GENETIC STUDIES OF HYPOSPADIAS

Louise Frisén

Stockholm 2002
Supervisors
Agneta Nordenskjöld, Associate Professor
Department of Molecular Medicine
Karolinska Institutet, Stockholm, Sweden

Holger Luthman, Professor
Department of Endocrinology
Lund University, University Hospital MAS, Malmö, Sweden

Opponent
Juha Kere, Professor
Department of Biosciences at Novum
Karolinska Institutet
Huddinge University Hospital, Huddinge, Sweden

Thesis committee
Göran Annerén, Professor
Department of Clinical Genetics
Uppsala University Hospital, Uppsala, Sweden

Jan Hillert, Professor
Department of Neuroscience
Karolinska Institutet
Huddinge University Hospital, Huddinge, Sweden

Olle Söder, Professor
Department of Woman and Child Health
Karolinska Institutet, Stockholm, Sweden

All previously published papers were reproduced with permission from the publisher.

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

Genetic studies of hypospadias
© Louise Frisén, 2002
ISBN 91-7349-397-X
In most cases the cause of hypospadias is unknown. In our case the family believes that the condition originated by imprinting, when the mother of patient IV-12, while being pregnant, stared at a camel (camels apparently urinate backwards).

_Frydman et al, American Journal of Medical Genetics, 1995_
CONTENTS

ABSTRACT .................................................. I
LIST OF PUBLICATIONS .............................. II
ABBREVIATIONS ........................................ III
INTRODUCTION ........................................... 1
    CHARACTERISTICS OF HYPOSPADIAS .......... 2
    MALE SEX DIFFERENTIATION ................. 4
    EPIDEMIOLOGY IN HYPOSPADIAS ............. 5
    GENETICS AND HYPOSPADIAS .................. 8
    COMPLEX TRAITS .................................. 11
    GENE MAPPING ..................................... 13
        Linkage analysis ............................. 13
        Parametric versus non-parametric linkage
          analysis .................................... 14
        Affective relative pair method ........... 14
        Power estimations .......................... 16
        Significance levels ......................... 17
        Association mapping ....................... 18
        Success so far and alternative strategies .. 19
AIMS OF THE STUDY ..................................... 21
MATERIALS ............................................ 22
ETHICAL PERMISSIONS .............................. 26
METHODS ............................................. 27
RESULTS AND DISCUSSION ....................... 32
GENERAL DISCUSSION ............................. 44
SUMMARY IN SWEDISH ............................. 46
ACKNOWLEDGEMENTS ............................... 48
REFERENCES .......................................... 51
PAPER I-V .............................................
ABSTRACT

Hypospadias is defined as an abnormal opening of the urethra on the underside of the penis. It is a frequently found malformation with an incidence of 3 per 1000 males. The aim of this thesis was to identify genetic and environmental factors in the pathogenesis of hypospadias. For this purpose, a variety of genetic methods were used in a nation-wide material corresponding to half of all registered cases of hypospadias in Sweden.

We identified 18 monozygotic twins discordant for hypospadias. In 16 of these, the twin with lowest birth weight was affected with hypospadias. This shows that low birth weight is important for hypospadias, regardless of the genetic constitution (paper I).

We investigated the familial rate and analyzed the association with low birth weight in 2138 families with at least one boy with hypospadias. There was a familial rate of 7% and a significant lower birth weight in cases with hypospadias compared with their respective brothers, used as controls (p=5x10⁻¹³). An increased frequency of dizygotic as well as monozygotic male-male twins was found, with a skewed distribution towards monozygotic twins. This paper (II) also includes a description of the ethnic background in the material and the distribution of phenotypes.

A complex segregation analysis was performed to define the mode of inheritance in a material consisting of 2005 pedigrees. We found a best fit for the multifactorial model and a heritability of 0.99. This is interpreted as monogenic effects acting in some of the families but a multifactorial cause in the majority (paper III).

A genome-wide linkage analysis based on a non-parametric affected relative pair method was used in 69 families. All available family members were genotyped with 360 polymorphic PCR based microsatellite markers with a mean interval of 9.5 cM. Five chromosomal regions reaching the level of suggestive significance were identified (paper IV). These need to be investigated further to identify hypospadias susceptibility genes.

Linkage analysis and subsequent mutation analysis in a family with autosomal dominant inheritance of hypospadias revealed a novel mutation in the HOXA13 gene (paper V). This suggests the diagnosis of hand-foot-genital syndrome although the phenotype in this family is atypical compared with previously reported families.

In this thesis, several lines of evidence suggesting genetic factors in the pathogenesis of hypospadias are presented, including the identification of five chromosomal regions in which genes for hypospadias are likely to be located and a novel mutation in the HOXA13 gene. It is also shown that low birth weight is an important risk factor for hypospadias.

Key words: hypospadias, genetic, complex trait, low birth weight, twin study, segregation analysis, genome-wide linkage analysis, mutation analysis, HOXA13 gene
LIST OF PUBLICATIONS

This thesis is based on the following papers, which will be referred to by their Roman numerals.

I Louise Fredell, Paul Lichtenstein, Nancy L. Pedersen, Jan Svensson and Agneta Nordenskjöld
Hypospadias is related to birth weight in discordant monozygotic twins

II Louise Fredell, Ingrid Kockum, Einar Hansson, Staffan Holmner, Lars Lundquist, Göran Läckgren, Jörgen Pedersen, Arne Stenberg, Gunnar Westbacke and Agneta Nordenskjöld
Heredity of hypospadias and the significance of low birth weight
The Journal of Urology 167, 1423-1427, 2002

III Louise Fredell, Lennart Iselius, Andy Collins, Einar Hansson, Staffan Holmner, Lars Lundquist, Göran Läckgren, Jörgen Pedersen, Arne Stenberg, Gunnar Westbacke and Agneta Nordenskjöld
Complex segregation analysis of hypospadias
Human Genetics 111, 231-234, 2002

IV Louise Frisén, Cilla Söderhäll, Margareta Tapper-Persson, Holger Luthman, Ingrid Kockum and Agneta Nordenskjöld
Genome-wide linkage analysis for hypospadias susceptibility genes
In manuscript

V Louise Frisén, Kristina Lagerstedt, Margareta Tapper-Persson, Ingrid Kockum and Agneta Nordenskjöld
A novel duplication in the HOXA13 gene in a family with atypical hand-foot-genital syndrome
Submitted
ABBREVIATIONS

AMH  anti Müllerian hormone
AR   the androgen receptor gene
bp   base pairs
cM   centiMorgan
ddNTP dideoxynucleotide
DNA  deoxyribonucleic acid
FSH  follicle stimulating hormone
hCG  human chorionic gonadotropin
HFGS hand-foot-genital syndrome
HOX  homeobox gene
IBD  identical by descent
IBS  identical by state
kb   kilobase pairs
LH   luteinizing hormone
LOD  logarithm of the odds
Mb   megabase pairs
PCR  polymerase chain reaction
RFLP restriction fragment length polymorphisms
SNP  single nucleotide polymorphism
SRD5A2 the 5-alpha-reductase gene
SRY  the sex-determining region Y gene


INTRODUCTION

The term hypospadias is derived from the Greek words hypo (under) and spad (something torn, from span to tear, pluck off) meaning that the urethral orifice is located on the underside of the penis (Merriam-Webster 2002).

Hypospadias was first described in the second century A.D. by Galen (c. 130-201 A.D.). At this time, as well as during the following 1000 years, the treatment of choice was amputation beyond the urethral orifice (Adams 1844-1847). The importance of chordee (curvature of the penis) was also recognized by Galen as “men inflicted with hypospadias find it impossible to beget children, the meatus being turned away from the extremity of the penis by the frenum, not because they lack fertile sperm, but because the curvature of the penis prevents its normal overflow from being conveyed forwards. This theory is confirmed by the ability to beget children if the frenum is divided” (Galen c. 130-201 A.D.). In the 16th century this was illustrated in the successful correction of the chordee of King Henry II of France, after which he fathered ten children in his marriage to Catherine de Medici (Smith 1997).

The husband to a Maltese woman, Mathia, did not manage as well during this era. The marriage was annulled by the Roman Catholic Church on request from Mathia, who described her husband’s hypospadias as “a defect in the configuration of his virile member on account of which he did not urinate in a natural way like other men”. The medical examination took place in court and resulted in the following report: “John’s male member was inept or incapable and also useless for deflorating or perforating because it was short and curved, this curvation tending to become more pronounced with rigidity of the penis” (Cassar 1974).
CHARACTERISTICS OF HYPOSPADIAS

“Eddings had no injuries except several old scars, mostly on his knee. But biology had dealt him an earlier blow called hypospadias, which meant his urethra opened on the underside of his penis instead of in the center. This moderate defect would have caused him a great deal of anxiety, especially as a boy. As a man he may have suffered sufficient shame that he was reluctant to have sex.”

*Patricia Cornwell, Cause of Death, 1996*

The main features of hypospadias are the abnormal opening of the urethral orifice and different degrees of curvature of the penis (chordee) (figure 1).

**Figure 1**

Hypospadias is classified into glandular, penile, penoscrotal, scrotal or perineal according to the localization of the urethral orifice. A cleaved prepuce is considered a mild variant of hypospadias. Cryptorchidism, bifid scrotum and micropenis are occasionally associated with the condition, increasing in frequency with the severity of hypospadias. In the most severe cases, there may be ambiguity regarding the sex of the newborn child, warranting further investigations in order to determine this.

Hypospadias is considered to be restricted to males but females with hypospadias-like manifestations have been reported (Knight et al. 1995; Ronzoni et al. 2001).
During normal male sex differentiation, the urethral folds close in fetal weeks 8-16 (figure 2). Hypospadias results from a failure in this fusion process. The severity of hypospadias is a continuum, and depends on when during the embryonic period the fusion fails.

**Figure 2**
The urethral folds gradually fuse during fetal weeks 8-16.

The development of the penis and the urethra is an androgen dependent process, relying on proper timing as well as quantity of the required hormones. This suggests that hypospadias may be due to impairment in the androgen pathway. During the first trimester, fetal androgen production is induced by maternal gonadotropins, preferably human chorionic gonadotropin (hCG) produced in the placenta, whereas later in gestation, this is provided by gonadotropins (i.e. LH, FSH) produced in the fetal pituitary.

Surgical correction of hypospadias is recommended before the age of school start. Preoperative treatment with hCG or testosterone result in decreased chordee and increased penile length, thus facilitating the surgical treatment (Davits et al. 1993; Koff and Jayanthi 1999). Severe hypospadias can be detected with prenatal ultrasonography (Sides et al. 1996; Meizner et al. 2002). However, in most cases, there is no indication for routinely carrying out prenatal diagnosis for hypospadias, since it is a mild malformation in which surgical correction is almost always successful.

As a group, boys with hypospadias undergo normal psychological and sexual development and exhibit no divergent gender identity (Sandberg et al. 1995; Sandberg et al. 2001). The sexual debut may be delayed but is in that case related to the individuals’ genital perception (Bracka 1999). Depending on the severity of hypospadias and on the underlying cause, the fertility is sometimes reduced. In the majority of cases with mild variants and satisfactory surgical correction, there is no evidence for reduced fertility.
MALE SEX DIFFERENTIATION

At fertilization, the fusion of an egg carrying 22 autosomal chromosomes and an X chromosome with a sperm with 22 autosomal chromosomes and either an X or a Y chromosome determines the sex (chromosomal sex). An XX individual develops into a female and an XY individual into a male (phenotypic sex). Until fetal week 6, the embryo is sexually undifferentiated. Two bipotential gonads will ultimately differentiate into testes or ovaries (gonadal sex). At this point, there are also two undifferentiated bilateral duct systems, the Müllerian and the Wolffian ducts, generating internal female or male genital organ, respectively.

