Association of Genetic Variants of Apoptosis Related Genes with Early Colorectal Tumor Lesions

DISSERTATION

Submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

Of the Ruperto-Carola University of Heidelberg, Germany

For the degree of

Doctor of Natural Sciences

Stefano Gerola

2012

Referees:

Univ.-Prof. Dr. rer. nat. Roland Eils

Gründungsdirektor BIOQUANT, Institut für Pharmazie und
Molekulare Biotechnologie

Univ.-Prof. Dr. med. Michael Neumaier

Lehrstuhl und Direktor des Instituts für Klinische Chemie Medizinische
Fakultät Mannheim der Universität Heidelberg

Index

Zusammenfassung

Die Deregulation der Apoptose ist eine häufig auftretende Veränderung in gutartigen, präkanzerogenen Läsionen der Kolonmukosa. Ihr Beitrag zur Entwicklung von kolorektalen Karzinomen wurde ausgiebig diskutiert. Individuelle Unterschiede bei der Regulation der Apoptose sind (epi-) genetisch festgelegt. Die Identifizierung von SNPs in der kodierenden/ flankierenden Region von Genen mit Apoptose-relevanter Funktion könnte daher dazu führen, das individuelle Risiko für die Entstehung solcher Läsionen bewerten zu können.

Um einen möglichen Zusammenhang zwischen genetischem Polymorphismus und Tumorgenese herstellen zu können, wählten wir 865 Gene aus, die nachgewiesenermaßen entweder eine Rolle im Apoptose-Signalweg oder einem verwandten (z.B. stressverbundenen) Signalweg spielen.

Unsere Untersuchungen wurden auf einem maßgefertigten Goldengate Illumina Chip mit 1536 Einzelnukleotid-Polymorphismen in zwei Schritten durchgeführt. In Schritt 1 wurden 272 Patienten mit hyperplastischen Polypen sowie 512 Kontrollen untersucht. Das Ziel von Schritt 2 war es, die vorläufigen Daten durch Screening der „Kandidaten"-SNPs ($p<0,01$) an einer unabhängigen Patienten- und Kontrollkohorte ($n=76$) zu validieren. Nach der Meta-Analyse der Daten von Schritt 1 und 2 wurde durch einen „false discovery rate"-Ansatz (FDR) der Signifikanz anzeigende „q-Wert" berechnet, um einen besseren Ausgleich zwischen Typ1 und Typ2 Fehlern zu erreichen.

Am Ende konnten 9 SNPs signifikant mit dem Auftreten von hyperplastischen Polypen in Zusammenhang gebracht werden. Von diesen acht Kandidaten zeigte der SNP rs4709583 (PARK2) das höchste Signifikanzlevel ($q=0,003$). Dieser Polymorphismus wurde daher zusätzlich *in-vitro* untersucht, um seinen möglichen Einfluss auf das Splicing des *PARK2* Gens zu bestimmen.

Abstract

Deregulation of apoptosis is a frequent alteration in benign, pre-cancerous lesions of the colon mucosa that has been extensively discussed as contributor to the development of colorectal cancer. Individual differences in the regulation of apoptosis are (epi-)genetically determined and identification of SNPs within coding/ flanking regions of genes with apoptosis relevant function could provide a basis to assess the individual risk to develop these lesions.

To identify a possible association between genetic polymorphisms and tumorigenesis we selected 865 genes with reported function within the apoptosis pathways or in related (e.g. stress related) pathways. Our screening was performed on a customized goldengate Illumina chip covering 1536 single nucleotide polymorphisms in a two stage approach. Stage I was performed on 272 patients harboring hyperplastic polyps and 512 controls. Stage II aimed to validate preliminary data by screening the candidate SNPs ($p<0.01$) on an independent cohort of patients and controls ($n=76$). After the meta-analysis between stage1 and stage2 the false discovery rate (FDR) approach has been used to calculate the "q" of significativity in order to have a better compromise between Type1 and Type2 errors.

Among the candidate polymorphisms 9 SNPs were significantly correlated with the occurrence of hyperplastic polyps. Out of these the variation rs4709583 (PARK2) presented the highest significativity level ($q=0.003$) and was therefore further analysed in-vitro to identify its potential influence on splicing in the *PARK2* gene.

INTRODUCTION

Epidemiological aspects

Colorectal cancer (CRC) represents one of the most common cancers worldwide and is the second cause of cancer-related deaths in Germany (Figure 1). With a lifetime risk of 5-6% CRC is one of the most common cancers in the Western world[1].

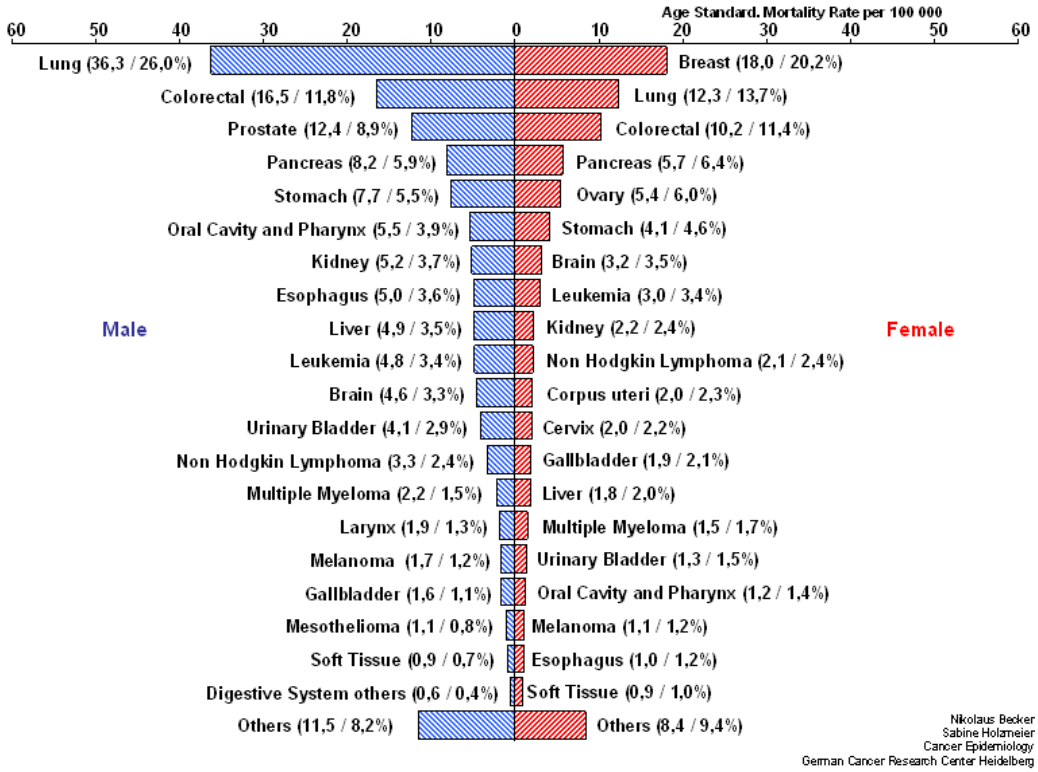


Figure 1: Causes of cancer-related deaths in Germany; the bars represents the standard ages divided by gender

CRC is a complex disease, where an interaction of genetic factors, life style factors and the environment occur in its aetiology. Over the past decades prevention of colon cancer has become a large public health issue. In fact, it has been shown that serial faecal occult blood testing and sigmoidoscopy have the potential to reduce colon cancer-related mortality. The possibility to prevent the outcome of the disease is therefore much higher in colorectal cancer than in many other neoplastic conditions in humans[2].

Among the common malignancies, colorectal cancer has one of the largest proportions of familial cases. It has been shown that approximately 30% of all CRC cases are familial forms of the disease, of which around 5% are associated with highly penetrant inherited mutations. The aetiology of the

remaining is still not completely understood. They are likely to be caused by alterations in single genes with lower penetrance but are more common than those associated with the well-characterized syndromes, like common polymorphisms in genes that regulate metabolism or in ones that are regulated by environmental factors [3].

There are several known risk factors contributing to the development of CRC. These can be divided in two main groups: factors that cannot be changed, and risk factors linked to the behaviour and life style. The first group includes factors like age, gender, ethnic background and genetic alterations (like the presence of polyps in the colon, having history of bowel disease or family history of CRC). As an evidence of correlation between age or genetic predisposition and colorectal cancer, it has been shown that more than 9 out of 10 people found to have CRC are indeed older than 50 years old and that some type of polyps can increase the risk of CRC, especially if they are large and in high number. All these aspects are considered as "environmental factor", since they do not depend on single person behaviour. In contrast to these risk factors, there are some that are influenced by personal behaviour and are considered "life styles factors", comprising factors like diet, alcohol, smoking, severe obesity and a general lack of exercise. As incremental risk factors for developing of CRC, it has been shown that meat cooked at high heat, frying and barbecue could create chemicals that increase the chance of developing of cancer. Moreover, it has been shown that tobacco could increase formation of polyps in the colon. Conversely, some factors like a diet rich in fibres or vitamin supplements, especially folate or mineral intake are capable to reduce the risk of cancer formation. Specifically, calcium can bind to fatty acids and bile acids further reducing the risk. All these aspects modulate the risk of carcinogenesis in the colon [4, 5].

It is very well known that colorectal tumors progress through a well recognised series of clinical and histopathological stages. These range from very early microscopic lesions to benign neoplastic lesions with increasing malignant potential to full-fledged carcinomas. Morphology and size as well as molecular characteristics are successfully used for clinical and molecular classification of these stages.

Clinical and morphological factors of non-malignant colon lesions

A colon polyp is an abnormal growth in the mucosal surface of the colon, projecting from the mucosal cell layer. Polyps can be divided in three different types: inflammatory, hyperplastic and adenomatous polyps.

The inflammatory polyps (also defined as pseudopolyps) are the reaction to a chronic inflammation of the colon wall. They are not cancerous and are frequent in patients affected with ulcerative colitis and Crohn's disease.

The hyperplastic polyps (nondysplastic polyps) are the most frequently encountered lesions and represent approximately 90% of all polyps. They are rather small, with a size less than 5mm in diameter.

The adenomatous polyps (dysplastic polyps) with a frequency around 10% are usually small in diameter (less than 1cm), but a small percentage can be also larger.

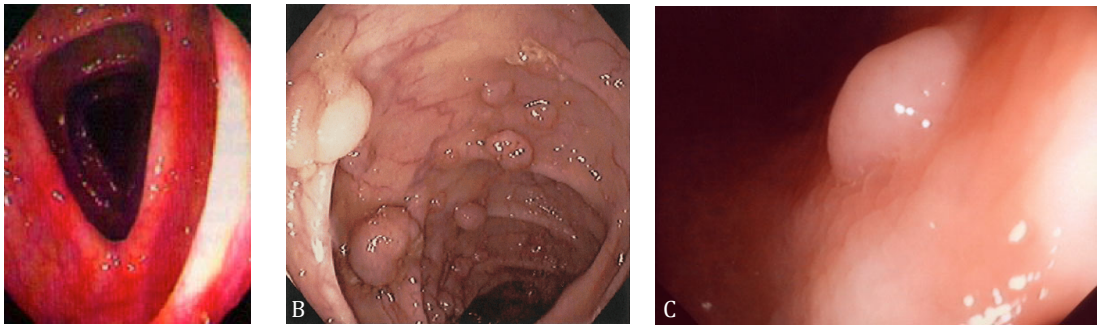


Figure 2: (A) Normal Mucosa; (B) Adenoma; (C) Hyperplastic Polyps

Although both types of polyps (hyperplastic and adenomatous) often appear identical at a macroscopic level, dysplastic polyps are distinguished by disruption of the colonic epithelial architecture and cytology and have a well-documented tendency to progress – through molecular transition - to malignancy, the risk of which increases with size and histological abnormalities. In contrast, hyperplastic polyps are thought to rarely progress[6].

Histologically, there are profound differences between hyperplastic polyps, adenomatous polyps and cancer tissue. In hyperplastic polyps, the morphological findings include crypt elongation and branching, an increased number of cells, a high number of goblet cells and regular shaped nuclei. However, the general architecture is unchanged, and cells show a normal mucosal differentiation. In contrast, in the adenoma there is a clear loss of regular architecture of the crypts, changes in cellular morphology like elongated nuclei, loss of goblet cells and a dense locking mucosal layer (Figure 3).

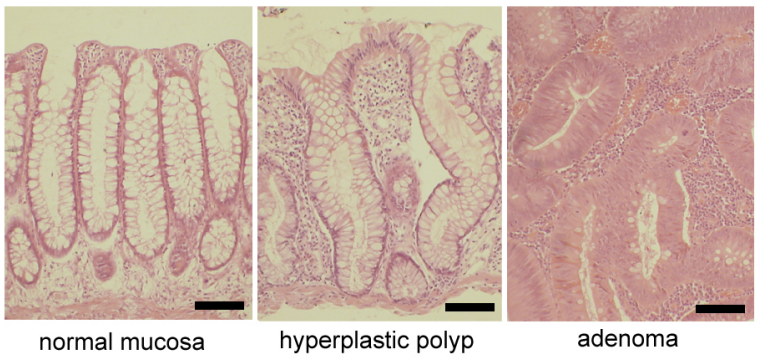


Figure 3: H&E staining of human colon mucosa. Note the reduced number of goblet cells. Scale bar 100µm

Normal colon mucosa (NM) of the gastrointestinal tract notoriously develops hyperplastic aberrant crypt foci (ACF) and hyperplastic polyps (HP)[7]; It has been shown that if a *KRAS* gene mutation has a chance to occur as first genetic event, a nondysplasic polyp will form which has little potential to progress. In contrast, if a mutation in the adenomatous polyposis coli (APC) gene occurs first a dysplasic polyp will result[6]. However, hyperplastic polyps rapidly regress because of the normal apoptosis control without the chance for gatekeeper defects to occur[8].

It has been shown that over 50% of the population will develop an adenomatous polyp by the age of 70, but only 1/10 of these polyps will proceed to carcinoma[9].

Genetic and epigenetic models of tumor development

It is accepted by now that the majority of colorectal cancers arise in pre-existing adenomas, and that a high percentage of adenomas is initiated by defects in the Wnt-signalling pathway. These are in most cases caused by mutation in the tumor suppressor gene *APC* (figure 4) and subsequent loss of heterozygosity. However, it was not immediately obvious that colorectal cancers would have been so diverse genetically. A systematic approach to molecular classification of the underlying genetic defects resulted in the concept of multistep carcinogenesis proposed by Fearon and Vogelstein in 1990.

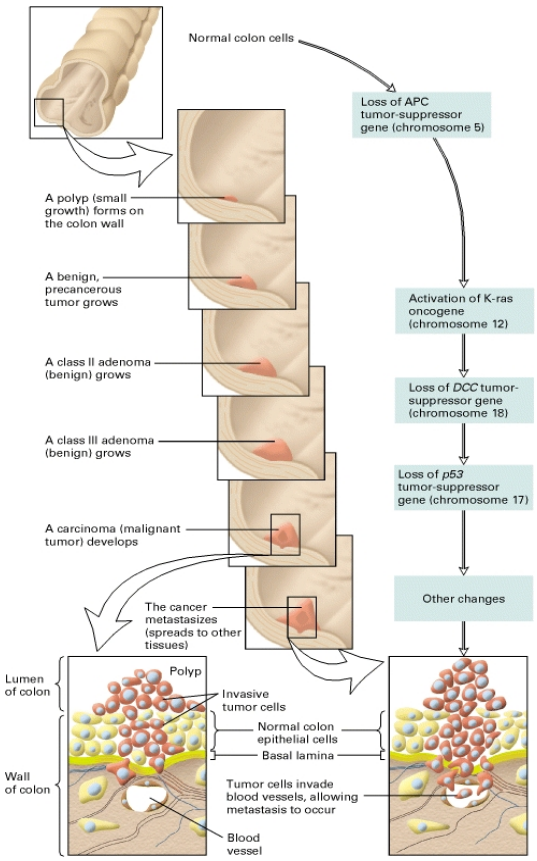


Figure 4: Model of tumor development; Molecular Cell Biology, 4th edition; Lodish H. et al; New York: W.H. Freeman; 2000

According to this model[8, 10], colorectal cancer is caused by an accumulation of genetic aberrations that need to occur in a particular order for a full-fledged cancer to develop. First step involves the defect of the adenomatous polyposis coli (APC)/Wnt signaling pathway observed in approx. 80% of CRC[11, 12]. Several studies indicate APC as molecular gatekeeper. APC (Adenomatous polyposis coli) is a protein encoded by the *APC* gene involved in many processes like cell migration, adhesion and apoptosis. The APC protein is a tumor suppressor acting as an antagonist of the growth-promoting Wnt signaling pathway. Its inactivation initiates neoplastic growth, a chromosomal instability phenotype and eventually leads to polypous CRC[13-15]. Developing of the neoplastic polyps along this pathway is due to mutation in the *APC* tumor-suppressor gene in a single epithelial cell. This mutation is one of the events that lead the cell to divide, forming a mass of localized benign cells, the adenomatous polyp. Following further critical defects within the multistep cascade, chromosomal instability (CIN) ensues. Eventually, malignant transformation occurs with newly mutated cells then transgressing the surrounding basal lamina, invading the surrounding tissue and vessels and finally leading to metastasis.

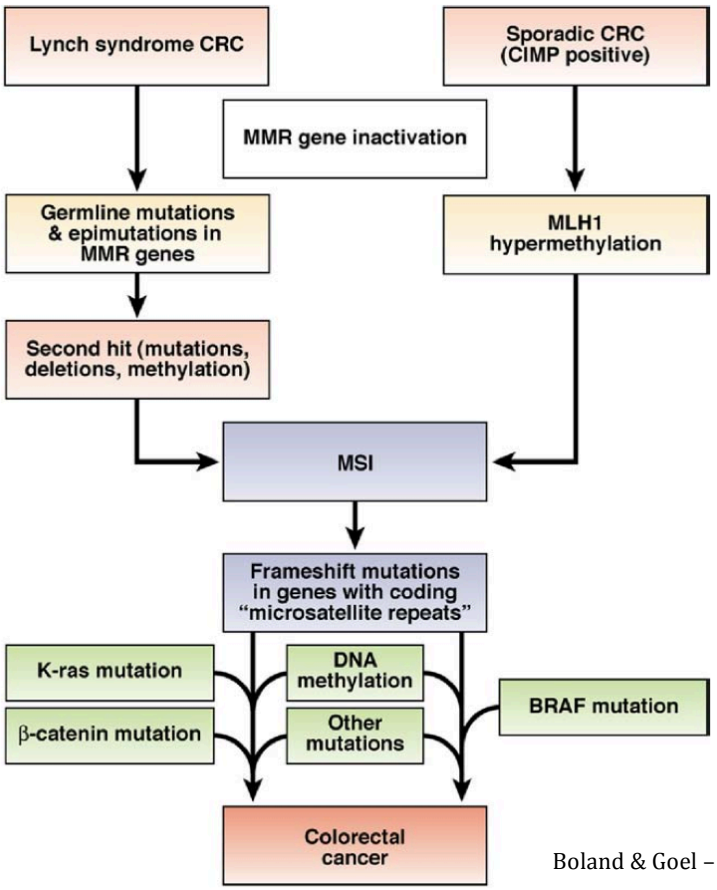
However, this model alone does not completely explain the complexity of early colorectal tumour development. There are two main pathways operative in the evolution of polyps to sporadic colorectal cancer.

As detailed above, the CIN pathway is the most common pathway for colorectal cancer, and involves chromosomal karyotypic abnormalities as well as pivotal mutations in specific oncogenes and tumor suppressor genes (*APC, TP53, SMAD4*).

The Microsatellite instability (MSI) pathway, which results in a hypermutable phenotype due to the loss of DNA mismatch repair activity, is responsible for15-20% of colorectal cancers. The tumour suppressor functions of mismatch repair genes (MMR) are inactivated in the hereditary form, i.e. the Lynch syndrome, or their activity is lost due to mutational defects or hypermethylations, thereby leading to the sporadic cases of MSI colon cancers. Particularly, epimutations have recently recognized to play a crucial role in the aetiology of colorectal tumour development[16].

Specifically, about half of the genes in the human genome have promoters that are embedded in clusters of cytosine-guanosine dinucleotide residues called CpG islands. Cytosines in CpG dinucleotides can be methylated by DNA methyltransferases leading to transcriptional silencing of the respective genes, e.g. tumour suppressor genes. As shown below, the so-called CpG island methylator phenotype (CIMP) has been proposed as an entity that represents the origin of the sporadic CRC as

characterized by MSI[16]. Defect in mismatch repair activity result in a further rapid accumulation of somatic mutations.



Boland & Goel – Gastroenterology V138, 2010

The last pathway, but whose reliability is still under debate, is the so-called "serrated pathway", which involves the transformation of sessile serrated polyps and traditional serrated adenomas to colorectal cancer. The characteristic feature of all serrated polyps is the "saw-toothed" infolding of the crypt epithelium and the term "serrated adenoma" indicates polyps showing a mixture of features of hyperplastic and adenomatous polyps[16].

Genome wide & association studies: different approaches

A maximum of 5% cases of colorectal cancer are caused by FAP, Lynch syndrome (caused by mutation in DNA mismatch repair genes) or the recessive MUTYH-associated polyposis (MAP). These genetic alterations are inherited in a monogenic mendelian fashion. In addition, up to 30% of the CRC risk is thought to be due to familial susceptibility inherited in a polygenic fashion. This can be explained by the occurrence of genetic variants of lower penetrance.

Initially, to identify these variations, linkage studies have been performed on cancer-affected families or were based on candidate gene approaches to search for associations between more common polymorphic variants and specific disease. The hypothesis was that either the discovered variant itself was disease causing, or that it was in linkage disequilibrium with the disease-causing variant (candidate association study). However, the advent of high-throughput technologies and the advance of genome-wide association studies (GWAS) have revolutionized research on genetic determinants of risk for common diseases.

GWAS studies are based on the tagging SNP model and have been used for the identification of low-penetrance genetic factors to explain the cancer cases in families, in which leading genetic causes have not been found (so-called heritability gap). For example, a recent multicenter study has shown that an accumulation of 20 different risk alleles would increase the odd ratio for colorectal cancer[17] (Figure 5). Independently, a linear association between the number of risk alleles and familial CRC has been identified showing that for each one risk allele increase, the odds of having familial cancer is increased by a factor of 1.16 (Figure 6) [13].

