Genomic and functional profiling of gastric cancer

Samuel Myllykangas
Helsinki 2008

Genomic and functional profiling of gastric cancer

Samuel Myllykangas

Department of Pathology
Haartman Institute and HUSLAB
University of Helsinki and Helsinki University Central Hospital
Helsinki, Finland

Helsinki Biomedical
Dissertations no. 113

Academic Dissertation
To be presented, with the permission
of the Faculty of Medicine, University of Helsinki,
for public examination in the Niilo Hallman
Hall, Childers's Hospital, Helsinki University
Central Hospital, Stenbäckinkatu 11, on September 19th,
2008, at 12 noon.

Supervised by
Professor **Sakari Knuutila,** PhD
Department of Pathology
Haartman Institute and HUSLAB
University of Helsinki and Helsinki University Central Hospital
Helsinki, Finland

Reviewed by
Professor **Pentti Sipponen,** MD, PhD
Division of Pathology
HUSLAB
Helsinki University Central Hospital Jorvi
Espoo, Finland

Research Professor **Matej Orešič,** PhD
Quantitative Biology and Bioinformatics
VTT Technical Research Centre of Finland
Espoo, Finland

Official opponent
Professor **Juha Kere,** MD, PhD
Department of Biosciences and Nutrition
Karolinska Institutet
Huddinge, Sweden

ISSN 1457-8433
ISBN 978-952-10-4927-9
ISBN 978-952-10-4928-6 (PDF)

”Tietä käyden tien on vanki.
Vapaa on vain umpihanki.”
Aaro Hellaakoski

Rakkailleni, **Ninalle**, **Saaralle** ja **Viljamille**

Table of contents

List of publications

This thesis consists of an introductory part and the following publications:

I Myllykangas S, Monni O, Nagy B, Rautelin H, Knuutila S. *Helicobacter pylori infection activates FOS and stress-response genes and alters expression of genes in gastric cancer-specific loci.* Genes Chromosomes Cancer 2004, 40:334–341.

II Myllykangas S, Saharinen J, Veckman V, Knuutila S. *Helicobacter pylori stimulation regulated cellular signaling in AGS human gastric cancer cells.* Submitted.

III Myllykangas S*, Junnila S*, Kokkola A, Autio R, Scheinin I, Kiviluoto T, Karjalainen-Lindsberg M-L, Hollmén J, Knuutila S, Puolakkainen P, Monni O. *Integrated gene copy number and expression microarray analysis of gastric cancer highlights potential target genes.* Int J Cancer 2008, 123:817–825.

IV Myllykangas S, Himberg J, Böhling T, Nagy B, Hollmén J, Knuutila S. *DNA copy number amplification profiling of human neoplasms.* Oncogene 2006, 25:7324–7332.

V Myllykangas S, Tikka J, Böhling T, Knuutila S, Hollmén J. *Classification of human cancers based on DNA copy number amplification modeling.* BMC Medical Genomics 2008, 1:15.

The roman numbers (I–V) are used in the text when referring to the publications.

* Equal contribution

Abbreviations

(a)CGH,	(Array) comparative genomic hybridization
BFB,	Breakage-fusion-bridge
CML,	Chronic myeloid leukemia
EMSA,	Electrophoretic mobility shift assay
HSR,	Homogeneously staining region
ICA,	Independent component analysis
PCA,	Principal component analysis
PCR,	Polymerase chain reaction
RNS,	Reactive nitrogen species
ROS,	Reactive oxygen species
ROC,	Receiver operating characteristics

Gene symbols are marked in italics and according to the guidelines of the Human Genome Organization nomenclature committee (HGNC). More detailed description can be found at http://www.genenames.org/.

Abstract

Background

Helicobacter pylori infection is a risk factor for gastric cancer, which is a major health issue worldwide. Gastric cancer has a poor prognosis due to the unnoticeable progression of the disease and surgery is the only available treatment in gastric cancer. Therefore, gastric cancer patients would greatly benefit from identifying biomarker genes that would improve diagnostic and prognostic prediction and provide targets for molecular therapies. DNA copy number amplifications are the hallmarks of cancers in various anatomical locations. Mechanisms of amplification predict that DNA double-strand breaks occur at the margins of the amplified region.

Objectives

The first objective of this thesis was to identify the genes that were differentially expressed in *H. pylori* infection as well as the transcription factors and signal transduction pathways that were associated with the gene expression changes. The second objective was to identify putative biomarker genes in gastric cancer with correlated expression and copy number, and the last objective was to characterize cancers based on DNA copy number amplifications.

Methods

DNA microarrays, an *in vitro* model and real-time polymerase chain reaction were used to measure gene expression changes in *H. pylori* infected AGS cells. In order to identify the transcription factors and signal transduction pathways that were activated after *H. pylori* infection, gene expression profiling data from the *H. pylori* experiments and a bioinformatics approach accompanied by experimental validation were used.

Genome-wide expression and copy number microarray analysis of clinical gastric cancer samples and immunohistochemistry on tissue microarray were used to identify putative gastric cancer genes. Data mining and machine learning techniques were applied to study amplifications in a cross-section of cancers.

Results

FOS and various stress response genes were regulated by *H. pylori* infection. *H. pylori* regulated genes were enriched in the chromosomal regions that are frequently changed in gastric cancer, suggesting that molecular pathways of gastric cancer and premalignant *H. pylori* infection that induces gastritis are interconnected. 16 transcription factors were identified as being associated with *H. pylori* infection induced changes in gene expression. NF-κB transcription factor and p50 and p65 subunits were verified using elecrophoretic mobility shift assays. *ERBB2* and other genes located in 17q12-q21 were found to be up-regulated in association with copy number amplification in gastric cancer. Cancers with similar cell type and origin clustered together based on the genomic localization of the amplifications. Cancer genes and large genes were co-localized with amplified regions and fragile sites, telomeres, centromeres and light chromosome bands were enriched at the amplification boundaries.

Conclusions

H. pylori activated transcription factors and signal transduction pathways function in cellular mechanisms that might be capable of promoting carcinogenesis of the stomach. Intestinal and diffuse type gastric cancers showed distinct molecular genetic profiles. Integration of gene expression and copy number microarray data allowed the identification of genes that might be involved in gastric carcinogenesis and have clinical relevance. Gene amplifications were demonstrated to be non-random genomic instabilities. Cell lineage, properties of precursor stem cells, tissue microenvironment and genomic map localization of specific oncogenes define the site specificity of DNA amplifications, whereas labile genomic features define the structures of amplicons. These conclusions suggest that the definition of genomic changes in cancer is based on the interplay between the cancer cell and the tumor microenvironment.

Abstract

Introduction

Understanding the mechanisms underlying diseases furthers the development of treatments. Throughout the history of human illness, technological progress has played a central role in determining causes of diseases and laying the foundations for treatment **(Table 1)**. A major trend in technological advancement in medicine has been that the causes of diseases are interrogated in increasing resolution. Hippocrates defined diseases as having natural causes and treated patients with natural products to restore the humoral balance, whereas the development of microscope enabled the identification of bacteria as disease causing agents that lead to the discovery of antibiotics to treat infectious diseases. Knowledge about DNA structure (Watson and Crick, 1953) started a new molecular era in biology. Molecular biology is based on the theory of sequential information flow in living organisms, "The central dogma of molecular biology" (Crick, 1970; Crick, 1958). The dogma presents the information transfer events between three main biomolecules: DNA, RNA (nucleic acids) and proteins **(Figure 1)**. Nucleic acids and proteins are biopolymers, which are made up of single building blocks. Nucleic acids are composed of nucleotides (Adenine, Guanine, Cytosine, Urasil, Thymine; also referred to as A, G, C, U, T or bases) and proteins are made up of 20 different amino acids. Genetic information is translated from nucleotide sequence into amino acid sequence according to the genetic code, in which nucleotide codons (words of three nucleotides or triplets) correspond to specific amino acids. The genetic information is stored and distributed in DNA, transmitted via RNA and to be put into practice by proteins. The information may be transferred from one biopolymer compartment to another according to the rules formulated by **Francis Crick** (Crick, 1970; Crick, 1958). General transfer operations, which occur in

Table 1.
Biomedical timeline.
Cancer related inventions are marked red.

Time	Developer	Technology		Cause	Treatment
400 B.C.	Hippocrates	Western medicine	RESOLUTION	Environmental factors, nutrition, living habits, benign and malignant tumors	Natural products
1000	Ibm al-Haytham	Scientific method			
1020	Ibm Sina				Surgery
1600	Dutch	Microscope			
1723	Antonie van Leeuwenhoek	Microbiology			
1750	Carl von Linné	Taxonomy		Symptoms	
1860	Rudolf Virchow	Pathology		Cell / Leukemia cell	
1874	Campbell De Morgan			Metastasis	
1860 – 1880	Robert Koch and Louis Pasteur			Bacteria	Vaccination
1895	Wilhelm Röntgen			Radiation	Radiation
1900	Carl Neuberg	Biochemistry			
1914	Theodor Boveri			Chromosomal changes	
1928	Alexander Fleming				Antibiotics
1948	Faber et al.				Chemotherapy
1953	Francis Crick and James Watson	Molecular biology	THROUGHPUT	Genes	
1960	Peter Novell Philadelphia			chromosome	
1970	Intel	Microprocessor			
1971	Alfred Knudson			Two-hit hypothesis	
1974	Janet Rowley			Ph-chromosome: 9; 22 translocation	
1977	Fredrik Sanger	Sequencing			
1979				TP53 and tyrosine kinases	
1982	Barry Marshall and Robin Warren			*H. pylori*	
1984 – 1990				*BCL-2, RB, ERBB2, BRCA1, BCR-ABL*	
1992	Olli Kallioniemi	Comparative genomic hybridization			
1995	Patric Brown	DNA microarray		Gene groups	
2001	IGC / Novartis	Human genome sequence			Imatinib mesylate
2007	Illumina, Roche / Agendia	Massively parallel sequencing		MammaPrint	

most of the living organisms, include replication, transcription and translation. Replication is a process where DNA sequence is copied. DNA sequence is usually in double-strand form, and in most cases this feature is preserved in replication. In transcription, single-strand RNA is synthesized using a DNA template. RNA code is complementary to DNA template, given that T is replaced with U. In turn, RNA sequence is used as a template in translation, where specific three nucleotide triplets, codons, correspond to specific amino acids. In addition, there are some information transfers that only occur in abnormal conditions, laboratory experiments or in some viruses.

The development of molecular biology techniques were of remarkable consequence in cancer research, since cancer is a disease where DNA sequence information is erroneously encoded. According to the fundamentals of information flow in eukaryotic cells, DNA sequence errors (also referred to as mutations) are transferred to

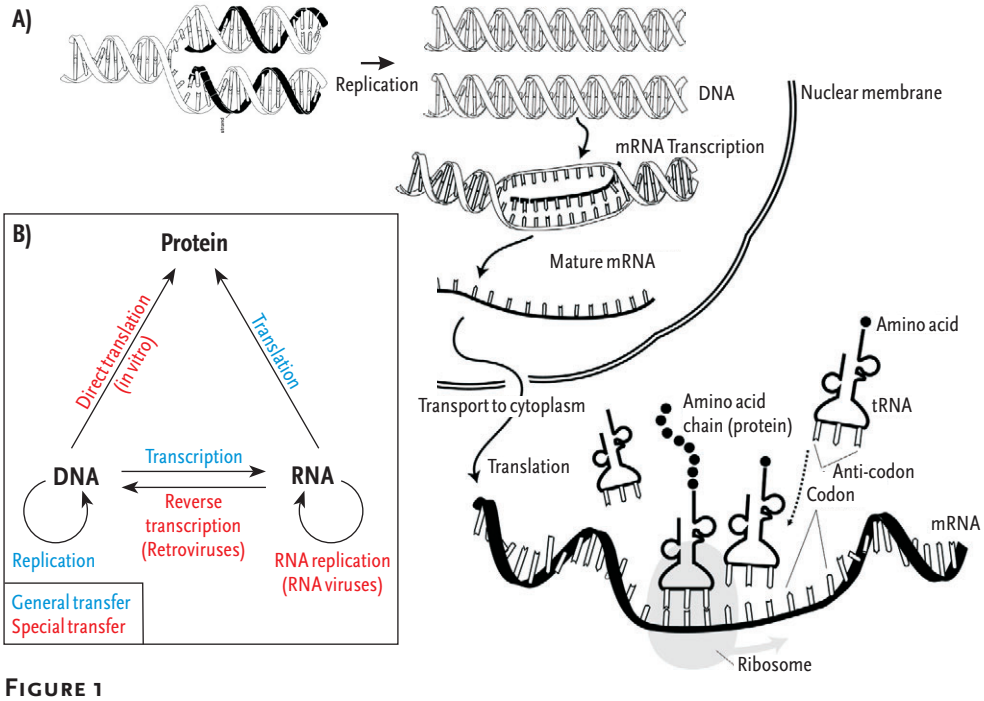


Figure 1

Central dogma in molecular biology

In living organisms, genetic information flows through DNA, RNA and proteins. **A)** An illustration of the information transfer events in eukaryotic cells. DNA doubles in replication in semi conservative manner as both of the complementary strands serve as a template for newly synthesized DNA. In transcription, the DNA code is reformatted to the RNA code. Mature messenger RNA (mRNA) translocates out from the nucleus and translation takes place in the ribosomes in the cytosol. The transfer RNA (tRNA) joins the amino acids (building blocks of proteins) according to the information encoded in the mRNA sequence. **B)** Schematic overview of the information transfer events. Biological data flows through general information transfer events in majority of living organisms. In special situations, additional information transfer events exist. The central dogma states that once the information is translated into protein it can not go back to the nucleic acid code. Figure produced after www.genome.gov/glossary.cfm with permission from the National Human Genome Research Institute.

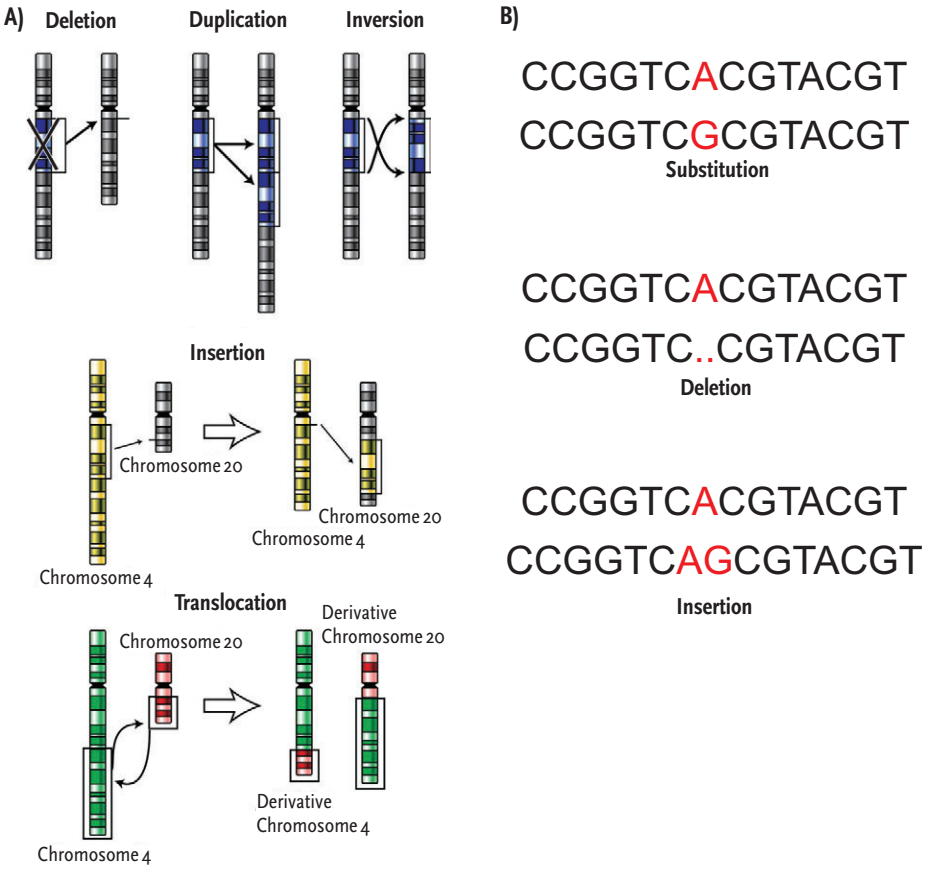


Figure 2
Mutations in cancer
A) Chromosomal rearrangements. Image is obtained from www.genome.gov/glossary.cfm with permission from the National Human Genome Research Institute and **B)** point mutations.

RNA in transcription and to daughter cells through DNA replication during cell division. Altered RNA sequence translates into altered amino acid sequence, and the abnormal protein structure has an altered function in the cell. Eventually, DNA mutation encoded perturbations in protein function lead to the transformation of a normal cell to a cancer cell. In healthy tissue, cell death and renewal are balanced and controlled. Tight surveillance of cell division and disposal enable the proper cell function and number in tissues and organs. In cancer cells, the mechanisms that execute the surveillance of cell existence have been turned off, which leads to uncontrolled growth, limitless replication, acquisition of nutrients and expansion. **Hanahan** and **Weinberg** have defined the physiologic changes, six hallmarks, that characterize the cancer cell phenotype (Hanahan and Weinberg, 2000). Uncontrolled growth requires that

cancer cells are (1) independent of growth supporting signaling and (2) notwithstanding growth restraints. Cancer cells acquire limitless replication potential by becoming (3) immune to programmed cell death (apoptosis) and (4) free from natural replication restraints (telomere shortening). In order to proliferate cancer cells need nutrients. Thereby, (5) activation of neovascularization and vascular maintenance are required for tumor growth. Because of this, cancer cells must possess the capability to promote angiogenesis, generation of new blood vessels into solid tumor mass. Malignant capacity is achieved when cancer cells (6) lose the control of their primary anatomical location and begin tissue invasion and to metastasize to distant sites.

Cancer cells acquire the hallmarks of cancer and the capacity to grow unsupervised when mutations change the function of cancer genes (Hanahan and Weinberg, 2000). Cancer genes are originally normal human genes that function in cell growth, replication, energy metabolism and microenvironment modeling. By definition, when mutated, they are capable of promoting and participating in cancerous cell growth. Cancer genes may be activated by gain-of-function mutations (oncogenes) and silenced by loss-of-function mutations (tumor suppressor genes). Even when inherited germ line mutations and their induced cancer syndromes are known to promote the emergence of a subset of cancers, mutations of cancer genes are usually somatic, sporadic changes in DNA sequence of non-gamete cells. The mechanisms that induce genomic changes include chromosomal rearrangements (Albertson et al., 2003) and nucleotide point mutations **(Figure 2)**. These mutations can cause both, gain and loss of function of genes, depending on the end point. The point mutations that alter DNA sequence at a specific nucleotide can activate an oncogene, if it encodes a change in oncoprotein that renders it immune to expression regulation or whether it has an increased activity. Nucleotide point mutations can alter the reading frame in the gene or introduce a new stop codon that destroys the original protein product. Structural chromosomal rearrangements, such as translocation and inversion, can produce gene fusions that activate oncogenes or break and suppress function of genes. Inversion is a situation in which a genomic segment flips around to another direction. In addition to translocations, fusion genes can emerge also when a genomic sequence is deleted and the ends of a broken chromosome are rejoined. Translocation refers to a situation where genetic materials from two different chromosomes are joined and form a novel genotype at the site of the junction. Translocation can be either balanced, when no DNA material is lost and no additional chromosomal fragments are present in the cell, or unbalanced, if genomic material is lost from the cell during the rearrangement. Alterations that change the DNA copy number are frequently observed in a variety of human cancers (Mitelman et al., 1994). Normal diploid human genome contains two copies of each gene and alterations in gene copy number are one of the main mechanisms that activate oncogenes.

Previously, the treatment of cancer has relied on surgery and radiation therapies as well as on chemotherapies. Although they are useful in treating locally occurring tumors, these techniques are ineffective against tumors that grow in unapproachable locations and metastasize to distant sites. It has become evitable that targeted therapies are needed to treat wide-spread disease. Advances in molecular biology provided new means to determine mutated genes and attack cancer cells by targeting the oncoproteins that they encode. Major breakthroughs have been established using molecular biology techniques in exploring genetic

basis of cancer, such as identification of *TP53* and *RB* tumor suppressor genes and molecular changes in chronic myeloid leukemia (CML). The use of molecular targeted drugs in CML is one of the greatest successes in the treatment of cancer. CML is characterized by a chromosomal rearrangement, Philadelphia chromosome, which is formed in translocation between chromosomes 9 and 22. In the site of the translocation a fusion of the *BCR* and *ABL* genes takes place. Gene fusion causes increased activation of the ABL tyrosine kinase. Imatinib mesylate (also known as Gleevec® or Clivec®) is a small molecule inhibitor of the kinase activity of ABL and in 2001 it was approved as a therapeutic agent for the treatment of CML. In addition, Imatinib mesylate has been successfully used as a therapeutic agent in the treatment of gastrointestinal stromal tumor, a mesenchymal tumor characterized by KIT positivity, as, besides ABL, Imatinib mesylate has been shown to inhibit the KIT tyrosine kinase activity (Miettinen and Lasota, 2006).

In spite of the great improvements that have been made in order to unreveal the molecular backgrounds of cancer and of the fact that some cancers with specific genetic changes can be treated using target drugs, most of the cancers still remain untreatable. The majority of the most lethal forms of cancer is multifactorial, and forms a heterogeneous set of diseases that do not have specific mutations that could be targeted with drugs. Instead of one specific target there might be diverse sets of genes that are perturbed and are causing disease. Traditional molecular biology provides a reductionist approach that tries to resolve the complexity

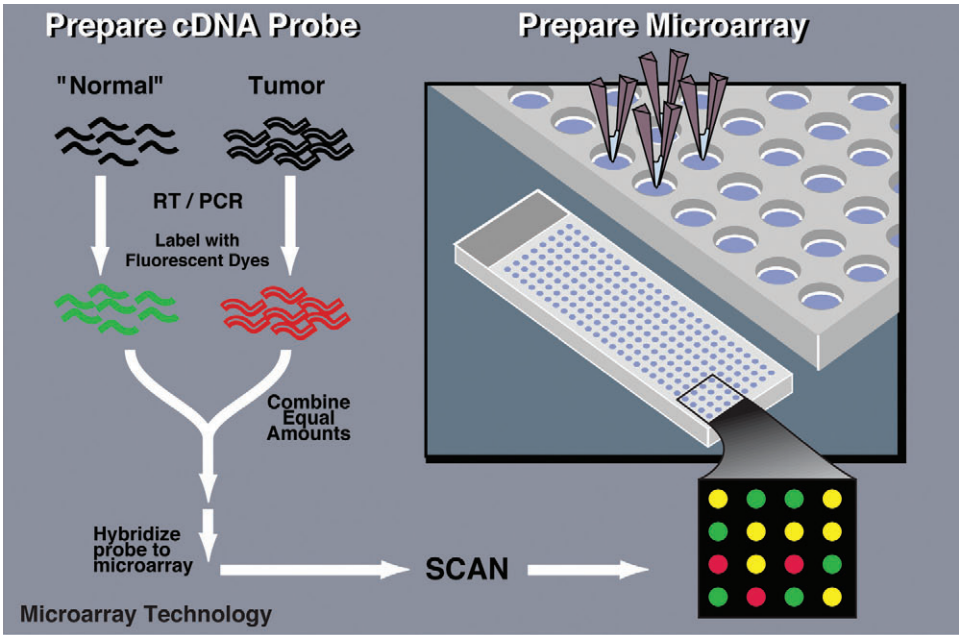


Figure 3
Principles of a microarray experiment
From www.genome.gov/glossary.cfm. Image is used with permission from the National Human Genome Research Institute.

in living organisms by studying individual molecular objects separately. It is not possible to identify groups of disease genes using traditional molecular biology techniques, since it is difficult to be sure whether the objects of research are the right ones or whether there are combinatorial effects between the biological components. To circumvent these limitations, a second major technological trend in biomedicine has been to increase the throughput of the experimental measurements. In this perspective, comparative genomic hybridization (CGH) (Kallioniemi et al., 1992) had an unprecedented impact on the field of cancer research because it introduced new technological concepts that allowed up-scaling of the measurement capacity. Firstly, CGH was a genome-wide technology for high-throughput examination of the DNA copy number. Secondly, in CGH, a hybridization probe was placed on a solid support and samples were labeled using fluorescent dyes. Thirdly, CGH was based on measuring relative changes in fluorescence intensities between the test sample and the reference sample. Further development of genome-wide technologies has been largely relying on these principles.

