

From Department of Women's and Children's Health
Karolinska Institutet, Stockholm, Sweden

HYPOSPADIAS: GENE MAPPING AND CANDIDATE GENE STUDIES

Trinh Thi Thai Hanh



**Karolinska
Institutet**

Stockholm 2009

Supervisors

Agneta Nordenskjöld, professor
Department of Women's and Children's Health, Karolinska Institutet

Kristina Lagerstedt Robinson, PhD
Department of Molecular Medicine and Surgery, Karolinska Institutet

Johanna Lundin, PhD
Department of Molecular Medicine and Surgery, Karolinska Institutet

Opponent

Olaf Hiort, professor
Pädiatrische Endokrinologie und Diabetologie
Klinik für Kinder- und Jugendmedizin
Universitätsklinikum Schleswig-Holstein, Campus Lübeck
Lübeck, Germany

Thesis committee

Ewa Ehrenborg, associate professor
Department of Medicine, Karolinska Institutet

Kate Abrahamsson, associate professor
Sahlgrenska Academy, Pediatric Surgery, Göteborg University

Lena Ekström, associate professor
Department of Laboratory Medicine, Clinical Pharmacology, Karolinska Institutet

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

Published by Karolinska Institutet. Printed by [Larserics Digital Print]

© Trinh Thi Thai Hanh, 2009
ISBN 978-91-7409-597-5

To my Parents

ABSTRACT

Hypospadias is a common congenital malformation in boys, characterized by incomplete fusion of the urethral folds, abnormal opening of urethra and different degrees of curvature of the penis. In Sweden, the incidence of hypospadias is 1.14 per 300 male live-births according to the annual Swedish Malformation Registry. Hypospadias is considered to be a complex genetic disorder caused by the interplay between environmental factors and the additive effects of multiple genes. Several observations suggest that hypospadias are under genetic influence. To identify disease genes in complex disorders, two main strategies have been used. We have performed a whole genome screening for linkage in families with additional cases of hypospadias and also performed DNA sequencing of candidate genes for hypospadias

To identify the chromosomal loci involved in the pathogenesis of hypospadias, a genome-wide linkage analysis in a three-generational family showing autosomal dominant inheritance of hypospadias was performed. Fifteen individuals, whereof seven affected, were genotyped within a total of 426 microsatellite markers and the genotyping results were analyzed using parametric and non-parametric linkage analyses. The genome-wide linkage analysis and subsequent fine mapping gave a maximum linkage in both parametric (LOD score 2.71) and non-parametric (NPL score 5.01) single-point analyses for marker D7S640. A susceptibility haplotype shared by all affected boys was identified by markers D7S2519 and D7S2442, respectively. This finding suggests a novel hypospadias locus (25 cM) at chromosome 7q32.2-q36.1 (Study I).

In a previous genome-wide scan for familial hypospadias, we have identified suggestive linkage in nine chromosomal regions. To extend this analysis new families and additional markers were included. The fine mapping analysis displayed an increased LOD score on chromosome 8q24.1 and 10p15 in altogether 82 families. From the chromosome region 10p15, we sequenced the *AKR1C3* and *KLF6* genes that have possible roles during male urethra development. Sequencing analysis showed one mutation (c.697A>G, p.A215T) in the *AKR1C3* gene and one mutation (c.496T>C, p.P166S) in the *KLF6* gene. In addition, three polymorphisms (rs3763676, rs12529 and rs7741) in the 5' end of the *AKR1C3* gene showed a significant association with hypospadias. These findings indicate that the *AKR1C3* and possibly also the *KLF6* genes function as genetic risk factors for hypospadias (Study II).

Several different pathways are implicated in the male genital development; of those the androgen pathway is crucially important. Therefore the androgen receptor (*AR*) and the 5 alpha reductase (*SRD5A2*) genes were sequenced in 38 isolated hypospadias cases. One mutation in the *AR* gene (p.Q798E) and another in the *SRD5A2* gene (p.K199S) was detected. Interestingly, also a high frequency of the rare leucine allele in the V89L polymorphism of the *SRD5A2* gene was found. The leucine allele has been shown to confer a decreased activity of enzyme, thus this allele may be a risk factor for hypospadias. This finding was confirmed in total 158 hypospadias cases compared to 96 controls (Study III).

The *MAMLD1* (*CXorf6*) gene is the first gene causing isolated hypospadias. The *MAMLD1* gene was sequenced in DNA from 97 sporadic hypospadias cases to elucidate the role of this gene in hypospadias. One new mutation, p.Q529K, that is predicted to affect the splicing process, was found in a boy with severe hypospadias. The variants, p.V432A and p.531ins3Q, have been reported previously and are indicated in this study as polymorphisms. Additionally, a significant association between the p.N589S polymorphism (rs2073043) that may alter the predicted protein structure and hypospadias was detected (p-value <0.05). The combination of two rare alleles of p.N589S and p.P286S (rs41313406) is over represented among cases. It may be that these two polymorphisms are inherited together as a haplotype (T-G), thus increasing the risk for hypospadias. These findings suggest that in a few cases hypospadias is caused by mutations in the *MAMLD1* gene, and that the polymorphism (p.N589S) is a genetic risk factor (Study IV).

LIST OF PUBLICATIONS

- I. Thai HT*, Söderhäll C*, Lagerstedt K, Omrani MD, Frisén L, Lundin J, Kockum I, Nordenskjöld A. **A new susceptibility locus for hypospadias on chromosome 7q32.2-q36.1.** Hum Genet. 2008 Sep;124(2):155-60. *Equal contribution
- II. Thai HT, Söderhäll C, Chen Y, Shulu Z, Frisén L, Lundin J, Kockum I, Nordenskjöld A. **Fine mapping of familial hypospadias identifies two new candidate genes: AKR1C3 and KLF6.** In manuscript
- III. Thai HT, Kalbasi M, Lagerstedt K, Frisén L, Kockum I, Nordenskjöld A. **The valine allele of the V89L polymorphism in the 5-alpha-reductase gene confers a reduced risk for hypospadias.** J Clin Endocrinol Metab. 2005 Dec;90(12):6695-8.
- IV. Thai HT*, Chen Y*, Lundin J, Lagerstedt K, Shengtian Z, Nordenskjöld A. **The p.N589S polymorphism in the MAMLD1-gene is associated with hypospadias.** Submitted. *Equal contribution.

CONTENTS

INTRODUCTION	1
COMPLEX DISEASE AND GENE MAPPING	1
Human genetic diseases	1
Complex diseases	2
Gene mapping in complex diseases	2
Linkage analysis	3
<i>Genetic linkage</i>	3
<i>Study design</i>	3
<i>Genetic markers</i>	4
<i>Mapping strategy</i>	6
<i>LOD score</i>	6
Association study	8
<i>Common applications</i>	8
<i>Basic method for association analysis</i>	10
Gene identification and characterization	11
HYPOSPADIAS	12
Male external genital embryology	12
Hypospadias definition	14
Epidemiology	15
Hypospadias is a complex disease	16
Environmental factors	16
Genetic factors	18
The molecular background of hypospadias	19
AIMS	21
MATERIALS AND METHODS	22
PATIENT MATERIALS	22
METHODS	23
Genotyping	23
Statistical analysis	24
Direct sequencing	25
Allele-specific PCR amplification	25
Taqman genotyping assay	26
Bioinformatics tools	27
RESULTS AND DISCUSSION	29
Study I: A new susceptibility locus for hypospadias on chromosome 7q32.2-q36.1	29
Study II: Fine mapping analysis in affected sib-pair study identifies two new candidate genes: the <i>AKRIC3</i> and <i>KLF6</i> gene	31
Study III: The Valine allele of the V89L polymorphism in the 5-alpha reductase gene confers a reduced risk for hypospadias	35
Study IV: The p.N589S polymorphism in the <i>MAMLD1</i> -gene is associated with hypospadias	37
CONCLUDING REMARKS	40
ACKNOWLEDGMENT	42
REFERENCES	45

LIST OF ABBREVIATIONS

AKR1C1-C4	aldo-ketoreductase family1, member C1-C4
AKR1D1	aldo-ketoreductase family1, member D1
AR	androgen receptor
ATF3	activating transcription factor 3
bp	base pair
cM	centi-Morgan
MAMLD1/ CXorf6	mastermind-like domain-containing protein 1/ chromosome X open reading frame 6
ddNTP	dideoxynucleotide
DNA	deoxyribonucleic acid
DHT	dihydrotestosterone
hCG	human chorionic gonadotropin
HSD17B3	hydroxysteroid (17-beta) dehydrogenase 3
IVF	<i>in-vitro</i> fertilization
KLF6	krueppel-like factor 6
LD	linkage disequilibrium
LOD	logarithm of the odds
PTN	pleiotrophin
RFLP	restriction fragment length polymorphism
SNP	single nucleotide polymorphism
SRD2A2	5-alpha reductase
SRY	sex determining region Y

INTRODUCTION

COMPLEX DISEASES AND GENE MAPPING

Human genetic diseases

The discovery of Mendel's laws of inheritance and their confirmation in 1900 opened a new era in human genetics. Mendel showed that certain features of an organism were determined by units of inheritance, later called genes, which were transmitted from one generation to the next with mathematical precision. By the 1960s, approximately 1500 different human genes inherited as Mendelian traits were recorded. This list has expanded dramatically with the development of powerful non-Mendelian strategies for identifying human genes. In the early 1990s, approximately 5500 genes had been described. However, with the sequencing of the entire human genome in 2001, the estimate of the number of human genes based on gene-prediction software was 25,000 to 35,000 (Venter et al, 2001). In 2003, 99.9% of the human genome was identified (The International Human Genome Sequencing Consortium, 2004).

Many studies have shown that, in contrast to earlier belief, the genetic background is of significance for many diseases (Hall et al, 1978; Carnevale et al, 1985; McCandless et al, 2004; Stevenson and Carey, 2004). Classically, genetic diseases are classified into three major categories: chromosomal, monogenic and multifactorial diseases. Chromosomal disorders are caused by errors in an entire chromosome or part of a chromosome. Since each chromosome contains thousands of genes, abnormalities on structure or quantity of chromosome often result in miscarriages, childhood deaths or severe phenotypes. Monogenic disorders are caused by the presence of disease alleles at one genetic locus. These disorders are inherited in a simple Mendelian fashion and are also referred to as Mendelian diseases. Therefore, they are characterized by a strong, usually one-to-one, relationship between defective gene and clinical diagnosis. More than 1800 human distinct disorders are now known or suspected to be monogenic diseases inherited in an autosomal dominant, autosomal recessive or X-linked fashion (Brinkman et al, 2006). However, some common diseases such as cardiovascular diseases, hypertension, diabetes, cancers and asthma that contribute much to the public health burden do not follow a Mendelian inheritance. Since these diseases are caused by both genetic and environmental factors, they are referred to as multifactorial

or complex disorders. To develop effective treatment and preventive measures, it is desirable to identify underlying genetic and environmental factors to understand the biology of these diseases.

Complex diseases

In genetics, the term “complex” usually refers to traits for which there is no single feature considered as disease-causing. Complex diseases are thought to involve multiple genes and environmental risk factors, as well as interaction between a gene and other genes and/or the environment. However, identification of the underlying causes is difficult for these diseases because each individual factor is likely to contribute only to a small amount toward the disease or the trait, and the risk factors also might vary among populations. Not only the distribution of these factors differs in different samples, but also the association between any one potential factor and the trait is likely to vary. Additionally, mechanisms of complex disease by which genetic factors influence the trait may deviate from our traditional understanding of genetic inheritance. Only a small proportion of diseases follow Mendelian patterns of inheritance but in most cases the relationship between genotype and phenotype is not one to one. In that sense, the work to determine genetic mechanism, even after a gene has been identified, can be difficult.

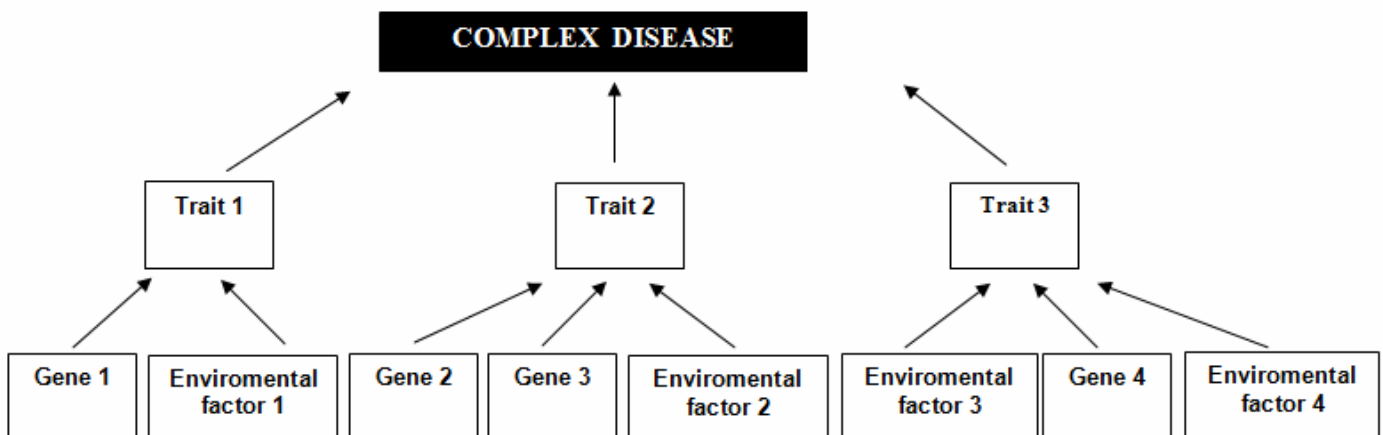


Figure 1: Genetic and environmental dissection of complex disease. Gene contributes singly or with other genes and environmental factors to the distribution of traits.

Gene mapping in complex diseases

Gene mapping aims to identify loci in the genome that are responsible for phenotypic variation and to identify which specific genetic variants cause the observed effect. This is a useful strategy because it

leads not only to identification of genetic variants but also to understanding the nature of variation and biological pathways that cause or influence disease. There are two main approaches to gene mapping: linkage mapping in pedigrees and linkage disequilibrium mapping in the studied population (association studies).

Linkage analysis

Genetic linkage

The principle of linkage analysis is based on the observed segregation of homologous chromosomes in families. In meiosis the maternal and paternal homologues of each chromosome align and recombine before segregation to one or the other daughter cell. It means that the recombined chromosome contains a mix of the two homologous chromosomes. Chiasma counts in human male meiosis show an average of 49 crossovers per cell (Morton et al, 1982). The basic of genetic linkage analysis is that recombination events between two genetic loci on the same chromosome occur at a rate related to the distance between them. Alleles at loci located close together on the same chromosome are always transmitted together all of the time and they are said to be completely linked. Otherwise, alleles at loci on different chromosomes or on the same chromosome but far apart are transmitted separately of each other and are separated in 50% of meioses. The probability of recombination to occur between loci is measured by the recombination fraction (θ) between them. When two loci are segregating independently, recombinant and non-recombinants are in equal proportion and $\theta=0.5$. Whereas, two loci are (tightly) linked and $\theta<0.5$ when recombinant and non-recombinants are not equal (Ott, 1999). This fraction is also used to calculate the genetic distance between any loci. In practice, $\theta = 1\%$ corresponds to a map distance of 1cM which approximately corresponds to 1Mb in physical distance.

Study design

To investigate the network of genetic background of complex diseases, a number of different approaches have been used. These studies include many steps to determine the evidence of genetic effect, identify the gene, measure the effect size of the gene, and then assess its functions (Terwilliger and Goring, 2000). There are three basic designs of ascertainment for a genetic analysis: extended families with multiple affected individuals, relative pair and single affected family members.

