MOLECULAR GENETIC STUDIES OF TWO FORMS OF X-LINKED HYPOHIDROTIC ECTODERMAL DYSPLASIAS

by

Judith Gault

A DISSERTATION

Presented to the Department of Molecular and Medical Genetics
and the Oregon Health Sciences University
School of Medicine
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

May 1995

APPROVED: ██████████████████████████

··   ·····················

(Professor in Charge of Thesis)

██████████████████████

··   ·····················

(Chairman, Graduate Council)

TABLE OF CONTENTS

Table of Contents continued

List of Illustrations

Introduction

Chapter 1

Chapter 2

Discussion

Acknowledgements

Throughout the years (two short and three long) numerous people have contributed significantly to my training as a graduate student. I would like to thank Jon Zonana for his patience, support, and thorough critique of my work. My appreciation goes to a favorite coworker Marilyn Jones for her advice and assistance. From the Litt lab I would like to thank Mike Litt, David Browne, and Thas Phromchotikul for their technical advice. All the people on the Molecular and Medical Genetics floor have been exceptionally generous in sharing knowledge, equipment, good food, and reagents. In particular I would like to thank Markus Grompe, Matt Thayer, Mike Boiling Prep Whitney, TK Hyatt, Jim Hejna, Eric Bronner, Todd Taylor, Shoko Hagen and the nicest guy around, Felix Munoz. I look forward to reading about all the successful results produced from this department.

Special thanks goes to Mt. Hood Meadows for the attitude adjustment which enabled me to put graduate school into proper perspective.

My gratitude goes to my family, especially Ted Gault, Jean Griffiths, Elizabeth Morningstar, my dad (Ed Gault) and my grandma (Majorie Morse) for their continuous love and support.

My thesis is dedicated to Karl Liedtke, my mom (Yvonne Kilbourne), and to the memory of my grandma, Irene Sandkulla, in recognition of my love, appreciation and admiration for them.

Abstract

Molecular genetic studies have been performed on two forms of X-linked hypohidrotic ectodermal dysplasias, EDA the classic form, and a rarer form which includes an immunodeficiency with hyper IgM. Hypohidrotic ectodermal dysplasia (EDA) is the most common form of ectodermal dysplasia. EDA is an X-linked disorder mapping to the Xq13.1 region. Efforts have been focused upon cloning the EDA gene using a positional cloning strategy. A mouse probe, pcos 169E/4, identifies one locus, DXCrc169, tightly linked to *tabby* in the mouse, and a second locus, DXS732E, in humans which is tightly linked to EDA. *Tabby* is the putative murine homolog to EDA. *Tabby* and EDA phenotypes are similar and the disorders map between homologous loci on syntenic regions of the human and mouse X chromosomes. The probe pcos 169E/4 and a homologous human genomic cosmid clone, ICRFc104HO41.43 detect a deletion in one EDA patient ED1015. In addition, mouse probe pcos 169E/4 is conserved in many species supporting the hypothesis that pcos 169E/4 represents an expressed sequence and is part of a candidate gene for EDA. In order to test this hypothesis, a subclone from human cosmid ICRFc104HO41.43, homologous to pcos 169E/4, was used as a probe to screen a human fetal brain cDNA library. Eight overlapping cDNA clones were isolated and characterized. Sequence from three cDNA clones and a genomic fragment from the putative 5' exon of the gene at the DXS732E locus, termed TED for Two Established Domains, was used in BLAST analyses and resulted in the identification of two domains of homology with the HUMMRNAC cDNA characterized by Li et al. 1993. The function of the TED gene has not been determined, however patient ED1015 is at a minimum deleted for the 3' exon of the TED gene. Seven additional deletion and translocation breakpoints identified in EDA patients do not disrupt the TED gene,

and the rearrangements map beyond its 3' end. Additional studies are necessary before the TED gene can be ruled out as a candidate for EDA.

Ectodermal dysplasias are a heterogeneous group of disorders with primary defects in ectodermal derivatives. Over 100 forms of ectodermal dysplasia have been described and classified according to mode of inheritance and phenotypic involvement of hair, teeth, nails and sweat glands. The grouping of EDs continues to evolve as new forms of ectodermal dysplasia are identified. Two families are presented with hypohidrotic ectodermal dysplasia similar to the EDA phenotype and an immunodeficiency with hyper IgM similar to the phenotype of X-linked immunodeficiency with hyper IgM (HIGM1). The pair of affected half-brothers from each of these families have different fathers and the same mother. Twenty-four polymorphic loci spanning the X chromosome, including loci linked to EDA and HIGM1, were analyzed for concordance in two pairs of maternally related half brothers. In family ED 1136, the half-brothers were discordant for loci tightly linked to EDA, and concordant at a locus tightly linked to HIGM1. In family ED 1082, the half-brothers were concordant at loci tightly linked to EDA but discordant at the HIGM1 locus (CD40L). If the mutations in the two families are allelic, then the disorder would be distinct from EDA and HIGM1 and would map to the regions where both pairs of half-brothers are concordant. Both pairs of half-brothers appear to be concordant at loci in the Xq21.1-q26.1 region and the Xq26.2-28 region.

Introduction

I   Human Ectodermal Dysplasias

The ectodermal dysplasias, genetic disorders with primary involvement of ectodermal derivatives, are both phenotypically and genetically heterogeneous. Over 100 ectodermal dysplasias, including those with malformation syndromes, have been identified and grouped according to mode of inheritance, and phenotypic involvement of various combinations of ectodermal derivatives (Freire-Maia and Pinheuro 1984). Classification of the ectodermal dysplasias continues to evolve as new cases are categorized and new forms of ectodermal dysplasias, EDs, continue to be discovered. A genetic etiology has been identified for most forms of ED based upon inheritance patterns. The most progress towards gene identification has been made for hypohidrotic ectodermal dysplasia, EDA, which maps to the Xq13.1 region. Forms of ED have been identified and mapped in the mouse, such as *naked, tabby, downless, sleek,* and *crinkled.* The genes causing most forms of ectodermal dysplasia have not been identified.

According to Freire-Maia and Pinheiro in 1984, 91% of EDs involve abnormalities of hair, 80% involve teeth abnormalities, 75% involve nail abnormalities, and 42% involve sweat gland abnormalities. Ectodermal dysplasias result from abnormal morphogenesis of ectodermal derivatives such as hair, teeth, nails and sweat glands. We are only beginning to understand the processes of hair, teeth, nail and sweat gland morphogenesis. Elucidation of both the developmental pathways involved in forming these cutaneous structures, and the etiologies of ectodermal dysplasias are reciprocal processes. Analysis and understanding of the genetic bases and pathophysiology of the ectodermal dysplasias can contribute greatly to our knowledge of the normal

1

organization, function, and regulation of the hair, nail, teeth and sweat gland developmental pathways.

II   Morphogenesis of Ectodermal Derivatives

a. Epithelial-mesenchymal interactions

Early in embryogenesis, after the cleavage stage where cells undergo rapid mitotic divisions, gastrulation occurs. During gastrulation, the cells rearrange to form three cell layers or germ layers. The germ layers include the outer ectoderm, the inner endoderm and the interstitial mesoderm. Each of these layers gives rise to different structures during the developmental stages to follow. The ectoderm is composed of epidermal and nervous system progenitor cells. The fate of the cells forming the ectodermal layer will be progressively determined through a series of cellular interactions. The complexities of induction pathways are not completely understood.

Induction is a common mechanism involved in cell differentiation and organogenesis. Induction is the determination of the developmental fate of one cell mass by another. Epithelio-mesenchymal interactions, involving epithelial sheets, the cells which develop into the tissues which line organs and adjacent mesenchyme, are known as secondary inductions. Skin is composed of both an epithelial component, the outer epidermis derived from ectoderm, and a mesenchymal component, the dermis derived from mesoderm. There are several fundamental steps to epithelial-mesenchymal interactions. The mesenchyme or dermis starts the induction process by sending a signal to the epidermis. There are three general mechanisms of cell-cell communication.

2

Communication can transpire through cell-cell contact, cell-matrix contact, and through diffusion of signal molecules. Any of these types of cell-cell communications could be used between the epidermis and the dermis. The message from the dermis is specific, determining the fate of the epidermis by directing its differentiation into one cutaneous structure. The epidermis must be competent to receive the signal and respond by both altering the expression of specific genes and by sending a signal back to the mesenchyme. The differentiation process is complex, requiring precise timing of intracellular and intercellular signaling pathways. The signaling pathways must coordinate the spatial arrangements, proliferation, death and movements of cells.

The continuous cellular interactions between the epidermis, and the dermis are essential for normal differentiation into a variety of cutaneous structures including hair, teeth, nails, sweat glands, and mammary glands in mammals. These cutaneous structures are malformed, reduced in number or missing in patients with ectodermal dysplasia. The process of induction is essential for initiating development of cutaneous structures. Absence of hair, teeth, nails and sweat glands may be the result of an error in the epitheliomesenchymal induction pathway in developing skin. Malformed structures could result from errors in morphologic signals throughout the development of these cutaneous structures.

b. Eccrine sweat gland morphogenesis

Karen Holbrook has summarized the development of human eccrine sweat glands in the book, Recent Advances in Ectodermal Dysplasias. The first visible sweat gland structure to form, the primary epidermal ridges, can be identified between the third and fourth month of gestation. Sweat ducts form by

3

growing from the epidermal ridges into the dermis between the third and fourth month of gestation on the volar surfaces of the digits, and in the fifth month on the rest of the body. Sweat glands are formed as a bulbous expansion of the duct between the fourth and fifth month of gestation on volar surfaces, and in the sixth month on other regions of the body. The sweat glands on the body do not complete their maturation process and become functional until after birth.

The hypohidrotic forms of ectodermal dysplasia result in diminished sweating frequently associated with a reduced number, or total absence of sweat glands. It is likely that a reduction in the number of sweat glands is the result of them never having developed. Blecher in 1986, upon examination of serial sections of volar and palmar surfaces on the paws of *tabby* mice, found no evidence of sweat gland formation. These results support the hypothesis that these structures failed to form. Failure of any step during induction, initiation of sweat gland formation, may result in reducing the number of sweat glands formed.

c. Hair follicle morphogenesis

There are four general stages in hair follicle formation defined by the histological formation of particular structures in human fetuses (Figure 1)(Muller et al. 1991). Hair follicle development occurs first on the scalp and face; trunk skin starts to form follicles around 80 days estimated gestational age (Salinas et al. 1988, p20). Induction processes takes place before these identifiable structures are formed. The first stage, hair bud formation, takes place at approximately 11 weeks, 77 days, of intrauterine life. The second stage which transforms the hair bud into the hair bulb takes place at 12-13 weeks of intrauterine life. The third stage takes place around the 15th week of intrauterine

4

life and is identified by the formation of the hair cone, keratinization of epithelial cells, and the rough formation of sebaceous glands. At the final stage, the sebaceous gland is differentiated and a hair is formed that comes to the skin surface at the 18th week of intrauterine life.

Many molecules known to be expressed during hair follicle formation have already been identified. Retinoic acid receptors (α,β,γ) are expressed in developing mouse vibrissa and hair follicles. Cytoplasmic retinoic acid binding protein (CRABP) is expressed in the hair follicle wall and is thought to sequester free retinoic acid (Hardy 1992). Retinoids interfere with hair follicle and vibrissa development by acting through mesenchymal cells. Recombinations with mouse retinoid-treated dermis and untreated epidermis, grafted to chicken, arrests hair follicle development and results in mucous gland development, suggesting retinoic acid response proteins may participate in sending or receiving the second dermal message. The second dermal message, transmitted during stage II in the developmental scheme presented earlier, is responsible for directing development towards one cutaneous structure. Retinoids disrupt a hair follicle signal, resulting in mucous gland formation (Hardy 1992).

Many growth factor families and their receptors are expressed during hair follicle development including, epidermal growth factor (EGF) and its receptor EGF-R, Notch (Weinmaster et al. 1991, and Kopan and Weintraub 1993), transforming growth factor alpha (TGFα), transforming growth factors beta 1-3 (TGFβ-1, TGFβ-2, TGFβ-3), bone morphogenetic protein-2 and 4, acidic fibroblast growth factor (aFGF), basic FGF, and insulin-like growth factor (Hardy 1992, du Cros 1993a and b). TGFα is involved in hair morphology as well as the pattern of hair follicle development. Disruption of TGFα in the mouse results in wavy hair and disoriented, misaligned hair follicles (Luetteke et al. 1993). BMP-4, EGF and its receptor, aFGF, and bFGF are implicated in induction of hair

5

follicles because they are expressed before the first developing hair follicle structure is visible. Injections of EGF, before and during hair follicle development, into newborn mice delayed or slowed hair follicle formation in the coat (du Cros 1993a). Injections of bFGF, before and during hair follicle morphogenesis, delayed formation of hair follicles surrounding the injection site (du Cros 1993a and b). These results support the hypothesis that EGF and bFGF may be involved in initiation and development of hair follicles.

Cell adhesion molecules such as E-cadherin, and P-cadherin in mouse (Hardy 1992) and tenascin, neural cell adhesion molecule and integrin in chicken (Jiang and Chuong 1991 and Chuong et al. 1993) are expressed in specific structures during hair development, but not during initiation of hair follicle formation (Hardy 1992). Hox 3.1, and Hoxc-8, are expressed during hair follicle morphogenesis, not during induction, but their function is unknown (Messenger et al. 1993; Hardy 1992).

Over expression of parathyroid hormone-related peptide (PTHrP) and keratinocyte growth factor in the skin of transgenic mice results in failure to develop ventral hair follicles (Wysolmerski et al. 1994 and Guo et al. 1993). *Tabby* mice lack ventral hair, so these molecules may interact in the same pathway as the *tabby* gene product. Despite the identification of multiple proteins expressed during hair follicle development, the function of these molecules, including the ones that may be involved in the induction of hair follicles, has not been determined.

Several abnormalities in hair are found in patients with an ectodermal dysplasia. Sparse or missing hair could result from a primary failure to induce hair follicle formation. Abnormal hair structure, and delayed growth could result from an error in mesenchymal influence, and hypopigmentation could result from the failure of melanocytes to synthesize pigment (Salinas et al. 1988, p18).

Molecules described here are good candidates for ectodermal dysplasias because they are expressed during hair follicle morphogenesis, and have been characterized as either growth factors, cell adhesion molecules, transcription factors or hormones . None of the molecules mentioned map to the X chromosome so they are not good candidates for the EDA gene. Structural proteins expressed during hair follicle development are the result of morphogenesis, and may not be involved in regulating morphogenesis.

d. Tooth morphogenesis

Humans have 20 primary teeth and 32 permanent teeth. Although each tooth is anatomically unique, teeth go through the same general developmental stages (Figure 2), (Avery 1994). Oral epithelium and underlying mesenchyme interact to induce the development of a bud, the first recognized structure in tooth formation. Buds are formed by proliferating oral epithelium which protrudes into the mesenchyme, forming a bud shaped structure. The bud stage occurs between the sixth and eight weeks in utero for developing primary teeth, at the fifth prenatal month for permanent incisors, at the tenth prenatal month for permanent premolars, at the fourth prenatal month for the first permanent molars and at 4 years for the second permanent molars (Avery 1994, p 73). Partial or total anodontia is thought to be the result of disruption at or before the bud stage in developing teeth. After the bud stage, the developing tooth enters the cap stage, a proliferative stage, when the oral epithelium continues to divide and a concave surface is formed surrounded by the dental papilla, a mesenchymal component. By the end of the cap stage three structures are formed, the enamel organ which later forms enamel, the dental papilla which later forms both the

7

dentin and the pulp, and the dental follicle which later forms supporting structures. During the bell stage, the enamel epithelial cells differentiate into ameloblasts, and the process of morphodifferentiation occurs, forming a crown outline of a differentiated adult-looking bell stage tooth. Cells from the prospective enamel continue to differentiate. Both the incremental deposition of predentin secreted by odontoblasts and the predentin mineralization to form dentin, are characteristic of the dentinogenesis stage. The incremental deposition of partially mineralized enamel matrix by ameloblasts and the continued mineralization to form enamel is known at the amelogenesis stage. After crown growth through continued dentin and enamel deposition and mineralization or calcification tooth development is completed.

Retinoid, fibronectin, tenascin and growth factors (EGF family members) are molecules that have been implicated in tooth development. BMP 4 is expressed during the cap and bell stages in the dental papilla (Heikinheimo, 1994). BMP 6 is expressed in dental lamina in the cap stage, odontoblasts and dental papilla during the bell stage (Heikinheimo, 1994). BMP 2 is present during odontoblast differentiation and dentin secretion. Heikinheimo, found bone morphogenetic protein, BMP 4 and 2 to be expressed in mandibular bone and the alveolar bone respectively. Notch is expressed in developing tooth buds, among other developing organs. It is thought to be involved in cell-fate decisions (Weinmaster et al. 1991).

III   Hypohidrotic (Anhidrotic) Ectodermal Dysplasia (EDA)

a. EDA Phenotype

Hypohidrotic (Anhidrotic) ectodermal dysplasia, EDA, (Mendelian Inheritance in Man (MIM) # 305100, 1995), results from abnormal morphogenesis of hair, teeth and eccrine sweat glands. It was first described by Thurnam in 1848 (MIM # 305100, 1995). EDA is synonymous with both Christ-Siemens-Touraine syndrome and hypohidrotic ectodermal dysplasia, HED. It is the most common form of ectodermal dysplasia, however it is still is a rare disorder. The frequency of the disorder has not been accurately estimated. However, one estimate of the prevalence of EDA is 1/100,000 births (Stevenson and Kerr et al. 1967). EDA exhibits an X-linked mode of inheritance which was first described by Darwin in 1875, and EDA was first recognized as an X-linked disorder by Thadini in 1921 (MIM #305100, 1995). Males who are hemizygous for the mutant EDA gene, fully manifest the EDA phenotype. EDA is associated with several widely recognized physical abnormalities. A cardinal feature of EDA is the reduction in the numbers or absence of eccrine sweat glands. In addition, scalp hair is sparse and thin. Primary and permanent dentition is absent or conically misshapen. Other features found in EDA patients include periorbital hyperpigmentation with wrinkling, a saddle-nose, short stature, and lacrimal gland hypoplasia (MIM #305100, 1995). Diagnosis of EDA at birth is difficult because the diagnostic features of the disease are subtle in infants. Infants normally lack teeth, hair, and don't sweat very much. Early diagnosis and preventative care is important, since affected individuals are susceptible to both severe infections and hyperthermia which may result in death (Clarke et al. 1987). In a study done by Clarke et al. in 1987, the mortality rated was estimated at 30% in infants with unrecognized EDA. Chronic upper respiratory tract infections may result from a reduced number of mucous glands in the pharynx, larynx, trachea, bronchi, and esophagus (Siegel et al. 1990). Females heterozygous for the mutant EDA gene, exhibit variable degrees of severity in the

9

EDA phenotype, presumably due to X-chromosome inactivation. In one study done by Clarke, et al. in 1987, 78% of women who were obligate carriers had some type of dental abnormality. In addition, many carrier females frequently have mosaic patterns hypohidrosis and lactation insufficiency (Happle and Frosch 1985)(Clarke et al. 1987, 1991).

b. EDA and X inactivation

X-chromosome inactivation, first hypothesized by Mary Lyon in 1961, is the means by which dosage compensation, or the regulation of the expression of X-linked genes, is achieved between mammalian females and males. X-chromosome inactivation is generally a random event resulting in inactivation of on average 50% of either the paternally or maternally inherited X chromosomes in a female individual. Presumably, the females manifesting the EDA phenotype as severely as affected males, have preferentially inactivated the normal X chromosome resulting in expression of the X chromosome with the mutant EDA gene on it. Skewing of X inactivation can be due to random chance, or to X-autosome translocations. Schmidt and Du Sart showed in 1992 that 77% of patients with balanced X-autosome translocations had inactive normal X chromosomes. Apparently, inactivation of the normal X chromosome alleviates the consequences of X inactivation spreading to an autosome. Inactivation of the X chromosome involved in the translocation may result in a functional autosomal monosomy, or the lack of complete X inactivation resulting in a partial functional X disomy. Cells which have an active normal X chromosome have either the derivative (X) inactivated or the derivative (autosome) inactivated, depending on the presence of the X inactivation center at Xq12-q13. So, these cells are

10

hypothesized to be selected against because they result in autosomal monosomy or an X disomy (Schmidt and Du Sart 1992).

c. Carrier detection

Diagnosis of EDA is based upon phenotypic findings, making the diagnosis of a carrier female difficult. The identification of closely linked markers to EDA is important for carrier detection and prenatal testing in families where there are at least two people known to have EDA (Zonana et al. 1989). When only one person in a family is affected, the case might be sporadic, so the risk to possible carriers can only be reduced if that person does not have the haplotype carrying the mutated EDA gene on it in the proband of that family. By determining the alleles at various loci linked to EDA, one can construct the haplotype which is segregating with the EDA phenotype. The prediction of carrier status based upon haplotype analysis is most accurate when female offspring of an obligate carrier are looked at. Daughters of affected males and women with affected sons, in addition to other affected relatives are obligate carriers. The most accurate way to determine carrier status is to look directly for the presence of a mutation. Carrier status of EDA has already been determined based upon direct mutation analysis (Zonana et al. in 1994). Carrier status was established by detecting a large genomic deletion, originally identified in a boy with EDA, in female relatives. Once the EDA gene and its mutations have been identified, the mutations can be used to screen potential carriers of the EDA mutation. This approach has been successful for carrier detection of the DF508 mutation which accounts for approximately 70% of the mutations causing cystic fibrosis. It is unlikely that one EDA mutation will be as common as 70% in affected individuals and carriers. The majority of disorders result from different

11

mutations. The identification and characterization of the EDA gene and its mutations will improve the accuracy of detecting both carriers and affected individuals in families where one affected individual has already been identified.

d. EDA and the *Tabby* Mouse

Tabby, a sex linked disorder in the mouse, was initially described by Falconer in 1953. Since the original *Ta* mutation, several other mutations arose at the same locus, 41 cM from the X centromere (Sundberg 1994). Based upon the syntenic relationship of genes flanking the EDA and *tabby* loci, and their similar phenotypes, EDA and *tabby* are reputed homologous disorders. The *tabby* and EDA loci are flanked by the androgen receptor proximally, and the PGK-1 gene distally (Keer 1990, Zonana et al. 1992). This syntenic region is fairly large (Figure 3). The *tabby* phenotype is very similar to EDA in humans. *Tabby* mice, Ta/Y males, lack sweat glands and dermal ridges on the volar and plantar surfaces of their paws (Blecher 1986). The hair of the *tabby* mouse is also abnormal, consisting of only one abnormally formed type of fur, the awl hairs, out of the four types of pelage hairs normally found on mice (Blecher et al. 1983). The mice have focal and ventral alopecia, or lack hair behind their ears, and absence of hair on their ventral surface. The tails, which completely lack hair and hair follicles, are frequently kinked at the tip. The number of secondary vibrissa, whiskers, are extraordinarily variable in *tabby* mice. The molars and incisors of the *tabby* mouse erupt late and are generally small or absent (Blecher et al. 1983). Many exocrine glands involved in secreting components of lacrimal fluid (tears) function abnormally or are absent in the *tabby* mouse. The glands affected include the meibomian glands, which contribute to conjunctival sac fluid and secrete fluid during the final stages of eyelid separation; the lacrimal gland,

which is likely to be involved in conjunctival sac formation; and the harderian glands found behind the eyeball (Gruneberg 1966, Kapalanga and Blecher 1991). In addition to the similar phenotype to EDA, the eyelids of *tabby* mice are delayed in opening, and the mice have narrow palpebral fissures.

The *tabby* mouse was named after the black striped coat of females heterozygous for the *tabby* mutation, which resembles the coat of a tabby cat (Sundberg 1994, p 455). It is interesting to note that these stripes are not produced by wild-type and *tabby*-type hair bands. *Tabby* and wild type hairs are found in both stripes, however there is more of the normal zigzag type of hairs in the light stripes (Kindred 1967). The $Ta/+$ skin is made up of mosaic patches of *Ta* and wild type dermis and epidermis, but this does not affect follicle density on the body or tail (Kindred 1967). One explanation for this result is that the normal *tabby* gene product diffuses to the X^{Ta} patches and is able to signal hair follicle formation. Scalp hair is not in mosaic patterns in female carriers of EDA.

Biochemically, *tabby* and *crinkled* mice are reported to have high sulfhydryl:disulfide (SH:S-S) ratios in ectodermally derived components of skin and hair in addition to elevated sulfhydryl levels in their tooth germs (Weeks and Blecher 1983). Sulfhydryl and disulfide groups were measured by quantitating fluorescent emission of N-(7-dimethylamino-4-methylcoumarinyl)-maleimide stain of tissue sections for sulfhydryl groups. High SH;S-S ratios may affect coenzyme catalysis, tertiary structure formation in proteins such as keratins, a major epidermal protein. This biochemical feature, the elevated SH:S-S ratio, was first described as being characteristic of patients with Clouston's hidrotic ectodermal dysplasia in humans (Gold and Scriver 1971). It is not clear whether this biochemical finding is characteristic of other forms of ectodermal dysplasias, such as EDA, because studies addressing this question have not been done. The significance of these biochemical findings will become clear as the

tabby/EDA genes and their function are characterized. In cultured mouse skin, it was shown that at a particular concentration of EGF in the medium, the SH;S-S ratio could be lowered, but at all the other levels of EGF tested in culture, the ratio was generally increased (Robertson and Blecher 1987). Therefore, the SH;S-S ratio is affected by the amount of EGF in the medium but the relation between EGF and thiols is variable.

The effects of EGF have been extensively studied in the *tabby* mouse. EGF maps to the 4q25-q27 region, and EGF receptor maps to chromosome 7. EGF is a 53 amino acid serum protein, a growth hormone, which is known to stimulate division in a wide variety of cells. Most of EGF is produced by the granular convoluted tubules, GCT, within the submandibular gland, a salivary gland. In 1983, Blecher et al. found that the proportion of GCT in the submandibular glands of *tabby* mice were reduced in number. Presumptive evidence suggested that EGF production may be deficient in the *tabby* mouse. In one laboratory it was shown that *tabby* mice, before the time sweat glands are formed, injected subcutaneously between the shoulders with EGF, developed some dermal ridges as well as some functional sweat glands (Blecher et al. 1990). EGF has been shown to promote eyelid opening and incisor eruption in mice, whereas the *tabby* mutation results in delayed eyelid opening and incisor eruption (Hoath 1986; Rhodes et al. 1987). In addition, neonatal *tabby* mice injected with EGF, had accelerated rates of both eyelid opening and the morphological changes associated with eyelid opening, including conjunctival sac formation and keratinization of the epidermal connection between the upper and lower eyelids. Conjunctival sac formation and eyelid separation may depend upon glandular development. The meibomian, lacrimal, and harderian glands are known to be malformed or absent in the *tabby* mouse. The hypothesis proposed in conclusion of all this EGF and *tabby* work is that the *Ta* mutation may interfere

with the role of EGF as it is involved in tooth bud, sweat gland, dermal ridge and eyelid development. It was suggested that the *Ta* gene may act by reducing the size of the submandibular gland and therefore cause a decrease in EGF production. Without direct evidence showing EGF levels are reduced in the *tabby* mouse this hypothesis remains speculative. More recent work by du Cros, 1993a, has shown that injection with EGF delays hair follicle development in mice, suggesting EGF can act similarly to *tabby*. Because EGF does not map to the *tabby* or EDA region it is not a candidate for these disorders. However, an EGF-related molecule mapping to the region would be an excellent candidate.

Tabby-25H (Ta^{25H}) is a radiation-induced X chromosome deletion spanning both the *tabby* and *testicular feminization* loci, *Ar*, in the mouse (Cattanach, et al. 1991). The deletion is a minimum of approximately 1.5 cM, the distance between the *Ar* and *Ta* loci, and a maximum of around 4 cM, the distance between the flanking loci, Zfx and Ccg-1, known to be present. The resulting phenotype described by Cattanach et al. in 1991 is complex. In hemizygous Ta^{25H} male mice, the phenotype includes both the *tabby* and *testicular feminization* phenotypes. Testicular feminization occurs when there is a defect in the androgen receptor. Mice chromosomally male, (40,XY), developed external female genitalia without developing internal female reproductive organs. Testes develop but were located in the abdomen. Unlike *Ta* males and females, both the Ta^{25H}/Y mice and the Ta^{25H}/+ mice were small, runted, and had reduced prenatal viability; some had exencephaly in utero. The Ta^{25H}/Y mice were sterile. The Ta^{25H}/+ females frequently exhibited abnormalities not found in either their Ta^{25H}/Y, siblings or in other *Ta* females. They occasionally had extra incisors, a shortening and doming of the skull, some with lateral distortions. Front or rear feet syndactyly involving the first or second digit was found. Polydactyly was seen on the rear feet. The cause of the unique

phenotype seen only in heterozygous, Ta^{25H}/+ females was not evident, and the only explanation offered was that X-chromosome inactivation might be involved. The chromosome with the Ta^{25H} deletion is subject to X-chromosome inactivation like normal X chromosomes are. The cause of the additional phenotype in general, not seen in testicular feminization or the *tabby* mouse, was attributed to the deletion of other genes in the region which have not yet been identified. In 1991, Brockdorff et al. identified two additional loci (DXCrc131 and DXCrc169) deleted for in the Ta^{25H} mouse. Eag I and Not I, single copy linking clones from an X chromosome linking library, were localized to the central span of the X chromosome through both somatic cell hybrids analysis containing known translocations and interspecific Mus spretus/Mus musculus domesticus backcross analysis segregating for the *Ta* locus (Brockdorff et al. 1990, 1991). Linking libraries are constructed so that DNA with CpG islands and the genes associated with them are enriched. It has been estimated, using sequence database information, that approximately 56% of genes in humans and 47% of genes in mice are associated with CpG islands (Antequera and Bird 1993). The DXCrc169 locus segregates with the *Ta* locus with no recombinants in over 100 informative meioses (Brockdorff et al 1991). The DXCrc169 locus was analyzed further because it was genetically inseparable from the *Ta* locus and therefore a good candidate for the *Ta* gene. A genomic mouse cosmid, pcos 169, was identified by the probe, pEM169A, in a library screen. The probe pEM169A, identifies the polymorphism at the DXCrc169 locus. A Pst I subclone from cosmid 169, pcos 169E/4, was found to be conserved in rabbit, dog, pig, cow, chicken, monkey and humans, suggesting the probe is an expressed sequence. In addition, the probe detected related sequences from autosomes and the mouse X chromosome, suggesting that this putative gene at the DXCrc169 locus is part of a multigene family.

e. Positional cloning

Positional cloning is a strategy employed in the isolation of a gene based upon its location in the genome. The hypothesis being tested when using a positional cloning strategy is that the underlying disease mechanism is genetic and the chromosome region containing the gene causing the phenotype will segregate with the disease. Mapping a disease gene to a chromosomal region limits the number of candidate genes one needs to consider for the disorder to those genes shown to map to the region. For example, without knowing where the EDA gene maps, one would have to consider all the genes expressed in developing hair follicles, sweat glands, and in developing teeth as possible candidates for the disorder. Positional cloning has been used in the efforts toward isolation of the EDA gene. EDA exhibits an X-linked mode of inheritance, simplifying the first step in positional cloning; mapping the gene to a particular chromosome. The process of linkage exclusion mapping of the X chromosome in the search for where the EDA gene maps was facilitated by the discovery of a patient with EDA and a translocation. In 1988, the significance of a fibroblast cell line, GM0705 or AnLy, established from a female with an X;autosome translocation, fully manifesting EDA in addition to having moderate mental retardation, was rediscovered (Zonana et al. 1988). This was the first EDA patient to be identified with an X-autosome balanced translocation. The hypothesis that the AnLy translocation breakpoint interrupts the EDA gene was first proposed by Dr. P. J. L. Cook in 1973 at the first Human Gene Mapping Workshop. In theory the X;autosome translocation results in the inactivation of the EDA allele on the normal X chromosome at the same time it disrupts the other allele of the EDA gene. Elaboration of this theory is discussed in the EDA

and X-chromosome inactivation section of this introduction. As a result of the translocation, neither the inactivated allele nor the disrupted allele is expressed, and the female fully manifests the EDA phenotype. The discovery of the AnLy translocation and the identification of the breakpoint in the Xq13.1 region quickly focused linkage analysis on this region. The next step in positional cloning is to close in on the disease locus using polymorphic markers for linkage analysis. Linkage analysis is a method used to determine the location of a locus relative to other loci with known locations on a genetic linkage map. Linkage analysis is a statistical method which determines the likelihood of two loci cosegregating through meiosis if the loci are linked versus if they are not genetically linked. When loci are linked, they segregate together through meiosis more often than would be expected from independent assortment. A linkage map can determine the physical order of linked loci, provided there is at least one recombination between the loci of interest. Linkage maps roughly correlate with physical distance between loci. 1 centiMorgan, cM, meaning one observes a recombination 1% of the time between the loci being tested, is approximately one million base pairs in humans (Hartley et al. 1984). Loci that are not genetically linked can be on the same chromosome if they are further than 50 cM apart. Confirmation that the EDA locus maps to the Xq13.1 region was shown with genetic mapping. Polymorphic marker loci in this region show linkage to the EDA locus in families with multiple members affected with EDA (Zonana et al. 1988, and 1992). Analysis of 10 marker loci in 41 families resulted in the identification of four loci, DXS159, DXS339, DXS732E, DXS348, that were tightly linked to the EDA locus at $\theta=0$ with Z_{max} ranging from 9.84 to 28.91 (Zonana et al. 1992). HPRT-deficient hamster fibroblasts were fused with AnLy fibroblasts to make a somatic cell hybrid cell line known to contain the derivative 9. This AnLy somatic cell hybrid cell line was used to show, by Southern analysis, that the

18

DXS732E locus is proximal to the AnLy breakpoint and that the DXS453 locus is distal to the AnLy breakpoint (Thomas et al 1993). The DXS732E locus was identified with the conserved mouse probe pcos169E/4 (see *tabby* mouse section for more details). The probe detects two polymorphic alleles (13.3 kb and 8.4 kb) with an Xba I digest in humans. One affected male, EDA 1015, was found to be deleted for the probe which identifies the DXS732E locus (Zonana et al. 1993). Identification of the first submicroscopic deletion in an EDA patient with the conserved probe pcos 169E/4 led to the hypothesis that the probe represented part of a candidate gene for EDA.

