1

INTRODUCTION

1.1 Genetics, genomics and functional genomics

1.1.1 The origin of genetics

Since prehistorical times, humankind has applied knowledge in heredity and variation for domestication and breeding of plants and animals, albeit the mechanism of inheritance itself remained obscure for long times. In the beginning of the twentieth century, the importance of Gregor Mendel's early work was rediscovered, and the term *genetics* (from the Greek γεννω = give birth) was coined for the science of genes, heredity, and the variation of organisms (William Bateson: Inaugural address to the Third International Conference on Plant Hybridization, London, 1906). Four years later, Thomas Hunt Morgan could show that genes reside on the chromosomes, organized structures of DNA and proteins in the cell nuclei, in a linked fashion, with alleles not segregating independently, as expected from Mendel's law of inheritance. Finally, scientists moved on towards the identification of the very nature of the genetic material in a series of pioneering experiments.

Finally, the Hershey-Chase experiment proved that the genetic information is encoded in the DNA (Hershey and Chase 1952). Before, protein was often thought to be the one material capable of encoding the complex genetic information due to its much higher variability in sequence and structure. In the 1950s, James Watson and Francis Crick elucidated the structure of the DNA double helix (Watson and Crick 1953) through the merging of data that had been accumulated by many other scientists, amongst them Erwin Chargaff's base pair rule (Chargaff 1950) and X-ray diffraction images of DNA by Rosalind Franklin (Franklin and Gosling 1953) and Maurice Wilkins (Wilkins *et al.* 1953). During the 1960s and 1970s, many laboratory techniques were developed to analyze and manipulate DNA on a molecular level, such as DNA cloning and sequencing. The number of scientists working in the field of molecular genetics greatly expanded, and we have learned a wealth of information about the blueprint of genetic information and the complex relationship between genes and traits.

1.1.2 The advent of genomics

In 1972, Walter Fiers and his team at University of Ghent were the first to determine the sequence of a gene, namely that for the bacteriophage MS2 coat protein (Min Jou *et al.* 1972). The first genome to be entirely sequenced was that of bacteriophage Φ-X174 (Sanger *et al.* 1977). Since then, the sequencing of genomes became faster every year. Before, the existence of a gene could only be proven when a mutant allele with an altered phenotype was found, and the total number of genes for a given organism could not be determined. Now, the enumeration of all open reading frames in the genome (ORFs; the coding parts that can be translated into protein) made it possible to estimate the number of genes in an organism without purifying and sequencing the proteins one by one. Consequently, genomics as the study of entire genomes appeared in the 1980s. The term "genomics" was coined in 1986 by Thomas Roderick to describe the scientific discipline of mapping, sequencing, and analyzing genomes. Beadle's and Tatum's "one gene, one enzyme" hypothesis (Beadle and Tatum 1941), after having been changed to "one gene, one polypeptide", now became something like "one gene, one chromosomal segment responsible for making a functional product".

The genomics field took off in the 1990s with the first large-scale project in the field of biology, namely the human genome project with the ambitious goal to sequence the entire three gigabases of the 22 autosomes and chromosomes X and Y. This project was significantly boosted by the enthusiastic commitment of James Watson, who stated that "only once would I have the opportunity to let my scientific life encompass the path from the double helix to the 3 billion steps of the human genome." (Watson 1990). The pace of DNA sequencing in large centers sped up from 1,000 bases per day in the mid-1980s to 1,000 bases per second in the year 2000 and is still climbing without end in sight. The first sequenced eukaryotic genome was that of *S. cerevisiae* (Goffeau *et al.* 1996), followed a year later by the first metazoan, the nematode *C. elegans*, in 1998 and by the fruit fly *Drosophila melanogaster* in 2000. Since 2001, a rough draft of the humane genome sequence has been completed further and further to reach greater overall coverage and accuracy (Gregory *et al.* 2006).

The sequence information coming from genome projects allowed scientists to bridge the cleft between genotype and phenotype in various aspects. The genomics field is currently undergoing an expansion from the mapping and sequencing of genomes to the analysis of gene functions and complex regulatory networks. To reflect this shift,

the terms "structural genomics" and "functional genomics" have emerged (Hieter and Boguski 1997). Structural genomics is the phase of genome analysis that has a clear end point: the construction of a high-resolution genetic, physical, and transcript map of an organism. Functional genomics describes the attempt to analyze the more dynamic and regulatory functions of the coding and non-coding elements of the genome.

1.1.3 The field of functional genomics

Functional genomics is a field of molecular biology that tries to make use of the wealth of data produced by structural genomic projects to describe gene and protein functions and interactions. The term is widely used, but has many different interpretations. A common basis is the development and application of genome- or system-wide experimental approaches to assess gene function, often using model organisms that allow for systematic variation of parameters. Functional genomics is characterized by high-throughput or large-scale experimental methodologies, combined with computational analysis of the results (McKusick 1997; Rastan and Beeley 1997; Ivakhno 2007). As shown in Figure 1-1, the objective of functional genomics analyses is to assess the various levels of the information flow during gene expression according to the central dogma of molecular biology.

Analysis at the DNA level:

Haplotype mapping identifies loci contributing to genetic disorders or underlying disease susceptibility due to single nucleotide polymorphisms (SNPs) or insertions/deletions (indels). Systematic generation of engineered animal strains lacking single genes allow for phenotypic analysis of the altered genotypes. Gene structures are used to predict stage- and tissue-specific alternative splicing events leading to proteins with varying domain structures.

Analysis at the transcript level:

Gene expression profiling tests when, where and for how long the expression of a gene is turned on. Well-established techniques include expression profiling by DNA or oligonucleotide microarrays, quantitative polymerase chain reaction (PCR), serial analysis of gene expression (SAGE) and RNA *in situ* hybridization. The resulting global view of expression patterns and changes in gene expression in response to

physiological shifts or manipulation of transcriptional regulators has already provided fundamental insights into the wiring scheme of complex organisms (Zou *et al.* 2000; Hahn *et al.* 2004; Wong *et al.* 2007). Knock-down of endogenous transcript levels using RNA interference (RNAi) or small interfering RNAs (siRNAs) is used to silence gene expression in cells, tissues or whole organisms (Fortunato and Fraser 2005) and to analyze phenotypic consequences at the level of transcription (Cullen and Arndt 2005).

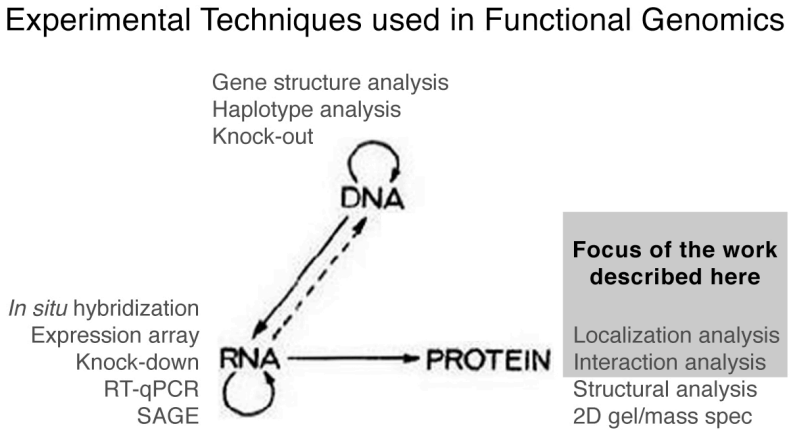

Figure 1-1. Commonly used functional genomics techniques for analysis of the different levels of genetic information flow. The ultimate goal would be to link all information obtained on an organism's genes, transcripts and proteins with their influences on biological processes and regulatory pathways in that organism to develop a detailed model of health and disease states. The work described here focused on analysis of protein localizations and protein-protein interactions of the proteins encoded on human chromosome 21. Picture modified from F.H. Crick (Crick 1958).

Analysis at the protein level:

Alterations in protein levels and post-translational modifications can be assessed through two-dimensional gel electrophoresis, followed by mass spectrometry analysis. After the complete 'parts list' of an organism has been assembled, the sequence information coming from the genomics projects can be used to clone genes into expression vectors for producing of the corresponding encoded proteins. These proteins can then be studied independent from the genome they originate from. Common goals include the elucidation of intracellular localizations through immunofluorescence detection and the identification of protein interaction partners through yeast two-hybrid (Y2H) screening.

1.1.4 Limitations and applications of functional genomics

It should be emphasized that functional genomics will not replace the established use of genetics, biochemistry, cell biology and structural studies in gaining a detailed understanding of biological mechanisms. Also, a variety of other techniques exist for the more detailed study of individual mechanisms. The extent to which any functional genomics approach actually defines the function of a protein varies depending on the method and gene involved. In general, the functional information gained can provide a framework and a starting point for further analyses (Hieter and Boguski 1997).

Functional genomics technologies have been successfully applied to the analysis of DNA and transcript sequences and their abundance in healthy and disease states of many different organisms and tissues. On the level of proteins, the achievements have been more modest, since technologies for protein analysis are intrinsically more complex due to the higher complexity of protein sequences in contrast to nucleic acid sequences. Whereas nucleic acid sequences can be readily detected using hybridization with synthetic probes (either during PCR or on microarrays), protein detection relies on the availability of specific antibodies or a marker sequence that has been artificially attached during cloning of the corresponding open reading frames (ORFs). Although the collections of antibodies and cloned ORFs are growing steadily (Brasch *et al.* 2004; Rual *et al.* 2004; Lamesch *et al.* 2007), researchers are still often limited to the analysis of smaller numbers of proteins.

A prime example for these bottlenecks is the molecular analysis of the pathogenesis of Down syndrome resulting from trisomy of human chromosome 21 (Hsa21). Genomic DNA sequencing allowed the elucidation of the DNA sequence of Hsa21, and transcriptome analyses in human trisomic cell lines and mouse models showed dysregulation of many transcripts (see paragraph 1.4.3), but analyses on the protein level are still in their infancy. In the following chapter, the diseases associated to Hsa21 will be presented and discussed. The subsequent chapter is dedicated to Down syndrome alone and describes the phenotype in more detail. Then, it will be highlighted how functional genomics analyses can contribute to the elucidation of gene functions and to the understanding of disease mechanisms in Down syndrome.

1.2 Human chromosome 21 (Hsa21)

1.2.1 Characteristics of Hsa21

Hsa21 is the smallest human autosome, containing only about 46.9 Megabases of genomic DNA. In 1959, Jérôme Lejeune reported the existence of an extra copy of Hsa21 in a child with Down syndrome (Lejeune *et al.* 1959). Since then, Hsa21 has been in the focus of investigation for researchers working on this congenital chromosome aberration. Chromosome 21 was the first autosome for which a dense linkage map (McInnis *et al.* 1993) and yeast artificial chromosome (YAC) physical maps were developed (Korenberg *et al.* 1995).

As a result of the Human Genome Project, the nucleotide sequence of Hsa21 was finished In May 2000, 41 years after Lejeune's discovery. 33,546,361 bases of high-quality sequence from the long arm of Hsa21 (called 21q) were published (Hattori *et al.* 2000), resulting in a coverage of 99.7% and an overall accuracy exceeding 99.995%. The sequence of Hsa21 contains ~38% interspersed repeats. Among these, there are 10.8% SINEs, 15.5% LINEs, 9.2% LTR elements and 2.4% other DNA elements. Only ~3% of the total sequence is coding for proteins.

1.2.2 Gene content and gene functions

The initial Hsa21 gene catalog estimated 225 genes (127 known genes and 98 predicted genes) and 59 pseudogenes on the long arm of the chromosome (21q). Based on additional information including sequencing of expressed sequence tags (ESTs), reverse transcription PCR (RT-PCR) and comparative genomics analyses, a recent estimate from our group now counts 284 protein-coding genes on Hsa21. Out of these, 206 contain a known full open reading frame, 37 genes have only a partial ORF, 15 genes are without ORF >50 bp and 26 genes belong to the clusters of keratin-associated proteins (KRTAPs; Shibuya *et al.* 2004).

The study of gene functions and diseases makes wide use of orthologous genes in model organisms such as *S. cerevisiae* (baker's yeast), *C. elegans* (nematode) and *D. melanogaster* (fruit fly). Unfortunately, only 15% of the 284 genes are conserved among humans and all these model organisms (16% have an ortholog in yeast, 32% in the nematode and 35% in the fruit fly). Many genes on Hsa21 have been analyzed for their function, but despite intense research, almost half of the genes cannot be

categorized according to their molecular function or biological process (Figure 1-2), mainly because the emphasis has been put on a limited number of 'popular' genes.

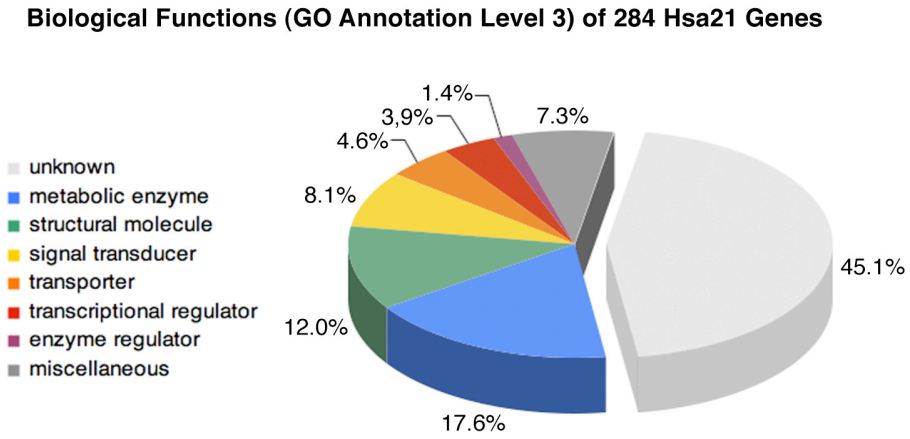

Figure 1-2. Biological functions of 284 Hsa21 genes according to Gene Ontology annotations. 128 genes from Hsa21 (45.1% of 284 genes) have not been functionally characterized. GO annotations for 'Biological Process' at GO level 3 were retrieved using RefSeq nucleotide accession numbers for the Hsa21 transcripts in the annotation tool FatiGO+ available at the Babelomics web site (ttp://babelomics.bioinfo.cipf.es).