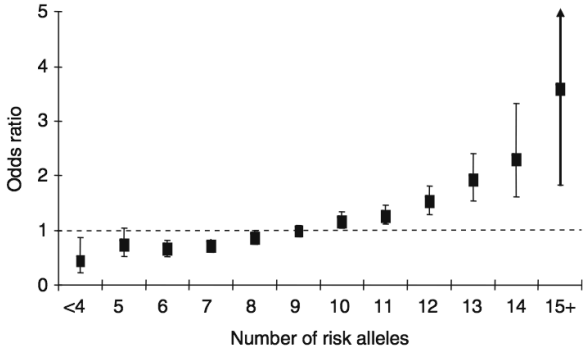


Figure 5: Plot showing the increase ORs for CRC associated with increasing number of risk alleles. (Tomlinson et al., 2010)

Figure 6: Odds ratio for familial cancer by number of risk alleles

Number of risk alleles	Familial cases n(%)	Sporadic cases n(%)	Odds[A]	OR if 10 ref.*
15	3 (3.1)	4 (0.6)	0.2	2.15
14	4 (4.1)	15 (2.2)	0.18	1.84
13	1 (1.0)	44 (6.4)	0.15	1.58
12	16 (16.5)	69 (10.0)	0.13	1.36
11	27 (27.8)	121 (17.5)	0.11	1.17
10	12 (12.4)	115 (16.6)	0.1	1
9	13 (13.4)	135 (19.5)	0.08	0.86
8	14 (14.4)	100 (14.5)	0.07	0.74
7	5 (5.2)	49 (7.1)	0.06	0.63
6	1 (1.0)	23 (3.3)	0.05	0.54
5	0	11 (1.6)	0.04	0.47
Total	97	691	OR=1.16 (1.04-1.30) per allele P_{Trend}=0.006	

[A]=Adjusted with age at diagnosis

*Odds ratio if 10 risk alleles (the most common number in all cases) set as a reference group

However, there are limitations in this approach concerning the SNP analysis. In all these loci, the best SNP markers exhibit very modest odds ratios for CRC predisposition themselves[18]. Even with large sample sizes of several thousands cases and controls, there is usually limited power to detect alleles of modest effect size (with odds ratio <1.20) and minimal power to detect risk allele odds ratios of <1.10 even for very common variants. GWAS studies provide pieces of information on common variants in the general population and cannot be compared with association studies.

If one compares the genetic susceptibility or risk loci identified in mendelian cancer syndromes, family studies, in pedigree analyses and finally in GWAS studies, it becomes apparent that these loci bear no resemblance between each other. Indeed, with one exception they show no congruence (Figure 7).

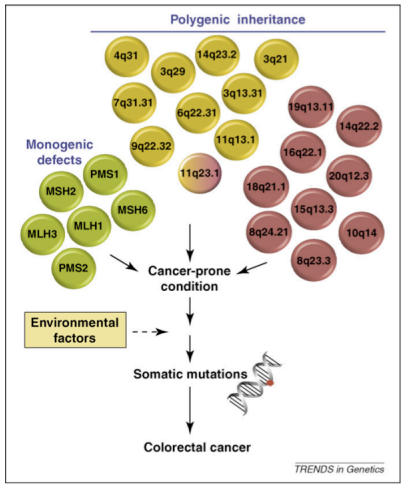


Figure 7: Several causative mutations at single genes (monogenic syndromes in green), as well as variations at several loci detected by association study (in yellow) or GWAS (in red) increase the risk of cancer. Only the 11q23.1 locus was common to association and GWA studies

The permissive mutational model

Classification of early premalignant colorectal lesions can be obtained on the basis of their morphology. Dysplastic (adenomatous) or non-dysplastic (hyperplastic) polyps often appear identical at a macroscopic level, but they present with different molecular defects. Jen et al.[7] already showed that it is possible to distinguish these different lesions on the basis of mutations in specific genes. These can therefore provide additional insight into the development of the lesions in respect to the model proposed by Fearon and Vogelstein.

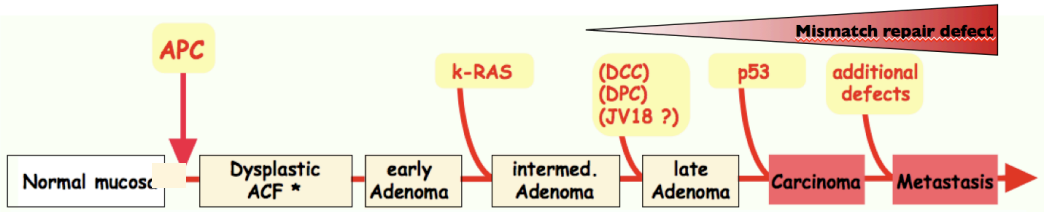


Figure 8: Model from Fearon & Vogelstein, 1990

Jen et al. could first verify that no *APC* mutations were identified in any of the non-dysplastic lesions examined, suggesting that the mutations of APC are closely associated with the advent of dysplasia. This has also been recently confirmed by Nittka et al., who compared Wnt-Status in microscopic and macroscopic hyperplasias and adenomatous polyps. Nittka et al (2004) propose that hyperplasia generally lack defects of the APC gatekeeper and therefore will not develop into neoplasia, while they present with tumour-associated gene defects. In contrast, Jen et al. propose a model where if the first genetic event occurring is a *RAS* gene mutation, non-dysplastic aberrant cript foci (ACF) or hyperplastic polyp will form and eventually regress. However, if an APC mutation occurs first, dysplastic ACF will result, which will then be able to progress to adenoma by subsequent mutations in *RAS* and other genes [7] (Figure 9).

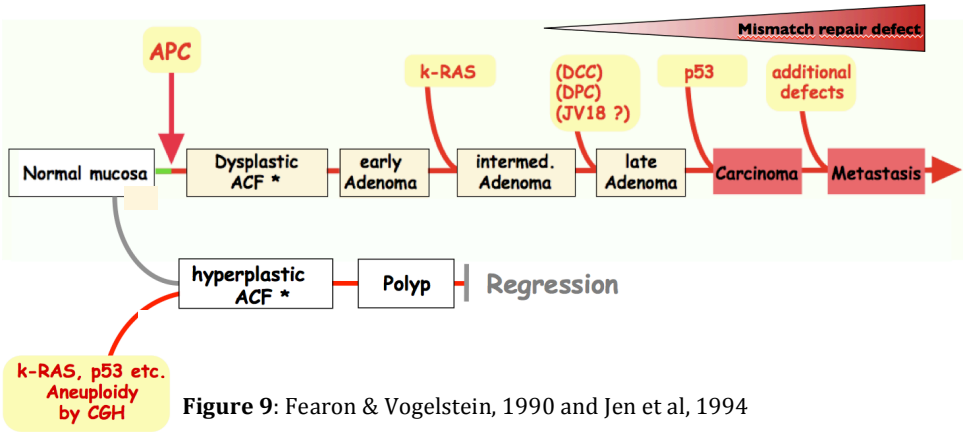


Figure 9: Fearon & Vogelstein, 1990 and Jen et al, 1994

New light on the events that precede the genetic alterations known to contribute to the development of colorectal cancer has been shed by the observation of a down-regulation of a tumor suppressor protein (CEACAM1) in hyperplastic lesions of the colon, observed by Nittka et al[19]. The authors present evidence that hyperplastic polyps and ACF have the same reduced levels of expression of the tumor suppressor protein (and inducer of apoptosis) CEACAM1 as found in adenomas and adenocarcinomas.

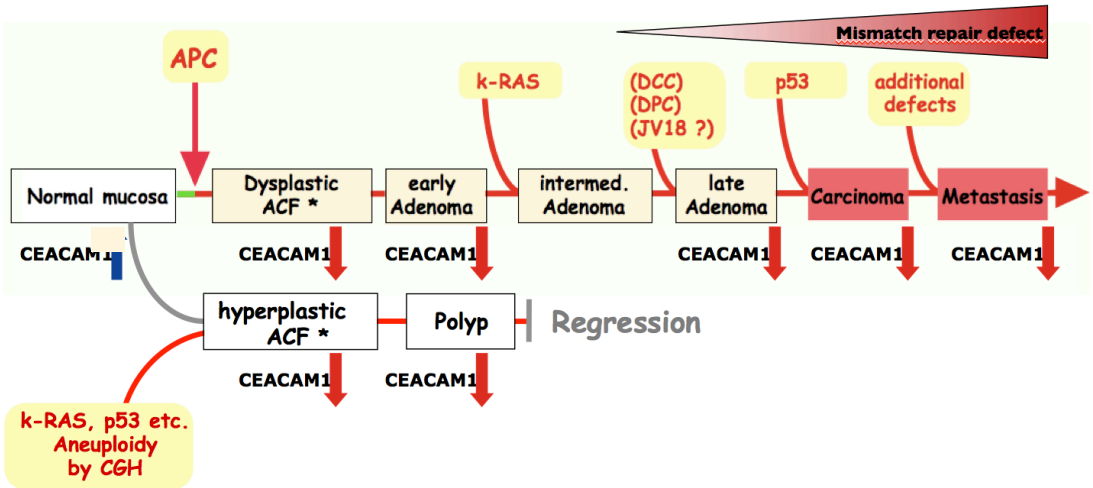


Figure 10: Loss of CEACAM1
Neumaier M et al., PNAS (1993)
Nollau P et al., AJP (1997) CR (1997)

ACF and hyperplastic polyps have therefore to be considered as significant pre-neoplastic lesions that may be involved in carcinogenesis as first steps at a preneoplastic level (Figure 10). Loss of CEACAM1 expression would lead to reduced levels of apoptosis, thus leading to an increase of the lesion cells. Even if the majority of the hyper-proliferative cells regress due to incapacitating mutations, a small portion may contract oncogenic mutations leading to cancer development.

In an Oncogene editorial, J.E. Shively[20] discussed this hypothesis, which argues that the loss of CEACAM1 expression contributes to the development of hyperplastic polyps. Central to this hypothesis is the observation that CEACAM1 expression modulates differentiation-dependent apoptosis. Furthermore, colon crypts devoid of CEACAM1 showed reduced apoptosis, while the few crypts still expressing CEACAM1 also showed unaffected apoptosis.

Shively pointed out that the growing hyperplastic polyp cells would lead to a changed crypt architecture that may expose its stem cell compartment to a wide variety of mutagens in the stool.

It is widely known that hyperplastic polyps harbor a plethora of genetic and epigenetic defects. However, none of the defects in so-called gatekeeper genes like *APC* or others are observed that would cause entry into a transition to colorectal neoplastic lesions.

The revised model proposed by Nittka et al. and shown in Figure 11 is consistent with this observation and defines two different kinds of genetic mutations (permissive or non-permissive) that may occur in the hyperplastic polyps, thus leading to two different pathways.

The various genetic defects that are most frequently encountered in HP supposedly lead to the regression of hyperplasia and are termed “non-permissive” mutations, thereby leading back to normal mucosa tissue.

In contrast, when a “permissive mutation” i.e. a gatekeeper mutation occurs, transition to neoplasm would occur within the HP, resulting in the formation of an adenoma. As a facultative precancerous lesion, the adenoma could then evolve in carcinoma and eventually form metastases.

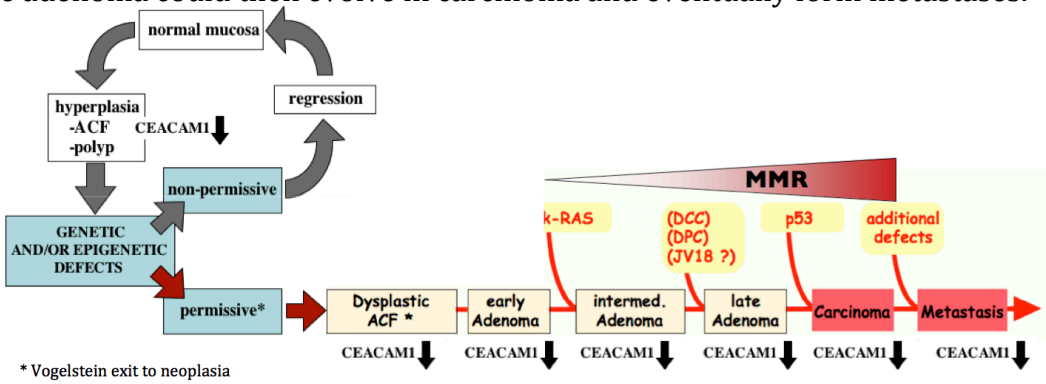


Figure 11: *Nittka S et al., Oncogene (2004)*

Thus there are two different pathways that originate from a common background, the hyperplastic polyp.

AIM OF THE STUDY

The study presented here originates from a clinical investigation (in preparation) that observed a significant difference in apoptosis response between 1) lymphocytes from patients bearing hyperplastic polyps and 2) lymphocytes from healthy control individuals. Indeed, patients bearing hyperplastic polyps showed slower and reduced responses to a set of standardized apoptosis stimuli. Together with the known fact that HP development is associated with reduced apoptosis in the colon mucosa, the results in these lymphocytes suggested a genetic cause for these differences.

This project detailed below focused in an "a priori" approach, where candidate genes involved (in the widest sense) in apoptosis response and regulation were selected for analysis of common and rare variants in order to evaluate potentially different prevalences compared to controls. This approach verified as the first association study on non-neoplastic colon tumors, would maintain an overall high power and allow the analysis of low odds ratio variations.

Objective of this thesis was to study Single Nucleotide Polymorphisms (SNPs) of different genes directly involved in apoptosis or in apoptosis related pathways, and to associate them to strong or weak levels of apoptosis. This study has been performed in two cohorts with a cases-controls approach. The "cases" population was composed of patients that, after colonoscopy, would present only hyperplastic polyps, while the control group was composed of people with no defects in the colon. My goal was to be able to identify markers associated with apoptosis efficiency, in order to use them in a clinical setting as marker for early stages of colorectal cancer.

MATERIALS & METHODS

Patients & Controls

Study population

Participants undergoing screening colonoscopy (a procedure that the German health-care system offers since October 2002 to average-risk individuals aged ≥ 55 years old as part of the general health plan) were used as a cohort to identify bearers of hyperplastic polyps. A total number of 310 patients and 550 controls were analysed in this study. Out of the 860 samples, 784 where representing the first stage of our study and 76 have been included into the second stage.

736 samples (224 patients and 512 controls) were recruited through the BliTz (BegleitendeEvaluierunginnovativerTestverfahrenzurDarmkrebsfrüherkennung) study, an ongoing screening project conducted since 2006 in cooperation with 20 gastroenterology private practices in southwestern Germany that aims to comparatively evaluate new tests for the early detection of CRC. The BliTz study contributed with 72% of the patient cohort and provided all the controls age-sex matched for the whole study.

The department for Endoscopy of the Surgery Dept. of the Universitätsmedizin Mannheim (Dr. Kähler) provided 48 patients used in this study.

76 samples (38 patients and 38 controls) required for the validation study were provided by an external department (Prof. H.K Seitz, Department of Medicine, Salem Medical Centre Heidelberg, Lab. Alcohol Research, Liver Disease and Nutrition).

Sample collection

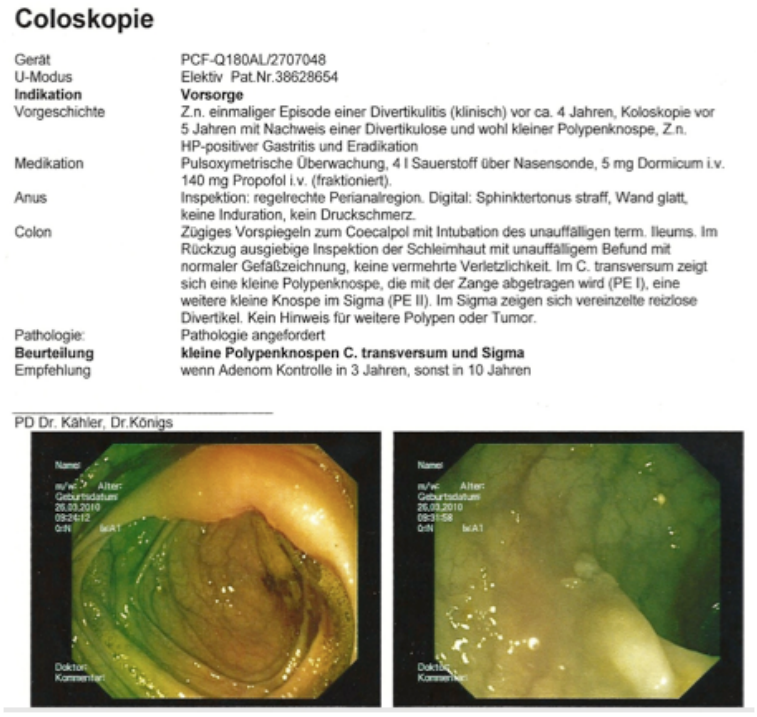
Prior to the start, the study was approved by the Ethics committee of the Medical Faculty Mannheim of the University of Heidelberg. Every patient signed an informed consent and completed a questionnaire to collect medically relevant context information. Questions focused on dietary habits, physical exercise and history of specific familial background (for details, see supplementum).

We only included patients with confirmed hyperplastic polyps following the colonoscopy. Patients with neoplasms i.e. presenting with colorectal adenomas of any stage or with CRC or reporting a previous history of cancer were excluded. Patients with incomplete colonoscopy reports, other gastrointestinal pathologies like inflammatory bowel disease, pseudopolyps or other lesions also were excluded.

Patients with no sign of polyps and no further pathological finding during colonoscopy were recruited in the control population.

Polyps identification

Identification of polyps was carried out during colonoscopy. EDTA blood was taken during this procedure from all patients. To classify samples, we established three different categories, (patients/control/excluded) according to the colonoscopy record (Figure 12).



Coloskopie

Gerät	PCF-Q180AL/2707048
U-Modus	Elektiv Pat.Nr.38628654
Indikation	**Vorsorge**
Vorgeschichte	Z.n. einmaliger Episode einer Divertikulitis (klinisch) vor ca. 4 Jahren, Koloskopie vor 5 Jahren mit Nachweis einer Divertikulose und wohl kleiner Polypenknospe, Z.n. HP-positiver Gastritis und Eradikation
Medikation	Pulsoxymetrische Überwachung, 4 l Sauerstoff über Nasensonde, 5 mg Dormicum i.v. 140 mg Propofol i.v. (fraktioniert).
Anus	Inspektion: regelrechte Perianalregion. Digital: Sphinktertonus straff, Wand glatt, keine Induration, kein Druckschmerz.
Colon	Zügiges Vorspiegeln zum Coecalpol mit Intubation des unauffälligen term. Ileums. Im Rückzug ausgiebige Inspektion der Schleimhaut mit unauffälligem Befund mit normaler Gefäßzeichnung, keine vermehrte Verletzlichkeit. Im C. transversum zeigt sich eine kleine Polypenknospe, die mit der Zange abgetragen wird (PE I), eine weitere kleine Knospe im Sigma (PE II). Im Sigma zeigen sich vereinzelte reizlose Divertikel. Kein Hinweis für weitere Polypen oder Tumor.
Pathologie:	Pathologie angefordert
Beurteilung	**kleine Polypenknospen C. transversum und Sigma**
Empfehlung	wenn Adenom Kontrolle in 3 Jahren, sonst in 10 Jahren

PD Dr. Kähler, Dr.Königs

Figure 12: Diagnosis used to identify patients and controls.
In this case the patient was affected by hyperplastic polyps only, and had no history for cancer, therefore he has been included in our case group.

Sample Preparation: Patients and controls were recruited from three different sources: the BLITZ study, the Endoscopy department of the Universitätsmedizin Mannheim and the Department of Medicine of the Salem Medical Centre Heidelberg.

a) The BLITZ study provided 6µl [10ng/µl] of DNA for each patient and control individual. Starting material for the amplification was 25ng of purified genomic DNA template. This DNA was subjected to whole genome amplification (WGA) using the REPLI-g® Midi Kit from

Qiagen according to the manufacturer´s protocol. The procedure started with the lysis of the sample material and subsequent denaturation of the DNA by adding of a specific denaturation buffer. After 3 minutes of denaturation, a neutralization buffer stopped the reaction and a master mix containing buffer and DNA polymerase was added. The isothermal amplification reaction was run for 16 hours at 30°C. This procedure provides a highly uniform amplification across the entire genome using a DNA polymerase capable of replicating up to 100kb without dissociating from the genomic DNA template. The polymerase has a 3'-5' exonuclease proofreading activity to maintain high fidelity during replications. WGA secured that sufficient DNA was available for subsequent array experiments.

b) DNA was isolated from EDTA blood using the AGOWA mag maxi kit (Agowa, Germany), which uses surface-modified superparamagnetic particles that bind nucleic acids via a classic polarity-based binding mechanism. The magnetic separation method consists in four different steps: 1) Cell lysis in 200µl of whole blood; 2) Binding of the released nucleic acid to the superparamagnetic beads; 3) Washing of the nucleic acid/particles complex to remove impurities and contaminants; 4) Elution of the purified DNA from the beads. The whole procedure was automated on a PerkinElmer MultiPROBE® II PLUS Ex Liquid Handling System. Eluted DNA was then amplified via Qiagen REPLI-g® Midi Kit.

c) DNA from Formalin-fixed-Paraffin-embedded (FFPE) tissue was purified with the QIAamp® DNA FFPE Tissue Handbook from Qiagen according to manufactories protocol. Briefly, the procedure consists in 6 steps: 1) Removal of paraffin by dissolving it in xylene; 2) Lysis of the sample by Proteinase K digestion; 3) Incubation at 90°C to reverse formalin crosslinking; 4) Binding of the DNA to the membranes with non-binding contaminants flowing through; 5) Washing off the residual contaminants from the membranes; 6) Elution of purified and concentrated DNA from the membrane. Eluted DNA was then amplified via Qiagen REPLI-g® Midi Kit.

Following whole genome amplification of each DNA, randomly chosen samples were analysed via human-specific multiplex short tandem repeats (STR) profiling which includes 13 polymorphic loci on 10 chromosomes to verify possible cross-contamination. No signs of DNA contamination where found.

DNA measurement using Picogreen staining

Each DNA sample was measured spectrophotometrically and adjusted to a minimum working concentration of 50ng/µl.
The Picogreen Kit from Invitrogen (Quant-iT™ PicoGreen® dsDNA Assay Kit *2000 assays*) was used in combination with the Tecan Infinite F200 multimode reader to measure the fluorescence intensity of the assay.
All the samples presented a high DNA concentration with an average > 120ng/µl.

DNA extraction from cells

A total of 38 different tissue cell lines were thawed and grown in culture using RPMI supplemented with 10% FCS and 1% Penicillin at 37° with 5% CO_2. Isolation of genomic DNA from cultured cells was performed using the Wizard® Genomic DNA purification Kit (Promega).
Briefly, after centrifuging and washing, the cells were lysed with Nuclei Lysis Solution and then mixed by pipetting. RNase solution was added to the lysate, and protein precipitation was then performed using a protein precipitation solution (provided in the kit). The supernatant containing the DNA was then moved to a fresh tube containing isopropanol, centrifuged for DNA precipitation and the supernatant discarded. Precipitated genomic DNA was washed with 70% ethanol in water, air dried and dissolved in 100µl DNA Rehydration Solution.

RNA purification

Purification of total RNA from tissue culture cells was performed via the RNeasy Mini Kit (250) from Qiagen. This kit provides a fast and simple method for preparing up to 100µg total RNA per sample with a technology that combines the selective binding properties of a silica-based membrane with the speed of microspin technology. The purified RNA is ready to be used for cDNA synthesis.
$1x10^7$ cells per sample were lysed with Buffer RTL and homogenized by sonication. The complete disruption of cells walls, plasma membranes and organelles is required to release all the RNA contained in the sample, while homogenization is necessary to reduce the viscosity of the lysates produced by disruption. To allow the selective binding of RNA to the RNeasy membrane, ethanol has been added to the lysate. The samples were then moved to the RNeasy Mini spin column. RNA was then eluted withRNase-free water analyzed via Nanodrop and then frozen to -80°C.