When using positional cloning it is important to constantly look for gene candidates, such as the genes known to map to the particular region, or genes known to be involved in relevant biological pathways. All the genes involved in hair, teeth and sweat gland morphogenesis are considered candidates for EDA unless they do not map to the EDA region at Xq13.1. To date none of the genes identified from cutaneous morphogenic pathways have mapped to the EDA region. However, these genes may be involved in one of the many other forms of ectodermal dysplasia. The rate limiting step in postitional cloning, unless of course the gene of interest has already been characterized, is the identification of expressed sequences from the region of interest. Closely linked flanking markers may be hundreds of kilobases to megabases apart. The most abundant size class of mammalian genes is between 2 and 10 kb in genomic length, the average length is 16.6 kb (Lewin 1994, p685 and 687). So, numerous genes may be found between two closely linked markers. In order to accurately determine the distance between the linked markers, a physical map of the region, generated from restriction mapping using pulsed-field gel electrophoresis, is necessary. Identification of closely linked flanking markers is ideal, because the markers delimit a region which contains the gene and enables one to build a

physical map of the region extending from both ends. These loci, DXS453, and DXS732E were used to screen an ICRF X chromosome enriched library, and several YACs (yeast artificial chromosomes) were identified which mapped back to the region (Thomas et al. 1993). Two YACs, clone 4757 and 4758, were shown to span the AnLy, GM0705, translocation breakpoint by use of situ hybridization. These YACs were used as probes to screen an ICRF flow-sorted X-chromosome cosmid library. Many cosmids were identified and shown to overlap, forming four noncontiguous groups. The cosmid groups were ordered through Southern analysis of DNA from both somatic cell hybrids containing various parts of the X chromosome including the derivative 9 chromosome of the AnLy translocation, and deletion patient 1015. A cosmid clone from group C was shown with in situ hybridization AnLy cells in metaphase to span the AnLy translocation (Thomas et al. 1993). Cosmids from each group were used to screen a panel of 80 unrelated males affected with EDA (Zonana et al. 1994). Three additional submicroscopic deletions were identified in EDA patients, and the extent of these deletions could be roughly estimated, based upon Southern analyses using cosmids from the EDA region (Figure 4 from Zonana et al. 1994). Independent physical mapping and identification of translocation breakpoints has roughly confirmed the limits of the critical region. Identifying and mapping translocation and deletion break points is extremely helpful in narrowing the critical region (Kere et al. 1993). Several other balanced translocations have been reported which have not been localized on a physical map but disrupt the Xq13.1 region based on cytogenetic findings (Turleau et al. 1989, Plougastel et al. 1992). The EDA critical region is still quite large, possibly around 300-400 kb. The EDA gene may span the entire critical region. The dystrophin gene spans approximately 2.3 Mb of sequence (Tennyson et al. 1995). Other more complex explanations must be invoked to explain how the EDA gene may not span the

entire EDA critical region, and these theories are explored in the discussion of this thesis.

f. Major thesis hypothesis

The conserved mouse clone, pcos 169E/4, identifies two loci, one human locus tightly linked to the EDA locus, DXS732E, and one mouse locus tightly linked to the *tabby* locus, *DXCrc169* (Zonana et al. 1992 and Brockdorff et al. 1991). One patient with EDA and no other abnormalities was found to be deleted for the homologous human genomic sequence normally identified by pcos 169E/4 using Southern analysis. The major hypothesis being tested in the first paper is that the clone pcos 169E/4 represents part of a gene which is partially deleted in at least one EDA patient, and is therefore a candidate gene for EDA.

IV   X-Linked Hypohidrotic Ectodermal Dysplasia and Immunodeficiency with Hyper IgM

a. Genetic heterogeneity

Of the over 100 types of ectodermal dysplasias, approximately half include hypohidrotic forms, resulting in decreased sweat production, according to Freire-Maia and Pinheiro. Of these the X-linked hypohidrotic forms of ectodermal dysplasia include focal dermal hypoplasia (FDH) syndrome (MIM#305600), dyskeratosis congenita, mapping to Xq28 (MIM#305000), EDA, Dermotrichic syndrome and Lenz-Passarge's dysplasia (Freire-Maia and Pinheiro 1988). It is unclear whether Camarena syndrome, another hypohidrotic form of ectodermal dysplasia, is X-linked or exhibits autosomal dominant inheritance. Focal dermal

hypoplasia has been shown to be an X-linked dominant disorder with in utero lethality in males, and females affected with apocrine gland anomalies, dysplastic nails, hypodontia and a number of other severe abnormalities. Dermotrichic syndrome is not listed in MIM and is extremely rare. Only one family including six affected men over two generations, suggesting X-linked recessive inheritance, has been reported with lack of hair, normal teeth, dystrophic fingernails, hypohidrosis among other abnormalities (Freire-Maia and Pinheiro 1988). Lenz-Passarge's dysplasia, not found in MIM, has been described in a few families, with men more severely affected than women. The phenotype includes hypohidrosis, hypodontia, precocious balding in men and nonspecific skin lesions. The phenotype is similar to EDA and one large family thought to have Lenz-Passarge's dysplasia by Freire-Maia, was diagnosed as having EDA by Settineri et al, 1976. Camarena syndrome has been described in one family and is not listed in MIM. The phenotype of this disorder includes thin brittle hair, anodontia, dysplastic nails, and anhidrosis. Descriptions of the disorders not listed in MIM, are from the book written by Freire-Maia and Pinheiro in 1984.

EDA is clearly an X-linked disorder; however, a rare autosomal recessive hypohidrotic form of ED, clinically identical to EDA, has been described supporting non-allelic genetic heterogeneity (Passarge et al. 1966 and 1977, Crump and Danks 1971). Autosomal recessive inheritance of an EDA phenotype has not been conclusively proven and remains controversial (Sybert 1989). Linkage studies of families with vertical transmission of EDA report no evidence of non-allelic heterogeneity (MacDermot et al. 1986, Kolvraa et al. 1986, Zonana et al. 1988 and 1992). However, families analyzed in the EDA linkage analyses have multiple affected individual in more than one generation and are consistent with X-linked patterns of inheritance. It would be unusual for an autosomal recessive disorder to have affected individuals in multiple generations. The

autosomal recessive form of EDA might affect more than one generation in a family with a high degree of inbreeding (Ellis and Ahmed 1993). Ellis and Ahmed describe a consanguineous family with apparent autosomal recessive inheritance affecting more than one generation. Without haplotype analysis, recessive inheritance cannot be proven in this family, due to variable expression of the EDA phenotype in carrier females. Families including fully manifesting females in addition to haplotype analysis of markers linked to the EDA locus to show that the phenotype is not segregating with a certain X chromosome haplotype, are particularly useful in identifying autosomal recessive forms of EDA. Markers from EDA linkage studies have been useful in haplotype analysis to determine carrier status and assess risks of EDA family members. In 1990, Goodship et al. described a family with an affected boy with phenotypic findings that were most consistent with EDA and three mildly affected women, his mom and two aunts, that were presumed to be carriers. Haplotype analysis of this family, using flanking polymorphic markers shown to be linked to EDA, DXS159 and PGK1, indicates that the disease is not segregating with the Xq13.1 region. Possible explanations invoke either non-allelic heterogeneity or misdiagnosis of the carrier females, meaning the affected boy's condition could be the result of a new mutation.

The study of mouse models of human genetic diseases frequently provides important clues to the genetics, and pathophysiology, and to effective treatment of the disorder. There are several mouse models that phenotypically mimic hypohidrotic ectodermal dysplasia including *tabby (Ta), downless (dl), sleek (Dl^{slk})*, and *crinkled (cr)*. *Downless* and *sleek* are allelic mutations in the mouse which map to chromosome 10 approximately 29 cM from the centromere. They map one cM from an area with a syntenic region to 10q22 in humans (Sundberg 1994, p 242). Affected mice are phenotypically indistinguishable from

23

the *tabby* mouse. *Downless* is a recessive mutation, whereas *sleek, Dl^{slk}* is a dominant mutation. Another autosomal recessive disorder with a similar phenotype, *crinkled*, has been mapped to chromosome 13, 17 cM from the centromere. This region is syntenic with the 1q43 region in humans (Sundberg 1994, p 222). Dietary copper supplements during gestation and lactation corrects both the decreased activity of superoxide dismutase, a copper containing enzyme in the liver, and the delay and lack of hair growth in the *crinkled* affected mice in a litter. Perhaps the *crinkled* defect is somehow linked to copper metabolism; mutations have not been found in superoxide dismutase.

dl and *cr* effects of hair follicle initiation were examined through a series of experiments involving grafting skin, before hair follicle development had started, composed of reciprocal recombinations between *dl* and *cr* mouse tail dermis and epidermis. Only epidermis homozygous for these mutations failed to produce hair follicles suggesting that the primary defect in *dl* and *cr* is likely to be in the epidermis (Sofaer 1973, Mayer et al. 1977). In 1984, Pennycuik and Raphael analyzed the effects of the *Ta* mutation on both hair follicle initiation and hair morphology. When a *Ta* epidermis and a wild-type dermis are combined, zigzag hairs fail to form. The results indicate that the *tabby* gene in the epidermis determines hair morphology. The effects of *Ta* on hair follicle density were not conclusive. It is not known whether *Ta* expression from the dermis or the epidermis changes hair follicle development (Pennycuik and Raphael 1984).

There are four genetically distinct disorders in the mouse that are phenotypically indistinguishable from *tabby*, the reputed mouse homologous disorder to EDA. It is possible that considerable locus heterogeneity exists for EDA in humans, with disorders homologous to each of the mouse mutations, *Ta, cr, dl, Dl^{slk}*.

Four families have been reported with many of the classic features of EDA in addition to an immunodeficiency with hyper IgM. It is not clear whether all four families share the same disorder because the severity of both the hypohidrotic ectodermal dysplasia and the immunodeficiency with hyper IgM varies. The mothers in three families are unaffected, with two affected sons that are maternally related half-brothers, supporting the hypothesis that the disorders in these three families are X-linked. Although there is clear evidence for genetic heterogeneity in hypohidrotic ectodermal dysplasia, there has not been any evidence for another locus on the X chromosome. Either the EDA locus is involved in causing the ectodermal dysplasia in these families, or some other hypohidrotic ectodermal dysplasia X-linked locus is involved.

b. EDA and Immunodeficiency

There are five classes of immunoglobulins, IgG, IgA, IgM, IgE and IgD. Serum levels of IgG, IgA, IgM and IgE are important indicators of primary and secondary immunodeficiencies. Humoral immunodeficiency, antibody deficiency disorders, are diagnosed when serum levels of IgG, IgA, or IgM are below the 95% confidence limits of age-matched controls and when the patients have suffered recurrent, severe infections. Immunologic defects have been found in a number of ectodermal dysplasia cases that have been associated with opportunistic infections (Brooks et al. 1994, Davis et al. 1967, Fulginjiti et al. 1967, Gatti et al. 1969, Lux et al. 1970, Jhaveri et al. 1975). Some EDA cases have been associated with different types of immunological defects (Clarke et al. 1987, Huntley and Ross 1981). EDA patients' increased susceptibility to respiratory infections is possibly due to decreased seromucous glands (DeJager 1965). In 1987, Clarke et al. sought to document the possibility of a primary

immunological defect in EDA patients as a possible cause of recurrent infections. Immunoglobulins E, A, M, and G (including subclasses of G) were measured in thirty-six boys affected with EDA. Nineteen out of twenty-eight individuals had raised IgE levels ranging from 109->1000 kU/l. Eight out of thirty-six had low IgA concentrations ranging from 0.025-0.79 g/l. IgM concentrations were high in three cases (2.39-2.79 g/l) and low in six (0.39-0.49 g/l). Ten patients had high IgG concentrations ranging from 14.5-48.2 g/l. The IgG subclasses, light chains (kappa and lambda) were normal except for one individual with reduced IgG3. The affected boys also had normal antibody titers to E. coli and C. albicans, except one boy did not have a titer for C. albicans. Cellular immunity was tested in some individuals. T cell number and subsets were normal in six out of eight boys indicating normal lymphoid immunity. Consistent primary immunological abnormalities were not found.

A case was reported of a boy with EDA having recurrent respiratory infections with transient hypogammaglobulinemia. Serum immunoglobulins were low at 8 months of age with defects identified in antibody production or cell-mediated immunity, but increased to normal levels at 11 months of age. His transient hypogammaglobulinemia could have been caused by either EDA or his mother's treatment with glucocorticoids during pregnancy (Huntley and Ross et al. 1981).

Immunodeficiency with hyper IgM is an unusual finding in patients with EDA. Therefore, if this disorder involves the EDA locus, an additional locus would probably be involved in causing an immunodeficiency with hyper IgM found in the three families with affected maternally related half-brothers with hypohidrotic ectodermal dysplasia. A sub-microscopic deletion affecting both an immunodeficiency locus and an ectodermal dysplasia locus could be the cause of the phenotype seen in these families.

c. Types of X-linked immunodeficiencies

There are seven well characterized types of genetic immunodeficiencies that map to the X chromosome, including chronic granulomatous disease at Xp21, properdin deficiency at Xp11, Weskit Aldrich syndrome at Xp11, X-linked severe combined immunodeficiency (XSCID) at Xq13.1, X-linked agammaglobulinemia (XLA) at Xq22, X-linked lymphoproliferative syndrome at Xq25, and X-linked hyper IgM at Xq28 (Conley 1994).

EDA and XSCID are both linked to the DXS159 locus. It is possible that a sub-microscopic deletion affecting both of these loci could cause the phenotype seen in the four families. The XSCID phenotype includes impairment of cellular and humoral immune function, causing severe infections and failure to thrive. Lymphopenia is mild, serum levels of IgG and IgA are low, and IgM levels are low to normal in range. XSCID is usually fatal unless treated with a bone marrow transplant (Puck et al. 1994).

X-linked hyper-IgM (HIGM1, MIM#308230) syndrome is another candidate locus to consider in the families with hypohidrotic ectodermal dysplasia and immunodeficiency with hyper-IgM. Unique mutations including three microdeletions in the human CD40 ligand gene have been identified in HIGM1 patients, proving that HIGM1 is caused by mutations in the CD40L gene. CD40L, expressed on T helper cells, and CD40, a B cell surface receptor, interact and activate B cells (Puck et al. 1994).

The affected half-brothers' immunodeficiency is consistent with the HIGM1 phenotype because they have recurrent infections, low serum IgA, IgE, and IgG with high or normal levels of hyper-IgM. Upper respiratory tract infections, (73%), frequently caused by pyogenic bacteria, and diarrhea, (58%), are frequently

27

found in primary HIGM (Notarangelo et al. 1992). Hypohidrotic ectodermal dysplasia is not a feature of HIGM1. In addition HIGM1 does not map near EDA. It is possible that this locus is involved as part of a submicroscopic deletion, or these families represent allelic heterogeneity at the HIGM1 locus.

The other types of immunodeficiencies are less obvious candidates for the hypohidrotic ectodermal dysplasia and immunodeficiencies seen in these families.

d. Minor thesis hypothesis

The hypothesis being tested in the second paper is whether either the EDA and X-linked SCID loci or the HIGM1 locus could be involved in causing the hypohidrotic ectodermal dysplasia and immunodeficiency with hyper IgM phenotype found in the maternally related half-brothers of two families.

Figure 1 Hair follicle morphogenesis

From Muller et al. 1991. A sketch of the visible morphological structures during the four general stages of hair follicle development, the estimated time these structure are formed in the skin of a developing human fetus.

STAGE I: 12th week of intrauterine life

Hair bud ──────────────

Accumulated mesenchymatous cells ──────
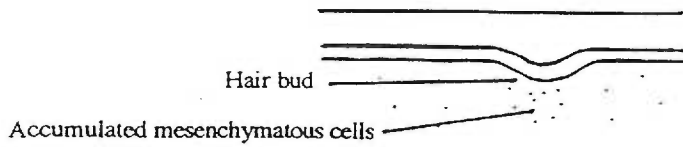
STAGE II: 14th week of intrauterine life

Hair bulb ──────────────

Mesenchymatous papilla ──────
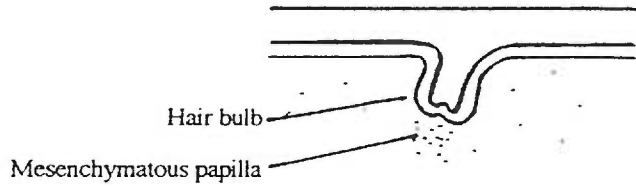
STAGE III: 16th week of intrauterine life

Roughly shaped sebaceous gland

Hair staff ──────────────────── Mesenchymatous papilla
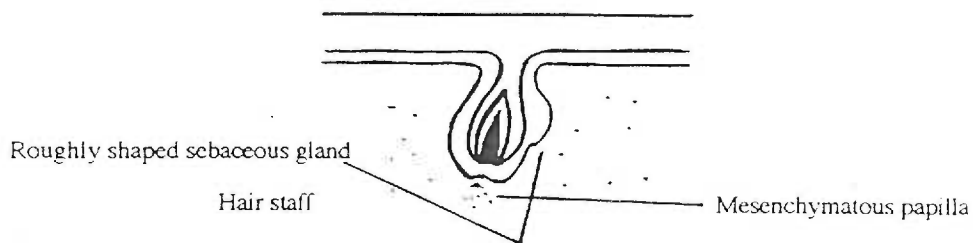
STAGE IV: 18th week of intrauterine life

Hair ──────────── Sebaceous gland

Connective tissue sheath ────── Bulb
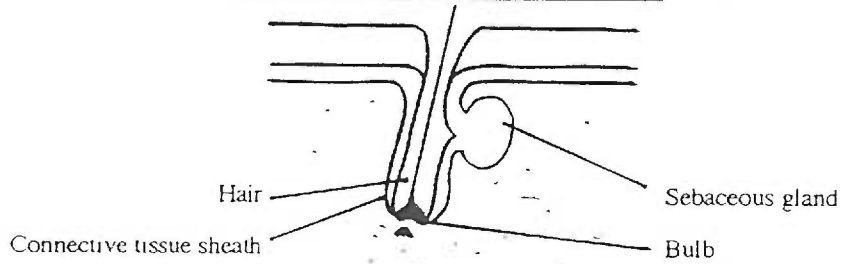
Figure 2 Tooth development

A diagram modified from the book <u>Oral Development and Histology</u> by James Avery depicting the stages of tooth development in humans (A-H).

Bud

Cap

Bell

H Enamel
Dentin
Pulp chamber

A

B

C

Dentinogenesis

D

G

F

E

Appositional
dentin and enamel

Amelogenesis
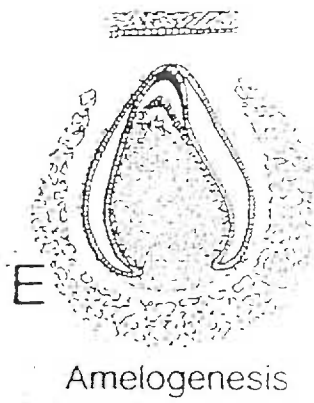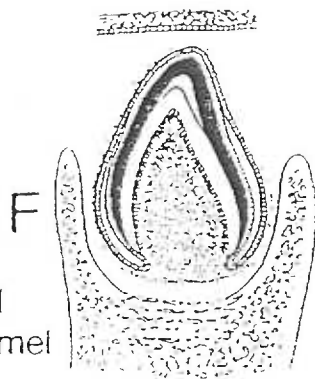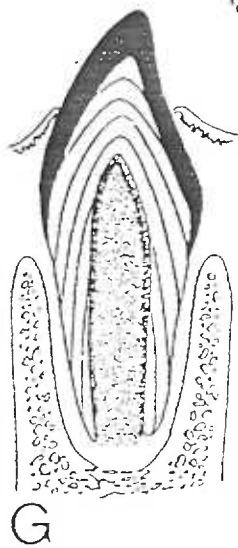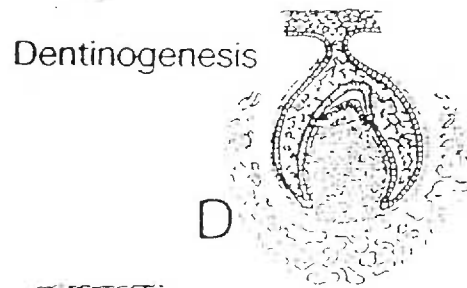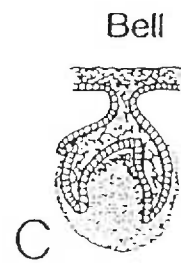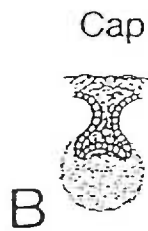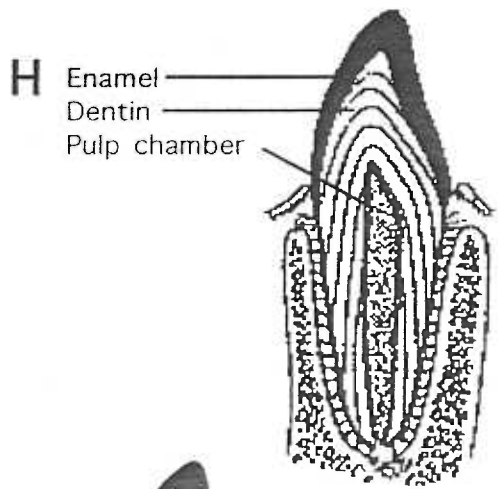
Figure 3 Syntenic regions of mouse and human X chromosomes

The human (left) and mouse (right) X chromosomes and their syntenic regions indicated by like patterns. Arrows indicate the *tabby* and EDA loci.

ZFX
DMD
CYBB

Cybb

Hprt

G6pd

AR
EDA ←
CCG1

Dmd

Zfx

Ar (Tfm)

Ta ←

PGK-1

Ccg1

Pgk-1

PLP

Plp

HPRT

G6PD
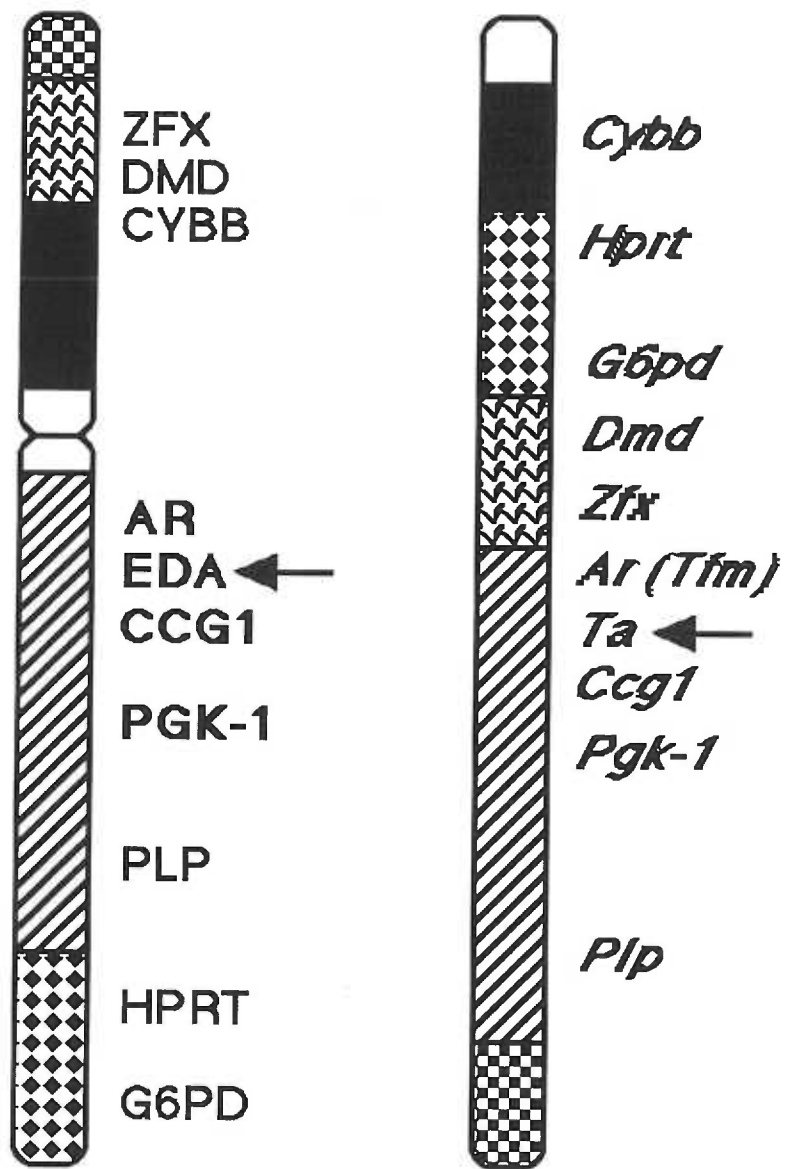
Figure 4 A map of the EDA region

From Zonana et al. 1994. Mapping of both cosmid clones, and EDA patient chromosomal aberrations within the EDA region. Cosmid clone C03.184 spans the AnLy translocation and is deleted in patient ED1015. Deletion breakpoints are indicated with a vertical bar, unknown breakpoints are indicated by a hatched bar.
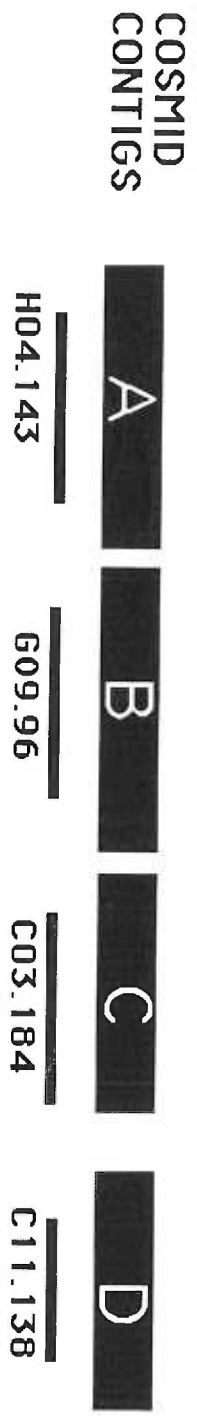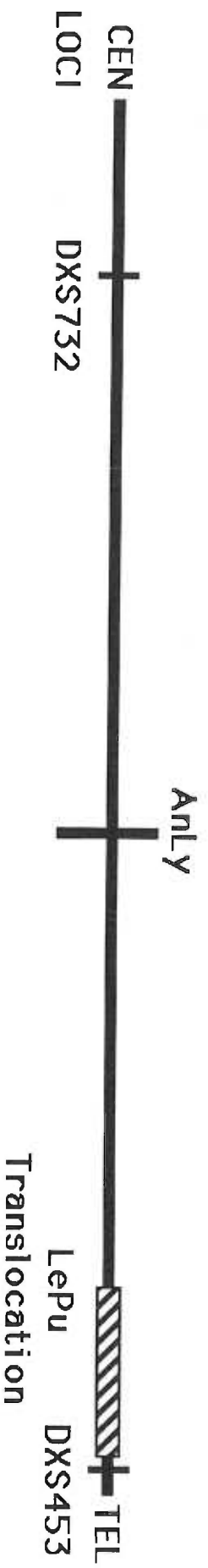
Isolation and characterization of a human X-linked gene mapping to the anhidrotic ectodermal dysplasia candidate region (Xq13.1)

Judith Gault*, Joan Zeltinger#, Neil Brockdorff^, Marilyn Jones*, and Jonathan Zonana*

*Depts of Molecular and Medical Genetics, Oregon Health Sciences University, Portland; ^Section of Comparative Biology, MRC Clinical Research Centre, Harrow (UK); #University of Washington, Seattle

Contact author:

> Jonathan Zonana, M.D.
> Dept. of Medical Genetics L-103
> Oregon Health Sciences University
> 3181 S.W. Sam Jackson Park Rd.
> Portland, OR 97201

Phone: (503) 494-4448

Fax: (503) 494-5738

Hypohidrotic ectodermal dysplasia (EDA) results in abnormal morphogenesis of hair, teeth and eccrine sweat glands. EDA is an X-linked disorder which maps to the Xq13.1 region. A positional cloning strategy towards isolating the EDA gene has been used and deletions in affected males and X-autosome translocations in fully manifesting females have been analyzed to define the EDA critical region. A mouse genomic clone, closely linked to the *tabby (Ta)* locus, the putative murine homologue to EDA, was used to identify a homologous human genomic clone, cosmid ICRFc104HO41.43, from the DXS732E locus. This human clone was subsequently used to isolate cDNAs from a human fetal brain library. Eight overlapping cDNAs were isolated. Sequence analysis, totaling 4038 base pairs, of three cDNAs and a genomic fragment containing the putative first exon, identified a 1422 base pair ORF. GRAIL analysis of the region supported the hypothesis that the putative ORF codes for a protein. A nucleic acid database search identified significant homology to two regions of the HUMMRNAC mRNA from a human frontal cortex brain library (Li et al. 1993). The gene, termed TED (Two Established Domains), at the DXS732E locus is expressed in human fetal kidney, intestine, brain, muscle, and in 77 day skin, a time when hair follicle formation begins. The TED gene, from the DXS732E locus, has been shown to be partially deleted in one EDA patient who does not appear to have any additional abnormalities. No large genomic rearrangements of the gene have been detected in 80 unrelated patients, including two females with X-autosome translocations and four additional males with submicroscopic molecular deletions. The additional deletions and translocations map

beyond the 3' end of the TED gene and it has not been determined if they interfere with TED gene expression. Until further analysis is performed, the TED gene remains a candidate for EDA.