Traditionally, large extended families with multiple affected individuals were used to identify genes. It is reasoned that if a gene causes a disease and it is likely that the disease is inherited in the family, this will lead to many affected family members, who share the risk alleles from a common ancestor. A disease-causing or a susceptibility variant will co-segregate with nearby genetic markers. Thus, by identifying genetic markers that are shared by affected individuals in a given family, it will be possible to identify chromosomal regions that may harbour the disease gene. Using this principle, many causative genes in the human genome, mainly for Mendelian traits have been identified (McKusick, 1998). This approach reduces the complexity of disease since the environment within the family is more similar than in random samples. The disadvantage is that such large families with multiple affected individuals are difficult to find and the identified genes tend to be rare so that they explain only a small and unique proportion of diseases. To avoid this problem, a larger number of smaller families, such as affected sibling pairs or parents and affected children can be recruited in relative pair studies. Since materials of these studies come from different families, the number of genetic variants in affected individuals as well as the number of environmental risk factors will increase. The principle of relative pair study, typically affected sibling design, is based upon the proportion of allele's identical-by-descent (IBD) that the sibs share. On average, sibs share 50% of their genes IBD at a particular location in the genome. If affected sib-pair share 0, 1 or 2 alleles from parents, the proportion of alleles shared IBD at and around the susceptibility locus will be 25%, 50% and 25% (Penrose, 1935). The most important characteristic of the affected sibling method is that if a number of different rare variants present at one or multiple disease susceptibility loci in the same genome region, then linkage can detect this region because excess IBD sharing is not with respect to a particular allele in the population. The last type of design, single affected family member approach uses samples that are collected from cases and both their parents, discordant sib pairs, where the siblings do not have the trait in question, or cases- controls. A disadvantage of this approach is that it is too difficult to perform on traditional linkage analysis since multiple affected individuals are needed to determine IBD sharing. This approach is typical for association study.

Genetic markers

Linkage refers to the mapping of a polymorphism or mutation at a genetic locus through the analysis of chromosomal segments transmitted to individuals with some known degree of relationship (Ott, 1991; Terwilliger and Ott, 1994). The principle for linkage mapping in humans is the use of known polymorphic DNA variants which are called *markers* (Botstein *et al*, 1980) to detect the correlated

inheritance of a particular trait with that of closely linked marker loci. To be useful in genetic mapping, markers must be polymorphic and their chromosomal location must be known.

Over the years, a variety of different genetic markers have been used for mapping purposes. In 1982, restriction fragment length polymorphisms (RFLPs) became the first modern genotyping markers to be used in a successful linkage (Gusella et al, 1983). Restriction enzymes are proteins that cleave at, or very close to, specific recognition sequences within DNA. These restriction sites are present in some people and absent in others, causing the DNA to be cut into different-sized fragments in different people for each respective restriction enzyme. The RFLPs commonly have two possible alleles, giving to only three possible combinations and therefore they are often not very informative. In addition, major disadvantage of this technique is that a large amount of DNA required and that it is time consuming. The inconvenience of using RFLPs was deciphered by the identification of the variable number of tandem repeats (VNTRs) (Nakamura et al, 1987). These are short, tandem repeated DNA sequences, where the number of repeats varies between individuals. The size of different repeated units varies from two up to ten nucleotides. Repeats of more than four nucleotides are called minisatellites and those of two to four nucleotides are called microsatellites (Litt and Luty, 1989; Weber and May, 1989). These repeats frequently vary in length in different individuals and therefore they can have high levels of heterozygosity. They rapidly replaced RFLPs in gene mapping studies due to these features, their easiness of typing and the small amount of template DNA required. Microsatellite markers have been used successfully in linkage analysis to detect susceptibility regions to many diseases (Chung et al, 1993; Mashfield et al, 1997; Scott et al, 2003; Moreno et al, 2004). It has been estimated that the human genome has 5000 – 10 000 such microsatellite repeats. Sources of information on microsatellite markers can be found in different databases such as Marshfield Clinic's Center for Medical Genetics (<http://research.marshfieldclinic.org/genetics>), DeCODE Genetics (<http://www.decode.com>), Ensembl Genome Browser (<http://www.ensembl.org/index.html>), Généthon (<http://www.genethon.fr/>) and National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

With development of modern molecular technologies, nowadays, single-nucleotide polymorphisms (SNPs) are used in gene mapping. A SNP represents a marker where a nucleotide differs at a specific place in the genome. SNPs are the most common type of human DNA sequence variation, with frequencies greater than 1%, occurring on average 1 per 500 to 1000 bp on a randomly selected

chromosome (Venter et al, 2001). Because these polymorphisms are biallelic, they are less informative than microsatellites. However, SNP assay systems are generally straightforward, inexpensive, and adaptable to automation.

Mapping strategy

The model for global search of human genome by linkage analysis contains two stages: a relatively scattered marker map covering the whole genome is used in the first step stage to detect linkage regions, followed by a second stage with a denser marker map in these regions. The most important step on the first stage, the genome wide scan, is selecting a marker set for linkage analysis.

The human nuclear genome consists of approximately 3600 cM in genetic distance (Kong et al, 2002); consequently to cover the genome with 10-cM resolution requires 360 genetic markers that are fully informative. Moreover, the basis of appropriate markers is: conveniently located in the gene-rich regions, easy to read and interpret, and highly polymorphic. Hence the general approach to genome wide scan is to perform a genotyping at approximately 10-cM density on the order of around 400 markers on familial samples. Following statistical analysis, potential regions of linkage are identified on various chromosomes. Each of these is subjected to genotyping of additional microsatellite markers with increased density.

Recently, mapping panels of single-nucleotide polymorphisms (SNPs) have been used for linkage mapping (Sellick et al, 2004). These sets are designed for family-based linkage analysis, and initially contained approximately 10 000 SNPs, roughly equivalent to a 5-cM microsatellite genome scan. A single microsatellite marker is often considered to have equivalent information content to 3–4 SNPs. For fine-mapping equivalent to microsatellites at about 1-cM resolution, hundreds of thousands to millions of potential SNPs are available in public databases (Sachidanandam et al, 2001; Holden, 2002). Because of the limitation of informativeness of SNPs, microsatellites probably will continue to have a useful role in fine-mapping stage in linkage studies with large susceptibility regions.

LOD score

In simple cases, the recombination fraction (θ) can be obtained by calculating the proportion of recombinants and non-recombinants, but it is often not possible to count this proportion directly. Since 1955, the likelihood-based methods have been usually used in linkage analysis (Morton, 1955). This method calculated a LOD score (logarithm of the odds), which is the logarithm of the ratio between the

probability that the two alleles are linked (H_A , $\theta > 0.5$) and the probability that they are not linked (H_0 , $\theta = 0.5$).

$$Z(\theta) = \log_{10} [L(\theta)/L(1/2)]$$

This test is usually performed by calculating the LOD score (Z) for all values of θ between 0 and 0.5, and determining the recombination fraction at which the LOD score is maximal. A LOD score of 3 or greater is taken as significant evidence that two loci are linked, for the reason that the linkage probability is 10^3 times more likely than the null-probability that the two loci are not linked. A LOD score below -2 are accepted as evidence that linkage has been excluded. Values between -2 and 3 are inconclusive and indicate that more samples or additional markers close by are needed. The most precious benefit of the LOD score method is that linkage result from different pedigrees can be calculated together.

In parametric or model-based linkage analysis, likelihood computations are simplified by assuming that the models describing both inherited trait and marker loci are known without error. This method requires certain parameters in a genetic model, such as the allele frequencies, dominance relationships among alleles, and the association between genotype and phenotype. If these parameters are correct, parametric linkage analysis is a powerful method (Amos and Williamson, 1993). Since the mode of inheritance of markers is codominant and relationship between genotype and phenotype is straightforward, the trait of disease is similar to Mendelian laws. However, if either the trait or marker locus models are misspecified in the analysis, then the power of parametric linkage is decreased (Williamson and Amos, 1990). The result of linkage analysis can be misleading, mainly because of incomplete penetrance, phenocopy, and misdiagnosis on phenotype ("affected" cases are not susceptibility gene carrier, whereas "unaffected" ones carry the susceptibility gene). Furthermore, the genetic model for the trait may also be misspecified by assuming a single trait locus when in fact multiple loci exist, such as in complex diseases. The misspecification of the trait model in parametric linkage analysis has motivated the utilization of a new method that relies less completely on the genetic model specification. Unlike parametric linkage analysis, nonparametric or model-free method uses allele-sharing approach (Lander and Schork, 1994). If there is a linkage between a disease locus and genetic marker, affected relative pairs should share same alleles identical by descent (IBD) that is inherited from a common ancestor. In comparison, two unrelated individuals can have the same genotype but share no alleles IBD. For these reasons, in nonparametric linkage analysis there is no need to specify a mode of inheritance for the trait being linked

to markers. Since the problem of trying multiple models is avoided in this approach, this analysis is more robust than the parametric method. This method can be used in several situations, such as for pedigrees with only affected sib-pair or families with more available relatives than sibling, but it is especially effective in a sib-pair study.

Both parametric and nonparametric analyses are used to calculate a LOD score between a trait locus and a single marker locus (two-point linkage) or between a trait locus and a map of multiple markers (multipoint linkage). In general, two-point linkage is less powerful than multipoint linkage (Lathrop et al , 1985) since multipoint linkage combines the information of the transmission of a haplotype of alleles at nearby loci to better track the transmission of alleles from parents to children at each location along a chromosome. However, multipoint linkage requires the order and distances between markers (Ott, 1999). Misspecification of the marker map order or large distance between markers will also decrease power to detect linkage in both parametric and nonparametric multipoint linkage (Ott and Lathrop, 1987; Ott, 1999). Therefore, both two-point and multipoint linkage analyses are performed in gene mapping. When linkage signal is found through two-point analysis, multipoint analysis is applied to maximize linkage information and to localize the susceptibility regions.

Association study

In contrast to linkage studies, association studies look for a correlation between a specific variant and disease status or quantitative trait in the population. Simply, the essential principle of this approach is that: if a DNA variant increases disease susceptibility, it is expected to be more frequent among those who are affected than those who are unaffected. The aim of association studies is to search for a significantly increased or decreased frequency of a marker allele, genotype, or haplotype with a disease trait than would be expected by chance if there is no association. It can be explained by the direct involvement of biological action of genetic variants with diseases or by linkage disequilibrium (LD) between markers and susceptibility genes. But one has to remember that association is a statistical rather than causal concept since it detects risk factors of diseases. Unlike linkage study, the conclusion for association of variant and disease is defined at the level of the population rather than that of the individual. Moreover, an association study requires a dense spacing of markers and therefore SNPs are the suitable markers of choice, since they occur on average once each 1000 base pairs, they are tractable to high-throughput scoring.

Common applications

Two main approaches have been applied in genetic research to detect and qualify an association between DNA variants and disease status in the population: gene mapping via linkage disequilibrium (LD) and candidate gene analyses for biological relevance.

In a population, a haplotype that contains linked loci is transmitted from one generation to the next but the length of inherited segment is reduced by recombination over generations. Many parts of the human genome exhibit these segments, called block structures (Reich et al, 2001; Crawford et al, 2004; McVean et al, 2004). If common DNA variants influence the risk of a disease, most affected individuals in the population may have inherited the mutation from ancestors in whom the mutation originated and shared the ancient haplotype with the next generation. It means that affected individuals carry the same disease-causing haplotypes that originally were transmitted from a common ancestor, even though they are not obvious family members. Whereas linkage analysis relies on recombination events in one or two generations, an association study concerns accumulated events over generations. Based on a relationship between linkage disequilibrium and the rate of recombination, LD mapping is used to estimate the location of a disease locus relative to a genetic marker (Hastbacka et al, 1992; Hill and Weir, 1994; Kaplan et al, 1995).

The whole-genome association (WGA) mapping study requires most of the known common variants in the entire genome, from several hundred thousand to several million SNPs, to be genotyped. Many studies suggest that the human genome is organized into haplotype blocks that show high LD, interspersed with shorter regions of high recombination and consequently low LD (Gabriel et al, 2002; Ardlie et al, 2002). Common haplotypes can represent most of the genetic variation across relatively large regions of the genome. These haplotypes can be genotyped by using a small number of SNPs that are tagged to untyped SNPs by virtue of the strong LD between them in a population of interest. As a result, it is possible to genotype only a subset of common tagging SNPs, instead of genotyping all common variants (Gabriel et al, 2002; Johnson et al, 2001; Carlson et al, 2004). In regional mapping approach, broad regions of the genome that are believed to contain the disease gene are examined for association. These regions are considered as susceptibility regions either because of known chromosomal abnormalities or from some earlier linkage analyses. Therefore LD mapping is regarded as a fine-mapping step for linkage study in order to eventually identify candidate genes. To fine map a region of positive linkage, random or tagging SNPs in an LD-based screening for signals of association or

functionally important SNPs of the most promising candidate genes in this region must be examined. This approach has a great power if all SNPs in the positive genes in a susceptibility region are analyzed.

Another application of association study is candidate gene analysis. The aim of this application is to check the hypothesis that DNA variants in certain genes affect disease susceptibility. These variants might have an effect on activity or expression level of the gene products or anonymous markers that are in LD with the variants of candidate genes. Moreover, association studies can be used to define the allele-specific risk such as the relationship between gene variants with environmental factors, age of onset, as well as the interactions between these variants and different variants in other genes. This approach is also used to characterize discovered disease genes by replicating studies in different populations to confirm the finding and identify a true causative variant.

Basic method for association analysis

In association studies, associations between marker alleles and disease phenotypes can be detected. There are two approaches: case-control and family-based test. A case-control study begins with the collection of affected individuals as well as unaffected individuals from the same population, and then allele or genotype frequencies in the two groups are compared. The later approach, family-based tests of association, uses allele frequencies in affected individuals and compares to family-based controls, typically parental controls or unaffected siblings.

One important step prior to association analysis is to test that genotyped markers are in Hardy–Weinberg equilibrium (HWE) in the control population, since HWE describes the expected relationship between allele and genotype frequencies under normal population conditions. For a case-control study, the evidence of association between a disease and alleles of a single SNP can be evaluated, in the simplest case, by comparing genotype or allele frequencies between cases and controls with a 2x2 chi-squared contingency table test with the null hypothesis “no association between case-control status and allele frequency”. In this analysis, the number of each allele type found in cases and controls is compared to its expected number. When a genotyped variable has more than two alleles, such as for the three genotypes at any locus in a diploid genome, a 2x3 contingency table is used.

Moreover, to check an association of haplotypes from multiple SNPs, “global” test is used. In this test, the null hypothesis is that the distribution of haplotype frequencies is the same in cases and controls, or

the frequency of a specific haplotype versus all other haplotypes considered together is the same in cases and controls.

Gene identification and characterization

The candidate-gene approach defined to study the genetic influences on a complex trait follows these steps (Tabor et al, 2002):

- identifying candidate genes that might have a role in the aetiology of the disease
- identifying variants in or near those genes that might either cause a change in the protein, its expression or function.
- genotyping these variants in a population
- using statistical methods to determine whether there is a correlation between those variants and the phenotype.

Once susceptibility regions of interest have been identified, the important task is to generate a list of genes that are located within these regions.