1.2.3 Diseases associated to Hsa21

To date, 32 disease loci with clinical synopsis have been mapped to Hsa21. Detailed descriptions for these can be found at the OMIM database (Online Mendelian Inheritance in Man, available at http://www.ncbi.nlm.nih.gov). The causes for development of pathologies range from mutations in single genes to chromosomal rearrangements and chromosomal aberrations, including monosomy and trisomy 21 (resulting in Down syndrome).

Diseases resulting from mutations in single genes

All diseases resulting from known mutations in single chromosome 21 genes are listed in Table 1-1.

Table 1-1. Diseases resulting from mutations in single Hsa21 genes (see also next page).

Disease (acronym)	Affected gene	Protein functions	Consequences of mutations
Alzheimer disease (AD)	APP	Diverse functions for APP have been proposed both as a membrane receptor and as a diffusible factor	Accumulation of neuro-fibrillary tangles and amyloid plaques associated with progressive dementia
Amyotrophic lateral sclerosis (ALS) or 'Lou Gehrig disease'	SOD1	Cu^{2+}/Zn^{2+} binding enzyme responsible for removal of free superoxide radicals	Death of motor neurons in brain, brainstem, and spinal cord resulting in paralysis

Disease (acronym)	Affected gene	Protein functions	Consequences of mutations
Autoimmune polyendocrinopathy - candidiasis - ectodermal dystrophy syndrome (APECED)	AIRE	Regulates ectopic gene expression in the thymus for tolerance of autoantigens	Autoimmune activity against more than one endocrine organ
Enterokinase deficiency	PRSS7	Activation of pancreatic proteolytic proenzymes	Malabsorption and growth failure
Formiminotransferase deficiency	FTCD	Channeling of 1-carbon units from histidine degradation to the folate pool	Formiminoglutamate in the urine, megaloblastic anemia and mental retardation
Glucocorticoid deficiency 2 (GCCD2)	MRAP	Melanocortin-2 receptor accessory protein	Severely impaired adrenal response to ACTH
Holocarboxylase synthetase (HLCS) deficiency	HLCS	Attachment of biotin to biotin-dependent metabolic carboxylase enzymes	Lactic acidosis, alopecia, keratoconjunctivitis, perioral erosions and seizures
Homocystinuria	CBS	Catalysis of the first step in the transsulfuration pathway, homocysteine-cystathionine	Defects in the eye, the central nervous system, skeletal and vascular systems
Knobloch syndrome (KNO)	COL18A1	Extracellular matrix protein and precursor of endostatin, a potent antiangiogenic protein	High myopia, retinal degeneration and macular abnormalities
Leukocyte adhesion deficiency (LAD)	ITGB2	Adhesion molecule mediating cell-cell and cell-matrix contacts during inflammation	Recurrent bacterial infections, impaired pus formation and wound healing
Long QT syndrome 5 (LQT5) or Jervell and Lange-Nielsen syndrome (JLNS)	KCNE1	Subunit of K^+ channels for the delayed-rectifier potassium current in cardiac myocytes	Cardiac arrhythmias resulting in recurrent syncopes, seizure or sudden death; deafness
Long QT syndrome 6 (LQT6)	KCNE2	Subunit of K^+ channels for the delayed-rectifier potassium current in cardiac myocytes	Cardiac arrhythmias resulting in recurrent syncopes, seizure or sudden death
Myoclonic epilepsy of Unverricht and Lundborg (ULD, or EPM1)	CSTB	Cysteine protease inhibitor protecting against proteinases leaking from lysosomes	Formation of amyloid fibrils, progressive epilepsy, ataxia and eventually dementia
Non-syndromic deafness types 8 and 10 (DFNB8 and DFNB10)	TMPRSS3	Transmembrane serine protease activating the epithelial Na^+ channel ENaC	Disruption of proteolytic activity resulting in non-syndromic deafness
Non-syndromic deafness type 29 (DFNB29)	CLDN14	Formation of tight junctions for transepithelial electrical resistance in the inner ear	Failure in tight junction formation resulting in non-syndromic deafness
Susceptibility to hypertriglyceridemia	LIPI	Lipase enzyme involved in lipid metabolic processes not yet characterized to detail	Hypertriglyceridemia promoting coronary heart disease and diabetes mellitus
Ullrich congenital muscular dystrophy (UCMD) and Bethlem myopathy	COL6A1, COL6A2	Extracellular matrix proteins thought to anchor the basal lamina to the interstitium	Congenital muscle weakness, proximal joint contractures, distal joint hyperextensibility

Information on the identity of the affected genes as well as protein functions and consequences of mutations has been retrieved from the Entrez Gene and OMIM (Online Mendelian Inheritance in Man) databases, available at http://www.ncbi.nlm.nih.gov.

Diseases resulting from chromosomal rearrangements

Presented below are the two main pathological chromosomal rearrangements involving Hsa21:

Acute lymphoblastic leukemia (ALL) is the most common malignancy seen in children. It is characterised by the overproduction and continuous multiplication of immature lymphoblasts in the bone marrow. ALL can result from malignantly transformed B cell progenitors. At least 25% of cases with a B progenitor

immunophenotype are associated with a translocation between chromosomes 12 and 21, leading to the formation of a chimeric gene (AML1-TEL fusion oncogene) that produces a transforming protein. The normal genes, now called RUNX1 and ETV6, respectively, are involved in hematopoietic development. The fusion protein consists of the N-terminal helix-loop-helix domain of ETV6 encoded on Hsa12p13 fused with a nearly complete RUNX1 transcription factor encoded on Hsa21q22.3 (Pui 1995).

Acute myeloid leukemia (AML) is a cancer of the myeloid line of white blood cells, characterized by the rapid proliferation of abnormal leukocytes, which accumulate in the bone marrow and interfere with the production of normal blood cells. The most frequent cause is a translocation between chromosomes 8 and 21, leading to the formation of the chimeric AML1-ETO oncogene. The substitution of the chromosome 8-derived ETO protein for the multifunctional C terminus of RUNX1 precludes targeting of RUNX1 to specific subnuclear domains. Instead, the AML1/ETO fusion protein is redirected by the ETO component to alternate nuclear matrix-associated foci, disrupring normal targeting of gene regulation (McNeil *et al.* 1999).

Disorders resulting from aneuploidy

Aneuploidy is a state in which one or more chromosomes have been gained or lost. Segmental aneuploidies involve only parts of chromosomes. Acquired aneuploidies are often associated with tumorigenesis, where extra copies of a complete or part of one or more chromosomes result in higher gene dosage of certain oncogenes in a subpopulation of cells. Congenital aneuploidies, on the other hand, are generally present in all cells of an organism and are a frequent cause for abnormal development. A large fraction of miscarriages is associated to non-viability of the embryo due to aneuploidy. Four types of aneuploidies have been described for Hsa21:

Monosomy 21 was believed to be the only autosomal monosomy compatible with life until 1976, when it was suggested that cases with apparently full monosomy 21 are in fact partial deletions resulting from either cryptic translocation or mosaicism (Schinzel 1976). Full monosomy 21 has been prenatally diagnosed, but seems not to be found in liveborns (Joosten *et al.* 1997). Phenotypic features that are frequently observed include pre- and postnatal growth retardation, low-set ears, hypertonia, heart defect and mental retardation (Chettouh *et al.* 1995).

Ring chromosome 21 (r21) is formed through the breakage and reunion of short- and long-arm regions of Hsa21, resulting in deletion or duplication of varying amounts of Hsa21q22.1 to 21qter, depending on the intermediate steps of (r21) formation, for which three mechanisms have been proposed (McGinniss *et al.* 1992). The phenotype of patients correlates with the extent of deletion or duplication of Hsa21 sequences and resembles either that of monosomy 21 or that of trisomy 21 (see below).

Partial trisomy 21 is a rare event responsible for only up to 5% of all cases of Down syndrome (DS, described below). One cause of partial trisomy 21 can be very rare events of *de novo* intrachromosomal duplications, for which only a few cases have been reported (Delabar *et al.* 1992). Most frequently, partial trisomy 21 arises from so-called Robertsonian translocations. These non-reciprocal whole-arm translocations occur when the long arms of two acrocentric chromosomes fuse at the centromere, and the two short arms are lost. This can happen in a parent carrying an asymptomatic balanced translocation, who can then transmit the translocated chromosome to his offspring, resulting in DS (familial form of DS). In humans, such translocations occur among the acrocentic chromosomes 13, 14, 15, 21 and 22. In Down syndrome, usually 14 and 21 are affected.

Full trisomy 21 is caused by a meiotic nondisjunction event, leading to a gamete with an extra copy of Hsa21. When combined with a normal gamete from the other parent, the child has 47 chromosomes, with three copies of Hsa21 (trisomy 21), and shows the characteristics of DS (see below). Most cases of trisomy 21 originate from nondisjunction in the maternal gamete, with the risk for a trisomic pregnancy increasing with maternal age. There also rare cases of Down syndrome parents having trisomy 21 children (James *et al.* 1998). In these cases, the ovaries of the mother were trisomic for Hsa21, leading to a secondary nondisjunction during gametogenesis. Such Down syndrome trisomies are indistinguishable from Down syndrome trisomies created through meiotic nondisjunction.

The phenotypic effects of an extra amount of Hsa21 DNA material vary from individual to individual, depending on the extent of the extra copy, the genetic background, environmental factors and also perhaps stochastic effects. A number of aspects concerning DS will be discussed in more detail in the following chapter.

1.3 Trisomy 21 and Down syndrome (DS)

1.3.1 Incidence and impact

Chromosome abnormalities occur in approximately one in 200 live births and account for a significant fraction of malformations and neonatal deaths. Trisomy 21, resulting in DS, accounts for about one third of these chromosome abnormalities. The median incidence rate of DS is estimated to be one in 700 live births, with an increase in frequency that shows a strong correlation with the mother's age. At maternal age 20 to 24, the risk is 1 in 1,490; at age 40 the risk is 1 in 60, and at age 49 the risk is 1 in 11 (Hook 1981a). Apart from maternal age, no other risk factors are known. Although the DS risk increases with maternal age, 80% of children with Down syndrome are born by women under the age of 35 (estimate from the National Down Syndrome Center, USA).

Despite the fact that 65-80% of all fetuses with trisomy 21 are lost spontaneously (Byrne and Ward 1994), there are currently more than two million people with Down syndrome worldwide, many of whom have to be institutionalized throughout their life time. A public calculation has estimated the 'lifetime per capita costs of a person affected with Down Syndrome' to range between $150,000 and $350,000 in the United States (http://www.epa.gov/oppt/coi/docs/III_8.pdf), which is a huge financial responsibility to take on for parents and families as well as society, not including individual suffering and other implications. Nevertheless, it should be stated that "today, individuals with Down syndrome can be active participants in the educational, social and recreational aspects of our communities. […] more teens and adults with Down syndrome each year are graduating from high school, going to college, finding employment and living independently" (National Down Syndrome Society (USA), http://www.ndss.org).

1.3.2 History and genetics

The history of Down syndrome

Possibly the earliest evidence that we have of Down syndrome is a terracotta dated to the Tolteca culture of Mexico that was created around the year 500 AD (Figure 1-3A). In this terracotta, it is possible to identify the short palpebral fissures, oblique eyes, midface hypoplasia and open mouth with macroglossia, findings that define the

face of a person with DS (Martinez-Frias 2005). For comparison, the characteristic craniofacial appearance of a contemporary individual with full trisomy 21 is shown in Figure 1-3B.

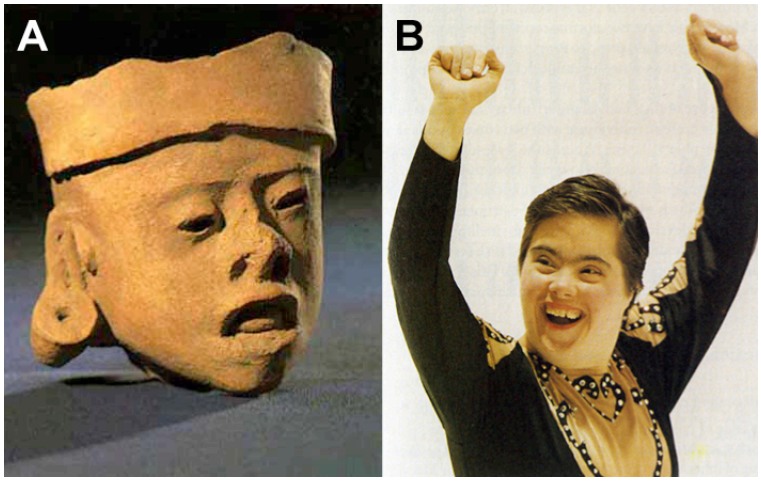

Figure 1-3. A presumed historical representation of an individual with Down syndrome resembles contemporary persons with DS. A: Terracotta from the Tolteca culture of Mexico (500 AD). B: Person with Down syndrome competing in the paralympic games (1996 AD). Pictures taken from Martinez-Frias (2005) and the textbook *'Genetics – analysis and principles'* (Brooker 1999).

