UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 448

Molecular Genetic Studies of ALSG, Kostmann Syndrome and a Novel Chromosome 10 Inversion

MIRIAM ENTESARIAN

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2009

ISSN 1651-6206
ISBN 978-91-554-7493-5
urn:nbn:se:uu:diva-100598

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbecklaboratoriet, Dag Hammarskjölds väg 20, Uppsala, Friday, May 15, 2009 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract
Entesarian, M. 2009. Molecular Genetic Studies of ALSG, Kostmann Syndrome and a Novel Chromosome 10 Inversion. Acta Universitatis Upsaliensis. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 448. 57 pp. Uppsala. ISBN 978-91-554-7493-5.

In summary, this thesis presents the localisation and identification of genetic variants of which some are disease associated and some considered to be neutral. Knowledge of the basic mechanisms behind human disorders is important both from a biological and medical point of view.

The thesis is based on four papers of which the first two clarify the genetic basis of autosomal dominant aplasia of lacrimal and salivary glands (ALSG). ALSG is a rare disorder with high penetrance and variable expressivity characterized by dry mouth and eyes. In paper I, we located the ALSG gene to a 22 centiMorgan region on chromosome 5 through a genome-wide linkage scan with microsatellite markers in two families. Mutations were found in the gene encoding fibroblast growth factor 10 (FGF10) situated in the linked chromosome 5 region. Mice having only one copy of the *FGF10* gene (*Fgf10*$^{+/-}$ mice) have a phenotype similar to ALSG, providing an animal model for the disorder. In paper II, we describe two additional patients with ALSG and missense mutations in *FGF10*, providing further genotype-phenotype correlations.

The aim of paper III was to identify a gene involved in autosomal recessive severe congenital neutropenia (SCN), also referred to as Kostmann syndrome. The disease is characterized by a very low absolute neutrophil count and recurrent bacterial infections. Affected individuals from the family with SCN originally described by Dr Kostmann were genotyped with whole-genome SNP arrays. Autozygosity mapping identified a shared haplotype spanning 1.2 Mb on chromosome 1q22. This region contained 37 known genes, of which several were associated with myelopoiesis. Our finding contributed to the identification of the gene mutated in Kostmann syndrome.

In paper IV a cytogenetic inversion on chromosome 10 was mapped and characterized. Sequence- and haplotype analysis of carriers from four non-related Swedish families revealed identical inversion breakpoints and established that the rearrangement was identical by descent. A retrospective study of karyotypes together with screening of large sample sets established that the inversion is a rare and inherited chromosome variant with a broad geographical distribution in Sweden. No consistent phenotype was found associated with the inversion.

Genetic research increases the understanding of our genomes and makes it possible to discover variants contributing to disease. Identification of such genetic variants further enables studies of gene function and pathogenesis. The finding of the disease associated variants in this thesis will eventually contribute to improved diagnosis, prognosis, risk assessment and a future treatment of patients.

Keywords: aplasia of lacrimal and salivary glands, FGF10, kostmann syndrome, HAX1, inversion, chromosome 10

Miriam Entesarian, Department of Genetics and Pathology, Uppsala University, SE-75185 Uppsala, Sweden

ISSN 1651-6206
ISBN 978-91-554-7493-5
urn:nbn:se:uu:diva-100598 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-100598)

Luck is infatuated with the efficient.

Persian proverb

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I **Entesarian M**, Matsson H, Klar J, Bergendal B, Olson L, Arakaki R, Hayashi Y, Ohuchi H, Falahat B, Bolstad AI, Jonsson R, Wahren-Herlenius M and Dahl N. Mutations in the gene encoding fibroblast growth factor 10 are associated with aplasia of lacrimal and salivary glands. *Nat Genet* **37**, 125-128 (2005).

II **Entesarian M**, Dahlqvist J, Shashi V, Stanley CS, Falahat B, Reardon W and Dahl N. FGF10 missense mutations in aplasia of lacrimal and salivary glands (ALSG). *Eur J Hum Genet* **15**, 379-82 (2007).

III Melin M, **Entesarian M**, Carlsson G, Garwicz D, Klein C, Fadeel B, Nordenskjöld M, Palmblad J, Henter JI and Dahl N. Assignment of the gene locus for severe congenital neutropenia to chromosome 1q22 in the original Kostmann family from Northern Sweden. *Biochem Biophys Res Commun* **353**, 571-5 (2007).

IV **Entesarian M**, Carlsson B, Mansouri MR, Stattin E-L, Holmberg E, Golovleva I, Stefansson H, Klar J and Dahl N. A chromosome 10 variant with a 12 Mb inversion [inv(10)(q11.22q21.1)] identical by descent and frequent in the Swedish population. *Am J Med Genet A* **149A**, 380-386 (2009).

Related papers

Orlen H, Melberg A, Raininko R, Kumlien E, **Entesarian M**, Soderberg P, Pahlman M, Darin N, Kyllerman M, Holmberg E, Engler H, Eriksson U & Dahl N. SPG11 mutations cause Kjellin syndrome, a hereditary spastic paraplegia with thin corpus callosum and central retinal degeneration. *Am J Med Genet B Neuropsychiatr Genet* (2009).

Carlsson G, van't Hooft I, Melin M, **Entesarian M**, Laurencikas E, Nennesmo I, Trebinska A, Grzybowska E, Palmblad J, Dahl N, Nordenskjold M, Fadeel B & Henter J I. Central nervous system involvement in severe congenital neutropenia: neurological and neuropsychological abnormalities associated with specific HAX1 mutations. *J Intern Med* **264**, 388-400 (2008).

Contents

Abbreviations

ANC	Absolute neutrophil count
ALSG	Aplasia of lacrimal and salivary glands
bp	Base pair
CGH	Comparative genome hybridization
CNV	Copy number variation
DNA	Deoxyribonucleic acid
ELA2	Elastase 2
ENCODE	Encyclopedia of DNA elements
FGF10	Fibroblast growth factor 10
FGFR2	Fibroblast growth factor receptor 2
FISH	Fluorescence *in situ* hybridization
GCSF	Granulocyte colony stimulating factor
GFI1	Growth factor independent 1
HAX1	HS1-associated protein X1
HGMD	Human gene mutation database
LINE	Long interspersed nuclear element
LADD	Lacrimo auriculo dento digital
LOD	Logarithm of the odds
LTR	Long terminal repeat
mRNA	Messenger ribonucleic acid
ncRNA	Non-coding RNA
nm	Nanometer
OMIM	Online mendelian inheritance in man
PCR	Polymerase chain reaction
rRNA	Ribosomal RNA
SCN	Severe congenital neutropenia
SINE	Short interspersed nuclear element
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
TE	Transposable element
tRNA	Transfer RNA

Introduction

The human genome

History of genetics

Gregor Mendel is considered by many as the founder of genetics. He was born in 1822 and resided as an Augustinian monk at St Thomas Monastery near Brünn, Austria (now Brno, Czech republic)[1]. Between the years 1856 and 1863 he conducted experiments in plant hybridization in the Monastery's garden[2]. He studied transmission of dominant and recessive characters in the annual garden pea, *Pisum*. A trait is dominant if it is expressed in the heterozygote and recessive if it is not. In 1866 Mendel published the paper "Experiments on Plant Hybrids" which contains statistical analysis of his hybridization data and mathematical models of the laws of heredity[2]. The two laws of heredity that he formulated (the law of segregation and the law of independent assortment) are today known as Mendel's laws. In 1944 Oswald Avery, Colin McLeod and MacLyn Mc Carty demonstrated through studies of bacterial transformation that genes are made up of the chemical substance deoxyribonucleic acid (DNA)[3]. James Watson and Francis Crick proposed the double helix structure of DNA in 1953[4]. Their model of specific base pairing suggested a possible replication mechanism for the genetic material.

Chromosome structure

With the exception of some terminally differentiated cells, all cells in the human body contain DNA. Most of the cell's DNA is present in the nucleus and a small amount is present in the mitochondria. Our DNA holds genetic instructions that are fundamental for development and organ functioning. A human cell contains about 2 meters of DNA[5], which is efficiently packed and condensed into the three dimensional structure of chromosomes (Figure 1). The middle of the chromosome is called the centromere and the regions of repetitive DNA at the distal tips are called telomeres. Each chromosome has two arms, named p (the short arm) and q (the long arm). The mixture of DNA and proteins comprising the chromosomes is called chromatin. The primary structure of chromatin consists of a fiber of 10 nm in diameter which resembles beads on a string[6]. The beads are called nucleosomes

containing octamers of histones onto which 147 base pairs (bp) of DNA is wrapped around[6]. Adjacent nucleosomes are connected by a short segment of spacer DNA. Mobilization and remodelling of nucleosomes are important for processes such as DNA replication, recombination, repair and transcription[7]. The nucleosomes are coiled into a condensed chromatin fiber called the 30 nm fiber[6]. Biochemical and electron micrograph studies indicate that the nucleosomes are arranged in a zigzag manner within the 30 nm fiber[7]. The chromatin is then even further compacted or "supercoiled" through the folding in the chromosomes. Regions of highly condensed chromatin are called heterochromatin whereas the open, extended conformation is known as euchromatin.

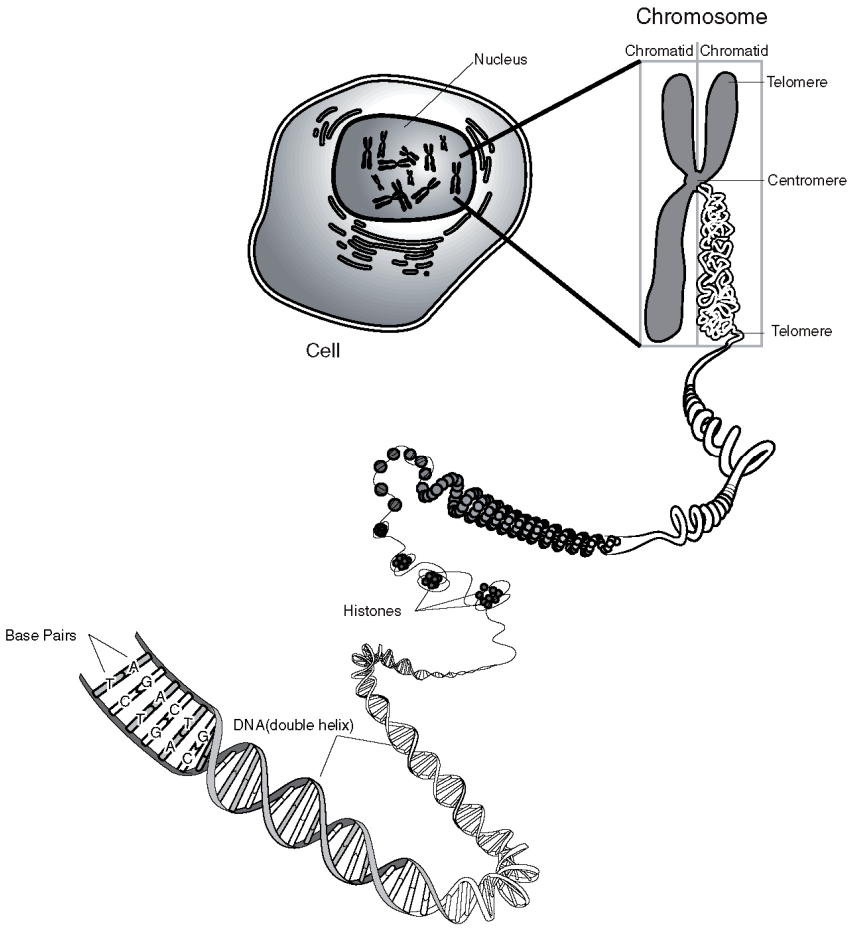

Figure 1. Packaging of DNA in the cell. The DNA in the nucleus is condensly packed at different levels (Illustration from Talking glossary of Genetic Terms, National Human Research Institute (NHGRI)).

The human karyotype

Most cells of our body are diploid and the normal human karyotype contains 22 pair of autosomes (nonsex chromosomes) and one pair of sex chromosomes (either XY or XX) (Figure 2). Before a cell divides the number of chromosomes is duplicated through DNA replication. There are two types of cell division: mitosis and meiosis. Mitosis is the process in which a cell separates the replicated chromosomes, into two identical sets in two daughter nuclei. Meiosis is necessary for sexual reproduction and involves two cell divisions to separate the replicated chromosomes into four haploid gametes. Pairing and genetic recombination between homologous chromosomes occur during meiosis. Sometimes chromosomal abnormalities occur in cells. These abnormalities can be divided into two groups: numerical and structural abnormalities. Numerical chromosomal abnormalities include gain or loss of entire chromosomes, caused by malsegregation during mitosis or meiosis. Structural chromosomal abnormalities involve one or several chromosomes with deviant shape and are caused by incorrect repair of chromosome breaks or recombination of mispaired chromosomes. Examples of structural chromosomal abnormalities are deletions, duplications, translocations, insertions and inversions.

