 2-O-sulfotransferase (Hs2st) an enzyme that attaches sulphate to sugar residues, causes absence or malformation of the kidneys, as well as cleft palate and defects in the eye and skeleton. A high proportion of mutants have postaxial polydactyly, and it was shown that the gene is highly expressed during skeletogenesis, suggesting that it plays a critical role during development (Bullock *et al.* 1998). It has been suggested that heparan sulphate proteoglycans can modulate the activity of growth factors through a number of mechanisms, such as facilitating their dimerisation and altering their effective concentration (Schlessinger *et al.* 1995). Several human disorders combine abnormalities of the kidney, eye, skeleton and palate, but so far, no mutations have been found in human HS2ST.

Syndactyly is often caused by defects in interdigital apoptosis. A gene that functions in this pathway is *Dad1*, the *defender against apoptotic cell death*. Disruption of the gene in mice led to soft-tissue syndactyly in some heterozygous mice. Homozygotes died shortly after apoptosis should have begun (Nishii *et al.* 1999). *Dad1* encodes a subunit of the oligosaccharyltransferase complex, which is well conserved among eukaryotes (Silberstein *et al.* 1995).

The work on glycosylation is far from complete, but it shows already that its function in modulation and modification of proteins and pathways is very important in development.

Hedgehog and the morphogen model
The morphogen model is a long standing model. The idea that a gradient of a signalling molecule is responsible for pattern formation is appealing, especially since several of these molecules have been found. But how do these proteins move from the cell that produces them to the target cells? The general assumption has been that these molecules

28

simply diffuse and attach to cells further away only after the cells closest to the source are satisfied. Recent reports suggest that for hedgehog signalling, this is indeed the case.

The signalling protein HH is produced as an inactive precursor, which needs to be cleaved to release the active N- terminal portion. This cleavage is performed by HH itself, it is 'autoproteolytic' (Lee *et al*. 1994). The N-terminus is then covalently bound to cholesterol and the protein is secreted (Porter *et al*. 1996). Recently it was shown that this cholesterol modification is necessary for Shh signalling, probably because the cholesterol moiety interacts with the hedgehog receptor Ptc (Lewis *et al*. 2001). Ptc contains a sterol-sensing domain, and it is therefore possible that Ptc binds to cholesterol with this domain, thereby sequestering Shh and restricting its motility (Beachy *et al*. 1997). Usually, lipids like cholesterol are inserted into the cell membrane, which would in theory tether the associated protein to the cell. The cholesterol modified Shh, however, is freely diffusible and forms a gradient across the anterior-posterior axis of the developing chick limb (Zeng *et al*. 2001). The protein is multimeric, and it was suggested that the cholesterol moieties are shielded from the aqueous environment by this multimerisation (Zeng *et al*. 2001).

Limb malformations
Defects in limb formation are not always due to genetic causes. Restriction of blood flow to the developing limb, for instance, leads to limb reduction abnormalities such as truncation or oligodactyly (Webster *et al*. 1987). One macabre way in which this may happen is when a monozygotic twin dies in an early stage of development and is reabsorbed by the other twin. Such a process is known to be accompanied by the formation of emboli (blood clots), which can obstruct the blood vessels of the live twin, leading to terminal limb abnormalities. Inhibitors of blood vessel formation (angiogenesis) cause limb malformations. One well known example is thalidomide (the Dutch brandname was Softenon), which was used in the late 1950s as a sedative. It was prescribed to pregnant women because it was effective against morning sickness, and this resulted in many cases of limb malformations (D'Amato *et al*. 1994).

In mice, limb mutations can be an unexpected result from knocking out a gene that is known to be involved in another process. Knockout mice are often generated to study the function of genes, and the resulting phenotype can be quite complex, with limb

malformations being only part of the phenotype. This nevertheless implies that the gene studied is involved in limb development. Angiogenesis is mediated by nitric oxide (NO), which is formed by nitric oxide synthase (NOS). When this enzyme is inhibited in rats during pregnancy, limb defects are seen in 40-50% of the offspring, and mice in which the *Nos* gene was knocked out also showed limb reduction defects (Gregg *et al.* 1998). It was suggested that these were caused by insufficient blood flow to the developing limb, as seen in the thalidomide cases.

Many of the genes involved in limb formation are known, but our insight into limb development is far from complete. Interaction studies with known proteins may reveal new components in a known pathway, but it will not shed light on other mechanisms involved. One way to find new candidates for limb development is by studying hereditary limb malformations and locating the genes responsible. In humans, this is a considerable task that involves locating affected family members and finding the genomic region they all share, which must contain the mutated gene. Despite this, such a project is rewarding because if the gene was not already known to be involved in limb formation, a new pathway or mechanism of limb formation may be revealed. Such a 'positional cloning' approach is independent of known pathways.

In mice, knockout phenotypes may give information on the function of the gene that was knocked out, such as the *NOS* gene described above. Spontaneous mutations also occur in mice, and then the process of finding the responsible gene is very similar to that used in humans. Occasionally, these 'spontaneous' mutations occur by insertion of a transgene into another gene. The disrupted gene cannot function properly, and the resulting phenotype is caused by this disrupted gene, not by the inserted transgene. Finding these genes is somewhat easier than locating spontaneous mutations, because the transgene can function as a tag, which can be located in the mouse genome. An example of such a study is described in Chapter 3 of this thesis. The chapter describes how the insertion of a mutated Rhodopsin gene in the mouse genome disrupts the *fibrillin 2* gene and thereby causes syndactyly.

Human limb malformations

In humans, malformations of the upper limb occur in approximately 1 in 626 newborns (Flatt 1994). These malformations can occur in isolation, meaning that no other abnormalities are seen in the patient, or they are found in combination with other hand and/or foot malformations, or as part of a syndrome. In all these cases, the cause of these malformations may be genetic, environmental, or a combination of these. Although most inherited hand malformations are single gene disorders, patients from one family may show a range of phenotypes indicating that modifier genes and/or the environment play a role in the penetrance of the disorder. One of the most frequently observed hand malformations is polydactyly, with a prevalence of between 5 and 17 per 10,000 live births (de Walle *et al*. 1992, EUROCAT 1991, Sesgin and Stark 1961). Depending on the location of the extra digit(s), polydactyly is divided into preaxial, postaxial and central polydactyly. In preaxial polydactyly, or PPD, the extra digits are located on the thumb side of the hand, and resemble thumbs and/or index fingers. Postaxial polydactyly (PAP) is a duplication of the little finger, and central polydactyly affects the middle digits.

PPD has been subdivided in four subtypes (Temtamy and McKusick 1978). Subtype I is a duplication of a biphalangeal (or normal) thumb. This type is usually not due to genetic causes and is usually unilateral, in contrast to the other three types, which usually segregate in families and are bilateral. Types II and III are polydactyly of a triphalangeal thumb and of an index finger, respectively, and type IV is polysyndactyly, in which the (extra) digits are webbed. Three of these four subtypes have been linked to the same region on chromosome 7q36 (MIM: 190605) (Heutink *et al*. 1994, Tsukurov *et al*. 1994). Later, a recessive limb reduction phenotype, acheiropodia, was linked to this region (MIM: 200500)(Escamilla *et al*. 2000). Chapter 5 of the thesis describes how a mutation causing acheiropodia was found in the LMBR1 gene in this region and Chapter 6 escribes why this gene is also implicated in the PPD and complex polysyndactyly (CPS) phenotypes (see below).

Different phenotypes are also caused by mutations in *GLI3*, one of the vertebrate *ci* homologs. Mutations in *GLI3* have been identified in Pallister-Hall syndrome (PHS)(MIM: 146510), Greig cephalopolysyndactyly (GCPS)(MIM: 175700), and in isolated post- and preaxial polydactyly and polysyndactyly (MIM: 174200 and 174700).

31

Limb malformations are associated with both PHS and GCPS, but are not seen in all patients. GCPS and PHS are quite different phenotypes, with PHS characterised by brain tumours and imperforate anus and GCPS characterised by craniofacial abnormalities. Many of the mutations found in *GLI3* cause truncation of the protein, and it has been suggested that the location of the mutation corresponds to the phenotype (Biesecker 1997). However, since this publication, other mutations have been found that do not correspond to this model, suggesting that the phenotypic outcome of *GLI3* mutations is due to additional factors (Radhakrishna *et al*. 1999).

Mutations in *HOXD13* cause polysyndactyly (Akarsu *et al*. 1996, Muragaki *et al*. 1996) (MIM: 186000). Many *Hox* genes contain a number of short sequences of repeated amino acid residues in their amino-terminal portion, which are thought to function in repression or activation of the gene. In polysyndactyly patients, an alanine stretch was expanded in HOXD13 and found to contain 7-10 more residues than the normal gene. The mutation is semidominant: homozygotes are more severely affected than heterozygotes. Homozygotes have very small hands and feet, because the bones are shortened. This is accompanied by polydactyly and syndactyly, a feature also seen in heterozygotes, who have normal sized hands and feet. The phenotype is probably not caused by a simple loss of function of the protein: when the gene was knocked out in mice, a related but distinct phenotype appeared, indicating that other *Hox* genes may compensate for the lost gene (Dolle *et al*. 1993). Indeed, when *Hoxd11* and *Hoxd12* were knocked out in addition to *Hoxd13*, a polysyndactyly phenotype was observed (Zakany and Duboule 1996). It was suggested that the polyalanine expansion in humans causes the loss of function of several 5' *HOXD* genes, possibly via their functional hierarchy. Later, a similar polyalanine expansion was found in mice with polysyndactyly (Johnson *et al*. 1998b). The phenotype is very similar to the human phenotype.

Mouse limb mutants
Many of the mutations known in human limb malformations cause a similar phenotype in mice (MGD, Mouse Genome Database, see references). Sometimes these phenotypes are induced by inserting a known human mutation in a mouse gene, or adding the mutated human gene to the mouse genome, and sometimes spontaneous mouse limb mutants are

found to have a mutation in a known gene. Additionally, mouse mutations are known that do not have a human counterpart. This is partially because mouse genetics is more controllable than human genetics: once a spontaneous hereditary limb malformation occurs, it is simple to keep the mouse line and search for the mutation.

Similar mutations do not always have the same outcome in both organisms: a mutation leading to syndactyly in mice may not have that effect in humans. Mutations in different parts of a gene sometimes have widely different phenotypes (see below) and even within a species, the phenotypic outcome of a similar mutation may vary, due to genetic modifiers in other parts of the genome. Nevertheless, studying mouse mutations has provided many insights in limb development in general. With its short generation time and easy maintenance, the mouse is an ideal model system for mammalian development. Mice have long been used for the study of limb formation at the molecular level and many key components were identified in this organism. The use of mouse limb mutants for finding genes has a big advantage over humans: even before the gene is found, it is possible to study the effects of the mutation in development: embryological studies allow close observation of the developing limb and the timing of failure can be monitored and it was found that in many cases, the function of the AER is disturbed and the genes necessary for outgrowth of the developing limb are downregulated or misexpressed.

In mice, four known phenotypes are associated with mutations in *Gli3* or with misexpression of the gene: *extra toes* (*xt*), *anterior digit pattern deformity* (*add*), *brachyphalangy* (*bph*) and the *polydactyly Nagoya* mutation (*pdn*) (MGI:95729). All four mutants show limb abnormalities with preaxial polydactyly and sometimes a postaxial rudimentary digit. The mutants have a variable degree of craniofacial abnormalities, but the *xt* mutant most resembles the human GCPS mutation. It is also the only phenotype in which a mutation in *Gli3* was shown. The other three phenotypes are allelic to *xt*: when heterozygotes of these lines are crossed with *xt* heterozygotes, one fourth of the offspring shows a phenotype resembling *xt* homozygotes. This means that in all phenotypes, the mutation is located in the same pathway, most likely in the same gene. However, no mutations have been reported for the *add*, *bph* and *pdn* phenotypes, and it is unlikely that no one has tried to find such mutations. It is possible that the three phenotypes are caused by mutations in regulatory elements of *Gli3*, preventing expression of the gene at a crucial

time during development. In *Gli3^{add}* mutants, a reduction in Gli3 expression was found, indicating that regulation of expression is indeed impaired in this mutant (Schimmang *et al.* 1993).

Jagged-1 and *–2* are vertebrate homologs of *Drosophila Serrate*, the gene expressed in dorsal ectoderm of the developing wing (Lindsell *et al.* 1995, Sidow *et al.* 1997). Like Serrate, Jagged-1 and –2 are ligands for the Notch receptor. Mutations in mouse *Jagged-2* are the cause of the *syndactylism (sm)* phenotype, also referred to as *Jag2^{sm}*. In the developing limbs of this mutant, interdigital apoptosis is reduced, which leads to webbing of the digits. Another *Jagged-2* mutant, with a deletion of a domain required for interaction with Notch family members, shows a more severe phenotype (Jiang *et al.* 1998). In this mutant, the syndactyly is more severe and mice die shortly after birth. In the mutant limb buds, the AER is malformed, *Fgf-8* expression in the AER is expanded and *Bmp-2* and *–7* expression in the interdigital mesenchyme are downregulated.

AER formation and maintenance is also affected in the *limb deformity (ld)* mutant, a recessive phenotype that involves severe syndactyly and oligodactyly of all four feet and reduced or absent kidneys. A gene was found at the *ld* locus, and it was named *Formin* (Woychik *et al.* 1985). The gene encodes multiple isoforms of the Formin protein, but its function is unknown. Six *ld* allelic phenotypes are found, but a mutation has been found in only three, two transgene insertions and one chromosomal rearrangement. In a fourth, one of the isoforms is absent, suggesting that a regulatory element of the *Formin* gene is mutated. This is similar to the findings in the *Gli3* mutant alleles described before, and indicates that mutations in regulatory elements may cause similar phenotypes to those caused by mutations in the open reading frame.

Zuniga *et al.* found that the *ld* phenotype can be restored by application of Gremlin to the developing limb buds (Zuniga *et al.* 1999). Gremlin is a Bmp antagonist, and Bmps repress AER activity. In normal limbs, Gremlin ensures outgrowth by blocking Bmps. *ld* mutants have no Gremlin activity, therefore Bmps are not downregulated in mutant limbs and outgrowth is impaired. The same authors showed that Gremlin expression is induced by ectopic Shh application, and they suggested that Formins function downstream of Shh and upstream of Gremlin, maintaining a positive feedback

loop between Shh and Fgf-8 in the AER. This is an example of a formerly unknown pathway important in limb development, which was found by positional cloning of a gene disrupted in a limb mutant.

Radiation induced mutagenesis has been used for finding locations of disease genes. Irradiating embryonic stem cells leads to random mutations, mostly large genomic deletions. In most cases this is lethal, but some of the cells survive, and retain the ability to develop into embryos. These early embryos are subsequently transplanted back into the uterus and allowed to undergo normal development. When the mice are born, some will have a phenotype, and it is possible to find the corresponding deletions in the genome. One such mutant is the *Shaker-with-syndactylism* (*sy*) mouse, which shows recessive syndactyly in combination with other skeletal abnormalities, as well as deafness and loss of balance, probably due to inner ear defects (MGI:98457). The heterozygote shows no phenotype, but most homozygotes die shortly after birth. The *sy* mutant lacks approximately 0.7 cM of chromosome 18, a part large enough to contain several genes (Johnson *et al.* 1998a). One of the missing genes has been identified as the mouse homolog of the human *SLC12A2* gene, which encodes a Na-K cotransporter (Dixon *et al.* 1999). To verify its function, the gene was knocked out in mice, and the resulting phenotype involved the circling behaviour and deafness also seen in the *sy* mouse (Delpire *et al.* 1999). However, the syndactylism was not seen in this mouse, indicating that another gene in the deleted region was responsible for the syndactyly phenotype. This hypothesis was strengthened by the finding of two allelic variants of *sy, fused phalanges (sy^{fp})* and *sy^{fp-2J}* . Both mutants show recessive syndactyly, but not deafness or circling behaviour, indicating that they are mutated in a 'limb specific' gene. Recently, this gene was shown to be *Fibrillin-2* (Chaudhry *et al.* 2001).

For many of the limb mutants known in mice, the underlying genetic cause has not yet been found. Two of these are the *hammertoe (Hm)* and *hemimelic extra toes (Hx)* mutants, localised on mouse chromosome 5p (MGI:96290). *Hm* mice show webbing of the digits of all four feet. In heterozygotes, the webbing is partial: it extends to the base of the most proximal phalanx. In homozygotes, webbing is complete. *Hx* mice have preaxial polydactyly which is also more severe in homozygotes. The mutations have been mapped close to *Sonic hedgehog* and *Engrailed-2* on mouse chromosome 5p (Knudsen and

Kochhar 1981). In humans, these two genes are located on chromosome 7q36, therefore mouse 5p is syntenic with human 7q36: it is very likely that the two regions contain the same genes. As described above, preaxial polydactyly (PPD) and complex polysyndactyly (CPS), the human PPD phenotypes, are located in this region, and *Hx* and *Hm* are remarkably similar to these phenotypes. Therefore, *Hx* and *Hm* are considered mouse models for PPD and CPS, respectively.

In 1999, another mouse mutation was linked to this region. In a study of *Hoxb1*, a gene not normally present in mouse developing limbs, a transgenic mouse was made containing Hoxb1 under a rhombomere specific promoter. One of the mouse lines showed PPD in the hindlimbs, indicating that the *Hoxb1* transgene insertion had disrupted a gene important in limb formation (Sharpe *et al.* 1999). The mutant was named *sasquatch (Ssq)* and the transgene insertion site was found to be close to *Shh*, in the *Hx* and *Hm* critical region and upstream of Shh. The expression pattern of Shh in *Ssq* limb buds is different from that seen in normal mice: in addition to its normal expression domain at the posterior side of the limb bud, expression was also seen at the anterior side. Such ectopic Shh expression is also seen in *Hx* and other polydactyly mutants (Masuya *et al.* 1995). As described above, ectopic expression of *Shh* at the anterior side of the limb bud leads to the formation of ectopic ZPAs and results in polydactyly.

It is likely that the transgene insertion in the *Ssq* mutant induces the ectopic expression of *Shh*, either directly or indirectly. Sharpe *et al.* suggested that the transgene interferes with an upstream regulatory element of *Shh* (Sharpe *et al.* 1999). Interestingly, the *Hoxb1* transgene was also expressed at the anterior and posterior sides of the limb bud in an expression pattern overlapping with *Shh*. However, the transgene was not expressed in other tissues that express Shh, indicating that the insertion has placed the construct under the influence of a limb-specific regulatory element that also acts upon Shh.

Scope of the thesis
To understand the genetics of limb development, it is important to identify genes important in limb formation. As described above, the study of limb mutants can be used to identify such genes. This thesis describes two studies of limb malformations, one in mice and one in humans.

In humans, the main way of identifying disease-causing genes is by positional cloning. This involves the localisation of the disease to a genomic region, the cloning of this region and the identification of genes in this region. Now that the Human Genome Project is nearly finished, a part of this process can be sped up tremendously. The advances in 'computer cloning' are described in the second part of the Introduction, **Chapter 2**.

Mouse mutants are somewhat easier to study than human subjects, and many of the genes involved in embryonal development have been identified in the mouse. We studied the *tipsy* mutant, a mouse that shows syndactyly in the hindlimbs. This mutation is recessive, and was caused by integration of a transgene. **Chapter 3** describes how we identified the gene *Fibrillin 2* at the integration site. Due to the transgene insertion, *Fibrillin 2* is disrupted and transcription is reduced. This indicates a function for *Fibrillin 2* in limb formation.

Limb malformations also occur in humans, and dominant mutations in genes can result in large families with hand and/or foot disorders. In 1994, PPD types II and III were linked to a region on chromosome 7q36, close to *Sonic hedgehog* (Heutink *et al.* 1994) (MIM: 190605). The phenotypes of the family members varied between opposable and non-opposable thumbs, thus between type II and III. This shows that although the clinical classification distinguishes between the phenotypes, the underlying mutation is probably the same. The phenotypic outcome of the mutation must be due to external factors, such as modifier genes or environmental influences. At the same time, type IV was linked to this region, and in 2000 a limb reduction phenotype, acheiropodia, was mapped to the region as well (Escamilla *et al.* 2000, Tsukurov *et al.* 1994) (MIM: 200500). This indicates that the same gene may be responsible for a wide range of limb phenotypes. We refined the candidate region to a 400 kb region, which is described in **Chapter 4a**. In this chapter we report three candidate genes that lie within the region, *C7orf2* (later renamed LMBR1), *C7orf3* and *Hlxb9*. With the help of computer analysis of the genome sequence in this region, we identified two additional transcripts, *C7orf13* and *RNF32*. This is described in **Chapter 4b**. **Chapter 5** of this thesis describes how a mutation causing acheiropodia was found in the *Lmbr1* gene, and **Chapter 6** describes why this gene is also implicated in the PPD and complex polysyndactyly (CPS)

phenotypes. In this chapter it is shown that the transgene insertion in *Ssq* resides in intron 5 of *Lmbr1*. In the same chapter, a translocation patient with PPD is described. The translocation breakpoint is in the same intron of LMBR1, indicating that a similar mutational mechanism is involved in the Ssq mouse and this patient. It is unclear if a regulatory element of *Shh* is disrupted or if the *Lmbr1* gene itself is involved in limb formation; various models are discussed in the chapter and in the Discussion (**Chapter 7**).

So far, no mutations have been found in the other mouse and human PPD phenotypes. However, Clark *et al.* found that *Lmbr1* was expressed during the formation of the limb bud in wildtype mice, and that it showed a reduction of expression levels in *Hx* mice (Clark *et al.* 2000). However, mutations in the gene have not been found, and this indicates that, similar to the mouse *Ssq*, *Gli3* and *ld* mutations, a regulatory element of the gene may be disrupted in PPD.

In this thesis, several limb phenotypes and mutations are discussed. Together, they show that genotype-phenotype relationships are not always clear, and that the molecular genetics of limb development may still hold a few surprises.

References

Akarsu, A. N., I. Stoilov, E. Yilmaz, B. S. Sayli, and M. Sarfarazi. 1996. Genomic structure of HOXD13 gene: a nine polyalanine duplication causes synpolydactyly in two unrelated families. *Hum Mol Genet* 5: 945-52.

Alcedo, J., M. Ayzenzon, T. Von Ohlen, M. Noll, and J. E. Hooper. 1996. The Drosophila smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell* 86: 221-32.

Alexandre, C., A. Jacinto, and P. W. Ingham. 1996. Transcriptional activation of hedgehog target genes in Drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes Dev* 10: 2003-13.

Aza-Blanc, P., and T. B. Kornberg. 1999. Ci: a complex transducer of the hedgehog signal. *Trends Genet* 15: 458-62.

Aza-Blanc, P., F. A. Ramirez-Weber, M. P. Laget, C. Schwartz, and T. B. Kornberg. 1997. Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* 89: 1043-53.

Basler, K., and G. Struhl. 1994. Compartment boundaries and the control of Drosophila limb pattern by hedgehog protein. *Nature* 368: 208-14.

Beachy, P. A., M. K. Cooper, K. E. Young, D. P. von Kessler, W. J. Park, T. M. Hall, D. J. Leahy, and J. A. Porter. 1997. Multiple roles of cholesterol in hedgehog protein biogenesis and signaling. *Cold Spring Harb Symp Quant Biol* 62: 191-204.

Bellaiche, Y., I. The, and N. Perrimon. 1998. Tout-velu is a Drosophila homologue of the putative tumour suppressor EXT-1 and is needed for Hh diffusion. *Nature* 394: 85-8.

Bernfield, M., R. Kokenyesi, M. Kato, M. T. Hinkes, J. Spring, R. L. Gallo, and E. J. Lose. 1992. Biology of the syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annu Rev Cell Biol* 8: 365-93.

Biesecker, L. G. 1997. Strike three for GLI3. *Nat Genet* 17: 259-60.

Blair, S. S. 2000. Notch signaling: Fringe really is a glycosyltransferase. *Curr Biol* 10: R608-12.

Blaumueller, C. M., H. Qi, P. Zagouras, and S. Artavanis-Tsakonas. 1997. Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane. *Cell* 90: 281-91.

Bullock, S. L., J. M. Fletcher, R. S. Beddington, and V. A. Wilson. 1998. Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase. *Genes Dev* 12: 1894-906.

Burrow, C. R. 2000. Regulatory molecules in kidney development. *Pediatr Nephrol* 14: 240-53.

Chang, D. T., A. Lopez, D. P. von Kessler, C. Chiang, B. K. Simandl, R. Zhao, M. F. Seldin, J. F. Fallon, and P. A. Beachy. 1994. Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development* 120: 3339-53.

Chapman, D. L., N. Garvey, S. Hancock, M. Alexiou, S. I. Agulnik, J. J. Gibson-Brown, J. Cebra-Thomas, R. J. Bollag, L. M. Silver, and V. E. Papaioannou. 1996. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev Dyn* 206: 379-90.

Chaudhry, S. S., J. Gazzard, C. Baldock, J. Dixon, M. J. Rock, G. C. Skinner, K. P. Steel, C. M. Kielty, and M. J. Dixon. 2001. Mutation of the gene encoding fibrillin-2 results in syndactyly in mice. *Hum Mol Genet* 10: 835-43.

Chen, Y., and G. Struhl. 1996. Dual roles for patched in sequestering and transducing Hedgehog. *Cell* 87: 553-63.

Chiang, C., Y. Litingtung, E. Lee, K. E. Young, J. L. Corden, H. Westphal, and P. A. Beachy. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383: 407-13.

Clark, R., P. Marker, and D. Kingsley. 2000. A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. *Genomics* 67: 19-27.

Crossley, P. H., G. Minowada, C. A. MacArthur, and G. R. Martin. 1996. Roles for FGF8 in the induction, initiation, and maintenance of chick limb development. *Cell* 84: 127-36.

Dahn, R. D., and J. F. Fallon. 2000. Interdigital regulation of digit identity and homeotic transformation by modulated BMP signaling. *Science* 289: 438-41.

D'Amato, R. J., M. S. Loughnan, E. Flynn, and J. Folkman. 1994. Thalidomide is an inhibitor of angiogenesis. *Proc Natl Acad Sci U S A* 91: 4082-5.

Davis, C. A., D. P. Holmyard, K. J. Millen, and A. L. Joyner. 1991. Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development* 111: 287-98.

Davis, C. A., and A. L. Joyner. 1988. Expression patterns of the homeo box-containing genes En-1 and En-2 and the proto-oncogene int-1 diverge during mouse development. *Genes Dev* 2: 1736-44.

de Celis, J. F., R. Barrio, and F. C. Kafatos. 1996. A gene complex acting downstream of dpp in Drosophila wing morphogenesis. *Nature* 381: 421-4.

de Walle, H. E. K., M. C. Cornel, T. M. Haverman, A. C. Breed, J. B. G. M. Verhey, and L. P. ten Kate. 1992. EUROCAT, registration of congenital anomalies North Netherlands, tables 1981-1990. Rijksuniversiteit, Department of Medical Genetics, Medical Faculty, Groningen, Groningen.

Dealy, C. N., A. Roth, D. Ferrari, A. M. Brown, and R. A. Kosher. 1993. Wnt-5a and Wnt-7a are expressed in the developing chick limb bud in a manner suggesting roles in pattern formation along the proximodistal and dorsoventral axes. *Mech Dev* 43: 175-86.

Delpire, E., J. Lu, R. England, C. Dull, and T. Thorne. 1999. Deafness and imbalance associated with inactivation of the secretory Na-K-2Cl co-transporter. *Nat Genet* 22: 192-5.

Dixon, M. J., J. Gazzard, S. S. Chaudhry, N. Sampson, B. A. Schulte, and K. P. Steel. 1999. Mutation of the Na-K-Cl co-transporter gene Slc12a2 results in deafness in mice. *Hum Mol Genet* 8: 1579-84.

Doherty, D., G. Feger, S. Younger-Shepherd, L. Y. Jan, and Y. N. Jan. 1996. Delta is a ventral to dorsal signal complementary to Serrate, another Notch ligand, in Drosophila wing formation. *Genes Dev* 10: 421-34.

Dolle, P., A. Dierich, M. LeMeur, T. Schimmang, B. Schuhbaur, P. Chambon, and D. Duboule. 1993. Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. *Cell* 75: 431-41.

Dolle, P., J. C. Izpisua-Belmonte, E. Boncinelli, and D. Duboule. 1991. The Hox-4.8 gene is localized at the 5' extremity of the Hox-4 complex and is expressed in the most posterior parts of the body during development. *Mech Dev* 36: 3-13.

Dolle, P., J. C. Izpisua-Belmonte, H. Falkenstein, A. Renucci, and D. Duboule. 1989. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. *Nature* 342: 767-72.

Dominguez, M., M. Brunner, E. Hafen, and K. Basler. 1996. Sending and receiving the hedgehog signal: control by the Drosophila Gli protein Cubitus interruptus. *Science* 272: 1621-5.

Dono, R., and R. Zeller. 1994. Cell-type-specific nuclear translocation of fibroblast growth factor-2 isoforms during chicken kidney and limb morphogenesis. *Dev Biol* 163: 316-30.

Driesch, H. 1892. Entwicklungsmechanisme Studien. I. Der Werth der beiden ersten Furchungszellen in der Echinodermentwicklung. *Zeitschrift für wissenschaftliche Zoologie* 53: 160-178.

Drossopoulou, G., K. E. Lewis, J. J. Sanz-Ezquerro, N. Nikbakht, A. P. McMahon, C. Hofmann, and C. Tickle. 2000. A model for anteroposterior patterning of the vertebrate limb based on sequential long- and short-range Shh signalling and Bmp signalling. *Development* 127: 1337-48.

Duboule, D. 1995. Vertebrate Hox genes and proliferation: an alternative pathway to homeosis? *Curr Opin Genet Dev* 5: 525-8.

Escamilla, M. A., M. C. DeMille, E. Benavides, E. Roche, L. Almasy, S. Pittman, J. Hauser, D. F. Lew, N. B. Freimer, and M. R. Whittle. 2000. A minimalist approach to gene mapping: locating the gene for acheiropodia, by homozygosity analysis. *Am J Hum Genet* 66: 1995-2000.

EUROCAT. 1991. Eurocat Report 4. Surveillance of congenital anomalies,1980-1988. EUROCAT central registry, Brussels.

Fallon, J. F., A. Lopez, M. A. Ros, M. P. Savage, B. B. Olwin, and B. K. Simandl. 1994. FGF-2: apical ectodermal ridge growth signal for chick limb development. *Science* 264: 104-7.

Fehon, R. G., P. J. Kooh, I. Rebay, C. L. Regan, T. Xu, M. A. Muskavitch, and S. Artavanis-Tsakonas. 1990. Molecular interactions between the protein products of the neurogenic loci Notch and Delta, two EGF-homologous genes in Drosophila. *Cell* 61: 523-34.

Ferguson, E. L., and K. V. Anderson. 1992. Localized enhancement and repression of the activity of the TGF-beta family member, decapentaplegic, is necessary for dorsal-ventral pattern formation in the Drosophila embryo. *Development* 114: 583-97.

Flatt, A. E. 1994. *The care of congenital hand anomalies.* Quality Medical Publishing, St. Louis.

Fleming, R. J., Y. Gu, and N. A. Hukriede. 1997. Serrate-mediated activation of Notch is specifically blocked by the product of the gene fringe in the dorsal compartment of the Drosophila wing imaginal disc. *Development* 124: 2973-81.

Ganan, Y., D. Macias, M. Duterque-Coquillaud, M. A. Ros, and J. M. Hurle. 1996. Role of TGF beta s and BMPs as signals controlling the position of the digits and the areas of interdigital cell death in the developing chick limb autopod. *Development* 122: 2349-57.

Garcia-Bellido, A., P. Ripoll, and G. Morata. 1973. Developmental compartmentalisation of the wing disk of Drosophila. *Nat New Biol* 245: 251-3.

Gardner, C. A., and K. F. Barald. 1992. Expression patterns of engrailed-like proteins in the chick embryo. *Dev Dyn* 193: 370-88.

Gibson-Brown, J. J., S. I. Agulnik, L. M. Silver, L. Niswander, and V. E. Papaioannou. 1998. Involvement of T-box genes Tbx2-Tbx5 in vertebrate limb specification and development. *Development* 125: 2499-509.

Gilbert, S. F. 2000. *Developmental Biology.* Sinauer Associates, Inc., Sunderland, MA.

Gregg, A. R., A. Schauer, O. Shi, Z. Liu, C. G. Lee, and W. E. O'Brien. 1998. Limb reduction defects in endothelial nitric oxide synthase-deficient mice. *Am J Physiol* 275: H2319-24.

Grimm, S., and G. O. Pflugfelder. 1996. Control of the gene optomotor-blind in Drosophila wing development by decapentaplegic and wingless. *Science* 271: 1601-4.

Haack, H., and P. Gruss. 1993. The establishment of murine Hox-1 expression domains during patterning of the limb. *Dev Biol* 157: 410-22.

Hall, B. K., and T. Miyake. 2000. All for one and one for all: condensations and the initiation of skeletal development. *Bioessays* 22: 138-47.

Haramis, A. G., J. M. Brown, and R. Zeller. 1995. The limb deformity mutation disrupts the SHH/FGF-4 feedback loop and regulation of 5' HoxD genes during limb pattern formation. *Development* 121: 4237-45.

Harrison, R. G. 1918. Experiments on the development of the fore limb of Amblystoma, a self-differentiating equipotential system. *J Exp Zool* 25: 413-61.

Heemskerk, J., and S. DiNardo. 1994. Drosophila hedgehog acts as a morphogen in cellular patterning. *Cell* 76: 449-60.

Heikinheimo, M., A. Lawshe, G. M. Shackleford, D. B. Wilson, and C. A. MacArthur. 1994. Fgf-8 expression in the post-gastrulation mouse suggests roles in the development of the face, limbs and central nervous system. *Mech Dev* 48: 129-38.

Heutink, P., J. Zguricas, L. van Oosterhout, G. J. Breedveld, L. Testers, L. A. Sandkuijl, P. J. Snijders, J. Weissenbach, D. Lindhout, S. E. Hovius, and *et al.* 1994. The gene for

triphalangeal thumb maps to the subtelomeric region of chromosome 7q. *Nat Genet* 6: 287-92.

Hurle, J. M., M. A. Ros, V. Climent, and V. Garcia-Martinez. 1996. Morphology and significance of programmed cell death in the developing limb bud of the vertebrate embryo. *Microsc Res Tech* 34: 236-46.

Irvine, K. D., and E. Wieschaus. 1994. fringe, a Boundary-specific signaling molecule, mediates interactions between dorsal and ventral cells during Drosophila wing development. *Cell* 79: 595-606.

Izpisua-Belmonte, J. C., C. Tickle, P. Dolle, L. Wolpert, and D. Duboule. 1991. Expression of the homeobox Hox-4 genes and the specification of position in chick wing development. *Nature* 350: 585-9.

Jackson, S. M., H. Nakato, M. Sugiura, A. Jannuzi, R. Oakes, V. Kaluza, C. Golden, and S. B. Selleck. 1997. dally, a Drosophila glypican, controls cellular responses to the TGF-beta-related morphogen, Dpp. *Development* 124: 4113-20.

Jiang, R., Y. Lan, H. D. Chapman, C. Shawber, C. R. Norton, D. V. Serreze, G. Weinmaster, and T. Gridley. 1998. Defects in limb, craniofacial, and thymic development in Jagged2 mutant mice. *Genes Dev* 12: 1046-57.

Johnson, K. R., S. A. Cook, and Q. Y. Zheng. 1998a. The original shaker-with-syndactylism mutation (sy) is a contiguous gene deletion syndrome. *Mamm Genome* 9: 889-92.

Johnson, K. R., H. O. Sweet, L. R. Donahue, P. Ward-Bailey, R. T. Bronson, and M. T. Davisson. 1998b. A new spontaneous mouse mutation of Hoxd13 with a polyalanine expansion and phenotype similar to human synpolydactyly. *Hum Mol Genet* 7: 1033-8.

Johnson, R. L., R. D. Riddle, E. Laufer, and C. Tabin. 1994. Sonic hedgehog: a key mediator of anterior-posterior patterning of the limb and dorso-ventral patterning of axial embryonic structures. *Biochem Soc Trans* 22: 569-74.

Johnson, R. L., and C. J. Tabin. 1997. Molecular models for vertebrate limb development. *Cell* 90: 979-90.

Johnston, S. H., C. Rauskolb, R. Wilson, B. Prabhakaran, K. D. Irvine, and T. F. Vogt. 1997. A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway. *Development* 124: 2245-54.

Kawakami, Y., J. Capdevila, D. Buscher, T. Itoh, C. Rodriguez Esteban, and J. C. Izpisua Belmonte. 2001. WNT signals control FGF-dependent limb initiation and AER induction in the chick embryo. *Cell* 104: 891-900.

Kimmel, R. A., D. H. Turnbull, V. Blanquet, W. Wurst, C. A. Loomis, and A. L. Joyner. 2000. Two lineage boundaries coordinate vertebrate apical ectodermal ridge formation. *Genes Dev* 14: 1377-89.

Klein, T., and A. M. Arias. 1998. Interactions among Delta, Serrate and Fringe modulate Notch activity during Drosophila wing development. *Development* 125: 2951-62.

Knudsen, T. B., and D. M. Kochhar. 1981. The role of morphogenetic cell death during abnormal limb-bud outgrowth in mice heterozygous for the dominant mutation Hemimelia-extra toe (Hmx). *J Embryol Exp Morphol* 65: 289-307.

Kornberg, T. 1981. Engrailed: a gene controlling compartment and segment formation in Drosophila. *Proc Natl Acad Sci U S A* 78: 1095-9.

Krabbenhoft, K. M., and J. F. Fallon. 1989. The formation of leg or wing specific structures by leg bud cells grafted to the wing bud is influenced by proximity to the apical ridge. *Dev Biol* 131: 373-82.

Laufer, E., R. Dahn, O. E. Orozco, C. Y. Yeo, J. Pisenti, D. Henrique, U. K. Abbott, J. F. Fallon, and C. Tabin. 1997. Expression of Radical fringe in limb-bud ectoderm regulates apical ectodermal ridge formation. *Nature* 386: 366-73.

Laufer, E., C. E. Nelson, R. L. Johnson, B. A. Morgan, and C. Tabin. 1994. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* 79: 993-1003.

42

Lawrence, P. A. 1990. Developmental biology. Compartments in vertebrates? *Nature* 344: 382-3.

Lecuit, T., W. J. Brook, M. Ng, M. Calleja, H. Sun, and S. M. Cohen. 1996. Two distinct mechanisms for long-range patterning by Decapentaplegic in the Drosophila wing. *Nature* 381: 387-93.

Lee, J. J., S. C. Ekker, D. P. von Kessler, J. A. Porter, B. I. Sun, and P. A. Beachy. 1994. Autoproteolysis in hedgehog protein biogenesis. *Science* 266: 1528-37.

Lettice, L., J. Hecksher-Sorensen, and R. E. Hill. 1999. The dominant hemimelia mutation uncouples epithelial-mesenchymal interactions and disrupts anterior mesenchyme formation in mouse hindlimbs. *Development* 126: 4729-36.

Lewis, E. B. 1995. *The Bithorax Complex: The First Fifty Years.*

Lewis, P. M., M. P. Dunn, J. A. McMahon, M. Logan, J. F. Martin, B. St-Jacques, and A. P. McMahon. 2001. Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by ptc1. *Cell* 105: 599-612.

Lin, X., N. Perrimon, M. Tsuda, K. Kamimura, H. Nakato, M. Archer, W. Staatz, B. Fox, M. Humphrey, S. Olson, T. Futch, V. Kaluza, E. Siegfried, L. Stam, and S. B. Selleck. 1999. Dally cooperates with Drosophila Frizzled 2 to transduce Wingless signalling. *Nature* 400: 281-4.

Lind, T., F. Tufaro, C. McCormick, U. Lindahl, and K. Lidholt. 1998. The putative tumor suppressors EXT1 and EXT2 are glycosyltransferases required for the biosynthesis of heparan sulfate. *J Biol Chem* 273: 26265-8.

Lindsell, C. E., C. J. Shawber, J. Boulter, and G. Weinmaster. 1995. Jagged: a mammalian ligand that activates Notch1. *Cell* 80: 909-17.

Logan, C., A. Hornbruch, I. Campbell, and A. Lumsden. 1997. The role of Engrailed in establishing the dorsoventral axis of the chick limb. *Development* 124: 2317-24.

Loomis, C. A., E. Harris, J. Michaud, W. Wurst, M. Hanks, and A. L. Joyner. 1996. The mouse Engrailed-1 gene and ventral limb patterning. *Nature* 382: 360-3.

Lopez-Martinez, A., D. T. Chang, C. Chiang, J. A. Porter, M. A. Ros, B. K. Simandl, P. A. Beachy, and J. F. Fallon. 1995. Limb-patterning activity and restricted posterior localization of the amino-terminal product of Sonic hedgehog cleavage. *Curr Biol* 5: 791-6.

Macias, D., Y. Ganan, T. K. Sampath, M. E. Piedra, M. A. Ros, and J. M. Hurle. 1997. Role of BMP-2 and OP-1 (BMP-7) in programmed cell death and skeletogenesis during chick limb development. *Development* 124: 1109-17.

Masuya, H., T. Sagai, S. Wakana, K. Moriwaki, and T. Shiroishi. 1995. A duplicated zone of polarizing activity in polydactylous mouse mutants. *Genes Dev* 9: 1645-53.

McCormick, C., Y. Leduc, D. Martindale, K. Mattison, L. E. Esford, A. P. Dyer, and F. Tufaro. 1998. The putative tumour suppressor EXT1 alters the expression of cell-surface heparan sulfate. *Nat Genet* 19: 158-61.

MGD. *Mouse Genome Database. http://www.informatics.jax.org/.*

Moloney, D. J., V. M. Panin, S. H. Johnston, J. Chen, L. Shao, R. Wilson, Y. Wang, P. Stanley, K. D. Irvine, R. S. Haltiwanger, and T. F. Vogt. 2000a. Fringe is a glycosyltransferase that modifies Notch. *Nature* 406: 369-75.

Moloney, D. J., L. H. Shair, F. M. Lu, J. Xia, R. Locke, K. L. Matta, and R. S. Haltiwanger. 2000b. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem* 275: 9604-11.

Muragaki, Y., S. Mundlos, J. Upton, and B. R. Olsen. 1996. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 272: 548-51.

Nellen, D., R. Burke, G. Struhl, and K. Basler. 1996. Direct and long-range action of a DPP morphogen gradient. *Cell* 85: 357-68.

Nishii, K., T. Tsuzuki, M. Kumai, N. Takeda, H. Koga, S. Aizawa, T. Nishimoto, and Y. Shibata. 1999. Abnormalities of developmental cell death in Dad1-deficient mice. *Genes Cells* 4: 243-52.

Niswander, L., S. Jeffrey, G. R. Martin, and C. Tickle. 1994. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature* 371: 609-12.

Niswander, L., and G. R. Martin. 1992. Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* 114: 755-68.

Niswander, L., C. Tickle, A. Vogel, I. Booth, and G. R. Martin. 1993. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* 75: 579-87.

Noji, S., T. Nohno, E. Koyama, K. Muto, K. Ohyama, Y. Aoki, K. Tamura, K. Ohsugi, H. Ide, S. Taniguchi, and *et al*. 1991. Retinoic acid induces polarizing activity but is unlikely to be a morphogen in the chick limb bud. *Nature* 350: 83-6.

Ohuchi, H., T. Nakagawa, A. Yamamoto, A. Araga, T. Ohata, Y. Ishimaru, H. Yoshioka, T. Kuwana, T. Nohno, M. Yamasaki, N. Itoh, and S. Noji. 1997. The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor. *Development* 124: 2235-44.

Ohuchi, H., H. Yoshioka, A. Tanaka, Y. Kawakami, T. Nohno, and S. Noji. 1994. Involvement of androgen-induced growth factor (FGF-8) gene in mouse embryogenesis and morphogenesis. *Biochem Biophys Res Commun* 204: 882-8.

Olwin, B. B., and A. Rapraeger. 1992. Repression of myogenic differentiation by aFGF, bFGF, and K-FGF is dependent on cellular heparan sulfate. *J Cell Biol* 118: 631-9.

Panin, V. M., V. Papayannopoulos, R. Wilson, and K. D. Irvine. 1997. Fringe modulates Notch-ligand interactions. *Nature* 387: 908-12.

Parr, B. A., and A. P. McMahon. 1995. Dorsalizing signal Wnt-7a required for normal polarity of D-V and A-P axes of mouse limb. *Nature* 374: 350-3.