INTRODUCTION

The ectodermal dysplasias are a group of genetic disorders that result from abnormal morphogenesis of some ectodermal derivatives including teeth, hair, and eccrine sweat glands. X-linked hypohidrotic ectodermal dysplasia (EDA) is the most common form of ectodermal dysplasia. The EDA phenotype includes sparse hair, missing and conical shaped teeth, and lack of sweat glands. Males that inherit this disorder exhibit full expression of the EDA phenotype whereas carrier females exhibit variable phenotypic expression, probably due to X-chromosome inactivation. One of the two copies of the EDA gene is inactivated in females and the severity of the phenotypic expression in carrier females depends upon the percentage of cells which inactivate the mutated gene in the relevant tissues. The EDA gene has been localized to the Xq13.1 region, using both genetic and physical mapping.

Tabby (Ta) is the putative mouse disorder homologous to EDA, based on phenotype, and the syntenic relationship between conserved genes flanking the EDA and *Ta* loci (Blecher 1986; Searle et al. 1987). The DXCrc169 locus is closely linked to the *tabby* locus with no recombinants in over 100 informative meioses (Brockdorff et al. 1991). The DXCrc169 locus is also deleted in the Ta^{25H} mouse. The Ta^{25H} mutation is a radiation induced deletion between 1.5 cM and 4 cM, which spans the *androgen receptor (AR)* and *tabby (Ta)* loci. In 1991, Brockdorff et al. described a highly conserved mouse genomic clone pcos169E/4, from the DXCrc169 region. Human genomic cosmid ICRFc104HO41.43, contains sequence homologous to pcos169E/4, and defines the DXS732E locus which is closely linked to EDA ($\theta_{max} = 0$, $Z_{max} = 21.2$) (Zonana et al. 1988). In addition, cosmid

ICRFc104HO41.43 is partially deleted in DNA from a patient, ED1015, with

EDA and no other apparent abnormalities (Zonana et al. 1993).

A positional cloning approach has been taken towards cloning the EDA gene

and transcripts are being sought within the EDA critical region. Mapping,

expression studies and the identification of a deletion in one patient, supports

the hypothesis that this conserved sequence represents a candidate gene at

the DXS732E locus. We identified and characterized a gene from the

DXS732E locus in order to investigate it as a candidate for the EDA gene.

RESULTS

Gene identification and mapping

Sequence analysis of a subclone, HO41.43-9, of human genomic cosmid

ICRFc104HO41.43 showed 82% identity, over an 800 bp region, with mouse

clone, pcos169E/4 (Figure 1). A 250 bp portion of this sequence, bold in

Figure 1, was rated as having excellent coding potential by GRAIL analysis.

Primers designed from this putative coding sequence were used in RTPCR

analysis and expression in human fetal brain was detected (Figure 1 and 2).

The human clone HO41.43-9 was used as a probe to screen a human fetal

brain library. Initially three unique clones, 3, 10, and 11 were identified. A

fragment from cDNA 11 was used to rescreen the library. A total of eight

unique overlapping cDNA clones have been identified that mapped back to

the Xq13.1 region (Figure 3). Four cDNAs were chimeric, indicated by a

failure of all of the EcoRI inserts to map back to the Xq13.1 region. Divergent

sequence identified in cDNA 11, failed to map to the X chromosome. This

clone was chimeric without the presence of an EcoRI site. Primers, not

shown, designed from cDNA sequence were used as probes to orient the

gene relative to the genomic map of cosmid ICRFc104HO41.43 (Figure 4).

The most 3' exon, and the putative 5' exon have been mapped as shown. However the position of the intervening exon(s) is not known. The cDNAs extend over approximately 15 kb of genomic sequence.

Sequence analysis and homology search

Three overlapping cDNA clones and one genomic clone containing the 5' untranslated region and the first exon, have been sequenced totaling 4038 base pairs. Intron/exon boundaries at the putative 5' exon and the most 3' exon have been identified based upon both sequence divergence between the cDNAs and genomic DNA, and splice site consensus sequence (Figure 5). The splice sites confirm that the clones are expressed sequences and not genomic contaminants in the library. The 5' end of the sequence maps within one of the two CpG islands mapped on cosmid ICRFc104HO41.43, recognized by the presence of multiple rare cutting restriction endonuclease sites (Figure 4). The cDNA sequence has a GC content of 58.4% overall, and is 69.8% GC with 9.2% occurence of the doublet CpG in the first 750 bp. At least six known Eag I sites, known to map within the upper 3 kb Pst I band, which is only detected in females, are hypermethylated on the inactive X chromosome. At least two of these six Eag I sites are hypomethylated on the active X chromosome producing the 306 bp band, the same size as the probe, seen in both males and females (Figure 6). The 3' end of the sequence contains a poly A addition site (ATAAA) 17 bps from a 33 bp poly A tail (Figure 7).

Both sequence and GRAIL analysis identified a 1468 base pair putative ORF (Figure 7). Using BLASTN, two segments of the putative ORF showed significant homology with the human HUMMRNAC (clone CTG-A4) mRNA sequence indicating potential domains of function (Figures 8 and 9). In

domain 1 there is 77% identity and 16% conservative substitutions over 43 amino acids. Domain 2 is 66% identical with 14% conservative substitutions over 118 amino acids. The domains are similarly spaced, having 19 and 13 amino acids 5' to domain 1, and having 172 and 197 amino acids between domains, in the genes at the DXS732E locus and the HUMMRNAC locus respectively. The HUMMRNAC Human (clone CTG-A4) cDNA, was cloned by Li et al. 1993 who were identifying cDNAs with triplet repeats from an adult human cerebral cortex cDNA library. Two short trinucleotide repeats coding for cysteine, (TGC), and proline, (GGC) have been identified in the TED gene. The HUMMRNAC gene contains a long interrupted CAG repeat coding for glutamine. The HUMMRNAC message is 5 kb in length and is expressed in cerebral and cerebellar cortex. The cDNA maps to chromosome 13 but the gene's function has not been determined. Southern analysis using a genomic Eag I fragment from cosmid ICRFc104HO41.43 as a probe containing domain 1, hybridized to the two bands from the DXS732E locus in addition to multiple related sequences not found at the DXS732E locus (Figure 6).

Expression studies

Expression analysis by Northern blot indicates that the gene at the DXS732E locus, is expressed in multiple tissues including human fetal brain, and adult heart, brain, kidney, and pancreas (Figure 10). The message is approximately 5 kilobases in length indicating that an additional one kilobase if cDNA has not been isolated or sequenced from the 5' untranslated region. RTPCR analyses (Figure 11) indicate the gene is expressed in 77 day skin, a time when hair follicle formation begins. The gene is not expressed in fibroblasts or lymphoblasts according to both Northern and RTPCR results.

Search for genomic rearrangements

Using Southern analysis, overlapping cDNA clones are either completely deleted (cDNA 11) or partially deleted (cDNAs 3 and 10) in patient (ED1015) who has EDA and no other apparent abnormalities (Figure 12). ED1015 is deleted for the most 3' exon of the TED gene including part of the putative ORF sequence. The centromeric deletion breakpoint lies somewhere in the first 1.1 kb of sequence as presented here, and could include domains 1 and 2. Expression studies of this gene in fibroblasts or lymphoblasts from patient ED1015 have not been possible since the gene is not expressed in these tissues. No deletions or rearrangements in these cDNAs have been identified in two female patients with X-autosome translocations or in four additional EDA patients with submicroscopic deletions. The translocation and additional deletions map beyond the 3' end of the gene in a telomeric direction (Figure 13). The position of the TED gene is based upon three findings. First, the TED gene lies centromeric to the AnLy translocation. Southern analysis of a somatic cell hybrid cell line known to contain the derivative 9 chromosome indicates TED is not on the translocated portion of the X chromosome. Second, patient ED1015 is deleted for the 3' end of the TED gene based upon Southern analysis. Third, there is a cosmid in contig C that spans the AnLy translocation according to in situ hybridization results, and is deleted in patient ED 1015. This cosmid is telomeric to cosmid ICRFc104HO41.43, the cosmid containing the TED gene. According to this data, the deletions and translocations map beyond the 3' end of the TED gene.

DISCUSSION

Assessment of the TED gene as an EDA candidate

There are several characteristics of the TED gene at the DXS732E locus described in this study which are consistent with it being a good candidate gene for EDA. The TED gene maps to the EDA critical region at Xq13.1. The TED gene at the DXS732E locus is conserved in mouse, which would be expected since *tabby* is the reputed homologous mouse disorder to EDA. It is likely that this gene is subject to X-chromosome inactivation because it is differentially methylated. At least five known Eag I sites are hypermethylated on the inactive X chromosome and at least two Eag I sites are hypomethylated on the active X chromosome. It has been hypothesized that the EDA gene will be subject to X chromosome inactivation because carrier females of EDA have variable expression of the EDA phenotype. It is likely, however, that several genes in the EDA region are subject to X inactivation, and conserved in the mouse. Genes from this region in humans probably map to the the *tabby* region in mice because of the syntenic relationship between the human and mouse X chromosomes in this region. Of greater significance, the TED gene at the DXS732E locus is expressed at the appropriate time in development, early during hair follicle development and during sweat gland morphogenesis (data not shown done by Joan Zeltinger). In addition, one EDA patient, ED1015, with no additional abnormalities, is deleted for the 3' end of this gene. The TED gene may be a less likely candidate gene for EDA because two translocation breakpoints and four additional submicroscopic deletions found in EDA patients, lie telomeric, beyond the 3' end of this gene. It is impossible to completely exclude the TED gene as the EDA gene at this time because the additional deletions and translocations may decrease expression of this gene resulting in EDA. The

41

POU3F4 gene, thought to be the gene causing X-linked mixed deafness, is not physically disrupted by four deletions mapping from 14 to 400 kb from the POU3F4 gene (de Kok et al. 1995). Mutations in the PAX6 gene are associated with aniridia. However, two families have chromosome deletions, segregating with aniridia, that do not disrupt the PAX6 gene, and map at least 85 kb away (Fantes et al. 1995).

Several theories have been proposed to explain how expression of the POU3F4 gene and the PAX6 gene may be affected by chromosome rearrangements that map beyond the genes. These theories explain why the TED gene is still an EDA candidate, even though seven chromosome aberrations identified in EDA patients do not physically disrupt the TED gene. An enhancer may be deleted or the aberrations may cause position effects, preventing normal expression of the gene. Deletions and translocations could bring a heterochromatic region close enough to the gene at the DXS732E locus to affect its expression as a result of heterochromatin spreading.

If mutations in the TED gene do not result in EDA, one would need to explain why patient ED 1015 does not appear to show a phenotype in addition to EDA. It is also possible that the deletion in patient ED1015 does not affect expression or function of this gene and its product, resulting in no phenotypic effect. Even if the deletion of the 3' end of this gene causes it to be nonfunctional it may not have a phenotypic effect. The lack of additional phenotypic finding, in this case, could be explained if the function of this gene is redundant and other genes are capable of acting in its place. Evidence in support of this hypothesis comes from the identification of two domains with a high degree of sequence homology and similar spacing with the HUMMRNAC gene (Accession # L10374). It is likely that the domain 1 probe from the TED

gene detects a HUMMRNAC gene band (Figure 6). The function of these two domains awaits further analysis but they may define a new multigene family. It is also possible that the deletion is deleterious, causing an abnormality that will become apparent in patient ED1015 at a latter date. Evidence supporting the hypothesis that mutations in the gene at the DXS732E locus may result in a phenotype other than EDA comes from analysis of the *tabby-25H* mouse. The mouse gene homologous to the TED gene has not been isolated. It is likely that at least part of the mouse homolog of the TED gene is deleted in the *Ta $25H$* mouse because the sequence normally detected by conserved genomic subclone pcos169E/4 is not present in the *Ta $25H$* mouse (Brockdorff et al. 1991). In addition to the *tabby* and *testicular feminization* phenotypes, the *Ta $25H$* phenotype includes increased neonatal and prenatal loss. Occasionally abnormalities including exencephaly (*Ta $25H$/Y* and *Ta $25H$/+*) were found. The carrier females (*Ta $25H$/+*) occasionally had additional abnormalities including, shortening and doming of the skull, extra incisors, syndactyly and polydactyly. The additional phenotype seen in the *Ta $25H$* mice maybe due to mutations in the mouse-homolog of the TED gene. The final proof of whether this is the EDA gene or not will come from mutation analysis of this and other genes in the EDA critical region.

MATERIALS AND METHODS

cDNA isolation A Stratagene human fetal brain cDNA library (18 weeks gestation), inserted into an EcoRI site of the Lambda ZAP II vector was screened by hybridization with a human genomic subclone, HO41.43-9, homologous to mouse clone pcos169 E/4. The standard Stratagene protocol was followed using Hybond-N, 0.45 Micron, nylon membranes from Amersham.

cDNA mapping Restriction maps of the cDNAs were constructed to first identify potential chimeric cDNA clones, detected upon EcoRI digestion. More detailed restrictions maps were constructed from EcoRI, BamHI, PstI, HindIII, single and double digests of DNA from the cDNA clones to determine their degree of overlap. Identified cDNA clones were mapped to the Xq13.1 region through Southern analysis of DNA from the cosmid ICRFc104HO41.43, an X-only human hamster hybrid cell line, and the patient ED1015 with a known deletion in the region. The map of cosmid ICRFc104HO41.43 (Nizetic et al. 1991) is based upon analysis of single and double digests of cosmid ICRFc104HO41.43, using Eag I, Sac II, Not I, Kpn I and Bam HI restriction enzymes. The approximate positions of cDNAs from the DXS732E locus were determined through Southern analysis of cosmid ICRFc104HO41.43 DNA using cDNA sequencing primers which were end-labeled as probes.

Sequence analysis The sequencing strategies employed for these clones involved exonuclease deletion analysis, primer walking, and cycle sequencing. Sequencing was done manually using Sequenase version 2 kits and a BRL double strand cycle sequencing kit. Exonuclease digestion was performed using standard methods outlined in the laboratory manual by Sambrook et al. 1989. The sequencing project was managed with IntelliGenetics computer program. The sequence was computer analyzed using GRAIL, Prosite, and BLAST programs (BLASTX, BLASTN, BLASTP, TBLASTN). GRAIL, Gene Recognition and Analysis Internet Link, recognizes coding potential of human sequence based upon human codon usage (Shah et al. 1994). BLAST, Basic Local Alignment Search Tool, programs search protein and nucleic acid databases for identity or homology to the query sequences (Altschul et al. 1990).

44

Methylation analysis Five Eag I restriction sites from the 5' region of this gene were identified from both genomic Eag I subclones from cosmid ICRFc104HO41.43, and cDNA sequence analysis. Methylation analysis of Eag I sites at the 5' end of this gene was done by Southern analysis using both Pst I and Eag I restriction digests of DNA extracted from control females and males. DNA digests were separated by gel electrophoresis on a 1.5 % agarose gel and alkaline transferred onto a hybond N+ nylon membrane. The probe, a 301 bp Eag I fragment from cosmid ICRFc104HO41.43, was isolated on a low melting point gel and oligo labelled.

Expression analysis Multiple tissue, poly A+, Clontech Northern blots from both fetal and adult tissues were hybridized using a random primed cDNA 11 probe and washed in 0.1% SDS, 0.1x SSC at 65°C for 15 minutes. (Isolation of human fetal RNA samples, preparation of Northern blots from the RNA, and hybridization of the blots to the relevant probes were performed by Dr. Joan Zeltinger, Post-doctoral fellow in the laboratory of Dr. Karen Holbrook at the Univ. of Washington). Northern blots on Hybond N nylon membranes from Amersham were made using the recommended procedures. Total RNAs were isolated from human fetal tissues using the method described by Puissant and Houdebine (1990). Polyadenylated RNA was isolated from the total RNAs using poly-AT-tract systems from Promega. For RTPCR analysis10 μg total RNA, was reverse transcribed with Superscript reverse transcriptase from GIBCO BRL. An aliquot was subseqently used in PCR with primers designed from sequence thought to be expressed.

Deletion patient Southern analysis DNAs from eighty unrelated patients with EDA and no other abnormalities were used in Southern analysis to detect

new deletions or rearrangements as described in Zonana et al. 1993. (This work was done by Research Associate Marilyn Jones.)

ACKNOWLEDGEMENTS

Figure 1 Sequence comparison of mouse genomic clone pcos 169E/4 and homologous human clone HO41.43-9

Comparison of homologous sequence between human genomic clone HO41.43-9, bottom sequence, and mouse genomic clone pcos 169E/4, upper underlined sequence. GRAIL rating of 250 bp region with coding potential in bold print. **CAG//** indicates the most 3' splice site shown in Figure 4, and **TGA** is the putative stop codon. * indicates homology between the mouse and human sequences. RTPCR human primers Z17F and Z17R are indicated with an arrow above the lower case human sequence.

```
AAGCTGGAGC-TCCACCGCGGTGCACCGGAAACCAAGTGCTGTGATGTGC
* * * * * *    *  * * * *        *  *     *  * * * * * * * * * * * * * * * * * * *   * * * *
CAG//GGTTGCTGGATACTTCACCAAAGCGT-CTGGAAACcaagtgctgtgacgtgc

AGTGGGTCTCCTGTGAGTCGGAGAAGAAGAAGTTCAAGGACTCTGAGCCC
* * * * * * * * * * * * * * * *    * *    * * * * * * * * * * * * * * * * * *    * * * * * *    * *
agtgGGTCTCCTGTGAGGCGAAGAAGAAGAAGTTCAAGGAGTCTGAGGCC

CCTAAAACCCACCACCAGCAATTCCACCACTCCTATTTTCACCACTACCA
* *    * * * * * * * * * *      * * * * * * * * * * * * * * * * * * * * * *    * * * * * * * * * * *
CCCAAAACCCACCAGCAGCAATTCCACCACTCCTATTTCCACCActacca

CCATCAGTACCACCACTACCATCCCCGCCATGAACCCCCAAGCCGAGTCA
* * *    * * * * * * * *    * * * * * * * *    * * * *      * * * * *    * * * * * *    * * * *    * * * *
ccaacagtaccatcactACCACCCCCATCATGATCCCCCAGGCCGTGTCA

GCAACAAGCCCTCCCTGCTGCCGGTCTCTGGGGGCTCCC--CTCAGTCCC
* * * * * * * * * * * .    * * * * * * * * * * * * * * * * * * * * * * * * * *      * * * * *    * * *
GCAACAAGCCCGCCCTGCTGCCGGTCTCTGGGGGCTCCCGCCTCAGCCCT

AGCAGGATCCGGCTCTGTGTCCTTGCTCTTATACTCCTCCATACAGTAGT
* * * * * * * * * * * * * * * * * *    * * * * * * *    * * *    * *    * * * * * * * * * * *    * *    * *
AGCAGGATCCGGCTCTGCGTCCTTGTTCTCATGCTCCTCCATACCGTGGT

GTCCTTCTCCAGCAGCCAGAGTGATGGGGGACTGGGACTGGAGACACTGC
* * * * * * * * * * * * *    * * * *    * * *    * * * * * * *    * * * *    * * * * * * * * * * * * *
GTCCTTCTCCAGCAACCAGGGTGGTGGGGGATTGGGGCTGGAGACACTGC

CTGCCCTCGAGGAGGGGTTGATGTAGGAAGAGTGACCACAGGGAGGGAGG
* * * * * * *    * * * * * * * *    * * *    * * * * * * * * * * *    * * * * * * * * * * *
CTGCCCTAGAGGAGGGCCTGACACGGGAAGAGTGACAGTAGGGAGGGAGG

ACAAACCTCCACCACATACTGACATCAGCTCCAGCCCAGCTCCAGGCTGG
* * *    * * * * * * * * * * * *    * * * * * * * * * * * * * * * * *    *    *    * * * * *    * * *
ACAGACCTCCACCACA--CTGACATCAGCTCCAGCTC--CCCCAGGTTGG

GGAAGGAA--GCATCTCCCACAGGAGGTGTAGGATCAGGTGGGGGTGGGG
* *    *      * *    * * * * * *    * * * * * * * * * * * * *    * * * * * * * * *    * * * *
GGGGGAGGGGGCTCCTCCCATGGGAGGTGTAGGATAAGGTGGGGGNGGGG

AGAAATGGGGGACCAG-ATATCGT-CCCAATCTACCCCTGCCTATGAAAG
 * *    * * * * * * *       *    * * *      * * * * *    * * *    * * *    * *      * * * *
NAAA-TGGGGGAATGACACATCCNNCCCAACCTANCCCCACCCCA-AAAG

CAGCTTCAACNAAATGAC-AGAGGGCCCCCAAGGAGCTATAAAACTCACC
* * * * *    * * * *    * * * *    *    * * * * * * * *    *    *    * * * * * *    * * * * * * * *    *
CAGCTCCAACAGAATGGCCAGAGGGCCTC-AGGGAGCTGCAAAACTCATC
```

CAGTAGGGAAAGGTAGGAGCAGCAGGTGTCCCCTCCCAAGC-CCCACCTT
*** **   ***** ********** *** ************ **** **
CAGGAGAAAAAGGCAGGAGCAGCAAGTGACCCCTCCCAAGCTCCCATCTA

TGGGGCTTAGTTCTNAAAAGGGGGA----GGGGCTGGAGTTGCCCACTCC
**********   *** ** ***   ********* ******* **
TGGGGCTTAGCAAAAAAAGGGAGGAAGTGGGGGCTGGATATGCCCACCCC

TGCCAAAAGCCCTGACCCCAGGGAAGGAGGCTGCTTGCTCAGCCTCAGCCA
*********************************** * ******   ***
TGCCAAAAGCCCTGACCCCAGGGAAGGAGGCTGCTCACCCAGCCTTGGCCC

TGCAGGGAATGTTGGGGGGCACAGAGGGGAGANNTCTNN

Figure 2 RTPCR results

RTPCR using primer set Z17F and Z17R (primer sequence shown in Figure 1). - lanes indicate negative controls to test for genomic DNA contamination. + lanes indicate reverse transcribed RNA from the tissues indicated. The expected product, 134 bp, is produced from 18 week EGA human fetal brain.

Figure 3 Alignment of cDNA clones

Alignment of eight cDNA clones, ▇▇▇, from the TED gene at the DXS732E locus. ▨▨▨ is chimeric portions of the cDNAs, ☐ is the genomic clone used as a probe, B= Bam HI sites, P= Pst I sites. Fragment sizes are indicated in kb.

Figure 4 Map of the TED gene on cosmid ICRFc104HO41.43

Distribution of the TED gene, ⬜➡ , over a map of genomic sequence contained in cosmid ICRFc104HO41.43.■ is TED gene coding sequence. The most 5', at least 910 bp, and 3', 2964 bp, exons have been mapped as shown; the intervening exon(s),▮ , totaling 169 bp, have not been localized within the eleven kb genomic fragment. K= Kpn I sites. The first CpG island contains at least six Eag I sites, two of which are also Not I sites, and at least three Sma I sites. The second CpG island contains at least one Eag I site and one Sac II site.

CpG island 1

CpG island 2

K   K   K   K

0   5   10   15   20   25   30   35   40   45kb
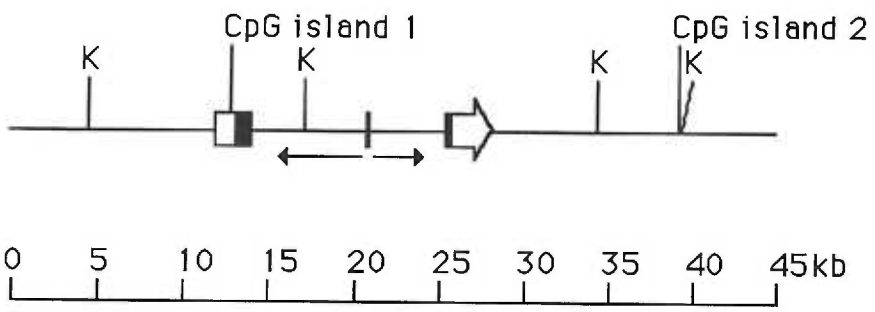
Figure 5 TED gene splice sites

Comparison of TED gene cDNA and genomic sequence at splice sites. Slice site consensus sequences listed below. A. The most 5' splice site identified at base 910. B. The most 3' spice site at base 1079. Additional splice sites may be found in the intervening 169 bp sequence. * indicates identity, / indicates predicted splicing cleavage site.

A.

```
cDNA seq    ...CTGTAAG/GCTGTCTACAAGGCCTGG...
                ******* *              * *
genomic     ...CTGTAAG/GTAAGGACTGGCTTCCGC...
                   ** *****
5' consensus      AG/GTAAGT
```

B.

```
cDNA seq    ...CCCTGGCTTTATCTGTACAG/GGTTG...
                ** *    * *    *   *********
genomic     ...CCTTCCTTATCCATTGGCAG/GGTTG...
                                  ***
3' consensus                     CAG/
```

Figure 6 Eag I digest of human genomic DNA

DNA from males and females as indicated, digested with Pst I and Eag I. The presence of the upper band, a 3 kb a Pst I fragment, indicates at least six Eag I sites are methylated. The lower 306 bp band indicates two of the Eag I sites are not methylated. The Southern blot was probed with the 306 bp Eag I genomic fragment.
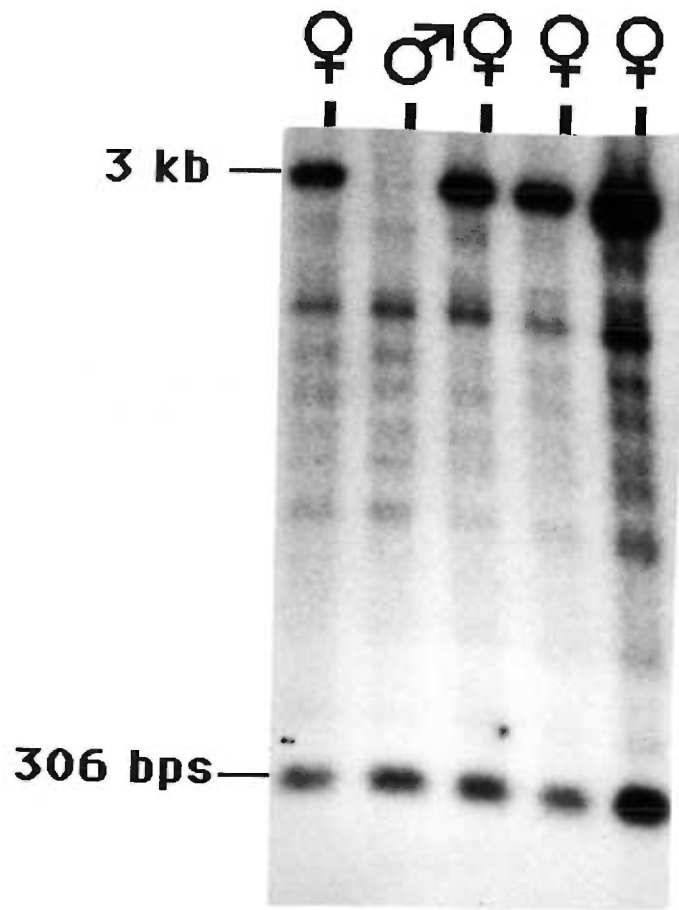
Figure 7 TED gene sequence

Sequence and predicted amino acid sequence from the TED gene at the DXS732E locus. **ATG's** are the two potential start codons. **/** indicate splice sites. **TGA** is the predicted stop codon. The **ATAAA** poly (A) signal sequence is indicated in bold. Underlined sequence indicates triplet repeats. **Domain 1** and **2** are in bold. **Ø** indicate the last base of an Eag I site.

