13

Long-Range Gene Control and Genetic Disease

Dirk A. Kleinjan and Laura A. Lettice
MRC Human Genetics Unit, Western General Hospital,
Edinburgh EH4 2XU, United Kingdom

I. From Genetic Disease to Long-Range Gene Regulation
II. Position Effect Revisited
A. Thalassemias and the α- and β-globin loci
III. Loss of a Positive Regulator
A. Van Buchem disease
B. Leri-Weill dyschondrosteosis
IV. TWIST, POU3F4, PITX2, SOX3, GLI3, and FOXP2
V. The "Bystander" Effect
A. Aniridia and PAX6
VI. MAF, SDC2, TGFB2, REEP3, and PLP1
VII. Two Position Effects—Different Outcomes
A. Sonic hedgehog, holoprosencephaly, and preaxial polydactyly
VIII. Phenotypes Resulting from Position Effects on More than One Gene
A. Split hand foot malformation locus 1
B. Combination of two position effects in SHH and RUNX2
IX. Global Control Regions; HOXD, Gremlin, and Limb Malformations
X. FOX Genes and Position Effects
A. FOX genes in eye anomalies
XI. SOX9 and Campomelic Displasia

Advances in Genetics, Vol. 61
Copyright 2008, Elsevier Inc. All rights reserved.
0065-2660/08 $35.00
DOI: 10.1016/S0065-2660(07)00013-2

XII. Facioscapulohumeral Dystrophy
XIII. Aberrant Creation of an Illegitimate siRNA Target Site
XIV. Genetic Disease Due to Aberrant Gene Transcription Can Be Caused by Many Different Mechanisms
A. The problem: How to find and assess regulatory mutations?
B. Long-range control and genome organization
C. Implications for (common) genetic disease
XV. Concluding Remarks
References

ABSTRACT

The past two decades have seen great progress in the elucidation of the genetic basis of human genetic disease. Many clinical phenotypes have been linked with mutations or deletions in specific causative genes. However, it is often less recognized that in addition to the integrity of the protein-coding sequences, human health critically also depends on the spatially, temporally, and quantitatively correct expression of those genes. Genetic disease can therefore equally be caused by disruption of the regulatory mechanisms that ensure proper gene expression. The term "position effect" is used in those situations where the expression level of a gene is deleteriously affected by an alteration in its chromosomal environment, while maintaining an intact transcription unit. Here, we review recent advances in our understanding of the possible mechanisms of a number of "position effect" disease cases and discuss the findings with respect to current models for genome organization and long-range control of gene expression. © 2008, Elsevier Inc.

I. FROM GENETIC DISEASE TO LONG-RANGE GENE REGULATION

Large advances have been made over the past two decades in determining the genetic basis of human genetic disease. Mutations in an impressive number of genes have been linked with specific clinical phenotypes (Valle, 2004). The completion of the human genome project has undoubtedly been invaluable in speeding up the search for new genes responsible for specific medical disorders. It has, however, also contributed to the recognition that in a small but significant number of patients the molecular lesions do not disrupt the transcribed region of the gene directly, but rather interfere with its transcriptional regulation. In many

instances, individual patient cases provided the first indications that long-range gene regulation might be involved in the control of expression of the genes underlying congenital malformations.

The spatially, temporally, and quantitatively correct activity of a gene requires the presence of not only intact coding sequence but also properly functioning regulatory control. There may be little difference between a coding region mutation that reduces the protein activity by half and a regulatory defect that causes a twofold reduction in expression. Thus, it comes as no surprise that apart from deleterious mutations in the protein encoding part of the gene, genetic disease can equally be caused by disruption of the regulatory mechanisms that ensure its proper expression. Due in part to the generally greater flexibility and robustness in the regulatory control sequences than in protein-coding sequences and in part to ascertainment bias, by far the majority of reported cases of disease-associated mutations alter the protein-coding sequence of the gene in some way. However, a number of different mechanisms can interfere with normal gene function by disrupting the proper regulatory controls of gene expression and lead to pathological states. In this chapter, we catalogue the various known mechanisms with examples of genetic diseases thought to be caused in those ways.

Understanding the mechanisms of gene expression control is one of the great challenges of the post-genome era. It is generally accepted that for most genes, transcriptional regulation is the main level at which control of gene expression takes place, although it is certainly true that in specific cases post-transcriptional events can also play an important role. Transcriptional control itself is influenced at two levels: (1) through modification of the chromatin structure of the gene locus and (2) through the action of *trans*-acting factors on *cis*-regulatory sequences. While often described as separate events, the two are intimately linked, with chromatin structure influencing accessibility of DNA to transcription factors, and the binding of *trans*-acting factors to the DNA involved in triggering chromatin modifications. It will be clear that when either the *cis*-trans regulatory system of a gene or the normal context of its chromatin structure is disrupted, the expression of the gene may be adversely affected, in some cases leading to disease. Conversely, in cases of genetic disease where the disease-causing gene product is known, but no mutations are found in the transcribed portion of the gene, disruption of its normal expression control can be suspected. Such cases have been termed “position effect” cases in reference to the phenomenon of position-effect variegation whereby expression of a gene is variably, but in a clonally heritable manner, inhibited because of juxtaposition with a region of heterochromatin (see Karpen, 1994). In the case of human genetic disease, we define the term “position effect” to refer to situations where the level of expression of a gene is deleteriously affected by an alteration in its chromosomal environment, while maintaining an intact transcription unit. Over the past years, an increasing number of such disease-related position effect

cases have come to light. The study of these cases has been valuable in highlighting the dependence on long-range gene control for the underlying gene, and in some cases has been instrumental in pinpointing the *cis*-regulatory elements.

An important aspect to come out of these studies is the observation that *cis*-regulatory sequences can be located at very large distances from their linked gene, and can even be found in locations beyond or within the introns of neighboring genes. As described below, severe genetic defects can be the result of even single nucleotide substitutions in such a long-range enhancer. Furthermore, it is important to be aware that a translocation or deletion could very well coincidentally disrupt the coding region of one gene as well as remove a distant enhancer from a second gene, and that the resultant alteration of expression of this second gene rather than the physical disruption of the first is the cause of the disease. As will be discussed later, regulatory mutations may disrupt only a spatiotemporally specific subset of the normal expression pattern, and hence the clinical phenotype observed in a patient with a position effect can be different from that due to a complete loss of the underlying gene. Thus, analysis of such patients may uncover novel functions for that gene.

While juxtaposition with heterochromatin is a potential mechanism in a small number of genetic diseases, it has become clear that in the majority of "position effect" cases studied to date the genetic defect appears to be caused by a disruption of the normal *cis*-regulatory architecture of the gene locus. Even so, a number of different mechanisms leading to such *cis*-regulatory disruption have been encountered as will be described below. We review recent advances in our understanding of the possible mechanisms of a number of "position effect" disease cases and try to incorporate the findings into a model for genome organization and long-range control of gene expression.

II. POSITION EFFECT REVISITED

A. Thalassemias and the α- and β-globin loci

Some of the earliest cases where it was recognized that disruption of normal gene regulation can lead to genetic disease stem from research on the blood disorder thalassemia. Thalassemias result from an imbalance in the levels of the α- and β-globin chains that make up the oxygen carrying hemoglobin in our red blood cells. Mostly this imbalance is caused by mutation or deletion of one or more of the globin genes, but, as in the case of the Spanish and Dutch thalassemias, it can also occur through translocations that remove the locus control region (LCR), a major upstream control region for the β-globin locus. The LCR consists of a set of *cis*-regulatory sites (the hypersensitive sites) with the ability to drive high-level expression of the β-globin genes in erythroid

cells (Grosveld *et al.*, 1987, and reviewed in the Palstra, de Laat, and Grosveld chapter in this volume). $\gamma\beta$-Thalassemias caused by translocations in the β-globin locus thus represent one of the earliest recognized position effect cases (Driscoll *et al.*, 1989; Kioussis *et al.*, 1983), and control of β-globin expression continues to be the subject of intense study and many ideas about mechanisms of gene regulation stem from these analyses. Recently, detailed studies of the murine β-globin locus using two novel techniques, 3C (Tolhuis *et al.*, 2002) and RNA-TRAP (Carter *et al.*, 2002), have shown evidence for a mechanism of long-range interaction that involves close contact between the enhancer and the promoter. The more versatile 3C technique has since been applied to a number of other gene loci and most of the published findings provide support for a looping model of enhancer–promoter interactions, whereby distal *cis*-elements and promoters come together in nuclear space with the intervening sequences looping out (Fraser, 2006).

The α-globin gene cluster also depends on long-range control elements for its expression, and deletions removing these elements from the locus are a cause of α-thalassemia (Hatton *et al.*, 1990; Viprakasit *et al.*, 2003). In contrast to the β-globin locus, the α-globin locus resides in a gene-rich region of open chromatin surrounded by a number of housekeeping genes (Vyas *et al.*, 1995). Continued analysis of new patient cases with small deletions outside the α-globin genes has homed in on the most highly conserved element, HS-40, as the dominant α-globin regulatory element in humans (Viprakasit *et al.*, 2006), even though deletion of the equivalent element in the mouse, HS-26, has a remarkably mild phenotype (Anguita *et al.*, 2002).

A novel type of position effect mechanism was revealed in a recently reported case of α-thalassemia (Tufarelli *et al.*, 2003). In this particular α-thalassemia case, an 18-kb deletion, encompassing the HBA1 and HBQ1 genes, was identified in a Polish family. In normal individuals, α-globin is transcribed from four major genes (HBA1 and HBA2 on both copies of chromosome 16) in the α-globin cluster, under the control of four regulatory regions (R1–R4) of which the 5′ HS-40 region (R2) plays the major role. The patients' HBA2 gene and control elements remained completely intact. However, while one copy of the HBA1 gene was indeed deleted in this family, the severity of their phenotype suggested that expression from the intact HBA2 gene on the deleted chromosome could also be affected. The absence of HBA2 expression from the abnormal chromosome was confirmed in an experiment using permissive cell hybrids. Furthermore, a 2-kb region including the HBA2 CpG island was found to be densely methylated in all tissues, while under normal circumstances the α-globin CpG islands always remain unmethylated, even in nonexpressing tissues. Further analysis of this case showed that rather than the initially suspected disruption of a control element, silencing and methylation of HBA2 were strongly correlated with

the presence of antisense RNA transcripts derived from the truncated neighboring LUC7L gene on the opposite strand. In addition to the HBA1 and HBQ genes, the 18-kb deletion had removed the final three exons of the LUC7L gene, including its polyadenylation signal, causing RNA polymerase to read through into the HBA2 promoter and coding region. The antisense-induced silencing and methylation of the promoter was confined to the deleted chromosome, indicating it occurred through a purely *cis*-acting mechanism (Tufarelli *et al.*, 2003).

This case presents a novel and currently unusual mechanism by which a "position effect" can occur, because it does not involve the disruption of normal long-range gene control required for normal expression. However, it can be envisaged that such a juxtaposition of a truncated, (highly) expressed gene, lacking a polyA signal, in close proximity to a disease gene on the opposite strand as a result of a deletion or translocation is a mechanism that could easily cause problems in many other gene loci, in particular in gene-dense regions of the genome.

Detailed study of the α-globin locus in a set of α-thalassemia patients from Melanesia has led to the discovery of yet another novel mechanism of human genetic disease. Analysis of these patients failed to find any of the previously described molecular defects known to cause α-thalassemia, such as mutations or deletions in the globin-coding regions or regulatory elements. The inheritance pattern suggested a codominant defect, and as linkage to the α-globin locus was confirmed it seemed the disease would most likely be due to a gain-of-function mutation that downregulates expression of the α-globin genes. Resequencing of a bacterial artificial chromosome (BAC) containing the globin locus derived from patient DNA identified a large number of single nucleotide polymorphisms (SNPs) compared to the database sequence. This raised the problem of discerning which of these SNPs might be functionally relevant. When the RNA expression profiles of the locus in normal and patient erythroblasts were compared, a major new peak of transcription was identified in the patient sample in the genomic region between the regulatory elements and the globin genes. The peaks of transcription of the globin genes were concomitantly reduced. Of the seven SNPs located underlying the new transcription peak, only one was unique to the patient and involved a T to C change that creates a new GATA factor binding site. GATA-1 is a key regulator of erythroid cell differentiation and binding of GATA-1 as well as other transcription factors, SCL, E2A, LMO2, and Ldb-1, which are often found in association with GATA-1 at erythroid-specific enhancers, was confirmed by chromatin immunoprecipitation. Thus, it appears that this single base pair mutation has created a new promoter, located between the upstream regulatory elements and their cognate promoters, that interferes with the normal regulatory mechanism, thereby causing downregulation of the α-globin genes.

III. LOSS OF A POSITIVE REGULATOR

Arguably the simplest way by which a position effect can occur is a situation where loss of a regulatory element leads to misexpression of the gene. Such situations have been identified for a number of human conditions and form the majority of currently recognized position effect cases. Mostly they entail the loss of one or more positive acting *cis*-elements, but they could also result from loss of a repressor or the gain of an inappropriate regulatory element. The latter mechanisms are known to play a part in certain cancers, but these will not be discussed here. The loss of *cis*-elements can lead to disease in both homozygous (recessive) and heterozygous [(semi-)dominant] states, though because of selection and ascertainment bias most cases are the result of haploinsufficiency of the target gene.

A. Van Buchem disease

The homozygous recessive disorder Van Buchem disease, a severe sclerosing bone dysplasia, has been mapped to chromosome 17p21 (Balemans *et al.*, 2002; Van Hul *et al.*, 1998). It is characterized by a progressive increase in bone density which results in facial distortion, head and mandible enlargement, entrapment of cranial nerves, and a general increase in bone strength and weight (Van Hul *et al.*, 1998). The closely related, but phenotypically more severe bone malformation sclerosteosis (MIM 269500) has also been mapped to the same region. A gene encoding a negative regulator of bone formation, sclerostin (SOST), was found in this region (Balemans *et al.*, 2001; Brunkow *et al.*, 2001) and its expression was shown to be affected in both sclerosteosis and Van Buchem disease. While sclerosteosis patients were shown to carry homozygous null SOST mutations, Van Buchem patients have an intact SOST-coding region but carry a homozygous deletion of 52 kb starting 35 kb downstream of the SOST polyadenylation signal and 10 kb upstream of the neighboring MEOX1 gene (Balemans *et al.*, 2001; Staehling-Hampton *et al.*, 2002). The similarities between sclerosteosis and Van Buchem disease strongly suggested that expression of the SOST gene was affected by the deletion, implicating sequences within the deletion in the regulation of SOST. A BAC carrying the human SOST locus faithfully reproduced the endogenous SOST expression pattern in transgenic mice, but mice carrying a BAC engineered to contain the 52-kb deletion had dramatically reduced levels of human SOST mRNA expression (Loots *et al.*, 2005). Using human/mouse sequence comparisons of the 52-kb region, seven highly conserved elements were identified and tested by luciferase assay in osteosarcoma cells. Only one, ECR5, was able to activate the human SOST promoter and this element was subsequently shown to drive reporter gene

expression in the skeleton of E14.5 transgenic mice. Thus, it is highly likely that the cause of Van Buchem disease is the removal of the 250 bp ECR5 element from the SOST locus (Loots *et al.*, 2005).

B. Leri-Weill dyschondrosteosis

The growth of bones in our body is attained by proliferation and differentiation of chondrocytes in the growth plates of the bones, a process that requires tight genetic (and environmental) control and misregulation can lead to skeletal dysplasias or short stature syndromes. One such syndrome, Leri-Weill dyschondrosteosis (OMIM 127300), is caused by haploinsufficiency of the short stature homeobox gene (SHOX) in about 50–70% of cases. Homozygous loss of SHOX results in the more severe Langer dysplasia (OMIM 249700). Both are characterized by mesomelic, disproportionate short stature and a characteristic curving of the radius (Madelung deformity). The SHOX gene is located on the pseudoautosomal region of the sex chromosomes and has therefore also been implicated in the skeletal deformities of Turner syndrome. While exonic mutations have been found in a large proportion of LWS cases, a number of studies have found patients with an intact SHOX intragenic region but carrying microdeletions in the region downstream from SHOX. These studies highlight a common deleted region of ~40 kb, (Fukami *et al.*, 2005), 30 kb (Benito-Sanz *et al.*, 2005), or as little as 10 kb (Huber *et al.*, 2006). As no other known genes are present in the 750 kb between SHOX and the next gene CSF2RA, a recent study conducted an *in vivo* analysis of putative enhancer activity in this area. Data from four patients with deletions of between 220 and 360 kb were used to define a 200 kb common deletion interval (Sabherwal *et al.*, 2007). Scanning of the region for the presence of multispecies conserved sequences revealed eight evolutionary conserved regions (ECRs) in comparison to dog, chicken, and frog. SHOX like other genes in the Xp22 region is absent in mouse and rat. Enhancer activity of the eight ECRs was therefore tested by electroporation of reporter constructs into the chicken limb bud *in ovo* and analyzed for expression at the appropriate stages of development. Three of the ECRs showed reporter gene expression in the proximal part of the limb, consistent with endogenous chicken SHOX expression. The most distal of those three ECRs lies within the small deletion intervals identified in two of the other LWS position effect studies and thus seems to represent a frequently deleted ECR in patients with LWS and downstream deletions.