Parr, B. A., M. J. Shea, G. Vassileva, and A. P. McMahon. 1993. Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* 119: 247-61.

Phillips, R. G., I. J. Roberts, P. W. Ingham, and J. R. Whittle. 1990. The Drosophila segment polarity gene patched is involved in a position-signalling mechanism in imaginal discs. *Development* 110: 105-14.

Pilia, G., R. M. Hughes-Benzie, A. MacKenzie, P. Baybayan, E. Y. Chen, R. Huber, G. Neri, A. Cao, A. Forabosco, and D. Schlessinger. 1996. Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nat Genet* 12: 241-7.

Porter, J. A., K. E. Young, and P. A. Beachy. 1996. Cholesterol modification of hedgehog signaling proteins in animal development. *Science* 274: 255-9.

Radhakrishna, U., D. Bornholdt, H. S. Scott, U. C. Patel, C. Rossier, H. Engel, A. Bottani, D. Chandal, J. L. Blouin, J. V. Solanki, K. H. Grzeschik, and S. E. Antonarakis. 1999. The phenotypic spectrum of GLI3 morphopathies includes autosomal dominant preaxial polydactyly type-IV and postaxial polydactyly type-A/B; No phenotype prediction from the position of GLI3 mutations. *Am J Hum Genet* 65: 645-55.

Rapraeger, A. C., A. Krufka, and B. B. Olwin. 1991. Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation. *Science* 252: 1705-8.

Rebay, I., R. J. Fleming, R. G. Fehon, L. Cherbas, P. Cherbas, and S. Artavanis-Tsakonas. 1991. Specific EGF repeats of Notch mediate interactions with Delta and Serrate: implications for Notch as a multifunctional receptor. *Cell* 67: 687-99.

Riddle, R. D., M. Ensini, C. Nelson, T. Tsuchida, T. M. Jessell, and C. Tabin. 1995. Induction of the LIM homeobox gene Lmx1 by WNT7a establishes dorsoventral pattern in the vertebrate limb. *Cell* 83: 631-40.

Riddle, R. D., R. L. Johnson, E. Laufer, and C. Tabin. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* 75: 1401-16.

Robbins, D. J., K. E. Nybakken, R. Kobayashi, J. C. Sisson, J. M. Bishop, and P. P. Therond. 1997. Hedgehog elicits signal transduction by means of a large complex containing the kinesin-related protein costal2. *Cell* 90: 225-34.

Rodriguez-Esteban, C., J. W. Schwabe, J. De La Pena, B. Foys, B. Eshelman, and J. C. Belmonte. 1997. Radical fringe positions the apical ectodermal ridge at the dorsoventral boundary of the vertebrate limb. *Nature* 386: 360-6.

Rulifson, E. J., and S. S. Blair. 1995. Notch regulates wingless expression and is not required for reception of the paracrine wingless signal during wing margin neurogenesis in Drosophila. *Development* 121: 2813-24.

Saldanha, G. 2001. The Hedgehog signalling pathway and cancer. *J Pathol* 193: 427-32.

Saunders, J. W. J. 1948. The proximo-distal sequence of origin of the parts of the chick wing and the role of the ectoderm. *J Exp Zool* 108: 363-403.

Saunders, J. W. J., and J. F. Fallon. 1967. *Cell death in morphogenesis.* Academic Press, New York.

Saunders, J. W. J., and M. T. Gasseling. 1968. Ectodermal-mesodermal interactions in the origin of limb symmetry. Pages 78-97 in R. E. Fleischmajer and R. Billingham, eds. *Epithlelial-Mesenchymal Interactions.* Williams and Wilkins, Baltimore.

Savage, M. P., C. E. Hart, B. B. Riley, J. Sasse, B. B. Olwin, and J. F. Fallon. 1993. Distribution of FGF-2 suggests it has a role in chick limb bud growth. *Dev Dyn* 198: 159-70.

Schimmang, T., F. van der Hoeven, and U. Ruther. 1993. Gli3 expression is affected in the morphogenetic mouse mutants add and Xt. *Prog Clin Biol Res* 383A: 153-61.

Schlessinger, J., I. Lax, and M. Lemmon. 1995. Regulation of growth factor activation by proteoglycans: what is the role of the low affinity receptors? *Cell* 83: 357-60.

Searls, R. L., and M. Y. Janners. 1971. The initiation of limb bud outgrowth in the embryonic chick. *Dev Biol* 24: 198-213.

Sekine, K., H. Ohuchi, M. Fujiwara, M. Yamasaki, T. Yoshizawa, T. Sato, N. Yagishita, D. Matsui, Y. Koga, N. Itoh, and S. Kato. 1999. Fgf10 is essential for limb and lung formation. *Nat Genet* 21: 138-41.

Selleck, S. B. 2000. Proteoglycans and pattern formation: sugar biochemistry meets developmental genetics. *Trends Genet* 16: 206-12.

Serrano, N., and P. H. O'Farrell. 1997. Limb morphogenesis: connections between patterning and growth. *Curr Biol* 7: R186-95.

Sesgin, M. Z., and R. B. Stark. 1961. The incidence of congenital defects. *Plast Reconst Surg* 27: 261-267.

Sharpe, J., L. Lettice, J. Hecksher-Sorensen, M. Fox, R. Hill, and R. Krumlauf. 1999. Identification of sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr Biol* 9: 97-100.

Shellenbarger, D. L., and J. D. Mohler. 1978. Temperature-sensitive periods and autonomy of pleiotropic effects of l(1)Nts1, a conditional notch lethal in Drosophila. *Dev Biol* 62: 432-46.

Sidow, A., M. S. Bulotsky, A. W. Kerrebrock, R. T. Bronson, M. J. Daly, M. P. Reeve, T. L. Hawkins, B. W. Birren, R. Jaenisch, and E. S. Lander. 1997. Serrate2 is disrupted in the mouse limb-development mutant syndactylism. *Nature* 389: 722-5.

Silberstein, S., P. G. Collins, D. J. Kelleher, and R. Gilmore. 1995. The essential OST2 gene encodes the 16-kD subunit of the yeast oligosaccharyltransferase, a highly conserved protein expressed in diverse eukaryotic organisms. *J Cell Biol* 131: 371-83.

Speicher, S. A., U. Thomas, U. Hinz, and E. Knust. 1994. The Serrate locus of Drosophila and its role in morphogenesis of the wing imaginal discs: control of cell proliferation. *Development* 120: 535-44.

St Johnston, R. D., and W. M. Gelbart. 1987. Decapentaplegic transcripts are localized along the dorsal-ventral axis of the Drosophila embryo. *EMBO J* 6: 2785-91.

Steller, H. 1995. Mechanisms and genes of cellular suicide. *Science* 267: 1445-9.

Struhl, G., and A. Adachi. 1998. Nuclear access and action of notch in vivo. *Cell* 93: 649-60.

Stryer, L. 1995. *Biochemistry.* W H Freeman & Co., New York.

Summerbell, D. 1974. A quantitative analysis of the effect of excision of the AER from the chick limb-bud. *J Embryol Exp Morphol* 32: 651-60.

Summerbell, D. 1979. The zone of polarizing activity: evidence for a role in normal chick limb morphogenesis. *J Embryol Exp Morphol* 50: 217-33.

Summerbell, D., J. H. Lewis, and L. Wolpert. 1973. Positional information in chick limb morphogenesis. *Nature* 244: 492-6.

Suzuki, H. R., H. Sakamoto, T. Yoshida, T. Sugimura, M. Terada, and M. Solursh. 1992. Localization of HstI transcripts to the apical ectodermal ridge in the mouse embryo. *Dev Biol* 150: 219-22.

Temtamy, S. A., and V. A. McKusick. 1978. The genetics of hand malformations. *Birth Defects Orig Artic Ser* 14: i-xviii, 1-619.

The, I., Y. Bellaiche, and N. Perrimon. 1999. Hedgehog movement is regulated through tout velu-dependent synthesis of a heparan sulfate proteoglycan. *Mol Cell* 4: 633-9.

Tickle, C., B. Alberts, L. Wolpert, and J. Lee. 1982. Local application of retinoic acid to the limb bond mimics the action of the polarizing region. *Nature* 296: 564-6.

Toyoda, H., A. Kinoshita-Toyoda, and S. B. Selleck. 2000. Structural analysis of glycosaminoglycans in Drosophila and Caenorhabditis elegans and demonstration that tout-velu, a Drosophila gene related to EXT tumor suppressors, affects heparan sulfate in vivo. *J Biol Chem* 275: 2269-75.

Tsuda, M., K. Kamimura, H. Nakato, M. Archer, W. Staatz, B. Fox, M. Humphrey, S. Olson, T. Futch, V. Kaluza, E. Siegfried, L. Stam, and S. B. Selleck. 1999. The cell-surface proteoglycan Dally regulates Wingless signalling in Drosophila. *Nature* 400: 276-80.

Tsukurov, O., A. Boehmer, J. Flynn, J. P. Nicolai, B. C. Hamel, S. Traill, D. Zaleske, H. J. Mankin, H. Yeon, C. Ho, and *et al.* 1994. A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. *Nat Genet* 6: 282-6.

Vogel, A., C. Rodriguez, and J. Izpisua-Belmonte. 1996. Involvement of FGF-8 in initiation, outgrowth and patterning of the vertebrate limb. *Development* 122: 1737-50.

Vogel, A., C. Rodriguez, W. Warnken, and J. C. Izpisua Belmonte. 1995. Dorsal cell fate specified by chick Lmx1 during vertebrate limb development. *Nature* 378: 716-20.

Vogt, T. F., and P. Leder. 1996. Polydactyly in the Strong's luxoid mouse is suppressed by limb deformity alleles. *Dev Genet* 19: 33-42.

Webster, W. S., A. H. Lipson, and P. D. Brown-Woodman. 1987. Uterine trauma and limb defects. *Teratology* 35: 253-60.

Wolpert, L. 1996. One hundred years of positional information. *Trends Genet* 12: 359-64.

Wolpert, L., I. Macpherson, and I. Todd. 1969. Positional information and the spatial pattern of cellular differentiation. Cell spreading and cell movement: an active or a passive process? *J Theor Biol* 25: 1-47.

Woychik, R. P., T. A. Stewart, L. G. Davis, P. D'Eustachio, and P. Leder. 1985. An inherited limb deformity created by insertional mutagenesis in a transgenic mouse. *Nature* 318: 36-40.

Yokouchi, Y., J. Sakiyama, T. Kameda, H. Iba, A. Suzuki, N. Ueno, and A. Kuroiwa. 1996. BMP-2/-4 mediate programmed cell death in chicken limb buds. *Development* 122: 3725-34.

Zakany, J., and D. Duboule. 1996. Synpolydactyly in mice with a targeted deficiency in the HoxD complex. *Nature* 384: 69-71.

Zeng, X., J. A. Goetz, L. M. Suber, W. J. Scott, Jr., C. M. Schreiner, and D. J. Robbins. 2001. A freely diffusible form of Sonic hedgehog mediates long-range signalling. *Nature* 411: 716-20.

Zuniga, A., A. P. Haramis, A. P. McMahon, and R. Zeller. 1999. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* 401: 598-602.

Chapter 2

Introduction into computer cloning
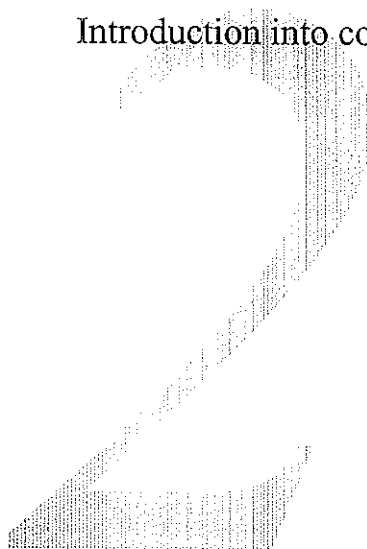
Computer cloning

Positional cloning

The most frequently used strategy for the identification of disease-causing genes in humans is positional cloning. In this procedure, the chromosomal localisation of the disease is determined first, and then the genes located in this region are characterised and used for mutation analysis in patients. The chromosomal localisation of the disease may be identified in several ways. Some diseases are associated with gene defects caused by chromosomal abnormalities, such as translocations, inversions and duplications, which can be observed under a microscope. Unfortunately, only a minority of cases show such obvious abnormalities. In most cases, the genomic localisation of the disease is determined by linkage analysis. This method is based on the rationale that all affected family members must share a genomic region which contains the mutated gene. This disease region will contain a number of polymorphic markers which co-segregate with the disease, and the identification of these markers allows the researchers to locate the disease region in the genome. The exact procedure of linkage analysis falls outside the scope of this thesis; excellent reviews can be found in most textbooks on human genetics (e.g. Human Molecular Genetics, Strachan and Read 1999). After the disease gene has been assigned to a chromosomal region, a physical clone-based map of this region is generated and several approaches are used to identify candidate genes within that region. Subsequently, the patients are analysed for mutations in these genes.

The identification of genes within the critical region is a laborious and time consuming step in the positional cloning procedure. Fortunately, for human disease-causing genes, this era has largely come to an end. In 2001, the draft sequence of the human genome was published (McPherson *et al.* 2001, Venter *et al.* 2001). The sequence was published by two different groups. The first group is the Human Genome Consortium (HCG), an international collaboration of 19 academic institutions that intended to determine the complete sequence of the human genome and to analyse this sequence. The aim of this project is to make the information available to scientists so that the sequence may be used for the public good. A second sequencing effort was undertaken by the private company

Celera Genomics, which only allowed full access to the data to subscribers of their database. The competition between the public and private sequencing efforts did greatly sped up the public sequencing, so that the draft sequence could be published already. It is expected that the HCG sequence will be finished by 2003, by which time all the remaining gaps should be filled and the sequence correctly ordered. The full annotation of the genome, with all the genes, single nucleotide polymorphisms (SNPs) and regulatory elements, will probably take much more time.

The human genome sequence can be used directly to identify genes in a candidate region, speeding up the positional cloning procedure. Because gene identification is not a time consuming step anymore, much larger critical regions can be used, usually obtained when affected families are relatively small. The identification of many genes in a region increases the need for selection of the most likely candidate gene, based on knowledge of the corresponding protein. The resulting approach is called the positional candidate approach.

In the following paragraphs, *in silico* methods for identification of candidate genes in a critical region are discussed, as well as procedures to identify the most likely candidate gene.

In silico cloning
The identification of candidate genes by computer relies on gathering as much information as possible. The final aim of the Human Genome Project is to provide a completely annotated version of the human genome, containing polymorphic markers, genes, predicted genes, single nucleotide polymorphisms and other data that may give information on the function of the sequence. In annotated genomic sequence, candidate genes for diseases are easily identified, and the best candidate can be chosen based on information available about these genes. Unfortunately, at this moment only a fraction of the human genome sequence is reliably annotated. This means that some genes have not been identified and others are not positioned correctly. Therefore, a number of programs are available to assist in annotation, allowing researchers to extract more information from the genomic sequence.

Two kinds of annotation programs exist: de novo gene prediction and sequence alignment programs. The gene prediction programs make use of the structural features of

genes to predict genes in a sequence. The sequence alignment programs compare the genomic sequence to sequences of known genes and transcripts. When sequences from the transcript databases are identical to sequences in the genomic region, the corresponding genes are located in this genomic region. With the help of these programs, a more or less accurate map of the region can be established, containing all the genes and possible genes in a candidate region.

Homology searches: FASTA and BLAST

Sequence alignment is necessary for the identification of candidate genes in critical regions, and for finding proteins that are homologous to predicted gene products (see below). Finding similar sequences in the available databases relies on alignment programs that compare a query sequence to all sequences in the selected database. The two programs most widely used are BLAST and FASTA.

FASTA (Pearson and Lipman 1988) is a global alignment tool which finds the best alignment over the entire length of the query sequence. When sequences are not completely identical, the program introduces gaps. This approach is useful for finding sequences with a high degree of sequence similarity.

BLAST (Basic Local Alignment Search Tool) (Altschul *et al.* 1997) is a local alignment tool, which finds the optimal alignment between subregions of the query sequence. When it encounters a region that does not fit, it stops the search instead of introducing big gaps. This means that it will often find several alignments to the same sequence. In contrast to the FASTA program, BLAST will find protein domains, short sequences of high similarity (see below).

All sequences in a database are compared to the query, and a score is assigned for each alignment (see Box 1). This score is positive for identical matches and negative for mismatches or gaps. In protein alignments, some amino acid substitutions are 'similar', such as leucine and isoleucine, and such matches receive an intermediate score. The total score of an alignment is calculated by adding the scores for all aligned pairs of residues and the highest scoring sequence are presented to the user.

Box 1. Path graphs and optimisation

FASTA and BLAST operate by finding the optimal 'path' through a graph where the axes are formed by the two sequences being compared. In this path graph, each co-ordinate represents an alignment between a residue from each sequence. In the plot, matching residues from both sequences are represented by a dot (fig. 1). A stretch of similarity can be seen as a diagonal line in the graph, and gaps and insertions are represented as horizontal or vertical lines. The difference between a global and a local alignment program is that the first will allow the alignment score to become negative during the search for an optimal alignment, while the latter does not. In effect this means that a local alignment program will stop searching when it encounters a region of low similarity whereas a global alignment program will try to incorporate it, lowering the total similarity score.In theory it is possible to consider all possible alignments of all sequences, but in practice this takes far too much time. The alignment programs have therefore been optimised for speed. This means that they have lost some sensitivity, but in effect they miss only weak similarities. The main restriction that both programs use is the k-tuple or word size: an alignment will only be considered if at least n subsequent residues in two sequences align. In FASTA the default k-tuple is 4 for DNA sequences, in BLAST it is 11. The FASTA program searches for the largest alignment within the window, as described before, the BLAST program extends the alignment in both directions to find the highest scoring segment pair or hsp. To reduce the number of extensions, the program has been refined to extend a hit only when another one is found on the same diagonal in close proximity.

Sequence databases

The BLAST and FASTA programs compare a protein or nucleotide sequence to the query sequence. These sequences are stored in several databases, and usually a BLAST or FASTA search is performed with a query sequence against one of these databases. An enormous amount of DNA and protein sequence has been generated in the last ten years, and the growth is still exponential. Some of it is plain sequence, generated in genome
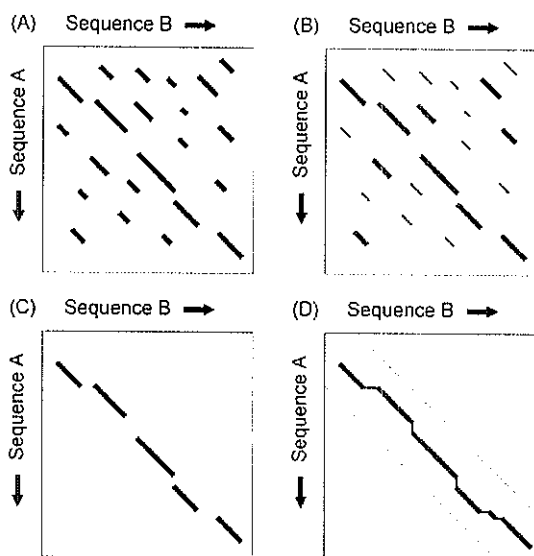
Figure 1. The use of path graphs: FASTA. **A.** During the first step, all regions of identity are identified. **B.** A score is assigned to each alignment, based on the level of identity. The best scoring regions are retained. **C.** The best scoring regions are joined if their combined score is higher than a set threshold level. **D.** The optimised alignment is calculated. Dotted lines indicate the boundaries set by the local alignment programs.

sequencing experiments or during the characterisation of cDNA libraries. Other sequences are annotated: the coding sequences are indicated and sometimes experimental data is attached. To cope with the vast amount of data, many databases have been set up.

Three databases aim for completeness in storing DNA and protein sequences: Genbank, run by the National Center for Biotechnology Information (NCBI) at the National Institute of Health (NIH), the DNA Data Bank of Japan (DDBJ), and the European Molecular Biology Laboratory (EMBL) database maintained by the European Bioinformatics Institute (EBI) in England. These three databases are based on nucleotide sequences; protein sequences are derived from the DNA coding sequences (see Box 2). They exchange information daily and every researcher can submit their data to each of these databases, knowing it will be available everywhere.

Because every researcher can submit their data to these databases, similar sequences are often submitted by different groups. The resulting redundancy is not necessarily a bad thing: it provides a quality control for individual sequences, and tags for

mapping of genomic sequence. However, in similarity searches, redundancy may result in numerous similar hits, which may then obscure weaker hits. To avoid this, non-redundant databases were constructed in which sequences are stored only once. In the DNA non-redundant database, only complete transcripts and finished genomic sequences are stored. Repetitive sequences, ESTs and unfinished genomic sequences are all contained in separate databases.

Box 2. Protein sequence databases.

The main protein databases are SWISS-PROT/ TrEMBL and PIR. SWISS-PROT maintains a high standard of annotation and redundancy is kept to a minimum. Because this delays the admission of protein sequences to the database, TrEMBL was created. This database contains the translation of all the sequences in the EMBL nucleotide sequence database, except the ones that are already available in SWISS-PROT. Therefore, these two databases are complementary. Eventually, most of the sequences in TrEMBL will be curated and stored in SWISS-PROT, but at the moment, TrEMBL contains twice as many entries as SWISS-PROT.

The Protein Information Resource (PIR) Protein Sequence Database provides extensive classification of its proteins. The classification into families and superfamilies is based on global sequence similarity (the homeomorphic family) or local sequence similarity (the domain family). Similar to the SWISS-PROT/TrEMBL division, the PIR database is divided in several sections to reduce redundancy.

Finding unknown genes by sequence alignment

While comparing translated genomic sequence to protein databases will reveal known genes, there is still a large possibility that unknown genes are situated in the candidate region. An important means of finding these is by comparing the genomic sequence to the Expressed Sequence Tag (EST) databases. ESTs are derived from mRNA, and therefore they represent the expressed genes in the tissue or cell type from which this mRNA was extracted. Since not every human gene is known, some ESTs represent unknown genes, and a BLAST hit with such an EST will therefore indicate a novel gene in the critical region.

Because ESTs do not contain introns, BLAST hits are usually gapped; a typical hit consists of 100-150 nucleotides (the average exon size in humans), a gap, another hit and so on. Since the EST database is highly redundant, several similar ESTs are usually found in a search. Sometimes hundreds of ESTs are found, if a gene is highly expressed in the tissues most often used for mRNA extraction, as is the case for some housekeeping genes. Not every EST is derived from a gene; about 10% of ESTs are unreliable due to intron retention, chimaeric clones or the incorporation of untranslated parts of genomic DNA (Hillier *et al.* 1996, Wolfsberg and Landsman 1997). Besides, ESTs are mostly derived from the 3' end of genes, which results in a bias in transcript representation. However, when used with caution, the EST databases represent a wealth of information on both well known and uncharacterised genes.

Gene prediction programs
A different approach for identification of genes in genomic sequence is gene prediction. Prediction programs rely on the common properties of genes, signature sequences that distinguish coding from noncoding sequences. One of the common properties of gene encoding regions is that they contain splice sites: GT at the 5' of an intron and AG at the 3' end (Mount 1982). Apart from the first and last exon of a gene, which do not have a 5' or 3' splice site, respectively, every exon lies between these splice sites. Another important feature of genes is that they mostly encode proteins. Coding exons must therefore contain at least one uninterrupted open reading frame (ORF), and a combination of coding exons must have a continuous ORF, without frameshifts. For a gene prediction program, the definition of an exon is therefore 'an open reading frame between two splice sites'. This means that programs are usually less suitable for finding first and last exons, as well as noncoding exons.

A third characteristic of genes is codon bias: although four different codons encode valine, in humans, GTG is found in 48% of cases and GTA is used in only 10% (http://bioinformatics.weizmann.ac.il/databases/codon/hum.cod). Gene prediction programs select for exons that use common codons: if too many infrequent codons are used, the presumptive exon is discarded.

Box 3. Nix analysis

The Nix program runs several sequence alignment and gene prediction programs and combines the results. The output is standardised into an interactive picture in which each coloured block is a link to the original data. The output of a NIX search is shown in *fig.* 2. Every hit is represented by a box, and combined hits, such as two exons connected by a gene prediction program, are linked by a horizontal bar. Colour intensity depicts the strength of the hit or the prediction and similar programs are grouped. The genomic sequence is represented by a bar in the centre of the picture. Blocks above this bar are the hits of the various programs on the forward strand, and below are the hits on the reverse strand.

Before sending the sequence to the various BLAST programs, low complexity repeats and the ALU, LINE and other repeats of which over 40% of the human genome consists, are masked by the RepeatMasker program (Smit, AFA & Green, P RepeatMasker at http://ftp.genome.washington. edu/RM/ RepeatMasker.html). This prevents numerous insignificant hits to repetitive sequences in the databases. NIX shows the position of these repeats in light blue. To ensure that no alien DNA contaminates the genomic sequence, the BLAST program is used for comparing the genomic sequence to databases containing vector and E. coli sequences. Other databases are screened for Sequence Tagged Sites (unique sequences used for mapping the genome), Genomic Survey Sequences (short sequences used for identification of clones in the human genome project) and High Throughput Sequences (unfinished human genome). For gene identification, the BLAST results are shown directly below the output of the gene prediction programs. Four databases are screened: the non redundant EMBL nucleotide database, the EST database and the SWISS-PROT and trEMBL databases. For the last two, the genomic sequence is translated in 6 reading frames.

The exon predictions are shown separate from the gene prediction results, because some exons are skipped in favour of others during gene prediction. The gene prediction favoured by a program may not always be the right prediction, therefore all the predicted exons are shown. It is obvious how well the prediction programs work: most of the BLAST hits from the SWISS-PROT and EMBL databases have a predicted counterpart. This is less obvious for the EST hits, mostly because of the genomic contamination of the EST libraries and because ESTs are biased towards 3' ends of coding regions, which are often untranslated and will not be predicted. In this case, prediction programs and EST data may be complementary.

NIX also runs a search for CpG islands, several polyadenylation site programs and algorithms that look for promoter-specific structures such as TATA boxes and binding sites of transcription factors. Together, these programs will find most of the genes localised in a critical region.

Figure 2. Nix analysis.

Gene prediction programs usually consist of an algorithm that predicts exons and a separate algorithm that connects the predicted exons so that a continuous open reading frame is formed. The latter may take other features into account, such as average exon and intron size and polyadenylation signals, which are found at the end of most genes. GC rich 'islands' are found near the start site of translation of housekeeping genes and can be used as tags for identification: it is likely that a coding sequence starts nearby. However, when too many general gene properties are taken into account, the selectivity of the program rises, but the sensitivity may decline, because genes with less common characteristics will be missed. Besides, most programs have been optimised for one species: intron and exon sizes as well as codon bias differ considerably between species.

Gene predictions in a genomic sequence are expected to overlap EST and mRNA hits. Several internet resources aim to combine the results of all the gene searching programs in a visual output. One of these is NIX (Nucleotide Identification of unknown sequences), accessible through the UK Human Genome Mapping Project Resource Centre (HGMP)(see Box 3). Eventually, all the human genome data will be annotated and programs like NIX will become obsolete. However, as long as genome sequencing efforts are undertaken, the parameters for gene identification programs may be optimised for different species: NIX may be adapted for use in plants or animals.

Function of candidate genes.
In principle, every gene in the critical region is a candidate gene for causing the disease. However, when the region is very large or when the gene density is high, many genes will be found. Mutation analysis is a time consuming process, and therefore the most likely candidate gene must be selected for mutation analysis first. This means that as much information as possible must be gathered for each candidate gene.

For known genes, the scientific literature will usually give some insight in function and site of expression. Another important source of information is the Online Mendelian Inheritance in Man (OMIM) database, which provides information on human inherited disorders, their chromosomal localisation and implicated genes. When one of the genes in the critical region is implicated in an unrelated disease, this gene is less likely to cause the disorder under study. However, genes can never be ruled out completely based on such information.

The function of predicted gene products may be deduced from their structure or similarity to known proteins. Many databases are set up specifically for comparison of a query sequence to the sequences stored in the databases, allowing researchers to find even weak similarities to proteins of which the function is known.

All-over sequence similarity of proteins is a good indication that the proteins are homologous, i.e. share a common ancestry. A further distinction can be made between paralogs, which share their last common ancestor at the time of a gene duplication, and orthologs, which share their last common ancestor at the time of speciation. Distinguishing between orthologs and paralogs is useful since orthologs have a much higher chance of being functionally related. It is difficult to determine what level of similarity between proteins indicates that they are homologous. In general, closely related species should have a high level of similarity while distantly related species may show low levels. This makes the identification of homologs in distantly related species more difficult.

Sometimes proteins share blocks of similarity, interrupted by regions of no homology. Such related sequences are probably functional domains: regions of sequence conservation that provide functions such as binding to proteins or DNA (see Box 4). Many different protein domains have been identified. Sometimes their consensus amino acids are spread over a few hundred amino acids. An example is the RING domain, which consists of 7 cysteines and 1 histidine. The spacing of these residues is more or less conserved, but the amino acids between the consensus residues differ widely. The function of many protein domains are known, and therefore the presence of such a domain in a predicted protein can give some indication about its function.

The three-dimensional structure of a protein is the key to its function. Although the exact conformation of a protein cannot be predicted from its constitutive amino acids, the conformation of some of its regions may be reconstructed. Polypeptide chains usually fold so that hydrophobic side chains are buried and the polar, charged chains are on the outside of the protein, extending into the aqueous environment. Longer stretches of hydrophobic residues may be transmembrane regions, indicating that the protein functions as a (co)receptor. Other amino acids conformations are the α helix and the β pleated sheet

Box 4. Searching for functional domains

Several domain databases are available for similarity searches. Examples are the BLOCKS and Pfam databases, searchable at the Swiss Institute for Bioinformatics (http://www. isrec. isb-sib.ch/software/PFSCAN_form.html). This server also contains a pattern search tool which allows the user to search protein databases with a self-made consensus sequence.

A novel domain may be found with position-specific iterated BLAST, or PSI-BLAST, available at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/). This program performs a normal BLAST against a protein database and then uses the output alignment for finding a consensus sequence. This 'profile' is used for a next BLAST round, after which it is refined. The second BLAST step is called an 'iteration' and can be repeated until no new hits are found. The resulting profile may represent a new functional domain, or a distantly related protein family. However, PSI-BLAST is error-prone, because inclusion of a protein that is part of a conserved family will often result in retrieval of the whole family, shifting the profile (Altschul and Koonin 1998).

(see Box 5). Specific combinations of these structures provide certain functions which may indicate the function of a protein.

Several other features may be predicted from the protein sequence. Examples are consensus sites for glycosylation and phosphorylation, which are possible regulatory sites for protein activity (see Chapter 1). Signal sequences to target the protein to the nucleus or the extracellular matrix give insight to the site of action of the protein, and so do consensus sites for the attachment of lipid anchors, which direct the protein the cell membrane.

The time and site of expression of a gene will give another indication to its function. For candidate gene selection, genes that are highly expressed in affected tissues are interesting candidates. The predominant site of expression may be inferred from the source of the ESTs derived from the gene: if all ESTs are descended from a brain library, the corresponding gene is expressed in brain. In this way, 'electronic Northern blots' may

be compiled, indicating the expression level per tissue. However, caution must be taken in this procedure, since some tissues, like fetal brain, are overrepresented in the EST databases.

Box 5. Three-dimensional structure comparison

In general, proteins may contain one or both of two general periodic structural motifs, the α helix and the β pleated sheet. The α helix is a rod-like structure stabilised by hydrogen bonds in which the polypeptide chain is tightly coiled, the β pleated sheet is an extended conformation consisting of adjacent polypeptide chains, which can either run parallel or antiparallel. Combinations of helices and sheets result in protein 'folds'. In fact, just ten different 'superfolds' account for 50% of the known structural similarities between protein superfamilies (Orengo *et al.* 1994). The formation of α helices and β sheets in a protein can be predicted because amino acid residues show a preference for one of these conformations, and the restricted number of possible protein folds has allowed the development of 'protein threading': prediction of the three-dimensional structure of a protein based on its sequence. However, currently the predicted 3-D structure is correct in only half of the cases, partially because the prediction of the α helices and β sheets agrees with the solved protein structure only in about 60% of the cases (Stryer 1995).

The three-dimensional structure of a protein may also be based on sequence similarities between the predicted protein and a protein of which the structure has been elucidated by nuclear magnetic resonance (NMR) or X-ray crystallography. The Protein Data Bank (PDB) contains 3-D structures.

Together, these searches will result in a general image of the function of a predicted gene. In most cases, it will not result in the complete exclusion of the gene as a candidate for the disease, but the knowledge gained from the databases will probably allow the researcher to make a priority list of candidate genes, which can be screened one by one for mutations. In general, computer cloning and candidate gene analysis will greatly speed up the positional cloning approach and allow for identification of disease causing genes in large candidate regions.

61

References

Altschul, S. F., and E. V. Koonin. 1998. Iterated profile searches with PSI-BLAST--a tool for discovery in protein databases. *Trends Biochem Sci* 23: 444-7.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-402.

Hillier, L. D., G. Lennon, M. Becker, M. F. Bonaldo, B. Chiapelli, S. Chissoe, N. Dietrich, T. DuBuque, A. Favello, W. Gish, M. Hawkins, M. Hultman, T. Kucaba, M. Lacy, M. Le, N. Le, E. Mardis, B. Moore, M. Morris, J. Parsons, C. Prange, L. Rifkin, T. Rohlfing, K. Schellenberg, M. Marra, and *et al.* 1996. Generation and analysis of 280,000 human expressed sequence tags. *Genome Res* 6: 807-28.

McPherson, J. D., M. Marra, L. Hillier, R. H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E. R. Mardis, R. K. Wilson, R. Fulton, T. A. Kucaba, C. Wagner-McPherson, W. B. Barbazuk, S. G. Gregory, S. J. Humphray, L. French, R. S. Evans, G. Bethel, A. Whittaker, J. L. Holden, O. T. McCann, A. Dunham, C. Soderlund, C. E. Scott, D. R. Bentley, G. Schuler, H. C. Chen, W. Jang, E. D. Green, J. R. Idol, V. V. Maduro, K. T. Montgomery, E. Lee, A. Miller, S. Emerling, Kucherlapati, R. Gibbs, S. Scherer, J. H. Gorrell, *et al.* 2001. A physical map of the human genome. *Nature* 409: 934-41.

Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res* 10: 459-72.

Orengo, C. A., D. T. Jones, and J. M. Thornton. 1994. Protein superfamilies and domain superfolds. *Nature* 372: 631-4.

Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444-8.

Strachan, T., and A. P. Read. 1999. *Human Molecular Genetics*. BIOS Scientific Publishers, Oxford.

Stryer, L. 1995. *Biochemistry*. W H Freeman & Co., New York.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. *et al.* 2001. The sequence of the human genome. *Science* 291: 1304-51.

Wolfsberg, T. G., and D. Landsman. 1997. A comparison of expressed sequence tags (ESTs) to human genomic sequences. *Nucleic Acids Res* 25: 1626-32.

Chapter 3

An insertional mutation in Fibrillin 2 leads to syndactyly and other skeletal abnormalities in the *sy* allele *tipsy*.

Marijke J. van Baren[1], Henk Heus[1], Guido J. Breedveld[1], Herma C. van der Linde[1], An Langeveld[2], Andreas Rump[3], Marcel Vermeij[4], Tiansen Li[5], Esther de Graaff, Ben A. Oostra[1] and Peter Heutink[1].

[1]Department of Clinical Genetics and [2]Department of Cell Biology and Genetics and [4]Department of Pathology, Erasmus University Rotterdam, the Netherlands
[3]Institute of Molecular Biotechnology, Department of Genome Analysis, Jena, Germany.
[5]Berman-Gund Laboratory for the Study of Retinal Degenerations, Harvard Medical School, Boston, USA.

Abstract

Tipsy (transgene induced phalange synostosis) is a recessive mouse mutant that shows webbing of the 3rd and 4th digit in combination with other skeletal abnormalities. We have mapped the integration site to mouse chromosome 18, close to the *Shaker-with-syndactylism* (*sy*) locus. *Sy* is a mutant that shows deafness and syndactyly, caused by a deletion of *Slc12a2* and *Fibrillin 2* (*Fbn2*), respectively. We cloned and sequenced the *tipsy* transgene integration site and found that the transgene had integrated in intron 7 of *Fbn2*. This results in a reduced expression of full length transcript during digit formation. Fbn2 is a major structural component of the microfibrils, which are abundant in elastic tissues. In humans, mutations in a specific region of Fbn2 cause dominant Congenital Contractural Arachnodactyly (CCA)(MIM: 121050). Recessive mutations, and mutations outside this region have not been reported thus far, and therefore Fbn2 is an excellent candidate for recessive syndactyly in humans.

Introduction

Limb development has been used extensively as a model for embryonal patterning and development. Indeed, many proteins that function in differentiation and outgrowth of the limb have been shown to function in other parts of the developing embryo as well, such as Sonic Hedgehog and the Fibroblast Growth Factors (for review see Cohn and Tickle 1996). Limb development starts with the formation of a bud along the axis of the developing embryo. Cells at the tip of the limb bud form the apical ectodermal ridge (AER), a thin, tightly packed band of epithelial cells which signals to the underlying mesoderm to promote outgrowth of the limb. Outgrowth progresses in a proximodistal way, whereby the humerus is formed first, the radius and ulna are formed subsequently and lastly the digits (Summerbell et al. 1973). Digits are formed by condensation of mesenchyme. The interdigital mesenchyme undergoes apoptosis, which is mediated by Bone Morphogenic Protein (Bmp) -2 and -7 and possibly -4, expressed in the limb mesoderm (Ganan et al. 1996, Macias et al. 1996, Macias et al. 1997, Yokouchi et al. 1996, Zou and Niswander 1996). Beads soaked in these Bmps accelerate apoptosis when implanted in the interdigital region, and, when implanted in the tip of a growing digit, lead to the formation of an ectopic area of apoptosis. Interestingly, the specification of digit identity is also mediated by Bmps, possibly by the formation of heterodimers between Bmp-2, -4 and −7 (Dahn and Fallon 2000, Drossopoulou et al. 2000). When Bmp-2 is absent, all digits have the same identity, corresponding to the function of Bmp-2 in digit specification (Laufer et al. 1994).

Syndactyly, the webbing of digits, is thought to be caused by failure of apoptosis between the digital rays during limb development (Winter and Tickle 1993). The webbing may involve soft tissues only, or may include the bones. In humans and mice, syndactyly occurs mostly between digits two and three, three and four, or all three of these. Several genes have been implicated in syndactyly. An example is *Dad1*, the defender against apoptotic cell death. Disruption of the gene in mice led to soft-tissue syndactyly in some heterozygous mice. The gene is important for apoptosis: homozygotes die shortly after apoptosis should have begun (Nishii et al. 1999). Mutations in the Fibroblast Growth

Factors (FGFs) and their receptors (FGFRs) result in syndromes of which syndactyly is a characteristic (Muenke et al. 1994, Partanen et al. 1998, Wilkie et al. 1995). The FGFs are involved in mediating apoptosis by regulation of the BMPs (Montero et al. 2001).

Identification of new genes and biochemical pathways in limb development has been aided by the study of mouse limb mutants. One of these is the *Shaker-with-syndactylism (sy)* mouse mutant, which was generated by radiation-induced mutagenesis (MGD). The mouse shows recessive syndactyly in combination with other skeletal abnormalities, as well as deafness and loss of balance, probably caused by inner ear defects. Most mice die shortly after birth. The mutant lacks approximately 0.7 cM of chromosome 18, a part that contains the *Slc12a2* gene, which encodes a Na-K cotransporter (Dixon et al. 1999, Johnson et al. 1998). Initially, only *Slc12a2* was found to be deleted. When this gene was knocked out, mice showed the circling behaviour and deafness of the *sy* mouse, but not the syndactylism, indicating that another gene in the deleted region is responsible for the syndactyly phenotype (Delpire et al. 1999). During the course of this work, this gene was identified as *Fbn2* (Chaudhry et al. 2001), a gene that has been implicated in the human connective tissue disorder Congenital Contractural Arachnodactyly (CCA) (MIM: 121050). Mutations in *Fbn2* were found in two allelic variants of *sy*: fused phalanges (sy^{fp}) and sy^{fp-2i} Both show recessive syndactyly but not deafness or circling behaviour and homozygotes are viable and fertile (MGD).

We report here another syndactyly mutant, the *tipsy* mouse (for *t*ransgene *i*nduced *p*halange *sy*nostosis). The *tipsy (ty)* mutation arose in a line transgenic for human *Rhodopsin* (Li et al. 1996). The mouse shows synostosis of the phalanges of digits 2 and 3 of the hindlimbs, fusion of the most caudal ribs and of the sacral vertebrae. In this report we present the *tipsy* phenotype and show that the transgene disrupted the gene for Fibrillin 2 on chromosome 18, which makes *tipsy* allelic to *sy*. The skeletal abnormalities in *tipsy* have not been reported for the other *sy* phenotypes, indicating that the mutational mechanism in this mutant is different from the phenotypes reported before.

Materials and Methods

Mice

Generation of mice carrying the human *Rhodopsin* P347S mutation has been described by Li *et al.* (Li et al. 1996). Mice carrying the sy[fp] mutation were obtained from the Jackson Laboratory. Mice were maintained on a mixed background including C57BL/6J and FVB.

Skeletal preparations

Mice were killed, skinned and eviscerated. Skeletons were fixed in 100% ethanol for 1 day followed by staining over night with alcian blue and alizarin red (15 mg/l alcian blue, 50 mg/l alizarin red in 80% ethanol/20% acetic acid). Skeletons were cleared in 3% KOH and stored in glycerol.

Microscope

Collected tissues were fixed in 10% buffered formalin, embedded in paraffin blocks, sectioned, and stained using standard methods, including haematoxylin/eosin staining and -BrdU immunohistochemistry. Sections were examined and photographed using a light microscope.

Fluorescent in situ hybridization

Preparation of metaphase chromosomes and fluorescent *in situ* hybridization were done as has been described by Mulder *et al.* (Mulder et al. 1995). A 13.3 kb *Rhodopsin* construct described by Li *et al.* (Li et al. 1996) was used as a probe.

Construction of genomic library

ty/ty mouse genomic DNA was partially digested with Sau3AI. The presence of restriction fragments sized 11-13 kb was confirmed on a 0.4% agarose gel by eletrophoresis. The restriction solution was ligated to lambdagem-11 phage vector (Promega) and packaged into phage particles using GigaPack Gold and GigaPack XL packaging extracts (Clontech). Packaged phages were used to infect *E. coli* cells (strain KW 251, Promega).

Library screening

1×10^6 plaques were plated, transferred to nylon filters (Amersham) and hybridized with a 263 nt PCR product located in the first exon of human *Rhodopsin* (primers rh1f: GGCACAGAAGGCCCTAACT and rh1r: GCTGGTGAAGCCACCTAG) and a 350 nt PCR product located in exon 5 (primers rh3'f: GAGGGCAGGAGCATGAGATA and rh3'r: GTCATCGCCAAGGTCTGATT).

Mapping of genomic phage clones

Phage clones were digested with SalI. The resulting DNA fragments were separated by electrophoresis on a 0.7% agarose gel, transferred to a nylon membrane and hybridized with individual SalI restriction fragments using standard procedures (Sambrook et al. 1989).

Genomic sequencing.

Nebulized fragments of the Lambda-clones were subcloned separately into M13mp18 vector (Yanisch-Perron et al. 1985). About 300 plaques were selected from each sublibrary, M13 DNA prepared and sequenced using dye-terminators, ThermoSequenase (Amersham) and universal M13-primer (MWG-Biotech). The gels were run on ABI-377 sequencers and data were assembled and edited using the GAP4 program (Bonfield et al. 1998). Genomic DNA sequence analysis was performed using the automated sequence annotation system RUMMAGE (Taudien et al. 2000).

cDNA production, PCR and sequencing

RNA was isolated from pooled limbs with RNAzol (Campro Scientific) and 5μg of RNA was used for reverse transcription with Superscript II RT (Gibco BRL) following manufacturer's instructions. PCR amplifications were performed in a total volume of 50 μl containing 2 μl cDNA, 20 mM Tris HCl pH 8.4, 1.5 mM $MgCl_2$, 50 mM KCl, 200 μM dNTPs, 0.5 U Taq polymerase (Gibco BRL) and 10 pmol forward and reverse primers. Cycling conditions were 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at T_{an} (4xG/C+2xA/T-5°C), 90 s at 72°C and a final step of 5 min at 72°C. The following primers were used:

For detection of chimaeric transcripts in *ty* homozygotes: Fbn6: ACTGGGACGACATGGAGAAGA; Rh5: CGTCTTGGACACGGTAGCAGAG; Fbn9: ACTGGGACGACATGGAGAAGA. For detection of *Fbn2* exons in *sy* mutants: Fbn1F:GGTTGTGTCTCCAGCCCTAC;   Fbn1R:CAGAGCGTCTTTGGTGGAG; Fbn9F:GATCAGAGGGCTGGAACATGC; Fbn9R: GGAACCTCTGACAGGACACC; Fbn65F:AGCAAATCAGCCTGGAGAGCG; Fbn65R:TTCCAAACACTTCTGAAGGCCG.