RNA reverse transcription

Purified RNA was reverse transcribed into cDNA as follows: 5µg of total RNA (concentration

measured via Nanodrop) was mixed with 1ul oligo(dT)$_{18}$ primer in nuclease-free water to a final volume of 11ul. The following components were then added:
5x reaction buffer 4ul – RiboLockRNase Inhibitor (20u/µl) 1µl – 10mM dNTP Mix 2µl – M-MuLV Reverse Transcriptase (20U/µl) 2µl. The samples were incubated for 60' at 37°C. Subsequently, the reaction was stopped by incubating for 5' at 70°C. Nucleic acids concentration was subsequently measured by spectrophotometry using Nanodrop.

Real Time RT-PCR analysis

Realtime RT-PCR analysis was performed using a LightCycler to quantitate mRNA expression levels of PARK2. PBGD was used as housekeeping gene as and internal control.

In a 1.5ml reaction tube placed on ice, the RT-PCR Mix was prepared by adding

8.2µl of Water with Mn(OAc)$_2$ 50mM,
1.3µl - PCR primer mix 10x concentrated
2.0µl – LightCycler® RNA Master SYBR Green 2.7x concentrated
7.5µl for a final volume of 19µl.
19µl – total volume

These 19µl of RT-PCR mix have been pipetted to a pre-cooled LightCycler Capillary and 1µl of RNA template was added. After brief centrifugation, the capillary was inserted in the LightCycler Sample Caroussel and then into the LightCycler Instrument ready to be cycled.

Chip Design

GoldenGate® assay

The GoldenGate assay is a custom DNA array that allows a high degree of locus multiplexing (1536-plex) during the extension and amplification steps. Use of this technology allows the analysis of many polymorphisms and patients at the same time. The assay runs on the BeadArray^{tm} platform for 1.536-plex in multiples of 96.

The Golden Gate array designs were carried out with the support from the manufacturer. The process was initiated by selecting and submitting a list of requested loci (refer to "Gene and SNPs selection chapter" for the criteria used in the selection process) to Illumina. Upon submission, Illumina

evaluated the list using the Illumina Assay Design Tool (ADT) to ensure successful assay development (Figure 13).

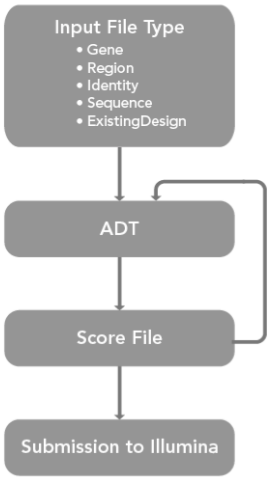


Figure 13: Custom GoldenGate genotyping assay design workflow

The list of genes has been formatted in comma-separated values (*.csv) format.
By submitting a Gene list we could query all loci within genes of interest and in the regions upstream and downstream from the indicated genes.
HUGO identifiers were used to interrogate currently supported build of the human genome (build 36). We decided in order to include eventual promoters and splicing regions, to include 1.000 bases in upstream and downstream regions in the query to the Assay Design Tool (Figure 14).

21	Gene_Name	Bases_Upstream	Bases_Downstream	Species
22	AATF	1000	1000	Homo Sapiens
23	ABCC8	1000	1000	Homo Sapiens
24	ABL1	1000	1000	Homo Sapiens
25	ACIN1	1000	1000	Homo Sapiens
26	ACVR1C	1000	1000	Homo Sapiens
27	ADAM12	1000	1000	Homo Sapiens
28	ADAMTSL4	1000	1000	Homo Sapiens
29	AES	1000	1000	Homo Sapiens
30	AIFM1	1000	1000	Homo Sapiens
31	AIFM2	1000	1000	Homo Sapiens
32	AIFM3	1000	1000	Homo Sapiens
33	...	...	...	...

Figure 14: excerpt from the .csv file submitted to the ADT

ADT generates the score output file, providing an important set of informative metrics for each scored locus requested. These metrics were used to preferentially select the assays that have a higher likelihood for success in the final product design.

Heading	Description
SNP_Name	RS unique universal identifier SNP code
Sequence	The bracketed site identified by the SNP_Name with >50 bases of flanking sequence
Genome_Build_Version	Genome build that was queried (36)
Chromosome	Chromosome on which the SNP is located
Coordinate	Chromosome coordinate of SNP
Source	Identify the source of the sequence and annotation data
Sequence_Orientation	Forward or Reverse
Region_Description	Description of the region of interest
Final_Score	Ranges from 0 to 1.1, with higher values reflecting greater ability to design a successful assay
Failure_Codes	If applicable, reasons why a successful assay at this SNP is unlikely
Validation_Class	Numerical representation of validation_bin
Validation_Bin	Manner in which designed assays have been validated
MAF_Caucasian	Minor allele frequency from the largest peer-reviewed study conducted, the study size in terms of number of chromosomes and the study type.
Chr_Count_Caucasian	
Gene_ID	Gene ID number from NCBI
Gene_Symbol	HUGO identifier
Accession	RefSeq Accession Number
Location	Structural location of the SNP: intron, coding, flanking_5UTR, flanking_3UTR, 5UTR, 3UTR
Location_relative_to_gene	If the SNP does not fall within an exon, the value is the actual base pair distance from gene start
Coding_status	NONSYN or SYNON. If the SNP falls within an exon, this field notes a synonymous or nonsynonymous amino acid change
Amino_acid_change	This field notes the actual change to the amino acid, followed by the Genbank ID

Pyrosequencer

To validate the data obtained from the first stage the PyroMark Q96 ID System has been used. It performs a quantitative SNP analysis. One of the main advantages of this technology is the fact that the short sequences of approximately 50 nt show the mutation of interest in the context of the neighbouring genetic sequence. It provides graphically the peak of the SNP in analysis and the ones of the surrounding sequence. This is considered as a “built-in quality control”, since gives the operator the opportunity to verify that the sequenced amplicon is the one of interest, and it guarantees that the assay worked correctly.

For each SNP forward, reverse and sequencing primers were designed. According to the pyrosequencing software, either the forward primer or the reverse one had to have a biotinylated tag (Table 1).

Table 1: Primer sequences used in pyrosequencing assay for SNP genotyping.

Primername	PrimerSequence	T annealing
F-4709583	BIO-TTGATCTACCTGCTGGAGAAGAAA	61°
R-4709583	GTTGGCAAGAGAGAGAAAAGATTT	
S-4709583	GCAGCATTATTAGCCAC	
F-4820294	ACCCTGGGCCATCTCCTAAG	63.5°
R-4820294	BIO-GCTCCTCGGGAAGGCTAAAGA	
S-4820294	CCTCAGGCTTGGCGCCC	
F-611251	BIO-GGGCAAGACACCCAAGTG	62°
R-611251	ACCATCACACCGGGAGTGTT	
S-611251	GGACGATGAAGAAGCTG	
F-3806689	AAGTTTTATTGCTGGCTTTGAGG	62°
R-3806689	BIO-GTCTTTCTCCTGCTCCACTTCTT	
S-3806689	AAGAGAAATTCTAGCTTCTG	
F-11016073	GCTGCTTCAGGGCTGAGA	63°
R-11016073	BIO-TATGGAACTGGGATGGAGAGGT	
S-11016073	TGGTGGAGATTTGCAG	
F-1053516	GGGTTCCCTCACTGCTAGTCA	63°
R-1053516	BIO-GCACACAGGGCAAAATGACC	
S-1053516	AGTTGTGAATTAATGGCA	
F-1053530	ACCAGCCCTCAGTTGTGTTT	59°
R-1053530	BIO-CCCTTTTGGCATTGCAGACT	
S-1053530	TAATACAGTTTCCAGAAAAT	
F-10476829	AATGGAGGGGAGGAAGTGA	61°
R-10476829	BIO-CAGTGCCTGACAGAAGAAGTTG	
S-10476829	AAGGTTCATGTTTGTGG	
F-8110238	BIO-CACTGTGCCTAGCCTGTCTTG	63°
R-8110238	ATGAGCTACACCGAAGCACCTA	
S-8110238	ATCCACGCAGGTCAA	
F-4959774	TTTCATCAAAGGGTTGCACATT	61°
R-4959774	BIO-CCTCCAGCTATAAAGGACACAAGT	
S-4959774	GCTCTGTTTCTGTCCCTA	
F-4674107	GGTTCCCAGGAGTTATGAGCT	64°
R-4674107	BIO-ATCTTTTCATCTCAGCCCTTGCTA	
S-4674107	GAAATGGCTTCCTTAAA	
F-7083622	BIO-CTGGGTCTGGTTGTGGAGATTT	62.5°
R-7083622	CCAGACACCAGGCACTGACA	
S-7083622	CCACGACTGATGAGAA	
F-1654495	BIO-CTTCTGGATGAGGGTGCTAAGT	62°
R-1654495	GTGGCCCCAGATCGAAAG	
S-1654495	GAAAGCACAGCTGCTC	
F-4674258	BIO-GAGGGAGCATGATAAAGCCATAAG	59°
R-4674258	TGTTGAATCCCCGCACCT	
S-4674258	ATCCCCGCACCTGAC	

The first step in the procedure was to perform a PCR with the forward and reverse primer in order to amplify the region harboring the SNPs of interest and to include the biotin in the amplicon.

Several conditions were tested to optimize the PCR reaction. Finally, the following conditions were found to be optimal for our subset of primers:

3‘ at 94°
30‘‘ at 94° \
30 ‘‘ at X° } repeated 45 times
60” at 72° /
10’ at 72°
4° o/n

where X is the annealing temperature, different for each SNP.

The protocol that has been used for the pyrosequencer is the following:

- Perform PCR using one biotinylated and one unlabeled primer to amplify the region containing the SNP

- Binding of the PCR product to the Beads:
  - Per PCR product: 37µl Binding Buffer + 3µl Beads
  - Add 20µl Water to the 40µl of the Beads mix
  - Add 20µl of the PCR product
  - Vortex on a mixer plate for 10min

- Primer mix:
  - Per PCR: add 38,4µl of Annealing Buffer + 1,6µl of Sequence Primer (10µM)
  - Vortex
  - Aliquot 40µl of this solution to the plate

- Separation of the primers with the vacuum:
  - Vacuum the beads with the proper tool
  - Wash the beads 5” in 70% Ethanol (washing step)
  - Air dry for 5”
  - 5” in 0,2N NaOH (denaturation step)
  - Air dry for 5”
  - 5” wash in washing Buffer
  - 30” air dry

- Remove the vacuum and release the primer-beads mix in the plate with the primer-sequence mix
- 80°C for 2’
- 5’ room temperature to cool down
- Proceed with the analysis with Pyrosequencer

Bioinformatics and Statistic

The data obtained were analyzed with R v2.11, a language and environment for statistical computing. It is an integrated suite of software facilities for data manipulation, calculation and graphical display. In addition, packages SnpMatrix, BioBase and Hexbin, all developed from the Bioconductor software project, were used to elaborate and analyze files provided by the Illumina software.

Original files:

Illumina software produces three different files as output: "Finalreport.txt" "SNPTable.txt" and "SampleTable.txt".

The final report is a text file containing the genotype information for each sample, clustered by SNP name.

```
[Data]
SNP Name        Sample ID       Sample Name     Allele1 - Top   Allele2 - Top   GC Score
rs7970377       SG09_01_A01     1-1043                  G               G       0.7488
rs7970377       SG09_01_A02     2-1034                  G               G       0.7488
rs7970377       SG09_01_A03     3-1048                  G               G       0.7488
rs7970377       SG09_01_A04     4-1036                  G               G       0.7488
rs7970377       SG09_01_A05     5-1038                  G               G       0.7488
rs7970377       SG09_01_A07     6-1052                  G               G       0.7488
rs7970377       SG09_01_A08     7-1030                  G               G       0.7488
rs7970377       SG09_01_A09     8-1044                  G               G       0.7488
rs7970377       SG09_01_A10     9-1049                  G               G       0.7488
rs7970377       SG09_01_A11     10-1064                 G               G       0.7488
rs7970377       SG09_01_A12     11-1015                 G               G       0.7488
...
...
...
```

The SNPTable is a text file containing several information about each SNP i.e. HW equilibrium, Minor Allele frequencies, Sequence, AA AB BB frequencies…

```
Index   Name    Chr     Position        Comment ChiTest100      Het Excess
        HW Equil        Minor Freq      # no calls      AA Freq AB Freq BB
Freq    Call Freq       Rep Errors      # Calls 10% GC  50% GC  SNP
Final Report Allele     ILMN Strand     Customer Strand Top Genomic
Sequence        Illumicode Seq  ASO A   ASO B   LSO
1464    rs4820294       22      36400989        Cluster not optimal
0.9100364       0.01129926      0.7661216       0.3455988       122
0.1168831       0.4574315       0.4256854       0.8503067       0       693
        0.2590025       0.7874008       [A/G]   [A/G]   TOP     TOP
CCTGAACGCGGCCAAGCCCGAGGCCTTAGCCAAGCCCTGCAGCCTCAGGCTTGGCGCCCG[A/
G]CCCAGCCTTTCTTTAGCCTTCCCGAGGAGCCAGCGAGAGAGCGCTGCAGCAGCCGCAGTC
TGGCACGCATGTCTTGTATAATGC         ACTTCGTCAGTAACGGACGCTCAGGCTTGGCGCCCGA
GAGTCGAGGTCATATCGTGCTCAGGCTTGGCGCCCGG
CCAGCCTTTCTTTAGCCTTCTGGCACGCATGTCTTGTATAATGCGTCTGCCTATAGTGAGTC
350     rs334558        3       121295972       Cluster not optimal
0.4631751       0.07336286      0.04523995      0.3234899       70
0.4416107       0.4697987       0.08859061      0.9141104       0       745
        0.2767291       0.6815096       [A/G]   [A/G]   TOP     TOP
ACTTGGCCCGGGCGGCGGCGGCGGCGGCGGCGGCACAAGCCCGCATTCGCCCGGGTCAGG[A/
G]GCTGCTCTGTGTGAGGAGCGCTGTCTGCGCAGCCGCCTTGCACTTCCCCACTCCTCCTCC
CAGAAGCGTTCAGTTGTCCGCT  ACTTCGTCAGTAACGGACCATTCGCCCGGGTCAGGA
GAGTCGAGGTCATATCGTCATTCGCCCGGGTCAGGG
CTGCTCTGTGTGAGGAGCCAGAAGCGTTCAGTTGTCCGCTGTCTGCCTATAGTGAGTC
388     rs9170  2       43303642        Cluster not optimal     0.8321798
        0.02119068      0.5616916       0.1306667       65      0.01466667
        0.232   0.7533333       0.9202454       1       750     0.2450657
        0.5571055       [A/G]   [A/G]   TOP     BOT
ACATGTCTTTATACTCGAGTATGTCACAGTAGAACTGGTGGAATAAGCAAACACTTTTTT[A/
G]CTAGTTTATAAAGTTGGAATTAGAAAAGCATGCCACATTTCAGCCTGATTGCAAAGTATG
TAGGCAGTTCACTCGACCGCGA  ACTTCGTCAGTAACGGACGTGGTGGAATAAGCAAACACTTTTTTA
```

Finally, the ‘SampleTable’ file is a text file that associates to each array position the correspondent sample name and provides information about the final call rate.

Index	Array Info.Sentrix ID	Array Info.Sentrix Position	Patient ID	Sample ID		Gender	p95 Grn	p95 Red	p10 GC	p50 GC	Rep Error Rate	Call Rate
398	4775508045	R002_C010	468-2949	SG09_05_B10	Unknown	6590	3926	0.5419981	0.82934			0.9696356
678	4775559002	R008_C012	834-1-026	SG09_09_H12	Unknown	4319	1974	0.5624446	0.8329252		0.002749146	0.9831309
688	4753051018	R001_C012	806-5542	SG09_08_A12	Unknown	5022	2109	0.5908464	0.8401741			0.9871795
758	4775559002	R002_C004	886-2379	SG09_09_B04	Unknown	3887	2087	0.6455711	0.8450144			0.9871795
825	4775559002	R008_C003	955-10521	SG09_09_H03	Unknown	3823	2084	0.5657619	0.8337487			0.9878542
754	4775559002	R001_C011	882-1987	SG09_09_A11	Unknown	3967	1117	0.5363206	0.8274181			0.988529
768	4775559002	R003_C002	896-3839	SG09_09_C02	Unknown	3589	1952	0.6123106	0.8418135			0.9892038
804	4775559002	R006_C003	932-3897	SG09_09_F03	Unknown	5621	3166	0.6486236	0.849135			0.9892038
126	4776244002	R003_C012	137-1562	SG09_02_C12	Unknown	1768	1232	0.5604988	0.8288589			0.9892038
828	4775559002	R008_C010	964-12005048	SG09_09_H10	Unknown	5213	2411	0.6560652	0.8481836			0.9892038
704	4753051018	R003_C010	831-1-017	SG09_08_C10	Unknown	4553	2390	0.6106665	0.8450144			0.9898785
639	4753051027	R007_C001	765-3617	SG09_07_G01	Unknown	6831	920	0.6036888	0.8431768			0.9898785
746	4775559002	R001_C001	872-1285	SG09_09_A01	Unknown	3541	1433	0.6112616	0.8409349			0.9905533
609	4753051027	R004_C007	728-2604	SG09_07_D07	Unknown	7603	4372	0.6654311	0.8501655			0.9905533

In addition to these, the file “Status_individuals” was created containing all the information on sex and age for each sample. Since the study was performed in a blinded fashion, this file was used at the end of the project to unblind the data.

TN_ID	status	Geschlecht	age
1002	nothing	m	60
1008	nothing	m	67
1015	nothing	m	66
1022	hyplast	w	70
1025	nothing	w	77
1027	nothing	m	71
1029	hyplast	w	67
1030	nothing	m	74
1034	hyplast	m	73
1036	nothing	m	60
1038	nothing	m	50
1041	nothing	m	64
1043	nothing	m	64
1044	nothing	m	53
1048	hyplast	w	70

The statistical procedure will be explained in more details in the result section in context with the data obtained.

RESULTS

Genes selection

Initial selection of the genes to include in the project was performed manually by analyzing public databases and searching the literature.

iHop database (www.ihop-net.org) was used to identify the function of a gene and possible participation in specific pathway. As another source, original articles and reviews (www.pubmed.com, www.highwire.org) were searched for additional background information. Pathways known to be important for apoptosis, cell cycle control and differentiation were extracted using the KEGG database (Kyoto Encyclopedia of Genes and Genomes, www.genome.jp/kegg) and gene ontology (GO, www.geneontology.org) resources.

After an initial screening, 826 genes were identified, and assigned to one of 9 functional categories. (table 2).

Table 2: functional categories of selected genes

Pathways	# of genes
Apoptosis	436
Anti-apoptosis	172
DNA repair	41
Inflammation	27
Cell cycle control	25
Proliferation	11
Adhesion / Cell survival	10
Autophagy	4
Other	100

190 genes from the above list where excluded from the study, since they do not contain SNPs fulfilling our SNP selection criteria (see SNP Selection ¶4). Therefore, the analysis was conducted on a total number of 636 genes (see supplementum).

Chip Design – Screening and identification of SNPs

The Score output file generated by the ADT provided a list of 160.000 SNPs that fell within our selected genes or in the flanking regions.

Accurate selection of single nucleotide polymorphisms is an issue of primary importance. In order to focus on the more significant SNPs, several criteria have been considered to filter them to a smaller number:

1. All SNPs presenting any of the following failure codes were removed from the list: Error codes 103; 104; 301; 302; 340; 360; 399 (see table 3 for detailed explanation)

Table 3: critical failures considered for selecting appropriate SNPs

Critical Failures	
101	Flanking sequence is too short
103	TOP/BOT strand cannot be determined due to low sequence complexity
104	SNP is not appropriate for Illumina platform due to Tri- or quad-allelic variant
301	SNP is in duplicated/repetitive region
302	Tm is outside assay limits
340	Another SNP is closer than 61 nucleotides away
360	SNP has a low score
399	Multiple contributing issues

2. Only SNPs with a Minor Allele Frequency (MAF) >0.03 were included in the study. All SNPs with smaller allele frequencies (<3%) in the Caucasian population where excluded.

3. Each SNP investigated during the chip design is given a score ranging from 0 to 1 provided by Illumina. A rank score <0.4 has a low success rate, from 0.4 to <0.6 has a moderate success rate, and >0.6 has a high success rate for the conversion of a SNP into a successful GoldenGate assay. SNPs resulting with a final score <0.6 were excluded, since they could have decreased the overall efficiency of the assay.

4. SNPs presenting "*Validation_Class = 1*" were removed from the list. This validation status description indicates that the polymorphism has not been validated; even if it has a high design score, there is an increased chance that it would be a monomorphic. SNP. Validation_Bin accepted for the study where the following:

a. TwoHitValidated: Both alleles have been seen in two independent methods and population

b. HapMapValidated: SNP has been genotyped by the HapMAp project

c. GoldenGate_Validated: SNP has been previously designed and successfully generated polymorphic results using the GoldenGate assay

5. Finally, SNPs were filtered accordingly to functional aspects.

Rules for including SNPs into the final gene selection list:

a. Among the variations located in coding exon sequences, only the nonsynonimous SNPs (NONSYN leading to an amino acid exchange) were included. All the synonymous or the three-quad allelic variations were excluded.

b. SNPs located in the promoter area were also included. As promoter area, we have chosen the intronic space of 1000bases length, both in the 3' and 5' flanking regions of each gene.

c. SNPs falling in putative splicing region were included. As splicing region, intronic sequences of 30bases length were chosen upstream and downstream of the acceptor and donor splice sites in all genes of the list, respectively.

Through this process of SNPs filtration the initial 160.000 polymorphisms provided by the Score output file were reduced to the 1536 most important SNPs (for full list of variations included in the chip, refer to the supplement).

Statistical analysis package and language R

R (http://www.r-project.org/) is a language for statistical computing providing a wide variety of statistical and graphical techniques. It can be expanded using "packages" available through Bioconductor (http://www.bioconductor.org/).

Out of the many packages available online, three were chosen for this project:

SnpMatrix, an R package for analysis of genome-wide association studies;
Hexbin for plotting the graphical output;
Biobase, which provides functions needed by other Bioconductor packages.

In order to make more clear the reading of the following chapters, the code used in R has been marked in blue and the output and the results achieved in italic.

Stage 1 – Analysis and File processing code

Setup and loading of the files

Three packages "snpMatrix, Biobase and Hexbin", have been used for this study:

```
> library(Biobase)
Welcome to Bioconductor
  Vignettes contain introductory material. To view, type
  'openVignette()'. To cite Bioconductor, see
  'citation("Biobase")' and for packages 'citation(pkgname)'.

> library(snpMatrix)
Loading required package: survival
Loading required package: splines

> library(hexbin)
Loading required package: grid
Loading required package: lattice
```

The three main object data required for the data analysis were created at first.