Genome browse such as Ensembl (<http://www.emsembl.org>), UCSC (<http://genome.ucsc.edu>) or the NCBI (<http://www.ncbi.nlm.nih.gov/mapview/>) are used to provide a graphical interface at the regions. Commonly, there are several hundred genes under a single peak. For that reason, a large scale gene-screening approach could become money and labour consuming. A traditional approach to find candidate genes is to assess genes that are involved in the developmental pathway for the disease and/or have expression that is significantly altered in affected tissue compared to normal tissue. This is the most difficult step since candidate-gene selection is based on the ability to predict functional candidate genes and in some cases, and current knowledge is insufficient to make these predictions.

There are numerous methods to detect causal genetic variants. Ultimately, several different physicochemical techniques such as denaturing HPLC, single stranded sequence polymorphism (SSCP) have been developed, but the golden standard remains direct DNA sequencing (Kristensen et al, 2001). Once a potential variant has been discovered, next step need to be taken to confirm that this variant results in obvious changes in gene function. Moreover, this sequence change is generally well accepted as causal for rare phenotypes especially when they are absent in a large number of random control individuals matched for ethnic background where necessary. Sequence alterations such as frameshift,

stop codons or missense mutations can cause changes in conserved or biochemically validated amino acid residues, or changes in conserved splice junction elements, etc. Studies in recent years have shown that a wide variety of diseases may be caused by mutations that affect RNA processing, as opposed to mutations that alter the protein-encoding sequence of the gene (D'souza et al, 1999; Percy, 2000; Pagani et al, 2003). Some missense and synonymous variants may occur in cis-acting elements that regulate splicing, resulting in exon skipping, and inefficient splicing of introns or usage of cryptic splice sites. In addition, mRNA stability may also be affected by variants in the intronic or exonic splicing enhancer or silencer, 3'UTR and 5' UTR (Cazzola and Skoda, 2000).

Another approach to identify a candidate gene is comparative genomics strategy that includes comparative functional genomics strategy and comparative structural genomics strategy. Genes are considered as candidate genes if they may be functionally conserved or structurally homologous genes between different species, and mutations in these gene cause same phenotype of interest (Zhu et al, 2007). Recently, a new method has been developed for candidate gene identification which is a computer facilitated candidate gene approach. This method objectively extracts, filters, assembles, and analyzes all possible resources available derived from the public web databases to prioritize the most likely candidate genes through a variety of web resource-based data sets. Application softwares or online tools such as GFSST (<http://gfsst.nci.nih.gov>), GENESEEKER (<http://www.cmbi.ru.nl/GeneSeeker/>), ENDEAVOUR (<http://www.esat.kuleuven.be/endeavour>), and POCUS (http://www.hgu.mrc.ac.uk/Users/Colin.Semple/Semple_Lab/Home.html) have been developed and released to the public.

HYPOSPADIAS

Male external genital embryology

The normal development of external genitalia in male occurs between the 8th and the 16th week of gestation under the influence of androgens secreted by the fetal testes. The genital tubercle elongates to form the shaft and glands of the penis. During this elongation, the genital folds form the lateral walls of the urethral groove. This groove extends along the caudal aspect of the elongated phallus and forms the urethral plate. In the 12th gestation week, the medial edges of the ectoderm urethral fold fuse to form the

penile urethra. The urethral canal does not reach the most distal part until the 16th week of gestation, when ectodermal cells from the tip of glands penetrate inward and form a short epithelial cord. This cord later obtains a lumen, thus forming the external urethral meatus. The formation of a complete prepuce with its final cutaneous fold surrounding the glans ends of this process (Figure 2).

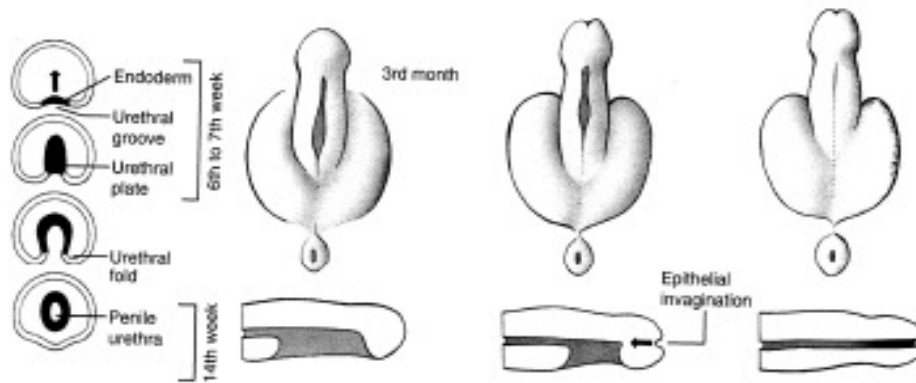


Figure 2: Formation of the external genitalia in male from the 6th week to the 14th week of gestation. (Adapted from *Human Embryology, Larsen, 1998*)

Normal sex development is a complex process that involves many genes. In brief, sex determination and differentiation consist of three sequential stages. Firstly, the genetic sex determination by the sex chromosome constitution in the zygote at the time of conception. Secondary, the genetic information determines whether an undifferentiated gonad differentiates into either a testis or an ovary. The *SRY* gene is the master gene for testes development and acts directly on the gonadal ridge and indirectly on the mesonephric ducts. The *SRY* protein is the testis-determining factor, since under its influence male development occurs, but in its absence female development is established. Lastly, phenotypic sex results in male or female differentiation under the control of hormones.

The formation of male genitalia needs the action of a specific androgen: testosterone (T). Placenta human chorionic gonadotropin (hCG) stimulates Leydig cells in fetal testes to produce testosterone. In Leydig cells, 17beta-hydroxysteroid dehydrogenase (*HSD17B3*) catalyzes the conversion of androstenedione to testosterone. Testosterone is converted to the more active dihydrotestosterone (DHT) by the enzyme steroid 5 α - reductase (*SRD5A2*). Both hormones (T and DHT) bind to the androgen receptor (AR) in the genital target tissue. Testosterone induces the differentiation of Wolffian duct into epididymis, vas deferens and seminal vesicles whereas dihydrotestosterone, bound to the same androgen receptor,

modulates the differentiation of the prostate gland, penis and scrotum (Figure 3). Premature arrest during the fusion of the urethral folds leads to different severity of hypospadias.

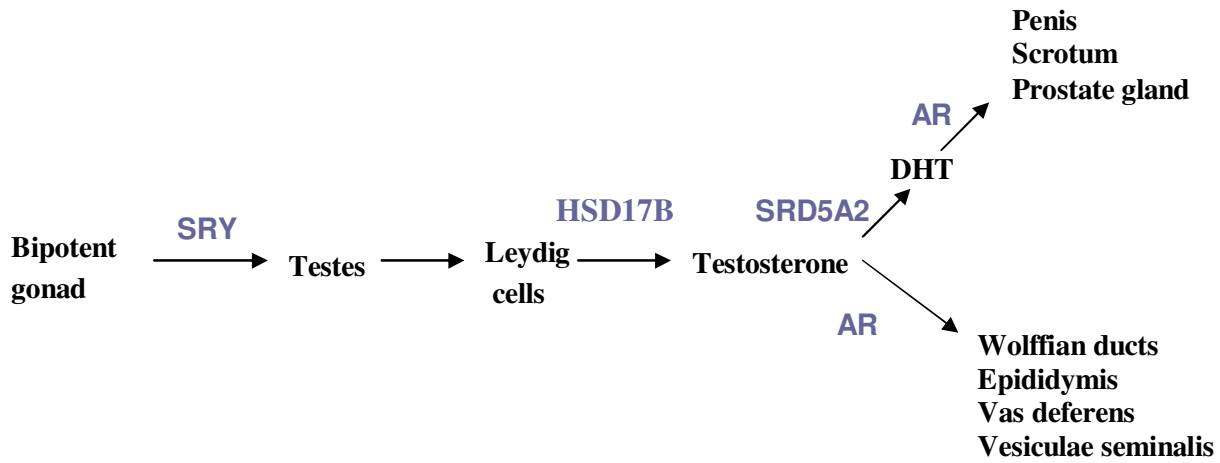


Figure 3: Genes involved in the development of male genitalia.

Hypospadias definition

Hypospadias is a malformation of the genital tract, characterized by incomplete fusion of the urethral folds, abnormal opening of urethra and different degrees of curvature of the penis.

Hypospadias classification	Location of urethral meatus
Anterior (60-70% cases)	
- Glandular	On the underside of the glands
- Coronal	At the level of the glandular- preputial ridge
- Distal penile	On the anterior third of the shaft of the penis
Middle (20% cases)	
- Middle penile	On the middle third of the shaft of the penis
Posterior (10% cases)	
- Proximal penile	On the posterior third of the shaft of the penis
- Penoscrotal	At the junction of the penis and scrotum
- Scrotal	At the level of the scrotum
- Perineal	At the level of the perineum

Table 1: Classification of hypospadias based on the location of urethral meatus

Although, hypospadias appears most often in isolated manner, other malformations like cryptorchidism, bifid scrotum and micropenis are associated with the condition, especially when the degree of severity of hypospadias is pronounced. Sexual ambiguity is associated with the most severe cases (Stokowski et al, 2004). Hypospadias is classified in different variants based on the position of the urethral meatus (Table 1).

Surgical treatment of hypospadias is performed to construct a normal looking penis and to reconstruct the urethra and position the meatus at the penile tip that enable the patient voiding while standing. The ideal age for surgical correction is considered before one or two years of age. Prenatal ultrasonography can detect moderate and severe cases but it is not usually indicated for anterior and middle hypospadias (Sides et al, 1996; Meizner et al, 2002).



A: Glandular



B: Penile



C: Penoscrotal



D: Scrotal



E: Perineal

Figure 4: Hypospadias. A-E, the severity and morphology of hypospadias depend on the location of meatus and degree of chordee.

Epidemiology

Hypospadias is a common congenital malformation with frequency ranging from 0.4% to 0.8% of male live births in Europe, Asian and North America (Hussain et al, 2002; Silver et al, 2000; Chong et al, 2006). A higher incidence of hypospadias has uniformly been reported in Caucasians than in other ethnicities (Gallentine, 2001). In Sweden, the incidence of hypospadias is 1.14 boys per 300 male live births according to the annual Swedish Malformation Registry (Kallén and Winberg, 1982; Paulozzi, 1999). The difference in prevalence rate between countries could be explained either by variations in case definition and incomplete ascertainment, or different genetic or environmental factors.

However, a series of scientific publications have reported an increasing incidence of hypospadias in several countries. Based upon the analysis of American survey of congenital malformations, Paulozzi et al observed that the rate of hypospadias almost doubled between 1970 and 1990 in the United States (Paulozzi, 1997). The same increase of hypospadias cases has also been indicated in some European countries including Norway, Denmark, Italy and France (Paulozzi, 1999). The reason for the rising incidence remains uncertain but the high frequency of endocrine disruption in industrialized countries is proposed as a possible explanation for this increase (Dolk et al, 1998; Wakefield et al, 2001).

Hypospadias is a complex disease

Since both environmental factors and genetic factors have a strong influence on the development of hypospadias, this malformation is considered as a complex disease.

Environmental factors

A relationship between hypospadias and low birth weight has been found in several studies (Källén and Winberg, 1982; Calzolari et al, 1986; Akre et al, 1999, Weidner, 1999). In a group of discordant monozygotic male twins, the incidence of the smaller twin having hypospadias is high comparing to concordance monozygotic male twins (Fredell et al, 1998). Moreover, the birth weight of non-twin sibling without hypospadias is significantly higher than that of proband with hypospadias (Fredell et al, 2002). Up to a ten-fold increase of hypospadias has been reported in infants small for gestational age (Gatti et al, 2001). The decrease in some growth parameters including birth weight and duration of

gestation in hypospadias infants suggested that the primary cause occurred early, during the first trimester of pregnancy. Therefore, the poor intrauterine growth is indicated as risk factor for hypospadias (Hussain et al, 2002). However, it is unclear whether growth retardation in itself has an impact on the formation of the urethra, or if other environmental factors that influence the intrauterine growth and morphogenesis of the urogenital tract are considered to be the cause. One hypothesis is that disturbance of placental function early in pregnancy is a key mechanism underlying both low birth weight and the improper closure of the urethra, since the placenta is involved in the differentiation and development of the fetal organs in this period (Källén 1988; Akre et al. 1999; Weidner et al. 1999; Hussain et al. 2002; Aschim et al. 2004; Boisen et al. 2005). Placental disorders such as infarction, hemangioma, and membranous placenta would reduce the functional volume of the placenta that may then have limited capacity for both hCG production and nutrition of the fetus. As discussed in more detail above, placental hCG plays an important role on the morphogenesis of the male external genitalia, including urethral development in early gestation. Therefore, reduced levels of human chorionic gonadotropin (hCG) have been suggested as a candidate in the aetiology of hypospadias (Czeizel et al. 1979). Moreover, iron supplementation in mothers immediately prior to conception and/ or during the first trimester of gestation has been suggested as a risk factor for hypospadias (North et al, 2000; Brouwers et al, 2006). A possible explanation is that this supplementation may increase blood viscosity, which subsequently impairs placenta blood flow and then results in malfunction of the placenta, leading to both low birth weight and hypospadias.

Another risk factor for hypospadias is parental subfertility (Sweet et al, 1974; Czeizel et al, 1985; Brouwers et al, 2006). Fathers with signs of subfertility such as decreased sperm density, motility, and morphology have a four- fold increased risk of giving birth to boys with hypospadias (Fritz and Czeizel, 1996). Because hypospadias and male subfertility may share the same embryonic origin with genetic and environmental components, the affected fathers may transmit a certain predisposition to their sons (Skakkebaek et al, 2001). Additional evidence that parental subfertility increases the risk for hypospadias has come from the report of a high incidence of hypospadias in boys whose parents had undergone fertility treatment. A five- fold increased risk for hypospadias has been found in infants conceived by in vitro fertilization (IVF) procedures in the United States (Silver et al, 1999). Although no increased risk for hypospadias was found after standard IVF in Sweden, intracytoplasmic sperm injection (ICSI), another specific IVF technique, resulted in a three-fold increased risk of hypospadias in boys (Wennerholm et al, 2000; Ericson and Källén, 2001). A possible explanation is that hormone

administration as part of pregnancy support interferes with androgen production in early gestation and thereby disturbs normal genital development.

Although the maternal use of oral contraceptive in early pregnancy is not associated with hypospadias (Källén et al, 1991; Carmichael et al, 2005), the role of other endocrine factors in the etiology of hypospadias is still questioned. Nowadays, endocrine disruptors such as pesticides, fungicides, industrial chemical products, detergents and material for the fabrication of plastics are commonly used. Natural substances from vegetables having similar properties as phyto-estrogen have also been classified as potential endocrine disruptors (Santti et al, 1998). Experimental studies have found an influence from endocrine disruptors on the development of the male genital tract. Administration of potentially anti-androgenic substances, finasteride (an inhibitor of 5- α -reductase type 2) and flutamide (an inhibitor of testosterone fixation on receptors) can induce hypospadias in rodents, rabbits and mice (Kurzrock et al, 2000; Kojima et al, 2002). Klip et al (2002) showed that boys born to mothers who had been exposed to diethylstilbestrol (DES) in utero had a higher risk of hypospadias. A five-fold increased risk of hypospadias has also been reported in boys whose mothers followed a strict vegetarian diet during pregnancy (North and Golding, 2000), even though the association between maternal dietary and hypospadias is still questioned (Pierik et al, 2004)

Genetic factors

Several observations suggest that hypospadias is influenced by genetic factors. Familiar clustering for hypospadias has been reported in 4% to 28% of cases (Sizzle et al, 1979; Stoll et al, 1990; Fredell et al, 2002). About 7% to 9% of the fathers of hypospadias boys have the same malformation (Bauer et al, 1979, Stokowski et al, 2004), whereas 5.4% of hypospadias boys have at least one other affected relative. The recurrence risk for a brother of an affected child also is 17% (Stoll et al, 1990). The more severe malformation of the index patient, the higher the incidence of hypospadias in the next male siblings. With the first degree of hypospadias, 3.5% have an affected brother. Whereas 10% to 19% hypospadias is inherited in the next brother if the index boy has a second or third degree of hypospadias (Bauer et al, 1981). An increased risk for hypospadias among twins has also been reported in some studies (Cheng et al, 1971, Roberts et al, 1973). Furthermore, the prevalence of hypospadias is higher among member of male-male twins and lower among male in male-female twins (Källén et al, 1986).