For mutation analysis of *Fbn2*, primersequences are available on request. Sequence analysis was performed on an ABI-377 automated fluorescence dye sequencer using Bigdye chemistry (Perkin Elmer).

Semiquantitative PCR

Ty homozygous, heterozygous and wt cDNA was obtained from 10.5 and 13.5 dpc embryo fore- and hindlimbs as described above. Semiquantitative PCR was performed in 10 ul, using the same conditions as described, but with addition of 0.1μl ^{32}P dCTP (0.37 MBq/μl) and limb cDNA as template. The following primers were used: actinF: AGGACAGGCCCGTGTTTCA, actinR: GGAACCTCTGACAGGACAGG, Fn6: (see before), Fbn9: ACTGGGACGACATGGAGAAGA and Rh3: ACGAGCTGCCCATAGCAGAA.

Samples were withdrawn every second cycle, starting at 10 cycles. After PCR, 4 μl loading buffer for denaturing gels was added to each sample and 2 μl of the resulting mix was run on a 4% non denaturing acrylamide gel (Amresco). A Phosphorimager (Molecular Dynamics) was used to obtain a digital image of the gels. The relative amount of products was determined with ImageQuant software (version 3.3, Molecular Dynamics).

For each PCR, the log phase of the reaction was determined first to ensure optimal amplification of the products. For actin, the log phase was between 14 and 30 cycles, for Fbn6/Fbn9 it was between 24 and 30 cycles, and for Fbn6/Rh3 no log phase was reached.

To obtain equalized amounts of cDNA in each limb sample, a dilution series was made for determination of actin levels at cycle 22. Subsequently, samples were diluted to

equal actin concentrations. For Fbn6/Fbn9, product levels were determined using the mean value of three points in the log phase (24-28 cycles), and these were standardized by the internal actin control.

Results

Description of the *tipsy* phenotype

The *tipsy (ty)* mouse was generated during *Rhodopsin* transgenesis. A construct containing 13.3 kb of human genomic DNA containing all five *Rhodopsin* exons and promoter sequences had been used for transgenesis (Li et al. 1996). One of five mouse lines showed syndactyly and we named this phenotype *tipsy, for transgene induced phalange synostosis.*

The homozygous *ty* mutant shows syndactyly of the 2nd, 3rd and sometimes 4th digit, varying between soft tissue syndactyly to partial or complete synostosis of the phalanges. In combination with syndactyly, the mouse shows other skeletal abnormalities: in the hindlimbs, an amorphous bone is attached to the calcaneum. The long bones of fore- and hindlimbs, the scapula- and pelvic girdle bones are normal in overall size and shape, but appear very thin and fragile. The distal ends of the most caudal ribs are fused and the sacral vertebrae are abnormal in shape (fig. 1). In addition, some mutants have a kinky tail. Microscopic evaluation showed no gross abnormalities of the inner organs.

Both heterozygous and homozygous *ty* mice are viable and fertile. Heterozygotes show no abnormalities, indicating that the phenotype inherits as a recessive trait.

Chromosomal localization of *ty*

Rhodopsin is a light sensing protein that functions exclusively in the retina. Because Rhodopsin does not have a function in limb development, we hypothesized that in *ty* mice the *Rhodopsin* construct had integrated in or near a gene involved in limb development, thereby disrupting it. In order to find genes near the integration site, we determined the chromosomal localization of the transgene insertion by fluorescence *in situ* hybridization. The 13.3 kb human *Rhodopsin* construct was used as a probe on metaphase chromosomes
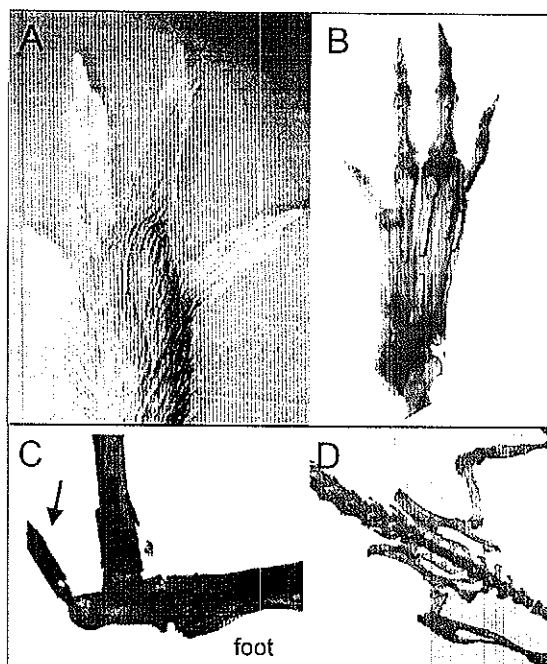
Figure 1. *Tipsy* phenotype. **A.** Right hindlimb, dorsal view. **B.** Bone staining of right hindlimb. **C.** An extra bone is apparent in the hindlimb. **D.** Abnormally shaped vertebrae.
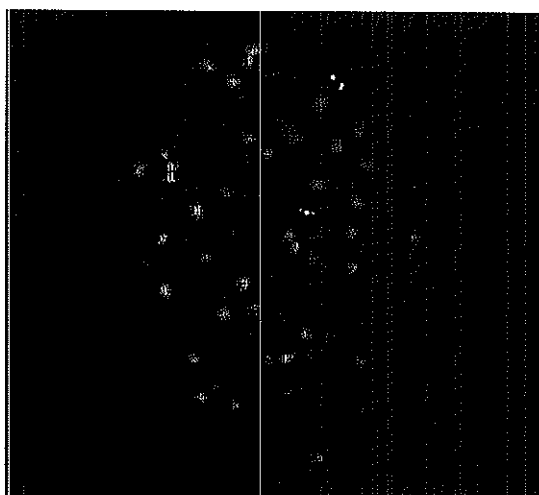

Figure 2. Fluorescence *in situ* hybridization on *ty* homozygous chromosomes. The transgene has inserted into chromosome 18qE1.

of a homozygous *ty* mutant. The probe hybridized to chromosome 18qE1 (fig. 2), indicating that the transgene had integrated there. This was confirmed by co-hybridization with the 92/A probe, located on 18qE1 (R. Delwel and N.G. Copeland, personal communication).

The 18qE1 region did not contain any genes known to be involved in limb formation, and therefore no obvious candidate genes for disruption were available. However, the mutation *Shaker-with syndactylism* (*sy*) and its allelic phenotypes *sy^{fp}* and *sy^{fp-2j}* had been mapped to this region (MGD). All three mutants show recessive syndactyly very similar to *ty*, indicating that these phenotypes may be allelic with *ty*. No candidate genes were known for the *sy* phenotypes, therefore we set out to clone the insertion site and identify the gene disrupted in *ty*.

Map construction
A genomic library of a homozygous *ty* mutant was constructed in phage vectors and screened with two human *Rhodopsin* probes, one at each side of the transgene construct. Positive clones were isolated and sizes were estimated on agarose gel. Two of the largest phages, one at each side of the integration point, were chosen for sequencing (fig. 3). On one of these clones, phage f65, two exons of *Fbn2* were identified: exon 6 and 7. These exons lie in the same orientation as the transgene (data not shown). Phage f11 did not contain any exons.
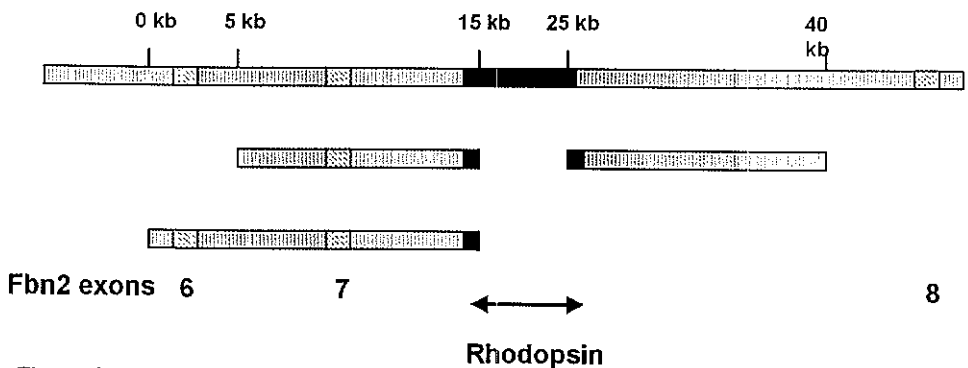

Figure 3. Map of insertion region. Phage clones f11 and f65 were sequenced. Two exons of *Fbn2* were found in phage 11.

Transcription of a Fbn-Rhodopsin chimaeric RNA in *ty* mice

Because the *Rhodopsin* construct had integrated in the same orientation as *Fbn2*, a transcript initiated at *Fbn2* exon 1 will contain *Rhodopsin* exons after exon 7 of *Fbn2*. During the splicing reactions, the alien exons will either be removed or incorporated into the mature mRNA. The splicing machinery will not recognize *Rhodopsin* exon 1 because it does not have a 5' splice site, but the 5' splice site of exon 2 will be recognized. A polyadenylation signal is present in *Rhodopsin* exon 5, so we speculated that a chimaeric product may arise, containing exon 1-7 of *Fbn2* and 2-5 of *Rhodopsin* (Rh) (fig. 4). To test this hypothesis, we designed primers in exon 6 of *Fbn2* (Fbn6) and exon 5 of *Rhodopsin* (Rh5) and used these for PCR on *ty* embryonic heart cDNA. A band of the expected size was produced (fig. 4). This band was sequenced and indeed contained the predicted chimaeric product. The fusion of *Rhodopsin* with *Fbn2* is in frame, therefore if the mRNA is translated, a true Fbn2/Rh fusion protein will form.

To test if any full length Fbn2 mRNA was still present in the homozygous *ty* mutant, we
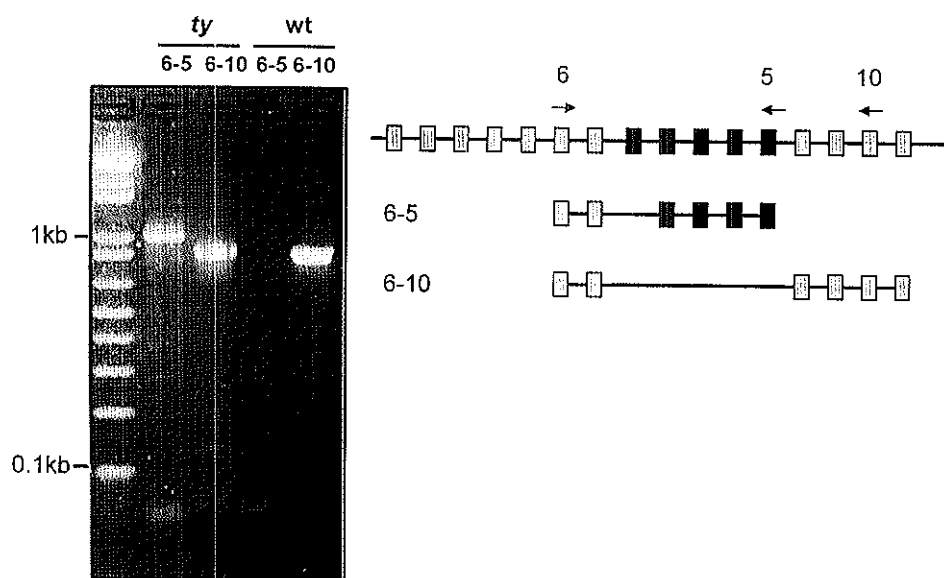
Figure 4. The transgene insertion leads to a chimaeric Rhodopsin-Fbn2 transcript in *ty* homozygotes. Wildtype Fbn2 was also found, indicating that the transgene is spliced out in some transcripts. A schematic representation of primer locations and PCR products is given. *Rhodopsin* exons are depicted in black, Fibrillin exons in grey.

73
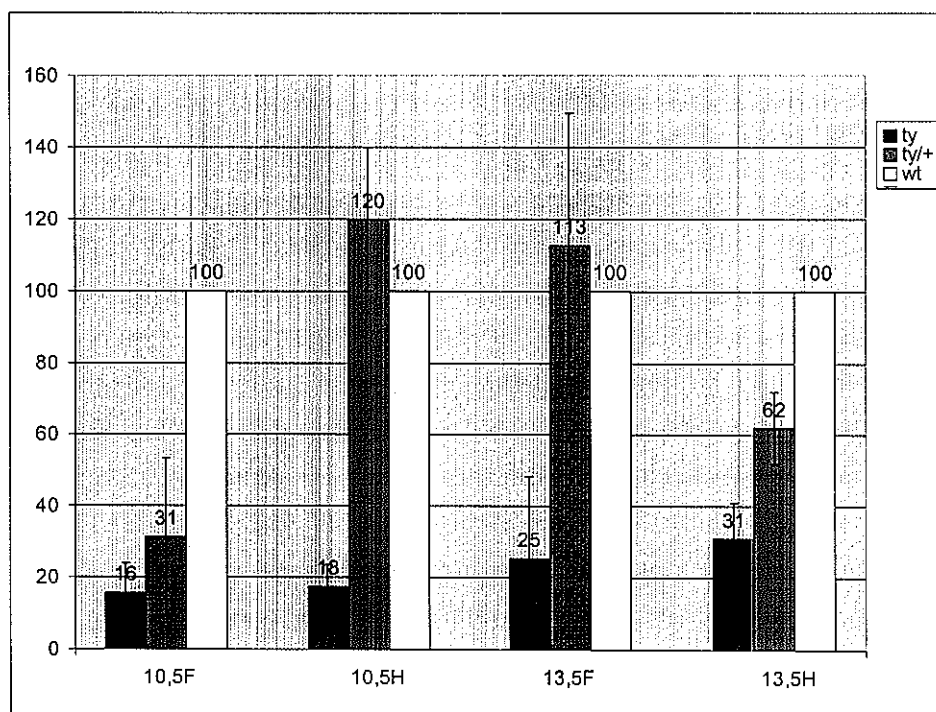
Figure 5. Semiquantitative PCR results. Fbn2 levels were compared in ty/ty, ty/+ and wt fore- and hindlimbs of 10.5 and 13.5 dpc. Fbn2 levels are shown as percentage of wildtype levels.

designed primers in exon 6 and 10 of *Fbn2* and used these for PCR on cDNA derived from both wt and *ty* homozygous embryos. Interestingly, normal Fbn2 was seen in both wt and *ty* homozygous embryos, indicating that the *Rhodopsin* exons are spliced out in some cases (fig. 4). Since *ty* mutants produce some normal *Fbn2* transcript, it is likely that there is some functional Fbn2 protein present in these mutants

Leakage of *Fbn2* transcription in *ty* homozygotes
The *ty* phenotype may be caused by a lack of functional Fbn2 protein or by interference of the chimaeric Fbn-Rh product. In order to gain insight in the amount of chimaeric and normal transcript in *ty* mutants, we performed semiquantitative RT-PCR. *Ty* homozygous, heterozygous and wt cDNA was obtained from 10.5 and 13.5 dpc embryo fore- and hindlimbs. cDNA concentrations were equalized (for details see Material and Methods) and used for PCR with primers in *Fbn2* exons 6 and 9 (Fbn6/Fbn9, product is across the

insertion site) and with primers in *Fbn2* exons 6 and *Rhodopsin* exon 3 (Fbn6/Rh3). Samples were withdrawn every two cycles to determine the log phase of the reaction. The *Fbn2-Rhodopsin* PCR did not reach a proper log phase, indicating that the amounts of chimaeric transcript are very low.

Fbn6/Fbn9 product amounts were determined using the mean value of three samples taken during the log phase (24-28 cycles), and standardized by an internal control (actin). Subsequently the Fbn2 levels in *ty/ty* and *ty/+* were compared to wildtype levels (fig. 5). In *ty* homozygotes, the Fbn6/Fbn9 PCR product represents the 'leakage' of *Fbn2* transcription, i.e. the amount in which the *Rhodopsin* exons are spliced out. In the tested samples, the amount of this product varies from 16% of wildtype levels in forelimbs at 10.5 dpc to31% in hindlimbs at 13.5 dpc, indicating that available Fbn2 is transcript in the mutant is a fraction of normal levels.

The Fbn2 amount in heterozygotes varies from 31% of wildtype levels in 10.5 dpc forelimbs to 120% in 10.5 dpc hindlimbs. This suggests that at some timepoints during limb development, the normal allele may compensate for the mutant allele in heterozygotes.

Fbn2 maps close to *Slc12a2*

The mutations in *Fbn2* in the other *sy* alleles were reported during the course of this work (Chaudhry et al. 2001). Because *Slc12a2* is also deleted in the *sy* mutant, this means that these two genes map close together, within the 0.7cM region deleted in the *sy* mutant. However, in the Mouse Genome Database, these two genes map 3 cM apart. The mapping of Fbn2 was mostly done by one backcross experiment (Goldstein et al. 1994), whereas sy was mapped in 22 backcrosses (MGD). The antiquitin and lamin genes are mapped to the same position as *Fbn2* in the mouse (MGD). By PCR we tested for deletion of these genes in the *sy* mutant, and both genes were still present (data not shown). This indicates that *Fbn2* was mapped inaccurately.

Discussion

In this report, we have shown that mutations in *Fibrillin 2* cause syndactyly in mice. The *tipsy* mutant contains a transgene insertion in intron 7 of *Fibrillin 2*, which leads to a decrease of *Fibrillin 2* expression and the formation of a chimaeric product containing

exons 1-7 of *Fbn2*. The *shaker-with-syndactylism* mutant is allelic with *ty*, and in this mutant, *Fbn2* is deleted completely. This finding is consistent with that of Chaudhry *et al.*, who reported deletion of *Fbn2* in the *sy* mutant and mutations in the *sy* alleles *sy^fp* and *sy^fp-2j* (Chaudhry et al. 2001). However, the *tipsy* phenotype differs from the other *sy* alleles because skeletal abnormalities occur outside the limbs. This indicates that the mutational mechanism in the *tipsy* mutant may be different from the other *sy* mutants.

Fbn2 is located near *Slc12a2*

The *sy* mutant contains a 0.7cM deletion on chromosome 18q, which encompasses the *Slc12a2* and *Fbn2* genes. Deletion of these genes accounts for most of the *sy* phenotype, since a mutation of *Slc12a2* in *sy^ns* mice causes deafness (Dixon et al. 1999) and disruption of *Fbn2* in *ty* mutants leads to syndactyly. However, homozygous *ty* and *sy^fp-2j* mice are viable and fertile, whereas most *sy* mice die shortly after birth (MGD). It is possible that one or more other genes are missing in the *sy* mutant, leading to a reduction in viability. However, no such gene has been found in the *sy* deletion region and we have not found deletion of genes mapping close to Fbn-2. An alternative explanation for the viability of *ty* and *sy^fp-2j* mutants is the presence of some functional Fbn2. It is unlikely that, in the *ty* mutant, the Fbn2-Rhodopsin chimaeric product is able to take over Fbn2 function, since only 7 of 65 *Fbn2* exons remain. Furthermore we suggest that Fbn-2 was mapped incorrectly in the mouse and should be placed near *Slc12a2* on chromosome 18.

Function of Fibrillins

Fibrillins 1 and 2 are structurally related proteins found throughout the connective tissue as major structural components of the 10 nm-diameter microfibrils (Sakai et al. 1986, Zhang et al. 1994). Microfibrils are abundant in skin, blood vessels, perichondrium, tendons, and the ciliary zonules of the eye, and they play an important role in elastic fiber formation (Mecham and Heuser 1991). Elastic fibers consist of an amorphous core, mainly consisting of elastin, surrounded by microfibrils (Cleary and Gibson 1983).

The Fibrillins are highly similar multidomain proteins consisting of five structurally distinct regions: A-E. Region C differs the most between the two proteins: it is glycine rich in Fbn1 and proline rich in Fbn2. Both proteins contain 47 EGF domains, of

which 43 are calcium binding (cbEGFs). The cbEGFs are thought to stabilize the protein conformation when calcium is bound (Reinhardt et al. 1997).

In humans, mutations in Fbn-1 and –2 cause the autosomal dominant connective tissue disorders Marfan syndrome and Congenital Contractural Arachnodactyly (CCA), respectively (Dietz et al. 1991, Putnam et al. 1995)(MIM: 154700 and 121050). Although the disorders are clinically distinct, they share the skeletal feature of long bone overgrowth. Marfan syndrome is more common than CCA, and is characterized by severe cardiovascular and ocular abnormalities. Patients have a high risk of dying from sudden disruption of the aorta. Mutations in Marfan are found throughout Fbn-1 with no clear genotype-phenotype relationship except that mutations in exons 24-32 cause neonatal Marfan syndrome, a severe form of the disease leading to premature death (Booms et al. 1999, Godfrey et al. 1995) Mutations occur mostly in the amino acids which constitute the calcium binding EGF domains. These mutations are thought to destabilize Fibrillin conformation (Robinson and Godfrey 2000).

CCA is a milder and less common phenotype, characterized by crumpled ears, joint contractures and minor cardiovascular abnormalities. The long bone overgrowth seen in both disorders are accompanied by long fingers (named arachnodactyly) in CCA. In CCA, only nine different mutations are known, all except one situated in exons 23-34 (Belleh et al. 2000, Park et al. 1998). This region overlaps the neonatal Marfan region in *Fbn1*. All *Fbn2* mutations are either base changes or exon skipping mutations; large deletions of *Fbn2* have not been found. Since deletion of *Fbn2* causes syndactyly in mice, it is likely that deletions or disruptions of *Fbn2* outside exons 23-34 in humans cause a similar, recessive, phenotype.

Fbn2 and syndactyly

How does deletion of *Fbn2* cause syndactyly in mice? *Fbn2* expression precedes that of *Fbn1* during development and it is widely expressed in the mesoderm at positions where morphogenesis is starting (Zhang et al. 1995). It has been suggested that expression of *Fbn2* directs the assembly of elastic fibers during early embryogenesis and that it transduces the forces that drive the morphologic movements associated with gut formation and segmental plate regression (Rongish et al. 1998, Zhang et al. 1995). In quail embryos, Fbn2 forms a cage-like structure around the somites and the notochord at the time of

differentiation (Rongish et al. 1998). Is Fbn2 the executing molecule for the specification of structures? If so, it could form a web around prospective digits, identifying them in response to patterning signals. In the mouse, formation of the digits starts around day 13 dpc, at a stage prior to apoptosis of the interdigital mesenchyme. *Fbn2* is expressed in the limb bud at 10.5 dpc, and *Fbn1* expression becomes abundant at 16.5 dpc, after digit specification (Zhang et al. 1995). A 'caging' function for Fbn-2 could explain the finding of elongated limbs in Marfan and CCA patients: a reduction in functional Fibrillin, as expected from the dominant mutations found in these patients, could lead to less rigid microfibrillar webs around the limb cartilage. The lower physical restriction could result in extreme outgrowth of the limbs.

Elastin-rich microfibrillar networks are present in the developing limb at the moment of digit formation (Hurle et al. 1994). The observed fibers run from the ectoderm towards the chondrogenic core of the limb, and later also to the interdigital spaces. Fibrillin was not associated with this elastic scaffold, but the antibodies used were against Fbn1. It is possible that Fbn2 functions in the establishment of the elastin scaffold in developing limbs.

Since *Fbn2* expression is widespread in the developing embryo, why are so few structures affected? In both *ty* mice and *sy* mice, somites do form, as well as the heart, gut and other *Fbn2* expressing tissues. During normal development, expression of *Fbn1* and *Fbn2* overlap, and it is possible that *Fbn1* is capable of taking over *Fbn2* function. Possibly, *Fbn1* expression is switched on earlier in *Fbn2* mutants, or it is expressed in time to allow for normal growth, except in the limbs and in the most caudal ribs and associated vertebrae.

Is Fibrillin modulated by glycosylation?
The mouse mutant *syndactylism* (*Jag^{sm}*) is characterized by a defect in interdigital cell death, leading to fusion of digits three and four and sometimes one and two of all four feet (MGD). The defect is caused by a missense mutation in the Jagged-2/Serrate gene, which encodes a ligand for the Notch receptor (Sidow et al. 1997). Both the Notch receptor and its ligands are large transmembrane proteins with multiple Epidermal Growth Factor (EGF) like repeats in their extracellular domains.

It has recently been shown that the Notch receptor is O-glycosylated on several of its 36 extracellular EGF domains (Moloney et al. 2000). Two types of sugar are attached: fucose and glucose. Fucose is attached to a serine or threonine between the second and third cysteine of the EGF domain, in the consensus sequence C_2XXGGS/TC_3 (where X represents any amino acid). The O-linked glucose consensus site lies between the first and second conserved cysteines (C_1XSXPC_2). The fucose moieties are modulated by Fringe, which causes Notch to become more sensitive to its ligand Delta and less sensitive to its Serrate/Jagged (Fleming et al. 1997, Klein and Arias 1998, Panin et al. 1997). Notch contains five consensus sites for O-fucose. Intriguingly, Delta and Serrate/Jagged homologs, the two ligands of Notch, contain several of these sites as well, all located between C2 and C3 of their EGF domains. Is it possible that recognition of Notch ligands is aided by the glycosyl modification?

The Fibrillins also contain O-glycosylation sites. Fbn1 contains an O-fucose consensus sequence in its second EGF domain and an O-glucose consensus site in cbEGF domain 9. Fbn2 contains the same glycosylation sites, and an additional O-glucose site in its third cbEGF motif. It is therefore possibile that the Fibrillins are modified by Fringe or another glycosyltransferase, and that they act as signaling molecules during development. At a crucial moment in development, Fbn2 may become glycosylated, or existing sugar residues are modified by Fringe or another glycosyltransferase. This modification may attract Delta or other factors to their receptors in the cell membrane, raising their local concentration. A lack of functional Fbn2 may fail to attract these factors and lead to syndactyly. In this respect it is intriguing that mutations in Jagged cause syndactyly as well.

We have shown that a mutation that leads to reduction of functional Fbn2 in mice leads to syndactyly and other skeletal abnormalities. This adds Fbn2 to the growing list of genes important in embryonal development and indicates a role for Fbn2 in bone formation.

References

Belleh, S., G. Zhou, M. Wang, V. M. Der Kaloustian, R. A. Pagon, and M. Godfrey. 2000. Two novel fibrillin-2 mutations in congenital contractural arachnodactyly. *Am J Med Genet* 92: 7-12.

Bonfield, J. K., C. Rada, and R. Staden. 1998. Automated detection of point mutations using fluorescent sequence trace subtraction. *Nucleic Acids Res* 26: 3404-9.

Booms, P., J. Cisler, K. R. Mathews, M. Godfrey, F. Tiecke, U. C. Kaufmann, U. Vetter, C. Hagemeier, and P. N. Robinson. 1999. Novel exon skipping mutation in the fibrillin-1 gene: two 'hot spots' for the neonatal Marfan syndrome. *Clin Genet* 55: 110-7.

Chaudhry, S. S., J. Gazzard, C. Baldock, J. Dixon, M. J. Rock, G. C. Skinner, K. P. Steel, C. M. Kielty, and M. J. Dixon. 2001. Mutation of the gene encoding fibrillin-2 results in syndactyly in mice. *Hum Mol Genet* 10: 835-43.

Cleary, E. G., and M. A. Gibson. 1983. Elastin-associated microfibrils and microfibrillar proteins. *Int Rev Connect Tissue Res* 10: 97-209.

Cohn, M. J., and C. Tickle. 1996. Limbs: a model for pattern formation within the vertebrate body plan. *Trends Genet* 12: 253-7.

Dahn, R. D., and J. F. Fallon. 2000. Interdigital regulation of digit identity and homeotic transformation by modulated BMP signaling. *Science* 289: 438-41.

Delpire, E., J. Lu, R. England, C. Dull, and T. Thorne. 1999. Deafness and imbalance associated with inactivation of the secretory Na-K-2Cl co-transporter. *Nat Genet* 22: 192-5.

Dietz, H. C., G. R. Cutting, R. E. Pyeritz, C. L. Maslen, L. Y. Sakai, G. M. Corson, E. G. Puffenberger, A. Hamosh, E. J. Nanthakumar, S. M. Curristin, and et al. 1991. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 352: 337-9.

Dixon, M. J., J. Gazzard, S. S. Chaudhry, N. Sampson, B. A. Schulte, and K. P. Steel. 1999. Mutation of the Na-K-Cl co-transporter gene Slc12a2 results in deafness in mice. *Hum Mol Genet* 8: 1579-84.

Drossopoulou, G., K. E. Lewis, J. J. Sanz-Ezquerro, N. Nikbakht, A. P. McMahon, C. Hofmann, and C. Tickle. 2000. A model for anteroposterior patterning of the vertebrate limb based on sequential long- and short-range Shh signalling and Bmp signalling. *Development* 127: 1337-48.

Fleming, R. J., Y. Gu, and N. A. Hukriede. 1997. Serrate-mediated activation of Notch is specifically blocked by the product of the gene fringe in the dorsal compartment of the Drosophila wing imaginal disc. *Development* 124: 2973-81.

Ganan, Y., D. Macias, M. Duterque-Coquillaud, M. A. Ros, and J. M. Hurle. 1996. Role of TGF beta s and BMPs as signals controlling the position of the digits and the areas of interdigital cell death in the developing chick limb autopod. *Development* 122: 2349-57.

Godfrey, M., M. Raghunath, J. Cisler, C. L. Bevins, A. DePaepe, M. Di Rocco, J. Gregoritch, K. Imaizumi, P. Kaplan, Y. Kuroki, and et al. 1995. Abnormal morphology of fibrillin microfibrils in fibroblast cultures from patients with neonatal Marfan syndrome. *Am J Pathol* 146: 1414-21.

Goldstein, C., P. Liaw, S. A. Jimenez, A. M. Buchberg, and L. D. Siracusa. 1994. Of mice and Marfan: genetic linkage analyses of the fibrillin genes, Fbn1 and Fbn2, in the mouse genome. *Mamm Genome* 5: 696-700.

Hurle, J. M., G. Corson, K. Daniels, R. S. Reiter, L. Y. Sakai, and M. Solursh. 1994. Elastin exhibits a distinctive temporal and spatial pattern of distribution in the developing chick limb in association with the establishment of the cartilaginous skeleton. *J Cell Sci* 107: 2623-34.

Johnson, K. R., S. A. Cook, and Q. Y. Zheng. 1998. The original shaker-with-syndactylism mutation (sy) is a contiguous gene deletion syndrome. *Mamm Genome* 9: 889-92.

Klein, T., and A. M. Arias. 1998. Interactions among Delta, Serrate and Fringe modulate Notch activity during Drosophila wing development. *Development* 125: 2951-62.

Laufer, E., C. E. Nelson, R. L. Johnson, B. A. Morgan, and C. Tabin. 1994. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* 79: 993-1003.

Li, T., W. K. Snyder, J. E. Olsson, and T. P. Dryja. 1996. Transgenic mice carrying the dominant rhodopsin mutation P347S: evidence for defective vectorial transport of rhodopsin to the outer segments. *Proc Natl Acad Sci U S A* 93: 14176-81.

Macias, D., Y. Ganan, M. A. Ros, and J. M. Hurle. 1996. In vivo inhibition of programmed cell death by local administration of FGF-2 and FGF-4 in the interdigital areas of the embryonic chick leg bud. *Anat Embryol (Berl)* 193: 533-41.

Macias, D., Y. Ganan, T. K. Sampath, M. E. Piedra, M. A. Ros, and J. M. Hurle. 1997. Role of BMP-2 and OP-1 (BMP-7) in programmed cell death and skeletogenesis during chick limb development. *Development* 124: 1109-17.

Mecham, R. P., and J. E. Heuser. 1991. The elastic fiber. Pages 79-109 in D. Hay, ed. *Cell Biology of the Extracellular Matrix.* Plenum Publishing Co., New York.

MGD. *Mouse Genome Database. http://www.informatics.jax.org/.*

Moloney, D. J., L. H. Shair, F. M. Lu, J. Xia, R. Locke, K. L. Matta, and R. S. Haltiwanger. 2000. Mammalian Notch1 is modified with two unusual forms of O-linked glycosylation found on epidermal growth factor-like modules. *J Biol Chem* 275: 9604-11.

Montero, J. A., Y. Ganan, D. Macias, J. Rodriguez-Leon, J. J. Sanz-Ezquerro, R. Merino, J. Chimal-Monroy, M. A. Nieto, and J. M. Hurle. 2001. Role of FGFs in the control of programmed cell death during limb development. *Development* 128: 2075-84.

Muenke, M., U. Schell, A. Hehr, N. H. Robin, H. W. Losken, A. Schinzel, L. J. Pulleyn, P. Rutland, W. Reardon, S. Malcolm, and et al. 1994. A common mutation in the fibroblast growth factor receptor 1 gene in Pfeiffer syndrome. *Nat Genet* 8: 269-74.

Mulder, M. P., M. Wilke, A. Langeveld, L. G. Wilming, A. Hagemeijer, E. van Drunen, E. C. Zwarthoff, P. H. Riegman, W. H. Deelen, A. M. van den Ouweland, and et al. 1995. Positional mapping of loci in the DiGeorge critical region at chromosome 22q11 using a new marker (D22S183). *Hum Genet* 96: 133-41.

Nishii, K., T. Tsuzuki, T. Kumai, N. Takeda, H. Koga, S. Aizawa, T. Nishimoto, and Y. Shibata. 1999. Abnormalities of developmental cell death in Dad1-deficient mice. *Genes Cells* 4: 243-52.

Panin, V. M., V. Papayannopoulos, R. Wilson, and K. D. Irvine. 1997. Fringe modulates Notch-ligand interactions. *Nature* 387: 908-12.

Park, E. S., E. A. Putnam, D. Chitayat, A. Child, and D. M. Milewicz. 1998. Clustering of FBN2 mutations in patients with congenital contractural arachnodactyly indicates

an important role of the domains encoded by exons 24 through 34 during human development. *Am J Med Genet* 78: 350-5.

Partanen, J., L. Schwartz, and J. Rossant. 1998. Opposite phenotypes of hypomorphic and Y766 phosphorylation site mutations reveal a function for Fgfr1 in anteroposterior patterning of mouse embryos. *Genes Dev* 12: 2332-44.

Putnam, E. A., H. Zhang, F. Ramirez, and D. M. Milewicz. 1995. Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder, congenital contractural arachnodactyly. *Nat Genet* 11: 456-8.

Reinhardt, D. P., D. E. Mechling, B. A. Boswell, D. R. Keene, L. Y. Sakai, and H. P. Bachinger. 1997. Calcium determines the shape of fibrillin. *J Biol Chem* 272: 7368-73.

Robinson, P. N., and M. Godfrey. 2000. The molecular genetics of Marfan syndrome and related microfibrillopathies. *J Med Genet* 37: 9-25.

Rongish, B. J., C. J. Drake, W. S. Argraves, and C. D. Little. 1998. Identification of the developmental marker, JB3-antigen, as fibrillin-2 and its de novo organization into embryonic microfibrous arrays. *Dev Dyn* 212: 461-71.

Sakai, L. Y., D. R. Keene, and E. Engvall. 1986. Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J Cell Biol* 103: 2499-509.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Sidow, A., M. S. Bulotsky, A. W. Kerrebrock, R. T. Bronson, M. J. Daly, M. P. Reeve, T. L. Hawkins, B. W. Birren, R. Jaenisch, and E. S. Lander. 1997. Serrate2 is disrupted in the mouse limb-development mutant syndactylism.*Nature*389:722-5.

Summerbell, D., J. H. Lewis, and L. Wolpert. 1973. Positional information in chick limb morphogenesis. *Nature* 244: 492-6.

Taudien, S., A. Rump, M. Platzer, B. Drescher, R. Schattevoy, G. Gloeckner, M. Dette, C. Baumgart, J. Weber, U. Menzel, and A. Rosenthal. 2000. RUMMAGE--a high-throughput sequence annotation system. *Trends Genet* 16: 519-20.

Wilkie, A. O., S. F. Slaney, M. Oldridge, M. D. Poole, G. J. Ashworth, A. D. Hockley, R. D. Hayward, D. J. David, L. J. Pulleyn, P. Rutland, and et al. 1995. Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome. *Nat Genet* 9: 165-72.

Winter, R. M., and C. Tickle. 1993. Syndactylies and polydactylies: embryological overview and suggested classification. *Eur J Hum Genet* 1: 96-104.

Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33: 103-19.

Yokouchi, Y., J. Sakiyama, T. Kameda, H. Iba, A. Suzuki, N. Ueno, and A. Kuroiwa. 1996. BMP-2/-4 mediate programmed cell death in chicken limb buds. *Development* 122: 3725-34.

Zhang, H., S. D. Apfelroth, W. Hu, E. C. Davis, C. Sanguineti, J. Bonadio, R. P. Mecham, and F. Ramirez. 1994. Structure and expression of fibrillin-2, a novel microfibrillar component preferentially located in elastic matrices. *J Cell Biol* 124: 855-63.

Zhang, H., W. Hu, and F. Ramirez. 1995. Developmental expression of fibrillin genes suggests heterogeneity of extracellular microfibrils. *J Cell Biol* 129: 1165-76.

Zou, H., and L. Niswander. 1996. Requirement for BMP signaling in interdigital apoptosis and scale formation. *Science* 272: 738-41.

Chapter 4a

A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36.

Henk C. Heus[1], Anne Hing[2], Marijke J. van Baren[1], Marijke Joosse[1],
Guido J. Breedveld[1], Jen. C. Wang[3], Andrea Burgess[3], Helen Donnis-Keller[3],
Cathleen Berglund[2], Julia Zguricas[1] Stephen W. Scherer[4], Johanna M. Rommens[4],
Ben A. Oostra[1] and Peter Heutink[1].

[1]Department of Clinical Genetics, Erasmus University Rotterdam, the Netherlands
[2]Department of Pediatrics and [3]Department of Surgery, Washington University
School of Medicine, St. Louis, Missouri
[4]Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto, Canada.
[5]Department of Molecular and Medical Genetics, University of Toronto, Cananda.

A Physical and Transcriptional Map of the Preaxial Polydactyly Locus on Chromosome 7q36

Henk C. Heus,* Anne Hing,† Marijke J. van Baren,* Marijke Joosse,* Guido J. Breedveld,*
Jen C. Wang,‡ Andrea Burgess,‡ Helen Donnis-Keller,‡ Cathleen Berglund,†
Julia Zguricas,* Stephen W. Scherer,§ Johanna M. Rommens,§¶
Ben A. Oostra,* and Peter Heutink*,1

*Department of Clinical Genetics, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands; †Department of Pediatrics
and ‡Department of Surgery, Washington University School of Medicine, St. Louis, Missouri; §Program in Genetics
and Genomic Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; and ¶Department
of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada

Preaxial polydactyly is a congenital hand malformation that includes duplicated thumbs, various forms of triphalangeal thumbs, and duplications of the index finger. A locus for preaxial polydactyly has been mapped to a region of 1.9 cM on chromosome 7q36 between polymorphic markers D7S550 and D7S2423. We constructed a detailed physical map of the preaxial polydactyly candidate region. With a combination of methods we identified and positioned 11 transcripts within this map. By recombination analysis on families with preaxial polydactyly, using newly developed polymorphic markers, we were able to reduce the candidate region to approximately 450 kb. The homeobox gene *HLXB9*, a putative receptor C7orf2, and two transcripts of unknown function, C7orf3 and C7orf4, map in the refined candidate region and have been subjected to mutation analysis in individuals with preaxial polydactyly. © 1999 Academic Press

INTRODUCTION

Human limb malformations are relatively common and are mostly associated with syndromes that display other congenital malformations as well. Of all congenital hand malformations, preaxial polydactyly (PPD) is most frequently observed and includes duplicated thumbs as well as various forms of triphalangeal thumbs and index finger duplications (OMIM 190605). Sporadic cases of isolated PPD have been described,

but most cases show an autosomal dominant mode of inheritance. By linkage analysis a PPD locus has been mapped near polymorphic marker D7S559 on human chromosome 7q36 (Heutink *et al.*, 1994). The critical region is flanked by polymorphic markers D7S550 and D7S2423 and spans approximately 1.9 cM (Zguricas *et al.*, 1999). The penetrance of the PPD phenotype is almost complete; 139 of 140 related individuals sharing the disease-associated haplotype for D7S2423 and D7S559 showed a polydactyly phenotype (>99%) (Zguricas *et al.*, 1999).

In the same region the gene responsible for complex polysyndactyly (CPS) has been mapped (Tsukurov *et al.*, 1994). CPS is a congenital hand malformation characterized by pre- and postaxial limb anomalies combined with webbing of the fingers. Sacral agenesis (OMIM 176450), an autosomal dominant condition that includes a presacral mass and urogenital and anorectal anomalies, has also been mapped close to polymorphic marker D7S559 on chromosome 7q36 (Lynch *et al.*, 1995).

A high-resolution map of a genomic area is essential for the positional cloning of disease genes. These maps are useful for the various transcript identification methods; they allow the development of new polymorphic markers and can serve as a template for large-scale genomic sequencing. The holoprosencephaly type 3 (HPE3) region is adjacent to the PPD candidate region. And previously, a contig of genomic YAC clones was constructed to allow positional cloning approaches on the PPD and HPE3 genes on chromosome 7q36 (Scherer *et al.*, 1992; Kunz *et al.*, 1994; Belloni *et al.*, 1996). This YAC contig was used as a starting point in our effort to identify the PPD gene.

To make the candidate region more suitable for experimental analysis we constructed a detailed contig of genomic P1, PAC, and cosmid clones. Using a combination of exon trapping, cDNA selection, and EST map-

342

TABLE 1

Sequence Tagged Sites Positioned in the Genomic Contig from Centromere to Telomere

STS name	Size (bp)	Forward primer 5'-3'	Reverse primer 5'-3'
HSC7E445(L)	150	gtttcctccatctgacttcacc	catggtgactgagtgaaaaagc
381E2(R)	132	aaagttcttcttctgcctaacaga	cgctctttaactgacgggac
99E12(T7)	110	gctggaggaggagcgtcatgaac	gacaccaaaaactctgccctgaa
HSC7E70(L)	100	gagagtctctgccagcctgt	aagagtagaattcttgctagtggtca
34E3(SP6)	85	gcttctgtgctgaaggggg	gtctctagggtacacaaactgcg
K0922(T7)	163	tgcctttacctgattcgtg	attatatgttgttgcttttc
B1120(SP6)	84	ttttcctctaaagaaaacgtattgc	tcttcagacttcaaacaagctcc
106C9(T7)	108	agtctgtgtcatttctgcctc	gaggacggtgacataaacggt
I123(T7)	183	tgagcactaagcatcctg	aaacaaaaaacccaacata
HSC7E739(R)	109	agatcccacaaatgagcgag	gttaatcacgagaataaacggacc
HSC7E158(L)	119	cgcagggtccataaatagaa	gaagaggggaagaagatccg
H0153(SP6)	150	aattacctattggaaccctacccatc	aacataatgcacaagatgatgcaa
HSC7E445(R)	123	accttgacagcagcagt	ttagcatagagctgtgtcactatgg
N2113(T7)	84	aaccccatctctacaaaa	tgcctcagcctcccaagt

ping methods, 11 transcripts were identified and precisely positioned in the genomic map. We developed 11 new polymorphic markers and used these for detailed recombination analysis on PPD families. This enabled us to localize the gene responsible for the PPD phenotype to a region of approximately 450 kb.

Four transcripts map to this refined PPD candidate region. These are the homeobox-containing transcription factor *HLXB9*, the putative receptor C7orf2, and two transcripts with an unknown function, called C7orf3 and C7orf4. All four transcripts have been analyzed and sequenced in PPD patients, but no pathogenic mutations were identified. Large-scale genomic sequencing of the entire PPD candidate region will hopefully lead to the identification of the PPD gene.