While female sex differentiation is the default pathway, male sex differentiation relies on a series of crucial events. First, the \textit{SRY} gene on the Y chromosome is solely responsible for initiating the male differentiation of the XY embryo in fetal week 6. The \textit{SRY} gene encodes a transcription factor, which induces the differentiation of the indifferent gonads into testes. Shortly after the onset of \textit{SRY} expression, cells in the developing testis differentiate into Sertoli cells, producing anti Müllerian hormone (AMH) and Leydig cells, producing testosterone. AMH induces the regression of the Müllerian ducts whereas testosterone induces the differentiation of Wolffian ducts into the epididymis, vas deferens and seminal vesicles. In target tissues, testosterone is converted intracellularly into dihydrotestosterone through the 5-alpha-reductase enzyme. Dihydrotestosterone modulates differentiation of the prostate gland, penis and scrotum. Testosterone as well as dihydrotestosterone binds to its intracellular receptor (the androgen receptor). Ultimately this hormone-receptor complex binds to DNA to regulate transcription.

In fetal week 8, male external genitalia start forming. As the phallus elongates it pulls the urethral folds forward so that they form the lateral walls of the urethral groove. The epithelial lining of the groove is of endodermal origin and forms the urethral plate. In fetal week 12, the two urethral folds close over the urethral plate, thus forming the penile urethra. Later, the most distal part of the urethra is formed when ectodermal cells from the tip of glans penetrate inward, forming a short epithelial cord. The formation of external genitalia, including closure of the prepuce, is completed in fetal week 16 (Sadler 1990).
Hypospadias is one of the most prevalent malformations in man and by far the most common urogenital malformation. There are large geographical differences in reported hypospadias rates, ranging from 0.3 per 1000 male births (Japan) to 7 per 1000 (The Netherlands) (Paulozzi 1999; Virtanen et al. 2001; Pierik et al. 2002). However, it is difficult to make comparisons between countries due to various inclusion criteria and incomplete ascertainment. There are about 5000 cases with hypospadias in the Swedish Malformation Registry. The incidence in Sweden has remained at a stable rate of 3 per 1000 male births since the beginning of the 1970s, when the registered incidence increased from 0.8 per 1000 infants in 1969 to 1.5 per 1000 infants in 1973 (Källén and Winberg 1982). The reason for this has not been explained. An overview of epidemiological studies of hypospadias is shown in table 1.

In some of these, an association for hypospadias with low birth weight is found (Chen and Woolley 1971; Monteleone Neto et al. 1981; Calzolari et al. 1986). Low weight at birth reflects a growth retardation throughout the gestation and has been shown to be correlated to sub-optimal first-trimester growth (Smith et al. 1998). Up to a 10-fold increase of hypospadias was found in infants small for gestational age (Akre et al. 1999; Weidner et al. 1999; Gatti et al. 2001; Hussain et al. 2002). The poor intrauterine growth was shown to be of early gestational cause (Hussain et al. 2002). Interestingly, the degree of growth retardation is not correlated to the severity of hypospadias, indicating that this is determined by genetic factors (Calzolari et al. 1986; Gatti et al. 2001; Hussain et al. 2002).

It is unclear whether the growth retardation in itself renders the fetus more susceptible for other predisposing factors (genetic or environmental), or if there is a common denominator in the genesis of the two conditions. A placentary malfunction is suggested by the association with a low weight of the placenta (Stoll et al. 1990) and severe preeclampsia (Akre et al. 1999). Abnormalities of the fetal-placental-maternal interaction may also explain the finding that women giving birth to boys with hypospadias had a higher rate of weak contractions during birth, induced deliveries and caesarean sections (Källén 1988).
Table 1 Epidemiological studies of hypospadias

<table>
<thead>
<tr>
<th>Author</th>
<th>Incidence (male births)</th>
<th>Data source</th>
<th>Numbers included</th>
<th>Familial rate</th>
<th>% of sibs affected</th>
<th>Recurrence risk for sib</th>
<th>Heritability</th>
<th>Twin rate</th>
<th>Birth weight</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sørensen 1953</td>
<td>3.3/1000</td>
<td>Hospital</td>
<td>103</td>
<td>28%**</td>
<td>9.6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75% glandular</td>
</tr>
<tr>
<td>Chen&amp;Wolley 1971</td>
<td>1.98/1000</td>
<td>Registry</td>
<td>93</td>
<td>6.5%</td>
<td>9.7%</td>
<td>10.4%</td>
<td></td>
<td></td>
<td></td>
<td>40% glandular</td>
</tr>
<tr>
<td>Roberts&amp;Lloyd 1973</td>
<td>1.98/1000</td>
<td>Registry</td>
<td>93</td>
<td>6.5%</td>
<td>9.7%</td>
<td>10.4%</td>
<td></td>
<td></td>
<td></td>
<td>46% penile</td>
</tr>
<tr>
<td>Sweet 1974</td>
<td>8.2/1000</td>
<td>Hospital</td>
<td>113</td>
<td>8%</td>
<td>8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1% severe</td>
</tr>
<tr>
<td>Czeizel 1979</td>
<td>4.4/1000</td>
<td>Hospital</td>
<td>294</td>
<td>4%</td>
<td>9.1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>75% glandular</td>
</tr>
<tr>
<td>Bauer 1981</td>
<td>Hospital</td>
<td>307</td>
<td>21%**</td>
<td>14%</td>
<td>12%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>32% glandular</td>
</tr>
<tr>
<td>Montelone 1981</td>
<td>0.76/1000*</td>
<td>Registry</td>
<td>324</td>
<td>6.1%</td>
<td>0.68</td>
<td>3.4% (male-male dominance)</td>
<td>low***</td>
<td>72% glandular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calzolari 1986</td>
<td>Registry</td>
<td>168</td>
<td>9.1%</td>
<td>0.67</td>
<td>5.4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21.4% penile</td>
</tr>
<tr>
<td>Källén 1986</td>
<td>1.53/1000*</td>
<td>Registry</td>
<td>262</td>
<td>4.2%</td>
<td>4.2%</td>
<td>(male-male dominance)</td>
<td>low</td>
<td>75% glandular</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stoll 1990</td>
<td>2.89/1000</td>
<td>Registry</td>
<td>176</td>
<td>17%</td>
<td>0.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>69% glandular</td>
</tr>
</tbody>
</table>

* per live births  ** the more severe the malformation, the higher the recurrence risk in relatives  *** p<0.001
An increased risk for hypospadias following *in vitro* fertilization and intracytoplasmic sperm injection has been found (Macnab and Zouves 1991; Silver et al. 1999b; Wennergren et al. 2000; Ericson and Källén 2001). This may be explained by the administration of high doses of gestagens interfering with androgen production in the male fetus (Silver et al. 1999a). Furthermore, children born after *in vitro* fertilization have lower birth weight (Bergh et al. 1999), an established risk factor for hypospadias. Older mothers run a higher risk of giving birth to a boy with hypospadias but it is unclear if this is a direct consequence of the mothers age or of other factors, such as a higher frequency of *in vitro* fertilization in older women (Fisch et al. 2001). The increased frequency of hypospadias after intracytoplasmic sperm injection may be related to paternal subfertility (Wennergren et al. 2000).

Hypospadias is not associated with maternal use of oral contraceptives in early pregnancy (Källén et al. 1991). An increased frequency of hypospadias in boys born to 2780 Swedish women with intake of loratadine (a non-prescription antihistamine) during pregnancy was recently reported in the daily press. However, the number of cases was small (15 observed boys with hypospadias to be compared with the expected 5-6) and a biological mechanism for this remains to be explained (Hellbom 2002).

The increasing incidences of hypospadias reported in some countries have raised speculations on the involvement of environmental endocrine disrupters, such as estrogenic and anti-androgenic chemicals, in the pathogenesis of hypospadias (Toppari et al. 1996; Paulozzi et al. 1997). Possibly relating to this is the finding that mothers with a vegetarian diet during pregnancy have a five-fold increased risk of giving birth to a boy with hypospadias (North and Golding 2000). This has been suggested to result from increased levels of phytoestrogens acting anti-androgenic in the developing male fetus. However, studies of boys born to women in gardening occupations, exposed to pesticides with presumed estrogenic or anti-androgenic properties, showed no increased frequency of hypospadias (Weidner et al. 1998).
Several observations in hypospadias suggest that genetic factors are involved in the pathogenesis. Familial clustering has been reported in 4% to 28% of cases (table 1). The more severe the malformation of the index patient, the higher is the recurrence risk for the next male sibling, ranging between 4% and 17% (table 1).

In some ethnic groups in which consanguinity is a common feature, hypospadias has a particularly high incidence (Frydman et al. 1985; Tsur et al. 1987). In these families recessive inheritance can be suspected. An autosomal dominant mode of inheritance has also been observed (Lowry and Kliman 1976; Page 1979).

Furthermore, hypospadias is a feature in more than one hundred genetically caused syndromes (McKusick 2002). In some of these, mutations have been identified in genes involved in sex differentiation, e.g. the X-linked partial androgen insensitivity syndrome caused by mutations in the androgen receptor gene (AR, Xq11-12) and the recessive 5-alpha-reductase deficiency due to mutations in the 5-alpha-reductase gene (SRD5A2, 2p23) (Imperato-McGinley et al. 1974; Wilson et al. 1993; Quigley et al. 1995). However, these syndromes are characterized by severe hypospadias in association with other genital malformations such as cryptorchidism, bifid scrotum and penoscrotal transposition. Infrequently, mutations in the androgen receptor gene (Hiort et al. 1994; Allera et al. 1995) and the 5-alpha-reductase gene (Silver and Russell 1999) have been identified in isolated hypospadias.

Although there is without doubt a familial aggregation in hypospadias that in some cases are caused by a mutation in a single gene, susceptibility genes in common forms of the malformation have not yet been identified. Before undertaking an investigation for disease-causing genes, it is important to estimate the contribution of genetic effects.

An overview of various strategies in identifying the genetic component of a disease is presented in figure 3.
Figure 3
Strategies in the identification of disease-causing genes

- Definition of phenotype
- Identification of genetic component
  - Relative risk
  - Segregation analysis
  - Pedigree analysis
  - Twin study
- Identification of mode of inheritance
  - Complex
    - Non-parametric linkage analysis in many families, few affected/family
  - Mutations analysis in candidate gene
  - Association studies
  - Monogenic
    - Parametric linkage analysis in large families, many affected
The relative risk ($\lambda$) is used to estimate the contribution of genetic factors by dividing the recurrence risk in family members ($\lambda_r$), or in siblings ($\lambda_s$), with the risk in the general population (Risch 1990a). Because $\lambda$ is a ratio, the prevalence in relatives as well as in the general population affects its size. Thus, a strong familial aggregation in a common disease results in a smaller $\lambda$ than the same degree of aggregation in a rare disease. In general, $\lambda>2$ indicate a significant genetic component and $\lambda>40$ suggests a strong genetic component (Kruglyak and Lander 1995b). With an incidence of 0.3% and a familial recurrence risk of about 10%, the $\lambda_r$ for hypospadias is 33.