In addition, some families have been described showing a Mendelian pattern of inheritance. Frydman reported the autosomal recessive mode of inheritance in a large consanguineous family with eight hypospadias members (Frydman et al, 1985). An autosomal dominant characteristic also has been suggested in some families that have affected males in at least two generations (Lowry et al, 1976; Page et al, 1979; Frisé et al, 2003). Altogether, hypospadias is mainly due to monogenic inheritance in only a small portion of affected cases (Fredell et al, 2002). Heritability for hypospadias with a multifactorial model is reported to range between 0.57 and 0.74 (Monteleone et al, 1981; Stoll et al, 1990). In a complex segregation analysis of 2005 hypospadias pedigrees, the heritability of 0.99 was obtained, thus confirming a previous study of 103 families (Harris and Beaty, 1993; Fredell et al, 2002). These findings indicated that there is autosomal dominant trait of inheritance in some families but usually a complex mode.

Hypospadias is also a feature in more than one hundred genetic syndromes (www.ncbi.nlm.nih.gov/omim) such as hand-foot-genital, Silver-Russell or Klinefelter syndromes and is part of ambiguous genitalia caused by hermaphroditism or gonadal dysgenesis, for example.

The molecular background of hypospadias

The exact molecular mechanisms that predispose for hypospadias remain largely unknown. Mapping hypospadias candidate genes has, until recently, little success. Genome-wide linkage gene mapping with sib-pair analysis has been performed and suggests linkage on 9 chromosomal regions: 1q23, 2p11, 6q25, 8p21-22, 8q24.1, 9q21-22, 10p15, 10q21, 18q21 (Frisé et al, 2004). Since each region contains hundreds of genes, this finding provides a basis for only outlining the complex genetic background of hypospadias and further studies concerning candidate genes on these regions need to be performed.

Recent studies of hypospadias have revealed a possible involvement of genes in the earlier genital tubercle development such as sonic hedgehog (*SHH*), fibroblast growth factors 8 and 10 (*FGF8* and *FGF10*), and homeobox A13 and D13 (*HOXA13* and *HOXD13*) as well as ephrin (*Eph*) (Cohn and Bright, 1999; Haraguchi et al, 2000; Perriton et al, 2002; Frisé et al, 2003; Morgan et al, 2003; Beleza et al, 2007a; Yucel et al, 2007). Nevertheless several different pathways are implicated in male genital development, but the main pathway driven by testosterone is remarkably important. Since this hormone plays an important role on the development of hypospadias, defects anywhere along the pathway of

androgen may interfere with proper functioning of testosterone. Mutations in the *AR*, *SRD5A2* and *HSD17B3* genes can therefore account for a higher risk for developing hypospadias (Thigpen et al, 1992; Batch et al, 1993; Geissler WM 1994; Allera et al, 1995; Boehmer et al 1999; Nordenskjöld et al, 1999). However, mutations in these genes are usually found in severe cases with other malformations such as micropenis, cryptorchidism and bifid scrotum, and only a few isolated hypospadias cases carrying mutations in the *AR* gene and the *SRD5A2* gene has been reported (Hiort et al, 1994; Allera et al, 1995; Silver and Russell, 1999). Moreover, studies on the functions of estrogen and estrogen receptor during male reproductive tract development as well as the balance between estrogen and androgen have shown that estrogens are also important in the etiology of hypospadias (Kim et al, 2004; Liu et al, 2005; Beleza et al, 2007b; Wang et al, 2007; Watanabe et al, 2007). Mutations in the Activating Transcription Factor 3 (*ATF-3*) gene, one estrogen responsive gene, have been detected in some isolated hypospadias cases (Beleza et al, 2008, Kalfa et al, 2008).

The Mastermind-like domain containing 1 (*MAMLD1*) gene, previously known as the Chromosome X open reading frame 6 (*CXorf6*) gene, has however been identified as the first gene for isolated hypospadias (Fukami et al, 2006). *In-situ* hybridization (ISH) analysis showed that *MAMLD1* mRNA is expressed in Müllerian ducts, forebrain, somite neural tube, and pancreas and especially, in fetal Sertoli and Leydig cells around the critical period for sex development. Transient knockdown of *Mamld1* results in significantly reduced testosterone production in mouse Leydig tumor cells (Ogata et al, 2008). Furthermore, this gene is co-expressed with steroidogenic factor (SF-1), which regulates the transcription of genes involved in sex development (Ogata et al, 2008; Sadovsky and Dorn, 2000). These findings indicate that *MAMLD1* seems to exhibit the sex development via the androgen metabolism. Subsequently, DNA sequencing analyses have shown mutations of the *MAMLD1* gene in hypospadias cases with a large range of phenotypes, varying from mild to severe hypospadias (Fukami et al, 2006; Kalfa et al, 2008).

AIMS

The aim of this thesis is to identify susceptibility genes that contribute to the pathogenesis of hypospadias. The specific aims of individual projects are:

- To identify predisposing loci for hypospadias in a large family (paper I)

- To fine-map and identify candidate genes for hypospadias in affected sib-pairs (paper II)

- To determine the contribution of genetic variants in candidate genes for hypospadias (paper III and IV)

MATERIAL AND METHODS

PATIENT MATERIAL

Blood samples from boys with non-syndromic hypospadias of different severity, both familial and sporadic cases, were collected. Of those, the family material is part of an on-going large project collected for a genetic study of hypospadias (Fredell et al, 1998).

In study I, one family including eight affected boys with autosomal dominant inheritance of hypospadias was recruited for genome wide linkage scan (Figure 5). Blood was collected from 15 available family members, including seven affected boys and eight unaffected family members. The affected boys had coronar hypospadias only. In addition to hypospadias, member III.3 was born with heart malformation (bicuspid tricuspidalis valve) while III.6 was born with heart malformation (left heart hypoplasia) but without hypospadias. These two boys were excluded in the genetic analysis due to lack of available tissue. Hence, all individuals included in the study as affected had isolated hypospadias.

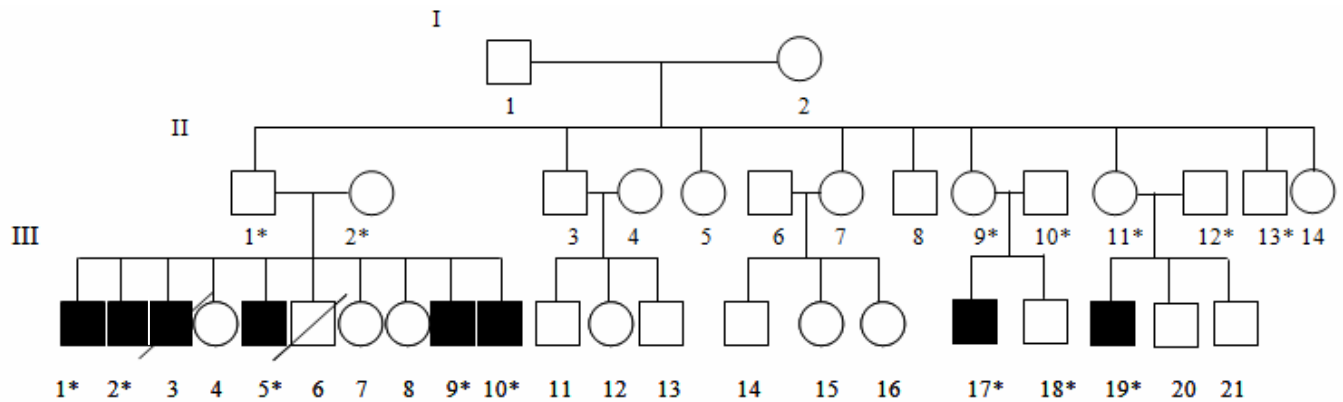


Figure 5: Pedigree of a large family with hypospadias. Affected individuals are shown with black squares while normal individuals are shown with white squares and circles. * indicates members from whom we obtained DNA.

In study II, families with at least two affected siblings were selected for sip-pair analysis. In the previous study, blood samples from 69 families with at least 2 affected members were collected and used for a

genome-wide linkage scan (Frisén et al, 2004). For fine mapping analysis, we recruited 13 additional families including 11 Swedish families (46 individuals) and 2 Iranian families (8 individuals). Altogether 363 individuals were genotyped, of which 180 were affected. The distribution of the degree of relationship of the entire 96 affected relative-pairs is listed in Table 2.

	Total	Swedish	Middle East
Affected sib pairs	65 (9)	51 (7)	14 (2)
Affected half-sib pairs	4 (1)	2 (1)	2
Proband with affected relative (degree)			
<i>Second</i>	6 (1)	6 (1)	0
<i>Third</i>	17 (2)	16 (2)	1
<i>Fourth</i>	4	4	0

Table 2: Degree of relationship in the affected relative pairs (whereof added pairs for fine mapping analysis)

In study III and IV, samples from sporadic cases recruited through medical records in Sweden have been used for sequencing analysis. As a control group, blood samples from healthy voluntary anonymous blood donors at Karolinska University Hospital were collected.

Genomic DNA was extracted from blood of cases and controls using a standard protocol. The Ethics Committee at the Karolinska Institute has approved these studies.

METHODS

Genotyping

A genome-wide scan for linkage was initially performed by using 360 microsatellite markers with an average spacing of 9.6cM (349 autosomal and 11 sex chromosomal markers). The markers were selected from the integrated maps of the Weber 6 screening set (Sheffield et al, 1995). The average intermarker distance across the fine-mapped regions was 2.5 cM. In regions with LOD score higher than suggestive

for linkage, new markers were added from Marshfield Medical Research foundation (<http://research.marshfieldclinic.org/genetics>), and Decode database (http://www.ensembl.org/Homo_sapiens/textview) for fine mapping. To detect microsatellites, fluorescent dyes (6-FAM, HEX, and TET) were attached to the 5' end of PCR primers. Normally, one microsatellite locus was amplified per PCR and different PCR products were pooled together following electrophoresis. PCR products were size fractionated on ABI 377 and ABI3130 machines (Applied Biosystems, Foster City, California). When activated by laser light, each dye of fluorescence emits a signal which can be detected. The genotyping results were analyzed by Genescan 2.1 and Genotyper 2.0 software (Applied Biosystems). These software programs recognize the standard peaks and estimated the sizes of the product peaks based on their migration relative to the standard peaks.

Statistical analysis

Family structures were examined by comparing the expected and observed average identity-by-state of genotyped markers in all sib-pairs using the zGenstat 1.128 software (Henric Zazzi, unpublished). Genotypes were checked for Mendelian inconsistencies using the same program. All inconsistencies were re-analyzed and incompatibilities were resolved unambiguously or individuals and/or pedigrees were excluded from linkage analyses. Since hypospadias is restricted by sex, males with hypospadias were defined as affected, whereas other males with normal phenotype were marked as unaffected and females were coded as unknown. Linkage analysis was performed using the information of all markers used in the genome-wide linkage scan and markers for fine mapping. For every marker, single-point and multi-point linkage analysis were obtained using Allegro software (Gudbjartsson et al. 2000). Corresponding p-values were interpreted, and p-values were estimated using the formula

$$p(\text{LOD}) = 0.5 \times (\chi^2_1 > 2 \ln 10 \times \text{LOD})$$

In study I, parametric and non-parametric linkage analyses were performed. For parametric linkage analysis, hypospadias was set as an autosomal dominant trait with reduced penetrance (70%). Allele frequencies of 0.999 for the wild-type allele and 0.001 for the mutant allele were assumed.

In study II, linkage analysis was performed on the whole set of 82 families with 96 affected pairs and for two separate subgroups (Swedish and Middle Eastern origin). Since the pedigrees show mixed patterns of disease inheritance and the true underlying inheritance model is unknown, a non-parametric model was used.

Direct sequencing

In 1977, Frederik Sanger and his colleagues devised one new method for sequencing: chain terminator sequencing (Sanger et al, 1977). This method is divided into four separated reactions and each of reaction contains of a single-stranded DNA template, a short complementary oligonucleotide primer, a DNA polymerase, four standard deoxynucleotides (dATP, dGTP, dCTP and dTTP) and only one of the four dideoxynucleotides (ddATP, ddGTP, ddCTP, or ddTTP). Because of lacking 3' hydroxyl grouping the structure, any ddNTP that is incorporated into a elonging DNA chain cannot participate in phosphodiester bonding at its 3' carbon atom, thereby causing termination of chain synthesis and resulting in DNA fragments of varying length. By setting the concentration of the ddNTP to be very much lower than that of its normal dNTP analog, chain termination will occur randomly at one of the many positions containing the base in question. When these reactions are completed, the DNA fragments are performed separately in electrophoresis. Direct sequencing used today is a modification of Sanger method. In dye-terminator sequencing, each different dideoxynucleotide is labelled with a different fluorophore, so the sequencing reaction can be performed in a single tube containing an unlabelled primer and all of the four different dideoxynucleotides. Then, DNA fragments with different length are size fractionated in an automated DNA sequencer and base pair substitution of sequences is recorded directly as unequal and colorful peaks.

In our study, primers for sequencing were designed in Primer3 software program following requirement: each pair of primers is determined from sequences that flank each exon or from the terminal ends of each exon, as well as they are 17-24 bases in length, with 40-60% GC content. To clean the PCR product, ExoSAP that can remove unconsumed dNTP and primers out of PCR products was used. Direct sequencing was performed on both strands using Big-Dye terminator v3.1 (AppliedBiosystems), and run on an ABI3130 Sequencer (AppliedBiosystems). Sequencing primers were the same as used in the PCR amplification. The sequencing reactions were carried out according to the manufacturer's recommendations. Chromatograms were analyzed using SeqScape v3.7 (Applied Biosystems).

Allele-specific PCR amplification

The principle of allele-specific PCR is using paired PCR reactions to detect alleles that differ at just a single nucleotide. One primer called a common primer is the same in both reactions, the others in two

slightly different versions, one specific for allele 1 and the other specific for allele2. Primer for allele 1 will bind perfectly to the complementary strand of the allele1 sequence, permitting amplification with the common primer. The primer for allele 2 mismatches with one nucleotide in allele 1 sequence, making amplification impossible. Similarly the primer for allele 2 can bind perfectly to allele 2 and initiate amplification, unlike primer for allele 1. Additional control (non-specific) primers are usually included, to amplify unrelated sequence from every sample to check that the PCR reaction has worked.