IV. TWIST, POU3F4, PITX2, SOX3, GLI3, AND FOXP2

There are now several other cases of human genetic disease where the removal of a positive acting regulatory element is strongly suspected, but for which the exact *cis*-element has not yet been identified and characterized. Examples of these are

found in patients with Saethre-Chotzen syndrome (TWIST), X-linked deafness (POU3F4), Rieger syndrome (PITX2), hypoparathyroidism (SOX3), and Greig cephalopolysyndactyly (GLI3) (see Table 13.1).

Saethre-Chotzen syndrome is a common autosomal dominant form of craniosynostosis, the premature fusion of the sutures of the calvarial bones of the skull. Haploinsufficiency of TWIST, a basic helix-loop-helix transcription factor, is implicated as the pathogenic mechanism by virtue of the identification of mutations and deletions in the coding region (Chun *et al.*, 2002). Using real-time polymerase chain reaction (PCR) to analyze allele dosage by "walking" across the critical region, translocation or inversion breakpoints were found in two patients located at least 260 kb downstream from the TWIST gene (Cai *et al.*, 2003).

The POU3F4 (Brn-4) gene is involved in the pathogenesis of X-linked deafness type 3 (DFN3), the most common form of X-linked inherited deafness. Clinical features include fixation of the stapes and a widening of the internal auditory canal, allowing entry of cerebrospinal fluid into the inner ear. In addition to a spectrum of missense and truncating mutations, a cohort of genomic deletion cases has been studied (de Kok *et al.*, 1995). The observed deletions either remove the POU3F4 gene itself or overlap in a small region ~900 kb upstream of the gene (de Kok *et al.*, 1995, 1996). The smallest of these deletions comprises an 8 kb small fragment containing a 2-kb sequence that is 80% conserved between mouse and human (Cremers and Cremers, 2004). Interestingly, a mouse mutant generated in a random mutagenesis screen with developmental malformations of the inner ear resulting in hearing loss and vestibular dysfunction, the sex-linked fidget (slf) mutant, was shown to be a regulatory mutant of Pou3f4/brn-4 (Phippard *et al.*, 2000). The mutant carries a large X-linked inversion with one breakpoint near but not in the Pou3f4/brn-4 transcription unit, with no gross structural rearrangements being detected within 6–10 kb of the coding region. Expression of Pou3f4/brn-4 was abolished in the embryonic inner ear of the mutant; however, the gene continued to be expressed normally within the neural tube, consistent with a study that mapped two neural tube enhancers within 6 kb upstream of the gene and which would be unaffected by the inversion (Heydemann *et al.*, 2001).

Rieger syndrome (RIEG) type 1 is an autosomal dominant disorder characterized by dental hypoplasia and malformation of the umbilicus and anterior segment of the eye. The main RIEG locus is mapped to chromosome 4q25–27, and mutations in PITX2, a paired-related homeobox gene with multiple isoforms, have been shown to cause Rieger syndrome (Alward *et al.*, 1998). In addition to deletions and mutations in the gene itself, translocation breakpoints 15 and 90 kb upstream of PITX2 have been identified in three separate patients (Flomen *et al.*, 1998; Trembath *et al.*, 2004).

Hypoparathyroidism (HPT) is an endocrine disorder due to deficiency of parathyroid hormone (PTH) leading to hypocalcemia and hyperphosphatemia. It can exist as part of larger syndromes or as an isolated endocrinopathy for

Table 13.1. Position Effect Genes in Human Genetic Disease

Gene	Gene function	Domains/Motifs	Disease	OMIM	Distance of furthest breakpoint	3′ or 5′ side	References
PAX6	TF	Paired box and homeodomain	Aniridia	106210	125 kb	3′	Fantes *et al.*, 1995; Kleinjan *et al.*, 2001
TWIST	TF		Saethre-Chotzen syndrome	101400	260 kb	3′	Cai *et al.*, 2003
POU3F4	TF	POU homeodomain	X-linked deafness	304400	900 kb	5′	de Kok *et al.*, 1996
PITX2	TF	Homeodomain	Rieger syndrome	180500	90 kb	5′	Trembath *et al.*, 2004
GLI3	TF	Zinc finger	Greig cephalopolysyndactyly	175700	10 kb	3′	Wild *et al.*, 1997
MAF	TF	bZIP	Cataract, ocular anterior segment dysgenesis, and coloboma	610202	1000 kb	5′	Jamieson *et al.*, 2002
FOXC1	TF	Forkhead	Glaucoma/autosomal-dominant iridogoniodysgenesis	601631	25 kb	5′	Davies *et al.*, 1999
FOXC2	TF	Forkhead	Lymphedema-distichiasis syndrome	153400	120 kb	3′	Fang *et al.*, 2000
FOXL2	TF	Forkhead	BPES (blepharophimosis-ptosis-epicanthus inversus syndrome)	110100	170 kb	5′	Crisponi *et al.*, 2004; Beysen *et al.*, 2005
FOXP2	TF	Forkhead	Speech-language disorder	602081	>680 kb	5′	Scherer *et al.*, 2003

(*Continues*)

Table 13.1. (*Continued*)

Gene	Gene function	Domains/Motifs	Disease	OMIM	Distance of furthest breakpoint	3′ or 5′ side	References
TGFβ2	Signaling		Peters anomaly		500 kb	3′	David *et al.*, 2003
SOX9	TF	HMG box	Campomelic dysplasia	114290	932 kb	5′	Leipoldt *et al.*, 2007;
					1300 kb	3′	Velagaleti *et al.*, 2005
SIX3	TF	Homeodomain	Holoprosencephaly (HPE2)	157170	<200 kb	5′	Wallis *et al.*, 1999
SHH	Signaling		Holoprosencephaly (HPE3)	142945	265 kb	5′	Roessler *et al.*, 1997; Fernandez *et al.*, 2005
SHH	Signaling		Preaxial polydactyly	174500	1000 kb	5′	Lettice *et al.*, 2003
RUNX2	TF		Cleidocranial dysplasia	119600	700 kb	5′	Fernandez *et al.*, 2005
SHFM1	TF	DLX5/DLX6?	Split hand split foot malformation	183600	~450 kb	5′/3′	Crackower *et al.*, 1996
ALX4	TF	Homeobox	Potocki-Schaffer syndrome	601224	>15 kb	3′	Wakui *et al.*, 2005
REEP3	TH receptor regulator		Autism		43 kb		Castermans *et al.*, 2007
PLP1	Proteolipid protein		Spastic paraplegia type 2 with axonal neuropathy	312920	136 kb	3′	Lee *et al.*, 2006;

(*Continues*)

Table 13.1. (*Continued*)

Gene	Gene function	Domains/Motifs	Disease	OMIM	Distance of furthest breakpoint	3′ or 5′ side	References
			Pelizaeus-Merzbacher disease	312080	70 kb	5′	Muncke *et al.*, 2004
FSHD	Unknown		Facioscapulo humeral dystrophy	158900	Unknown	3′	Gabellini *et al.*, 2002 Jiang *et al.*, 2003 Masny *et al.*, 2004
HBB	Oxygen carrier	Globin	gamma;β-Thalassemia	141900	50 kb	5′	Kioussis *et al.*, 1983 Driscoll *et al.*, 1989
HBA	Oxygen carrier	Globin	α-Thalassemia	141800	30 kb	5′	Viprakasit *et al.*, 2003
				141850	18 kb	3′	Tufarelli *et al.*, 2003
HOX D complex	TF	Homeodomain	Mesomelic dysplasia and vertebral defects		56 kb	3′ 950 kb	Spitz *et al.*, 2002 Dlugaszewska *et al.*, 2006
RET	Receptor Tyr kinase		Increased risk of Hirsch-sprung disease	142623	Intronic enhancer	Intronic	Emison *et al.*, 2005
SHOX	TF		Langer mesomelic dysplasia Leri-Weill dyschondrosteosis	249700 127300	Micro-deletion 200 kb	3′	Sabherwal *et al.*, 2007; Fukami *et al.*, 2005
SOST	Sclerostin	BMP antagonists	Van Buchem disease	239100	35 kb		Loots *et al.*, 2005
SALL1	TF		Townes-Brocks syndrome	107480	>180 kb	5′	Marlin *et al.*, 1999

(*Continues*)

Table 13.1. (*Continued*)

Gene	Gene function	Domains/Motifs	Disease	OMIM	Distance of furthest breakpoint	3′ or 5′ side	References
SOX3	TF	HMG box	X-linked recessive hypoparathyroidism	307700	67 kb	3′	Bowl *et al.*, 2005
LCT	Enzyme	Lactase	Lactase persistence	223100	15 kb	5′	Enattah *et al.*, 2002; Tishkoff *et al.*, 2007

In case of 3′ breakpoints the distance refers to the distance from the breakpoint to the 3′ end of the gene or complex.

which there are several etiologies including autosomal dominant and recessive forms as well as an X-linked recessive form (OMIM 307700), mapped to a 900-kb region on Xq27. A recent study of a multigeneration family with recessive X-linked HPT revealed a deletion of a 25-kb fragment located around 67 kb downstream from the SOX3 gene, which was replaced by a 340-kb insertion derived from chromosome 2p25. Although the insertion contains a portion of a gene, SNTG2, the inserted fragment lacks an open reading frame and trisomy of the segment is unlikely to be the cause of the disease as patients with 2p trisomy do not suffer from HPT. Instead, it is thought that the 25 kb deleted fragment contains *cis*-regulatory sequences for SOX3 or that *cis*-elements beyond the deletion fragment have been moved beyond the reach of the promoter by the 340-kb insertion (Bowl *et al.*, 2005).

The zinc finger gene GLI3 has a role in the embryonic development of limbs and the skull, and is an important effector of the hedgehog signaling network. Greig cephalo-polysyndactyly syndrome (GCPS), characterized by PPD, syndactyly, and craniofacial abnormalities, is caused by haploinsufficiency of GLI3 on chromosome 7p13 (Wild *et al.*, 1997). All identified mutations leading to GCPS are thought to be loss of function while truncating mutations within Gli3 lead to Pallister Hall syndrome, which is characterized by postaxial/central polydactyly, hypothalamic hamartoma, and bifid epiglottis. A probable position effect translocation has been described in a GCPS patient with a breakpoint 10 kb downstream from the last exon of GLI3 (Vortkamp *et al.*, 1991). In the mouse, two dominant mutant alleles, the extra toes (Xt and Xt^J) phenotypes, are caused by mutations within the murine Gli3 gene (Vortkamp *et al.*, 1992). However, the weak recessive Xt allele, anterior digit deformity (add), is caused by a transgene integration combined with the deletion of an 80-kb region at ~40 kb upstream of Gli3 (van der Hoeven *et al.*, 1993). Another mouse mutant, the polydactyly/arhinencephaly mouse (Polydactyly Nagoya-Pdn/Pdn), shows phenotypic similarity to GCPS. It is caused by insertion of a transposon into intron 3 of the Gli3 gene, resulting in suppression of Gli3 expression (Ueta *et al.*, 2002).

Two translocation cases with a diagnosis of autism have been implicated to be due to position effects. In the first case, the translocation breakpoint mapped near the *FOXP2* gene, which has previously been shown to cause a form of speech and language disorder, suggesting that *FOXP2* might also be involved in autism (Scherer *et al.*, 2003). The child inherited the translocation from the mother, who had speech delay. Seven isoforms of *FOXP2* (spanning 545 kb) were characterized, but all mapped at least 680 kb 3′ of the translocation breakpoint in a gene-poor region, raising the possibility of a position-effect mutation. In the other case, the breakpoint was nearest to the neuronal pentraxin 2 (*NPTX2)* gene, which is thought to be involved in excitatory synaptogenesis and could therefore be considered a functional candidate for autism (Scherer *et al.*, 2003).

V. THE "BYSTANDER" EFFECT

The situation is slightly more complicated in a second group of position effect cases, where the gene disrupted by the chromosomal rearrangements is not directly involved in causing the disease, but rather appears as a bystander, implicated by virtue of harboring regulatory elements for an adjacent gene located some distance away.

A. Aniridia and PAX6

Among the first human genetic diseases where a disruption of the transcriptional control of the gene was recognized as a cause of disease is the congenital eye malformation aniridia. Aniridia (OMIM 106210) is characterized by severe hypoplasia of the iris, usually accompanied by foveal hypoplasia, cataracts, and corneal opacification (Fig. 13.1). It is caused by haploinsufficiency of the PAX6 gene at chromosome 11p13 as shown by deletion cases (Ton *et al.*, 1991), and through loss-of-function point mutations (van Heyningen and Williamson, 2002). In the mouse, heterozygosity for a number of mutation and deletion alleles of Pax6 is the cause of the small eye (Sey) phenotype. A number of aniridia patients with translocation breakpoints mapping downstream of the PAX6 gene have been described (Crolla and van Heyningen, 2002; Fantes *et al.*, 1995; Lauderdale *et al.*, 2000). Detailed mapping of these breakpoints placed them at various positions downstream from an intact PAX6-coding region, with the furthest located 125 kb beyond the final exon (Fantes *et al.*, 1995). Analysis of the sequence downstream of PAX6 revealed the presence of a ubiquitously expressed gene, ELP4 (a subunit of Elongator, a protein complex associated with the elongating RNA polymerase II). Even though all breakpoints map within the final intron of this gene, presumably interfering with its activity, it was shown through transgenic yeast artificial chromosome (YAC) rescue of a Sey deletion mutant that heterozygosity of Elp4 was unlikely to be the cause of the eye phenotype (Kleinjan *et al.*, 2002). A naturally occurring Pax6 mutant carrying a truncating mutation (Hill *et al.*, 1991), the Sey^{Ed} mutant, was employed to further study the PAX6 position effect. A 420-kb human *PAX6* YAC-containing sequence extending a further 80 kb beyond the position of the most distal patient breakpoint was shown to rescue the Sey lethality and give full phenotypic correction in both heterozygous and homozygous *Sey* mice (Schedl *et al.*, 1996). In contrast, a shorter YAC extending only as far as the breakpoint failed to rescue or correct the mutant phenotype (Kleinjan *et al.*, 2001). Essential regulatory elements were identified in the sequence between the ends of the shorter and longer YACs using DNaseI hypersensitive site mapping and evolutionary sequence comparison. Reporter studies in transgenic mice revealed that some of these elements direct eye- and brain-specific expression, providing strong

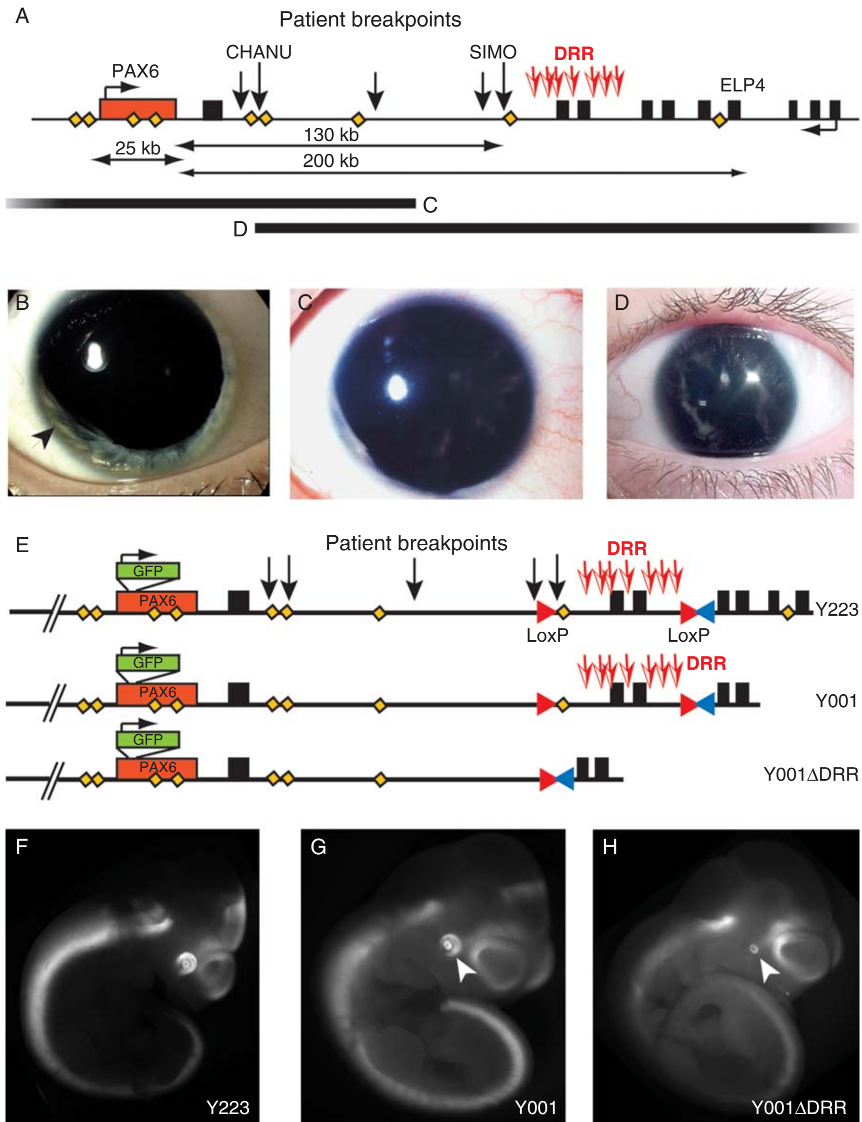

Figure 13.1. The PAX6 position effect. Chromosomal rearrangements with a breakpoint in the region downstream of PAX6 give rise to the congenital eye malformation aniridia (absence of iris). The eye phenotype of the position effect patients is indistinguishable from aniridia due to deletion or mutation of the PAX6 coding region itself. (A) Schematic representation of the PAX6 locus on chromosome 11p13. The 13 exon PAX6 gene is shown as a single box. The exons of the adjacent ubiquitously expressed ELP4 gene are shown as black boxes. Downward solid arrows show the

evidence that the PAX6 position effect is due to the removal of these elements, collectively termed the downstream regulatory region (DRR), from the PAX6 locus (Kleinjan *et al.*, 2001). To obtain further evidence that removal of the DRR would lead to aniridia, YAC transgenic mice were engineered in which a green fluorescent protein (GFP) reporter gene was inserted into the PAX6-coding region and LoxP sites were placed around a 30-kb region containing the DRR. Strong GFP signal was observed in a proper PAX6 expression pattern in full-length YAC transgenics, but after Cre mediated deletion of the DRR, expression was no longer found in the neuroretina, iris, and ciliary body (Fig. 13.1). As multiple enhancers for these expression sites have been identified, not only within the DRR but also in upstream and intronic regions of the PAX6 locus, this study not only demonstrates the essential nature of enhancers in the DRR but also highlights the interdependence of multiple *cis*-elements spread throughout the gene locus.