MATERIALS AND METHODS

Construction of a High-Resolution Genomic Contig

Genomic library screening. The ICRF cosmid library (113-L4/FS7), the PAC library (RPCI1) constructed by P. J. de Jong (Ioannou *et al.*, 1994), and a P1 library (Shepherd *et al.*, 1994) were hybridized with long-range inter-*Alu* products (Expand Long Template PCR System; Boehringer Mannheim) derived from YAC clones HSC7E445, HSC7E158, and HSC7E70. These YAC clones were obtained from The Hospital for Sick Children chromosome 7-specific library and previously used for the construction of a physical map of the HPE3 critical region (Belloni *et al.*, 1996).

Sequence tagged site (STS) assays from the ends of each P1 clone were used to identify overlapping clones. All resulting clones were mapped back to the contig of YAC clones by hybridization, using standard techniques as described by Sambrook *et al.* (1989). The positions of individual clones are depicted in Fig. 1.

YAC-specific cosmid library construction. DNA of YAC clones HSC7E445 and HSC7E158 was obtained by growing single yeast colonies in Achilles heel cleavage medium at 30°C for 48 h. Total yeast DNA was prepared in 100 µL Seaplaque (FMC, Rockland, ME) agarose plugs. Agarose plugs were incubated overnight in *Mbo*I restriction buffer containing 0.1 U *Mbo*I restriction enzyme at 4°C. Restriction of the DNA was performed at 37°C for 20 min and was stopped by addition of EDTA to a final concentration of 50 mM. The presence of 30- to 35-kb restriction fragments was checked by size separation on a pulsed-field agarose gel as has been described by van Ommen (1986). The remaining agarose plugs were treated with

agarase (Boehringer Mannheim) according to the manufacturer's protocol. DNA fragments were cleaned by phenol extraction, precipitated with ethanol, and ligated into the *Bam*HI site of the sCOGH2 cosmid vector (Datsun *et al.*, 1996) using standard techniques as described by Sambrook *et al.* (1989). The resulting cosmid library was packaged into phage particles using the GigapackII XL extract (Stratagene) according to the manufacturer's protocol. The cosmid library was plated, transferred to nylon filters, and hybridized with total human genomic DNA to identify the clones that contain a human genomic DNA insert. We picked, isolated DNA from, and fingerprinted 68 cosmid clones that contained a human DNA insert using standard techniques (Sambrook *et al.*, 1989). Cosmid clones have been designated FRS followed by a number. The positions of the individual clones are depicted in Fig. 1.

Map integration. Sequence tagged sites were generated from vector–insert junction sequence from YAC, P1, and cosmid clones using the ligation-mediated PCR technique as described in Mueller and Wold (1989). Table 1 lists the primer sequences for the STSs.

Sequencing. All sequencing in this study was done on a ABI-377 automated fluorescence dye sequencer using Bigdye chemistry (Perkin Elmer) or Dye terminators (Amersham).

Identification of Polymorphic Markers

Genomic clones were partially restricted with *Sau*3A, shotgun cloned into *Bam*HI-linearized pBluescript vector (Stratagene), and plated. Colonies were transferred to nylon filters and hybridized with the following set of di-, tri-, and tetranucleotide oligos: (CA)₁₅, (CTG)₁₀, (AATT)₉, (AAAT)₉, (AACT)₆, (AAGT)₉, (AGAT)₉, and (ACAT)₉. Plasmid inserts from positive colonies were sequenced, and PCR amplification primers flanking the repeated regions were designed. Each marker was tested on several unrelated individuals to test whether it was polymorphic. In this way 11 new polymorphic markers were integrated into the map (see Table 2). The positions of individual markers are depicted in Fig. 1. Existing markers, D7S559 and D7S2465, were also integrated into the map (Green *et al.*, 1991). Positions of the polymorphic markers in the map are indicated in Fig. 1. Additional markers that were used for recombination analysis were D7S550, D7S3037, D7S3036, D7S104, and D7S2423 (Green *et al.*, 1991; Belloni *et al.*, 1996).

Amplification of Polymorphic Markers

Amplification of polymorphic markers was performed in a 15-µL reaction containing 20 mM Tris–HCl, pH 8.4, 1.5 mM MgCl₂, 50 mM KCl, 200 µM dNTPs, 0.5 U *Taq* DNA polymerase (Gibco BRL), and 10 pmol forward and reverse primers. Cycling conditions were 10 min at 94°C, 30 cycles of 30 s at 94°C; 30 s at 55°C, 90 s at 72°C, and a final step of 5 min at 72°C.

TABLE 2

New Polymorphic Markers in the Genomic Contig from Centromere to Telomere

Marker	Parent clone	Size (bp)	Forward primer 5'–3'	Reverse primer 5'–3'
CGR13	264E9	154	cctctatggcaatctcctgc	cttgggtggcatacagtgtg
CGR12	140E19	212	ccaggagaactgtgttttcc	tggccacagttgaactttca
CGR6	K0922	112	aaaagacagtgtttccaaagca	gtggccattcattgagcc
HING1	B1120	169	gtaggctgacaaagaagaataac	actagataataatatctaactac
CGR5	G2238	295	ggaggcagaggtttcagtga	acttcagtagttttctaggttt
CGR17	255D6	168	cactccagcctgggtgac	accttctgccccaagaattc
CGR16	N1914	132	actccactacctcaattttaata	ttggattgtattgttatttaaatg
CGR3	1280C3	300	gccaagcatgttggcttta	cagaaactcctggcttgag
CGR8	I1338	194	tgacaagagcaaaactctgtca	ttcctgcctttcatcagcct
CGR10	FRS11	109	ctctgcctctctctgtagcca	tgcatgggctggttttaaat
CGR2	P1317a1	100	gcataggccccacacacac	gtaaaaactcagcactgaaggc

cDNA Selection

DNA from genomic clones was transferred to a nylon membrane. The membrane was hybridized with a cDNA mixture made of poly(A)⁺ RNA from fetal brain, fetal liver, fetal kidney, placenta, adult testis, and adult brain (frontal cortex). To block repetitive sequences cDNA made from total RNA isolated from the Caco-2 (ATTC HTB37) cell line and sonicated human placental DNA were added to the cDNA mixture. Hybridization was carried out in Church hybridization solution at 60°C (Church and Gilbert, 1984). The final stringency for washing was 0.2× SSC/0.1% SDS at 60°C. After two consecutive cycles of hybridization cDNA fragments were eluted from the filters and amplified by PCR. PCR products were cloned into the pBluescript vector (Stratagene), sequenced, and subjected to further analysis.

Exon-Trapping Analysis

Exon trapping was done as described by Buckler et al. (1991). After the secondary amplification step, PCR products of the putative exons were cloned into the pAMP10 vector (Gibco BRL) and plated. Plasmid inserts were amplified by PCR directly from individual colonies using dUSD2/dUSA4 primers. Insert size was checked by size separation on a 2% agarose gel. Putative exons were sequenced, and sequences were screened for repeats using REPEATMASKER (Smith et al., 1996). A nonredundant set of putative exons was mapped back to genomic clones from the candidate region by hybridization. Exon trapping was performed on the hatched clones depicted in Fig. 1.

EST Mapping

Based on the localization of ESTs to the radiation hybrid bins of the Human Transcript Mapping Project (http://www.ncbi.nlm.nih.gov/science96/) and the sequence nucleotide polymorphisms (SNPs) map (Wang et al. 1998), 16 primer pairs of transcripts that potentially map between markers D7S550 and D7S2423 were identified. These primer pairs were tested by PCR on DNA from a selection of genomic clones covering the region between polymorphic markers CGR2 and D7S2465. Total genomic DNA served as control.

cDNA Library Screening

For each library 1 × 10⁶ plaques were plated, transferred to nylon filters, and hybridized with the probes of interest using standard procedures (Sambrook et al., 1989). The following cDNA libraries were used: human fetal brain (Clontech; 5'-stretch library HL1149x), human placenta (Clontech; 5'-stretch library HL1144x), and human testis (Clontech; HL1010b).

5'RACE

All 5'RACE experiments were done with the human fetal brain Marathon-Ready cDNA kit (Clontech) as recommended by the manufacturer.

Northern Blot Analysis

Multiple-tissue Northern blots were purchased from Clontech. Hybridization in ExpressHyb solution (Clontech) and washes were done as recommended by the manufacturer. The Northern contains RNA from the following tissues: heart, brain, placenta, lung, liver, muscle, kidney, and pancreas. Hybridization probes were prepared by PCR, and the nucleotide sequence was confirmed. The following probes were used: a 558-nucleotide probe from C7orf3(1), a 431-nucleotide probe from C7orf2(3), and a 149-nucleotide probe from C7orf4(1). See Table 3 for the primer sequences of these PCR products.

Mutation Analysis

The families used in this study have been described before (Heutink et al., 1994; Zguricas et al., 1999) and are unrelated. Affected individuals do not share a common haplotype across the candidate region. PCR products from either genomic DNA or reverse-transcribed (RT) products were amplified (see Table 3 for primer sequences), cleaned with the QIAquick PCR purification kit (Qiagen), and sequenced. Mutation analysis for the HLXB9 and C7orf4 transcripts was performed on genomic DNA. HLXB9 primers were designed to amplify all three exons of the HLXB9 gene including their intron–exon boundaries. Exon 1 has been amplified with 286-bp upstream and 166-bp downstream sequences. Exons 2 and 3 have been amplified including the 598-bp intron that separates them. The amplification of these exons included 58 bp of sequence upstream of exon 2 and 89 bp of sequence downstream of exon 3. Mutation analysis on the HLXB9 and C7orf4 transcripts was carried out on an affected individual from each of the following four families: Dutch, Cuban A, UK, and Turkish (Heutink et al., 1994; Zguricas et al., 1999). C7orf4 primer pairs were designed to amplify the known transcript sequence. Control sequence was obtained from unaffected individuals from the above-mentioned families and YAC clone HSC7E445.

Mutation analysis on the partly known C7orf2 and C7orf3 transcripts was carried out on three affected persons from the Dutch and Cuban A and B families (Heutink et al., 1994; Zguricas et al., 1999). Transcripts were amplified from RT-PCR products made from Epstein–Barr virus-transformed lymphocyte culture RNA. Control sequence was obtained from unaffected persons. Primer pairs for the C7orf2 and C7orf3 transcripts were designed to cover all known sequence of the transcripts.

RESULTS

Construction of a Genomic Contig

YAC clones HSC7E445, HSC7E158, and HSC7E70 span a large part of the region between markers D7S550 and D7S2423 and have been used as a starting point to construct a detailed genomic contig. In gen-

TABLE 3

Primer Sequences Used in Mutation Analysis of the Candidate Genes

PCR product	Forward primer 5'–3'	Reverse primer 5'–3'
HLXB9(1)	cagcccgggctgcttaggac	atttgccaatoatcaaagtcgc
HLXB9(2)	agggctgcagcgcaaagaa	ctctgcgcagaggcggctc
HLXB9(3)	aaccaccataacaataccgagc	ctgcaggctgccgctngtc
HLXB9(4)	gagcgcctctgcgcagag	goagcccggcttgggcag
HLXB9(5)	tgcgcgccgagagcccgt	cagggtgcagccccagcgc
HLXB9(6)	ccaccaccacgcgcatc	gccactggtccngctggaag
HLXB9(7)	acgcgcatccggcgcagc	cggggtgcgcgccttgca
HLXB9(8)	ggcgctctcctactcgtaccc	ataattcgctgcccgtctcatg
HLXB9(9)	tgtagtggtacaagccagcaacgg	gcaaaggtaacngtgtcccctg
HLXB9(10)	gggacttgngggacagtgac	ggaggccgcggccgnatcc
HLXB9(11)	ctccggggacaggoagc	ctcgtcctccgnggagcagt
HLXB9(12)	agagcaggctcgaccgcc	ccagcagtttgaacgctcg
HLXB9(13)	cttgaaaccgcctctgggg	gactccgattttcacccctcagc
HLXB9(14)	gcccaagcgcttcgaggtg	gggacttttctaccccgcgcc
HLXB9(15)	ggccgcattcgagnaattttgttc	ttcatccgccggttctggaa
C7orf3(1)	agctggagcgttgcctcgg	cactcaacctgttaagtagacc
C7orf3(2)	attcctcggaggacggtgac	ggagagactgtaccacgtgg
C7orf3(3)	tggtgaggangttcgatgcc	gacccgtctgctccgcac
C7orf3(4)	tgcagagagctttggtccg	ggcaagtttttccnagtcccg
C7orf3(5)	cctggctagcaaattctgtg	ttacngtcanagggccacc
C7orf3(6)	ccttgcaacgaagtgtctgc	tgcctcagccacctcagtagc
C7orf2(1)	tgctgaggcgacggcggt	aaggcaaaggggcatcaatac
C7orf2(2)	ctcaganctactatattcag	acgagaaaggccaaactggtgtac
C7orf2(3)	gagttctatctaccctattt	gttccttttggcattgctg
C7orf2(4)	atctcggtcctcttggtg	gccaagttcttcgtagatta
C7orf2(5)	cgtcttctgccaccangag	tacangcagttgccattgc
C7orf2(6)	ctttattccagactgtagaagc	cctataaggtnaaaatccaag
C7orf2(7)	gangtatcatgtaatcacagt	cagtggncttanagccang
C7orf4(1)	cttctatcctttcccaactctcctccac	cgctattaaataccatgcaggaaattggg

eral, two strategies were applied. First, we screened cosmid, PAC, and P1 libraries with long-range inter-*Alu* products of these YAC clones. Second, we constructed a YAC-specific cosmid library from YAC clones HSC7E445 and HSC7E158 and positioned 67 additional cosmid clones in the map. End sequencing of several clones allowed the development of STSs (see Table 1) that were used to rescreen the P1 library for overlapping clones. The 20 ICRF cosmids, 21 P1 clones, and 13 PAC clones we identified were positioned in the map and form a contig with a total size of approximately 1200 kb. This contig is continuous, except for a gap between the YAC-derived cosmid FRS7 and the P1 clone 181C4. Based on the size and overlap of YAC clones HSC7E445 and HSC7E158, the gap is estimated to be less than 100 kb. Clones covering the gap were not present in any of the genomic cosmid, PAC, and P1 libraries screened. A schematic overview of the high-resolution genomic contig and the positions of the STSs is presented in Fig. 1.

Transcript Identification and Characterization

To identify transcripts within the genomic contig we used a combination of cDNA selection, exon-trapping analysis, and public database searching. Several transcripts were identified by more than one method. Table 4 gives an overview of the identified transcripts, and Fig. 1 indicates their precise location in the genomic contig.

cDNA Selection

cDNA selection was performed on clones HSC7E158, HSC7E445, and P1317A1, which span almost the entire genomic contig. This yielded 18 unique cDNA fragments that were positioned in the contig and sequenced. By a search of publicly available databases with these sequences 9 fragments obtained from P1317A1 were observed to encode fragments of phogrin, a tyrosine-phosphatase-like protein that is expressed in brain and pancreas (Kawasaki *et al.*, 1996). Five other fragments are part of KIAA0010 (Nomura *et al.*, 1994). This 5160-nucleotide transcript encodes a protein that contains a ubiquitin-ligase domain. Northern analysis of KIAA0010 showed bands of 5.1 kb in all tissues, with an additional band of 3.2 kb in skeletal muscle (data not shown). A single cDNA selection fragment that maps to N2113 showed homology with ubiquitin. Fragment 2AD21 (AF107458) maps to cosmid clone N2113 and shows homology with several mouse ESTs of unknown function (AA516622, AA516843, and AA522017). Fragments GT725 (AF107459) and GT727 (AF107460) both map to cosmid clone N1042 and do not show homology with any known sequences. Screening of cDNA libraries yielded no clones that would extend these two cDNA selection fragments.

The C7orf2 transcript (AF107454) was identified by a separate direct cDNA selection experiment on genomic clones from chromosome 7 performed by one of
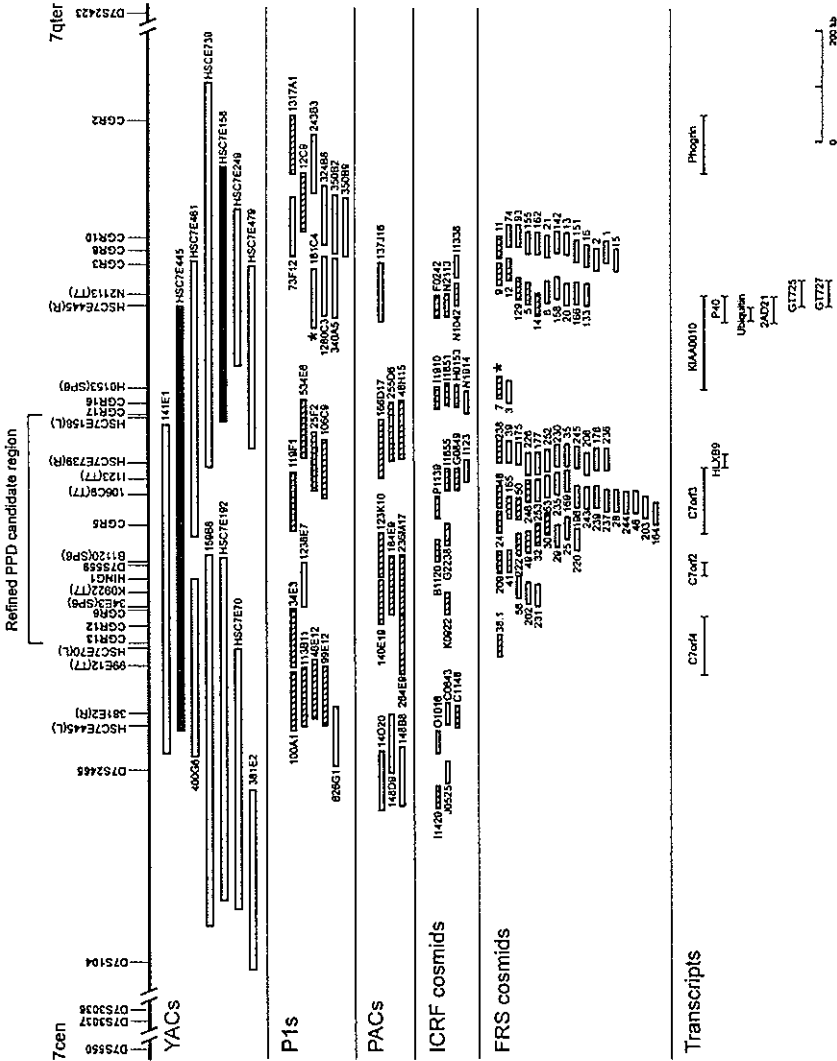
FIG. 1. Schematic representation of the detailed physical map around the PPD candidate region. The positions of STSs and polymorphic markers are indicated at the top. For markers D7S2423, D7S104, D7S3036, D7S3037, and D7S550 the relative positions are indicated. Genomic YAC clones HSC7E445 and HSC7E158, which were used to make the FRS cosmid library, are depicted in gray. Clones FRS7 and 181C4 flank a small gap in the contig and have been marked with an asterisk. The hatched P1, PAC, and cosmid clones were used for exon trapping. Additional information on the YAC clones can be found at http://www.genet.sickkids.on.ca/chromosome7/.

TABLE 4

Overview of Identified Transcripts between Markers CGR2 and D7S2465

Name	GB ID	Method	Proposed function	Expression	Reference
Phogrin	U66702	cs	Tyrosin phosphatase-like	Brain, pancreas	Kawasaki et al., 1996
P40	Y11395	db	G protein-coupled receptor	Predominantly heart, brain	Mayer et al., 1998
Ubiquitin	X63237	cs	Ubiquitin	Not done	
KIAA0010	D13635	cs et db	Ubiquitin ligase	All tissues MTN	Nomura et al., 1994
HLXB9	AF107457	db	Homeobox transcription factor	Lymphoid tissues, pancreas	Harrison et al., 1994
C7orf3	AF107455	et	Unknown function	All tissues MTN	
C7orf2	AF107454	cs et db	Transmembrane receptor	All tissues MTN	
C7orf4	AF107456	db	Unknown function	No signal	
2AD21	AF107458	cs	Unknown function	No signal	
GT725	AF107559	cs	Unknown function	Not done	
GT727	AF107460	cs	Unknown function	Not done	

Note. The methods that have been used for identification of transcripts are cs, cDNA selection; et, exon trapping; and db, database searching.

the authors (unpublished results). Northern analysis of the C7orf2 transcript showed bands of 1.9 and 4.8 kb in all tested tissues, with highest expression in heart and pancreas (see Fig. 2). By using 5'RACE and screening of cDNA libraries a sequence of 4849 nucleotides was obtained. When the Northern blot was exposed for 7 days, an additional faint band of 5.2 kb was visible in brain, indicating a possible third transcript, predicting an additional 400 nucleotides of sequence that was not found in clones we isolated from cDNA libraries. The C7orf2 transcript sequence encodes a 492-amino-acid open reading frame (ORF) that starts with a methionine residue. The ORF is preceded by a 171-nucleotide 5'-UTR that contains a stop codon positioned three triplets before the start of the ORF. The ORF is followed by a relatively long 3202-nucleotide 3'-UTR that does not contain any extensive ORFs. The C7orf2 ORF predicts a protein that shows homology to proteins in Fugu rubripes (AF056116) and Caenorhabditis elegans (P35535). Both are hypothetical proteins and of unknown function. Because the C7orf2 protein contains nine transmembrane domains and a coiled-coil domain, and is probably located in the plasma membrane (Nakai and Kanehisa, 1992), it might encode a receptor protein.

Exon-Trapping Analysis

Exon trapping of 33 cosmid, 11 P1, and 8 PAC clones, covering the entire genomic contig, yielded 68 unique putative exons. Several exons identified a transcript (C7orf3) that included the EST02091. A total of 45 putative exons could not be assigned to any previously identified transcript, and we were unable to find corresponding cDNA clones in several libraries. The sequences of these putative exons have been submitted to GenBank (AF107407–AF107451).

The C7orf3 transcript (AF107455) is highly expressed in heart and skeletal muscle with two major bands of 1.4 and 2.4 kb and a number of minor bands up to 7.3 kb in size (see Fig. 2). After 5'RACE experiments and screening of human fetal brain, fetal limb, testis, and placenta cDNA libraries we obtained a sequence of 2754 nucleotides. This sequence codes for a
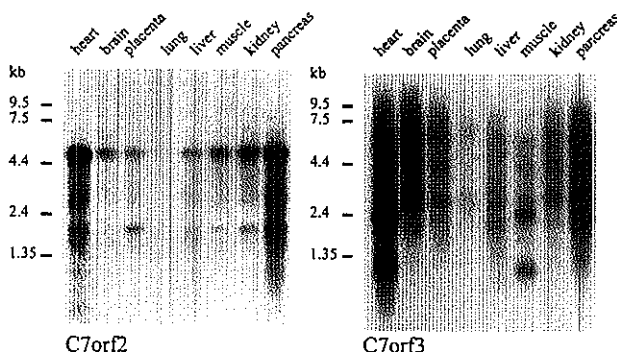

FIG. 2. Northern blot analysis. (A) C7orf2 transcript after 1 day of exposure showing 1.9- and 4.8-kb bands. (B) C7orf3 transcript after 3 days of exposure.

738-amino-acid ORF that is followed by a 540-nucleotide 3'-UTR containing a possible polyadenylation signal. The ORF starts at the first nucleotide of the known sequence, and is not complete, but we were unable to find additional cDNA clones or ESTs that extend the transcript farther 5'. The partial C7orf3 protein shows homology to proteins of unknown function from *Schizosaccharomyces pombe* (Z98601) and *Saccharomyces cerevisiae* (S51341).

EST Mapping

As a third strategy to identify transcripts we searched several public databases for ESTs that potentially map within the PPD critical region and mapped three transcripts in the region.

A 4544-nucleotide transcript called P40 encodes a 399-amino-acid seven-transmembrane domain protein, which is predicted to be a putative G-protein-coupled receptor. P40 was initially identified by isolating a 40-kDa integral membrane protein from human erythrocytes. Northern blot analysis of P40 shows a major 4.8-kb transcript that is predominantly expressed in heart and brain (Mayer *et al.*, 1998).

The homeobox motif-containing *HLXB9* gene was originally mapped to chromosome 1q41 (Harrison *et al.*, 1994), but was reported in the Human Transcript Map to map to chromosome 7q36 (Deloukas *et al.*, 1998). We could confirm this localization with fluorescence *in situ* hybridization (data not shown). The *HLXB9* gene consists of three exons spread over 6 kb of genomic DNA and encodes a 2.2-kb transcript that is expressed in lymphoid and pancreatic tissues (Harrison *et al.*, 1994). The homeodomain of the *HLXB9* protein is distantly related to that of the *Drosophila* gene *proboscipedia* (*pb*). The most closely related human homeodomain is that of *HOXB2* (Deguchi and Kehrl, 1991; Harrison *et al.*, 1994).

The STS sTSG8150, which contains the SNP WIAF2185, has been mapped on chromosome 7q36 in-between polymorphic markers D7S550 and D7S2423. sTSG8150 identified a contig of four overlapping cDNA clones in dbEST, with the 630-nucleotide ZI25H08 as the most extended cDNA clone. All four ESTs came from the same Soares fetal liver/spleen cDNA library (1NFLS S1), suggesting a tissue-specific expression of the transcript. This transcript has been called C7orf4 and showed no signals on Northern blot, and cDNA library screening and 5'RACE experiments did not result in clones that extended the transcript any farther. The known sequence encodes a 108-amino-acid ORF, followed by a 304-nucleotide 3'-UTR containing a putative polyadenylation site. No homologies to other known proteins were identified with the partial C7orf4 protein sequence.

The following ESTs could be excluded from the PPD refined candidate region (see below) by hybridization and PCR: cda0ba09, SGC34464, A002Q20, WI-6320, StSg2621, bda72a07, and a008429.

TABLE 5

Haplotype Analysis of Individuals in Cuban and Dutch PPD Families

Markers	Cuba-A, affected	Dutch, affected	Dutch, not affected	Dutch, not affected
D7S550	■	O	■	■
D7S3037	·	O	■	■
D7S3036	·	I	■	■
D7S104	I	■	■	■
D7S2465	■	■	■	I
CGR13	■	■	■	O
CGR12	·	■	I	·
CGR6	·	■	O	·
D7S559	■	■	I	O
CGR5	■	■	O	O
CGR17	O	■	O	·
CGR16	O	■	O	·
CGR3	I	I	O	I
CGR10	I	■	O	·
CGR2	O	■	O	O
D7S2423	O	■	O	O

Note. The polymorphic markers that were not informative in all four individuals have been left out for reasons of clarity. O, unaffected allele; ■, affected allele; I, not informative; ·, not done.

Refinement of the Candidate Region

During the construction of a detailed physical map we also attempted to reduce the size of the candidate region. Several new polymorphic markers within the candidate region were developed (see Table 2), and additional markers became available. The order of polymorphic markers in the region starting from the centromeric side is D7S550–D7S3037–D7S3036–D7S104–D7S2465–CGR13–CGR12–CGR6–HING1–D7S559–CGR5–CGR17–CGR16–CGR3–CGR8–CGR10–CGR2–D7S2423. To detect possible recombination events between the disease phenotype and these genetic markers we performed haplotype analysis on individuals from Cuban and Dutch PPD families. Previously, we reported recombinations in these families with D7S550 on the centromeric side and with D7S2423 on the telomeric side, an interval of 1.9 cM (Zguricas *et al.*, 1999). By looking for recombination events in these affected persons we were able to refine further the PPD candidate region between CGR17 on the telomeric site and D7S3037 on the centromeric site, an interval of 1 cM that excludes *Sonic hedgehog* as a candidate gene (see Table 5). In addition we identified two recombination events on the centromeric site in unaffected persons with markers D7S104 and CGR13. Because penetrance of PPD is almost complete (>99%) we have focused our attention on a region of approximately 450 kb in-between markers CGR17 and CGR13.

Mutation Analysis on Candidate Genes

After reduction of the PPD candidate region four transcripts remained as potential candidate genes: *HLXB9, C7orf2, C7orf3,* and *C7orf4*. We performed

mutation analysis on these transcripts, but no sequence alterations specific for affected individuals were found. Several heterozygous sequence alterations were detected for the *HLXB9* gene and the C7orf2 and C7orf3 transcripts, but these alterations were also found in healthy family members and unrelated controls and should therefore be regarded as nonpathogenic polymorphisms. None of the sequence alterations influenced amino acid coding potential of the *HLXB9* gene and the C7orf2 and C7orf3 transcripts.

The *HLXB9* sequence we obtained from human genomic DNA, and the YAC clone HSC7E445, differs somewhat from the published genomic sequence, and the corrected sequence has been submitted to GenBank (AF107452, AF107453, and AF107457).

DISCUSSION

Generation of high-resolution genomic maps of the human genome is the first step toward characterization of the genes within a region. We focused our attention on the region between polymorphic markers CGR2 and D7S2465 on chromosome 7q36. This region contains the gene responsible for PPD and is directly adjacent to that described for HPE3 (Belloni *et al.*, 1996). Using an available contig of YAC clones, a high-resolution contig of P1, PAC, and cosmid clones was constructed. This contig is continuous except for a gap that could not be covered by any of the genomic cosmid, PAC, and P1 libraries used. The gap is bridged by the KIAA0010 transcript, which maps on clones flanking the gap. Since the genomic libraries were constructed from independent DNA sources we assume that this gap contains sequences that are difficult to clone, such as highly repeated, or extremely GC-rich sequences.

The assembly of an ordered high-resolution contig of this region has enabled the precise location of several genes. A combination of cDNA selection, exon trapping, and database searching was employed to make the transcription map as complete as possible. The 11 independent transcripts that were identified are a tyrosine-phosphatase-like protein called phogrin, a ubiquitin ligase KIAA0010, a transmembrane receptor C7orf2, a G-protein-coupled receptor P40, the homeobox-containing transcription factor *HLXB9*, 5 transcripts of unknown function, and a fragment that shows homology with ubiquitin. These transcripts are candidate genes for congenital malformations such as PPD, CPS, and sacral agenesis that have been mapped within this region.

Using clones from the genomic contig, new polymorphic markers were developed that allowed us to refine recombination events in PPD families. The first polymorphic marker that recombined with the PPD phenotype of an affected individual on the centromeric side is D7S3037. A second recombination event on the centromeric side was identified with polymorphic marker CGR13. This recombination event was found in an unaffected individual and could be a case of nonpen-

etrance of the PPD phenotype. However, we previously reported that penetrance of the PPD phenotype is almost complete (Zguricas *et al.*, 1999), and further detailed clinical examination by one of us with no prior knowledge of the genetic status of this individual showed no abnormalities of hands and feet. Therefore the possibility that this person is a case of nonpenetrance is small. In addition, we found a third recombination event on the centromeric side, also in an unaffected person, with polymorphic marker D7S104, which lies somewhat centromeric of CGR13. The recombination event that defines the PPD candidate region on the telomeric side is with polymorphic marker CGR17 and was identified in an affected individual. In this way the PPD candidate region was reduced to approximately 450 kb between markers CGR13 and CGR17. In the refined candidate region of 450 kb, four genes remained as candidates for PPD.

HLXB9 is the most plausible candidate gene in the region. Homeobox-containing transcription factors play an important role in regulatory processes during embryonic development and are shown to be involved in a number of other congenital hand malformations. Examples are HOXD13 in synpolydactyly (Muragaki *et al.*, 1996) and HOXA13 in hand–foot–genital syndrome (Mortlock and Innis, 1997). Recently, it was found that mutations in the *HLXB9* gene are involved in the pathogenesis of sacral agenesis (Ross *et al.*, 1998). The C7orf2 transcript encodes a protein that might function as a receptor located in the plasma membrane. Receptors play an important role in signaling cascades that function during embryonic patterning and are therefore potential candidates for congenital malformations like PPD. The remaining two transcripts in the refined PPD candidate region are C7orf3 and C7orf4. At this moment it is not known what the function of these transcripts is or whether they are biologically plausible candidate genes. It is also unknown whether any of the candidate genes is expressed during embryonic limb development, something that is expected for the gene responsible for the PPD phenotype.

Although we tested PPD patients for mutations in these four candidate genes, no pathogenic sequence alterations were found. The heterozygous sequence alterations that have been identified in the *HLXB9* gene, and the C7orf2 and C7orf3 transcripts, were also found in unaffected individuals and rule out deletions of whole genes. Comparable peak heights at the variable positions in the C7orf2 and C7orf3 transcripts indicated that both forms of the transcripts are expressed at similar levels, ruling out instabilities of these transcripts as a cause of the PPD phenotype. There is still a possibility that candidate genes are mutated in parts of the genes that we were not able to investigate; C7orf4 and C7orf3 transcript sequences are not complete. The transcript size of the full-length C7orf4 transcript is unknown, and from Northern analysis we estimate that the C7orf3 transcript has an additional 4.5 kb of sequence that has not been investigated.

Furthermore, it is possible that the PPD phenotype is caused by mutations at intron–exon boundaries or in regulatory elements, such as promoter regions or enhancer sequences. Mutation analysis for the C7orf2 and C7orf3 transcripts were done on reverse-transcribed mRNA. We will not be able to investigate mutations at intron–exon boundaries or in regulatory elements until the genomic organization of the candidate genes is known. Finally, based on the exon-trapping results there is a possibility that there are additional genes present in the PPD candidate region.

The high-resolution contig between CGR2 and D7S2465 has been submitted to a large-scale sequencing center where it will be sequenced as a part of the human genome project. The data will help identify possible additional transcripts, complete the sequence of known transcripts, and resolve the complete genomic organization of the candidate genes. Hopefully this will lead to the identification and characterization of the gene responsible for PPD and encourage the identification of the gene responsible for CPS. Ultimately, we hope that identification of the PPD gene will give us more insight into embryonic limb development and the underlying processes leading to congenital malformations of the hand.

ACKNOWLEDGMENTS

REFERENCES

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H. F., Donis-Keller, H., Helms, C., Hing, A. V., Heng, H. H. Q., Koop, B., Martindale, D., Rommens, J. M., Tsui, L. C., and Scherer, S. W. (1996). Identification of *Sonic hedgehog* as a candidate gene responsible for holoprosencephaly. *Nat. Genet.* 14: 353–356.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* 88: 4005–4009.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. *Proc. Natl. Acad. Sci. USA* 81: 1991–1995.

Datsun, N., van de Vosse, E., Dauwerse, H. G., Bout, M., van Ommen, G. J. B., and den Dunnen, J. T. (1996). Scanning for genes in large genomic regions: Cosmid-based exon trapping of multiple exons in a single product. *Nucleic Acids Res.* 6: 1105–1111.

Deguchi, Y., and Kehrl, J. H. (1991). Nucleotide sequence of a novel diverged human homeobox gene encodes a DNA binding protein. *Nucleic Acids Res.* 19: 3742.

Deloukas, P., Schuler, G. D., Gyapay, G., Beasley, E. M., Soderlund, C., Rodriguez-Tome, P., Hui, L., Matise, T. C., McKusick, K. B., Beckmann, J. S., Bentolila, S., Bihoreau, M., Birren, B. B., Browne, J., Butler, A., Castle, A. B., Chiannilkulchai, N., Clee, C., Day, P. J. R., Dehejia, A., Dibling, T., Drouot, N., Duprat, S., Fizames, C., Fox, S., *et al.* (1998). A physical map of 30,000 human genes. *Science* 282: 744–746.

Green, E. D., Mohr, R. M., Idol, J. R., Jones, M., Buckingham, J. M., Deaven, L. L., Moyzis, R. K., and Olson, M. V. (1991). Systematic generation of sequence-tagged sites for physical mapping of human chromosomes: Application to the mapping of human chromosome 7 using yeast artificial chromosomes. *Genomics* 3: 548–564.

Harrison, K. A., Druey, K. M., Deguchi, Y., Tuscano, J. M., and Kehrl, J. H. (1994). A novel homeobox gene distantly related to *proboscipedia* is expressed in lymphoid and pancreatic tissues. *J. Biol. Chem.* 269: 19968–19975.

Heutink, P., Zguricas, J., van Oosterhout, L., Breedveld, G. J., Testers, L., Sandkuijl, L. A., Snijders, P. J., Weissenbach, J., Lindhout, D., Hovius, S. E. R., and Oostra, B. A. (1994). The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. *Nat. Genet.* 63: 287–292.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A., and de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat. Genet.* 6: 84–89.

Kawasaki, E., Hutton, J. C., and Eisenbarth, G. S. (1996). Molecular cloning and characterisation of the human transmembrane protein tyrosine phosphatase homologue phogrin, an autoantigen of type 1 diabetes. *Biochem. Biophys. Res. Commun.* 227: 440–447.

Kunz, J., Scherer, S. W., Klawitz, I., Soder, S., Du, Y.-Z., Speich, N., Kalff-Suske, M., Heng, H., Tsui, L.-C., and Grzeschik, K.-H. (1994). Regional localization of 725 human chromosome 7-specific yeast artificial chromosome (YAC) clones. *Genomics* 22: 439–448.

Lynch, S. A., Bond, P. M., Copp, A. J., Kirwan, W. O., Nour, S., Balling, R., Mariman, E., Burn, J., and Strachan, T. (1995). A gene for autosomal dominant sacral agenesis maps to the holoprosencephaly region at 7q36. *Nat. Genet.* 11: 93–95.

Mayer, H., Salzer, U., Breuss, J., Ziegler, S., Marchler-Bauer, A., and Prohaska, R. (1998). Isolation, molecular characterisation, and tissue-specific expression of a novel putative G protein-coupled receptor. *Biochem. Phys. Acta* 1395: 301–308.

Mortlock, D. P., and Innis, J. W. (1997). Mutation of HOXA13 in hand–foot–genital syndrome. *Nat. Genet.* 15: 179–180.

Mueller, P. R., and Wold, B. (1989). In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* 10: 780–786.

Muragaki, Y., Mundlos, S., Upton, J., and Olsen, B. O. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* 272: 548–551.

Nakai, K., and Kanehisa, M. (1992). A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14: 897–911.

Nomura, N., Miyajima, N., Sazuka, T., Tanaka, A., Kawarabayasi, Y., Sato, S., Nagase, T., Seki, N., Ishikawa, K., and Tabata, S. (1994). Prediction of the coding sequences of unidentified human genes. *DNA Res.* 1: 27–35.

Ross, A. J., Ruiz-Perez, V., Wang, Y., Hagan, D. M., Scherer, S., Lynch, S. A., Lindsay, S., Custard, E., Belloni, E., Wilson, D. I., Wadey, R., Goodman, F., Orstavik, K. H., Monclair, T., Robson, S., Reardon, W., Burn, J., Scambler, P., and Strachan, T. (1998). A homeobox gene, HLXB9, is the major locus for dominantly inherited sacral agenesis. *Nat. Genet.* 20: 358–361.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scherer, S. W., Tompkins, B. J. F., and Tsui, L.-C. (1992). A human chromosome 7-specific genomic DNA library in yeast artificial chromosomes. *Mamm. Genome* 3: 179–181.

Shepherd, N. S., Pfrogner, B. D., Coulby, J. N., Ackerman, S. L., Vaidyanathan, G., Sauer, R. H., Balkenhol, T. C., and Sternberg, N. (1994). Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. Natl. Acad. Sci. USA* 91: 2629–2633.

Smith, R. F., Wiese, B. A., Wojzynski, M. K., Davison, D. B., and Worley, K. C. (1996). BCM search launcher: An integrated interface to molecular biology database search and analysis services available on the World Wide Web. *Genome Res.* 6: 454–462.

Tsukurov, O., Boehmer, A., Flynn, J., Nicolai, J. P., Hamel, B. C., Traill, S., Zaleske, D., Mankin, H. J., Yeon, H., Ho, C., Tabin, C., Seidman, J. G., and Seidman, C. (1994). A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. *Nat. Genet.* 63: 282–286.

van Ommen, G. J. B. (1986). Restriction analysis of chromosomal DNA in a size range up to two million base pairs by pulsed field gradient electrophoresis. *In* "Human Genetic Diseases" (K. E. Davies, Ed.), pp. 111–132, IRL Press, Oxford/Washington, DC.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M., and Lander, E. S. (1998). Large-scale identification mapping and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077–1082.

Zguricas, J., Heus, H. C., Morales-Peralta, E., Breedveld, G. J., Kuyt, B., Mumcu, E. F., Bakker, W., Akarsu, N., Kay, S. P. J., Hovius, S. E. R., Heredero-Baute, L., Oostra, B. A., and Heutink, P. (1999). Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localisation of the responsible gene to a 1.9 cM region on chromosome 7q36. *J. Med. Genet.* 36: 32–40.

Chapter 4b

A double RING-H2 domain in *RNF32*,

a gene expressed during sperm formation.

Marijke J. van Baren[1], Herma C. van der Linde[1], Guido J. Breedveld[1],
Willy M. Baarends[2], Esther de Graaff[1], Ben A. Oostra[1], and Peter Heutink[1]

[1]Department of Clinical Genetics and [2]Endocrinology and Reproduction, Erasmus University
Rotterdam, Rotterdam, 3000 DR, The Netherlands [1]Department of Clinical Genetics and
[2]Department of Urology and [4]Department of Pathology, Erasmus University Rotterdam,
the Netherlands

Abstract

The RING domain is a cysteine rich zinc-binding motif, which is found in a wide variety of proteins, among which are several proto-oncogenes and the gene implicated in autosomal recessive juvenile parkinsonism, *Parkin*. The domain mediates binding to other proteins, either via their RING domains or other motifs. In several proteins, RING domains are found in combination with other cysteine rich binding motifs and some proteins contain two RING domains. Recent evidence suggests that RING finger proteins function in the ubiquitin pathway as E3 ligases. A variant of the RING domain is the RING-H2 domain, in which one of the cysteines is replaced by a histidine. We have cloned and characterised a novel gene, *RNF32*, located on chromosome 7q36. *RNF32* is contained in 37 kb of genomic DNA and consists of 9 constitutive and 8 alternatively spliced exons, most of which are alternative first exons. A long and a short transcript of the gene are expressed; the short transcript containing exons 1-4 only. This gene encodes two RING-H2 domains separated by an IQ domain of unknown function. This is the first reported gene with a double RING-H2 domain. In humans, *RNF32* overlaps with a processed retroposon located on the opposite strand, *C7orf13*. RNF32 is specifically expressed in testis and ovary, whereas *C7orf13* is testis-specific, suggesting that its expression may be regulated by elements in the *RNF32* promoter region. *RNF32* is expressed during spermatogenesis, most likely in spermatocytes and/or in spermatids, suggesting a possible role in sperm formation.

Introduction

RING finger domains (for Really Interesting New Gene) have generated a lot of interest in recent years, because of their widespread occurrence and their implication in human disease. The structure of this domain resembles that of the classical zinc finger, but instead of four cysteines, it contains seven cysteines (C1-C7) and one histidine (H1) (see fig. 1) (Freemont *et al.* 1991). This motif adopts a 'cross brace' structure in which one zinc atom is bound to C1, C2, C4 and C5 and another to C3, H1, C6 and C7 (Barlow *et al.* 1994, Borden *et al.* 1995). A subclass of the RING finger family is the RING-H2 domain, in which the fourth cysteine of the motif is replaced by a histidine residue (fig. 1) (Freemont 1993). The domains have been found in animal, plant and virus proteins, proteins located in both the nucleus and the cytoplasm.

Several RING containing genes are involved in the ubiquitin proteasome pathway as E3 ligases, the proteins that specify targets for ubiquitination. One of these is the CBL protein, which is necessary for desensitisation of the Epidermal Growth Factor (EGF) receptor. Mutations in the RING domain of CBL abolish ubiquitination of the EGF receptor (Waterman *et al.* 1999, Yokouchi *et al.* 1999). These and other findings have led to the proposal that all RING containing proteins function in ubiquitination (Freemont
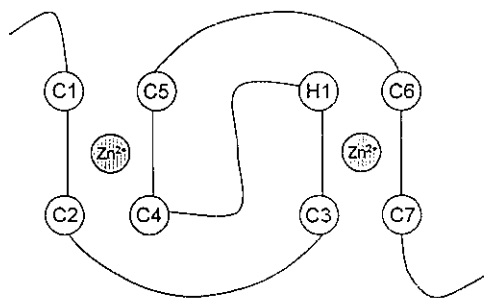
Figure 1. Schematic representation of the RING domain. Two zinc ions are bound by four ligands each: Four cysteines bind the first zinc ion, three cysteines and one histidine bind the second zinc ion. In the RING-H2 domain, the fourth cysteine of the domain is replaced by a histidine.