It was not until the late nineteenth century, however, that Down syndrome was recognized as an entity separate from other forms of mental retardation. The first accurate description was published by John Langdon Down in a scholarly work that earned him the reputation as the "father" of this syndrome. The original essay "Observations on an ethnic classification of idiots" from 1866 has been reprinted in the journal Mental Retardation (Down 1995). Since the beginning of the twentieth century, many people with Down syndrome were excluded from the society and housed in institutions or colonies; the associated medical problems were left untreated, and many people died early in their life. It was assumed that DS was caused by certain combinations of inheritable factors, or from trauma during pregnancy.

Genetics of Down syndrome

Finally, in 1959, after the diploid human chromosome number had been determined to be 46 (Ford and Hamerton 1956), it was recognized that cells from people with Down syndrome contained an additional copy of chromosome 21. This so-called trisomy 21 was then the first chromosomal abnormality found associated with a human disease (Jacobs *et al.* 1959; Lejeune *et al.* 1959).

As described above, Down syndrome can be caused by a full trisomy 21 (95% of cases), by a partial trisomy 21 (5% of all cases), or by a ring chromosome 21 (very rare).

For **full trisomy 21**, the meiotic origin of Hsa21 nondisjunction could be elucidated with the help of DNA polymorphisms at loci in the pericentromeric region on the long arm of Hsa21 in 200 families (Antonarakis *et al.* 1992). Most cases of nondisjunction were of maternal origin (95%), only few were derived from the father (5%). In maternal nondisjunction during oogenesis, errors in meiosis I were responsible for about three quarters, whereas errors in meiosis II are associated to about one quarter of nondisjunctions. These figures are reversed in the case of paternal nondisjunction during spermatogenesis.

Only 1-2% of all DS cases were found to be due to **mosaicism**, where the Hsa21 nondisjunction occurred during one of the early cell divisions after fertilization, so that both normal and trisomic cells coexist in the affected individual (Mikkelsen 1977). The DS phenotype due to mosaicism depends on the extent and localization of trisomic cells in the body.

Inherited **partial trisomy 21** (3-4% of DS cases) was found to be mostly due to unbalanced Robertsonian translocation in parental germ cells (Hook 1981b). In all cases analyzed, the *de novo* translocation t(14q;21q) was of maternal origin. In DS cases with *de novo* dup(21q) chromosomes, these were mostly isochromosomes (genetically identical arms due to transversal split of the centromere during cell dicision) and not translocations between different chromosomes (Antonarakis *et al.* 1990).

Very few cases of *de novo* **microduplications** leading to DS have been reported, leading to extra copy of some genes on Hsa21 within the same chromosome. These cases can show some, but not all characteristics of DS, depending on the duplicated region. Only a few cases have been reported and analyzed (Huret *et al.* 1987; Delabar *et al.* 1992).

1.3.3 Clinical characteristics

The phenotype resulting from trisomy 21 is associated with several clinically recognizable features, symptoms and characteristics, which often occur together. It is therefore called Down *syndrome*. Most organs in the body can be affected, and more

than 80 known clinical hallmarks have been described (Epstein 1990). Each DS patient shows a unique combination of these signs with different degrees of severity. Some features of DS are constantly present in all affected individuals, while other variable signs appear only in subpopulations.

Constant features include characteristic appearance of the face, reduced size and altered morphology of the brain, and mental retardation of variable severity. Variable signs include musculoskeletal and extremity abnormalities, gastrointestinal anomalies, speech and motor delay, conductive type hearing loss (in 65-90% of cases), nystagmus and strabismus (in 40-60%), and cardiac malformations (in 40%). Moreover, individuals with DS have a higher susceptibility for a number of health problems, including respiratory trouble, infections, specific leukemias, and early onset Alzheimer disease.

Since each of these signs could be found individually in the non-DS population, the clinical diagnosis of DS, physicians evaluate the simultaneous presence of a number of phenotypic features that discriminate single symptoms from the syndrome. Indexing different numbers and types of features has been proposed, for example the use of ten informative signs for non-overlapping classification into DS and non-DS (Jackson *et al.* 1976), or the use of eight phenotypic findings of DS to categorize individual cases with 99.9% confidence (Rex and Preus 1982). These eight features include three dermatoglyphic traits (hallucal and forefinger pattern, and palmar triradius), two measurements of physical traits (ear length and internipple distance), and three other clinical findings (Brushfield spots in the eye, wide-spaced first toe, and excess back neck skin).

In the following paragraphs, the main morphological and clinical features associated with DS will be described in more detail. These features are ranging from altered appearance over severe malformations of organs and disturbance of the endocrine system to more subtle changes in brain morphology and body biochemistry resulting in a number of associated disorders.

Physical appearance

Individuals with Down syndrome may have some or all of the following morphological characteristics. A distinctive craniofacial appearance can be observed (Figure 1-4A) due to microbrachycephaly with flat facial profile, small nose, depressed nasal bridge and upslanted palebral fissures with epicanthal fold (Figure 1-4B), a protruding

tongue, small ears (Figure 1-4C), and a short neck (Figure 1-4D). The stature in general is short and often obese, with sloping shoulders. Characteristic features of the extremities include broad hands and feet with short digits, incurved fifth fingers and transverse palmar creases (Figure 1-4E), and a wide space between first and second toe.

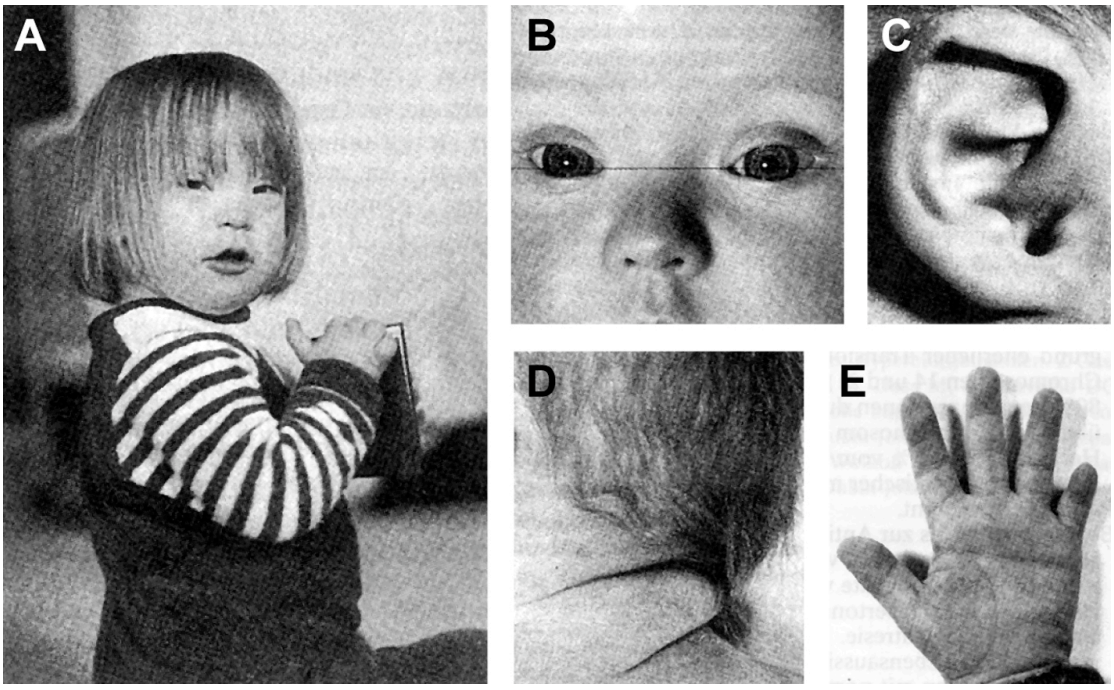

Figure 1-4. Physical appearance of individuals with DS. A: 20 months old girl with typical facial expression; B: depressed nasal bridge and almond shape to the eyes caused by an epicanthic fold of the eyelid, eyes with Brushfield spots (white spots in outer third of the iris); C: small ears; D: short neck with skin creases; E: hand with short digits and transverse palmar crease. (Pictures taken from "Die klinischen Syndrome", Bernfried Leiber and Gertrud Olbrich 1996, Bd. 1 "Krankheitsbilder").

Cardiovascular defects

About half of all affected individuals with DS have a congenital heart defect. The most common malformation is an atrioventricular septal defect (AVSD), a cardiac anomaly that occurs when the superior and inferior endocardial cushions fail to close completely, resulting in incomplete formation of the atrial and ventricular valves and septa. So far, only the cell adhesion molecule CRELD1 was identified and characterized as a candidate gene for the AVSD locus mapping to chromosome 3p25 (Robinson *et al.* 2003; Zatyka *et al.* 2005). Analysis of the CRELD1 gene from individuals with non-trisomy 21-associated AVSD identified heterozygous missense mutations in 6% of this population (Robinson *et al.* 2003). Also, it has been proposed

that CRELD1 mutations contribute at least in part to the occurrence of AVSD in some DS individuals (Maslen *et al.* 2006).

Other cardiac defects observed in DS include patent ductus arteriosus (failure to establish the mature circulatory pattern at birth by closing the lung bypass through the ductus arteriosus), and tetralogy of Fallot (ToF), the most common cause of blue baby syndrome. The four heart formations in ToF are: (1) ventricular septal defect, (2) pulmonic stenosis, (3) overriding aorta, and (4) right ventricular hypertrophy.

Gastrointestinal malformations

Down Syndrome is one of the most common predisposing conditions for a group of serious gastrointestinal anomalies. Duodenal obstruction with or without pyloric stenosis (5%), tracheo-esophageal fistula (3%), imperforate anus (3%) and Hirschsprung's disease (2%) are the most prevalent lesions (Buchin *et al.* 1986). Associated congenital heart disease, especially endocardial cushion defects, and the frequent occurrence of pneumonia contributes to a relatively high mortality rate in the patient group with duodenal stenosis. A possible common pathway to the observed anomalies might be the enhanced epithelial adhesiveness demonstrated by *in vitro* experiments with fibroblasts, but the understanding of the morphogenetic mechanisms responsible for this range of abnormalities is far from being clear (Levy 1991).

Endocrine abnormalities

The prevalence of thyroid disease is increased in Down syndrome. Thyroid function tests reveal hypothyroidism in 16-35% of DS patients (Pueschel *et al.* 1991a). Patients with hypothyroidism before the age of eight years usually do not have thyroid autoantibodies at diagnosis, whereas most patients who develop hypothyroidism after this age have thyroid autoantibodies (Karlsson *et al.* 1998). Although these findings hint at an autoimmune basis of the DS-associated thyroid disorder, many DS individuals with autoantibodies nevertheless show a normal thyroid function.

General neurological phenotype

The major neurological deficits in DS individuals are mental retardation, hypotonia and neuropathological changes leading to an early onset of Alzheimer disease

(Wisniewski *et al.* 1985; Epstein 1990), but many other neurological dysfunctions can be found. Among these are difficulties in recalling auditorially presented verbal material (Marcell and Armstrong 1982; Marcell and Weeks 1988), whereas non-retarded individuals typically show better short-term memory for brief sequences of auditory than visual information (called the *modality effect*). Epileptic seizures occur with increased incidence (8%) in DS adolescents, with 40% of seizure patients beginning to have seizures before the age of one year, and another 40% starting with seizure activity in their third decade of life (Pueschel *et al.* 1991b). Another interesting finding was that individuals with Down syndrome express pain or discomfort more slowly and less precisely than the general population (Hennequin *et al.* 2000), which may be associated with impaired peripheral somatosensory function (Brandt and Rosen 1995). Numerous other cellular and systemic abnormalities in the DS nervous system have been reported, but no known specific set of factors could so far explain the range of neurological changes in DS.

Pathophysiology of the brain

For some time, researchers have been analyzing the size and morphology of brains from DS individuals in order to correlate the cognitive impairments with altered structures in their brains. Through morphometric studies, it has been noted early that brains of individuals with DS show a reduced number of neurons, potentially due to arrest of neurogenesis and synaptogenesis in early stages of fetal brain development, and there is evidence for abnormalities in neuronal differentiation and migration in fetal and infant brain (Wisniewski *et al.* 1984). While early post-mortem analyses suggested general brain atrophy, more recent *in vivo* anatomic imaging studies using MRI (magnetic resonance imaging) do not indicate gross atrophy or failed brain development. Rather, morphological changes in the brain appear to be variable from patient to patient, similar to other phenotypic consequences of DS. Caution must be taken in interpreting studies of brain structure in subjects with DS, as many have employed small numbers of subjects and relatively low resolution image acquisition techniques.

What can be consistently observed in persons with DS are smaller volumes for the whole brain, cerebral cortex, white matter, and especially the cerebellum (Weis *et al.* 1991). DS subjects have substantially smaller cerebral and cerebellar hemispheres, ventral pons, mammillary bodies, and hippocampal formations (Raz *et al.* 1995). Shown in Figure 1-5 are the results of a voxel-based morphometric study using MRIs

of 19 non-demented individuals with DS and 11 age-matched controls. The DS group showed reduced grey matter volume predominately in the cerebellum, left medial frontal lobe, right superior/middle temporal lobe, and throughout a large portion of the cingulate gyrus. Significant increases in grey matter volume were observed in a superior/caudal portion of the brainstem, which extended superiorly and laterally into the left parahippocampal gyrus (Figure 1-5, left panel). Regarding the white matter, the DS group showed less white matter volume throughout the brainstem and more white matter volume in both the left and right parahippocampal gyrus (Figure 1-5, right panel).