VI. MAF, SDC2, TGFB2, REEP3, AND PLP1

Awareness of the possibility that an adjacent gene rather than the directly disrupted gene could be the true cause of the disease was helpful in the case of a family with lens and ocular anterior segment anomalies with a t(5;16) translocation (Jamieson *et al.*, 2002). Breakpoint analysis indicated that the translocation occurred in an intron of the WWOX gene (Paige *et al.*, 2000). However, since WWOX is a widely expressed putative tumor suppressor gene and the phenotype was observed in both balanced and unbalanced forms of the translocation, WWOX seemed an unlikely candidate to cause an eye phenotype. Based on its expression pattern and known involvement in eye development, a more likely candidate was the bZIP transcription factor gene MAF, located ~1 Mb

position of a number of patient breakpoints. Grey/open arrows indicate a region beyond the furthest patient breakpoint termed the downstream regulatory region (DRR), containing a number of hypersensitive sites, some of which have been shown to have enhancer activity in transgenic mice. Diamond shapes indicate positions of several tissue-specific enhancers characterized in transgenic mice. (B) Eye of an aniridia patient with a point mutation in PAX6 (G36R). Arrow indicates a bit of residual iris tissue (C) Eye of a patient carrying a large deletion encompassing PAX6. (D) Eye of a patient (CHANU) with a PAX6 position effect due to an inversion of a large chromosome 11 segment (E) Reporter YAC transgenic experiment. GFP was inserted into the PAX6-coding region and LoxP sites were introduced surrounding the DRR. The full-length YAC largely reproduces the endogenous PAX6 expression pattern (F), while a slightly shorter YAC insertion lacks part of the expression in diencephalon. (G) Consistent with the position effect cases, Cre-mediated deletion of the DRR from this line results in loss of expression in the retina (white arrowheads) as well as iris and ciliary body at later embryonic stages.

telomeric of the breakpoint. Subsequent identification of a missense mutation within MAF in another family with lens and iris anomalies confirmed its role in lens development (Jamieson *et al.*, 2002). The identification of affected individuals carrying both balanced and unbalanced forms of the t(5,16) translocation suggests the phenotype may be caused by the dominant misregulation of MAF due to sequences on the der(5) chromosome.

A similar situation occurred with the breakpoint mapping of an X;8 reciprocal translocation in a patient manifesting multiple exostoses and autism with mental retardation and epilepsy (Ishikawa-Brush *et al.*, 1997). The X-chromosomal breakpoint was shown to locate within the first intron of the gastrin-releasing peptide receptor (GRPR), while the breakpoint on chromosome 8 occurred near the syndecan-2 (SDC2) gene, located 30 kb from the breakpoint, but left intact by the translocation. Gastrin-releasing peptide (GRP) has various neurobiological activities in the brain, lungs, and gastrointestinal tract, and while it appears to escape X-inactivation, a lower dosage of GRPR could well be a cause of autism. It is, however, unlikely to be a cause for exostoses. Awareness that breakpoints near but not directly disrupting the coding region should also be considered led to closer examination of the chromosome 8 breakpoint region and the SDC2 gene. SDC2 is a member of a family of cell surface heparan sulphate proteoglycans with a role in bone formation, and a position effect on SDC2 is therefore the most likely reason for the exostosis phenotype (Ishikawa-Brush *et al.*, 1997).

A third example is found in a case of Peters anomaly, which affects the anterior chamber of the eye. The translocation between chromosomes 1 and 7 disrupts the HDAC9 gene at 7p21.1, while the breakpoint on chromosome 1 is located 500 kb from TGFβ2. Nevertheless, since the knockout of TGFβ2 in the mouse has a very similar eye phenotype to Peters anomaly, TGFβ2 must be considered to be the causal gene (David *et al.*, 2003).

A further example of such a situation occurs in a patient with autism and a *de novo* balanced paracentric inversion 46,XY,inv(10)(q11.1;q21.3). The distal breakpoint directly disrupts the TRIP8 gene, a transcriptional regulator associated with nuclear thyroid hormone receptors, but no link between thyroid gland and autism has ever been reported. A nearby gene, receptor expression-enhancing protein (REEP)3, located 43 kb distal to the breakpoint with a probable role in regulating cellular vesicle trafficking between the ER and Golgi, would seem a much more likely candidate. Using an SNP in the 3′UTR of this gene expression analysis in patient- and control-derived cell lines indicated a normal biallelic expression in the controls, but a distinctly monoallelic expression of REEP3 in the patient cell line (Castermans *et al.*, 2007).

Finally, position effects have been found in two clinically distinct allelic disorders Pelizaeus-Merzbacher disease (PMD) and spastic paraplegia type 2 (SPG2), both caused by mutations in the proteolipid protein gene (PLP1),

which codes for the major myelin component in the central nervous system (CNS). PMD (OMIM 312080) is an X-linked disorder of varying severity characterized by dysmyelination of the CNS. SPG2 (OMIM 312920) is generally less severe and presents as progressive weakness and spasticity of the lower extremities. PMD-causing mutations fall into two categories, duplications of PLP1, which account for the majority of cases, and sequence variations within the gene. A male PMD patient carrying an inversion on the X-chromosome (inv (X) (p22.3; q22) displayed a subset of PMD symptoms. His Xq22 breakpoint was mapped about 70 kb upstream from the PLP1 gene, within GLRA4, a putative glycine receptor family pseudogene. Because of the established link between PMD and PLP1 and failure to identify intragenic mutations in GLRA4 in a cohort of comparable patients, a position effect on PLP1 is considered the most likely cause of the disease. The PMD phenotype of the patient most closely resembles that of patients carrying PLP1 duplications. As loss of PLP1 is expected to give a different phenotype, the most likely result of this position effect is the upregulation of PLP1 (Muncke *et al.*, 2004). In a separate study, semiquantitative PCR was used to detect a 150-kb duplication located 136 kb downstream from PLP1 in a patient with SPG2, at the milder end of the PMD/SPG2 spectrum. In this case, the duplication is therefore thought to silence PLP1 repression (Lee *et al.*, 2006).

VII. TWO POSITION EFFECTS—DIFFERENT OUTCOMES

A. Sonic hedgehog, holoprosencephaly, and preaxial polydactyly

In the case of the Sonic hedgehog (SHH) gene on chromosome 7q36, transcriptional misregulation through disruption of long-range control of the same gene, but via different mechanisms, is the cause of two very different human genetic disorders (Fig. 13.2). The first, holoprosencephaly type 3 (HPE3; OMIM 142945), is caused by deletions and point mutations in the SHH gene itself, but can also be caused by translocations up to 265 kb upstream of the gene (Belloni *et al.*, 1996; Roessler *et al.*, 1996, 1997). Phenotypic expressivity is variable, ranging from a single cerebral ventricle and cyclopia to clinically unaffected carriers in some familial HPE cases. In humans, HPE3 is caused by haploinsufficiency due to the loss of one allele of SHH, whereas in the mouse both alleles need to be inactivated to produce a similar phenotype (Chiang *et al.*, 1996), indicating a more critical role for correct SHH dosage in humans. The clinical variability is especially prominent in two familial position effect cases with translocation breakpoints at 235 and 265 kb upstream of SHH, where some family members are phenotypically normal or mildly affected, while others are much more severely affected (Roessler *et al.*, 1997). It is thought that the removal

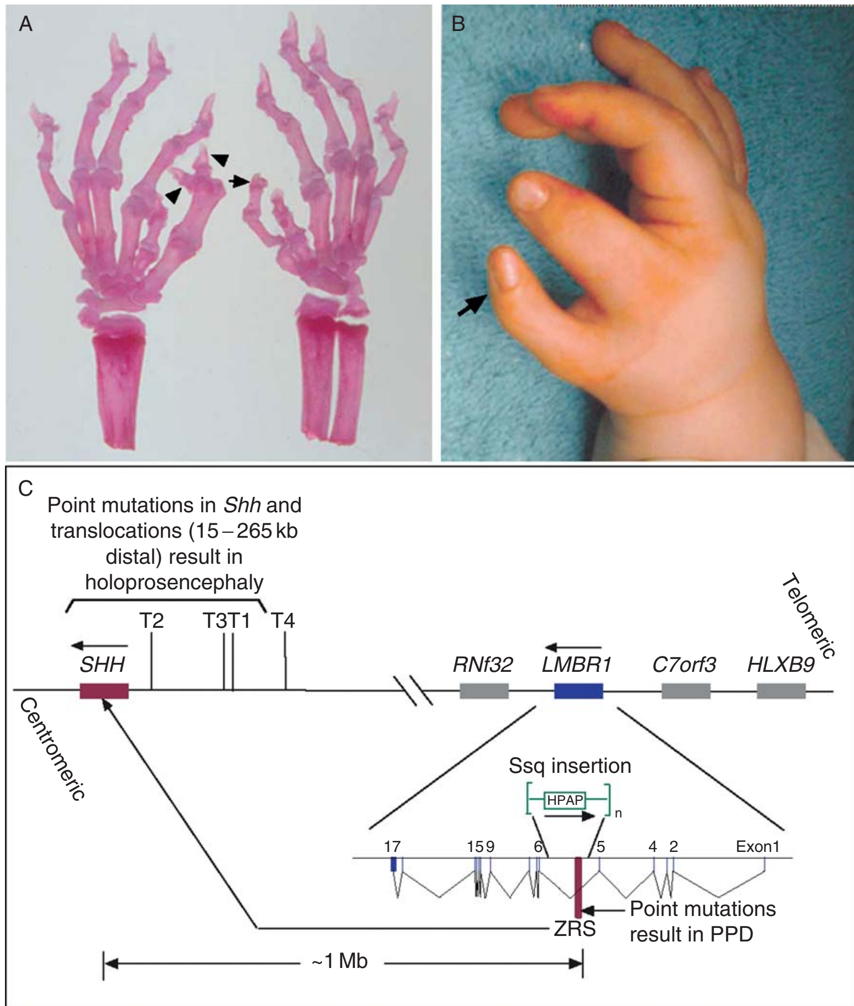

Figure 13.2. Limb-specific malformation by a long-range position effect on the Sonic Hedgehog (SHH) gene a chromosome 7q36. (A) Alizarin red stained bones from the fore limbs of an adult *Ssq* homozygous mouse. Left paw has two sets of terminal phalanges (arrowheads) from a single metacarpal, while the right paw exhibits complete duplication of all the bones of digit one (arrow). (B) Photograph of the hand of a patient with preaxial polydactyly (PPD) showing duplication of the thumb (arrow). (C) Schematic depiction of the genomic organization of the mouse *Shh* locus from *Hlxb9* to *Shh* itself. Positions of the translocations which result in holoprosencephaly (Belloni *et al.*, 1996; Roessler *et al.*, 1996, 1997) are indicated (T1-T4). Also marked is the position of the zone of polarizing activity regulatory sequence (ZRS), located 1 Mb upstream of SHH, beyond RNF32 and within an intron of the LMBR1 gene.

of a *cis*-regulatory element located more than 265 kb upstream leads to a reduced SHH expression level that fluctuates around the critical level. A further translocation, located 315 kb upstream, has not produced a detectable HPE phenotype in the seven individuals examined, thus limiting the region in which to look for the putative control element(s) (Roessler *et al.*, 1997). A functional screen, conducted in mouse (Jeong *et al.*, 2006), has identified a number of Shh enhancers which drive expression in the developing forebrain. Some lie at distances of over 400 kb from the promoter and would predictably be disrupted by translocations such as these.

The second disorder linked to misregulation of SHH expression is the limb developmental malformation preaxial polydactyly (PPD OMIM 174500). Much progress has been made in recent years on long-range gene regulation of SHH in the limb through the analysis of a mouse limb mutant, the polydactylous *sasquatch* mouse (*Ssq*) (Sharpe *et al.*, 1999). The *Ssq* mutation arose as a result of a random insertion of a reporter cassette during transgenic studies analyzing HoxB1 regulatory sequences. Mice carrying the *Ssq* insertion display PPD, with homozygotes being more affected than heterozygotes. Genetic analysis of the mutant showed that the reporter transgene segregates with the limb phenotype and that the insertion site was physically linked to the *Shh* gene but situated almost 1 Mb away. The effect could not be attributed to the reporter cassette itself as none of the other transgenic lines with the same construct showed a similar phenotype (Sharpe *et al.*, 1999). The linkage to *Shh* was interesting because the nature of the *Ssq* phenotype was consistent with the limb phenotypes of a group of mouse mutants called the hemimelia-luxate group which all exhibit misexpression of *Shh* during limb development (Hill *et al.*, 2003). Analysis of *Shh* expression in the limb buds of mutant embryos indeed showed an additional region of ectopic expression in the anterior of the limb bud. During normal limb development, *Shh* is believed to act as a morphogen, emanating from a region (called the zone of polarizing activity, ZPA) at the posterior end of the limb bud to set up a concentration gradient across the anterior/posterior axis of the limb, with high levels of SHH resulting in digits with a posterior identity and lower levels more anterior digits. Ectopic anterior expression will clearly disrupt such gradient, providing an explanation for the observed phenotype. Interestingly, when the insertion site was cloned it proved to lie 1 Mb upstream of Shh, beyond the adjacent testis- and ovary-specific Rnf32 gene, within intron 5 of a gene identified as *Lmbr1* (Limb region 1) (Lettice *et al.*, 2002). The human LMBR1 gene was originally identified as a candidate gene within the critical region for PPD,

The ZRS drives *Shh* expression within the limb and point mutations in this *Shh* limb-specific regulatory element are associated with human PPD and with extra digits in mouse polydactyly models.

one of the most common human congenital limb malformations (Heus *et al.*, 1999). Analysis of a *de novo* chromosomal translocation in a PPD patient also indicated a breakpoint in intron 5 of LMBR1, close to the *Ssq* insertion site (Lettice *et al.*, 2002). Furthermore, *Lmbr1* was shown to be located in the critical region for the closely linked mouse limb mutants, the polydactylous *Hemimelia extra-toes* (*Hx*) and the syndactylous *Hammertoe* (*Hm*) (Clark *et al.*, 2000), and its expression was found to be downregulated in Hx embryos at the time when the phenotype is first observed (Clark *et al.*, 2000). However, examination of the LMBR1 and Lmbr1 structural genes in the various patients and mouse mutants failed to uncover any coding region mutations. The observed *Shh* misexpression and known importance of *Shh* in limb development suggested that the *Ssq* phenotype could be due to disruption of long-range regulation of *Shh* and that the linkage with *Lmbr1* is merely coincidental. A genetic *cis*-trans experiment was set up to generate by recombination a chromosome in which the *Ssq* insertion was located in *cis* to an *Shh* null allele. Analysis of mice carrying the recombinant chromosome showed no abnormal limb phenotype (as is the case for the heterozygous *Shh* null by itself), indicating that *Ssq* is a dominant mutation that interferes with limb-specific expression of *Shh* (Lettice *et al.*, 2002). A similar *cis*-trans test was subsequently done for the *Hx* mutant with the same result (Sagai *et al.*, 2004). Using phylogenetic sequence comparisons, a small, highly conserved region, termed the ZRS (for Zone of polarising activity Regulatory Sequence), was identified in LMBR1 intron 5 that was shown to act as a limb-specific regulatory element in transgenic reporter mice (Lettice *et al.*, 2003). To assess whether mutations in the ZRS could account for the PPD phenotype in families displaying a normal karyotype, the element was sequenced in a large number of affected and unaffected family members as well as controls. Amazingly single-point mutations in the ZRS were found in four PPD families, which were observed in all affected and none of the unaffected individuals. Analysis of the ZRS in the Hx mouse also identified a single base pair change segregating with the phenotype (Lettice *et al.*, 2003). Subsequent analyses in a number of ethylnitrosourea (ENU)-induced mouse mutants (Sagai *et al.*, 2004, Masuya *et al.*, 2007) and in other PPD families have identified more mutations (Gurnett *et al.*, 2007). The ZRS is thought to have a dual function in the regulation of SHH, (1) driving the initiation of expression in the limb bud and (2) restricting this expression to the posterior margin. Expression studies using ZRS sequences to drive reporter expression (Maas and Fallon, 2005; Masuya *et al.*, 2007) demonstrate that the mutations affect the latter but not the former activity of the ZRS, and lead to ectopic SHH expression in the anterior. All of the mutations lie in different parts of the ZRS, and are scattered throughout its length. The lack of clustering of the point mutations suggests that the phenotype cannot simply arise from the inactivation of a single transcription factor binding site. In an interesting aside, it was shown that the ZRS enhancer is

conserved in species with limbs, wings, and fins (Lettice *et al.*, 2003), but no conservation is seen in several limbless species, such as snakes and a limbless newt (Sagai *et al.*, 2004).