The object "**gwas.data**" containing genotype information for each SNP and sample is created by reading the information from the file *"*SG01_to_SG09_Analysis_FinalReport_with_name.txt*"*. The object

"**snp.info**" containing SNP information as "Index, Name, Chromosome, Position. HW equilibrium…." is created from the file *"SG01_to_SG09_Analysis_SNPTable_completeAnno.txt"* while the object "**sample.info**" containing information on each sample, as Patient ID, call rate and error rate was create from the file *"SG01_to_SG09_Analysis_SamplesTable.txt"*

```
>gwas.data = read.table("SG01_to_SG09_Analysis_FinalReport_withName.txt", sep="\t", header=T, quote="", skip=9, as.is=T)

>snp.info = read.table("SG01_to_SG09_Analysis_SNPTable_completeAnno.txt", header=T, quote="", sep="\t", comment.char="")

> sample.info = read.table("SG01_to_SG09_Analysis_SamplesTable.txt", sep="\t",
header=T, quote="")
```

Call rate and Minor Allele Frequency verification

The code "summary" provides information on the object in argument to verify the correct format while the "dim" code gives information of the size of sample.info object; a matrix composed by 815 rows and 12 columns. 815 is the number of patients and controls that have been used (784 patients and controls plus 31 internal controls) 12 are the columns as shown in the summary.

> summary(sample.info)

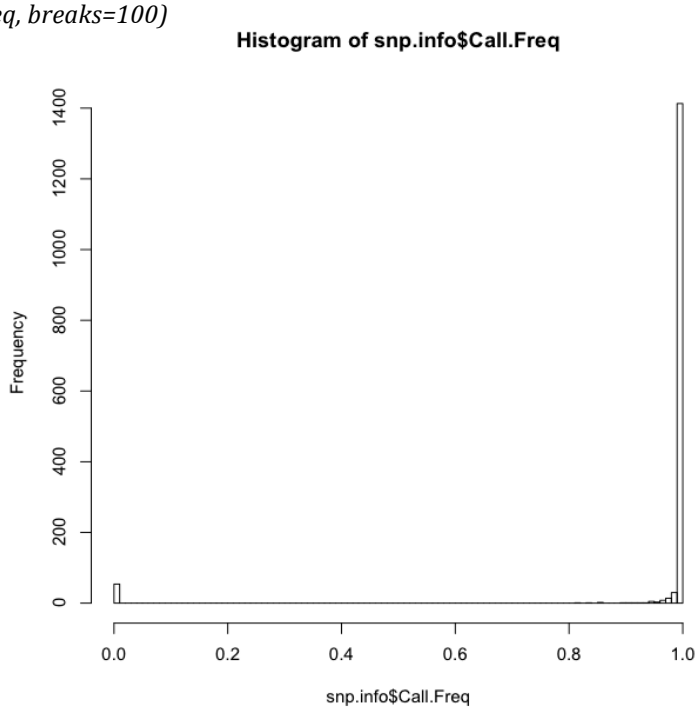
```
     Index       Array.Info.Sentrix.ID Array.Info.Sentrix.Position     Patient.ID        Sample.ID       Gender       p95.Grn
 Min.   :  1.0   Min.   :4.753e+09     R001_C001:  9                794-5038 :  3   SG09_01_A01:  1   Unknown:815   Min.   : 1768
 1st Qu.:216.5   1st Qu.:4.755e+09     R001_C002:  9                834-1-026:  3   SG09_01_A02:  1                 1st Qu.: 5151
 Median :420.0   Median :4.770e+09     R001_C005:  9                101-1323 :  2   SG09_01_A03:  1                 Median : 6668
 Mean   :419.0   Mean   :4.767e+09     R001_C007:  9                129-1565 :  2   SG09_01_A04:  1                 Mean   : 6830
 3rd Qu.:623.5   3rd Qu.:4.776e+09     R001_C008:  9                14-1098  :  2   SG09_01_A05:  1                 3rd Qu.: 8066
 Max.   :829.0   Max.   :4.776e+09     R001_C009:  9                183-1725 :  2   SG09_01_A07:  1                 Max.   :14608
                                       (Other)  :761                (Other)  :801   (Other)    :809
    p95.Red         p10.GC           p50.GC       Rep.Error.Rate        Call.Rate
 Min.   : 803   Min.   :0.5363   Min.   :0.8252   Min.   :0.000e+00   Min.   :0.9696
 1st Qu.:2633   1st Qu.:0.6565   1st Qu.:0.8483   1st Qu.:0.000e+00   1st Qu.:0.9973
 Median :3479   Median :0.6608   Median :0.8495   Median :0.000e+00   Median :0.9987
 Mean   :3535   Mean   :0.6575   Mean   :0.8488   Mean   :2.183e-04   Mean   :0.9978
 3rd Qu.:4292   3rd Qu.:0.6676   3rd Qu.:0.8502   3rd Qu.:3.375e-04   3rd Qu.:0.9993
 Max.   :6947   Max.   :0.6779   Max.   :0.8509   Max.   :2.749e-03   Max.   :1.0000
                                                  NA's   :7.550e+02
```

> dim(sample.info)
[1] 815 12

With the "sum" code referring to the sample.info object related to a specific Call rate, it is possible to evaluate how many samples are falling within a specific call rate (=0 or <0.95 or <0.99) and visualize their distribution.

> sum(sample.info$Call.Rate==0)
[1] 0
> sum(sample.info$Call.Rate<0.95)
[1] 0
> sum(sample.info$Call.Rate<0.99)
[1] 12

>hist(snp.info$Call.Freq, breaks=100)



As the histogram shows there are some SNPs with a call frequency equal to 0, these SNPs have been excluded from the study

As controls, the Hardy-Weinberg Equilibrium in snp.info has to be tested, and the distribution of the minor allele frequency is represented via histogram

```
> summary(snp.info$HW.Equil)
   Min.         1st Qu.          Median          Mean         3rd Qu.          Max.
0.0000          0.1892          0.4614          0.4726          0.7503          1.0000
> sum(snp.info$HW.Equil< 0.01)[1] 105
> sum(snp.info$HW.Equil< 0.05)
[1] 180
>hist(snp.info$Minor.Freq)
```

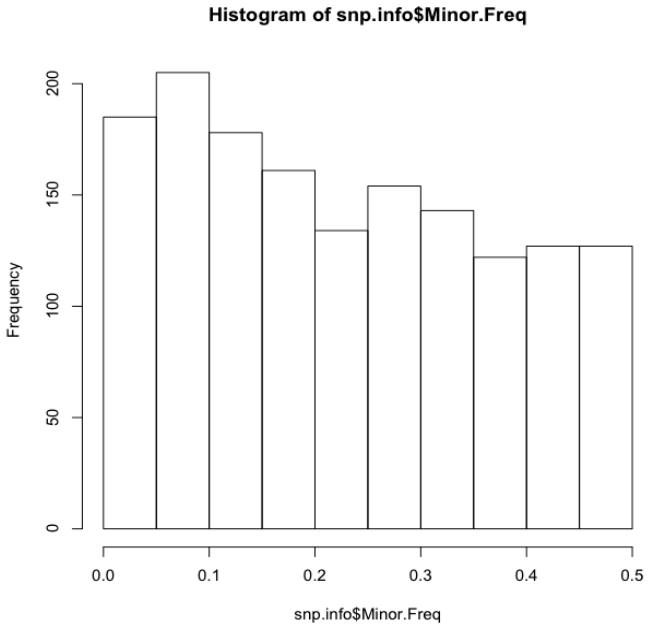


Figure 14: In the histogram is represented the distribution of the minor allele frequency detected in all the SNP included in the study. On the y axe is represented the number of SNPs with a specific MAF (x axe).

Association of SNPs information and patients status

The object gwas.data is composed of several columns containing information about SNPs and Samples; these information have been used to create the object "snp.names" and "sample.names"

```
> head(gwas.data)
SNP.Name     Sample.ID   Sample.Name   Allele1...Top   Allele2...Top   GC.Score
1 rs7970377  SG09_01_A01    1-1043          G               G            0.7488
2 rs7970377 SG09_01_A02     2-1034          G               G       0.7488
3 rs7970377 SG09_01_A03     3-1048          G               G            0.7488
4 rs7970377 SG09_01_A04     4-1036       G                  G            0.7488
5 rs7970377 SG09_01_A05     5-1038   G              G              0.7488
6 rs7970377 SG09_01_A07     6-1052          G             G              0.7488
>snp.names = unique(as.character(gwas.data$SNP.Name))
>sample.names = unique(as.character(gwas.data$Sample.Name))
```

The structure of the object "snp.names" is shown as single identifiers "rs codes" while the one of the "sample.names" is composed by the specific number of each sample

```
>str(snp.names)
chr [1:1482]"rs7970377" "rs4253211" "rs318475" "rs13182512" "rs9859413" "rs3733875" "rs901746" "rs4674260"
"rs3918270" ...
>str(sample.names)
chr [1:784] "1-1043" "2-1034" "3-1048" "4-1036" "5-1038" "6-1052" "7-1030" "8-1044" "9-1049" "10-1064" "11-1015" "12-
1067" ...
```

The object “status.sample” is created from the file “status_individuals”. It contains information of all the samples and provides 4 variables: id number, sex (“Geschlecht” m or w), age and status (“hyperplastic patient” or “control”). These information are required to unblind the study and proceed with statistical analysis

```
>status.sample = read.table("Status_individuals_v5.txt", sep="\t", header=T)
>str(status.sample)
'data.frame':       784 obs. of  4 variables:
 $ TN_ID     : Factor w/ 784 levels "1-004","1-0049",..: 49 55 60 63 65 67 68 69 72 73 ...
 $ status    : Factor w/ 2 levels "hyplast","nothing": 2 22 1 2 2 1 2 1 2 ...
 $ Geschlecht: Factor w/ 2 levels "m","w": 1 11 2 2 1 2 1 11 ...
 $ age       : int  60 67 66 70 77 71 67 74 73 60 ...
> head(status.sample)
  TN_ID  status Geschlecht age
1  1002 nothing         m  60
2  1008 nothing         m  67
3  1015 nothing         m  66
4  1022 hyplast         w  70
5  1025 nothing         w  77
```

The resulting status.sample is composed of 272 Hyperplastic polyps patients and 512 Controls for a total number of 784 single samples successfully analyzed in this study.

```
> table(status.sample$status)
hyplast nothing
    272     512
> sum(table(status.sample$status))
[1] 784

> table(status.sample$Geschlecht)
  m   w
441 343
> sum(table(status.sample$Geschlecht))
[1] 784
```

The object “STATUS” is created to properly format all the information: “hyperplastic” is now considered “patient” and “nothing” is “control”.

```
>STATUS = status.sample$status[match(sub("^[0-9]+-", "",sample.names), status.sample$TN_ID)]
>str(STATUS)
 Factor w/ 2 levels "hyplast","nothing": 2 1 1 2 222222 ...
> levels(STATUS)
[1] "hyplast" "nothing"
> levels(STATUS) = c("case", "control")
> table(STATUS)
STATUS
case control
272     512
> sum(table(STATUS))
[1] 784
```

Chi square test

The snp.mat.new object is created by reading all the genotype information and associating them to each sample.

>snp.mat.new = read.snps.long(files="SG01_to_SG09_Analysis_FinalReport_withName.txt", sep="\t", skip=10, sample.id=sample.names, snp.id=snp.names, fields=c(sample=3, snp=1, allele1=4, allele2=5), codes="nucleotide", in.order=F)
1205209 genotypes successfully read
2621 genotypes were not called
Warning message:
In read.snps.long(files = "SG01_to_SG09_Analysis_FinalReport_withName.txt", :
10 lines of input file(s) were skipped

A resulting matrix is now created

> show(snp.mat.new)

A snp.matrix with 784 rows and 1482 columns
Row names: 1-1043 ... 965-12014497
Col names: rs7970377 ... rs9527024

>snp.mat.new@.Data[1:10,1:10]

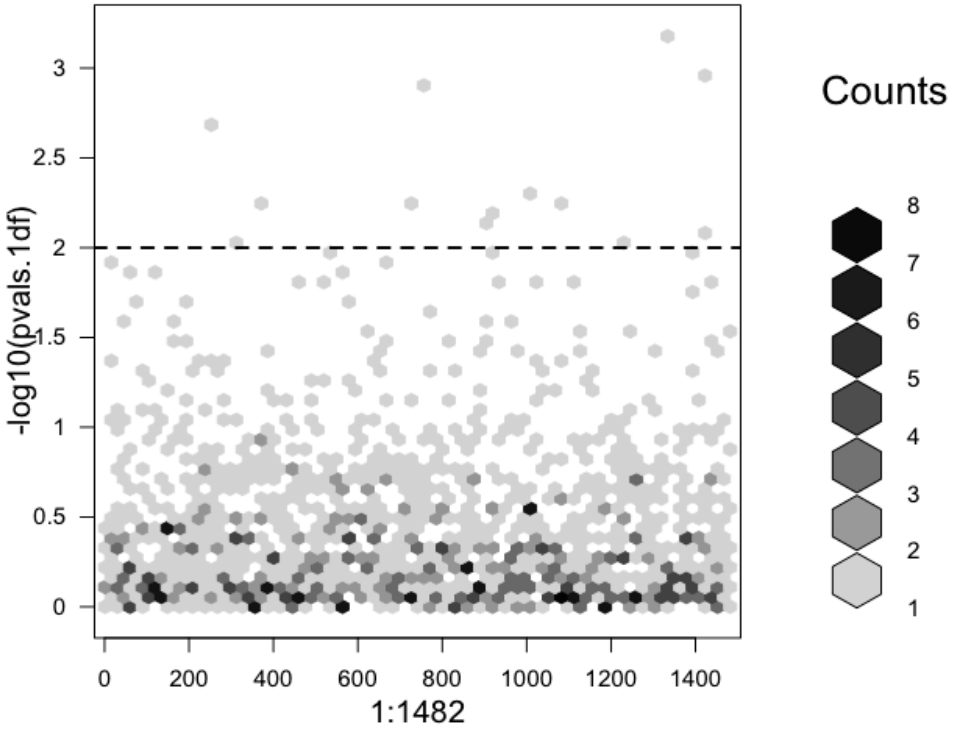
	rs7970377	rs4253211	rs318475	rs13182512	rs9859413	rs3733875	rs901746	rs4674260	rs3918270	rs2069391
1-1043	03	02	03	03	03	03	01	01	03	03
2-1034	03	01	02	02	02	03	01	02	02	03
3-1048	03	01	02	01	02	03	02	03	02	03
4-1036	03	01	02	01	03	03	01	02	03	03
5-1038	03	01	02	01	03	03	02	02	03	03
6-1052	03	01	01	01	03	03	01	03	03	03
7-1030	03	01	03	01	03	02	01	03	03	02
8-1044	03	02	02	03	03	03	02	03	03	03
9-1049	03	01	01	02	03	03	03	01	02	03
10-1064	03	01	01	03	03	02	02	01	03	03

Chi square test with 1 and 2 degree of freedom is now performed and the package hexbin is loaded for graphical representation of the SNPs.

> test.allSNPs.1df = single.snp.tests(STATUS, snp.data=snp.mat.new)
> summary(test.allSNPs.1df)

N	Chi.squared.1.df	Chi.squared.2.df	P.1df	P.2df
Min. :640.0	Min. : 0.00000	Min. :5.286e-04	Min. : 0.000687	Min. :2.857e-04
1st Qu.:784.0	1st Qu.: 0.08697	1st Qu.:5.743e-01	1st Qu.: 0.246271	1st Qu.:2.501e-01
Median :784.0	Median : 0.45550	Median :1.452e+00	Median : 0.499734	Median :4.838e-01
Mean :782.4	Mean : 0.96715	Mean :2.070e+00	Mean : 0.507223	Mean :4.950e-01
3rd Qu.:784.0	3rd Qu.: 1.34434	3rd Qu.:2.772e+00	3rd Qu.: 0.768061	3rd Qu.:7.504e-01
Max. :784.0	Max. :11.52400	Max. :1.632e+01	Max. : 1.000000	Max. :9.997e-01
	NA's :30.00000	NA's :9.200e+01	NA's :30.000000	NA's :9.200e+01

> pvals.1df = p.value(test.allSNPs.1df, df=1)
> plot(hexbin(1:1482, -log10(pvals.1df), xbin=50))



The line represents a p value of 0.01. All the SNP over this line are associated with a p<0.01

> sp2 = plot(hexbin(1:1482, -log10(pvals.1df), xbin=50))
>hexVP.abline(sp2$plot.vp, h=2, lty=2)

These are the rs codes and the p values that identifies the 14 SNPs associated (p<0.01)

> sum(pvals.1df < 0.01, na.rm=T)
[1] 14

> pvals.1df[which(pvals.1df < 0.01,)]

rs3806689	*0.0019950478*	***rs611251***	*0.0012251791*	***rs7083622***	*0.0089261534*
rs1654495	*0.0092208981*	***rs4959774***	*0.0069063909*	***rs4709583***	*0.0006870336*
rs1053516	*0.0054163600*	***rs8110238***	*0.0063290929*	***rs4820294***	*0.0010579281*
rs4674258	*0.0099687868*	***rs11016073***	*0.0048065344*	***rs4674107***	*0.0083418650*
rs10476829	*0.0056928364*	***rs1053530***	*0.0054163600*		

Evaluation of Sample Clustering

After all the samples have been analyzed with the Illumina GoldenGate, verification of optimal clustering has been performed. Each SNP presenting with error codes “not optimal clustering” (figure 15) or “not analyzable” (figure 16) was manually verified.

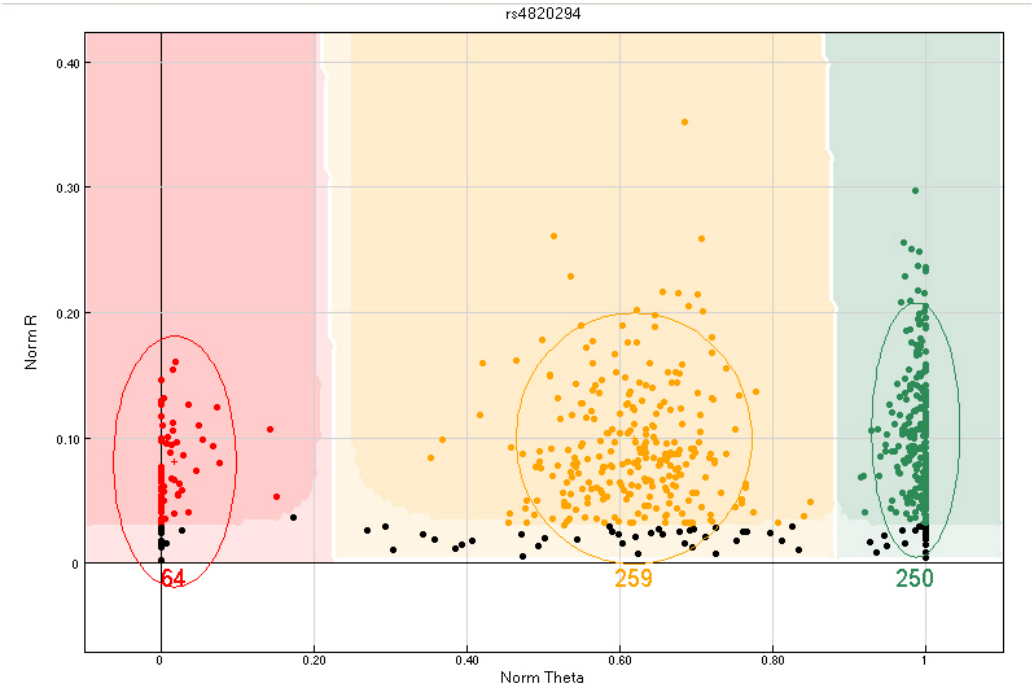


Figure 15: Cluster not optimal; in this event only the samples below the threshold have not been included in the study.

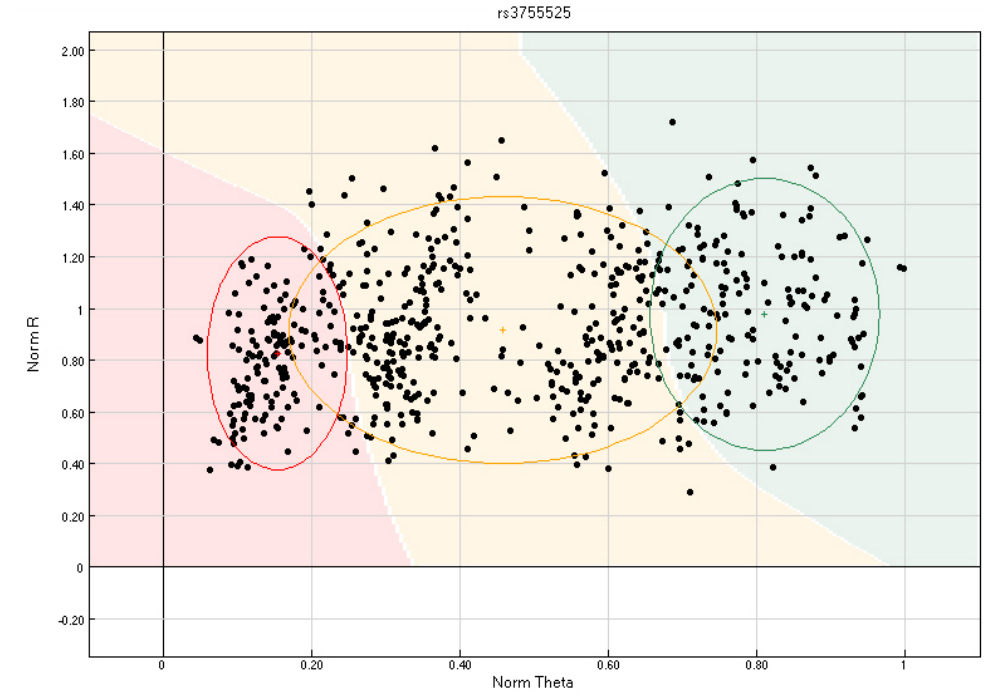


Figure 16: Cluster not analyzable; it has not been possible to cluster properly all the samples. The SNP has therefore been excluded from the study;

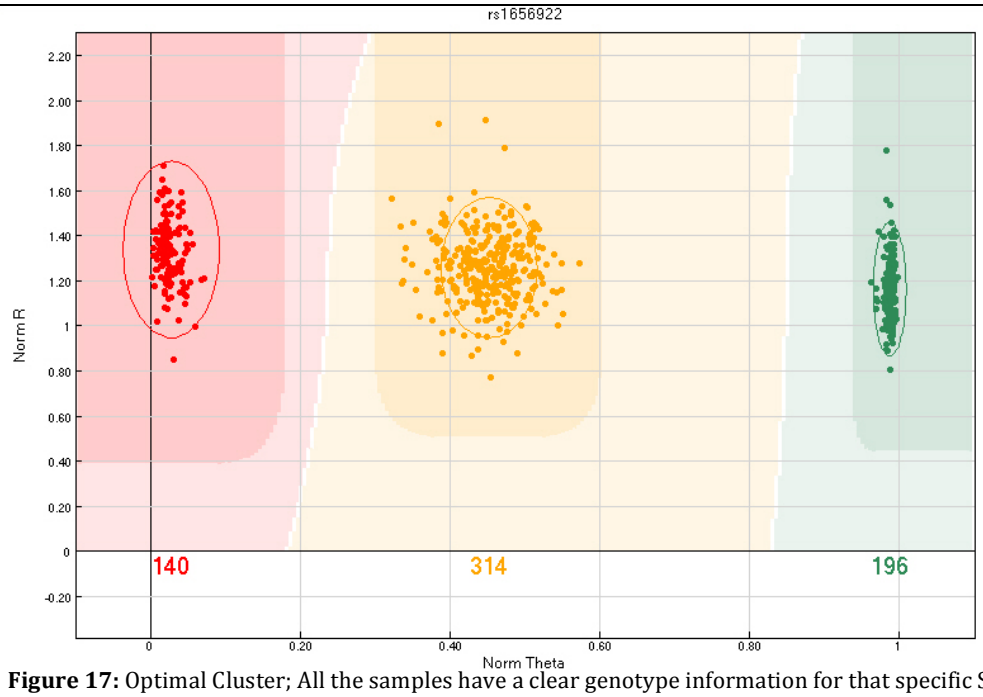


Figure 17: Optimal Cluster; All the samples have a clear genotype information for that specific SNP

The panels in Figures 15-16-17 represent examples of the genotype assignments of the patients/control samples of the study to a single SNP. Specifically, every dot is a sample with the color defining the sample as being homozygous for the minor allele frequency (in red) and the major allele frequency (in green) or heterozygous (in yellow).