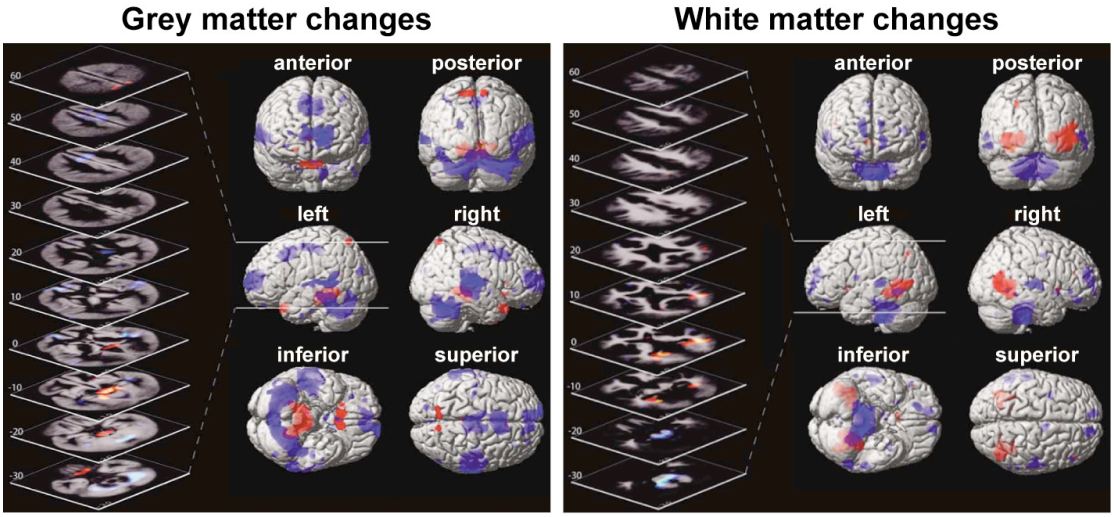

Figure 1-5. Regionally specific changes in volumes of grey and white matter in the brains of individuals with Down syndrome. On the left-hand side of each figure are a series of 10-mm axial cuts though the brain with respective z coordinates. Surface renderings (in six orthogonal views) of regional changes are shown on the right-hand side of the figure. Significant decreases and increases in DS regional brain volume are shown in blue and red, respectively. Left panel: regionally specific changes in grey matter volume (includes nerve cell bodies, glial cells, and capillaries). Right panel: regionally specific changes in white matter volume (composed of myelinated axons). Pictures modified from White *et al.* 2003.

Ultrastructural analysis of brains from DS children showed retardation of brain growth, maturation delay, and cortical dysgenesis. Newborns showed fewer neurons (20-50% less), lower neuronal densities, and neuronal distribution, especially of cortical layers II and IV (Wisniewski 1990).

The molecular basis responsible for developmental brain abnormalities in DS still remains largely elusive. There have been reports suggesting abnormalities reflecting increased risk of apoptosis in DS brains. Cultured neurons from both patients and model animals are more vulnerable to apoptosis. Accordingly, increased expression of several apoptosis-related genes (p53, fas, ratio of bax to bcl-2, GAPDH) in DS

brains has been reported (Sawa 1999). It has been shown in primary neuronal cultures from ETS2 transgenic mice that moderate overexpression of this Hsa21-encoded transcription factor induces apoptosis that is dependent on p53 (Wolvetang *et al.* 2003). Overproduction of reactive oxygen species could also cause increased apoptosis in DS tissues (Sawa 1999). Besides, there is evidence that the Hsa21-encoded APP gene regulates neural stem cell biology in the brain, and that altered APP metabolism in DS may have implications for brain abnormalities (Kwak *et al.* 2006). Various studies have suggested a role of other Hsa21 genes as causing agents for pathological changes in the DS brain, but a consistent picture of the overall development of brain abnormalities in DS still has to emerge. Clearly, basic neurophysiological and functional information is needed to understand the mechanisms of mental retardation in DS.

Mental retardation

DS is the most common non-heritable genetic cause of mental retardation, and contributes to about 30% of moderate to severe cases of mental retardation. Cognitive dysfunctions include significant delay in nonverbal cognitive development, accompanied by additional, specific deficits in speech, language production, and auditory short-term memory in infancy and childhood (Chapman and Hesketh 2000). The IQ declines during childhood and continues to decrease during adolescence and adulthood (Carr 1994). IQ scores in adults range from 20 (severe mental retardation) to 85 (low normal), with an average IQ of 45. This means that the overall learning abilities of an adult DS individual are equivalent to a six-year-old child without Down syndrome.

Early-onset Alzheimer disease

It is well known that many adults with DS over the age of 40 years display Alzheimer's disease (AD). Characteristic features observed in the brain include senile plaques, Aβ amyloid deposits, neurofibrillary tangles and granulovascular degeneration. It has been found that up to the age of 60, the prevalence of dementia doubles with each 5-year interval. Up to the age of 49, the prevalence is 9%, from 50 to 54, it is 18%, and from 55 to 59, it is 32%. In the age category of 60 and above, there is a decrease in prevalence of dementia to 26% due to four times higher mortality among elderly demented DS patients in comparison with non-demented patients (Coppus *et al.* 2006). Other studies find even higher prevalence rates of up

to 75% at the age of 60 (Zigman *et al.* 1997). Taken together, these findings classify the dementia observed in DS as early-onset (or presenile) Alzheimer's disease.

The triplication of the gene encoding the amyloid precursor protein (APP) in DS has been proposed as the main cause for the close association between DS and AD, but since increased levels of APP are not present in non-DS Alzheimer's disease, additional factors must account for the amyloid deposition (Rumble *et al.* 1989). Other genes associated with AD pathogenesis in DS patients include the E4 allele of ApoE, which is associated with the late-onset familial and sporadic forms of Alzheimer's disease (Corder *et al.* 1993), the involvement of the triplicated beta-site APP-cleaving enzyme 2 (BACE2, encoded on Hsa21q22.3) which shows elevated expression in brains of patients with DS (Motonaga *et al.* 2002), and a prion protein (PRNP) polymorphism associated with faster decline of cognitive ability in elderly DS patients (Del Bo *et al.* 2003).

Premature aging

Apart from mental deficiency and anatomical abnormalities, premature aging is characteristic for DS. Individuals with DS show age-related declines (reduced skin elasticity, fenestration of cardiac valves, premature cataracts) up to ten years earlier than the general population (Brown 1987). Regarding the brain, persons with DS have features of premature aging detectable at routine MRI evaluated for presence of three markers of brain aging: atrophy, white matter lesions, and T2 hypointensity of the basal ganglia (Roth *et al.* 1996). Cognitive function often declines with age in individuals with DS, especially for tasks that require attention and planning. Also, thymus integrity and function declines with age. Premature aging of the thymus-dependent immune system and the presence of 'premature T-cell aging' may predispose Down syndrome individuals to organ-specific autoimmunity and age-related disorders (Seger *et al.* 1977; Rabinowe *et al.* 1989).

Since DS cells are more sensitive to reactive ogygen species, it has been suggested that a disturbance of balance of reactive oxygen species may result in premature aging, possibly due to the location of the gene for (Cu,Zn)-superoxide dismutase (SOD1) on Hsa21 (Kedziora and Bartosz 1988). This abundant copper/zinc enzyme found in the cytoplasm converts superoxide into hydrogen peroxide and molecular oxygen, a reaction that is essential not only for protection against reactive oxygen species, but also for H_2O_2-mediated oxidation and inactivation of phosphatases in growth factor signaling (Juarez *et al.* 2008).

1.3.4 Prenatal diagnosis and screening

Prenatal diagnosis of DS

Diagnostic testing for DS by maternal age started three decades ago, when **amniocentesis** was offered to older women. This test, performed between weeks 16 and 20 of pregnancy, involves sampling of amniotic fluid via a needle inserted through the abdomen. Subsequently, isolated cells can be analyzed for the presence of chromosome aberrations. Later on, **chorionic villus sampling**, which can already be performed between weeks 8 and 12, was added to the repertoire of prenatal DS diagnosis. Here, a placenta biopsy is taken through the cervix or the abdomen. If the pregnancy is already beyond week 20, **percutaneous umbilical blood sampling** is used where a blood sample is retrieved from the umbilical chord via a needle. All invasive diagnostic tests carry a risk of complications (preterm labor and miscarriage) of about 1:100 (chorionic villus sampling even higher). For these reasons, prenatal screening procedures have been established to pre-estimate the risk of a fetus having DS before proceeding to the invasive diagnostic tests.

Prenatal screening for DS

Biomarkers found in the serum of pregnant women can serve as indicators of trisomic pregnancies. After it was observed that fetal chromosome abnormalities were often associated to low maternal serum level of alpha-fetoprotein (AFP; Merkatz *et al.* 1984), AFP became a biomarker used in a **single-marker screening test** along with maternal age. Other markers were subsequently discovered: the association between elevated serum chorionic gonadotropin (hCG) and DS led to the **double test** (AFP and hCG), and a third marker for DS was unconjugated estriol (uE3), found to be lower in trisomic pregnancies (Canick *et al.* 1988), leading to the so-called **triple test**, which is commonly used (Wald *et al.* 1997). A more recent addition to the second trimester serum markers has been Inhibin-A, which is present at higher levels in DS pregnancies (Van Lith *et al.* 1992). The **quadruple test** (the combination of AFP, hCG, uE3 and Inhibin) is currently the most popular second trimester screening test in the USA (D'Alton and Cleary-Goldman 2005).

Generally, serum tests are combined with **nuchal translucency testing** (performed during weeks 11 and 14), where the clear space in the folds of tissue at the developing neck is measured via ultrasound imaging. Fetuses with chromosomal abnormalities accumulate more fluid there, letting the space appear larger. The

combined diagnostic power of quadruple test and nuchal translucency can reach a 85% detection rate of DS, with a false positive rate of only 0.9%.

1.4 Molecular analysis of Down syndrome

1.4.1 Defining a Down syndrome critical region (DSCR)

Most cases of DS are caused by complete trisomy 21. There is, however, a small fraction of cases where a partial trisomy 21 results in many of the phenotypic changes present in full DS. As a consequence of this observation, researchers have tried to narrow down the region on Hsa21 in order to identify a Down syndrome critical region (DSCR) that is responsible for all or the majority of the clinical phenotypes associated with DS. During the 1980s, the first molecular markers became available for cytogenetic analysis, together with high-resolution banding techniques. Several groups then reported a number of studies on a total of about 30 patients with partial trisomy 21 (Antonarakis *et al.* 1989; Rahmani *et al.* 1989; Korenberg *et al.* 1994).

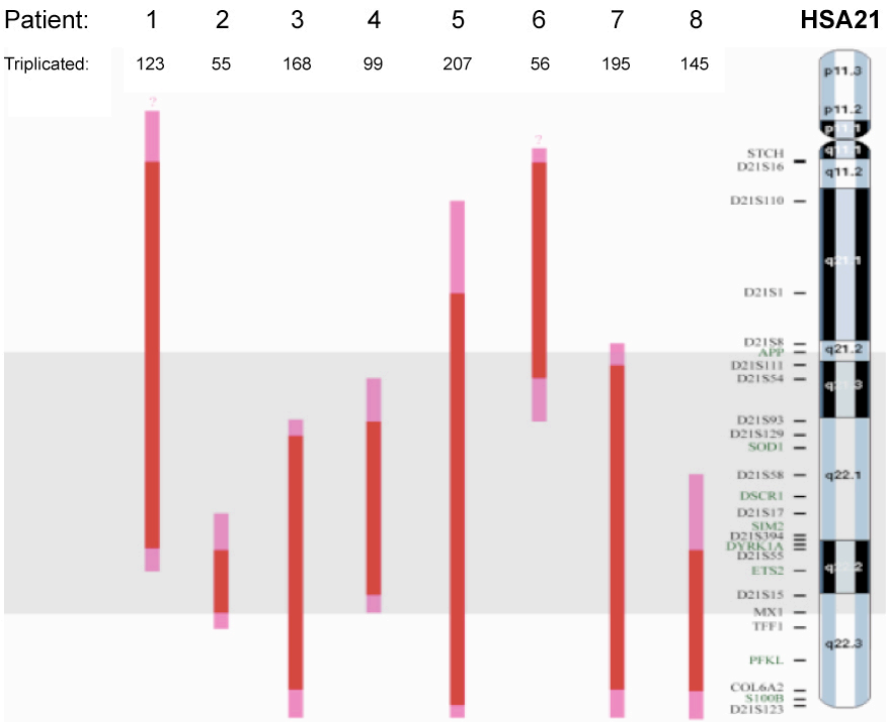

Figure 1-6. Partial trisomies of eight patients used for defining a DSCR. Right: Ideogram of human chromosome 21 shows the locations of important genes and genetic markers. Left: Partial trisomies of eight patients with the corresponding numbers of triplicated genes. Patients 1 and 2 correspond to cases FG and IG (Rahmani *et al.* 1989); patients 3-6 correspond to cases DUP21SOL, DUP21KJ, DUP21DS and DUP21GY (Korenberg *et al.* 1994); patients 7 and 8 correspond to cases GM1413 and MP01 (Antonarakis *et al.* 1989). Red stretches represent regions between triplicated markers, whereas the pink stretches represent segments not assessed for gene dosage, since they are located between a triplicated marker and a marker with normal gene dosage. This figure is based on Figure 2-3 from the dissertation of M. Sultan (accessible online at http://www.diss.fu-berlin.de/2007/244/).