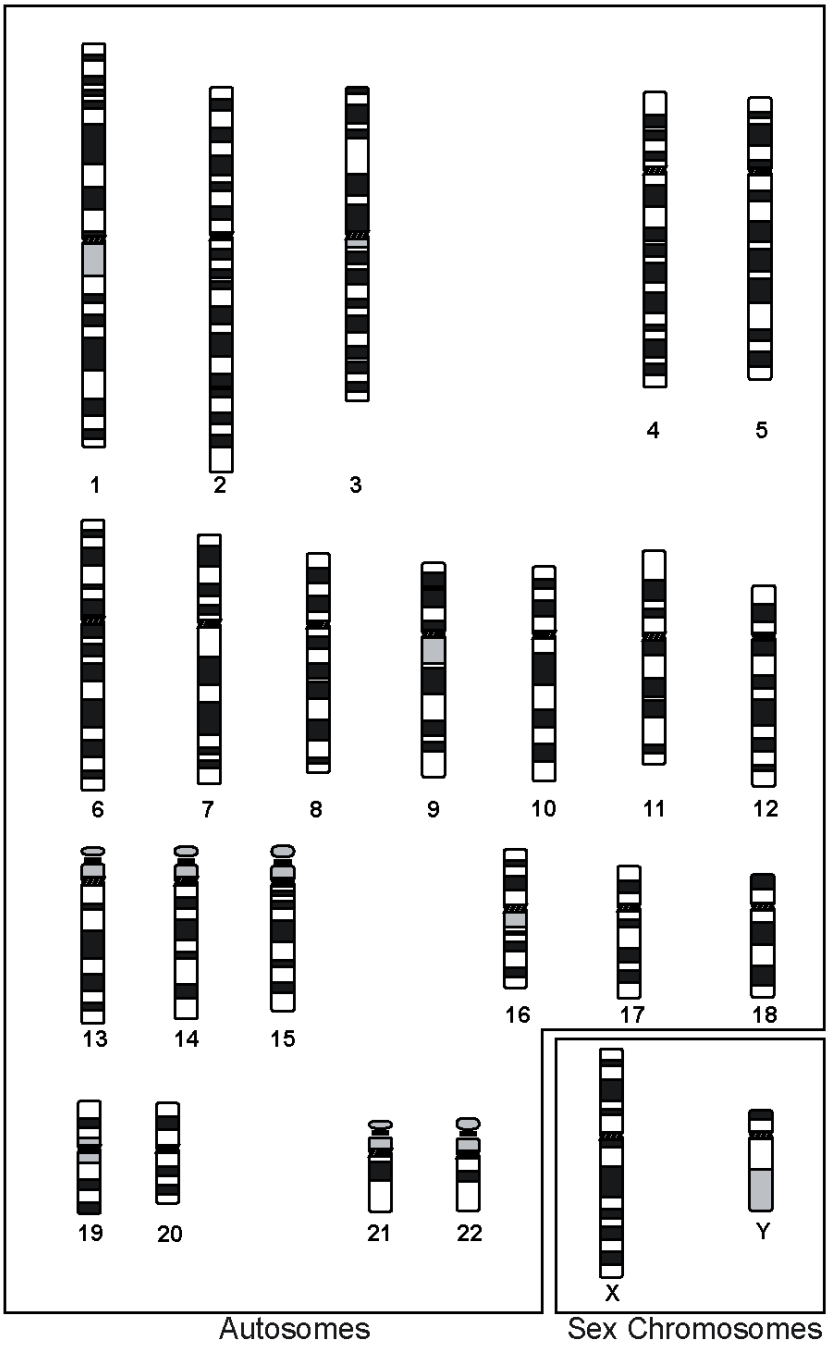

Figure 2. Human karyotype. Humans have 22 pair of autosomes and one pair of sex chromosomes. DNA staining enables studies of the number, type, shape and banding of the chromosomes (Illustration from Talking glossary of Genetic Terms, National Human Research Institute (NHGRI)).

Ribonucleic acid

The human genome is pervasively transcribed, which means that the majority of its bases can be found in primary transcripts[8]. Transcription is the synthesis of ribonucleic acid (RNA) from DNA by the enzyme RNA polymerase. RNA transcripts have many different roles in the cell. They can be biologically active molecules [eg ribosomal RNA (rRNA) or transfer RNA (tRNA)] or be involved indirectly by encoding other active molecules [eg messenger RNA (mRNA)][8]. Mammalian genes are commonly composed of several relatively small exons that are interspersed between much longer stretches of non-coding DNA, introns. To accurately identify and join together RNA sequences that code for proteins, the introns must be excised from the pre-mRNA and the exons joined together. This is called splicing and is catalyzed by the spliceosome which is a large RNA-protein complex. Small nuclear RNAs (snRNAs) guide splicing of pre-mRNAs[9]. Some genes can be alternatively spliced, which means that they can use different sets of splice junction sequences to produce alternative transcripts. This gives rise to different protein isoforms from the same gene. Alternative splicing is responsible for much of the complexity of the proteome[10]. Translation is the process in which a protein is synthesized through the assembly of amino acids using the information in mRNA as a template. Ribosomes are complexes of proteins and rRNAs involved in translation. The rRNAs have a catalytic function and provide a mechanism for decoding mRNA into amino acids and also interact with the tRNAs during translation. The tRNAs guide amino acids to the ribosomal site of protein synthesis during translation so that a polypeptide is assembled.

Although there is a broad amount of transcription across the human genome, a significant portion of the transcriptome has little or no protein-coding capacity[11]. Transcripts which are not translated are called non-protein coding RNAs (ncRNAs)[8]. These include structural RNAs (as the aforementioned snRNAs, rRNAs and tRNAs) and more recently discovered regulatory RNAs [eg microRNA and small interfering RNA (siRNA)]. MicroRNAs and siRNAs down-regulate expression of genes; microRNAs mostly by binding to their mRNAs and thereby inhibiting their translation and siRNAs through targeting mRNAs for degradation[9]. Twenty to thirty per cent of animal mRNAs are considered to be targets of post-transcriptional gene regulation by microRNAs, and individual microRNAs often have more than 100 targets[12]. Pseudogenes are considered to be non-functional copies of genes, some of which are transcribed[8]. Pseudogenes can arise by gene duplication or retroposition[13]. Nonprocessed pseudogenes contain sequences corresponding to exons and introns whereas processed pseudogenes contain the exonic sequences of an active gene. Pseudogenes drift from their ancestral sequence more slowly than expected by chance, implying some pseudogene sequences are under evolutionary selection to retain the ability

to produce antisense transcripts targeting their cognate genes[14]. Another kind of ncRNAs described in literature are long non-coding RNAs. These ncRNAs are longer than 200 nucleotides and can regulate gene-expression through a diversity of mechanisms, including chromatin modification, transcription and post-transcriptional processing[11]. There are probably many more RNA species yet to be discovered and classified.

Gene definition

The human genome project was initiated in 1990 with the goal of obtaining a very accurate sequence of the euchromatic portion of the human genome serving as a permanent foundation for biomedical research[15]. The sequencing of the human genome has proceeded in phases. The current genome sequence consists of 3.09 Gbp (NCBI build 36.3) and encodes approximately 20,000-25,000 protein-coding genes[15]. The concept of the gene has been modified many times during the last century. The classical view of a gene as a discrete element in the genome (each coding for one protein) has been challenged by non-coding RNA and the ENCyclopedia Of DNA Elements (ENCODE) project. The ENCODE project characterizes the transcriptional activity and regulation of the genome using tiling arrays[16] and was launched in 2003 with the intention to indentify all functional elements in the human genome[17]. Aligning and comparing sequences makes it possible to identify conserved sequence signatures and enrich for coding and noncoding functional regions[18]. Sequence conservation across large evolutionary distances is often associated with a functional role. The new findings give a much more complex picture of the human genome and create the need for an updated gene definition. A proposed updated definition of a gene is: “A gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products”[16]. With this new definition, genotype still determines phenotype which means that the DNA sequence determines the sequence of the functional molecule. In the simplest case where the gene is continuous or there are no overlapping products, one DNA sequence still codes for one protein or RNA molecule. According to the new definition, different RNAs or proteins which overlap in their usage of DNA sequence can be considered as one gene.

Interspersed repetitive DNA

Interspersed repeats, also known as transposable elements (TEs), are mobile DNA sequences found throughout the genome. TEs were first discovered by Barbara McClintock through her studies of corn in the 1940s and 1950s[19]. Transposable elements constitute large fractions of most eukaryotic genomes, composing nearly 50% of the human genome[20]. TEs affect the genome by their ability to move and replicate. The high density of TEs

jeopardizes our genome by causing genomic mutations and genomic alterations when inserted. They can also cause genomic rearrangements through recombination between nonallelic homologous TE sequences[21]. TEs can also act on neighbouring genes by altering splicing and polyadenylation patterns, or by working as enhancers or promoters[20]. TEs can be classified by the presence or absence of an RNA transposition intermediate[22]. They can also be classified according to their degree of mechanistic self-sufficiency[23]. Autonomous transposable elements produce all the proteins that are required for transposition. Non-autonomous transposable elements are dependent on the proteins produced by autonomous elements of the same element family to transpose. Class I TEs all transpose via an RNA intermediate[22]. The original transposon is maintained *in situ*, where it is transcribed. Its RNA transcript is then reverse transcribed into DNA and integrated into a new genomic position. Class I can be divided into five orders: long terminal repeat (LTR) retrotransposons, DIRS-like elements, Penelope-like elements, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs)[22]. LTR retrotransposons range from a few hundred base pairs up to, sometimes, 25 kb. They have flanking LTRs from a few hundred base pairs up to 5 kb in size. LINEs can reach several kilobases in length and do not have LTRs. SINEs are small (80-500 bp) and non-autonomous. Class II TEs (DNA transposons) move through a cut and paste mechanism.

DNA variation in humans

Genetic variation

New estimates of interchromosomal difference in humans reveal that only 99.5% similarity exists between the two chromosomal copies of an individual[24]. Differences in our genome are called genetic variations. Genetic variants with a population frequency above 1% are considered as polymorphisms. Variation in the human genome is present in many forms (Figure 3). The most common polymorphism in the human genome is the single nucleotide polymorphism (SNP). Almost all SNPs are diallelic and the alleles are referred to as the “major allele” or “minor allele” based on their observed population frequency. On average there is one SNP every 200 bp in the human genome, although many of them are rare[25]. Most SNPs reside outside the coding regions of genes or in intergenic regions. There are approximately 6 million common SNPs (minor allele frequency of 5-20%) in the human genome[25]. The International HapMap Project was launched in 2002 with the aim of providing a public resource of common genetic diversity to accelerate medical genetic research[26]. By mapping and understanding the patterns of common genetic diversity in the human genome the search for genetic causes of human disease can be facilitated[27].

HapMap currently contains SNP genotypes and common haplotypes from eleven different populations (http://www.hapmap.org).

Microsatellites are another type of highly abundant polymorphism in eukaryotic genomes[28]. They have high levels of heterozygosity and are present at more than 100 000 regions in the genome[29]. Microsatellites are tandem repeats of 1-6 bp motifs. Microsatellites are highly polymorphic and have a high mutation rate (10^{-4} to 10^{-2} per locus per generation) due to polymerase template slippage during DNA replication of adjacent repeat motifs[28].

Besides SNPs and microsatellites there is also structural variation in the human genome. Structural variants are defined as genomic alterations that involve DNA segments larger than 1 kb[30]. Structural variation can be divided into microscopic and submicroscopic variants. Variants which are 3 Mb or larger in size are considered as microscopic structural variants while submicroscopic variants range from 1 kb to 3 Mb in size[30]. Microscopic structural variants can be identified through chromosome banding or fluorescence *in situ* hybridization (FISH) while submicroscopic variants demands for techniques with higher resolution like comparative genome hybridization (CGH) arrays or whole-genome SNP arrays. Examples of structural variants are segmental duplications and copy-number variants (CNVs). A segmental duplication is defined as duplication of a DNA segment longer than 1 kb with >90% sequence identity[31]. Segmental duplications constitute approximately 4% of the human genome[31]. A CNV is a DNA segment of 1 kb or more present at a variable copy number in comparison with a reference genome[30]. CNVs can be insertions, deletions or duplications. CNVs can cause disease if the different copy number of a gene influences the quantity of the gene product[29]. Studies of CNVs in mice show that expression levels of genes within CNVs tends to correlate with copy number changes and that the CNVs also have an effect on nearby genes[32]. A population analysis of large CNVs established that variants larger than 500 kb are present in 5-10% of individuals, and variants greater than 1 Mb in 1-2%[33]. The same study found that rare CNVs contain more genes than common ones and that homozygous deletions are especially gene poor. This is consistent with the theory that large CNVs are deleterious in relation to both their size and gene content[33].