Oligo pool assay and quality control characteristics

The chip consisted of 1536 selected SNPs. Out of these, 72 SNPs where excluded because they where "not analyzable " (n=52), due to multiclustering (n=13) or because the MAF calculated did not match the one present in the dbSNP database (n=7). Therefore the final Oligo Pool Assay (OPA) contained 1464 (95,3%) successfully genotyped SNPs. 864 wells were analyzed with 815 samples successfully processed. As quality control, 31 blinded samples were processed in duplicate and all of them had a reproducibility frequency >99.7%. Data have been generated using Genotyping Module 1.6.3 of Genome Studio Version V2010.1 (Illumina).

Stage 2 - Validation

After Stage 1 was completed, genotype data for 1464 SNPs in 784 patients were collected. 14 SNPs presented association score with a p value <0.01.

SNP_Name	p value	Chr	MAF	Gene	Location	Position	Codingstat	AA
rs4709583	0.0006870336	6	0.0417	PARK2	intron	-20		
rs4820294	0.0010579281	22	325	LGALS1	flanking_5UTR	-570		
rs611251	0.0012251791	19	158	PPP1R15A	coding	[595/1069]	NONSYN	V199A
rs3806689	0.0019950478	3	0.2174	KNG1	flanking_5UTR	-43		
rs11016073	0.0048065344	10	225	MKI67	coding	[2959/3885]	NONSYN	I2101T
rs1053516	0.0054163600	3	259	WDR48	flanking_3UTR	-1		
rs1053530	0.0054163600	3	258	GORASP1	flanking_3UTR	-9		
rs10476829	0.0056928364	5	0.05	HDAC3	flanking_3UTR	-381		
rs8110238	0.0063290929	19	0.5	NLRP2	flanking_3UTR	-785		
rs4959774	0.0069063909	6	0.1136	RIPK1	flanking_5UTR	-722		
rs4674107	0.0083418650	2	277	IGFBP5	flanking_3UTR	-474		
rs7083622	0.0089261534	10	33	MKI67	coding	[2251/4593]	NONSYN	T2337N
rs1654495	0.00922008981	19	0.5	NLRP2	intron	-19		
rs4674258	0.0099687868	2	0.4762	IL8RB	flanking_5UTR	-160		

All these polymorphisms were selected for analysis in a different cohort of samples for validation purposes. 38 patients affected from hyperplastic polyps were genotyped together with 38 sex-age matched controls via pyrosequencing. All genotype data have been collected in the "genotype_replication_1-14.txt" file.

```
>data.snp.replication = read.table("genotype_replication_1-14.txt", header=T, sep="\t", row.names=1)
> head(data.snp.replication)
          rs4709583  rs4820294  rs611251  rs3806689  rs11016073  rs1053516
PATIENT1          3          2         3          1           3          1
PATIENT2          3          2         2          3           3          2
...
          rs1053530  rs10476829  rs8110238  rs4959774  rs4674107  rs7083622
PATIENT1          1           3          2          1          3          3
PATIENT2          2           3          3          1          3          3
...
          rs1654495  rs4674258
PATIENT1          2          2
>snpMat.replication
A snp.matrix with  76 rows and  14 columns
Row names:  PATIENT1 ... CONTROL38
Col names:  rs4709583 ... rs4674258

> table(pheno.snp.replication)
pheno.snp.replication
  case control
    38      38
```

Chi square test was performed on the validation set

>single.snp.tests(phenotype=factor(pheno.snp.replication), snp.data=snpMat.replication)

	N	Chi.squared.1.df	Chi.squared.2.df	P.1df	P.2df
rs4709583	*76*	*1.38248848*	*NA*	*0.23967772*	*NA*
rs4820294	*76*	*0.34562212*	*NA*	*0.55660187*	*NA*
rs611251	*76*	*3.14776275*	*4.235268*	*0.07603114*	*0.12031598*
rs3806689	*76*	*0.03671072*	*1.281285*	*0.84805522*	*0.52695388*
rs11016073	*76*	*0.51724138*	*4.526316*	*0.47202081*	*0.10402148*
rs1053516	*76*	*0.02846300*	*1.623788*	*0.86602491*	*0.44401639*
rs1053530	*76*	*0.02846300*	*1.623788*	*0.86602491*	*0.44401639*
rs10476829	*76*	*0.55147059*	*NA*	*0.45771743*	*NA*
rs8110238	*76*	*0.23413111*	*1.390469*	*0.62847745*	*0.49895732*
rs4959774	*76*	*0.07723996*	*1.076555*	*0.78107317*	*0.58375289*
rs4674107	*76*	*0.00000000*	*1.127010*	*1.00000000*	*0.56921034*
rs7083622	*76*	*0.34246575*	*NA*	*0.55840939*	*NA*
rs1654495	*76*	*0.11682243*	*1.019984*	*0.73250671*	*0.60050033*
rs4674258	*76*	*0.44117647*	*7.697368*	*0.50655517*	*0.02130775*

Subsequently, the data from the two set of different chi square test were combined in a meta-analysis

>meta.set = pool(snp.discovery.tests, snp.replication.tests)>meta.set
>str(meta.set)
Formal class 'snp.tests.single' [package "snpMatrix"] with 4 slots
..@ snp.names: chr [1:14] "rs4709583" "rs4820294" "rs611251" "rs3806689" ...
..@ chisq : num [1:14, 1:2] 12.9 9.52 4.73 9.1 6.12 ...
.. ..- attr(, "dimnames")=List of 2*
.. .. ..$: NULL
.. .. ..$: chr [1:2] "1 df" "2 df"
..@ N : int [1:14] 860 749 860 860860860860860860860 ...
..@ N.r2 : num(0)

Combined p values have been calculated

>meta.set

	N	Chi.squared.1.df	Chi.squared.2.df	P.1df	P.2df
rs4709583	860	12.896463	13.079910	0.0003292037	0.0014445534
rs4820294	749	9.516305	11.494979	0.0020365421	0.0031907811
rs611251	860	4.729152	12.112057	0.0296555095	0.0023436900
rs3806689	860	9.104347	10.575772	0.0025500270	0.0050524300
rs11016073	860	6.120372	12.475439	0.0133632712	0.0019543070
rs1053516	860	7.197419	7.212210	0.0073008519	0.0271574179
rs1053530	860	7.197419	7.212210	0.0073008519	0.0271574179
rs10476829	860	8.173442	16.671572	0.0042508169	0.0002397806
rs8110238	860	7.540036	9.787198	0.0060342850	0.0074944015
rs4959774	860	7.084398	8.700201	0.0077757938	0.0129055164
rs4674107	860	6.513431	6.787297	0.0107062747	0.0335859204
rs7083622	857	7.145838	7.469649	0.0075138174	0.0238773671
rs1654495	860	5.710033	7.218430	0.0168682244	0.0270730907
rs4674258	859	5.087192	13.460167	0.0241031841	0.0011944330

Meta-analysis

After the validation study was completed, 9/14 SNPs possessed p values of <0.01, i.e. the differences between genotypes of patients and controls has a lower than 1% probability to be caused by chance. A variation that has been associated with a pathology can have a protective effect or a causative effect.

For a protective variant, the chances to develop the disease are reduced, whereas the opposite happens, when a causative variation is found.
When dealing with a meta-analysis between two different cohorts of patients and controls and two different association studies with the same polymorphisms the "effect scenario" has to be considered. It is possible that a SNP that has been associated with a strong p (p<0.01) with a pathology in both the first stage and the replication analysis could have a different distribution of the minor allele in the two population, being more frequent for example in the patient in one case and in the controls in the other. If this happens, the SNP is considered having a causative effect in one study and a protective effect in the replication. It was verified, that the "sign effect" between stage one and stage two for these SNPs would have been the same.
Only variations that present the same "sign effect" in both the studies can be considered positively associated.
Identification of the effect sign for each SNP in the two stages was performed as follow:

```
>effect.sign(snp.replication.tests)
 rs4709583      rs4820294       rs611251       rs3806689      rs11016073      rs1053516
        -1              -  1              1              1               1              1
rs1053530      rs10476829      rs8110238      rs4959774      rs4674107      rs7083622
        1              -1             -1              1               0              1
rs1654495      rs4674258
       -1              1

> which(effect.sign(snp.discovery.tests) == effect.sign(snp.replication.tests))
 rs4709583      rs3806689      rs11016073      rs1053516      rs1053530      rs10476829
         1              4               5              6              7               8
rs8110238      rs4959774      rs7083622
         9             10              12
```

Five SNPs presented a different effect sign between the first stage and the second and therefore have been excluded from the final list.

As our study includes a large number of variables we decided to control the proportion of rejecting true Hypothesis via the "false discovery rate" (FDR) approach as suggested by Nakagawa[21] and highlighted by García[22], instead of controlling the probability of obtaining even one false rejection of H_0 using the Bonferroni correction.

Since controlling FDR provides a much better compromise between Type I and Type II errors when

multiple testing is necessary[21], we performed this test on the 9 SNPs resulted from the meta-analysis, according to Benjamini and Hochberg description[23].

The new p of significativity according to FDR (q value) obtained is shown in the table:

SNP	*original p*	significativity limit FDR	"*q* value"
rs4709583	0.00033	0.005556	0.002963
rs3806689	0.00255	0.011111	0.008748
rs10476829	0.00425	0.016667	0.008748
rs8110238	0.00603	0.022222	0.008748
rs1053530	0.00730	0.027778	0.008748
rs1053516	0.00730	0.033333	0.008748
rs7083622	0.00751	0.038889	0.008748
rs4959774	0.00778	0.044444	0.008748
rs11016073	0.01336	0.05	0.013363

8 SNPs resulted associated with the presence of hyperplastic polyps with a significativity level <0.01 and rs11016073 presented a significatifity level <0.05.

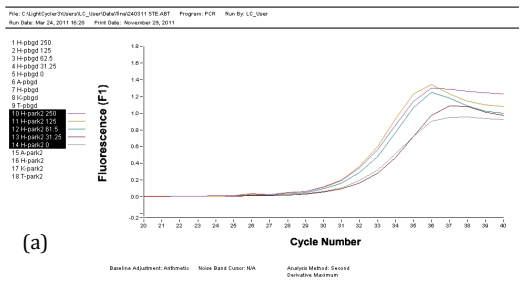
Out of all, rs4709583 had the highest association with a q value of 0.0029.

Pyrosequencing Analysis from Cell Lines

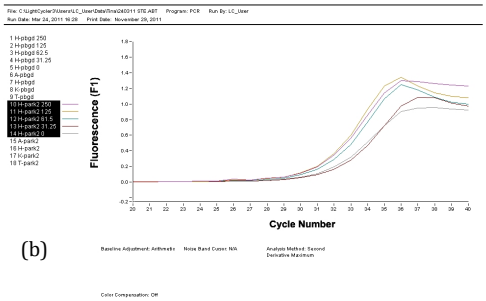
DNA from 38 different cell lines was analysed using pyrosequencing to identify the potential presence of the mutated allele in rs4709583 (*PARK2)*. Out of all the cell lines, only HL60 (acute myeloid leukemia) and THP-1 (acute monocytic leukemia) showed a heterozygosityT/C in rs4709583. As control, wild type C/C AR230 and K562 (chronic myeloid leukemia) cells were used.

Real Time and PCR analysis

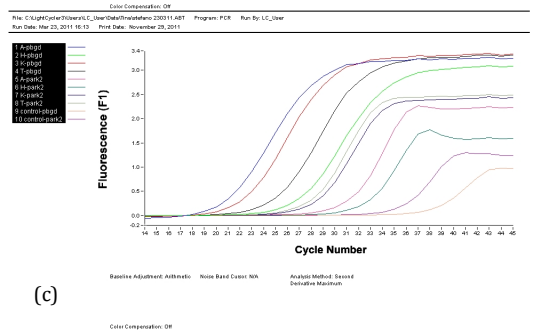
Due to its close position to a splicing site, the variation rs4709583 was further analysed in-vitro to identify its potential influence on splicing in the *Park2* gene that would result in changes of the protein structure. Specifically, alternative splicing would eliminate exon 4. Real time analysis has been performed to compare RNA expression between cell lines with wild type variation (AR230 and K562) and cell lines with heterozygous mutation (HL60 and THP1). Analysing the ratio between the RNA expression of PARK2 and PBGD genes in the four different heterozygous cell lines did not show any differences in *Park2* mRNA expression.



(a)



(b)



(c)

Cell lines-gene	Norm. Expression ratio
AR230-park2	0
AR230-pbgd	
K562-park2	0.09
K562-pbgd	
HL60-park2	0.17
HL60-pbgd	
THP1-park2	0.68
THP1-pbgd	

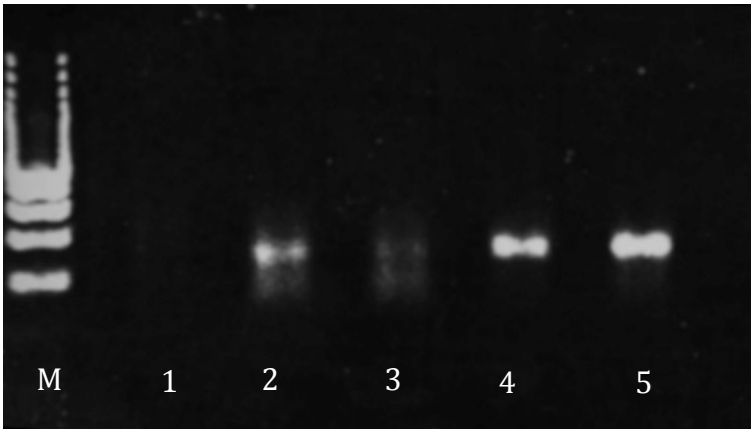
Figure 18: Standard curve for PARK2 (a), PBGD (b) and amplification curve for PARK2 and PBGD (c)

PCR have been performed to assess alternative splicing in PARK2 mRNA. Nucleic acids were extracted from cell lines AR230, K562, HL60 and THP1 and were subjected to RT-PCR. 5µl of PCR products were loaded on Agarose gel (Figure 19). No differences in band size indicative of alternatively spliced mRNA were noticeable. Specifically, if the SNP would affect splicing, the exon 4 would be eliminated resulting in a 262 bp PCR amplicon on the gel composed of exon 3 and 5, only. However, all band sizes observed were the one of 384bp total size, as expected for an amplicon including all the three exons. The amplicon of 384bp were obtained as shown in figure 9 demonstrating specific PARK2 product. No alternative splicing was caused by rs4709583.

Figure 19:

Amplification of PARK2 gene.

M – Marker
1 – H20
2 – AR230
3 – HL60
4 – K562
5 – THP1



384bp

DISCUSSION

In the last decades, studies on the genetics of tumor development have dramatically increased. The advent of new technologies like microarrays and genome-wide studies has provided the scientific community with the opportunity to analyse the genetic aspects involved in tumor formation in increasing detail. In the field of colorectal cancer development, the research is focusing more on the aspects of development and alterations necessary for its initiation rather than on the analysis of changes in late stage tumors. This approach proposes to identify molecular defects that are important for tumor formation at early stages with the aim to avoid all epiphenomena occurring in advanced cancer. Thus, it would be consequential to search lesions not yet neoplastic.

When we genotype the DNA of a patient affected by cancer, it is easy to identify many genetic mutations and copy number variations. Nevertheless, it is almost impossible to define if there is one, or a group of features, that are somehow causative and not only a consequence of other genetic variations. Mutations in the DNA sequence can cause misregulation in the proteomics balance or severe modifications and instability in the chromosomes, thus it is pivotal to understand their role in order to better define the steps causative of tumor formation.

In our project we focused on the analysis of thousands of different mutations, identified according to strict criteria. Our goal was to be able to select the ones that, when misregulated, could result into a cascade of factors, which would eventually lead to the development of tumor and possibly to all other subsequent modifications.

Hyperplastic polyps display an abnormal increment of the mucosal colon epithelium without participation of lower stromal structures. According to our hypothesis, their appearance is a key step for colorectal tumor development.

They could be non-permissive and allow the polyps to regress via apoptosis, as they normally do, or permissive. This latter case leads to a change in size and aspect of the hyperplastic polyps and allows its development as adenomatous polyps.

In one of our recent projects (publication ongoing) we observed diminished responses to apoptosis stimuli in lymphocytes of patients bearing hyperplastic polyps as compared to healthy control individuals. These results, and the fact that the development of hyperplastic polyps is associated with reduced apoptosis in the colon mucosa, suggest a genetic rather than a tissue-specific cause for these differences. We believe that apoptosis plays a crucial role, since it is the mechanism that would allow the organism to eliminate hyperplastic polyps and therefore we decided to investigate genes that are

directly connected with apoptosis pathways.

In order to identify possible risk loci, we analyzed the DNA of patients bearing hyperplastic polyps rather than the ones bearing adenomatous polyps or cancer patients. In this way we could focus our attention on the very first variations that may occur since patients affected by adenomatous polyps already present several independent genetic variations and somatic mutations. Their analysis would therefore not be determinant to identify the key variations responsible for the advance of the hyperplastic polyp to the adenomatous tumors.

In this study we performed a two-stage approach for genotyping polymorphisms in apoptosis-related pathways. The technique that has been used was the Goldengate technology from Illumina[24]. Instead of a genome wide approach (GWAS), we opted for a pre-defined approach. Our intention was to select candidate genes and defined polymorphisms in order to achieve a higher statistical power to detect alleles with a modest effect size[25, 26]. The "a priori" approach has been used in order to verify if the presence of certain mutations in genes of interest selected "a priori" could directly impact the function of these genes. The genes have been selected based on our hypothesis of apoptosis as modulator of hyperplastic polyps. All genes belonging to the apoptosis pathways or secondarily related to them have been considered. These included genes modulating the cell-cycle control, proliferation, adhesion, inflammation, DNA repair and anti-apoptosis. A total of 636 genes have been selected and analysed for SNPs.

Different SNPs can have very different effects depending on where they occur within the genome[27] (Figure 20).

Random SNPs (rSNPs) are the most common ones. They are considered unlikely to have any perceivable effect on the phenotype, since they are located in silent regions of our genome.

Gene-associated SNPs (gSNPs), located in introns alongside genes. They are useful for association studies between gene variants and certain phenotypes.

Coding SNPs (cSNPs). These are SNPs located in the coding regions of the genes. Possibly leading to non-synonymous changes, they could have a major influence on the function of proteins.

Phenotype-relevant SNPs (pSNPs). Those gSNPs and cSNPs that can influence the phenotype[27].

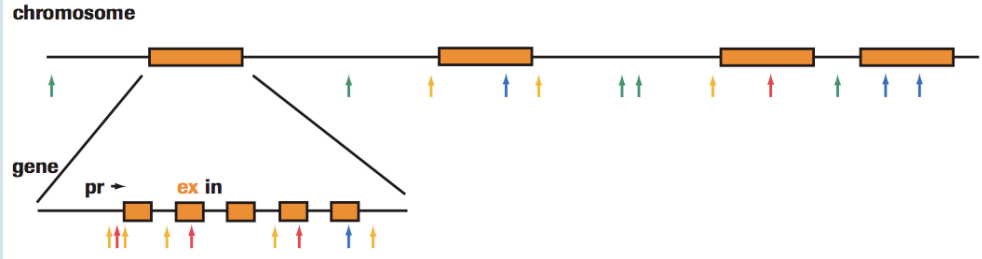


Figure 20: Schematic representation of the four SNP groups. rSNPs (green), gSNPs (yellow), cSNPs (blue) and pSNPs (red). (from: SNPs: the great importance of small differences. Roche)

Since our study was not a GWAS, but has been developed as an "a priori" approach, all the random SNPs (generally included in the GWAS approach) have not been considered.

The immediate upstream region (200-500 bps) of a transcribed sequence is considered the proximal promoter area[28], where the binding of multiple transcription factor proteins triggers expression. To be on a safe side and analyzing all the gSNPs we decided to include all the polymorphisms present in a region of 1000pb upstream and downstream each gene, according to the maximum number of bases accepted in the Assay Design Tool default settings in a GoldenGate study. In addition, we considered an area of 30bp upstream and downstream each exon to include all the gSNPs present in possible splicing areas and branch points[29], since a change in a single nucleotide in within the exon can lead to missense mutations resulting in a non-synonymous amino acid change, thus leading to modification in the protein structure and function. Otherwise they may represent nonsense mutations creating additional stop codons with protein truncation as consequence. Because of the change in apoptosis susceptibility we noticed in earlier studies, we figured that it would be prudent to focus on likely phenotypically relevant SNPS and eventually analyze all the cSNPs by including all the variations present in the exons.

These preliminary criteria were used for the development of a chip of selected SNPs to be genotyped in patients affected by hyperplastic polyps and in controls. In total 784 samples were successfully genotyped for 1464 unique SNPs.

Fourteen SNPs showed an initial association at the $p<0.01$ significance level. However, there is always the possibility that some of these associations identified could be based on serendipity, resulting in the identification of peculiarities of our specific sample rather than characteristics of the population. To encompass this problem, a validation is required. We refrained from using the cross-validation model (consisting in the splitting of the samples in two random groups with subsequent re-analysis) as we had the opportunity to validate the results in a completely new set of individuals.

The problem on the interpretation of results that include multiple statistical test is that, as showed by Zaykin et al[30], the probability of achieving at least one significant result when multiple test at the same significance level are taken, is greater than the significance level. This would lead to an increase probability of rejecting a null hypothesis when it would be inappropriate to do so. To avoid this usually the Bonferroni correction is used. However, Moran et al[31] argued for the rejection of the sequential Bonferroni adjustment as a solution to this problem while Nyholt suggested to rely only on replication as a means of verification[32]. Accordingly to these observations we decided to control the False Discovery Rate as it provides a much better compromise between Type I and Type II errors[21].

Nine single polymorphisms were positively identified after validation and FDR control as statistically associated with the presence of hyperplastic polyps: rs4709583 ($q=3x10^{-3}$); rs3806689 (q=0.0087); rs10476829 (q=0.0087); rs8110238 (q=0.087); rs1053516 (q=0.0087); rs1053530 (q=0.0087); rs7083622 (q=0.0087); rs4959774 (q=0.0087); rs11016073 (q=0.013). All these SNPs showed a significant association (q<0.01) with hyperplasic polyps presence.

The SNP with the ID Rs4709583 ($q=3x10^{-3}$) is located 20bp upstream exon 3 in the gene PARK2.