When these initial reports were revisited later on, it was found that only eight of these cases carried a partial trisomy 21 as sole chromosomal abnormality (Patients 1-8 in Figure 1-6)(Antonarakis *et al.* 1989; Rahmani *et al.* 1989; Korenberg *et al.* 1994), while the other cases were shown to carry additional chromosomal aberrations, preventing formal conclusions on genotype-phenotype correlations. The smallest DSCR was proposed by a study on two DS patients (Rahmani *et al.* 1989) and was estimated to span only between 400 and 3,000 kilobases around the marker D21S55 (Patients 1 and 2 in Figure 1-6). Moreover, different studies identified other subregions as "minimal" or "critical" sites necessary and sufficient to produce the clinical condition of DS. As apparent from Figure 1-6, it has not been possible to univocally map the phenotypic features of DS to a region common to all patients with partial trisomy 21.

Also, when a phenotypic map was constructed that included 25 features and assigned regions of 2-20 Megabases as likely to contain the responsible genes, evidence arose for a significant contribution of genes outside the D21S55 region to the DS phenotypes, which speaks against a single DS chromosomal region responsible for most of the DS phenotypic features (Korenberg *et al.* 1994). Another hint came from the study of mouse models with trisomy for the DSCR (Olson *et al.* 2007), which found that the DSCR is necessary but not sufficient for brain phenotypes of trisomic mice (see also below). However, one should keep in mind that there are significant differences between mice and humans in terms of developmental times and physiologic aspects, making data transfer difficult.

To date, it must be stated that there is not sufficient evidence to conclude that individual loci on Hsa21 are responsible for specific phenotypic aspects of DS.

1.4.2 Mouse models for trisomy 21

It has become almost routine to use transgenic or knock-out mice (*Mus musculus*) for the study of human genetic disorders, since the relatively close phylogenetic relationship between mice and men (both *Eutheria/Euarchontoglires*) often allows to recapitulate specific aspects of a given disease.

The underlying assumption or using DS mouse models is that basic genetic processes disrupted by gene-dosage imbalance will be largely conserved, while the observed phenotypes will vary in some species-specific ways. Consequently, the phenotypic parallels between a viable mouse model and DS will result in a catalog of

traits that can be compared across species. While most genes show synteny in content and order between man and mouse, exceptions have been reported (Gardiner *et al.* 2003).

For the study of DS, there have been some difficulties in using this approach, since in mouse, there are three chromosomal regions that are orthologous to Hsa21: from the human centromere to the q-telomere, about 23.2 Mb are homologous to Mmu16, about 1.1 Mb to Mmu17, and 2.3 Mb to Mmu10. Therefore, it has been a difficult task to construct mouse models that reflect the chromosome aberration resulting in DS.

The first mouse model introduced for DS was a mouse trisomic for Mmu16, called **Ts16** (Epstein *et al.* 1985). Although Ts16 mice show several features of DS, this early mouse model had three major drawbacks: (1) the trisomic fetuses do not survive as live-born animals, (2) the Ts16 mice are trisomic for many other genes from Hsa3, 8, 16 and 22, and (3) the Ts16 are not trisomic for all orthologous genes of Hsa21. Other groups tried to compensate for these drawbacks by constructing mouse models with different kinds of segmental trisomy 16, which proved valuable for gaining insights into the molecular mechanisms of trisomy 21.

The most studied model for DS is the **Ts65Dn** mouse, which was obtained using translocation chromosomes and contains a segmental trisomy for orthologs of 128 Hsa21 genes, from APP to ZNF295, which is almost half of the genes annotated on Hsa21. While the mice do not appear to have all of the features characteristic for DS, they represent a mouse model that survives to adulthood and is thus useful to study also DS features that develop later in life (Davisson *et al.* 1993). Among the parallels between DS individuals and Ts65Dn mice are craniofacial abnormalities (Richtsmeier *et al.* 2000), reduced cerebellar volume and decreased density in granule and Purkinje cells (Baxter *et al.* 2000), cardiac anomalies (Moore 2006), and developmental delays as well as deficits in spatial learning and memory (Reeves *et al.* 1995; Demas *et al.* 1996).

Another well-studied model is the **Ts1Cje** mouse, which carries a triplicated Mmu16 with 95 genes orthologous to Hsa21 (Sago *et al.* 1998), from SOD1 to ZNF295. These mice are very similar to Ts65Dn, but do not carry the triplication from APP to SOD1. Although Ts1Cje is less comparable to human trisomy 21 than Ts65Dn, it shows many phenotypic features of DS, among them craniofacial and cerebellar phenotypes (Richtsmeier *et al.* 2002; Olson *et al.* 2004) as well as behavioural and learning abnormalities (Sago *et al.* 1998).

The most recent addition to the family of DS mouse models is a mouse carrying a freely segregating, almost complete copy of human chromosome 21 (O'Doherty *et al.* 2005), which was called **Tc1**. These mice have phenotypic alterations in behavior, synaptic plasticity, cerebellar neuronal number, heart development, and mandible size that relate to human DS. Many cells in adult Tc1 mice show mosaicism, increasing the complexity of analyses using this model. Nevertheless, Tc1 mice may provide a powerful tool for investigation of the pathogenesis of trisomy 21, as well as a platform for analysis of similarities and differences in the evolution of gene regulation (Reeves 2006).

Although none of the described mouse models perfectly mimics the chromosome aberration resulting in DS, they have proven to be a valuable tool in DS research. As described in the next chapter, some interesting results have been generated through gene expression studies performed in these mouse models.

1.4.3 Dysregulation of transcript levels

The primary change resulting from increased gene dosage in trisomy 21 is at the level of the transcriptome. Most comparative gene expression studies for Hsa21 have been performed with mouse models, since availability of human tissues is usually limited. Moreover, the expression patterns of Hsa21 gene orthologs have been well studied in wild-type mice. Gene expression patterns of Hsa21 mouse orthologs were systematically analyzed by mRNA *in situ* hybridization on whole mount embryos at critical stages of embryonic and brain development. These studies resulted in valuable data sets for targeting candidates that participate in the pathogenesis of DS (Gitton *et al.* 2002; Reymond *et al.* 2002), especially the large fraction of Hsa21 mouse orthologs that are found expressed in the brain.

Microarray analysis of transcriptome levels in the cerebellum of Ts65Dn mice showed that a dosage imbalance of 124 genes alters the expression of thousands of genes (Saran *et al.* 2003). Most Hsa21 orthologous genes were found overexpressed by approximately 1.5-fold. Studies of the Hsa21 orthologs using macroarrays and RT-PCR confirmed this gene-dosage effect of the triplication, with exceptions (Kahlem *et al.* 2004; Lyle *et al.* 2004). Analysis of the Ts1Cje cerebellar transcriptome demonstrated a similar global overexpression of triplicated genes (Dauphinot *et al.* 2005). In all of these studies, several genes escaped the ‘1.5-fold

overexpression' rule, suggesting that they may be controlled by additional tissue-specific regulatory mechanisms.

More recently, novel insights have been gained via expression studies examining material from individual mice rather than pooled material from different individuals. These studies showed that inter-individual gene expression levels span a broad range of values, with three categories of genes: (1) genes with expression levels consistently higher in trisomic mice than in euploids; (2) genes whose expression levels partially overlap between the two groups; and (3) genes with intermingled expression, which cannot be used to differentiate trisomics (Sultan *et al.* 2007). A similar study used trisomic and normal human lymphoblastoid and fibroblast cell lines and showed that, according to the degree of overlap in expression levels, genes can be classified into three groups: (A) nonoverlapping, (B) partially overlapping, and (C) extensively overlapping expression distributions between normal and DS samples (Prandini *et al.* 2007). Moreover, it was found in trisomic amniocytes that there are genes which shows greater variance of expression in DS tissues than in euploid tissues (Chou *et al.* 2008). In summary, the study of inter-individual variations in gene expression will further help to identify sets of candidate genes which may contribute to the variable phenotypic abnormalities observed in DS.

How does the data from mouse models fit to that derived from transcriptome analysis in humans? Gene expression analysis in human T lymphocytes from patients with DS showed a global upregulation of Hsa21 gene expression, reinforcing the gene dosage hypothesis (Giannone *et al.* 2004). Transcriptome analysis of primary cultures of human fetal amniocytes derived from trisomy 21 cases showed that the average level of transcription on the trisomic chromosome was increased only approximately 1.1-fold compared to normal cells, while >95% of misregulated genes did not map to the trisomic chromosome (FitzPatrick *et al.* 2002). And Hsa21 gene expression was found globally up-regulated in human fetal trisomy 21 cases, both in cerebral cortex extracts and in astrocytic cell lines cultured from cerebral cortex (Mao *et al.* 2003).

It is still debated if the relatively subtle changes in gene expression levels are the main cause for the deregulation of brain development and function observed in DS. It has become clear that transcriptome analyses must be carried out on series of DS individuals and controls to take into account relatively large inter-individual differences.

1.4.4 Protein dosage imbalances

Transcriptome analysis can only give part of the answers regarding molecular mechanisms underlying DS pathogenesis. Proteins are the actual workhorses running most biological processes, and their concentration and activity is not exclusively dependent on the transcript levels of the corresponding genes. Therefore, deciphering the molecular basis of DS necessites to include the analysis of the levels (and activities) of all Hsa21 proteins whose gene dosage is altered in DS. To date, such an analysis has not been carried out due to lack of experimental techniques for sensitive and specific quantification of hundreds of proteins in parallel. The analysis of protein activities is even less advanced in respect to throughput, since biochemical experiments differ widely among different types of enzymes and can usually not be carried out in a systematic fashion.

Some groups have tried to analyze proteome-wide changes in DS protein expression using 2D gel electrophoresis and subsequent mass spectrometric identification of proteins with altered expression. For example, one study measured protein levels in fetal cerebral cortex (Bajo *et al.* 2002). The authors inferred a derangement of enzymes involved in intermediary metabolism during prenatal development of DS individuals, because five enzymes from this large cellular network were over- or underexpressed. Another proteomic study reported the results of an *in vitro* neuronal differentiation system of mouse ES cells containing a single human chromosome 21 with parental ES cells as control (Kadota *et al.* 2004). Here, a total of 18 significantly altered proteins were found in cells with an extra copy of Hsa21, including SOD1 and CCT8 encoded on Hsa21. The other 16 proteins fall in various functional categories and provide a set that differs from that identified in adult or fetal brain with DS.

Concerning the expression levels of Hsa21 proteins themselves, a multitude of studies have been carried out measuring the level and activity of single Hsa21 proteins in trisomic versus normal human and mouse samples (Engidawork and Lubec 2001). Care must be taken when comparing these results with each other and to those of other studies, but some lessons can be learned: (1) protein expression levels change with developmental stage, (2) protein levels appear to be generally higher with increased gene dosage, but (3) a significant number of proteins escape these rules.

In summary, the data available for Hsa21 proteins in DS suggest that the DS phenotype cannot alone be explained alone by general overexpression of all Hsa21 genes and subsequently, proteins. It appears that the trisomic situation can cause pertubation of expression in different directions, thus affecting proteins involved in brain development and function. The observed age-related changes in protein expression may be of importance because (i) those proteins that are deregulated in fetal DS brain may be proteins that are important for normal development and could be responsible for neurodevelopmental disorders in DS, while (ii) proteins unchanged in fetal DS brain but altered in adults may be responsible for age-related and neurodegenerative changes in DS.

1.4.5 Disturbed signal transduction in Down syndrome

The molecular hypothesis guiding DS research is that (1) the gene dosage effect caused by trisomy 21 leads to increased expression of Hsa21 genes not only at the mRNA level, but also at the protein level, that (2) at least some of the increases result in pertubation of the cellular processes and pathways these proteins function in, and (3) the additive effect of many small perturbations will result in the neurodevelopmental and cognitive abnormalities that characterize the mental retardation in DS (Gardiner *et al.* 2004). Since there are a large number of candidate genes, multiple pathways are likely to be involved. But even our knowledge on the best-studied genes is incomplete and surely misses the variety of protein isoforms, expression patterns and alternative protein functions and interactions. Nevertheless, it is of major importance to build hypotheses relating Hsa21 genes via pathways to phenotypic consequences, because single-gene analyses will not be able to explain many (or most) anomalies in DS, while multiple pathway analysis is more likely to provide an integrated model of the pathophysiology of DS.

To date, there have been only few reports trying to integrate Hsa21 proteins into cellular pathways. Some proposed effects of Hsa21 protein overexpression on signaling pathways are discussed in the following paragraphs. Obviously, any effect of overexpression will be of a far more complex nature than described here; even so, some lessons can be learned from examination of the signal transduction pathways connected to Hsa21 proteins.