Another true milestone in the development of biomedical technologies was the sequencing of the human genome (Lander et al., 2001; Venter et al., 2001). In addition to revealing the organisms' building instructions, a detailed knowledge of genome sequences has facilitated the genome-wide analysis of gene expression (RNA synthesis) patterns in the field of functional genomics using DNA microarrays (DeRisi et al., 1996; Schena et al., 1995). Whereas in CGH, in which metaphase chromosomes were used as probes, DNA microarrays consist of collections of single-strand DNA probes spotted on a glass surface (Monni et al., 2002). Specific probes with sequences that are complementary to different genomic counterparts, genes or genomic map locations, can be produced. Oligonucleotide probes from 20 to 60 nucleotides (Agilent Technologies and Affymetrix, Palo Alto, Ca) can be synthesized and cDNA probes were collected from DNA libraries (Monni et al., 2002). Microarray experiment is based on the hybridization of complementary, single-strand DNA fragments and detection is made possible by labeling the sample using fluorescent dyes **(Figure 3)**. Microarray experiments are semi-quantitative as the sample is co-hybridized with a reference, and changes in fluorescence intensity ratios are measured. Either DNA or RNA may be used as a sample in microarray experiment, although, RNA has to be reverse-transcribed to complementary DNA (cDNA) using specific viral enzymes (Monni et al., 2002). The amount of RNA in a given condition is a measure for gene expression and functionality. While post-transcriptional regulation of gene expression exists, the intensity of transcription is one of the main regulators of gene function. Similarly, the amount of DNA corresponds to the gene copy number. In addition to the basic research, microarray technologies are starting to penetrate the clinical practice as products like MammaPrint®, the Food and Drug Administration of the United States approved diagnostic test, which uses measurements of some 70 genes to predict the metastatic behavior of breast tumors, are being developed (Buyse et al., 2006). Genome-wide technologies have exponentially increased the amount of quantitative data produced by biomedical research. Large-scale datasets can not be processed using human resources, but the development of microprocessor and propagation of computers have allowed high-throughput analysis of genome-wide biomedical data. There are databases for storing and managing genome-wide datasets (Edgar et al., 2002) and genomic information (Flicek et al., 2008) as well as bioinformatics methods are needed in order to analyze genome-wide measurements and to interpret multifactorial biological phenomena. ❦

Review of the literature

Gastric cancer is a major health issue in Finland and worldwide. Close to one million new gastric cancer cases, 9% of all cancers, were diagnosed in the year 2002 alone (Ferlay et al., 2004). As an example of the scope of the problem, gastric cancer was the second most common cause of cancer-related death. In Finland, gastric cancer ranked sixth in mortality with close to 600 annual deaths and eighth in prevalence (over 700 new cases diagnosed yearly). Gastric cancer is the fourth most common cancer (after lung, breast, and colorectal cancers) (Ferlay et al., 2004). The highest incidence areas are Asia, Eastern Europe and the Andean region in South America, whereas low rates are found in America, Northern Europe, South-East Asia and Africa **(Figure 4)**. Incidence and risk in high risk areas are usually as much as ten times larger than in low risk areas (Figure 4 and **Table 2)**. Gastric cancer is more frequent in males than in females, and usually affects elderly, as 75% of gastric cancer patients are over 54 years old **(Figure 5)**.

There are two distinct subtypes of gastric cancer, intestinal and diffuse (Lauren, 1965). Undifferentiated diffuse type is characterized by non-cohesive and scattered cancer cells that grow by infiltrating deep into the stroma (Werner et al., 2001). Intestinal gastric cancer forms distinguishable and distorted glandular structures and grows by expansion (Werner et al., 2001). The Intestinal type gastric cancer progresses through sequence of premalignant lesions **(Figure 6)**. Generally, diffuse type gastric cancer is not characterized by stepwise progression and intermediate steps but there is evidence that premalignant polyps are found in gastric crypts preceding hereditary diffuse type gastric cancer (Oliveira et al., 2005). Moreover, premalignant stages, chronic gastritis, atrophy and achlorhydric stomach and intestinal metaplasia, are risk factors in the pathogenesis of intestinal gastric cancer, while simple inflammatory stress is associated with the diffuse type (Fenoglio-Preiser et al., 2000). The intestinal type gastric cancer affects older patients and it has slightly better prognosis than the diffuse type, which is more prevalent in younger patients (Fenoglio-Preiser et al., 2000). The intestinal type is more common in males than in females but similar gender bias is not observed in the diffuse type (Teh and Lee, 1987). Hereditary gastric cancer, associated with *CDH1* mutations,

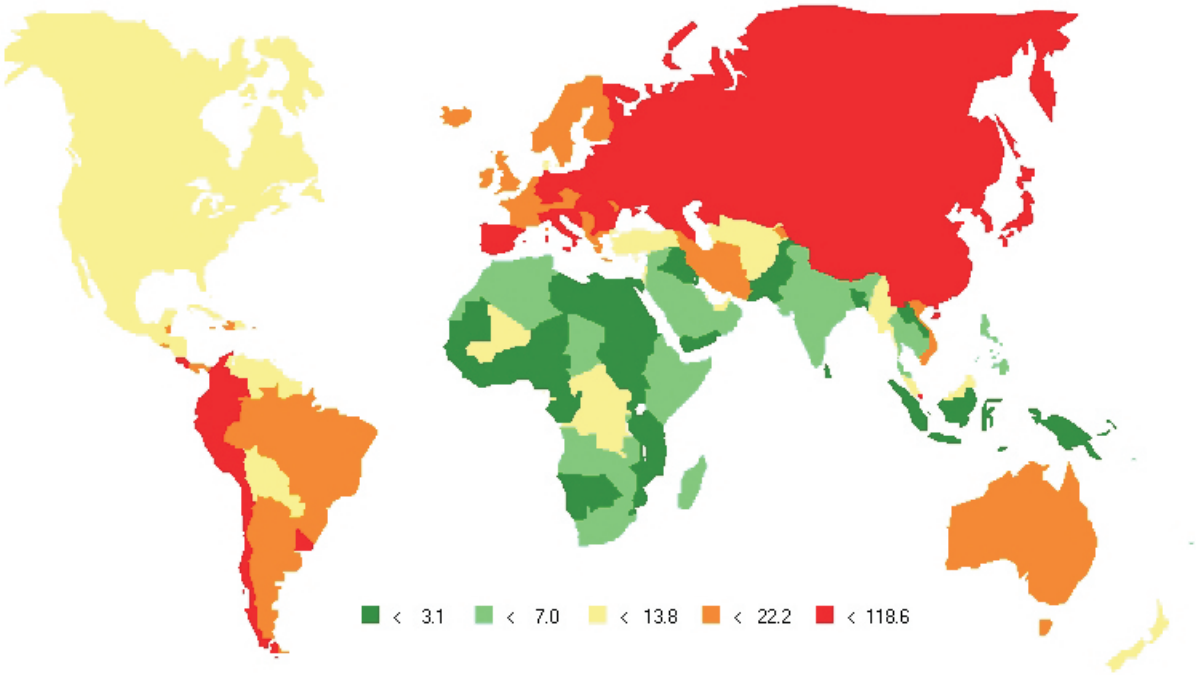


Figure 4
Demographic distribution of gatric cancer incidence
Gastric cancer incidence is shown in males (per 100 000).
Figure was produced using Globocan 2002 software (Ferlay et al., 2004)

manifests solely as diffuse type (Oliveira et al., 2006). All in all, different subtypes of gastric cancer have both biological and clinical distinctive features.

Gastric cancer is an aggressive disease, and 5-year survival rates are only 10 to 30% (Green et al., 2002; Harrison et al., 1998; Msika et al., 2000). The high mortality rate of gastric cancer is explained by the fact that the early stages of gastric cancer are often asymptomatic and left undetected. Thus, the majority of the diagnosed tumors in the stomach are malignant gastric adenocarcinomas (Schwartz, 1996), and the clinical panorama of gastric cancer is dominated by tumors that invade lymph nodes and metastasize to distant locations (Hundahl et al., 2000). The treatment of gastric cancer is restricted to gastrectomy (removal of the stomach), endoscopical surgery and lymphadenectomy (removal of lymph nodes). Patients may therefore greatly benefit from predictive diagnosis based on the biomarkers that would guide clinical cancer management and from the administration of targeted therapies.

Molecular markers for diagnostic and prognostic purposes as well as therapeutic targets are required in the treatment of gastric cancer, because of the inconspicuous progression of the disease and its poor response to therapy in later stages. Furthermore, gastric cancer comprises two different subtypes, intestinal and diffuse, that may have different molecular properties and treatment requirements. Several epigenetic (methylation), genetic (polymorphisms, mutations, amplifications and deletions) and function-

al (overexpression) alterations have been associated with gastric carcinogenesis, transformation and cancer progression **(Table 3)**. Even when many genetic and epigenetic defects have been identified in gastric cancer, molecular biology based applications in clinical practice used to treat gastric cancer are still few.

Diet shows the most significant association with gastric cancer in many epidemiological studies (Fenoglio-Preiser et al., 2000). Alcohol consumption and smoking are the main risk factors for gastric cancer (Salaspuro, 2003). Heavy alcohol usage is particularly associated with gastric cancer in Asian individuals, because of the frequent genetic inability to detoxify acetaldehyde, ethanol metabolite. Acetaldehyde has been experimentally shown to be mutagenic and carcinogenic. Moreover, the microbes in the stomach are able to endogenously produce acetaldehyde from ethanol. A high concentration of acetaldehyde in the stomach is associated with pathogenesis of gastric cancer. Sufficient consumption of fresh fruits and vegetables lowers the risk of gastric cancer, whereas intake of salt, smoked or pickled food and chili peppers increases the risk of gastric carcinogenesis (Fenoglio-Preiser et al., 2000). The protective effect of fruits and vegetables is believed to be associated with the antioxidant activity found in these foods. Gastric cancer risk elevates after bile reflux which causes gastric irritation because of the bile fluids. Nonetheless, the most significant development in the etiology of gastric adenocarcinoma was when *H. pylori* infection was identified as a causative factor in gastric carcinogenesis. While diet is considered to be associated with the majority of cancers, *H. pylori* infection seems to be specific to gastric cancer. Because of the strong epidemiological evidence *H. pylori* infection was identified as a risk factor for gastric cancer by the International Agency for Research on Cancer (IARC, 1994).

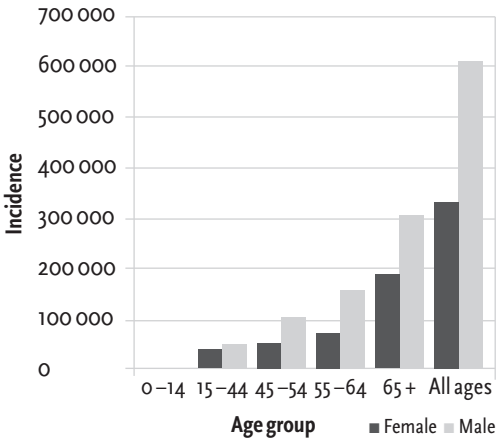


Figure 5
Age distribution of gastric cancer patients
Worldwide incidence of gastric cancer.
Data was collected using Globocan 2002 software (Ferlay et al., 2004).

Table 2
Population-based differences in the incidence and risk of gastric cancer.

Population	Gender	Crude rate	Cases
Finland	Male	15,3	387
Finland	Female	12,3	326
Japan	Male	118,6	73 785
Japan	Female	55,4	35 994

Crude rate; cases per 100 000

H. pylori infection in gastric cancer pathogenesis

The 2005 Nobel Prize winners for Medicine, **Barry Marshall** and **Robin Warren**, demonstrated in 1982 that *H. pylori* infection causes stomach inflammation as well as duodenal and gastric ulcers (Marshall and Warren, 1984). After the observations made by Marshall and Warren, *H. pylori* infection, and subsequent atrophic gastritis,

Table 3
Gastric cancer associated genes. LOH; loss of heterozygocity, GC; gastric cancer.

Gene	Gene function	Role in cancer	Changes	Reference
APC	wnt/β-catenin and TGFβ-pathways	Tumor suppressor gene	Mutations	(Fang et al., 2002)
AURKA	Cell cycle, mitosis	Progression	Functional polymorphism	(Ju et al., 2006; Kamada et al., 2004)
BTRC	wnt/β-catenin pathway		Mutations	(Kim et al., 2007)
CCND1	Cell cycle	High levels are associated with alcohol consumption	Amplified and overexpressed, polymorphisms	(Bizari et al., 2006)
CCNE1	Cell cycle		Amplification and overexpression	(Varis et al., 2003)
CDH1	TGFβ-pathway	Hereditary diffuse gastric cancer	Mutations	(Peek and Reddy, 2007)
CDKN2A	Cell cycle		Inactivation by methylation or LOH	(Zhang et al., 2003; Zhao et al., 2007)
CTNNB1	wnt/β-catenin/ TGFβ-pathway, cell cycle	Associated with invasive and aggressive disease	Mutations	(Cheng et al., 2005)
RHOBTB2		Tumor suppressor gene that has growth inhibitory function	Mutations and LOH	(Cho et al., 2007a)
EGFR	Transmembrane receptor protein kinase	Associated with poor survival, target for tyrosine kinase inhibitors	Increased expression	(Galizia et al., 2007)
ERBB2	Transmembrane receptor protein kinase	Target for inhibitors	Mutations, amplification	(Lee et al., 2006)
GAST	Gastrin hormone, mitogen	Associated with intestinal metaplasia and gastric carcinoma	Amplified and elevated expression	(Dockray, 2004)
HRAS	Signal transduction		Overexpression	(Kim et al., 2000)
KRAS	Signal transduction		Mutations	(Lee et al., 2006)
MET	Tyrosine kinase	Activated in GC and intestinal metaplasia, induces proliferation	Differential expression	(Inoue et al., 2004; Tang et al., 2004)
NRAS	Membrane protein		Mutation	(Sasaki et al., 2004)

Gene	Gene function	Role in cancer	Changes	Reference
PIK3CA	PTEN pathway		Mutations, up-regulation	(Li et al., 2005a)
PPP1R1B	Kinase or phosphatase inhibitor, regulation of apoptosis		Amplified and overexpressed	(Belkhiri et al., 2005; Varis et al., 2004)
PTEN	Cell cycle	Tumor suppressor gene	Mutations, down-regulation	(Li et al., 2005b)
PTGS2	prostaglandin synthesis, inflammation, mitogenesis	Target for inhibitor, prognostic factor	Polymorphisms, over-expression	Numerous
PTPN11	Membrane protein tyrosine phosphatase, cell growth, differentiation, mitosis	*H. pylori* target	Tyrosine phosphorylation of CagA toxin	(Yamazaki et al., 2003)
RB1		Cell cycle, tumor suppressor gene, *H. pylori* target	Mutations, differential expression	(Lan et al., 2003)
RUNX3	Transcription factor	Tumor suppressor gene, *H. pylori* related, apoptosis, potential prognostic factor	Inactive by down-regulation or methylated	(Homma et al., 2006; Ito et al., 2005)
STK11	Protein kinase	Tumor suppressor gene, Peutz-Jeghers syndrome associated gastric cancer	Mutations	(Shinmura et al., 2005)
TP53	DNA binding, transcriptional activation, DNA repair, cell cycle arrest, apoptosis, senescence	Tumor suppressor gene, Li-Fraumeni syndrome	Mutations	Numerous
ZFHX3	Transcription factor	Aggressive form	Genetic alterations	(Cho et al., 2007b)

has been identified as the most significant single environmental factor associated with the increased risk for developing gastric adenocarcinoma in many epidemiological, case and animal model studies (Forman et al., 1991; Parsonnet et al., 1991a; Parsonnet et al., 1991b; Recavarren-Arce et al., 1991; Watanabe et al., 1998). *H. pylori* infection has been shown to cause changes that lead to development of pre-cancerous conditions and lesions, i.e., chronic gastritis, gastric atrophy, intestinal metaplasia and dysplasia (Craanen et al., 1992; Rugge et al., 1996; Watanabe et al., 1998). *H. pylori* strains, which carry the virulence factor *cagA* and produce vacuolating cytotoxin A,

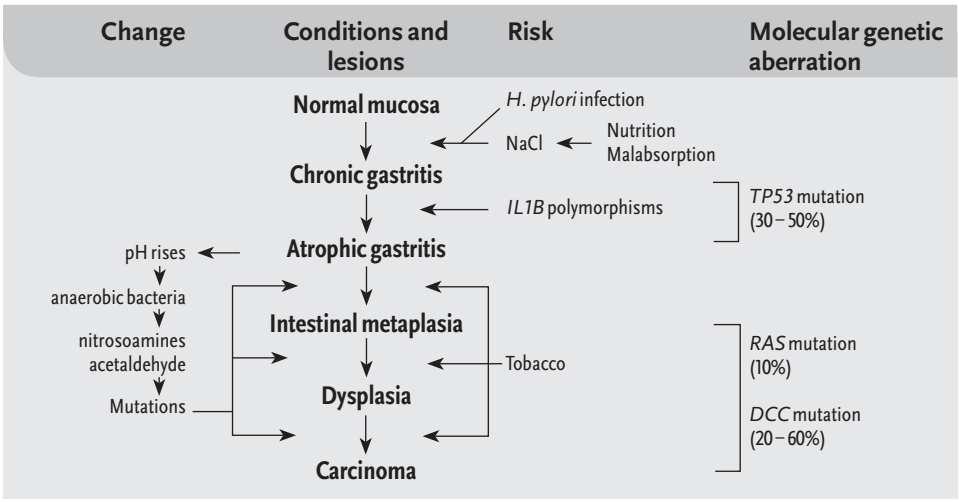


Figure 6
Sequential progression of gastric cancer
Figure after (Correa, 1992; Peek and Blaser, 2002).

are particularly associated with gastric cancer (Atherton, 1997). Although the association of *H. pylori* infection with gastric cancer is undisputable, the molecular genetic mechanisms of how *H. pylori* infection promotes the gastric carcinogenesis still remain unknown.

Prolonged *H. pylori* infection and sequential precancerous process precedes gastric cancer. A multistep progression of gastric cancer includes chronic gastritis, atrophic gastritis, intestinal metaplasia, dysplasia and adenocarcinoma (Correa, 1992) **(Figure 6)**. Gastritis and gastric atrophy eliminate gastric mucosa, which reduces gastric acid secretion and neutralizes gastric juice. Elevated gastric pH allows changes in the gastric flora and it also allows anaerobic bacteria to colonize in the stomach. Anaerobic bacteria produce reductase enzymes that catalyze the nitrite synthesis from food nitrate. Nitrite reacts with amines and urea to produce carcinogenic N-nitroso compounds, which are capable of interacting with DNA. Carcinogenic agents initiate the progression from metaplasia to adenocarcinoma as DNA damage accumulates in the cells. *H. pylori* infection is involved in gastric carcinogenesis as it is the most frequent cause of chronic gastritis and it decreases acid-pepsin secretion and ascorbic acid (dietary antioxidant) concentration in the stomach along with progressing mucosal atrophy. Furthermore, *H. pylori* colonies are observed overlaying and preceding lesions of the intestinal metaplasia.

H. pylori can adapt to the acidic environment of the stomach by secreting urease enzyme that metabolizes urea into ammonium, which neutralizes hydrochloric acid and produces a neutral microenvironment around the organism (Dunn et al., 1990). What is more, *H. pylori* bacteria secrete virulence factors, which damage gastric epithelial cells. Urease not only is a survival factor, it also functions as a virulence factor by inducing an inflammatory reaction (Harris et al., 1996) and toxic effect in gastric epithelial cells (Smoot et al., 1990). In addition to urease, *H. pylori* virulence factors

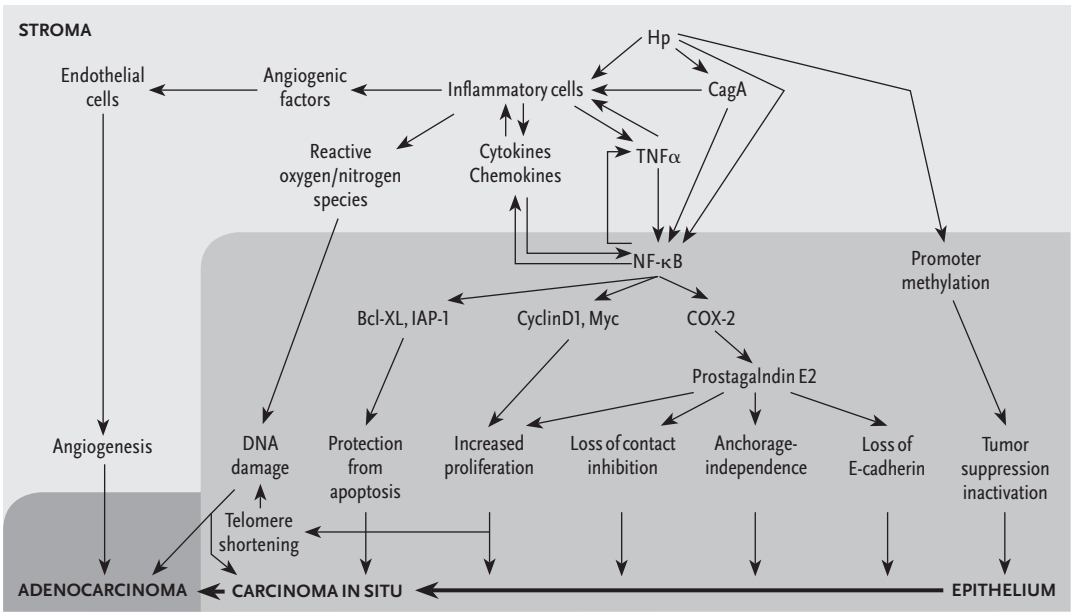
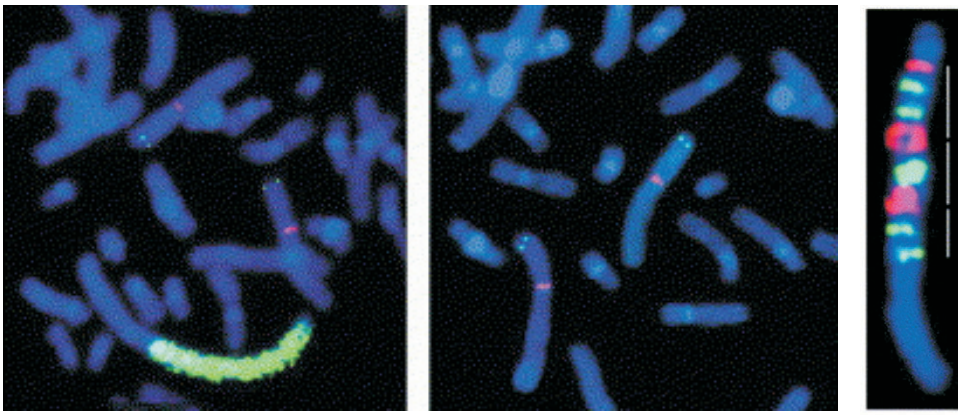


Figure 7
A model of *H. pylori* infection induced inflammation and gastric cancer promotion

include vacuolating cytotoxin A (VacA) and genes in the cytotoxicity associated gene pathogenicity island (*cagPAI*). VacA induces mucosal damage and is associated with gastric carcinogenesis (de Figueiredo Soares et al., 1998). *CagPAI* genes encode a secretion system (Covacci et al., 1999) that transfers an effector protein, cagA, into a host cell (Odenbreit et al., 2000). The cellular membrane tyrosine kinase phosphorylates cagA (Covacci and Rappuoli, 2000). Phosphorylated cagA interacts with host signaling molecules, including SHP-2 (Higashi et al., 2002), and causes morphological changes in the epithelial cells (Moese et al., 2004). Especially *cagA*-positive strains are associated with gastric adenocarcinoma (Blaser et al., 1995). Nevertheless, bacterial strains carrying the *CagPAI* genes induce more intense inflammation than strains that lack cag genes (Kolho et al., 1999; Yamaoka et al., 1997). The interaction between cag secretory system and host cell induces the inflammatory reaction (gastritis). Notwithstanding, production of pro-inflammatory cytokines (interleukin-8) in epithelial cells (Odenbreit et al., 2000) also occur without *cagA* involvement (Fischer et al., 2001).