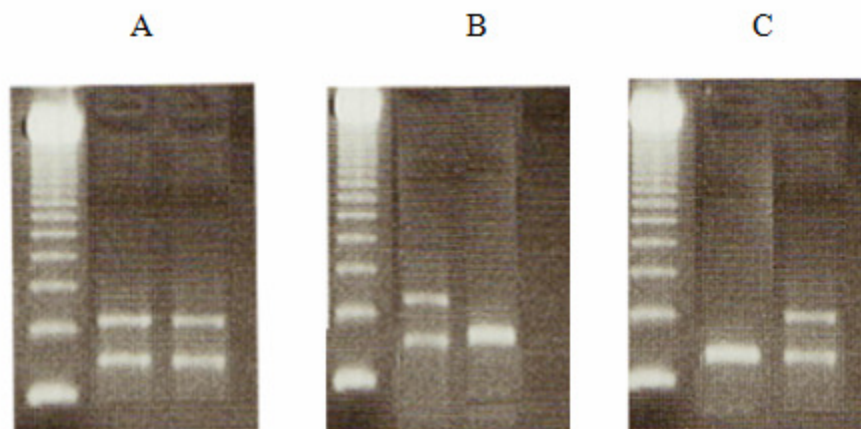


Figure 6: Allele specific PCR. PCR product of each individual is performed in two rows, the left for valine and the right for leucine allele. The smaller band is indicated for non-specific PCR product, the larger band is indicated for specific PCR product. A: heterozygous Leucine/Valine; B: homozygous for Valine allele; C: homozygous for Leucine. A 123 bp size ladder is in the lane to the left.

In paper III, the *SRD5A2* gene was analyzed for the V89L polymorphism using allele specific PCR. The primers used for PCR amplification are the forward primer, the non-specific reverse primer, and either the valine-specific reverse primer or the leucine-specific reverse primer. PCR products are resolved on a 3% agarose-gel, stained with ethidium bromide, visualized under UV light, and photographed.

Taqman genotyping assay

SNP typing using 5'-nuclease allele discrimination TaqMan assay is a robust allele discrimination method. For one reaction, two TaqMan probes that are specific for each allele are used. The fluorogenic probe consists of an oligonucleotide labeled with both a fluorescent reporter dye on the 5' end (FAM or VIC) and a quencher dye on the 3' end (TAMRA). The first criterias for selecting probes are: probes must be as short as possible, normally less than 20- to 24-mer sequence, and the polymorphic site of

probes can not locate within the first two or last two nucleotides of the probe. Moreover, proximity of the quencher reduces the fluorescence signal observed from the reporter dye. PCR amplification and detection with fluorogenic probes in the 5' nuclease assay follow three main steps:

- Polymerization: the reporter and a quencher are attached to the fluorogenic probe.
- Strand displacement: reporter dye emission is quenched when both dyes are attached to the probe.
- Cleavage: the DNA polymerase cleaves the reporter dye from the probe. Once the reporter dye is separated from the quencher, the fluorescence intensity of this dye is increased.

These steps result in substantial increase in FAM or VIC fluorescent signal and indicate homozygosity for the FAM- or VIC-specific allele. An increase in both signals indicates heterozygosity. ABI 7900HT (Applied Biosystems) was used for thermal cycling. Single marker and multiple markers association test were performed using Haploview both with and without missing data estimation. The p-value was corrected for multiple testing by performing 10 000 permutations using Haploview. Differences in haplotype frequencies between cases and controls were calculated using Chi-square test and Fisher test on Statistica 7.0 software.

Bioinformatics tools

Genetic variants including point mutations, insertions, deletions, translocations, and duplications, play critical roles in determining biological phenotypes and disease susceptibilities. Some coding variants may affect protein structure such as protein folding, active sites, protein–protein interactions, protein stability and mRNA expression. If these variants are conserved they can be identified through the multiple sequence alignment with the orthologous sequences from different species. Moreover, the impact of a missense mutation can also be modelled if the protein or motif structure is available in the mutation region. Nonsense and frame-shift variants can introduce an early termination and lead to nonsense-mediated decay or a truncated protein with obvious functional changes. Synonymous variants are thought as “silent”, but these changes can be part of exonic or intronic splicing regulatory elements that affect the gene expression via altered splicing process. Variants in the regulatory regions may alter the consensus of transcription factor-binding sites, splicing efficiency or promoter elements, as well as variants in the untranslated regions (5'UTR or 3'UTR) may change mRNA stability. Therefore, investigating the functional roles of these variations provides a great opportunity for understanding

complex common diseases. The *in silico* approach that plays a critical role in prediction of structural and functional consequences of a particular DNA variant greatly enhances the ability to characterize functional DNA variants. The principle of bioinformatics tools is based on analysing wild-type and mutant sequences and looking for an alteration in the predicted outcome. Genetic variants can also be evaluated at a simple level by looking at physical considerations of the properties of genes and proteins, or they can be evaluated in the context of a variant within a family of homologous or orthologous genes or proteins. Furthermore, *in silico* analysis can identify the function of promoter DNA variants at the conserved sites and potential binding motifs for transcription factors.

In study III and IV, the impact of the gene variants in 5' and 3' splice sites were predicted using three *in silico* tools: NetGene 2.0 (available at <http://www.cbs.dtu.dk/services/NetGene2/>), NNSplice (available at http://www.fruitfly.org/seq_tools/splice.html) and HSF 2.4 (available at <http://www.umd.be/HSF/>). HSF 2.4 was created by integration of two different matrices, one from SSF (www.genet.sickkids.on.ca/ali/splicesitefinder) and the other from MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html). NetGene 2.0, NNSplice and SSF can identify the most probable splice site in a given sequence and then assign a score of splice site strength. In contrast, MaxEntScan is unable to read through a sequence and the exon-intron boundary has to be indicated. Since different matrix was used in each tool, our conclusion was based upon consistency of result from at least two of these tools. It is recommended that the score from mutant type should be at least 20% lower than the wild type score in order to be a positive and deleterious prediction (Houdayer et al, 2008). The alteration of potential branch points, splicing enhancers and silencers were also checked using HSF 2.4 program. Promoter prediction was performed with Regulatory analysis of Variation in Enhancers (RAVEN) created by CGB, Karolinska Institutet (<http://burgundy.cmmt.ubc.ca/cgi-bin/RAVEN/a?rm=home>). A putative homolog of human KLF6 was detected by HomoloGene and alignments were made with ClustalW software (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>). The three-dimensional structure was predicted at Protein Homology/analogy recognition Engine from the Structure bioinformatics Group, Imperial College, London (<http://www.sbg.bio.ic.ac.uk/~phyre/>).

RESULTS AND DISCUSSION

STUDY I: A new susceptibility locus for hypospadias on chromosome 7q32.2-q36.1

Genome-wide linkage analysis using 360 polymorphic microsatellite markers was performed on a large family with eight hypospadias boys (Figure 5). By using an autosomal dominant model, potential linkage was detected in nine regions distributed on six chromosomes: 1q41-1q42.3, 3p24.2-3p21.2, 6p12.2-6q16.1, 7p22.2-7p21.3, 7q32.2-7q36.1, 12p13.31-12p12.3, 12q21.1-12q24.33, 15q15.3-15q22.31 and 15p26.2-15p26.3. Sixty-six additional microsatellite markers were added for fine mapping of these regions, which confirmed the peak on 7q32.2-q36.1 and resulted in loss of the peaks on chromosomes 1, 3, 6, 12 and 15 (Table 3). We obtained linkage with maximum single-point LOD score on chromosome 7 at the locus for marker D7S640, both in the parametric (LOD score 2.71) and non-parametric (NPL score of 5.01) analysis. The absence of increase in the LOD score after multipoint analysis (LOD= 2.37) compared to single-point analysis (LOD=2.71) is most likely explained by the fact that the information value is already very high in the single-point analysis and therefore the small increase in information value in the multi-point analysis does not increase the LOD score.

Chromosome region	Marker	Peak location (cM)	Single-point			Multipoint		
			LOD	p-value	NPL	LOD	p-value	NPL
1q41-1q42.3	D1S2785	243.64	0.48	0.068	1.35	-0.22	0.5	1.38
3p24.2-3p21.2	D3S1768	57.76	-0.742	0.5	0.12	-2.54	0.5	0.62
6p12.2-6q16.1	D6S1040	99.7	0.05	0.5	-0.14	-1.98	0.5	0.36
7p22.2-7p21.3	D7S641	8.72	0.65	0.041	1.13	0.64	0.043	1.4
7q32.2-7q36.1	D7S640	137.9	2.71	0.0002	5.01	2.37	0.0004	11.09
12p13.31-12p12.3	D12S364	22.6	-3.05	0.5	-0.15	-4.77	0.5	-0.21
12q21.1-12q24.33	D12S395	131.63	0.17	0.188	-0.26	-2.2	0.5	-0.24
15q15.3-15q22.31	D15S1033	51.29	0.31	0.116	-0.99	-0.56	0.5	-0.23
15p26.2-15p26.3	D15S657	109.34	0.19	0.174	0.69	-0.21	0.5	0.29

Table 4: The resulting maximum single-point and multipoint LOD-score from the fine mapping linkage analysis of chromosome 1, 3, 6, 7, 12 and 15

Twenty-three markers were used to construct a haplotype confining the critical interval of the linked region spanning from 116 cM to 146 cM (Figure 7). On the basis of these recombination events, the maximum linked region is determined to a 25 cM genetic interval at 7q32.2-q36.1, corresponding to an 18.2 Mb region on the physical map (www.ensembl.org). All seven affected members have inherited the same haplotype in this region, but none of the unaffected male individuals.

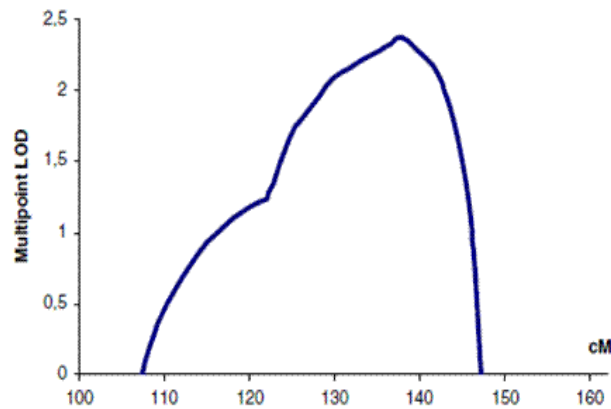


Figure 7: Multipoint LOD score results on chromosome 7q32.2-36.1. The LOD score is given on the y-axis and cM from pter on the x-axis

The *AKR1D1* and *PTN* genes located in the candidate region were chosen for sequencing analysis since their function and expression are involved in the androgen pathway. *AKR1D1* can catalyze 5-beta-reduction of steroid hormones and inactivates active androgens like testosterone (Kondo et al 1994, Charbonneau and The, 2001). *PTN* is expressed in the mesenchyme of the reproductive tract during development and its expression is up-regulated by estrogen and testosterone (Vanderwinden et al. 1992, Vacherot et al. 1995). Mutation analysis was performed on all exons and flanking intronic sequences of these genes but no mutation was found.

In this study, we provide evidence of a novel susceptibility locus for familial hypospadias on chromosome 7q32.2-q36. This is the first chromosomal region recognized for isolated hypospadias. Since the 7q32.2-q36.1 region contains around 300 known genes, further studies concerning candidate gene identification need to be done. Once a causative gene is identified in this specific family, it may suggest a new biological pathway and hence additional candidate genes in hypospadias.

STUDY II: Fine mapping of familial hypospadias identifies two new candidate genes: AKRIC3 and KLF6

Our previously published genome-wide linkage scan in 69 families had revealed evidence for suggestive linkage on chromosome 2p11, 9q21-22, 10p15 and 10q21 and weaker linkage signal on chromosome 1q23, 6q25 in the sub-group of 58 Swedish families and on chromosome 8q24 and 18q21 in the sub-group of 11 families with Middle Eastern origin (Frisén et al, 2004). For fine-mapping with the inclusion of eighty-five additional markers and 13 extra families (11 Swedish and 2 Iranian families), we could confirm and strengthen the evidence of linkage for chromosome 8q24.1 (LOD = 1.32, $p < 0.03$) and 10p15 (LOD = 1.67, $p < 0.02$) (Figure 8).

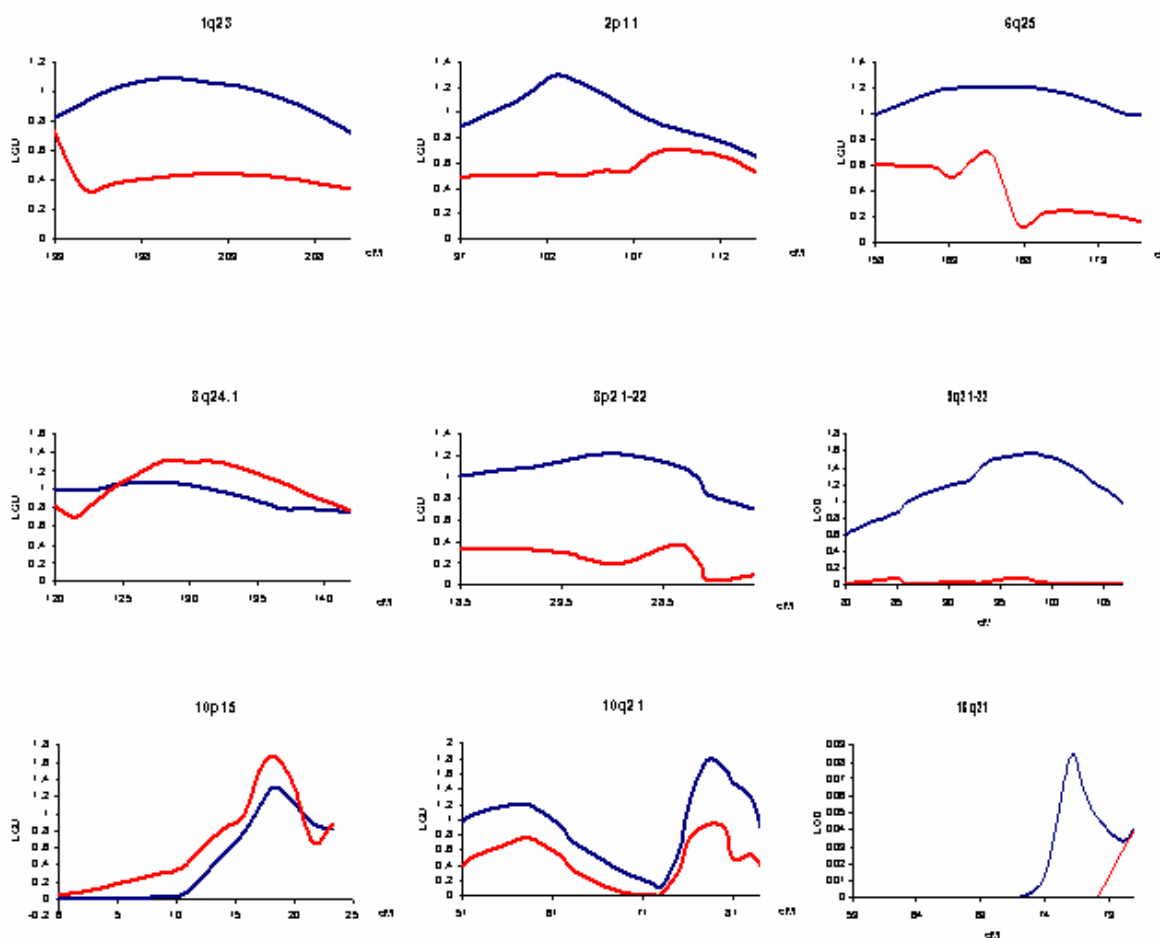


Figure 8: Comparison between multi-point LOD score in the initial genome-wide linkage scan (69 families, blue line) and the current fine mapping study (82 families, red line). LOD score is shown on the y-axis and cM from the pter on x-axis.

Interestingly, these regions showed evidence of linkage in different subgroups in the previous study, the 10p15 region in the Swedish subgroup and the 8q24.1 region in the Middle Eastern group. The peaks were also confirmed on 1q23 (LOD= 1.27, $p < 0.02$) and 10p15 (LOD= 1.14, $p < 0.04$) in the Swedish subgroup, as well as on 8q24 (LOD=1.04, $p < 0.03$) and 18q21 (LOD= 1.1, $p < 0.03$) in the group of Middle Eastern families (Table 4). We chose two genes in the 10p15- region that showed highest LOD score, namely the *AKRIC3* and *KLF6* genes, to perform sequencing analysis.