Segregation analysis is used to evaluate whether a major gene contributes to the phenotype, and if that is the case, to determine the mode of inheritance. Ideally, the segregation analysis results in correct estimation of the genetic model, the penetrance of the disease allele and its frequency. These parameters can then be used in a linkage analysis. A successful example of this is the segregation analysis in breast cancer (Iselius et al. 1991), eventually resulting in the identification of the BRCA1 and BRCA2 genes (Miki et al. 1994; Wooster et al. 1995). If no major genes are involved and the data instead suggest a polygenic model, the segregation analysis is useful to quantify the degree of genetic contribution, i.e. heritability (Khoury et al. 1993).

Heritability is defined as the proportion of the phenotypic variance due to the additive effects of many genes (the polygenic component) (Haines and Pericak-Vance 1998). Heritability for hypospadias has been estimated to be between 0.57 and 0.74 using simple multifactorial threshold models (Chen and Woolley 1971; Czeizel et al. 1979; Monteleone Neto et al. 1981; Calzolari et al. 1986; Stoll et al. 1990). In a complex segregation analysis of 103 families, the heritability was 0.99 (Harris and Beaty 1993).

One per 80 pregnancies results in twins. One third of twins are monozygotic and two thirds are dizygotic. Studies in twins are valuable in order to disentangle genetic and environmental factors. This is based on the fact that monozygotic twins are genetically identical whereas dizygotic twins, like siblings, share on average half of their genetic material. Higher concordance rates (i.e. both twins affected) in monozygotic twins than in dizygotic twins suggest that genetic factors are important in the pathogenesis of a disease.

Several studies have reported an increased frequency of hypospadias in twins, with a skewed distribution towards monozygotic twins (Sørensen 1953; Roberts and Lloyd 1973; Czeizel et al. 1979). The increase in frequency is related to the sex distribution in the twins, with a higher incidence of hypospadias in male-male twins (regardless of

Twins that are monozygotic and discordant constitute a powerful model for estimating the role of environmental factors. Monozygotic twins discordant for hypospadias were first described by Lamy (Lamy 1952). Sørensen identified five twin pairs discordant for hypospadias, in which the twin with lowest birth weight was affected. By outward resemblance and blood groups, zygosity could be established in four of these. Three of the twin couples were monozygotic and one was dizygotic (Sørensen 1953).

Although several lines of evidence suggest that genetic factors are involved in the pathogenesis of hypospadias, environmental factors are also important. Hypospadias is therefore considered a complex trait caused by the combined influence of genes and environment.

COMPLEX TRAITS

Monogenic diseases are consequences of mutations in a single gene, transmitted according to Mendelian models of inheritance (e.g. autosomal dominant, autosomal recessive or X-linked). In contrast, complex traits do not exhibit Mendelian inheritance attributable to a single gene locus but is rather believed to result from interactions of several genes (i.e. susceptibility genes). The more genes involved, the smaller the contribution of each of them, making them difficult to detect (Konig et al. 2001). Also, individuals with the same phenotype may have different genetic background (genetic heterogeneity) and susceptibility genes are believed to interact at a genetic level (epistasis) as well as with environmental factors. In some individuals, the phenotype results from environmental factors only (phenocopies), whereas others are unaffected despite carrying the susceptibility allele(s) (reduced penetrance).

Thus, many different genes acting together and in various combinations, with or without environmental factors, comprise the liability, or predisposition, to a complex trait. Although the phenotype is qualitative (i.e. affected or unaffected), liability to the disease is measured on a quantitative scale (figure 4). The proportion of affected relatives will be highest among severely affected persons, since their liability is further beyond the threshold than that for mildly affected persons. In line with this, the risk is higher for closely related family members and increases with the number of cases in the family.
Figure 4
A multifactorial threshold model describing the situation in which a genetically predisposed individual is affected when exceeding a threshold of genetic and/or environmental factors (Falconer 1965). The lower right-shifted curve illustrates an increased liability compared to the top curve.
GENE MAPPING

In about 1500 monogenic diseases, the mutated gene has been identified (Peltonen and McKusick 2001). For complex traits, the situation is completely different and few susceptibility genes have been found so far. Figure 3 illustrates the various strategies that can be used to identify the gene(s) involved in a disease. The candidate gene approach can be used directly in monogenic as well as complex traits. However, in some cases the prior identification of chromosomal regions is required. For this purpose, linkage analysis is used.

**Linkage analysis**
The term linkage refers to the tendency for two loci on the same chromosome to be inherited together. During meiosis the two homologous chromosomes exchange genetic material, a process known as recombination. This occurs at least once per chromosome arm in each pair of homologous chromosomes. Loci close to each other are separated by recombination less often than loci far apart. A probability of 1% for recombination between two loci corresponds to a genetic distance of 1 cM. This is roughly equal to a physical distance of 1 Mb. The probability of recombination to occur is called the recombination fraction (\( \theta \)). Recombination enables the construction of genetic maps with the use of genetic markers. Genetic markers can be any part of the DNA reflecting normal sequence variations. Genetic maps based on direct measurements of DNA sequence variation was first constructed using restriction fragment length polymorphisms (RFLP) (Botstein et al. 1980). The basis for the polymorphism in a RFLP marker is a single base pair change that introduces or abolishes a cleavage site for restriction enzymes. These variabilities are now known as single nucleotide polymorphisms (SNP) scattered throughout the genome (1 per 1250 base pair) and potentially useful for genome-wide association mapping (Lander et al. 2001; Venter et al. 2001).

A breakthrough for the mapping of genetic diseases was achieved in the 1990s by the accessibility of microsatellite markers (Sheffield et al. 1995). These repetitive segments of DNA (e.g. di, tri or tetrancleotiderepeats) have a high variability between individuals (i.e. heterozygosity), which makes them useful in linkage analysis. The microsatellite markers are spread all over the genome, approximately one every 2 kb (Lander et al. 2001) and are flanked by unique sequences that enable amplification with Polymerase Chain Reaction (PCR). This permits a high-throughput methodology, which is essential since many microsatellite markers are used in a genome-wide linkage analysis. A minimum of 300 microsatellite markers giving a mean distance of 10 cM is required to cover the whole genome, but more markers generally increase the power.
In linkage analysis, genetic markers are used to trace the gene(s) involved in a disease. A genetic marker close to the disease locus will be inherited together with the disease more often than expected by random segregation (i.e. it is linked). The probability of linkage is given as the logarithm of the odds (LOD). This is derived from the log10 of the ratio between the two likelihoods under the null and the alternative hypothesis. The null hypothesis is that the two loci are not linked and the alternative hypothesis that the two loci are linked (Morton 1955). A LOD score of 3 means that the odds that the loci are linked are 1000 times greater than the odds that they are not linked. This corresponds to a significance level of 5%, taking into consideration the inherent probability of linkage (≈1/50) (Ott 1991).

Linkage analysis can be performed in a single-point or a multipoint fashion. In single-point analysis, LOD score calculations are performed for each marker in relation to the disease locus, independent of the surrounding markers. Multipoint analysis refers to the simultaneous analysis of several markers with known location on a genetic map, which increases information in the region.

**Parametric versus non-parametric linkage analysis**

Linkage analysis is used to identify loci in monogenic as well as in complex traits. In monogenic diseases, parametric linkage analysis is the method of choice. This involves the assumption of certain parameters such as the mode of inheritance, the penetrance of the disease-causing allele and its frequency in the population. The chance of success is dependent on the recruitment of families large enough to give significant evidence for linkage. However, if that is impossible, LOD scores can be added between different families with the same phenotype.

Although the power of parametric linkage analysis greatly exceeds that of non-parametric linkage analysis, the latter method is preferred in the mapping of susceptibility genes in common, genetically complex, traits. This is due to the fact that no assumptions about the essential parameters can be reliably made. Instead, a non-parametric linkage analysis including many affected relative pairs is preferred (Lander and Schork 1994).

**Affective relative pair method**

The simplest form of an affected relative pair method is the affected sib pair analysis first proposed by Penrose in 1935 (Penrose 1935). This is based on the assumption that two affected siblings share the chromosomal segment carrying the disease locus more often than expected by chance alone. Since they will show excess allele sharing even in the presence of incomplete penetrance, phenocopies, genetic heterogeneity and high-
frequency disease alleles, this type of linkage analysis is more robust than classical parametric linkage analysis (Lander and Schork 1994). It is, however, less powerful than a correctly specified linkage model.

With the use of genetic markers spread all over the genome, shared regions can be identified (Kruglyak and Lander 1995a). The method relies on calculations based on deviations from allele sharing expected by Mendelian laws of inheritance (figure 5). The null hypothesis is that the sharing is consistent with the expected distribution by Mendelian laws of inheritance and the alternative hypothesis that there is excess sharing of alleles identical by descent (IBD). The affected relative pair LOD score is calculated by the log$_{10}$ of the ratio between the two likelihoods under the null and the alternative hypothesis. It should be noted that whereas in parametric linkage analysis the alternative hypothesis is that two loci are linked, in the affected relative pair method, the alternative hypothesis is excess sharing of alleles IBD.

**Figure 5**
According to Mendelian laws of inheritance, the probabilities of a sib pair sharing one marker allele identical by descent (IBD) is 50%, two alleles IBD is 25% and no allele IBD is 25%.

![Diagram showing Mendelian probabilities](image)

Although siblings appear to have inherited identical alleles for a marker, they might just be identical by state (IBS). Parents as well as unaffected siblings can therefore be included in the genotyping to enable the distinction between alleles that are IBD and IBS (Haines and Pericak-Vance 1998). The proportion of genes shared IBD is shown in table 2.
Table 2 shows the degree of relationship and corresponding proportion of IBD sharing. It should be emphasized that although sibling pairs and parent-child pairs each share 50% of their genetic material, the type of sharing is different. Sibling pairs share on average 50% of their genetic material, of which some is derived from the mother and some from the father, whereas a parent transmits exactly 50% of his or her genetic material to each offspring. For this reason, parent-child pairs are not informative in an affected relative pair method. Parents are included only to enable the distinction between alleles that are IBD and IBS.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Degree of relationship</th>
<th>Proportion of IBD shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monozygotic twins</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dizygotic twins</td>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>Sibs</td>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>Parent-child</td>
<td>1</td>
<td>1/2</td>
</tr>
<tr>
<td>Grandparent-grandchild</td>
<td>2</td>
<td>1/4</td>
</tr>
<tr>
<td>Avuncular</td>
<td>2</td>
<td>1/4</td>
</tr>
<tr>
<td>Half-sibs</td>
<td>2</td>
<td>1/4</td>
</tr>
<tr>
<td>Cousins</td>
<td>3</td>
<td>1/8</td>
</tr>
<tr>
<td>Great grandparent-great grandchild</td>
<td>3</td>
<td>1/8</td>
</tr>
<tr>
<td>Great avuncular</td>
<td>3</td>
<td>1/8</td>
</tr>
</tbody>
</table>

**Power estimations**

Power is the probability of correctly concluding that linkage exists. Since the underlying genetic model for a complex trait is unknown, as well as the number of susceptibility loci and their λ, it is difficult to accurately predict the power of a non-parametric linkage analysis. It has been estimated that, for a locus with λ 3.0, a 10 cM interval flanked with markers with heterozygosity value of 0.75 and a sample size of 100 affected sibling pairs can generate a LOD score of 3.0 with a power of 0.45. As the number of affected sibling pairs or the λ increases, so does the power. In this example, increasing the sample size to 200 affected sibling pairs results in an increase of power to 0.90 (Hauser et al. 1996). As a rule, any sample of fewer than 40 sib pairs is unlikely to detect even relatively strong genetic effects (Haines and Pericak-Vance 1998).
Significance levels

The LOD scores achieved from a linkage analysis including many markers must be corrected for multiple testing. This results in different significance levels that must be kept apart. Point-wise significance level is the probability of randomly exceeding the observed value at one specific locus, whereas genome-wide significance level is the probability of randomly exceeding the observed value anywhere in the genome (Lander and Kruglyak 1995). Lander and Kruglyak suggested the following definitions of genome-wide significance levels: Suggestive linkage is defined as statistical evidence expected to randomly occur once per genome-wide linkage analysis. Significant linkage is defined as statistical evidence expected to randomly occur 0.05 times in a genome-wide linkage analysis. Highly significant linkage is defined as statistical evidence expected to randomly occur 0.001 times in a genome-wide linkage analysis.