Several more congenital limb abnormalities map to the same 7q36 region. Patients for one of these, acheiropodia, an autosomal recessive disease in which all bones of the hand and feet are missing and the tibia are truncated, were shown to carry a small deletion on both alleles of LMBR1 (Ianakiev *et al.*, 2001). The deletion, which includes LMBR1 exon 4 and about 6 kb of surrounding sequence, is slightly more distal than the site of the ZRS, but could similarly be hypothesized to disrupt long-range regulation of SHH. In that case however, rather than result in ectopic misexpression, the deletion would be predicted to disrupt a positive regulatory activity driving the normal ZPA-specific expression of SHH. In accordance, the phenotype observed in the acheiropodia families very closely resembles the limb truncations observed in the *Shh* loss-of-function mice (Chiang *et al.*, 1996). A similar phenotype has been identified inthe chick mutant *oligozeugodactyly* (*ozd*) (Ros *et al.*, 2003) also hypothesized to result from a defect in a regulatory element that controls limb-specific expression of *Shh*.

Two further mouse mutations, both carrying radiation-induced large-scale inversions of chromosome 5, implicate the long-range disregulation of *Shh* in a wide range of developmental defects. Replicated anterior zeugopod (raz) (Krebs *et al.*, 2003) mice exhibit two anterior skeletal elements in the zeugopod and symmetrical central polydactyly. It results from a huge inversion of much of proximal chromosome 5, covering both *Shh* and ZRS, which when homozygous downregulates expression of Shh mRNA and protein in the limb bud to about 20% of wild-type levels (expression in other tissues is unaffected). *Short-digits* (*Dsh*) mice (Niedermaier *et al.*, 2005) carry a much smaller deletion of 11.7 Mb with one breakpoint lying just over 13 kb upstream of *Shh* (between it and a number of putative enhancers including ZRS). Homozygous mice exhibit a number of defects and are phenotypically nearly identical to the *Shh* loss-of-function mice; the two mutations also fail to complement each other, demonstrating that *Dsh* is a regulatory mutation of *Shh* resulting in almost complete downregulation of Shh during E9.5–12.5. Heterozygous mice, however, show a phenotype similar to human brachydactyly type 1A with a fusion and shortening of the proximal and middle phalanges in all digits. Analysis shows that *Shh* is ectopically expressed in a skeletogenic domain typically occupied by a closely related gene, Indian hedgehog (*Ihh*), within the developing digits at E13.5–14.5. *Ihh* is normally downregulated in regions that will become the joint space, but in *Dsh*/ + mice, *Shh* bypasses this regulatory control and persists; accordingly, cells maintain their chondrogenic fate and the developed digits are shorter than normal. This suggests that perhaps the role of one of the endogenous *Shh* regulatory elements is to act as repressor segregating the activity of *Shh* from

Ihh (de la Fuente and Helms, 2005). Direct misregulation of *Ihh* itself has been implicated in the Doublefoot mutant that shows PPD and craniofacial abnormalities (Hayes *et al.*, 2001; Yang *et al.*, 1998).

VIII. PHENOTYPES RESULTING FROM POSITION EFFECTS ON MORE THAN ONE GENE

In a small but growing number of cases the observed phenotypes cannot easily be explained by hypothesizing the involvement of a single gene, but are more likely to be the result of a combination of effects on two or more genes, either neighboring genes under common control or artificially juxtaposed by a translocation.

A. Split hand foot malformation locus 1

Split hand foot malformation type I (SHFM1) (OMIM 183600), a limb developmental disorder characterized by missing digits, fusion of remaining digits, and a deep median cleft in the hands and feet, is a genetically heterogeneous human disorder also referred to as ectrodactyly or lobster claw deformity. The SHFM1 critical region has been mapped on the basis of chromosomal rearrangements to ~1.5-Mb interval on 7q21.3 (Crackower *et al.*, 1996; Ignatius *et al.*, 1996; Scherer *et al.*, 1994). Despite a decade-long search, no disease gene has so far been identified. Three candidate genes, DSS1 and the distalless homeobox genes DLX5 and DLX6, are located in this region. Chromosomal breakpoints from 12 unrelated ectrodactyly patients were found to be scattered throughout the critical region and do not collectively interrupt any single gene. No missense mutations within the genes have been identified in patients. In mouse, inactivation of either Dlx5 or Dlx6 alone does not produce a limb phenotype (Robledo *et al.*, 2002). However, the targeted double inactivation of Dlx5 and Dlx6, by a targeted 17-kb deletion covering both coding regions and the intervening DNA, causes bilateral ectrodactyly with a severe defect of the central ray of the hindlimbs, a malformation typical of SHFM1. Dss1 continues to be normally expressed in these double knockout mice. Even though there are differences in the two phenotypes—the human condition is believed to be inherited in an autosomal dominant manner, while affected mice must be homozygous for the loss-of-function mutation and in man both hands and feet are affected, while the phenotype is observed only on murine hindlimbs—the overall similarities in phenotype suggest the possibility that SHFM1 is due to a position effect on the two DLX genes, and possibly DSS1 (Merlo *et al.*, 2002). The simplest mechanism would involve the chromosomal breakpoints located up to 1 Mb centromeric of human *DLX5* and *DLX6* separating critical shared regulatory elements from the genes and lead to their misregulation during development.

B. Combination of two position effects in SHH and RUNX2

Interestingly, there is one report in the literature of a translocation resulting in two proposed position effects on two different genes which both contribute to the observed phenotype. The patient carries a *de novo* reciprocal translocation with breakpoints at 6p21.1 and 7q36 and displays premaxillary agenesis, skeletal abnormalities, and impacted teeth (Fernandez *et al.*, 2005). The 7q36 breakpoint has been mapped 15 kb telomeric to the 5′ of the SHH gene (Belloni *et al.*, 1996) and is thought to contribute to the premaxillary agenesis (part of the HPE spectrum). However, fine mapping of the chromosome 6 breakpoint positions it further than 700 kb of the 5′ end of the osteoblast-specific transcription factor CBFA/RUNX2 (with one gene, SUPT3H, a probable transcriptional activator, lying in the interval). Mutations in CBFA/RUNX2 result in cleidocranial dysplasia, a dominant disorder characterized by dental, pelvic, and clavicular disorders; some of which show distinct similarities to the phenotype displayed by this patient.

IX. GLOBAL CONTROL REGIONS; HOXD, GREMLIN, AND LIMB MALFORMATIONS

A particularly interesting locus depending on long-range control is the HOXD gene locus on human chromosome 2q31 (see also the Spitz and Duboule chapter in this volume). The locus contains a number of genes, including the HOXD cluster and the EVX2 and lunapark (LNP) genes, involved in patterning of the axial skeleton and in limb development where they are required for correct digit formation (Spitz *et al.*, 2003). The HoxD cluster, with its characteristic collinearity in spatiotemporal expression of the genes correlating with gene order along the chromosome, has been extensively studied in mouse models. These studies have led to the notion that the ancestral role of the cluster was the specification of morphogenesis along the main body axis, while later in evolution the HoxD genes were co-opted to function in the development of novel structures such as the limbs (Spitz *et al.*, 2001). Consistent with this scenario, it was shown that the regulatory controls for the ancestral and collinear expression are located within the HoxD cluster, while the co-opted expression domains depend on enhancers located at remote positions outside the cluster. During limb development, HoxD10–13, Evx2, and Lnp are coexpressed in the presumptive digits with very similar profiles, suggesting the possibility that a single enhancer would control digit expression for all genes (van der Hoeven *et al.*, 1996). A search for such enhancer using BAC transgenics and sequence comparisons subsequently led to the identification of a region far upstream of the locus that controls tissue-specific expression in multiple tissues of a contiguous set of genes in the

locus. This region, termed the global control region (GCR), is proposed to create a widespread regulatory landscape, sharing its enhancing activity over a defined number of genes in a tissue-specific manner (Spitz *et al.*, 2003). Digit activity of the GCR spreads over six genes and at least 240 kb, while CNS activity is limited to the Lnp and Evx2 genes. A direct demonstration of the role of the GCR was provided by analysis of a semidominant mouse limb mutant, *Ulnaless*. *Ulnaless* carries a paracentric inversion with one breakpoint in the Lnp gene and the other 770 kb more telomeric, thus resulting in an inversion of Evx2 and the HoxD cluster. When bred against a targeted allele with a deletion of the Evx2 to HoxD11 region, expression of Evx2 and HoxD13 in distal limb was shown to be lost from the *Ulnaless* allele (Herault *et al.*, 1997; Spitz *et al.*, 2003). This suggests that the inversion either has moved the genes out of reach of the GCR (about 700 kb) or has introduced some insulating activity between the genes and the GCR. Interestingly, while no human cases with chromosomal disruptions between the GCR and the HoxD genes have been described, one human translocation patient has been identified with a translocation centromeric to the GCR (Dlugaszewska *et al.*, 2006) The patient, who exhibits severe brachydactyly and syndactyly among other abnormalities, carries an apparently balanced t(2;10)(q31.1;q26.3) translocation. The chromosome 2 breakpoint lies ~390 kb from the end of HOXD13 and no known genes have been identified on a breakpoint spanning BAC, implicating a role of misregulation of the HoxD genes in the phenotype.

An elegant series of deletion and duplication experiments in mouse (Tarchini and Duboule, 2006) has demonstrated that an early limb bud-activating element (ELCR) lies telomeric to the HoxD cluster, an idea first hypothesized by Zakany *et al.* (2004). The element controls Hox gene expression such that the closer a gene is positioned to the telomeric end of the cluster, either naturally or by artificial means, the earlier it is activated during development and the more anterior within the limb bud it is expressed. This early expression pattern appears to correlate with subsequent expression in the developing forearm and while the element responsible has not been located, it is interesting to note that a patient suffering from mesomelic dysplasia (shortening of the forearms and forelegs) with vertebral defects was found to have a balanced translocation 56 kb telomeric from the HoxD1 gene (Spitz *et al.*, 2002). Two other patients with abnormalities restricted to the limbs have been identified carrying balanced *de novo* translocations with breakpoints 450 and 950 kb telomeric to the HoxD complex (Dlugaszewska *et al.*, 2006).

Over the years, much work has gone into studying the mouse limb deformity (ld) mutations. Mice display disrupted epithelial–mesenchymal interactions between the polarizing region and the apical ectodermal ridge, which alters the pattern of the distal limb and results in fusion and loss of digits

(see Zeller *et al.*, 1999). A number of ld alleles have been described and as these disrupt the C terminal domain of formin, this was proposed as the causative mutation (Maas *et al.*, 1990; Woychik *et al.*, 1990). However, this was rather at odds with the discovery that the remaining two alleles (ld^{OR} and ld^{J}) disrupt the coding region of the neighboring gene Gremlin (the intragenic distance is 38 kb) and are allelic to loss-of-function mutations within it. This discrepancy has been resolved by the discovery that a GCR, required for correct expression of gremlin (and formin) within the posterior limb bud mesenchyme, lies within the region of exons 19–24 of formin and is deleted or disturbed in the relevant ld alleles (Zuniga *et al.*, 2004). Taken together, the GCRs from the HoxD complex and the Gremlin locus reveal a mechanism whereby expression of a number of genes within a region can be controlled in concert.

X. FOX GENES AND POSITION EFFECTS

A. FOX genes in eye anomalies

A number of position effect cases are associated with genes of the forkhead/winged helix family of transcription factors, indicating the importance of gene dosage for members of this family. Forkhead genes are involved in a diverse range of developmental pathways, but a large number of them appear to be involved in eye development (Lehmann *et al.*, 2003). FOXC1 (previously named FKHL7) lies in a cluster of forkhead genes on chromosome 6p25. Mutations cause glaucoma-associated ocular developmental anomalies with varying degrees of iris and extraocular abnormality, in addition to abnormalities in a range of other organ systems. Segmental duplication or deletion of the 6p25 region results in developmental defects of the anterior segment of the eye, indicating that precise gene dosage is critical for normal eye development. In addition to mutations and deletions in the FOXC1 gene itself in patients with Axenfeld Rieger anomaly and iris hypoplasia, a balanced translocation mapping 25 kb from the gene was found in a patient with primary congenital glaucoma (Nishimura *et al.*, 1998). A further patient, with glaucoma- and autosomal-dominant iridogoniodysgenesis, was shown to carry an interstitial deletion of 6p24-p25 with the proximal breakpoint estimated to lie at least 1200 kb proximal to the FOXC1 locus (Davies *et al.*, 1999). The case for a position effect in this patient is possible but speculative considering the large distance and the presence of a number of other possible candidate genes, most notably AP2α, in the deletion interval.

A typical case of a position effect is provided by a closely related forkhead member, the FOXC2 gene on 16q24 (formerly MFH-1). FOXC2-inactivating mutations have been found in patients with lymphedema-distichiasis (LD; OMIM 153400), an autosomal-dominant disorder that classically presents

as lymphedema of the limbs and double rows of eyelashes (distichiasis). Other complications may include cardiac defects, cleft palate, extradural cysts, and photophobia, highlighting the pleiotrophic effects of FOXC2 during development. A t(Y;16) (q12;q24.3) translocation with the breakpoint mapping 120 kb 3′ of the FOXC2 gene was found in a patient with neonatal lymphedema. The translocation did not appear to interrupt a gene on chromosome 16, nor were any candidate genes found on the Y chromosome, making a clear case for a position effect on the FOXC2 gene. Interestingly, the FOXL1 gene maps in between FOXC2 and the breakpoint, and thus could also be inactivated and have phenotypic effects in this patient (Fang *et al.*, 2000).

A third forkhead family member implicated in a position effect scenario is the FOXL2 gene on chromosome 3q23. Coding region mutations in FOXL2 have been shown to cause blepharophimosis-ptosis-epicanthus inversus syndrome (BPES; OMIM 110100), an eyelid and forehead dysmorphology in both sexes, often associated with gonadal dysgenesis and premature ovarian failure (POF) in women (Crisponi *et al.*, 2001). Consistent with the disease phenotypes, FOXL2 is selectively expressed in the mesenchyme of developing mouse eyelids and in adult ovarian follicles. Three reported translocations that cause BPES, mapping 170 kb from FOXL2, all fall within intron 6 of MRPS22, a ubiquitously expressed gene located upstream of FOXL2 (Crisponi *et al.*, 2004). Sequence comparisons between human and mouse reveal the presence of three highly conserved segments beyond the furthest breakpoint, in introns 6, 11, and 12 of MRPS22. In a more recent study, MLPA was used to specifically detect deletions in the FOXL2 region. In addition to cases removing the FOXL2 gene itself, five deletion cases with a phenotype identical to intragenic mutations were found mapping both upstream (four cases) and downstream (one case) of the intact FOXL2 gene (Beysen *et al.*, 2005). The smallest of the upstream deletion cases involves a 126-kb region that includes the human orthologous region to the goat polled intersex syndrome (PIS) locus. Polled intersex syndrome (PIS) is a genetic syndrome in goats, which combines a craniofacial defect resulting in polledness (an absence of horns), female infertility, and XX sex reversal, and is considered an animal model for BPES. It is caused by an 11.7-kb deletion, located 280 kb upstream of goat FOXL2, distal of the three translocation breakpoints (Pailhoux *et al.*, 2001). It encompasses a block of several conserved sequences in intron 11 of MRPS22, making these strong candidates to be distant *cis*-regulatory elements affecting FOXL2 expression.

XI. SOX9 AND CAMPOMELIC DISPLASIA

A locus with strong evidence for long-range regulatory elements located at extremely large distances from its cognate gene is the SOX9 locus on chromosome 7q24.3. SOX9 mutation has been identified as the cause of campomelic

dysplasia (CD; OMIM 114290), an autosomal-dominant osteochondrodysplasia, characterized by congenital shortening and bowing of the long bones (campomelia) in combination with other skeletal anomalies, such as hypoplasia of the scapular and pelvic bones, lack of mineralization of thoracic pedicles, a missing pair of ribs, and clubbed feet. Lung hypoplasia and a malformed thorax, resulting in severe respiratory distress, may lead to the neonatal death of some patients. XY sex reversal is found in about two thirds of karyotypically male CD patients. In most CD cases, heterozygous loss-of-function mutations are found within the coding region of SOX9 on chromosome 17q, implying that CD results from haploinsufficiency for SOX9. However, as with other position effect cases already described, CD patients have been identified who carry two intact copies of the compact 5.4 kb SOX9 transcription unit, with chromosomal rearrangements in the vicinity of one SOX9 allele. In all but one case, the rearrangements are found upstream of SOX9, and the breakpoints have been mapped at distances from 50 kb up to 950 kb (Hill-Harfe *et al.*, 2005; Pfeifer *et al.*, 1999). In a single case with a complex, balanced translocation, the 17q breakpoint was mapped 1.3 Mb downstream from Sox9 (Velagaleti *et al.*, 2005). Examination of a large genomic interval around SOX9 has failed to find other genes, giving rise to the notion that SOX9 resides in a so-called "gene desert." The lack of any protein-coding gene in the region combined with the similarity, though generally less severe, of the phenotype of these position effect patients to SOX9 loss-of-function cases strongly suggests that the chromosomal rearrangements remove one or more *cis*-regulatory elements. Detailed analysis of the breakpoints of 15 CD patients suggests that they are found in two clusters: ten translocation breakpoints and one inversion fall in a proximal cluster between 50 and 375 kb upstream, while four more breakpoints lie in a distal cluster between 789 and 932 kb away (Leipoldt *et al.*, 2007). A further patient with CD presented with a large deletion from 380 to 1860 kb upstream (Pop *et al.*, 2004). In accordance with the human phenotype caused by SOX9 mutations, studies in the mouse have shown that Sox9 functions as an essential developmental regulator at various steps of chondrogenesis and during the initial phase of testis determination and differentiation. Furthermore, heterozygous Sox9 knockout mice recapitulate essentially all the symptoms seen in CD patients except for the sex reversal (Kist *et al.*, 2002). The involvement of long-range gene control is supported by the fact that mice transgenic for human SOX9 spanning YACs showed transgene expression patterns (except in gonads) that were similar to endogenous Sox9 only when the YAC transgene contained 350 kb of sequence upstream of SOX9, but not with a truncated YAC containing only 75 kb of SOX9 5′ flanking sequence (Wunderle *et al.*, 1998). Comparative sequence analysis of the SOX9 genomic region between human, mouse, chicken, and fugu has revealed the presence of many conserved elements, some of which have been shown to drive reporter expression in a SOX9 sub-pattern (Bagheri-Fam *et al.*, 2001, 2006; Qin *et al.*, 2004).