↓

```
GTC TGG CTC GCG CCT GCC TGT TCC CTC CAG CCC GGA CCC CCC TGA AAT ATG TTC  54
                                                                  MET Phe

AGG GGC GCT TGG ATG TGG CCC GGG AAA GAC GCC GCC GCG CTG ACT ATC TGC TGC  108
Arg Gly Ala Trp MET Trp Pro Gly Lys Asp Ala Ala Ala Leu Thr Ile Cys Cys

TGC TGC TGC TGC TGG GCT CCC AGG CCG AGC GAC AAA CCT TGC GCC GAC TCC GAG  162
Cys Cys Cys Cys Trp Ala Pro Arg Pro Ser Asp Lys Pro Cys Ala Asp Ser Glu

CGG GCG CAG CGA TGG CGA CTG TCC CTG GCG TCC CTG CTC TTC TTC ACC GTG CTG  216
Arg Ala Gln Arg Trp Arg Leu Ser Leu Ala Ser Leu Leu Phe Phe Thr Val Leu

CTC GCT GAC CAT CTG TGG CTG TGC GCG GGG GCC CGG CCC CGG GCC AGG GAG CTG  270
Leu Ala Asp His Leu Trp Leu Cys Ala Gly Ala Arg Pro Arg Ala Arg Glu Leu
                                           ↓

AGC AGC GCC ATG CGG CCC CCA TGG GGG GCC GGC CGG GAG CGG CAG CCG GTG CCT  324
Ser Ser Ala MET Arg Pro Pro Trp Gly Ala Gly Arg Glu Arg Gln Pro Val Pro

CCT CGC GCG GTG CTG CCC GTG CCG CCG CCG CCG CCC GGC GAG CCC AGC GCG CCC  378
Pro Arg Ala Val Leu Pro Val Pro Pro Pro Pro Pro Gly Glu Pro Ser Ala Pro

CCA GGC ACC TGC GGC CCC AGA TAC AGC AAC CTG ACC AAA GCC GCC CCC GCC GCC  432
Pro Gly Thr Cys Gly Pro Arg Tyr Ser Asn Leu Thr Lys Ala Ala Pro Ala Ala
              ↓

GGC TCT CGG CCG GTC TGC GGC GGC GTC CCA GAG CCC ACG GGG CTG GAC GCA GCT  486
Gly Ser Arg Pro Val Cys Gly Gly Val Pro Glu Pro Thr Gly Leu Asp Ala Ala

TGC ACC AAA TTG CAA TCT TTG CAG AGA CTT TTC GAA CCG ACT ACT CCG GCC CCC  540
Cys Thr Lys Leu Gln Ser Leu Gln Arg Leu Phe Glu Pro Thr Thr Pro Ala Pro
                                                   ↓

CCT CTG CGG CCC CCT GAC TCC CCT TCC CGT GCC CCG GCC GAG TTC CCC TCC GCC  594
Pro Leu Arg Pro Pro Asp Ser Pro Ser Arg Ala Pro Ala Glu Phe Pro Ser Ala

AAA AAA AAC TTG CTC AAA GGC CAC TTT CGG AAC TTC ACT CTC TCC TTT TGC GAC  648
Lys Lys Asn Leu Leu Lys Gly His Phe Arg Asn Phe Thr Leu Ser Phe Cys Asp

ACC TAC ACG GTC TGG GAC TTG CTG CTG GGC ATG GAC CGC CCC GAC AGC CTG GAC  702
Thr Tyr Thr Val Trp Asp Leu Leu Leu Gly MET Asp Arg Pro Asp Ser Leu Asp

TGT AGC CTG GAC ACC CTG ATG GGG GAC CTG CTG GCC GTG GTG GCC AGC CCG GGC  756
Cys Ser Leu Asp Thr Leu MET Gly Asp Leu Leu Ala Val Val Ala Ser Pro Gly

TCC GGG GCC TGG GAG GCG TGT AGC AAC TGT ATC GAG GCG TAC CAG CGG CTG GAC  810
Ser Gly Ala Trp Glu Ala Cys Ser Asn Cys Ile Glu Ala Tyr Gln Arg Leu Asp

CGA CAC GCT CAG GAA AAA TAT GAC GAG TTC GAC CTC GTG CTG CAT AAA TAC TTA  864
Arg His Ala Gln Glu Lys Tyr Asp Glu Phe Asp Leu Val Leu His Lys Tyr Leu

CAG GCG GAA GAG TAC TCA ATC CGG TCC TGC ACG AAA GGC TGT AAG/GCT GTC TAC  918
Gln Ala Glu Glu Tyr Ser Ile Arg Ser Cys Thr Lys Gly Cys Lys Ala Val Tyr

AAG GCC TGG CTG TGC TCA GAA TAC TTC AGC GTG ACC CAG CAG GAA TGC AGC CGC  972
Lys Ala Trp Leu Cys Ser Glu Tyr Phe Ser Val Thr Gln Gln Glu Cys Gln Arg

TGG GTG CCC TGC AAG CAA TAC TGC CTG GAG GTG CAG ACC CGG TGC CCC TTT ATA  1026
Trp Val Pro Cys Lys Gln Tyr Cys Leu Glu Val Gln Thr Arg Cys Pro Phe Ile
```

CTC CCC GAC AAT GAG GAA ATG GTG TAC GGA GGG CTC CCT GGC TTT ATC TGT ACA   1080
Leu Pro Asp Asn Glu Glu MET Val Tyr Gly Gly Leu Pro Gly Phe Ile Cys Thr

G/GG TTG CTG GAT ACT TCA CCA AAG CGT CTG GAA ACC AAG TGC TGT GAC GTG CAG   1134
Gly Leu Leu Asp Thr Ser Pro Lys Arg Leu Glu Thr Lys Cys Cys Asp Val Gln

TGG GTC TCC TGT GAG GCG AAG AAG AAG AAG AAG TTC AAG GAG TCT GAG GCC CCC   1188
Trp Val Ser Cys Glu Ala Lys Lys Lys Lys Lys Phe Lys Glu Ser Glu Ala Pro

AAA ACC CAC CAG CAG CAA TTC CAC CAC TCC TAT TTC CAC CAC TAC CAC CAA CAG   1242
Lys Thr His Gln Gln Gln Phe His His Ser Tyr Phe His His Tyr His Gln Gln

TAC CAT CAC TAC CAC CCC CAT CAT GAT CCC CCA GGC CGT GTC AGC AAC AAG CCC   1296
Tyr His His Tyr His Pro His His Asp Pro Pro Gly Arg Val Ser Asn Lys Pro

GCC CTG CTG CCG GTC TCT GGG GGC TCC CGC CTC AGC CCT AGC AGG ATC CGG CTC   1350
Ala Leu Leu Pro Val Ser Gly Gly Ser Arg Leu Ser Pro Ser Arg Ile Arg Leu

TGC GTC CTT GTT CTC ATG CTC CTC CAT ACC GTG GTG TCC TTC TCC AGC AAC CAG   1404
Cys Val Leu Val Leu MET Leu Leu His Thr Val Val Ser Phe Ser Ser Asn Gln

GGT GGT GGG GGA TTG GGG CTG GAG ACA CTG CCT GCC CTA GAG GAG GGC CTG ACA   1458
Gly Gly Gly Gly Leu Gly Leu Glu Thr Leu Pro Ala Leu Glu Glu Gly Leu Thr

CGG GAA GAG **TGA** CAG TAG GGA GGG AGG ACA GAC CTC CAC CAC ACT GAC ATC AGC   1512
Arg Glu Glu   .

TCC AGC TCC CCC AGG TTG GGG GGG AGG GGC TCC TCC CAT GGG AGG TGT AGG ATA   1566
AGG TGG GGG CGG GGG AAA TGG GGG AAT GAC ACA TCC CCC CCA ACC TAC CCC CAC   1620
CCC AAA AGC AGC TCC AAC AGA ATG GCC AGA GGG CCT CAG GGA GCT GCA AAA CTC   1674
ATC CAG GAG AAA AAG GCA GGA GCA GCA AGT GAC CCC TCC CAA GCT CCC ATC TAT   1728
GGG GCT TAG CAA AAA AAG GGA GGA AGT GGG GGC TGG ATA TGC CCA CCC CTG CCA   1782
AAA GCC CTG ACC CCA GGG AAG GAG GCT GCT CAC CCA GCC TTG GCC CTG CAG GGA   1836
ATG TTG GGG GGC ACA GAG GGG AGA AGC TCT TTC CCC CAC TCC ACA TTC CTT TTG   1890
TTG CCA AGA TCC TTA ATT CCC TCC TGC CCA TAT CCC TAC AGG CGA CGG CAG ACA   1944
GTG CAA TGG CCC TCC TGC CAC TTC AGC ACA CCT TGC CCC ACC TGG GAC CAT CAC   1998
ACG TGA AGC ACC AGG CTG GGA GAT GAG GTG CAC ACA GTT GCA GCT AGG TCG GGG   2052
CCC CAG TTA AGC TGT GCC CCA CCA CCC TAG GCA ATG AGG GGC AGG AAA GGG GTA   2106
CAG AAT GAA TGG TGA AAG AGA GTG GAG ATA CGG AAG GGG GAG AGG AGA GGA GAA   2160
ATG TGG ACA GAG GGC TTG AAG ACA TGG CCG AAT AGG CTA CTG CCA CCC GGC TTT   2214
GGG GAG ATG GGC GAT AAG TGT TGA GGT AGG CTC GAG AAC TGC TCC CCA AAT CAG   2268
TGA GAA TTG CAT TAG GAG CCT CTG GGG AAA GAG CAC AGG AAC TGA GTT GCT GGG   2322
CCT CAA AAC AAA AGG CTG CAG CAC ATG TTA GAC TGC CGC TGT CTG AAG ACC ATA   2376
TCT TAC AGA AGG TGC ACA TAC TCC CAA GGG CCA CTC TTG CCC TGG AGA TCT TGG   2430
CCT TGG GGC AAA GAC ATT TCC ATT TCC GAT CTC CAT GGG GAG GAG GGA CTT AAG   2484
CCC AAG TGG GTC AGG CCT GGC CCA GGT CCT TAT ACT CCC CTA TCT TGT CCT AGC   2538
CCA GGC CTC CAT GAA GGA GAA CCT GGT GGC ATT GCC CAG AGC TCT CCA GAG ATG   2592
AGC CTC CCC TAT CCC CAC CCT GTC CCC GTC CCC AGC CTA CCA AAG GTA TTC ACA   2646
CCT TTC CAT CCC TGA CTC TAT GGG TTA CTG TCC CAA TTG CGA AAC ACC CAT ACT   2700
CCT CTT CAC CCT GTA GCA GGG TCT TCT GCT CCA GTC ACC CCG GCC TCT GCA AAG   2754
AGA AAG TCC AGG GAC TGG CAA GAA GAC CCT TTG ACT AAA GAT AAT GTG CTG TCA   2808
TAT CTT GGC TGA GAA ACT GTT GAG TCA GGC TAG AGG ACT GGA CAA GAG AAG AGG   2862
CAT TGA GAG CCC CCT TGG GAG CCA TTA GCC TTG CTC TGG TCT GCT GTG TGG GTT   2916
GGG GAG TAG GGG GGA TCC CAC ATT GCC TGG GAC ATC TGA ATG CTG ATA CGA ATG   2970
GAC AGG AAA GGG AGC CAG AGC TGT ATA CCT GGG ACA GGC CAG GGA GCC TTT AGT   3024
GCC AGC AGG GCC TGT GTC CTG GGA TGT GAT GCT AGT GAG TGA ATG GGA GGG CGG   3078
GCC TGG AGC TGC GTG AGA GTG AAA CCA AGA AGC GGA TGG GGA AGG GAG AGC TGG   3132
CCT TGG TCA GCT TAG GGA TGA ATG GAG AAG GAA TCA TTC ATG GGC TCG GGC TAG   3186

```
GTC CCA GAC TGA TTA CAT GAG ACT GGG GAG TCC TTC CCT GCC AGT TTT CTC ATT   3240
TAT TAC AGC CTA TCC CCC AAC CCT ACC CCT GGG CCC TGC CTA CAC ACT CCC GTT   3294
TGG CCC TCA GCA CTT TGG GCC TGG TCA CAT ACC CCA TGC CCT AGA GTG CTC TTT   3348
TCA GCT GGG GAG GGG AAG AGT GCT GGC TCC TTT CTA GAC ATG CTT ATG TAT AAG   3402
TAA AGG GGC AGG TGC TTG GGC TCC AGA GAG ACC CTA GGT ACT GCC TTC TAA CTC   3456
CTG CAT CCT CAG AAG GGG AAA AGG GAC TGG GAA GTA AAG GGA AAC CAT ATG GCC   3510
CCA GGG AAT GGA CCC TTG GAG ACT CCC ATC CTC TCT TCC CCC TCA CCA AGT GCC   3564
CAG GAG ACC CCT GGC TCA GAC CAG CCC AAG GCC TTG CCC AGT GGC AGG GGA AAG   3618
GGG CAC CTC ACT CCA CCT CAA AGT CAT TTG ACT GCC CCC ATC CAC CCC ACC ACC   3672
CTA ATC TTC CAC CAC CTC TCC GGC TGC CCC TGC CTC CCA TCA CAG ACA GCA GAG   3726
CCG GGC AGC TTT CTT ATG CCA TTT TCT ACA CTG TGC TTC ATG AGT AGG ACT TTC   3780
TTG CAC TAG TTC CTA TGA CTG AGT CTC CAA ACT GGT TTC CTA GTA GTC CCC CAT   3834
CCC TTC CTC CCT TAC CCA GCT ATG ATT CAG TTG TCT CTG CCC TCC CTC TTA CCC   3888
TGC CTC TGT GTT TCG GTG AGA GTC TCT GTA TAT GTA CTG CTT TCT GTT TTG TTG   3942
CAT TGT GGG GCA GAG GCC TCT CTG CTC TTG TTT AAC CCC **ATA AA**G AAA TAA ATG   3996
GTA AGA CTT GAT GAT AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA          4044
```

Figure 8 Alignment of the TED and HUMMRNAC cDNAs

■ indicates domain 1, ■ indicates domain 2. Both messages are approximately 5 kb in length.

triplet
repeats

TED GENE

MET   Stop   ATAAA

Domain 1

Domain 2

HUMMRNAC

MET   triplet   Stop
        repeat

(A)n
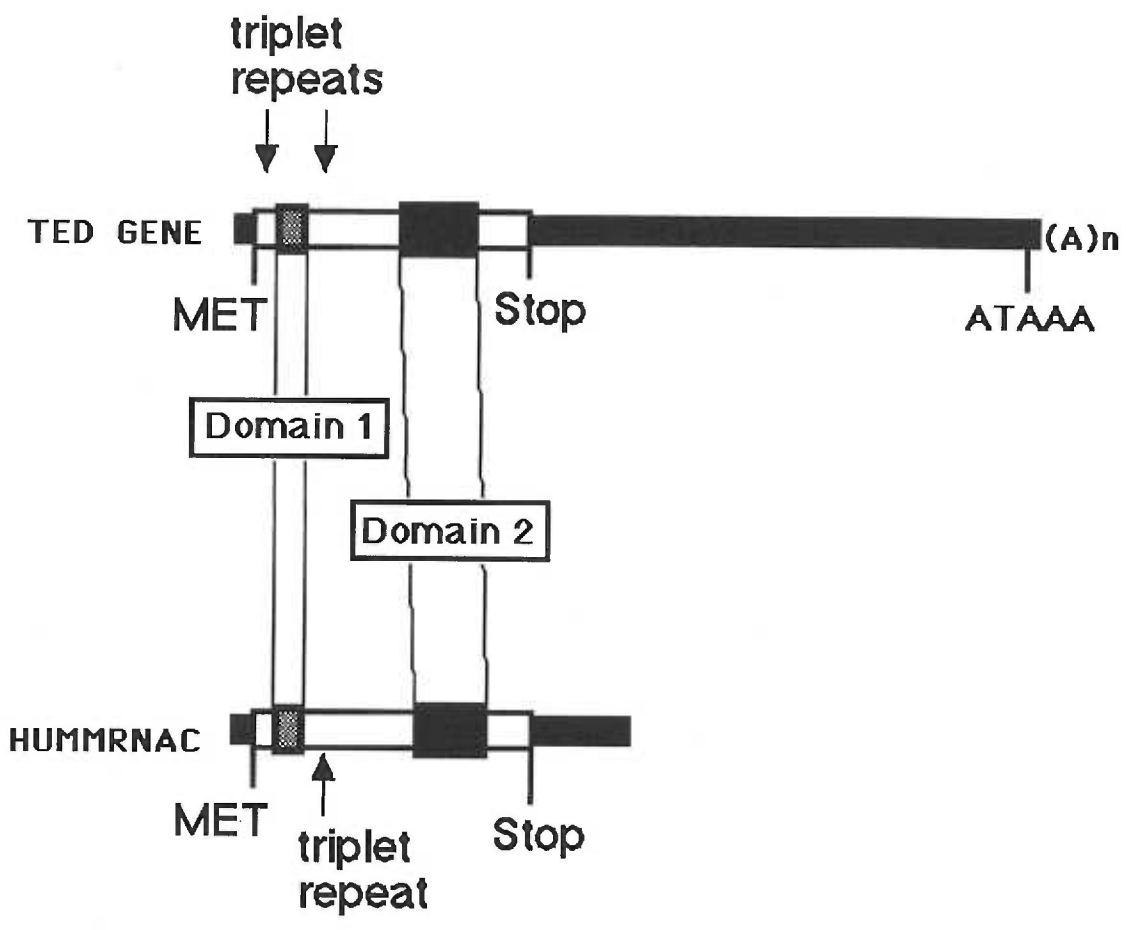
Figure 9 Nucleic acid and amino acid alignment of the domains of the TED gene and HUMMRNAC cDNA.

A. Domain 1 is 43 amino acids long with 77% identical amino acids, and 16% conservative substitutions. Nucleic acids are 72% identical over 93 bp. B. Domain 2 is 118 amino acids long with 66% identical amino acids, and 14% conservative substitutions. Nucleic acids are 68% identical over 354 bp. * indicates identity, + indicates conservative substitution.

A.

```
              123
TED gene    GCT CCC AGG CCG AGC GAC AAA CCT TGC GCC GAC TCC GAG CGG
            Ala Pro Arg Pro Ser Asp Lys Pro Cys Ala Asp Ser Glu Arg
             *   *   *           +   +   *   *       +   *   *   *   *
HUMMRNAC    GCA CCC CGA GAG AAC GAG AAA CCG TTC ATC GAT TCC GAG AGG
            Ala Pro Arg Glu Asn Glu Lys Pro Phe Ile Asp Ser Glu Arg

            GCG CAG CGA TGG CGA CTG TCC CTG GCG TCC CTG CTC TTC TTC ACC GTG CTG CTC
            Ala Gln Arg Trp Arg Leu Ser Leu Ala Ser Leu Leu Phe Phe Thr Val Leu Leu
             *   *   +   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
            GCT CAG AAA TGG CGA CTG TCT CTG GCA TCT CTC TTG TTT TTC ACA GTC CTG CTC
            Ala Gln Lys Trp Arg Leu Ser Leu Ala Ser Leu Leu Phe Phe Thr Val Leu Leu

            GCT GAC CAT CTG TGG CTG TGC GCG GGG GCC CGG
            Ala Asp His Leu Trp Leu Cys Ala Gly Ala Arg
             *   *   *   *   +   *   *       *   +   +
            TCT GAT CAC TTG TGG TTC TGC GCC GAG GCC AAG
            Ser Asp His Leu Trp Phe Cys Ala Glu Ala Lys
```

B.

```
              774
TED         CGT GTA GCA ACT GTA TCG AGG CGT ACC AGC GGC TGG ACC GAC ACG
            Cys Ser Asn Cys Ile Glu Ala Tyr Gln Arg Leu Asp Arg His Ala
             *       +   *   +   *   *   *   *           *   +   *   *
HUMMRNAC    TGC AGG CAG TGC GTC GAG GCT TAC CAG GAC TAT GAC CAC CAT GCT
            Cys Arg Gln Cys Val Glu Ala Tyr Gln Asp Tyr Asp His His Ala

            CTC AGG AAA AAT ATG ACG AGT TCG ACC TCG TGC TGC ATA AAT ACT TAC AGG CGG
            Gln Glu Lys Tyr Asp Glu Phe Asp Leu Val Leu His Lys Tyr Leu Gln Ala Glu
             *   *   *   *   +   *   *   +   *   *   *   *   *   *   *       *
            CAG GAG AAA TAC GAA GAG TTT GAA AGC GTG CTC CAC AAA TAT TTA CAG TCG GAG
            Gln Glu Lys Tyr Glu Glu Phe Glu Ser Val Leu His Lys Tyr Leu Gln Ser Glu

            AAG AGT ACT CAA TCC GGT CCT GCA CGA AAG GCT GTA AGG CTG TCT ACA AGG CCT
            Glu Tyr Ser Ile Arg Ser Cys Thr Lys Gly Cys Lys Ala Val Tyr Lys Ala Trp
             *   *   *   +   +   *   *           *   *   +   *   *   *   *   *
            GAG TAC TCG GTG AAA TCC TGT CCT GAA GAC TGT AAG ATT GTC TAC AAA GCC TGG
            Glu Tyr Ser Val Lys Ser Cys Pro Glu Asp Cys Lys Ile Val Tyr Lys Ala Trp

            GGC TGT GCT CAG AAT ACT TCA GCG TGA CCC AGC AGG AAT GCC AGC GCT GGG TGC
            Leu Cys Ser Glu Tyr Phe Ser Val Thr Gln Gln Glu Cys Gln Arg Trp Val Pro
             *   *   *       *   *       *   *   *       +   *       +   *   +   *
            CTC TGT TCC CAG TAT TTT GAA GTC ACA CAG TTT AAC TGC AGA AAG ACA ATT CCT
            Leu Cys Ser Gln Tyr Phe Glu Val Thr Gln Phe Asn Cys Arg Lys Thr Ile Pro

            CCT GCA AGC AAT ACT GCC TGG AGG TGC AGA CCC GGT GCC CCT TTA TAC TCC CCG
            Cys Lys Gln Tyr Cys Leu Glu Val Gln Thr Arg Cys Pro Phe Ile Leu Pro Asp
             *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
            TGC AAG CAA TAC TGT TTG GAG GTT CAG ACG AGG TGT CCA TTT ATA TTG CCC GAC
            Cys Lys Gln Tyr Cys Leu Glu Val Gln Thr Arg Cys Pro Phe Ile Leu Pro Asp

            ACA ATG AGG AAA TGG TGT ACG GAG GGC TCC CTG GCT TTA TCT GTA CAG GGT TGC
            Asn Glu Glu MET Val Tyr Gly Gly Leu Pro Gly Phe Ile Cys Thr Gly Leu Leu
             *   +   *   +   +   *   *   *   *       +   *   *   *   *   *   *   *
            AAT GAT GAA GTC ATC TAC GGA GGC CTC TCC AGT TTC ATC TGT ACA GGG CTT TAT
            Asn Asp Glu Val Ile Tyr Gly Gly Leu Ser Ser Phe Ile Cys Thr Gly Leu Tyr

            TGG ATA CTT CAC CAA AGC GTC TGG AAA CCA AGT GCT GTG
            Asp Thr Ser Pro Lys Arg Leu Glu Thr Lys Cys Cys Asp
             +   *   +               *           *   *   *
            GAA ACC TTT CTA ACC AAT GAT GAA CCA GAA TGC TGT GAC
            Glu Thr Phe Leu Thr Asn Asp Glu Pro Glu Cys Cys Asp
```

Figure 10 Result from probing Clontech multiple tissue Northern blots with cDNA 11

A. Human fetal tissues from 20-26 week EGA. Brain 5 kb band (1), lung (2), liver (3), kidney, (4). B. Human adult tissues. Heart 5 kb band (1), brain 5 kb band (2), placenta (3), lung (4), liver (5), skeletal muscle faint 5 kb band (6), kidney 5 kb band (7), pancreas 5 kb band (8).
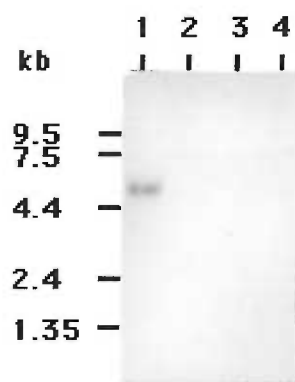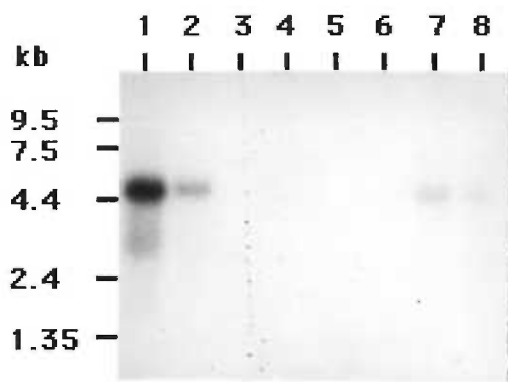
Figure 11 RTPCR analysis of TED gene in human fetal skin

RTPCR results using TED cDNA specific primers that span an intron. Primers Z-25 F (ACACCCTGATGGGGGACCTG) and Z-26 F (AAGTATCCAGCAACCCTGTAC). Lanes A, 77 day EGA human fetal skin. Lanes B, 94 day EGA skin. Lane C, 123 bp ladder. Lanes D, 120 day EGA skin. Lane E, H_2O control. Lane F, 10^7 dilution of TED cDNA. Lane G, 10^6 dilution of TED cDNA. + lanes indicate reverse transcription of RNA samples, - (negative control) lanes contain all the components of the reverse transcription reaction except reverse transcriptase. The expected 400 bp band indicates TED expression in the 77 day human fetal skin.
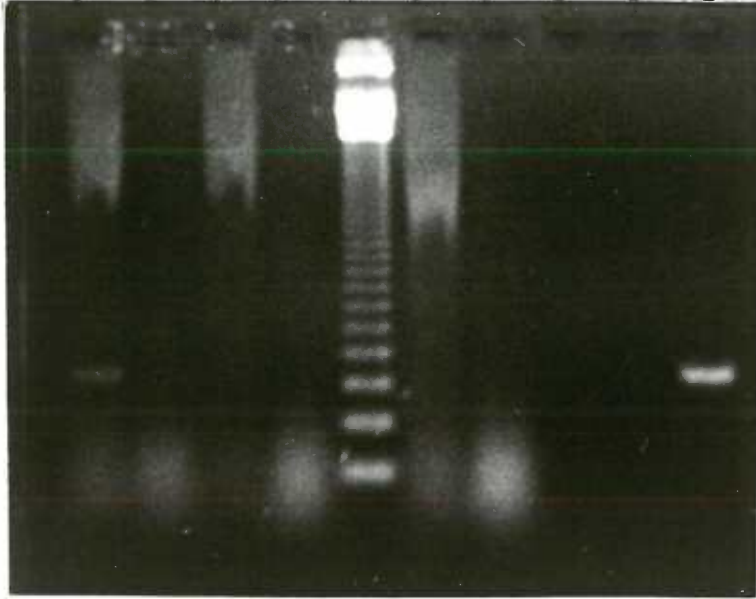
Figure 12 Southern blot showing the junctional fragment in patient ED 1015 EcoRI Southern blot hybridized with cDNA 3 from the TED gene. Lanes A, B, D, E, are EDA males with the normal pattern. Lane C contains DNA from EDA patient 1015. ED1015 is missing the upper EcoRI band and the arrow indicates the detection of a junctional fragment.
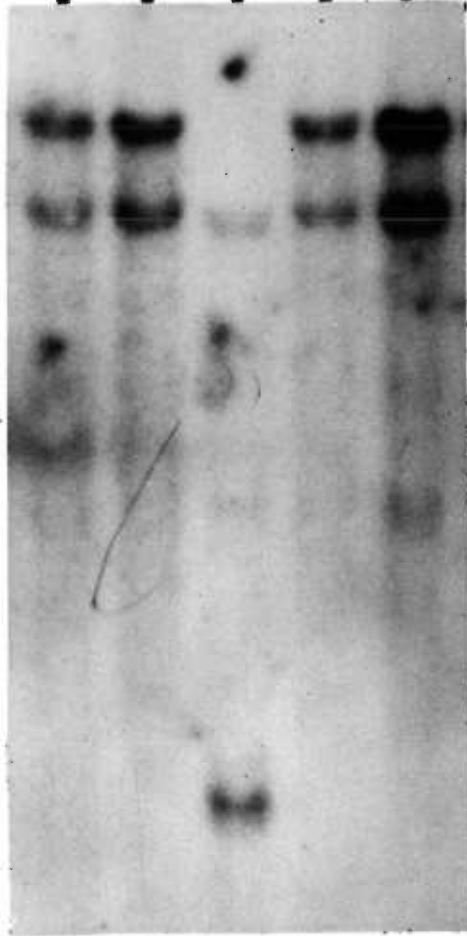
Figure 13 Map of EDA region

A scale at the top estimates the distance between the DXS732E locus and the DXS453 locus. The TED gene is indicated by the arrow above the DXS732E locus. Translocation breakpoints are indicated by a vertical bar if localized to a cosmid, and a bracket if the exact location is not known. Cosmid contigs (A-D) were constructed by our collaborators (Jamel Chelly and Nick Thomas) using the indicated YACs as a probe. A total of five patients with deletion have been identified. Breakpoints known to map within a particular cosmid are indicated with a vertical bar at the end. Unknown breakpoint are indicated by an open bar.
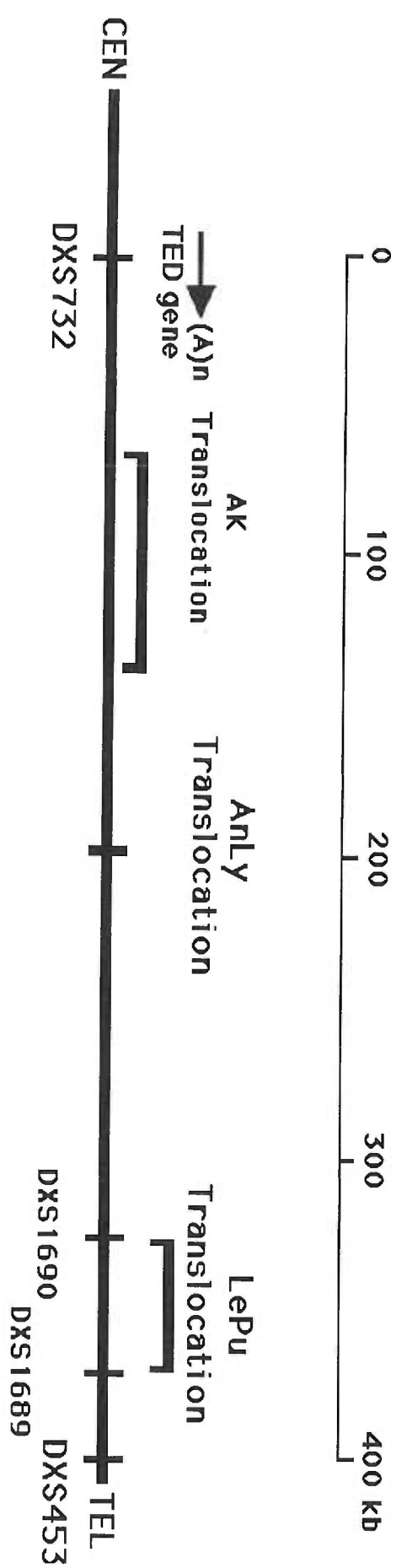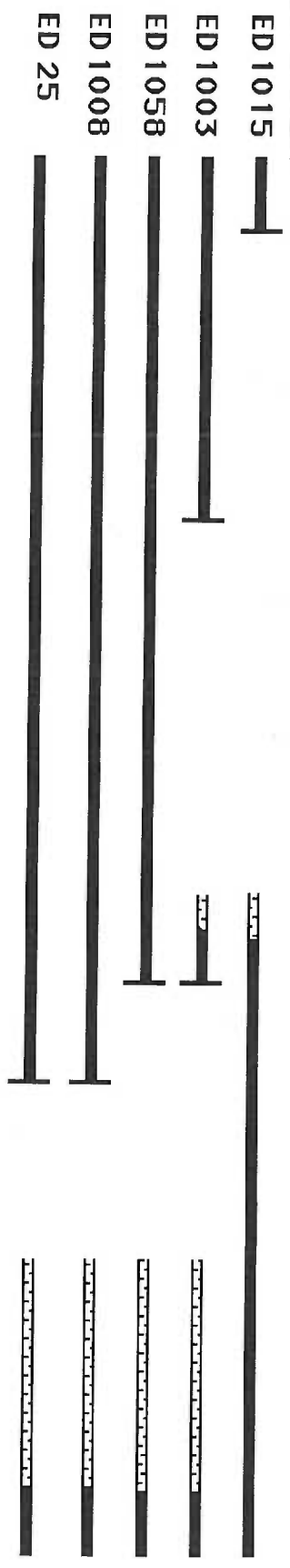
HYPOHIDROTIC ECTODERMAL DYSPLASIA AND IMMUNODEFICIENCY WITH HYPER IgM ; AN X-LINKED DISORDER POTENTIALLY DISTINCT FROM EITHER THE EDA OR HIGM1 GENE LOCI

Judith Gault[1], Seth J. Orlow[2], Lynda Schneider[3], Marilyn Jones[1], Ramsay Fuleihan[3], Barbara Pober[4], Stuart Shapira[5], Michael Litt[1], and Jonathan Zonana[1]

[1]Dept. of Molecular and Medical Genetics, Oregon Health Sciences Univ., Portland

[2]Depts. of Dermatology and Pediatrics, New York University School of Medicine, New York

[3]Dept. of Pediatrics, Harvard Medical School, Div. of Immunology, Children's Hospital, Boston, MA

[4]Dept. of Genetics, Yale Univ. School of Medicine, New Haven

[5]Institute for Molecular Genetics, Baylor College of Medicine, Houston

Contact author:

       Jonathan Zonana, M.D.