A signal transduction pathway that has been associated with trisomy 21 for a long time is interferon (IFN) signaling:

Interferon signaling

The IFN-alpha/beta/omega receptor chains **IFNAR1** and **IFNAR2** as well as the IFN-gamma receptor beta chain **IFNGR2** are located on Hsa21. As early as in 1974, it was shown that cultured human trisomy 21 fibroblasts demonstrated an increased sensitivity to human fibroblast interferon (Tan *et al.* 1974). Later on, it was reported that in trisomic lymphocytes from patients with DS, the 50% increase in gene dosage for the IFN-alpha and -beta receptor molecules results in higher basal activity of 2',5'-oligoadenylate synthetase. This enzyme polymerizes ATP into pppA oligomers, which activate a latent Rnase F that degrades mRNA, leading to inhibition of protein synthesis and virus growth (Gerdes *et al.* 1993). Then it was shown that anti-interferon immunoglobulins can improve the phenotype of mice with trisomy 16 (Ts16), one model for human trisomy 21. Maternal anti-interferon treatment was found to provide significant return-toward-normal values observed for overall fetal growth, eye opening, and back curvature (Maroun 1995). Also, reduction of the gene dosage of IFN-alpha/beta and IFN-gamma receptors in Ts16 mouse fetuses showed improved growth and cultured neuron viability (Maroun *et al.* 2000). Aberrant expression of interferon-related protein RACK1 in fetal Down syndrome brain has been shown by 2D-gel and MALDI analysis (Peyrl *et al.* 2002). Moreover, IFN-gamma plays a crucial role in the induction of Abeta 40 and Abeta 42 production from the beta-amyloid precursor protein **APP**, another protein encoded on Hsa21 (Blasko *et al.* 1999). A more recently reported cross-talk to the Notch pathway in DS by activation of Notch and Hes1 through APP signaling might affect brain development, since the Notch pathway plays a pivotal role in neuron-glia differentiation (Fischer *et al.* 2005). In the DS thymus, overexpression of IFN-gamma and TNF-alpha has been demonstrated to be linked to abnormal thymic anatomy (Murphy *et al.* 1995). Taken together, these findings support a connection between interferon signaling and several aspects of the DS phenotype.

Three other signal transduction pathways have been proposed to be affected by deregulated Hsa21 proteins (Gardiner *et al.* 2004; Gardiner and Costa 2006):

Estrogen and glucocorticoid signaling

The Hsa21 nuclear receptor inhibitor protein **NRIP1** inhibits estradiol-induced transcription (Teyssier *et al.* 2003) and was found to be overexpressed in hippocampi from aged DS individuals (Gardiner 2006). Estrogen maintains normal function of

basal forebrain cholinergic neurons. Thus, effects of normal estrogen levels might be diminished in DS persons, potentially leading to the early menopause and cognitive decline observed in DS. Ts65Dn mice develop cholinergic degeneration similar to young adults with DS, while estrogen replacement therapy restored cell size and total number of cholinergic neurons in the medial septum of Ts65Dn mice (Granholm *et al.* 2002). Also, NRIP1 was shown to antagonize glucocorticoid receptor (GR)-mediated responses by direct interaction with the receptor. Interestingly, mouse models with decreased GR levels display impaired learning ability (Oitzl *et al.* 2001), an effect that might be further strengthened by another Hsa21 protein, **SUMO3**, which also inhibits transcription by the GR (Holmstrom *et al.* 2003).

MAP kinase and calcineurin signaling

Mitogen-activated protein (MAP) kinases regulate the function of many proteins via phosphorylation, and calcineurin dephosphorylates a partially overlapping set of proteins. Among the targets of these signaling enzymes are a number of proteins whose phosphorylation state can affect learning of hippocampal-based tasks, such as Ras, ERK1/2, CREB, GR and ELK (Sweatt 2004). Hippocampal impairments have been shown in both children with DS and the Ts65Dn mouse model of DS (Pennington *et al.* 2003). It is noteworthy that eleven Hsa21 proteins (marked in bold letters) interact with components of the MAPK and calcineurin pathways:

ITSN1, a molecular scaffold involved in regulating endocytosis and mitogenic signaling, activates Ras and stimulates transcription through an independent c-Jun N-terminal kinase (JNK) pathway (Mohney *et al.* 2003). **TIAM1** is a guanine-nucleotide exchange factor for Rac, leading to Ras and JNK activation (Michiels *et al.* 1997). **DSCR1** is a direct inhibitor of calcineurin and of critical importance to the calcineurin pathway (Rothermel *et al.* 2000), but restoration of DSCR1 to disomic levels in the Ts16 mouse model of DS did not correct cardiac or craniofacial development anomalies in embryos (stages E11.5 and E13.5)(Lange *et al.* 2005). The superoxide dismutase **SOD1** and calmodulin-binding peptide **PCP4** may also inhibit calcineurin activity via elevating the level of hydrogen peroxide – which induces DSCR1 expression – and regulate calmodulin activity, respectively (Johanson *et al.* 2000; Lin *et al.* 2003). The dual-specificity tyrosine phosphorylation-regulated kinase **DYRK1A** (see also below) directly phosphorylates CREB, leading to the stimulation of CRE-mediated gene transcription during neuronal differentiation, while blockade of DYRK1A activation significantly inhibits neurite outgrowth (Yang *et*

al. 2001a). DYRK1A is also involved in regulating the recruitment activity but not the phosphatase activity of inositol 5'-phosphatase **SYNJ1** (Adayev *et al.* 2006). Another important point of connection between MAPK and calcineurin pathways is the beta-amyloid precursor protein **APP**, whose transcription can be activated by a Ras/MAP kinase signaling pathway (Ruiz-Leon and Pascual 2001), and APP overexpression activates ERK1/2. When APP is overexpressed, more beta-amyloid peptide is formed. This peptide can directly induce increased expression of DSCR1, which may lead to mRNA levels of DSCR1 three times higher in patients with extensive neurofibrillary tangles than in controls (Ermak *et al.* 2001). Finally, there are at least three transcription factors encoded on Hsa21 that are targets of the MAPK pathway: the **ETS2** oncogene (Yordy and Muise-Helmericks 2000), the GA repeat binding protein **GABPA** (Fromm and Burden 2001), and the runt-related transcription factor **RUNX1** (Tanaka *et al.* 1996). These transcriptional regulators may add a multitude of downstream effects to pertubations of the MAPK pathway.

RNA processing pathways

It has been suggested that subtle defects in RNA processing might have consequences on the neurological and other abnormalities observed in DS. Eight proteins from Hsa21 (marked in bold letters) are thought to participate in the processing of pre-mRNA. The arginine/serine-rich splicing factor **SFRS15** is the human ortholog of rat rA4 protein that interacts with the C-terminal domain of RNA polymerase II. Since it also contains a RNA recognition motif commonly seen in splicing factors, it has been proposed that SFRS15 may physically and functionally link transcription and pre-mRNA processing (Yuryev *et al.* 1996). The small subunit of the essential U2 auxiliary factor, **U2AF1**, is a non-snRNP protein required for selection of splice sites with weak polypyrimidine tracts during alternative splicing (Pacheco *et al.* 2006), whose deregulation could influence a specific subset of cellular transcripts. The RNA-specific adenosine deaminase **ADARB1** is an enzyme responsible for essential pre-mRNA editing of glutamate and serotonine receptor subunits as well as a potassium channel in human brain by site-specific deamination of adenosines (Hoopengardner *et al.* 2003; Kawahara *et al.* 2003). ADARB1 therefore has the potential to severely alter neurophysiological functions of the brain. Accordingly, ADARB1 has been linked to depression associated with interferon-alpha therapy in patients treated for chronic viral hepatitis and certain malignancies (Yang *et al.* 2004). The dual-specificity tyrosine-(Y)-phosphorylation regulated kinase

DYRK1A is found in nuclear speckles belonging to the splicing compartiment, where it plays a role in the assembly of these speckles (Alvarez *et al.* 2003). This gene is a homolog of the Drosophila *minibrain* gene and is expressed in the developing brain where it seems to play a role in proliferation of neural progenitor cells, neurogenesis, and neuronal differentiation (Hammerle *et al.* 2003).

Another important Hsa21 protein found to localize to the splicing compartment is the phosphorylated intracellular Cgamma fragment of the beta-amyloid precursor protein **APP** (Muresan and Muresan 2004), which co-localizes in nuclear speckles with the APP binding scaffolding protein APBB1 (Nakaya and Suzuki 2006). The latter – also called Fe65 – has recently been shown to be involved in alternative splicing of APP exons 7 and 8 (Donev *et al.* 2007). The poly(rC) binding protein **PCBP3** is a member of a family of RNA-binding proteins with functions in post-transcriptional control (Makeyev and Liebhaber 2000). The organization of the three hnRNP K homology (KH) domains of PCBP3 – two nearby the N-terminus, one at the C terminus – is similar to that of neuro-oncological ventral antigen NOVA1, a protein that regulates alternative splicing of glycine and GABA receptor subunits and a set of other components of the inhibitory synapses (Ule *et al.* 2003). The last two Hsa21 proteins that were found to be involved in splicing are the anonymous ORFs **C21orf66** and **C21orf70**, which were both identified as spliceosome components using a mass spectrometric approach (Rappsilber *et al.* 2002). C21orf66 was also identified in another similar analysis (Zhou *et al.* 2002).

There are a number of neurological abnormalities associated with alterations of the normal splicing patterns (Faustino and Cooper 2003). The question remains if an altered expression of some or many of these Hsa21 proteins can result in pertubation of either constitutive or alternative splicing patterns, leading to some of the phenotypic alterations observed in DS.

From pathophysiology to genomics analyses

To elucidate the contribution of each triplicated gene to the phenotypic and molecular alterations described above, it is important to know the functional classification of the corresponding proteins. The work described here was dedicated to unravelling new functional information for the proteins encoded on chromosome 21 through functional genomics analyses, as described in the next chapter.

1.5 Functional Genomics Analyses of Hsa21 Proteins

1.5.1 Cloning of open reading frames

Functional genomics and proteomics aim at examining the roles of large sets of genes and proteins of an organism in a high-throughput format. Most applications depend on the availability of cloned copies of the transcripts in a format allowing protein expression. The first comprehensive open reading frame (ORF) clone collections were constructed using gap-repair cloning via homologous recombination in the yeast *S. cerevisiae* (Uetz *et al.* 2000; Ito *et al.* 2001). More recently, site-specific recombination-based cloning has emerged as an alternative and as a more general method for constructing large ORF clone collections (Walhout *et al.* 2000). Although recombination-based cloning is of higher cost due to the recombinase enzymes used, ultimately these systems are more practical and cost-effective for transferring ORFs to many different expression vectors.

Recombinatorial cloning strategies

Recombinational cloning systems use approaches in which master clones (“entry clones”, propagated as plasmids in *E. coli*) can be used to create corresponding expression clones. The site-specific recombination reactions utilize "cut and paste" strategies where the transferred DNA is conserved in its integrity, rendering sequence validation of expression constructs unnecessary (Marsischky and LaBaer 2004). Commercial recombination-based systems are supplied by Invitrogen (“Gateway” cloning system) and by Clontech (“Creator”/”In-Fusion” cloning systems).

Creating ORF clones requires PCR amplification of the coding sequence with ORF-specific primers. Four types of starting material can be used as template for amplification: (1) genomic DNA, (2) first-strand cDNA, (3) cDNA libraries, or (4) available sequence-validated full-length cDNA clones. The use of genomic DNA is limited to organisms with no or only little splicing activity (prokaryotes and simple eukaryotes) and intron-less genes. The remaining sources are all biased towards the most abundant transcripts in the cells used for their generation.

Generation of entry clones

The most failure-prone step in building large ORF clone collections is the process of creating master clones. In a pilot study directly comparing Gateway and In-Fusion

cloning, ORFs smaller than 2 kb were cloned with ~90% efficiency in both systems (Marsischky and LaBaer 2004). For ORFs larger than 2 kb, the In-Fusion system was somewhat more efficient than Gateway BP cloning (76% vs. 50% for 2-3 kb, and 63% vs. 29% for 3-4 kb). Both cloning systems were found to be functionally equivalent or at least competitive for many of the key parameters important in cloning large ORF collections.

The primary step for creating Gateway entry clones is a reaction in which the ORF with flanking recombination (attB) sites (usually attached via PCR) is recombined into a vector with the corresponding attP sites (Figure 1-7A). Recombination is accomplished by an *in vitro* recombination reaction that requires the λ Int and IHF proteins, a mixture termed "BP Clonase" by Invitrogen. Gateway BP recombination is both efficient and relatively insensitive to target DNA concentration.

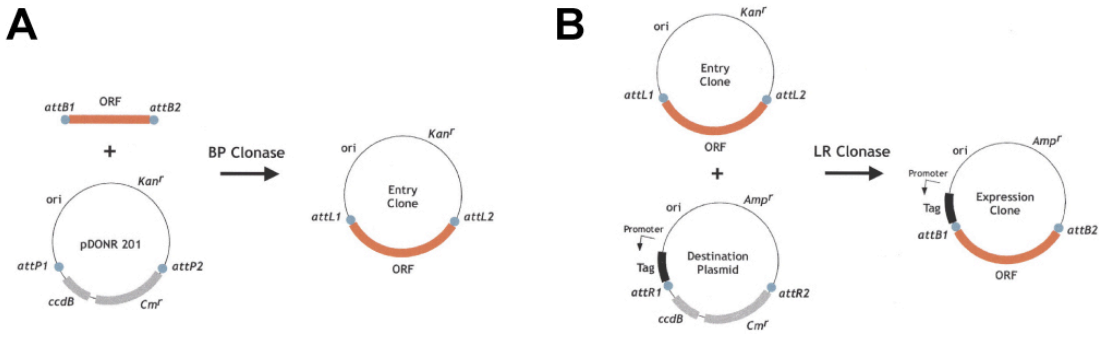

Figure 1-7. Schematic overview of the Gateway site-specific recombination cloning system. A: Cloning of ORF attB-PCR products by Gateway BP clonase-mediated recombination. The blue circles represent λ att recombination sites. B: Transfer of ORFs from the entry vector to create an expression clone by Gateway LR clonase-mediated recombination. (Pictures modified from Marsischky and LaBaer 2004).