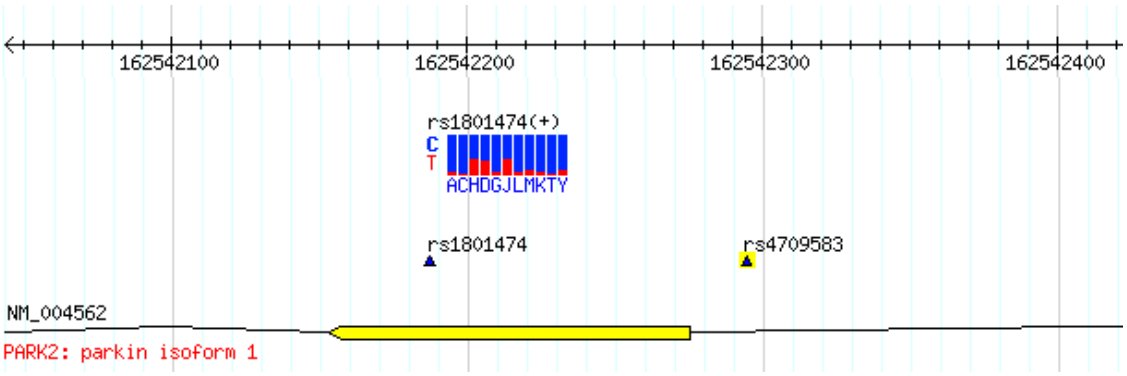


Figure 21: Hapmap visualization of rs4709583 in chr6, position 162622304 with a 500bp spanning region view. In yellow is visible the exon 3 of PARK2.

PARK2 is a gene that has been previously associated with Parkinson's disease[33]. It encodes Parkin, one of the E3 protein-ubiquitin ligases, and it is also a tumor suppressor gene[34, 35]. Mutations of this gene have been shown to lead to a functional loss of Parkin[36, 37] whereas deletion in PARK2 has recently been found to occur with higher frequency in sporadic colorectal cancer[38]. Poulogiannis and colleagues[38] showed that deficiency of *PARK2* expression is associated with adenomatous polyposis coli (APC) deficiency in human colorectal cancer.

Out of the 9 SNPs resulting positively associated, 6 have been found located in the 3'UTR or 5'UTR flanking region of their genes. These are:

- Rs1053516 located in the 3'UTR flanking region of *WDR48* (wd repeat domain 48).

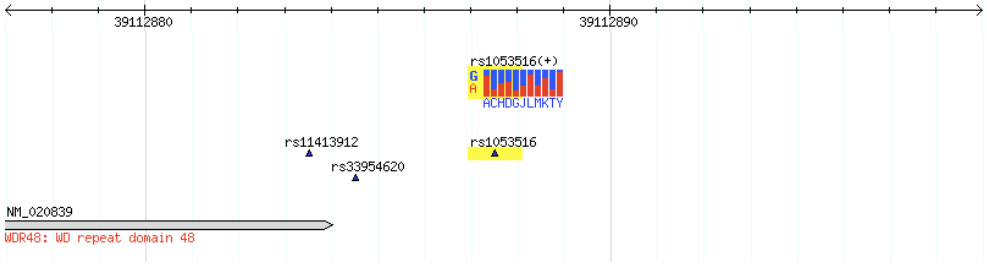


Figure 22: Hapmap visualization of rs1053516 in chr3 position 39137883 with a 21bp spanning region view. In gray is visible WDR48 3'UTR region

- Rs1053530 located in the flanking 3'UTR region of *GORASP1* (Golgi reassembly stacking protein 1); *GORASP1* codes for a membrane protein involved in establishing the stacked structure of the Golgi apparatus. It is a caspase-3 substrate, and cleavage of this protein contributes to Golgi fragmentation in apoptosis.

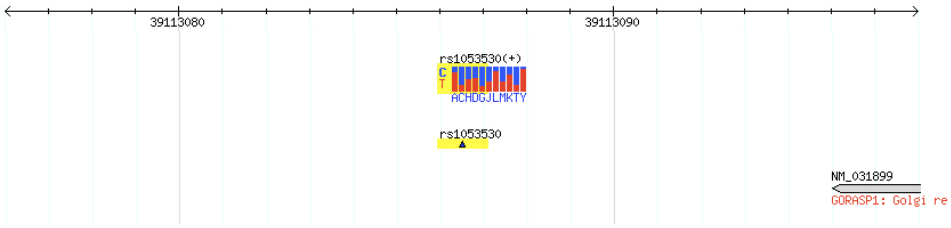


Figure 23: Hapmap visualization of rs1053530 in chr3 position 39138082 with a 21bp spanning region view. In gray is visible GORASP1 3'UTR region

- Rs8110238 flanking the 3'UTR of *NLRP2*, a gene coding for NALP proteins that are involved in the activation of caspase-1 via Toll-like receptors . When associated with PYCARD, NLRP2 activates CASP1, leading to the secretion of mature proinflammatory cytokine interleukin 1 beta IL1B. Interleukin 1 stimulates B-cell maturation and proliferation, and IL-1 proteins are involved in the inflammatory response.

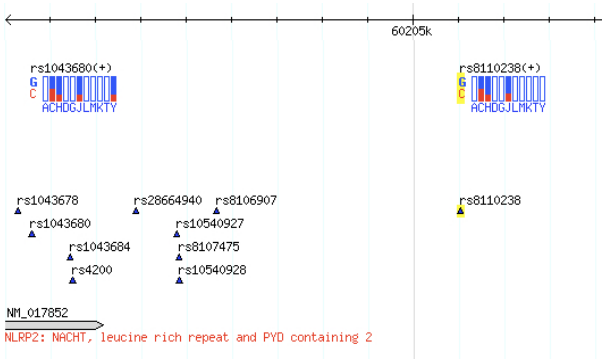


Figure 24: Hapmap visualization of rs8110238 in chr19 position 60205103 with a 2 Kbp spanning region view. In gray is visible NLRP2 3'UTR location

- Rs10476829 located in the flanking 3' UTR region of *HDAC3*.

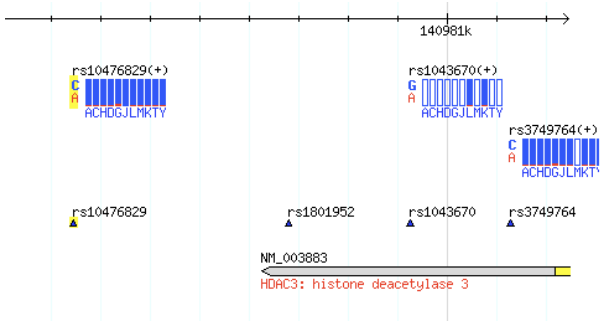


Figure 25: Hapmap visualization of rs10476829 in chr5 position 1410000062 with a 2 Kbp spanning region view. In gray is visible HDAC3 3'UTR region

HDAC3 is the gene encoding for Histone Deacetylase 3 (HDAC3), the enzyme which removes the acetyl groups from histones leading to chromatin condensation and transcriptionally inactive heterochromatin. HDACs have become one of the emerging targets for cancer therapy, and HDAC inhibitors (HDACi) have shown promising anticancer activities[28]. In fact inhibitors of this enzyme can induce cell cycle arrest and apoptosis in a variety of cancer cells[29, 30, 31].

The family of the HDAC can be subdivided into four classes. The class I HDACs, which includes HDAC1, HDAC2, HDAC3 and HDAC8 has been reported to be highly expressed in colon cancer[32, 33].

- Rs3806689 is located in the 5'UTR of kininogen 1 (*KNG1*), a gene that plays an important role in assembly of the kallikrein-kinin system.

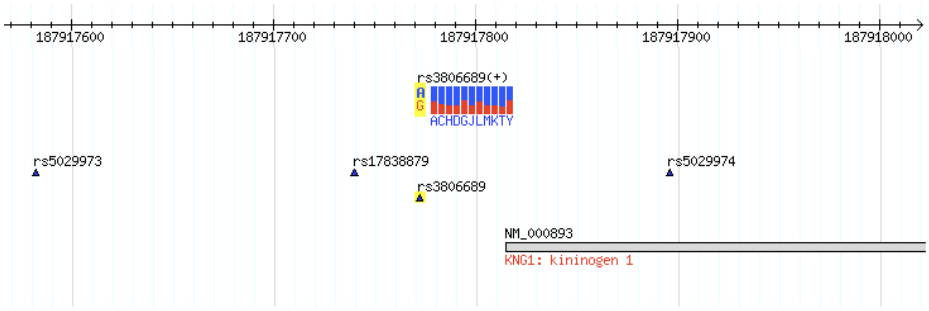


Figure 26: Hapmap visualization of rs3806689 in chr3 position 187917771 with a 500bp spanning region view. In gray is visible KNG1 5'UTR region

It has been shown that an alternative splicing of *KNG1* generates two different proteins, "high molecular weight kininogen" (HMWK), essential for blood coagulation and assembly of the kallikrein-kinin system, and "low molecular weight kininogen" (LMWK)[34]. The kallikrein-kinin system is known to modulate both the generation of bradykinin (BK), a potent mediator of inflammation, and the release of kallikrein in the intestinal extracellular space, associated with inflammatory bowel disease[35]. Moreover, bradykinin directly stimulates transcription and translation of TLR2 receptor by activation of protein kinase C and AKT[36] as demonstrated by Venegas et al**.**

- rs4959774 flanking 5' UTR of *RIPK1* (receptor (TNFRSF)-interacting serine-threonine kinase 1); the death domain kinase, RIP (receptor-interacting protein), is important for DNA damage-induced, p53-independent cell death as reported by Hur et al[48] .

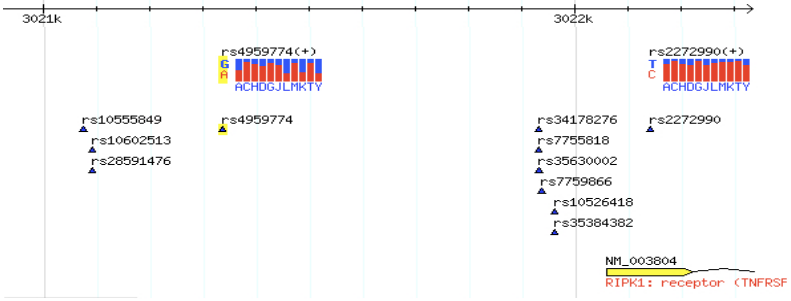


Figure 27: Hapmap visualization of rs4959774 in chr6 position 3021335 with a 2 Kbp spanning region view. In yellow is visible RIPk1 exon 1 location

Furthermore, in our project we could identify two cSNPs to be significantly associated with hyperplastic polyps. These cSNPs are rs7083622 and rs11016073 and both are non-synonymous SNPs located in exon 3 of *MKI67* (antigen identified by monoclonal antibody Ki-67), a gene located in chromosome 10. Rs7083622 leads to a change from Threonine to Asparagine, T [Thr] –> N [Asn]. Rs11016073 leads to a change from Isoleucine to Threonine I [Ile] ⇒ T [Thr] resulting in the loss of the hydrophobic side chain of Isoleucine.

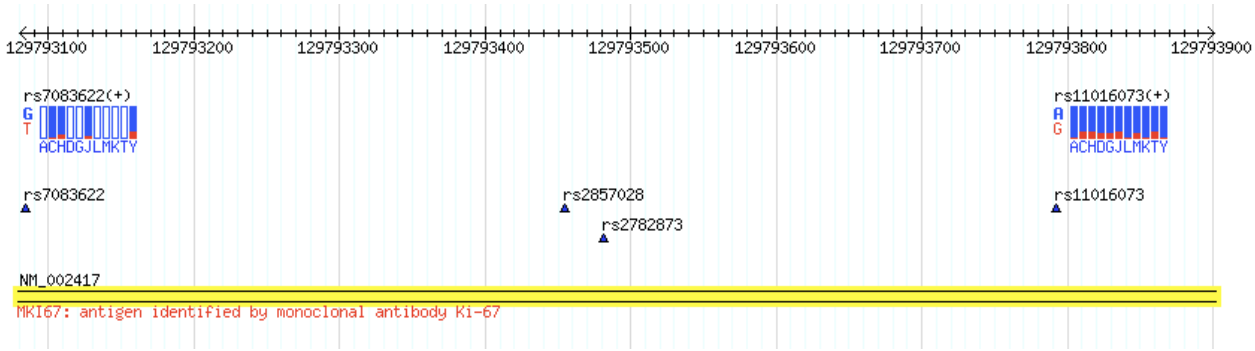


Figure 28: Hapmap visualization of rs7083622 and rs11016073 in chr10 positions 129793084 and 129793792 respectively, with a 821bp spanning region view. In yellow is visible MKI67 exon 3.

MKI67, broadly used as a diagnostic marker in various cancers[49], has been recently associated with prostate cancer[50]. It encodes the nuclear 359-KDa protein Ki-67, an absolute requirement for cell growth[51]. High Ki-67 expression is associated with cellular proliferation, multiple liver metastases with associated higher mortality and with a decreased interval from primary disease to metastasis[52,53].

Out of all the SNPs identified, we decided to first start the analysis of the SNP with the lowest q value, namely Rs4709583. This SNP, as already mentioned, is located in *PARK2* few bases upstream an acceptor splicing area. Usually, introns start with a GU sequence and end with an AG sequence in 5' to 3' direction. These are designated the splice donor and splice acceptor site, respectively. Together with the branch site that is located 20-50pb upstream of the acceptor site (identified by the sequence CU(A/G)A(C/U) where A is conserved in all genes) they constitute the splicing area. Rs4709583 was therefore studied for a possible role as splicing modulator. However, no evidence was found to suggest that Rs4709583 does affect splicing as transcripts analysed in cell lines harbouring the SNP do not express different transcript variants (exon skipping) or expression level as shown in comparison to house keeping gene expression. We therefore conclude that this SNP has no functional effects on gene expression or will affect protein size.

A second significant SNP, rs10476829, is being also investigated as possible modulator for *HDAC3* gene expression, as it is located in the 3'UTR. In-silico analysis of the 3' UTR region of the *HDAC3* gene revealed a second, (cryptic) polyadenylation signal which usage was confirmed by mRNA analysis.

So far expression screening and recombinant 3'UTR expression in a heterozygous cell line do not indicate differential regulation of the respective mRNAs. On the other hand, ongoing work focusing on analysis of HDAC3 mRNA synthesis and stability under cellular stress indicates an altered half life of the respective mRNAs. This result clearly shows the need to test this SNP in a homozygous cell line to unmask moderate and possibly stress related effects

So far we could not establish evidence suggesting that the *HDAC3* genotypes influence protein expression between wild type controls and mutated samples; however, since preliminary data suggest a role for rs10476829 on RNA stabilization, promising studies are still on going.

All of the SNPs identified in this study have no confirmed or suggested functional effects from previous association studies and there is no evidence in literature; they have never been associated or related to any pathology and therefore further studies will have to be performed to investigate their possible biological roles as well as possible interaction between the genes involved.

Our aim was to identify possible linking markers between apoptosis and early colorectal tumor lesions. The results we obtained so far suggest with a high significant value a role for SNP rs4709583 and eight other SNPs as key variations. To our knowledge, these SNPs have never been associated with any pathology in general, and more specifically related to the development of hyperplastic polyps. These findings open new prospective in this field, as we believe that these mutated variations could lead to modifications in the apoptosis machinery, resulting therefore in an increase of the chances to proceed from hyperplastic polyp down in the cancer pathway to the development of colorectal cancer. Further studies in order to understand the possible role of these SNPs and their genes products in the development of colorectal cancer would be therefore strongly recommended.

Bibliography

1. Becker N. and Holzmeier S.: Cancer Epidemiology German Cancer Research Center Heidelberg.

2. Yang V.W. et al; Colon Cancer: An Update and Future Directions; Gastroenterology; 2010: 138:2027-2028

3. Jasperson K.W. et al; Hereditary and Familial Colon Cancer; Gastroenterology; 2010: 138:2044-2058

4. Lipkin M. et al; Dietary Factors in Human Colorectal Cancer;
   Annu. Rev. Nutr; 1999: 19:545-86

5. Chan A.T. and Giovannucci E.L.; Primary Prevention of Colorectal Cancer; Gastroenterology; 2010: 138:2029-2043

6. Bastian B.C.; Understanding the progression of melanocytic neoplasia using genomic analysis: from fields to cancer;
   Oncogene; 2003: 3081-3086

7. Jen J. et al; Molecular Determinants of Dysplasia in Colorectal Lesions; Cancer Research; 1994: 54:5523-5526

8. Roncucci et al; Aberrant crypt foci in colorectal carcinogenesis. Cell and crypt dynamics;
   Cell Prolif; 2000: 33(1): 1-18

9. Arnold C.N. et al; Molecular Pathogenesis of Colorectal Cancer;
   Cancer; 2005: 104;10; 2035-2047

10. Knöpnadel et al; Epidemiologie und gesundheitsökonomischeBedeutung des Darm-krebses in Deutschalnd.
    Internist; 2003

11. Kinzler K.W. and Vogelstein B.; Lessons from Hereditary Colorectal Cancer;
    Cell; 1996: Vol. 87, 159-170

12. Tenesa A. and Dunlop M.G.; New Insights into the aetiology of colorectal cancer from genome-wide association studies;
    Nature Reviews; 2009: 10

13. Niittymäki et al; Low-Penetrance Susceptibility Variants in Familial Colorectal Cancer;
    Cancer Epidemiol Biomarkers Prev; 2010: 19(6)

14. Zisman A.L. et al; Associations Between the Age at Diagnosis and Location of Colorectal Cancer and the Use of Alcohol and Tobacco;
    Arch Intern Med; 2006: 166

15. Geboes et al; Pathology of early lower GI cancer;
    Best Pract Res ClinGastroent; 2005: 19:6

16. Boland C.R. and Goel A.; Microsatellite Instability in Colorectal Cancer; Gastroenterology; 2010: 138:2073-2087

17. Tomlinson I.P.M. et al; COGENT (COlorectal cancer GENeTics): an international consortium to study the role of polymorphic variation on the risk of colorectal cancer; British Journal of Cancer; 2010: 102: 447-454

18. Leggett B. and Whitehall V.; Role of the Serrated Pathway in Colorectal Cancer Pathogenesis. Gastroenterology; 2010: 138:2088-2100

19. Nittka S. et al; The human tumor suppressor CEACAM1 modulates apoptosis and is implicated in early colorectal tumorigenesis; Oncogene; 2004: 23:9306-9313

20. Shively J.E.; CEACAM1 and hyperplastic polyps: new links in the chain of events leading to colon cancer; Oncogene; 2004: 23**:**9303–9305

21. Nakagawa S.; A farewell to Bonferroni: the problems of low statistical power and publication bias; Behavioral Ecology; 2004: 15 (6): 1044–1045

22. García L.V.; Controlling the false discovery rate in ecological research; Trends in Ecology and Evolution; 2003: 18: 553-554

23. Benjamini Y. and Hochberg Y.; Controlling the false discovery rate: a practical and powerful approach to multiple testing; Journal of the Royal Statistical Society; 1995: 57: 289-300

24. Fan J.B. et al; Highly parallel genomic assays; Nat. Rev. Genet.; 2006: 7(8): 632-633

25. Wu M.C. et al; Powerful SNP-Set Analysis for Case-Control Genome-wide Association Studies; Am. J. Hum. Genet.; 2010: 86(6):929-942

26. Zhu M. and Zhao S.; Candidate gene identification approach: progress and challenges; Int. J. Biol. Sci.; 2007; 3(7):420-427

27. Roche; SNPs: the great importance of small differences

28. Goñi J.R. et al; Determining promoter location based on DNA structure first-principles calculations; Genome Biology; 2007: 8:R263

29. Clancy S.; RNA splicing: introns, exons and spliceosome; Nature Education; 2008: 1(1)

30. Zaykin D.V. et al; Truncated product method for combining p-values. Genet. Epidemiol.; 2002: 22: 170-185.

31. Moran M.D.; Arguments for rejecting the sequential Bonferroni in ecological studies; Oikos; 2003: 100:403-405

32. Nyholt D.R.; Genetic case-control association studies – correcting for multiple testing; Hum Genet; 2001: 109:564-565

33. Kitada T. et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature; 1998: 392:605– 608

34. Cesari R. et al; Parkin, a gene implicated in autosomal recessive juvenile parkinsonism, is a candidate tumor suppressor gene on chromosome 6q25- q27. Proc Natl Acad Sci; 2003: 100:5956 –5961.

35. Veeriah S et al; Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. Nat Genet; 2010:42: 77– 82.

36. Yamamura Y.; The long journey to the discovery of PARK2. Neuropathology; 2010: 30:495-500

37. Chung S.J. et al; Common variants in PARK loci and related genes and Parkinson's disease. Movement Disorders; 2011: 26:2

38. Poulogiannis G. et al; PARK2 deletions occur frequently in sporadic colorectal cancer and accelerate adenoma development in Apc mutant mice. PNAS; 2010: 107:34

39. Archer S.Y. and Hodin R.A.; Histone acetylation and cancer; Curr Opin Genet Dev; 1999: 9

40. Minucci S. and Pelicci P.G.; Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. Nat Rev Cancer; 2006: 6

41. Mariadason J.M.; HDACs and HDAC inhibitors in colon cancer. Epigenetics: official journal of the DNA Methylation Society; 2008 :3

42. Spurling C.C. et al; HDAC3 Overexpression and Colon Cancer Cell Proliferation and Differentiation; Molecular Carcinogenesis; 2008: 47:137-147

43. Bolden J.E., Peart M.J. and Johnstone R.W.; Anticancer activities of histone deacetylase inhibitors; Nature reviews Drug discovery; 2006: 5:769–784

44. Kay et al; A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2; Neurology; 2010

45. Cheung P.P. et al; Chromosomal mapping of human kininogen gene (KNG) to 3q26-qter; Cytogenet Cell Genet; 1992: 59: 24-26

46. Devani M. et al; Kallikrein-kinin system in inflammatory bowel diseases: Intestinal involvement and correlation with the degree of tissue inflammation; Dig Liver Dis; 2005: 37(9):665-73

47. Gutiérrez-Venegas G. and Arreguín-Cano J.A.; Bradykinin promotes TLR2 expression in human gingival fibroblasts; Int Immunopharmacol; 2011: 11(12):2079-85

48. Hur G.M. et al; The death domain kinase RIP has an important role in DNA damage-induced, p53-independent cell death; J Biol Chem; 2006: 281(35):25011-7

49. Hou, Y.Y. et al; MicroRNA-519d targets MKi67 and suppresses cell growth in the hepatocellular carcinoma cell line QGY-7703; Cancer Lett; 2011: 307:182-190

50. Mitra A.V. et al; Elevated expression of Ki-67 identifies aggressive prostate cancers but does not distinguish BRCA1 or BRCA2 mutation carriers; Oncology reports; 2010: 23(2):2

51. Schluter C. et al; The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins; J Cell Biol; 1993: 123:513-522

52. Petrowsky H. et al; Relevance of Ki-67 antigen expression and K-ras mutation in colorectal liver metastases; Eur J Surg Oncol; 2001: 27:80–7.