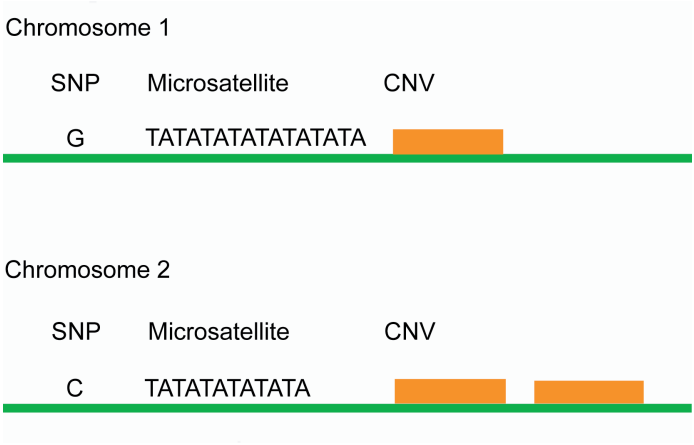

Figure 3. Examples of human genetic variation. Chromosome 1 has the G allele for the SNP while chromosome 2 has the C allele. Chromosome 1 has a longer microsatellite (TA-repeat) allele than chromosome 2. In chromosome 2 the orange sequence is duplicated, compared to chromosome 1.

Whole genome sequencing

The development of new sequencing techniques like the 454 pyrosequencing technology, Solexa sequencing-by-synthesis technology and the Supported Oligonucleotide Ligation and Detection (SOLiD) platform technology can make whole genome sequencing a standard component of biomedical research and patient care[34]. Whole genome sequencing of human genomes will also give a clearer picture of the genetic variation in our genomes and allows for detection of rare variants. The new techniques have improved speed, accuracy, efficiency and cost-effectiveness compared to Sanger sequencing[35]. They allow for cloning-free amplification, and the use of single-molecule templates enables the detection of heterogeneity in a DNA sample[36]. A limitation with the next-generation sequencing techniques are the short read lengths which are difficult to align if the read is repeated elsewhere in the genome or if the read harbours variations compared to the reference sequence[35].

Different forms of mutations

Mutations are changes in the nucleotide sequence of human DNA which can be responsible for both normal DNA variation of our species and disease[37]. Over the past decades the development of novel DNA technologies has led to a great advance in the analysis and diagnosis of inherited human disorders through the cloning of many new disease genes. Various types of mutations have been detected and characterized in a large number of different human genes[37]. Databases of mutations that summarize information and contain or

point to original or other sources of data play a relevant role in research as well as in diagnostic and general health care[38]. There are two large databases that record mutations associated with human phenotypes: Online Mendelian Inheritance in Man (OMIM)[39] (http://www.ncbi.nlm.nih.gov/omim) and the Human Gene Mutation Database (HGMD)[40] (www.hgmd.org). McKusick's Online Mendelian Inheritance in Man, a knowledgebase of human genes and phenotypes, originated with a book, *Mendelian Inheritance in Man* published in 1966[39]. The content of OMIM is based exclusively on published biomedical research literature and is updated daily[39]. HGMD contains data from more than 500 different life-science and medical journals, of which *Human Mutation* and *The American Journal of Human Genetics* make up the major part[40]. Single base pair substitutions (missense, nonsense, splicing and regulatory mutations) represent the majority (68%) of mutations recorded in the HGMD as of February 6th, 2009 (Professional version). Of these, missense nucleotide substitutions comprise 45% of the total entries. Small deletions (16.5%) and small insertions (6.5%) are also frequent. Gross lesions (deletions, insertions and duplications), repeat variations and complex rearrangements are also found in the database.

The correct classification of mutations is important for understanding the structure-function relationships in the affected protein, for estimating the phenotypic risk in individuals with familial disease predispositions and for developing new therapies[10]. For most genes the correct coding of each nucleotide in the DNA sequence is essential for the correct assembly, property and function of the gene product. Single nucleotide substitutions in the coding regions that do not alter splice site consensus sequences are generally considered as missense, nonsense or silent mutations. Single nucleotide substitutions that occur in introns and that affect the classical consensus splice-site signals are regarded as splicing mutations[10]. Splicing mutations can induce exon skipping, intron retention, activation of cryptic splice sites or alter the balance of alternatively spliced isoforms and thereby cause disease phenotypes[41]. Mutations in regulatory regions can also affect transcription, splicing and translation. Mutations that modify the pre-mRNA secondary structure can alter the display of target RNA sequences and thereby have an affect on the splicing efficiency[41]. Nonsense mutations are commonly assumed to produce truncated proteins, whereas missense mutations are presumed to identify functionally or structurally important amino acids[10]. Small insertions and deletions can cause frameshift mutations, which alter the normal translational reading frame if they are not a multiple of three. Premature termination codons, whether they arise from nonsense or frameshift mutations or as a result of exon skipping usually trigger nonsense-mediated mRNA decay[10]. This cellular mechanism ensures mRNA quality by degrading mRNAs that contain a premature termination codon.

Epigenetics

Epigenetics is the study of heritable changes in gene expression that do not involve a change in DNA sequence. The changes can be transmitted through meiosis or mitosis to daughter and progeny cells. Examples of epigenetic mechanisms are DNA methylation, imprinting and changes in chromatin conformation.

In mammals cytosine methylation is almost exclusively limited to CpG dinculeotides[7]. About 80% of CpG dinucleotides in mammals are methylated at the carbon atom 5 of cytosine. Unmethylated CpG dinucleotides are mainly found in CpG islands close to promoters[42]. Cytosine methylation patterns have roles in many different processes including development and silencing of parasitic elements and at least a portion of cytosine methylation is heritable[6]. After DNA replication the newly synthesized strand will receive the same CpG methylation pattern as the parental strand, making it possible to transmit the methylation pattern to the daughter cell.

Genomic imprinting involves differences in allele expression according to parent of origin. Either the maternally or paternally inherited allele is repressed. This leads to unequal expression of the maternal and paternal alleles for a diploid locus[42]. An important factor in maintaining the imprinted status during cell division is allele-specific DNA methylation. In addition to serving as an allelic mark to distinguish parental alleles, DNA methylation can also repress transcription and play a role in allele-specific silencing of imprinted genes[42].

Chromatin conformation can also affect gene transcription. In general, acetylation of the nucleosomal histones is associated with unfolding and accessibility of chromatin making it transcriptionally active[42]. In transcriptionally active chromatin the gene promoter regions are typically unmethylated. Deacetylation of histones promotes repression of gene expression through condensation of chromatin. Condensed and transcriptionally inactive chromatin is often also associated with DNA methylation.

Monogenic disorders

Monogenic disorders, also called Mendelian disorders, are caused by mutation in a single gene and follow the laws of inheritance described by Gregor Mendel. There are five basic Mendelian inheritance patterns (Figure 4):

Autosomal dominant disorders affect both males and females and can be transmitted through either sex. Affected individuals are seen in each generation and these are heterozygous for the mutated allele. Examples of

autosomal dominant disorders are achondroplasia[43] and Huntington´s disease[44].

Autosomal recessive disorders also affect either sex. Affected individuals are usually born to healthy parents who are carriers. Only individuals homozygous for the mutated allele are affected. Consanguinity increases the risk of a recessive disorder. Cystic fibrosis[45], phenylketonuria[46] and Kjellin syndrome[47] are examples of recessive disorders.

X-linked recessive disorders affect mainly males and the disorder is transmitted through healthy female carriers. None of the offspring of an affected male are affected but all his daughters are obligate carriers. Anhidrotic ectodermal dysplasia[48] and hemophilia[49] are inherited in an X-linked recessive fashion.

An X-linked dominant disorder affects either sex, but more females than males. An affected male passes the disorder to all his daughters. An example of X-linked dominant disorder is vitamin-D resistant rickets[50].

In Y-linked disorders only males are affected, with transmission from father to son. There are probably no Y-linked disorders apart from disorders of male sex determination[51] and function[52].

Most Mendelian disorders are rare but since they are so many different ones, the monogenic phenotypes have a tremendous power in helping to classify and understand human diseases[53]. The identification of genes involved in human disease in combination with the study of their regulation as well as their pathogenic mutations help elucidate their function. This also enables better diagnosis and treatment for the patients. Many disorders that were initially characterized as being monogenic have later proven to be either caused or modified by a small number of loci[54]. There probably exists a conceptual continuum between classical Mendelian disorders and complex traits[54]. In complex traits both genes and environment are involved. A mutant gene that gives rise to a monogenic disorder can contribute to the understanding of a similar complex disease. One example is the discovery of mutations in the adenomatous polyposis coli gene in hereditary colon cancer, which led to the discovery of mutant alleles of additional genes that also cause or predispose to colon cancer[53]. Common genetic variations detected in genes causing rare genetic disorders (often severe and early-onset types) have been found to be risk factors for common diseases with similar phenotypes. One example is autosomal dominant mutations in six genes causing maturity-onset diabetes of the young, where variations in the same genes are associated with the susceptibility to type 2 diabetes[29].

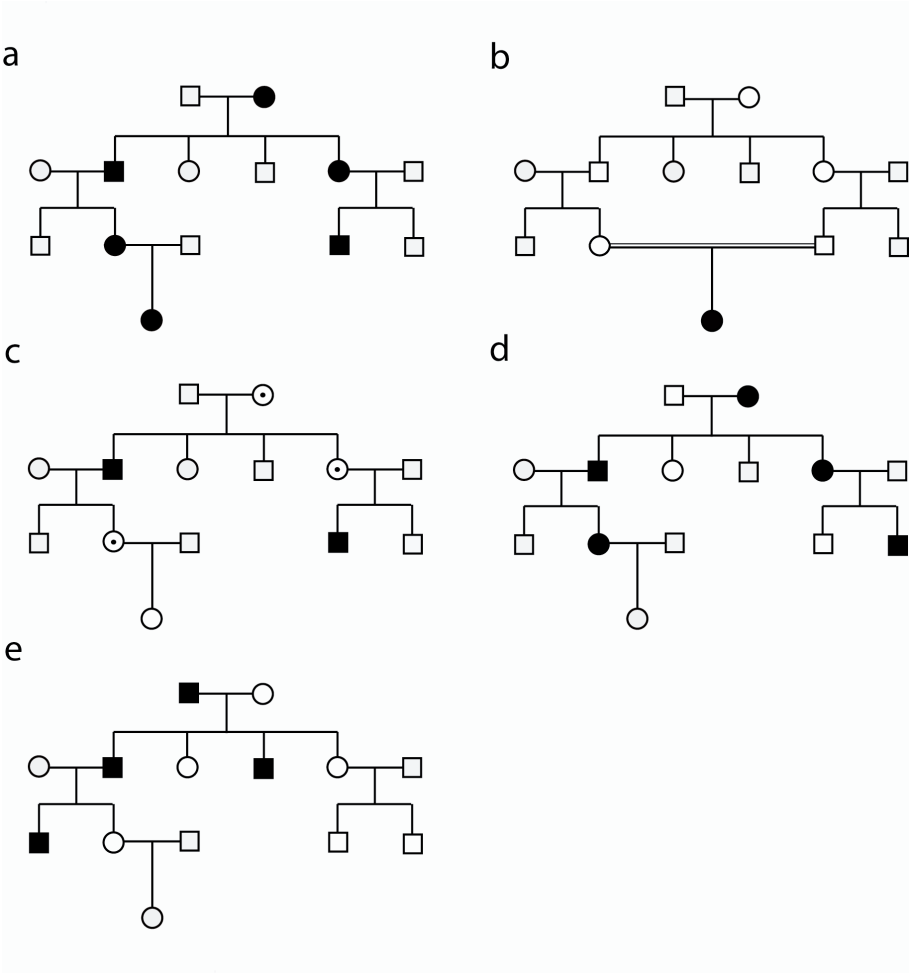

Figure 4. Mendelian inheritance patterns. Squares depict males; circles depict females. Filled symbols denote affected individuals; open symbols denote unaffected individuals; spotted symbols denote carriers. (a) Autosomal dominant. (b) Autosomal recessive. (c) X-linked recessive. (d) X-linked dominant. (e) Y-linked.

Methods

Linkage analysis

Genetic markers can be used to map an inherited disorder to a specific chromosomal location. The inheritance of markers in healthy and affected members of a family with the disorder is then studied. Polymorphic microsatellites (di-, tri- or tetranucleotide repeats) and single nucleotide polymorphisms (SNPs) can be used as genetic markers. These are numerous in the human genome, often informative and easy to genotype. If a marker is close enough to the disease locus it will be inherited together with it, because of the low chance of recombination between the loci. This phenomenon is called linkage. In the case of an autosomal recessive disorder you search for a marker for which the affected individuals are homozygous and all the carriers are heterozygous. When mapping an autosomal dominant disorder, a marker allele present in all the affected family members but not in the healthy is required for linkage. There are statistical tests for linkage that use assumptions such as mode of inheritance, allele frequencies and penetrance[54]. Logarithm of the odds (LOD) score analysis is a popular and efficient way to analyze pedigrees for linkage. The method originated in an article by Morton in 1955[55]. LOD score (Z) is a measure of the likelihood of genetic linkage between two loci. Two loci are linked if they are inherited together more often than expected by chance. LOD scores are a function of the recombination fraction (θ). The best estimate of θ is that which maximises the lod score function, the maximum lod score[56]. Positive LOD scores suggest linkage while negative LOD scores suggest no linkage. Typically, a $Z = 3.0$ is the threshold for accepting linkage with a 5% chance of error ($Z = 3.0$ corresponds to 1000:1 odds that two loci are linked) whereas linkage can be excluded if $Z < -2.0$. Values between -2.0 and 3.0 are considered inconclusive.