GROUP	Chomosomal region	Multipoint		Marker	Single-point	
		LOD	p		LOD	p
All	8q24.1	1.32	< 0.03	D8S514	1.02	<0.06
	10p15	1.67	<0.02	D10S1779	1.50	<0.02
Middle	8q24.1	1.04	<0.03	D8S514	0.61	<0.07
Eastern	18q21	1.10	<0.03	D18S858	0.95	<0.05
Swedish	1q23	1.27	<0.02	D1S518	1.02	<0.07
	10p15	1.14	<0.04	D10S1412	1.03	<0.06

Table 4: Summary of all linkage peaks for fine mapping with maximum multipoint LOD score greater than 1 and their corresponding single-point LOD score obtained in linkage analysis in all set of families and in subgroups.

AKRIC3

The *AKRIC3* gene is an interesting candidate gene for hypospadias since *AKRIC3* interconverts active androgen and estrogen with their inactive metabolites and it could also regulate the ligand occupancy of androgen receptor and estrogen receptor in target tissues (Khanna et al, 1995; Penning et al, 2004). One novel heterozygous mutation in the *AKRIC3* gene, p.A215T, was detected in one boy with penile hypospadias. The mutant threonine allele of p.A215T causes a break of a silencer site and instead creates a new site for enhancer binding, and thus the mutation can affect the translation of *AKRIC3*, which might therefore be pathogenic. We also report additional *AKRIC3* sequence alterations that can affect enhancer or silencer splice sites (Table 5). Moreover, we were able to show an association between polymorphisms in the *AKRIC3* gene and hypospadias (Table 6). The most significant association was detected with rs12529 (allele G, $p < 0.01$). By creating additional sites for enhancer, the rare allele of

rs12529 may increase the reductase activity of *AKR1C3* and result in low level of 5 α -DHT. In contrast, the rare alleles of rs3763676 (g) and rs7741 (A) decrease this activity and may reduce the risks for hypospadias. These effects could also explain the relationship of the aGG and gCA haplotypes to hypospadias. Haplotype aGG was over-represented among cases, whereas haplotype gCA was over-represented in the control group (with the protective alleles allele g of rs3763676 and allele A of rs7741). We hypothesize that by affecting enhancer or silencer splice sites in the splicing process, the *AKR1C3* gene variants might exhibit or inhibit the transcription. Through this mechanism the concentration of 5 α -DHT in target tissue as well as the balance between androgens and estrogens might be altered. Our results indicate that sequence variants of *AKR1C3* can affect the risk for hypospadias.

Gene	Gene variants	Branch point site	Enhancer site	Silencer site
<i>AKR1C3</i>	c.15C>G(p.Q5H)		New site	
	c.90G>A (p.P30)		Broken site	New site
	c.697A>G (p.A215T)		New site	Broken site
<i>KLF6</i>	c.496T>C (p.P166S)		New site	Broken site
	c.802+21c>t			New site
	*46a>t	Damaged site		New site
	*62g>a		New site	Broken site

Table 5: The impact of gene variants on splice regulation factors were predicted using three *in silico* tools: NetGene 2.0, NNSplice and HSF 2.4.

KLF6

Kruppel-like factor 6 (*KLF6*/COPEB) belongs to a large family of Sp1-like/KLF transcription factors that plays a crucial role in many cellular processes associated with differentiation and tissues development including growth of the genital tract. This indicates that the *KLF6* gene might be of importance for the development of hypospadias. We have also found novel sequence variants in the *KLF6* gene, of those one variant is a missense mutation (p.P166S). The patient carrying this missense mutation however has a 45,X/46,XY karyotype which makes the evaluation of this finding difficult, since 45,X/46,XY mosaics have a wide spectrum of phenotypes from sex reversal to hypospadias and normal male. We also report four novel intronic variants of *KLF6* that are found in seven cases but not in

controls. Splice site analysis predicted that these variants might have an impact on splicing process by affecting on enhancer, silencer and potential branch point site (Table 5). We hypothesize that *KLF6* can induce hypospadias either through the expression of *KLF6* in placenta or through an up-regulated function on the *ATF3* gene, one of the risk genes for hypospadias (Blanchon et al, 2001, Huang X et al, 2008).

	Alleles	Allele frequencies (n)		p-value
		Cases	Controls	
rs3763676(c1-138a>g)	a	0.69 (475)	0.65 (1155)	n.s
	g	0.31 (213)	0.35 (623)	n.s
rs12529 (Q5H)	C	0.54 (367)	0.60 (1064)	n.s
	G	0.46 (315)	0.40 (706)	p<0.01
rs7741(P30)	G	0.69 (502)	0.65 (1163)	p<0.03
	A	0.31 (224)	0.35 (637)	n.s
		Case (%)	Controls (%)	p- value
Haplotype	aGG	0.46	0.40	0.0094
	gCA	0.30	0.35	0.0324
	aCG	0.23	0.25	0.2969

Table 6: SNP genotyping results in the *AKRIC3* gene. Significant differences in allele frequencies ($p<0.05$) between cases and controls were observed for rs12529 and rs7741. Haplotype analysis for rs3763676, rs 12529 and rs7741 (in order) detected one block using Haploview 4.0 software. The haplotype aGG is over-represented in patients ($p=0.0094$) and the haplotype gCA is over-represented in controls ($p=0.0324$). n.s: not significant, n: number of chromosomes

In summary, we report in this study two new susceptibility genes for hypospadias, *AKRIC3* and *KLF6*, based on our findings from linkage analysis and fine mapping in 82 families as well as result of DNA sequencing and association studies.

STUDY 3: The Valine allele of the V89L polymorphism in the 5-alpha reductase gene confers a reduced risk for hypospadias

The role of the *AR*, *SRD5A2* and *HSD17B3* genes on the androgen pathway make them all obvious candidate genes for hypospadias. However, mutations in these genes are usually found in severe cases of hypospadias and often associated with other genital malformations. We initially analyzed the DNA sequences of the *AR* gene and *SRD5A2* gene in 37 cases with isolated severe hypospadias. In one boy with scrotal hypospadias we detected a point mutation in exon 6 in the *AR* gene in codon798 (p. Q798E). This mutation in the *AR* gene has been described previously in cases with different severity of under masculinisation but this is the first case of Q798E mutation was found in isolated hypospadias (Bevan et al, 1996; Wang et al, 1998).

For the *HSD17B3* gene, we chose to analyze index cases from the 19 families participating to the chromosomal 9q22.32 peak in the initial sib-pair analysis (Frisén et al, 2004). No mutations were found in the coding regions of *HSD17B3* in that subgroup of familial hypospadias.

Patient group (no. individuals)	Val-allele frequency (%)	OR (95% CI)	Chi Sqr	p	pc*
Hypospadias (158)	48	0.38 (0.26-0.6)	24.2	9. 10 ⁻⁷	2.10 ⁻⁵
Sporadic (69)	49	0.39 (0.25-0.61)	15.9	0.0001	0.003
Familial (89)	48	0.38 (0.25-0.58)	19.5	0.00002	0.0006
Swedish (99)	53	0.46 (0.31-0.69)	13.0	0.0003	0.0008
Turkey (28)	34	0.21 (0.11-0.39)	23.5	2.10 ⁻⁶	0.00004
Middle-East (26)	44	0.33 (0.18-0.66)	11.6	0.0007	0.02
Mild phenotype (60)	51	0.34 (0.17-0.66)	9.0	0.003	n.s
Middle phenotype(32)	44	0.37 (0.15-0.94)	3.4	n.s	n.s
Severe phenotype (31)	45	0.27 (0.11-0.67)	6.6	0.02	n.s
Control (96)	71				

**Table 7: Association of the Val allele of the V89L polymorphism in the *SRD5A2* gene with hypospadias pc* = p-value corrected for the number of comparisons made (n=27)
n.s: not significant**

DNA sequencing of the *SRD5A2* gene revealed one mutation in heterozygous form (p.K199S). While sequencing the *SRD5A2* gene in the initial 37 hypospadias boys, we detected an over representation of leucine allele in the V89L polymorphism of this gene in cases as compared to earlier published population result (Ntais et al, 2003). We therefore analyzed additional 121 hypospadias cases and 96 controls by performing an allele-specific PCR for the V89L polymorphism. Totally, in the 158 hypospadias cases the allele frequencies were 48% for valine (V89) and 52% for leucine (L89) as compared to controls, 71% and 29%, respectively. This finding can be explained by the fact that leucine in codon 89 (L89) results in almost 30% reduced enzyme activity and thus lower concentration of DHT (Makridakis et al, 2000; Allen et al, 2001; Hsing et al, 2001).

Patient group (no. individuals)	Val/Val %	Val/Leu %	Leu/Leu %
Hypospadias (158)	27.8 ¹	40.5 ³	31.6
Sporadic (69)	27.6 ²	42.0	30.4
Familial (89)	28.1 ²	39.3	32.6
Swedish (99)	0.3	44.4 ³	25.3
Turkey (28)	21.4 ³	53.6	23.1 ³
Middle-East	23.1 ³	42.3	34.6
Mild phenotype (60)	31.7 ³	38.3	30.0
Middle phenotype (32)	28.1 ³	31.3	40.8
Severe phenotype (31)	25.8 ³	38.7	32.4
Control (96)	61.5	18.8	19.8

Table 8: Association of V89L polymorphism with hypospadias.

¹ Comparison with controls resulted in $p < 3 \cdot 10^{-7}$, $pc^* < 7 \cdot 10^{-6}$

² Comparison with controls resulted in $p < 5 \cdot 10^{-5}$, $pc^* < 0.001$

³ Comparison with controls resulted in $p < 0.002$, $pc^* < 0.05$

⁴ Comparison with controls resulted in $p < 0.05$, pc^* : n.s

Boys homozygous for valine allele (V89) had an odds ratio (OR) for having hypospadias of 0.24 (95% CI 0.14-0.41), pc -value= 7×10^{-6} , while carriers of the leucine allele (L89) in homozygous or heterozygous form had an OR of about 2 (1.9 and 2.9, respectively). Furthermore, the protective effect of being homozygous for the valine allele was more significant in mild phenotypes although it was also

evident in severe hypospadias. In the group with severe hypospadias (penoscrotal, scrotal and perineal) 26% are homozygous for valine, as compared to our control material with 61.5%.

We hypothesize that the Val89 allele of the SRD5A2 gene in homozygous form reduces the risk for hypospadias in itself, or is in linkage equilibrium with another susceptibility factor.

STUDY 4: The p.N589S polymorphism in the MAMLD1-gene is associated with hypospadias

The Mastermind-like domain containing 1 (*MAMLD1*) gene, previously known as the Chromosome X open reading frame 6 (*CXorf6*) gene, was identified as the first causing gene for isolated hypospadias (Fukami et al, 2006). Subsequently previous DNA sequencing analysis has shown mutations of the *MAMLD1* gene in a few cases hypospadias, ranging from mild to severe hypospadias (Fukami et al, 2006; Kalfa et al, 2008). In this study, we performed direct DNA sequencing of the *MAMLD1* gene in a large material of sporadic hypospadias cases in order to elucidate how common these mutations are in the pathogenesis of hypospadias.

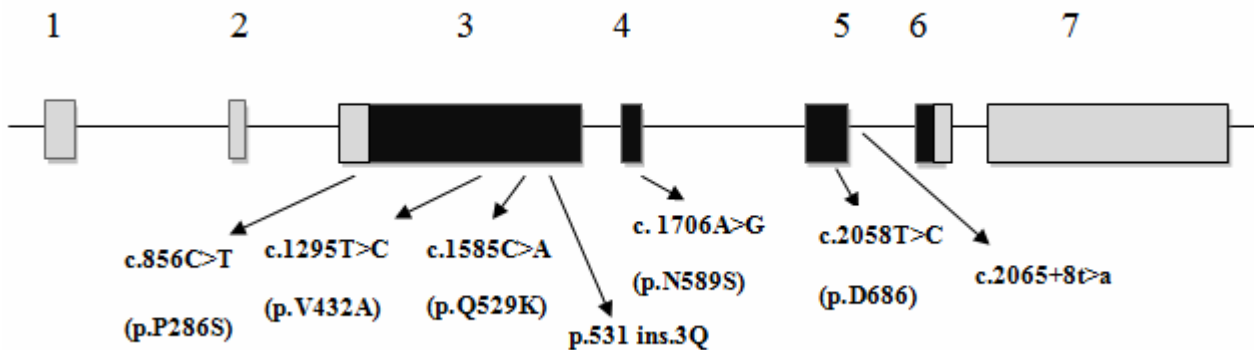


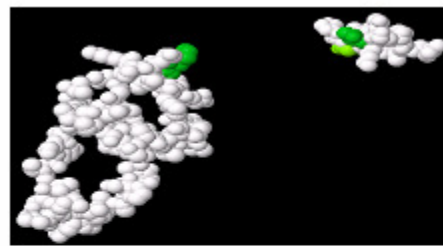
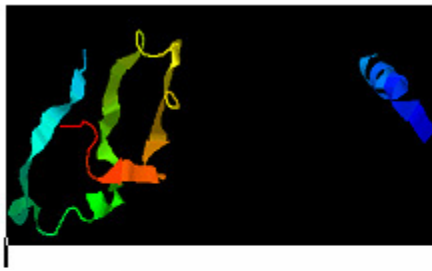
Figure 9: Seven gene variants, mainly in exon 3, were found in the *MAMLD1* gene in 99 hypospadias cases. The gene variant p.Q529K has not been previously described. Black boxes represent the nonsense-mediated decay region and gray boxes represent the stable region.

Our sequencing analysis in 99 sporadic cases identified seven gene variants (Figure 9). Of those, a novel missense variant (p.Q529K) in exon 3 of the *MAMLD1* gene was detected, that was also absent among

controls. According to data from splicing site detection, p.Q529K might change many motifs of exonic splicing enhancers that are necessary for splicing, and seems to determine splicing pathways of *MAMLD1*. Moreover, this mutation is located in the region that can trigger nonsense-mediated mRNA decay and produce truncated protein, which will be harmful or nonfunctional. Therefore, p.Q529K may be a pathogenic mutation.

Two other gene variants, p.V432A and insertion (CAG₁₀-CAG₁₃), are reported as mutations in a previous study (Kalfa et al, 2008) but they also appear in our control group. Moreover, allele frequencies of these variants do not differ between cases and controls. Thus, these variants seem to be very rare polymorphisms. The silent variant c.2058T>C and the intronic variant c.2065+8t>a were not detected among ninety-five control males and have not been previously described, and may be risk factors.

Wild- type



N589S

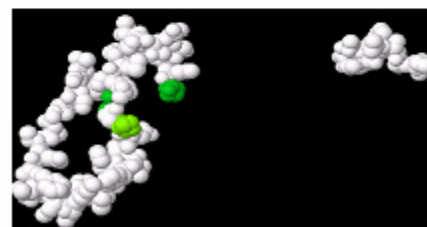
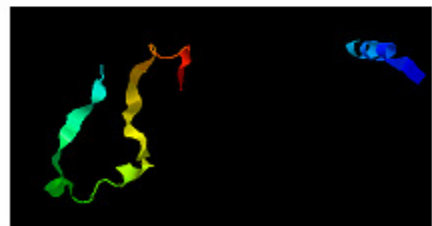


Figure 10: Three- dimensional protein structure (A) and secondary structure (B) prediction of wild-type and mutant MAMLD1 using Protein Homology/ analogY Recognition Engine (PHYRE).