2000). However, not every reported RING protein contains a true RING domain; the 'RING' domains of two E3 ligases, RBX1 and APC2, do not follow the consensus C3HC4: RBX1 contains a cysteine-rich domain that can be described as C2HC3H2C2H, and the many cysteines and histidines in its homolog APC2 are even less conserved. It is unclear if these domains function in the same way as true RING domains.

RING proteins are often found in large protein complexes, and they are difficult to analyze because they tend to aggregate. This has led to the suggestion that they are involved in the formation of macromolecular complexes (Borden 2000). RING domains can bind to other RING domains and thereby result in homo- or heterodimerization, or they can associate with non-RING domains. For example, the RING of BRCA1 interacts with the BARD1 (BRCA1-associated RING domain 1) RING finger protein and with the BAP1 (BRCA1 associated protein-1) protein, which does not contain RING domains (Jensen *et al.* 1998, Meza *et al.* 1999).

Mutations in the RING domain of BRCA1 have been found in breast cancer patients (Miki *et al.* 1994, Takahashi *et al.* 1995). Several other human proto-oncogenes contain RING domains, for example human c-Cbl becomes oncogenic after deletion or disruption of its RING finger (Blake *et al.* 1991, Langdon *et al.* 1989). The domain is also found in the promyelocytic leukaemia protein PML and the transcriptional intermediary factor TIF1, which become oncogenic after chromosomal translocation resulting in their fusion to other proteins (de The *et al.* 1990, Le Douarin *et al.* 1995). Not all RING-containing disease genes are proto-oncogenes: for example, mutations in the RING domains of Parkin result in autosomal recessive juvenile parkinsonism (MIM 600116) (Kitada *et al.* 1998). Parkin also functions as an E3 ligase, although its target is unknown (Zhang *et al.* 2000).

Parkin contains two RING domains separated by another cysteine rich motif, the IBR (for In Between Ring domains) (Morett and Bork 1999). This motif is also found in the mouse Rbck1 (for RBCC protein interacting with Pkc1) protein and in the *Drosophila melanogaster* ariadne protein. The function of the IBR domain is unknown, but the integrity of the motif is necessary for Parkin function (Morett and Bork 1999). Proteins with more than one RING domain all contain the IBR box. No proteins with a double RING-H2 domain have been reported.

We report the cloning and characterisation of a novel human gene, *RNF32*. This gene is mainly expressed in testis. The putative protein of *RNF32* contains a novel motif: two RING-H2 domains separated by an IQ domain.

Results and Discussion

In a screen for genes localised around marker *D7S559* on chromosome 7q36, we used exon trapping and cDNA selection (Heus *et al*. 1999). We extended this search by analysis of approximately 400 kb of genomic sequence, using several gene prediction programs in combinations with dbEST (see Material and Methods). Here we describe two novel genes, *RNF32* and *C7orf13*. The genes are localised on opposite strands and overlap partially (fig. 2).

Genomic organisation of *RNF32*

RNF32 (AF441222, AF441223, AF441224 and AF441225) comprises 37 kb of genomic DNA between markers *D7S168* and *D7S559* and consists of 9 consistent and 8 alternatively spliced exons (see below) (fig. 2). Of the alternative exons, 7 are alternative first exons and the other is an exon that is sometimes found between exons 3 and 4, hence named exon 3a. All splice sites match the consensus (5' AG/N, 3' N/GT) with a polymorphism in exon 8; in 4 out of 42 alleles the less frequently used N/GC was found (Jackson 1991). A polyadenylation signal (AATAAA) was found 1372 nt into exon 9. The size of the putative cDNA is between 1658 and 1844 bp, with a 1086 nt open reading frame (ORF), starting in exon 2. The sequence around the putative start site, GGCATGT, contains the important purine in the −3 position, but does not otherwise fit the Kozak sequence (Kozak 1987).

Alternative splices of *RNF32*

Based on the predicted sequence, we designed primersets in exons 3 and 8 and performed PCR on cDNA derived from human lymphocytes. With the resulting PCR products we screened a human testis cDNA library and selected 19 clones that hybridized with both probes. These were sequenced and seven alternative first exons were found, all of which
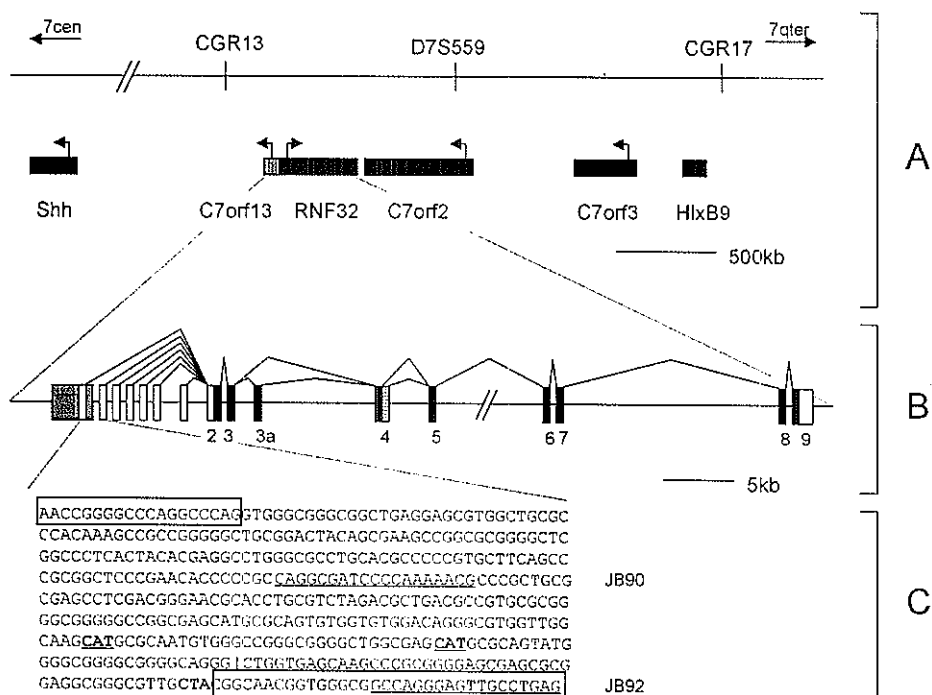
Figure 2. Schematic representation of *RNF32* and *C7orf13*.

A. Localisation of the genes on chromosome 7q36 with respect to other genes reported in this region (Heus *et al*. 1999). Orientation of genes is indicated by arrows. The orientation of HlxB9 is unknown. *C7orf13* overlaps one of the alternative exons 1 of *RNF32* and is indicated by a grey box.

B. Genomic organisaton of *RNF32* and *C7orf13*. The longer version of *RNF32* exon 4 is indicated by a hatched box; coding exons are in black, noncoding exons in white. For reasons of clarity, distances between alternative exons 1 are not drawn to scale; in reality they lie close together. *C7orf13*, in grey, overlaps the first alternative exon 1 of *RNF32* and is located on the opposite strand.

C. Sequence around the 5' end of *C7orf13*. *RNF32* exons are boxed and primer sequences are underlined. The sequence between the two primers is shown with the in frame methionine codons on the opposite strand in bold and underlined, and the preceding in frame stop codon in bold. The putative splice site is indicated by a vertical line.

could be aligned to the genomic sequence. In addition, several Expressed Sequence Tags (ESTs) aligned with the predicted *RNF32* (see Materials and Methods). In two of the ESTs (aa968469 and aa758018) and one of our own clones, we found an additional exon 3a of 107 bp (AF441224). The two ESTs end after exon 4, including 219 and 250

100

nucleotides after the usual splice site and both contain a polyadenylation signal (AATAAA) 196 bp after the splice site. The clone contains all the normal RNF32 exons, including one of the alternative exons 1. To confirm that exon 3a is transcribed, we performed PCR on cDNA with primers in exons 3 and 3a and found a product of the correct size (data not shown).

Two other clones from the cDNA library (neither containing exon 3a) contain an additional 17 bp after exon 4, where an alternative splice site is located (AF441225). These clones do not contain a polyadenylation signal in exon 4 and they contain the other RNF32 exons.

Structure of the RNF32 protein

The predicted protein is 362 aa in size and contains two RING-H2 domains separated by an IQ domain (fig. 3A). Both RING domains follow the consensus $CX_2CX_{9-39}CX_{1-3}HX_{2-3}HX_2CX_{4-48}CX_2C$ (Saurin $et\ al.$ 1996). This is the first report of a gene that encodes a double RING-H2 protein. It is not known what the functional difference between the RING and RING-H2 domains is, but since RING containing proteins are important in protein-protein interactions, it is possible that a protein with a double RING domain acts as a scaffold for binding several proteins that function in the same pathway. This would bring these proteins close together, allowing them to act cooperatively.

The IQ domain (consensus: IQX_3RGX_3R) most resembles one of five IQ domains in mouse myosin heavy chain dilute, but in contrast to unconventional myosins such as this one, RNF32 contains only one IQ domain. Single IQ domains are also found in neuromodulin and neurogranin and, as in unconventional myosins, in these proteins the domain binds calmodulin (Cheney and Mooseker 1992). The motif has been found in other proteins that interact with calmodulin, but also in many other proteins that do not. Some of the non-calmodulin interacting proteins are GTP regulatory and cell cycle proteins, others are receptors or channel proteins (Rhoads and Friedberg 1997). It is not known what function, if any, the IQ domain has in these non-calmodulin interacting proteins. It remains to be determined whether the IQ domain in RNF32 binds calmodulin, and what the functional importance of this domain may be.

A

```
cagaagaatagaaggaaggtgataggatgtgatgatagaatttgtgatagccaagcaacaac
                 2|3
ttttcctaattcggcatgttaaaaaataag|ggtcactcatctaagaaagataacttggca
                                  M   L   K   N   K   G   H   S   S   S   K   K   D   N   L   A      15
gtcaatgcagttgctttacaagatcacatttttacatgatcttcaacttcgaaatctttca
 V   N   A   V   A   L   Q   D   H   I   L   H   D   L   Q   L   R   N   L   S      35
gttgcagatcattctaagacacaagtacaaaagaaagagaacaaatctctaaaaagagat
 V   A   D   H   S   K   T   Q   V   Q   K   K   E   N   K   S   L   K   R   D      55
acaaaggcaataatagatactggacttaaaaaaactacacagtgccccaaactagaagac
 T   K   A   I   I   D   T   G   L   K   K   T   T   Q   C   P   K   L   E   D      75
                                                       3|4
tcagaaaaagaatatgttcttgatcccaaaccgccgccgttgactttgg|cacagaagttg
 S   E   K   E   Y   V   L   D   P   K   P   P   P   L   T   L   A   Q   K   L      95
ggcctcattgggcctccaccacctccactgtcatcagatgaatgggagaaggtgaaacag
 G   L   I   G   P   P   P   P   P   L   S   S   D   E   W   E   K   V   K   Q      115
cgctctctcctgcaagggggactccgtgcaaccatgccccatctgtaaagaagaattcgag
 R   S   L   L   Q   G   D   S   V   Q   P   C   P   T   C   T   E   K   F   E      135
       4|5                                       5|6
cttcgtcctcag|gtgctgctttcatgctcccatgtgttccacaaa|gcatgtcttcaggct
 R   P   Q   V   L   L   S   C   S   H   V   F   H   K   A   C   L   Q   R      155
tttgaaaagttcacaaataagaaaacctgtcctctctgtagaaagaaccagtatcaaacc
 F   E   K   F   T   N   K   K   T   C   P   L   C   R   K   N   Q   Y   Q   T      175
                                                     6|7
cgagtgatacacgatggggcccgcctgttcagaatcaagtgtgtgaccag|aatccaagcc
 R   V   I   H   D   G   A   R   L   F   R   I   K   C   V   T   R   I   Q   A      195
tactggagaggatgtgttgttagaaagtggtacagaaacctgaggaaaacagtacctccc
 Y   W   R   G   C   V   V   R   K   W   Y   R   N   L   R   K   T   V   P   P      215
                                                     7|8
acagatgccaagttaagaaaaaaattctttgaaaaaaag|ttcacagaaatcagccaccgc
 T   D   A   K   L   R   K   K   F   F   E   K   K   F   T   E   I   S   H   R      235
atcctgtgctcatacaacaccaacattgaagagctctttgcagaaatcgatcagtgcttg
 I   L   C   S   Y   N   T   N   I   E   E   L   F   A   E   I   D   Q   C   L      255
gccataaatcgaagtgttcttcagcagttggaagaaaaatgtggccatgagatcacagaa
 A   I   N   R   S   V   L   Q   Q   L   E   E   K   C   G   H   E   I   T   E      275
                                 8|9
gaggaatgggagaaaatccaagtgcag|gctctgcgccgggagacccacgagtgctccatc
 E   E   W   E   K   I   Q   V   Q   A   L   R   R   E   T   H   E   C   S   I      295
tgcctggcccctctctccgctgctggcggtcagcgcgtgggtgcaggcaggcgttccaga
 P   A   P   S   A   R   C   Q   R   V   E   A   G   R   R      315
gagatggcccctgtcctgctcacatgtgttccaccatgcgtgtctgctggcactagag
 M   A   P   S   V   S   S   V   R   H   A   K   L   E   A   L      335
gagttctccgtgggagacaggcctcctttccatgctgctcctctctgccgctcctgctac
 C   S   V   G   R   P   P   F   H   A   C   S   L   R   S   C   Y      355
cagaagaagattcttgaatgttgaattcatagtcaaggaaagttaggtaattctgaggaa
 Q   K   K   I   L   E   C   -                                    362
aaaagtttaccatcattttggatgaactgcatgagttctgggttaagtactacaatgtaa
tctgtttcycagggaaataagctattggtagttgtaggaaatcttagtatattttaaaag
ctgacatcccacctaattttaatctttggtctctaaaaagtaaatttcaaatttatgagt
ttaatcacttcaaatatgaatagcaaaaaatgagagcttgcttacttctaaaaaaytgagg
ttaagatatagctagtgtctgaacgacactccttaaagtaagttccaaatgtaaaacact
ccttaagttccaaatgtttttccgctaatagtctgtcctaaagcctttgccattcctaata
ctcgttttgtaataattgctgtatttctgtgtaataaaatataaaaataaaatattcagt
ggtattcaacatcaaaaaa
```

B

```
human  MLKNKGSSKKDNLAVNAVALQDHLHDLQLRNLSVADHSKISVQKKE--NKSLKRDIKA  58
mouse  MLRNKDRSSNKGNLAVTAVALQDHLHDLQLGLSVADPCKIRARKKEKKSKSLKRDATA  60

human  IIDTGLKKTCCPKMEDSEKEYVLDPKPPPLTLAQKLGLGPPPPPLSSDEWEKVKQRSL  118
mouse  IIDTGLRKSTGPNMEDPEKEYVLDPTPPPLTLAQKLGLPPPPLPLSSDEWERVKQRSL  120

human  LQGDSWQPCPICKEEFELKPQVLLSCSHVFHKACLQAFEKFTNKKTCPLCRKNQYQTRVI  178
mouse  LQGDSIQPCPICKEEFELKPQVLLSCSHVFHRACLQAFEKFTNKKTCPLCRKNQYQTRVI  180

human  HDGARLFRIKCVTRIQAYWRGCNVRKWYRNLRKTVPPIDAKLRIKFFEIKFTEISHRILC  238
mouse  HDGARLFRVKCATRIQAYWRGYIVRKWYRNLRKILPPIDAKLRIKFFEIKFTEISHRILC  240

human  SYNTNIEELFIEIDQCLANRSVLQQLIEKCGIEITIEIWEKIQVQALIRETIECSICLI  298
mouse  SYNTNIDELFSEIDVCLAVNRSTLQQLIERCGQITIIEDWEKIQAQAAIIEIMECSICLI  300

human  PLSAAGGQR---MG-AGRRSREMALLSCIHIFHIACLLALEEFSIGDRPPFHACPLCRSC  354
mouse  PLSFHGDGRQAAIGTISSORPRETVLLSCAHIFHNACLLALEEFSIGDNAPFHVCPLCRSC  360

human  YQKKIIEC  362
mouse  YQKKIIEC  368
```

Figure 3A. Human RNF32 protein and DNA sequence, starting with exon 2. RING–H2 domains are shaded, the IQ domain is underlined and the consensus residues are in bold. Splice sites are indicated by a vertical line.
B. Conservation of the predicted amino acid sequence between *RNF32* and mouse *Rnf32*. Identical residues are in black, similar residues are in grey.

The exon 1 splice variants of *RNF32* do not alter the predicted ORF, since the start codon is located in exon 2. However, exon 3a contains a stop codon, leading to truncation of the protein before the first RING domain. In the transcript that ends with the extended exon 4, a stop codon is present 16 bp after the 3' splice site used in the normal exon 4, truncating the protein within the first RING-H2 domain This leaves an ORF of 145 aa (AF441223). The transcripts that use the alternative splice site after exon 4 result in a frameshift, which results in a stop codon in exon 7 and a 210 aa ORF. The frameshift occurs in the first RING-H2 domain.

RNF32 homologs

The mouse ortholog of *RNF32* is *Lmbr2* (Clark *et al.* 2000), which is now renamed *Rnf32*. Alignment of the two predicted protein sequences is shown in fig. 3B. The mouse protein is 76.6% identical to the human protein, the DNA sequence from the ORFs is 79.3%

identical. No significant homology was found in the untranslated regions (UTRs). Using mouse primersets on cDNA from rat brain and kidney tumours, we identified part of rat Rnf32 (data not shown).

A database search for sequences homologous to the RNF32 sequence revealed only one other sequence with a predicted double RING-H2/IQ structure. This sequence is an incomplete, single sequence read from *Giardia intestinalis*, a parasite in the human intestine (AC037512). Translation of this sequence predicts two RING-H2 domains separated by a single IQ domain, the same conformation as found in RNF32. There is no sequence homology between RNF32 and this predicted protein outside the domains.

Expression of *RNF32*

The expression pattern of RNF32 was analysed using Northern blots containing multiple human adult tissues. Two transcripts of 0.8 and 1.9 kb were detected in testis and, less abundant, in ovary (fig. 4A). No expression was detected in any other tissue (see Material and Methods). Exons 2-4 hybridised to both bands and exons 5-8 only to the larger 1.9 kb band, indicating that indeed a short and a long transcript of *RNF32* exist. The intensities of the bands seen in fig. 4A indicate that both transcripts are present in equal amounts.

The found transcript sizes correspond to the size of the bands on the Northern blots, indicating that we have determined the full length transcript. We did not isolate the shorter transcript from the cDNA library because we did specifically select for clones that contained both exon 3 and exon 8, however, the finding of the smaller band is consistent with the ESTs that contain a polyadenylation signal after exon 4.

Additionally, a stop codon and a polyadenylation signal were found in exon 9, and we did not find any clone or EST with exons beyond exon 9, and no such exons were predicted by the gene prediction programs.

The expected size differences due to alternative splicing of the *RNF32* transcript are too small to detect on these Northern blots (107-250 additional nucleotides), therefore it is not possible to see if exon 3a is inserted and which version of exon 4 is used.

To further characterise testis expression of *RNF32*, we made use of its mouse homolog. In adult mice, spermatogenesis occurs continuously, with all stages of sperm development present at each time point. The first wave of spermatogenesis starts at birth. Meiosis is initiated around day 7 and the first haploid round spermatids can be detected around day 21. A Northern blot containing testis RNA of mice aged 7, 14, 21 and 36 days was used
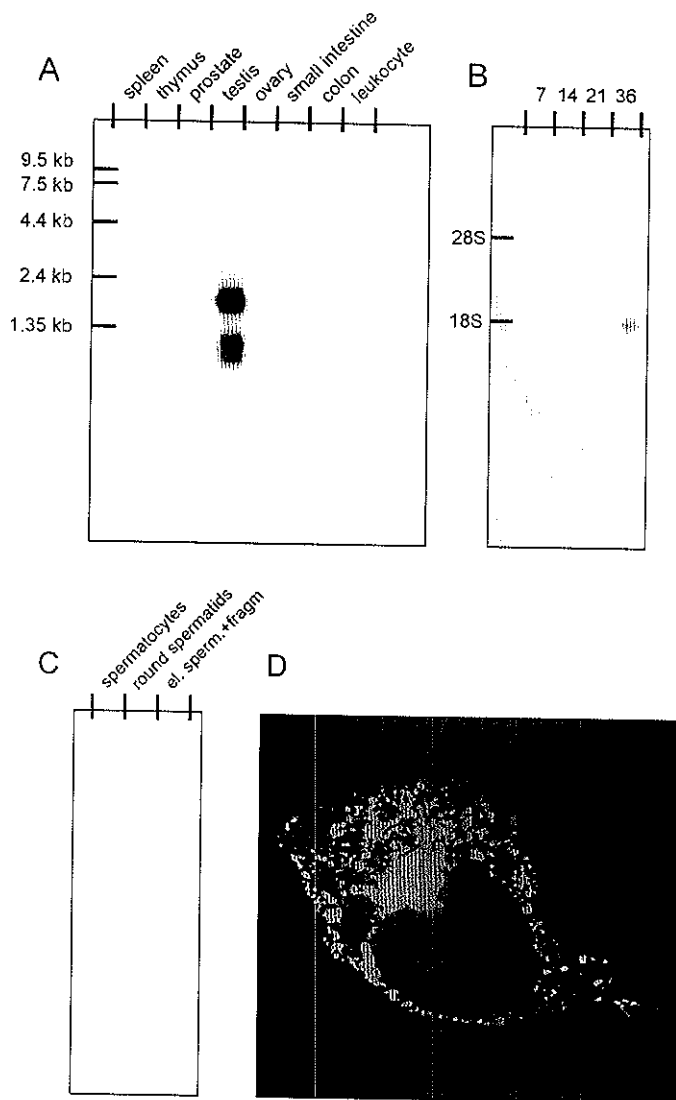
Figure 4. Expression of *RNF32*.
A. Hybridisation with exon 3 on human multiple tissue Northern blot. Two transcripts of *RNF32* are seen in testis. After prolonged exposure, a band in ovary is seen (not shown).
B. On a blot containing testis RNA of mice aged 7, 14, 21 and 36 days, *Rnf32* signal is seen in testis from 21- and 36-day-old-mice.
C. *Rnf32* is expressed in all three germ cell fractions: spermatocytes, round spermatids and the elongated spermatocytes/cytoplasmic fragments (el. sperm.+fragm).
D. His tagged *RNF32* localises to the cytoplasm and forms aggregates in cultured COS cells.

for hybridisation with a mouse *Rnf32* probe. Signal was detected in testis aged 21 and 36 days (fig. 4B), indicating that transcription of *Rnf32* starts late in the meiotic prophase or shortly after meiosis. Based on these results, it cannot be excluded that *Rnf32* is expressed only in somatic cells in the testis from postnatal day 21 onwards. Therefore, we isolated total RNA from different testicular cell fractions: one fraction highly enriched for spermatocytes going through the prophase of meiosis, one fraction containing postmeiotic round spermatids and one fraction containing elongated spermatocytes and cytoplasmic fragments. Northern blot analysis using the same Rnf32 probe, resulted in a signal in all fractions with the strongest signal in round spermatids (fig. 4C). This shows that *Rnf32* is indeed expressed in germ cells.

To determine the intracellular localisation of RNF32, we transfected COS and CHO cells with a RNF32-His tag construct containing the long transcript. After 24 h, the fusion protein was detected mainly in aggregates in the cytoplasm (fig. 4D). These results confirm the findings of other groups that RING containing proteins are often found in aggregates (see Borden *et al.* 2000). It remains to be determined if RNF32 functions in the ubiquitin pathway.

C7orf13

Several ESTs map proximal to *RNF32*. The sequence of these clones does not align with *RNF32* and they align to the genomic sequence without gaps, indicating that no intron is present. An open reading frame is predicted in the opposite direction from *RNF32*, indicating that a separate transcript maps to the other DNA strand. We designated this transcript *C7orf13* (AF439977)(fig. 5).

To identify the start of the coding sequence, we used the genomic sequence to design primers proximal to the known sequence and used these to perform PCR on cDNA derived from human lymphocytes, with a primer located within the known sequence, JB69 as reverse primer. The most proximal primer that still gave a product was JB90, whereas primer JB92, located within the second alternative exon 1 of *RNF32,* did not produce a band (data not shown), indicating that the exon starts somewhere between those primers. Interestingly, JB90 is located within the genomic region of *RNF32*, after the most proximal alternative exon 1 (fig. 2C). This means that *C7orf13* overlaps *RNF32.*

```
atgctcgccagccccgcccggcccacattgcgcatgcttgccaaccacgccctgtccaca
 M   L   A   S   P   A   R   P   T   L   R   M   L   A   N   H   A   L   S   T        20
ccacactgcgcatgctcgccggcccccgccccgcgcacggcgtcagcgtctagacgcagg
 P   H   C   A   S   P   A   P   A   P   R   T   A   S   A   S   R   R   R        40
tgcgttcccgtcgaggctcgcgcagcgggcgtttttggggatcgcctggcgggggtgttc
 C   V   P   V   E   A   R   A   A   G   V   F   G   D   R   L   A   G   V   F        60
gggagccgcgggctgaagcacggggggcgtgcaggcgcccaggcctcgtgtagtgagggcc
 G   S   R   G   L   K   H   G   G   V   Q   A   P   R   P   R   V   V   R   A        80
gagccccgcgccggcttcgctgtagtccgcagccccggcggctttgtgggcgcagccac
 E   P   R   A   G   F   A   V   V   R   S   P   R   R   L   C   G   R   S   H        100
gctcctcagccgcccgcccacctgggcctgggccccggttgtttccctgcggtggctgtg
 A   P   Q   P   P   A   H   L   G   L   G   P   G   C   F   P   A   V   A   V        120
gtcgttcccgttccagggtctcgagcgcaccgcccttcgctgcgctgttagtcgagggc
 V   V   P   V   P   G   S   R   A   H   R   P   F   A   A   L   L   V   E   G        140
agtttttctggggggatcccccgatcccccgcgccgctcaggcgtcttggctcgcggaagc
 S   F   L   G   D   P   P   I   P   P   R   R   S   G   V   L   A   R   G   S        160
gcaggcgccgactgcctggcgtcgtccggtgactcctggcccttctctgtggatcccttg
 A   G   A   D   C   L   A   S   S   V   T   P   G   P   S   L   W   I   P   L        180
cttctcgttgcgggttgtgtttcctgcttcgtgggcctggccgtttgcgtttggatgcag
 L   L   V   A   G   C   V   S   C   F   V   G   L   A   V   C   V   W   M   Q        200
gcccgggtgagccctgcgtggcccgctggccttttcctgctccctaggtgactgagccgc
 A   R   V   S   P   A   W   P   A   G   L   F   L   L   P   R        216
gtgtgtgacacatcggagtgaattgaggcaacgcgttcctttccaggcttgcagcgggcg
gtgtcgttcacgctcctgagcagagccttccggtgctttactcgacgccacaaatagtgc
agggttcctcccagggggctgtcgcgagctagaactcttccccgccctgtgagctcgggc
tgttcttcctcctccgcgtggctcactccagcctcggatgtttcctcgctgcacagacct
gctcgggaagactcgagaaggccacccccaggactccagctctccccggaagggcttcccc
tcggattgcgtcttggcccctggaaccctccctctgtatcctctgctgttcctctcttgg
cctccatttgggttgtggcctggatactctggctcgctgcttcccttctcttaaggatca
cagccctgcactgcctgtccggtgcctggaaaccattgtatttttttggtctgatttt
gtagttgtttgagaacttaagtctggtccctcacactccatgcccagaagcagaaatgag
taaagtttgtttatcttttcttaatatgacaattattgtgttggttcaacttatgttgt
actttaattagaagaaatttggccgaaaatacaaggaaaatatacaaatgcaagtaattt
tttttaaacttccctgaaagcagggtctaaagaaattaccaaccaacttagactggatct
agaagaaaaggaagggtctttgcagtcttaggactcttccgttccgcgacgtaggtgtta
ggataacagccatatatggttgtaagactttgggctcagataagtagacttaagttcaaa
ttttgacttattttacaagtgtgtgattttttggcaagctcatcttcctaaaccatgagct
ccttatttgtaaaggggacattagccactctccagcaacagccctggtacttcttcagtc
ctgggatgggacgtatgattagcctaagcgaaccagaaaatccaggcccgtgaccagtga
cttgatcagggctggcgatatatctaggtaggcaaccaggtggactcagtattttttgtg
ggtgctactggaaaatttatttaattctaactgaatgtagaaacagcaacagacatgaaa
tggcagttgtattgctgtcttatcatgaggtgagggcctgaagctatggtagccaccctg
tgaaccttggaaggagggttctacaggaactggcagagctgagactgggacgcaacccat
gtcctggtgacgtatttgagtcctgaggctggccctcctgttagtcttttctcttactgg
cttcagccagatggggtctggttttctgttacttgcaacaggaaaaattggaatgatatc
tatgcatgctattaccgccttatggctgtggtgagattaaaacatgatactgtgggtgaa
ctgcttagtacagtgcctgctattctttatgttggttttaaggtattttttcttgaacttt
aaagttctaagaaaaattaccgttttaaaattcctaccctctgttcacatgtttaaattt
tcatgatagttttctgtctgcatgttttctattcaa<u>aataaa</u>ttggtcattcttcagtt
```

Figure 5. C7orf13 DNA sequence and predicted protein, starting from the first methionine after the stop codon (see Figure 2C). The polyadenylation signal is underlined.

Two methionines are present between JB90 and JB92, and a stop codon is preceding the most proximal methionine. The sequence around both ATGs match the Kozak sequence in the important −3 position, indicating that one of these methionines may be the starting point of translation. However, a putative splice site lies between the most proximal ATG and the stop codon and therefore it is possible that an additional exon exists, preceding the known ORF. We tried to isolate additional 5' sequence with race PCR, but we did not find additional exons.

The 216 aa ORF of *C7orf13* does not contain known domains and does not resemble any other proteins. A polyadenylation signal is present 2257 nt downstream of the first methionine and all ESTs contain a polyA tail 17-114 bp after this signal.

A *C7orf13* probe hybridised to a human multiple tissue Northern blot showed a band of 2.3 kb in testis only (fig. 6), indicating that we have determined the full-length transcript. The C7orf13 probe did not show any bands on a developing mouse Northern blot and we have not found any C7orf13 homologs in other species in Genbank or in our own libraries. Analysis of the mouse genomic sequence around exon 1 of *Rnf32* did not show a predicted *C7orf13* homolog and no homologies are present between the mouse and human sequences in this region. Because of this, and because C7orf13 is intronless, it could very well be derived from a retroposon (Vanin 1985). Still, the intact reading frame of *C7orf13* suggests that this is not a functionless pseudogene. It should be noted that the testis appears to express a relatively high number of functional retroposons (see Sedlacek *et al.* 1999 and references therein). Several of these genes encode variants of X-chromosomal genes, to compensate for the silencing of X (and Y) chromosomal genes in spermatocytes. In addition, a switch to the spermatogenic expression of retroposed genes is also observed for autosomal-to-autosomal retroposition (Kleene *et al.* 1998).

Both *RNF32* and *C7orf13* are mainly expressed in testis. Regulatory sequences are often found around the first exon of genes, and the physical overlap of *RNF32* and *C7orf13* may therefore cause the transcription of *C7orf13* in testis. The function of the alternative exons 1 of *RNF32* is unknown, but it is possible that they constitute a mechanism of regulation. In this respect, it will be interesting to see if the mouse homolog of *RNF32* also contains alternative splices.

RNF32 is the first protein reported with a double RING-H2 domain. Conventional RING domains have frequently been found in larger motifs that may contain other cysteine/histidine rich domains. Examples of these are the RBCC and TRAF motifs
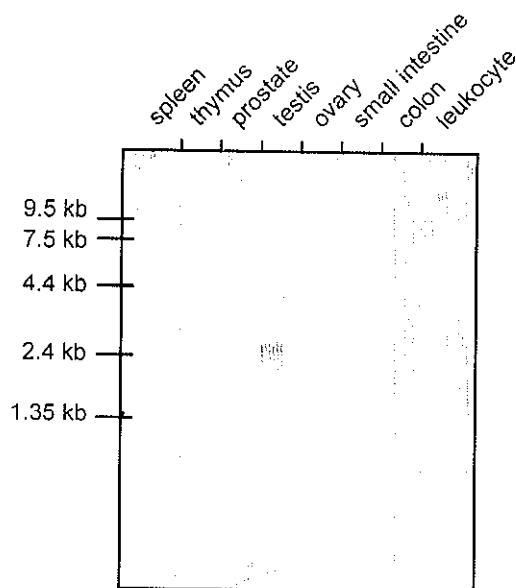
Figure 6. Testis-specific expression of *C7orf13* mRNA in humans.

and the RING-IBR-RING motif (Cheng *et al.* 1995, Kastner *et al.* 1992, Morett and Bork 1999, Reddy and Etkin 1991, Reddy *et al.* 1992). All proteins with more than one RING domain reported so far contain the IBR box (Morett and Bork 1999). In contrast, RNF32 contains and IQ domain between its RING-H2 fingers. This suggests that despite the similarities between RING-H2 and RING domains, these domains may have different functions.

Acknowledgements

Materials and Methods

Databases and software

PAC clones AC005534, AC007075, AC007097and AC006357 were used for gene prediction, which was performed using the Nix analysis software of HGMP (http://www.hgmp.mrc.ac.uk/). This program uses the prediction programs GRAIL, Fex, Hexon, MZEF, Genemark, Genefinder and Fgene. Predicted genes were aligned with ESTs found in the region using the Vector NTI program package, which was also used for *in silico* translation of the transcripts. Putative proteins were screened for domains using Pfam, accessible through ISREC (http://www.isrec.ch/). Sequences homologous to the putative proteins were found using the BLAST program available at NCBI (http://www.ncbi.nlm.nih.gov/).

ESTs assigned to *RNF32* are AA968469, AA758018, AA412403, AI024391, AA904554, AA412402 AA581432, N51460, N34954, R60655, N48673, AA626330, N53600, R60654, N44021, AI190316, AI472192 and AI217423 (UniGene cluster Hs.26791)

C7orf13 ESTs are AA631112, AI693868, AA868311, AI492920, AI168080, AI139964, AA912620, AI468759, AA378476, AA826489, AI971242, AI827946 and AW104902 (UniGene cluster Hs.124854).

cDNA library screening

The cDNA libraries screened were derived from Human Testis (Clontech, HL1010b) and Mouse Developing Limb, kindly provided by J.P. Evans. For each library, 1×10^6 plaques were plated, transferred to nylon filters and hybridized with the probes of interest using standard procedures (Sambrook *et al.* 1989).

Sequencing

Sequence data were obtained using the Ready Reaction Dye Terminator Cycle sequencig kit (Perkin-Elmer) on an ABI 377 automatic sequencer. Sequences were aligned and analysed using the DNASTAR program package.

RNF32 construct and transfection

A PCR product containing the ORF of *RNF32* was generated with primers that contained a *Bam*H1 recognition sequence

(ringbamF: TGAAACCGCGGATCCAAATAAGGGTCACTCATCTAAGAAAGA and ringbamR: TGAAACCGCGGATCCATTCAACATTCAAGAATCTTCTTCTG). The product was cloned into

the pCRII vector using the TA cloning kit of Invitrogen, digested with BamH1 and cloned into the pcDNA3.1HisB vector (Invitrogen).

COS cells were cultured in Dulbecco's minimal Eagle's medium (DMEM), and CHO cells were grown in α-MEM (Gibco BRL). Media were supplemented with 10% (v/v) fetal calf serum (Clontech), 100 IU/ml penicillin and 100 μg/ml streptomycin (both from Gibco BRL) and kept at 37°C in 5% CO_2.

Cells were transiently transfected with the *RNF32* construct using Lipofectamine reagent (Gibco BRL) according to the manufacturer's instructions.

Immunofluorescence

Cells were rinsed twice in cold phosphate-buffered saline (PBS), fixed and permeabilized in 100% methanol, 15 min at –20°C. The fixed cells were then incubated with mouse monoclonal antibodies against the Xpress epitope (1:100 dilution)(Invitrogen) for 1 hr at room temperature and, after washing in PBS with 0.5% bovine serum albumine and 0.15% glycine, incubated with secondary fluorescein-conjugated rabbit anti-mouse antibody (1:100)(Sigma). Cells were mounted in Vectashield mounting media (Vector) and examined with a Leica DMXRA fluorescence microscope equipped with a digital camera.

Northern blots

Human multiple tissue Northern blots MTN-I and MTN-II and Mouse Embryo (Clontech) were hybridized with $α^{32}P$-labeled probes and washed following manufacturer's instructions. The blots contained total RNA from the following tissues: MTN-I: heart, brain, lung, liver, muscle, kidney and pancreas; MTN-2: spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocyte; Mouse Embryo: Total RNA from 7, 11, 15 and 17 dpc embryo.

For the mouse testes blot, mouse testes (c57/B16) were isolated at different days after birth (7, 14, 21 and 36 days) and RNA was extracted using TRIzol reagent (Life Technologies). For the testis fractions blot, spermatocytes, round spermatids, and elongated spermatids/cytoplasmic fractions were isolated from mouse testes (C57Bl/6) after collagenase and trypsin treatment, followed by sedimentation at unit gravity (StaPut procedure) (Grootegoed *et al.* 1984). RNA was isolated by the LiCl/urea method (Sambrook *et al.* 1989). The amount of RNA was measured by UV spectrometry and 20 μg of total RNA was applied to eacht lane on 1% agarose gels; the separated RNA was then transferred to Hybond-N+ membranes and hybridized with the labeled probe.

Probes were prepared by PCR on cDNA (see below) and genomic DNA. Remaining template was removed by the Concert Rapid Gel Extraction System (Gibco BRL), running the PCR products on a 0.5% TAE gel, cutting out the bands and cleaning the DNA with glass beads (Bio101). *RNF32* probes were: an exon 3 probe (primers 3F: GTTGTATGTGCCATCAAGG and 3R: GCCCTAATAAGCCAAAAGATGA) and an exon 5-8 probe (primers 5F: CTTGTGCTGTGCTAAACAAACA and 8R: GCACTGATCGATTTCTGCAA). Probes from *Rnf32* and *C7orf13* were produced with T3 and T7 primers on IMAGE clones 515614 and 1158395 as templates, respectively.

cDNA production and amplification reactions

RNA was isolated from human lymphocytes with RNAzol (Campro Scientific) and 5µg of RNA was used for reverse transcription with Superscript II RT (Gibco BRL) following manufacturer's instructions.

PCR amplifications were performed in a total volume of 50 µl containing 50ng genomic DNA or 2 µl cDNA, 20 mM Tris HCl pH 8.4, 1.5 mM MgCl$_2$, 50 mM KCl, 200 µM dNTPs, 0.5 U Taq polymerase (Gibco BRL) and 10 pmol forward and reverse primers. Cycling conditions were 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at T$_{an}$ (4xG/C+2xA/T°C), 90 s at 72°C and a final step of 5 min at 72°C.

For confirmation of exon 3a, primers 3aF (GCCCTAATAAGCCAAAAGATGA) and 3aR (CTGGAACCTCTGGATGTTG) were used. Primersets designed on mouse Rnf32 were used to identify a rat RNF32 homolog. These primers were 4F (TGCTTGGATTTTCTCCCAGT) and 8R (GCAGTTGCCTTACAGGACCA) in mouse exons 4 and 8, respectively.

For 5' RACE PCR, the Marathon-Ready kit of Clontech was used following manufacturer's instructions. The probe specific pimer for *C7orf13* was ACGGGAACGCACCTGCGTCTAGA, and the cDNA template was derived from Human Fetal Brain.

112

References

Barlow, P. N., B. Luisi, A. Milner, M. Elliott, and R. Everett. 1994. Structure of the C3HC4 domain by 1H-nuclear magnetic resonance spectroscopy. A new structural class of zinc-finger. *J Mol Biol* 237: 201-11.

Blake, T. J., M. Shapiro, H. C. Morse, 3rd, and W. Y. Langdon. 1991. The sequences of the human and mouse c-cbl proto-oncogenes show v-cbl was generated by a large truncation encompassing a proline-rich domain and a leucine zipper-like motif. *Oncogene* 6: 653-7.

Borden, K. L. 2000. RING domains: master builders of molecular scaffolds? *J Mol Biol* 295: 1103-12.

Borden, K. L., M. N. Boddy, J. Lally, N. J. O'Reilly, S. Martin, K. Howe, E. Solomon, and P. S. Freemont. 1995. The solution structure of the RING finger domain from the acute promyelocytic leukaemia proto-oncoprotein PML. *EMBO J* 14: 1532-41.

Cheney, R. E., and M. S. Mooseker. 1992. Unconventional myosins. *Curr Opin Cell Biol* 4: 27-35.

Cheng, G., A. M. Cleary, Z. S. Ye, D. I. Hong, S. Lederman, and D. Baltimore. 1995. Involvement of CRAF1, a relative of TRAF, in CD40 signaling. *Science* 267: 1494-8.

Clark, R., P. Marker, and D. Kingsley. 2000. A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. *Genomics* 67: 19-27.

de The, H., C. Chomienne, M. Lanotte, L. Degos, and A. Dejean. 1990. The t(15;17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor alpha gene to a novel transcribed locus. *Nature* 347: 558-61.

Freemont, P. S. 1993. The RING finger. A novel protein sequence motif related to the zinc finger. *Ann N Y Acad Sci* 684: 174-92.

Freemont, P. S. 2000. RING for destruction? *Curr Biol* 10: R84-7.

Freemont, P. S., I. M. Hanson, and J. Trowsdale. 1991. A novel cysteine-rich sequence motif. *Cell* 64: 483-4.

Grootegoed, J. A., R. Jansen, and H. J. Van der Molen. 1984. The role of glucose, pyruvate and lactate in ATP production by rat spermatocytes and spermatids. *Biochim Biophys Acta* 767: 248-56.

Heus, H. C., A. Hing, M. J. van Baren, M. Joosse, G. J. Breedveld, J. C. Wang, A. Burgess, H. Donnis-Keller, C. Berglund, J. Zguricas, S. W. Scherer, J. M. Rommens, B. A. Oostra, and P. Heutink. 1999. A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. *Genomics* 57: 342-51.

Jackson, I. J. 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 19: 3795-8.

Jensen, D. E., M. Proctor, S. T. Marquis, H. P. Gardner, S. I. Ha, L. A. Chodosh, A. M. Ishov, N. Tommerup, H. Vissing, Y. Sekido, J. Minna, A. Borodovsky, D. C. Schultz, K. D. Wilkinson, G. G. Maul, N. Barlev, S. L. Berger, G. C. Prendergast, and F. J. Rauscher. 1998. BAP1: a novel ubiquitin hydrolase which binds to the BRCA1 RING finger and enhances BRCA1-mediated cell growth suppression. *Oncogene* 16: 1097-112.

Kastner, P., A. Perez, Y. Lutz, C. Rochette-Egly, M. P. Gaub, B. Durand, M. Lanotte, R. Berger, and P. Chambon. 1992. Structure, localization and transcriptional properties of two classes of retinoic acid receptor alpha fusion proteins in acute promyelocytic leukemia (APL): structural similarities with a new family of oncoproteins. *EMBO J* 11: 629-42.

Kitada, T., S. Asakawa, N. Hattori, H. Matsumine, Y. Yamamura, S. Minoshima, M. Yokochi, Y. Mizuno, and N. Shimizu. 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 392: 605-8.

Kleene, K. C., E. Mulligan, D. Steiger, K. Donohue, and M. A. Mastrangelo. 1998. The mouse gene encoding the testis-specific isoform of Poly(A) binding protein (Pabp2) is an expressed

retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes. *J Mol Evol* 47: 275-81.

Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* 15: 8125-48.

Langdon, W. Y., J. W. Hartley, S. P. Klinken, S. K. Ruscetti, and H. C. Morse, 3rd. 1989. v-cbl, an oncogene from a dual-recombinant murine retrovirus that induces early B-lineage lymphomas. *Proc Natl Acad Sci U S A* 86: 1168-72.

Le Douarin, B., C. Zechel, J. M. Garnier, Y. Lutz, L. Tora, P. Pierrat, D. Heery, H. Gronemeyer, P. Chambon, and R. Losson. 1995. The N-terminal part of TIF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18. *EMBO J* 14: 2020-33.

Meza, J. E., P. S. Brzovic, M. C. King, and R. E. Klevit. 1999. Mapping the functional domains of BRCA1. Interaction of the ring finger domains of BRCA1 and BARD1. *J Biol Chem* 274: 5659-65.