       Dept. of Molecular and Medical Genetics L-103

       Oregon Health Sciences University

       3181 S.W. Sam Jackson Park Rd.

       Portland, Oregon 97201

       Phone: (503) 494-4448 Fax: (503) 494-6886

Running Head: Ectodermal dysplasia and immunodeficiency

SUMMARY

We describe two families exhibiting X-linked inheritance of a disorder with the general features of hypohidrotic ectodermal dysplasia and an immunodeficiency with hyper IgM. These families show differences in the severity of the features of hypohidrotic ectodermal dysplasia and immunodeficiency. The disorder is so severe in one family, ED 1082, that the affected members died. In the other family the affected individuals are managing relatively well. These differences may be due to variable expressivity of the same disorder or these families may have different disorders. In both families the disorders are most probably X-linked. The affected half-brothers in each family have different fathers and their mothers appear normal.

The phenotypes found in both families are most consistent with two well characterized X-linked disorders, hypohidrotic (anhidrotic) ectodermal dysplasia, EDA, and immunodeficiency with hyper-IgM, HIGM. The phenotype includes significant hypohidrosis, hypodontia, and immunodeficiency with hyper-IgM. Hypohidrotic ectodermal dysplasia (EDA) has been localized to the region Xq12-q13.1 by both genetic analyses and by the location of the chromosomal breakpoints in females with EDA and X;autosome balanced translocations. Severe immunodeficiency is not a known feature of EDA, and hypohidrotic ectodermal dysplasia is not a recognized feature of HIGM1. We wished to see if the disorder(s) might represent a contiguous gene deletion syndrome involving either the EDA or HIGM1 loci, or if these disorders segregate with markers flanking the EDA locus, HIGM1 locus, or other X-linked loci associated with immunodeficiency.

The segregation of alleles from 24 polymorphic loci spanning the X chromosome were analyzed. In family ED 1136, the brothers had discordant

alleles at loci extending from Xp22.33 (DXYS17) to Xq21.1 (DXS72), including the DXS732E locus, located at or very near the EDA locus. In family ED 1082, the brothers were concordant for all informative loci tested between Xp22.33 to Xq24, including six loci flanking the EDA locus. However, discordant alleles were present at the HPRT and CD40 ligand loci (Xq26). The CD40 ligand, is a candidate gene for HIGM1, and has been demonstrated to be mutated in multiple patients with this disorder. No deletions were observed at any of the loci tested. If both families are affected with the same disorder, then the disorder must be distinct from both the EDA and HIGM1 loci and may map to the Xq21.1-Xq26.1 or Xq26.2-Xq28 regions.

INTRODUCTION

The X-linked locus for hypohidrotic (anhidrotic) ectodermal dysplasia (EDA), characterized by abnormal morphogenesis of hair, teeth and eccrine sweat glands (Clarke et al. 1987), has been mapped to the region Xq12-q13.1 by both genetic and physical mapping methods (Zonana et al. 1992; Plougastel et al. 1992; Thomas, et al. 1993, Kere et al. 1993). Although there has been controversy about the existence and frequency of an autosomal recessive genocopy of the disorder (Sybert 1989), there has been no convincing evidence for a second X-linked locus. Linkage-based heterogeneity analyses (Zonana et al. 1992), and the localization of common breakpoints in several females with EDA and X;autosome balanced translocations (Plougastel et al. 1992; Thomas, et al. 1993, Kere et al. 1993), have supported a single X-locus, with the exception of a couple of atypical families (Goodship et al. 1990; Zonana et al. 1992).

Immunodeficiency with hyper-IgM, a disorder characterized by recurrent infections with low levels of IgG and IgA and normal-to- increased levels of IgM, may be genetically heterogeneous with both X-linked and autosomal loci (Notarangelo et al. 1992). The X-linked locus (HIGM1) has been mapped to Xq26 without evidence for a second X-linked locus (Padayachee et al. 1992). Recently, the CD40 ligand (CD40L), a candidate gene for HIGM1, has been cloned and mapped to the Xq26 region, and shown to be mutated in multiple affected individuals (Korthauer et al. 1993; Allen et al. 1993; Arrufo et al. 1993, DiSanto et al. 1993, Ramesh et al. 1993).

We describe two families, each with a pair of affected half-brothers with different fathers, who have an ectodermal dysplasia with hypohidrosis, and an immunodeficiency with hyper-IgM. The disorder may be due to either a contiguous gene deletion syndrome involving either the EDA or HIGM1 locus, or to an intragenic mutation at either of these loci with unusual pleiotropic affects.

Alternatively, the disorder may represent a new entity at a distinct locus on the X chromosome. In order to distinguish among these possibilities, we have tested multiple polymorphic marker loci spanning the X chromosome in an attempt to exclude involvement of either the EDA or HIGM1 locus in each family, and to identify candidate regions for a potentially distinct X-linked locus.

PATIENTS, MATERIAL, and METHODS

Families studied

Family ED 1082: The full clinical and laboratory details of the immunodeficiency in this family are described elsewhere (Schneider et al. 1992; Wang et al. submitted). The proband (III-1) (Figures 1 and 2) was hospitalized multiple times during his first year of life with bacterial infections of both soft tissue and bone. Laboratory studies demonstrated immunodeficiency with low levels of IgG (143 mg/dl) and IgA (8 mg/dl), and abnormally elevated levels of IgM (868 mg/dl). Studies of cellular immunity and a karyotype were normal. At 17 months of age, he developed diffuse mycobacterial granulomatosis, and despite aggressive treatment, he expired at two years of age. A post-mortem examination was refused. In addition to his immunodeficiency, he was noted to sweat abnormally, and a quantitative sweat test demonstrated a very low amount of sweat production (13 mg). He had sparse, thin scalp hair and eyebrows, frontal bossing, deep-set eyes, and mild mid-facial hypoplasia (Fig. 2). Broad fingernails with splayed tips, and deep-set toe nails with slight cracking and brittleness were also noted. His dentition was abnormal, and at 20 months of age only two widely-spaced conical upper incisors had erupted. Dental X-rays revealed two conically-shaped upper incisors, and absence or extreme hypoplasia of the four lower incisors.

The proband's younger half-brother (III-2), from a different unrelated father, presented with bacterial sepsis and meningitis at three months of age. He

subsequently developed Pneumocystis carinii pneumonia, and disseminated Mycobacterium avium intracellulare infection. Immunologic work-up showed hypo-gammaglobulinemia (IgG - 154 mg/dl, IgA - 15 mg/dl) except for elevated levels of IgM (251 mg/dl). He was noted to have sparse, fine scalp hair, frontal bossing, deep-set eyes, and mild mid-facial hypoplasia with two conical shaped maxillary central incisors. Normal laboratory studies included a karyotype, blood lactate and pyruvate levels, quantitative plasma amino acids, and urine organic acids. A skin biopsy from his thigh demonstrated parakeratosis and dyskeratosis, and a quantitative sweat test was abnormally low. He died at 35 months of age. Family history was unremarkable without signs or symptoms of ectodermal dysplasia or immunodeficiency. Their mother (II-2) had normal dentition, scalp and body hair, and breast development.

Family ED 1136: The proband (III-1) is a 12 year old white male who presented for evaluation of recurrent aphthous stomatitis (Figures 1 and 3a). His past medical history was significant for recurrent infections including otitis media, sinusitis, and purulent nasal discharge. He had been hospitalized four times for pneumonia, once for osteomyelitis of the femur, and once for an abscess of the biceps muscle. A diagnosis of immunodeficiency with hyper-IgM was made at ten months of age based on persistent low levels of IgG (<200 mg/dl - normal range of 294-1069 mg/dl), low-to-normal IgA, and elevated levels of IgM (1100 mg/dl - normal range 41-149 mg/dl). He has been maintained on monthly intravenous gamma globulin therapy, and oral antibiotics. Of significance, he was noted to be unable to sweat as an infant, necessitating air conditioning and aggressive cooling measures. He had conical shaped maxillary lateral incisors, and was missing four of his primary and seven of his secondary dentition, excluding the third molars. Based on these findings he had received a diagnosis

65

of hypohidrotic ectodermal dysplasia. Laboratory studies included a normal 46,XY karyotype. On examination at twelve years of age he showed periorbital wrinkling, and hypodontia with widely spaced conical teeth. Mild eczematous changes were present on the trunk, and superficial erosions were present on the tongue and buccal mucosa. Scrapings failed to demonstrate the presence of multinucleated giant cells.

The proband's six year old half-brother (III-2) (Figures 1 and 3b), with a different unrelated father, was studied during infancy because of his brother's disorder, and found to have a similar immunodeficiency with elevated levels of IgM. He was begun on treatment with immunoglobulin and had recurrent episodes of otitis media, but was hospitalized only once, due to an abscess of the right thigh. A typical trough level on therapy at the age of two showed an IgG level of 603 mg/dl (normal range 538-1400 mg/dl), and an IgM level of 723 mg/dl (normal range of 54-206 mg/dl). Similar to his brother, he did not sweat and needed air conditioning and cooling measures. He had conical shaped primary incisors, and was missing four primary teeth. The absence of four of his secondary dentition, excluding the third molars, was demonstrated radiologically. Family history was unremarkable. Their mother (II-2) had normal dentition, sweating, scalp and body hair, breast development, and immunoglobulin levels.

DNA Samples

DNA samples were obtained from the mother and her two sons in family ED 1082, but neither father was available for testing. Since both boys were deceased, DNA samples were extracted from cultured skin fibroblasts by standard methods (Miller et al. 1988). In family ED 1136, blood samples were obtained from the affected boys, their mother, and the father (II-3) of the proband's half-brother (III-2).

Nucleated cell pellets were prepared from 10-15 ml of whole blood (anticoagulated with ACD or EDTA) using a blood cell lysis buffer (BCL) (0.32 M sucrose, 0.01M Tris HCl pH 7.5, 0.005 M $MgCl_2$, 1% triton X-100). Cold BCL was added to the blood to a total volume of 50 ml, incubated on ice for 10 minutes, and centrifuged at 850xg for 15 minutes at 4° C. The supernatant was removed, the pellet resuspended in 30 ml cold BCL, and re-centrifuged. The supernatant was removed and DNA was extracted from the pellet using previously described methods (Miller et al. 1988).

Polymorphic Loci Analyzed

Twenty-four polymorphic marker loci spanning the X chromosome were analyzed in both families (Table 1). The concordance or discordance of alleles at informative loci was scored for each set of brothers. This method has been used previously in the analysis of Rett syndrome when the number of informative meioses was too small for conventional linkage analysis (Ellison et al. 1992). Particular emphasis in our study, was placed on loci in the Xq12-q13.1 and Xq26 region, at or near the EDA and HIGM1 (CD40L) loci, respectively. The location and order of the loci are as reported in the current X chromosome consensus map (Schlessinger et al. 1993) and the data of Weissenbach et al. (1992a and 1992b).

PCR reactions were performed in a 13 µl volume including 25 ng of genomic DNA, 0.26 units of Taq polymerase (Ampli-Taq), 10 pmoles of each primer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 50 mM KCl, 10 mM Tris-HCl pH 8.3, 0.25 mM spermidine, and 0.01% gelatin. Published sequence and annealing temperatures were used for each locus (Table 1), except for the CD40L locus. Primers (TCCCCCAGTCTCTCTTCTCA; TTTCGTAATGAGGAGTGGGC) flanking the polymorphic $(CA)_n$ dinucleotide repeat at this locus (DiSanto et al.

1993) were designed from sequence deposited in GenBank to give a product of 216 bp, using an annealing temperature of 58°C. "Touchdown" PCR amplification was performed in a MJ Research PTC-100 thermocycler, using an initial temperature 10°C higher than the annealing temperature of the primers, and decreased 2°C in each cycle for the first 5 cycles. The temperature was then held constant at the annealing temperature of the primers for 30 additional cycles. Each cycle consisted of 30 seconds denaturing at 94°C, 30 seconds at the specific annealing temperature, and a 30 second extension at 74°C. PCR products were resolved on a 32% formamide-7% polyacrylamide sequencing gel, capillary blotted on to Hybond N+ membrane (Amersham), and probed with [^{32}P] end-labeled (CA)$_{15}$ oligonucleotide (1 pmole/ml of hybridization solution). Blots were hybridized in 5X SSPE/0.1% SDS at 65°C for one hour, and subsequently washed 2 to 3 times at 60°C in 6x SSC/0.1% SDS. Autoradiography was performed using Kodak X-OMAT AR film with an intensifying screen for 1 to 3 days at -80°C.

The AR and HPRT loci were detected by internal labeling using 0.5 µCi of alpha [^{32}P] dCTP and 2 µM cold dCTP, in a 15 µl reaction mixture as described above (Edwards et al. 1992). DMSO was added to the reaction mixture to a final concentration of 10% for the analysis of the AR locus. The internally labeled PCR products were resolved on a 7% polyacrylamide gel, which was subsequently dried, and autoradiography was performed.

Southern analyses were used to detect polymorphisms at the DXS255, DXS732E, and DXS72 loci. These analyses could not be performed on family ED 1082 due to insufficient DNA secondary to the loss of all of the fibroblast cell lines on individual III-2. Genomic DNA was digested with the appropriate restriction enzyme, electrophoresed on 0.8% agarose gels, transferred to Hybond N+ membranes using 0.4N NaOH, and probed with the appropriate oligo-labeled

68

probe. Hybridizations were performed in 7% PEG, 10% SDS, 50 mM PO_4, 200µg/ml human placental DNA, and $1.5 - 2 \times 10^6$ cpm per ml probe, at 65°C overnight. Blots were washed sequentially in 2X SSC/0.1% SDS followed by 0.1X SSC/0.1% SDS for 15 minutes at room temperature. A final wash was done in 0.1% SSC/0.1% SDS at 60°C, and autoradiography was performed.

RESULTS

Nine out of twenty-four loci were informative in both families, ten were informative in only one family, and five were uninformative in both families. No deletions were observed for any of the loci tested in either family, including the CD40 ligand and DXS732E loci. The concordance or discordance for alleles at each of the loci tested is shown in Figure 4. In family EDA 1136, the brothers were discordant for alleles at nine loci extending from Xp22.33 (DXYS17) to Xq21.1 (DXS72). Of significance, they were discordant for alleles both proximal (AR, DXS732E) and distal (DXS72) to the EDA locus. Barring gene conversion or double recombination within a very small region, the results seen in family EDA 1136 excludes the disorder from the Xq12-q13 region, wherein the EDA locus has been localized. The discordant alleles at the DXS732E locus are of significance, since results of previous two-point linkage analyses demonstrate a θ_{max} of 0.00 (Z_{max} 21.2) between the EDA and the DXS732E loci, and the DXS732E locus has been shown to be deleted in an otherwise normal male with EDA and a submicroscopic molecular deletion (Zonana et al. 1993). The DXS732E locus is also present on a yeast artificial chromosome shown to span the breakpoint in a female with EDA and an X;autosome translocation (Thomas et al. 1993). The brothers were concordant at five loci within the interval Xq21.3-Xq28.

In family EDA 1082, the brothers were concordant at twelve loci from Xp22.33 to q24, and at the DXS998 locus (Xq27.3). Of significance, they were concordant at the DXS339 locus, which on a two-point analysis has a θ_{max} of 0.00 (Z_{max} 28.19) between it and the EDA locus (Zonana et al. 1992). However, they were discordant for alleles at the HPRT locus, and for the (CA)n dinucleotide repeat located within the 3' untranslated region of the CD40 ligand, both located at Xq26. This would likely exclude involvement of the HIGM1 locus. To further support this conclusion, one of the brothers (III-2) has been shown to have normal CD40 ligand expression on the surface of his T cells (R. Fuleihan, unpublished data), which has been shown to be deficient in males with X-linked immunodeficiency and hyper-IgM (Fuleihan et al. 1993). Neither family showed any evidence for submicroscopic deletions involving either the HIGM1 or EDA loci. Insufficient DNA was available for RFLP analysis at the DXS732E locus in family EDA 1082. However, results of a previous study that included the proband as part of a patient panel, showed no evidence for a deletion upon Southern blotting with cosmid ICRFc104HO41.43 (DXS732E) used as a probe (Zonana et al. 1993).

Two half-brothers in each family were stated to have been fathered by different unrelated men, and had clinically unaffected mothers. Analysis of DNA polymorphisms in family ED 1136 at the D11S533 (Litt et al. 1992) and DXYS17 loci (Decorte et al. submitted) were consistent with individual II-3 being the father of III-2, but not of III-1 (Fig. 5). Sixteen autosomal loci, heterozygosity 0.75, at least 50 cM apart were analyzed in the mother and her two sons of family ED 1082, to test the hypothesis that the boys had the same father (loci analysis done by Todd Taylor). The boys shared alleles at 4 of the loci and were discordant at 12 of the loci. Using Chi-square probability, the boys have a 1% probability of having the same fathers.

DISCUSSION

Confirmation that the brothers in family ED 1136 had different fathers supports inheritance of the disorder as a probable X-linked recessive trait. The immunodeficiency in both families was characterized by low levels of IgG and IgA, but abnormally elevated levels of IgM. Although minor abnormalities of immunoglobulin levels have been reported in males with X-linked hypohidrotic ectodermal dysplasia (Clarke et al. 1987), they have been clinically insignificant, and have not been characterized by markedly elevated levels of IgM. The clinical and laboratory findings in families ED 1082 and 1136 are compatible with a diagnosis of immunodeficiency with hyper IgM, a causally heterogeneous disorder (Notarangelo 1992). The immunodeficiency was more severe in family ED 1082, resulting in infection with opportunistic organisms and early death. Both marked variability in clinical severity, and opportunistic infections have been described in other patients with immunodeficiency and hyper-IgM, despite the inability to demonstrate deficient cell-mediated immunity in the laboratory (Notarangelo 1992).

The males in both families showed clear signs of hypohidrosis and hypodontia. Family ED 1082 appeared to have more marked manifestations of ectodermal dysplasia, with sparser hair and mid-face hypoplasia. The hair and dental involvement in family ED 1136 was less severe than that usually seen in X-linked hypohidrotic ectodermal dysplasia (EDA). If the disorder in the two families are allelic, the difference in the severity of manifestations could be explained by the site and nature of their respective mutations. One cannot prove that the disorders are allelic until the basic defect is identified.

Therefore, if the mutations in both families involve the same locus, the evidence suggests a locus for the disorder on the X chromosome distinct from

71

either the EDA or HIGM1 loci. Although the density of the informative marker loci

tested is insufficient to completely exclude all of the regions between discordant

loci, the data support a possible location for such a locus within either the

Xq21.1-q26.1 or Xq26.2-q28 intervals. At least three disorders with

immunodeficiency or ectodermal dysplasia have been localized to the Xq21.1-

q26.1 and Xq26.2-q28 regions. The gene responsible for agammaglobulinemia

(AGMX1) has been cloned and localized to Xq21.3-q22 (Tsukada et al. 1993;

Vetrie et al. 1993), however, these patients do not have elevated levels of IgM

nor ectodermal dysplasia. Elevated levels of IgM can be seen in the X-linked

lymphoproliferative syndrome (LYP) at Xq25 (Grierson et al. 1991), but males

with this disorder and large microscopic and submicroscopic deletions (2-4 Mb)

have been described, and have no signs of ectodermal dysplasia (Wu et al.

1993). Dyskeratosis congenita maps to Xq28 and may have manifestations of

immunodeficiency and hypohidrosis (Davidson et al. 1988; Arnagrimsson et al.

1993). However, neither hypodontia nor immunodeficiency with hyper-IgM has

been associated with this disorder, and other major clinical features of the

disorder are absent in our patients. Additional families with immunodeficiency

and hypohidrotic ectodermal dysplasia may be required to further localize and

define this disorder.

ACKNOWLEDGMENTS

Figure 1 Pedigrees of families ED 1082 and ED 1136

Blackened squares represent affected males, dotted circles denote obligate carrier females, and arrows indicate the probands.

EDA 1082
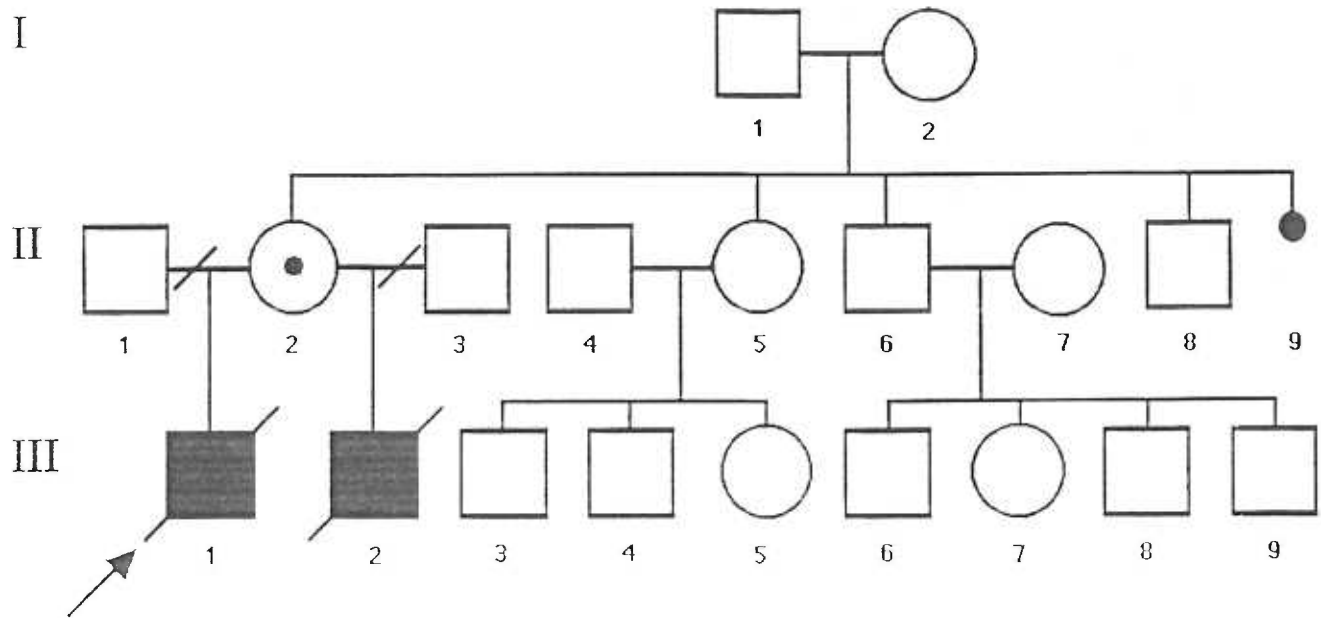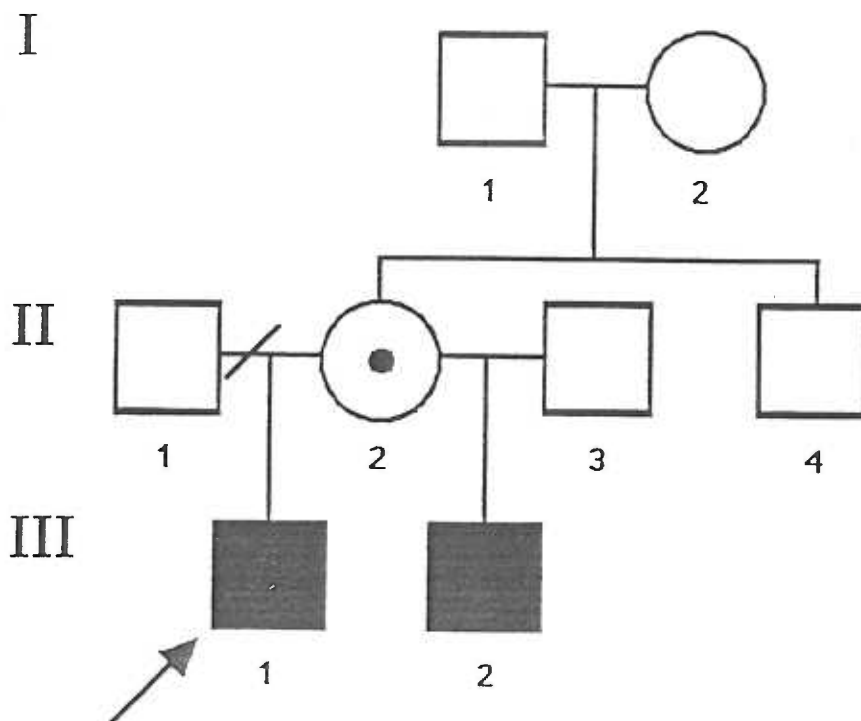
ED   1136

Figure 2 Proband (III-1) from family ED 1082

Note his sparse scalp hair and eyebrows, and mild mid-facial hypoplasia.

Figure 3 The proband from family ED 1136 and his half-brother

(a) Proband (III-1) from family ED 1136 (left) (b) his affected half-brother (III-2)

(right). Note his missing and conical shaped teeth.
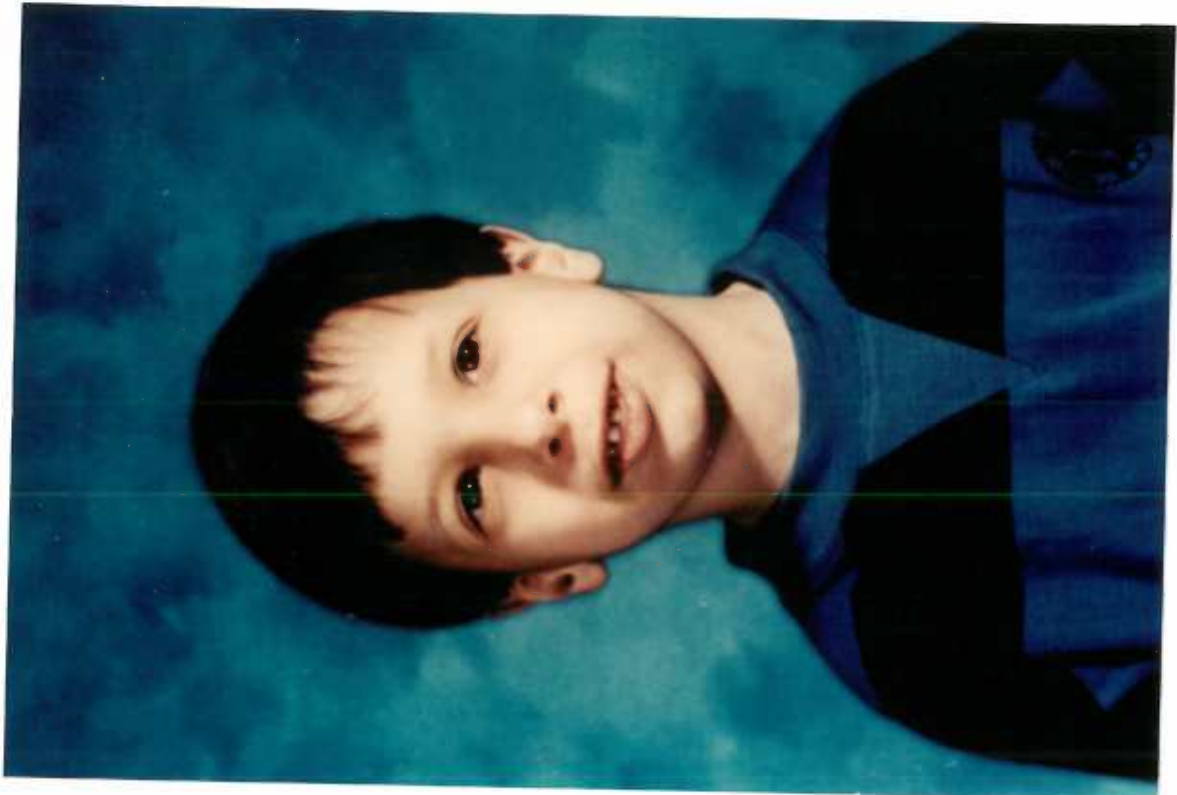
Figure 4 Results displaying concordance/discordance in each pair of half-brothers

Concordance/discordance analysis of twenty-four polymorphic loci spanning the X chromosome for the pairs of maternally related half-brothers from families ED 1136 and ED 1082.