Polymerase chain reaction

The polymerase chain reaction (PCR) was invented by Dr Kary Mullis[57] for which he was rewarded with the Nobel Prize in chemistry 1993. PCR is probably the most widely used molecular genetic technique. It has numerous applications and almost all molecular tools include a PCR step in one way or

another. The technique is robust, sensitive and rapid. It allows selective amplification of a specific target sequence, and needs only small amounts (in the nanogram range) of sample DNA. PCR requires template DNA, buffer, deoxynucleotides, oligonucleotide primers complementary to the sequence of interest and DNA polymerase. The usage of two primers permits exponential amplification. PCR is often composed of three steps repeated 25-35 times in a thermal cycler. The first step is DNA denaturation at about 95°C which allows dissociation of double-stranded DNA to single-stranded DNA. In the second step the primers are allowed to anneal to the single-stranded DNA, usually at temperatures between 50°C and 60°C. In the final step DNA synthesis takes place at a temperature of approximately 72°C, which is optimal for a heat-stable DNA polymerase. In theory, the number of DNA molecules is doubled after each cycle.

Sanger sequencing

Sequencing of PCR products can be done with cycle sequencing. This approach is similar to regular PCR but with two differences, only one primer is used and the reaction includes a fraction of dideoxynucleotides (which lack the 3′ hydroxyl group). Dideoxynucleotides are terminators of chain extension, which means that when they are incorporated into the growing strand no further synthesis is possible. DNA sequencing with chain-terminating inhibitors was first described by Sanger et al in 1977[58]. The four types of dideoxynucleotides can be labelled with different fluorophores. Cycle sequencing then gives a mixture of fragments of different lengths and fluorophores. These fragments are separated according to size by gel or capillary electrophoresis. During this separation, a laser excitation causes the four fluorophores to emit light of different wavelengths detected by a monitor. The output is a chromatogram displaying the sequence of individual bases with their corresponding colour code. Two common applications of DNA sequencing are SNP genotyping and mutation screening.

Mouse models

The mouse has been used as a model organism in biomedical research since the early 1900s, serving as a powerful tool in clarifying the genetic basis of human physiology and pathogenesis[59]. One reason for the use of the mouse as a model is that there are relatively small genomic differences between human and mice. They also exhibit inherited disorders, both Mendelian and polygenic[59]. Other advantages with using mouse models include short generation times and easy maintenance. Mouse models are useful in the study of gene expression and function, understanding of human disorders,

and drug development. It is however important to keep in mind that human and mice are different in many ways. A good mouse model should nonetheless resemble the human disease phenotype in most aspects to confirm that the correct gene has been identified. To prove that a gene is involved in a human disorder, a transgenic mouse model or knockout mouse can be constructed. A transgenic animal contains artificially introduced exogenous DNA, whereas a knockout mouse has a targeted gene inactivation. Three major mouse knockout programs (KnockOut Mouse Project, European Conditional Mouse Mutagenesis Program and North American Conditional Mouse Mutagenesis Project) are ongoing worldwide, working together to mutate all protein-encoding genes in the mouse and so far about 9000 mouse genes in total have been knocked out[60]. Mario R Capecchi, Martin J Evans and Oliver Smithies received the Nobel Prize in medicine or physiology 2007 for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells.

Autozygosity mapping with SNP arrays

Autosomal recessive disease gene loci can be identified by autozygosity mapping in consanguineous families where segregation of a founder mutation is expected[61]. One then searches for a chromosomal region where all the affected individuals are homozygous for an allele identical by descent. Whole-genome SNP genotyping arrays are an efficient and rapid method for autozygosity mapping. These are high-density synthetic oligonucleotide microarrays developed to generate large quantities of genetic information in a single experiment. Through a one-primer assay requiring only a small amount of DNA, hundreds of thousands SNPs can be genotyped simultaneously[62,63].

Fluorescence *in situ* hybridization

Metaphase or interphase chromosomes can be visualised in a fluorescence microscope by DNA-binding fluorescent agents[64]. FISH is a powerful cytogenetic technique in which a labelled DNA probe is hybridized to a microscope slide prepared with e.g. denatured metaphase chromosomes. The results are then scored by eye using a fluorescence microscope. FISH has high sensitivity and resolution compared to e.g. Giemsa (G) banding. It enables both chromosome classification and detection of structural and numerical chromosome aberrations[65]. Bacterial artificial chromosomes (BACs) with human inserts or PCR products in the 3-5 kb range (mini-FISH)[66] can be used as probes for FISH analysis.

Southern blot hybridization

Southern blotting is an excellent method to test for DNA rearrangements in the 1-25 kb range. The method is named after its inventor Edwin Southern[61]. Genomic DNA is digested with one or several restriction endonucleases, size-fractionated by agarose gel electrophoresis, denatured and transferred by capillary blotting to a nitrocellulose or nylon membrane. The membrane containing the DNA is subsequently hybridized with a radioactively or fluorescently labelled probe specific for the DNA region of interest. After washing, the pattern of hybridization is visualized on X-ray film by autoradiography, or in the case of a fluorescently labelled probe by a fluorescence scanner. By simultaneous use of patient and control samples, restriction fragments altered by DNA rearrangements can be identified.

Aims of the thesis

Paper I	To map autosomal dominant aplasia of lacrimal and salivary glands (ALSG) to a chromosomal region. To identify, evaluate and screen candidate genes in the linked region for mutations. To search for mutations in the gene causing ALSG in individuals with similar or overlapping phenotypes, such as dry eyes and/or dry mouth. To study the phenotype of mice heterozygous for the gene mutated in ALSG.
Paper II	To analyze the mutation spectrum in the *FGF10* gene in patients with ALSG. To predict how these mutations cause disease.
Paper III	To identify the chromosomal region for Kostmann syndrome by autozygosity mapping. To identify and evaluate potential genes in the candidate region.
Paper IV	To map and characterize an inversion of chromosome 10 and establish its frequency in the Swedish population. To analyze the breakpoint sequences and identify possible genes disrupted by the inversion breakpoints. To develop a PCR based assay to screen other populations for the same rearrangement.

Aplasia of lacrimal and salivary glands

Background

Aplasia of lacrimal and major salivary glands (ALSG; OMIM 180920) is a rare congenital disorder. Familial occurrence of absent salivary glands was first described by Ramsey in 1924[67]. The following year, Blackmar reported the first case of congenital absence of salivary glands associated with atresia of lacrimal puncta, aptyalism (deficiency or absence of the saliva) and decreased lacrimation[68]. ALSG is inherited as an autosomal dominant trait with variable expressivity[69]. Aplasia of the major salivary glands and lacrimal glands may be associated with absence of lacrimal puncta (Figure 5). The clinical manifestations vary, but involvement of the lacrimal glands results in irritable eyes and recurrent infections as well as epiphora (constant tearing) if the nasolacrimal ducts or lacrimal puncta are missing. Aplasia or hypoplasia of the major salivary glands causes xerostomia (dryness of the mouth) and increases the risk of dental erosion and dental caries[70]. Other complications include periodontal diseases, oral soft tissue inflammation and disorders of smell, chewing and swallowing[69]. The incidence of aplasia/hypoplasia of lacrimal and/or major salivary glands is unknown and difficult to estimate since symptoms in many cases are considered mild. Typically, the patients are first observed and diagnosed by a dentist or by an ophthalmologist. Individuals affected with ALSG are sometimes confused or misdiagnosed with the more prevalent disorder Sjögren's syndrome, an autoimmune disorder affecting exocrine glands which is characterized by keratoconjunctivitis sicca and xerostomia[71].

When we started our genetic analysis of ALSG, no previous linkage analyses or molecular investigations had been performed on ALSG patients, although several sporadic cases and families had been described in the literature[69,70,72-75]

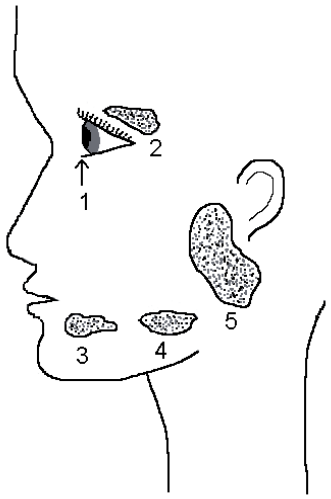

Figure 5. The lacrimal apparatus and the three major salivary glands. 1, lacrimal punctum; 2, lacrimal gland; 3, sublingual gland; 4, submandibular gland; 5, parotid gland.

Paper I: Subjects

We have identified two Swedish multi-generation families affected by autosomal dominant ALSG with no additional abnormalities. The clinical examination consisted of oral examination, ophthalmologic examination and magnetic resonance imaging (MRI). Individuals were considered as affected when they presented symptoms from both the lacrimal apparatus (lacrimal glands and/or lacrimal puncta) and the salivary glands.

Paper I: Results and discussion

We performed a genome-wide linkage scan with approximately 400 microsatellite markers in family 1. Evidence of significant linkage was found for ALSG to a continuous pericentric region on chromosome 5. The same region was linked to ALSG in family 2. We obtained a maximum cumulative LOD score of 5.72 ($\theta = 0$) at the marker locus D5S398 for both families. Haplotype analyses of the two families restricted the linked region to a 22 centiMorgan (cM) interval on chromosome 5p13.2-q13.1 flanked by markers D5S395 and D5S2046. Different haplotypes were inherited with the disorder in the two families.

The gene encoding fibroblast growth factor 10 (FGF10) is situated in the linked region. *FGF10* maps to 5p12-p13 in human[76] and is necessary for the formation of several organs in mouse, including lacrimal- and salivary glands[77,78]. The FGFs are important regulators of cellular proliferation, differentiation, migration and survival. They interact with fibroblast growth factor receptors (FGFRs), which are members of the tyrosine kinase receptor family[78]. The FGF protein family comprises at least 20 members[78]. *FGF10* consists of three exons and encodes a 208 amino acid protein which binds to fibroblast growth factor receptor 2b (FGFR2b) with high affinity[79]. *Fgf10*$^{-/-}$ mice display a complex phenotype and die at birth due to lack of lung development[80]. These mice also have absent lacrimal- and salivary glands, absent fore- and hind limbs, agenesis of pituitary and thyroid glands, dysgenic teeth, kidney, thymus, stomach, pancreas and inner ear as well as abnormal hair and skin[78]. Abnormal external genitalia development[81], anorectal malformations[82] and abnormal mammary gland formation[83] have also been reported.

From our linkage analysis and from the previously described *Fgf10*$^{-/-}$ mice, the human gene encoding FGF10 became an obvious candidate for ALSG. Mutation screening of *FGF10* revealed a 53 kb deletion including exon 2 and 3 in family 1 and a nonsense mutation in exon 3 (p.Arg193X; c.577C>T) in family 2. If a truncated protein is synthesized at all in family 2, two predicted sites for post-translational modification[84] and two amino acid residues involved in interaction with FGFR2b[85] are abrogated. We propose that ALSG in these two families is caused by haploinsufficiency of FGF10 and that the level of FGF10 derived from one allele is sufficient for the development and homeostasis of other organs dependent on it. This may explain the specific phenotype restricted to the lacrimal and salivary glands of these patients. No abnormalities had previously been described in *Fgf10*$^{+/-}$ mice so we decided to thoroughly examine these. Macroscopical and histological examination demonstrated aplasia of lacrimal glands and hypoplasia of salivary glands in *Fgf10*$^{+/-}$ mice. Other internal organs were macroscopically normal. We therefore propose that the response to FGF10 is dosage-sensitive, at least at the embryonic stage and at the site of lacrimal and salivary gland formation.

To investigate whether mutations in *FGF10* are a common cause of dry eyes/dry mouth, we screened DNA samples from 74 patients with dry mouth and/or dry eyes not fulfilling the criteria for Sjögren´s syndrome[86,87]. No sequence alterations in the *FGF10* gene were found in these patients, which implies that mutations in *FGF10* are uncommon in patients with unspecific sicca syndromes.

Since our initial publication of *FGF10* mutations associated with ALSG, independent studies have confirmed our findings. A disorder quite similar to ALSG is Lacrimo Auriculo Dento Digital Syndrome (LADD syndrome) (OMIM 149730) which is an autosomal dominant disorder characterized by

abnormalities of the face, ears, eyes, mouth, teeth, digits and genitourinary organs[88]. There is a significant clinical overlap between ALSG and LADD syndrome, through the aplasia of lacrimal and salivary glands and absence of lacrimal puncta. In 2006 it was established that LADD syndrome is genetically heterogeneous and may be caused by heterozygous missense mutations in *FGF10* as well as mutations in the fibroblast growth factor receptors 2 and 3 (*FGFR2, FGFR3*)[89,90]. The same year a nonsense mutation in *FGF10* was reported in a mother with ALSG and her daughter with LADD syndrome[90].