Miki, Y., J. Swensen, D. Shattuck-Eidens, P. A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L. M. Bennett, W. Ding, and e. al. 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266: 66-71.

Morett, E., and P. Bork. 1999. A novel transactivation domain in parkin. *Trends Biochem Sci* 24: 229-31.

Reddy, B. A., and L. D. Etkin. 1991. A unique bipartite cysteine-histidine motif defines a subfamily of potential zinc-finger proteins. *Nucleic Acids Res* 19: 6330.

Reddy, B. A., L. D. Etkin, and P. S. Freemont. 1992. A novel zinc finger coiled-coil domain in a family of nuclear proteins. *Trends Biochem Sci* 17: 344-5.

Rhoads, A. R., and F. Friedberg. 1997. Sequence motifs for calmodulin recognition. *Faseb J* 11: 331-40.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Saurin, A. J., K. L. Borden, M. N. Boddy, and P. S. Freemont. 1996. Does this have a familiar RING? *Trends Biochem Sci* 21: 208-14.

Sedlacek, Z., E. Munstermann, S. Dhorne-Pollet, C. Otto, D. Bock, G. Schutz, and A. Poustka. 1999. Human and mouse XAP-5 and XAP-5-like (X5L) genes: identification of an ancient functional retroposon differentially expressed in testis. *Genomics* 61: 125-32.

Takahashi, H., K. Behbakht, P. E. McGovern, H. C. Chiu, F. J. Couch, B. L. Weber, L. S. Friedman, M. C. King, M. Furusato, V. A. LiVolsi, and e. al. 1995. Mutation analysis of the BRCA1 gene in ovarian cancers. *Cancer Res* 55: 2998-3002.

Vanin, E. F. 1985. Processed pseudogenes: characteristics and evolution. *Annu Rev Genet* 19: 253-72.

Waterman, H., G. Levkowitz, I. Alroy, and Y. Yarden. 1999. The RING finger of c-Cbl mediates desensitization of the epidermal growth factor receptor. *J Biol Chem* 274: 22151-4.

Yokouchi, M., T. Kondo, A. Houghton, M. Bartkiewicz, W. C. Horne, H. Zhang, A. Yoshimura, and R. Baron. 1999. Ligand-induced ubiquitination of the epidermal growth factor receptor involves the interaction of the c-Cbl RING finger and UbcH7. *J Biol Chem* 274: 31707-12.

Zhang, Y., J. Gao, K. K. Chung, H. Huang, V. L. Dawson, and T. M. Dawson. 2000. Parkin functions as an E2-dependent ubiquitin- protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. *Proc Natl Acad Sci U S A* 97: 13354-9.

Chapter 5

Acheiropodia is caused by a genomic deletion in C7orf2

Marijke J. van Baren[1], Henk Heus[1], Guido J. Breedveld[1], Herma C. van der Linde[1], Ann Hagemeijer[2], Andreas Rump[3], Marcel Vermeij[4], Tiansen Li[5], Esther de Graaff, Ben A. Oostra[1] and Peter Heutink[1].

P. Ianakiev[1], M. J. van Baren[2], M. J. Daly[3], S. P.A. Toledo[4], M. G. Cavalcanti[5], J. Correra. Neto[5], E. Lemos Silveira[6], A. Freire-Maia[7], P. Heutink[2], M. W. Kilpatrick[1], and P. Tsipouras[1].

[1]Department of Pediatrics, University of Connecticut Health Center, Farmington, CT, USA
[2]Department of Clinical Genetics, Erasmus University Rotterdam, the Netherlands
[3]Whitehead Institute for Biomedical Research, Cambridge, MA, USA
[4]LIM/25-D, University of Sao Paulo School of Medicine and 5Private Practice, Sao Paulo, Brazil.
[6]Private Practice, Porto Alegre, Brazil.
[7]Department of Genetics, UNESP-Universidade Estadual Paulista, Botucatu SP, Brazil.

Acheiropodia Is Caused by a Genomic Deletion in C7orf2, the Human Orthologue of the *Lmbr1* Gene

P. Ianakiev,[1] M. J. van Baren,[2] M. J. Daly,[3] S. P. A. Toledo,[4] M. G. Cavalcanti,[4] J. Correa Neto,[5] E. Lemos Silveira,[6] A. Freire-Maia,[7] P. Heutink,[2] M. W. Kilpatrick,[1] and P. Tsipouras[1]

[1]Department of Pediatrics, University of Connecticut Health Center, Farmington, CT; [2]Department of Clinical Genetics, Erasmus University, Rotterdam; [3]Whitehead Institute for Biomedical Research, Cambridge, MA; [4]LIM/25-D, University of Sao Paulo School of Medicine, and [5]Private Practice, Sao Paulo; [6]Private Practice, Porto Alegre, Brazil; and [7]Department of Genetics, UNESP-Universidade Estadual Paulista, Botucatu SP, Brazil

Acheiropodia is an autosomal recessive developmental disorder presenting with bilateral congenital amputations of the upper and lower extremities and aplasia of the hands and feet. This severely handicapping condition appears to affect only the extremities, with no other systemic manifestations reported. Recently, a locus for acheiropodia was mapped on chromosome 7q36. Herein we report the narrowing of the critical region for the acheiropodia gene and the subsequent identification of a common mutation in C7orf2—the human orthologue of the mouse *Lmbr1* gene—that is responsible for the disease. Analysis of five families with acheiropodia, by means of 15 polymorphic markers, narrowed the critical region to 1.3 cM, on the basis of identity by descent, and to <0.5 Mb, on the basis of physical mapping. Analysis of C7orf2, the human orthologue of the mouse *Lmbr1* gene, identified a deletion in all five families, thus identifying a common acheiropodia mutation. The deletion was identified at both the genomic-DNA and mRNA level. It leads to the production of a C7orf2 transcript lacking exon 4 and introduces a premature stop codon downstream of exon 3. Given the nature of the acheiropodia phenotype, it appears likely that the *Lmbr1* gene plays an important role in limb development.

Introduction

Acheiropodia (MIM 200500) is a unique condition presenting with bilateral congenital amputations of the upper and lower extremities and with aplasia of the hands and feet (Toledo and Saldanha 1969). It is distinguished from other hemimelias by a specific pattern of malformations, consisting of a complete amputation of the distal epiphysis of the humerus, amputation of the distal part of the tibial diaphysis, and aplasia of the radius, ulna, fibula and of the carpal, metacarpal, tarsal, metatarsal, and phalangeal bones (Toledo and Saldanha 1969, 1972). There appears to be little variability of expression (Grimaldi et al. 1983); however, in some affected individuals an ectopic bone (Bohomoletz bone) has been found at the distal end of the humerus (Toledo and Saldanha 1969). This severely handicapping condition appears to affect only the extremities, with no other systemic manifestations reported. With the exception of two affected siblings in Puerto Rico, the reported

cases are of Brazilian origin (Kruger and Kumar 1994). The incidence of acheiropodia in Brazil has been estimated to be ~1/250,000 births (Freire-Maia et al. 1975a). Acheiropodia is inherited as an autosomal recessive trait, and heterozygotes are phenotypically normal (Freire-Maia et al. 1975a). The vast majority of affected individuals are the offspring of consanguineous matings. The etiology of acheiropodia has remained obscure. Earlier studies ruled out abnormalities in glycosaminoglycan metabolism (Mourao et al. 1977). In a recent report, a locus for acheiropodia was mapped on chromosome 7q36 (Escamilla et al. 2000), to a region that overlaps with the preaxial polydactyly (PPD [MIM 174500]) and triphalangeal thumb loci (MIM 190605) (Heutink et al. 1994; Tsukurov et al. 1994). Herein we report the narrowing of the critical region for the acheiropodia gene and the subsequent identification of a common mutation in C7orf2 (GenBank accession number AF107454), the human orthologue of the mouse *Lmbr1* gene responsible for the disease.

Subjects and Methods

Clinical Phenotype

The family panel comprises five families with acheiropodia, each consisting of an affected individual and

38

his or her unaffected relatives. The families originate from the states of Sao Paulo (families 1 and 3–5) and Rio Grande do Sul (family 2) in Brazil (fig. 1). The phenotype of the proband in family 1 was characterized by truncation of the distal humeri, absence of the Bohomoletz bone, truncation of the tibiae, and absence of the fibulae. The phenotype of the proband in family 2 has been described elsewhere and is defined by bilateral truncations of the distal humeri and the tibiae, absence of the fibulae, and presence of the Bohomoletz bone at the distal aspect of the humeri (Lemos Silveira and Freire-Maia 1998). The phenotype of the proband in family 3 is similar to that of family 1. Pedigree 4 has been de-

scribed elsewhere (Toledo and Saldanha 1969). The affected individuals have no feet, hands, or forearms. Some of the affected individuals possess one digit implanted into each upper extremity, either unilaterally or bilaterally (fig. 2). The phenotype of the proband in family 5 is characterized by the presence of small fingerlike appendages, bilaterally, which, on x-ray have been recognized as the Bohomoletz bone.

Haplotype Analysis

Peripheral blood from individuals affected with acheiropodia and from their relatives was taken after in-
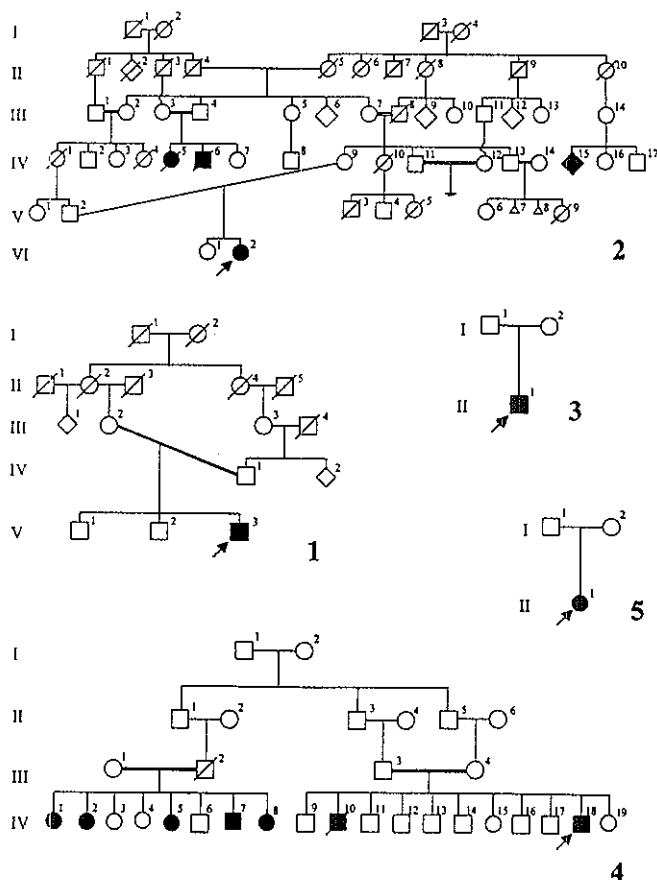
Figure 1   Pedigrees of panel of families with acheiropodia that were used in this study
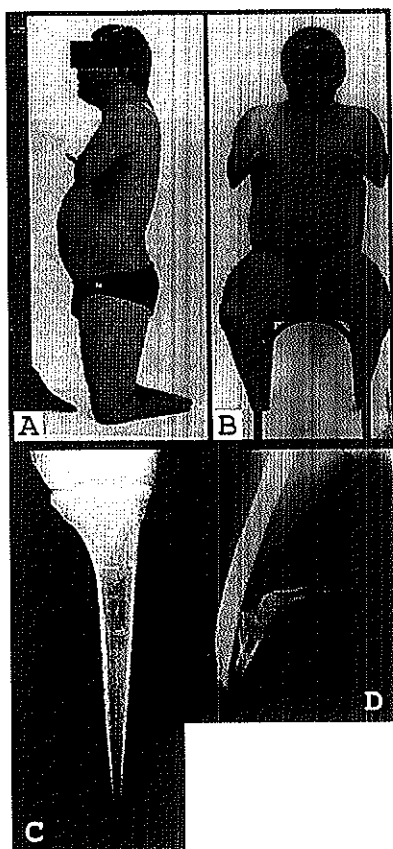
Figure 2 Affected individual (IV-18) from family 4. Note the small fingerlike appendages at the end of the arm, which are present bilaterally (A and B). C, Radiograph showing tapered amputation of the distal tibia. The proximal tibial epiphysis is well preserved. D, Radiograph showing dysplastic distal humerus articulating with a rudimentary forearm composed of three dysplastic long bones.

formed consent was obtained, and genomic DNA was extracted. The DNA was genotyped for 15 polymorphic markers from the 7q36 region, by means of radioactive PCR amplification followed by denaturing PAGE. Conditions for PCR reactions were as follows: 10 ng of genomic DNA, 0.2 mM of each dNTP, 1 × PCR buffer (Boehringer), 1 μM nonlabeled primer, 0.1 μM [³²P]–end-labeled primer, and 0.2 U of *Taq* polymerase were incubated at 97°C for 20 s, 55°C for 30 s, and 72°C for 30 s, for 30 cycles. Sequences of the primers used for amplification have been published elsewhere (The Ge-

nome Database). Allele sizes for the markers in the acheiropodia critical region, used for estimation of the age and ancestry of the acheiropodia mutation, were determined, and they were numbered according to designations used by The Genome Database.

Mutation Analysis

Multiplex PCR.—Simultaneous amplification primer C7orf2-35 (5'-tgt tgg ctg tta tca ccc aga aaa t-3') with primer C7orf2-36 (5'-gag aat tga tga ctt cag aat cag-3') (exon 3), primer C7orf2-93 (5'-ttt gat gac ttt cgt gtt cat tg-3') with primer C7orf2-94 (5'-gac cca aat cgt tta ctg gag ga-3') (exon 4), and primer C7orf2-95 (5'-ttc cca ata ggt aca agt tac aat g-3') with primer C7orf2-96 (5'-tgc ttt ttc tag ttc aat gtg aat g-3') (exon 5) (Zguricas et al. 1999) was performed under standard conditions. In brief, 10 ng of genomic DNA, 0.2 mM of each dNTP, 1 μM of each primer, and 0.2 U of *Taq* polymerase, in 1 × PCR buffer (Boehringer), were incubated at 97°C for 20 s, 62°C for 30 s, and 72°C for 30 s, for 30 cycles. PCR products were resolved on an agarose gel and were visualized by staining with ethidium bromide.

Reverse Transcriptase–PCR (RT-PCR).—Total RNA was extracted from whole blood from individuals affected with acheiropodia and from their parents, according to existing protocols. Ten micrograms of total RNA were subjected to reverse transcription using random hexamers and a first-strand–synthesis kit (Gibco BRL). The resulting cDNA was then PCR amplified by primers C7orf2-F3 (5'-gca tga agg agg atg gaa gg-3') and C7orf2-R1 (5'-tga gtg ctg aag cta ccc aca-3'), under the conditions described above, which produced a 521-bp fragment that spans exons 1–6 of the cDNA.

Sequence Analysis.—RT-PCR product from unaffected and affected individuals was column purified (Qiagen) and was sequenced by means of an internal primer (C7orf2-F2 [5'-agt cca cga tat gtt tcc ttc-3']). Sequencing reactions were performed by means of a Beckman Coulter sequencing kit for automated sequencing, under the conditions recommended by the manufacturer. Products from the sequencing reaction were separated and then were analyzed on a Beckman Coulter model SEQ 2000 automated sequencer.

Results

Fine Mapping of the Acheiropodia Locus

To facilitate the identification of the gene responsible for this disorder, we refined the acheiropodia critical region by haplotype analysis of the five families with acheiropodia. Genotyping of these families for a series of 15 highly polymorphic 7q36 markers enabled us to compute, by using the GENEHUNTER package (Kruglyak et al. 1996), a LOD score of 4.04 for the three families (families 1, 2, and 4; see fig. 1) in which the offspring

was the result of a known consanguineous mating. Furthermore, a long region of homozygosity was observed in affected individuals from all five families (fig. 3), and these regions overlapped in identical alleles at D7S3037–D7S3036, suggesting that the most likely region for the acheiropodia locus can be reduced to the interval between D7S550 and D7S2465 (fig. 3). Thus, analysis of this family panel reduced the acheiropodia critical region from 8.4 cM (Escamilla et al. 2000) to 1.3 cM, on the basis of identity by descent (fig. 4) (Kruglyak et al. 1995; Ianakiev et al. 2000). The physical map of the region (see the Integrated Chromosome 7 Database [The Chromosome 7 Project]) suggests that it is wholly contained within a single YAC (c655a11) and that the size of the region is <0.5 Mb (fig. 4).
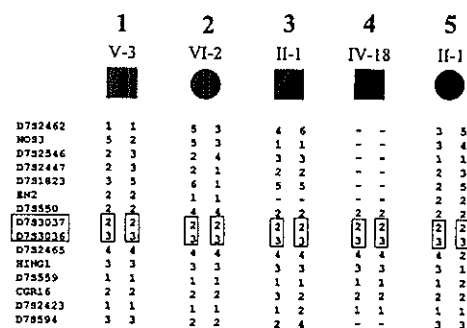
Genomic Structure of C7orf2

Physical and transcriptional mapping of the PPD locus on chromosome 7q36, which partially overlaps with the acheiropodia critical region, identified three transcripts, designated "C7orf2," "C7orf3," and "C7orf4" (Heus et al. 1999). C7orf2 is the human orthologue of the mouse *Lmbr1* gene and encodes a putative receptor, whereas C7orf3 and C7orf4 encode proteins of unknown function that are not expressed in the developing limb (Heus et al. 1999; Clark et al. 2000; M.J.v.B. and P.H., unpublished data). The mouse *Lmbr1* gene shows striking alterations of expression in the limbs of *Hemimelic extra toes* (*Hx*) mice (Clark et al. 2000). The *Hx* mutation causes hemimelia of the radius and tibia and preaxial polydactyly on both forelimbs and hind limbs (Knudsen and Kochhar 1981). The involvement of the mouse *Lmbr1* gene in a limb phenotype, as well as its genomic location, suggest that C7orf2 is a good candidate gene for acheiropodia.

The 17 exons of C7orf2 encompass ~200 kb of genomic DNA (M.J.v.B. and P.H., unpublished data). All splice sites follow the consensus (5' AG/N, 3' N/GT), except for the 3' splice site of exon 2, which uses the less common N/GC. The C7orf2 gene encodes a 1,470-nucleotide open reading frame preceded by a 176-nucleotide 5'UTR and followed by a 3,203-nucleotide 3'UTR (M.J.v.B. and P.H., unpublished data). The open reading frame predicts a 490-amino-acid protein containing nine putative transmembrane domains and a coiled-coil domain, which has led to the speculation that it might be located in the plasma membrane and might be a receptor protein. The predicted protein is 95% identical to the mouse *Lmbr1* protein (Clark et al. 2000).

Mutation Analysis of C7orf2

To test the possibility that mutations in the C7orf2 gene are responsible for acheiropodia, genomic DNA from affected individuals from each of the five families with acheiropodia was amplified by sets of primers for

|	1 V-3		2 VI-2		3 II-1		4 IV-18		5 II-1	
D7S2462	1	1	5	3	4	6	–	–	3	5
H093	5	2	5	3	1	1	–	–	3	4
D7S2546	2	3	2	4	3	3	–	–	1	1
D7S2447	2	3	2	1	2	2	–	–	2	3
D7S1823	3	5	6	1	5	5	–	–	2	5
IN2	2	2	1	1	–	–	–	–	2	2
D7S550	2	2								
D7S3037	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2	2 2
D7S3036	3 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3	3 3
D7S2465	4	4	4	4	4	4	4	4	4	2
HING1	3	3	3	3	3	3	3	3	3	1
D7S559	1	1	1	1	1	1	1	1	1	2
CGRL6	2	2	1	1	3	2	2	2	2	2
D7S2423	1	1	2	2	1	2	1	1	1	1
D7S94	3	3	2	2	2	4	–	–	3	3

Figure 3   Haplotypes derived by analysis of the five nuclear families. The putative ancestral haplotype shared by all affected individuals is boxed.

all 17 exons of the C7orf2 gene. All exons amplified successfully, with the exception of exon 4—which consistently failed to amplify in affected individuals, in contrast to its reproducible amplification in both obligate carriers and normal controls. Multiplex PCR with primer sets for exons 3–5 clearly demonstrated the lack of exon 4 in affected individuals but not in obligate carriers (fig. 5A).

To determine the status of exon 4 at the RNA level, total RNA from an affected individual, obligate carriers, and normal controls was analyzed by RT-PCR using primers spanning exons 1–6. RNA from normal controls and obligate carriers produced the expected 521-bp fragment; however, RNA from the individual with acheiropodia produced a 381-bp fragment, 140 bp smaller than the normal fragment and exactly the size of exon 4. RNA from the obligate-carrier parents of the affected individual produced both the 521-bp fragment and the 381-bp fragment, consistent with their heterozygous state (fig. 5B). Analysis by RT-PCR of RNA from a sample of 16 normal individuals from an ethnically mixed population showed only the normal-size fragment (data not shown).

Sequence analysis of the fragments produced by RT-PCR confirmed the lack of exon 4 of C7orf2 mRNA in individuals with acheiropodia (fig. 5C). Sequencing of the 381-bp fragment showed a clean transition from exon 3 to exon 5 in acheiropodia individuals (fig. 5C). In contrast, sequencing of the 521-bp fragment from normal controls showed the expected sequence for exons 3–5. Thus, the acheiropodia mutation leads to the production of a C7orf2 transcript lacking exon 4 of the gene and introduces a frameshift leading to a premature stop codon in exon 6. It therefore appears to be a true null mutation, consistent with the lack of any phenotypic effect in heterozygotes.
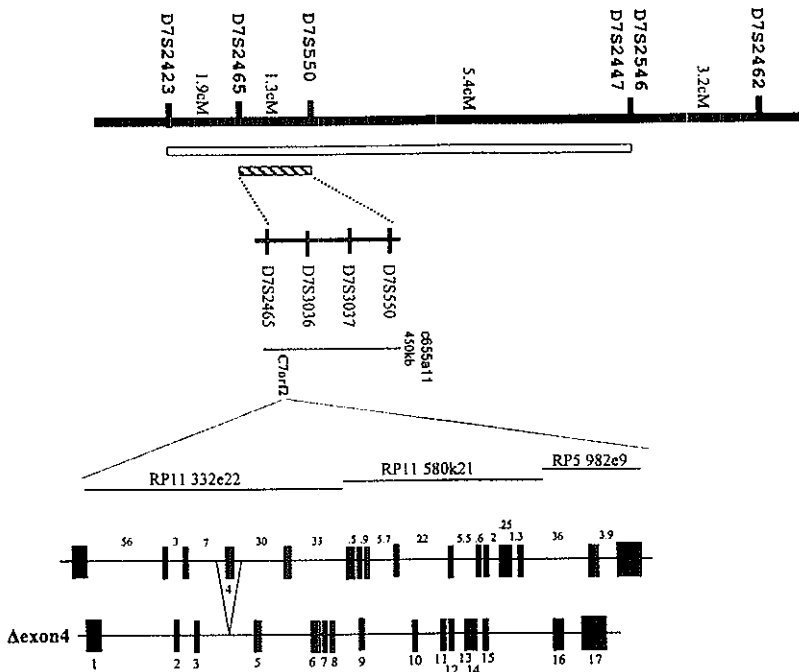
Figure 4 Map of the 7q36 region containing the acheiropodia locus. Distances (in cM) between markers are shown. The acheiropodia critical region derived by Escamilla et al. (2000) is represented the white vertical bar, and the critical region identified in the present study is represented by the hatched vertical bar. A physical map of the D7S550-D7S2465 minimal critical region is shown, including the c655a11 YAC that spans the region and the positions of three BAC clones (RP11 332e22, RP11 580k21, and RP5 982e9) that contain the C7orf2 gene. Graphic representations of both the C7orf2 gene and the Δexon4 acheiropodia mutation are shown below the physical map; sizes (in kb) of introns are shown above representation of the C7orf2 gene, and exon designations are shown below the representation of the Δexon4 acheiropodia mutation.

PCR-based analysis of genomic DNA from individuals with acheiropodia, by sets of primers spaced throughout introns 3 and 4, delineated the deletion's boundaries as being 1.2–2.5 kb 5′ of exon 4 and 2.7–3.5 kb 3′ of exon 4. These data indicate a deletion of 4–6 kb (data not shown).

Discussion

After the initial mapping of acheiropodia to 7q36 (Escamilla et al. 2000), we attempted to confirm and refine the mapping of the acheiropodia locus, by studying an affected individual from each of five Brazilian families. Confirmation was straightforwardly achieved, since three of the five individuals were the offspring of known consanguineous marriages. We computed a LOD score of 4.04 in these families—well in excess of the statistical

proof required to confirm significant linkage. Furthermore, in each of the five affected individuals, homozygosity was observed in no fewer than eight consecutive microsatellite markers. Within these overlapping regions of homozygosity, we observed a two-marker haplotype that was shared identically among all these patients and that was not seen in homozygous form in any unaffected individuals in these families. The rarity of certain alleles strongly indicates that this segment, which spans <1.3 cM, was inherited by descent from a single ancestor. Interestingly, this 1.3-cM region is completely contained within the 8.4-cM region implicated by Escamilla et al.'s (2000) report, and the shared alleles near this region match the shared haplotype in that study. This observation strongly suggests that all the cases in the two studies share homozygosity for a common haplotype and confirms the expectation that a single common ancestral
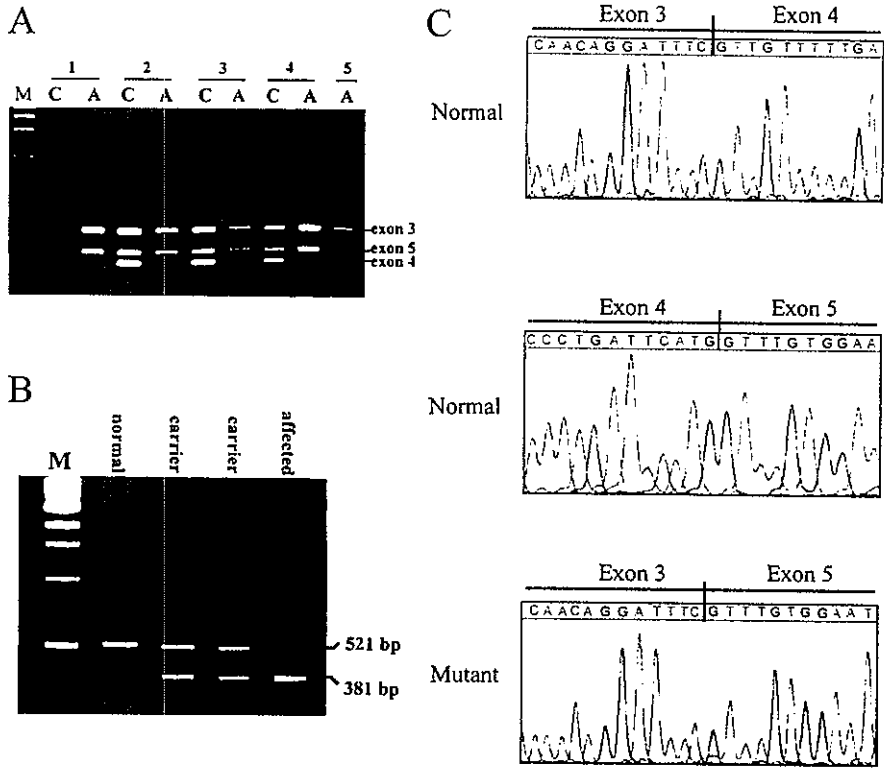
Figure 5   A, Results of multiplex PCR analysis of genomic DNA. For each of the five families with acheiropodia, results of analysis of exons 3–5 of the C7orf2 gene, both in affected individuals (lanes A) and in one of their obligate-carrier parents (lanes C), are shown. *Lane M*, Size markers and positions of bands from each of the three exons (which are identified to the right of the gel). *B*, Results of RT-PCR analysis of total RNA, using primers spanning exons 1–6 from an affected individual and obligate-carrier parents in family 4 and from a normal control individual. *Lane M*, Size markers and sizes of RT-PCR fragments (which are denoted to the right of the gel). *C*, Results of sequence analysis of cDNA from an affected individual in family 4 and from a normal control individual. The sequence from the normal individual shows the exon 3/exon 4 and exon 4/exon 5 boundaries. The sequence from the affected individual shows the exon 3/exon 5 boundary created by the Δexon4 mutation.

mutation (Freire-Maia 1975*b*) is the cause of nearly all cases of acheiropodia in Brazil.

The length (~1.3 cM) of the 7q36 region shared in all five patients in the present study suggests a relatively recent age (31 generations) for this common mutation; however, because only five individuals (assumed to represent only five independent paths to the ancestor, with very recent consanguinity causing duplication) were examined, simulations indicated that a mutation arising in a common ancestor anywhere from 5 to 95 generations ago would also be consistent with this extent of

haplotype sharing. If we include in this analysis the families studied by Escamilla et al. (2000), simulations indicate a 20-generation-old mutation, with a 95% confidence interval of 4–60 generations.

Analysis of the C7orf2 gene identified a common acheiropodia mutation in all five unrelated affected individuals studied. C7orf2 originally had been identified as one of the candidate genes for PPD (Heus et al. 1999) and is the human orthologue of the mouse *Lmbr1* gene. *Lmbr1* maps to the candidate interval defined for the mouse *Hx* and *Hammertoe* (*Hm*) mutants. These are

generally considered to be the mouse homologous phenotypes for PPD and complex polysyndactyly, respectively, suggesting C7orf2 and *Lmbr1* as candidate genes for these phenotypes. Although no coding-sequence alterations or genomic rearrangements of C7orf2 were detected in five unrelated families with PPD (M.J.v.B. and P.H., unpublished data) and no mutations were found in *Hx* or *Hm* mice, *Lmbr1* expression has been found to be altered in developing *Hx* limbs (Clark et al. 2000). *Lmbr1* is ubiquitously expressed, including the developing limbs, but in the *Hx* mutant the level of expression is significantly reduced during a specific period of limb development (E10.5–E12.5) (Clark et al. 2000). This aberrant expression coincides with the ectopic expression of *Shh* at the anterior aspect of the apical ectodermal ridge, as opposed to its normal expression, which restricted to the posterior aspect (Riddle et al. 1993). *Shh* is one of the key signaling molecules involved in specification of the antero-posterior and proximo-distal axes of the limb. A possible explanation for the observed ectopic expression of *Shh* in the *Hx* mutant could be that C7orf2/*Lmbr1* acts as a repressor of *Shh*. Whether this effect occurs through direct interaction with *Shh* or is mediated by other proteins remains to be clarified. Given the nature of the acheiropodia and *Hx* phenotypes, it is clear that C7orf2/*Lmbr1* plays a crucial role in distal-limb formation and outgrowth. Dissection of the function of this gene will foster our understanding of the processes involved in pattern formation during embryogenesis.

Acknowledgments

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

Chromosome 7 Project, The, http://www.genet.sickkids.on.ca/chromosome7 (for the Integrated Chromosome 7 Database)
GenBank Overview, http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html (for C7orf2 [accession number AF107454])
Genome Database, The, http://gdbwww.gdb.org (for allele sizes and amplification sequences)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for acheiropodia [MIM 200500] and triphalangeal thumb locus [MIM 190605])

References

Clark RM, Marker PC, Kingsley DM (2000) A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of hemimelic extra-toes mutant mice. Genomics 67:19–27

Escamilla MA, DeMille MC, Benavides E, Roche E, Almasy L, Pittman S, Hauser J, Lew DF, Freimer B, Whittle MR (2000) A minimalist approach to gene mapping: locating the gene fora acheiropodia by homozygosity analysis. Am J Hum Genet 66:1995–2000

Freire-Maia A, Freire-Maia N, Morton NE, Azevedo ES, Quelce-Salgado A (1975a) Genetics of acheiropodia (the handless and footless families of Brazil). VI. Formal genetic analysis. Am J Hum Genet 27:521–527

Freire-Maia A, Li WH, Maruyama T (1975b) Genetics of acheiropodia (the handless and footless families of Brazil) VII. Population dynamics. Am J Hum Genet 27:665–675

Grimaldi A, Masiero D, Richieri-Costa A, Freire-Maia A (1983) Variable expressivity of the acheiropodia gene. Am J Med Genet 16:631–634

Heus HC, Hing A, van Baren MJ, Joosse M, Breedveld GJ, Wang JC, Burgess A, Donnis-Keller H, Berglund C, Zguricas J, Scherer SW, Rommens JM, Oostra BA, Heutink P (1999) A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. Genomics 57:342–351

Heutink P, Zguricas J, van Oosterhout L, Breedveld GJ, Testers L, Sandkuijl LA, Snijders PJ, Weissenbach J, Lindhout D, Hovius SE (1994) The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. Nat Genet 6:287–292

Ianakiev P, Kilpatrick MW, Daly MJ, Zolindaki A, Bagley D, Beighton G, Beighton P, Tsipouras P (2000) Localization of an acromesomelic dysplasia on chromosome 9 by homozygosity mapping. Clin Genet 57:278–283

Knudsen TB, Kochhar DM (1981) The role of morphogenetic cell death during abnormal limb-bud outgrowth in mice heterozygous for the dominant mutation hemimelia-extra toe (Hmx). J Embryol Exp Morphol Suppl 65:289–307

Kruger LM, Kumar A (1994) Acheiropody: a report of two cases. J Bone Joint Surg 76A:1557–1560

Kruglyak L, Daly MJ, Lander ES (1995) Rapid multipoint linkage analysis of recessive traits in nuclear families including homozygosity mapping. Am J Hum Genet 56:519–517

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lemos Silveira E, Freire-Maia A (1998) Acheiropodia: new cases from Brazil. Clin Genet 54:256–257

Mourao PAS, Toledo SPA, Dietrich CP (1977) Urinary mucopolysaccharides in acheiropodia. Acta Genet Med Gemellol (Roma) 26:92–94

Riddle RD, Johnson RL, Laufer E, Tabin C (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. Cell 75:1401–1416

Toledo SPA, Saldanha PH (1969) A radiological and genetic investigation of acheiropody in a kindred including six cases. J Genet Hum 17:81–94

——— (1972) Further data on acheiropody. J Genet Hum 20: 253–258

Tsukurov O, Boehmer A, Flynn J, Nicolai JP, Hamel BC, Traill S, Zaleske D, Mankin HJ, Yeon H, Ho C (1994) A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. Nat Genet 6:282–286

Zguricas J, Heus H, Morales-Peralta E, Breedveld G, Kuyt B, Mumcu EF, Bakker W, Akarsu N, Kay SP, Hovius SE, Heredero-Baute L, Oostra BA, Heutink P (1999) Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localization of the gene responsible to a 1.9 cM region on chromosome 7q36. J Med Genet 36:32–40

Chapter 6

Preaxial Polydactyly Due to a Long Range *Cis*-Acting

Regulatory Element for *Shh*

Laura A. Lettice[a,1], Taizo Horikoshi[b,c,1], Simon Heaney[a,1], Marijke J. Van Baren[d,1], Herma C. van der Linde[d], Guido J. Breedveld[d], Marijke Joosse[d], A. N. Akarsu[e], Ben A. Oostra[d], N. Endo[c], M. Shibata[f], M. Suzuki[g], E. Takahashi[g], T. Shinka[h], Y. Nakahori[h], D. Ayusawa[i], K. Nakabayashi[j], S. W. Scherer[j], Peter Heutink[d], Robert E. Hill[a], and Sumihare Noji[b]

[a] MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh UK EH4 2XU
[b] Department of Biological Science and Technology, Faculty of Engineering, University of Tokushima, Tokushima, Japan
[c] Department of Orthopedic Surgery, School of Medicine, University of Niigata, Niigata, Japan
[d] Department of Clinical Genetics, Erasmus University, P.O.Box 1738, 3000 DR Rotterdam, the Netherlands.
[e] Gene Mapping Laboratory, Basic and Applied Research Center of Children's Hospital, Hacettepe University, Ankara, Turkey.
[f] Division of Plastic & Reconstructive Surgery, School of Medicine, University of Niigata, Niigata, Japan
[g] Otsuka GEN Research Institute, Otsuka Pharmaceutical Co., Ltd, Tokushima, Japan
[h] Department of Public Health, School of Medicine, University of Tokushima, Tokushima, Japan
[i] Kihara Institute for Biological Research, Graduate School of Integrated Science, Yokohama City University, Yokohama, Japan
[j] Departments of Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Canada

[1] Contributed equally to this work

Abstract

Preaxial polydactyly (PPD) is a common limb malformation linked to chromosome 7q36. PPD maps to a complex locus that includes the other limb disorders acheiropodia and complex polysyndactyly. Here, we identify a translocation breakpoint in a PPD patient and a transgenic insertion site in the polydactylous mouse mutant sasquatch (*Ssq*). The genetic lesions in both mouse and human lie within the same intron of the *LMBR1/Lmbr1* gene. However, genetic analysis of *Ssq* shows that the mutation interrupts a long-range *cis*-acting regulatory element which operates on *Shh* ~1Mb away. Thus PPD in human most likely results from similar acting mutations and supports the prospect that *Shh* regulatory elements underlie this complex locus.

Introduction

Preaxial Polydactyly (PPD, MIM190605) is one of the most frequently observed human congenital limb malformations. Sporadic cases of PPD have been described, but most show an autosomal dominant mode of inheritance. The phenotype shows large variation within families and varies from triphalangeal thumb or duplication of the thumb to tibial aplasia. Using several large families, a PPD locus was mapped to a 450 kb region on chromosome 7q36 and all families described so far are linked to this locus (Heus *et al.* 1999, Heutink *et al.* 1994, Hing *et al.* 1995, Radhakrishna *et al.* 1997, Zguricas *et al.* 1999). The other distinct but limb specific mutations, complex polysyndactyly (CPS) (Tsukurov *et al.* 1994) and acheiropodia (Ianakiev *et al.* 2001), map to this region suggesting that elements essential for limb development are located in this locus.

Sasquatch (*Ssq*) is a mouse mutation that arose recently through a transgenic insertion (Sharpe *et al.* 1999). The mutation is semidominant resulting in supernumerary preaxial (anterior) digits on the hindfeet in the heterozygotes. In homozygotes both fore- and hindlimbs show additional preaxial digits and in some cases the long bones are shortened such that the limbs appear twisted. The insertion site responsible for the *Ssq* phenotype is physically linked to within ~1Mb of *Shh*. Here, we show that *Ssq* maps to the region on mouse chromosome 5 that corresponds to the human PPD locus. In addition *Ssq* and the mouse mutant hemimelic extra toe (*Hx*) share a similar limb phenotype and are most likely allelic (Clark *et al.* 2000, Knudsen and Kochhar 1981).

We combined genetic approaches in human and mouse to identify mutations in a PPD patient and in the *Ssq* mutant to determine the basis for formation of preaxial supernumerary digits. The patient carries a *de novo* translocation with a breakpoint within the critical region identified for PPD. Isolation of the PPD translocation breakpoint and the *Ssq* transgene insertion site revealed a similar location for these genetic disruptions embedded in a chromosomal region surrounded by genes. We provide genetic analysis that shows the *Ssq* mutation is not acting locally but in fact, interrupts a long-range *cis*-acting regulatory element. This element operates on *Shh* residing 1.8cM, a physical distance of ~1Mb, away. Consequently disruption of *Shh* regulation is most likely the
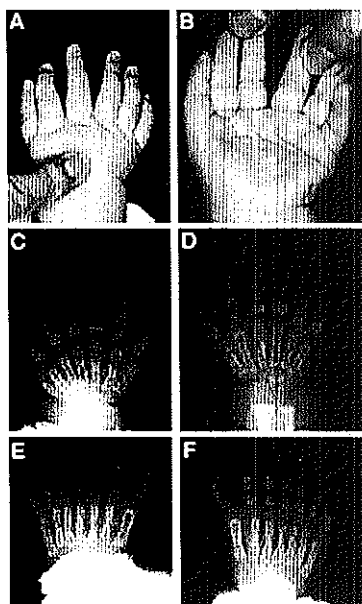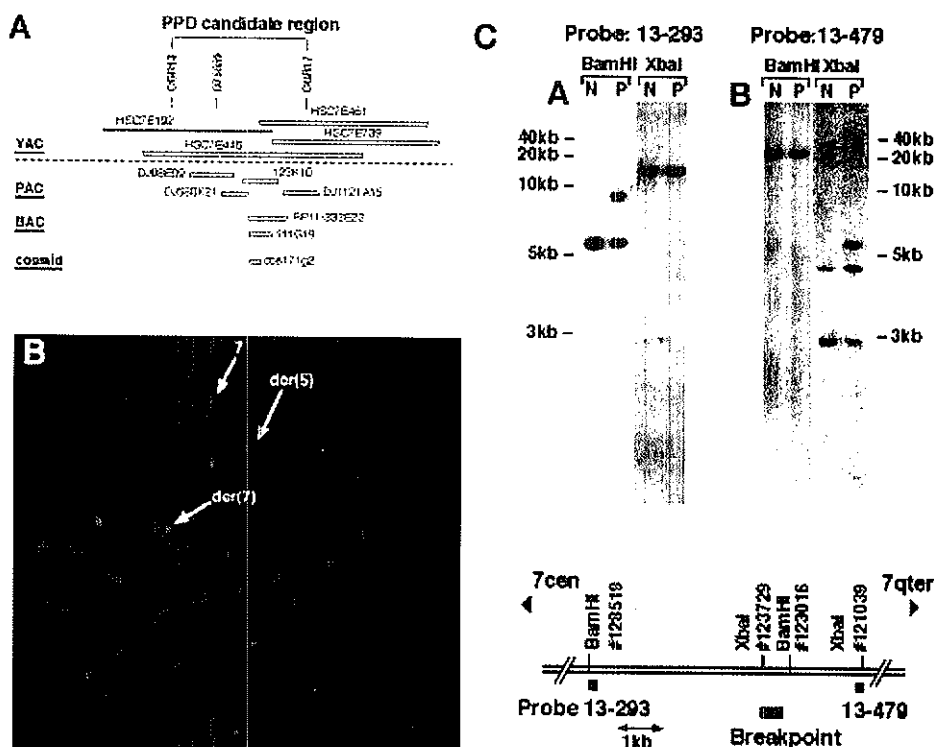
Figure 1
Polydactyly in the hands and feet of the t(5,7)(q11,q36) translocation patient. Dorsal (**A**) and ventral (**B**) photographs of the left hand of the PPD translocation patient showing duplication of the triphalangeal thumb. Radiogram of the hands (**C** and **D**) showing bilateral duplications and of the feet (E and F) showing triplication of the great toes

basis for PPD in humans. We present a model in which long range Shh regulation may also be responsible for the acheiropodia phenotype.