Locus	Location	ED 1136	ED 1082
DXYS17	Xp22.33	●	○
DXS996	Xp22.33-p22.32	◍	○
KAL	Xp22.31	●	○
DXS987	Xp22.12	●	◍
DXS451	Xp22.11	●	○
DXS997	Xp22.12-p11.3	●	○
DXS1068	Xp22.12-p11.3	◍	◍
DXS426	Xp11.23	◍	○
DXS255	Xp11.22	●	—
AR	Xq12	●	○
PGKP1	Xq12	◍	◍
DXS339	Xq12	◍	○
DXS732E	Xq12	●	—
DXS453	Xq13.1	◍	◍
DXS441	Xq13.3	◍	◍
DXS72	Xq21.1	●	—
DXYS1X	Xq21.3	◍	○
DXS456	Xq21-q22	○	○
DXS424	Xq24	○	○
HPRT	Xq26.1	○	●
CD40L	Xq26	◍	●
DXS998	Xq27.3	○	○
F8C	Xq28	◍	◍
DXS1108	Xq28	○	◍
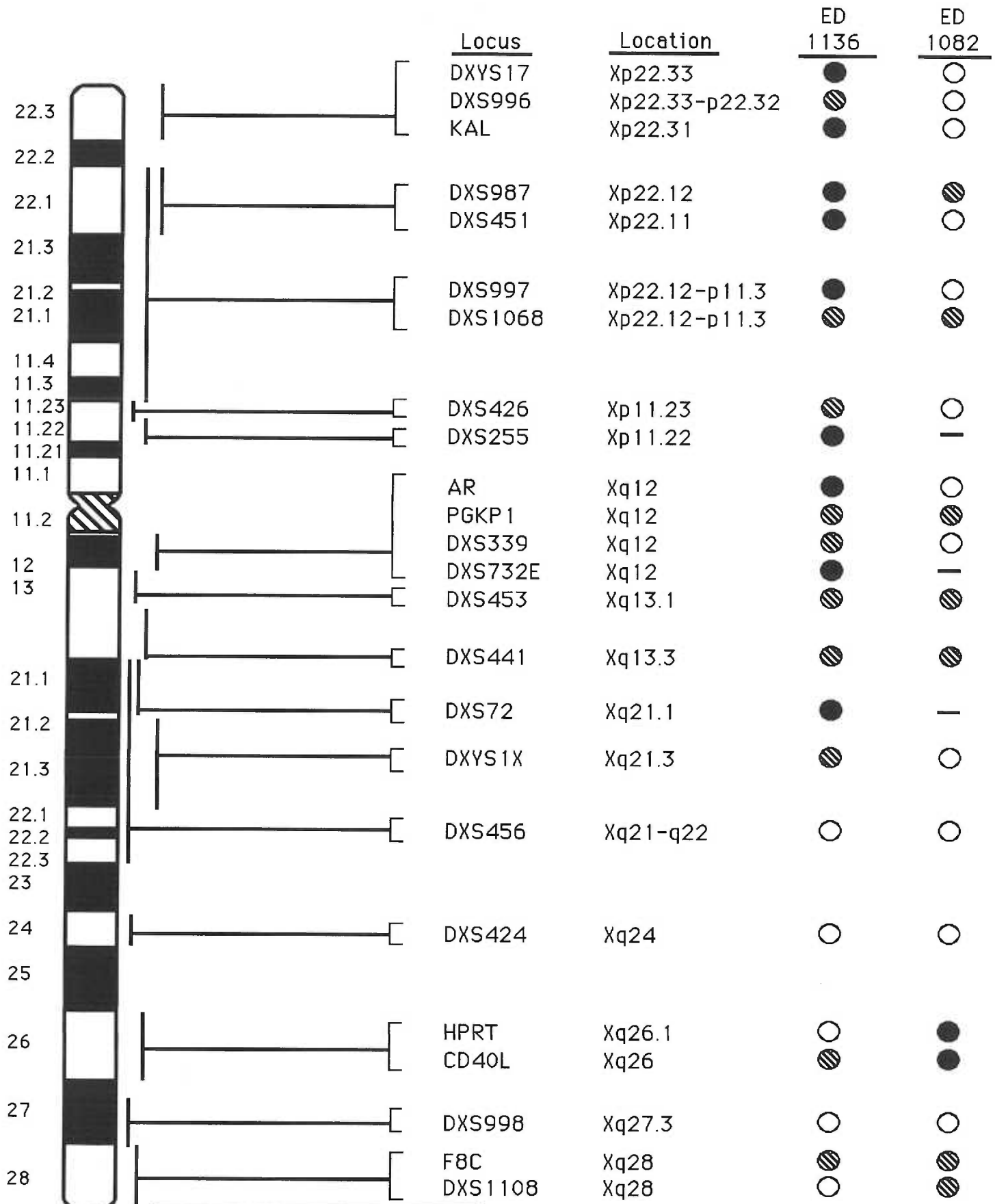
KEY ● = Discordant
○ = Concordant
◍ = Not Informative

Figure 5 Results of analysis of the DXYS17 and D11S533 loci in family ED 1136

Alleles at the DXYS17 and D11S533 loci in family ED 1136, consistent with individual II-3 being the father of one of the affected boys (III-2), but not of his affected half-brother (III-1).
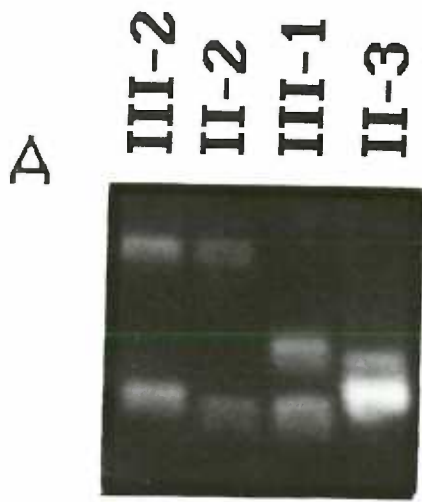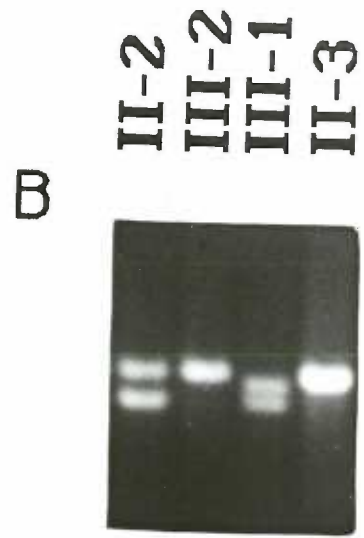
A — DXYS17

B — D11S533

Locus	Location	Polymorphism	Het.	References
DXYS17	Xp22.33	VNTR	0.84	Decorte et al. submitted
DXS996	Xp22.33-p22.32	(CA)n	0.82	Weissenbach et al. 1992
KAL	Xp22.31	(CA)n	0.76	Bouloux et al. 1991
DXS987	Xp22.12	(CA)n	0.83	Weissenbach et al. 1992
DXS451	Xp22.11	(CA)n	0.80	Browne et al 1992
DXS997	Xp22.3-p11.3	(CA)n	0.64	Weissenbach et al. 1992
DXS1068	Xp22.3-p11.3	(CA)n	0.80	Weissenbach 1993
DXS426	Xp11.23	(CA)n	0.52	Luty et al. 1990
DXS255	Xp11.22	RFLP	0.90	Fraser et al. 1987
AR	Xq12	(GGC)n	0.90	Edwards et al. 1992
PGKP1	Xq12	(CA)n	0.43	Browne et al 1992
DXS339	Xq12	(CA)n	0.73	Zonana et al. 1992
DXS732	Xq12	RFLP	0.48	Zonana et al. 1992
DXS453	Xq13.1	(CA)n	0.68	Weber et al. 1990
DXS441	Xq13.3	(CA)n	0.76	Porteous et al. 1992
DXS72	Xq21.1	RFLP	0.50	Zonana et al. 1988
DXYS1X	Xq21.3	(CA)n	0.53	Browne et al. 1991
DXS456	Xq21-q22	(CA)n	0.77	Luty et al. 1990
DXS424	Xq24	(CA)n	0.83	Luty et al. 1990
HPRT	Xq26.1	(AGAT)n	0.77	Edwards et al. 1992
CD40L	Xq26	(CA)n	0.50	This study*
DXS998	Xq27.3	(CA)n	0.55	Weissenbach et al. 1992
F8C	Xq28	(CA)n	0.71	Lalloz et al. 1991
DXS1108	Xq28	(CA)n	0.71	Pattatucci et al. 1993

Discussion and Conclusions

Discussion of chapter 1

Characterization of the TED gene at the DXS732E locus

A novel gene, termed TED, mapping to the DXS732E locus has been isolated and characterized. Sequence analysis of three cDNAs and part of a Not I genomic subclone, containing part of the putative first exon, totals 4038 base pairs. Sequence analysis using a primer walking approach produced 3.4 kb of single stranded non-contiguous sequence. Sequence analysis using a nested deletion approach, cDNAs were prepared with exonuclease digestion, resulted in 0.9 kb of double stranded sequence and 3.2 kb of single stranded sequence. Other sequencing methods used include sequencing cDNA subclones and cycle sequencing. In total 4038 base pairs were sequenced by different methods with approximately 3200 bases of sequence from both strands. The most abundant transcript of the gene is estimated to be 5 kb in length according to Northern analysis. The entire 5' untranslated region has not been cloned. Primer extension experiments could be used to estimate the distance between the 5' end of the mRNA and the current sequence. 5' rapid amplification of cDNA ends, RACE, experiments could be used to clone the 5' sequence. In addition the fetal brain library could be re-screened by hybridization using the most 5' sequence as a probe. The 5' end of the gene that has not been cloned is approximately 800 bp from the current 5' end based upon comparing the message size to the total length of the contiguous cDNA clones. The size of the poly A tail may account for part of the additional sequence. Only 34 A's were sequence from cDNA 10 and poly A tails are approximately 200-250 bp in length (Lewin 1994, p875). It is likely that the remaining sequence is noncoding because it lies 5' of the two putative start methionines and GRAIL does not rate sequence beyond the two start methionines as being coding.

The putative open reading frame, 1422 base pairs, was identified through sequence analysis and confirmed with GRAIL analysis (Shah et al. 1994). The ORF has been rated as having "excellent" coding potential by GRAIL analysis. GRAIL predictions of "excellent" coding potential have been accurate throughout this project. Initially, GRAIL rated a 250 bp portion of an 800 bp human genomic sequence as having "excellent" coding potential. Forty-two base pairs 5' of the "excellent" human sequence is a splice site and 94 base pairs 3' of this sequence is the putative stop codon from the identified ORF (Figure 1). In this example, GRAIL roughly estimated an intron exon boundary and the stop codon. The homologous mouse genomic sequence was rated as "marginal" for only 10 bp over the homologous region. GRAIL is designed for the analysis of human sequences, but reportedly has successfully identified coding regions in other mammalian DNA. It is possible that GRAIL is not as sensitive at detecting coding regions in mouse, even if the coding region is 82% homologous to the human sequence. There are at least two suspected sequence errors in the putative coding mouse genomic sequence indicated by gaps in the alignment of human and mouse sequence (Figure 1). One gap of one base pair in the alignment is 13 bp from the splice site and a second gap of two base pairs is 243 bp from the splice site. The clone pcos 169E/4 was sequenced by Neil Brockdorff, and the autorads are not available to check the sequence in these regions. These gaps could have decreased the sensitivity of GRAIL analysis by changing the frame (Figure 1).

GRAIL was extremely useful in determining which frame of sequence was coding, allowing one to focus in on potential sequence errors indicated by GRAIL ratings as "excellent" in different frames. GRAIL analysis gave the first indication that the ORF continued beyond the cDNA clones. The 5' sequences of cDNA clones 3 and 10 start at the same base, supporting the hypothesis that both

80

contained the 5' end. This hypothesis turned out to be false, and the sequences may have the same start as a result of secondary structure in the RNA causing the reverse transcriptase to terminate at the same position in both cDNAs. The two cDNAs start after the third triplet repeat of an (ACG)x6 triplet repeat. This repeat may promote the formation of secondary structures that are difficult to transcribe through. GRAIL analysis recognized coding potential in one frame of sequence near the most 5' cDNA sequence. The sequence was rated "moderate" but the sequence in that frame did not contain any start methionines. Identification of domain 1, a stretch of sequence thought to be coding with homology to the HUMMRNAC cDNA sequence, supported the hypothesis that the start methionine was 5' of the cDNA clones. If the structure between the HUMMRNAC gene and the gene from the DXS732E locus were similar, the start methionine would not be far beyond the cDNA clones. After cloning and sequencing the Not I genomic fragment from cosmid ICRFc104HO41.43 which contains the most 5' sequence of the cDNAs, the putative start codon(s) were identified.

There are two potential start methionines in the same frame only 41 and 65 bases 5' of the sites were cDNA 10 and 3 are cloned into the polylinker. The putative ORF starts just 3 or 21 bases after a stop codon depending on which start methionine is used, supporting the hypothesis that one of two close methionines is the start methionine (Figure 2). The identified stop codon at the end of this putative ORF is probably correct since no errors were found in sequence from both strands. In addition, multiple gaps, not in multiples of three, are found in the mouse and human sequence alignment in the region following the putative stop codon. Before the stop codon (TGA in bold, Figure 1) only two gaps are present. One may be due to a compression in the mouse sequence. Multiple sequencing errors would have to be present near the putative stop

codon causing the frame to shift a number of times in order to extend the ORF significantly (Figure 3). Grail rated two sequences beyond the putative stop codon as having some coding potential. One short stretch of 12 bases was rated as marginal in one frame approximately 700 bases from the putative stop codon, and one 50 base pair sequence in one frame was rated as "good" approximately 1660 bases beyond the putative stop. In a study described in the GRAIL manual sequences rated as "marginal" and "good" turned out to be coding 16% and 69% of the time respectively.

Two intron-exon boundaries have been identified based upon divergent sequence between cDNAs and genomic clones. Restriction maps of cDNAs and genomic clones coincide before the first splice site over 910 base pairs and after the last splice site over 2965 base pairs, supporting the hypothesis that the last exon of this gene has been identified and part of the first exon may be identified. Final proof that the 5' exon has been identified will come with cloning the 5' region including the promoter. The putative ORF begins in the first identified exon and ends in the last exon containing a poly (A) addition site based upon cDNA sequence analysis. Additional splice sites have not been identified and may exist in the 169 bp intervening coding sequence or in the region beyond (5') of the current sequence (Figure 4).

A restriction site discrepancy between the genomic Not I clone and cDNA 3 was identified within the first exon. Genomic and cDNA sequence were in agreement on either side of this discrepant region. The genomic sequence contained an Eag I site at base 301, while cDNA 3 does not based upon Southern analysis of Eag I digested DNA. Southern analysis of DNA from normal males and females was used to test for the presence of this Eag I site. Ten X chromosomes tested were in agreement with the cosmid restriction map which shows the Eag I site at position 301. The loss of this Eag I site may represent a

82

rare polymorphism or a mutation in cDNA 3. Sequence through this region from cDNA 3 is not complete, despite multiple sequencing attempts, due to the high GC content of this region. Therefore, the exact nucleotides which differ have not been identified.

Sequence from the TED gene has been used to search sequence and protein databases for identical or homologous sequences. BLAST programs have been the most efficient method for database searching (Altschul et al. 1990). No identical sequences have been found, but one sequence with a high degree of homology in two domains was identified. Two homologous regions of sequence are shared between the TED gene and the HUMMRNAC gene that maps to chromosome 13 (Li et al., 1993). These domains are compared in detail in Chapter 1 of this thesis (Figures 8 and 9). These two domains have been identified as a result of cloning and sequencing the gene at the DXS732E locus. The gene at the DXS732E locus has been termed TED for Two Established Domains. The genes have similar transcript sizes and spacing of the two domains. Both genes contain expressed triplet repeat sequences. The HUMMRNAC gene has an interrupted repeat of CAGs coding for Gln (GlnX4, GlnX9, Gln X6) 3' of domain 1. The repeat is not very polymorphic (Li et al., 1993). Two alleles were identified giving an average heterozygosity of 20%. The TED gene has two short repeat sequences one either side of domain 1 including 6 TGCs and 4 CCGs that code for cysteine and proline respectively. The expressed triplet repeats may not be characteristic of the gene family. It is difficult to determine the significance of these triplet repeats since the repeats do not code for the same amino acid and are not in the same relative positions.

It is possible that the homologous domains are characteristic of multiple genes, members of the same gene family, that have not yet been entered into the sequence databases. Hybridization with a probe containing domain 1 to human

genomic DNA shows multiple bands in addition to the expected bands from the DXS732E locus even after stringent washing (0.1X SSC, 0.1% SDS at 60 C for 10 minutes).

The entire TED gene isolated to date maps within cosmid ICRFc104HO41.43 (Figure 4). A map of the cosmid was constructed with Kpn I, Eag I, Not I, and Sac II single and double digests. The TED gene was placed on the map through a series of primer hybridizations to Southern blots of cosmid DNA digested with single and double digests of Kpn I, Eag I, Not I, Bam HI, and Sac II (Figure 4). Identification of the most 5' and most 3' splice sites of clones were essential to the construction of the map. A detailed map of the CpG island that includes the 5' exon of the TED gene was mapped by cloning the 1.3 kb Not I fragment from ICRFc104HO41.43, and then sequencing subcloned Eag I fragments from the Not I clone (Figure 4). The efficiency of sequencing GC rich regions in CpG islands can be increased by subcloning fragments generated from rare cutting restriction digests. Few compressions were present upon sequencing the six Eag I subclones. The most difficult compressions to get sequence from seemed to correspond to Eag I sites when using primer walking, exonuclease sequencing or cycle sequencing. By subcloning at the Eag I sites compressions were reduced perhaps by circumventing the problem of secondary structure formation at these GC rich palindromic sequences. In addition, sequencing from vector primers consistently resulted in cleaner sequence. The CpG island also contains at least three Sma I sites. Additional Eag I sites from this island may map outside of the Not I fragment and could not be detected with this analysis. The 169 bp exon(s) maps somewhere in between the 5' and 3' exon. This 169 bp fragment could be localized to a specific region by using primers designed from this region as probes against a Southern with single and double digests of the cosmid. The second CpG island, 3' to the TED gene,

contains at least one Eag I site and Sac II sites. These enzymes could cut multiple times in this island and not be detected in this analysis.

Is the TED gene a candidate gene for EDA?

Several lines of evidence support the hypothesis that the TED gene might be involved in EDA. This gene maps to the DXS732E locus at Xq13.1. Recombinations have not been detected between the DXS732E locus and the EDA locus, $\theta_{max}=0(Z_{max}=21.2)$ (Zonana et al. 1992). It is likely that many genes will map to or between the DXS732E and DXS453 loci, the two tightly linked markers flanking the EDA locus, since they are about 400 kb apart. One recombination between the DXS453 locus and the EDA locus was found in a family used in the analysis by Zonana et al. 1992. This family may be useful in mapping genes in the region between the EDA and DSX453 loci. The TED gene is homologous to a gene at the DXCrc169 locus in mouse which is tightly linked to the *tabby* locus. Every gene that maps to the EDA or *tabby* region should initially be considered a candidate gene and worth further investigation.

The EDA gene should be conserved in the mouse, and subject to X-chromosome inactivation, like the TED gene is. Although these characteristics are essential to an EDA candidate gene, they are less discriminating and it is likely that several genes in this region share these characteristics. The differential methylation pattern found in the putative first exon, showing hypermethylation of five Eag I sites on the inactive X chromosome and hypomethylation of the active X chromosome, is characteristic of genes that are subject to X inactivation. There is no simple way to prove that this gene is not expressed from the inactive X chromosome because the gene is not expressed in fibroblasts or lymphoblasts. A gene subject to X inactivation would be differentially expressed in somatic cell hybrids, constructed from fibroblasts, one

retaining an active X chromosome, and the other an inactive X chromosome (Brown and Willard 1990). The TED gene and the putative homologous mouse gene at the DXCrc169 locus are highly conserved based upon hybridization results and partial sequence comparisons between the two genes. Even the putative three prime untranslated regions are highly conserved between mouse and humans based upon comparison of cDNA sequence from the TED gene and genomic sequence from the putative mouse gene.

Expression analysis results are consistent with the TED gene being an good EDA candidate. Northern analysis done by our collaborator, Joan Zeltinger working in the laboratory of Karen Holbrook, PhD. at the University of Washington, shows the TED gene is expressed at low levels at 55 and 96 days estimated gestational age (EGA), and higher levels at 77 and 84 days EGA in human fetal skin. At 77 days EGA, the first developing hair follicle structures become visibly identifiable. Many hair follicles are missing in EDA patients, probably resulting from a failure to induce hair follicle formation. If the EDA gene is involved in the induction of hair follicle formation, one would expect it to be expressed in human skin before 77 days EGA. Based upon Northern analysis, probed with cDNAs 11 and 3, the TED gene is expressed at low levels at 55 days EGA and either increases consistently throughout the intervening 20 days to the levels seen at 77 days EGA, or the levels remain low and are increased to the levels seen at 77 days EGA at some point between 55 and 77 days EGA. It is possible that the TED gene is expressed at relatively high levels before the first sign of hair follicle development is visible. The first visible structure in developing eccrine sweat glands forms at 90-120 days EGA. A weaker signal is detected at 96 day whole body skin, 105 day scalp, 120 day scalp, and adult skin according to Joan Zeltinger's Northern analysis results. Induction of eccrine sweat gland formation must occur some time before 90 days and expression of the TED gene

was relatively high at 84 days EGA. The expression pattern of the TED gene is consistent with a gene involved in both the induction of hair follicle and sweat gland formation. In situ hybridization in necessary to prove the TED gene is expressed in presumptive and developing hair follicles and sweat glands. Expression analysis of tissues from presumptive tooth germs was not done.

The most convincing evidence in support of the TED gene being a good EDA candidate was the finding that one patient ED1015, with EDA and no other apparent abnormalities, was deleted for the 3' end of the TED gene. The deletion involves at least 388 bases of coding sequence and the entire 3' untranslated region including the poly (A) addition site based upon Southern analysis (Figure 5). The function of poly (A) tails is not known, but histone messages are processed without poly (A) tails. Lack of a poly (A) addition site could interfere with transcription, the stability of the transcript, and translation. Normally RNA polymerase transcription is terminated downstream of the poly (A) signal (Darnell et al. 1990). The signal to terminate transcription is not known but may involve the poly (A) site. Poly (A) tails are thought to stabilize mRNAs since it has been shown that they are removed before degradation of certain messages (Darnell et al. 1990). Removing poly A tails from transcripts has been shown to inhibit in vitro translation (Darnell et al. 1990). Modification, a two base pair deletion, of the poly(A) addition site has been shown to cause disease (thalassaemia) by Harteveld et al. 1994. The TED gene could be inefficiently transcribed, unstable as a transcript or incapable of being translated in ED1015 because it lacks the poly(A) addition site.

Deletion of the 3' untranslated region could interfere with posttranscriptional regulation of translation. Iron responsive element-binding protein, has been shown to bind to the 3' untranslated region of members of the IRE family messages to protect the mRNA from degradation (Theil, 1994).

87

ED1015 is deleted for the entire 3' untranslated region of the TED gene, 2576 bp in size.

Domain 2 may be important for protein functioning, and is at least partially deleted for in patient ED 1015. The breakpoint of the deletion through the TED gene has been estimated to lie within a 10 kb of genomic DNA between the two identified splice sites. A more precise estimate of the deletion breakpoint could be determined by sequentially probing a Southern blot of DNA from ED1015 with small subclones of the cDNAs or using PCR primer sets along the region which is potentially deleted. It seems likely that the deletion would disrupt transcription, processing, or function of the TED gene. The gene is not expressed in lymphoblasts or fibroblasts, so we have been unable to determine if the TED gene is transcribed in patient ED1015. One might expect to see a phenotype associated with the loss of function of the TED gene. The expected phenotype should be related to the expression pattern of the TED gene. The TED gene is expressed in skin during hair follicle and sweat gland formation so one might see a phenotype involving these ectodermal derivatives. This gene is expressed in other fetal tissues including brain (18 and 21 weeks EGA), kidney (18 weeks EGA), and heart (15 weeks J. Zeltinger). The gene is also expressed in adult brain, heart (highest levels), kidney, skin (J. Zeltinger), and at low levels in skeletal muscle and pancreas. ED 1015 does not exhibit known phenotypic abnormalities involving the tissues TED is expressed in except those abnormalities associated with EDA. It is possible that abnormalities in addition to EDA will become apparent at a later date. It is also possible that the TED gene does not have an essential role in other tissues where it is expressed. The function of the TED gene may be redundant in various tissues. Probes from the TED gene cross hybridize with genomic DNA that is not from the DXS732E locus invoking the hypothesis that the TED gene is part of a multigene family. The

function of the TED gene may be redundant, and other members of this family that are expressed in the same tissues as the TED gene, may be able to compensate for the loss of TED function in these tissues. One gene HUMMRNAC, is a member of this gene family and is known to be expressed in the brain.

Additional evidence supporting the hypothesis that other members of this family are expressed in some of the same tissues that the TED gene is, comes from Joan Zeltinger's Northern analysis. She detected two messages (2.52-2.9 kb and 1.6-1.9 kb) in addition to the most abundant 5 kb message when using cRNA labeled probes generated from cDNA 3 and 11. These bands were not seen when cDNA 11 was used as an oligo-labeled DNA probe on Clontech Northern blots. It is possible that her cRNA labeled probes were hotter and that's why she detected additional messages. It is also possible that the additional bands were specific to the age and tissue types that she tested. This could be determined by probing her Northern with a oligo-labeled cDNA 11 probe. The fainter bands detected by the riboprobes, could represent less abundant alternately spliced TED gene messages, or they could be the result of hybridization that is not as robust because the messages are related but not identical to the probe. These messages could be from homologous genes expressed in some of the same tissues as the TED gene. The only way to prove that these band represented either alternate splicing or a member of a multigene family with similar expression, would be to clone and characterize additional cDNAs from the specific tissues with these messages. We have the HUMMRNAC clone but it is expected to hybridize to a five kb band, approximately the same size as the major band detected by cDNAs from the TED gene.

89

If the deletion in ED1015 results in juxtapositioning the TED gene to another gene, a chimeric message could be formed that allows normal processing and expression of the TED gene. This possibility seems unlikely, but would easily explain why no phenotype in addition to EDA is seen in ED1015, if the TED gene is not responsible for the EDA phenotype. If this explanation were true and a chimeric gene was formed one would still have to explain why disruption of the second gene involved in forming the chimera does not produce a phenotype in addition to EDA. The second gene could be the EDA gene, a nonfunctional pseudogene, or its product could have a redundant function.

The most convincing evidence that the TED gene may not be the EDA gene comes from the physical mapping data and the placement of patients' deletion and translocation breakpoints within the physical map. The orientation of the TED gene (Figure 6) is based upon the identification of the poly (A) tail, the results indicating that the 3' end of the TED gene is deleted for in patient ED1015, the results indicating that this gene is proximal to the AnLy translocation, and the identification of a cosmid, C03.184, which is deleted for in patient ED1015 and identifies the AnLy translocation breakpoint (Thomas et al. 1993). Additional deletions found in EDA patients also map 3' of the TED gene approximately 100-300 kb away.

The ideal candidate for the EDA gene is one that is disrupted by all the deletions and translocations. This would mean the EDA gene and its regulatory elements, span a minimum distance between the AnLy translocation breakpoint, and the LePu translocation breakpoint. The distance between these breakpoints has not been determined because the LePu translocation breakpoint has not been localized. These breakpoints may be as much as 150-200 kb apart based upon the estimated distance from the AnLy breakpoint to the DXS453 locus which maps beyond the translocation and deletion breakpoints (work done by our

collaborators). An additional translocation has been shown to map proximal to the AnLy breakpoint, AK, (Kere et al. 1993) suggesting the minimum EDA critical region is closer to 200-550 kb. AK and AnLy breakpoints are a minimum of 50 kb apart and a maximum of 350 kb apart according to Kere et al. 1993.

de Kok et al. 1995, discovered point mutations in the POU3F4 gene in multiple patients with X-linked mixed deafness. They identified a 500 kb critical region based upon characterization of eight patients' deletions. The POU3F4 gene spans approximately 120 kb of this critical region and three deletions and one duplication do not physically disrupt the POU3F4 gene. It is thought that the four rearrangements disrupt regulatory regions mapping between 15 and 400 kb away, the distances the POU3F4 gene maps from the deletions. Another explanation could be that expression of the POU3F4 gene is silenced as a result of the aberrations changing the chromosomal structure in the region. Recently, Fantes et al. in 1995 describe two families with aniridia and chromosomal rearrangements that map at least 85 kb beyond the 3' end of the PAX6 gene. Multiple intragenic PAX6 mutations have been described in patients with aniridia supporting the theory that mutations in PAX6 cause aniridia. Mutations have not been detected in the PAX6 open reading frame in the two families with rearrangements segregating with aniridia. However they cannot confirm that both alleles of the PAX6 gene were analyzed and the promoter, and untranslated region of the gene have not been analyzed. The hypothesis that PAX6 expression is silenced through changes in chromatin structure caused by the two described deletions has been suggested as the underlying mechanism of disease in these two families.

The EDA gene may only span a portion of the EDA critical region, and may not be disrupted by all of the chromosomal aberrations found in EDA patients. It is possible that some translocations and deletions disrupt an EDA

gene enhancer. In addition, deletions or translocations could result in position effects where constitutive heterochromatin is juxtapositioned close enough to heterochromatin or a silencer to affect expression of genes that are normally capable of expression. Clearly, the TED gene does not span the translocation and deletion breakpoints and the majority of these aberrations map beyond its 3' end. Only one of the nine aberrations detected in this region disrupts the TED gene. It is possible that the aberrations disrupt a regulatory element, or the chromatin structure of the TED gene. This theory could be tested by looking at whether the TED gene is expressed in the patients with chromosomal aberrations. However, the TED gene is not expressed in fibroblasts or lymphoblasts making this analysis dependent upon illegitimate transcription. Since TED expression apparently correlates with methylation status of at least two Eag I sites, looking at these sites might give an indication of whether the TED gene is expressed in patients with aberrations which don't physically disrupt the TED gene. If the TED gene is methylated at the six Eag I sites in affected males with known chromosome deletions then there would be significant evidence supporting the hypothesis that the deletions 100-300 kb away may silence the TED gene, perhaps through changes in chromatin structure as a result of enhancer elements being deleted.

The EDA critical region, defined by translocation and deletion breakpoints, is quite large and it is possible that a single gene does not span the entire length. The large EDA critical region could contain a cluster of genes involved in hair, tooth and sweat gland morphogenesis. Keratins, intermediate filaments of the cytoskeleton of epithelial cells, are encoded by a large family of approximately 20 genes. There are two intermediate filament keratin gene families, the acidic type I family and the neutral-basic type II keratins. Members of the type I keratins are clustered in sheep (Powell et al. 1986). The type II keratins are clustered in

92

sheep and in humans in the 12q11-q13 region (Wanner et al. 1993). Typically only one type I keratin and one type II keratin are expressed in a particular phenotype, suggesting coordinated regulation of the two clusters of genes. Keratins are structural proteins in hair and are expressed in hair follicles. It is possible that there are multiple genes involved in hair follicle, sweat gland, and tooth morphogenesis that are clustered in the Xq13.1 region. There are two types of clustered genes, one type involves genes that arose by gene duplication events and include the keratin genes, as well as the globin (Goodman et al. 1987) and homeobox genes (Kessel and Gruss, 1990, and Akam 1989). If the TED gene is part of the cluster of genes in the EDA region, it is unlikely that the cluster is composed of homologous genes. Genes with homology to the TED gene have not been shown to map to the EDA critical region. It is possible that one of the genes homologous to the TED gene maps to the X chromosome since two X loci were detected by pcos 169E/4 in mouse (Brockdorff et al. 1991). One would expect this homologous X sequence to be deleted in the Ta^{25H} mouse if it maps to the *tabby* region and is part of a gene cluster. The second X band appears to be present in the Ta^{25H}. No additional X loci that map to the EDA region have been detected with Southern analysis using cDNAs from the TED gene. However, the studies addressing the hypothesis that there is a cluster of homologous genes in the EDA critical region have not been done. Ideally, small probes would be used covering the entire length of the TED gene in order to detect small regions of homology between genes. Additional bands identified would then have to be map to the EDA critical region.

It is possible that a cluster of genes in the EDA region would have related functions but do not have sequence or protein homology. An example of clusters of gene with related function without homology include the housekeeping genes at the mouse surfeit locus (Colombo et al. 1992). Genes in this type of cluster,

93

are thought to remain clustered because of cis-acting elements that act on the cluster. Regulation of expression would be one reason why genes might cluster at the Xq13.1 region. In order to prove this hypothesis one would have to clone additional genes in the EDA region, and show that they are expressed during hair follicle, and sweat gland morphogenesis. If this hypothesis is correct, one might find mutations in the various genes within the cluster, including the TED gene, that result in the EDA phenotype.