Generation of expression clones

Once a library of entry clones is established, the ORFs can be transferred to any expression vector that has been modified for use with the Gateway system. The transfer of ORFs into Gateway expression vectors is accomplished by "LR recombination" in an *in vitro* reaction that requires the phage λ Int and Xis proteins together with the IHF protein (LR Clonase mix), an expression plasmid with attR sites and an entry ORF clone in which the ORF to be transferred is flanked by attL sites (Figure 1-7B).

Resources of human entry clones

A large academic initiative has reported the generation of a first version of the human ORFeome using a Gateway recombinational cloning approach (Rual *et al.* 2004). In the human ORFeome v1.1 published in 2004, there were ORFs cloned for 7,263 different human genes, among them 72 ORFs from Hsa21. The current version (ORFeome 3.1) contains 99 ORFs from Hsa21, which is still only 35% of all genes on this chromosome. Commercial vendors, such as Invitrogen and GeneCopoeia, offer larger clone sets (ca. 16,000 Invitrogen human Ultimate ORF Clones and ca. 20,000 GeneCopoeia human OmicsLink ORF expression clones at the moment), but most academic research projects will not have the budget to cover the enormous costs associated to these products, and many genes are still missing in these collections, which are redundant for different splice variants.

Applications of ORF clones

Once ORFs are cloned into entry vectors, they can be subsequently used for a variety of functional genomics analyses, depending on the experimental assays and suitable expression vectors available (Brasch *et al.* 2004). In the following chapters, two methods will be discussed that allow for both initial characterization of unknown proteins as well as expanding the knowledge for known proteins. Protein localization studies can provide initial information about protein function, since compartimentalization is typically related to the biological role of proteins. And the study of protein-protein interactions can give crucial hints towards the interaction networks in which a protein is involved, often allowing for a ‘guilt by association’-type of functional annotation.

1.5.2 Subcellular protein localization

Despite the rapid expansion of sequence databases and the development of *in silico* analysis methods, it has become clear that computational analyses show limitations for assigning functions to many unknown genes and the encoded proteins. Since a DNA sequence in itself is not always functionally informative, experimental assays have to be developed that can provide additional hints towards protein functions. A wealth of knowledge from cell biological studies has revealed that the broad variety of eukaryotic cell organelles and structures are the places of specific cell functions (O'Rourke *et al.* 2005). For example, transcription factors can only exert their

regulatory role when localized to particular compartments of the nucleus, and transmembrane receptors have to be exposed to the extracellular environment to bind their soluble ligands. The categorization of proteins by their subcellular localization therefore is one of the essential ways for functional annotation of the human genome.

Conventional approaches for subcellular localization studies

Traditionally, the localization of proteins has often been studied by (1) immunofluorescent staining of the endogenous protein, (2) construction of fusion proteins with autofluorescent proteins such as GFP and its derivatives or (3) construction of fusion proteins with epitope tags. These techniques have been used to confirm protein interactions obtained from yeast two-hybrid screens (Wong and Naumovski 1997). Usually, *in vivo* analyses are performed on single genes where either antibodies are available or recombinant DNA constructs encoding fusion proteins are transfected into cells, leading to overproduction of gene products. These cell-based assays are typically carried out in Petri dishes or multi-well plates, limiting the throughput due to high reagent consumption and requirement for sophisticated automation equipment.

Computational approaches for protein localization prediction

A multitude of software has been developed over the last decade to predict the subcellular localization of proteins. These programs apply various biological concepts and computational methods to the localization problem. The four leading methods underlying these solutions are:

(1) Prediction based on overall amino acid composition. The program SubLoc predicts protein localizations based on the finding that oftentimes, proteins with different localizations show distinctive amino acid compositions (Hua and Sun 2001);

(2) Prediction based on known targeting sequences. The presence of targeting signals, such as signal peptides, mitochondrial targeting peptides, nuclear localization signals and transmembrane helices leads proteins to different cellular organelles. These signals are recognized by programs such as SignalP, TargetP and TMHMM (Emanuelsson *et al.* 2007);

(3) Prediction based on sequence homology and/or motifs. Programs such as Proteome Analyst use database annotations for homologs in combination with machine learning to predict subcellular localizations (Lu *et al.* 2004);

(4) Prediction using a combination of different information obtained through the methods described in (1)-(3). Recently, the available localization prediction methods CELLO, MultiLoc, Proteome Analyst, pTarget and WoLF PSORT were compared using large evaluation datasets. It turned out that no individual method had a sufficient level of sensitivity across the evaluation sets that would enable reliable application to hypothetical proteins (Sprenger *et al.* 2006). Nuclear and extracellular proteins were predicted with a high sensitivity, whereas proteins localized to the secretory pathway were the most difficult to predict. The enormous complexity of cellular protein sorting still limits the application of computational in subcellular localization prediction.

High-throughput localization studies

In order to acquire more comprehensive experimental data sets on protein localizations, efforts on a larger scale have been carried out using cells from various organisms. The first large-scale study reported the localizations of 250 proteins in fission yeast by the use of a GFP-fusion genomic DNA library (Ding *et al.* 2000). This study was followed by larger projects with 2,744 proteins (Kumar *et al.* 2002) and 4,156 proteins (Huh *et al.* 2003), representing 75% of the yeast proteome. The first large-scale study with mammalian cells was reported by the German cDNA consortium, which used GFP fusions of 107 human genes in a microplate format (Simpson *et al.* 2000). Currently, there are now over 900 entries in their localization dataset, which is publicly available at http://www.dkfz.de/LIFEdb. For chromosome 21, there are only two entries for the uncharacterized proteins C21orf7 and C21orf56.

Use of transfected cell arrays for the study of protein localizations

Microarray technology has become one of the most widely applied methodologies in functional genomics. Initially developed as a technique for large-scale DNA mapping (Poustka *et al.* 1986), it has been adapted to and is now mostly used for determination of transcript levels of thousands of genes in parallel (Schena *et al.* 1995). The concept of miniaturization by high-density arraying has expanded into many areas, ranging from SNP genotyping over measurement of DNA copy number changes to protein arrays, antibody arrays, tissue arrays and others. When combined

with computational analysis, the large data sets derived from these techniques provide the opportunity to get insights into complex functional networks and regulatory mechanisms.

Motivated by the success of microarrays, this technology was developed also to miniaturize cell-based functional assays. The so-called *transfected cell array* (TCA) emerged as an alternative for high-throughput analysis of gene functions in mammalian cells (Ziauddin and Sabatini 2001). In principle, nucleic acid molecules are spotted together with a transfection agent onto a solid surface in a microarray format (Figure 1-8A). After spotting, a lawn of cells is grown on top of the surface. The cells located on the spots are transfected in a localized way, while the cells between the spots stay non-transfected (Figure 1-8B). Using subsequent cell-based assays, the phenotypic consequences of transfection can be monitored for up to thousands of different molecules in parallel. Different applications included loss-of-function type of TCA experiments, where siRNAs are used to silence genes (Vanhecke and Janitz 2004), and gain-of-function experiments using PCR products or expression plasmids for overexpression of proteins (Liebel *et al.* 2003; Hu *et al.* 2005).

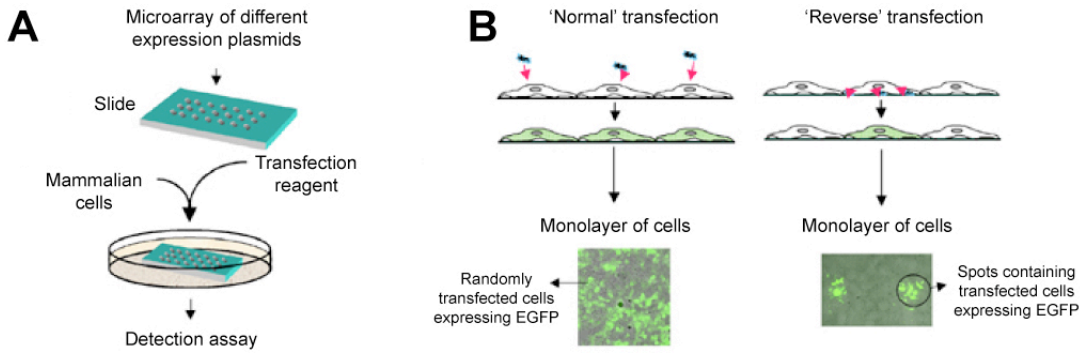

Figure 1-8. The transfected cell array (TCA) technique combines microarray technology with living cells. A: Nucleic acids are spotted onto DNA microarrays and transfected into a monolayer of cells grown on top of the slide. B: Reverse transfection allows for multiple, spatially separated transfections of nucleic acids in a single experiment. (Pictures modified from Vanhecke and Janitz 2004).

The transfected cell array technique was available and used during the work presented here for studies on the subcellular localizations of Hsa21 proteins, as described in the Results.

1.5.3 Protein-protein interactions

The analysis of protein-protein interactions (PPIs) has proven indispensable for the understanding of the biological role of proteins. Intermolecular associations are essential for a wide range of cellular processes, forming a network of extraordinary complexity.

Experimental methods for PPI detection

Experimental techniques used for detection of PPIs can be categorized into different methodological classes, consisting of screening approaches and verification approaches, carried out either *in vivo* or *in vitro*. Important factors that govern the choice of a specific technique include the availability, purity and concentration of the proteins to be analyzed, the requirement for native state, post-translational modifications and co-factors as well as pH and ionic strength of the surrounding environment.

Table 1-2 lists twelve common methods for PPI detection, classified according to the two criteria mentioned above. The upper part of the table recalls available screening approaches, while the lower part is dedicated to methods used mainly for subsequent verification and further analysis of PPIs. Some main advantages and drawbacks of each method are listed, among with the original works describing the technique, where available.

Until some years ago, knowledge of cellular PPI networks was fairly limited. This has been changed mainly by adapting two PPI detection techniques to enable large-scale *de novo* interaction screening. Since then, the field of PPI discovery has been dominated by yeast two-hybrid (Y2H) screening using cloned libraries of bait and prey proteins (Figure 1-9A) and by tandem affinity purification (TAP) of protein complexes, followed by mass spectrometric identification of the components of the complexes (Figure 1-9B). Verification of a subset of newly identified PPIs is then often carried out using co-immunoprecipitation (Figure 1-9C) or pull-down assays, and/or immunofluorescence co-localization of the potential interactors in transfected or non-transfected cells (Figure 1-9D).

Table 1-2. Experimental methods commonly used for identification, verification and analysis of protein-protein interactions.

Set-up	Method	Advantages	Drawbacks	References
Screening approaches				
in vivo	Yeast two-hybrid (Y2H) and other split systems	highly sensitive; performed in intact cells; has been fully automated	heterologous system; nuclear import required (size limitation); not always specific (false positives)	(Fields and Song 1989)
	Protein crosslinking	can detect weak interactions; can be done *in vivo* and cell stage specific	not always specific (false positives from neighboring proteins)	(Peretz and Elson 1976)
in vivo / *in vitro*	Protein affinity chromatography	very sensitive; results are specific due to competition with all cellular proteins	pure protein and relatively large lysate volumes required	(Formosa *et al.* 1991)
	Tandem affinity purification (TAP)	detection of existing endogenous complexes; not limited to binary interactions	false negatives due to low abundance and transient interactions	(Rigaut *et al.* 1999)
in vitro	Phage display	large libraries can be screened; can be used to increase binding affinities	bacterial host system; secretion required; size limitation	(Smith 1985)
	Protein array	large protein sets can be screened for interaction with proteins, lipids and compounds	pure proteins required; proteins can denature on array surface	(Lueking *et al.* 1999)
Verification approaches				
in vivo	Subcellular co-localization	both partners in natural state; can be done cell stage specific	not very sensitive; not necessarily direct interaction	(Brelje *et al.* 1993)
	Fluorescence resonance energy transfer (FRET)	can be done *in vivo* and cell stage specific; can detect direct interactions	not very sensitive; protein labeling required	(Herman 1989)
in vivo / *in vitro*	Co-immuno-precipitation (coIP)	both partners in natural state; specific due to competition with other cellular proteins	not very sensitive; not necessarily direct interaction	(Phizicky and Fields 1995)
in vitro	Pull-down assay	straightforward; can show direct physical interaction if performed with pure partners	pure protein required; one partner present in large excess	(Phizicky and Fields 1995)
	Surface plasmon resonance (SPR, Biacore)	fast method, requires little material; real-time measurement of association/dissociation curves possible	pure protein required; proteins can denature on sensor surface	(Jonsson *et al.* 1991)
	Far Western Blot/ Blot Overlay	simple and sensitive analysis of complex mixtures	pure protein required; partners can denature on blot membrane	(Kremer *et al.* 1988)

This list is not intended to be comprehensive; rather, the most commonly applied experimental techniques are compared regarding the assay purpose (screening for new PPIs or verification/analysis of candidate PPIs), the assay environment (*in vivo*, *in vitro* or a combination of both) and some key advantages and drawbacks. References in the table cite original publications of the techniques or, when not applicable, comprehensive reviews on the subject. This table is mostly based on Phizicky and Fields 1995.