In addition to direct *H. pylori* assault, endogenous host responses are crucial in determining the progression of *H. pylori* infection in to pre-neoplastic lesions and ultimately gastric adenocarcinoma (Correa and Houghton, 2007) **(Figure 7)**. *H. pylori* infection induces activation of NF-κB (Maeda et al., 2000), which then activates growth factors (e.g., Cyclin D1 and Myc) and suppressors of apoptosis (e.g., BCL-XL) (Karin et al., 2002). Although increase in cell turn-over is not directly associated with malignancy, excess propagation of cells increases the probability of mutations by inducing DNA damage from shortened

A)



HSR Normal Ladder

B)

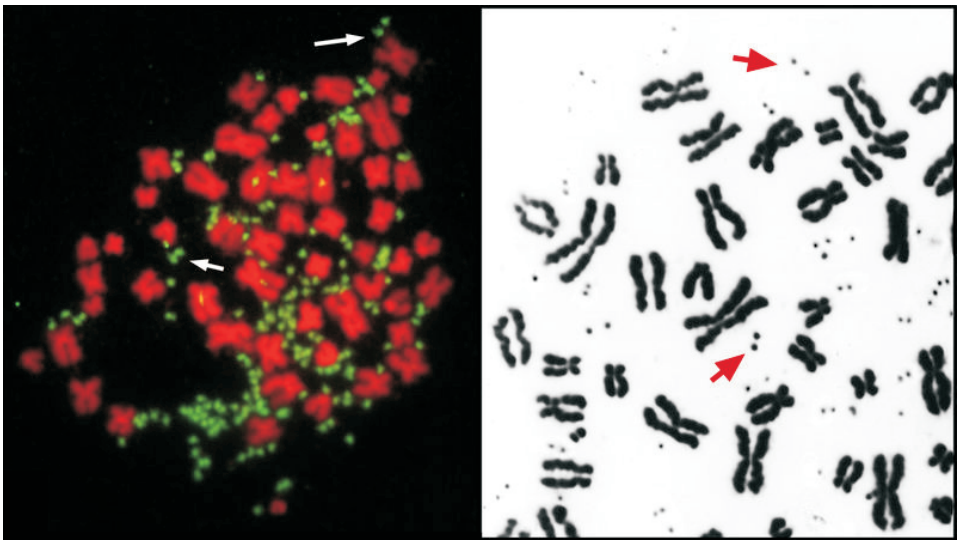


Figure 8
Manifestations of DNA copy number amplification
A) Homogeneously staining region and **B)** Double minute chromosomes. Images from (Schwab, 1998) and www.wikipedia.org.

telomeres. Furthermore, NF-κB activation initiates COX-2 and prostaglandin pathways (Williams et al., 1999) that introduce morphological changes in the epithelium. COX-2 expression induces production of prostaglandin E2, which is a strong inducer of volatile and detached phenotype that is characteristic of cancer cells. PGE2 production leads to the suppression of E-cadherin function and induces anchorage independence and contact inhibition (Weinberg, 2007).

Pathogenic insult usually leads to an acute inflammation response and to a complete clearance of the microbe, but in *H. pylori* infection, inflammatory process fails to eradicate the microbe, infection sustains

and a state of chronic inflammation develops (Correa and Houghton, 2007). A prolonged inflammation (chronic gastritis) and the subsequent atrophic gastritis serve as a platform for carcinogenesis by providing a suitable microenvironment for inducing DNA damage. The inflammation induces mutagenic hits by causing excess cell proliferation (De Luca et al., 2003; De Luca et al., 2004), changes in the epigenetic regulation of the host gene expression (Nardone et al., 2007) and oxidative/nitrosative stress by inducing the production of reactive oxygen and nitrogen species (Ernst, 1999). The inflammatory cells produce cytokines and chemokines (e.g., TNF-α) that act via NF-κB pathway (Karin et al., 2002; Weinberg, 2007) and increase the proliferation of epithelial cells. In the chronic *H. pylori* induced gastritis, the CpG islands of the promoter regions of several tumor suppressor genes have been found to be hypermethylated (Kang et al., 2003). The hypermethylation of a promoter region is an epigenetic change that inhibits transcription and induces gene silencing by blocking the binding of a transcription factor to its target sequence in the promoter. Activated leukocytes produce reactive oxygen species, namely oxygen ions, free radicals and peroxides, and reactive nitrogen species as a means to attack colonizing pathogens (Ernst, 1999). Oxygen and nitrogen metabolites are highly reactive due to the presence of an unpaired electron and they cause damage by reacting randomly and extremely rapidly with various biomolecules, such as DNA, proteins and lipids. *H. pylori* infection promoted inflammation is amplified by the expansion of the inflammatory cell pool via NF-κB–TNF-α–cytokine pathways and DNA damage induced feedback loops.

DNA copy number amplifications in human cancers

Many human cancers of different anatomical locations are characterized by DNA copy number amplifications. Amplification is a chromosomal change that results in an increase of the copy number of a specific DNA region (Albertson et al., 2003; Lengauer et al., 1998). Even hundred-fold elevation in the gene copy number may be present in some tumors. Such high-level amplifications of *MYC* and *EGFR* oncogenes occur in neuroblastoma (Schwab et al., 2003) and glioma (Vogt et al., 2004), respectively. DNA copy number amplification is a local, intra-chromosomal mutation that affects a DNA segment of less than 20 million bp in length. DNA copies generated in amplification manifest as concatenate homogenously staining chromosomal regions (HSRs) and extra-chromosomal acentric DNA fragments (double minutes and episomes) (Albertson et al., 2003; Schwab, 1998). HSRs form a ladder-like structure of inverted repeats within chromosomes (Schwab, 1998) **(Figure 8A)**. Double minute chromosome bodies are extra-chromosomal, circular DNA segments (Hahn, 1993) **(Figure 8B)**. An excess chromosome or chromosome segments, HSRs and double minute chromosomes can be detected using standard microscopic techniques (Schwab et al., 2003; Vogt et al., 2004) and comparative genomic hybridization (CGH) (Kallioniemi et al., 1992). Episomes are submicroscopic extra-chromosomal DNA segments of ~250 bp in length, which can be detected with molecular biology methods, such as fluorescent *in situ* hybridization and microarray CGH (Graux et al., 2004; Maurer et al., 1987). The amplified region may contain gene fusions, i.e., DNA from different genomic sites (Graux et al., 2004; Guan et al., 1994).

DNA copy number amplification resulting in the formation of a homogeneously staining region has been proposed to have occurred according to the Breakage-Fusion-Bridge (BFB) model **(Figure 9)**. The event initiating in BFB sequence is a DNA double-strand break. Before cell division the uncapped chromosome

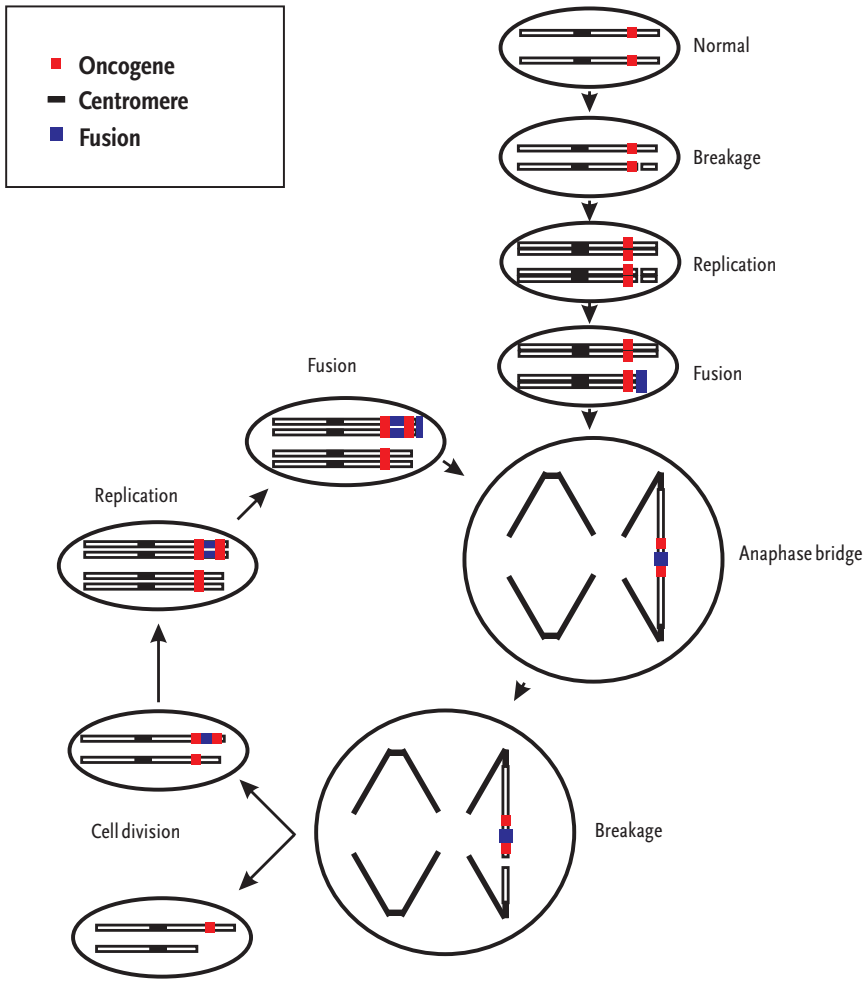


Figure 9
Breakage-Fusion-Bridge model of DNA amplification
Figure produced after Schwab, 1998 (Schwab, 1998)

replicates. BFB sequence progresses if fusion of the uncapped sister chromosomes occur before chromosome segregation. During mitosis the two centromeres of the dicentric fusion chromosome are drawn to the opposite poles of the mitotic spindle forming an anaphase bridge. DNA double-strand break in the new site during mitosis inflicts inverted duplication and loss of the genomic region between the two breakpoints in the daughter cells. The BFB cycle proceeds with every cell division, and the duplicated DNA region proliferates until the open chromosome ends are sealed. The steps of the BFB sequence, DNA double-strands break due to telomere loss (Murnane and Sabatier, 2004) and chromosomal breakage (Coquelle et al., 1997), sister chromatin fusion and breaking of the anaphase bridge (Shimizu et al., 2005) as well as ladder-

like structure of inverted repeats (Toledo et al., 1992), have been proven to exist *in vivo*.

The disintegration of circular extra-chromosomal chromatin bodies (double minutes and episomes) from chromosome, unequal segregation in cell division and selection in tumor tissue is another mechanism which results in DNA copy number amplification (Graux et al., 2004; Hahn, 1993). Extra-chomosomal amplification most likely occurs by unequal segregation in cell division and selection in tumor tissue, since double minutes and episomes rarely contain replication origins or centrosome that would enable propagation by replication and mitotic segregation. Extra-chromosomal DNA elements are thought to rise either before or after replication. Chromosomal deletion occurs if a DNA segment loops out from the chromosome before replication and chromosome ends are joined. There is evidence that supports the model that extra-chromosomal DNA is looped out from the genome, since original genomic architecture has been shown to remain together in the *MYCN* amplification (Schneider et al., 1992) and excised double minute (Toledo et al., 1993). It has also been shown that episomes (Graux et al., 2004) may carry DNA that is deleted in the original genomic region. If double minute excision happens after replication, the excised DNA segment stays also in the chromosome. There are two models for postreplicative multiplication of double minutes (Vogt et al., 2004) **(Figure 10A and B)**. Double minutes may unfold via breakage of the replication bubbles and multiply due to unequal segregation during mitosis. Furthermore, double minutes may contain inverted, repeated sequences suggesting that double minute formation may happen by HSR breakdown and the circularization of the derivative DNA fragments (Fakharzadeh et al., 1993; Singer et al.,

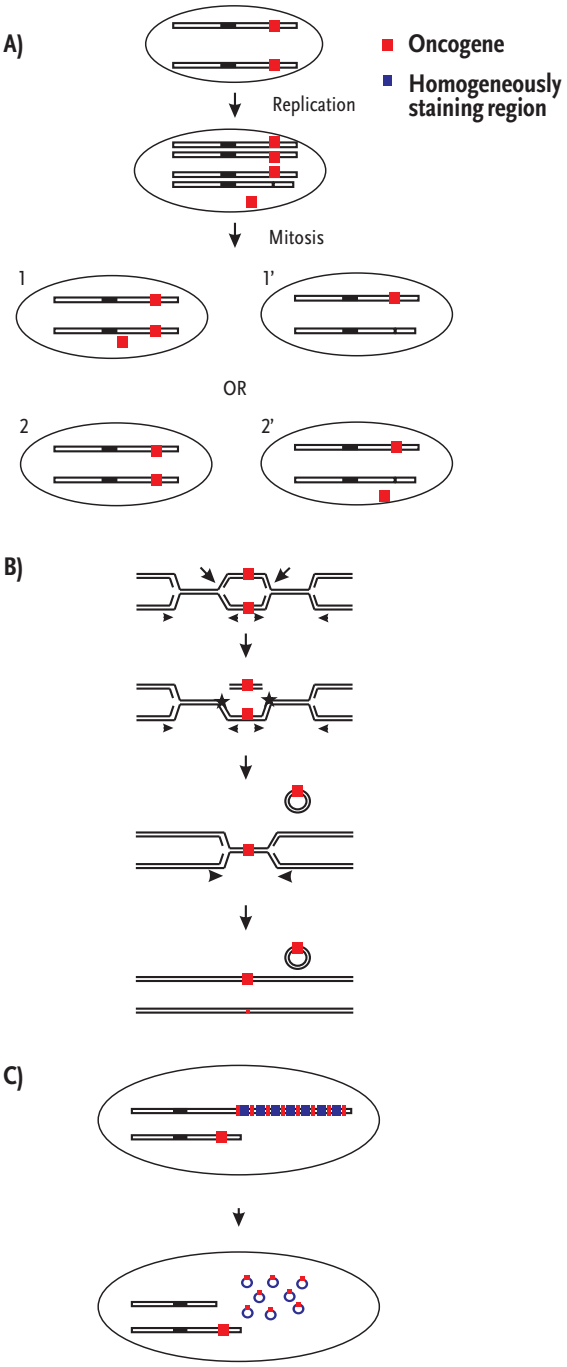


Figure 10
Models of double minute formation
A) Segregation after replication, **B)** Breakage of the replication bubles and **C)** HSR breakdown. Figure produced after Vogt et al., 2004 (Vogt et al., 2004).

2000) **(Figure 10C)**. In proportion, DNA double-strand breaks may induce genomic relocation of double minutes and episomes resulting in formation of HSRs or distributed insertions (Coquelle et al., 1998). Even though current data supports the excision-segregation-selection models of extrachromosomal amplification, also episomal proliferation (autonomous plasmid-like replication of sub-microscopic circular molecules) (Von Hoff et al., 1988) and recombination to form double minutes that reinsert into the chromosomes (Von Hoff et al., 1990) have been proposed to result in DNA amplification.

Models of DNA amplification mechanisms, BFB sequence and excision of extrachromosomal DNA segments, stipulate that two independent DNA double-strand breaks that flank the amplified region are required to occur in order to make DNA amplification possible. There are numerous agents and processes that have been shown to induce DNA double-strand breaks. Tobacco, ethanol and caffeine contain clastogenic chemicals that cause chromosomal breaks (Ban et al., 1995; Kuwano and Kajii, 1987; Rao et al., 1988). Asbestos is a group of silicate fiber, which is associated with the risk of developing lung cancer. Asbestos fibers block cell division, puncture into the nucleus and physically irritate the chromatin, suggesting a mechanism for inducing chromosome breaks (Jensen et al., 1996). Exposure to radiation is an established extrinsic cause of chromosomal breaks (Ban et al., 1995). The *Vpr* gene encoded protein induces chromosomal breaks leading to amplifications in patients infected with human immunodeficiency virus (Shimura et al., 1999). Folate and oxygen deficiencies (Blount et al., 1997; Coquelle et al., 1998) and release of oxygen-free radicals (Marnett, 2000) are intrinsic cellular processes that have been associated with the DNA damage.

Human genome is not consistently durable and resistance to DNA breaks varies between different regions. Specific DNA sequence features have been proposed to increase damage susceptibility. Fragile sites are genomic regions, which are damage-prone when cells are treated with chemicals that interfere with replication (Schwartz et al., 2006). There are 120 fragile sites in the human genome according to the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/). Eighty-eight fragile sites are classified as common, since they are found in all individuals, and 31 fragile sites, which are found in less than 5 % of the population, as rare. Damage susceptibility of the fragile sites varies and it has been shown that 15 % of all the common fragile sites contained 75 % of the DNA breaks (Glover et al., 1984). Fragile sites harbor large genes. *WWOX* and *FHIT* genes are 1.1 Mb and 0.8 Mb in length and located in two most damage-prone common fragile sites (FRA16D in 16q23.1 and FRA3B in 3p14.2) (Iliopoulos et al., 2006). In addition, fragile sites in general are enriched with extremely large genes (Smith et al., 2006). Chromosomal fragility induced breaks have been proposed to initiate the amplification pathway (Coquelle et al., 1997). Specifically, the common fragile sites were shown to coincide with amplification boundaries in a cell line experiment (Hellman et al., 2002). Furthermore, *MET* oncogene amplifications in six primary esophageal adenocarcinomas contained at least one break point that was located within a fragile site (Miller et al., 2006).

According to the central dogma, DNA information, including copy number, is transferred to RNA. The information encoded in the DNA copy number does not transfer into RNA quantity in a strict proportion because transcription is in many ways regulated. However, increased gene copy number generally induces elevated gene expression but the impact of gene amplification

on gene expression differs between cancer types, and it also depends on the structure of the amplicon. It has been estimated that a 2-fold change in DNA copy number causes a 1.5-fold difference in the mRNA level (Pollack et al., 2002). In studies of primary breast tumors (Hyman et al., 2002; Pollack et al., 2002) 44% to 62% of highly amplified genes were found to be overexpressed. In contrast, only 19.3% of the amplified genes in prostate cancer cell lines were shown to be overexpressed (Wolf et al., 2004).

Oncogene amplification promotes carcinogenesis as an amplified *ERBB2* gene was found to induce abnormal development of mammary glands and eventually initiate the tumorigenesis of the breast in a mouse model (Andrechek et al., 2000). According to a very conservative estimate, there are over 40 oncogenes that have been identified to be activated by copy number amplification in human cancers (causes increased gene expression or specifically associates with clinical outcome or tumor biology) (Myllykangas et al., 2007). Most studied amplified genes in human cancers include *AKT2* in ovarian cancer, *ERBB2* in breast and ovarian tumors, *MYCL1* in small cell lung cancer, *MYCN* in neuroblastoma, *REL* in Hodgkin Lymphoma, *EGFR* in glioma and non-small cell lung cancer and *MYC* in various cancers (Futreal et al., 2004).

Drug target gene amplifications frequently induce resistance to drug therapy as asserted in laboratory experiments using cell lines as well as in clinical studies (Goker et al., 1995; Johnston et al., 1983; Shannon, 2002; Wahl et al., 1979). The *CAD* gene amplification is a common mechanism in conducting resistance to N-(phosphonacetyl)-L-aspartate mediated genotoxicity (Wahl et al., 1979). Methotrexate is an inhibitor of the DHFR enzyme that is used in treatment of various cancers. Methotrexate resistance has been shown to result from the *DHFR* gene amplification in hamster cell lines (Johnston et al., 1983) as well as in patients with acute lymphoplastic leukemia (Goker et al., 1995). Imatinib mesylate is a BCR-ABL fusion protein inhibitor, which has been successfully applied in the treatment of CML patients. Imatinib mesylate treatment resistance occurs in some of the patients due to *BCR-ABL* fusion gene amplification or mutations in the *ABL* gene (Shannon, 2002). Additionally, changes in metabolite, substrate and competitive enzyme fluxes may underlie some cases of Imatinib mesylate resistance (Boren et al., 2001; Boros et al., 2002).