Many different computer programs, based on different statistical algorithms, are available for non-parametric linkage analysis but LOD scores are not easily comparable between those. It has therefore been suggested that, besides the LOD scores, the appropriate significance thresholds and corresponding p-value should be reported (Nyholt 2000). These recommendations are summarized in table 3.

Table 3
Significance levels and corresponding p-values as suggested by Lander and Kruglyak (1995). LOD scores according to Nyholt (2000) are shown for Allegro (Gudbjartsson et al. 2000) and MAPMAKER/SIBS (MMS, for the X chromosome) (Kruglyak and Lander 1995a) since these programs are used in this thesis. The corresponding p-values denotes allele-sharing methods in sibs or half-sibs. For studies involving a mixture of relative pairs, the thresholds are roughly in the range of 10^{-4}-5\times10^{-4} for suggestive linkage and 5\times10^{-5}-10^{-5} for significant linkage (Lander and Kruglyak 1995).

<table>
<thead>
<tr>
<th>Significance level</th>
<th>Corresponding p-value (sibs or half-sibs)</th>
<th>LOD scores (Allegro)</th>
<th>LOD scores (MMS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>p&lt;0.05</td>
<td>0.59</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>p&lt;0.01</td>
<td>1.18</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>p&lt;0.005</td>
<td>1.44</td>
<td>2.21</td>
<td></td>
</tr>
<tr>
<td>p&lt;0.001</td>
<td>2.07</td>
<td>2.93</td>
<td></td>
</tr>
<tr>
<td>Suggestive linkage</td>
<td>p&lt;7.4\times10^{-4}</td>
<td>2.19</td>
<td>3.06</td>
</tr>
<tr>
<td>Significant linkage</td>
<td>p&lt;2.2\times10^{-5}</td>
<td>3.63</td>
<td>4.62</td>
</tr>
<tr>
<td>Highly significant linkage</td>
<td>p&lt;3\times10^{-7}</td>
<td>5.30</td>
<td>6.52</td>
</tr>
</tbody>
</table>
**Association mapping**

The regions that can be identified in a genome-wide linkage analysis of a complex trait are relatively large, often spanning more than 20 cM. A chromosomal region of that size harbors about 200-300 genes. In most cases, the identification of a disease-causing allele will require that the region is restricted using more markers and association mapping (Todd 2001).

Whereas linkage analysis relies on the identification of markers shared between affected individuals due to their proximity to the disease-causing gene, association implies that the specific allele is identical in apparently unrelated affected individuals. The associated allele may be related to the pathogenesis in itself, or it may be in linkage disequilibrium with the disease-causing allele. Linkage disequilibrium refers to the fact that certain alleles in adjacent loci are co-inherited more often than expected by random segregation. This reflects a preserved chromosomal segment through generations and is exploited in association mapping. Association can be found using case-control studies, in which allele frequencies are compared between affected individuals and matched controls, or within families (tests of transmission distortion). In the latter case, problems with selection bias are overcome. Association mapping has been essential in the few examples of successful identification of susceptibility genes for complex traits.

It has been proposed that the identification of SNPs would facilitate association mapping. Although less than one per cent of SNPs have an impact on protein function (Venter et al. 2001) they may be valuable tools in the detection of linkage disequilibrium. However, there are several limitations in the use of SNPs. Since they are biallelic, their informativity is greatly reduced compared to microsatellite markers. This implies the testing of many markers, thus reducing the possibility to reach statistical significance, as corrections for multiple testing must be made. Due to the magnitude of markers needed, high-throughput methodology is required and this is now increasingly being made available to the research community.
Success so far and alternative strategies

Despite the great expectations cherished in the mid 1990s, there has been but a few identified genes in human complex traits so far. These include the identification of an association of the \textit{CAPN10} gene with type 2 diabetes (Horikawa et al. 2000), the \textit{NOD2} gene with Crohn’s disease (Hugot et al. 2001; Ogura et al. 2001), the \textit{ADAM33} gene with asthma (Van Eerdewegh et al. 2002) and the \textit{PCDC1} gene with systemic lupus erythematosus (Prokunina et al. 2002).

Most of these findings have relied on the collection of large materials. In type 2 diabetes, the association with the \textit{CAPN10} gene was initially detected in 170 families with Mexican-American origin (Horikawa et al. 2000), but has since then been replicated in other populations as well (Cassell et al. 2002; Malecki et al. 2002). The association of asthma with the \textit{ADAM33} gene was found using 460 affected sib-pairs (Van Eerdewegh et al. 2002). The \textit{NOD2} gene in Crohn’s disease was identified using 235 and 416 families, respectively (Hugot et al. 2001; Ogura et al. 2001). However, only 78 families were included in the initial linkage analysis identifying the susceptibility locus for Crohn’s disease at chromosome 16 (Hugot et al. 1996). In systemic lupus erythematosus, the association with the \textit{PCDC1} gene was detected using 2510 affected individuals, but in the initial linkage analysis no more than 19 multiplex families were included (Lindqvist et al. 2000).

In short segment Hirschsprung disease, three susceptibility loci were recently identified (Gabriel et al. 2002). By estimating each locus-specific risk ratio ($\lambda$), the combined effect of the three loci were shown to be both necessary and sufficient for the disease aggregation. Although it was concluded that the susceptibility gene at one of the loci is the \textit{RET} gene (known to be mutated in long segment Hirschsprung disease), \textit{RET} mutations could only be identified in 40% of linked families. This suggests that non-coding mutations or mutations in regulatory regions in the \textit{RET} gene contribute to short segment Hirschsprung disease, illustrating yet another obstacle in the dissection of complex traits.

In Hirschsprung disease, using subphenotypes of the disease turned out to be a valuable strategy. The susceptibility locus for systemic lupus erythematosus at chromosome 2q37 was initially identified using 19 multiplex families in a parametric linkage analysis. Other possible approaches in order to restrict the phenotype are the use of severe cases or cases with early age of onset of disease (Lander and Schork 1994).
Genetically homogenous populations, such as the Finnish or the Icelandic populations, may be advantageous for several reasons. It is assumed that the number of segregating susceptibility alleles is reduced compared to a heterogeneous population and that these as a consequence bring about a stronger genetic effect. Also, there is an increased likelihood for linkage disequilibrium in a homogenous population. In particular the Finnish population has the advantage of relatively recent bottlenecks. Although it is without doubt that this facilitates the mapping of monogenic traits by linkage disequilibrium techniques, it has been questioned whether it holds true for complex traits (Kruglyak 1999; Eaves et al. 2000; Johnson and Todd 2000; Altmuller et al. 2001). Nevertheless, evidences for linkage in the Finnish population have been found in for example asthma, multiple sclerosis, schizophrenia and autism-spectrum disorders (Ekelund et al. 2001; Laitinen et al. 2001; Paunio et al. 2001; Auranen et al. 2002; Saarela et al. 2002).

The Icelandic population, although a small and homogenous population, has not undergone recent bottlenecks (Edwards 1999). However, gene mapping in Iceland is carried out in an industrialized fashion using extensive genealogical and medical records and this strategy may be beneficial in the mapping of susceptibility genes (Kong et al. 1999). Evidences for linkage in the Icelandic population have been reported for example in asthma, hypertension, preeclampsia, schizophrenia and stroke (Arngrimsson et al. 1999; Gretarsdottir et al. 2002; Hakonarson et al. 2002; Kristjansson et al. 2002; Stefansson et al. 2002).

In spite of intense efforts in these apparently suitable populations, they have not yet resulted in the identification of susceptibility genes. Although the finding of susceptibility genes in a particular population may not generally be extrapolated to other populations, it would nevertheless give valuable insights in patophysiological mechanisms.
AIMS

The aim of this thesis was to identify genetic and environmental factors in the pathogenesis of hypospadias. The specific aims of the individual projects were to:

1. Establish the impact of low birth weight in hypospadias (paper I-II)
2. Estimate the role of genetic factors in hypospadias (paper II-III)
3. Localize susceptibility genes for hypospadias (paper IV)
4. Identify the gene responsible for dominant inherited hypospadias in a large family (paper V)
MATERIALS

Questionnaires were mailed to 2503 boys admitted for surgical correction of hypospadias at the departments of pediatric surgery in Stockholm, Uppsala, Lund, Gothenburg and the departments of urology and plastic surgery in Umeå and Örebro during the years 1970-1997. A final reply rate of 88% was attained. We asked for additional family members with hypospadias, the number of brothers and birth weight of the boys with hypospadias and their brothers. Patients in whom hypospadias was suspected to be part of a syndrome were excluded (e.g. identifiable syndromes, multiple malformations). All patients that reported relatives with hypospadias were contacted by telephone in order to obtain full information regarding the pedigree. In the present study, familial cases are defined as patients with one or more first, second, or third-degree relatives with hypospadias (table 2). A flow chart overview of the collection of families included in papers II-IV is shown in figure 6.

Overview of included material

In paper I, we included 28 twin pairs discordant for hypospadias. They were identified through the questionnaires and through the Department of Pediatric Surgery at Astrid Lindgren Children Hospital, Karolinska Hospital.

In paper II, we describe 2138 families with at least one boy with hypospadias ascertained through the departments in Sweden where boys with hypospadias undergo surgery. In eight families, there were four or more affected (pedigrees in figure 7).

In paper III, 2005 families with complete information of the pedigree were included, ascertained through the departments in Sweden where boys with hypospadias undergo surgery.

In paper IV, 69 families with at least two affected members were included after telephone interviews and blood sampling. They were identified through the questionnaires and through records from the Swedish Malformation Registry. In 58 of the included families both parents were born in Sweden, whereas 11 families originated from Middle Eastern countries (Turkey, Syria, Lebanon, Iran or Iraq). In seven families, there were three affected family members.