In the mouse, a complex position effect has been reported as the cause of the dominant insertional mutation *Odd sex* (ocular degeneration with sex reversal, *Ods*) (Bishop *et al.*, 2000). *Ods* is the result of a transgenic insertion of a tyrosinase minigene driven by the dopachrome tautomerase (Dct) promoter accompanied by a 134-kb deletion at ~980 kb upstream of Sox9. In contrast to the male to female sex reversal found in many human CD patients, *Ods* mice show female to male sex reversal, as well as microphthalmia with pigmentation defects and cataracts. The XX sex reversal phenotype of these mice is thought to be the result of *Sox9* misexpression. Normally *Sox9* expression is repressed in the XX fetal gonad at the time of sex determination, and this repression is counteracted by the presence of *Sry* in the male gonads. Initially, it was proposed that the *Ods* deletion had removed gonad-specific long-range regulatory element(s) that normally mediate the female-specific repression of *Sox9*, resulting in upregulation of *Sox9* in the absence of *Sry* and the consequent male development (Bishop *et al.*, 2000). However, more recent experiments have shown that the 134-kb deletion in itself is insufficient to cause the sex reversal (Qin *et al.*, 2004). When a double gene targeting strategy was used to recreate the *Ods* deletion combined with the introduction of a tyrosinase minigene driven by its own promoter, no eye or sex reversal phenotype was observed, suggesting instead a mechanism of long-range interaction between the *Dct* promoter and *Sox9* in the *Ods* mutant. The *Ods* eye phenotype was also generated in transgenic mice with a *Dct-Sox9* minigene cassette, and a temporal misexpression of *Sox9* under the control of the *Dct* promoter was demonstrated. This suggests that in the eyes of *Ods* mice, the *Dct* promoter acts as a long-range activator on the *Sox9* promoter over a distance of 980 kb. The mechanism behind the sex reversal phenotype is more complex. No sex-reversal was seen in *Dct-Sox9* transgenics, suggesting that in the gonads the *Dct* promoter interacts with *Sox9* via an indirect mechanism possibly involving endogenous gonad-specific *Sox9* enhancers and chromosomal conformation changes due to the deletion (Qin *et al.*, 2004).

XII. FACIOSCAPULOHUMERAL DYSTROPHY

Facioscapulohumeral dystrophy (FSHD) is a neuromuscular disorder affecting predominantly the facial and shoulder girdle muscles. It is inherited as an autosomal dominant trait, but the onset and severity of the disease can be highly variable even within patient families carrying the same genetic lesion, and about 10–30% of patients carry *de novo* mutations, some of them as somatic mosaics. At present the mechanism underlying the disease, although still not very well understood, appears to have a unique epigenetic etiology. The characteristic molecular event is the deletion of an integral number of 3.3 kb tandem repeats from the subtelomeric region of the long arm of chromosome 4 (Wijmenga *et al.*, 1992). Unaffected individuals carry between 11 and 150 copies of this repeat,

named D4Z4, while patients have 10 or fewer on one of their chromosomes (van Deutekom *et al*., 1993). In general, the severity and onset of the disease correlates with the number of repeats, with a lower repeat number associating with more severe disease manifestations. Interestingly, no FSHD patients have been reported completely lacking all D4Z4 repeats, and individuals carrying an unbalanced translocation where the entire 4q35 region is lost do not suffer from FSHD. This suggests that the loss of repeats results in a gain-of-function mutation (Tupler *et al*., 1996), but much debate continues over the mechanism. The D4Z4 repeat contains internal VNTR-type repeats, has CpG island characteristics, and contains a putative open reading frame (DUX4), but no protein-coding transcripts have been identified from the repeat sequence despite intense efforts. The short D4Z4 repeat at 4q35 appears to cause FSHD indirectly by some unknown *cis*-interaction, because an almost identical repeat array (plus further homology on both sides of the repeat) is present on chromosome 10q26, but does not cause the disease when shortened (Bakker *et al*., 1995). Thus, FSHD is the result of a mutation outside the gene or genes underlying the disease, suggesting some kind of position effect mechanism. Furthermore, in addition to the D4Z4 contraction at least one other, so far unidentified *cis*-element is required to develop the disease. The 4q subtelomeric region was found to exist in two allelic variants, 4qA and 4qB, which are almost equally common in the population. FSHD alleles are always of the 4qA subtype, suggesting additional elements on 4qA are necessary to cause the disease or elements on 4qB can act preventively. Some FSHD families have been reported with healthy individuals carrying FSHD-sized repeat contractions on a 4qB allele (Lemmers *et al*., 2004). A number of genes or putative genes have been identified in the region centromeric to the D4Z4 repeats, ANT1, FRG1, DUX4C, and FRG2. ANT1, though located at 3.5 Mb distance, was considered the most likely candidate because it encodes an adenine nucleotide translocator that has been implicated in myopathy and is predominantly expressed in heart and skeletal muscle (Gabellini *et al*., 2002). A multiprotein repressor complex consisting of HMG2B, YY1, and nucleolin has been identified that binds to the D4Z4 repeat (Gabellini *et al*., 2002). This complex was hypothesized to act as a repressor, negatively regulating gene expression of the 4q35 region, possibly through promoting the spreading of heterochromatin in *cis* throughout the region. D4Z4 deletions remove binding sites for the repressor complex to below a certain threshold, thus allowing local decondensation of chromatin and the consequent derepression of the genes in the region (Gabellini *et al*., 2002). In support of the hypothesis, all the genes in the region were found to be upregulated in FSHD. However, several follow-up studies have failed to reproduce the upregulation of ANT1 and FRG1 in muscle tissue of FSHD patients compared to unaffected individuals. Furthermore, histone H4 acetylation levels over the various gene promoters suggested that the region adopts a nonexpressed euchromatin-like structure both in control

individuals and in FSHD patients (Jiang *et al.*, 2003). Instead, a model was proposed in which a short array of D4Z4 repeats forms a long distance loop to interact directly with an as yet unknown gene or genes on 4q35, while longer arrays of D4Z4 repeats form intra-array loops, thus sequestering the array (Jiang *et al.*, 2003).

The case for involvement of several genes in the region, notably ANT1, PDLIM3, and FRG2, has weakened by recent overexpression studies in transgenic mice where none of the genes appeared to cause muscular dystrophy. Mice overexpressing FRG1, a putative spliceosomal protein gene located 120 kb from the D4Z4 repeats, on the other hand develop muscular dystrophy with a severity proportional to the level of overexpression (Gabellini *et al.*, 2006). Nonetheless, the role of FRG1 remains speculative. Another avenue of exploration that is pursued relates to the observation that the 4q telomere uniquely always localizes at the nuclear periphery, requiring Lamin A/C to do so, unlike all other chromosome ends studied including the highly homologous 10qter (Masny *et al.*, 2004). How this links with alterations in gene expression in FSHD patients is not clear, but disturbed myogenic differentiation detectable by gene expression profiling suggests possible chromatin effects.

XIII. ABERRANT CREATION OF AN ILLEGITIMATE siRNA TARGET SITE

In recent years, small nuclear RNAs have become recognized as important modulators of transcriptional and translational mechanisms of gene expression. A study to identify quantitative trait loci underlying the economically important feature of meatiness of Texel sheep pointed toward involvement of the myostatin (GDF8) gene locus. Detailed analysis has led to the identification of a single base pair substitution in the 3′UTR of the gene (Clop *et al.*, 2006). This mutation creates a target site for *mir1* and *mir206*, two miRNAs that are highly expressed in skeletal muscle, and causes translational inhibition of the myostatin gene leading to muscular hypertrophy. As miRNAs are undoubtedly also involved in regulatory control in humans, such a mechanism could equally well be involved in the etiology of human genetic disease.

One of the first examples of such a disease mechanism is the aberrant strengthening of a weak miRNA site in the 3′UTR of a gene called SLITRK1, as a possible cause of Tourette syndrome (TS). TS is a developmental neuropsychiatric disorder characterized by chronic vocal and motor tics. In this case the gene involved is the Slit and Trk like family member 1 (SLITRK1) gene. Interestingly, SLITRK1 was first identified as a potential TS syndrome candidate from a position effect case in a boy with a *de novo* chromosome 13 inversion (inv (13)(q31.1;q33.1)). The breakpoint mapped well outside the coding region of

the gene, but subsequent sequencing of the single coding exon of SLITRK1 in a TS patient cohort identified a truncating mutation in one of them. More interesting was the finding of an identical sequence variant in the 3′UTR of two further, unrelated patients. The single base pair change mapped to a highly conserved nucleotide within a predicted target site for human miRNA hs-miR-189, and would be predicted to strengthen the binding site. SLITRK1 and miR-189 show substantial overlap in expression pattern in the brain, and luciferase assays suggest that repression of the mutant 3′UTR variant of SLITRK1 by the miRNA is significantly increased over the wild-type version (Abelson *et al.*, 2005).

XIV. GENETIC DISEASE DUE TO ABERRANT GENE TRANSCRIPTION CAN BE CAUSED BY MANY DIFFERENT MECHANISMS

As highlighted by the cases described above, transcriptional control can be disrupted by many different mechanisms. So far the following mechanisms have been encountered: (1) separation of *cis*-regulatory elements from the gene promoters through chromosomal translocations or inversions, (2) deletion of long range *cis*-elements, (3) deleterious mutations in *cis*-elements, (4) disturbing the normal interactions of promoters and *cis*-acting enhancers through appearance of a new promoter, (5) alteration of local chromatin structure through interference by an antisense transcript, and (6) disturbance of more global chromatin structure through loss of microsatellite repeats.

A further, but at this stage rather speculative, mechanism, with a potential involvement in genetic disease is transvection. Transvection refers to the effect on a gene's expression from one allele through the influence of regulatory elements on the opposite allele through the pairing of homologous chromosomes. In some disease cases, particularly involving dosage-sensitive genes, differences in severity have been observed between inactivating mutations and larger gene deletions. One possible mechanism to explain that phenomenon could be an ameliorating influence through regulatory effects of distal control elements on the remaining intact copy of the gene on the other chromosome. Such a model has been proposed for the observed differences in penetrance of craniofacial anomalies in mouse models of Smith-Magenis syndrome (SMS OMIM 182290) (Yan *et al.*, 2007). Craniofacial abnormality is one of the major clinical manifestations of SMS, and includes midface hypoplasia, broad nasal bridge, and prognathia. SMS is a multiple congenital anomaly and mental retardation condition due to a heterozygous 3.7-Mb deletion on chromosome 17p11.2 in the majority of cases. Frameshift and nonsense mutations have been identified in RAI1, a retinoic acid inducible gene, in some patients with SMS, suggesting it is the major gene involved through haploinsufficiency. Whereas

penetrance of the craniofacial phenotype in isogenic strains of mice carrying various sized deletion around the Rai1 gene was more or less complete, it was much lower in a Rai1 insertional inactivation strain which has retained the rest of the locus. Considering the dosage-sensitivity of Rai1, an upregulation of the wild-type copy of the gene through a *trans*-regulatory mechanism could be a plausible mechanism (Yan *et al.*, 2007). Evidence, though mainly anecdotal, exists for similar differences in severity and penetrance between locus-deletion- and gene-mutation-derived disease cases and deeper investigation of such situations should bring new insights into the phenomenon of transvection.

A. The problem: How to find and assess regulatory mutations?

Determining that a genetic disease diagnosed in a patient is caused by a regulatory mutation or deletion is difficult for a number of reasons: Not only is the possibility of regulatory impairment rather than protein product malfunction not always considered, but for most genes too little is known about their regulatory sequences and mechanisms to easily further investigate the possibility. At present there are no simple ways to discern regulatory elements from nonfunctional sequence (though interspecies sequence comparison is proving very useful in this respect) or to know what regulatory and/or evolutionary conserved elements may be relevant to the disease etiology. Furthermore, while it is usually clear when a mutation is found in a protein-coding sequence whether that mutation causes a faulty protein, this is much more problematic for mutations in regulatory elements and requires extensive functional analysis. Finally, it is likely that many regulatory element mutations will be associated with phenotypes distinct from those identified for coding region mutations. For instance, if the regulatory mutation occurs in a tissue-specific element the phenotype may affect just a subset of the full spectrum seen with inactivation of the protein. The SHH gene forms an example where different diseases can be associated with the same gene: Disruption of the protein or loss of expression through proximal rearrangements causes HPE, while mutation of the distal, limb-specific ZRS enhancer leads to limb malformation (Fig. 13.2) (Lettice *et al.*, 2003). Similarly, if a mutation affects a later acting *cis*-element the regulatory mutant phenotype could be completely different as it would normally be masked by earlier effects from complete knockout of the gene. In general phenotypes due to regulatory mutations will be milder, because genes, in particular developmental control genes, are often regulated by multiple enhancers with partly overlapping activities. While in some cases these enhancers will act additively or synergistically, they can also act hierarchically. Both these options could be seen in a YAC transgenic experiment where a PAX6 DRR was deleted via a Cre/LoxP strategy (Fig. 13.1). The DRR contains a number of *cis*-elements including ones for expression of Pax6 in the developing lens and retina (Kleinjan *et al*, 2001). However, several

further enhancers for these expression sites are also found outside the DRR and these are not removed by the deletion. While after DRR deletion reporter expression was maintained in the lens (at least at early embryonic stages and possibly at lower level), expression in the retina, iris, and ciliary body was completely abolished (Kleinjan *et al.*, 2006), indicating that elements within the DRR are critically important for PAX6 expression in these sites.

The cases described above all highlight genes where long-range transcriptional control has been identified through the analysis of patients with genetic malformations. There are of course many more genes for which long-range gene regulation undoubtedly plays an important role, but which have so far not been implicated in a position effect-type genetic disease (e.g., DiLeone *et al.*, 2000; Hadchouel *et al.*, 2003; Kimura-Yoshida *et al.*, 2004; Uchikawa *et al.*, 2003). Many of the disease cases described in this chapter are caused by semidominant *de novo* chromosomal abnormalities. Because expression levels are often critical for many developmental regulators, haploinsufficiency acts as a selection mechanism for the link between these genes and their involvement in genetic disease. For other genes, however, dosage may not be critical; therefore, disruption of long-range control on one allele would not lead to a recognizable phenotype. In other cases, the disruption may have only a subtle effect or give rise to a phenotype that is quite different from the phenotype shown upon complete gene inactivation. Therefore one can safely assume that more genetic diseases caused by regulatory mutations will exist that have thus far not been associated with the correct causative gene.

As is clear from the many cases described above, long-range enhancers can be essential for correct expression of many genes. Mechanisms through which these distal enhancers interact with the promoters of their target genes have been discussed in other chapters. It is tempting to speculate that mutations in factors involved in such long-range enhancer–promoter interactions should also be a cause of genetic disease. One disorder where this may indeed be the case is Cornelia de Lange syndrome, (CDLS, OMIM 122470) (Krantz *et al.*, 2004; Tonkin *et al.*, 2004). CDLS is a multiple malformation disorder, characterized by facial dysmorphology, mental retardation, growth delay, and limb reduction. *NIPBL*, the human homologue of *Drosophila Nipped-B*, was identified as the gene mutated in individuals with CDLS. *Nipped-B* was first identified as a facilitator of enhancer promoter communication in *Drosophila* in a screen for long-range interactions with the *Drosophila cut* gene wing margin enhancer located 85 kb upstream of its cognate promoter (Dorsett, 1999; Rollins *et al.*, 1999). In keeping with the putative general role of Nipped-B in facilitating long-range regulation of multiple genes, compromised long-range regulation of several developmental control genes would fit with the diverse phenotypic anomalies seen in CDLS. Based on homology with yeast *Scc2* protein, with a role in sister chromatid cohesion, Nipped-B has been proposed to have a dual role, where a

more ancient function facilitating *trans*-interactions between sequences on sister chromatids has been adapted to include an additional role in long-range transcriptional regulation (Rollins *et al.*, 2004). The recent findings that two other forms of the disorder, CDLS2 (OMIM 300590) and CDLS3 (OMIM 610759), are caused by mutations in SMC1A and SMC3, both of which encode other components of the cohesin complex, points to a role of the cohesin complex in the regulation of multiple genes.