Oncogene amplifications occur frequently in advanced cancers and in the majority of organs and they usually are associated with poor prognosis (Myllykangas et al., 2007). Amplifications are more common in solid tumors than hematologic malignancies (Myllykangas et al., 2007). Although the gain-of-function effect of gene amplifications is very a appealing target for modern small molecule inhibitors, only therapeutic agents targeted at *ERBB2* and *EGFR* amplifications are currently available (Myllykangas et al., 2007). The *EGFR* gene is frequently amplified in colorectal cancers and there is a monoclonal antibody (panitumab) that binds to EGFR and inhibits its function. The *RAS* oncogene encodes a signal transduction protein that is downstream of EGFR. Nevertheless, patients with mutations in the *RAS* gene do not respond to EGFR antibody treatment and, in fact, the latest reports indicate that patients with *RAS* mutations have worse prognosis using EGFR antibodies than with traditional therapies. ❦

Hypotheses and objectives of the research

The treatment of cancer ultimately requires a detailed knowledge about the molecular properties of tumors. The complexity and heterogeneity of gastric cancer suggests that the genome-wide approach is well justified to study its molecular properties. The genome-wide methods can be used to detect genomic changes in cancer and to link them with cellular functionality and cancer phenotype. Novel microarray technologies for measuring gene expression and copy number allow molecular-level dissection of the functional and structural genomic changes in cancer. It is possible to define the cancer biomarkers and genes associated with the cancer phenotype using a multi-level, high-throughput genomic profiling, clinically and biologically relevant background data and computational data mining. Gastric cancer is a well-suited model for studying cancer since it is fairly common, it has well-documented epidemiological and biological background and it manifests many characteristics that are common to human cancer. More explicitly, the association of gastric cancer with *H. pylori* infection allows the research of carcinogenesis and cancer initiation. The biological background of cancer can only be understood if different cancer types can simultaneously be studied. Exploring DNA copy number amplifications in a landscape of cancers allows the uncovering of biomolecular fundamentals behind cancer and the confluences between different cancer types. ❦

The scientific objectives of the thesis were to:

- Identify gene expression changes and gene regulatory networks in *H. pylori* infected cells (I and II)
- Identify putative target genes in gastric cancer (III)
- Discover biological fundamentals behind DNA copy number amplifications in cancer (IV and V)

Materials and Methods

A BRIEF OVERVIEW of the materials and methods is presented in the following sub-sections. More detailed technical descriptions can be found in the corresponding publications.

In vitro model of *H. pylori* infection (I and II)

Laboratory experiments (*in vitro*) are needed in order to study the specific properties in biological systems. Laboratory experiments allow the researchers to have control over different biological, physical, and chemical properties, and focus the measurements on the specific biological properties. *H. pylori* infection induced gene expression changes in human epithelial cells and they were studied *in vitro* using *H. pylori* strain NCTC 11637 and AGS cell line (I). *H. pylori* strain NCTC 11637 met the requirements of a pathogen in the experiment, since the virulence factors *cagA* and vacA were present (JIANG ET AL., 1996). AGS cell line (the American Type Culture Collection, Manassas, MA) was chosen to be the host in the *in vitro* experiments as its growth characteristics, monolayer, resembled the gastric epithelium. Moreover, AGS cell line contains less genomic changes compared to other stomach cancer cell lines. AGS cell line has previously been used as a model host in *H. pylori* studies. Therefore, it is possible to compare results with the previous data.

Clinical gastric cancer samples (III)

Experiments carried out in the laboratory are not equivalent to the natural phenomenon. That is why it is also important to study the original samples. Clinical gastric cancer samples were collected prospectively in the Department of Surgery (Helsinki University Central Hospital, Helsinki, Finland) (III). Ethical issues had to be met so that the privacy of the patients who participated in the study was

protected. The study was authorized by the Clinical Review Board of Helsinki University Central Hospital and the research project was reviewed and approved by the Ethical Committee of the Department of Medical Genetics. Each participant signed a letter of consent prior to entering the study. Gastric cancer samples were collected between 1999 and 2005 during surgical removal of the stomach (gastrectomy). Extra precautions were taken in the sample preparation, since downstream analysis using microarrays is very sensitive. Sample quality determines the quality of the measurement data. In order to preserve the integrity of the nucleic acids, samples were taken immediately after the gastrectomy. Cancer samples were taken from the tumor site and normal tissue was obtained as far away from the tumor site as possible. Subsequently, samples were fresh frozen in liquid nitrogen and stored in -80 °C in as to preserve DNA and RNA from degradation. The collection of stomach specimens included a total of 375 samples. In order to ensure the anonymity of the participating patients, generic sample codes were given to each collected tissue sample by the attending clinician and all personal data was eliminated when tissue samples were subjected to research.

Classification of gastric cancer histology (III)

The classification of gastric cancer was defined according to Laurén (Laurén, 1965). The tissue samples used in the histopathology were surgical specimens from gastric cancer patients (see previous section). The classification was performed by a professional pathologist in Helsinki University Central Hospital (HUSLAB). The histopathological classification was performed using the frozen section preparations from the samples. The main stains used in the classification experiments were HE and Alcian blue-PAS. The patient records at the department of pathology were used to collect the clinical data.

Microarray experiments (I, III)

Nucleic acid extraction (I, III)

Total RNA was extracted from infected and non-infected cell lines using the Qiagen RNeasy method (I). DNA was extracted from the homogenized clinical gastric samples using the DNeasy tissue extraction kit (Qiagen) and total RNA was extracted using the RNeasy midi kit (III). The same homogenates were used in RNA and DNA extraction. The concentrations of DNA and RNA samples were measured using Eppendorf Biophotometer (Eppendorf AG, Hamburg, Germany) and the nucleic acid quality was measured using gel electrophoresis and Agilent's 2100 Bioanalyzer (Agilent Technologies) (I and III).

Hybridization (I, III)

The gene expression profile of *H. pylori* NCTC 11637 infected AGS cell line was measured using custom-printed microarray containing 12,000 cDNA probes (I). For each hybridization, 50 µg of the total RNA was extracted from *H. pylori* infected cell lines, and non-infected cell lines were labeled with fluorescent dyes using reverse transcription enzyme. The labeled cDNAs were hybridized on a cDNA microarray as previously described (Mousses et al., 2000).

The integration of gene copy number and the expression of microarray data is also required to understand the effects of gene regulation and transcription in the manifestation of genomic copy number changes. Forty-six gastric tissue samples were selected based on the sample quality and were analyzed using array comparative genomic hybridization (aCGH) and gene expression microarrays (III). The gene expression in 46 gastric tissues (8 normal and 38 cancer samples) was studied using a commercial microarray kit with 44,000 oligonucleotide probes. Twenty µg of total RNA was used in

hybridization. An RNA pool of 10 non-gastric cancer cell lines was used as a reference. Using a universal reference that contains a genome-wide coverage of transcribed genes allows the measuring of more genes. If the gene that is expressed in the test sample is not expressed in the reference sample, it is not possible to get a result, since ratio cannot be calculated if the reference is zero. The labeling of the test and reference RNAs, hybridizations and wa shings were carried out according to the manufacturer's instructions.

For performing aCGH, DNA samples were labeled as previously described (Hyman et al., 2002; Wolf et al., 2004). Pooled DNA extracted from the gender-matched healthy individuals' blood's buffy coat fraction obtained from the Finnish Red Cross was used as a reference. Twenty µg of genomic DNA from test sample and reference sample were used in the hybridization experiment. DNAs were labeled using fluorescent dyes and reverse transcription enzyme. The manufacturer's instructions were followed in the hybridization and washing of the slides.

Microarray data analysis (I, III)

Microarray data preprocessing (I, III)

After the hybridization experiments, microarray slides were scanned using a confocal laser scanner and the fluorescence intensities of the DNA spots were measured. Two scans were performed for each microarray slide to produce fluorescence intensity images for both, test and reference channels. Image analysis software is required in order to interpret scanned images from TIFF-format into numeric values as well as to obtain a text file of raw intensity values for test (Cy5) and reference (Cy3) channels. The image analysis was performed using DEARRAY software (Chen et al., 1997) (I) or Feature Extraction software 8.1 (Agilent Technologies) (III).

Data from different measurements and arrays had to be normalized to be able to compare the different data from the different samples. All normalizations were done using GeneSpring software. Lowess normalization was applied on the data set using GeneSpring 5.0 data analysis software (Silicon Genetics, Redwood City, CA) (I). aCGH and gene expression microarray normalizations were carried out using the Feature Extraction data import plug-in with default parameters for 2-color microarrays (Agilent Technologies) (III). Outlier features were disregarded in the normalization.

Filtering the results by quality parameters is required to make certain that all the data that enters the downstream analysis is of good quality. Intensity data were imported into the GeneSpring data analysis software (Agilent Technologies). Data were filtered by requiring the measurement quality score of the DEARRAY software to be more than 0.5 (Chen et al., 2002) (I). The genes that passed the quality control had spot fluorescence intensity over two times higher than the variance of the background, the intensity of the test sample was over 200, and the area of the spot was over 25. Microarray results in publication III were quality filtered using outlier and control flags determined by the Feature Extraction Software.

In order to be able to interpret the results from the microarray experiments it is important to know exactly what the arrayed probes are measuring. One efficient and stable technique is to match the probe sequences in the human genome sequence and extract gene annotations through the genomic databases. In publication III, the DNA sequence information of the probes was provided by the manufacturers of the microarrays that were used in the study. The corresponding genomic map positions of the microarray probes were retrieved using MegaBlast analysis against the human genome sequence. Each microarray probe was assigned with an Ensembl gene identi-

fier according to the map positions of the probe sequences. The gene identifications of Ensembl were applied to collect gene annotation data.

Identifying H. pylori infection regulated genes (I)

The genes that were regulated in AGS cells by co-culture experiments of *H. pylori* stimulation were identified by filtering data in different time points (I). A cut-off of two-fold change in gene expression ratios was used to select genes regulated during the *H. pylori* stimulation time series. In addition to this, t-test ($p<0.05$) was used to select differentially expressed genes in individual time points. The accumulation of *H. pylori* regulated genes in chromosomal regions that contain frequent changes in gastric cancer was assessed using a statistical hypothesis testing. P-value correction (e.g., False Discovery Rate) was not used when determining statistical significance in the analysis of the result genes since the identified genes were used for hypothesis generation and the individual findings were further validated using alternative methods. On the other hand, downstream-analysis was carried out using the whole set of genes.

Identifying gene copy number alterations in gastric cancer samples (III)

In microarray-based comparative genomic hybridization, the copy number of the gene is quantified and reported by a relative and continuous scale. Microarray data is noisy and when thousands of genes are measured simultaneously, individual measurements can vary unintentionally. Consequently, it is difficult to determine gene copy number changes based on a single measurement. In aCGH, measurements are not independent but interconnected by genomic location. As a result the measurements of the neighboring genomic positions have a higher probability of representing the same amount of genomic DNA. The dependency between probes representing close-by genomic elements can be taken into account when the genomic regions of copy number changes are determined. The CGH Explorer software (Lingjaerde et al., 2005) was applied to identify DNA copy number changes in 38 gastric tumors (III). Analysis of copy errors algorithm and false discovery rate of 0.002 were used in the analysis. Copy number gains and losses were determined for each tumor sample and the frequency profile across the genome was reported.

Classifying gastric cancer samples based on gene copy number aberrations (III)

The classification of samples is an elemental problem in clinical practice. Proper classification and diagnosis are the foundation of any effective treatment, as it is the informative follow-up of the disease. General concept in diagnostics is to determine, based on some continuous measurement value, whether a patient has a given condition or not. For example, what bodily temperature is considered as fever or does a patient have a specific gastric cancer sub-type based on the genomic properties of the tumor. There are four different outcomes of a two-class classification problem, such as intestinal or diffuse gastric cancer sub-type. True outcomes are achieved, when prediction and actual value are the same. This can be the case when an actual intestinal type sample is predicted to be intestinal or if a diffuse type is predicted to be diffuse. Conversely, false outcomes ensue, if an actual value is different from the predicted value, when intestinal type is predicted to be diffuse type and vice versa.

Receiver operating characteristics (ROC) analysis (Swets, 1998) was used to assess the values of gene copy number changes as classifiers of intestinal (n=25) or diffuse (n=13) type gastric tumors and cancers located in corpus (n=19) or antrum (n=19) (III).

The gains and losses were treated separately in the analysis. ROC curve can be drawn by calculating the true and false positive rates. The area under the ROC curve was calculated and these values were used to estimate the diagnostic value of a given gene.

ROC analysis identified gene groups that were too large for diagnostics use, since in clinical applications, patterns of no more than a few genes are sought after. Moreover, identifying a small subset of classifier genes would be more informative in determining the biological significance of the genomic changes that determine the phenotypes of different subtypes. Forward selection algorithm and a Naïve Bayes classifier (Duda et al., 2001) were applied to identify the patterns of genes that would be ideal when classifying gastric cancer patients based on the tumor location or subtype. Ten-fold cross-validation procedure was performed and repeated 50 times to ensure the stability of the identified patterns. Ten-fold cross-validation was chosen as it is optimal for such a small sample size. The number of genes in each pattern was determined by minimizing the validation error of 500 repetitions with different subsets of data in the framework of maximum likelihood. The significance of the identified patterns was assessed by comparing them to randomly selected variables. Ten thousand Naïve Bayes classifers were trained with randomly selected variables and the performance of the true classifier was evaluated using a statistical hypothesis test.

Integration of gene expression and copy number data from paired gastric cancer samples (III)

DNA copy number change does not necessarily mean that all the genes inside the changed region would be implicated in cancer. It is more likely that there is a subset of genes within the changed region that contributes to the manifestation of the cancerous phenotype. Integrating the knowledge of gene expression intensity with the copy number is a powerful tool for separating the biologically relevant driving genes from the by-standers. Paired gene expression and copy number data from 38 gastric cancer tissues were integrated in order to identify genes that were differentially expressed due to copy number aberration according to the method reported by Hautaniemi et al. (Hautaniemi et al., 2004) (III). Array CGH and gene expression microarray probes were combined using the genomic map positions of the arrayed probes. Gains and losses were treated separately in the analysis. The copy number aberration labels were used to classify samples into two groups. Subsequently the mean gene expression levels were calculated for each group. Signal-to-noise statistics and 10,000 random permutations of the label vectors were used to assess the significance of the difference in the mean gene expression intensities between the compared groups.

Validation of the microarray results (I, III)

Gene discovery approach using microarrays is often disturbed by noise in the data and the candidate genes need to be further examined in order to validate the results and estimate the relevance of the findings. Gene expression changes in AGS cells caused by *H. pylori* stimulation were verified using real-time PCR and target genes in gastric cancer were validated in protein level by immunohistochemistry and tissue microarray.

Quantitative real-time polymerase chain reaction analysis (I)

Quantitative real-time polymerase chain reaction (PCR) is a technique that can be used to detect and quantify the amount of a specific DNA sequence in a sample. The procedure follows basic PCR but the accumulation of a PCR product is measured after each amplifi-

cation cycle. When RNA is used as a template it has to be reversely transcribed into cDNA before quantitative PCR reaction. Quantitative real-time PCR was applied to validate the expression of ten genes regulated by *H. pylori* infection (I). Gene expression was measured in two time points and the untreated cell line was used as a control. For each analyzed gene a specific pair of primers were designed and synthesized. Quantitative real-time PCR was run using Light Cycler thermal cycler and SYBR green detection system. Duplicate PCR reactions were executed for each gene. A negative control accompanied each run. Standard curves were obtained by performing serial dilutions of beta-globulin gene. The *ACTB* gene (beta actin) was used as an internal control and for the normalization of RNA quantity and quality differences in all samples.

Immunohistochemistry using tissue microarray (III)

Immunohistochemistry is a laboratory technique, which utilizes antibodies to detect specific antigens in tissues. Antibodies (immunoglobulins) are proteins that are produced by cells of the immune system to identify and neutralize foreign agents. Each immunoglobulin binds to a specific antigen. Physiological antigens are often biological components that are specific to the invasion of bacteria and viruses, but specific antibodies against various antigens can be produced. Tissue microarrays consist of hundreds of separate tissue specimens punctured into a paraffin block. Using immunohistochemistry on tissue microarray enables simultaneous analysis of specific antigens in multiple samples. The specific antibodies that attach to ERBB2 and MUC-1 proteins were used on gastric cancer tissue microarray of 78 specimens to validate cancer biomarker discoveries made using integrated gene copy number and expression analysis of the clinical gastric cancer samples (III). The tissue microarray used in the study was independent of the samples that were used in the microarray profiling.

Analysis of *H. pylori* infection regulated transcription factors and signaling pathways in AGS cells (II)

Rapid changes in the gene expression of AGS cells after *H. pylori* infection suggested that the host responses are controlled by signal transduction pathways that regulate the activity of specific transcription factors. Transcription factors are proteins that bind to specific DNA sequence after being activated by specific signals and regulate, either by increasing or decreasing, the expression of specific target genes. Transcription factor activation is a rapid mechanism for cells to respond to environmental stimuli by accordingly changing their gene expression. Consequently, the hypothesis was that using the gene expression data from *H. pylori* experiments and a bioinformatics approach, it would be possible to reverse engineer the gene regulatory networks activated after *H. pylori* infection. The transcription factors and signal transduction pathways that were associated with *H. pylori* infection were charted by making use of phylogenetic footprinting and transcription factor binding site mapping on the regulated genes´ promoter regions (II). More reliable conclusions about the regulatory backgrounds of the gene expression changes were allowed to be made by studying the entire *H. pylori* transcriptome.

Computational analysis of H. pylori infection regulated signal transduction pathways in AGS cells (II)

H. pylori target genes (n=200) in AGS cells were previously identified using an *in vitro* model and gene expression microarrays (I). Two kilobases of DNA upstream from the predicted transcription start site were

extracted in order to collect the regulatory DNA elements, promoters, of the *H. pylori* infection target genes. Similarly, the putative promoter regions were collected from 400 randomly picked reference genes. Orthologous mouse genes were selected for *H. pylori* target and reference genes and their promoter regions were extracted using Ensembl database (Birney et al., 2004). Phylogenetic footprinting was applied to identify the conserved regions in the paired human-mouse promoter sequences (Lenhard et al., 2003). The DNA sequences that are evolutionary conserved are likely to have function and, thus, contain regulatory DNA elements (Lenhard et al., 2003). Transcription factor binding sites were mapped to the promoter sites that were evolutionary conserved using the application on Consite website (Lenhard et al., 2003). The transcription factor binding sites were collected from the JASPAR database (Sandelin et al., 2004). The frequencies of transcription factor binding sites in *H. pylori* infection regulated transcriptome and reference gene set were counted and the statistical hypothesis testing was applied in order to test the enrichment of the binding sites of specific transcription factors. BioCarta pathway database (http://www.biocarta.com/genes/index.asp) and database of Santa Cruz Biotechnology, Inc. (Santa Cruz, Ca) were used to collect annotation data for transcription factors.

Validation of NF-κB transcription factor activation after H. pylori stimulation of AGS cells using electrophoretic mobility shift assay (II)

An electrophoretic mobility shift assay (EMSA) is a laboratory technique that is employed when studying protein-DNA interactions. Proteins and radio-labeled DNA are mixed, and radioactivity is detected using gel electrophoresis separation. Specific antibodies are used to identify proteins that bind to radio-labeled DNA. EMSA technique was used to identify binding of NF-κB transcription factor to its target DNA sequence after *H. pylori* stimulation of AGS cells. A similar *in vitro* model was utilized as previously described (I). Upon activation, transcription factors translocate from cytosol to the nucleus, where they bind to a specific DNA sequence. The nuclear proteins of *H. pylori* infected cells were extracted in order to study the transcription factor activation of NF-κB. Nuclear protein extracts were mixed with radio-labeled target DNA sequence of NF-κB. Shift assays and specific antibodies against p50 and p65 subu nits were used to determine NF-κB transcription factor complex activation.

In silico analysis of DNA copy number amplifications in human cancers (IV, V)

The *in silico* analysis of DNA copy number amplifications in human cancer presented in this thesis can be classified as applied computational biology. Various computational tasks were executed in order to discover novel information from a data collection of DNA copy number amplifications in a cross-section of human cancers.

Collection of DNA copy number amplifications in human cancer (IV, V)

DNA copy number amplifications were obtained from a publicly available data collection (Knuutila et al., 2000) (IV and V). The data was curated from 838 original peer-reviewed research articles of chromosomal comparative genomic hybridization studies in cancer that were published between 1992 and 2002. The data collection contained 23 284 cases of which 4 590 cases had amplifications. The amplifications were mapped to chromosome sub-band resolution (n=393) (Shaffer and Tommerup, 2005). The amplifications were presented for each case by binary vectors of amplifi-

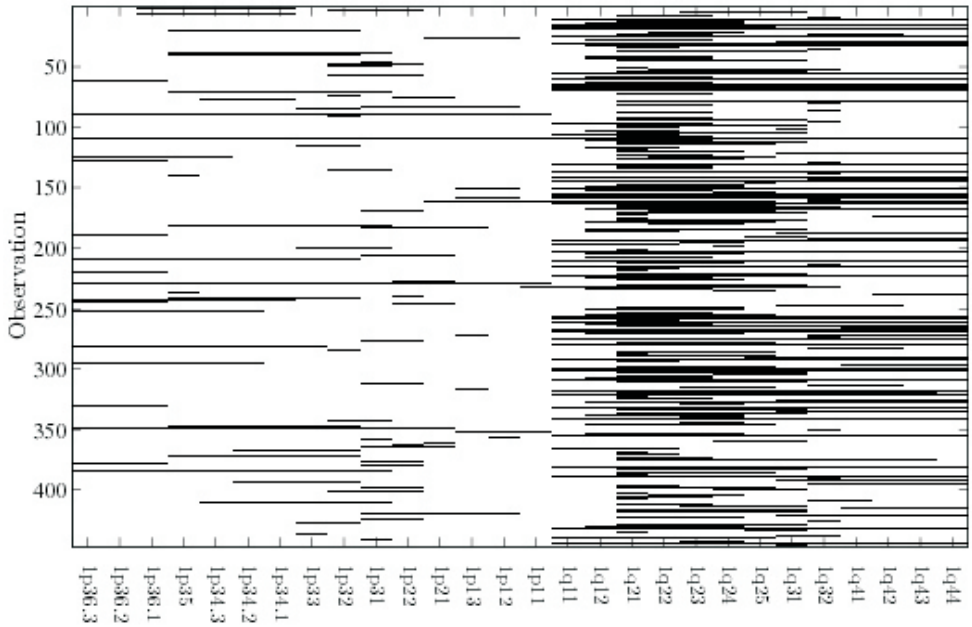


Figure 11
DNA copy number amplifications in chromosome 1
Original data from over 400 cases with amplifications that were collected from the literature (IV). Cancer cases are in random order.

cation observations, where zero denoted a non-amplified chromosome band and one amplified chromosome band. Data matrix was assembled by piling data vectors.

DNA copy number amplification profiling (IV)
The cancer cases were classified according to the World Health Organization guidelines (Kleihues and Sobin, 2000–2006) and neoplasm groups (n=73) were formed in relation to the collected classification data (IV). The genome-wide amplification frequency profiles were defined for each neoplasm type. A hierarchical clustering was applied to evaluate the similarities of amplification frequency profiles between the different neoplasms. Average amplification frequencies for each neoplasm group were used in the clustering.

Identification of amplification hot spots in the human genome (IV)
Computational methods can be very useful when extracting meaningful data from large data sources that are often affected by noise. The database of DNA copy number amplification is somewhat affected by noise due to inconsistencies in the literature and limitations in the traditional CGH method. The independent component analysis (ICA) (Hyvärinen et al., 2001) is a computational method that aims at finding statistically independent components in the data. The independent components are the multivariate factors that are composed of multiple signal sources. The components are

formed by maximizing their statistical independence. The ICA method was applied in order to identify preferential amplification sites in the human genome as it was presented earlier by Himberg et al. (Himberg et al., 2004) (IV). For an optimal performance, it is important to reduce noise by lowering the dimensionality in the data. The principal component analysis (PCA) (Duda et al., 2001) was used to reduce noise and to select the 60 components with the largest eigenvalues. After PCA whitening, the ICA analysis was performed and, the 30 most stable components in the data were selected and sorted using the stability indexes. The identified ICA components represented the regions in the human genome that are preferentially amplified.