In paper V, one family with autosomal dominant inheritance of hypospadias was ascertained through the Department of Pediatric Surgery at Astrid Lindgren Children Hospital, Karolinska Hospital (pedigree in figure 8).
Figure 6
Flow chart showing the collection of families included in paper II-IV

2503 hypospadias cases in Sweden were mailed a questionnaire regarding additional family members with hypospadias and birth weight

2211 (88%) responded

73 refused participation, stated wrong diagnosis or adoption

2138 included in the analysis of heredity, birth weight and ethnic origin (paper II)

2005 with complete information on pedigree included in the complex segregation analysis (paper III)

144 (7%) familial cases

98 familial cases not included in the genome-wide linkage analysis

DNA from 46 complete families

DNA from 23 complete families ascertained through the Swedish Malformation Registry

69 families included in the genome-wide linkage analysis (paper IV)
Figure 7 Eight families with four or more affected with hypospadias (black squares)
Figure 8 Pedigree for a family with autosomal dominant inheritance of hypospadias, clinodactyly and feet abnormalities. Phenotypic features as indicated: black shading of the left half: hypospadias in males, urogenital abnormalities in females; stripes in right half: hand and/or foot abnormalities.
ETHICAL PERMISSIONS

The Ethics Committee at the Karolinska Hospital as well as the local Ethics Committees at each included department approved ethical permissions for this project. We obtained permission from the Swedish Data Inspection Board regarding the computer-based filing of the material. This information was publicly spread by advertisements in a national newspaper (Dagens Nyheter) according to the guidelines from the Swedish Data Inspection Board.
METHODS

Paper I, IV and V
DNA isolation
Genomic DNA was extracted from peripheral venous blood, skin or nails using standard protocols.

Paper I
Twin zygosity
Zygosity of twins can be established by outward resemblance, histopathological evaluation of the placenta and fetal membranes or DNA analysis. An evaluation based on outward resemblance, such as parents describing the twins as “alike as berries”, is considered to correspond to a significance level of 95% (Nancy Pedersen, the Swedish Twin Registry, personal communication). Histopathological examination of the placenta and fetal membranes has an estimated significance level of about 80%, whereas DNA analysis can reach significance levels of 99.999%, if many markers are included.

Here, we used data from the histopathological examination of placenta and fetal membranes in 2 twin pairs and DNA analysis in 15 pairs. In one pair, previous DNA analysis had been performed elsewhere. For the DNA analysis, eleven polymorphic microsatellite markers (on different chromosomes) were used for PCR amplification. PCR products were visualized by autoradiography after size fractionation by electrophoresis. Parents were included to enable the identification of shared alleles. According to allele sharing at each locus, probability ratios for monozygosity were calculated (p<0.005 considered statistically significant). Dizygosity was established when each twin had a different allele for at least two markers. Differences in birth weight were analyzed using a paired Student’s t-test.

Paper II
Analysis of heredity, ethnic origin, birth weight, twin rate and phenotype
Data concerning the pedigree and country of birth were supplemented from Statistics Sweden. The Swedish National Board of Health and Social Welfare provided information of birth weight and gestational age for 1722 boys with hypospadias and for 1417 of their unaffected brothers, serving as controls. Birth weight of the cases was compared with the mean value of their respective brother(s) using a paired Student’s t-test (n=946). Since birth weight depends on gestational age, we stratified cases and controls based on gestational age in four groups (w. 25-29, w. 30-35, w. 36-39, w. 40-45). Student’s t-test was used to compare cases and controls within each group.
Twin zygosity was either established by DNA analysis, as described in paper I, or by information given by the parents. Aχ² test was used to compare the observed twin rate with the expected.

Experienced pediatric urologists phenotyped 531 index cases and 145 familial cases through medical records. The severity of hypospadias was estimated according to statements in the records regarding the localization of the urethral orifice and the surgical method used. The degree of chordee and the presence of cryptorchidism were also evaluated. Hypospadias was classified as glandular with the orifice on the glans or in the sulcus, penile with the urethral orifice anywhere along the penile shaft or penoscrotal/perineal with the urethral orifice on the scrotum or in the perineum. Differences in phenotype between sporadic and familial cases were analyzed using a χ² test.

**Paper III**
Complex segregation analysis was performed to evaluate the genetic background in hypospadias in this material. A fundamental problem in segregation analysis involves the ascertainment through affected individuals, resulting in a biased segregation proportion. For this purpose, the probability of an affected individual to be ascertained (π) was estimated to 0.75. Affection status, sex and relationship for each member in 2005 pedigrees, forming 2080 nuclear families, were stated in a file used for the segregation analysis. Liability was set to 0.003 for males. Since hypospadias is a sex-limited trait that only affects males, the liability for females was set to 0. Four different genetic models (additive, multifactorial, dominant and recessive) were tested versus a sporadic model. The sporadic model corresponds to no familial resemblance. The complex segregation analysis involved a maximum likelihood analysis to find the best fit for the included pedigrees. The analysis was performed in POINTER (Lalouel and Morton 1981). Only nuclear families can be analyzed in this program and it therefore involves the usage of pointers, meaning through whom the family was selected.

**Paper IV-V**
**Genotyping**
PCR amplifications of 377 microsatellite markers were carried out as single reactions in 96-well plates. Each forward primer was fluorescently labeled in blue, green or yellow. Since the PCR products also differed in sizes, this enabled the simultaneous size fractionation of several (up to 15) markers on an ABI377 (Applied Biosystems). The resulting genotype data were analyzed with Genescan 2.1 and Genotyper 2.0 software.
(Applied Biosystems). For each marker, allele numbers were assigned and their sizes standardized using a control DNA with known alleles included on each 96-well plate and on each gel.

The basis for the microsatellite markers was the Weber 6 screening set (Sheffield et al. 1995). In sparsely covered regions, new markers were added from the Genome Database (http://www.gdb.org/) and the Marshfield Medical Research foundation (http://research.marshfieldclinic.org/genetics/). Mean heterozygosity for the autosomal markers was 0.76 and for the X chromosome 0.71. All genotyped markers were analyzed for Mendelian incompatibilities using zGenStat 1.126 software (Henric Zazzi, unpublished). All inconsistencies were reanalyzed and incompatibilities were either resolved unambiguously, or individuals and/or pedigrees were excluded from linkage analyses. To identify markers with allelic dropout or other problems, the expected number of homozygotes was calculated based on the estimated allele frequencies and compared with the observed numbers of homozygotes. For this the Pearson $\chi^2$- test as implemented in the zGenStat 1.126 software was used. Any marker showing significant deviation from expected homozygosity frequency ($p<0.001$) was reanalyzed, resulting in the exclusion of seven markers. A success rate less than 30% resulted in the exclusion of 10 markers. Thus, after quality assessments 17 of the 377 markers were excluded, resulting in 360 markers included in the genotyping with a mean average intermarker distance of 9.5 cM.

Family structures were verified using the SibError software (Ehm and Wagner 1998) based on genotype data of 128 markers spaced at 30 cM interval. We identified one monozygotic twin pair, which was excluded from the linkage analysis.

**Linkage analysis (paper IV)**

As the mode of inheritance is unclear, we used a non-parametric affected relative pair based linkage analysis to detect linkage. This was performed in the Allegro software (Gudbjartsson et al. 2000), which is capable of analyzing allele-sharing between more distant relatives (Nyholt 2000). The analyzed families differ in sizes and each family was therefore power weighted. Since hypospadias is restricted to males, females were coded as unknown. As suggested by Nyholt (2000), linkage analysis for the X chromosome was performed using MAPMAKER/SIBS (Kruglyak and Lander 1995a) through the HGMP web site (http://www.hgmp.mrc.ac.uk/). Corresponding p-values were interpreted according to Lander and Kruglyak (Lander and Kruglyak 1995), as summarized by Nyholt (2000). Allele frequencies were estimated from all genotyped individuals using the zGenStat 1.126 software (Henric Zazzi, unpublished). For the analyses in each
subgroup (i.e. families with Swedish and Middle Eastern origin) the allele frequencies were derived from each group. The map distances were based on the Marshfield map (http://research.marshfieldclinic.org/genetics/).

Single-point and multipoint linkage analyses were performed. In single-point analysis, LOD score calculations are performed for each marker in relation to the disease locus, independent of the surrounding markers, whereas multipoint analysis refers to the simultaneous analysis of several markers with known location on a genetic map, which increases information in the region (Ott 1996).

**Linkage analysis (paper V)**
Parametric linkage analysis was used in this family with apparent dominant inheritance of hypospadias. After an initial genome-wide linkage analysis resulting in evidence for linkage to the short arm of chromosome 7, markers D7S2514, D7S641, D7S2464, D7S664, D7S2557, D7S2508, D7S507, D7S503, D7S488, D7S2551, D7S493 and D7S673 were added. Two-point linkage analysis was performed using MLINK in the FASTLINK package through the HGMP web site (http://www.hgmp.mrc.ac.uk/), assuming an autosomal dominant model with full penetrance and the gene frequency 0.001.

**Association studies (paper IV)**
A transmission disequilibrium test was done in the pedigree disequilibrium test (Pdt) which can include several affected individuals in a single pedigree (Martin et al. 2000). The Pdt analyses were performed through the HGMP web site (http://www.hgmp.mrc.ac.uk/).

**Simulation analysis (paper IV)**
We performed simulation analyses to evaluate the results from the genome-wide linkage analysis and to obtain significance levels in all 69 families and in the subgroups consisting of Swedish and Middle Eastern families. Five thousand simulations were generated in Allegro, using the same family structures, the same observed allele frequencies and the same mean success rates for the autosomal markers as in the genome-wide linkage analysis. Genotypes were only generated for individuals that actually were genotyped in the genome-wide linkage analysis. Multipoint analysis was performed using a linear model weighted for each family.
Mutation analysis (paper V)

*HOXA13* is located in the region at chromosome 7p15 ([http://www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) with evidence for linkage in this family and was therefore subject to mutation analysis. The most sensitive method to detect an unknown mutation is by direct sequence analysis. Here, we used cycle sequencing with fluorescently labeled dideoxynucleotides (ddNTPs). The sequencing reaction involves PCR amplification using one primer and ddNTP chain terminators, resulting in randomly occurring stops in the amplified sequence. The resulting fragments of different lengths are size fractionated by electrophoresis and transferred into a chromatogram, in which the sequence can be interpreted.

We used PCR primers amplifying the whole coding region of *HOXA13* (Kosaki et al. 2002). DNA sequence analysis was performed on both strands of amplified and purified PCR products using the ABI PRISM BigDye Terminator CycleSequencing kit 2.0 (Applied Biosystems). The sequencing reactions were carried out according to the manufacturer’s recommendations and analyzed on an ABI310 DNA sequencer.
RESULTS AND DISCUSSION

Paper I
In this study including 28 twin pairs discordant for hypospadias we found 18 monozygotic twin pairs. In 16 of these, the twin with lowest birth weight was affected with hypospadias. The mean intra-pair difference in birth weight was 498 g (p<0.01). This difference was more pronounced than in a normal population of monozygotic twins (p<0.001).

Given the identical genetic background in monozygotic twins, this shows that low birth weight in itself is an important risk factor for hypospadias. Although post-zygotic events may cause discordance in monozygotic twins, this is an unlikely explanation in as many as 16 of 18 discordant twins. Low weight at birth reflects a growth retardation throughout gestation and has been shown to be correlated to sub-optimal first-trimester growth (Smith et al. 1998). In twins, it is well recognized that the twin with lowest birth weight is the smaller throughout gestation, i.e. also at the time of male external genitalia development (T-H Bui, personal communication).