53. Garrett M.N. et al; KRAS Mutation Correlates With Accelerated Metastatic Progression in Patients With Colorectal Liver Metastases; Ann Surg Oncol: 2010: 17:572–578

54. Nittka S. et al; The CEACAM1-mediated apoptosis pathway is activated by CEA and triggers dual cleavage of CEACAM1; Oncogene; 2008: 27:3721-3728

55. Jass J.R.; Hyperplastic Polyps of the Colorectum – Innocent or Guilty?; Dis Colon Rectum; Feb 2001

56. Pretlow T.P. and Pretlow T.G.; Mutant KRAS in aberrant crypt foci (ACF): initiation of colorectal cancer? Biochim Biophys Acta; 2005: 25:1756(2):83-96

57. Galvan A., Ioannidis J.P.A. and Dragani T.A.; Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer; Trends in Genetics; 2010: 26:3

58. Jass JR et al; Neoplastic progression occurs through mutator pathways in hyperplastic polyposis of the colorectum;
Gut; 2000: 47:43-49

59. Jass JR; Hyperplastic polyps and colorectal cancer: is there a link?;
Clin Gastroenterol Hepatol; 2004: 2:1

60. Fearon E. R. & Vogelstein B. A; Genetic model for colorectal tumorigenesis.
Cell; 1990: 61:759-767

61. Powell S. M. et al; APC mutations occur early during colorectal tumorigenesis.
Nature; 1992: 359:235-237

62. Goss K. H. and Groden J.; Biology of the adenomatous polyposis coli tumor suppressor;
J. Clin. Oncol.; 2000: 18:1967-1979

63. Esteller M. et al; Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer;
Cancer Res; 2000: 60:4366:4371

64. Fearnhead N., Britton M. and Bodmer W; The ABC of APC;
Hum. Mol. Genet.; 2001: 10:721

65. Haug et al; Sensitivity and specificity of faecal tumour M2 pyruvate kinase for detection of colorectal adenomas in a large screening study;
Br J Cancer; 2008 :8:99

66. Hundt et al; Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection;
Ann Intern Med; 2009: 150:162-9

67. Tateyama H. et al; Apoptosis Index and Apoptosis-Related Antigen Expression in Serrated Adenoma of the Colorectum;
The American Journal of Surgical Pathology; 2002: 26(2):249-256

68. Jass J.R. et al; Evolution of colorectal cancer: Change of pace and change of direction;
Journal of Gastroenterology and Hepatology; 2002: 17:17-26

69. Zhao X. et al; An Integrated View of Copy Number and Allelic Alterations in the Cancer Genome Using Single Nucleotide Polymorphism Arrays; Cancer Research; 2004: 64:3060-3071

70. Brockton N. et al; *N*-Acetyltransferase Polymorphisms and Colorectal Cancer: A HuGE Review;
American Journal of Epidemiology; 1999: 151:9

71. Bhatti P. et al; Candidate Single Nucleotide Polymorphism Selection using Publicly Available Tools: A Guide for Epidemiologists;
American Journal of Epidemiology; 2006: 164:8

72. Webb E.L. et al; Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives;
Human Molecular Genetic; 2006: 15:21

73. Longmate J.A.; Complexity and Power in Case-Control Association Studies;
Am. J. Hum. Genet. 2001: 68:1229-1237

74. Pechlivanis S. et al; Insulin pathway related genes and risk of colorectal cancer: INSR promoter polymorphism shows a protective effect; Endocrine-Related Cancer; 2007 :14:733-740

75. Teo Y.Y. et al; A genotype calling algorithm for the IlluminaBead Array platform.
Bioinformatics; 2007: 23:20

76. Tomlinson I.P.M. et al; A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3; Nature Genetics; 2008: 40:5

77. Tenesa A. et al; Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21;
Nature Genetics; 2008: 40:5

78. Grant S.F.A. and Hakonarson H.; Microarray Technology and Applications in the Arena of Genome-Wide Association.
Clinical Chemistry; 2008: 54:7

79. Fleshner P., Slater G. and Aufses A.H.A.; Age and Sex Distribution of Patients with Colorectal Cancer;
Dis. Col. & Rect.; 1989

80. Cacev T. et al; Influence of interleukin-8 and interleukin-10 on sporadic colon cancer development and progression.
Carcinogenesis; 2008: 29:8

81. Pande M. et al; Genetic Variation in Genes for the Xenobiotic-Metabolizing Enzymes CYP1A1, EPHX1, GSTM1, GSTT1, and GSTP1 and Susceptibility to Colorectal Cancer in Lynch Syndrome.
Cancer Epidemiol Biomarkers Prev; 2008: 17(9)

82. COGENT Study; Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer;
Nature Genetics; 2008: 40:12

83. Markowitz S.D. and Bertagnolli M.M.; Molecular Basis of Colorectal Cancer.
The New England Journal of Medicine; 2009: 361:2449-60

84. Bodmer W. and Tomlinson I.; Rare genetic variants and the risk of cancer;
Current Opinion in Genetics & Development; 20:262-267; 2010

85. Center M.M. et al; Worldwide Variations in Colorectal Cancer.
CA Cancer J Clin; 2009: 59:366-378

86. Chou C.W. et al; HDAC Inhibition Decreases the Expression of EGFR in Colorectal Cancer Cells. PLoS ONE; 2011: 6:3

87. Chung C.C. et al; Genome-wide association studies in cancer - current and future directions. Carcinogenesis; 2010: 31:1

88. Ghavami S. et al; Apoptosis and cancer: mutations within caspase genes. J Med Genet; 2009: 46:497-510

89. Gibson G.; Hints of hidden heritability in GWAS. Nature Genetics; 2010: 42:558-560

90. Goodman J.E. et al; Exploring SNP-SNP interactions and colon cancer risk using polymorphism interaction analysis. Int. J. Cancer; 2006: 118:1790-1797

91. Imyanitov E.N.; Gene polymorphisms, apoptotic capacity and cancer risk; Hum Genet; 2009: 125:239-246

92. Jubb A.M. et al; Methylation and colorectal cancer; Journal of Pathology; 195:111-134; 2001

93. Terzic J. et al; Inflammation and Colon Cancer; Gastroenterology; 2010: 138:2101-2114

94. Wallace M.B. and Kiesslich R.; Advances in Endoscopic Imaging of Colorectal Neoplasia; Gastroenterology; 2010: 138:2140-2150

95. Liebermann D.; A Call to Action — Measuring the Quality of Colonoscopy; N. Engl. J. Med.; 2006: 355:24

96. West N.J. et al; Apoptosis in the Colonic Crypt, Colorectal Adenomata, and Manipulation by Chemoprevention; Cancer Epidemiol Biomarkers Prev; 2009: 18:6

97. Hazra A. et al; Large-Scale Evaluation of Genetic Variants in Candidate Genes for Colorectal Cancer Risk in the Nurses' Health Study and the Health Professionals' Follow-up Study; Cancer Epidemiol Biomarkers Prev; 2008: 17:311-319

98. Perneger T.V.; What's wrong with Bonferroni adjustments; Brit. Med. J.; 1998: 316: 1236-1238

Appendices

Complete list of the 1536 SNPs included in the chip:

rs1001814	rs10455909	rs10873081	rs11247641	rs1191778	rs12681513	rs13423990	rs1566464
rs10022693	rs1045976	rs10876450	rs1127575	rs11920551	rs1268371	rs1345186	rs1570920
rs10041762	rs10463292	rs10879312	rs1127717	rs11933240	rs1274495	rs1353252	rs1573132
rs10051026	rs1046864	rs10890800	rs1127732	rs11936535	rs12747154	rs13571	rs1574220
rs1005165	rs10471960	rs10924968	rs1129474	rs11940114	rs12757286	rs135750	rs157704
rs10063294	rs10476829	rs10925027	rs1130409	rs12027359	rs12760299	rs13706	rs157705
rs1007381	rs10476907	rs10925028	rs1130861	rs12040042	rs12765065	rs1372438	rs157706
rs10082391	rs1047972	rs10946282	rs1131877	rs12040523	rs12770335	rs1382573	rs158688
rs1009977	rs10484575	rs10967717	rs1133782	rs1206196	rs12777740	rs1385699	rs159153
rs1010231	rs1048906	rs10967719	rs1135983	rs12068365	rs12792390	rs13881	rs1614984
rs10133354	rs1048945	rs10999152	rs1136138	rs12076604	rs12793024	rs1389723	rs1624701
rs10133460	rs1049174	rs11016071	rs1136410	rs12088254	rs12818627	rs1397529	rs1635576
rs10137191	rs1049306	rs11016073	rs1138272	rs12095268	rs12819075	rs139998	rs1654495
rs10170774	rs1049493	rs11016076	rs1138294	rs12109242	rs12821290	rs140054	rs1656922
rs10180391	rs1049728	rs11016850	rs1138357	rs12121499	rs12830368	rs140056	rs1661309
rs10180445	rs10498475	rs11022395	rs1138358	rs12123647	rs12836234	rs1400657	rs16802
rs10190751	rs1050026	rs11024148	rs1138374	rs12123975	rs12872310	rs140504	rs1681663
rs10192752	rs10500880	rs11033019	rs1139139	rs12137901	rs12878534	rs140692	rs16829782
rs10215499	rs10502001	rs11047892	rs1143627	rs12141641	rs1291863	rs1407300	rs1683356
rs10239188	rs10502045	rs11053646	rs1144393	rs12143231	rs12918952	rs1409837	rs16836789
rs1028488	rs10505477	rs11054704	rs1150258	rs1214615	rs12953717	rs1413218	rs16845711
rs1034528	rs10506287	rs1105769	rs1151845	rs12152411	rs12969690	rs1413219	rs16861474
rs1035140	rs10506591	rs11063099	rs11556218	rs12155747	rs12973968	rs14133	rs16883235
rs1035281	rs1050767	rs1106966	rs11568943	rs12233352	rs12993800	rs1413902	rs16886463
rs1039559	rs1051006	rs11078476	rs11569017	rs12240031	rs12998837	rs1421896	rs16900793
rs1039560	rs10512203	rs11084419	rs11574685	rs12240532	rs13006529	rs1431195	rs16924899
rs10399805	rs10512747	rs11089197	rs11593766	rs12249629	rs13009282	rs1431196	rs1692821
rs10399931	rs10521711	rs11102930	rs11598673	rs12257763	rs13010627	rs1441204	rs16933823
rs10401178	rs1052576	rs1110337	rs11600510	rs12269215	rs13023958	rs1446963	rs16940
rs10402677	rs1053516	rs11106	rs11601671	rs12289513	rs13047599	rs1467558	rs16942
rs10403164	rs1053530	rs11127	rs11613534	rs12298173	rs13081248	rs1474593	rs16944
rs10405032	rs1055333	rs11129565	rs11615947	rs12299278	rs13097796	rs1474594	rs16945
rs10410487	rs1055690	rs11129566	rs11626603	rs12316284	rs13115636	rs1483977	rs16947383
rs10411771	rs1056719	rs11130575	rs11643855	rs12320259	rs13156380	rs1483978	rs16949649
rs10412761	rs1059831	rs11134720	rs11666350	rs12338	rs13167280	rs1494555	rs16949924
rs10417652	rs1060742	rs11135703	rs11672373	rs12344755	rs13168328	rs1494558	rs1695
rs10417878	rs1061622	rs11145047	rs11681451	rs12410036	rs13181	rs1503392	rs16972197
rs1041935	rs1062808	rs11152093	rs11708753	rs12425275	rs13182512	rs1508146	rs16975132
rs1041981	rs1063535	rs11169690	rs11709791	rs12461110	rs13189180	rs1516790	rs16975865
rs10422063	rs1065381	rs11169691	rs11720451	rs12462372	rs13211268	rs1534904	rs1699089
rs1042381	rs10711	rs11169696	rs11738611	rs12491108	rs1322178	rs1545228	rs17010021
rs10424046	rs10732438	rs11177386	rs11739823	rs12500797	rs13223993	rs1545762	rs17012739
rs1042838	rs10787227	rs11187083	rs11741807	rs12562182	rs1322997	rs1547012	rs17032278
rs1043673	rs10787984	rs11187120	rs11743001	rs12565967	rs13242	rs1548915	rs17032283
rs10438355	rs1080199	rs11196418	rs11767165	rs125701	rs13254954	rs1549760	rs17057444
rs1044141	rs10802438	rs11203366	rs11780874	rs12575005	rs13265018	rs1550659	rs17057718
rs1044169	rs10823258	rs11203367	rs11788639	rs12575838	rs13269700	rs1553850	rs17066616
rs1044240	rs10825492	rs1122470	rs11810295	rs12600392	rs13272847	rs1554338	rs1707321

rs1044243	rs10825493	rs11226613	rs1182136	rs12616005	rs1329600	rs1554753	rs1707322
rs10444404	rs10845552	rs1123669	rs11837144	rs12624329	rs13306449	rs1558254	rs17073227
rs1044813	rs10846094	rs11243438	rs11870749	rs12631504	rs13312735	rs1558355	rs17092079
rs1045411	rs10851733	rs11244787	rs11874761	rs12657663	rs13343612	rs1558529	rs17103597
rs1045485	rs10864367	rs11246048	rs11894993	rs1266379	rs1334811	rs1559195	rs17107764
rs17108008	rs1785878	rs1984003	rs2076363	rs2236338	rs2295612	rs2452993	rs291085
rs17108046	rs1792773	rs1990760	rs2076377	rs2237051	rs2295614	rs2455013	rs2921787
rs17115479	rs1799845	rs1994209	rs2076389	rs2239364	rs229566	rs2457426	rs2929971
rs17125286	rs1799854	rs2000466	rs2077781	rs2239666	rs2295814	rs2464810	rs2929972
rs17126153	rs1799933	rs2003334	rs2084651	rs2240	rs2296636	rs2471551	rs293796
rs17128916	rs1799950	rs2004414	rs210134	rs2240108	rs2297370	rs2486966	rs2952271
rs17162549	rs1799964	rs2004888	rs2108411	rs2241669	rs2297743	rs2493392	rs2977995
rs171649	rs1799966	rs2007	rs210936	rs2243191	rs2298019	rs2494732	rs2986754
rs17177522	rs1799977	rs2013146	rs210937	rs2243525	rs2298821	rs2494750	rs299163
rs17178365	rs1800058	rs2020902	rs2112834	rs2243791	rs2301134	rs2511837	rs2997922
rs17188012	rs1800629	rs2020911	rs2125078	rs2244063	rs2301241	rs25487	rs3006488
rs17192524	rs1800630	rs2027790	rs2126053	rs2246266	rs2301242	rs2575390	rs300917
rs17209237	rs1800682	rs2035653	rs2142097	rs2247491	rs2301251	rs2583759	rs3019294
rs17225749	rs1800683	rs2041296	rs2147503	rs2247510	rs2301522	rs2589515	rs302453
rs17234734	rs1800793	rs20417	rs2152143	rs2250889	rs2301686	rs2594257	rs302462
rs17246395	rs1800794	rs2043211	rs2155548	rs2267172	rs2302196	rs2602141	rs302475
rs17257293	rs1800795	rs2044693	rs2155587	rs227053	rs2302197	rs263439	rs304077
rs17257357	rs1800796	rs2050329	rs2158467	rs2270953	rs2302524	rs263476	rs306457
rs17259252	rs1801278	rs20576	rs2174670	rs2270954	rs2302685	rs2651639	rs306496
rs17286683	rs1801516	rs2066496	rs2217900	rs2271660	rs2302757	rs2655060	rs306506
rs17287758	rs1803919	rs2066804	rs2227262	rs2272653	rs2302759	rs267172	rs306507
rs17289745	rs1805075	rs2066906	rs2227310	rs2273535	rs2302860	rs26762	rs306508
rs17293823	rs1805318	rs2066964	rs2227928	rs2273715	rs2303191	rs2684941	rs3092921
rs17323082	rs1805321	rs2069186	rs2228059	rs2274649	rs2303377	rs269933	rs3092946
rs17349	rs1805405	rs2069391	rs2228139	rs2274756	rs2303428	rs2699887	rs3092993
rs1735475	rs1805419	rs2069442	rs2228526	rs2275053	rs2304277	rs2700990	rs3093059
rs1735476	rs1805794	rs2069456	rs2228528	rs2275073	rs2304466	rs2705897	rs3093668
rs17357206	rs1805800	rs2069492	rs2228529	rs2275075	rs2305229	rs2715555	rs3094124
rs1736088	rs1809238	rs2069707	rs2229032	rs2275913	rs2305382	rs2717548	rs310224
rs17374170	rs1823035	rs2069727	rs2229080	rs2276864	rs2306536	rs2725832	rs310229
rs17375948	rs1835959	rs2069728	rs2229094	rs2277562	rs2306696	rs2736726	rs3104233
rs17419586	rs1862430	rs2069762	rs2229603	rs2278845	rs2307272	rs2742112	rs3118865
rs17420378	rs1863332	rs2069776	rs2230229	rs2280167	rs2308962	rs2748249	rs3118866
rs17447587	rs1864223	rs2069861	rs2230600	rs2280845	rs231150	rs2782877	rs3128495
rs1748020	rs1866140	rs2070179	rs2230804	rs2281082	rs2318785	rs2793289	rs3131713
rs17499125	rs1869258	rs2070180	rs2231292	rs2281860	rs2328537	rs2799511	rs3132291
rs17506789	rs1870584	rs2070584	rs2231439	rs2282537	rs2341468	rs2802268	rs3136228
rs17512051	rs1871042	rs2070724	rs2231440	rs2287153	rs2341469	rs2805910	rs3136586
rs17561	rs1871499	rs2070766	rs2233113	rs2287250	rs2355676	rs281505	rs3136614
rs17576	rs1883847	rs2070864	rs2233336	rs2288068	rs2366807	rs281506	rs3136618
rs17649845	rs1885097	rs2071229	rs2233406	rs2289043	rs2366894	rs2817801	rs3138055
rs17697978	rs1887429	rs2071368	rs2233659	rs2289539	rs2394644	rs28200	rs3143
rs17698002	rs190108	rs2071431	rs2233660	rs2289851	rs2400509	rs2838692	rs3176352
rs17699561	rs1932437	rs2071496	rs2233667	rs2290087	rs2400510	rs2853550	rs3181309
rs17699678	rs1946518	rs2071498	rs2233873	rs2290207	rs2403590	rs2853887	rs3181318
rs17700406	rs1946612	rs2071590	rs2234167	rs2290842	rs2412504	rs2862833	rs3181338
rs17710008	rs1946613	rs2072239	rs2234212	rs2290854	rs241697	rs28683503	rs3181354
rs17750575	rs196939	rs2072668	rs2234391	rs2290889	rs2426130	rs28763973	rs318473
rs17775180	rs197388	rs2073498	rs2235582	rs2291967	rs242725	rs2887501	rs318475
rs17794612	rs197412	rs2074451	rs2236028	rs2293399	rs2437770	rs2907749	rs3194051
rs17800727	rs197414	rs2075241	rs2236137	rs2293889	rs2439593	rs290986	rs3203777
rs17807642	rs1977420	rs2075267	rs2236141	rs2294998	rs2439831	rs290987	rs3212153

rs17829281	rs1980191	rs2075820	rs2236142	rs2295402	rs2440683	rs291084	rs3212245
rs3212247	rs3742483	rs3801238	rs4253042	rs4648065	rs4820294	rs556052	rs6151533
rs3212249	rs3742704	rs3806232	rs4253047	rs4648072	rs4821494	rs557806	rs6151535
rs3212254	rs3743135	rs3806689	rs4253055	rs4648085	rs483345	rs558289	rs6151586
rs3213176	rs3743477	rs3807186	rs4253067	rs4648099	rs483638	rs560191	rs6151588
rs3213182	rs3744000	rs3808916	rs4253077	rs4648306	rs4841601	rs567193	rs6151589
rs3213183	rs3744549	rs3809470	rs4253079	rs4648308	rs4842196	rs5742612	rs619030
rs3213403	rs3745231	rs3810366	rs4253099	rs4652	rs4845144	rs5742620	rs619129
rs3218076	rs3745980	rs3810816	rs4253101	rs4658574	rs4858828	rs5742725	rs619881
rs322351	rs3746006	rs3811699	rs4253106	rs4674107	rs4861358	rs5742729	rs623383
rs324131	rs3747517	rs3811740	rs4253121	rs4674258	rs486907	rs5742875	rs627839
rs334558	rs3749164	rs3814342	rs4253126	rs4674260	rs4871857	rs5742878	rs627928
rs3398	rs3750511	rs3814394	rs4253132	rs4678620	rs490934	rs5742881	rs629849
rs344781	rs3750512	rs3814395	rs4253162	rs4679087	rs4910577	rs5742914	rs630852
rs347033	rs3750751	rs3815029	rs4253165	rs4686904	rs4910852	rs5742915	rs630923
rs348337	rs3751093	rs3817085	rs4253166	rs4698803	rs4925648	rs5743806	rs6330
rs348390	rs3751362	rs3819024	rs4253193	rs4704552	rs4930702	rs5743810	rs634742
rs350886	rs3752383	rs3819025	rs4253211	rs4708955	rs4939827	rs5743827	rs635596
rs350911	rs3753036	rs3820390	rs4253234	rs4709390	rs4940109	rs5744168	rs638050
rs35091611	rs3753869	rs3821815	rs4253441	rs4709583	rs4950928	rs5744171	rs638203
rs361072	rs3755525	rs3822328	rs4253444	rs471028	rs4959774	rs5745568	rs641320
rs365238	rs3757095	rs3824886	rs4253449	rs4711998	rs497900	rs5745608	rs6427628
rs367825	rs3757462	rs3825073	rs4280262	rs471692	rs4986850	rs5745687	rs6429593
rs3730089	rs3759880	rs3826007	rs4292025	rs4719799	rs4986852	rs5746026	rs6432774
rs3730446	rs3760468	rs3826803	rs4295944	rs4720265	rs4987082	rs5749529	rs643780
rs3730447	rs3760469	rs3827519	rs4301834	rs4720672	rs4987872	rs5750813	rs6458556
rs3730448	rs3760895	rs3829062	rs433429	rs4724729	rs4988372	rs5754315	rs6462719
rs3731249	rs3761055	rs3829490	rs4351280	rs4727666	rs4988453	rs581000	rs6462720
rs3732183	rs3761846	rs3844283	rs4392990	rs4729673	rs5029973	rs584427	rs6462722
rs3732294	rs3762797	rs386830	rs441827	rs4730204	rs5030335	rs5945776	rs646558
rs3732413	rs3764340	rs3917330	rs4448378	rs4733047	rs5030416	rs595429	rs6478109
rs3733681	rs3764342	rs3917332	rs4451921	rs4739139	rs5030606	rs5956583	rs6484746
rs3733875	rs3764482	rs3917365	rs4464148	rs4745580	rs5031036	rs5963153	rs648802
rs3734000	rs3764795	rs3917444	rs448162	rs475007	rs506504	rs5995133	rs649101
rs3734028	rs3765545	rs3917449	rs4525607	rs475043	rs507438	rs600739	rs652311
rs3734763	rs3765730	rs3917452	rs4558496	rs4750685	rs507879	rs6007806	rs6537355
rs3735135	rs3768235	rs3918270	rs456671	rs4752825	rs510432	rs6008684	rs6557634
rs3736149	rs3768889	rs39509	rs4570508	rs4752983	rs514492	rs6008749	rs6570593
rs3736235	rs3769148	rs396991	rs4579555	rs4755392	rs514552	rs6011441	rs6576776
rs3736242	rs3772985	rs40076	rs458906	rs4756196	rs51500	rs6011456	rs657867
rs3736430	rs3773840	rs412050	rs4595501	rs4757947	rs522616	rs6011458	rs658560
rs3737498	rs3781053	rs4130513	rs459552	rs4760169	rs523104	rs6011471	rs6587756
rs3738200	rs3781684	rs4134994	rs4611608	rs4760653	rs527245	rs6011473	rs6592630
rs3738372	rs3783299	rs4145212	rs461970	rs476632	rs533558	rs6011478	rs6602472
rs3738952	rs3783508	rs4149371	rs4627097	rs4779584	rs5361	rs6011488	rs6603883
rs3740199	rs3783521	rs4149570	rs4644	rs4783121	rs5362	rs6020846	rs661561
rs3740383	rs3783557	rs4149584	rs4645943	rs4785963	rs5367	rs6024836	rs663625
rs3740423	rs3792594	rs4150196	rs4645959	rs4794056	rs5368	rs6060022	rs6659104
rs3740458	rs3793784	rs4151636	rs4646759	rs4801386	rs537112	rs6060031	rs6663474
rs3740469	rs3794422	rs41741	rs4646761	rs4801387	rs542571	rs6065779	rs6663989
rs3740720	rs3794456	rs4242182	rs4646762	rs4802307	rs544115	rs6089810	rs6672420
rs3740834	rs3795005	rs4245719	rs4646999	rs4811191	rs546220	rs610308	rs6674775
rs3740961	rs3795024	rs4251545	rs4647297	rs481736	rs546449	rs611251	rs6683830
rs3741883	rs3796504	rs4252591	rs4647950	rs4817580	rs552130	rs6151532	rs6687850
rs6734462	rs7115499	rs7636357	rs813591	rs9386924	rs7110845	rs762550	rs8110904
rs6737601	rs7125607	rs763944	rs8176322	rs9399468	rs7114563	rs762624	rs8124620
rs6747918	rs713358	rs7643235	rs8176323	rs9399686	rs935888	rs9993838	rs9373592