Modeling and clustering DNA copy number amplifications (V)

In science, models are often used to explore new ideas and concepts and to disseminate information. The models are often helpful when it is difficult to comprehend the properties of the studied phenomenon, or if it is difficult to circulate the findings among peers. In traditional molecular biology, where single isolated events were studied, it was possible to present the measurement data or experiment outcomes in raw format, as simple tables or images, which were comprehensible and informative as a basis of interpretation. Modern genomics applications measure tens of thousands of data points simultaneously, and raw measurement data is presented in high-resolution images of arrayed hybridization signals and data files of tens of megabytes in size. It is evident that these array images and result tables no longer are comprehensible and useful for humans when interpreting the results. As a result, by definition, genome-wide microarray data are incomprehensible and impossible to distribute in its original format and data needs to be described with the appropriate visual or exploratory models. Fortunately, computers are fast and precise in handling large images and data files. It is therefore natural that computational methods have become instrumental in modeling genomic -scale data.

DNA copy number amplifications **(Figure 11)** were modeled in order to explore the biological fundamentals in the amplifications and form a classification based on amplifications (V). Machine learning techniques were applied to model inherent structures of the DNA copy number amplifications. DNA copy number amplifications were modeled for each chromosome separately. Finite mixture models of Bernoulli distributions were used in the modeling experiment. Expection-Maximization algorithm (Everitt and Hand, 1981) was applied in estimating the models and in selecting the complexity of the models in a chromosome (Tikka et al., 2007). DNA copy number amplification models can be used to determine sample groups with a similar amplification pattern. The clustering of the cancer cases based on the amplification models was executed as presented earlier.

Finite descriptions of the DNA copy number amplification models (V)

Even when probabilistic models are effective in describing the data, they may not be optimal when interpreting or distributing the findings. The models are represented as arrays of probability values that may be readily visualized. In order to fully interpret them, it is required for the models to be described in the original language of the chromosome bands. This ensures that further exploration becomes possible and relevant background data can be integrated in the models. Moreover, after translating the models into finite and comprehensible descriptions, the cancer specialists are able to relate to them. Maximal frequent itemset mining (Hollmén and Tikka, 2007) was

used in the identification of the finite and comprehensible descriptions of the identified DNA copy number amplification models (V). Amplification patterns depict the ranges of the amplified areas and the general structure of the amplicons.

Data mining of the DNA copy number amplifications (IV, V)

Data mining is a computational discipline for extracting meaningful knowledge from large data sets (Hand et al., 2001). In addition to the original quantitative measurements and observations, data miners use metadata (data about the data or background data) to improve their inferences. Meaningfulness of the information is usually measured using statistical methods and exploring the non-random dependencies in the data.

The amplification hot spots (IV) and amplification patterns (V) were fused with relevant cytogenetic and genomic metadata and statistical hypothesis testing methods were applied to enlighten the elemental factors contributing to the mechanisms and the selectivity of amplification. Chromosomal features and genomic annotations were collected from the literature and databases. In general, the frequencies of chromosomal and genomic features on amplification hot spots and amplification patterns were compared to a random reference. The significance of the differences in the frequencies was assessed using a statistical hypothesis test and permutation analysis.

An inherent structure in the amplification model-based clustering of cancer cases was compared to the clinicohistopathological classification of cancers (V). The classification data were collected from World Health Organization (Kleihues and Sobin, 2000 – 2006) and a data set of 101 attributes and 29 different parameters from 95 different neoplasms and 82 cancer types was compiled. A comparison of the proportion of an attribute in a specific cluster to a random reference was used to test the accumulation of specific attributes in individual clusters. A statistical hypothesis test was used to measure the significance of the difference in the proportions and random permutation analysis was applied to determine the empirical p-value.

Summary of the Methods (I – V)

The methods have been summarized in **Figure 12.** ❦

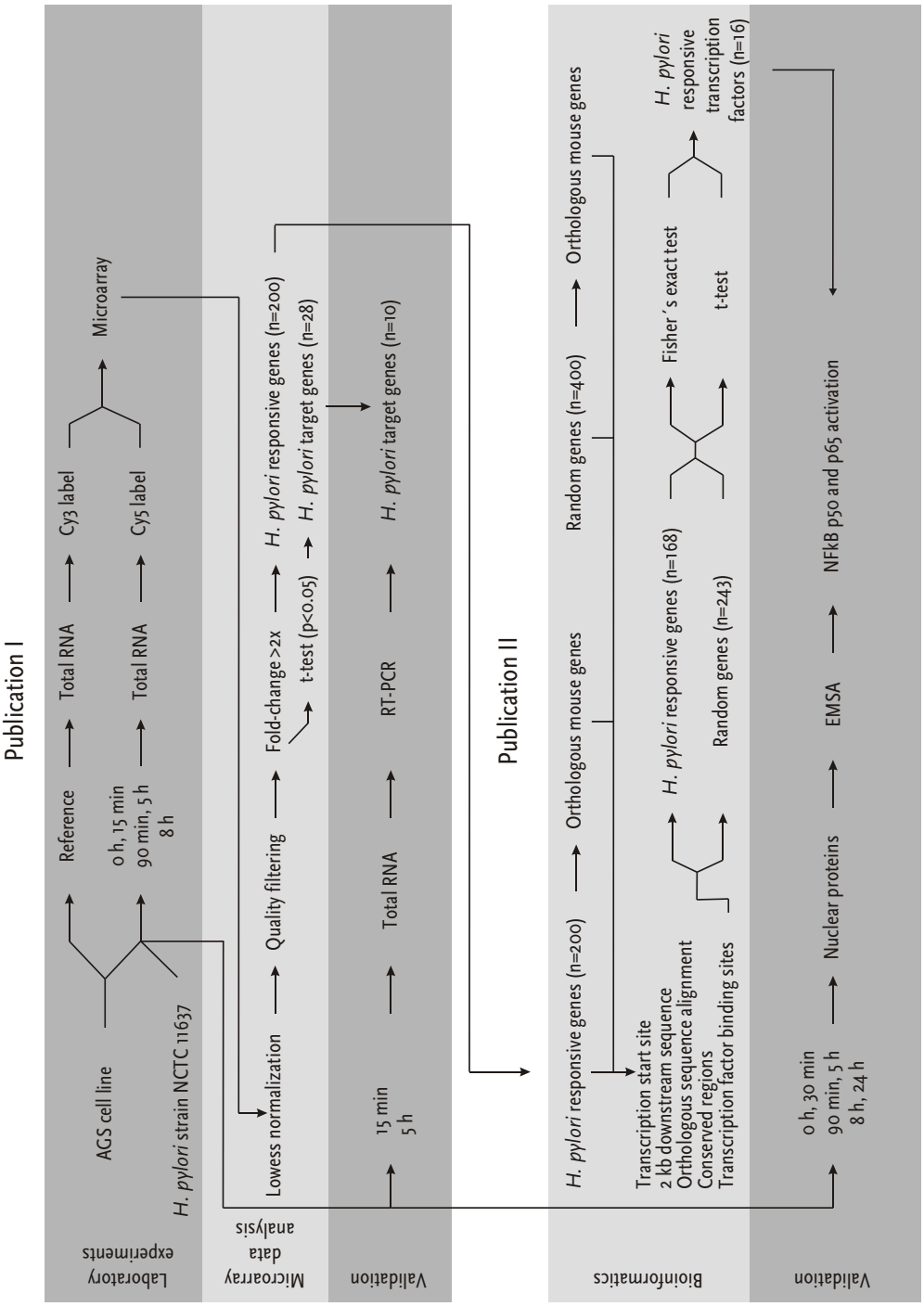


Figure 12
Summary of workflow

Figure 12
Summary of workflow

Figure 12
Summary of workflow

Results and Discussion

Gene expression changes and regulatory networks in *H. pylori* infected cells

The pathogenesis of the *H. pylori* partly comes across via the host responses, and gene expression changes in the host cells are likely to contribute to the promotion of the malignant phenotype. *In vitro* model and DNA microarrays were used to screen the target genes of *H. pylori* in the host. *H. pylori* infection induced differential expression of several target genes. According to the time series experiments and microarray analysis, *FOS* was the primary target in the infection and was up-regulated after 30 min of co-culture (Table 1 in I). Furthermore, *FOSL1* and *ATF3* were rapidly up-regulated in *H. pylori* stimulated AGS cells (Table 1 in I). Many genes encoding heat shock (e.g., *DNAJB1* and *HSPA1A*) and DNA damage response proteins (e.g., *GADD45A*) were up-regulated by *H. pylori* infection (I). Heat shock proteins are expressed when cells are exposed to stress. The transcription of *GADD45A* has been shown to increase when cells are exposed to DNA damaging agents in environmental stress, and to go to growth arrest (Jackman et al., 1994) and ATF3 protein level increases rapidly as a response to diverse stress conditions (Jiang et al., 2004). *ATF3* expression has been shown to be associated with *CagA* positivity (Zhang et al., 2007). The up-regulation of *ATF3* and *GADD45A* has been validated using a *H. pylori* infection model on the human gastric epithelial immortalized GES-1 cells (Liu et al., 2006). *H. pylori* regulated gene network functions in mechanisms that are activated in response to cellular damage and stress.

Gene expression changes induced by *H. pylori* infection are regulated by the action of specific transcription factors. The modulation of the transcription factor activity is the main mechanism by which *H. pylori* is able to hijack cellular functions and abuse the host. A gene expression profile data and a bioinformatics approach

were used to decipher the transcription factors that are activated after *H. pylori* stimulation. Computational analysis revealed 16 transcription factors that were significantly ($p<0.01$) enriched in *H. pylori* regulated genes' promoter regions (Table 1 in II). EMSA assay demonstrated that *H. pylori* stimulation of AGS cells resulted in an increased activation of NF-κB transcription factor and p50 and p65 subunits in the 3 h time point (Figure 1 in II). Previously, the activation of FOS (Meyer-ter-Vehn et al., 2000), NF-κB (p50 and p65 subunits) (Maeda et al., 2000), USF and CREB (Juttner et al., 2003) transcription factors have been shown to associate with *H. pylori* infection. Moreover, transcription factors that have not been linked with *H. pylori* infection were identified (ARNT, ELF1, EPAS1, E2F, MAX, MYC, MYC-MAX, SNAI1, TBP, and TCF3). *H. pylori* activated transcription factors function in cellular processes that are associated with carcinogenesis. NF-κB and FOS function in the production of the pro-inflammatory cytokines and in the activation of primary epithelial innate immunity response to pathogens (Eferl and Wagner, 2003; Karin et al., 2002) as well as the regulation of the anti-apoptotic and proliferative signaling (Karin et al., 2002). MYC is involved in cell-cycle progression, it promotes cell proliferation and is also associated with cancer initiation (Pelengaris et al., 2002). ARNT and CREB regulate cell response to hypoxia and promote angiogenesis (Harris, 2002). E2F controls cell-cycle and is associated with cell growth (Nevins, 2001). SNAI1 promotes tumorigenesis by repressing E-cadherin, increasing cell motility and by promoting invasive growth (Thiery, 2002). USF1/2 transcription factors control telomerase reverse transcriptase (TERT) activity, which is required for immortalization and continuous replication (Goueli and Janknecht, 2003). The transcription factors of *H. pylori* response are involved in cancer-related cellular mechanisms.

In eukaryotic cells, transcription activity is controlled by the signal transduction pathways that consist of proteins that are sequentially activated so as to transfer signals from outside the cell into the nucleus, where response is executed at the level of gene expression. Cells have many receptors that they use to monitor their surroundings. Some of these receptors are used to detect pathogens. For example, toll-like receptor 4 binds to lipopolysaccharide of *H. pylori* (Kawahara et al., 2001). In addition to the surveillance proteins of the host, pathogens have evolved proteins that mimic cellular signaling proteins and evoke phony cell responses. For instance, *H. pylori* secreted cagA toxin interacts with the host cell's signaling proteins (Atherton et al., 2001; Hatakeyama, 2004). Even when cross-talk between different signaling pathways might exist, both, pertinent physiological responses to bacteria as well as *H. pylori* secreted disturbance signals are likely to contribute to specific pathway activation in *H. pylori* infected cells. Due to the fact that *H. pylori* strains that carry bacterial toxins are more carcinogenic, there might be specific pathway targets for the toxins that are related to the progression of the malignancy. *H. pylori* activated transcription factors were used to chart the signal transduction pathways that were associated with the infection (II). *H. pylori* stimulation regulated transcription factors were shown to bind to specific target sequences and function as downstream targets of Ahr signal transduction pathway, MAPK signaling pathway, NF-κB signaling pathway, RB pathway, and Wnt signaling pathway (Table 1 in II). Additionally, using the information from Biocarta, a model explaining the gene regulatory network of *H. pylori* infection in a host cell could be drafted (Figure 2 in II). The activation of MAPK and NF-κB signal transduction pathways have been shown to occur after

H. pylori infection (Juttner et al., 2003; Kawahara et al., 2001; Maeda et al., 2000; Meyer-ter-Vehn et al., 2000; Mitsuno et al., 2001; Nakayama et al., 2004; Su et al., 2003). NF-κB signaling is most likely activated through lipopolysaccharide sensing by Toll-like receptor 4 (Kawahara et al., 2001), whereas cagA and VacA toxins are taught to activate MAPK signaling (Atherton et al., 2001; Hatakeyama, 2004). Apart from the known *H. pylori* regulated pathways, novel signal transduction pathways, namely Ahr, WNT and E2F, were identified. *H. pylori* pathogenecity probably acts through specific signaling pathways and the identification of cellular responses with malignant potential would serve as biomarkers for predicting the outcome of the infection.

H. pylori induced reactive changes in the host cell's gene expression are not explicitly cancer-related; rather, these changes affect cellular mechanisms that have the potential to lead to perturbations that cause cancer. *H. pylori* stimulation activated NF-κB transcription factor, which is a key regulator of inflammation and cancer promoting cellular mechanisms. Among others, *H. pylori* target genes and signal transduction pathways regulate cell-proliferation, apoptosis and cell motility. Increased cell turnover can lead to increased mutation rate and carcinogenesis. Similarly, failures in executing apoptosis in control may have cancer-promoting implications. Cell proliferation involves DNA replication, which has a low but significant error rate, and *H. pylori* infection, which favors cell proliferation, is indirectly mutagenic. In the view of the fact that *H. pylori* infection regulated pathways that are constitutively active in gastritis and operate critical cellular functions, the genes encoding for proteins involved in these activities are likely targets for mutations. It could be further postulated that those pathways that are hijacked in *H. pylori* infection and the genes that they control might be preferential targets for genomic changes in gastric cancer initiation. *H. pylori* infection regulated genes in AGS cells were enriched in the chromosomal regions that are frequently implicated in gastric cancer (Table 2 in I). The transcriptional activation of *H. pylori* infection target genes involves introducing changes in the chromatin structure. Histone coiling needs to be disassembled in order to get the transcription protein complex to access DNA. In open formation, the physical protection of chromatin reduces and DNA gets exposed to mutagenic agents. DNA in the open chromatin formation is more likely to be damaged than DNA in the inactive regions of the genome. All in all, *H. pylori* infection leads to activation of the cellular mechanisms that increase the probability of DNA damage, mutations and cancer promotion.

Putative gastric cancer target genes

Biomarkers are measurable properties that can be used to indicate a biological characteristic of the sample object. Genome-wide microarray technologies are excellent tools for screening the genome for anomalies that could indicate a malignancy state or a certain biological process that is involved in cancer. Microarrays were exploited to identify the putative biomarkers and the biologically relevant target genes in gastric cancer. Putative gastric cancer biomarker genes were identified by integrating the gene expression and the copy number microarray data (III). The rationale behind integrating the gene expression with the copy number is that those genes that are differentially expressed due to a change in the copy number are more likely to be involved in cancer promoting processes. The integration of gene expression and copy number measurements revealed genes that were up-regulated in association with copy number gain or down-regulated in association with copy number loss (Figure 2 and Table 1 in III).

The 17q12-q21 region contained many genes that were up-regulated in association with amplification (Figure 2 in III). Amplification and expression of *ERBB2* as well as of other genes in the 17q12-q21 region have previously been identified (Maqani et al., 2006; Varis et al., 2004). In addition, several novel gastric cancer candidate genes were identified. These novel putative gastric cancer biomarker genes provide a prioritized selection of candidate genes for further studies.

Gastric cancer is divided into two main histological subtypes, namely intestinal and diffuse, which have different clinical and biological features (Lauren, 1965). The information of the histology of the tumors was used in fusion with gene copy number profiling data in order to identify subtype specific changes. Intestinal type gastric cancer was characterized by amplification of *ERBB2* and gains at 20q13 and Xp (III). Previously, 17q12-q21 amplifications have been linked to intestinal type gastric cancer (Kokkola et al., 1997; Maqani et al., 2006; Vauhkonen et al., 2006). More specifically, the *ERBB2* gene was shown to be activated, in RNA (Figure 2 in III) and protein levels (Figure 3 in III), by an amplification in the intestinal type gastric cancer (III). ERBB2 is a transmembrane receptor tyrosine kinase, which functions in cellular signal transduction and regulates processes that are frequently perturbed in cancer. There is a new category of cancer therapies which target protein kinases. *ERBB2* amplification and overexpression have been proposed as a therapeutic biomarker in gastric cancer and Trastuzumab (ERBB2 tyrosine kinase domain binding monoclonal antibody) treatment was shown to inhibit tumor growth in gastric cancer cell lines and in a patient (Gong et al., 2004; Rebischung et al., 2005). According to amplification modeling analysis, the 17q12-q21 region, which contains the *ERBB2* gene, is amplified in gastric and esophageal adenocarcinomas (Figure 5A in V). 17q12-q21 amplification was also typical for cancers of epithelial origin, more specifically adenocarcinomas. These data suggest that the 17q12-21 amplifications are specific for epithelial type cancer of the gastroesophageal tract and amplification and over-expression of *ERBB2* could be used as a biomarker for intestinal type gastric cancer in the proximal stomach.

There were interconnections between *H. pylori* infection regulated genes and gastric cancer genome. *MUC1* was shown to be amplified and over-expressed in intestinal type gastric cancer (Figures 2 and 3 in III). ERBB2 has been shown to phosphorylate MUC1 (Li et al., 2001; Schroeder et al., 2001) and regulate its interaction with β-catenin (Yamamoto et al., 1997). In conjunction, *H. pylori* binds to MUC1 on the surface of gastric epithelium (Vinall et al., 2002) and activates β-catenin in gastric epithelial cells (Franco et al., 2005). β-catenin has been shown to enhance invasiveness and promote proliferation in gastric cancer cell lines (Lowy et al., 2006). Based on these observations, there are multiple mechanisms – gene copy number gain, ERBB2 kinase phosphorylation and *H. pylori* binding – that are capable of activating *MUC1*. Furthermore, *MUC1* might be involved in gastric carcinogenesis by promoting proliferation and invasion through interaction with β-catenin. *PPP1R1B* was one of the target genes at 17q21.1 that was up-regulated by a copy number gain (Table 1 and Figure 2 in III). *PPP1R1B* transcription was also up-regulated by *H. pylori* infection (Supplementary data in I). Because *H. pylori* infection and gene copy number gain are both potential activators of *PPP1R1B* that further strengthens its central role in gastric carcinogenesis. Depending on the phosphorylation pattern of specific amino acid residues, PPP1R1B functions as a kinase or phosphatase inhibitor (Bibb et al., 1999;

Greengard et al., 1999). Based on its function, *PPP1R1B* is a potent regulator of cellular signal transduction that could promote carcinogenic processes in gastric cancer. Confluences in genomic functions of gastric cancer and its premalignant condition, *H. pylori* infection, suggest that there is a connection between molecular pathways in the different steps of gastric carcinogenesis.

DNA copy number amplifications in cancer

DNA copy number amplifications were analyzed from a large set of cancers to elucidate the confluences and relationships between different cancer types. The clustering of neoplasms based on DNA copy number amplification frequency profiles showed that cancers with similar cell type or origin grouped together (Figures 1 and 2 in IV). For example, cancers of the gastrointestinal tract, esophageal, gastric and colorectal adenocarcinomas, formed a tight cluster. Correspondingly, sarcomas (mesenchymal origin) of different anatomical locations clustered together. Carcinomas, epithelial cancers, formed separate cluster with a tight subcluster of squamous cell carcinomas that originate from columnar epithelium. Clustering based on amplification frequencies suggests that the type of the cell-of-origin and/or tumor microenvironment determines the site specificity of DNA copy number amplifications.

Amplification profiling analysis was not able to link specific cancer types with specific amplifications. DNA copy number amplification modeling and data fusing of cancer classification information were applied in order to study the background of the amplification-based clustering of human cancers (V). DNA copy number amplifications were modeled using finite mixtures of the Bernoulli Distributions (Figure 1 in V). Amplification models were used to cluster cancer cases (Figure 2 in V). Classification attributes (Figure 3 in V) and etiological factors (Figure 4 in V) were collected. Specific cancer types and classifications were shown to be associated with specific amplification models (Figure 5 in V). For example, 17q12-21 amplification was associated with esophageal (Barret) and gastric adenocarcinomas (Figure 5A in V). In addition, 17q12-q21 amplicon was over-represented in cancers that were related with tobacco, obesity, diet, bacteria, inflammation, gastrointestinal tract, esophagus, stomach, adenocarcinoma and epithelial background. It is noteworthy that the 17q12-21 amplification was identified in microarray analysis of gastric cancer (III). Molecular pathophysiology of cancer can be dissected by cell lineage-specificity (Garraway and Sellers, 2006). Cell lineage models presume that specific properties of the precursor stem cells promote the evolution of molecular changes in cancer. During embryonic development, epithelial, hematological, mesenchymal and neuroepithelial cell-lineages emerge. The development of specific organs and tissues unfolds through the differentiation of the specified stem cells. The amplification spectrum in cancers seems to be determined by cell-lineage, seeing that specific amplifications were identified for different cell lineages of embryonic, tissue and organ development (IV and V). Site-specificity of amplifications in different cell lineages might reflect the properties of the cancer stem cells or of the selective microenvironment in the tissue-of-origin. Oncogenes were shown to frequently map to amplification target loci **(Figure 13)** (V).