It is here shown that the association with low birth weight is independent of genetic factors, raising some interesting questions. It can be assumed that the twins share intrauterine environmental factors, in all but one respect, the blood supply by the placenta. This suggests an inadequacy of the placenta to provide the fetus with nutrients and/or hormones. A placentary malfunction has previously been implied in hypospadias by the association with a low weight of the placenta (Stoll et al. 1990), severe preeclampsia (Akre et al. 1999) and dystocia (Källén 1988). During the first trimester, fetal androgen production is induced by maternal gonadotropins, preferably hCG produced in the placenta. Thus, a relative lack of hCG in the smaller twin can explain the increased susceptibility for hypospadias. Another explanation may be hypoxia in genital tissue as suggested by the observation of hypospadias in several cases with congenital anemia (e.g. homozygous alpha-thalassemia and hypotransferrinemia) (Dame et al. 1999; Fung et al. 1999; Goldwurm and Biondi 2000). Alternatively, the growth impairment in itself may render the fetus more vulnerable to inadequate endogenous hormone levels (e.g. androgen) or to deleterious exogenous environmental factors (e.g. estrogenic and anti-androgenic chemicals).
Paper II

In this study of a large number of families with at least one member with hypospadias, we found a familial rate of 7% (to be compared to 3% in the general population). Since we have relied on information given by family members, this is likely to be an underrating and the actual familial rate is probably higher. This finding supports genetic factors in the pathogenesis of hypospadias.

We obtained additional evidence for the impact of low birth weight, since the boys with hypospadias differed significantly in birth weight compared to their brothers (p=5x10^{-13}).

Since hypospadias is a sex-limited trait, we only asked for male siblings and were consequently only able to detect male-male twins. An increased frequency of twins was nevertheless found in this material. Given a twin rate of 1 per 80 pregnancies and the occurrence of male-male twins in 1/3 of these (figure 9), we could expect nine male-male twins. Dizygosity could be expected in six of these and monozygosity in three.

**Figure 9**

Distribution of zygosity and sex in twins

- 1/3 monozygotic
  - 1/6 monozygotic male-male
  - 1/6 monozygotic female-female

- 1/3 dizygotic same-sex
  - 1/6 dizygotic male-male
  - 1/6 dizygotic female-female

- 1/3 dizygotic unlike sex
  - 1/3 dizygotic male-female

We observed 40 male-male twins (p=4x10^{-25}). Zygosity was established in 33 twins, of which 11 were dizygotic and 22 monozygotic. Thus, we observed an increased frequency of dizygotic as well as monozygotic male-male twins in this material. The finding that two-thirds were monozygotic deviates from the expected 50:50 distribution in twins of same sex (figure 9).
An over-representation of monozygotic twins has previously been described and attributed either to a common denominator in monozygosity and hypospadias or to some predisposing factor in association with monozygotic twinning (Roberts and Lloyd 1973). That the risk for hypospadias seems to be higher in male-male twins, regardless of zygosity, again suggests a relative lack of androgen-inducing hormones (i.e. hCG) as one cause for hypospadias.

The high twin rate may be due to the use of assisted reproduction, however this was not investigated here. An increased risk for hypospadias following in vitro fertilization and intracytoplasmic sperm injection has been reported (Macnab and Zouves 1991; Silver et al. 1999b; Wennerholm et al. 2000; Ericson and Källén 2001).

With regards to the distribution of phenotype, we found a higher proportion of intermediately affected cases (i.e. penile) than in most previous studies (table 1) (Sørensen 1953; Roberts and Lloyd 1973; Sweet et al. 1974; Monteleone Neto et al. 1981; Calzolari et al. 1986; Källén et al. 1986). Since we defined hypospadias as penile as soon as the urethra opened anywhere along the penile shaft (including juxta-coronal variants), many of the penile cases are relatively mild variants with the urethral orifice located only a few millimeters proximal to the corona. A similarly high frequency of penile cases was described in the patients studied by Bauer (table 1) (Bauer et al. 1981). However, this population consisted mostly of referrals from pediatricians and urologists, whereas we consider our group of patients to be representative for a general hypospadias population. We found significant differences between familial and sporadic cases with regards to glandular and penoscrotal/perineal variants. Severe variants were less common in familial cases than in sporadic cases. An explanation for this may be a reduced fertility in severe forms of hypospadias.

Interestingly, 6% (n=134) of the cases originated from the Middle Eastern region (Turkey, Syria, Lebanon, Iran or Iraq), to be compared with 2% in the general Swedish population (Susanne Dällöf, Statistics Sweden, personal communication). Of the 134 subjects from Middle Eastern countries, 22% reported additional family members with hypospadias. In the 144 familial cases, 20% originated from Middle Eastern countries. These observations speak in favor of genetic factors in hypospadias. Consanguinity is frequent in this region, suggesting that recessive genes are involved in hypospadias in this population.
Paper III

In the complex segregation analysis of 2005 pedigrees we obtained further evidence for a familial aggregation of hypospadias. The high heritability of 0.99 suggests that there are monogenic effects in this material; however, the best fit was achieved for a multifactorial model. This indicates that there are major genes acting in a minor proportion of the families but that there is a multifactorial cause in the majority of cases. Some of these may even be phenocopies, in which hypospadias is caused by environmental factors only (e.g. low birth weight). Similar conclusions were drawn from the segregation analysis (performed in POINTER) of 103 Danish families (Harris and Beaty 1993). The heritability in that study was equal to the one in this study (0.99).

In this study, more than 2000 families were ascertained through a nation-wide scan for hypospadias cases. We initially contacted 2500 patients, corresponding to half of all registered cases in Sweden, and included 80% of these in the segregation analysis. There are no simple rules with regards to the size of the material for segregation analysis. The power depends on the structures of the pedigrees and on the underlying genetic model, which by definition is unknown, but a larger material will nevertheless increase the power.

Paper IV

In the genome-wide linkage analysis, we found five chromosomal regions with suggestive linkage (figure 10). According to the definition of suggestive linkage, one randomly occurring peak per genome-wide linkage analysis is expected. It is still possible that each of these five peaks represent true susceptibility genes for hypospadias. Interestingly, there were no peaks in regions harboring the genes most frequently found in monogenic variants of hypospadias (e.g. the androgen receptor gene \(AR\), Xq11-12) and the recessive 5-alpha-reductase deficiency \(SRD5A2\), 2p23). This suggests that the susceptibility genes for hypospadias may be different from the ones mutated in monogenic variants of hypospadias.
Figure 10 shows the results from the multipoint linkage analysis for hypospadias in all 69 families (black line), in the 58 families with Swedish origin (blue line) and in the 11 families with Middle Eastern origin (red line). The LOD score is given on the y-axis and cM from pter on the x-axis. The linkage analysis was performed in Allegro (Gudbjartsson et al. 2000) for the autosomes and in MAPMAKER/SIBS (Kruglyak and Lander 1995a) for the X chromosome. Multipoint analysis was performed using a linear model weighted for each family. The horizontal line denotes LOD score=1.44.
How can the search for susceptibility genes be pursued? Lander and Kruglyak (1995) recommend that suggestive linkage should be followed up in extension studies. (The term extension study is preferred for the testing of additional families in the search for significant linkage, whereas replication study should be reserved for situations in which significant linkage has already been obtained). In the extended study, they recommend that data are pooled and the entire dataset reanalyzed (Lander and Kruglyak 1995). An alternative strategy is to study founder populations. Families with several affected may also be useful, as demonstrated in the recent identification of the PCDC1 gene in systemic lupus erythematosus (Prokunina et al. 2002).

Estimates of the desired number of families (i.e. power calculations) are difficult to perform, and to evaluate, since it requires specifications of the model and this cannot be reliably made. It was originally believed that several hundred families would be needed for the identification of susceptibility genes in common traits and that these numbers would be correlated to the relative risk (Risch 1990b). However, it should be noted that only 78 families were included in the original linkage analysis for Crohn’s disease, eventually resulting in the identification of the NOD2 gene (Hugot et al. 1996; Hugot et al. 2001).

Adding more markers has the advantage of increasing information, given that all previously markers were not fully informative (Haines and Pericak-Vance 1998). There is, however, no endpoint equivalent to the situation in the mapping of monogenic traits in which recombinations restrict the region (Kruglyak and Lander 1995b). Instead, association mapping may be used to further narrow the region. Fine mapping of the candidate region, before proceeding with investigations in candidate genes, has been strongly recommended (Lander and Kruglyak 1995).
The identification of hand-foot-genital syndrome (HFGS) in this family with autosomal dominant inherited hypospadias was unexpected, since the skeletal malformations are much less severe than in previously reported HFGS families. However, the finding of a novel polyalanine insertion in the \textit{HOXA13} gene clearly speaks in favor of this diagnosis, albeit an atypical variant of the syndrome.

HFGS is an autosomal dominant syndrome characterized by skeletal anomalies in the distal limbs and urogenital malformations (McKusick 1986, 1990). Typical skeletal manifestations in the hand include short, proximally placed thumbs and clinodactyly of the fifth finger and in the feet short, medially deviated great toes and fusion defects of the bones. The radiographic pattern is characteristic for the syndrome (Poznanski et al. 1970). The skeletal manifestations are invariable and highly penetrant, whereas the urogenital abnormalities show reduced penetrance and variable expression. Females have Müllerian duct fusion defects such as uterus bicornis, vaginal septum and ectopic localization of ureteric and urethral orifices. Males often present with hypospadias.

Eight families and four sporadic cases with HFGS have previously been described (Stern et al. 1970; Poznanski et al. 1975; Giedion and Prader 1976; Goeminne 1981; Verp et al. 1983; Halal 1988; Cleveland and Holmes 1990; Donnenfeld et al. 1992; Fryns et al. 1993; Devriendt et al. 1999; Goodman et al. 2000; Utsch et al. 2002). The phenotype varies both within and between these families (table 4).