We also detected an association between the p.N589S polymorphism (rs2073043) and hypospadias, but not for the p.P286S polymorphism (rs41313406). Notably, the alteration in codon 589 was predicted to make a change on the structure of protein (Figure 10). Hence, we hypothesize that p.N589S represent a

regulatory variant of *MAMLD1* and this is the first polymorphism of the *MAMLD1* gene associated with hypospadias. Moreover, the combination of two rare alleles of p.N589S and p.P286S polymorphisms (T and G) is over represented among cases. It may be that these two polymorphisms are inherited together as a haplotype (T-G), thus increasing the risk for hypospadias.

Our findings suggest that hypospadias is caused by mutations in the *MAMLD1* gene in a few cases and the polymorphism (p.N589S) is a risk factor for hypospadias.

CONCLUDING REMARKS

With the aim to identify the genes for hypospadias, gene mapping with linkage analysis was performed. The susceptibility chromosome region 7q32.2-q36.1 was identified in one large family with multiple affected boys. The major problem to select candidate genes for hypospadias in this family is that the region contains many genes, and in some cases, current knowledge is insufficient to make prediction for their function. Even if we did not identify the candidate genes in this region, our finding provides a platform for further investigation. The next step will be to perform an association study between DNA variants in genes that are located in this region and sporadic hypospadias. If hypospadias is associated with DNA variants in an index gene, some affected individuals may have mutations in that gene.

The results from linkage analysis in the affected sib-pair study helped us to define two new candidate genes for familial hypospadias: the *AKRIC3* and *KLF6* genes. The functions of these two genes in the androgen pathway strengthen once more the importance of this pathway in the development of hypospadias. In this study, we also detected the impact of some *AKRIC3* and *KLF6* variants in the splicing process through *in silico* tools. These findings need to be confirmed *in vitro* experiments. In a complex genetic disease, not only mutations located within candidate genes but also mutations in distant genes that can regulate the expression of candidate gene can be pathogenic (Kleinjan et al, 2005). We assumed that the association of different genes from different candidate regions might induce the development of disease. Our future work should center on developmental genes that may regulate candidate genes located within our susceptibility regions.

Several findings in sequencing analysis on genes involved in hypospadias (the *AR*, *SRD5A2* and *MAMLD1* genes) also suggest that mutations in candidate genes might cause hypospadias in a few severe cases. Our result from association study of common variants in the *SRD5A2*, *MAMLD1* and *AKRIC3* genes indicate that variants in candidate genes can be susceptibility factors. Moreover, these variants are inherited together as a haplotype, thus increasing the risk for hypospadias.

We carried out these studies with the hope to elucidate the genetic background of hypospadias. A better understanding of the pathogenesis of hypospadias as well as a better diagnosis and information will be

provided. This is especially important in cases with severe hypospadias that the sex at birth is difficult to assign. In addition, new knowledge on normal genital development will be added.

ACKNOWLEDGEMENTS

During my time as a PhD student I have learnt a lot from many people and I would like to express my sincere gratitude to all of you who have supported, taught and helped me. Especially, I would like to thank:

Agneta Nordenskjöld, my main supervisor, for accepting me to be a member in her group. Thank you very much for your enthusiasm, support, endless patience, and most of all for giving me the opportunity to grow and to express my opinion and idea.

Kristina Lagerstedt, my co-supervisor, for many enthusiastic discussions about human genetics and for always answering my questions about molecular techniques. I hope that we can continue our “journal club” after my dissertation.

Johanna Lundin, my co-supervisor, for having valuable discussion and comments. Thanks for always supporting me.

The Hung Bui, my mentor: Thanks for introducing me to Sweden, for always having time for me and for treating me as your daughter. I have learnt a lot from you.

Ingrid Kockum, my former co-supervisor, and Cilla Söderhäll for a lot of assistance, for being helpful and for your valuable guidance in gene mapping and genetic statistics.

Magnus Nordenskjöld for creating an enjoyable atmosphere and stimulating research environment.

My friends and colleagues in the *Malformation Genetics* group:

Fredrik, Anna, Magdalena, Ana- Beleza, and Christina: for much laughter, for good company, for your help with all kind of problems and for creating an excellent work environment. I will always remember the fun chats, the “tårta” lab meetings and my birthday parties in our lab. Yougen and Zu: for your admirable work on sequencing project. Agne, for teaching me a lot about association study and for joining me in MAF project.

My friends in writing room: thank you so much for your friendship and for “cookie” time. Husam, for fun chats and for the help with immunohistochemistry. Tian Ling, for chats about fashion and for cocktail book. Simone and Sanna, for always answering my question about association study.

All present and former member of CMM 02, especially:

Professor Anna Wedell, Professor Annika Lindblom, Günther, Michela, Jessica, Virpi (thank you so much for slides), Petra, Paula, Tanja, Fabio, Keng- Ling, Feng-qing, Bying, Xiaoli, Anya, Jaqueline, Annika, Tiina, Emma, Clara, Enikö and Andor, Katja, Lovisa, and all the others that I might have forgotten but whom should be here and mentioned: thank you for everything.

Laboratory technicians, Anki, Sigrid, Barbro and Anna-Lena: thanks for all your help and support.

Margareta Tapper-Persson, you are always in my heart.

Yvonne Cowan, Kerstin Florell, Majalena Granqvist, Christina Bremer, Delphi Post and Astrid Häggblad: thanks for administrative work.

Lennart Helleday for always having time to solve my computer problems.

To all my friends and colleges from Hanoi Medical University, Department of Biology and Genetics: thanks for supporting me and continuing my father’s work. To all my Vietnamese friends in Stockholm, particularly anh Chuyên, chị Ngọc, anh Đức, Thơm, Lan Anh, Hà , Hiếu, Phương, anh Đô, chị Hương, chị Dung, anh Hòa, Ngọc Anh, chị Hương (cao), thank you for all your help and for making my life here more joyful.

To my father in-law and my family in Stockholm, mẹ Hằng, Kristian, Agnes, Trâm, Yến Nhi, Ken, thank you for welcoming me into our family, for all the encouragement and inspiration, for always being beside me when I need.

To my dear sister, brother in-law, Bồng and Tí: for always believing me, for your warm spirit, for very fun calls and for helping me to take care of our parents.

To my mother and my late father: Con không biết phải nói những lời cảm ơn thế nào đến bố mẹ nhưng con biết rằng con không thể trưởng thành đến ngày hôm nay nếu không có tình yêu thương vô bờ bến và sự chỉ dạy cẩn thận của bố mẹ. Con vô cùng tự hào là con gái của bố mẹ và xin dành tặng bản luận văn này như một sự tri ân đến những người mà con yêu kính nhất.

To Maja, my angel, for your love, for your morning kisses and evening hugs.

To my “Minh”. Before I met you, I was a girl with a boring life. You came and have shone so much happiness, love and wonderful thing on my life. This thesis would not have been finished without your help and support. You and Maja mean everything to me. I love both of you plenty!

REFERENCES

- Akre O, Lipworth L, Cnattingius S, Sparen P, Ekblom A (1999) Risk factor patterns for cryptorchidism and hypospadias. *Epidemiology* 10: 364-9
- Allen NE, Forrest MS, Key TJ (2001) The association between polymorphisms in the CYP17 and 5alpha-reductase (SRD5A2) genes and serum androgen concentrations in men. *Cancer Epidemiol Biomarkers Prev* 10: 185-9
- Allera A, Herbst MA, Griffin JE, Wilson JD, Schweikert HU, McPhaul MJ (1995) Mutations of the androgen receptor coding sequence are infrequent in patients with isolated hypospadias. *J Clin Endocrinol Metab* 80: 2697-9
- Amos CI, Williamson JA (1993) Robustness of the maximum-likelihood (LOD) method for detecting linkage. *Am J Hum Genet* 52: 213-4
- Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3: 299-309
- Aschim EL, Haugen TB, Tretli S, Daltveit AK, Grotmol T (2004) Risk factors for hypospadias in Norwegian boys - association with testicular dysgenesis syndrome? *Int J Androl* 27: 213-21
- Batch JA, Evans BA, Hughes IA, Patterson MN (1993) Mutations of the androgen receptor gene identified in perineal hypospadias. *J Med Genet* 30: 198-201
- Bauer SB, Bull MJ, Retik AB (1979) Hypospadias: a familial study. *J Urol* 121: 474-7
- Bauer SB, Retik AB, Colodny AH (1981) Genetic aspects of hypospadias. *Urol Clin North Am* 8: 559-64
- Beleza-Meireles A, Kockum I, Lundberg F, Soderhall C, Nordenskjold A (2007a) Risk factors for hypospadias in the estrogen receptor 2 gene. *J Clin Endocrinol Metab* 92: 3712-8
- Beleza-Meireles A, Lundberg F, Lagerstedt K, Zhou X, Omrani D, Frisen L, Nordenskjold A (2007b) FGFR2, FGF8, FGF10 and BMP7 as candidate genes for hypospadias. *Eur J Hum Genet* 15: 405-10
- Beleza-Meireles A, Tohonen V, Soderhall C, Schwentner C, Radmayr C, Kockum I, Nordenskjold A (2008) Activating transcription factor 3: a hormone responsive gene in the etiology of hypospadias. *Eur J Endocrinol* 158: 729-39
- Bevan CL, Brown BB, Davies HR, Evans BA, Hughes IA, Patterson MN (1996) Functional analysis of six androgen receptor mutations identified in patients with partial androgen insensitivity syndrome. *Hum Mol Genet* 5: 265-73
- Blanchon L, Bocco JL, Gallot D, Gachon AM, Lemery D, Dechelotte P, Dastugue B, Sapin V (2001) Co-localization of KLF6 and KLF4 with pregnancy-specific glycoproteins during human placenta development. *Mech Dev* 105: 185-9
- Boisen KA, Chellakooty M, Schmidt IM, Kai CM, Damgaard IN, Suomi AM, Toppari J, Skakkebaek NE, Main KM (2005) Hypospadias in a cohort of 1072 Danish newborn boys: prevalence and relationship to placental weight, anthropometrical measurements at birth, and reproductive hormone levels at three months of age. *J Clin Endocrinol Metab* 90: 4041-6
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32: 314-31
- Brinkman RR, Dube MP, Rouleau GA, Orr AC, Samuels ME (2006) Human monogenic disorders - a source of novel drug targets. *Nat Rev Genet* 7: 249-60
- Brouwers MM, Feitz WF, Roelofs LA, Kiemeneij LA, de Gier RP, Roeleveld N (2006) Hypospadias: a transgenerational effect of diethylstilbestrol? *Hum Reprod* 21: 666-9

- Calzolari E, Contiero MR, Roncarati E, Mattiuz PL, Volpato S (1986) Aetiological factors in hypospadias. *J Med Genet* 23: 333-7
- Carlson CS, Eberle MA, Kruglyak L, Nickerson DA (2004) Mapping complex disease loci in whole-genome association studies. *Nature* 429: 446-52
- Carmichael SL, Shaw GM, Laurent C, Croughan MS, Olney RS, Lammer EJ (2005) Maternal progestin intake and risk of hypospadias. *Arch Pediatr Adolesc Med* 159: 957-62
- Carnevale A, Hernandez M, Reyes R, Paz F, Sosa C (1985) The frequency and economic burden of genetic disease in a pediatric hospital in Mexico City. *Am J Med Genet* 20: 665-75
- Cazzola M, Skoda RC (2000) Translational pathophysiology: a novel molecular mechanism of human disease. *Blood* 95: 3280-8
- Charbonneau A, The VL (2001) Genomic organization of a human 5beta-reductase and its pseudogene and substrate selectivity of the expressed enzyme. *Biochim Biophys Acta* 1517: 228-35
- Chen YC, Woolley PV, Jr. (1971) Genetic studies on hypospadias in males. *J Med Genet* 8: 153-9
- Chong JH, Wee CK, Ho SK, Chan DK (2006) Factors associated with hypospadias in Asian newborn babies. *J Perinat Med* 34: 497-500
- Chung E, Coffey R, Parker K, Tam P, Pembrey ME, Gardiner RM (1993) Linkage analysis of infantile pyloric stenosis and markers from chromosome 9q11-q33: no evidence for a major gene in this candidate region. *J Med Genet* 30: 393-5
- Cohn MJ, Bright PE (1999) Molecular control of vertebrate limb development, evolution and congenital malformations. *Cell Tissue Res* 296: 3-17
- Crawford DC, Bhangale T, Li N, Hellenthal G, Rieder MJ, Nickerson DA, Stephens M (2004) Evidence for substantial fine-scale variation in recombination rates across the human genome. *Nat Genet* 36: 700-6
- Czeizel A (1985) Increasing trends in congenital malformations of male external genitalia. *Lancet* 1: 462-3
- Czeizel A, Toth J, Erodi E (1979) Aetiological studies of hypospadias in Hungary. *Hum Hered* 29: 166-71
- Dolk H (1998) Rise in prevalence of hypospadias. *Lancet* 351: 770
- D'Souza I, Poorkaj P, Hong M, Nochlin D, Lee VM, Bird TD, Schellenberg GD (1999) Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc Natl Acad Sci U S A* 96: 5598-603
- Ericson A, Källén B (2001) Congenital malformations in infants born after IVF: a population-based study. *Hum Reprod* 16: 504-9
- Fredell L, Iselius L, Collins A, Hansson E, Holmner S, Lundquist L, Lackgren G, Pedersen J, Stenberg A, Westbacke G, Nordenskjöld A (2002a) Complex segregation analysis of hypospadias. *Hum Genet* 111: 231-4
- Fredell L, Kockum I, Hansson E, Holmner S, Lundquist L, Lackgren G, Pedersen J, Stenberg A, Westbacke G, Nordenskjöld A (2002b) Heredity of hypospadias and the significance of low birth weight. *J Urol* 167: 1423-7
- Fredell L, Lichtenstein P, Pedersen NL, Svensson J, Nordenskjöld A (1998) Hypospadias is related to birth weight in discordant monozygotic twins. *J Urol* 160: 2197-9
- Frisen L, Lagerstedt K, Tapper-Persson M, Kockum I, Nordenskjöld A (2003) A novel duplication in the HOXA13 gene in a family with atypical hand-foot-genital syndrome. *J Med Genet* 40: e49
- Frisen L, Soderhall C, Tapper-Persson M, Luthman H, Kockum I, Nordenskjöld A (2004) Genome-wide linkage analysis for hypospadias susceptibility genes. *J Urol* 172: 1460-3