In conclusion, the TED gene at the DXS732E locus cannot be ruled out as a candidate gene for EDA at this time. Its expression pattern in skin coincides with hair follicle and sweat gland morphogenesis. It is partially deleted in one EDA patient, and it seems that the deletion would interfere with normal TED gene processing. Most of the deletion and translocation breakpoints do not physically disrupt the TED gene and map about 100-300 kb beyond the 3' end. It is difficult to determine whether expression of the TED gene is silenced as a result of deletions and translocation because the TED gene is not expressed in fibroblasts or lymphoblasts. Nested primers could be used in an attempt to detect illegitimate transcription in these tissues. Identification of point mutations is necessary to provide further evidence for or against the hypothesis that the TED gene is an EDA candidate. Chemical mismatch cleavage might be the most efficient method to detect mutations in the TED gene given the large size of at least two exons (Grompe et al. 1989). PCR through the GC rich region of the putative 5' exon might be a problem since multiple attempts at PCR in this region have failed. If mutations were not found in the known sequences of the TED gene it would most likely not be the EDA gene. However, the TED gene could still not be ruled out because mutations could be found in the five prime untranslated or promoter regions that have not yet been characterized. Other candidate genes expressed at the appropriate time in fetal skin from the EDA

critical region may be identified which are disrupted by all or most of the genomic rearrangements.

Characterization of the TED gene contributes to our knowledge of the human genome. One objective of the Human Genome Project is to develop various maps of the entire human genome including a map of all expressed genes (a haploid human genome is estimated to consist of three billion base pairs, and code for approximately 100,000 expressed genes). This ambitious goal will ultimately lead to biomedical applications which prevent and treat human diseases and disorders. There are at least four thousand genetic disorders which are caused by defects in single genes. If the TED gene is not the EDA gene, it may be involved in another genetic disorder, perhaps a multifactorial disorder involving more than one locus. While the function of the TED gene is not currently known, the analyses described here likely will be useful in future investigations.

Discussion of chapter 2

The ectodermal dysplasia described in both families is most consistent with an EDA phenotype. The half-brothers from both families had abnormally low levels of sweat production, hypohidrosis, a cardinal feature of EDA. In addition, the boys had a few small or conical shaped teeth with multiple missing teeth. The half-brothers from family ED 1082 had thin, sparse, hair on both their scalp and eyebrows, well known features of EDA that were not part of the phenotype seen in the boys from family ED 1136. The boys from family ED 1082 were noted to have a characteristic facies found in EDA patients, which included frontal bossing, deep-set eyes, with mild mid-face hypoplasia. One of the boys from family ED 1136 was noted as having periorbital wrinkling, sometimes described in EDA patients. In addition to the EDA phenotype, one boy from family ED 1082 showed nail abnormalities. In general, the symptoms found in family ED 1136 were less severe than both those found in family ED 1082 and in most male patients affected with EDA. However, the severity of the phenotype in males with EDA is variable. Studies performed by Angus Clarke showed inter- and intra-familial variation. One family had affected males with no secondary teeth and males with as many as 28 teeth. Hair is usually sparse but this is not always an obvious feature. Sweat pores can be absent or less than 50% of normal, but all the patients in Clarke's study had a reduced number of sweat glands. Clearly, patients with EDA manifest the phenotype to various degrees.

The affected boys from both families had an immunodeficiency characterized by low levels of IgG and IgA, but with high levels of IgM and recurrent infections. This type of immunodeficiency is consistent with HIGM1 and has not been described in EDA patients. Affected members of family ED 1082 had a severe immunodeficiency resulting in opportunistic infections which led to

their early death despite intravenous immunoglobulin infusions and antibiotic treatment. Patients with HIGM1, frequently suffer from life threatening infections despite immunoglobulin and antibiotic treatments and the early mortality rate is approximately 10% (Kroczek et al. 1994). Treatment with intravenous gamma globulin therapy and oral antibiotic was successful for the boys from family ED 1136, suggesting that their immunodeficiency is less severe than in the boys from family 1082. Both marked variability in clinical severity and in opportunistic infections have been described in patients with HIGM.

The affected half-brothers in each family were stated to have different fathers and the same mother who was unaffected, suggesting X-linked recessive inheritance of their disorder(s). Analysis of polymorphic markers in one father, the unaffected mother and the two affected boys, were consistent with the half-brothers having different fathers. Analysis of multiple markers in the mother and her two sons from family ED 1136 were consistent with the boys having different fathers (Markers run by Todd Taylor).

There are other possible ways this disorder(s) was inherited. A new autosomal dominant mutation could have occurred in the mothers' germline resulting in mosaicism. In this case only the mothers' offspring would be affected. If the germline itself was mosaic, the mother could have both normal and affected children. Excluding germline mosaicism, the mutated allele(s) and the loci linked to the disorder are shared by the half-brothers and their mother, since it is unlikely that new mutations at the same locus occurred independently. Half-brothers share on average a quarter of their DNA with each other and half of their DNA with their mother. It is possible, that this disorder(s) is autosomal dominant with incomplete penetrance. Autosomal dominant disorders with reduced penetrance can skip generations, so the mother could be carrying the

disorder but not manifesting it. With these forms of inheritance in mind, it is most likely that these families exhibit X-linked inheritance.

Although both families had affected individuals with similar general features of an X-linked hypohidrotic ectodermal dysplasia, in addition to an immunodeficiency with hyper-IgM, it cannot be proven, without identifying the molecular defects, that these families share the same disorder. The disorders could be distinct, in which case the disorder in family 1082 could be allelic to the EDA locus and the disorder in family ED 1136 could involve the CD40L (HIGM1) locus, based upon the concordance data presented. The fact that deletions of either loci flanking the EDA locus or the HIGM1 locus were not detected in the families, is evidence against the hypothesis that these disorders result from a sub-microscopic deletion syndrome involving either the EDA or HIGM1 locus. However, family ED 1082 could have a rare mutation distinct from all the mutations that result in the EDA phenotype only. Based upon the concordance data the disorder segregating in family ED 1136 could map to the Xp21.3-q28, and the disorder segregating in family ED 1082 could map to either the Xq22.3-q26 or Xq26-q28 interval.

The differences in phenotypic severity seen in the two families could be the result of allelic heterogeneity. A correlation between allelic mutations and the severity of a disorder has been found in various disorders including Duchenne and Becker muscular dystrophy and Hurler-Scheie syndrome. If the clinical heterogeneity found in these two families results from allelic mutations and the disorder is X-linked, then the discordance data presented suggests that this immunodeficiency and hypohidrotic ectodermal dysplasia disorder is distinct from both the EDA and HIGM1. Markers at or near the EDA locus were discordant in the boys from family 1136, and markers at or near the HIGM1 locus were discordant in the boys from family 1082. If the boys have the same disorder, one

would expect each pair of boys to be concordant at and around the disease locus. The loci tested were spaced approximately 10 cM apart so that the chance of a double recombination occurring between two of these loci is about one percent. Some loci are less than 10 cM apart including the AR, PGKP1, DSX339, DXS732E and DXS453 loci which are tightly linked to the EDA locus, and the CD40L and HPRT loci which are tightly linked to the HIGM1. Not all the markers tested were informative, so the interval between some markers is greater than 10 cM. The density of informative markers is insufficient to completely exclude regions of the X chromosome based upon discordant loci. A double recombination could occur between the loci, resulting in concordant regions that would not be detected by this analysis. On average there are 1.5 recombinations along the X chromosome during meiosis. A recombination between the mother's X chromosomes can be detected by using discordance/concordance analysis of loci on the X chromosomes of her sons. One recombination in the Xq21-q22 region was indicated by a transition from concordance to discordance in the boys from family ED 1136 (Figure 7). Two recombinations were detected in the Xq24-q26 region and in the Xq26-q27 region in the boys from family ED 1082 (Figure 7). It is possible that double recombination occurred between loci in either of these families but was not detected in this analysis. The concordant/discordant analysis identified overlapping concordant regions between the pairs of boys from family ED 1136 and ED1082. The disease locus may map to the Xq21.1-q26.1 and Xq26.2-q28 regions because both pairs of half-brothers are concordant in these intervals. Both families are concordant for the DXS456, and DXS424 loci in the Xq21.1-q26.1 and for the DXS998 locus in the Xq26.2-q28 interval.

One form of ectodermal dysplasia, dyskeratosis congenita (MIM #305000), maps to the Xq28 region. Dyskeratosis congenita, (DKC), is genetically

heterogeneous, with the most common form being the X-linked recessive disorder, characterized by cutaneous pigmentation, nail dystrophy, lacrimal duct obstruction, decreased blood platelets, bone marrow hypofunction, and malignancy predisposition (Arngrimsson et al. 1993). DKC may have hypohidrotic manifestations. Immunodeficiency with hyper IgM, and hypodontia are not characteristic of DKC. Other forms of hypohidrotic ectodermal dysplasia that have not been mapped or well characterized may map to this interval.

Two immunodeficiency disorders, X-linked agammaglobulinemia, (XLA, MIM#300300), at Xq22 and X-linked lymphoproliferative disease, (XLPD, MIM#308240), at Xq25, map to the regions of concordance between the two pairs of maternally related half-brothers. Phenotypic expression of XLPD can be triggered by an Epstein-Barr viral (EBV) infection, although immune defects are present before EBV infection (Grierson et al. 1991). Following EBV infection, affected males may develop hypogammaglobulinemia, or non-Hodgkin's lymphoma, but most affected individuals develop a fatal infectious mononucleosis. Hyper IgM has been reported to occur after EBV infection in a few individuals, and Grierson et al. in 1991 found IgM levels to be elevated at times in eight out of thirteen XLPD males that were not infected with EBV. Hypohidrotic ectodermal dysplasia is not part of the XLPD disorder. However, a deletion could affect both this locus and an ED locus, presently unknown to map to this region, resulting in the phenotype seen in the two families. Three large overlapping deletions have been identified in unrelated males affected with XLPD with both Southern analysis and chromosome banding (Skare et al. 1993). Two of the three boys were healthy with elevated levels of IgM, the third boy died following EBV infection. It is felt that the two boys will remain healthy as long as they escape EBV infection. The two brothers were identified because they had brothers with XLPD that died following EBV infection. The deletions in these

boys were cytogenetically detectable, yet there was no evidence for a contiguous gene deletion syndrome. The families studied with a hypohidrotic ectodermal dysplasia and immunodeficiency with hyper IgM had normal karyotypes. Assuming the karyotypes were done at the resolution level needed to visualize Xq25 deletions in these families, it is unlikely that a cytogenetically undetectable deletion in these families could result in a contiguous gene syndrome when there are three affected boys with cytogenetically detectable deletions without any phenotype in addition to XLPD. Markers known to be linked to the XLPD locus were not typed in the two families, however surrounding loci are concordant. Without typing the loci linked to XLPD one cannot rule out a microdeletion.

X-linked agammaglobulinemia is characterized by profound hypogammaglobulinemia and lack of mature B cells. The disorder is treated effectively with immunoglobulin infusions. Patients are prone to bacterial but not viral infections. The Bruton's agammaglobulinemia tyrosine kinase gene, BTK, was found to have deficient expression and to be mutated in patients with XLA (Tsukada et al. 1993, Vetrie et al. 1993). Ectodermal dysplasia and hyper IgM are not part of the XLA phenotype.

There are a few mouse disorders that include immunodeficiency and abnormalities in hair, but none of them map to the X chromosome. The *motheaten*, *me*, mutation maps to chromosome 6 (Green et al. 1976). The *me* phenotype includes skin lesions displacing hair follicles and causing areas of hypopigmentation (Sundberg 1994, p 351. The *nude*, *nu*, maps to chromosome 11 (Sundberg 1994, p 371). Nude mice develop short thin hair with a reduced number of whiskers (Sundberg 1994, p379). The *hairless*, *hr*, mutation maps to chromosome 14 in the mouse (Morrissey et al. 1980). *Hairless* mice loose their hair at about 14 days, and their immunodeficiency is less severe than the other mutations mentioned here (Sundberg 1994, p291).

101

Allelic heterogeneity could explain involvement of any of the DKC, XLPD or XLA loci. Analysis of more families with phenotypes similar to those in the families described here are needed in order both to establish whether these families have the same disorder, and to determine which locus or loci are involved. Two additional families with a similar phenotype have been reported to us, supporting the hypothesis that the phenotype seen results from a single disorder. One family seen at UCSF reffered to us, has two affected maternally related half-brothers. The other family was seen in France had two deceased affected brothers. Results from DNA analysis of X-linked polymorphic markers in these families should help us narrow the candidate regions.

Figure 1 Sequence comparison of mouse genomic clone pcos 169E/4 and homologous human clone HO41.43-9

Comparison of homologous sequence between human genomic clone HO41.43-9, bottom sequence, and mouse genomic clone pcos 169E/4, upper underlined sequence. GRAIL rating of 250 bp region with coding potential in bold print. **CAG//** indicates the most 3' splice site and **TGA** is the putative stop codon. * indicates homology between the mouse and human sequences. RTPCR human primers Z17F and Z17R are indicated with a line above the lower case human sequence.

103

```
    AAGCTGGAGC-TCCACCGCGGTGCACCGGAAACCAAGTGCTGTGATGTGC
    * * * * * *     *  * * * *        *  *    *  * * * * * * * * * * * * * * * *    * * * *
CAG//GGTTGCTGGATACTTCACCAAAGCGT-CTGGAAACcaagtgctgtgacgtgc

AGTGGGTCTCCTGTGAGTCGGAGAAGAAGAAGTTCAAGGACTCTGAGCCC
* * * * * * * * * * * * * * * *  * *  * * * * * * * * * * * * * * * *  * * * * * *  * *
agtgGGTCTCCTGTGAGGCGAAGAAGAAGAAGTTCAAGGAGTCTGAGGCC

CCTAAAACCCACCACCAGCAATTCCACCACTCCTATTTTCACCACTACCA
* *    * * * * * * * * * *    * * * * * * * * * * * * * * * * * * * * * * *    * * * * * * * * * * *
CCCAAAACCCACCAGCAGCAATTCCACCACTCCTATTTCCACCActacca

CCATCAGTACCACCACTACCATCCCCGCCATGAACCCCCAAGCCGAGTCA
* * *    * * * * * * * *    * * * * * * * *    * * * *     * * * * *   * * * * * *    * * * *    * * * *
ccaacagtaccatcactACCACCCCCATCATGATCCCCCAGGCCGTGTCA

GCAACAAGCCCTCCCTGCTGCCGGTCTCTGGGGGCTCCC--CTCAGTCCC
* * * * * * * * * *    * * * * * * * * * * * * * * * * * * * * * * * * *    * * * * *    * * *
GCAACAAGCCCGCCCTGCTGCCGGTCTCTGGGGGCTCCCGCCTCAGCCCT

AGCAGGATCCGGCTCTGTGTCCTTGCTCTTATACTCCTCCATACAGTAGT
* * * * * * * * * * * * * * * * * *  * * * * * * *    * * *    * *    * * * * * * * * * *    * *    * *
AGCAGGATCCGGCTCTGCGTCCTTGTTCTCATGCTCCTCCATACCGTGGT

GTCCTTCTCCAGCAGCCAGAGTGATGGGGGACTGGGACTGGAGACACTGC
* * * * * * * * * * * * *  * * * *   * * *   * * * * * * *   * * * *   * * * * * * * * * * * * *
GTCCTTCTCCAGCAACCAGGGTGGTGGGGGATTGGGGCTGGAGACACTGC

CTGCCCTCGAGGAGGGGTTGATGTAGGAAGAGTGACCACAGGGAGGGAGG
* * * * * * *  * * * * * * * *    * * *    * * * * * * * * * * *    * * * * * * * * * * *
CTGCCCTAGAGGAGGGCCTGACACGGGAAGAGTGACAGTAGGGAGGGAGG

ACAAACCTCCACCACATACTGACATCAGCTCCAGCCCAGCTCCAGGCTGG
* * *   * * * * * * * * * * *    * * * * * * * * * * * * * * * *    *    *   * * * * *   * * *
ACAGACCTCCACCACA--CTGACATCAGCTCCAGCTC--CCCCAGGTTGG

GGAAGGAA--GCATCTCCCACAGGAGGTGTAGGATCAGGTGGGGGTGGGG
* *    *       * *    * * * * * *    * * * * * * * * * * * * *   * * * * * * * * *   * * * *
GGGGGAGGGGGCTCCTCCCATGGGAGGTGTAGGATAAGGTGGGGGNGGGG

AGAAATGGGGGACCAG-ATATCGT-CCCAATCTACCCCTGCCTATGAAAG
  * *   * * * * * * *       *   * * *     * * * * *   * * *   * * *    * *         * * * *
NAAA-TGGGGGAATGACACATCCNNCCCAACCTANCCCCACCCCA-AAAG

CAGCTTCAACNAAATGAC-AGAGGGCCCCCAAGGAGCTATAAAACTCACC
* * * * *  * * * *    * * * *  *  * * * * * * * *   *   *   * * * * * *   * * * * * * * *   *
CAGCTCCAACAGAATGGCCAGAGGGCCTC-AGGGAGCTGCAAAACTCATC
```

CAGTAGGGAAAGGTAGGAGCAGCAGGTGTCCCCTCCCAAGC-CCCACCTT
*** ** ***** ********** *** ************ **** **
CAGGAGAAAAAGGCAGGAGCAGCAAGTGACCCCTCCCAAGCTCCCATCTA

TGGGGCTTAGTTCTNAAAAGGGGGA----GGGGCTGGAGTTGCCCACTCC
********* *** ** *** ********* ******* **
TGGGGCTTAGCAAAAAAAGGGAGGAAGTGGGGGCTGGATATGCCCACCCC

TGCCAAAAGCCCTGACCCCAGGGAAGGAGGCTGCTTGCTCAGCCTCAGCCA
********************************** * ****** ***
TGCCAAAAGCCCTGACCCCAGGGAAGGAGGCTGCTCACCCAGCCTTGGCCC

TGCAGGGAATGTTGGGGGGCACAGAGGGGAGANNTCTNN

Figure 2 Start codon analysis

Translation in three frames near the putative start methionines in bold and underlined. A stop codon in the same frame before the start is in bold and underlined.

Start Codon Analysis

<pre>
                           27   54
GTC TGG CTC GCG CCT GCC TGT TCC CTC CAG CCC GGA CCC CCC **TGA** AAT **ATG** TTC
Val Trp Leu Ala Pro Ala Cys Ser Leu Gln Pro Gly Pro Pro . Asn <u>**MET** Phe</u>
 Ser Gly Ser Arg Leu Pro Val Pro Ser Ser Pro Asp Pro Pro Glu Ile Cys Ser
  Leu Ala Arg Ala Cys Leu Phe Pro Pro Ala Arg Thr Pro Leu Lys Tyr Val Gln

                           81   108
AGG GGC GCT TGG **ATG** TGG CCC GGG AAA GAC GCC GCC GCG CTG ACT ATC TGC TGC
<u>Arg Gly Ala Trp **MET** Trp Pro Gly Lys Asp Ala Ala Ala Leu Thr Ile Cys Cys</u>
 Gly Ala Leu Gly Cys Gly Pro Gly Lys Thr Pro Pro Arg . Leu Ser Ala Ala
  Gly Arg Leu Asp Val Ala Arg Glu Arg Arg Arg Ala Asp Tyr Leu Leu Leu
</pre>

Two Potential Start Methionines

Underlined Frame Rated Excellent By GRAIL

Start Methionines Preceeded By Stop Codon

Figure 3 Stop codon analysis

Analysis of the putative stop codon in bold. The putative ORF is underlined.
Notice the stops in frame 2 before and after the putative stop codon. Frame 3
contains several stops.

```
                1431                                                1458
GGT GGT GGG GGA TTG GGG CTG GAG ACA CTG CCT GCC CTA GAG GAG GGC CTG ACA
CCA CCA CCC CCT AAC CCC GAC CTC TGT GAC GGA CGG GAT CTC CTC CCG GAC TGT
Gly Gly Gly Gly Leu Gly Leu Glu Thr Leu Pro Ala Leu Glu Glu Gly Leu Thr
Val Val Gly Asp Trp Gly Trp Arg His Cys Leu Pro  .  Arg Arg Ala  .  His
Trp Trp Gly Ile Gly Ala Gly Asp Thr Ala Cys Pro Arg Gly Gly Pro Asp Thr


                1485                                                1512
CGG GAA GAG TGA CAG TAG GGA GGG AGG ACA GAC CTC CAC CAC ACT GAC ATC AGC
GCC CTT CTC ACT GTC ATC CCT CCC TCC TGT CTG GAG GTG GTG TGA CTG TAG TCG
Arg Glu Glu  .  Gln  .  Gly Gly Arg Thr Asp Leu His His Thr Asp Ile Ser
Gly Lys Ser Asp Ser Arg Glu Gly Gly Gln Thr Ser Thr Thr Leu Thr Ser Ala
Gly Arg Val Thr Val Gly Arg Glu Asp Arg Pro Pro Pro His  .  His Gln Leu


                1539                                                1566
TCC AGC TCC CCC AGG TTG GGG GGG AGG GGC TCC TCC CAT GGG AGG TGT AGG ATA
AGG TCG AGG GGG TCC AAC CCC CCC TCC CCG AGG AGG GTA CCC TCC ACA TCC TAT
Ser Ser Ser Pro Arg Leu Gly Gly Arg Gly Ser Ser His Gly Arg Cys Arg Ile
Pro Ala Pro Pro Gly Trp Gly Gly Gly Ala Pro Pro MET Gly Gly Val Gly  .
Gln Leu Pro Gln Val Gly Gly Glu Gly Leu Leu Pro Trp Glu Val  .  Asp Lys


                1593                                                1620
AGG TGG GGG CGG GGG AAA TGG GGG AAT GAC ACA TCC CCC CCA ACC TAC CCC CAC
TCC ACC CCC GCC CCC TTT ACC CCC TTA CTG TGT AGG GGG GGT TGG ATG GGG GTG
Arg Trp Gly Arg Gly Lys Trp Gly Asn Asp Thr Ser Pro Pro Thr Tyr Pro His
Gly Gly Gly Gly Gly Asn Gly Gly MET Thr His Pro Pro Gln Pro Thr Pro Thr
Val Gly Ala Gly Glu MET Gly Glu  .  His Ile Pro Pro Asn Leu Pro Pro Pro


                1647                                                1674
CCC AAA AGC AGC TCC AAC AGA ATG GCC AGA GGG CCT CAG GGA GCT GCA AAA CTC
GGG TTT TCG TCG AGG TTG TCT TAC CGG TCT CCC GGA GTC CCT CGA CGT TTT GAG
Pro Lys Ser Ser Ser Asn Arg MET Ala Arg Gly Pro Gln Gly Ala Ala Lys Leu
Pro Lys Ala Ala Pro Thr Glu Trp Pro Glu Gly Leu Arg Glu Leu Gln Asn Ser
Gln Lys Gln Leu Gln Gln Asn Gly Gln Arg Ala Ser Gly Ser Cys Lys Thr His


                1701                                                1728
ATC CAG GAG AAA AAG GCA GGA GCA GCA AGT GAC CCC TCC CAA GCT CCC ATC TAT
TAG GTC CTC TTT TTC CGT CCT CGT CGT TCA CTG GGG AGG GTT CGA GGG TAG ATA
Ile Gln Glu Lys Lys Ala Gly Ala Ala Ser Asp Pro Ser Gln Ala Pro Ile Tyr
Ser Arg Arg Lys Arg Gln Glu Gln Gln Val Thr Pro Pro Lys Leu Pro Ser MET
Pro Gly Glu Lys Gly Arg Ser Ser Lys  .  Pro Leu Pro Ser Ser His Leu Trp


                1755                                                1782
GGG GCT TAG CAA AAA AAG GGA GGA AGT GGG GGC TGG ATA TGC CCA CCC CTG CCA
CCC CGA ATC GTT TTT TTC CCT CCT TCA CCC CCG ACC TAT ACG GGT GGG GAC GGT
Gly Ala  .  Gln Lys Lys Gly Gly Ser Gly Gly Trp Ile Cys Pro Pro Leu Pro
Gly Leu Ser Lys Lys Arg Glu Glu Val Gly Ala Gly Tyr Ala His Pro Cys Gln
Gly Leu Ala Lys Lys Gly Arg Lys Trp Gly Leu Asp MET Pro Thr Pro Ala Lys
```
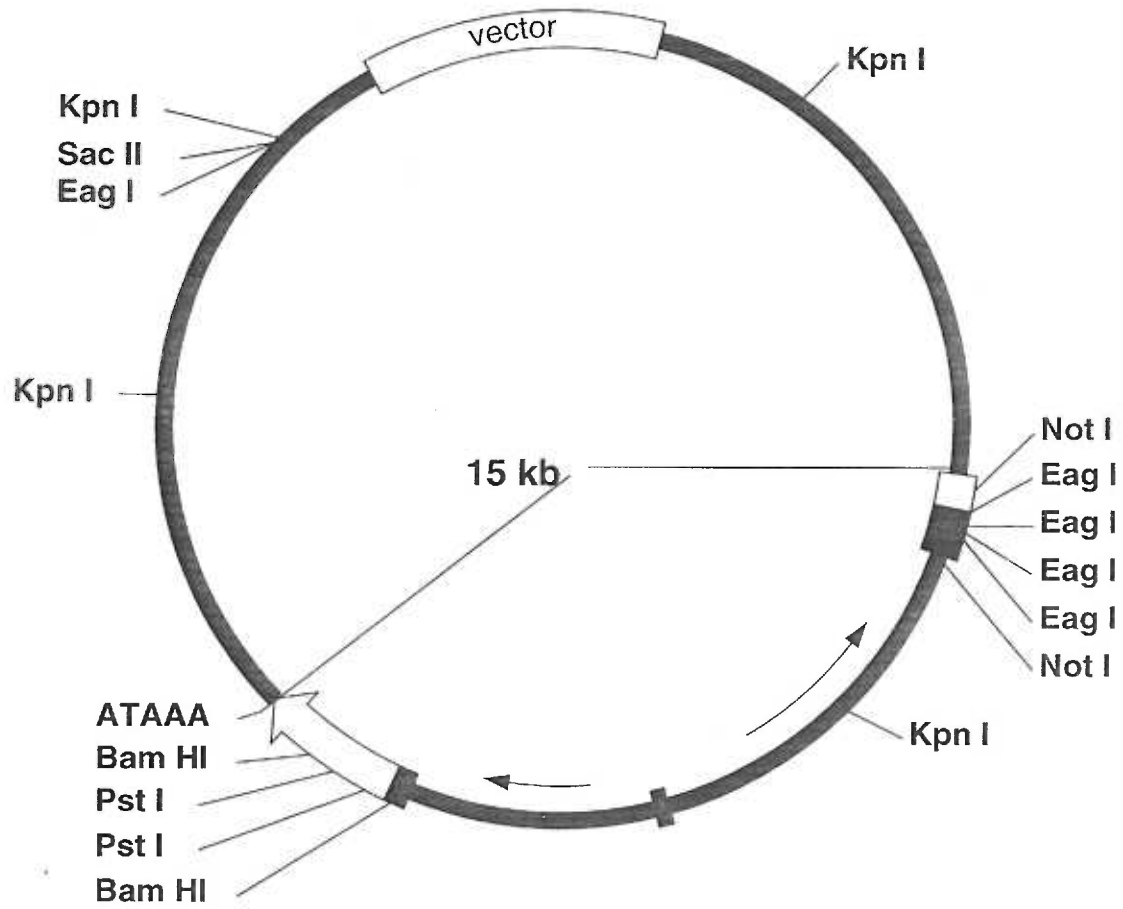
Figure 4 Map of TED gene on cosmid ICRFc104HO41.43

The positions in () of restriction sites relative to the Eag I site in the vector sequence. The cosmid is approximately 40 kb, and the TED gene isolated to date, spans approximately 15 kb of genomic sequence. The TED gene, . is TED gene coding sequence. The 5' exon, at least 910 bp, and most 3' exon, 2964 bp, have been mapped as shown; the intervening exon(s), , totaling 169 bp, have not been localized within the eleven kb genomic fragment. K= Kpn I sites. The first CpG island contains at least six Eag I sites, two of which are also Not I sites, and at least three Sma I sites. The second CpG island contains at least one Eag I site and one Sac II site.

The TED gene map within cosmid
ICRFc104HO41.43

vector

Kpn I

Kpn I
Sac II
Eag I

Kpn I

15 kb

Not I
Eag I
Eag I
Eag I
Eag I
Not I

Kpn I

ATAAA
Bam HI
Pst I
Pst I
Bam HI

Figure 5 Map of Deletion in patient ED 1015

Top of figure indicates patient ED 1015 is not deleted for the first identified exon. His deletion breakpoint, indicated by dashed line, lies within the ten kb of genomic sequence between the first identified exon and the last identified exon. This patient is deleted for the entire 3' exon including part of domain 2, 388 bp of coding sequence, the 3' untranslated sequence and the poly (A) addition site. K = Kpn I sites. E = Eag I sites, B= BamHI sites and P= Pst I sites.