Three additional families with some of the characteristic features of HFGS have been reported (Longmuir et al. 1986; Hennekam 1989; Guttmacher 1993). The syndrome described by Longmuir (1986) consists mainly of distal skeletal malformations. The family reported by Hennekam (1989) has atypical features of HFGS including Müllerian duct fusion defects and ear malformations, but no hand malformations. The syndrome designated as Guttmacher (1993) show some features of HFGS, such as short thumbs and great toes as well as hypospadias in males, but also includes postaxial polydactyly.
**Table 4** Phenotype and mutation data in the 9 families and 4 sporadic cases with HFGS reported thus far

<table>
<thead>
<tr>
<th>Author and year</th>
<th>Affected</th>
<th>Hand phenotype</th>
<th>Foot phenotype</th>
<th>Genital phenotype</th>
<th>Type of mutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stern 1970</td>
<td>18</td>
<td>short thumbs, clinodactyly</td>
<td>small feet, tarsal fusion, short calcaneus, short toe 1, hallux varus</td>
<td>double uterus, vaginal septum</td>
<td>nonsense</td>
</tr>
<tr>
<td>Poznanski 1975</td>
<td>10</td>
<td>short thumbs, clinodactyly</td>
<td>tarsal fusion, short calcaneus, short toe 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gideon 1976</td>
<td>3 brothers</td>
<td>short thumbs, clinodactyly</td>
<td>short calcaneus, hallux varus</td>
<td>hypospadias</td>
<td></td>
</tr>
<tr>
<td>Goeminne 1981</td>
<td>sporadic male</td>
<td>clinodactyly</td>
<td>short toe 1, hallux varus</td>
<td>uterine bicornis, vaginal septum, hypospadias</td>
<td></td>
</tr>
<tr>
<td>Verp 1983</td>
<td>7 affected in 4 generations</td>
<td>short thumbs, clinodactyly</td>
<td>short toe 1, hallux varus</td>
<td>uterus bicornis, vaginal septum, hypospadias</td>
<td>polyalanine insertion (+8) in third tract</td>
</tr>
<tr>
<td>Donnenfeld 1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halal 1986</td>
<td>6</td>
<td>short thumbs, clinodactyly</td>
<td>tarsal fusion, short calcaneus, short toe 1, hallux varus, toe syndactyly</td>
<td>uterus bicornis, vaginal septum, ectopic ureter and urethra</td>
<td></td>
</tr>
<tr>
<td>Cleveland 1990</td>
<td>3</td>
<td>short thumbs, clinodactyly</td>
<td>tarsal fusion, short toe 1, hallux varus</td>
<td>uterus bicornis, hypospadias</td>
<td>nonsense</td>
</tr>
<tr>
<td>Fryns 1993</td>
<td>4 (father and 3 sons)</td>
<td>short thumbs, clinodactyly</td>
<td>small feet, short toe 1</td>
<td>hypospadias</td>
<td>nonsense</td>
</tr>
<tr>
<td>Devriendt 1999</td>
<td>sporadic male</td>
<td>short thumbs, clinodactyly</td>
<td>tarsal fusion, short toe 1, hallux valgus</td>
<td>cryptorchidism, chordee</td>
<td>deletion of 7p14</td>
</tr>
<tr>
<td>Goodman 2000 (1)</td>
<td>sporadic male</td>
<td>short thumbs, clinodactyly</td>
<td>short toe 1, hallux varus</td>
<td>short penis</td>
<td>nonsense</td>
</tr>
<tr>
<td>Goodman 2000 (5)</td>
<td>sporadic male</td>
<td>extremely short thumbs, hypoplasia of phalanges</td>
<td>tarsal fusion, absence of toe 1, hypoplasia/absence of phalanges</td>
<td>glandular hypospadias</td>
<td>missense</td>
</tr>
<tr>
<td>Utsch 2002</td>
<td>6 affected in 5 generations</td>
<td>short thumbs, clinodactyly</td>
<td>small feet, short toe 1, hallux varus</td>
<td>uterus bicornis, double cervix, vaginal septum</td>
<td>polyalanine insertion (+6) in third tract</td>
</tr>
<tr>
<td>Frisén 2002</td>
<td>28 affected in 6 generations</td>
<td>clinodactyly</td>
<td>small feet, gap between toe 1 and 2, short toe 2, hallux varus</td>
<td>hypospadias (glandular in 7, penile in 3)</td>
<td>polyalanine insertion (+6) in second tract</td>
</tr>
</tbody>
</table>
In 1997, it was reported that HFGS is caused by mutations in the \textit{HOXA13} gene (Mortlock and Innis 1997). An overview of the mutations that have since then been identified in HFGS is shown in figure 11. The phenotype associated with the previously reported mutations is essentially the same, except for that produced by the missense mutation, which affects the hands and feet much more severely. In addition, an interstitial deletion removing the entire \textit{HOXA} cluster was found in the family reported by Devriendt (1999). Interestingly, this patient displays a phenotype that is relatively mild, especially with regards to the genital manifestations. In Guttmacher syndrome, both a missense mutation in the homeobox region of the \textit{HOXA13} gene and a dinucleotide deletion in the promoter was recently found (Innis et al. 2002).

\textbf{Figure 11}

Mutations in \textit{HOXA13} in HFGS families

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure11}
\caption{Mutations in \textit{HOXA13} in HFGS families}
\end{figure}

The polyalanine insertions consist of cryptic expansions (i.e. GCA, GCC, GCG, GCT) rather than trinucleotide repeats, are stable through generations and are believed to have originated through unequal crossing-over (Warren 1997). Similar polyalanine tract expansions have been described in \textit{HOXD13} in several families with synpolydactyly (Muragaki et al. 1996; Goodman et al. 1997; Goodman et al. 2002).
HOX genes are highly conserved through evolution and were first identified in Drosophila as key regulators of patterning of the body plan. Mutations in Drosophila result in a certain body segment loosing its positional identity and being transformed to another body structure, i.e. a homeotic transformation. In the C-terminal region of HOX genes there is a 180 bp stretch of highly conserved DNA coding for a 60 amino acids homeodomain. This domain binds to DNA and regulate transcription in interaction with other proteins. The potential target genes and interacting proteins are numerous and not fully characterized (Veraksa et al. 2000).

In all species, genes within a HOX cluster are arranged according to their temporal and spatial expression during development, a phenomenon known as colinearity. In man, there are four paralogous clusters (HOXA, HOXB, HOXC, HOXD), which probably originated through two successive gene duplications. A certain HOX gene is functionally more similar to its paralogue in another cluster than to a neighboring gene in the same cluster. For example, the paralogues HOXA13 and HOXD13 are located at the 5’ end of their respective cluster and are both important for the development of the limbs and the lower urogenital tract. Thus, they are likely to have partly overlapping functions with respect to the embryological development of digits and genitalia (Fromental-Ramain et al. 1996; Warot et al. 1997). Both these structures are regions of apical growth and represent morphogenetic end organs. In line with this, it has been suggested that digits and penis have a related phylogenetic history (Kondo et al. 1997).

The mutations in HOXA13 and HOXD13, together with a mutation in HOXA11 resulting in amegakaryocytic thrombocytopenia and radio-ulnar synostosis (Thompson and Nguyen 2000), are the only mutations in HOX genes described in man to date. Characteristically, these abnormalities are discrete and may easily escape medical attention.

The family described here displays a milder phenotype than previously reported HFGS families. Affected individuals are heterozygous for a polyalanine expansion in the second polyalanine tract in the first exon of the HOXA13 gene. This is in contrast to the two previously reported polyalanine expansions in the HOXA13 gene localized in the third tract. It is likely that the discrepancies in phenotype between this and previously reported HFGS families are caused by the different localization of the insertion.
It remains unclear whether the mutated HOXA13 protein acts by haploinsufficiency or a dominant negative mechanism. The patient with HFGS carrying a heterozygous deletion of the entire HOXA cluster is an example of haploinsufficiency of the HOXA13 gene (Devriendt et al. 1999). Interestingly, he has a relatively mild variant of HFGS, in particular with regards to the genital malformations that are limited to cryptorchidism and chordee. In contrast, the patient with a missense mutation has an unusually severe phenotype (Goodman et al. 2000), suggesting a dominant negative effect. This is further supported by the finding that mice heterozygous for a minor deletion within the Hoxa13 gene have a more severe limb phenotype than mice homozygous for a Hoxa13 null mutation (Fromental-Ramain et al. 1996; Mortlock et al. 1996).

Several lines of evidence indicate that the polyalanine expansions described in HOXD13 may as well render a dominant negative effect. In synpolydactyly, the severity has been shown to correlate with the size of the polyalanine expansion and in the family with the largest expansion (14 additional alanines) affected males have hypospadias (Goodman et al. 1997). Mice homozygous for a spontaneous polyalanine expansion (expanding the stretch from 15 to 22 alanines) in the Hoxd13 gene have a much more severe phenotype than mice with complete absence of Hoxd13 function (Zakany and Duboule 1996; Johnson et al. 1998). Genetic complementation studies in this mouse model indicate that the mutated protein exerts a "super" dominant negative effect, by interfering with the function of the remaining wild-type Hoxd13 and other 5' Hoxd proteins (Bruneau et al. 2001). Moreover, mice with homozygous deletions of Hoxd11, Hoxd12 and Hoxd13 have a less severe phenotype than mice and humans with polyalanine tract expansions in HOXD13 (Zakany and Duboule 1996; Johnson et al. 1998).
GENERAL DISCUSSION

In this thesis, data suggesting a significant impact of low birth weight on the pathogenesis of hypospadias is presented. Hypospadias is known to be strongly associated with low birth weight, but it remains unclear whether it is the low birth weight in itself that renders the fetus more susceptible for other predisposing factors (genetic or environmental) or if there is a common denominator in the genesis of the two conditions. Here, we show that the association for hypospadias with low birth weight is independent of genetic factors in discordant monozygotic twins. Since monozygotic twins are genetically identical and can be assumed to share intrauterine environmental factors, this points to a role of the placenta in the pathogenesis of hypospadias. This exemplifies a situation in which hypospadias results from environmental factors only, although it is still possible that the condition arose in a genetically predisposed individual surpassing a threshold.

Several findings in this thesis also suggest a genetic cause for isolated hypospadias: a 7% familial rate, a high rate of sporadic as well as familial hypospadias in Middle Eastern populations (in which consanguinity is frequent) and a heritability of 0.99. However, there is little support for monogenic effects in hypospadias, at least in the majority of this material. Nevertheless, in one family, we identified a mutation in a HOX gene causing autosomal dominant inheritance of hypospadias. This illustrates the extreme end of the spectrum in which a single genetic alteration is sufficient to cause the malformation. It is likely that the background for hypospadias constitutes a continuum from environmental factors only to single genetic effects, with the majority of cases caused by the interaction of several susceptibility genes with or without environmental factors.

With the aim to identify the susceptibility genes for hypospadias, a genome-wide linkage analysis was performed. Five loci with suggestive evidence for linkage were identified. These regions need to be investigated further, preferably in an extended population. But it is not improbable that there are in fact five genes behind the susceptibility for hypospadias and that these eventually will be identified. Two different models may explain the genetic background for hypospadias as well as other complex traits. One theory is that common alleles at a few loci interact to cause the disease, i.e. the common disease/common variant hypothesis. The other concept is that rare alleles at numerous loci each on its own cause the disease, i.e. genetic heterogeneity (Pritchard and Cox 2002; Smith and Lusis 2002). To date, the available sample of known complex disease genes is too small to draw any general conclusions. In the end, both these viewpoints may turn out to be accurate, within a disease as well as in different diseases. The results from the
genome-wide linkage analysis presented here may suggest a common disease/common variant model in hypospadias, since it is unlikely that we would have detected five loci unless they all contribute to the phenotype in some way.

It has been suggested that the polarized view of monogenic versus complex traits is outdated (Badano and Katsanis 2002). Some classical monogenic diseases have turned out to be not so simple (e.g. lack of genotype-phenotype correlation in some families with cystic fibrosis) (Dipple and McCabe 2000). This has in part been attributed to the influence of modifier genes. In other diseases, an unexpected mode of inheritance has been found (e.g. triallelic inheritance in Bardet-Biedl syndrome) (Katsanis et al. 2001). Although complex traits are generally believed to result from the interaction between several genes and environmental factors, in some individuals the phenotype is caused by mutations in one gene. These observations speak in favor of a transition from a segmented view of human genetic diseases to a conceptual continuum between monogenic and complex traits.

The overall aim of this thesis was to gain an increased understanding of the pathogenesis for hypospadias. An additional benefit would be to shed light on key mechanisms in sex differentiation. In that respect, this is basic research with no immediate clinical application, although the importance of increased information to affected families should not be undervalued. It is nevertheless possible that the future identification of the molecular basis for hypospadias may enable causal treatment.
SAMMANSATTNING PÅ SVENSKA

Hypospadi är en av de vanligaste missbildningarna med en förekomst av cirka 0.3 % och innebär att urinrörsmyningen är lokaliserad på undersidan av penis. Mycket talar för att hypospadi har en genetisk bakgrund i kombination med miljöfaktorer. Den övergripande målsättningen med denna avhandling var att identifiera orsaker till hypospadi.