Paper II: Subjects

We identified two patients with symptoms concerning their lacrimal and salivary glands. Patient 1 had dry mouth, reduced lacrimal fluid production, absent lacrimal puncta and caries at a very young age. The father of the patient had similar features and went through MRI showing absence of the lacrimal glands and several of the major salivary glands. Patient 2 had dry mouth and caries. He was born with absent inferior lacrimal puncta and did not produce tears when crying. He also had hypospadias. MRI revealed hypoplastic lacrimal, parotid and submandibular glands. Sublingual glands were present and of normal size.

Paper II: Results and discussion

Sequence analysis of the coding region of *FGF10* as well as the exon-intron boundaries was performed in the patients and their parents. *FGF10* gene analysis of patient 1 revealed two missense mutations. The first was a heterozygous c.240A>C nucleotide transversion in exon 1 that was also present in the affected father. This nucleotide transversion is predicted to result in a chemically non-conservative (basic to neutral) amino acid substitution from arginine to serine at position 80. The arginine at position 80 in FGF10 is known to interact with the D3 region of FGFR2b[85]. Second, a heterozygous c.620A>C nucleotide transversion was identified in DNA of the proband that was inherited from the mother. This nucleotide transversion is located in exon 3 of *FGF10* and is predicted to result in a non-conservative amino acid change from histidine to proline (basic to neutral) at position 207. This residue is not predicted to be evolutionary conserved. Sequence analysis of *FGF10* in patient 2 and his parents disclosed a heterozygous *de novo* nucleotide transition, c.413G>A, in exon 2, resulting in a non-conservative substitution from glycine to glutamic acid (neutral to acidic) at position 138 of FGF10. The glycine at position 138 in FGF10 is a highly conserved amino acid residue of predicted functional importance.

These findings indicate that ALSG may be associated with missense mutations as well as nonsense mutations or deletions.

Future perspectives

The number and types of described *FGF10* mutations continues to increase. Recently, Scheckenbach et al. identified a splicing mutation in intron 2 (c.430-1G>A) in two brothers with ALSG[91]. Identification of *FGF10* as the gene causing ALSG and some cases of LADD syndrome will hopefully result in increased diagnostic accuracy and will reduce the number of undiagnosed or possibly miss-diagnosed patients. We therefore suggest mutation screening of *FGF10* when any of these two syndromes are suspected. Comparison of three FGF10 LADD syndrome mutants (C106F, I156R and K137X) to wildtype FGF10 revealed that haploinsufficiency due to impaired biological activity causes the disorder[92]. The C106F mutant has reduced protein stability, the I156R mutant has decreased FGFR2b binding affinity and there was no expression of the K137X mutant. The reason for why some patients with *FGF10* mutations have ALSG and some have LADD syndrome remains to be elucidated. It is possible that modifier genes, gene-gene interactions, environment or stochastic events play a role.

Kostmann syndrome

Background

Severe congenital neutropenia (SCN) was first described as an autosomal recessive disorder by the paediatrician Rolf Kostmann in 1956. He reported a large inbred Swedish family where the disorder was inherited[93]. There is evidence that a single founder from the parish of Överkalix is responsible for the genetic defect in this family[94]. Kostmann syndrome is characterized by a very low absolute neutrophil count (ANC) ($<0.2x10^9$), caused by maturation arrest of the myelopoiesis at the promyelocyte/myelocyte stage in the bone marrow[95]. A common symptom of SCN is recurrent bacterial infections which appear early in life. Before the discovery of antibiotics, affected children died during the first year of life. Treatment with granulocyte colony stimulating factor (GCSF) bypasses the maturation arrest and thereby restores the ANC to normal levels. This reduces the number of infections in the patients[95]. GCSF injections must be administered throughout life. The only cure for the syndrome is haematopoietic stem cell transplantation. SCN has an estimated prevalence of 1-2 cases per million individuals[96]. SCN is a heterogenous disorder with different forms of pathogenesis and inheritance. Many patients with sporadic or dominant SCN have mutations in elastase 2 (*ELA2*) or in growth factor independent 1 (*GFI1*)[97,98]. When this study started no gene had yet been associated with autosomal recessive SCN.

Paper III: Subjects

We have obtained DNA from members of the original Kostmann family. A second family with SCN not known to be connected to the Kostmann pedigree but from the same geographical region was also included in the study. Diagnosis was established in the patients within the first 6 months of life. All patients had an ANC lower than $0.2x10^9$. Bone marrow examination confirmed the typical maturation block at the promyelocyte/myelocyte stage. The clinical findings in the patients were skin abscesses, otitis, gingivitis, tonsillitis, sinusitus, pneumonia and septic infection.

Paper III: Results and discussion

ELA2, *GFI1*, *GCSF* and granulocyte colony stimulating factor receptor (*GCSFR*) had previously been excluded through mutation analysis, as the gene causing the disorder in this family. A genome scan with 370 polymorphic microsatellite markers (Weber screening set version 6/Cooperative Human Linkage Center) had also been performed. No linkage was found and the reason may be that the average intermarker distance of 10 cM was too small to detect linkage. The theoretic estimate of the region shared by the affected individuals is 1.9 cM. This value is based on the number of meioses separating the individuals in the pedigree. To get a better genomic coverage we first used the GeneChip Human Mapping 10K array and then the GeneChip Human Mapping 100K set from Affymetrix[62,63]. The 100K set allowed simultaneous genotyping of 116 000 SNPs on two oligonucleotide arrays and only required a total of 500 ng genomic DNA. The mean intermarker distance was 23.6 kb. Four patients from the original Kostmann family were genotyped with the arrays. The resolution of the 100K set allowed us to identify a region on chromosome 1q22 that was confirmed with microsatellite markers in three of the four patients. The fourth patient did not share a continuous homozygous SNP pattern within this region and was later found to have a heterozygous *de novo ELA2* mutation[99]. At this time of the study DNA was extracted from formalin fixed paraffin embedded autopsy specimens from a fifth patient from the original Kostmann family. The minimal haplotype shared by the patients covered 1.2 Mb and contained 37 known genes of which several were associated with myelopoiesis. The haplotype also segregated with the disease in the second family not connected to the original Kostmann family. This indicated a common founder mutation for SCN in the two families.

Future perspectives

In 2007, HS1-associated protein X1 (*HAX1*), localized to chromosome 1q21.3, was identified as the gene causing SCN in the original Kostmann family as well as in three Kurdish pedigrees and a cohort of sporadic and familial individuals[100]. HAX1 participates in B cell receptor mediated signal transduction[101] and is associated with the actin cytoskeleton[102]. Quantitative reverse-transcriptase PCR established that transcript variants I and II of *HAX1* are ubiquitously expressed[103]. HAX1 is needed for maintaining the inner mitochondrial membrane potential and protects against apoptosis in myeloid cells which suggests that it is a major regulator of myeloid cell homeostasis[100]. The affected individuals from the original Kostmann family and the second family, linked to chromosome 1q22, were homozygous for a nonsense mutation (p.Q190X; c.586C>T) affecting both transcript

variants[100,103]. These individuals developed neurological disease with decreased cognitive function and three of the four individuals developed epilepsy[103]. The clinical phenotype of HAX1 deficiency seems to depend on the localization of the mutation and its effect on the transcript variants[104]. It is now established that mutations affecting only transcript variant I are associated with SCN whereas mutations affecting both transcript variants cause SCN and neurological symptoms including developmental delay and epilepsy[103-107]. This suggests a pivotal role of the HAX1 isoform b in neural development. Mutation screening of *HAX1* is recommended in all individuals with autosomal recessive SCN. Identification of *HAX1* as the gene causing Kostmann syndrome will help in the understanding of the genes and factors involved in myelopoiesis as well as neurodevelopment. Hopefully in the long run it will also provide better care and therapy for patients.

Inversion mapping

Background

Information about the human genomic sequence in combination with techniques such as FISH has made molecular cytogenetics an evolving and easily accessible approach to identify structural and numerical chromosome aberrations. Inversions are the most common type of human constitutional karyotype alteration[108]. A chromosomal inversion involves two breaks and the reinsertion of the chromosome segment in the opposite direction compared to the original chromosome[109]. There are two forms of inversions: 1) pericentric inversions where a break and reunion occur between the short and long chromosome arms and 2) paracentric inversions where both the break and the reunion occur within the same arm[110] (Figure 6). Inversions are relatively common in the human population with estimated frequencies of 1-2% for pericentric inversions and 0.009-0.049% for paracentric inversions[111]. A large proportion of inversions have recurrent break-points and could either have arisen through multiple unrelated events or be identical by descent[112]. Most inversions (both paracentric and pericentric) are without direct phenotypic effects unless they disrupt genes[112].

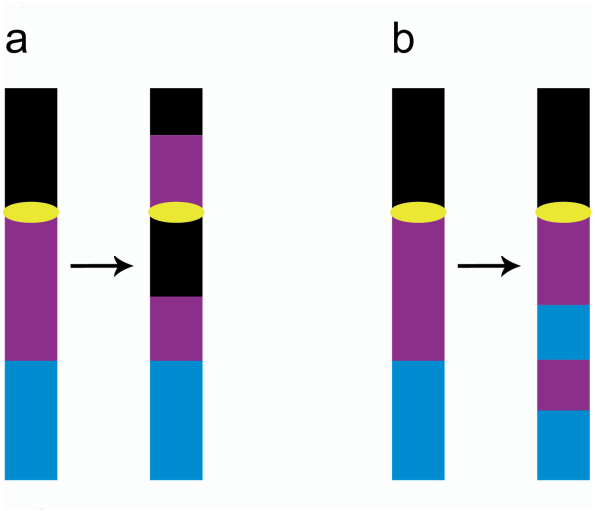

Figure 6. Chromosome inversions. Black, p arm; yellow, centromere; purple and blue, q arm. (a) Pericentric inversion. (b) Paracentric inversion.

Paper IV: Subjects

Four non-related Swedish patients with a chromosome 10 inversion were identified through standard G-banding. They were each referred to cytogenetic analysis because of a clinical phenotype. Patient 1 had ADHD and learning deficits, patient 2 had developmental delay and autism, patient 3 had mild mental retardation and patient 4 had Feingold syndrome. Cell lines were established for FISH analysis from three of the subjects through Epstein-Barr virus (EBV)-transformation of B-lymphocytes.

Paper IV: Results and discussion

FISH analysis with bacterial artificial chromosome clones from 10q demonstrated a 12 Mb inversion between bands 10q11.22 and 10q21.1. The presence of large intra-chromosomal segmental duplications on 10q11.22 made the mapping difficult. Through mini-FISH analysis the distal breakpoint was further restricted to a 36 kb region on 10q21.1. This region was narrowed down to a 3 kb region by Southern blot analysis. Once the 3 kb sequence surrounding the distal breakpoint was identified inverse PCR was used to clone the proximal breakpoint. Inverse PCR allows rapid amplification of regions of unknown sequence flanking a DNA segment of known sequence[113]. The sequences immediately spanning the breakpoints were determined and this revealed that the breakpoints were identical in the three subjects. The inversion was approximately 12 Mb in size. No genes or predicted transcripts were disrupted by the inversion. The proximal breakpoint was situated within a LTR and the distal breakpoint is situated 900 bp downstream of a SINE. Sequence analysis of 10 kb of DNA at each breakpoint revealed that the breakpoint regions were AT-rich and enriched for interspersed repeats. An assay based on multiplex PCR was developed for easy detection of the inv(10)(q11.22;q21.1). The assay confirmed that the four subjects were heterozygous for the inversion. Samples from five parents were available for screening with the assay and three of them were identified as carriers. The PCR assay was also used to screen 154 Swedish blood donors, 2400 Icelandic blood donors and 454 patients with unspecific mental retardation. These samples were all negative for the inv(10)(q11.22;q21.1).

Rearrangements that are identical by descent can be distinguished from recurrent events by studying polymorphic markers that are close to the sites of breakage and exchange. In independent events, the marker alleles would be expected to be different whereas if the rearrangement is identical by descent the alleles would be expected to be identical[114]. Genotyping of 15 microsatellite markers along the inversion was done on DNA from the four subjects and available parents. Haplotype analysis indicated that the chromosome 10 inversion was identical by descent.

Table 1. *Incidence of inv(10)(q11.22q21.1) in Swedish blood samples and amniocenteses.*

	Inversion 10	Normal 10
Blood samples	11	7428
Amniocenteses	7	8889

A retrospective study of karyotypes performed between 1996 and 2007 at two Swedish cytogenetic laboratories was carried out and the inv(10)(q11.22;q21.1) was found in 11 out of 7439 blood specimens and in 7 out of 8896 amniocenteses (Table 1). The incidence of the chromosome inversion was not significantly different between the blood specimens and the amniocenteses (chi square test, $p = 0.18$). This gave an overall carrier frequency of 0.11%, suggesting the rearrangement to be widely distributed in the Swedish population. The inversion carrier frequency in the Icelandic samples did not differ significantly from the carrier frequency in the Swedish samples (chi square test, $p = 0.10$) (Table 2). Nevertheless, the absence of the inversion in the Icelandic cohort may imply a variation in frequency among Europeans.