B. Long-range control and genome organization

The position effect cases presented above highlight a number of intriguing points: First, among the genes thus far shown to depend on long-range regulatory control, both through human disease-related position effects and work in experimental organisms, a large proportion consists of genes encoding developmental regulators. Similarly, it has been noted when using bioinformatic tools to map multispecies conserved elements in a genome-wide manner, that many of these conserved elements cluster near genes with key developmental functions (e.g., Woolfe *et al.*, 2005). We argue that this is not coincidental, but stems from the complex and strictly critical expression patterns these proteins require.

Developmental regulator genes need to be active in specific tissues at defined time-points in development, often at critically defined levels, and have to be strictly inactive in all other tissues and time-points. To achieve such sophisticated expression profiles, these genes require multiple enhancer elements and these regulatory elements all need to be fitted in *cis* in the region surrounding the gene. Moreover, despite the common use of terminology that suggests some sort of engineered design in the structure of gene loci with respect to transcription and other biological processes, the acquisition and loss of regulatory elements has not occurred through conscious design, but rather as a result of evolutionary tinkering (Carroll, 2001; Duboule and Wilkins, 1998). The redeployment of developmental regulatory genes in the development of new and other tissues and pathways has become a recognized feature in the evolution of greater complexity in higher organisms. To a large extent, this depends on the chance appearance of a new combination of sequences with regulatory activity within the vicinity of the appropriate promoter. As long as the activity can "reach" the promoter, does not interfere with the existing regulatory control in a disadvantageous manner, and presents some kind of evolutionary advantage itself, the new *cis*-element might become fixed. The further appearance of elements that synergize with existing ones to more reliably achieve the optimal expression level will similarly be of evolutionary advantage. One example where there is evidence for the redeployment of regulatory factors in more recent evolutionary adaptations is the HoxD cluster, and interestingly a correlation has been noted between the

more proximal location of enhancers involved in the ancestral function of the gene and the more distal location of regulatory elements required for expression in tissues in which the gene has been co-opted. As described above, the HoxD cluster is essential for axial patterning along the main body axis as well as for limb development. Control elements for expression in the trunk, the ancestral site of HoxD function, are located within the HoxD cluster, while control elements for limb expression are located in more distal positions (Spitz *et al.*, 2001). However, the separation between proximal, more ancient and distal, more recent *cis*-elements in the HoxD cluster is probably a somewhat special case where the appearance of new *cis*-elements is strongly influenced by the clustered, coregulated nature of the genes in the complex, and the formation of a new element within the cluster would be likely to interfere adversely with the regulatory phenomenon of collinearity that is the hallmark of the Hox complexes (Tarchini *et al.*, 2006) However, when the distant positions of the limb-specific control elements in the HoxD complex as well as in the Shh and Gremlin loci are considered together, it is plausible that their separation from the coding regions represents a more recent add-on to the ancestral axial pattern in concert with the evolution of the limb.

Second, the analysis of several "position effect" cases has highlighted the fact that *cis*-regulatory elements that drive expression of one gene may be located within introns of another, neighboring gene, or can in some cases even be located beyond the adjacent gene. This not only presents a problem for investigators trying to find the causative gene for a disease of interest through mapping of patient breakpoints, but has also initiated new ideas about the concept of genes and gene domains. Eukaryotic genomes are made up of a large number of genes and gene loci that are regulated independently, and often genes/loci with quite different patterns of expression are located in close proximity to one another on the chromosome. Traditionally the concept of "structural gene domains" was widely established to explain how adjacent gene loci maintain their independence. In the structural domain model, genes enjoy functional autonomy through the physical separation from neighboring domains that carry a different local chromatin structure sometimes in combination with or brought about by specific functional sequences such as boundary or insulator elements (Dillon and Sabbattini, 2000). The chicken β-globin locus is a strong example to support this model (Felsenfeld *et al.*, 2004). However, the fact that genes can overlap and their *cis*-regulatory elements can be found within or beyond neighboring unrelated genes puts the universal applicability of the structural domain model into question. Detailed analysis of some of the disease genes described has highlighted that the structural domain model cannot apply to all genes. For instance, the main critical regulatory element in the human α-globin locus, the −40-kb enhancer, is located within an intron of a

neighboring, ubiquitously expressed housekeeping gene located in a large region of open chromatin (Vyas *et al*., 1995). In the case of PAX6, a number of regulatory elements have been found spread throughout at least three of the introns of the neighboring ELP4 gene (Kleinjan *et al*., 2001). In the human growth hormone cluster, elements required for tissue-specific expression of the pituitary-specific GHN gene and the placenta-specific CSL, CSA, GHV, and CSB genes are located beyond a B-cell-specific gene, within introns of the muscle-specific SCN4A gene (Bennani-Baiti *et al*., 1998). Many more examples are now known. These observations provide a compelling argument against a fundamental requirement for the physical isolation of a gene and its regulatory sequences. Rather it has been proposed that for many loci, specificity of enhancer–promoter interactions is the key in maintaining the functional autonomy of adjacent genes (Dillon and Sabbattini, 2000). Thus, while the concept of a gene as a physical entity with a distinct map position on the chromosome is still useful in many instances, a new concept has been proposed, based on the old definition of the gene as a "unit of inheritance" (de Laat and Grosveld, 2003; Dillon, 2003). In that concept genes are defined in functional terms as "functional expression modules," encompassing both the transcribed regions and their *cis*-regulatory control systems comprising of the specific *cis*-acting sequences in conjunction with the local chromatin structure. This model works independent of the physical structure of the gene locus, so that a functional expression module can derive equally well from a gene and its *cis*-elements that are interdigitated with adjacent genes, or located in its own separate domain or in a gene desert.

The completion of the human (and other) genome sequence has confirmed the highly uneven distribution of genes across the genome into gene-rich and gene-poor areas. It has been estimated that around 25% of the human genome consists of long regions lacking any protein-coding sequences or obvious biological function, and these regions have been termed gene deserts. (Venter *et al*., 2001). Based on comparative conservation levels with the chicken genome, gene deserts can be classed into two categories, stable and variable (Ovcharenko *et al*., 2005). Some of the gene deserts have been shown to contain long-range *cis*-regulatory sequences acting on neighboring genes (Kimura-Yoshida *et al*., 2004; Nobrega *et al*., 2003; Uchikawa *et al*., 2003) and these are overrepresented in the stable class of gene deserts. Gene ontology categories of genes neighboring stable gene deserts show a strong bias toward transcriptional and developmental regulators (Ovcharenko *et al*., 2005), suggesting the presence of long-range enhancers plays a role in the evolutionary maintenance of the stable gene deserts. In contrast, gene deserts of the variable class may be nonessential to genome function, since they can be deleted without obvious phenotypic consequence (Nobrega *et al*., 2004). The location of long-distance enhancers within or beyond adjacent genes is also thought to be a major force in the conservation of blocks of synteny over large evolutionary timespans (Mackenzie *et al*., 2004).

In conclusion, the fact that many developmental regulators in higher organisms have extended regulatory landscapes is a consequence of their acquisition of multiple roles in various spatiotemporal defined sites in the embryo. A direct and obvious consequence of having large gene domains is an increased risk of its disruption by chromosomal rearrangements such as translocations, deletions, and inversions.

C. Implications for (common) genetic disease

It is clear from the "position effect" cases reviewed above that long-range gene regulation plays a critical role in many developmental processes and that its disruption can lead to severe congenital disease. While the occurrence of distal control elements may be more common to developmental regulatory proteins due to their more complex spatiotemporal expression requirements, long-range gene control undoubtedly plays a role at many more gene loci. The effects of disruption/mutation of distal regulatory elements in those genes will be more subtle, but may be a significant factor in common disease and quantative traits.

It is now well accepted that a large percentage of quantitative traits are caused by variants that affect gene expression, so-called expression Quantitative trait loci (QTL). In some well-studied cases, the QT is shown to be the result of single nucleotide changes in distal regulatory elements. A genetic trait where such a single base pair substitution has had a dramatic effect is lactase persistence. While worldwide in most humans the ability to digest the milk sugar lactose declines rapidly after weaning because of decreasing levels of the enzyme lactase phlorizin hydrolase, this ability is maintained with high frequency in certain populations, particularly those whose ancestors have traditionally practiced cattle domestication. Lactase persistence in Europeans has been strongly associated with an SNP (C/T-13910) located 13.9 kb upstream of the lactase gene promoter, with a second correlative SNP found 22 kb upstream (Enattah *et al.*, 2002). The lactose tolerance SNP lies in a *cis*-element that enhances promoter activity (Olds and Sibley, 2003; Lewinsky *et al.*, 2005; Troelsen *et al.*, 2003). In a further study of this trait, three more SNPs in the same putative *cis*-element, SNPs (G/C-14010, T/G-13915, C/G-13907), were found to correlate with lactase persistence in pastoral African populations (Tishkoff *et al.*, 2007). Association studies demonstrate that the different SNP alleles have arisen independently and have spread rapidly to high frequency because of the strong selective force of adult milk consumption (Tishkoff *et al.*, 2007).

Quantitative traits caused by regulatory SNPs affecting levels of gene expression are likely to lie at the heart of many common diseases with a genetic component, such as diabetes, heart disease, hypertension, and obesity. In a recent study (Herbert *et al.*, 2006), a link was found between a common genetic variant located 10 kb upstream of the insulin-induced gene 2 (INSIG2) gene and risk of

obesity. INSIG2 protein can inhibit cholesterol and fatty acid synthesis, and thus makes a plausible QTL candidate gene affecting body mass index. The predisposing genotype was present in 10% of the test population, suggesting it is an ancient allele that has only recently become deleterious.

Quantitative traits can obviously be caused by variants in all kinds of *cis*-regulatory elements including distal enhancers, but also in promoters and intronic *cis*-elements. An example of the former is the association of a genetic predisposition to autism with an SNP in the promoter of the pleiotropic MET receptor tyrosine kinase gene on chromosome 7q31 (Campbell *et al.*, 2006). MET signaling has a role in neocortical and cerebellar growth and maturation, immune function, and gastrointestinal repair, consistent with reported medical complications in some children with autism. Functional assays showed that the autism-associated allele results in a twofold decrease in MET promoter activity.

A mutation in an intronic enhancer for the RET proto-oncogene is associated with an increased risk of Hirschsprung disease (HSCR). HSCR, or congenital aganglionosis with megacolon, is a multifactorial, non-mendelian disorder. The RET gene is the main HSCR gene implicated, but mutations also occur in many other genes involved in enteric development. Resequencing of conserved elements in the RET locus of HSCR families with demonstrated RET linkage, but no identified coding sequence mutations, led to the identification of an SNP (RET + 3) in a conserved element in intron 1 of the gene (Emison *et al.*, 2005). This common variant SNP was shown to decrease enhancer activity of the element in an *in vitro* enhancer assay. This case shows that not all mutations for rare diseases need to be rare themselves or fully penetrant, but that in complex genetic disorders noncoding mutations can conspire with mutations at additional sites or genes for disease to occur. Thus, the customary practice of validating a potential mutation by applying the criterion that it must be absent from nonaffected controls may not always be appropriate (Emison *et al.*, 2005).

XV. CONCLUDING REMARKS

Over the last 40 years, it has become well established that the genes within our genome contain the code to transform genetic information into the amino acid sequence of our proteins. In this post genome era, we are beginning to understand that the genome contains many more layers of biological information, not least of which are the codes that tell the transcriptional machinery where, when, and how much of those proteins to produce. While over the past decades, we have learned much about these latter codes, we are still a long way off fully understanding the mechanisms of gene regulation. Nevertheless, detailed analysis of human genetic malformations has provided invaluable information, first in

highlighting the occurrence and scope of long-distance transcriptional regulation in the genome, and second in the elucidation of a number of different mechanisms of long-range gene control. It seems likely that long-range transcriptional control is essential for a relatively small but substantial subset of our genes, and the genes and diseases discussed in this chapter form just a small example of that group. They do, however, serve as paradigms in our efforts to unravel the regulatory codes embedded in our genome and understand the (long-range) regulatory mechanisms that underlie our healthy development and homeostasis and which, when disrupted, sadly lead to genetic disorders.

References

Abelson, J. F., Kwan, K. Y., O'Roak, B. J., Baek, D. Y., Stillman, A. A., Morgan, T. M., Mathews, C. A., Pauls, D. L., Rasin, M. R., Gunel, M., Davis, N. R., Ercan-Sencicek, A. G., *et al.* (2005). Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* **310** (5746), 317–320.

Alward, W. L., Semina, E. V., Kalenak, J. W., Heon, E., Sheth, B. P., Stone, E. M., and Murray, J. C. (1998). Autosomal dominant iris hypoplasia is caused by a mutation in the Rieger syndrome (RIEG/PITX2) gene. *Am. J. Ophthalmol.* **125,** 98–100.

Anguita, E., Sharpe, J. A., Sloane-Stanley, J. A., Tufarelli, C., Higgs, D. R., and Wood, W. G. (2002). Deletion of the mouse alpha-globin regulatory element (HS-26) has an unexpectedly mild phenotype. *Blood* **100**(10), 3450–3456.

Bagheri-Fam, S., Ferraz, C., Demaille, J., Scherer, G., and Pfeifer, D. (2001). Comparative genomics of the SOX9 region in human and Fugu rubripes: Conservation of short regulatory sequence elements within large intergenic regions. *Genomics* **78,** 73–82.

Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Gunther, T., Schule, R., Kemler, R., Mallo, M., Kanzler, B., and Scherer, G. (2006). Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Dev. Biol.* **291**(2), 382–397.

Bakker, E., Wijmenga, C., Vossen, R. H., Padberg, G. W., Hewitt, J., van der, W. M., Rasmussen, K., and Frants, R. R. (1995). The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* **2,** 39–44.

Balemans, W., Ebeling, M., Patel, N., Van Hul, E., Olson, P., Dioszegi, M., Lacza, C., Wuyts, W., Van Den Ende, J., Willems, P., Paes-Alves, A. F., Hill, S., *et al.* (2001). Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST). *Hum. Mol. Genet.* **10,** 537–543.

Balemans, W., Patel, N., Ebeling, M., Van Hul, E., Wuyts, W., Lacza, C., Dioszegi, M., Dikkers, F. G., Hildering, P., Willems, P. J., Verheij, J. B., Lindpaintner, K., *et al.* (2002). Identification of a 52 kb deletion downstream of the SOST gene in patients with van Buchem disease. *J. Med. Genet.* **39,** 91–97.

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., Koop, B., *et al.* (1996). Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nat. Genet.* **14,** 353–356.

Benito-Sanz, S., Thomas, N. S., Huber, C., Gorbenko del Blanco, D., Aza-Carmona, M., Crolla, J. A., Maloney, V., Rappold, G., Argente, J., Campos-Barros, A., Cormier-Daire, V., and Heath, K. E. (2005). A novel class of Pseudoautosomal region 1 deletions downstream of SHOX is associated with Leri-Weill dyschondrosteosis. *Am. J. Hum. Genet.* **77,** 533–544.

Bennani-Baiti, I. M., Cooke, N. E., and Liebhaber, S. A. (1998). Physical linkage of the human growth hormone gene cluster and the CD79b (Ig beta/B29) gene. *Genomics* **48,** 258–264.

Beysen, D., Raes, J., Leroy, B. P., Lucassen, A., Yates, J. R., Clayton-Smith, J., Ilyina, H., Brooks, S. S., Christin-Maitre, S., Fellous, M., Fryns, J. P., Kim, J. R., *et al.* (2005). Deletions involving long-range conserved nongenic sequences upstream and downstream of FOXL2 as a novel disease-causing mechanism in blepharophimosis syndrome. *Am. J. Hum. Genet.* **77,** 205–218.

Bishop, C. E., Whitworth, D. J., Qin, Y., Agoulnik, A. I., Agoulnik, I. U., Harrison, W. R., Behringer, R. R., and Overbeek, P. A. (2000). A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. *Nat. Genet.* **26,** 490–494.

Bowl, M. R., Nesbit, M. A., Harding, B., Levy, E., Jefferson, A., Volpi, E., Rizzoti, K., Lovell-Badge, R., Schlessinger, D., Whyte, M. P., and Thakker, R. V. (2005). An interstitial deletion-insertion involving chromosomes 2p25.3 and Xq27.1, near SOX3, causes X-linked recessive hypoparathyroidism. *J. Clin. Invest.* **115**(10), 2822–2831.

Brunkow, M. E., Gardner, J. C., Van Ness, J., Paeper, B. W., Kovacevich, B. R., Proll, S., Skonier, J. E., Zhao, L., Sabo, P. J., Fu, Y., Alisch, R. S., Gillett, L., *et al.* (2001). Bone dysplasia sclerosteosis results from loss of the SOST gene product, a novel cystine knot-containing protein. *Am. J. Hum. Genet.* **68**(3), 577–589.

Cai, J., Goodman, B. K., Patel, A. S., Mulliken, J. B., Van Maldergem, L., Hoganson, G. E., Paznekas, W. A., Ben Neriah, Z., Sheffer, R., Cunningham, M. L., Daentl, D. L., and Jabs, E. W. (2001). Increased risk for developmental delay in Saethre-Chotzen syndrome is associated with TWIST deletions: An improved strategy for TWIST mutation screening. *Hum. Genet.* **114,** 68–76.