In addition to the cell-lineage determination and the selection guided by oncogene attraction, genomic architecture is involved in the generation of amplifications. The structural outlines of amplicons depend on genomic properties that affect chromosomal durability and genomic integrity. The data mining of the amplification hot spots

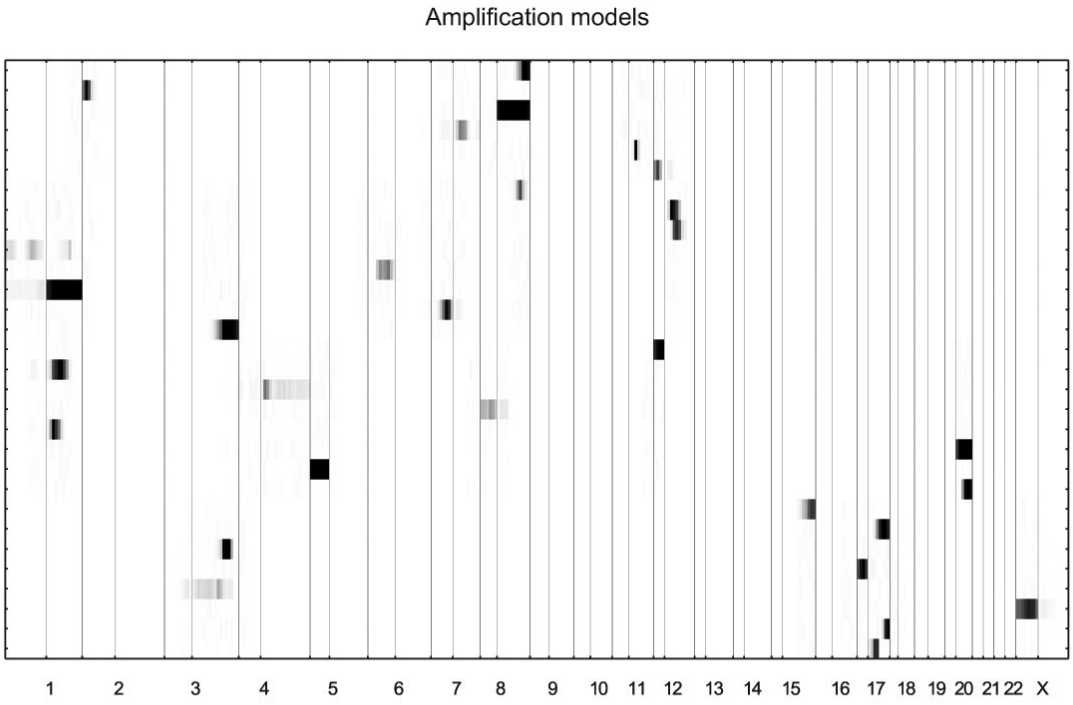
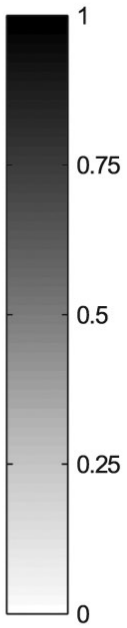


Figure 13
Amplification models and co-localized oncogenes
30 most frequently observed DNA copy number amplification patterns and co-localized cancer genes. Mode was extracted from the amplification models to depict the most probable element in the amplification model. Cancer genes that co-localize with the mode loci were collected from the CancerGenes database (Higgins et al., 2007).

(IV) and maximal frequent itemsets of the amplification models (amplification patterns) (V) revealed properties in the genomic architecture that were associated with mechanisms of DNA amplification. The mechanistic models of amplification predict that DNA double-strand breaks initiate the amplification pathway. The statistical evidence showed that there are specific labile chromosomal regions that are enriched in the margins of amplification patterns. Fragile sites, telomeres, light chromosome bands and centromeres were enriched at chromosomal sites that are often implicated in the breakpoints of amplicons. Large genes (in hundred kilobase range) were statistically significantly over-represented in the amplification hot spots. Correspondingly, fragile sites were preferentially co-localized with exceptionally large genes. Large genes may contribute to chromosomal fragility and promote amplicongenesis by increasing the probability of break-induction. One mechanism that has been proposed to inflict chromosomal breakage is the collision of transcription and replication machineries. In plasmid reconstruction experiments, it has been shown that DNA double-strand breaks

Mode	Candidate target genes
8q24.2	*MYC*
2p24	*MYCN*
8q11.2-8q24.2	Numerous
7q21	*CDK6, ERVWE1*
11q13	*MEN1, CCND1, NUMA1, GSTP1*
12p12	*ZNF384, KRAS*
8q23	*COX6C*
12q13	*ATF1, CDK2, DDIT3, GLI1, HOXC13, HOXC11, CDK4*
12q15	*HMGA2*
None	
6p12	*HSP90AB1*
1q21-1q42	Numerous
7p12	*EGFR, ZNFN1A1*
3q26.1-3q27	*MDS1, TERC, RPL22P1, PIK3CA, BCL6*
12p11.2-12p11.1	
1q24	*PRRX1, ABL2*
4q12	*KIT, PDGFRA, FIP1L1*
None	
1q21	*SDHC, ARNT, BCL9, MUC1, MLLT11, FCRL4, PRCC, TPM3, NTRK1*
20q11.2-20q13.2	*MAFB, TOP1, SS18L1, GNAS*
5p15.2	*FKSG14*
20q13.2	
15q25	*NTRK3*
17q23-17q24	*BCL5, PRKAR1A, AXIN2*
3q26..2-3q26.3	*PIK3CA*
17p11.2	*MAP2K4, SPECC1, FLCN*
None	
Xp11.4	*WAS*
17q25	*ASPSCR1, KIAA1618*
17q21	*BRCA1, RARA, ETV4, MLLT6, COL1A1*



occurred and amplification process was initiated by synchronized and site-directed transcription and replication (Shimizu et al., 2003). The transcription and replication of large genes considerably takes more time, which widens the window for an occasional collision between the two processes. Light chromosome bands contained multiple amplicon boundaries. Light chromosome bands are gene rich and usually in open conformation (Gilbert et al., 2004). The enrichment of amplification boundaries on light chromosome bands suggests that open chromatin would be more vulnerable and DNA damage prone than inactive closed chromatin. Telomeres were also among the chromosomal regions that were over-represented by amplification breakpoints. Telomere erosion in uncontrollably replicating cancer cells is a mechanism that can produce DNA double-strand breaks in the telomere regions. Non-random distribution of labile genomic features indicated that the human genome is not evenly resilient, and that specific sites are more prone to DNA double-strand breaks. ❦

Conclusions and future prospects

***H. pylori* activated the genes,** transcription factors and signal transduction pathways that were identified in studies I and II. The molecular targets of *H. pylori* infection were associated with the cellular functions that might be capable of promoting carcinogenesis of the stomach. These targets could be useful when predicting the outcome of the infection and the designing novel treatments. Intestinal and diffuse type gastric cancers showed distinct molecular genetic profiles and the integration of gene expression and copy number microarray data allowed the identification of genes that might be involved in gastric carcinogenesis and have clinical relevance (III). Even when some subtypes with concordant genomic changes were identified in gastric cancer, it is likely that most of the carcinomas of the stomach are unique diseases that do not share a specific genomic change. Gene amplifications were shown to be non-random genomic instabilities, and cell lineage, properties of precursor stem cells, tissue microenvironment and genomic map localization of specific oncogenes to define the site specificity of DNA amplifications, whereas labile genomic features defined the structures of the amplicons (IV and V). These conclusions suggest that the definition of genomic changes in cancer is based on the interplay between cancer cell and the tumor microenvironment. Categorizing cancers according to their underlying genetic properties, rather than looking at the appearance or behavior of the tumor, is likely to improve cancer treatment. Morphologically similar tumors may arise in the same anatomical location and still have entirely different genetic anomalies. Conversely, distinct genomic changes can cause cancers in different parts of the body as well as of diverse morphologies. Regardless of the site and behavior of the tumor, genetically coherent cancer subtypes

are likely to depend on the same molecular perturbations, and the understanding of the genetic underpinnings of cancers facilitates the development of the targeted treatments.

In the future, the identified putative biomarkers for gastric cancer and *H. pylori* infection related carcinogenic processes need to be validated in a lager set of samples using alternative approaches and their roles in gastric cancer should be explored using functional experiments. Moreover, there might be a concordance among gastric cancers in higher hierarchy of genomic functionality, namely in the level of biological processes and pathways, which could be revealed using a systems biology approach. In order to treat heterogeneous disease, such as gastric cancer, genomic properties of individual tumors need to be decoded. Novel, high-throughput, massively parallel sequencing technologies enable rapid and near genome-wide examination of the tumor DNA sequence, which is fundamental for the development of personalized treatments for gastric cancer. To fully understand the mechanisms and backgrounds of genomic changes, a large-scale datamining effort using DNA microarray or sequencing data should be executed. ❦

Acknowledgements

This work was carried out at the Department of Pathology in Haartman Institute and HUSLAB, University of Helsinki and Helsinki University Central Hospital during the years 2003–2008. I am grateful for the financial support received from the Academy of Finland (SYSBIO Research Program), the State Appropriations of Helsinki University Central Hospital, the Helsinki University grants, the Sigrid Jusélius Foundation and the Finnish Cancer Organizations. I wish to express my gratitude to all those who have contributed to the accomplishment of this study, and I especially acknowledge:

Professors **Veli-Pekka Lehto** and **Leif Andersson**, present and former Heads of the Department of Pathology, Haartman Institute, University of Helsinki, for their positive attitude towards my studies and for providing excellent research facilities. The former and present Heads of Department of Medical Genetics, Haartman Institute, University of Helsinki and Department of Pathology, HUSLAB, Helsinki University Central Hospital, are also acknowledged.

Professor **Sakari Knuutila**, the supervisor of this thesis, is highly appreciated for his caring and positive attitude, encouragement, and not only for providing resources but also inspiration, and for sharing his intellect and passion for cancer research and cycling. His ideas, such as the one of riding around the Bodom Lake, or going to work over the frozen sea, reflect a freedom of thoughts and a willingness to take detours – state-of-mind – that I have learned to embrace when facing scientific challenges. I respect his ability to lead by setting an example, rather than by having expectations and being enthusiastic and not overly keen on following the traditional ways. A novice researcher could not have hoped for a better mentor than him. Above all, I value his friendship and rich personality.

Professor **Pentti Sipponen** and Docent **Ari Ristimäki**, my thesis committee members, provided great help by overseeing my research activities, reviewing my work as well as the manuscript of this thesis, and by providing much appreciated critique and constructive comments over the years. I thank them for their positive attitude, warm support and ample attention. I am especially grateful to Professor Sipponen for being an official reviewer of this thesis, for the enlightening lessons given on gastric cancer and for providing wise comments that truly helped to improve my thesis. Research Professor **Matej Orešič**, another official reviewer of this thesis, is thanked for his all-round scientific expertise, which promoted an emergence of great revelations and new ideas when finishing this thesis. I was also fortunate to obtain his insight regarding the presentational clarity of the thesis.

I have had the privilege of having two talented researchers as additional, unofficial supervisors. Docent **Outi Monni** is acknowledged for mentoring and teaching DNA microarray experiments (I and III). As the senior author, she was essential in the unraveling of publication III. What is more, she has taught me in great lengths about scientific writing and conducting genomic-scale research, which I really appreciate. I would like to thank her for her encouragement and endorsement in the search for new research opportunities. **Jaakko Hollmén** is acknowledged for bringing top-of-the-line statistical knowledge for publications III, IV and V. Moreover, he was the senior author in publication V and did an excellent job in navigating that work. I am thankful for his outstanding teaching about probabilistic methods and statistical analysis, reviewing various research papers – especially this thesis – and for always giving such positive, constructive and almost Rainmanishly detailed feedback. His quick-witted humor and observant nature have made our collaboration an incredibly pleasant experience.

I am grateful to my co-authors for their commitment, scientific and intellectual inputs and for providing a comfortable and supportive work atmosphere. I would like to sincerely acknowledge their contributions to the completion of this thesis work: **Bálint Nagy** for carrying out real-time PCR experiments (I) and for collecting data (IV); Professor **Hilpi Rautelin** for her guidance and training in *H. pylori* culture experiments (I) and for her encouragement when I had my first international presentations; **Ville Veckman** for tutoring and assisting the EMSA experiments (II), the traditional publication lunches and friendship; **Juha Saharinen** for his collaboration in bioinformatics studies (II), as well as for the analytical discussions and opinions regarding both, professional and everyday matters; **Siina Junnila** for her excellent collaboration in gastric cancer study (III), good spirits in the lab, amusing coffee breaks and friendship; Professor **Pauli Puolakkainen** and Docent **Arto Kokkola** for providing gastric cancer samples, clinical data and clinicians' point of view for the research (III); **Reija Autio**, for integration analysis (III); **Ilari Scheinin** for microarray probe matching (III), fruitful scientific off-thesis collaborations, iPhone help and continuous bioinformatic support; **Tuula Kiviluoto** for providing gastric cancer specimens (III); Docent **Marja-Liisa Karjalainen-Lindsberg** for histopathological evaluation of the gastric preparations (III) and for teaching gastric histology; Johan Himberg for bioinformatic analyses (IV); Docent **Tom Böhling** for reviewing cancer classifications (IV and V) and for educating me regarding cancer pathology; and **Jarkko Tikka** for executing computational analyses (V).

There are a number of people without whom this thesis would not have been possible. **Pirjo Kosonen** is thanked for her practical assistance in *H. pylori* culturing and

Ulla Kiiski for the preparation of the gastric cancer slides. The personnel of the Molecular Pathology Laboratory, especially **Tiina Wirtanen**, **Leena Rossi**, **Paula Kvick**, **Tarja Salonen**, **Mikael Lindlöf** and **Kirsi Autio**, are thanked for their aid in the lab and good company at work, as well as in cycling and skiing trips, Christmas parties and other recreational events. I am in gratitude to **Pirjo Pennanen** for the language revision of individual manuscripts and linguistic assistance when drafting numerous reports, grant applications and various obscure writings. **Laura Villella-Tolonen** deserves my gratitude for the prompt language revision when finalizing the thesis and **Tomi Metsä-Heikkilä** for designing the layout of the thesis.

My colleagues, especially former and present CMG-group members, **Ioana Borze**, **Mohamed Guled**, **Kowan Jee**, **Sippy Kaur**, **Eeva Kettunen**, **Leo Lahti**, **Pamela Lindholm**, **Tuija Lundan**, **Tarja Nieminen**, **Tarja Niini**, **Shinsuke Ninomiya**, **Penny Nymark**, **Fabrizio Passador-Santos**, **Salla Ruosaari**, **Suvi Savola**, **Linda Siggberg**, **Anu Usvasalo**, **Hanna Vauhkonen**, and **Harriet Wikman** are thanked for their exciting scientific discussions, aid in the lab, creating an inspiring working atmosphere and pleasant company in different work trips. My gratitude goes to the people in the SYSBIO research consortium, especially to Professor **Samuel Kaski** for heading the program. I am grateful for **Jukka Vakkila** for inspiring lessons on cancer immunology and inflammation-related diseases. Special thanks go to **Perttu Salmenperä** for his good company and for engaging in interesting scientific conversations at lunch and coffee breaks as well as in Snowbird. **Massimiliano Gentile** is thanked for all those hours of patiently and straightforwardly teaching me how to use GeneSpring in microarray data analysis. In addition to that, our scientific side projects have been interesting, educational and rewarding. Most importantly, enjoyable lunches at the Mountain restaurant and those at-least-twice-a-week doses of Nepalese food have been important ingredients for this thesis. His thorough and wide knowledge about mountain bike and camera gear truly amazes me, and I'm glad that he has introduced me to the mountain biking grounds in the Keskuspuisto. I wish to thank him for the skiing photos in Snowbird and especially the one that covers this thesis. Thanks are due to "the Microbes": **Anu Jääskeläinen**, **Anna Katz**, **Tiina Partanen**, **Hanna Sinkko** and Ville Veckman for good times during studying and ever since. I greatly appreciate your friendship and all the peculiar traditions we have managed to maintain.

I wish to thank all my friends, particularly **Tommi Sneck**, **Ville** and **Kristiina Laakso**, **Tuomas** and **Monika Kivimäki**, **Lotta Haikkola** and **Jussi Kulonpalo**, **Niina Sihto** and **Panu Laaksonen**, **Pauliina Arvela** and **Mark Fletcher**, **Mari Kiviharju** and **Kari Rantavuori**, **Sari Lankinen**, Laura Villella-Tolonen and **Mikko Tolonen**, **Risto Sarvas** and **Tuula Savola**, **Mari Erma**, **Tania Ruutu** and **Anna Elmgren** for showing an interest in my scientific undertakings as well as good times during the years. Special words of thanks go to the band members, **Timo Erma**, **Olli Ruutu** and **Jaakko Vehanen**, for their almost life-long friendship and continuous support. Even when domestic beer has been replaced by fancy exports and sessions often begin with salmon canapé servings, I still think that we occasionally truly ROCK! It is a privilege to have been able to play and share that passion for music (and whatever comes along with it) with them.

Sincere thanks to my wife's family, **Timo** and **Riikka Elvo**, **Annikki Elvo**, **Jenni** and **Luke Bland**, **Sanna**, **Petri**, **Patrik**, **Anni** and **Emma Viro**, and **Heidi**, **Sami**, **Topias**, **Ella** and **Iris Saurén**, for caring and having faith in me.

I am grateful for my parents, **Hannele** and **Antti Myllykangas**, for their unconditional love, support and encouragement. I think my mother is the one that set the spark for my interest in writing and pursuing a career in a field that enables me to express myself with words. I believe it was because of my father's influence, endless curiosity and desire of knowledge, that I have developed a fascination for research. I thank my parents, for their intuition and guidance in many occasions when making decisions seemed difficult. I wish to thank my brothers for their love and support. I respect **Jani's** no-nonsense attitude and I look up to **Miika** for his courage and upright search for his own way in life. In many ways, their examples have been comforting to follow.

Foremost thanks go to my wife Nina, for her love and for being such a strong companion, since it has been her who had to put up with my non-productive loitering on the living room sofa. To me she is a tower of strength. Without her endless effort in running our family, my accomplishments would not be worth anything. I think that she is by far the most efficient person that I know and it is altogether her merit that I generally get anything done. I thank my children, Viljami and Saara, for their love. From them I get the most rewarding feedback, as well as the most intense criticism, and I truly cherish them for that. ❦

Helsinki, August 29th, 2008

Concepts

Achlorhydric:
A state when production of gastric acid in the stomach is low.

Adenocarcinoma:
A type of cancer that originates from glandular epithelial tissue.

Amplicon:
A genomic region containing multiple DNA copies.

Anaerobic organism:
An organism that does not need oxygen to grow.

Anaphase:
A late stage in cell division where sister chromatin moves towards opposite poles of the dividing nucleus.

Antioxidant:
A molecule that is capable of scavenging electrons and preventing oxidation of other molecules. Antioxidants protect cells from free radicals that are generated in oxidation reactions. Free radicals can damage cells.

Apoptosis:
A programmed cell death of a complex system of biological events leading to disposal of the cell.

Bernoulli distribution:
A discrete probability distribution with two outcomes. Given outcome 1 has a probability of p, outcome 0 has a probability of $p' = 1-p$.

Biomarker:
An indicator of a specific biological state.

Carcinogenesis:
A process where a normal cell is transformed into a cancer cell.

Centromere:
A region of specific and repetitive DNA in the chromosomes that function in chromosomal segregation during cell division.

Chromatin:
A composite of DNA and proteins forming the chromosome.

Confocal:
An optic system that includes one lens for both focusing of the emission laser to the object and fluorescence to the image detector.

Cross-validation:
A procedure where data is divided into subsets, training and validation sets, that are used to executing the initial analysis and validating and conforming the result.

Data mining:
"Analysis of large observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner."
(Hand et al., 2001)

DNA hybridization:
A process of combining single-stranded complementary DNAs into one molecule.

DNA sequencing:
A method of determining the order of the nucleotides in a DNA molecule.

Electrophoretic mobility shift assay:
A laboratory technique that is used to study protein-DNA interactions. Proteins and radio-labeled DNA are mixed and radioactivity is detected using gel electrophoresis separation. Specific antibodies are used to identify proteins that bind to radio-labeled DNA.

Ensembl:
A publicly available database for genomic information that can be accessed at www.ensembl.org.

Epidemiology:
A discipline of studying factors that affect health and illnesses in populations.

Epigenetic:
A change in the gene function without changing the basic structure of the underlying DNA sequence. Epigenetic changes can remain after cell division.

Epithelial cell:
In general, epithelial cells may originate from ectodermal (e.g., the epidermis), endodermal (e.g., the intestinal tract) and mesodermal (e.g. wall of gut) germ layers of the developing embryo. Epithelium is a classification of the tissue forming glands and layers that line the cavities and structures of the body, for example skin, mouth, esophagus, gastrointestinal tract as well as reproductive and urinary tracts. Cancers that originate from epithelium are called carcinomas.

Expection-Maximization algorithm:
An iterative maximum likelihood estimation algorithm, to be used, for instance, in mixture models.

Finite mixture model:
In statistics and machine learning, a mixture model is a linear superposition of simple distributions (components), that together form a more complex distribution. Mixture models are used for example in clustering.

Forward selection algorithm:
An algorithm that proceeds by combining the currently best set of variables with the remaining single variables and chooses the best combination as the new current best set.

Fusion gene:
A combination of previously separate genes.

Gamete:
A haploid egg and sperm cell.

Gastric atrophy:
Partial or complete disappearance of the gastric mucosa, glands and gastric pits.

Gastric ulcer:
A gastric wound that derives from the necrotic tissue removal.

Gastritis:
An inflammation of the mucous membrane layer of the stomach.

Genome:
Organism's hereditary information encoded in DNA.

Glioma:
A brain tumor arising from glial cells.

Hybridization probe:
A fragment of DNA used to detect presence of complementary DNA or RNA in the sample.

Immunohistochemistry:
A laboratory technique, which utilizes antibodies to detect specific antigens in tissues.

Intestinal metaplasia:
A premalignant condition where gastric epithelium has been replaced with cells of intestinal morphology.

Kinase:
An enzyme that catalyzes a reaction called phosphorylation, the transfer of a phosphate group from a donor molecule, e.g., ATP, to a target (substrate). Protein kinases are modifying the function of target proteins by phosphorylation.

Lesion:
A pathologic change in tissue structure and function.

Leukocyte:
A cell of the immune system. White blood cells are specialized to neutralize microbe infections and foreign material. Leukocytes derive from hematopoietic stem cells in the bone marrow.

Machine learning:
A computational discipline of identifying patterns and extracting rules from large data sets using statistical methods and algorithms that are able to learn from the data.

Massively parallel sequencing:
A method that allows high-throughput DNA sequencing. The sequencing by extension platforms (Thomas et al., 2006) from 454 Life Sciences (Bradford, CT) and Illumina (Hayward, CA) can sequence over 100 megabase s per day.

Maximum likelihood:
A principle that estimates parameters to maximize the probability of the data.

Mesenchymal cell:
A type of cell originating from the mesodermal germ layer of the developing embryo. Given a proper extracellular environment, the mesenchymal cells are able to migrate individually. This class of cells includes typically the cells in the connective tissue, bone, muscle and fat. Mesenchymal tumors arise from mesenchymal stem cells.

Metadata:
Background data about the observational measurement data.

Metaphase:
A stage in mitosis, in the eukaryotic cell cycle, in which chromosomes are condensed and aligned in the middle of the nucleous.

Metaplasia:
Replacement of the original differentiated cell type with another differentiated cell type.

Metastasis:
Spreading of the malignant tumor cells from a primary anatomical location to another site that is not adjacent to the original location.

Monoclonal antibody:
Antibodies, also called immunoglobulins, are proteins that are produced by specialized immune system cells, B-cells. Antibodies bind to specific microbial structures, antigens, and initiate the immune response to neutralize the infection. Each cell produces a unique antibody and monoclonal antibodies refer to proteins produced by one type of B-cell that has been clonally multiplied. Monoclonal antibodies are used in cancer therapy to neutralize deregulated oncoproteins.