Table 2. *Incidence of inv(10)(q11.22q21.1) in the Swedish and Icelandic populations.*

	Inversion 10	Normal 10
Sweden	18	16317
Iceland	0	2400

Future perspectives

The inv(10)(q11.22;q21.1) was not found to disrupt any gene or transcribed sequence and it does not seem to be associated with any consistent phenotype. The intrachromosomal segmental duplications at 10q11.22, the high percentage of repetitive sequence at the breakpoints and/or the presence of the LTR at the 10q11.22 breakpoint might once have mediated the inversion. The majority of chromosomal rearrangements are non-random events, which may result from predisposing elements or sequence motifs in the genome[115]. Analysis of inversion breakpoint regions shows that repetitive elements, high sequence similarity to the opposite breakpoint, pseudogenes, gene deserts, segmental duplications and co-localization with fragile sites are over-represented[108]. Many prevalent inversions are regarded as polymorphic variants since they do not disrupt any genes. These are stably transmitted and are without any disease association or detected phenotype. Still a subtle effect on the carrier can not be ruled out, for example by the inversion

influencing the expression of nearby genes. Further studies are required to shed light on any subtle genotype-phenotype correlations for the inversion. It remains to be determined how abundant inv(10)(q11.22;q21.1) is in additional European populations and other ethnic groups. The developed inv(10)(q11.22;q21.1) screening assay allows for efficient and accurate analysis of large sample sets, which can be used for this purpose.

Concluding remarks

In the last decades there have been tremendous breakthroughs in human genetics. The completion of the human genome sequence, targeted resequencing, discovery of transcription factor binding sites, ncRNA expression profiling, CGH arrays, the HapMap project and chromatin profiling are just a few of the things that have increased our understanding of the very dynamic human genome. The breakthroughs have also changed our way of thinking when it comes to what can be considered as normal genetic variation compared to disease related deviations. Genetic variation plays a role in phenotypic diversity and genome evolution. Future studies will give a clearer view of the genetic variation in our genomes and allow for detection of rare variants. The recent sequencing of the entire genome of individuals[24,116-118] has opened the doors for personal genomics where genome sequences may become a standard form of personal information. The development of novel DNA technologies has also accelerated the cloning of many new genes associated with inherited human disorders. The identification of human disease genes makes it possible to study their function and learn more about the pathophysiology. In the long term view this can improve diagnostics, risk assessment and therapy, and sometimes also help to develop medicine. It will be very interesting to see how the human genetic analysis and biomedical research will advance in the years to come.

Populärvetenskaplig sammanfattning

I alla våra celler finns långa DNA molekyler som bildar strukturer som kallas för kromosomer. I varje cell har man normalt 46 kromosomer, 22 par autosomer och ett par könskromosomer (XX eller XY). DNA i våra celler bär på livsviktig information för kroppens funktioner. Denna utgörs av den genetiska koden som beskriver hur informationen överförs från nukleotidsekvens i mRNA till aminosyrasekvens i proteiner. Fel i den genetiska koden (mutationer) samt fel i antalet eller strukturen på kromosomer kan orsaka sjukdom. Ett lokus är en specifik plats på en kromosom där en gen (arvsanlag) eller annan typ av DNA-sekvens är lokaliserad. För varje lokus har en individ två alleler (alternativa anlag), en ärvd från mamman och en ärvd från pappan. En egenskap är dominant när det räcker att individen har en allel av den för att egenskapen ska uttryckas medan en egenskap är recessiv om det krävs att individen har två alleler för att den ska uttryckas. Två lokus som sitter nära varandra på en kromosom har en tendens att nedärvas tillsammans, detta kallas för koppling. En genetisk markör är en DNA-sekvens med känd kromosomposition som har flera olika alleler. Genetiska markörer kan användas för att lokalisera anlaget och i bästa fall identifiera genen för en genetisk sjukdom. Man studerar då hur de genetiska markörerna nedärvs i friska och sjuka individer inom en familj. Man letar efter ett kromosomområde där de sjuka individerna ser likadana ut på DNA-nivå för en eller flera genetiska markörer. Om detta område ligger tillräckligt nära sjukdomslokuset för att de ska nedärvas tillsammans, kan man genom att hitta ett kromosområde kopplat till sjukdomen också hitta sjukdomsorsaken.

Bättre förståelse av orsakerna bakom ärftliga sjukdomar är viktiga både ur medicinsk och ur biologisk synvinkel. Genom att identifiera genen för en ärftlig sjukdom kan man studera dess funktion och på så sätt få kunskap om hur genen är inblandad i sjukdomen. Identifiering av sjukdomsgener kan på lång sikt ge förbättrad diagnostik, prognos, riskbedömning samt behandling av patienter och i bästa fall också leda till ett botemedel. I denna avhandling presenterar jag arbetet för identifiering av de genetiska orsakerna bakom avsaknad av tår- och spottkörtlar (ALSG) och Kostmanns sjukdom samt identifieringen av en ny kromosomvariant.

ALSG är en mycket ovanlig autosomalt dominant sjukdom. Normalt avsöndrar spottkörtlarna saliv som håller slemhinnorna i munnen fuktiga och underlättar tuggande och sväljande av mat. Saliven minskar också

bakteriefloran i munnen. Tårkörteln producerar tårvätska som skyddar, renar och fuktar ögat. Symptom på avsaknad av tår- och spottkörtlar är muntorrhet, karies, tandlossning och torra ögon. Genom analys med genetiska markörer i två stora svenska familjer med ALSG kunde vi identifiera en region på kromosom 5 kopplad till sjukdomen. Genen *FGF10* i detta område valdes ut för mutationsanalys och visade sig vara muterad. Efter den ursprungliga studien fick vi tag på ytterligare två patienter med ALSG och sekvensanalys visade att också dessa hade mutationer i *FGF10*.

Kostmanns sjukdom är en autosomalt recessiv neutropeni. Denna karakteriseras av ett väldigt lågt antal vita blodkroppar. Vita blodkroppar är en viktig del av immunförsvaret som hjälper kroppen att bekämpa infektioner. Bristen på vita blodkroppar gör att patienterna drabbas av upprepade bakterieinfektioner. Innan det fanns antibiotika dog patienterna i spädbarnsåldern. Idag behandlas patienterna med en tillväxtfaktor som återställer neutrofilnivåerna till det normala och därmed minskar antalet infektioner. Patienter från den ursprungliga familjen, beskriven av Dr. Kostmann 1956, deltog i studien och ett område på kromosom 1 kunde kopplas till sjukdomen. I en annan studie med familjer och patienter med samma sjukdom identifierades HAX1 som sjukdomsgenen. HAX1 ligger på kromosom 1 och visade sig vara den defekta genen även i Kostmann familjen.

En inversion är en strukturell kromosomförändring som innebär att en bit av kromosomen brutits av på två ställen och återinsatts i motsatt riktning. Med hjälp av olika cytogenetiska metoder kunde en inversion på kromosom 10 och dess brottpunkter identifieras och karakteriseras. Ingen specifik sjukdom eller egenskap kunde associeras till inversionen. Ett diagnostiskt test utvecklades för inversionen som kan användas till snabb och enkel analys av stora prover för att fastställa dess förekomst i en population. Genom en retrospektiv studie av flera tusen kromosomuppsättningar kunde det fastställas att inversionen är en ovanlig kromosomvariant med bred spridning i Sverige.

Studierna i denna avhandling leder till ökad förståelse av genetiken bakom bildandet av tår- och spottkörtlar samt neutropeni. En annan lärdom är att kromosomförändringar utan negativ inverkan på bäraren är vanligt förekommande.

Acknowledgements

The work presented in this thesis was performed at the Department of Genetics and Pathology at Uppsala University. I wish to express my genuine gratitude to the following people for making this thesis happen:

My supervisor, Professor *Niklas Dahl* for giving me the opportunity to do research in the field of clinical genetics. Thank you for believing in my judgement and skills and for giving me valuable advice when it comes to reading and writing.

Present and former members of the "Dahl group": *Birgit* for nice tip-offs about FISH and life. *Anne-Sophie*, *Malin*, *Hanna Ö* and *Johanna* for great friendship and fun talks about both science and daily life. It has been nice that at least one of you have been working at the same time as me during the years. *Jocke* for help with computers and statistics and for throwing parties in your and Jenny's fabulous apartment. Former post-docs *Lena*, *Ed* and *Maria* for nice discussions and opinions on things. *Larry*, our meeting with the Queen at the Royal Palace in Stockholm is a precious PhD student memory. *Jens*, *Jitendra*, *Jimmy* and the newcomers *Dana*, *Sara* and *Chikari* for company at Friday fika. Our Pakistani co-workers Dr *Baig*, *Mahmood*, *Sadia* and *Aysha* thanks for oriental influences. Degree project and summer school students over the years are also acknowledged.

The Wadelius boys; *Mattias*, *Benjamin*, *Fredrik*, *Alvaro*, *Ola*, *Mehdi* and The Bondesson girls; *Maja*, *Christina* and *Sara* for company in the lab and at kick-offs, dinners and parties. *Sofie*, *Dan* and *Anna H* my teaching colleagues.

The personnel at Uppsala Genome Center for runs on the ABI machines. Dr *Anders Isaksson*, *Hanna Göransson* and *Maria Rydåker* at Uppsala Array Platform for help with SNP array genotyping. Associate professor *Lena Åslund* thanks for the "Advanced techniques in molecular medicine" course and Professor *Claes Wadelius* for the "Transcription factors" literature course. Professor *Göran Annerén*, Docent *Marie-Louise Bondeson*, Professor *Lore Zech*, Professor *Karl-Henrik Gustavsson*, the personnel at Clinical Genetics, the administrative staff and all the other people at the Rudbeck laboratory for the friendly and scientific atmosphere.

Birgitta Bergendal, *Lena Olson* and *Babak Falahat* thank you for the interesting and fruitful collaboration on the ALSG project.

My beloved family: *Hasse* (stay-at-home dad) for reading my thesis, working and taking care of our children all at the same time. You are the best! ♥ *Christian* a.k.a. Babo my chubby-faced, blue-eyed boy and *Hera* my little princess. You make me so happy! ☼ Mom and dad, you have always been there for me. The house in Berthåga will always feel like home. *Tony*, my darling brother, I wish you the best of luck with your university studies and your music. 2 unlimited forever! Baby brother *Philip*, not so little anymore. You have your own apartment, car and a "real job". I am really proud of you. *Michelle* (mini-me), you are such a warm and kind person. I am sure whatever you decide to become you will be great at it.

My old friends "kompisar från förr": *Ami*, we have now been best friends for seventeen years, so many wonderful memories. *Ida*, good luck with your PhD studies. *Meysam*, *Linnea* and *Annso* and everyone else I know from S-hagen, those were the days!

My fellow biology students: *Kattis*, our friendship has evolved from Friday nights at Värmlands nation to couple dinners to children's birthday parties. *Maria*, I was so happy to find an Uppsala girl with similar interests at the first biology course. *Gina*, *Emma* and *Jenny* thanks for all the fun during our studies. *Emmeli* and *Rita* my degree project colleagues. Rita, whenever I hear a Shakira song I still think of you. *Jonas Krantz* thank you for teaching us girls how to play Dungeons & Dragons. I still have my pink role-play dice. *Johan*, *Dariush*, *Adam*, *Mårten*, *Ulrika*, *Anna M*, *Åsa*, *Eva*, *Olle*, *Dieter*, *Edith*, *Karolina*, *Anna-Karin* and all the others good luck with your future plans!

Mammagruppen (*Pernilla*, *Sofia*, *Helena*, *Lotta* and *Hanna S*) thank you for caffe lattes, long walks, play-dates and dinners at Ekocaféet.

A special thank goes to the co-authors of my papers and all the patients who participated in the studies.

The following foundations are acknowledged for financial support: Uppsala University, Uppsala University hospital, Swedish Research Council, The Thuréus Foundation, The Swedish Children's Cancer Foundation, The Swedish Cancer Society, The National Institutes of Health, Torsten och Ragnar Söderbergs Stiftelser, Birgit och Bengt Hedströms Stiftelse för barnneurologisk forskning, Stiftelsen Sävstaholm and Stiftelsen Drottning Silvias Jubileumsfond.