- Fritz G, Czeizel AE (1996) Abnormal sperm morphology and function in the fathers of hypospadiacs. *J Reprod Fertil* 106: 63-6
- Frydman M, Greiber C, Cohen HA (1985) Uncomplicated familial hypospadias: evidence for autosomal recessive inheritance. *Am J Med Genet* 21: 51-60
- Fukami M, Wada Y, Miyabayashi K, Nishino I, Hasegawa T, Nordenskjold A, Camerino G, Kretz C, Buj-Bello A, Laporte J, Yamada G, Morohashi K, Ogata T (2006) *CXorf6* is a causative gene for hypospadias. *Nat Genet* 38: 1369-71
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296: 2225-9
- Gallentine ML, Morey AF, Thompson IM, Jr. (2001) Hypospadias: a contemporary epidemiologic assessment. *Urology* 57: 788-90
- Gatti JM, Kirsch AJ, Troyer WA, Perez-Brayfield MR, Smith EA, Scherz HC (2001) Increased incidence of hypospadias in small-for-gestational age infants in a neonatal intensive-care unit. *BJU Int* 87: 548-50
- Geissler WM, Davis DL, Wu L, Bradshaw KD, Patel S, Mendonca BB, Elliston KO, Wilson JD, Russell DW, Andersson S (1994) Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3. *Nat Genet* 7: 34-9
- Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 25: 12-3
- Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, Tanzi RE, Watkins PC, Ottina K, Wallace MR, Sakaguchi AY, et al. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306: 234-8
- Hall JG, Powers EK, McLlvaine RT, Ean VH (1978) The frequency and financial burden of genetic disease in a pediatric hospital. *Am J Med Genet* 1: 417-36
- Haraguchi R, Suzuki K, Murakami R, Sakai M, Kamikawa M, Kengaku M, Sekine K, Kawano H, Kato S, Ueno N, Yamada G (2000) Molecular analysis of external genitalia formation: the role of fibroblast growth factor (Fgf) genes during genital tubercle formation. *Development* 27: 2471-9
- Harris EL, Beaty TH (1993) Segregation analysis of hypospadias: a reanalysis of published pedigree data. *Am J Med Genet* 45: 420-5
- Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 2: 204-11
- Hill WG, Weir BS (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. *Am J Hum Genet* 54: 705-14
- Hiort O, Klauber G, Cendron M, Sinnecker GH, Keim L, Schwinger E, Wolfe HJ, Yandell DW (1994) Molecular characterization of the androgen receptor gene in boys with hypospadias. *Eur J Pediatr* 153: 317-21
- Holden AL (2002) The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. *Biotechniques Suppl*: 22-4, 26
- Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, d'Enghien CD, Lauge A, Castera L, Gauthier-Villars M, Stoppa-Lyonnet D (2008) Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat* 29: 975-82
- Hsing AW, Chen C, Chokkalingam AP, Gao YT, Dightman DA, Nguyen HT, Deng J, Cheng J, Sesterhenn IA, Mostofi FK, Stanczyk FZ, Reichardt JK (2001) Polymorphic markers in the

- SRD5A2 gene and prostate cancer risk: a population-based case-control study. *Cancer Epidemiol Biomarkers Prev* 10: 1077-82
- Huang X, Li X, Guo B (2008) KLF6 induces apoptosis in prostate cancer cells through up-regulation of ATF3. *J Biol Chem* 283: 29795-801
- Hussain N, Chaghtai A, Herndon CD, Herson VC, Rosenkrantz TS, McKenna PH (2002) Hypospadias and early gestation growth restriction in infants. *Pediatrics* 109: 473-8
- Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. *Nat Genet* 29: 233-7
- Kalfa N, Liu B, Klein O, Wang MH, Cao M, Baskin LS (2008a) Genomic variants of ATF3 in patients with hypospadias. *J Urol* 180: 2183-8; discussion 2188
- Kalfa N, Liu B, Ophir K, Audran F, Wang MH, Mei C, Sultan C, Baskin LS (2008b) Mutations of CXorf6 are associated with a range of severities of hypospadias. *Eur J Endocrinol* 159: 453-8
- Källén B, Winberg J (1982) An epidemiological study of hypospadias in Sweden. *Acta Paediatr Scand Suppl* 293: 1-21
- Källén B, Bertollini R, Castilla E, Czeizel A, Knudsen LB, Martinez-Frias ML, Mastroiacovo P, Mutchinick O (1986) A joint international study on the epidemiology of hypospadias. *Acta Paediatr Scand Suppl* 324: 1-52
- Källén B (1988) Case control study of hypospadias, based on registry information. *Teratology* 38: 45-50
- Källén B, Mastroiacovo P, Lancaster PA, Mutchinick O, Kringelbach M, Martinez-Frias ML, Robert E, Castilla EE (1991) Oral contraceptives in the etiology of isolated hypospadias. *Contraception* 44: 173-82
- Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. *Am J Hum Genet* 56: 18-32
- Khanna M, Qin KN, Wang RW, Cheng KC (1995) Substrate specificity, gene structure, and tissue-specific distribution of multiple human 3 alpha-hydroxysteroid dehydrogenases. *J Biol Chem* 270: 20162-8
- Kim KS, Torres CR, Jr., Yucel S, Raimondo K, Cunha GR, Baskin LS (2004) Induction of hypospadias in a murine model by maternal exposure to synthetic estrogens. *Environ Res* 94: 267-75
- Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 76: 8-32
- Kojima Y, Hayashi Y, Mizuno K, Mogami M, Sasaki S, Kohri K (2002) Spermatogenesis, fertility and sexual behavior in a hypospadiac mouse model. *J Urol* 167: 1532-7
- Kondo KH, Kai MH, Setoguchi Y, Eggertsen G, Sjoblom P, Setoguchi T, Okuda KI, Bjorkhem I (1994) Cloning and expression of cDNA of human delta 4-3-oxosteroid 5 beta-reductase and substrate specificity of the expressed enzyme. *Eur J Biochem* 219: 357-63
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31: 241-7
- Kristensen VN, Kelefiotis D, Kristensen T, Borresen-Dale AL (2001) High-throughput methods for detection of genetic variation. *Biotechniques* 30: 318-22, 324, 326 passim
- Kurzrock EA, Jegatheesan P, Cunha GR, Baskin LS (2000) Urethral development in the fetal rabbit and induction of hypospadias: a model for human development. *J Urol* 164: 1786-92
- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265: 2037-48

- Lathrop GM, Lalouel JM, Julier C, Ott J (1985) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* 37: 482-98
- Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44: 397-401
- Liu B, Wang Z, Lin G, Agras K, Ebbers M, Willingham E, Baskin LS (2005) Activating transcription factor 3 is up-regulated in patients with hypospadias. *Pediatr Res* 58: 1280-3
- Lowry RB, Kliman MR (1976) Hypospadias in successive generations - possible dominant gene inheritance. *Clin Genet* 9: 285-8
- Makridakis NM, di Salle E, Reichardt JK (2000) Biochemical and pharmacogenetic dissection of human steroid 5 alpha-reductase type II. *Pharmacogenetics* 10: 407-13
- Mansfield TA, Simon DB, Farfel Z, Bia M, Tucci JR, Lebel M, Gutkin M, Vialettes B, Christofilis MA, Kauppinen-Makelin R, Mayan H, Risch N, Lifton RP (1997) Multilocus linkage of familial hyperkalaemia and hypertension, pseudohypoaldosteronism type II, to chromosomes 1q31-42 and 17p11-q21. *Nat Genet* 16: 202-5
- McCandless SE, Brunger JW, Cassidy SB (2004) The burden of genetic disease on inpatient care in a children's hospital. *Am J Hum Genet* 74: 121-7
- McKusick VA (1998) On the naming of clinical disorders, with particular reference to eponyms. *Medicine (Baltimore)* 77: 1-2
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* 304: 581-4
- Meizner I, Mashlach R, Shalev J, Efrat Z, Feldberg D (2002) The 'tulip sign': a sonographic clue for in-utero diagnosis of severe hypospadias. *Ultrasound Obstet Gynecol* 19: 250-3
- Monteleone Neto R, Castilla EE, Paz JE (1981) Hypospadias: an epidemiological study in Latin America. *Am J Med Genet* 10: 5-19
- Moreno LM, Arcos-Burgos M, Marazita ML, Krahn K, Maher BS, Cooper ME, Valencia-Ramirez CR, Lidral AC (2004) Genetic analysis of candidate loci in non-syndromic cleft lip families from Antioquia-Colombia and Ohio. *Am J Med Genet A* 125A: 135-44
- Morgan EA, Nguyen SB, Scott V, Stadler HS (2003) Loss of Bmp7 and Fgf8 signaling in Hoxa13-mutant mice causes hypospadias. *Development* 130: 3095-109
- Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7: 277-318
- Morton NE, Lindsten J, Iselius L, Yee S (1982) Data and theory for a revised chiasma map of man. *Hum Genet* 62: 266-270
- Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, et al. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616-22
- Nordenskjold A, Friedman E, Tapper-Persson M, Soderhall C, Leviav A, Svensson J, Anvret M (1999) Screening for mutations in candidate genes for hypospadias. *Urol Res* 27: 49-55
- North K, Golding J (2000) A maternal vegetarian diet in pregnancy is associated with hypospadias. The ALSPAC Study Team. *Avon Longitudinal Study of Pregnancy and Childhood. BJU Int* 85: 107-13
- Ntais C, Polycarpou A, Ioannidis JP (2003) SRD5A2 gene polymorphisms and the risk of prostate cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 12: 618-24
- Ogata T, Wada Y, Fukami M (2008) MAMLD1 (CXorf6): a new gene for hypospadias. *Sex Dev* 2: 244-50
- Ott J, Bhat A (1999) Linkage analysis in heterogeneous and complex traits. *Eur Child Adolesc Psychiatry* 8 Suppl 3: 43-6

- Ott J, Lathrop GM (1987) Goodness-of-fit tests for locus order in three-point mapping. *Genet Epidemiol* 4: 51-7
- Pagani F, Stuani C, Tzetis M, Kanavakis E, Efthymiadou A, Doudounakis S, Casals T, Baralle FE (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 12: 1111-20
- Page LA (1979) Inheritance of uncomplicated hypospadias. *Pediatrics* 63: 788-90
- Paulozzi LJ (1999) International trends in rates of hypospadias and cryptorchidism. *Environ Health Perspect* 107: 297-302
- Paulozzi LJ, Erickson JD, Jackson RJ (1997) Hypospadias trends in two US surveillance systems. *Pediatrics* 100: 831-4
- Percy AK (2000) Pelizaeus-Merzbacher disease: splice sites are nice sites for disease expression. *Neurology* 55: 1072-3
- Perriton CL, Powles N, Chiang C, Maconochie MK, Cohn MJ (2002) Sonic hedgehog signaling from the urethral epithelium controls external genital development. *Dev Biol* 247: 26-46
- Penning TM, Jin Y, Steckelbroeck S, Lanisnik Rizner T, Lewis M (2004) Structure-function of human 3 alpha-hydroxysteroid dehydrogenases: genes and proteins. *Mol Cell Endocrinol* 215: 63-72
- Pierik FH, Burdorf A, Deddens JA, Juttman RE, Weber RF (2004) Maternal and paternal risk factors for cryptorchidism and hypospadias: a case-control study in newborn boys. *Environ Health Perspect* 112: 1570-6
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. *Nature* 411: 199-204
- Roberts CJ, Lloyd S (1973) Observations on the epidemiology of simple hypospadias. *Br Med J* 1: 768-70
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-33
- Sadovsky Y, Dorn C (2000) Function of steroidogenic factor 1 during development and differentiation of the reproductive system. *Rev Reprod* 5: 136-42
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-7
- Santti R, Makela S, Strauss L, Korkman J, Kostian ML (1998) Phytoestrogens: potential endocrine disruptors in males. *Toxicol Ind Health* 14: 223-37
- Scott WK, Hauser ER, Schmechel DE, Welsh-Bohmer KA, Small GW, Roses AD, Saunders AM, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA (2003) Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22. *Am J Hum Genet* 73: 1041-51
- Sellick GS, Longman C, Tolmie J, Newbury-Ecob R, Geenhalgh L, Hughes S, Whiteford M, Garrett C, Houlston RS (2004) Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. *Nucleic Acids Res* 32: e164
- Sides D, Goldstein RB, Baskin L, Kleiner BC (1996) Prenatal diagnosis of hypospadias. *J Ultrasound Med* 15: 741-6

- Silver RI (2000) What is the etiology of hypospadias? A review of recent research. *Del Med J* 72: 343-7
- Silver RI, Rodriguez R, Chang TS, Gearhart JP (1999) In vitro fertilization is associated with an increased risk of hypospadias. *J Urol* 161: 1954-7
- Silver RI, Russell DW (1999) 5alpha-reductase type 2 mutations are present in some boys with isolated hypospadias. *J Urol* 162: 1142-5
- Skakkebaek NE, Rajpert-De Meyts E, Main KM (2001) Testicular dysgenesis syndrome: an increasingly common developmental disorder with environmental aspects. *Hum Reprod* 16: 972-8
- Stevenson DA, Carey JC (2004) Contribution of malformations and genetic disorders to mortality in a children's hospital. *Am J Med Genet A* 126A: 393-7
- Stokowski LA (2004) Hypospadias in the neonate. *Adv Neonatal Care* 4: 206-15
- Stoll C, Alembik Y, Roth MP, Dott B (1990) Genetic and environmental factors in hypospadias. *J Med Genet* 27: 559-63
- Sweet RA, Schrott HG, Kurland R, Culp OS (1974) Study of the incidence of hypospadias in Rochester, Minnesota, 1940-1970, and a case-control comparison of possible etiologic factors. *Mayo Clin Proc* 49: 52-8
- Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 3: 391-7
- Terwilliger JD, Goring HH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* 72: 63-132
- Terwilliger JD, Ott J (1993) A novel polylocus method for linkage analysis using the lod-score or affected sib-pair method. *Genet Epidemiol* 10: 477-82
- The International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431: 931-45
- Thigpen AE, Davis DL, Milatovich A, Mendonca BB, Imperato-McGinley J, Griffin JE, Francke U, Wilson JD, Russell DW (1992) Molecular genetics of steroid 5 alpha-reductase 2 deficiency. *J Clin Invest* 90: 799-809
- Vacherot F, Laaroubi K, Caruelle D, Delbe J, Barritault D, Caruelle JP, Courty J (1995) Upregulation of heparin-affin regulatory peptide by androgen. *In Vitro Cell Dev Biol Anim* 31: 647-8
- Wakefield J (2001) Bad news for boys: linking hypospadias and endocrine disruptors. *Environ Health Perspect* 109: A542-3
- Vanderwinden JM, Mailleux P, Schiffmann SN, Vanderhaeghen JJ (1992) Cellular distribution of the new growth factor pleiotrophin (HB-GAM) mRNA in developing and adult rat tissues. *Anat Embryol (Berl)* 186: 387-406
- Wang Q, Ghadessy FJ, Yong EL (1998) Analysis of the transactivation domain of the androgen receptor in patients with male infertility. *Clin Genet* 54: 185-92
- Wang Z, Liu BC, Lin GT, Lin CS, Lue TF, Willingham E, Baskin LS (2007) Up-regulation of estrogen responsive genes in hypospadias: microarray analysis. *J Urol* 177: 1939-46
- Watanabe M, Yoshida R, Ueoka K, Aoki K, Sasagawa I, Hasegawa T, Sueoka K, Kamatani N, Yoshimura Y, Ogata T (2007) Haplotype analysis of the estrogen receptor 1 gene in male genital and reproductive abnormalities. *Hum Reprod* 22: 1279-84
- Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44: 388-96
- Weidner IS, Moller H, Jensen TK, Skakkebaek NE (1999) Risk factors for cryptorchidism and hypospadias. *J Urol* 161: 1606-9
- Wennerholm UB, Bergh C, Hamberger L, Lundin K, Nilsson L, Wikland M, Kallen B (2000) Incidence of congenital malformations in children born after ICSI. *Hum Reprod* 15: 944-8

- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, et al. (2001) The sequence of the human genome. *Science* 291: 1304-51
- Williamson JA, Amos CI (1990) On the asymptotic behavior of the estimate of the recombination fraction under the null hypothesis of no linkage when the model is misspecified. *Genet Epidemiol* 7: 309-18
- Yucel S, Dravis C, Garcia N, Henkemeyer M, Baker LA (2007) Hypospadias and anorectal malformations mediated by Eph/ephrin signaling. *J Pediatr Urol* 3: 354-63
- Zhu M, Zhao S (2007) Candidate gene identification approach: progress and challenges. *Int J Biol Sci* 3: 420-7