Results

Identification of a PPD translocation breakpoint

A three-year-old girl presented with bilateral duplication of the triphalangeal thumb and triplication of the great toe without any associated abnormality (fig. 1). She was found to carry a *de novo* reciprocal translocation t(5,7)(q11,q36). Fine mapping of the translocation breakpoint by use of FISH analysis in the lymphoblastoid cell line derived from the patient identified a BAC clone Rp11-332E22 (Genbank accession number AC007097) that spanned the breakpoint. Using fragments from the BAC, a cosmid clone 171g2 was identified which spanned the breakpoint (fig 2B). Sequencing the ends of the cosmid

Figure 2
Identification of the breakpoint in the t(5,7)(q11,q36) translocation patient.
A. shows the YACs, PACs, BACs and a cosmid used to define the translocation breakpoint relative to the polymorphic markers within the PPD candidate region (from ref. 5). B. shows FISH analysis of metaphase chromosomes from a lymphoblastoid cell line from the patient, using cosmid clone171G2 as a probe. Hybridisation signals are marked by arrows on the wildtype chromosome 7 and on the products of the translocation, the derived chromosome 5 (der 5) and derived chromosome 7 (der 7). C. Southern blot analysis with genomic DNA from the lymphoblastoid cell line from the patient (P) and a normal control (N). DNA was digested with restriction endonucleases *BamHI* and *XbaI*. In the first panel STS probe 13-293 was used and in the second panel STS probe 13-479 was used. The relative positions of the STS probes, restriction sites and the translocation breakpoint are indicated in the diagram below.

insert identified the genomic fragment as lying inside one of the genes (originally given the designation C7orf2) within the genetically defined critical region for PPD (fig. 2A). This gene is the orthologue of the mouse *Lmbr1* gene (Clark *et al.* 2000). Southern blot analysis of DNA from the lymphoblastoid cell line further refined the site of the breakpoint (fig. 2C). A number of repeat free STSs including 13-293 and 13-479 were PCR-amplified and used as probes. This analysis showed that the translocation breakpoint is located in a 714 bp *BamHI/XbaI* fragment residing between exon 5 and 6 of *LMBR1* without an accompanying large deletion (fig. 2C). This DNA fragment consists of 95% LINES and 5% simple repeated sequence, and seems to neither encode an open reading frame nor contain known regulatory sequences.
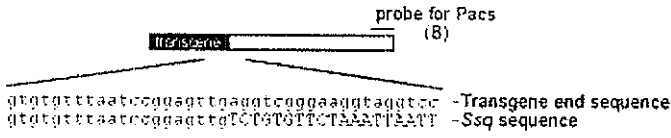
Isolation of the Ssq insertion site

To isolate the transgenic insertion site from the *Ssq* mutant mouse, we made a λ genomic library made from *Ssq/Ssq* mice. Clones were isolated using part of the transgene as a probe and one contained the junction of the transgene and the insertion site (fig. 3A). Analysis of the insertion site showed that multiple copies of the transgene had incorporated into a genomic region that is predominantly L1 high copy repeat. Less than 50bp is deleted from genomic DNA. Repetitive DNA extends for more than 2kb on each side of the insertion site (fig. 3D). The genomic end fragment of this λ clone (fig. 3B) was

Figure 3. Identification of the mouse *Ssq* insertion site.
A. shows the *Ssq* λ genomic clone identified as containing the insertion site junction fragment. Sequence across this join is shown, indicating that 19 nucleotides was trimmed from the end of the transgene upon insertion. The position of the probe used to identify the PACs is also shown and the sequence of this fragment is shown in **B.** FISH data in **C.** confirms the co-localization of the PACs with the transgene. PAC RPCI21-542N10 is labeled in green and the HPAP containing transgene is labeled in red. The first panel shows an interphase nucleus and the second a portion of a metaphase spread. In **D.** the position of the mouse PACs containing *Lmbr1* that were identified are shown relative to the known genomic structure of the human *LMBR1* gene as the mouse sequence is unfinished in the database. The integration site of the transgene is depicted as a close up between exons 5 and 6 (shown as black boxes), the transgene has inserted within 4.5kb of exon 6 and in the opposite orientation to the *Lmbr1* cDNA. Most of the intron has been sequenced, but there remains a gap (shown by the parallel lines); L1 repetitive DNA is depicted by the hatched bars and non-repetitive DNA by the open bars.
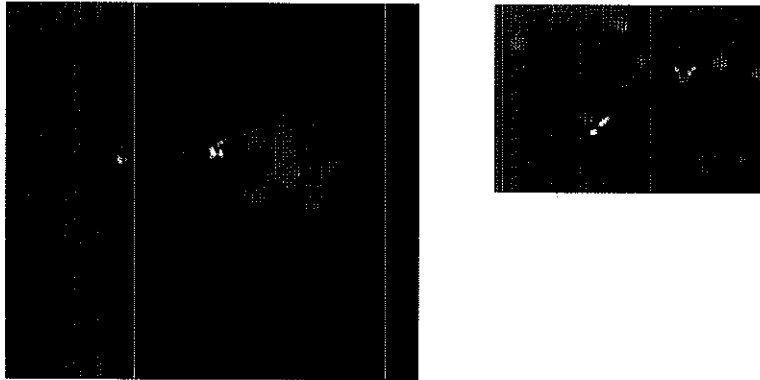
132

A.

probe for Pacs
(B)

transgene

gtgtgtttaatccggagttgaggtcggaaggtagatcc -Transgene end sequence
gtgtgtttaatccggagttgTCTGTGTTCTAAATTAATT -Ssq sequence

B.
```
tctaattgaa cacaggattc agttccgtta cacttcctgg tatcccttta
tcacttcaac tatgcaatc atattgcttc ctttctcagc ttgctccttt
tcatttaaat ttaatttaag cctagagtct gtttttgtaa tattttggtg
aagaacatgt ctatattttg cattgtccga agagtttgcc tgaggtggag
ggatttagct taatttctct gaca
```

C.

D.

Mouse Pacs

RPCI21-542N10

RPCI21-576B18

Human *LMBR1*

exon1   2 4   5   6   91015   17

Transgene integration in Ssq

Lmbr1
exon 5

Lmbr1
exon 6

3'   HPAF   5' n

133
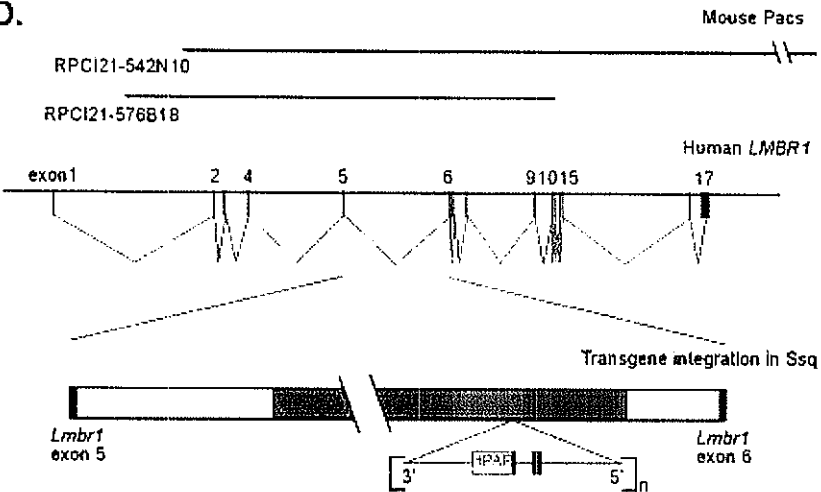
used to identify mouse PAC clones and subsequent exon trapping (Nehls *et al.* 1994) identified a surrounding gene (fig. 3D). Chromosomal FISH analysis showed that the PAC clones and the transgene co-localise (fig 3C). The complete sequence of the mouse cDNA was determined and was found to be identical to mouse *Lmbr1* (ref 10). The exact position of the insertion site was determined by sequencing the surrounding region until coding sequence was identified. The insertion site lies within intron 5 (fig 3D), the same intron as that interrupted by the human translocation breakpoint. The *Lmbr1* gene is highly conserved and found in species as diverse as *Drosophila* (accession no. AE003749) and *C.elegans* (accession no. P34535). Our analysis of the secondary sequence topology (Cserzo *et al.* 1997) agrees with previous predictions (Clark *et al.* 2000, Heus *et al.* 1999) that the *Lmbr1* gene encodes a protein with multiple (possibly up to nine) transmembrane domains.

Analysis of the mutational mechanism

We furthered our investigations to define the mechanism by which the preaxial polydactyly mutational events were operating. It is clear that PPD is not a simple loss of function mutation. Large chromosomal deletions on human 7q36 that remove *LMBR1* and surrounding genes (Belloni *et al.* 1996) do not result in limb abnormalities, thus suggesting a dominant gain-of-function mechanism. In the analysis of *Lmbr1* transcription in the *Ssq* mouse we found a high level of premature termination resulting in truncated transcripts. At the 5' end of the mutant allele (fig 4, top panel), *Lmbr1* transcript levels appear similar to wildtype, but transcription of exons 3' of the insertion site is detectably lower, representing about 10-30% of wildtype levels (fig. 4, middle panel). This indicates that the transgene insertion leads to termination of transcription between exons 5 and 6. In human the t(5,7)(q11,q36) translocation breakpoint would predictably have the capacity to generate a similar truncated product. These data are consistent with a dominant gain of function mechanism; a truncated *Lmbr1* protein product (if produced) acquiring dominant activity perhaps as an activated receptor.

However, our analysis of the *LMBR1* gene in five PPD families (Heutink *et al.* 1994, Zguricas *et al.* 1999) provides contradictory evidence. We tested for *LMBR1*
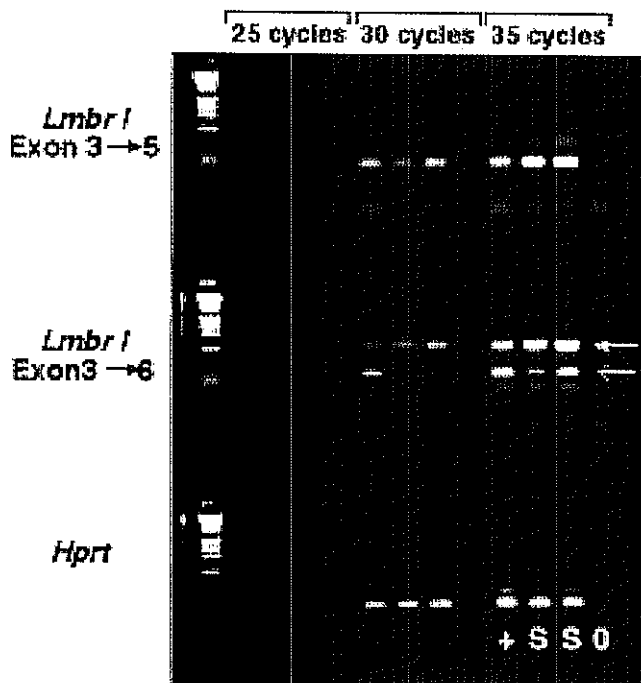
Figure 4.
Transcripts from the *Lmbr1* gene in *Ssq* are truncated. RT-PCR samples were removed after 25, 30 and 35 cycles. RNA is from 11.5 day embryos of wildtype (+) and Ssq/Ssq (S) genotypes, a no RNA control (0) is also included. The top two panels make use of primer pairs specific to *Lmbr1*; the first is from exon 3 to exon 5 and examines expression levels 5' of the transgene integration. Comparable levels are detected in wildtype and *Ssq/Ssq* embryos. The second panel uses primers that cross the integration site spanning exons 3 to 6. With these, at 30 cycles a product is only detected in the wild type sample. The primers used in this experiment cross-hybridise with the mouse glycosyl-phosphatidyl-inositol-anchored protein homologue (closed arrowhead) which serves as positive control for the experiment and to normalizes the samples. In the bottom panel primers from the HPRT locus were used as positive control for RNA integrity and controls for reverse transcription.

mutations in the exons and the intron/exon boundaries of these families and no pathogenic mutations were found. In addition, we amplified *LMBR1* transcripts from RT-PCR products made from RNA from a lymphoblastoid cell line of Dutch and Cuban patients, and sequenced the products. Polymorphisms in a Cuban (in exons 13 and 17), and in two related Dutch (one in exon 5 and the other in 14) patients indicated that both *LMBR1* alleles, at the 5' end and the 3' end, were expressed (data not shown). Furthermore, an

overlapping series of three PCR products covering the whole *LMBR1* transcript showed no difference in band size between patients and controls (data not shown), indicating that no exons are skipped in patients. These data do not support the hypothesis that mutations in the coding region of *LMBR1* or mutations disrupting *LMBR1* transcription commonly associate with the PPD phenotype.

In the *Ssq* mutation, we previously showed an intriguing relationship between *Shh* and the *Ssq* mutation (Sharpe *et al*. 1999). *Shh* is normally expressed in a posterior domain of the developing limb called the ZPA and is fundamental to patterning of the digits (Pearse and Tabin 1998). In *Ssq*, as in other unrelated preaxial polydactylous mutants (Lettice *et al*. 1999, Masuya *et al*. 1997, Qu *et al*. 1997), *Shh* is detected in an additional anterior ectopic site. An ectopic site of signaling activity is essential for the formation of extra digits. In addition, we showed that the reporter gene contained within the *Ssq* transgene insertion had acquired a limb-specific *Shh*-like expression pattern; not only in the appropriate posterior ZPA, but also at the anterior ectopic site (Sharpe *et al*. 1999). We previously suggested the possibility that mutations in *Ssq* lie within regulatory elements of *Shh*, causing limb specific mis-expression.

To examine the possibility that preaxial polydactyly is caused by disruption of a long range regulatory element, we devised a *cis-trans* genetic test. We predicted that the interruption of such an element regulating *Shh* would be functioning in a *cis*-acting manner. To test this possibility a mouse cross was set up to derive a recombinant chromosome (Sharpe *et al*. 1999) in which the *Ssq* mutation was located in *cis* to an easily distinguishable *Shh* allele (fig. 5A). Towards this aim, we crossed *Ssq/Ssq* male mice with females carrying the mutant *Shh* null allele (*Shh^{null}*) (Chiang *et al*. 1996) to produce the F1 generation The F1 generation (n=14) carrying these two alleles in *trans* showed complete penetrance of the *Ssq* mutant phenotype (*Shh^{null}* shows no phenotype in the heterozygous state). These F1 mice were mated to wildtype and assayed for the recombinant chromosome 5 containing the *Ssq* insertion and the *Shh^{null}* allele in *cis*. Each F2 mouse was assayed by PCR for the presence of the *Ssq* transgene insertion and the neo^{r} gene contained in the null Shh allele (fig. 5B). In this cross we produced 446 F2 offspring of which eight were recombinants representing a recombination frequency of 1.8%.
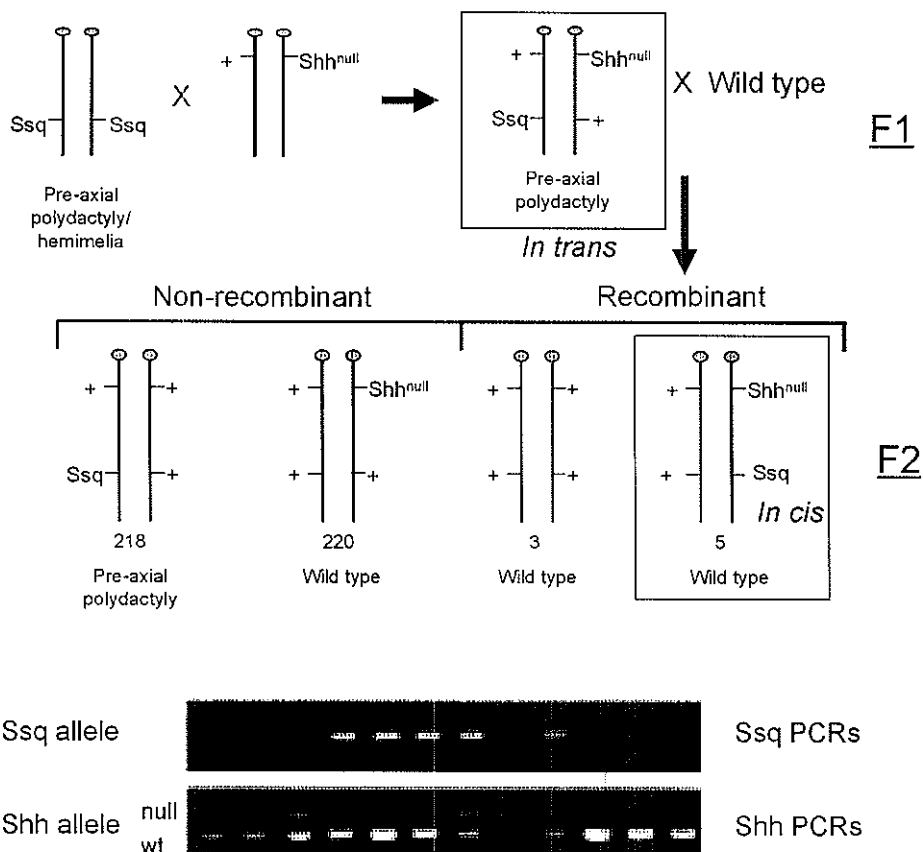
Figure 5. *Cis- trans* genetic test.
A. shows the genetic cross employed. A homozygous *Ssq* mouse was crossed to a *Shh^null^*
heterozygote. The F1 mice which carry both mutations in *trans*, all of which exhibit preaxial
polydactyly, were crossed to wild type mice. The bottom row depicts the outcome of the cross.
The majority of F2 progeny were, as expected, found to be heterozygous for either the *Ssq*
allele or the *Shh^null^* (218 and 220 mice respectively). The mice of interest were those identified
as carrying a chromosome 5 which has recombined between the *Shh* locus and the region of the
Ssq insertion. Three had inherited the wildtype alleles at both loci and five carried the *Ssq* allele
and the *Shh^null^* allele in *cis*. These mice show no additional digits and have wildtype paws. This
cross results in a genetic distance of 1.8 cM between *Shh* and *Ssq*. **B.** shows examples of the
genotyping PCRs used to analyze the F2 progeny. Lanes 1-3 are mice heterozygous for the
*Shh^null^* allele. Lanes 4-6 are wildtype for *Shh* but carry the *Ssq* insertion. Mice in lanes 7-12
carry the recombinant chromosomes; 7-9 show F2's with both *Shh^null^* and *Ssq* while 10-12 are
wild type at both loci.

137

Three of the recombinants were uninformative since these contained the wildtype alleles (fig.5A). The other five recombinants carried both the *Ssq* insertion and *Shh^null* allele; i.e., situated in *cis*. Analysis of the limb phenotype in these five recombinants showed no preaxial polydactyly or other detectable limb phenotypes. In addition we bred two males carrying the recombinant chromosome 5 (*Ssq/Shh^null*) with wildtype females and showed that a second generation had no limb abnormalities (21 F2 mice carrying the recombinant chromosome out of 37 offspring). The *Ssq* heterozygous mutation is not fully penetrant, the degree depending on the genetic background. In this cross 7% of the *Ssq/+* mice generated showed no detectable limb phenotype. Two *Ssq/+* males which displayed no limb phenotype were bred further and were shown to transmit the phenotype (4 with additional preaxial digits in 12 offspring).

Thus the data show that the *Shh* null allele inactivates the *Ssq* mutation when located in *cis* (but not in *trans*) on chromosome 5. The data are consistent with a long range limb-specific regulatory element residing within or near the *Lmbr1* gene being disrupted by the *Ssq* insertion. The disruption leads to a dominant acting mis-expression of the *Shh* gene. By inference the same mechanism is operating in the human translocation to cause PPD.

Discussion

PPD usually segregates as a simple fully penetrant autosomal dominant disorder. Large families are available and a positional cloning approach allowed reduction of the candidate region in humans to 450 kb in which five genes are located (Heus *et al.* 1999). Identification of the mechanism responsible for the phenotype, however, required a combination of molecular and cytogenetic analysis of human patients and detailed analysis of the *Ssq* transgenic mouse model, underscoring the value of examining model systems in parallel to studying the genetics of human disease.

In this report we show that a translocation breakpoint in intron 5 of human *LMBR1* and a transgene insertion in the same intron of mouse *Lmbr1* causes a similar preaxial polydactyly phenotype in both species. However, in a straightforward analysis of familial PPD cases we found no pathogenic mutations, suggesting that mutations outside the *Lmbr1* coding region are involved in limb dysgenesis. This is consistent with the

138

finding in *Hx* mice, where *Lmbr1* was implicated in polydactyly; but no causal mutation in the coding region was identified (Clark *et al.* 2000).

The lack of mutations within the *Lmbr1* gene led to further investigation of the mutational mechanism and an alternative was suggested by studies on *Ssq*. A direct effect on the *Shh* gene by the *Ssq* mutation (Sharpe *et al.* 1999) was initially inferred by the following: 1.) *Ssq* and *Shh* are physically linked, although at a genetic distance of 1.8cM, 2.) *Ssq* mutation affects *Shh* expression and 3.) the *Ssq* reporter gene has a *Shh*-like expression pattern. We predicted that the *Ssq* phenotype may be due to interruption of a long range regulatory element of *Shh* and if so, would therefore be functioning in a *cis*-acting manner. In order to test this hypothesis, we recombined *Ssq* and the *Shh* null allele onto the same chromosome. All five recombinant mice generated showed complete suppression of the *Ssq* phenotype strongly arguing that the *Ssq* transgene insertion disrupted a *cis*-acting *Shh* regulatory element. (*Ssq* should be now considered an allele of *Shh*, and be designated Shh^{Ssq}.) Furthermore, the relative proximity of the *Ssq* insertion site and the PPD translocation breakpoint suggests that in human supernumerary preaxial digits correspondingly result from similar disruption of *Shh* regulatory elements.

PPD appears to constitute only a part of a more complex locus affecting limb morphogenesis. In addition to PPD, two other limb defects, acheiropodia (Fett-Conte and Richieri-Costa 1990, Ianakiev *et al.* 2001, Toledo and Saldanha 1969, Toledo *et al.* 1972) and complex polysyndactyly (CPS) (Tsukurov *et al.* 1994), with distinctly different characteristics map to the region. Little is known about the molecular etiology of CPS; however, a recent report (Ianakiev *et al.* 2001) examining a group of Brazilian families showed that the autosomal recessive acheiropodia results from a small deletion within the *LMBR1* gene. Acheiropodia is a severe limb-specific phenotype and, in contrast to PPD, patients present with bilateral congenital amputations of the upper and lower extremities. The result is that all bones of the hands and feet are absent, the tibia is truncated distally and the radius, ulna, and fibula are lost (Fett-Conte and Richieri-Costa 1990, Toledo and Saldanha 1969, Toledo *et al.* 1972). Affected individuals (which carry the same ancestral haplotype) show deletions in both *LMBR1* alleles that remove exon 4 and approximately 5-6kb of surrounding genomic DNA (Ianakiev *et al.* 2001).

If extra digits result from ectopic *Shh* expression then loss of *Shh* expression in the limb may be expected to result in severe distal limb truncations similar to those seen in acheiropodia patients. There are no mouse models for acheiropodia; however, analysis of the limb in the *Shh* targeted deletion has been reported (Chiang *et al*. 1996, Kraus *et al*. 2001). In accord, although the overall mouse phenotype is complex and severe, in *Shh*[-/-] mutant mice the limbs show loss of all bones of the feet and truncations of the long bones. The phenotype is very similar to that seen in the acheiropodia patients. We suggest that, in contrast to *Ssq* and PPD, acheiropodia results from a limb-specific loss of *Shh* expression. Thus the ~5kb acheiropodia deletion from within the *Lmbr1* gene may remove elements essential for the limb-specific *cis* acting regulatory activity.

Shh embryonic expression is a dynamic, complex pattern. We propose a model (fig. 6) in which *Shh* limb specific regulatory elements arose as a discrete unit and reside at a considerable distance in the *Lmbr1* gene. Depending on the mode of perturbation in this region the regulatory element is either inactivated as in the recessive acheiropodia condition or modified leading to misexpression as in the *Ssq* insertion. Since PPD is relatively common in the human population, multiple, independent mutations interrupt the regulatory element allowing mis-expression. This is consistent with a model in which an element normally a part of the regulatory domain drives expression in both the anterior and the posterior limb regions. The anterior expression is actively repressed and it is disruption of this repression that causes PPD.

We have shown that mutational events affect a gene that resides at a physical distance of almost 1Mb. Genetically, in mouse, the distance between mutation and affected gene is 1.8cM. In most cases such a genetic distance would argue against analysis of a gene so far removed from the mutation and would lead to the analysis of the nearest gene carrying the insertion or the translocation breakpoint. In the case of *Ssq* however we were fortunate to have reporter gene expression that hinted at a *Shh* regulatory element residing nearby. This raises the disturbing prospect that other situations exist in which genetic analysis has lead to interest in a linked but unaffected gene.
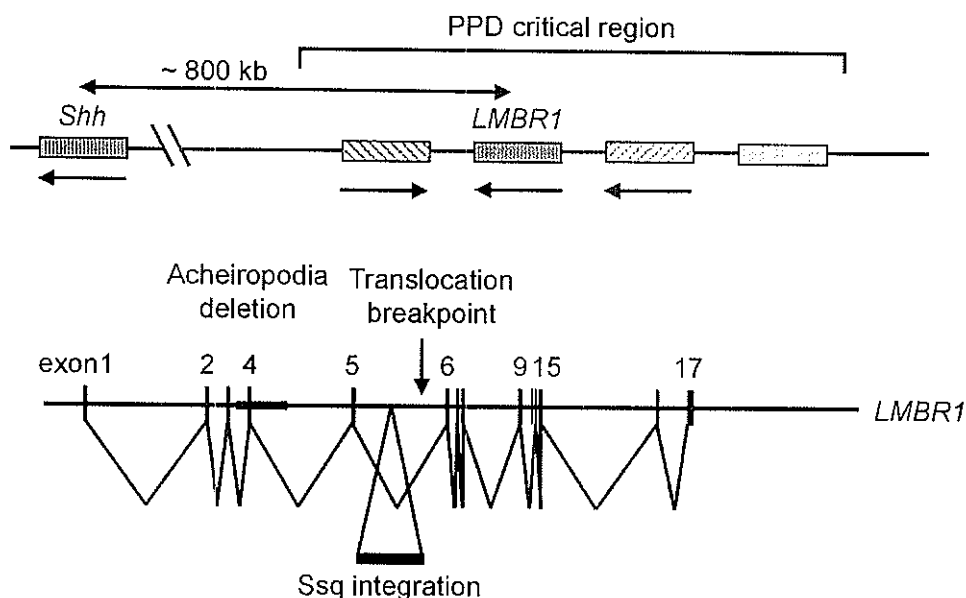
Figure 6. Diagram of the human PPD critical region on chromosome 7q36 and the relationship with *Shh*. The corresponding region in mouse contains *Lmbr1* surrounded by the orthologues of C7orf3 and C7orf4 and is ~800kb from *Shh*. No other genes are known to lie between *Shh* and the PPD critical region, suggesting that this is a gene poor region. The arrows indicate the transcriptional orientation. The structure of the 17 exon containing *LMBR1* gene is depicted at the bottom and shows the relative position of each mutation. The double headed arrows indicate the gene regions responsible for the opposing acheiropodia and PPD (*Ssq*) phenotypes. It is predicted that mutations between exon 5 and 6 release normal repression of the ectopic anterior expression pattern; whereas, deletions around exon 4 inactivates limb specific expression.

Materials and methods

Patient material

The translocation patient was clinically examined, and a member of her family was interviewed for family history at the Niigata University Hospital in Niigata. All the studies were approved by the local ethics committee. A member of the family gave written informed consent on behalf of the patient.

null alleles. Primers for the *Ssq* insertion were CTCTGTTTCCTTTTCCTCTATC and GTATGGGATTAATTAAATCTTGTGTC which generate a 180bp product in the presence of the insertion. The *Shh* PCR's were carried out with the primers described by Chiang *et al.* (1996) and generated a 550bp product for the null allele and 230 bp product for the wild type allele.

References

Belloni, E., M. Muenke, E. Roessler, G. Traverso, J. Siegel-Bartelt, A. Frumkin, H. F. Mitchell, H. Donis-Keller, C. Helms, A. V. Hing, H. H. Heng, B. Koop, D. Martindale, J. M. Rommens, L. C. Tsui, and S. W. Scherer. 1996. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet* 14: 353-6.

Chiang, C., Y. Litingtung, E. Lee, K. E. Young, J. L. Corden, H. Westphal, and P. A. Beachy. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383: 407-13.

Clark, R., P. Marker, and D. Kingsley. 2000. A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. *Genomics* 67: 19-27.

Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson. 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* 10: 673-6.

Fett-Conte, A. C., and A. Richieri-Costa. 1990. Acheiropodia: report on four new Brazilian patients. *Am J Med Genet* 36: 341-4.

Heus, H. C., A. Hing, M. J. van Baren, M. Joosse, G. J. Breedveld, J. C. Wang, A. Burgess, H. Donnis-Keller, C. Berglund, J. Zguricas, S. W. Scherer, J. M. Rommens, B. A. Oostra, and P. Heutink. 1999. A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. *Genomics* 57: 342-51.

Heutink, P., J. Zguricas, L. van Oosterhout, G. J. Breedveld, L. Testers, L. A. Sandkuijl, P. J. Snijders, J. Weissenbach, D. Lindhout, S. E. Hovius, and et al. 1994. The gene for triphalangeal thumb maps to the subtelomeric region of chromosome 7q. *Nat Genet* 6: 287-92.

Hing, A. V., C. Helms, R. Slaugh, A. Burgess, J. C. Wang, T. Herman, S. B. Dowton, and H. Donis-Keller. 1995. Linkage of preaxial polydactyly type 2 to 7q36. *Am J Med Genet* 58: 128-35.

Ianakiev, P., M. J. van Baren, M. J. Daly, S. P. Toledo, M. G. Cavalcanti, J. C. Neto, E. L. Silveira, A. Freire-Maia, P. Heutink, M. W. Kilpatrick, and P. Tsipouras. 2001. Acheiropodia Is Caused by a Genomic Deletion in C7orf2, the Human Orthologue of the Lmbr1 Gene. *Am J Hum Genet* 68: 38-45.

Knudsen, T. B., and D. M. Kochhar. 1981. The role of morphogenetic cell death during abnormal limb-bud outgrowth in mice heterozygous for the dominant mutation Hemimelia-extra toe (Hmx). *J Embryol Exp Morphol* 65: 289-307.

Kraus, P., D. Fraidenraich, and C. A. Loomis. 2001. Some distal limb structures develop in mice lacking Sonic hedgehog signaling. *Mech Dev* 100: 45-58.

Lettice, L., J. Hecksher-Sorensen, and R. E. Hill. 1999. The dominant hemimelia mutation uncouples epithelial-mesenchymal interactions and disrupts anterior mesenchyme formation in mouse hindlimbs. *Development* 126: 4729-36.

Masuya, H., T. Sagai, K. Moriwaki, and T. Shiroishi. 1997. Multigenic control of the localization of the zone of polarizing activity in limb morphogenesis in the mouse. *Dev Biol* 182: 42-51.

Nehls, M., D. Pfeifer, G. Micklem, C. Schmoor, and T. Boehm. 1994. The sequence complexity of exons trapped from the mouse genome. *Curr Biol* 4: 983-9.

Pearse, R. V., 2nd, and C. J. Tabin. 1998. The molecular ZPA. *J Exp Zool* 282: 677-90.

Qu, S., K. D. Niswender, Q. Ji, R. van der Meer, D. Keeney, M. A. Magnuson, and R. Wisdom. 1997. Polydactyly and ectopic ZPA formation in Alx-4 mutant mice. *Development* 124: 3999-4008.

Radhakrishna, U., J. L. Blouin, H. Mehenni, U. C. Patel, M. N. Patel, J. V. Solanki, and S. E. Antonarakis. 1997. Mapping one form of autosomal dominant postaxial polydactyly type A to chromosome 7p15-q11.23 by linkage analysis. *Am J Hum Genet* 60: 597-604.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY.

Sharpe, J., L. Lettice, J. Hecksher-Sorensen, M. Fox, R. Hill, and R. Krumlauf. 1999. Identification of sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr Biol* 9: 97-100.

Takahashi, E., T. Hori, P. O'Connell, M. Leppert, and R. White. 1990. R-banding and nonisotopic in situ hybridization: precise localization of the human type II collagen gene (COL2A1). *Hum Genet* 86: 14-6.

Toledo, S. P., and P. H. Saldanha. 1969. A radiological and genetic investigation of acheiropody in a kindred including six cases. *J Genet Hum* 17: 81-94.

Toledo, S. P., P. H. Saldanha, A. Borelli, and A. B. Cintra. 1972. Further data on acheiropody. *J Genet Hum* 20: 253-8.

Tsukurov, O., A. Boehmer, J. Flynn, J. P. Nicolai, B. C. Hamel, S. Traill, D. Zaleske, H. J. Mankin, H. Yeon, C. Ho, and *et al.* 1994. A complex bilateral polysyndactyly disease locus maps to chromosome 7q36. *Nat Genet* 6: 282-6.

Zguricas, J., H. Heus, E. Morales-Peralta, G. Breedveld, B. Kuyt, E. F. Mumcu, W. Bakker, N. Akarsu, S. P. Kay, S. E. Hovius, L. Heredero-Baute, B. A. Oostra, and P. Heutink. 1999. Clinical and genetic studies on 12 preaxial polydactyly families and refinement of the localisation of the gene responsible to a 1.9 cM region on chromosome 7q36. *J Med Genet* 36: 32-40.

Chapter 7

General Discussion
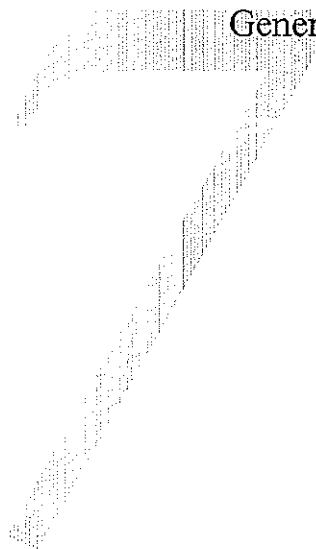
Discussion

Limb development is a model system for embryonal development. In order to identify genes important in limb development, we studied several hereditary limb phenotypes in mouse and humans and tried to identify the underlying genetic causes. Two phenotypes were studied. The first was the mouse *tipsy* phenotype, a syndactyly of the second and third digits in hindlimbs. This phenotype arose after transgene insertion, and we found that the transgene had inserted in another gene, *Fibrillin 2*. The second phenotype is a number of related limb malformations in human and mouse that map to human 7q36 and the mouse syntenic region on chromosome 5p. We found that at least some of these phenotypes are due to mutations in an intron of *Lmbr1/LMBR1*, and that these mutations influence the transcription of the Shh gene that lies upstream of this region.

In finding genes responsible for mutant phenotypes, we use the positional candidate cloning approach. This method is based on the assumption that we know enough about the function of genes to predict which genes are the most likely candidates in a critical region. But do we? Chapter 3 and 6 of this thesis can be used as an argument that we should be cautious. This will be explained in the next two paragraphs.

Tipsy

As discussed in Chapter 3, the *Fibrillin 2* mutations in the *tipsy* and the *shaker with syndactylism* mice were an unexpected finding. The function of the Fibrillins was thought to be known: they function in the extracellular matrix as part of the microfibrils that provide strength and elasticity (Sakai *et al.* 1986, Zhang *et al.* 1994). Mutations in *Fibrillin-1* lead to a phenotype that is compatible with this function, with cardiovascular and ocular abnormalities (Dietz *et al.* 1991) (MIM: 154700). The mutations found in *Fibrillin 2* lead to a related, but distinct phenotype, including joint contractures (Putnam *et al.* 1995) (MIM: 121050). It is not known why these joint contractures occur, and what the function of Fibrillin 2 is in the joints. The overgrowth of the long bones seen in both Marfan and CCA patients, and the arachnodactyly in CCA patients are not easy to explain either. We do know that most of the mutations in these patients disrupt an EGF like binding domain that binds calcium, and it is thought that this binding stabilizes the

Fibrillin protein (Robinson and Godfrey 2000). Mutations would then lead to destabilisation of microfibrils. Indeed, abnormalities in the microfibrils of Marfan patients have been observed by immunofluorescence (Godfrey *et al.* 1990, Hollister *et al.* 1990) and electron microscopy (Robinson and Godfrey 2000).

The Fibrillins are thought to assemble into long stretches of aligned proteins (Baldock *et al.* 2001). Together with other proteins, these multimers form a scaffold for elastin deposition, resulting in formation of the microfibril. The mutations in CCA and Marfan are dominant, and it is thought that the conformation of the Fibrillins change due to these mutations (Downing *et al.* 1996). This conformational change may actively disrupt the Fibrillin multimer: If every other Fibrillin 2 molecule is wobbly and does not fit in, it is likely that the whole complex is less stable, resulting in weak microfibrils. If this is the result of single nucleotide changes or exon deletions, then what will happen if the whole gene is deleted? Recessive mutations, such as deletion of the entire *Fibrillin* gene or transgene insertions that result in a lower expression of the *Fibrillin* gene, cause the absence of protein, or less available protein. This will probably also have an effect on the microfibrils.

The finding of *Fibrillin 2* mutations in syndactylous mice shed a new light on the function of Fibrillin and its associated proteins in limb development. Elastic fibers contain Elastin and Fibrillin, and Elastin-containing fibers have been detected in developing limbs, radiating from the core of the digits towards the ectoderm (Hurle *et al.* 1994). It is possible that the fibers have a function here in separating the digits. Fibrillin was not found in the developing limb until the formation of the joints, but the detection method was probably specific for Fibrillin-1. It is possible that Fibrillin 2 does function in the formation of this Elastin scaffold. If this is true, mutations in Fibrillin 2 may cause syndactyly because the elastic fibers cannot function properly.

There is a wealth of information on elastic fibers, but currently, not much is known about the function of Fibrillin 2. It is known that expression of Fibrillin 2 precedes that of Fibrillin-1 in most tissues during development, suggesting that Fibrillin 2 has a function in the formation of these tissues (Zhang *et al.* 1995). In quail embryos, the molecule forms a network around prospective somites and a 'caging' function of the protein was proposed, whereby the network separates new structures from surrounding

tissue (Rongish *et al.* 1998). When the *Fibrillin 2* gene was cloned in rat, antisense probes were used to block translation of the protein (Yang *et al.* 1999). This resulted in reduction of branching in the lung. A logical assumption would therefore be that absence of Fibrillin 2 leads to lung abnormalities, but no such phenotype was observed in any of the *ty* and *sy* mice. This may be due to the difference in technique, but more detailed research is necessary to explain this difference.

The syndactyly in the mouse Fibrillin mutants directs us to a possible new role for Fibrillins. We propose a model in which they provide a structural role in digit separation, or in active definition of the digits. Here, microfibrils, and hence Fibrillins, are the effectors of a signalling process. Too little Fibrillin 2 results in too few or too weak microfibrils, and the web is not strong enough to keep the developing digits apart. A similar mutation in Fibrillin-1 will probably not result in such a phenotype, because Fibrillin-1 is a more structural protein which is expressed after digits have formed.

PPD

Phenotype-genotype relationships are not always obvious, and this can also be said for the study of preaxial polydactyly. This study started with a large family, an almost completely penetrant, dominant mutation; a textbook case for linkage analysis. However, as time went by, other phenotypes were linked to the region, and different mutations were found in mice and humans, and currently there is not a single model that explains satisfactorily all the phenotypes linked to chromsome 7q36 and its syntenic region in the mouse, chromsome 5p.

When we started the current study, the genes responsible for two different phenotypes were localized on human chromosome 7q36, upstream of *Shh*: complex polysyndactyly (CPS) and preaxial polydactyly (PPD). On the syntenic region in mice, at chromosome 5p, two similar phenotypes were localized: the syndactyly phenotype *hammertoe (Hm)* and the polydactyly phenotype *hemimelic extra toes (Hx)*, most likely the mouse counterparts of the human CPS and PPD phenotypes. The simplest model to explain these four phenotypes was that they resulted from mutations in a single gene, with the syndactyly phenotypes caused by one type of mutation and the polydactyly phenotypes by a different kind (Heus *et al.* 2001). However, at the start of this study, no mutations were known.

151

We narrowed down the critical region of PPD to three genes, *Hlxb9*, *C7orf2* and *C7orf3*. None of the PPD patients tested showed any mutations in the coding region of these genes (see Chapter 4a). In the current study, two more genes were found in the region with the help of in silico cloning procedures, *RNF32* and *C7orf13* (see Chapter 4b). Both of these genes are expressed in testis, which makes them unlikely candidate genes and indeed neither contains mutations in PPD patients.

The mouse critical region was narrowed down to two genes, *Lmbr1* and *Lmbr2*, by Clark *et al.* (Clark *et al.* 2000). Both genes are located in the region syntenic with the human PPD critical region, and, similar to the human PPD patients, neither *Hx* nor *Hm* mutants showed any mutations within the coding regions of these genes (see Chapter 4b). *Lmbr2* is the mouse homolog of *RNF32*, and it was assumed to be an unlikely candidate gene for similar reasons. *Lmbr1* is the mouse ortholog of *C7orf2* and this gene was renamed *LMBR1*. Clark *et al.* showed that the gene is expressed in developing limbs of wildtype and mutant mice. They claimed that in developing *Hx* limbs the expression of *Lmbr1* was reduced for a short time, and indeed a band of lower intensity is seen on the Northern blot. However, this blot contained an unusually high amount of RNA and the control probe is also reduced in expression in *Hx* limbs. Furthermore, this is the only indication that *Lmbr1* functions in limb development, since attempts to show *Lmbr1* expression by whole mount *in situ* hybridization failed (Clark *et al.* 2000 and our own unpublished results).

Subsequently, another human limb malformation was linked to the region on 7q36: acheiropodia, a recessive disorder that causes truncation of all four limbs (Escamilla *et al.* 2000). Because mouse *Lmbr1* was implicated in the mouse limb mutants (Clark *et al.* 2000), *LMBR1* was the most likely candidate gene in the region for acheiropodia. Therefore, we screened acheiropodia patients for mutations in this gene, and found a deletion of exon 4 and flanking sequences, leading to a frameshift and a premature stop codon in exon 6 (see Chapter 5). This finding suggests that the *LMBR1* gene has indeed a function in limb development.

A syndrome with PPD, F syndrome, was subsequently linked to chromosome 7q36 (Dundar *et al.* 2001). In association with PPD, patients show syndactyly of hands and feet, and deformities of the sternum. It was suggested that the causal mutation affects

name	mutation	phenotype
Sasquatch mouse	transgene insertion	PPD dominant
PPD human	translocation 5:7	PPD dominant
acheiropody human	missing exon	truncation recessive
Lmbr1^{ATG} mouse	no Lmbr1 transcript	some reduction recessive
Hdh^{dl4j} mouse	no 7q36	none dominant[1]
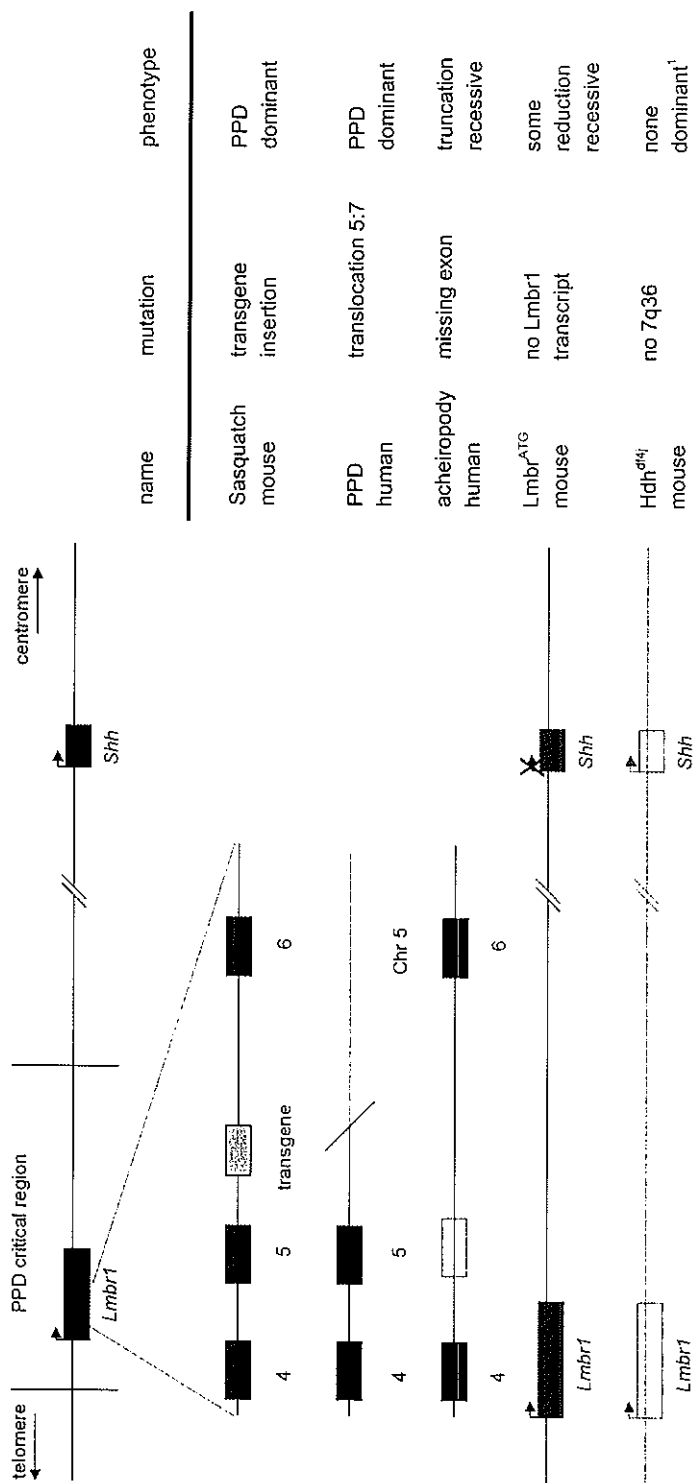
Figure 1. Mutations in Lmbr1 region.
[1] homozygous deletion of Shh is not viable
PPD is preaxial polydactyly

Shh expression directly or indirectly, since ectopic expression of *Shh* in lateral plate mesoderm leads to sternal defects in mice. No mutations were reported.