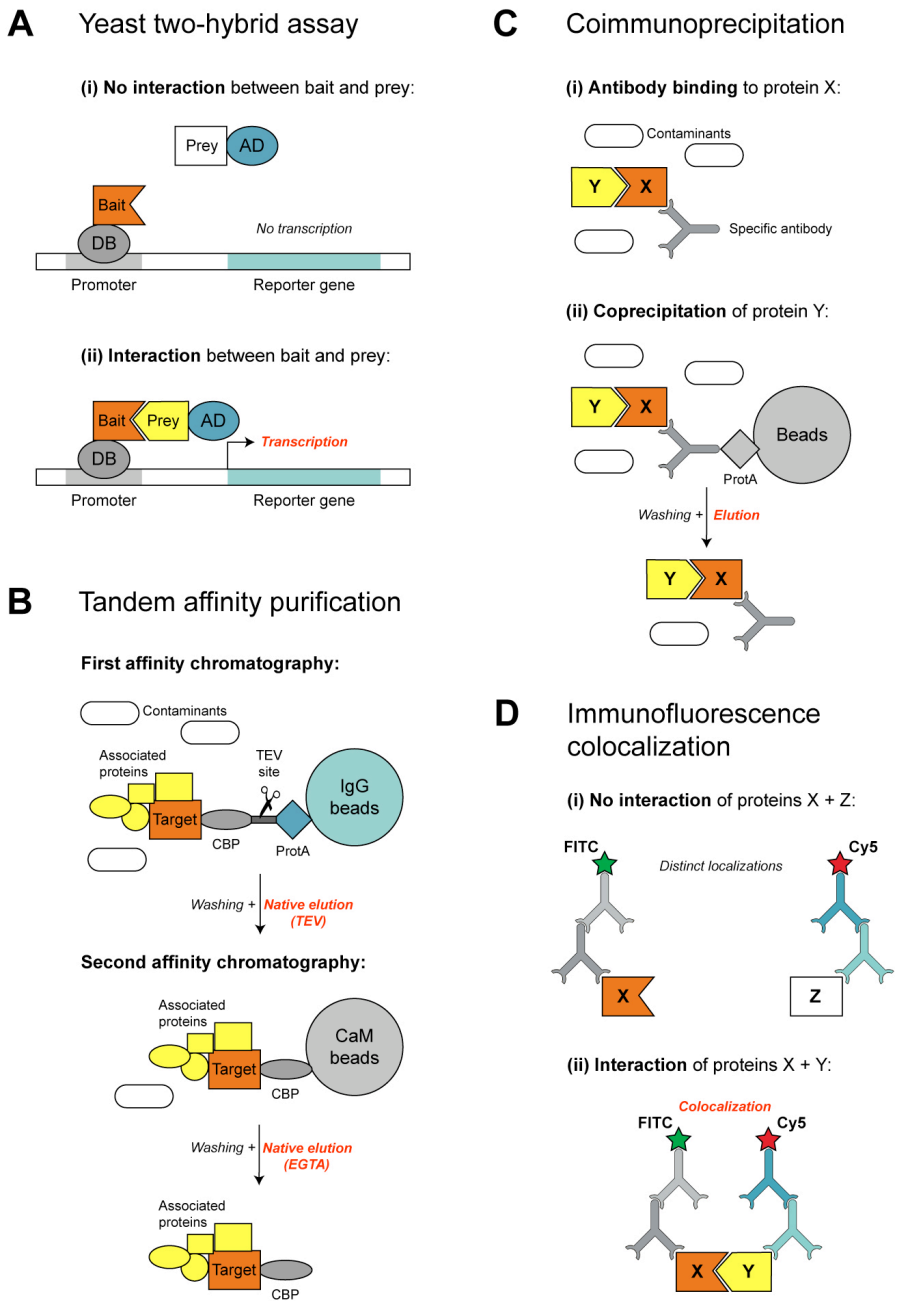

Figure 1-9. Schematic illustration of the principles underlying commonly used methods for detection of protein interactions. A: Yeast two-hybrid (Y2H) analysis works by separating the coding sequences for the DNA-binding (DB) and activation domain (AD) of a transcriptional activator and fusing them separately with a candidate protein (DB-bait) and potential interaction partners (AD-preys). Baits and preys can be transformed into yeast strains of opposite sex (a and α) and combined in a single cell by mating. If bait and prey are not interacting (i), reporter genes stay inactive. If bait and prey physically interact (ii), the reconstituted transcriptional activator mediates expression of a reporter gene, which can be monitored by auxotrophic growth or a color reaction. B: Tandem affinity purification (TAP) works by fusing the target protein with a TAP cassette consisting of a low-affinity tag/enzymatic cleavage site/high-affinity tag for stringent purification of protein complexes from cell lysates. After high-affinity binding of protein A (ProtA) to IgG beads and washing, the target complex is proteolytically released with TEV protease and subjected to a second purification step using the low-affinity binding of calmodulin-binding peptide (CBP) to calmodulin beads, which can be released using EGTA under mild conditions. The combination of two rounds of affinity chromatography usually results in relatively pure protein complexes, whose components can be identified using mass spectrometry after separation by gel electrophoresis. C: Coimmunoprecipitation is performed by adding an antibody specific for a protein X (or an epitope tag fused to it) to a cell lysate (i) and precipitating the protein-antibody complex using

affinity beads (ii). After thorough washing, the antibody and bound proteins are eluted and analyzed. If protein Y binds to protein X either directly (or indirectly), it can be visualized e.g. by Western blotting and immunodetection. D: Immunofluorescence colocalization works by fixation and permeabilization of cells expressing the two proteins of interest. Specific primary antibodies are bound to the proteins (or epitope tags fused to them) in the cells and detected by secondary antibodies coupled to fluorescent dyes, which can be visualized through fluorescence microscopy. If the two proteins do not interact, distinct localizations of the two dyes are observed (i), whereas a colocalization is seen if the two proteins interact either directly or indirectly (ii). This figure is partly based on "Proteins and proteomics: a laboratory manual" by Richard J. Simpson, Cold Spring Harbor Laboratory Press (Simpson 2002).

1.5.4 Protein network analysis

Insights from the first PPI maps generated for model organisms

Using PPI detection methods applied on a large scale, first insights into global interaction maps came from TAP and Y2H studies for the model organisms *S. cerevisiae* (Uetz *et al.* 2000; Ito *et al.* 2001; Gavin *et al.* 2002), *D. melanogaster* (Giot *et al.* 2003) and *C. elegans* (Li *et al.* 2004b). It was found that high-throughput networks as well as literature-curated networks share some important topological features:

• **Scale-free connectivity**: Most nodes (genes or proteins) in biological networks are sparsely connected, whereas a few nodes, called hubs, are highly connected. The probability P(k) that a node in the network interacts with k other nodes decays as an apparent power law, following $P(k) \sim k^{-\gamma}$ (Barabasi and Albert 1999). The evolution of this topology can be explained by the preferential attachment of new nodes to ones that already have many links, in a process related to gene duplication (Barabasi and Oltvai 2004);

• **Robustness**: as a consequence of protein network topology, random loss of a gene function mostly affects one of the many proteins with only a few partners, rather than one of the small number of hub proteins, rendering the network robust to changes (Barabasi and Oltvai 2004);

• **Small-world organization**: locally dense regions of connected proteins are sparsely connected to other such regions, but the overall average path length is relatively short (Watts and Strogatz 1998). Such characteristics are also seen in other networks, such as the World Wide Web and social networks.

• **Network motifs:** recurring patterns of interconnection often form network themes, i.e. classes of higher-order recurring interconnection patterns that encompass multiple occurrences of network motifs (Zhang *et al.* 2005); and

• **Correlation of lethality and centrality**: essential proteins are enriched for PPIs with other essential proteins, and the likelihood that removal of a protein will prove lethal correlates with the protein's connectivity (Jeong *et al.* 2001).

Notwithstanding their common features, critical comparison of PPI maps revealed surprising little overlap between different data sets (Bader and Hogue 2002). For example, the overlap between PPIs from the large Y2H screens for yeast was only six interactions (Ito *et al.* 2001), and the overlap between the *Drosophila* screens was only 28 PPIs, or 2% of the smaller data set (Stanyon *et al.* 2004). Data in high-throughput screens also fail to overlap significantly with published studies on single genes. Among the main reasons for these observations may be sampling effects. Every method and every screen picks out only one or part of all interactions that a protein can have *in vivo*. Also, relatively high false negative and false positive rates of both the Y2H and the TAP approach influence the outcome. Nevertheless, high-throughput screens have already contributed a wealth of data to the field of interaction mapping, and advances in experimental and computational validation of newly identified interactions are likely to reduce the fraction of false or artificial PPIs in future data sets.

Assembly of human PPI networks

After initial efforts dedicated to model organisms, the focus has shifted towards the analysis of human PPI networks. First large-scale human pair-wise PPI maps have been assembled using strategies belonging to one of the following three classes (Futschik *et al.* 2007):

1. **maps derived from literature search**, either through manual curation (Bader *et al.* 2001; Peri *et al.* 2003) or literature-mining algorithms (Ramani *et al.* 2005);
2. **maps obtained from interologs**, i.e. interactions between orthologous proteins in other organisms (Lehner and Fraser 2004; Brown and Jurisica 2005; Persico *et al.* 2005); and
3. **maps from large Y2H screens** (Rual *et al.* 2005; Stelzl *et al.* 2005).

A recent effort to unify these eight large-scale interaction maps combines three literature-based, three orthology-based and two Y2H-based maps by conversion of all proteins to EntrezGene IDs (Futschik *et al.* 2007). The sizes of these PPI maps

vary between 2,754 and 15,658 interactions, with average degrees (PPIs/protein) of 2.9-7.8, which is in the range of previous estimates of 3-10 (Bork *et al.* 2004).

It was found that the current PPI maps have only a small, but significant overlap. Although most proteins can be found in multiple maps, this is the case for less than 10% of the corresponding PPIs. Thus, different PPI maps are largely complementary at the moment.

1.5.5 Connecting interaction data with signaling pathways

An interesting question is how proteins interact in regulatory networks to form signal transduction pathways that influence the physiological state of cells and tissues according to changes in the micro- and macroenvironment. In the case of Down syndrome, the question is how the gene dosage effects resulting from trisomy 21 affect and perturb cellular signal transduction pathways, so that pathophysiological changes arise which lead to the observed phenotypic effects.

In recent years, a number of databases have been established that try to accumulate data and knowledge about cellular signaling pathways. Unfortunately, when looking for involvement of chromosome 21 proteins in pathways using these databases, only few annotations can be found. In the three major pathway databases, namely TransPath, KEGG and STKE (Gough 2002; Kanehisa *et al.* 2006; Krull *et al.* 2006), of all Hsa21 proteins, only five gene products (IL10RB, IFNAR2, IFNGR2, IFNAR1 and APP), which are all membrane-bound proteins, were directly annotated as parts of signaling cascades, namely the interleukin (IL), interferon (IFN) and c-Jun N-terminal kinase (JNK) pathways.

To overcome this current limitation in pathway annotation, which is also a direct consequence of poor functional annotation of many human proteins, it can be reasoned that new pathway associations may be established via existing pathway annotations of direct interactors of Hsa21 proteins. In other words, Hsa21 proteins that associate with other proteins that are part of a signal transduction pathway have a significant potential to alter the state of that pathway. Although not every protein interaction will be informative enough, it is essential to assemble a list of all PPIs that connect Hsa21 proteins with proteins annotated to be part of a signaling cascade.

Such a ‘pathway connection list’ will of course be limited in its comprehensiveness due to the currently limited knowledge about all protein interactions and the different

activation states of signaling proteins. Nevertheless, as will be shown in this work, a number of new connections arise from such a pathway connection analysis, which provide valuable information for the identification of pathways that might be of relevance in the study of trisomy 21.

1.6 Objectives of this study

During the last decades, researchers working on genes and their interaction with other genes in health and disease have accumulated a vast amount of data (and knowledge) on the functions of many human genes. Still, there is an astoundingly high number of 'orphan' genes whose function has not been analyzed due to lack of immediate interest in these genes, e.g. because there is no obvious associated phenotype or pathology. For example, there are only 2,427 known gene loci associated to human pathologies according to the OMIM database (Hamosh *et al.* 2005). Consequently, this relatively large fraction of functionally uncharacterized genes prevents researchers working on diseases with multifactorial causes, such as trisomy 21, to formulate molecular hypotheses on a more system-wide level, including modeling of complex functional and regulatory networks.

The aim of the work presented here was to apply a functional genomics approach using clone-based proteomics analyses (1) to contribute to the functional characterization of the proteins encoded on human chromosome 21, and (2) to analyze protein networks in regard to their possible involvement in the development of the phenotypes observed in Down syndrome.

As resource for further clone-based functional studies, a comprehensive collection of all the open reading frames encoded on Hsa21 was to be cloned as starting point for protein expression and characterization. Then, the applicability of high-throughput assays for functional genomic analysis should be tested using two experimental approaches, (a) to identify the subcellular localizations of Hsa21 proteins using organelle co-localization assays on transfected cell arrays, and (b) to identify new PPIs of Hsa21 proteins using an automated mating array-based yeast two-hybrid screen. After consolidation with all previously reported data sets into a comprehensive collection of protein-protein interactions for Hsa21 proteins, all identified interactions should be analyzed for their involvement in functional protein modules. Moreover, the entire Hsa21 protein interaction data set was to be analyzed for previously uncharacterized connections of Hsa21 proteins to regulatory networks and signal transduction pathways.

Finally, all assembled functional data sets were to be made accessible to the scientific community to enable further functional analyses as well as future expansion of systems biology models on the complex etiology of Down syndrome.