References

1. Dunn, P. M. Gregor Mendel, OSA (1822-1884), founder of scientific genetics. *Arch Dis Child Fetal Neonatal Ed* **88**, F537-9 (2003).
2. Fairbanks, D. J. & Rytting, B. Mendelian controversies: a botanical and historical review. *Am J Bot* **88**, 737-752 (2001).
3. Avery, O. T., MacLeod, C. M. & McCarty, M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. *J Exp Med* **149**, 297-326 (1979).
4. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
5. Ross, D. W. The human genome: information content and structure. *Hosp Pract (Minneap)* **34**, 49-54, 56-60, 65 (1999).
6. Rando, O. J. Chromatin structure in the genomics era. *Trends Genet* **23**, 67-73 (2007).
7. Khorasanizadeh, S. The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259-72 (2004).
8. Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J.,

Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Xu, M., Haidar, J. N., Yu, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., de Bakker, P. I., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. & de Jong, P. J. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).

9. Huttenhofer, A. & Schattner, P. The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat Rev Genet* **7**, 475-82 (2006).
10. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-98 (2002).
11. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat Rev Genet* **10**, 155-9 (2009).
12. Flynt, A. S. & Lai, E. C. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* **9**, 831-42 (2008).

13. Long, M., Betran, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* **4**, 865-75 (2003).
14. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**, 94-108 (2009).
15. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).
16. Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S. & Snyder, M. What is a gene, post-ENCODE? History and updated definition. *Genome Res* **17**, 669-81 (2007).
17. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-40 (2004).
18. Elgar, G. & Vavouri, T. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet* **24**, 344-52 (2008).
19. McClintock, B. Induction of Instability at Selected Loci in Maize. *Genetics* **38**, 579-99 (1953).
20. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272-85 (2007).
21. Cordaux, R. The human genome in the LINE of fire. *Proc Natl Acad Sci U S A* **105**, 19033-4 (2008).
22. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. & Schulman, A. H. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**, 973-82 (2007).
23. Prak, E. T. & Kazazian, H. H., Jr. Mobile elements and the human genome. *Nat Rev Genet* **1**, 134-44 (2000).
24. Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., Lin, Y., MacDonald, J. R., Pang, A. W., Shago, M., Stockwell, T. B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S. A., Busam, D. A., Beeson, K. Y., McIntosh, T. C., Remington, K. A., Abril, J. F., Gill, J., Borman, J., Rogers, Y. H., Frazier, M. E., Scherer, S. W., Strausberg, R. L. & Venter, J. C. The diploid genome sequence of an individual human. *PLoS Biol* **5**, e254 (2007).
25. Crawford, D. C. & Nickerson, D. A. Definition and clinical importance of haplotypes. *Annu Rev Med* **56**, 303-20 (2005).
26. Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Sun, W., Wang, H., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier,

J. F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Yakub, I., Birren, B. W., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archeveque, P., Bellemare, G., Saeki, K., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R. & Stewart, J. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-61 (2007).

27. Thorisson, G. A., Smith, A. V., Krishnan, L. & Stein, L. D. The International HapMap Project Web site. *Genome Res* **15**, 1592-3 (2005).
28. Lai, Y. & Sun, F. The relationship between microsatellite slippage mutation rate and the number of repeat units. *Mol Biol Evol* **20**, 2123-31 (2003).
29. Nakamura, Y. DNA variations in human and medical genetics: 25 years of my experience. *J Hum Genet* **54**, 1-8 (2009).
30. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nat Rev Genet* **7**, 85-97 (2006).
31. Zhang, L., Lu, H. H., Chung, W. Y., Yang, J. & Li, W. H. Patterns of segmental duplication in the human genome. *Mol Biol Evol* **22**, 135-41 (2005).
32. Henrichsen, C. N., Vinckenbosch, N., Zollner, S., Chaignat, E., Pradervand, S., Schutz, F., Ruedi, M., Kaessmann, H. & Reymond, A. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet* (2009).
33. Itsara, A., Cooper, G. M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R. M., Myers, R. M., Ridker, P. M., Chasman, D. I., Mefford, H.,

Ying, P., Nickerson, D. A. & Eichler, E. E. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* **84**, 148-61 (2009).
34. Strausberg, R. L., Levy, S. & Rogers, Y. H. Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov Today* **13**, 569-77 (2008).
35. ten Bosch, J. R. & Grody, W. W. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn* **10**, 484-92 (2008).
36. Morozova, O. & Marra, M. A. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-64 (2008).
37. Antonarakis, S. E., Krawczak, M. & Cooper, D. N. Disease-causing mutations in the human genome. *Eur J Pediatr* **159 Suppl 3**, S173-8 (2000).
38. George, R. A., Smith, T. D., Callaghan, S., Hardman, L., Pierides, C., Horaitis, O., Wouters, M. A. & Cotton, R. G. General mutation databases: analysis and review. *J Med Genet* **45**, 65-70 (2008).
39. Amberger, J., Bocchini, C. A., Scott, A. F. & Hamosh, A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* **37**, D793-6 (2009).
40. Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., Abeysinghe, S., Krawczak, M. & Cooper, D. N. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* **21**, 577-81 (2003).
41. Pagani, F. & Baralle, F. E. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet* **5**, 389-96 (2004).
42. Jiang, Y. H., Bressler, J. & Beaudet, A. L. Epigenetics and human disease. *Annu Rev Genomics Hum Genet* **5**, 479-510 (2004).
43. Le Merrer, M., Rousseau, F., Legeai-Mallet, L., Landais, J. C., Pelet, A., Bonaventure, J., Sanak, M., Weissenbach, J., Stoll, C., Munnich, A. & et al. A gene for achondroplasia-hypochondroplasia maps to chromosome 4p. *Nat Genet* **6**, 318-21 (1994).
44. Gusella, J. F., Wexler, N. S., Conneally, P. M., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P. C., Ottina, K., Wallace, M. R., Sakaguchi, A. Y. & et al. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* **306**, 234-8 (1983).
45. Wainwright, B. J., Scambler, P. J., Schmidtke, J., Watson, E. A., Law, H. Y., Farrall, M., Cooke, H. J., Eiberg, H. & Williamson, R. Localization of cystic fibrosis locus to human chromosome 7cen-q22. *Nature* **318**, 384-5 (1985).
46. Woo, S. L., Lidsky, A. S., Guttler, F., Chandra, T. & Robson, K. J. Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature* **306**, 151-5 (1983).
47. Orlen, H., Melberg, A., Raininko, R., Kumlien, E., Entesarian, M., Soderberg, P., Pahlman, M., Darin, N., Kyllerman, M., Holmberg, E., Engler, H., Eriksson, U. & Dahl, N. SPG11 mutations cause Kjellin syndrome, a hereditary spastic paraplegia with thin corpus callosum and

central retinal degeneration. *Am J Med Genet B Neuropsychiatr Genet* (2009).
48. Kere, J., Srivastava, A. K., Montonen, O., Zonana, J., Thomas, N., Ferguson, B., Munoz, F., Morgan, D., Clarke, A., Baybayan, P., Chen, E. Y., Ezer, S., Saarialho-Kere, U., de la Chapelle, A. & Schlessinger, D. X-linked anhidrotic (hypohidrotic) ectodermal dysplasia is caused by mutation in a novel transmembrane protein. *Nat Genet* **13**, 409-16 (1996).
49. Robertson, J. H. & Trueman, R. G. Combined Hemophilia and Christmas Disease. *Blood* **24**, 281-8 (1964).
50. Read, A. P., Thakker, R. V., Davies, K. E., Mountford, R. C., Brenton, D. P., Davies, M., Glorieux, F., Harris, R., Hendy, G. N., King, A. & et al. Mapping of human X-linked hypophosphataemic rickets by multilocus linkage analysis. *Hum Genet* **73**, 267-70 (1986).
51. Hawkins, J. R. Mutational analysis of SRY in XY females. *Hum Mutat* **2**, 347-50 (1993).
52. Tiepolo, L. & Zuffardi, O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet* **34**, 119-24 (1976).
53. Antonarakis, S. E. & Beckmann, J. S. Mendelian disorders deserve more attention. *Nat Rev Genet* **7**, 277-82 (2006).
54. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* **3**, 779-89 (2002).
55. Morton, N. E. Sequential tests for the detection of linkage. *Am J Hum Genet* **7**, 277-318 (1955).
56. Dawn Teare, M. & Barrett, J. H. Genetic linkage studies. *Lancet* **366**, 1036-44 (2005).
57. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**, 263-73 (1986).
58. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-7 (1977).
59. Peters, L. L., Robledo, R. F., Bult, C. J., Churchill, G. A., Paigen, B. J. & Svenson, K. L. The mouse as a model for human biology: a resource guide for complex trait analysis. *Nat Rev Genet* **8**, 58-69 (2007).
60. Collins, F. S., Rossant, J. & Wurst, W. A mouse for all reasons. *Cell* **128**, 9-13 (2007).
61. Carr, I. M., Flintoff, K. J., Taylor, G. R., Markham, A. F. & Bonthron, D. T. Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. *Hum Mutat* **27**, 1041-6 (2006).
62. Matsuzaki, H., Loi, H., Dong, S., Tsai, Y. Y., Fang, J., Law, J., Di, X., Liu, W. M., Yang, G., Liu, G., Huang, J., Kennedy, G. C., Ryder, T. B., Marcus, G. A., Walsh, P. S., Shriver, M. D., Puck, J. M., Jones, K. W. & Mei, R. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res* **14**, 414-25 (2004).
63. Matsuzaki, H., Dong, S., Loi, H., Di, X., Liu, G., Hubbell, E., Law, J., Berntsen, T., Chadha, M., Hui, H., Yang, G., Kennedy, G. C., Webster, T. A., Cawley, S., Walsh, P. S., Jones, K. W., Fodor, S. P. & Mei, R.

Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* **1**, 109-11 (2004).
64. Caspersson, T., Zech, L., Johansson, C. & Modest, E. J. Identification of human chromosomes by DNA-binding fluorescent agents. *Chromosoma* **30**, 215-27 (1970).
65. Pinkel, D., Straume, T. & Gray, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc Natl Acad Sci U S A* **83**, 2934-8 (1986).
66. Rauch, A., Schellmoser, S., Kraus, C., Dorr, H. G., Trautmann, U., Altherr, M. R., Pfeiffer, R. A. & Reis, A. First known microdeletion within the Wolf-Hirschhorn syndrome critical region refines genotype-phenotype correlation. *Am J Med Genet* **99**, 338-42 (2001).
67. Ramsey, W. R. A case of hereditary congenital absence of the salivary glands. *Am J Dis Child* **28**, 440 (1924).
68. Blackmar, F. B. Congenital atresia of all lacrimal puncta with absence of salivary glands. *Am J Ophthalmol* **8**, 139-140 (1925).
69. Ferreira, A. P., Gomez, R. S., Castro, W. H., Calixto, N. S., Silva, R. A. & Aguiar, M. J. Congenital absence of lacrimal puncta and salivary glands: report of a Brazilian family and review. *Am J Med Genet* **94**, 32-4 (2000).
70. Young, W., Khan, F., Brandt, R., Savage, N., Razek, A. A. & Huang, Q. Syndromes with salivary dysfunction predispose to tooth wear: Case reports of congenital dysfunction of major salivary glands, Prader-Willi, congenital rubella, and Sjogren's syndromes. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* **92**, 38-48 (2001).
71. Jonsson, R., Moen, K., Vestrheim, D. & Szodoray, P. Current issues in Sjogren's syndrome. *Oral Dis* **8**, 130-40 (2002).
72. Caccamise, W. C. & Townes, P. L. Congenital absence of the lacrimal puncta associated with alacrima and aptyalism. *Am J Ophthalmol* **89**, 62-5 (1980).
73. Higashino, H., Horii, T., Ohkusa, Y., Ohkuma, H., Ino, C., Nakazawa, M., Izumi, H. & Kobayashi, Y. Congenital absence of lacrimal puncta and of all major salivary glands: case report and literature review. *Clin Pediatr (Phila)* **26**, 366-8 (1987).
74. Milunsky, J. M., Lee, V. W., Siegel, B. S. & Milunsky, A. Agenesis or hypoplasia of major salivary and lacrimal glands. *Am J Med Genet* **37**, 371-4 (1990).
75. Myers, M. A., Youngberg, R. A. & Bauman, J. M. Congenital absence of the major salivary glands and impaired lacrimal secretion in a child: case report. *J Am Dent Assoc* **125**, 210-2 (1994).
76. Emoto, H., Tagashira, S., Mattei, M. G., Yamasaki, M., Hashimoto, G., Katsumata, T., Negoro, T., Nakatsuka, M., Birnbaum, D., Coulier, F. & Itoh, N. Structure and expression of human fibroblast growth factor-10. *J Biol Chem* **272**, 23191-4 (1997).
77. Makarenkova, H. P., Ito, M., Govindarajan, V., Faber, S. C., Sun, L., McMahon, G., Overbeek, P. A. & Lang, R. A. FGF10 is an inducer and Pax6 a competence factor for lacrimal gland development. *Development* **127**, 2563-72 (2000).