The Proximal Deletion Breakpoint in Patient ED 1015

ED 1015

1st CpG island   2nd CpG island

K   K   K   K

0   5 kb

Genomic Map

TED cDNA

E   E E E   B   P   P   B

splice
sites

0   0.5 kb

ED 1015

PROBES

cDNA 3

cDNA 11

pcos 169 E/4

Scale (top): 0 — 100 — 200 — 300 — 400 kb

CEN — DXS732 — DXS1690 — DXS1689 — DXS453 — TEL

TED gene → (A)n

AK Translocation

AnLy Translocation

LePu Translocation

CONTIGS: A B C D

YACS: 4757

DELETIONS:
ED 1015
ED 1003
ED 1058
ED 1008
ED 25

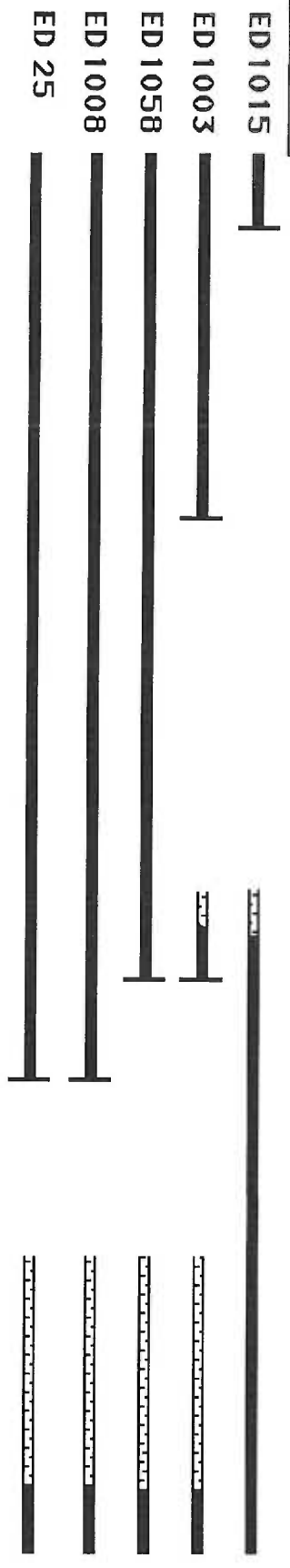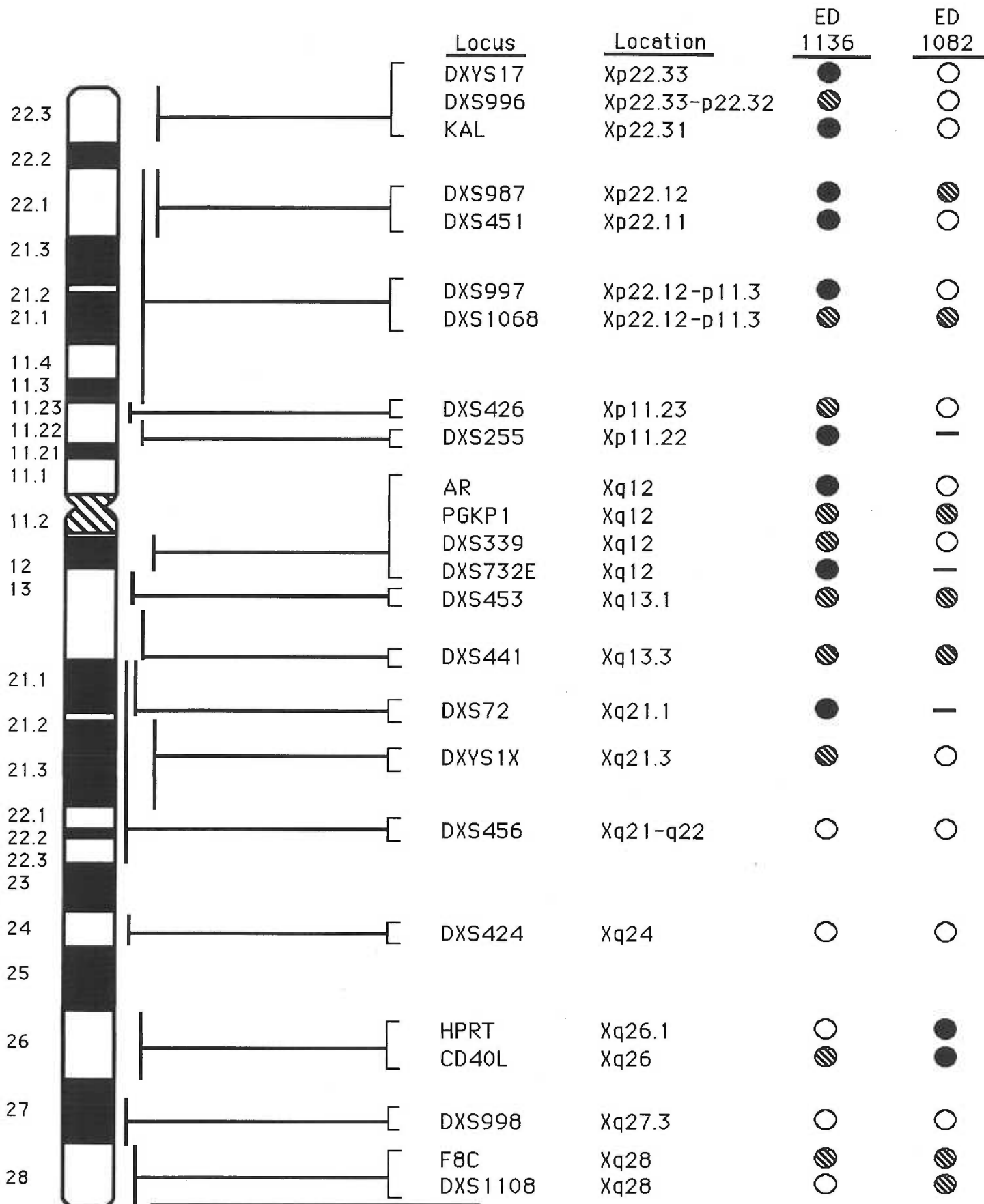| Locus	Location	ED 1136	ED 1082
DXYS17	Xp22.33	●	○
DXS996	Xp22.33-p22.32	◍	○
KAL	Xp22.31	●	○
DXS987	Xp22.12	●	◍
DXS451	Xp22.11	●	○
DXS997	Xp22.12-p11.3	●	○
DXS1068	Xp22.12-p11.3	◍	◍
DXS426	Xp11.23	◍	○
DXS255	Xp11.22	●	—
AR	Xq12	●	○
PGKP1	Xq12	◍	◍
DXS339	Xq12	◍	○
DXS732E	Xq12	●	—
DXS453	Xq13.1	◍	◍
DXS441	Xq13.3	◍	◍
DXS72	Xq21.1	●	—
DXYS1X	Xq21.3	◍	○
DXS456	Xq21-q22	○	○
DXS424	Xq24	○	○
HPRT	Xq26.1	○	●
CD40L	Xq26	◍	●
DXS998	Xq27.3	○	○
F8C	Xq28	◍	◍
DXS1108	Xq28	○	◍

KEY ● = Discordant
○ = Concordant
◍ = Not Informative

References

Akam, M (1989) Hox and HOM: homologous gene clusters in insects and vertebrates. Cell 57:347-9.

Allen RC, R.J. Armitage, M.E. Conley, H. Rosenblatt, N.A. Jenkins, N.G. Copeland, M.A. Bedell, et al (1993) CD40 ligand gene defects responsible for X-linked hyper-IgM syndrome. Science 259:990-993

Altschul SF, W. Gish, W. Miller, E.W. Myers, D.J. Lipman (1990) Basic local alignment search tool. J Mol Biol 215:403-10.

Antequera F, A. Bird (1993) Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A 90:11995-9.

Arngrimsson R, I. Dokal, L. Luzzatto, J.M. Connor (1993) Dyskeratosis congenita: three additional families show linkage to a locus in Xq28. J Med Genet 30:618-619

Aruffo A, M. Farrington, D. Hollenbaugh, X. Li, A. Milatovich, S. Nonoyama, J. Bajorath, et al (1993) The CD40 Ligand, gp39, is defective in activated T cells from patients with X-linked hyper-IgM syndrome. Cell 72:291-300

Avery JK (1994) Oral Developement and Histology 2nd edition. Thieme Medical Publishers, Inc. New York 70-93.

Blecher S R, M. Debertin, J.S. Murphy, (1983) Pleiotropic effect of Tabby gene on epidermal growth factor-containing cells of mouse submandibular gland. Anat Rec 207:25-9.

Blecher SR (1986) Anhidrosis and absence of sweat glands in mice hemizygous for the Tabby gene: supportive evidence for the hypothesis of homology between Tabby and human anhidrotic (hypohidrotic) ectodermal dysplasia (Christ-Siemens-Touraine syndrome). J Invest Dermatol 87:720-2.

Blecher SR, J. Kapalanga, D. Lalonde (1990) Induction of sweat glands by epidermal growth factor in murine X-linked anhidrotic ectodermal dysplasia. Nature 345:542-4.

Bouloux PMG, J.P. Hardelin, P. Munroe, J.M.W. Kirk, R. Legouis, J. Levilliers, J. Hazan, et al (1991) A dinucleotide repeat polymorphism at the Kallmann locus (Xp22.3). Nucleic Acids Res 19:5453

Brockdorff N, G. Kay, B.M. Cattanach, S. Rastan (1991) Molecular genetic analysis of the Ta25H deletion: evidence for additional deleted loci. Mamm Genome 1:152-7.

Brockdorff N, M. Montague, S. Smith, S. Rastan (1990) Construction and analysis of linking libraries from the mouse X chromosome. Genomics 7:573-8.

Brooks EG, G.R. Klimpel, S.E. Vaidya, S.E. Keeney, S. Raimer, A.S. Goldman, F.C. Schmalstieg, (1994) Thymic hypoplasia and T-cell deficiency in ectodermal dysplasia: case report and review of the literature. Clin Immunol Immunopathol 71:44-52.

Brown CJ, H.F. Willard, (1990) Localization of a gene that escapes inactivation to the X chromosome proximal short arm: implications for X inactivation. Am J Hum Genet 46:273-279.

Browne D, J. Zonana, M. Litt (1992a) Dinucleotide repeat polymorphism at the PGK1P1. Nucleic Acids Res 20:1169

Browne D, M. Litt (1992b) Dinucleotide repeat polymorphisms at the DXS365, DXS443 and DXS451 loci. Hum Mol Gen 1:213

Browne DL, J. Zonana, M. Litt (1991) Dinucleotide repeat polymorphism at the DXYS1X locus. Nucleic Acids Res 19:1721

Cattanach BM, C. Rasberry, E.P. Evans, L. Dandolo, M.C. Simmler, P. Avner (1991) Genetic and molecular evidence of an X-chromosome deletion

spanning the tabby (Ta) and testicular feminization (Tfm) loci in the mouse. Cytogenet Cell Genet 56:137-43.

Chuong CM, H.M. Chen, T.X. Jiang, J. Chia (1991) Adhesion molecules in skin development: morphogenesis of feather and hair. Ann N Y Acad Sci 642:263-80.

Chuong CM, R.B. Widelitz, T.X. Jiang (1993) Adhesion molecules and homeoproteins in the phenotypic determination of skin appendages. J Invest Dermatol 101:10S-15S.

Clarke A, D.I.M. Phillips, R. Brown, P.S. Harper (1987) Clinical aspects of X-linked hypohidrotic ectodermal dysplasia. Arch Dis Child 62:989-996

Colombo P, J. Yon, K. Garson, M. Fried (1992) Conservation of the organization of five tightly clustered genes over 600 million years of divergent evolution. Proc Natl Acad Sci U S A 89:6358-62.

Conley, M.E. (1994) X-linked immunodeficiencies. Curr Opin Genet Dev 4:401-6.

Crump IA, D.M. Danks (1971) Hypohidrotic ectodermal dysplasia. J Pediatr 78:466-473.

Darnell J, H. Lodish, D. Baltimore (1990) Molecular Cell Biology. W. H. Freeman and Company, New York

Davidson R, J.M. Connor (1988) Dyskeratosis congenita. J Med Genet 25:843-846

Davis JR, L.M. Solomon (1976) Cellular Immunodefieciency in Anhidrotic Ectoderma Dysplasia. Acta dermatovener (Stockholm) 56:115-120.

De Jager, H. (1965) Congenital anhidrotic ectodermal dysplasia: case report. J Pathol Bacteriol 90:321-322.

Decorte, R. Identification of internal variation in the psuedoautosomal VNTR DXYS17 with non-random distribution of the alleles on the X and the Y chromosome. (submitted)

DiSanto JP, J.Y. Bonnefoy, J.F. Gauchat, A. Fischer, G. de Saint Basile (1993) CD40 ligand mutations in X-linked immunodeficiency with hyper-IgM. Nature 361:541-543

du-Cros, D.L. (1993a) Fibroblast growth factor and epidermal growth factor in hair development. J Invest Dermatol 101:106S-113S.

du-Cros, D.L. (1993b) Fibroblast growth factor influences the development and cycling of murine hair follicles. Dev Biol 156:444-53.

Edwards A, H.A. Hammond, L. Jin, T. Caskey, R. Chakraborty (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241-253

Ellis SG, H. Ahmed (1993) Hypohydrotic ectodermal dysplasia affecting a female patient. Dent Update 20:447-50.

Ellison KA, C.P. Fill, J. Terwilliger, L.J. DeGennaro, A. Martin-Gallardo, M. Anvret, A.K.Percy, et al. (1992) Examination of X chromosome markers in Rett Syndrome: exclusion mapping with a novel variation on multilocus linkage analysis. Am J Hum Genet 50:278-287

Falconer, D.S. (1952) A totally sex-linked gene in the house mouse. Nature 169:664-665.

Fantes J, B. Redeker, M. Breen, S. Boyle, J. Brown, J. Fletcher, S. Jones, et al. (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum Mol Genet 4:415-422.

Fraser NJ, Y. Boyd, G.G. Brownlee, I.W. Craig (1987) Multi-allelic RFLP for M27ß, an anomymous single copy genomic clone at Xp11.3-Xcen [HGM9 provisional no. DXS255]. Nucleic Acids Res 15:9616

Freire-Maia N, M. Pinheriro (1984) Ectodermal Dysplasias: A Clinical and Genetic Study. Alan R. Liss, Inc., New York

Frosch P.J. (1985) Manifestation of the lines of Blaschdo in women heterozygous for X-linked hypohidrotic ectodermal dysplasia. Clin Genet 27:468-471.

Fuleihan R, N. Ramesh, R. Loh, H. Jabara, F.S Rosen, T. Chatila, S.M. Fu, et al (1993) Defective expression of the CD40 ligand in X chromosome-linked immunoglobulin deficiency with normal or elevated IgM. Proc Natl Acad Sci USA 90:2170-2173

Fulginjiti V.A., W.E. Hathaway, D.S. Pearlman, C.H. Kempe (1967) Agammaglobulinemia and achondroplasia. Br M ed J 2:242.

Gatti R.A., N. Platt, H.H. Pomerance, R. Hong, L.O. Langer, H.E.M. Kay, RA. Good (1969) Hereditary lymphopenic agammaglobulinemia associated with a distinctive form of short-limbed dwarfism and ectodermal dysplasia. J Pediatr 75:675-684.

Gold RJM, C.R. Scriver (1971) The characterization of hereditary abnormalities of keratin: Clouston's ectodermal dysplasia. Birth Defects, Orig. Artic. Ser. 7:91-95.

Goodman M, J. Czelusniak, B. F. Koop, D. A. Tagle, J. L. Slightom (1987) Globins: a case study in molecular phylogeny. Cold Spring Harb Symp Quant Biol 52:875-90.

Goodship J, S. Malcolm, A. Clarke, M. E. Pembrey (1990) Possible genetic heterogeneity in X linked hypohidrotic ectodermal dysplasia. J Med Genet 27:422-5.

Goodship J, S. Malcolm, A. Clarke, M.E. Pembrey (1990) Possible genetic heterogeneity in X linked hypohidrotic ectodermal dysplasia. J Med Genet 27:422-425

Green MC, L.D. Shultz (1975) Motheaten, an immunodeficient mutant of the mouse. I. Genetics and pathology. J Hered 66:250-258.

Grierson HL, J. Skare, J. Hawk, M. Pauza, D.T. Purtilo (1991) Immunoglobulin class and subclass deficiencies prior to Epstein-Barr virus infection in males with X-linked lymphoproliferative disease. Am J Med Genet 40:294-297.

Grompe M, D.M. Muzny, C.T. Caskey, (1989) Scanning detection of mutations in human ornithine transcarbamoylase by chemical mismatch cleavage. Proc Natl Acad Sci USA 86:5888-5892.

Gruneberg, H. (1966) The molars of the tabby mouse, and a test of the 'single-active X-chromosome' hypothesis. J Embryol Exp Morphol 15:223-44.

Guo L, Q. C. Yu, E. Fuchs (1993) Targeting expression of keratinocyte growth factor to keratinocytes elicits striking changes in epithelial differentiation in transgenic mice. Embo J 12:973-86.

Happle, R., B. Kindred (1967) Some observations on the skin and hair of Tabby mice. J Hered 58:197-199.

Hardy, M. H. (1992) The secret life of the hair follicle. Trends Genet 8:55-61.

Harteveld CL, M. Losekoot, H. Haak, J.G.A.M. Heister, P.C. Giordando and LF. Bernini (1994) A novel polyadenylation signal mutation in the a_2-globin gene causing a thalassaemia. Br J Haemat 87, 139-143.

Hartley DA, K. E. Davies, D. Drayna, R. L. White, R. Williamson (1984) A cytological map of the human X chromosome--evidence for non-random recombination. Nucleic Acids Res 12:5277-85.

Heikinheimo, K. (1994) Stage-specific expression of decapentaplegic-Vg-related genes 2, 4, and 6 (bone morphogenetic proteins 2, 4, and 6) during human tooth morphogenesis. J Dent Res 73:590-7.

Hoath, S.B. (1985) Treatment of the neonatal rat with epidermal growth factor: Differences in time and organ response. Pediatr Res 20:468-72.

Huang THM, R.W. Cottingham, D.H. Ledbetter, H.Y. Zoghbi (1992) Genetic mapping of four dinucleotide repeat loci, DXS453, DXS458, DXS454, DXS424, on the X chromosome using multiplex polymerase chain reaction. Genomics 13:375-380

Huntley CC, R. M. Ross (1981) Anhidrotic ectodermal dysplasia with transient hypogammaglobulinemia. Cutis 28:417-9.

Jhaveri M, S.H. Barnett, M. Gertn (1975) Ectrodactyly, ectodermal dysplasia, and clefting associated withthymic aplasia. Am J Dis Child 143:12.

Kapalanga J, S. R. Blecher (1991) Histological studies on eyelid opening in normal male mice and hemizygotes for the mutant gene Tabby (Ta) with and without epidermal growth factor treatment. Exp Eye Res 52:155-66.

Keer JT, R. M. Hamvas, N. Brockdorff, D. Page, S. Rastan, S. D. Brown (1990) Genetic mapping in the region of the mouse X-inactivation center. Genomics 7:566-72.

Kere J, K.H. Grzeschik, J. Limon, M. Gremaud, D. Schlessinger, A. de La Chapelle (1993) Anhidrotic ectodermal dysplasia gene region cloned in yeast artificial chromosomes. Genomics 16:305-310

Kessel M, P. Gruss (1990) Murine developmental control genes. Science 249:374-9.

Kolvraa S, T.A. Kruse, P.K.A. Jensen, K.H. Linde, S.R. Vestergaard, L. Bolund (1986) Close linkage between X-linked ectodermal dysplasia and a cloned DNA polymorphism in the region Xp11-p12. Hum Genet 74:284-7.

Kopan R, H. Weintraub (1993) Mouse notch: expression in hair follicles correlates with cell fate determination. J Cell Biol 121:631-41.

Korthauer U, D. Graf, H.W. Mages, F. Briere, M. Padayachee, S. Malcolm, A.G. Ugazio, et al (1993) Defective expression of T-cell CD40 ligand causes X-linked immunodeficiency with hyper-IgM. Nature 361:539-540

Kroczek RA, D. Graf, D. Brugnoni, S. Giliani, U. Korthuer, A. Ugazio, G. Senger, et al. (1994) Defective expression of CD40 ligand on T cells causes "X-linked immunodeficiency with hyper-IgM (HIGM1)". Immunol Rev 138:39-59.

Lalloz M, J.H. McVey, J.K. Pattinson, E. Tuddenham (1991) Haemophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene. The Lancet 338:207-211

Lewin, B. (1994) Genes V. Oxford University Press Inc., New York

Li S, M.G. McInnis, R.L. Margolis, S.E. Antonarakis, C.A. Ross (1993) Novel Triplet Repeat Containing Genes in Human Brain: Cloning, Expression, and Length Polymorphisms. Genomics 16:572-579.

Litt M., J.H. Eubanks, G.A. Evans, T. Phromchotikul (1992) Regional localization of the highly polymorphic locus D11S533 on the linkage map of human chromosome 11q. Genomics 14:207-211

Luetteke NC, T. H. Qiu, R. L. Peiffer, P. Oliver, O. Smithies, D. C. Lee (1993) TGF alpha deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice. Cell 73:263-78.

Luty J.A., Z. Guo, H.F. Willard, D.H. Ledbetter, S. Ledbetter, M. Litt (1990) Five polymorphic microsatellite VNTRs on the human X chromosome. Am J Hum Genet 46:776-783

Lux S.E., R.B. Johnston, C.S. August, B. Say, V.B. Penchaszadeh, F.S. Rosen, V.A. McKusick (1970) Chronic neutropenia and abnormal cellular immunity in cartilage-hair hypoplasia. N Engl J Med 282:231-236.

MacDermot KD, M. Hult'en (1990) Female with hypohidrotic ectodermal dysplasia and de novo (X;9) translocation. Clinical documentation of the AnLy cell line case. Hum Genet 84:577-9.

117

Mayer TC, C. K. Miller, M. C. Green (1977) Site of action of the crinkled (cr) locus in the mouse. Dev Biol 55:397-401.

Messenger, A. G. (1993) The control of hair growth: an overview. J Invest Dermatol 101:4S-9S.

Miller SA, D.D. Dykes, and H.F. Polesky (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215

Morrissey P.J., D.R. Parkinson, R.S. Schwartz, S.D. Waksal (1980) Immunological abnormalities in HRS/J mice. I. Specific deficit in T lymphocyte helper function in a mutant mouse. J Immunol 125:1558-1562.

Muller M, J. R. Jasmin, R. A. Monteil, R. Loubiere (1991) Embryology of the hair follicle. Early Hum Dev 26:159-66.

Nizetic D, G. Zehetner, A.P. Monaco, L. Gellen, B.D. Young, H. Lehrach (1991) Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: Their potential use as reference libraries. Proc Natl Acad Sci USA 88:3233-3237.

Notarangelo LD, M. Duse, and A.G. Ugazio (1992) Immunodeficiency with hyper-IgM (HIM). Immunodeficiency Reviews 3:101-122

Padayachee M, R.J. Levinsky, C. Kinnon, A. Finn, C. McKeown, C. Feighery, L.D. Notarangelo, et al (1992) Mapping of the X linked form of hyper IgM syndrome (HIGM1). J Med Genet 30:202-205

Passarge E, C.T. Nuzum, W.K. Schubert (1966) Anhidrotic ectodermal dysplasia as autosomal recessive traits in an inbred kindred. Humangenetik 3:181-5.

Pennycuik PR, K. A. Raphael (1984) The tabby locus (Ta) in the mouse: its site of action in tail and body skin. Genet Res 43:51-63.

Plougastel B, P. Couillin, V. Blanquet, E. Le Guern, E. Bakker, C. Turleau, J. De Grouchy, et al (1992) Mapping around the Xq13.1 breakpoints of two X/A

translocations in hypohidrotic ectodermal dysplasia (EDA) female patients. Genomics 14:523-525

Porteous M, A. Curtis, S. Lindsay, O. Williams, D. Goudie, S. Kamarkari, S. Bhattacharya (1992) The gene for Aarskog syndrome is located between DXS255 and DXS566 (Xp11.2-Xq13). Genomics 14:298-301

Powell BC, G. R. Cam, M. J. Fietz, G. E. Rogers (1986) Clustered arrangement of keratin intermediate filament genes. Proc Natl Acad Sci U S A 83:5048-52.

Puck JM, R.L. Nussbaum, D.L. Smead, M.E. Conley (1989) X-linked severe combined immunodeficiency: localization within the region Xq13.1-q21.1 by linkage and deletion analysis. Am J Hum Genet 44:724-730

Puck, J. M. (1994) Molecular and genetic basis of X-linked immunodeficiency disorders. J Clin Immunol 14:81-9.

Puissant C, L. M. Houdebine (1990) An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Biotechniques 8:148-9.

Ramesh N, R. Fulelhan, V. Ramesh, S. Lederman, M.J. Yellin, S. Sharma, L. Chese, et al. (1993) Deletions in the ligand for CD40 in X-linked immunoglobulin deficiency with normal or elevated IgM (HIGMX-1). Internation Immunology 5:769-773

Rhodes JA, D.H. Fitzgibbon, P.A. Macchiarulo, R.A. Murphy (1987) Epidermal growht factor-induced precocious incisor eruption is associated with decteased tooth size. Dev Biol 122:247-52.

Richards B, R. Heilig, I. Oberle, L. Storjohann, G.T. Horn (1991) Rapid PCR analysis of the St14 (DXS52) VNTR. Nucleic Acids Res 19:1944

Robertson N. W., S. R. Blecher (1987) Epidermal growth factor (EGF) affects sulphydryl and disulphide levels in cultured mouse skin: possible

relationship between effects of EGF and of the tabby gene on thiols. Biochem Cell Biol 65:658-67.

Salinas CF, J.M. Opitz, N.W. Paul (1988) Recent Advances in Ectodermal Dysplasias. Alan R.Liss, Inc., New York

Sambrook J, E.F. Fritsch, T. Maniatis (1989) Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Harbor Press, New York

Schlessinger D, J.L. Mandel, A.P. Monaco, D.L. Nelson, H.F. Willard (1993) Report of the fourth international workshop on human X chromosome mapping 1993. Cytogenet Cell Genet 64:147-194

Schmidt M, D. Du-Sart (1992) Functional disomies of the X chromosome influence the cell selection and hence the X inactivation pattern in females with balanced X-autosome translocations: a review of 122 cases. Am J Med Genet 42:161-9.

Schneider LC, C.T. Moody, S. Gellis, R.S. Geha (1992) Ectodermal dysplasia, immunodeficiency and mycobacterium avium intracellularae. J of Allergy and Clinical Immunology 89:271

Settineri WMF, F.M. Salzano, M.J. Melo e Freitas (1976) X-linked anhidrotic ectodermal dysplasia with some unusual features. J Med Genet 13:212-216.

Shah MB, X. Guan, JR. Einstein, S. Matis, Y. Xu, RJ. Mural, EC. Uberbacher (1994) User's Guide to GRAIL and GENQUEST and XGENQUEST Client-Server Systems. Available by anonymous ftp to arthur.epm.ornl. gov

Siegel MB,W. P. Potsic (1990) Ectodermal dysplasia: the otolaryngologic manifestations and management. Int J Pediatr Otorhinolaryngol 19:265-71.

Skare J, B.L. Wu, S. Madan, V. Pulijaal, D. Purtilo, D. Haber, D. Nelson, B. Sylla, H. Grierson, H. Nitowsky, J. Glaser, J. Wissink, B. White, J. Holden, D.

Housman, G. Lenoir, H. Wyandt and A. Milunsky (1993) Characterization of three overlapping deletions causing X-linked lymphoproliferative disease. Genomics 16:254-255.

Sofaer, JA. (1973) Hair follicle initiation in reciprocal recombinations of downless homozygote and heterozygote mouse tail epidermis and dermis. Dev Biol 34:289-96.

Stevenson AC, C.B. Kerr, (1967) On the distributions of frequencies of mutation to genes determining harmful traits in man. Mutat Res 4: 339-52.

Sundberg, JP (1994) Handbook of Mouse Mutations with Skin and Hair Abnormalities. CRC Press Inc. Florida

Sybert, VP (1989) Hypohidrotic ectodermal dysplasia: argument against an autosomal recessive form clinically indistinguishable from Christ-Siemens-Touraine syndrome. Ped Dermatol 6:76-81

Tennyson CN, H.J. Klamut, R.G. Worton (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nature Genet 9:184-185.

Theil, E. C. (1994) Iron regulatory elements (IREs): a family of mRNA non-coding sequences. Biochem J 304:1-11.

Thomas NS, J. Chelly, J. Zonana, K. J. Davies, S. Morgan, J. Gault, K. A. Rack, et al. (1993) Characterisation of molecular DNA rearrangements within the Xq12-q13.1 region, in three patients with X-linked hypohidrotic ectodermal dysplasia (EDA). Hum Mol Genet 2:1679-85.

Tsukada S, D. C. Saffran, D. J. Rawlings, O. Parolini, R. C. Allen, I. Klisak, R. S. Sparkes, et al. (1993) Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. Cell 72:279-90.

Turleau C, P. Niaudet, M. O. Cabanis, G. Plessis, D. Cau, J. de-Grouchy (1989) X-linked hypohidrotic ectodermal dysplasia and t(X;12) in a female. Clin Genet 35:462-6.

Vetrie D, I. Vorechovsky, P. Sideras, J. Holland, A. Davies, F. Flinter, L. Hammarstrom, et al. (1993) The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. Nature 361:226-233

Wang A, L. Schneider, S. Gellis, R. Geha MAI and immune deficiency in ectodermal dysplasia. (submitted)

Wanner R, H. H. Forster, I. Tilmans, D. Mischke (1993) Allelic variations of human keratins K4 and K5 provide polymorphic markers within the type II keratin gene cluster on chromosome 12. J Invest Dermatol 100:735-41.

Weeks N L, S. R. Blecher (1983) Evidence from thiol histochemistry for homology between the Tabby-crinkled syndrome in mice and human ectodermal dysplasia. J Histochem Cytochem 31:1407-11.

Weinmaster G, VJ. Roberts, G. Lemke (1991) A homolog of Drosophila Notch expressed during mammalian development. Development 113:199-205.

Weissenbach J, G. Gyapay, C. Dib, A. Vignal, J. Morissette, P. Millasseau, G. Vaysseix, et al. (1992a) A second-generation linkage map of the human genome. Nature 359:794-802

Weissenbach, J (1992b) The genethon microsatellite map catalogue. Genethon Human Genome Research Centre, Evry France pp116-121

Wu BL, A. Milunsky, D. Nelson, B. Schmeckeper, G. Porta, D. Schlessinger and J. Skare (1993) High-resolution mapping of probes near the X-linked lymphoproliferative disease (XLP) locus. Genomics 17:163-170

Wysolmerski, J. J., A. E. Broadus, J. Zhou, E. Fuchs, L. M. Milstone, W. M. Philbrick (1994) Overexpression of parathyroid hormone-related protein in

the skin of transgenic mice interferes with hair follicle development. Proc Natl Acad Sci U S A 91:1133-7.

Zonana J, A. Clarke, M. Sarfarazi, N.S.T. T homas , K. Roberts, K. Marymee, P.S. Harper (1988) X-linked hypohidrotic ectodermal dysplasia: localization within the region Xq11-21.1 by linkage analysis and implications for carrier detection and prenatal diagnosis. Am J Hum Genet 43:75-85

Zonana J, J. Gault, K.J.P. Davies, M. Jones, D. Browne, M. Litt, N. Brockdorff, et al. (1993) Detection of a molecular deletion at the DXS732 locus in a patient with X-linked hypohidrotic ectodermal dysplasia (EDA); with the identification of a unique junctional fragment. Am J Hum Genet 52:78-84

Zonana J, M. Jones, D. Browne, M. Litt, P. Kramer, H.W. Becker, Brockdorff N, et al. (1992) High resolution mapping of the X-linked hypohidrotic ectodermal dysplasia (EDA) locus. Am J Hum Genet 51:1036-1046

Zonana, J., J. Gault, K. J. Davies, M. Jones, D. Browne, M. Litt, N. Brockdorff, et al. (1993) Detection of a molecular deletion at the DXS732 locus in a patient with X-linked hypohidrotic ectodermal dysplasia (EDA), with the identification of a unique junctional fragment. Am J Hum Genet 52:78-84.

Zonana, J., M. Jones, A. Clarke, J. Gault, B. Muller, N. S. Thomas (1994) Detection of de novo mutations and analysis of their origin in families with X linked hypohidrotic ectodermal dysplasia. J Med Genet 31:287-92.

Zonana, J., M. Sarfarazi, N. S. Thomas, A. Clarke, K. Marymee, P. S. Harper (1989) Improved definition of carrier status in X-linked hypohidrotic ectodermal dysplasia by use of restriction fragment length polymorphism-based linkage analysis. J Pediatr 114:392-9.

Zonana, J., S. H. Roberts, N. S. Thomas, P. S. Harper (1988) Recognition and reanalysis of a cell line from a manifesting female with X linked

hypohidrotic ectodermal dysplasia and an X; autosome balanced translocation. J Med Genet 25:383-6.