Another mouse preaxial polydactyly mutant was found to be localized close to the *Hx/Hm* phenotypes on mouse chromosome 5. The *Sasquatch (Ssq)* mutation is an insertion of a transgene upstream of *Shh*, resulting in ectopic expression of *Shh* at the anterior side of the limb bud. Ectopic expression of *Shh* leads to the formation of extra digits, and such expression has also been observed in other polydactyly mutants (Masuya *et al.* 1995). It is likely that the polydactyly observed in *Ssq* is the result of the ectopic *Shh* expression, and that the transgene insertion causes in this aberrant expression pattern. Interestingly, the integrated transgene, *Hoxb1*, is also expressed in the posterior and anterior side of the limb bud, following the outline of the *Shh* expression, whereas in wildtype mice, expression of *Hoxb1* is restricted to the rhombomeres. This suggests that *Hoxb1* expression in the *Ssq* mutant is influenced by a regulatory element of *Shh*. It was proposed that the *Hoxb1* transgene had integrated into an inhibitory element for anterior *Shh* expression, thereby disrupting it, while an activating element would be located close to the transgene insertion site, exerting its function both on *Shh* and the integrated *Hoxb1* gene (Sharpe *et al.* 1999). As described in Chapter 3, the *Hoxb1* transgene was subsequently found to have integrated in intron 5 of *Lmbr1*.

The other mutation reported in Chapter 3 is a translocation which occurred in a human PPD patient. Interestingly, the translocation breakpoint is in intron 5, suggesting that the mutational mechanism leading to PPD in this patient and in the *Ssq* mutant is the same.

An overview of the mutations is shown in fig. 1.

There are two models for the explanation of these phenotypes. In the first model, LMBR1 itself has a function in limb development, and disruption of the gene, or a change in its expression, results in limb malformations. This model is called the "protein hypothesis". The other model is that *Shh* is misexpressed in all these limb mutants, and *LMBR1* contains regulatory elements of *Shh*; the "element hypothesis".

The protein hypothesis

The acheiropodia mutation agrees best with the protein hypothesis. The mutation removes an exon of *LMBR1*, which results in a frameshift and a premature stop codon, presumably leading to a truncated protein (see Chapter 5). Acheiropodia is a recessive condition, and it is unlikely that it results from a gain of function of the protein. This means that the acheiropodia mutation is probably a loss of function allele of *LMBR1*.

Another argument for the protein hypothesis is the expression of *Lmbr1* during limb formation, and its temporarily reduced expression in homozygous *Hx* mutants. This suggests that the gene has a function in that stage of development, and that lower expression of *Lmbr1* results in the polydactyly phenotype.

Truncations of the *Lmbr1* transcript are found in the *Ssq* mutant and are likely to be present in the PPD patient with the translocation breakpoint in intron 5 (see Chapter 6). These mutations lead to a dominant polydactyly phenotype. In acheiropodia patients, the removal of exon 4 introduces a frameshift and a premature stop codon. This mutation is recessive and leads to a reduction phenotype. If all three mutant transcripts are translated into (partial) Lmbr1/LMBR1 protein, the question then arises if this truncated protein is functional, and how the differences between these phenotypes can be explained.

In a mouse mutant with a chromosomal deletion that includes *Lmbr1* and *Shh, Hdh^{df4j}* , no polydactyly or syndactyly was observed in heterozygotes (Schimenti *et al.* 2000). However, in 25% of these mice, fusion of wristbones was seen (Clark *et al.* 2001). When *Lmbr1* was mutated by replacement of the first exon with a Neo cassette, a low incidence of limb abnormalities were seen, including carpal fusions, reduced length of phalanges, and rare digit reductions (Clark *et al.* 2001). The mutation was named *Lmbr1^{ATG}*. Such a replacement mutation usually results in a knocked out allele, but in this case, the knockout was incomplete: Some *Lmbr1* transcripts were found in the homozygous mutant, possibly initiated in the Neo cassette that was used for generating the knockout. Interestingly, when mice with this mutation were crossed with the *Lmbr1* deletion mutant, a high incidence of limb abnormalities were seen in the hemizygote, most of which were reductions of postaxial digits (Clark *et al.* 2001). The authors suggest that *Lmbr1* itself is necessary for limb formation.

155

Clark *et al.* also suggest that the dominant mouse and human mutations are gain of function regulatory changes resulting in upregulation of Lmbr1 (Clark *et al.* 2001). This overexpression leads to increased digit number. Loss of function or decreased activity of Lmbr1 produces a reciprocal phenotype of reduced digit number and loss of distal structures (in the acheiropodia and mouse knockout phenotypes). However, this does not explain why a protein truncation in acheiropodia leads to a reduction phenotype whereas a truncation in the transcript (and, possibly, in the protein) in *Ssq* mice and the translocation patient leads to PPD.

The element hypothesis

Developing limbs of *Hx* and *Ssq* mice contain a duplicated zone of *Shh* expression (Masuya *et al.* 1995, Sharpe *et al.* 1999), similar to that seen in two other polydactyly mutants, *Rim4* and *Xt* (Masuya *et al.* 1995). It is likely that the polydactyly in humans is also the result of anterior expression of *Shh*. But what is the mechanism? Does the Lmbr1 protein have a role in regulating expression of *Shh*, or does the *Lmbr1* gene contain one or several elements that guide limb specific *Shh* expression? The latter explanation is somewhat unusual, but in light of the current evidence it is an attractive hypothesis.

As described in Chapter 1, Shh is very important in patterning of the early embryo. It acts as an inductive signal in patterning of the ventral neural tube, the antero-posterior limb axis and the ventral somites (Echelard *et al.* 1993, Johnson *et al.* 1994, Riddle *et al.* 1993). In humans, mutations in the *Shh* gene result in holoprosencephaly (HPE), a heterogeneous disorder involving the forebrain and midface (MIM: 142945) (Roessler *et al.* 1996). HPE can also be caused by chromosomal rearrangements upstream of *Shh*, presumably involving a *cis* regulatory element located >250 kb upstream of the gene (Belloni *et al.* 1996). Deletion of this element was proposed to lead to aberrant expression of *Shh* during development. A mutation in a regulatory element of *Shh* was also suggested to cause preaxial polydactyly (Masuya *et al.* 1995). Such upstream regulatory elements have also been proposed for SOX9 in campomelic dysplasia and downstream elements have been implicated in PAX6 in aniridia (Lauderdale *et al.* 2000, Wunderle *et al.* 1998). Furthermore, mutations upstream of the human β-globin locus

have led to the identification of the upstream Locus Control Region (LCR) (see Box 1) (Grosveld *et al.* 1987).

It appears that *cis* regulatory elements can be far upstream from the start site of transcription, and the PPD phenotypes may be a new example. In the *Hx* and *Hm* mutations, different elements may be mutated, leading to PPD and syndactyly, respectively. PPD is due to ectopic expression of Shh, and syndactyly may be caused by insufficient Shh. The PPD element may be located in intron 5 of *Lmbr1*, the location of the transgene insertion in *Ssq* and of the translocation breakpoint in the human PPD patient (see Chapter 6). The *cis-trans* genetic test described in Chapter 6 shows that the *Ssq* mutation must be in *cis* to *Shh* in order to show PPD. This strongly suggests that a regulatory element for *Shh* resides in intron 5 of *Lmbr1*. This intron may also harbour mutations in the familial PPD cases. The location of the regulatory element mutated in *Hm* mice and in CPS patients is unknown, but the acheiropodia mutation may indicate its position. Broadly speaking, both CPS and acheiropodia are reduction phenotypes, and the same regulatory element may be involved. In acheiropodia, some flanking intronic sequence of *LMBR1* is deleted together with exon 4, and it is possible that it is not the missing exon of *LMBR1*, but the deletion of a regulatory element in this flanking sequence that is the cause of the condition. It is feasible that a gene like *Shh*, with many functions during embryonic development, is regulated on several different levels. There is no physical reason why regulatory elements have to be located close to, or within a gene. On the contrary, many enhancer elements have been found outside the coding regions (see Box 1) and it is easy to imagine how loops in the DNA position enhancers and promoter of a gene in close proximity.

This model also explains the fact that the heterozygous *Hdh*^{df4j} mutant (with deletion of *Lmbr1*) hardly shows a phenotype: this mutation involves *Shh*, therefore both the element and the gene it act on is missing. The resulting low incidence of wrist bone fusion may be caused by insufficient levels of *Shh* on the other allele.

Unfortunately, this model does not explain the phenotype of Clark *et al.*'s *Lmbr1* mutant. The low incidence of limb reductions seen in this mutant, with the Neo cassette in place of exon 1, may be explained if some *Shh* regulatory elements close to exon 1 are

157

Box 1. Distant *cis* regulatory elements

Mutations outside coding regions have been implicated in many hereditary diseases. In some cases, mutations have been found relatively far upstream or downstream of coding regions, indicating that regulatory elements are located in these positions.

SOX9 mutations have been found in campomelic dysplasia, a disorder characterized by congenital bowing and angulation of long bones, together with other skeletal and extraskeletal defects (MIM: 114290). In addition to frameshift and premature termination mutations in the coding region of *SOX9*, chromosomal rearrangements 88-950kb upstream of the gene were found in patients (Pfeifer *et al.* 1999, Young *et al.* 1992). Experiments in mice showed that important regulatory elements are scattered over a large region upstream of *SOX9* and that a rearrangement upstream of *SOX9*, similar to those observed in campomelic dysplasia patients, leads to a substantial reduction of *SOX9* expression (Wunderle *et al.* 1998).

PAX6 is important in development of the vertebrate eye. Haploinsufficiency of *PAX6* in humans causes aniridia, which is incomplete formation of the iris, whereas homozygous *PAX6* mutations leads to absence of the eyes, underdevelopment of the nose and central nervous system defects (Glaser *et al.* 1994) (MIM: 106210). Interestingly, a paracentric inversion and a translocation with breakpoints 85-124 kb downstream of *PAX6* are correlated with aniridia in several families (Fantes *et al.* 1995, Ton *et al.* 1991). Similarly, two small deletions >11 kb downstream of *PAX6* were reported to cause a phenotype that was indistinguishable from that caused by mutations in the *PAX6* coding region and it was shown that *PAX6* was transcribed only from the normal allele, suggesting that remote 3-prime regulatory elements are required for initiation of *PAX6* expression (Lauderdale *et al.* 2000).

The Locus Control Region (LCR) of the β-globin gene cluster was identified in a study of thalassemia, a hereditary blood disorder. Thalassemia is a distortion of the α/β globin ratio, due to reduced expression of one of the globins. Many mutations have been found in thalassemia patients, both single base mutations and deletions in the gene and/or promoter region. A Dutch type β thalassemic patient had an intact adult β globin promoter and gene sequences, but the gene was not transcribed (Kioussis *et al.* 1983, Van der Ploeg *et al.* 1980). This proved to be due to a deletion in an upstream region that contained an important regulatory elements, the LCR (Grosveld *et al.* 1987). This LCR has been studied extensively and appeared to be important in the order and rate of transcription of the β globin genes. Since its discovery, several other LCRs have been found in the human genome.

removed, but this suggests the existence of a third element. Furthermore, the phenotype of the hemizygote (the cross between Hdh^{df4j} and $Lmbr1^{ATG}$) can not be explained.

However, there is a strong argument in favour of the element hypothesis: An ortholog of *Lmbr1* is present in *Fugu rubripes*, upstream of *Shh*, and a paralog of *Lmbr1* (see below) lies upstream of human and mouse *Desert hedgehog (Dhh)*, a *Shh* paralog. Therefore, not only the gene, but also its localisation with respect to *hedgehog* has been conserved during vertebrate evolution. This could be due to non-divergence during evolution, but this is unlikely as none of the other genes close to *LMBR1* on chromosome 7q36 has a paralog close to *Dhh*. Therefore, it is likely that the gene order is conserved because regulatory elements of *hedgehog* genes are located in *Lmbr1* homologs.

Which model is true?

Currently, the data are conflicting, and the choice of 'protein' versus 'element' model is a hard to make. However, the majority of the data agrees with, or is explainable within, the 'element' hypothesis. In fact, the 'protein' model is only supported by the reduction of expression of *Lmbr1* in homozygous *Hx* developing limbs, and the acheiropodia mutation. The latter can be explained within the 'element' model, if we assume that there are regulatory elements for Shh located close to exon 4, and that these are deleted in acheiropodia.

What if both models are true? *Lmbr1* is strongly conserved between mouse and human, and even *Drosophila, C. Elegans* and *Fugu* orthologs exist, all with the same putative 9-transmembrane conformation. This strong conservation suggests that the protein has an important function in these animals. It is difficult to argue that mutation of *Lmbr1*, as seen in acheiropodia, only leads to a phenotype because the gene contains elements important for *Shh* expression. On the other hand, if the Lmbr1 protein is important for limb development, then why is its location with respect to *Shh* conserved? The localisation of an *Lmbr1* homolog close to *Dhh* is an indication that these two genes need to be located close together.

I propose the following model: The *Lmbr1* region contains regulatory elements for *Shh* expression, and possibly also for *Lmbr1* expression. Mutations in this region lead to a change in either or both of these expression patterns. Furthermore, Lmbr1 functions in the Shh pathway, possibly as a receptor for Shh. The existence of such a second hedgehog

159

receptor has recently been proposed (Ramirez-Weber et al. 2000). In addition, one of the few known nine-transmembrane proteins in human is Delta7-sterol reductase, which binds 7-dehydrocholesterol and reduces it to cholesterol (Moebius et al. 1998). Mutations in the gene for Delta7-sterol reductase cause Smith-Lemli-Opitz syndrome, a complex disorder of which polydactyly is a feature (Fitzky et al. 1998). Since Shh is cholesterol-modified (Porter et al. 1996), it is possible that Lmbr binds Shh on this cholesterol moiety.

A paralog of *LMBR*, called confusingly *LIMR*, was recently cloned (Wojnar et al. 2001). The protein has nine putative transmembrane regions and binds Lipocalin, a protein that binds many hydrophobic molecules, one of which is cholesterol (Glasgow et al. 1995). It was suggested that LIMR is a receptor for the lipophilic ligands of lipocalin (Wojnar et al. 2001).

If Lmbr1 is a receptor for Shh, mutations leading to absent or nonfunctional Lmbr1 protein may result in improper transduction of the Shh signal, resulting in truncation of the limbs. Such truncations are also seen in *Shh* knockout mice (Chiang et al. 1996).

The argument can only be settled when the other PPD causing mutations are identified and the function of Lmbr1 is determined. Interaction studies with Lmbr1 may reveal show binding to Shh or it may reveal another ligand or other associated proteins, and these may elucidate the function of Lmbr1. To study the PPD region, we should try to identify regulatory elements. This may be done by comparison of mouse and human genomic DNA and the identification of conserved regions. Such regions are likely to contain important elements, however, with this method smaller elements will not be found. Transcription factors often bind to short sequences in unwound DNA, which renders the DNA sensitive to restriction enzymes. Localisation of DNase1 hypersensitive sites indicates possible transcription binding sites, and could be used in the PPD critical region. Another means of regulation is DNA methylation, whereby one allele is inactivated. Since there is some evidence that *Lmbr1* mutations acts on the *Shh* gene in *cis* (see Chapter 6), the methylation status of the PPD critical region may differ between the alleles.

In the mouse, engineered mutations in the PPD critical region may reveal the regulation mechanism of the locus. Such experiments may show if there is a specific element located in intron 5 of *Lmbr1*, or if the mutations are position effects. Together, these studies will elucidate the mutation mechanism that causes the limb phenotypes located on chromosome 7q36.

In silico cloning

Chapter 4b describes the finding of two genes that were found by in silico cloning procedures. These genes had not been previously found, by traditional methods. In contrast, all the genes found before in the PPD critical region were easily identified by computer, showing that the identification of genes by computer is a reliable procedure.

The Human Genome Sequence becomes increasingly more annotated, and experimental data are integrated with gene predictions. This results in better gene models, and consequently better candidate disease gene predictions. For the PPD project described in this thesis, the most important advantage of computer cloning is the identification of all genes in the region. Selection of the best candidate gene is important at the start of the project, because if the selected candidate genes do not contain mutations, the other genes in the region will subsequently be sequenced, even if they are unlikely candidate genes. It can be argued that this is not necessary, as in the case of C7orf13 and RNF32, genes that are expressed only in testis, but since we do not know the pattern of expression during limb development these genes can not be ruled out until they have been screened for mutations in patients.

This poses a problem in larger candidate regions, with dozens of known genes, predicted genes and unidentified transcripts. Known gene functions and phenotypes resulting from mutations in these genes do not always show a clear relationship, therefore every gene in the critical region remains a candidate. What should we do when the obvious candidate genes do not contain mutations? We should probably look at the remaining candidate genes in a bit more detail.

Important information can be derived from expression patterns of candidate genes. If a gene is expressed during embryogenesis, it is a candidate for all congenital malformations, and it is worthwile to screen the affected tissues for expression of the gene. An example is PPD: a true candidate gene should be expressed in developing limbs.

161

Testing expression of candidate genes may generate a similar amount of work to screening the gene for mutations. The EST databases contain information on the tissue of which the transcripts were derived and will therefore give some insight into the expression pattern. Unfortunately, most of the libraries are normalized to reduce the representation of abundant gene transcripts, therefore the amount of ESTs derived from a library does not represent the amount of transcripts in the corresponding tissue. At this moment it is only possible to use this information to exclude genes: candidate genes for brain diseases that are not found in any brain cDNA library can be excluded. Therefore it is necessary to establish libraries derived from different developmental stages, or to use differential display or other methods to detect changes in expression levels of genes.

In the near future it will be hard to exclude candidate genes without mutation analysis, and we may end up screening all the genes in a region. On the other hand, if we do not screen unlikely candidate genes, we may never identify their function. Now that the Human Genome Project is almost finished, large scale functional analysis of the genes is the next challenge. Until this is achieved, the main advantage of computer cloning is the speed with which candidate genes are identified in positional cloning efforts. This allows us to screen much larger candidate regions than before, and it opens the road to finding genes for multifactorial disorders.

References

Baldock, C., A. J. Koster, U. Ziese, M. J. Rock, M. J. Sherratt, K. E. Kadler, C. A. Shuttleworth, and C. M. Kielty. 2001. The supramolecular organization of fibrillin-rich microfibrils. *J Cell Biol* 152: 1045-56.

Belloni, E., M. Muenke, E. Roessler, G. Traverso, J. Siegel-Bartelt, A. Frumkin, H. F. Mitchell, H. Donis-Keller, C. Helms, A. V. Hing, H. H. Heng, B. Koop, D. Martindale, J. M. Rommens, L. C. Tsui, and S. W. Scherer. 1996. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat Genet* 14: 353-6.

Chiang, C., Y. Litingtung, E. Lee, K. E. Young, J. L. Corden, H. Westphal, and P. A. Beachy. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383: 407-13.

Clark, R., P. Marker, and D. Kingsley. 2000. A novel candidate gene for mouse and human preaxial polydactyly with altered expression in limbs of Hemimelic extra-toes mutant mice. *Genomics* 67: 19-27.

Clark, R. M., P. C. Marker, E. Roessler, A. Dutra, J. C. Schimenti, M. Muenke, and D. M. Kingsley. 2001. Reciprocal mouse and human limb phenotypes caused by gain- and loss-of-function mutations affecting lmbr1. *Genetics* 159: 715-26.

Dietz, H. C., G. R. Cutting, R. E. Pyeritz, C. L. Maslen, L. Y. Sakai, G. M. Corson, E. G. Puffenberger, A. Hamosh, E. J. Nanthakumar, S. M. Curristin, and *et al.* 1991. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 352: 337-9.

Downing, A. K., V. Knott, J. M. Werner, C. M. Cardy, I. D. Campbell, and P. A. Handford. 1996. Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* 85: 597-605.

Dundar, M., T. M. Gordon, I. Ozyazgan, F. Oguzkaya, Y. Ozkul, A. Cooke, A. G. Wilkinson, S. Holloway, F. R. Goodman, and J. L. Tolmie. 2001. A novel acropectoral syndrome maps to chromosome 7q36. *J Med Genet* 38: 304-9.

Echelard, Y., D. J. Epstein, B. St-Jacques, L. Shen, J. Mohler, J. A. McMahon, and A. P. McMahon. 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* 75: 1417-30.

Escamilla, M. A., M. C. DeMille, E. Benavides, E. Roche, L. Almasy, S. Pittman, J. Hauser, D. F. Lew, N. B. Freimer, and M. R. Whittle. 2000. A minimalist approach to gene mapping: locating the gene for acheiropodia, by homozygosity analysis. *Am J Hum Genet* 66: 1995-2000.

Fantes, J., B. Redeker, M. Breen, S. Boyle, J. Brown, J. Fletcher, S. Jones, W. Bickmore, Y. Fukushima, M. Mannens, and *et al.* 1995. Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum Mol Genet* 4: 415-22.

Fitzky, B. U., M. Witsch-Baumgartner, M. Erdel, J. N. Lee, Y. K. Paik, H. Glossmann, G. Utermann, and F. F. Moebius. 1998. Mutations in the Delta7-sterol reductase gene in patients with the Smith-Lemli-Opitz syndrome. *Proc Natl Acad Sci U S A* 95: 8181-6.

Glaser, T., L. Jepeal, J. G. Edwards, S. R. Young, J. Favor, and R. L. Maas. 1994. PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects. *Nat Genet* 7: 463-71.

Glasgow, B. J., A. R. Abduragimov, Z. T. Farahbakhsh, K. F. Faull, and W. L. Hubbell. 1995. Tear lipocalins bind a broad array of lipid ligands. *Curr Eye Res* 14: 363-72.

Godfrey, M., V. Menashe, R. G. Weleber, R. D. Koler, R. H. Bigley, E. Lovrien, J. Zonana, and D. W. Hollister. 1990. Cosegregation of elastin-associated microfibrillar abnormalities with the Marfan phenotype in families. *Am J Hum Genet* 46: 652-60.

Grosveld, F., G. B. van Assendelft, D. R. Greaves, and G. Kollias. 1987. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 51: 975-85.

Heus, H. C., A. J. Luijsterburg, M. J. van Baren, G. J. Breedveld, M. N. Joosse, I. M. Nieuwenhuizen, C. Vermeij-Keers, B. A. Oostra, and P. Heutink. 2001. Hemimelic extra toes and Hammer toe are distinct mutations that show a genetic interaction. *Mamm Genome* 12: 77-9.

Hollister, D. W., M. Godfrey, L. Y. Sakai, and R. E. Pyeritz. 1990. Immunohistologic abnormalities of the microfibrillar-fiber system in the Marfan syndrome. *N Engl J Med* 323: 152-9.

Hurle, J. M., G. Corson, K. Daniels, R. S. Reiter, L. Y. Sakai, and M. Solursh. 1994. Elastin exhibits a distinctive temporal and spatial pattern of distribution in the developing chick limb in association with the establishment of the cartilaginous skeleton. *J Cell Sci* 107: 2623-34.

Johnson, R. L., R. D. Riddle, E. Laufer, and C. Tabin. 1994. Sonic hedgehog: a key mediator of anterior-posterior patterning of the limb and dorso-ventral patterning of axial embryonic structures. *Biochem Soc Trans* 22: 569-74.

Kioussis, D., E. Vanin, T. deLange, R. A. Flavell, and F. G. Grosveld. 1983. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* 306: 662-6.

Lauderdale, J. D., J. S. Wilensky, E. R. Oliver, D. S. Walton, and T. Glaser. 2000. 3' deletions cause aniridia by preventing PAX6 gene expression. *Proc Natl Acad Sci U S A* 97: 13755-9.

Masuya, H., T. Sagai, S. Wakana, K. Moriwaki, and T. Shiroishi. 1995. A duplicated zone of polarizing activity in polydactylous mouse mutants. *Genes Dev* 9: 1645-53.

Moebius, F. F., B. U. Fitzky, J. N. Lee, Y. K. Paik, and H. Glossmann. 1998. Molecular cloning and expression of the human delta7-sterol reductase. *Proc Natl Acad Sci U S A* 95: 1899-902.

Pfeifer, D., R. Kist, K. Dewar, K. Devon, E. S. Lander, B. Birren, L. Korniszewski, E. Back, and G. Scherer. 1999. Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *Am J Hum Genet* 65: 111-24.

Porter, J. A., K. E. Young, and P. A. Beachy. 1996. Cholesterol modification of hedgehog signaling proteins in animal development. *Science* 274: 255-9.

Putnam, E. A., H. Zhang, F. Ramirez, and D. M. Milewicz. 1995. Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder, congenital contractural arachnodactyly. *Nat Genet* 11: 456-8.

Ramirez-Weber, F. A., D. J. Casso, P. Aza-Blanc, T. Tabata, and T. B. Kornberg. 2000. Hedgehog signal transduction in the posterior compartment of the Drosophila wing imaginal disc. *Mol Cell* 6: 479-85.

Riddle, R. D., R. L. Johnson, E. Laufer, and C. Tabin. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* 75: 1401-16.

Robinson, P. N., and M. Godfrey. 2000. The molecular genetics of Marfan syndrome and related microfibrillopathies. *J Med Genet* 37: 9-25.

Roessler, E., E. Belloni, K. Gaudenz, P. Jay, P. Berta, S. W. Scherer, L. C. Tsui, and M. Muenke. 1996. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet* 14: 357-60.

Rongish, B. J., C. J. Drake, W. S. Argraves, and C. D. Little. 1998. Identification of the developmental marker, JB3-antigen, as fibrillin-2 and its de novo organization into embryonic microfibrous arrays. *Dev Dyn* 212: 461-71.

Sakai, L. Y., D. R. Keene, and E. Engvall. 1986. Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J Cell Biol* 103: 2499-509.

Schimenti, J. C., B. J. Libby, R. A. Bergstrom, L. A. Wilson, D. Naf, L. M. Tarantino, A. Alavizadeh, A. Lengeling, and M. Bucan. 2000. Interdigitated deletion complexes on mouse chromosome 5 induced by irradiation of embryonic stem cells. *Genome Res* 10: 1043-50.

Sharpe, J., L. Lettice, J. Hecksher-Sorensen, M. Fox, R. Hill, and R. Krumlauf. 1999. Identification of sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr Biol* 9: 97-100.

Ton, C. C., H. Hirvonen, H. Miwa, M. M. Weil, P. Monaghan, T. Jordan, V. van Heyningen, N. D. Hastie, H. Meijers-Heijboer, M. Drechsler, and *et al*. 1991. Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* 67: 1059-74.

Van der Ploeg, L. H., A. Konings, M. Oort, D. Roos, L. Bernini, and R. A. Flavell. 1980. gamma-beta-Thalassaemia studies showing that deletion of the gamma- and delta-genes influences beta-globin gene expression in man. *Nature* 283: 637-42.

Wojnar, P., M. Lechner, P. Merschak, and B. Redl. 2001. Molecular cloning of a novel lipocalin-1 interacting human cell membrane receptor using phage display. *J Biol Chem* 276: 20206-12.

Wunderle, V. M., R. Critcher, N. Hastie, P. N. Goodfellow, and A. Schedl. 1998. Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc Natl Acad Sci U S A* 95: 10649-54.

Yang, Q., K. Ota, Y. Tian, A. Kumar, J. Wada, N. Kashihara, E. Wallner, and Y. S. Kanwar. 1999. Cloning of rat fibrillin-2 cDNA and its role in branching morphogenesis of embryonic lung. *Dev Biol* 212: 229-42.

Young, I. D., J. M. Zuccollo, E. L. Maltby, and N. J. Broderick. 1992. Campomelic dysplasia associated with a de novo 2q;17q reciprocal translocation. *J Med Genet* 29: 251-2.

Zhang, H., S. D. Apfelroth, W. Hu, E. C. Davis, C. Sanguineti, J. Bonadio, R. P. Mecham, and F. Ramirez. 1994. Structure and expression of fibrillin-2, a novel microfibrillar component preferentially located in elastic matrices. *J Cell Biol* 124: 855-63.

Zhang, H., W. Hu, and F. Ramirez. 1995. Developmental expression of fibrillin genes suggests heterogeneity of extracellular microfibrils. *J Cell Biol* 129: 1165-76.

Summary

Embryonic development has been studied for many decades and on many levels. In recent years, the molecular genetics of embryogenesis has come into focus and much effort has been put in finding genes that regulate the formation of the embryo. The formation of the limb mirrors embryonic development: in many cases the same genes are expressed and processes of induction and pattern formations are similar. This is why the limb has long been used as a model system for development. To better understand embryonic development, genes involved in limb formation may be identified and characterised. This can be achieved by studying hereditary limb malformations and identifying the genetic cause.

This thesis covers three topics. Two of these are genes, or genomic regions, that function in limb development. The third is a discussion of how 'computer cloning' has sped up the process of finding and characterising genes.

The first gene discussed is *Fibrillin 2*, the deletion of which causes syndactyly in mice. We studied a mutant that shows syndactyly (webbing of the digits) between the second and third digits of the hindlimb. The mutation is recessive, and caused by integration of a transgene. The transgene itself does not cause the phenotype, since other transgenic mice with this gene do not show syndactyly. We used this transgene to identify the insertion site, on mouse chromosome 18, and identified a gene that was interrupted by the transgene: *Fibrillin 2*. This gene was known to function in the formation of connective tissue, and mutations in *Fibrillin 2* had been found in patients with Congenital Contractural Arachnodactyly, a connective tissue disorder. The fact that this gene is also important in limb formation was a new finding, and it shows that *Fibrillin 2* functions in several pathways during embryonic development.

The second topic of the thesis is the genomic region encompassing *Lmbr1*, which contains several mutations that cause different limb malformations. This study started originally with a large Dutch family with autosomal dominant pre-axial polydactyly (PPD), or duplications of the thumbs and/or index fingers. We found that this condition was linked to chromosome 7q36. This region is syntenic to the tip of mouse chromosome 5 (meaning that it contains the same genes), to which a mouse pre-axial polydactyly phenotype was linked, *Hemimelic extra toes* (*Hx*). Another limb phenotype, *Hammertoe* (*Hm*), was also mapped to this region in mouse. This mutant shows syndactyly. And

indeed a human syndactyly phenotype, complex polysyndactyly (CPS) was linked to human 7q36.

Subsequently, more limb phenotypes were linked to this region on human 7q36. One of these was acheiropodia, a recessive truncation of all four limbs. By this time we had identified several genes in the human candidate region. We started a cooperation with the authors of the acheiropodia linkage paper and showed deletion of an exon of *LMBR1* in these patients, leading to a frameshift and a truncation of the protein. This indicates that the LMBR1 protein functions in limb development. However, neither the PPD and CPS patients nor the mouse mutants *Hx* and *Hm* showed any mutations in the coding sequence of this gene.

Subsequently a mouse mutant was published with a transgene integration in this region, causing pre-axial polydactyly, similar to *Hx*. We collaborated and showed that the integration site lies in intron 5 of *Lmbr1*. Thereafter, a human translocation patient with a breakpoint in intron 5 of *LMBR1* was found, showing pre-axial polydactyly. It seems that there is an important regulatory element in this intron. This element may not be important for *Lmbr1* itself, as expression of *Lmbr1* can hardly be seen in limbs of normal and mutant mice. Rather, the element may have its influence on the gene *Sonic hedgehog*, which lies 1 Mb upstream and is known to be involved in limb formation. All pre-axial polydactyly mouse mutants tested so far show ectopic expression of this gene.

This means that expression of a gene may be governed by a faraway element and that a functional gene 'unit' may overlap other genes. It also shows the beautiful complexity of the genome.

This thesis also describes how computer analysis aids the discovery of novel genes and their function. A massive amount of data has become available with the genome projects of different organisms, and comparing and organising the information on the similar and different genes of these organisms will add considerably to our knowledge and understanding of genetics.

Samenvatting

De ontwikkeling van het embryo wordt al decennialang op allerlei manieren bestudeerd. De laatste jaren is er veel interesse voor de moleculaire genetica van de embryogenese en er wordt veel moeite gedaan om 'nieuwe' genen te vinden die de vorming van het embryo reguleren. De ontwikkeling van de ledemaat weerspiegelt die van het embryo: in veel gevallen komen dezelfde genen tot expressie en de processen van inductie en patroonvorming zijn vergelijkbaar. Dit is de reden dat de ledemaat al lange tijd gebruikt wordt als modelsysteem voor embryogenese. De identificatie en karakterisatie van genen die een rol spelen bij ledemaatvorming kan dus een ingang zijn tot beter begrip van embryonale ontwikkeling. Een manier om dit te bereiken is door erfelijke ledemaatafwijkingen te bestuderen en de genetische oorzaak te bepalen.

Dit proefschrift omvat drie onderwerpen. Twee hiervan zijn genen, of genomische regio's, die een functie hebben in ledemaatontwikkeling. De derde is een bespreking van 'kloneren per computer' en hoe dit het vinden en het karakteriseren van genen versnelt.

Het eerste gen dat besproken wordt is *Fibrilline 2*. Deletie van dit gen veroorzaakt syndactylie (fusering van tenen) bij muizen. We onderzochten een mutant die syndactylie vertoont tussen de tweede en derde teen van de achterpoot. Deze mutatie is recessief, en wordt veroorzaakt door integratie van een transgen. Dit transgen zelf veroorzaakt de afwijking niet, want andere muizen die dit transgen hebben, vertonen geen syndactylie. We gebruikten het transgen om het integratiepunt te vinden, op muischromosoom 18, en vonden daar een gen wat onderbroken werd door het transgen: *Fibrilline 2*. Van dit gen was bekend dat het een rol speelde bij de vorming van bindweefsel, en mutaties in *Fibrilline 2* waren gevonden in patiënten met 'Congenital Contractural Arachnodactyly', een bindweefselafwijking. Dat dit gen ook een rol speelt bij ledemaatvorming was een nieuwe ontdekking, en het laat zien dat *Fibrilline 2* een rol speelt in verschillende processen van de embryogenese.

Het tweede onderwerp van het proefschrift is het genomische gebied waarbinnen het gen *Lmbr1* ligt. In dit gebied werd een aantal mutaties gevonden die verschillende ledemaatafwijkingen veroorzaken. Deze studie begon met een grote Nederlandse familie waarbinnen autosomaal dominante preaxiale-polydactylie (meervingerigheid aan de duimkant) voorkomt. We ontdekten dat deze aandoening gekoppeld is aan chromosoom 7q36. Dit gebied is 'synteen' met (bevat dezelfde genen als) het uiteinde van muis

chromosoom 5, waaraan het polydactylie-fenotype *Hemimelic extra toes* (*Hx*) is gekoppeld. Een andere pootafwijking, *Hammertoe* (*Hm*) is ook hier gelokaliseerd. Deze mutant vertoont syndactylie. En inderdaad werd een humaan syndactylie fenotype, complex polysyndactyly (CPS) gelokaliseerd op humaan 7q36.

Vervolgens werden meer ledemaatafwijkingen in dit gebied gepubliceerd. Een hiervan was acheiropodie, een recessieve afknotting van alle vier de ledematen. Rond die tijd hadden we een aantal genen gevonden in het humane kandidaatgebied, waaronder *LMBR1*. We werkten samen met de auteurs van het acheiropodie-artikel en vonden een deletie van één exon van *LMBR1* in deze patiënten. Dit veroorzaakt een 'frameshift' (verspringing van de drie-letter code) en resulteert in een verkort eiwit. Dit duidt erop dat het LMBR1 eiwit een functie heeft in ledemaatontwikkeling. Echter, noch in de PPD en CPS patiënten, noch in de *Hx* en *Hm* mutanten werden mutaties gevonden in het coderende gedeelte van dit gen.

Vervolgens werd een andere muis-mutant met polydactylie gepubliceerd. De polydactylie lijkt op die van de Hx muis, en is veroorzaakt door een transgen integratie in dit gebied. We startten een samenwerking en vonden het integratiepunt in intron 5 van *Lmbr1*. Daarna werd een patiënt gevonden met een chromosomale translocatie (een proces waarbij chromosoomdelen worden uitgewisseld) met het breukpunt in intron 5 van *LMBR1*. Ook deze patiënt had preaxiale polydactylie. Het lijkt erop dat dit intron een belangrijk regulatie-element bevat. Dit element is misschien niet nodig voor de expressie van *LMBR1* zelf, aangezien dit nauwelijks tot expressie komt in ledematen van muize-embryo's. Het is mogelijk dat het element invloed uitoefent op de expressie van het *Sonic hedgehog* gen, wat 1 Mb (megabase) voor het *LMBR1* gen ligt, en waarvan bekend is dat het een functie heeft in ledemaat vorming. Alle muizen met preaxiale polydactyly die totnogtoe getest zijn, bleken een afwijkende expressie van dit gen te vertonen.

Dit betekent dat de expressie van een gen onder invloed kan staan van een element dat ver van het gen ligt, en dat een functionele gen 'unit' een aantal andere genen kan overlappen. Het laat ook zien hoe prachtig complex het genoom in elkaar zit.

Dit proefschift beschrijft ook hoe computeranalyse helpt bij de ontdekking van nieuwe genen en hun functie. Een enorme hoeveelheid data is beschikbaar gekomen met de genoomprojecten van verschillende organismen, en het vergelijken en organiseren van alle informatie over alle genen van deze organismen zal veel bijdragen aan de kennis en het begrip van de genetica.

List of abbreviations

phenotypes are in italics

A/P	anteroposterior
aa	amino acid
add	*anterior digit pattern deformity*
AER	apical ectodermal ridge
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
Bmp	Bone Morphogenic Proteins
bp	base pair
bph	*brachyphalangy*
(c)DNA	(copy) deoxyribonucleic acid
ci	*cubitus interruptus*
CPS	complex polysyndactyly
D/V	dorsoventral
Dl	*Delta*
DPP	Decapentaplegic
EGF	Epidermal Growth Factor receptor
EMBL	European Molecular Biology Laboratory
en	*engrailed*
En1	*Engrailed 1*
EST	Expressed Sequence Tag
Fgf	*Fibroblast Growth Factor*
fng	*fringe*
GCPS	Greig cephalopolysyndactyly
GPI	glycosylphosphatidylinositol
HGMP	Human Genome Mapping Project Resource Centre
HH	Hedgehog
Hm	*hammertoe*
HPE	holoprosencephaly
Hs2st	Heparan sulphate 2-O-sulfotransferase
Hx	*hemimelic extra toes*
kb	kilobase
ld	*limb deformity*
Mb	megabase
MGI	Mouse Genome Informatics
(m)RNA	(messenger) ribonucleic acid
N	Notch
NCBI	National Center for Biotechnology Information
NO	nitric oxide
NOS	nitric oxide synthase

nt	nucleotide
omb	*optomotor-blind*
MIM	Mendelian Inheritance in Man, online version is OMIM
ORF	open reading frame
PAP	postaxial polydactyly
pdn	*polydactyly Nagoya*
PHS	Pallister-Hall syndrome
PPD	preaxial polydactyly
PTC	Patched
Rfng	*Radical fringe*
Ser	*Serrate*
Shh	*Sonic hedgehog*
sm	*syndactylism*
SMO	Smoothened
SNP	single nucleotide polymorphism
Ssq	*sasquatch*
sy	*Shaker-with-syndactylism*
ttv	*tout-velu*
ty	*tipsy*
wg	*wingless*
xt	*extra toes*
YAC	yeast artificial chromosome
ZPA	zone of polarising activity

Curriculum Vitae

Naam: Marijke Jeltje van Baren

Geboortedatum: 20 april 1973

Geboorteplaats: Amersfoort

1985-1991	VWO opleiding op de Christelijke Scholengemeenschap voor Zuidoost Drente te Emmen.
1991-1996	Biologiestudie aan de Rijksuniversiteit Groningen. Afstudeerrichtingen Klinische- en Celgenetica.
1996-1997	Wetenschappelijk medewerker bij de afdeling Klinische Genetica, Erasmus Universiteit Rotterdam. Onderwerp: 'Identification of the gene responsible for familial frontotemporal dementia'
1997-2001	Onderzoeker in Opleiding bij de afdeling Klinische Genetica, Erasmus Universiteit Rotterdam. Onderwerp: 'Identification and characterization of the gene for preaxial polydactyly'.
Juli 2001	Post-doc in de werkgroep Bioinformatica van de afdeling Klinische Genetica aan de Erasmus Universiteit Rotterdam. Onderwerp: Identificatie van kandidaatgenen voor diverse aandoeningen met behulp van computeranalyse.

Dankwoord

Gestaag doorwerken, dan komt het vanzelf af. Gelukkig heb ik dit werk niet alleen gedaan. Graag wil ik iedereen bedanken voor de steun, suggesties en natuurlijk ook de gezelligheid die mij de afgelopen jaren ten deel zijn gevallen. Een paar mensen wil ik met name noemen.

Allereerst mijn copromotor, Peter Heutink. Peter, het is aan jouw enthousiasme te danken dat ik überhaupt begonnen ben op de afdeling Klinische Genetica, jaren geleden, als stagiaire. Ik ben blij dat ik kon blijven, en heb mede door jouw energie geen moment spijt gehad van die keuze. Zowel binnen als buiten het lab hebben we veel overlegd, ons veel verbaasd en nu en dan ook veel ('macho') bier gedronken. Ik hoop dat je al je plannen ten uitvoer kunt brengen.

Mijn promotor, Ben Oostra, ben ik dankbaar voor zijn steun en ideeën. Beste Ben, bedankt dat je me deze mogelijkheid gaf. Dank ook voor alle adviezen en voor het delen van je grote kennis.

Professor Bob Hill and Laura Lettice, thank you letting me stay in your lab and for helpful discussions. Professor Rolf Zeller and Antonella Galli, thank you for helping me out with the whole mount *in situ*'s. Ik dank de professoren Zeller, Dzierzak en Hovius voor het lezen en corrigeren van het manuscript.

Guido en Marijke, excuses voor alle domme vragen, de mislukte proeven en de dingen die ik veel beter aan jullie had kunnen overlaten. Ik heb mij er inmiddels bij neergelegd dat ik jullie kwaliteit nooit zal bereiken en ben dus nederig iets anders gaan doen... Vele malen bedankt, voor de wijze lessen, het geduld en vooral ook voor de humor.

Herma, je kwam, zag en overwon. In alle rust en op je eigen manier, gelukkig een heel stuk systematischer dan ik (met als direct gevolg dat je werkelijk van alles voor me kon opzoeken). Ik ben blij dat jij mijn paranimf bent.

Mijn andere paranimf, Mark, wil ik bedanken voor alle dingen die binnen en ook buiten dit boekje vallen. Jouw overtuiging dat het allemaal wel goed zou komen en de etentjes na het 'harde werk' hebben, zoals je ziet, zin gehad. En ik ben blij dat we het over sommige dingen nooit eens zullen worden.

Henk, mijn voorganger, en Esther, mijn opvolgster, wil ik bedanken voor hun gezichtspunten, en voor het feit dat ze de dingen op een andere manier aanpakken dan ik

Stellingen behorende bij het proefschrift

Thumbs up
Limb mutants elucidating mechanisms of development.

Preaxiale polydactylie wordt veroorzaakt door afwijkende expressie van het *Sonic hedgehog* gen, hoewel recombinatie-analyse het uitsluit als kandidaatgen.
dit proefschrift

De betekenis van de term 'mutaties in een gen' is aan vernieuwing toe.
dit proefschrift

Systematische naamgeving van genen leidt tot spraakverwarring.
dit proefschrift

Het perfecte kandidaatgen bestaat alleen achteraf.
dit proefschrift

Het bestuderen van pathogene mutaties geeft een te beperkt beeld
van de functie van een gen.
dit proefschrift

"Een gen is een reeks coderende sequenties tussen splice sites" is een te beperkte definitie
voor het voorspellen van het aantal humane genen.
Cell 2001: 106:413-5

LINE elementen *in trans* van actieve genen modelleren het genoom
door pseudogenen te produceren.
Nat. Genet. 2001: 24:363-7

Pathogene 'missense' mutaties treden voornamelijk op
in sterk geconserveerde aminozuren.
Hum. Mol. Gen. 2001: 10:2319-28

De genetische code is een taal waarvan we alleen
sommige zelfstandige naamwoorden kennen.
Nature 2001: 409: 934-41, Science 2001: 291:1304-51

"Vergeet het maar" is een uitvoerbare opdracht.
Nature 2001: 410:366-9

Horizonverbreding dient niet uitsluitend te worden nagestreefd vanaf de bovenste
verdieping van een hoog gebouw.

Marijke Jeltje van Baren, Rotterdam, 6 maart 2002