78. Ohuchi, H., Hori, Y., Yamasaki, M., Harada, H., Sekine, K., Kato, S. & Itoh, N. FGF10 acts as a major ligand for FGF receptor 2 IIIb in mouse multi-organ development. *Biochem Biophys Res Commun* **277**, 643-9 (2000).
79. Igarashi, M., Finch, P. W. & Aaronson, S. A. Characterization of recombinant human fibroblast growth factor (FGF)-10 reveals functional similarities with keratinocyte growth factor (FGF-7). *J Biol Chem* **273**, 13230-5 (1998).
80. Sekine, K., Ohuchi, H., Fujiwara, M., Yamasaki, M., Yoshizawa, T., Sato, T., Yagishita, N., Matsui, D., Koga, Y., Itoh, N. & Kato, S. Fgf10 is essential for limb and lung formation. *Nat Genet* **21**, 138-41 (1999).
81. Haraguchi, R., Suzuki, K., Murakami, R., Sakai, M., Kamikawa, M., Kengaku, M., Sekine, K., Kawano, H., Kato, S., Ueno, N. & Yamada, G. Molecular analysis of external genitalia formation: the role of fibroblast growth factor (Fgf) genes during genital tubercle formation. *Development* **127**, 2471-9 (2000).
82. Fairbanks, T. J., De Langhe, S., Sala, F. G., Warburton, D., Anderson, K. D., Bellusci, S. & Burns, R. C. Fibroblast growth factor 10 (Fgf10) invalidation results in anorectal malformation in mice. *J Pediatr Surg* **39**, 360-5; discussion 360-5 (2004).
83. Mailleux, A. A., Spencer-Dene, B., Dillon, C., Ndiaye, D., Savona-Baron, C., Itoh, N., Kato, S., Dickson, C., Thiery, J. P. & Bellusci, S. Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. *Development* **129**, 53-60 (2002).
84. Bagai, S., Rubio, E., Cheng, J. F., Sweet, R., Thomas, R., Fuchs, E., Grady, R., Mitchell, M. & Bassuk, J. A. Fibroblast growth factor-10 is a mitogen for urothelial cells. *J Biol Chem* **277**, 23828-37 (2002).
85. Yeh, B. K., Igarashi, M., Eliseenkova, A. V., Plotnikov, A. N., Sher, I., Ron, D., Aaronson, S. A. & Mohammadi, M. Structural basis by which alternative splicing confers specificity in fibroblast growth factor receptors. *Proc Natl Acad Sci U S A* **100**, 2266-71 (2003).
86. Vitali, C., Bombardieri, S., Moutsopoulos, H. M., Balestrieri, G., Bencivelli, W., Bernstein, R. M., Bjerrum, K. B., Braga, S., Coll, J., de Vita, S. & et al. Preliminary criteria for the classification of Sjogren's syndrome. Results of a prospective concerted action supported by the European Community. *Arthritis Rheum* **36**, 340-7 (1993).
87. Vitali, C., Bombardieri, S., Jonsson, R., Moutsopoulos, H. M., Alexander, E. L., Carsons, S. E., Daniels, T. E., Fox, P. C., Fox, R. I., Kassan, S. S., Pillemer, S. R., Talal, N. & Weisman, M. H. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* **61**, 554-8 (2002).
88. Bamforth, J. S. & Kaurah, P. Lacrimo-auriculo-dento-digital syndrome: evidence for lower limb involvement and severe congenital renal anomalies. *Am J Med Genet* **43**, 932-7 (1992).
89. Rohmann, E., Brunner, H. G., Kayserili, H., Uyguner, O., Nurnberg, G., Lew, E. D., Dobbie, A., Eswarakumar, V. P., Uzumcu, A., Ulubil-

Emeroglu, M., Leroy, J. G., Li, Y., Becker, C., Lehnerdt, K., Cremers, C. W., Yuksel-Apak, M., Nurnberg, P., Kubisch, C., Schlessinger, J., van Bokhoven, H. & Wollnik, B. Mutations in different components of FGF signaling in LADD syndrome. *Nat Genet* **38**, 414-7 (2006).
90. Milunsky, J. M., Zhao, G., Maher, T. A., Colby, R. & Everman, D. B. LADD syndrome is caused by FGF10 mutations. *Clin Genet* **69**, 349-54 (2006).
91. Scheckenbach, K., Balz, V., Wagenmann, M. & Hoffmann, T. K. An intronic alteration of the fibroblast growth factor 10 gene causing ALSG-(aplasia of lacrimal and salivary glands) syndrome. *BMC Med Genet* **9**, 114 (2008).
92. Shams, I., Rohmann, E., Eswarakumar, V. P., Lew, E. D., Yuzawa, S., Wollnik, B., Schlessinger, J. & Lax, I. Lacrimo-Auriculo-Dento-Digital Syndrome Is Caused by Reduced Activity of the Fibroblast Growth Factor 10 (FGF10)-FGF Receptor 2 Signaling Pathway. *Mol Cell Biol* **27**, 6903-12 (2007).
93. Kostmann, R. Infantile genetic agranulocytosis; agranulocytosis infantilis hereditaria. *Acta Paediatr* **45**, 1-78 (1956).
94. Iselius, L. & Gustavson, K. H. Spatial distribution of the gene for infantile genetic agranulocytosis. *Hum Hered* **34**, 358-63 (1984).
95. Carlsson, G. & Fasth, A. Infantile genetic agranulocytosis, morbus Kostmann: presentation of six cases from the original "Kostmann family" and a review. *Acta Paediatr* **90**, 757-64 (2001).
96. Dale, D. C., Cottle, T. E., Fier, C. J., Bolyard, A. A., Bonilla, M. A., Boxer, L. A., Cham, B., Freedman, M. H., Kannourakis, G., Kinsey, S. E., Davis, R., Scarlata, D., Schwinzer, B., Zeidler, C. & Welte, K. Severe chronic neutropenia: treatment and follow-up of patients in the Severe Chronic Neutropenia International Registry. *Am J Hematol* **72**, 82-93 (2003).
97. Person, R. E., Li, F. Q., Duan, Z., Benson, K. F., Wechsler, J., Papadaki, H. A., Eliopoulos, G., Kaufman, C., Bertolone, S. J., Nakamoto, B., Papayannopoulou, T., Grimes, H. L. & Horwitz, M. Mutations in proto-oncogene GFI1 cause human neutropenia and target ELA2. *Nat Genet* **34**, 308-12 (2003).
98. Horwitz, M., Benson, K. F., Person, R. E., Aprikyan, A. G. & Dale, D. C. Mutations in ELA2, encoding neutrophil elastase, define a 21-day biological clock in cyclic haematopoiesis. *Nat Genet* **23**, 433-6 (1999).
99. Carlsson, G., Aprikyan, A. A., Ericson, K. G., Stein, S., Makaryan, V., Dale, D. C., Nordenskjold, M., Fadeel, B., Palmblad, J. & Hentera, J. I. Neutrophil elastase and granulocyte colony-stimulating factor receptor mutation analyses and leukemia evolution in severe congenital neutropenia patients belonging to the original Kostmann family in northern Sweden. *Haematologica* **91**, 589-95 (2006).
100. Klein, C., Grudzien, M., Appaswamy, G., Germeshausen, M., Sandrock, I., Schaffer, A. A., Rathinam, C., Boztug, K., Schwinzer, B., Rezaei, N., Bohn, G., Melin, M., Carlsson, G., Fadeel, B., Dahl, N., Palmblad, J., Henter, J. I., Zeidler, C., Grimbacher, B. & Welte, K. HAX1 deficiency

causes autosomal recessive severe congenital neutropenia (Kostmann disease). *Nat Genet* **39**, 86-92 (2007).
101. Suzuki, Y., Demoliere, C., Kitamura, D., Takeshita, H., Deuschle, U. & Watanabe, T. HAX-1, a novel intracellular protein, localized on mitochondria, directly associates with HS1, a substrate of Src family tyrosine kinases. *J Immunol* **158**, 2736-44 (1997).
102. Gallagher, A. R., Cedzich, A., Gretz, N., Somlo, S. & Witzgall, R. The polycystic kidney disease protein PKD2 interacts with Hax-1, a protein associated with the actin cytoskeleton. *Proc Natl Acad Sci U S A* **97**, 4017-22 (2000).
103. Carlsson, G., van't Hooft, I., Melin, M., Entesarian, M., Laurencikas, E., Nennesmo, I., Trebinska, A., Grzybowska, E., Palmblad, J., Dahl, N., Nordenskjold, M., Fadeel, B. & Henter, J. I. Central nervous system involvement in severe congenital neutropenia: neurological and neuropsychological abnormalities associated with specific HAX1 mutations. *J Intern Med* **264**, 388-400 (2008).
104. Germeshausen, M., Grudzien, M., Zeidler, C., Abdollahpour, H., Yetgin, S., Rezaei, N., Ballmaier, M., Grimbacher, B., Welte, K. & Klein, C. Novel HAX1 mutations in patients with severe congenital neutropenia reveal isoform-dependent genotype-phenotype associations. *Blood* **111**, 4954-7 (2008).
105. Ishikawa, N., Okada, S., Miki, M., Shirao, K., Kihara, H., Tsumura, M., Nakamura, K., Kawaguchi, H., Ohtsubo, M., Yasunaga, S., Matsubara, K., Sako, M., Hara, J., Shiohara, M., Kojima, S., Sato, T., Takihara, Y. & Kobayashi, M. Neurodevelopmental abnormalities associated with severe congenital neutropenia due to the R86X mutation in the HAX1 gene. *J Med Genet* **45**, 802-7 (2008).
106. Matsubara, K., Imai, K., Okada, S., Miki, M., Ishikawa, N., Tsumura, M., Kato, T., Ohara, O., Nonoyama, S. & Kobayashi, M. Severe developmental delay and epilepsy in a Japanese patient with severe congenital neutropenia due to HAX1 deficiency. *Haematologica* **92**, e123-5 (2007).
107. Smith, B. N., Ancliff, P. J., Pizzey, A., Khwaja, A., Linch, D. C. & Gale, R. E. Homozygous HAX1 mutations in severe congenital neutropenia patients with sporadic disease: a novel mutation in two unrelated British kindreds. *Br J Haematol* **144**, 762-70 (2009).
108. Schmidt, S., Claussen, U., Liehr, T. & Weise, A. Evolution versus constitution: differences in chromosomal inversion. *Hum Genet* **117**, 213-9 (2005).
109. Madan, K. Paracentric inversions: a review. *Hum Genet* **96**, 503-15 (1995).
110. Pettenati, M. J., Rao, P. N., Phelan, M. C., Grass, F., Rao, K. W., Cosper, P., Carroll, A. J., Elder, F., Smith, J. L., Higgins, M. D. & et al. Paracentric inversions in humans: a review of 446 paracentric inversions with presentation of 120 new cases. *Am J Med Genet* **55**, 171-87 (1995).
111. Mikhaail-Philips, M. M., Ko, E., Chernos, J., Greene, C., Rademaker, A. & Martin, R. H. Analysis of chromosome segregation in sperm from a chromosome 2 inversion heterozygote and assessment of an interchromosomal effect. *Am J Med Genet A* **127A**, 139-43 (2004).

112. Thomas, N. S., Bryant, V., Maloney, V., Cockwell, A. E. & Jacobs, P. A. Investigation of the origins of human autosomal inversions. *Hum Genet* **123**, 607-16 (2008).
113. Ochman, H., Gerber, A. S. & Hartl, D. L. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-3 (1988).
114. Youings, S., Ellis, K., Ennis, S., Barber, J. & Jacobs, P. A study of reciprocal translocations and inversions detected by light microscopy with special reference to origin, segregation, and recurrent abnormalities. *Am J Med Genet A* **126**, 46-60 (2004).
115. Shaw, C. J. & Lupski, J. R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet* **13 Spec No 1**, R57-64 (2004).
116. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara, E. C. M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V., Johnson, M. Q., James, T., Huw Jones, T. A., Kang, G. D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele,

S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R. & Smith, A. J. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-9 (2008).

117. Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Guo, Y., Feng, B., Li, H., Lu, Y., Fang, X., Liang, H., Du, Z., Li, D., Zhao, Y., Hu, Y., Yang, Z., Zheng, H., Hellmann, I., Inouye, M., Pool, J., Yi, X., Zhao, J., Duan, J., Zhou, Y., Qin, J., Ma, L., Li, G., Zhang, G., Yang, B., Yu, C., Liang, F., Li, W., Li, S., Ni, P., Ruan, J., Li, Q., Zhu, H., Liu, D., Lu, Z., Li, N., Guo, G., Ye, J., Fang, L., Hao, Q., Chen, Q., Liang, Y., Su, Y., San, A., Ping, C., Yang, S., Chen, F., Li, L., Zhou, K., Ren, Y., Yang, L., Gao, Y., Yang, G., Li, Z., Feng, X., Kristiansen, K., Wong, G. K., Nielsen, R., Durbin, R., Bolund, L., Zhang, X. & Yang, H. The diploid genome sequence of an Asian individual. *Nature* **456**, 60-5 (2008).

118. Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y. J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X. Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A. & Rothberg, J. M. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872-6 (2008).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 448

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-100598

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2009