Campbell, D. B., Sutcliffe, J. S., Ebert, P. J., Militerni, R., Bravaccio, C., Trillo, S., Elia, M., Schneider, C., Melmed, R., Sacco, R., Persico, A. M., and Levitt, P. (2006). A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl. Acad. Sci. USA* **103**(45), 16834–16839.

Carroll, S. B. (2001). Chance and necessity: The evolution of morphological complexity and diversity. *Nature* **409,** 1102–1109.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F., and Fraser, P. (2002). Long-range chromatin regulatory interactions *in vivo*. *Nat. Genet.* **32,** 623–626.

Castermans, D., Vermeesch, J. R., Fryns, J. P., Steyaert, J. G., Van de Ven, W. J., Creemers, J. W., and Devriendt, K. (2007). Identification and characterization of the TRIP8 and REEP3 genes on chromosome 10q21.3 as novel candidate genes for autism. *Eur. J. Hum. Genet.* **15**(4), 422–431.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., and Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383,** 407–413.

Chun, K., Teebi, A. S., Jung, J. H., Kennedy, S., Laframboise, R., Meschino, W. S., Nakabayashi, K., Scherer, S. W., Ray, P. N., and Teshima, I. (2002). Genetic analysis of patients with the Saethre-Chotzen phenotype. *Am. J. Med. Genet.* **110,** 136–143.

Clark, R. M., Marker, P. C., and Kingsley, D. M. (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.

Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibe, B., Bouix, J., Caiment, F., Elsen, J. M., Eychenne, F., Larzul, C., Laville, E., *et al.* (2006). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* **38,** 813–818.

Crackower, M. A., Scherer, S. W., Rommens, J. M., Hui, C. C., Poorkaj, P., Soder, S., Cobben, J. M., Hudgins, L., Evans, J. P., and Tsui, L. C. (1996). Characterization of the split hand/split foot malformation locus SHFM1 at 7q21.3-q22.1 and analysis of a candidate gene for its expression during limb development. *Hum. Mol. Genet.* **5,** 571–579.

Cremers, F. P., and Cremers, C. W. (2004). *In* "Inborn Errors of Development: The Molecular Basis of Clinical Disorders of Morphogenesis" (C. J. Epstein, R. P. Erickson, and A. Wynshaw-Boris, eds.). Oxford University Press, New York.

Crisponi, L., Deiana, M., Loi, A., Chiappe, F., Uda, M., Amati, P., Bisceglia, L., Zelante, L., Nagaraja, R., Porcu, S., Ristaldi, M. S., Marzella, R., *et al.* (2001). The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/ epicanthus inversus syndrome. *Nat. Genet.* **27,** 159–166.

Crisponi, L., Uda, M., Deiana, M., Loi, A., Nagaraja, R., Chiappe, F., Schlessinger, D., Cao, A., and Pilia, G. (2004). FOXL2 inactivation by a translocation 171 kb away: Analysis of 500 kb of chromosome 3 for candidate long-range regulatory sequences. *Genomics* **83,** 757–764.

Crolla, J. A., and van Heyningen, V. (2002). Frequent chromosome aberrations revealed by molecular cytogenetic studies in patients with aniridia. *Am. J. Hum. Genet.* **71,** 1138–1149.

David, D., Cardoso, J., Marques, B., Marques, R., Silva, E. D., Santos, H., and Boavida, M. G. (2003). Molecular characterization of a familial translocation implicates disruption of HDAC9 and possible position effect on TGFbeta2 in the pathogenesis of Peters' anomaly. *Genomics* **81,** 489–503.

Davies, A. F., Mirza, G., Flinter, F., and Ragoussis, J. (1999). An interstitial deletion of 6p24-p25 proximal to the FKHL7 locus and including AP-2alpha that affects anterior eye chamber development. *J. Med. Genet.* **36,** 708–710.

de Kok, Y., Merkx, G. F., van der Maarel, S. M., Huber, I., Malcolm, S., Ropers, H. H., and Cremers, F. P. (1995). A duplication/paracentric inversion associated with familial X-linked deafness (DFN3) suggests the presence of a regulatory element more than 400 kb upstream of the POU3F4 gene. *Hum. Mol. Genet.* **4,** 2145–2150.

de Kok, Y., Vossenaar, E. R., Cremers, C. W., Dahl, N., Laporte, J., Hu, L. J., Lacombe, D., Fischel-Ghodsian, N., Friedman, R. A., Parnes, L. S., Thorpe, P., Bitner-Glindzicz, M., *et al.* (1996). Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. *Hum. Mol. Genet.* **5,** 1229–1235.

de Laat, W., and Grosveld, F. (2003). Spatial organization of gene expression: The active chromatin hub. *Chromosome Res.* **11**(5), 447–459.

de la Fuente, L., and Helms, J. A. (2005). The fickle finger of fate. *J. Clin. Invest.* **15,** 833–836.

DiLeone, R. J., Marcus, G. A., Johnson, M. D., and Kingsley, D. M. (2000). Efficient studies of long-distance Bmp5 gene regulation using bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **97,** 1612–1617.

Dillon, N. (2003). Gene autonomy: Positions, please. *Nature* **425,** 457.

Dillon, N., and Sabbattini, P. (2000). Functional gene expression domains: Defining the functional unit of eukaryotic gene regulation. *Bioessays* **22,** 657–665.

Dlugaszewska, B., Silahtaroglu, A., Menzel, C., Kubart, S., Cohen, M., Mundlos, S., Tumer, Z., Kjaer, K., Friedrich, U., Ropers, H. H., Tommerup, N., Neitzel, H., *et al.* (2006). Breakpoints around the HOXD cluster result in various limb malformations. *J. Med. Genet.* **43,** 111–118.

Dorsett, D. (1999). Distant liaisons: Long-range enhancer-promoter interactions in *Drosophila*. *Curr. Opin. Genet. Dev.* **9,** 505–514.

Driscoll, M. C., Dobkin, C. S., and Alter, B. P. (1989). Gamma delta beta-thalassemia due to a de novo mutation deleting the 5′ beta-globin gene activation-region hypersensitive sites. *Proc. Natl. Acad. Sci. USA* **86,** 7470–7474.

Duboule, D., and Wilkins, A. S. (1998). The evolution of 'bricolage'. *Trends Genet.* **14,** 54–59.

Emison, E. S., McCallion, A. S., Kashuk, C. S., Bush, R. T., Grice, E., Lin, S., Portnoy, M. E., Cutler, D. J., Green, E. D., and Chakravarti, A. (2005). A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* **434**(7035), 857–863.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., and Jarvela, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30,** 233–237.

Fang, J., Dagenais, S. L., Erickson, R. P., Arlt, M. F., Glynn, M. W., Gorski, J. L., Seaver, L. H., and Glover, T. W. (2000). Mutations in FOXC2 (MFH-1), a forkhead family transcription factor, are responsible for the hereditary lymphedema-distichiasis syndrome. *Am. J. Hum. Genet.* **67,** 1382–1388.

Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., Danes, S., van Heyningen, V., and Hanson, I. (1995). Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. *Hum. Mol. Genet.* **4,** 415–422.

Felsenfeld, G., Burgess-Beusse, B., Farrell, C., Gaszner, M., Ghirlando, R., Huang, S., Jin, C., Litt, M., Magdinier, F., Mutskov, V., Nakatani, Y., Tagami, H., *et al.* (2004). Chromatin boundaries and chromatin domains. *Cold Spring Harb. Symp. Quant. Biol.* **69,** 245–250.

Fernandez, B. A., Siegel-Bartelt, J., Herbrick, J. A. S., Teshima, I., and Scherer, S. W. (2005). Holoprosencephaly and cleidocranial dysplasia in a patient due to two position-effect mutations: Case report and review of the literature. *Clin. Genet.* **68**(4), 349–359.

Flomen, R. H., Vatcheva, R., Gorman, P. A., Baptista, P. R., Groet, J., Barisic, I., Ligutic, I., and Nizetic, D. (1998). Construction and analysis of a sequence-ready map in 4q25: Rieger syndrome can be caused by haploinsufficiency of RIEG, but also by chromosome breaks approximately 90 kb upstream of this gene. *Genomics* **47,** 409–413.

Fraser, P. (2006). Transcriptional control thrown for a loop. *Curr. Opin. Genet. Dev.* **16**(5), 490–495.

Fukami, M., Okuyama, T., Yamamori, S., Nishimura, G., and Ogata, T. (2005). Microdeletion in the SHOX 3′ region associated with skeletal phenotypes of Langer mesomelic dysplasia in a 45,X/46, X,r(X) infant and Leri-Weill dyschondrosteosis in her 46,XX mother: Implication for the SHOX enhancer. *Am. J. Med. Genet. A* **137,** 72–76.

Gabellini, D., Green, M. R., and Tupler, R. (2002). Inappropriate gene activation in FSHD: A repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110,** 339–348.

Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M. A., Bottinelli, R., Green, M. R., and Tupler, R. (2006). Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* **439**(7079), 973–977.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51,** 975–985.

Gurnett, C. A., Bowcock, A. M., Dietz, F. R., Morcuende, J. A., Murray, J. C., and Dobbs, M. B. (2007). Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *Am. J. Med. Genet. A* **143,** 27–32.

Hadchouel, J., Carvajal, J. J., Daubas, P., Bajard, L., Chang, T., Rocancourt, D., Cox, D., Summerbell, D., Tajbakhsh, S., Rigby, P. W., and Buckingham, M. (2003). Analysis of a key regulatory region upstream of the Myf5 gene reveals multiple phases of myogenesis, orchestrated at each site by a combination of elements dispersed throughout the locus. *Development* **130,** 3415–3426.

Hatton, C. S., Wilkie, A. O., Drysdale, H. C., Wood, W. G., Vickers, M. A., Sharpe, J., Ayyub, H., Pretorius, I. M., Buckle, V. J., and Higgs, D. R. (1990). Alpha-thalassemia caused by a large (62 kb) deletion upstream of the human alpha globin gene cluster. *Blood* **76**(1), 221–227.

Hayes, C., Rump, A., Cadman, M. R., Harrison, M., Evans, E. P., Lyon, M. F., Morriss-Kay, G. M., Rosenthal, A., and Brown, S. D. (2001). A high-resolution genetic, physical, and comparative gene map of the doublefoot (Dbf) region of mouse chromosome 1 and the region of conserved synteny on human chromosome 2q35. *Genomics* **78,** 197–205.

Herault, Y., Fraudeau, N., Zakany, J., and Duboule, D. (1997). Ulnaless (Ul), a regulatory mutation inducing both loss-of-function and gain-of-function of posterior Hoxd genes. *Development* **124,** 3493–3500.

Herbert, A., Gerry, N. P., McQueen, M. B., Heid, I. M., Pfeufer, A., Illig, T., Wichmann, H. E., Meitinger, T., Hunter, D., Hu, F. B., Colditz, G., Hinney, A., *et al.* (2006). A common genetic variant is associated with adult and childhood obesity. *Science* **312**(5771), 279–283.

Heus, H. C., Hing, A., van Baren, M. J., Joosse, M., Breedveld, G. J., Wang, J. C., Burgess, A., Donnis-Keller, H., Berglund, C., Zguricas, J., Scherer, S. W., Rommens, J. M., *et al.* (1999). A physical and transcriptional map of the preaxial polydactyly locus on chromosome 7q36. *Genomics* **57,** 342–351.

Heydemann, A., Nguyen, L. C., and Crenshaw, E. B., III (2001). Regulatory regions from the Brn4 promoter direct LACZ expression to the developing forebrain and neural tube. *Brain Res. Dev. Brain Res.* **128,** 83–90.

Hill, R. E., Favor, J., Hogan, B. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D., and van Heyningen, V. (1991). Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature* **354,** 522–525.

Hill, R. E., Heaney, S. J., and Lettice, L. A. (2003). Sonic hedgehog: Restricted expression and limb dysmorphologies. *J. Anat.* **202,** 13–20.

Hill-Harfe, K. L., Kaplan, L., Stalker, H. J., Zori, R. T., Pop, R., Scherer, G., and Wallace, M. R. (2005). Fine mapping of chromosome 17 translocation breakpoints > or = 900 Kb upstream of SOX9 in acampomelic campomelic dysplasia and a mild, familial skeletal dysplasia. *Am. J. Hum. Genet.* **76**(4), 663–671.

Huber, C., Rosilio, M., Munnich, A., and Cormier-Daire, V. (2006). High incidence of SHOX anomalies in individuals with short stature. *J. Med. Genet.* **43**(9), 735–739.

Ianakiev, P., van Baren, M. J., Daly, M. J., Toledo, S. P., Cavalcanti, M. G., Neto, J. C., Silveira, E. L., Freire-Maia, A., Heutink, P., Kilpatrick, M. W., and Tsipouras, P. (2001). Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. *Am. J. Hum. Genet.* **68,** 38–45.

Ignatius, J., Knuutila, S., Scherer, S. W., Trask, B., and Kere, J. (1996). Split hand/split foot malformation, deafness, and mental retardation with a complex cytogenetic rearrangement involving 7q21.3. *J. Med. Genet.* **33,** 507–510.

Ishikawa-Brush, Y., Powell, J. F., Bolton, P., Miller, A. P., Francis, F., Willard, H. F., Lehrach, H., and Monaco, A. P. (1997). Autism and multiple exostoses associated with an X;8 translocation occurring within the GRPR gene and 3′ to the SDC2 gene. *Hum. Mol. Genet.* **6,** 1241–1250.

Jamieson, R. V., Perveen, R., Kerr, B., Carette, M., Yardley, J., Heon, E., Wirth, M. G., van Heyningen, V., Donnai, D., Munier, F., and Black, G. C. (2002). Domain disruption and mutation of the bZIP transcription factor, MAF, associated with cataract, ocular anterior segment dysgenesis and coloboma. *Hum. Mol. Genet.* **11,** 33–42.

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D. J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* **133,** 761–772.

Jiang, G., Yang, F., Van Overveld, P. G., Vedanarayanan, V., van der, M. S., and Ehrlich, M. (2003). Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum. Mol. Genet.* **12,** 2909–2921.

Karpen, G. H. (1994). Position-effect variegation and the new biology of heterochromatin. *Curr. Opin. Genet. Dev.* **4,** 281–291.

Kimura-Yoshida, C., Kitajima, K., Oda-Ishii, I., Tian, E., Suzuki, M., Yamamoto, M., Suzuki, T., Kobayashi, M., Aizawa, S., and Matsuo, I. (2004). Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification. *Development* **131,** 57–71.

Kioussis, D., Vanin, E., deLange, T., Flavell, R. A., and Grosveld, F. G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature* **306,** 662–666.

Kist, R., Schrewe, H., Balling, R., and Scherer, G. (2002). Conditional inactivation of Sox9: A mouse model for campomelic dysplasia. *Genesis* **32,** 121–123.

Kleinjan, D. A., Seawright, A., Elgar, G., and van Heyningen, V. (2002). Characterization of a novel gene adjacent to PAX6, revealing synteny conservation with functional significance. *Mamm. Genome* **13,** 102–107.

Kleinjan, D. A., Seawright, A., Schedl, A., Quinlan, R. A., Danes, S., and van Heyningen, V. (2001). Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6. *Hum. Mol. Genet.* **10,** 2049–2059.

Kleinjan, D. A., Seawright, A., Mella, S., Carr, C. B., Tyas, D. A., Simpson, T. I., Mason, J. O., Price, D. J., and van Heyningen, V. (2006). Long-range downstream enhancers are essential for Pax6 expression. *Dev. Biol.* **299**(2), 563–581.

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., Nowaczyk, M. J., Toriello, H., *et al.* (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila* melanogaster Nipped-B. *Nat. Genet.* **36,** 631–635.

Krebs, O., Schreiner, C. M., Scott, W. J., Jr, Bell, S. M., Robbins, D. J., Goetz, J. A., Alt, H., Hawes, N., Wolf, E., and Favor, J. (2003). Replicated anterior zeugopod (raz): A polydactylous mouse mutant with lowered Shh signaling in the limb bud. *Development* **130,** 6037–6047.

Lauderdale, J. D., Wilensky, J. S., Oliver, E. R., Walton, D. S., and Glaser, T. (2000). 3′ deletions cause aniridia by preventing PAX6 gene expression. *Proc. Natl. Acad. Sci. USA* **97,** 13755–13759.

Lee, J. A., Madrid, R. E., Sperle, K., Ritterson, C. M., Hobson, G. M., Garbern, J., Lupski, J. R., and Inoue, K. (2006). Spastic paraplegia type 2 associated with axonal neuropathy and apparent PLP1 position effect. *Ann. Neurol.* **59**(2), 398–403.

Lehmann, O. J., Sowden, J. C., Carlsson, P., Jordan, T., and Bhattacharya, S. S. (2003). Fox's in development and disease. *Trends Genet.* **19,** 339–344.

Leipoldt, M., Erdel, M., Bien-Willner, G. A., Smyk, M., Theurl, M., Yatsenko, S. A., Lupski, J. R., Lane, A. H., Shanske, A. L., Stankiewicz, P., and Scherer, G. (2007). Two novel translocation breakpoints upstream of SOX9 define borders of the proximal and distal breakpoint cluster region in campomelic dysplasia. *Clin. Genet.* **71**(1), 67–75.

Lemmers, R. J., Wohlgemuth, M., Frants, R. R., Padberg, G. W., Morava, E., and van der Maarel, S. M. (2004). Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **75**(6), 1124–1130.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12,** 1725–1735.