Flera grundläggande genetiska metoder utnyttjades i ett landsomfattande material, motsvarande cirka hälften av alla registrerade fall i Sverige. Drygt 2500 individer med hypospadi erhöll en enkät bestående av frågor om ytterligare fall av hypospadi i släkten och födelsevikt. Detta material utgör grunden för fem delarbeten i avhandlingen.

I. Tvillingstudie i syfte att utvärdera miljöfaktorers inverkan vid uppkomsten av hypospadi. Enäggstvillingar är genetiskt identiska, medan tvåäggstvillingar delar hälften av det genetiska materialet, precis som syskon. Studier av enäggstvillingar där bara den ena är sjuk kan ge värdefull information om miljöfaktorers inverkan. I materialet identifierades 18 enäggstvillingpar, där enbart en av tvillingarna hade hypospadi. I 16 av dessa par var det den mindre tvillingen som drabbats. Detta visar att låg födelsevikt, oberoende av den genetiska bakgrunden, har betydelse för uppkomsten av hypospadi.

II. Epidemiologisk studie i syfte att dels undersöka andelen familjära fall i materialet, dels att fortsatt analysera sambandet mellan hypospadi och låg födelsevikt. Ett eller flera ytterligare fall av hypospadi i släkten rapporterades av 7 % (n=144). Det förelåg en signifikant skillnad i födelsevikt hos pojkar med hypospadi jämfört med sina respektive bröder (p=5x10^-13). Vi fann en ökad frekvens av såväl enägg- som tvåäggstvillingar. Dessutom kartlade vi den etniska bakgrunden i hela materialet och karaktäriserade graden av hypospadi i en tredjedel av materialet.

III. Segregationsanalys i syfte att definiera nedärvningsmönstret för hypospadi. En multifaktoriell modell var mest förenlig med nedärvningsmönstret i vårt material. Heritabiliteten beräknades till 0.99. Detta tolkas som att hypospadi har en multifaktoriell bakgrund i de flesta fall, medan mutationer i enskilda gener orsakar en mindre andel.

IV. Syskonparanalys med målsättning att identifiera gener som bidrar till uppkomsten av hypospadi. Syskonparanalys innebär att man jämför markörer mellan sjuka individer inom många familjer för att identifiera kromosomregioner som de delar i större utsträckning än förväntat. I denna studie inkluderades 69 familjer med två eller flera pojkar med hypospadi. Alla familjemedlemmar inklusive föräldrar genotypades med
360 fluorescensmärkta polymorfa mikrosatellitmarkörer spridda över hela genomet med i genomsnitt 9.5 cM avstånd. Vi fann fem intressanta kromosomregioner som är föremål för utvidgade studier.

V. Mutationsanalys i en familj med autosomalt dominant nedärvning av hypospadi.
Med kopplingsanalys lokaliserades genen till ett mindre område på kromosom 7, där \( \textit{HOXA13} \) genen är belägen. Mutationer i \( \textit{HOXA13} \) ger upphov till ett mycket ovanligt syndrom, hand-foot-genital syndrome (HFGS). Endast 12 familjer i världen finns beskrivna tidigare och hos 8 av dessa har man hittat en mutation i \( \textit{HOXA13} \) genen. Vi identifierade en ny mutation i \( \textit{HOXA13} \) genen i denna familj med atypisk variant av HFGS. Det är den hittills största beskrivna HFGS-familjen med 28 fall i 6 generationer. HFGS karaktäriseras av milda utvecklingsstörningar i skelett och genitalia. Typiska skelettavvikelser är små fotter, små stortår, små tummar och krokiga lillfingerar. Kvinnor med HFGS kan ha hjärtformad livmoder, skiljevägg i vagina, inkontinens och upprepade urininfektioner på grund av felplacerad mynning av urinrörelse och urinledare. Män med HFGS har ibland hypospadi. I denna familj är hypospadi ett vanligare fenomen än i tidigare beskrivna familjer och skelettavvikelserna är betydligt mildare. Mutationen utgörs av en insertion av sex extra alanin inom ett område där inga tidigare mutationer finns beskrivna. Den atypiska bilden i denna familj kan förklaras av mutationens lokalisation.

\section*{Slutsats}

Den övergripande målsättningen med arbetet var att öka förståelsen för bakgrunden till hypospadi, vilket dessutom kan ge insikter i grundläggande mekanismer för könsutvecklingen. Det kan också medföra ökad information till drabbade familjer och framtidiga identifiering av den molekylära orsaken till hypospadi kan möjliggöra riktad behandling.
ACKNOWLEDGEMENTS

I have been encouraged and supported by many people, to whom I am very thankful:

Agneta Nordenskjöld, my supervisor, for being so caring and patient during all these years and for limitless stamina, numerous ideas and vast knowledge on hypospadias

Holger Luthman, my co-supervisor, for encouragement and consideration, for enthusiasm and many stimulating discussions

Ingrid Kockum, my almost-supervisor, for lots of assistance, for always being helpful and for patiently explaining statistics in complex traits

Margareta Tapper-Persson for being the laboratory technician every PhD student dream of, for being so caring and kind since the very first day and for unbelievable amount of practical help

Magnus Nordenskjöld for creating an enjoyable atmosphere as well as a thorough research education and for engagement in this project, but most of all for being such a considerate person and for bringing many laughs into the lab

Lennart Iselius for generous and competent collaboration in the segregation analysis

Paul Lichtenstein and Nancy Pedersen for cooperation in the twin study and for their excellent research course in genetic epidemiology

Einar Hansson, Staffan Holmner, Lars Lundquist, Göran Läckgren, Jörgen Pedersen, Arne Stenberg and Gunnar Westbacke for providing records of their patients with hypospadias

Kicki Lagerstedt for being so energetic and helpful, in particular for invaluable sequencing assistance

Cilla Söderhäll for lots of practical advice along the way and many fruitful discussions, for valuable comments on this thesis and for being a good friend

Maria Bradley for shared struggles with complex genetics and for continuous close friendship

My fellow PhD students throughout the years: Britt-Marie Anderlid, Erik Björck, Lovisa Bylund, Kim Ericson, Heléne Fischer, Pernilla Holm, Erik Iwarsson, Shideh Khodaei-O’Brien, Charlotta Lindvall, Ann Nordgren, Sofia Persson, Tiina Robins, Jacqueline Schoumans and Emma Tham for lots of fun and friendship

Magdalena Fossum and Mina Kalbasi for good spirits and continued work in the hypospadias field

Former and present group leaders and post docs in the department: Maria Anvret, Jan Dumanski, Catharina Larsson, Nils-Göran Larsson, Annika Lindblom, Helena Malmgren, Kevin O’Brien, Martin Schalling, Keng-Ling Wallin, Günther Weber, Anna Wedell and Gabrielle Åhlberg for being such nice and helpful people and for showing interest in this work
Yvonne Cowan, Ann-Mari Dumanski and Kerstin Florell for lots of help with administrative matters

Laboratory technicians Sigrid Sahlén, Margareta Tapper-Persson, Anki Thelander and Barbro Werelius for providing everything including the joyful atmosphere

Sivonne Arvidsson and Ingrid Delin for teaching me ABI gel running and the interpretation of genotypes

Annika Röhl for invaluable help in the preparation of the sex differentiation presentation and for sharing numerous pictures as well as advice on layout

Lennart Nilsson for generously letting me use his extraordinary photographs

Susanne Dahllöf at Statistics Sweden for efficiently providing invaluable information

I am indebted to several people for helping out with computer-related obstacles: Carl Bruder for once introducing me to the world of computers, The-Hung Bui for patiently converting personal computer based items, Jacci Fredell, Godfather and uncle, for always taking time to help and in addition being better at it than any Excel-manual and Lennart Helleday, Rudolf Matousek and Dagmar Vejsicka for patience with a computer-illiterate and for being so nice and helpful

Komplexa gruppen for struggling with basic concepts in complex traits, in particular Ingrid Kockum for guidance

Karin Kindberg and colleagues at the Clinical Genetics laboratory for being helpful and friendly, in particular Ulla Grandell for practical help

Clinical geneticists Elisabeth Blennow for willingly answering questions and The-Hung Bui for valuable comments on the twin study

Barbro Högrell for advice and help during my first year in the lab

Summer students Enes Efendic, Madeleine Frederich, Linn Hjortsberg, Åsa Kallas, Philip Waldenström and Lisa Örtqvist for hard work

Anna Lundh and Carl Johan Sundberg for providing me with all kinds of projects inside KI, outside the lab, and for being good friends and fellow teachers

Rolf Heuman and colleagues at the surgical clinic, Mora lasarett, for giving me the opportunity to spend holidays and weekends on call and for being clinical role models

Anna Barnéus, Per Borgström and Tony Frisk for great times in medical school and for continued close friendship since then

Isabelle Dahlborg Lidström for lots of practical help and advice on layout

All friends outside the lab for being so good at having fun and for sweeping me away too
Ann Erling and my parents in law Marianne Frisén and Lars Frisén for showing surprisingly much interest in this work, for many good advice and helpful comments and for welcoming me so warmly into the family

All cousins, aunts and uncles for life long friendship and for many enjoyable family gatherings

Ebba von Rosen, Godmother and close friend throughout life, for generosity and support and for companionship at the Karolinska

Lottie and Nicolas for being the best imaginable siblings, for many laughs and lessons along the way and for being such good advisers

My parents for continuous support and trust, for being immensely caring, generous and fun and my role models on the brighter side of life

Jonas for innumerable lunches during the first part of this thesis and for marriage in the end of it, for being a wonderful life-companion and Kraken’s master, for providing inspiration in science as well as in all other aspects of life and for infinite patience and support

My deep gratitude to all boys with hypospadias and their families for participation

This work was supported by scholarships and grants from HRH Crown Princess Lovisa Foundation, the Swedish Research Council, Åke Wiberg Foundation, Magnus Bergvall Foundation, Marcus Borgström Foundation, Förenade Liv, Karolinska Institutet, Ingabritt och Arne Lundbergs forskningsstiftelse, Ronald McDonald Child Foundation, Erik Rönnerbergs donationer (Riksbankens jubileumsfond), Stiftelsen Frimurare Barnhuset, Stiftelsen Samariten, the Swedish Society of Medicine and Sällskapet Barnavård.
REFERENCES

Adams F (1844-1847) The Seven Books of Paulus Aegineta. Translated from the Greek. With a Commentary Embracing a Complete View of the Knowledge Possessed by the Greeks, Romans and Arabians on All Subjects Connected with Medicine and Surgery. vol 2, London, pp 347


Galen (c. 130-201 A.D.) Opera Omnia, vol 10. C.G. Kühn, Leipzig, pp 1001


Källén B (1988) Case control study of hypospadias, based on registry information. Teratology 38: 45-50


Lamy M (1952) Lecture. University Institute of Human Genetics, Copenhagen, Denmark


Penrose LS (1935) The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. Annual Eugenics 6: 133-8


Risch N (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46: 229-41


Sadler TW (1990) Langman`s medical embryology, Sixth edn. Williams & Wilkins, Baltimore, Maryland, USA