Naïve Bayes classifier:
A probabilistic classifier, which assumes, given the class variable, that the variables in the model are independent.

Neuroblastoma:
The most common solid tumor in children arising from sympathetic nervous system.

Oncogenes:
Class of genes that encode proteins that promote transformation of cells to cancer cells when mutations increase or dysregulate their activity.

Orthologous:
Corresponding elements between two species.

Permutation analysis:
A method of generating the empirical probability distribution by randomly mixing data. The obtained permuted probability distribution can be then used to estimate the significance of the true observation.

Promoter:
A DNA region of specific sequence that is involved in controlling the expression of the adjacent gene by binding specific transcription factors.

P-value:
Probability of obtaining a value by random that is at least as extreme as the measured value.

P-value correction:
When multiple hypotheses are measured, some p-values may become significant by chance. There are different methods that can be used to correct the p-values obtained from a statistical test repeated many times.

Real-time polymerase chain reaction:
A technique that can be used to detect and quantify the amount of a specific DNA sequence in a sample. The procedure follows basic PCR but the accumulation of a PCR product is measured after each amplification cycle. When RNA is used as a template it has to be reverse transcribed into cDNA before quantitative PCR reaction.

Receiver operating characteristic:
A curve depicting the trade-off between the false positives and true positives for a diagnostic system for every cut-off point for decision.

Somatic cells:
All other cells in the body except the cells of the germ line.

Stromal cells:
Cells that form the supportive framework of connective tissue.

Systems biology:
An approach to study the complex interactions between components of the biological system.

Telomere:
A region of repetitive DNA in the ends of the chromosomes. Telomeres function, for example, in protecting chromosomes from breakage during replication and cell division.

Transcription factor:
A protein that binds specific DNA sequence and is part of the protein complex that controls transcription.

T-test:
A statistical test that is used to compare mean values in two groups of normally distributed random variables.

Tumor suppressor genes:
A class of genes that encode proteins that function to prevent cells from transformation into cancer cells.

References

Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. (2003). *Chromosome aberrations in solid tumors.* Nat Genet 34, 369–376.

Andrechek, E. R., Hardy, W. R., Siegel, P. M., Rudnicki, M. A., Cardiff, R. D., and Muller, W. J. (2000). *Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis.* Proc Natl Acad Sci U S A 97, 3444–3449.

Atherton, J. C. (1997). *The clinical relevance of strain types of Helicobacter pylori.* Gut 40, 701-703.

Atherton, J. C., Timothy, L. C., Papini, E., and Telford, J. L. (2001). *Vacuolating cytotoxin, In Helicobacter pylori: physiology and genetics, H. L. T. Mobley, G. L. Mendz, and S. L. Hazell, eds.* (Washington: ASM Press), pp. 97–110.

Ban, S., Cologne, J. B., and Neriishi, K. (1995). *Effect of radiation and cigarette smoking on expression of FUdR-inducible common fragile sites in human peripheral lymphocytes.* Mutat Res 334, 197–203.

Belkhiri, A., Zaika, A., Pidkovka, N., Knuutila, S., Moskaluk, C., and El-Rifai, W. (2005). *Darpp-32: a novel antiapoptotic gene in upper gastrointestinal carcinomas.* Cancer Res 65, 6583–6592.

Bibb, J. A., Snyder, G. L., Nishi, A., Yan, Z., Meijer, L., Fienberg, A. A., Tsai, L. H., Kwon, Y. T., Girault, J. A., Czernik, A. J., et al. (1999). *Phosphorylation of DARPP-32 by Cdk5 modulates dopamine signalling in neurons.* Nature 402, 669–671.

Birney, E., Andrews, T. D., Bevan, P., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cuff, J., Curwen, V., Cutts, T., et al. (2004). *An overview of Ensembl.* Genome Res 14, 925–928.

Bizari, L., Borim, A. A., Leite, K. R., Goncalves Fde, T., Cury, P. M., Tajara, E. H., and Silva, A. E. (2006). *Alterations of the CCND1 and HER-2/neu (ERBB2) proteins in esophageal and gastric cancers.* Cancer Genet Cytogenet 165, 41–50.

Blaser, M. J., Perez-Perez, G. I., Kleanthous, H., Cover, T. L., Peek, R. M., Chyou, P. H., Stemmermann, G. N., and Nomura, A. (1995). *Infection with Helicobacter pylori strains possessing cagA is associated with an increased risk of developing adenocarcinoma of the stomach.* Cancer Res 55, 2111–2115.

Blount, B. C., Mack, M. M., Wehr, C. M., MacGregor, J. T., Hiatt, R. A., Wang, G., Wickramasinghe, S. N., Everson, R. B., and Ames, B. N. (1997). *Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage.* Proc Natl Acad Sci U S A 94, 3290–3295.

Boren, J., Cascante, M., Marin, S., Comin-Anduix, B., Centelles, J. J., Lim, S., Bassilian, S., Ahmed, S., Lee, W. N., and Boros, L. G. (2001). *Gleevec (STI571) influences metabolic enzyme activities and glucose carbon flow toward nucleic acid and fatty acid synthesis in myeloid tumor cells.* J Biol Chem 276, 37747–37753.

Boros, L. G., Lee, W. N., and Cascante, M. (2002). *Imatinib and chronic-phase leukemias*. N Engl J Med 347,67–68.

Buyse, M., Loi, S., van't Veer, L., Viale, G., Delorenzi, M., Glas, A. M., d'Assignies, M. S., Bergh, J., Lidereau, R., Ellis, P., et al. (2006). *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst 98, 1183–1192.

Chen, Y., Dougherty, E. R., and Bittner, M. L. (1997). *Ratio-based decisions and the quantitative analysis of cDNA microarray images.* J Biomed Opt 2, 364–374.

Chen, Y., Kamat, V., Dougherty, E. R., Bittner, M. L., Meltzer, P. S., and Trent, J. M. (2002). *Ratio statistics of gene expression levels and applications to microarray data analysis.* Bioinformatics 18, 1207–1215.

Cheng, X. X., Wang, Z. C., Chen, X. Y., Sun, Y., Kong, Q. Y., Liu, J., and Li, H. (2005). *Correlation of Wnt-2 expression and beta-catenin intracellular accumulation in Chinese gastric cancers: relevance with tumour dissemination.* Cancer Lett 223, 339–347.

Cho, Y. G., Choi, B. J., Kim, C. J., Song, J. H., Zhang, C., Nam, S. W., Lee, J. Y., and Park, W. S. (2007a). *Genetic analysis of the DBC2 gene in gastric cancer.* Acta Oncol, 1–6.

Cho, Y. G., Song, J. H., Kim, C. J., Lee, Y. S., Kim, S. Y., Nam, S. W., Lee, J. Y., and Park, W. S. (2007b). *Genetic alterations of the ATBF1 gene in gastric cancer.* Clin Cancer Res 13, 4355–4359.

Coquelle, A., Pipiras, E., Toledo, F., Buttin, G., and Debatisse, M. (1997). *Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons.* Cell 89, 215–225.

Coquelle, A., Toledo, F., Stern, S., Bieth, A., and Debatisse, M. (1998). *A new role for hypoxia in tumor progression: induction of fragile site triggering genomic rearrangements and formation of complex DMs and HSRs.* Mol Cell 2, 259–265.

Correa, P. (1992). *Human gastric carcinogenesis: a multistep and multifactorial process–First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention.* Cancer Res 52, 6735–6740.

Correa, P., and Houghton, J. (2007). *Carcinogenesis of Helicobacter pylori.* Gastroenterology 133, 659–672.

Covacci, A., and Rappuoli, R. (2000). *Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell.* J Exp Med 191, 587–592.

Covacci, A., Telford, J. L., Del Giudice, G., Parsonnet, J., and Rappuoli, R. (1999). *Helicobacter pylori virulence and genetic geography.* Science 284, 1328–1333.

Craanen, M. E., Blok, P., Dekker, W., Ferwerda, J., and Tytgat, G. N. (1992). Subtypes of intestinal metaplasia and Helicobacter pylori. Gut 33, 597–600.

Crick, F. (1970). *Central dogma of molecular biology.* Nature 227, 561–563.

——— (1958). *On protein synthesis.* Symp Soc Exp Biol 12, 138–163.

de Figueiredo Soares, T., de Magalhaes Queiroz, D. M., Mendes, E. N., Rocha, G. A., Rocha Oliveira, A. M., Alvares Cabral, M. M., and de Oliveira, C. A. (1998). *The interrelationship between Helicobacter pylori vacuolating cytotoxin and gastric carcinoma.* Am J Gastroenterol 93, 1841–1847.

De Luca, A., Baldi, A., Russo, P., Todisco, A., Altucci, L., Giardullo, N., Pasquale, L., Iaquinto, S., D'Onofrio, V., Parodi, M. C., et al. (2003). Coexpression of Helicobacter pylori's proteins CagA and HspB induces cell proliferation in AGS gastric epithelial cells, independently from the bacterial infection. Cancer Res 63, 6350–6356.

De Luca, A., De Falco, M., Iaquinto, S., and Iaquinto, G. (2004). *Effects of Helicobacter pylori infection on cell cycle progression and the expression of cell cycle regulatory proteins.* J Cell Physiol 200, 334–342.

DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A., and Trent, J. M. (1996). *Use of a cDNA microarray to analyse gene expression patterns in human cancer.* Nat Genet 14, 457–460.

Dockray, G. J. (2004). *Clinical endocrinology and metabolism. Gastrin.* Best Pract Res Clin Endocrinol Metab 18, 555–568.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification.* (Hobokan: John Wiley & Sons).

Dunn, B. E., Campbell, G. P., Perez-Perez, G. I., and Blaser, M. J. (1990). *Purification and characterization of urease from Helicobacter pylori.* J Biol Chem 265, 9464–9469.

Edgar, R., Domrachev, M., and Lash, A. E. (2002). *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Res 30, 207–210.

Eferl, R., and Wagner, E. F. (2003). *AP-1: a double-edged sword in tumorigenesis.* Nat Rev Cancer 3, 859–868.

Ernst, P. (1999). *Review article: the role of inflammation in the pathogenesis of gastric cancer.* Aliment Pharmacol Ther 13 Suppl 1, 13–18.

Everitt, B. S., and Hand, D. J. (1981). *Finite Mixture Distributions (Monographs on Applied Probability and Statistics)* (Boca Raton: Chapman & Hall).

Fakharzadeh, S. S., Rosenblum-Vos, L., Murphy, M., Hoffman, E. K., and George, D. L. (1993). *Structure and organization of amplified DNA on double minutes containing the mdm2 oncogene.* Genomics 15, 283–290.

Fang, D. C., Luo, Y. H., Yang, S. M., Li, X. A., Ling, X. L., and Fang, L. (2002). *Mutation analysis of APC gene in gastric cancer with microsatellite instability.* World J Gastroenterol 8, 787–791.

Fenoglio-Preiser, C., Carneiro, F., Correa, P., Guilford, P., Lambert, R., Megraud, F., Muñoz, N., Powell, S. M., Rugge, M., Sasako., M., et al. (2000). *Gastric carcinoma, In Pathology and Genetics.* Tumors of the Digestive System, S. R. Hamilton, and L. A. Aaltonen, eds. (Lyon: IARC Press), pp. 39–52.

Ferlay, J., Bray, F., Pisani, P., and Parkin, D. (2004). *GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide.* IARC CancerBase No. 5. version 2.0. (Lyon: IARC Press).

Fischer, W., Puls, J., Buhrdorf, R., Gebert, B., Odenbreit, S., and Haas, R. (2001). *Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8.* Mol Microbiol 42, 1337–1348.

Flicek, P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., et al. (2008). *Ensembl 2008.* Nucleic Acids Res 36, D707–714.

Forman, D., Newell, D. G., Fullerton, F., Yarnell, J. W., Stacey, A. R., Wald, N., and Sitas, F. (1991). *Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation.* Bmj 302, 1302–1305.

Franco, A. T., Israel, D. A., Washington, M. K., Krishna, U., Fox, J. G., Rogers, A. B., Neish, A. S., Collier-Hyams, L., Perez-Perez, G. I., Hatakeyama, M., et al. (2005). *Activation of beta-catenin by carcinogenic Helicobacter pylori.* Proc Natl Acad Sci U S A 102, 10646–10651.

Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M. R. (2004). *A census of human cancer genes.* Nat Rev Cancer 4, 177–183.

Galizia, G., Lieto, E., Orditura, M., Castellano, P., Mura, A. L., Imperatore, V., Pinto, M., Zamboli, A., De Vita, F., and Ferraraccio, F. (2007). *Epidermal growth factor receptor (EGFR) expression is associated with a worse prognosis in gastric cancer patients undergoing curative surgery.* World J Surg 31, 1458–1468.

Garraway, L. A., and Sellers, W. R. (2006). *Lineage dependency and lineage-survival oncogenes in human cancer.* Nat Rev Cancer 6, 593–602.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P., and Bickmore, W. A. (2004). *Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers.* Cell 118, 555–566.

Glover, T. W., Berger, C., Coyle, J., and Echo, B. (1984). *DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes.* Hum Genet 67, 136–142.

Goker, E., Waltham, M., Kheradpour, A., Trippett, T., Mazumdar, M., Elisseyeff, Y., Schnieders, B., Steinherz, P., Tan, C., Berman, E., and et al. (1995). *Amplification of the dihydrofolate reductase gene is a mechanism of acquired resistance to methotrexate in patients with acute lymphoblastic leukemia and is correlated with p53 gene mutations.* Blood 86, 677–684.

Gong, S. J., Jin, C. J., Rha, S. Y., and Chung, H. C. (2004). *Growth inhibitory effects of trastuzumab and chemotherapeutic drugs in gastric cancer cell lines.* Cancer Lett 214, 215–224.

Goueli, B. S., and Janknecht, R. (2003). *Regulation of telomerase reverse transcriptase gene activity by upstream stimulatory factor.* Oncogene 22, 8042–8047.

Graux, C., Cools, J., Melotte, C., Quentmeier, H., Ferrando, A., Levine, R., Vermeesch, J. R., Stul, M., Dutta, B., Boeckx, N., et al. (2004). *Fusion of NUP214 to ABL1 on amplified episomes in T-cell acute lymphoblastic leukemia.* Nat Genet 36, 1084–1089.

Green, D., Ponce de Leon, S., Leon-Rodriguez, E., and Sosa-Sanchez, R. (2002). *Adenocarcinoma of the stomach: univariate and multivariate analysis of factors associated with survival.* Am J Clin Oncol 25, 84–89.

Greengard, P., Allen, P. B., and Nairn, A. C. (1999). *Beyond the dopamine receptor: the DARPP-32/protein phosphatase-1 cascade.* Neuron 23, 435–447.

Guan, X. Y., Meltzer, P. S., Dalton, W. S., and Trent, J. M. (1994). *Identification of cryptic sites of DNA sequence amplification in human breast cancer by chromosome microdissection.* Nat Genet 8, 155–161.

Hahn, P. J. (1993). *Molecular biology of double-minute chromosomes.* Bioessays 15, 477–484.

Hanahan, D., and Weinberg, R. A. (2000). *The hallmarks of cancer.* Cell 100, 57–70.

Hand, D., Mannila, H., and Smyth, P. (2001). *Principles of Data Minin.* (Cambridge: MIT Press).

Harris, A. L. (2002). *Hypoxia–a key regulatory factor in tumour growth.* Nat Rev Cancer 2, 38–47.

Harris, P. R., Mobley, H. L., Perez-Perez, G. I., Blaser, M. J., and Smith, P. D. (1996). *Helicobacter pylori urease is a potent stimulus of mononuclear phagocyte activation and inflammatory cytokine production.* Gastroenterology 111, 419–425.

Harrison, L. E., Karpeh, M. S., and Brennan, M. F. (1998). *Extended lymphadenectomy is associated with a survival benefit for node-negative gastric cancer.* J Gastrointest Surg 2, 126–131.

Hatakeyama, M. (2004). *Oncogenic mechanisms of the Helicobacter pylori CagA protein.* Nat Rev Cancer 4, 688–694.

Hautaniemi, S., Ringnèr, M., Kauraniemi, P., Autio, R., Edgren, H., Yli-Harja, O., Astola, J., Kallioniemi, A., and Kallioniemi, O.-P. (2004). *A strategy for identifying putative causes of gene expression variation in human cancers.* J Franklin Institute 341, 77–88.

Hellman, A., Zlotorynski, E., Scherer, S. W., Cheung, J., Vincent, J. B., Smith, D. I., Trakhtenbrot, L., and Kerem, B. (2002). *A role for common fragile site induction in amplification of human oncogenes.* Cancer Cell 1, 89–97.

Higashi, H., Tsutsumi, R., Muto, S., Sugiyama, T., Azuma, T., Asaka, M., and Hatakeyama, M. (2002). *SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein.* Science 295, 683–686.

Higgins, M. E., Claremont, M., Major, J. E., Sander, C., and Lash, A. E. (2007). *CancerGenes: a gene selection resource for cancer genome projects.* Nucleic Acids Res 35, D721–726.

Himberg, J., Hyvärinen, A., and Esposito, F. (2004). *Validating the independent components of neuroimaging time-series via clustering and visualization.* Neuroimage 22, 1214–1222.

Hollmén, J., and Tikka, J. (2007). *Compact and Understandable Descriptions of Mixtures of Bernoulli Distributions, In Proceedings of the 7th International Symposium on Intelligent Data Analysis (IDA 2007): September 2007; Ljubljana, M. R. Berthold, J. Shawe-Taylor, and N. Lavrac, eds.* (Heidelberg: Springer-Verlag), pp. 1–12.

Homma, N., Tamura, G., Honda, T., Matsumoto, Y., Nishizuka, S., Kawata, S., and Motoyama, T. (2006). *Spreading of methylation within RUNX3 CpG island in gastric cancer.* Cancer Sci 97, 51–56.

Hundahl, S. A., Phillips, J. L., and Menck, H. R. (2000). *The National Cancer Data Base Report on poor survival of U.S. gastric carcinoma patients treated with gastrectomy: Fifth Edition American Joint Committee on Cancer staging, proximal disease, and the "different disease" hypothesis.* Cancer 88, 921–932.

Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringner, M., Sauter, G., Monni, O., Elkahloun, A., et al. (2002). *Impact of DNA amplification on gene expression patterns in breast cancer.* Cancer Res 62, 6240–6245.

Hyvärinen, A., Oja, E., and Karhunen, J. (2001). *Independent Component Analysis* (New York: John Wiley & Sons).

IARC (1994). *Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7–14 June 1994.* IARC Monogr Eval Carcinog Risks Hum 61, 1–241.

Iliopoulos, D., Guler, G., Han, S. Y., Druck, T., Ottey, M., McCorkell, K. A., and Huebner, K. (2006). *Roles of FHIT and WWOX fragile genes in cancer.* Cancer Lett 232, 27–36.

Inoue, T., Kataoka, H., Goto, K., Nagaike, K., Igami, K., Naka, D., Kitamura, N., and Miyazawa, K. (2004). *Activation of c-Met (hepatocyte growth factor receptor) in human gastric cancer tissue.* Cancer Sci 95, 803–808.

Ito, K., Liu, Q., Salto-Tellez, M., Yano, T., Tada, K., Ida, H., Huang, C., Shah, N., Inoue, M., Rajnakova, A., et al. (2005). *RUNX3, a novel tumor suppressor, is frequently inactivated in gastric cancer by protein mislocalization.* Cancer Res 65, 7743–7750.

Jackman, J., Alamo, I., Jr., and Fornace, A. J., Jr. (1994). *Genotoxic stress confers preferential and coordinate messenger RNA stability on the five gadd genes.* Cancer Res 54, 5656–5662.

Jensen, C. G., Jensen, L. C., Rieder, C. L., Cole, R. W., and Ault, J. G. (1996). *Long crocidolite asbestos fibers cause polyploidy by sterically blocking cytokinesis.* Carcinogenesis 17, 2013–2021.

Jiang, H. Y., Wek, S. A., McGrath, B. C., Lu, D., Hai, T., Harding, H. P., Wang, X., Ron, D., Cavener, D. R., and Wek, R. C. (2004). *Activating transcription factor 3 is integral to the eukaryotic initiation factor 2 kinase stress response.* Mol Cell Biol 24, 1365–1377.

Jiang, Q., Hiratsuka, K., and Taylor, D. E. (1996). *Variability of gene order in different Helicobacter pylori strains contributes to genome diversity.* Mol Microbiol 20, 833–842.

Johnston, R. N., Beverley, S. M., and Schimke, R. T. (1983). *Rapid spontaneous dihydrofolate reductase gene amplification shown by fluorescence-activated cell sorting.* Proc Natl Acad Sci U S A 80, 3711–3715.

Ju, H., Cho, H., Kim, Y. S., Kim, W. H., Ihm, C., Noh, S. M., Kim, J. B., Hahn, D. S., Choi, B. Y., and Kang, C. (2006). *Functional polymorphism 57Val>Ile of aurora kinase A associated with increased risk of gastric cancer progression.* Cancer Lett 242, 273–279.

Juttner, S., Cramer, T., Wessler, S., Walduck, A., Gao, F., Schmitz, F., Wunder, C., Weber, M., Fischer, S. M., Schmidt, W. E., et al. (2003). *Helicobacter pylori stimulates host cyclooxygenase-2 gene transcription: critical importance of MEK/ERK-dependent activation of USF1/-2 and CREB transcription factors.* Cell Microbiol 5, 821–834.

Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. (1992). *Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors.* Science 258, 818–821.

Kamada, K., Yamada, Y., Hirao, T., Fujimoto, H., Takahama, Y., Ueno, M., Takayama, T., Naito, A., Hirao, S., and Nakajima, Y. (2004). *Amplification/overexpression of Aurora-A in human gastric carcinoma: potential role in differentiated type gastric carcinogenesis.* Oncol Rep 12, 593–599.

Kang, G. H., Lee, H. J., Hwang, K. S., Lee, S., Kim, J. H., and Kim, J. S. (2003). *Aberrant CpG island hypermethylation of chronic gastritis, in relation to aging, gender, intestinal metaplasia, and chronic inflammation.* Am J Pathol 163, 1551–1556.

Karin, M., Cao, Y., Greten, F. R., and Li, Z. W. (2002). *NF-kappaB in cancer: from innocent bystander to major culprit.* Nat Rev Cancer 2, 301–310.

Kawahara, T., Kuwano, Y., Teshima-Kondo, S., Sugiyama, T., Kawai, T., Nikawa, T., Kishi, K., and Rokutan, K. (2001). *Helicobacter pylori lipopolysaccharide from type I, but not type II strains, stimulates apoptosis of cultured gastric mucosal cells.* J Med Invest 48, 167–174.

Kim, C. J., Song, J. H., Cho, Y. G., Kim, Y. S., Kim, S. Y., Nam, S. W., Yoo, N. J., Lee, J. Y., and Park, W. S. (2007). *Somatic mutations of the beta-TrCP gene in gastric cancer.* Apmis 115, 127–133.