Lettice, L. A., Horikoshi, T., Heaney, S. J., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N., Oostra, B. A., Endo, N., *et al.* (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl. Acad. Sci. USA* **99,** 7548–7553.

Lewinsky, R. H., Jensen, T. G., Moller, J., Stensballe, A., Olsen, J., and Troelsen, J. T. (2005). T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. *Hum. Mol. Genet.* **14**(24), 3945–3953.

Loots, G. G., Kneissel, M., Keller, H., Baptist, M., Chang, J., Collette, N. M., Ovcharenko, D., Plajzer-Frick, I., and Rubin, E. M. (2005). Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Res.* **15**(7), 928–935.

Maas, R. L., Zeller, R., Woychik, R. P., Vogt, T. F., and Leder, P. (1990). Disruption of formin-encoding transcripts in two mutant limb deformity alleles. *Nature* **346,** 853–855.

Maas, S. A., and Fallon, J. F. (2005). Single base pair change in the long-range Sonic hedgehog limb-specific enhancer is a genetic basis for preaxial polydactyly. *Dev. Dyn.* **232,** 345–348.

Mackenzie, A., Miller, K. A., and Collinson, J. M. (2004). Is there a functional link between gene interdigitation and multi-species conservation of synteny blocks? *Bioessays* **26**(11), 1217–1224.

Marlin, S., Blanchard, S., Slim, R., Lacombe, D., Denoyelle, F., Alessandri, J. L., Calzolari, E., Drouin-Garraud, V., Ferraz, F. G., Fourmaintraux, A., Philip, N., Toublanc, J. E., and Petit, C. (1999). Townes-Brocks syndrome: detection of a SALL1 mutation hot spot and evidence for a position effect in one patient. *Hum. Mutat.* **14**(5), 377–386.

Masny, P. S., Bengtsson, U., Chung, S. A., Martin, J. H., van Engelen, B., van der Maarel, S. M., and Winokur, S. T. (2004). Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum. Mol. Genet.* **13**(17), 1857–1871.

Masuya, H., Sezutsu, H., Sakuraba, Y., Sagai, T., Hosoya, M., Kaneda, H., Miura, I., Kobayashi, K., Sumiyama, K., Shimizu, A., Nagano, J., Yokoyama, H., *et al.* (2007). A series of ENU-induced single-base substitutions in a long-range cis-element altering Sonic hedgehog expression in the developing mouse limb bud. *Genomics* **89,** 207–214.

Merlo, G. R., Paleari, L., Mantero, S., Genova, F., Beverdam, A., Palmisano, G. L., Barbieri, O., and Levi, G. (2002). Mouse model of split hand/foot malformation type I. *Genesis* **33,** 97–101.

Muncke, N., Wogatzky, B. S., Breuning, M., Sistermans, E. A., Endris, V., Ross, M., Vetrie, D., Catsman-Berrevoets, C. E., and Rappold, G. (2004). Position effect on PLP1 may cause a subset of Pelizaeus-Merzbacher disease symptoms. *J. Med. Genet.* **41**(12), e121.

Niedermaier, M., Schwabe, G. C., Fees, S., Helmrich, A., Brieske, N., Seemann, P., Hecht, J., Seitz, V., Stricker, S., Leschik, G., Schrock, E., Selby, P. B., *et al.* (2005). An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression. *J. Clin. Invest.* **115,** 900–909.

Nishimura, D. Y., Swiderski, R. E., Alward, W. L., Searby, C. C., Patil, S. R., Bennet, S. R., Kanis, A. B., Gastier, J. M., Stone, E. M., and Sheffield, V. C. (1998). The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nat. Genet.* **19,** 140–147.

Nobrega, M. A., Ovcharenko, I., Afzal, V., and Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. *Science* **302**(5644), 413.

Nobrega, M. A., Zhu, Y., Plajzer-Frick, I., Afzal, V., and Rubin, E. M. (2004). Megabase deletions of gene deserts result in viable mice. *Nature* **431**(7011), 988–993.

Olds, L. C., and Sibley, E. (2003). Lactase persistence DNA variant enhances lactase promoter activity *in vitro*: Functional role as a cis regulatory element. *Hum. Mol. Genet.* **12,** 2333–2340.

Ovcharenko, I., Loots, G. G., Nobrega, M. A., Hardison, R. C., Miller, W., and Stubbs, L. (2005). Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **15**(1), 137–145.

Paige, A. J., Taylor, K. J., Stewart, A., Sgouros, J. G., Gabra, H., Sellar, G. C., Smyth, J. F., Porteous, D. J., and Watson, J. E. (2000). A 700-kb physical map of a region of 16q23.2 homozygously deleted in multiple cancers and spanning the common fragile site FRA16D. *Cancer Res.* **60,** 1690–1697.

Pailhoux, E., Vigier, B., Chaffaux, S., Servel, N., Taourit, S., Furet, J. P., Fellous, M., Grosclaude, F., Cribiu, E. P., Cotinot, C., and Vaiman, D. (2001). A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nat. Genet.* **29,** 453–458.

Pfeifer, D., Kist, R., Dewar, K., Devon, K., Lander, E. S., Birren, B., Korniszewski, L., Back, E., and Scherer, G. (1999). Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: Evidence for an extended control region. *Am. J. Hum. Genet.* **65,** 111–124.

Phippard, D., Boyd, Y., Reed, V., Fisher, G., Masson, W. K., Evans, E. P., Saunders, J. C., and Crenshaw, E. B., III (2000). The sex-linked fidget mutation abolishes Brn4/Pou3f4 gene expression in the embryonic inner ear. *Hum. Mol. Genet.* **9,** 79–85.

Pop, R., Conz, C., Lindenberg, K. S., Blesson, S., Schmalenberger, B., Briault, S., Pfeifer, D., and Scherer, G. (2004). Screening of the 1 Mb SOX9 5′ control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. *J. Med. Genet.* **41,** e47.

Qin, Y., Kong, L. K., Poirier, C., Truong, C., Overbeek, P. A., and Bishop, C. E. (2004). Long-range activation of Sox9 in Odd Sex (Ods) mice. *Hum. Mol. Genet.* **13,** 1213–1218.

Robledo, R. F., Rajan, L., Li, X., and Lufkin, T. (2002). The Dlx5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes Dev.* **16,** 1089–1101.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S. W., Tsui, L. C., and Muenke, M. (1996). Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat. Genet.* **14,** 357–360.

Roessler, E., Ward, D. E., Gaudenz, K., Belloni, E., Scherer, S. W., Donnai, D., Siegel-Bartelt, J., Tsui, L. C., and Muenke, M. (1997). Cytogenetic rearrangements involving the loss of the Sonic Hedgehog gene at 7q36 cause holoprosencephaly. *Hum. Genet.* **100,** 172–181.

Rollins, R. A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. *Mol. Cell. Biol.* **24,** 3100–3111.

Rollins, R. A., Morcillo, P., and Dorsett, D. (1999). Nipped-B, a *Drosophila* homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152,** 577–593.

Ros, M. A., Dahn, R. D., Fernandez-Teran, M., Rashka, K., Caruccio, N. C., Hasso, S. M., Bitgood, J. J., Lancman, J. J., and Fallon, J. F. (2003). The chick oligozeugodactyly (ozd) mutant lacks sonic hedgehog function in the limb. *Development* **130,** 527–537.

Sabherwal, N., Bangs, F., Roth, R., Weiss, B., Jantz, K., Tiecke, E., Hinkel, G. K., Spaich, C., Hauffa, B. P., van der Kamp, H., Kapeller, J., Tickle, C., and Rappold, G. (2007). Long-range conserved non-coding SHOX sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients. *Hum. Mol. Genet.* **16,** 210–222.

Sagai, T., Masuya, H., Tamura, M., Shimizu, K., Yada, Y., Wakana, S., Gondo, Y., Noda, T., and Shiroishi, T. (2004). Phylogenetic conservation of a limb-specific, cis-acting regulator of Sonic hedgehog (Shh). *Mamm. Genome* **15,** 23–34.

Schedl, A., Ross, A., Lee, M., Engelkamp, D., Rashbass, P., van Heyningen, V., and Hastie, N. D. (1996). Influence of PAX6 gene dosage on development: Overexpression causes severe eye abnormalities. *Cell* **86,** 71–82.

Scherer, S. W., Poorkaj, P., Massa, H., Soder, S., Allen, T., Nunes, M., Geshuri, D., Wong, E., Belloni, E., and Little, S. (1994). Physical mapping of the split hand/split foot locus on chromosome 7 and implication in syndromic ectrodactyly. *Hum. Mol. Genet.* **3,** 1345–1354.

Scherer, S. W., Cheung, J., MacDonald, J. R., Osborne, L. R., Nakabayashi, K., Herbrick, J. A., Carson, A. R., Parker-Katiraee, L., Skaug, J., Khaja, R., Zhang, J., *et al.* (2003). Human chromosome 7: DNA sequence and biology. *Science* **300**(5620), 767–772.

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

Spitz, F., Gonzalez, F., and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113,** 405–417.

Spitz, F., Gonzalez, F., Peichel, C., Vogt, T. F., Duboule, D., and Zakany, J. (2001). Large scale transgenic and cluster deletion analysis of the HoxD complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev.* **15,** 2209–2214.

Spitz, F., Montavon, T., Monso-Hinard, C., Morris, M., Ventruto, M. L., Antonarakis, S., Ventruto, V., and Duboule, D. (2002). A t(2;8) balanced translocation with breakpoints near the human HOXD complex causes mesomelic dysplasia and vertebral defects. *Genomics* **79,** 493–498.

Staehling-Hampton, K., Proll, S., Paeper, B. W., Zhao, L., Charmley, P., Brown, A., Gardner, J. C., Galas, D., Schatzman, R. C., Beighton, P., Papapoulos, S., Hamersma, H., *et al.* (2002). A 52-kb deletion in the SOST-MEOX1 intergenic region on 17q12-q21 is associated with van Buchem disease in the Dutch population. *Am. J. Med. Genet.* **110,** 144–152.

Tarchini, B., and Duboule, D. (2006). Control of Hoxd genes' collinearity during early limb development. *Dev. Cell* **10,** 93–103.

Tarchini, B., Duboule, D., and Kmita, M. (2006). Regulatory constraints in the evolution of the tetrapod limb anterior-posterior polarity. *Nature* **443,** 985–988.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., *et al.* (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**(1), 31–40.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10,** 1453–1465.

Ton, C. C., Hirvonen, H., Miwa, H., Weil, M. M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N. D., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., *et al.* (1991). Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell* **67,** 1059–1074.

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., and Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36,** 636–641.

Trembath, D. G., Semina, E. V., Jones, D. H., Patil, S. R., Qian, Q., Amendt, B. A., Russo, A. F., and Murray, J. C. (2004). Analysis of two translocation breakpoints and identification of a negative regulatory element in patients with Rieger's syndrome. *Birth Defects Res. Part A Clin. Mol. Teratol.* **70,** 82–91.

Troelsen, J. T., Olsen, J., Moller, J., and Sjostrom, H. (2003). An upstream polymorphism associated with lactase persistence has increased enhancer activity. *Gastroenterology* **125**(6), 1686–1694.

Tufarelli, C., Stanley, J. A., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G., and Higgs, D. R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* **34,** 157–165.

Tupler, R., Berardinelli, A., Barbierato, L., Frants, R., Hewitt, J. E., Lanzi, G., Maraschio, P., and Tiepolo, L. (1996). Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J. Med. Genet.* **33,** 366–370.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y., and Kondoh, H. (2003). Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Dev. Cell* **4,** 509–519.

Ueta, E., Nanba, E., and Naruse, I. (2002). Integration of a transposon into the Gli3 gene in the Pdn mouse. *Congenit Anom (Kyoto)* **42**(4), 318–322.

Valle, D. (2004). Genetics, individuality, and medicine in the 21st century. *Am. J. Hum. Genet.* **74,** 374–381.

van der Hoeven, F., Schimmang, T., Vortkamp, A., and Ruther, U. (1993). Molecular linkage of the morphogenetic mutation add and the zinc finger gene Gli3. *Mamm. Genome* **4,** 276–277.

van der Hoeven, F., Zakany, J., and Duboule, D. (1996). Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. *Cell* **85,** 1025–1035.

van Deutekom, J. C., Wijmenga, C., van Tienhoven, E. A., Gruter, A. M., Hewitt, J. E., Padberg, G. W., van Ommen, G. J., Hofker, M. H., and Frants, R. R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* **2,** 2037–2042.

van Heyningen, V., and Williamson, K. A. (2002). PAX6 in sensory development. *Hum. Mol. Genet.* **11,** 1161–1167.

Van Hul, W., Balemans, W., Van Hul, E., Dikkers, F. G., Obee, H., Stokroos, R. J., Hildering, P., Vanhoenacker, F., Van Camp, G., and Willems, P. J. (1998). Van Buchem disease (hyperostosis corticalis generalisata) maps to chromosome 17q12-q21. *Am. J. Hum. Genet.* **62,** 391–399.

Velagaleti, G. V., Bien-Willner, G. A., Northup, J. K., Lockhart, L. H., Hawkins, J. C., Jalal, S. M., Withers, M., Lupski, J. R., and Stankiewicz, P. (2005). Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. *Am. J. Hum. Genet.* **76**(4), 652–662.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., and Holt, R. A. (2001). The sequence of the human genome. *Science* **291**(5507), 1304–1351.

Viprakasit, V., Kidd, A. M., Ayyub, H., Horsley, S., Hughes, J., and Higgs, D. R. (2003). *De novo* deletion within the telomeric region flanking the human alpha globin locus as a cause of alpha thalassaemia. *Br. J. Haematol.* **20,** 867–875.

Viprakasit, V., Harteveld, C. L., Ayyub, H., Stanley, J. S., Giordano, P. C., Wood, W. G., and Higgs, D. R. (2006). A novel deletion causing alpha thalassemia clarifies the importance of the major human alpha globin regulatory element. *Blood* **107,** 3811–3812.

Vortkamp, A., Gessler, M., and Grzeschik, K. H. (1991). GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* **352,** 539–540.

Vortkamp, A., Franz, T., Gessler, M., and Grzeschik, K. H. (1992). Deletion of GLI3 supports the homology of the human Greig cephalopolysyndactyly syndrome (GCPS) and the mouse mutant extra toes (Xt). *Mamm. Genome* **3,** 461–463.

Vyas, P., Vickers, M. A., Picketts, D. J., and Higgs, D. R. (1995). Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. *Genomics* **29,** 679–689.

Wakui, K., Gregato, G., Ballif, B. C., Glotzbach, C. D., Bailey, K. A., Kuo, P. L., Sue, W. C., Sheffield, L. J., Irons, M., Gomez, E. G., Hecht, J. T., Potocki, L., *et al.* (2005). Construction of a natural panel of 11p11.2 deletions and further delineation of the critical region involved in Potocki-Shaffer syndrome. *Eur. J. Hum. Genet.* **13**(5), 528–540.

Wallis, D. E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E. H., Rommens, J., and Muenke, M. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. *Nat. Genet.* **22,** 196–198.

Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., Dauwerse, H. G., Gruter, A. M., Hofker, M. H., Moerer, P., and Williamson, R. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* **2,** 26–30.

Wild, A., Kalff-Suske, M., Vortkamp, A., Bornholdt, D., Konig, R., and Grzeschik, K. H. (1997). Point mutations in human GLI3 cause Greig syndrome. *Hum. Mol. Genet.* **6,** 1979–1984.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., *et al.* (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* **3**(1), e7.

Woychik, R. P., Generoso, W. M., Russell, L. B., Cain, K. T., Cacheiro, N. L., Bultman, S. J., Selby, P. B., Dickinson, M. E., Hogan, B. L., and Rutledge, J. C. (1990). Molecular and genetic characterization of a radiation-induced structural rearrangement in mouse chromosome 2 causing mutations at the limb deformity and agouti loci. *Proc. Natl. Acad. Sci. USA* **87,** 2588–2592.

Wunderle, V. M., Critcher, R., Hastie, N., Goodfellow, P. N., and Schedl, A. (1998). Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc. Natl. Acad. Sci. USA* **95,** 10649–10654.

Yan, J., Bi, W., and Lupski, J. R. (2007). Penetrance of craniofacial anomalies in mouse models of Smith-Magenis syndrome is modified by genomic sequence surrounding Rai1: Not all null alleles are alike. *Am. J. Hum. Genet.* **80,** 518–525.

Yang, Y., Guillot, P., Boyd, Y., Lyon, M. F., and McMahon, A. P. (1998). Evidence that preaxial polydactyly in the Doublefoot mutant is due to ectopic Indian Hedgehog signaling. *Development* **125,** 3123–3132.

Zakany, J., Kmita, M., and Duboule, D. (2004). A dual role for Hox genes in limb anterior-posterior asymmetry. *Science* **304,** 1669–1672.

Zeller, R., Haramis, A. G., Zuniga, A., McGuigan, C., Dono, R., Davidson, G., Chabanis, S., and Gibson, T. (1999). Formin defines a large family of morphoregulatory genes and functions in establishment of the polarising region. *Cell Tissue Res.* **296,** 85–93.

Zuniga, A., Michos, O., Spitz, F., Haramis, A. P., Panman, L., Galli, A., Vintersten, K., Klasen, C., Mansfield, W., Kuc, S., Duboule, D., Dono, R., *et al.* (2004). Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev.* **18,** 1553–1564.