Kim, Y. B., Han, J. Y., Kim, T. S., Kim, P. S., and Chu, Y. C. (2000). *Overexpression of c-H-ras p21 is correlated with vascular endothelial growth factor expression and neovascularization in advanced gastric carcinoma.* J Gastroenterol Hepatol 15, 1393–1399.

Kleihues, P., and Sobin, L. H., eds. (2000–2006). *World Health Organization Classification of Tumours* (Lyon: IARCPress).

Knuutila, S., Autio, K., and Aalto, Y. (2000). *Online access to CGH data of DNA sequence copy number changes.* Am J Pathol 157, 689.

Kokkola, A., Monni, O., Puolakkainen, P., Larramendy, M. L., Victorzon, M., Nordling, S., Haapiainen, R., Kivilaakso, E., and Knuutila, S. (1997). *17q12–21 amplicon, a novel recurrent genetic change in intestinal type of gastric carcinoma: a comparative genomic hybridization study.* Genes Chromosomes Cancer 20, 38–43.

Kolho, K. L., Karttunen, R., Heikkila, P., Lindahl, H., and Rautelin, H. (1999). *Gastric inflammation is enhanced in children with CagA-positive Helicobacter pylori infection.* Pediatr Infect Dis J 18, 337–341.

Kuwano, A., and Kajii, T. (1987). *Synergistic effect of aphidicolin and ethanol on the induction of common fragile sites.* Hum Genet 75, 75–78.

Lan, J., Xiong, Y. Y., Lin, Y. X., Wang, B. C., Gong, L. L., Xu, H. S., and Guo, G. S. (2003). *Helicobacter pylori infection generated gastric cancer through p53-Rb tumor-suppressor system mutation and telomerase reactivation.* World J Gastroenterol 9, 54–58.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). *Initial sequencing and analysis of the human genome.* Nature 409, 860–921.

Lauren, P. (1965). *The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification.* Acta Pathol Microbiol Scand 64, 31–49.

Lee, J. W., Soung, Y. H., Seo, S. H., Kim, S. Y., Park, C. H., Wang, Y. P., Park, K., Nam, S. W., Park, W. S., Kim, S. H., et al. (2006). *Somatic mutations of ERBB2 kinase domain in gastric, colorectal, and breast carcinomas.* Clin Cancer Res 12, 57–61.

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). *Genetic instabilities in human cancers.* Nature 396, 643–649.

Lenhard, B., Sandelin, A., Mendoza, L., Engstrom, P., Jareborg, N., and Wasserman, W. W. (2003). *Identification of conserved regulatory elements by comparative genome analysis.* J Biol 2, 13.

Li, V. S., Wong, C. W., Chan, T. L., Chan, A. S., Zhao, W., Chu, K. M., So, S., Chen, X., Yuen, S. T., and Leung, S. Y. (2005a). *Mutations of PIK3CA in gastric adenocarcinoma.* BMC Cancer 5, 29.

Li, Y., Ren, J., Yu, W., Li, Q., Kuwahara, H., Yin, L., Carraway, K. L., 3rd, and Kufe, D. (2001). *The epidermal growth factor receptor regulates interaction of the human DF3/MUC1 carcinoma antigen with c-Src and beta-catenin.* J Biol Chem 276, 35239–35242.

Li, Y. L., Tian, Z., Wu, D. Y., Fu, B. Y., and Xin, Y. (2005b). *Loss of heterozygosity on 10q23.3 and mutation of tumor suppressor gene PTEN in gastric cancer and precancerous lesions.* World J Gastroenterol 11, 285–288.

Lingjaerde, O. C., Baumbusch, L. O., Liestol, K., Glad, I. K., and Borresen-Dale, A. L. (2005). *CGH-Explorer: a program for analysis of array-CGH data.* Bioinformatics 21, 821–822.

Liu, Z. F., Chen, C. Y., Tang, W., Zhang, J. Y., Gong, Y. Q., and Jia, J. H. (2006). *Gene-expression profiles in gastric epithelial cells stimulated with spiral and coccoid Helicobacter pylori.* J Med Microbiol 55, 1009–1015.

Lowy, A. M., Clements, W. M., Bishop, J., Kong, L., Bonney, T., Sisco, K., Aronow, B., Fenoglio-Preiser, C., and Groden, J. (2006). *beta-Catenin/Wnt signaling regulates expression of the membrane type 3 matrix metalloproteinase in gastric cancer.* Cancer Res 66, 4734–4741.

Maeda, S., Yoshida, H., Ogura, K., Mitsuno, Y., Hirata, Y., Yamaji, Y., Akanuma, M., Shiratori, Y., and Omata, M. (2000). *H. pylori activates NF-kappaB through a signaling pathway involving IkappaB kinases, NF-kappaB-inducing kinase, TRAF2, and TRAF6 in gastric cancer cells.* Gastroenterology 119, 97–108.

Maqani, N., Belkhiri, A., Moskaluk, C., Knuutila, S., Dar, A. A., and El-Rifai, W. (2006). *Molecular dissection of 17q12 amplicon in upper gastrointestinal adenocarcinomas.* Mol Cancer Res 4, 449–455.

Marnett, L. J. (2000). *Oxyradicals and DNA damage.* Carcinogenesis 21, 361–370.

Marshall, B. J., and Warren, J. R. (1984). *Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration.* Lancet 1, 1311–1315.

Maurer, B. J., Lai, E., Hamkalo, B. A., Hood, L., and Attardi, G. (1987). *Novel submicroscopic extrachromosomal elements containing amplified genes in human cells.* Nature 327, 434–437.

Meyer-ter-Vehn, T., Covacci, A., Kist, M., and Pahl, H. L. (2000). *Helicobacter pylori activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes c-fos and c-jun.* J Biol Chem 275, 16064–16072.

Miettinen, M., and Lasota, J. (2006). *Gastrointestinal stromal tumors: review on morphology, molecular pathology, prognosis, and differential diagnosis.* Arch Pathol Lab Med 130, 1466–1478.

Miller, C. T., Lin, L., Casper, A. M., Lim, J., Thomas, D. G., Orringer, M. B., Chang, A. C., Chambers, A. F., Giordano, T. J., Glover, T. W., and Beer, D. G. (2006). *Genomic amplification of MET with boundaries within fragile site FRA7G and upregulation of MET pathways in esophageal adenocarcinoma.* Oncogene 25, 409–418.

Mitelman, F., Johansson, B., and Mertens, F. (1994). *Catalog of Chromosome Aberrations in Cancer, Vol 2* (New York: Wiley-Liss).

Mitsuno, Y., Yoshida, H., Maeda, S., Ogura, K., Hirata, Y., Kawabe, T., Shiratori, Y., and Omata, M. (2001). *Helicobacter pylori induced transactivation of SRE and AP-1 through the ERK signalling pathway in gastric cancer cells.* Gut 49, 18–22.

Moese, S., Selbach, M., Kwok, T., Brinkmann, V., Konig, W., Meyer, T. F., and Backert, S. (2004). *Helicobacter pylori induces AGS cell motility and elongation via independent signaling pathways.* Infect Immun 72, 3646–3649.

Monni, O., Hautaniemi, S., and Kallioniemi, O. (2002). *Geenisirutekniikka ja siihen liittyvä bioinformatiikka.* Duodecim 118, 1157–1166.

Mousses, S., Bittner, M. L., Chen, Y., Dougherty, E. R., Baxevanis, A., Meltzer, P. S., and Trent, J. M. (2000). *Gene expression analysis by cDNA microarrays, In Functional genomics, F. J. Livesey, and S. P. Hunt, eds.* (Oxford: Oxford University Press), pp. 113–137.

Msika, S., Benhamiche, A. M., Jouve, J. L., Rat, P., and Faivre, J. (2000). *Prognostic factors after curative resection for gastric cancer. A population-based study.* Eur J Cancer 36, 390–396.

Murnane, J. P., and Sabatier, L. (2004). *Chromosome rearrangements resulting from telomere dysfunction and their role in cancer.* Bioessays 26, 1164–1174.

Myllykangas, S., Böhling, T., and Knuutila, S. (2007). *Specificity, selection and significance of gene amplifications in cancer.* Semin Cancer Biol 17, 42–55.

Nakayama, M., Kimura, M., Wada, A., Yahiro, K., Ogushi, K., Niidome, T., Fujikawa, A., Shirasaka, D., Aoyama, N., Kurazono, H., et al. (2004). *Helicobacter pylori VacA activates the p38/activating transcription factor 2-mediated signal pathway in AZ-521 cells.* J Biol Chem 279, 7024–7028.

Nardone, G., Compare, D., De Colibus, P., de Nucci, G., and Rocco, A. (2007). *Helicobacter pylori and epigenetic mechanisms underlying gastric carcinogenesis.* Dig Dis 25, 225–229.

Nevins, J. R. (2001). *The Rb/E2F pathway and cancer.* Hum Mol Genet 10, 699–703.

Odenbreit, S., Puls, J., Sedlmaier, B., Gerland, E., Fischer, W., and Haas, R. (2000). *Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion.* Science 287, 1497–1500.

Oliveira, C., Moreira, H., Seruca, R., de Oliveira, M. C., and Carneiro, F. (2005). *Role of pathology in the identification of hereditary diffuse gastric cancer: report of a Portuguese family.* Virchows Arch 446, 181–184.

Oliveira, C., Seruca, R., and Carneiro, F. (2006). *Genetics, pathology, and clinics of familial gastric cancer.* Int J Surg Pathol 14, 21–33.

Parsonnet, J., Friedman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N., and Sibley, R. K. (1991a). *Helicobacter pylori infection and the risk of gastric carcinoma.* N Engl J Med 325, 1127–1131.

Parsonnet, J., Vandersteen, D., Goates, J., Sibley, R. K., Pritikin, J., and Chang, Y. (1991b). *Helicobacter pylori infection in intestinal- and diffuse-type gastric adenocarcinomas.* J Natl Cancer Inst 83, 640–643.

Peek, R., and Reddy, K. R. (2007). *Genetic mutations identified for hereditary diffuse gastric cancer.* Gastroenterology 133, 379–380.

Peek, R. M., Jr., and Blaser, M. J. (2002). *Helicobacter pylori and gastrointestinal tract adenocarcinomas.* Nat Rev Cancer 2, 28–37.

Pelengaris, S., Khan, M., and Evan, G. (2002). *c-MYC: more than just a matter of life and death.* Nat Rev Cancer 2, 764–776.

Pollack, J. R., Sorlie, T., Perou, C. M., Rees, C. A., Jeffrey, S. S., Lonning, P. E., Tibshirani, R., Botstein, D., Borresen-Dale, A. L., and Brown, P. O. (2002). *Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors.* Proc Natl Acad Sci U S A 99, 12963–12968.

Rao, P. N., Heerema, N. A., and Palmer, C. G. (1988). *Fragile sites induced by FUdR, caffeine, and aphidicolin. Their frequency, distribution, and analysis.* Hum Genet 78, 21–26.

Rebischung, C., Barnoud, R., Stefani, L., Faucheron, J. L., and Mousseau, M. (2005). *The effectiveness of trastuzumab (Herceptin) combined with chemotherapy for gastric carcinoma with overexpression of the c-erbB-2 protein.* Gastric Cancer 8, 249–252.

Recavarren-Arce, S., Leon-Barua, R., Cok, J., Berendson, R., Gilman, R. H., Ramirez-Ramos, A., Rodriguez, C., and Spira, W. M. (1991). *Helicobacter pylori and progressive gastric pathology that predisposes to gastric cancer.* Scand J Gastroenterol Suppl 181, 51–57.

Rugge, M., Cassaro, M., Leandro, G., Baffa, R., Avellini, C., Bufo, P., Stracca, V., Battaglia, G., Fabiano, A., Guerini, A., and Di Mario, F. (1996). *Helicobacter pylori in promotion of gastric carcinogenesis.* Dig Dis Sci 41, 950–955.

Salaspuro, M. P. (2003). *Acetaldehyde,microbes, and cancer of the digestive tract.* Crit Rev Clin Lab Sci 40, 183–208.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W., and Lenhard, B. (2004). *JASPAR: an open-access database for eukaryotic transcription factor binding profiles.* Nucleic Acids Res 32 Database issue, D91–94.

Sasaki, Y., Niu, C., Makino, R., Kudo, C., Sun, C., Watanabe, H., Matsunaga, J., Takahashi, K., Tagami, H., Aiba, S., and Horii, A. (2004). *BRAF point mutations in primary melanoma show different prevalences by subtype.* J Invest Dermatol 123, 177–183.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science 270, 467–470.

Schneider, S. S., Hiemstra, J. L., Zehnbauer, B. A., Taillon-Miller, P., Le Paslier, D. L., Vogelstein, B., and Brodeur, G. M. (1992). *Isolation and structural analysis of a 1.2-megabase N-myc amplicon from a human neuroblastoma.* Mol Cell Biol 12, 5563–5570.

Schroeder, J. A., Thompson, M. C., Gardner, M. M., and Gendler, S. J. (2001). *Transgenic MUC1 interacts with epidermal growth factor receptor and correlates with mitogen-activated protein kinase activation in the mouse mammary gland.* J Biol Chem 276, 13057–13064.

Schwab, M. (1998). *Amplification of oncogenes in human cancer cells.* Bioessays 20, 473–479.

Schwab, M., Westermann, F., Hero, B., and Berthold, F. (2003). *Neuroblastoma: biology and molecular and chromosomal pathology.* Lancet Oncol 4, 472–480.

Schwartz, G. K. (1996). *Invasion and metastases in gastric cancer: in vitro and in vivo models with clinical correlations.* Semin Oncol 23, 316–324.

Schwartz, M., Zlotorynski, E., and Kerem, B. (2006). *The molecular basis of common and rare fragile sites.* Cancer Lett 232, 13–26.

Shaffer, L., and Tommerup, N., eds. (2005). ISCN (2005): *An International System for Human Cytogenetic Nomenclature* (Basel: S. Karger).

Shannon, K. M. (2002). *Resistance in the land of molecular cancer therapeutics.* Cancer Cell 2, 99–102.

Shimizu, N., Hashizume, T., Shingaki, K., and Kawamoto, J. K. (2003). *Amplification of plasmids containing a mammalian replication initiation region is mediated by controllable conflict between replication and transcription.* Cancer Res 63, 5281–5290.

Shimizu, N., Shingaki, K., Kaneko-Sasaguri, Y., Hashizume, T., and Kanda, T. (2005). *When, where and how the bridge breaks: anaphase bridge breakage plays a crucial role in gene amplification and HSR generation.* Exp Cell Res 302, 233–243.

Shimura, M., Onozuka, Y., Yamaguchi, T., Hatake, K., Takaku, F., and Ishizaka, Y. (1999). *Micronuclei formation with chromosome breaks and gene amplification caused by Vpr, an accessory gene of human immunodeficiency virus.* Cancer Res 59, 2259–2264.

Shinmura, K., Goto, M., Tao, H., Shimizu, S., Otsuki, Y., Kobayashi, H., Ushida, S., Suzuki, K., Tsuneyoshi, T., and Sugimura, H. (2005). *A novel STK11 germline mutation in two siblings with Peutz-Jeghers syndrome complicated by primary gastric cancer.* Clin Genet 67, 81–86.

Singer, M. J., Mesner, L. D., Friedman, C. L., Trask, B. J., and Hamlin, J. L. (2000). *Amplification of the human dihydrofolate reductase gene via double minutes is initiated by chromosome breaks*. Proc Natl Acad Sci U S A 97, 7921–7926.

Smith, D. I., Zhu, Y., McAvoy, S., and Kuhn, R. (2006). *Common fragile sites, extremely large genes, neural development and cancer.* Cancer Lett 232, 48–57.

Smoot, D. T., Mobley, H. L., Chippendale, G. R., Lewison, J. F., and Resau, J. H. (1990). *Helicobacter pylori urease activity is toxic to human gastric epithelial cells.* Infect Immun 58, 1992–1994.

Su, B., Ceponis, P. J., Lebel, S., Huynh, H., and Sherman, P. M. (2003). *Helicobacter pylori activates Toll-like receptor 4 expression in gastrointestinal epithelial cells.* Infect Immun 71, 3496–3502.

Swets, J. A. (1998). *Measuring the accuracy of diagnostic systems.* Science 240, 1285–1293.

Tang, Z., Zhao, M., Ji, J., Yang, G., Hu, F., He, J., Shen, H., Gao, Z., Zhao, A., Li, J., and Lu, Y. (2004). *Overexpression of gastrin and c-met protein involved in human gastric carcinomas and intestinal metaplasia.* Oncol Rep 11, 333–339.

Teh, M., and Lee, Y. S. (1987). *Intestinal and diffuse carcinoma of the stomach among the ethnic and dialect groups in Singapore.* Cancer 60, 921–925.

Thiery, J. P. (2002). *Epithelial-mesenchymal transitions in tumour progression.* Nat Rev Cancer 2, 442–454.

Thomas, R. K., Nickerson, E., Simons, J. F., Janne, P. A., Tengs, T., Yuza, Y., Garraway, L. A., LaFramboise, T., Lee, J. C., Shah, K., et al. (2006). *Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing.* Nat Med 12, 852–855.

Tikka, J., Hollmén, J., and Myllykangas, S. (2007). *Mixture modeling of DNA copy number amplification patterns in cancer, In Proceedings of the 9th International Work-Conference on Artificial Neural Networks (IWANN'2007): June 2007; San Sebastián, F. Sandoval, A. Prieto, J. Cabestany, and M. Graña, eds.* (Heidelberg: Springer-Verlag), pp. 972–979.

Toledo, F., Buttin, G., and Debatisse, M. (1993). *The origin of chromosome rearrangements at early stages of AMPD2 gene amplification in Chinese hamster cells.* Curr Biol 3, 255–264.

Toledo, F., Le Roscouet, D., Buttin, G., and Debatisse, M. (1992). *Co-amplified markers alternate in megabase long chromosomal inverted repeats and cluster independently in interphase nuclei at early steps of mammalian gene amplification.* Embo J 11, 2665–2673.

Wahl, G. M., Padgett, R. A., and Stark, G. R. (1979). *Gene amplification causes overproduction of the first three enzymes of UMP synthesis in N-(phosphonacetyl)-L-aspartate-resistant hamster cells.* J Biol Chem 254, 8679–8689.

Varis, A., van Rees, B., Weterman, M., Ristimaki, A., Offerhaus, J., and Knuutila, S. (2003). *DNA copy number changes in young gastric cancer patients with special reference to chromosome 19.* Br J Cancer 88, 1914–1919.

Varis, A., Zaika, A., Puolakkainen, P., Nagy, B., Madrigal, I., Kokkola, A., Vayrynen, A., Karkkainen, P., Moskaluk, C., El-Rifai, W., and Knuutila, S. (2004). *Coamplified and overexpressed genes at ERBB2 locus in gastric cancer.* Int J Cancer 109, 548–553.

Watanabe, T., Tada, M., Nagai, H., Sasaki, S., and Nakao, M. (1998). *Helicobacter pylori infection induces gastric cancer in mongolian gerbils.* Gastroenterology 115, 642–648.

Watson, J. D., and Crick, F. H. (1953). *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature 171, 737–738.

Vauhkonen, H., Vauhkonen, M., Sajantila, A., Sipponen, P., and Knuutila, S. (2006). *DNA copy number aberrations in intestinal-type gastric cancer revealed by array-based comparative genomic hybridization.* Cancer Genet Cytogenet 167, 150–154.

Weinberg, W. A. (2007). *In The biology of cancer* (New York: Garland Science), pp. 441–453.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001). *The sequence of the human genome.* Science 291, 1304–1351.

Werner, M., Becker, K. F., Keller, G., and Hofler, H. (2001). *Gastric adenocarcinoma: pathomorphology and molecular pathology.* J Cancer Res Clin Oncol 127, 207–216.

Williams, C. S., Mann, M., and DuBois, R. N. (1999). *The role of cyclooxygenases in inflammation, cancer, and development.* Oncogene 18, 7908–7916.

Vinall, L. E., King, M., Novelli, M., Green, C. A., Daniels, G., Hilkens, J., Sarner, M., and Swallow, D. M. (2002). *Altered expression and allelic association of the hypervariable membrane mucin MUC1 in Helicobacter pylori gastritis.* Gastroenterology 123, 41–49.

Vogt, N., Lefevre, S. H., Apiou, F., Dutrillaux, A. M., Cor, A., Leuraud, P., Poupon, M. F., Dutrillaux, B., Debatisse, M., and Malfoy, B. (2004). *Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas.* Proc Natl Acad Sci U S A 101, 11368–11373.

Wolf, M., Mousses, S., Hautaniemi, S., Karhu, R., Huusko, P., Allinen, M., Elkahloun, A., Monni, O., Chen, Y., Kallioniemi, A., and Kallioniemi, O. P. (2004). *High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression.* Neoplasia 6, 240–247.

Von Hoff, D. D., Forseth, B., Clare, C. N., Hansen, K. L., and VanDevanter, D. (1990). *Double minutes arise from circular extrachromosomal DNA intermediates which integrate into chromosomal sites in human HL-60 leukemia cells.* J Clin Invest 85, 1887–1895.

Von Hoff, D. D., Needham-VanDevanter, D. R., Yucel, J., Windle, B. E., and Wahl, G. M. (1988). *Amplified human MYC oncogenes localized to replicating submicroscopic circular DNA molecules.* Proc Natl Acad Sci U S A 85, 4804–4808.

Yamamoto, M., Bharti, A., Li, Y., and Kufe, D. (1997). *Interaction of the DF3/MUC1 breast carcinoma-associated antigen and beta-catenin in cell adhesion.* J Biol Chem 272, 12492–12494.

Yamaoka, Y., Kita, M., Kodama, T., Sawai, N., Kashima, K., and Imanishi, J. (1997). *Induction of various cytokines and development of severe mucosal inflammation by cagA gene positive Helicobacter pylori strains.* Gut 41, 442–451.

Yamazaki, S., Yamakawa, A., Ito, Y., Ohtani, M., Higashi, H., Hatakeyama, M., and Azuma, T. (2003). *The CagA protein of Helicobacter pylori is translocated into epithelial cells and binds to SHP-2 in human gastric mucosa.* J Infect Dis 187, 334–337.

Zhang, Q. X., Ding, Y., Le, X. P., and Du, P. (2003). *Studies on microsatellite instability in p16 gene and expression of hMSH2 mRNA in human gastric cancer tissues.* World J Gastroenterol 9, 437–441.

Zhang, Y., Takeuchi, H., Nishioka, M., Morimoto, N., Kamioka, M., Kumon, Y., and Sugiura, T. (2007). *Relationship of IL-8 production and the CagA status in AGS cells infected with Helicobacter pylori exposed to low pH and activating transcription factor 3 (ATF3).* Microbiol Res.

Zhao, Y. F., Zhang, Y. G., Tian, X. X., Juan, D., and Jie, Z. (2007). *Aberrant methylation of multiple genes in gastric carcinomas.* Int J Surg Pathol 15, 242–251.

Graphic design
Tomi Metsä-Heikkilä

Cover photography
Massimiliano Gentile

Typography: Scala Pro, Scala Sans

Paper: Navigator 90 g

ISSN 1457-8433
ISBN 978-952-10-4927-9 ja
ISBN 978-952-10-4928-6 (PDF)

Yliopistopaino
Helsinki, 2008
Finland