rs6758025	rs713671	rs769214	rs8177762	rs947895
rs6759852	rs7144366	rs769242	rs8192660	rs9487736
rs6768093	rs7149482	rs7694723	rs8192919	rs949963
rs6780995	rs7157634	rs770087	rs8192920	rs9510777
rs6789961	rs717599	rs7701280	rs8192921	rs9514828
rs679620	rs717948	rs7712617	rs8192922	rs9527024
rs6803501	rs719725	rs7753051	rs8193036	rs9527032
rs6804482	rs7203193	rs775886	rs832582	rs9536234
rs682632	rs7208804	rs7760247	rs836478	rs9536314
rs683878	rs7216758	rs7804092	rs8473	rs9550987
rs685724	rs721933	rs7809365	rs848216	rs955461
rs686050	rs7221626	rs7820245	rs849370	rs9595305
rs6871703	rs723082	rs7823775	rs849397	rs9610645
rs6874525	rs7248144	rs783039	rs85276	rs9610775
rs688644	rs7249075	rs7831	rs869800	rs9619311
rs6893184	rs7251	rs7845219	rs870876	rs965823
rs6896163	rs7252927	rs7848647	rs871775	rs9658786
rs689647	rs7254617	rs785467	rs874881	rs9666607
rs6897932	rs7254951	rs7872802	rs876064	rs9668896
rs6916861	rs7256321	rs7903146	rs876581	rs9676881
rs6948467	rs7259764	rs7918199	rs880633	rs971667
rs6950861	rs7271169	rs7920256	rs881150	rs9787810
rs6981737	rs7316466	rs7928549	rs884195	rs9803935
rs6983267	rs732072	rs7943716	rs884209	rs9811027
rs698893	rs733455	rs796050	rs887829	rs9825571
rs6989717	rs7337765	rs7962818	rs888197	rs9831563
rs6997891	rs734249	rs7964641	rs895808	rs9834996
rs7002972	rs738304	rs7970377	rs900358	rs9838117
rs7005244	rs741233	rs7980903	rs900818	rs9842580
rs7008482	rs743855	rs7985423	rs901746	rs9857232
rs701848	rs746978	rs7997728	rs903357	rs9859259
rs7022613	rs7479101	rs799917	rs903358	rs9859413
rs702689	rs751128	rs8003631	rs904453	rs9866051
rs703473	rs7515776	rs8024471	rs905238	rs987106
rs704073	rs7523416	rs8027765	rs909691	rs9876116
rs7061360	rs7538978	rs8028241	rs9170	rs9879080
rs706714	rs7544152	rs8033436	rs919400	rs9886732
rs706931	rs7545293	rs8045	rs923370	rs9886857
rs7071768	rs7556852	rs8049033	rs9245	rs9887812
rs7073830	rs755740	rs8054314	rs92535	rs9894946
rs708035	rs755804	rs8056349	rs926328	rs9897994
rs7080373	rs757110	rs8056505	rs9282734	rs989902
rs7081726	rs7594480	rs805657	rs9292617	rs991967
rs7083622	rs7595492	rs8064138	rs9306198	rs9944966
rs709222	rs760783	rs809821	rs931643	rs9951523
rs7095325	rs7613767	rs8101863	rs9325604	rs9957673
rs7103750	rs7621556	rs8109165	rs9326069	rs997983
rs7104301	rs7624750	rs8110238	rs9344968	rs998074

Survey provided to the patients before attending colonoscopy:

Patientenerfassungsbogen

Hier Aufkleber:

Allgemeine Angaben

Geschlecht	○ männlich	○ weiblich
Geburtsjahr		

Fragen zu Vorsorgeuntersuchungen

Wurde bei Ihnen früher schon einmal eine Darmspiegelung gemacht?	○ Nein	○ Ja
Wenn ja.....	Wann zuletzt?	
	Wie oft bisher?	
	Wurden bei Ihnen jemals Darmpolypen festgestellt?	○ Ja ○ Nein ○ weiß nicht

Lassen Sie diese Darmspiegelung aufgrund von Beschwerden oder auffälligen Veränderungen durchführen?	○ Nein	○ Ja
Wenn ja, welcher Art sind diese Beschwerden?	○ Blut im Stuhl/ positiver Bluttest	
	○ Veränderungen der früheren Stuhlgewohnheiten (z.B. Durchfall, Verstopfung, Beschaffenheit)	
	○ Anhaltende Bauchschmerzen/ -krämpfe	
	○ Sonstige Beschwerden	

Medikamenteneinnahme

Nehmen Sie regelmäßig folgende Medikamente ein?		
Schmerzmittel und/ oder Rheumamittel?	○ nein	○ ja
Welche oder welches Mittel?		
Seit wann?		
Dosis:		
Nehmen Sie Thrombozytenaggregationshemmer? (z.B. ASS, Godamed, Herz-ASS, Plavix)	○ nein	○ ja
Wenn ja, welches Mittel war es?		
Seit wann?		
Dosis:		

Vorerkrankungen in der Familie

Hatte Ihre Mutter /Ihr Vater jemals Darmkrebs?					
Mutter	○ nein	○ Weiß nicht	**Vater**	○ nein	○ Weiß nicht
○ Ja, Alter bei Diagnose:			○ Ja, Alter bei Diagnose:		

Hatte Ihre Mutter/ Ihr Vater jemals eine andere Krebserkrankung?		
Mutter	○nein	○ ja
Wenn ja, welche?	○ Magen ○ Pankreas ○ Speiseröhre ○ Sonstige (bitte Eintragen) ✎	
Vater	○ nein	○ ja
Wenn ja, welche?	○ Magen ○ Pankreas ○ Speiseröhre ○ Sonstige (bitte Eintragen) ✎	
Kommentare		

Risikofaktoren und Lebensgewohnheiten

Haben Sie in Ihrem Leben jemals regelmäßig, d.h. täglich über mindestens 1 Jahr, geraucht?			
○ nein	○ ja		
Wie lange/ Seit wann?		Zahl der Päckchen pro Tag?	

An wie vielen Tagen pro Woche haben Sie in den letzten 12 Monaten üblicherweise Alkohol getrunken?
✎ (Tage pro Woche):

Wie oft haben Sie <u>in den letzten 12 Monaten</u> im Durchschnitt folgende Nahrungsmittel zu sich genommen?						
	mehrmals täglich	1x pro Tag	mehrmals pro Woche	1x pro Woche	seltener	nie
Wurstwaren						
Geflügelfleisch						
Sonst. Fleisch (Schwein, Rind, Lamm)						
Vollkornbrot						
Obst						
Gemüse, Salat						

Körperliche Betätigung: Wie viele Stunden pro Woche haben Sie sich <u>in den letzen 12 Monaten</u> durchschnittlich folgendermaßen körperlich betätigt?	
Körperlich anstrengende Arbeit (z.B. Landwirtschaft, als Bauarbeiter, Alten- und Krankenpflege……)	Stunden
Körperlich anstrengende sportliche Tätigkeit (z.B. Fußball, Schwimmen, Skifahren, sportl. Radfahren, Joggen….)	Stunden
Leichtere, vorwiegend gehende oder stehende Arbeit (z.B. Hausarbeit, Gartenarbeit, Verkäufer……)	Stunden
Leichte Betätigung (z.B. zu Fuß, zur Arbeit gehen, zum Einkaufen gehen, Spaziergänge, Radfahren…….)	Stunden
✎ Kommentare?	

List of genes analysed in the chip

Approved symbol	Approved Name
AATF	apoptosis antagonizing transcription factor
ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
ABL1	c-abl oncogene 1, non-receptor tyrosine kinase
ACIN1	apoptotic chromatin condensation inducer 1
ACVR1C	activin A receptor, type IC
ADAM12	ADAM metallopeptidase domain 12
ADAMTSL4	ADAMTS-like 4
AES	amino-terminal enhancer of split
AIFM2	apoptosis-inducing factor, mitochondrion-associated, 2
AIFM3	apoptosis-inducing factor, mitochondrion-associated, 3
AKR1C1	aldo-keto reductase family 1, member C1
AKT1	v-akt murine thymoma viral oncogene homolog 1
AKT2	v-akt murine thymoma viral oncogene homolog 2
AKT3	v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)
ALCAM	activated leukocyte cell adhesion molecule
ALDH1L1	aldehyde dehydrogenase 1 family, member L1
ALS2CR2	STE20-related kinase adaptor beta
ANKHD1	ankyrin repeat and KH domain containing 1
APBB2	amyloid beta (A4) precursor protein-binding, family B, member 2
APC	adenomatous polyposis coli
APCS	amyloid P component, serum
APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1
APIP	APAF1 interacting protein
APITD1	apoptosis-inducing, TAF9-like domain 1
APOL6	apolipoprotein L, 6
APPBP1	NEDD8 activating enzyme E1 subunit 1
APR-2	family with sequence similarity 215, member A (non-protein coding)
ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta
ARID4B	AT rich interactive domain 4B (RBP1-like)
ASAH2	N-acylsphingosine amidohydrolase (non-lysosomal ceramidase) 2
ATF3	activating transcription factor 3
ATF6	activating transcription factor 6
ATG5	ATG5 autophagy related 5 homolog (S. cerevisiae)
ATM	ataxia telangiectasia mutated
ATOH1	atonal homolog 1 (Drosophila)
ATP2B4	ATPase, Ca++ transporting, plasma membrane 4
ATP6V0A2	ATPase, H+ transporting, lysosomal V0 subunit a2
ATR	ataxia telangiectasia and Rad3 related
AURKA	aurora kinase A
AVEN	apoptosis, caspase activation inhibitor
AXIN1	axin 1
AXL	AXL receptor tyrosine kinase
AZI2	5-azacytidine induced 2
BAD	BCL2-associated agonist of cell death
BAK1	BCL2-antagonist/killer 1

BAX	BCL2-associated X protein
BCL10	B-cell CLL/lymphoma 10
BCL2	B-cell CLL/lymphoma 2
BCL2A1	BCL2-related protein A1
BCL2L	BCL2-like 1
BCL2L12	BCL2-like 12 (proline rich)
BCL2L13	BCL2-like 13 (apoptosis facilitator)
BCL2L14	BCL2-like 14 (apoptosis facilitator)
BCL2L7P1	BCL2-antagonist/killer 1 pseudogene 1
BCL6	B-cell CLL/lymphoma 6
BCOR	BCL6 corepressor
BCR	breakpoint cluster region
BIK	BCL2-interacting killer (apoptosis-inducing)
BIRC2	baculoviral IAP repeat containing 2
BIRC4	X-linked inhibitor of apoptosis
BIRC5	baculoviral IAP repeat containing 5
BIRC6	baculoviral IAP repeat containing 6
BIRC8	baculoviral IAP repeat containing 8
BLR1	chemokine (C-X-C motif) receptor 5
BMF	Bcl2 modifying factor
BMP6	bone morphogenetic protein 6
BNIP1	BCL2/adenovirus E1B 19kDa interacting protein 1
BNIP2	BCL2/adenovirus E1B 19kDa interacting protein 2
BNIP3L	BCL2/adenovirus E1B 19kDa interacting protein 3-like
BNIPL	BCL2/adenovirus E1B 19kD interacting protein like
BPTF	bromodomain PHD finger transcription factor
BRCA1	breast cancer 1, early onset
BRCC2	BH3-like motif containing, cell death inducer
BRD7	bromodomain containing 7
BRE	brain and reproductive organ-expressed (TNFRSF1A modulator)
BREA2	breast cancer estrogen-induced apoptosis
BTG1	B-cell translocation gene 1, anti-proliferative
BTK	Bruton agammaglobulinemia tyrosine kinase
C12orf5	chromosome 12 open reading frame 5
CAMLG	calcium modulating ligand
CAPN10	calpain 10
CAPN2	calpain 2, (m/II) large subunit
CAPN5	calpain 5
CARD10	caspase recruitment domain family, member 10
CARD11	caspase recruitment domain family, member 11
CARD14	caspase recruitment domain family, member 14
CARD6	caspase recruitment domain family, member 6
CARD8	caspase recruitment domain family, member 8
CASP1	caspase 1, apoptosis-related cysteine peptidase
CASP10	caspase 10, apoptosis-related cysteine peptidase
CASP14	caspase 14, apoptosis-related cysteine peptidase
CASP2	caspase 2, apoptosis-related cysteine peptidase
CASP3	caspase 3, apoptosis-related cysteine peptidase
CASP4	caspase 4 apoptosis-related cysteine peptidase

CASP5	caspase 5, apoptosis-related cysteine peptidase
CASP6	caspase 6, apoptosis-related cysteine peptidase
CASP7	caspase 7, apoptosis-related cysteine peptidase
CASP8	caspase 8, apoptosis-related cysteine peptidase
CASP8AP2	caspase 8 associated protein 2
CASP9	caspase 9, apoptosis-related cysteine peptidase
CAT	catalase
CBL	Cbl proto-oncogene, E3 ubiquitin protein ligase
CCAR1	cell division cycle and apoptosis regulator 1
CCL19	chemokine (C-C motif) ligand 19
CCNA1	cyclin A1
CCNA2	cyclin A2
CCNE1	cyclin E1
CCNK	cyclin K
CD247	CD247 molecule
CD3E	CD3e molecule, epsilon (CD3-TCR complex)
CD40LG	CD40 ligand
CD44	CD44 molecule (Indian blood group)
CD53	CD53 molecule
CD70	CD70 molecule
CD74	CD74 molecule, major histocompatibility complex, class II invariant chain
CDC2	cyclin-dependent kinase 1
CDC2L1	cyclin-dependent kinase 11B
CDC2L2	cyclin-dependent kinase 11A
CDC2L5	cyclin-dependent kinase 13
CDC2L6	cyclin-dependent kinase 19
CDC6	cell division cycle 6 homolog (S. cerevisiae)
CDGAP	Rho GTPase activating protein 31
CDK2	cyclin-dependent kinase 2
CDK5	cyclin-dependent kinase 5
CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)
CDKN2A	cyclin-dependent kinase inhibitor 2A
CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
CENTG1	ArfGAP with GTPase domain, ankyrin repeat and PH domain 2
CFLAR	CASP8 and FADD-like apoptosis regulator
CHEK1	checkpoint kinase 1
CHEK2	checkpoint kinase 2
CHI3L1	chitinase 3-like 1 (cartilage glycoprotein-39)
CHUK	conserved helix-loop-helix ubiquitous kinase
CIAPIN1	cytokine induced apoptosis inhibitor 1
CIB1	calcium and integrin binding 1 (calmyrin)
CLN6	ceroid-lipofuscinosis, neuronal 6, late infantile, variant
CLU	clusterin
CNOT2	CCR4-NOT transcription complex, subunit 2
CNR2	cannabinoid receptor 2 (macrophage)
COP1	caspase recruitment domain family, member 16
CRADD	CASP2 and RIPK1 domain containing adaptor with death domain
CRP	C-reactive protein, pentraxin-related

CRYAB	crystallin, alpha B
CSE1L	CSE1 chromosome segregation 1-like (yeast)
CSNK1A1	casein kinase 1, alpha 1
CSNK1E	casein kinase 1, epsilon
CTNNBL1	catenin, beta like 1
CTNND1	catenin (cadherin-associated protein), delta 1
CTSB	cathepsin B
CTSL1	cathepsin L1
CUGBP2	CUGBP, Elav-like family member 2
CUL3	cullin 3
CUL4A	cullin 4A
CUL5	cullin 5
CXCL13	chemokine (C-X-C motif) ligand 13
CYC1	cytochrome c-1
CYCS	cytochrome c, somatic
CYFIP2	cytoplasmic FMR1 interacting protein 2
CYLD	cylindromatosis (turban tumor syndrome)
CYR61	cysteine-rich, angiogenic inducer, 61
DAD1	defender against cell death 1
DAPK1	death-associated protein kinase 1
DAPK2	death-associated protein kinase 2
DAPK3	death-associated protein kinase 3
DBNL	drebrin-like
DCC	deleted in colorectal carcinoma
DDB2	damage-specific DNA binding protein 2, 48kDa
DDX20	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20
DEDD	death effector domain containing
DEDD2	death effector domain containing 2
DIABLO	diablo, IAP-binding mitochondrial protein
DIDO1	death inducer-obliterator 1
DIP	GRAM domain containing 4
DLL1	delta-like 1 (Drosophila)
DNAJA3	DnaJ (Hsp40) homolog, subfamily A, member 3
DNASE1	deoxyribonuclease I
DNASE1L3	deoxyribonuclease I-like 3
DNM2	dynamin 2
DOCK1	dedicator of cytokinesis 1
DUSP1	dual specificity phosphatase 1
DUSP16	dual specificity phosphatase 16
DUSP4	dual specificity phosphatase 4
DUSP6	dual specificity phosphatase 6
DYRK2	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2
E2F1	E2F transcription factor 1
EAF2	ELL associated factor 2
EDA2R	ectodysplasin A2 receptor
EDAR	ectodysplasin A receptor
EEF1E1	eukaryotic translation elongation factor 1 epsilon 1
EGF	epidermal growth factor
EGLN3	egl nine homolog 3 (C. elegans)

EGR1	early growth response 1
EGR2	early growth response 2
EI24	etoposide induced 2.4 mRNA
ELMO1	engulfment and cell motility 1
ENDOG	endonuclease G
EP400	E1A binding protein p400
EPHA2	EPH receptor A2
ERCC2	excision repair cross-complementing rodent repair deficiency, complementation grou
ERCC6	excision repair cross-complementing rodent repair deficiency, complementation grou
ERN1	endoplasmic reticulum to nucleus signaling 1
ERN2	endoplasmic reticulum to nucleus signaling 2
FADD	Fas (TNFRSF6)-associated via death domain
FAF1	Fas (TNFRSF6) associated factor 1
FAIM	Fas apoptotic inhibitory molecule
FAIM2	Fas apoptotic inhibitory molecule 2
FAIM3	Fas apoptotic inhibitory molecule 3
FAM130A1	cysteine-serine-rich nuclear protein 2
FANCD2	Fanconi anemia, complementation group D2
FAS	Fas (TNF receptor superfamily, member 6)
FASLG	Fas ligand (TNF superfamily, member 6)
FGFR3	fibroblast growth factor receptor 3
FHIT	fragile histidine triad
FIS1	fission 1 (mitochondrial outer membrane) homolog (S. cerevisiae)
FOXO1	forkhead box O1
FOXO3	forkhead box O3
FOXO4	forkhead box O4
FRAG1	ATPase family, AAA domain containing 5
FRAP1	mechanistic target of rapamycin (serine/threonine kinase)
FXN	frataxin
FYN	FYN oncogene related to SRC, FGR, YES
GAB1	GRB2-associated binding protein 1
GADD45A	growth arrest and DNA-damage-inducible, alpha
GADD45B	growth arrest and DNA-damage-inducible, beta
GADD45G	growth arrest and DNA-damage-inducible, gamma
GATA6	GATA binding protein 6
GLRX	glutaredoxin (thioltransferase)
GML	glycosylphosphatidylinositol anchored molecule like protein
GNLY	granulysin
GORASP1	golgi reassembly stacking protein 1, 65kDa
GPR132	G protein-coupled receptor 132
GPR65	G protein-coupled receptor 65
GPX1	glutathione peroxidase 1
GPX4	glutathione peroxidase 4
GSK3B	glycogen synthase kinase 3 beta
GSTM1	glutathione S-transferase mu 1
GSTP1	glutathione S-transferase pi 1
GTSE1	G-2 and S-phase expressed 1
GULP1	GULP, engulfment adaptor PTB domain containing 1
GZMA	granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3)

GZMB	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)
HBEGF	heparin-binding EGF-like growth factor
HBXIP	late endosomal/lysosomal adaptor, MAPK and MTOR activator 5
HCLS1	hematopoietic cell-specific Lyn substrate 1
HDAC3	histone deacetylase 3
HDAC4	histone deacetylase 4
HDAC7A	histone deacetylase 7
HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
HIPK3	homeodomain interacting protein kinase 3
HIST4H4	histone cluster 4, H4
HMGB1	high mobility group box 1
HSP90AB1	heat shock protein 90kDa alpha (cytosolic), class B member 1
HSPA1B	heat shock 70kDa protein 1B
HSPB1	heat shock 27kDa protein 1
HSPB2	heat shock 27kDa protein 2
HSPBP1	HSPA (heat shock 70kDa) binding protein, cytoplasmic cochaperone 1
HTATIP2	HIV-1 Tat interactive protein 2, 30kDa
HTRA2	HtrA serine peptidase 2
HUWE1	HECT, UBA and WWE domain containing 1
HYOU1	hypoxia up-regulated 1
ICEBERG	caspase recruitment domain family, member 18
IFIH1	interferon induced with helicase C domain 1
IFN1	interferon, type 1, cluster
IFNG	interferon, gamma
IGF1	insulin-like growth factor 1 (somatomedin C)
IGF2R	insulin-like growth factor 2 receptor
IGFBP3	insulin-like growth factor binding protein 3
IGFBP5	insulin-like growth factor binding protein 5
IHPK2	inositol hexakisphosphate kinase 2
IKIP	IKBKB interacting protein
IL15RA	interleukin 15 receptor, alpha
IL17A	interleukin 17A
IL17RD	interleukin 17 receptor D
IL18	interleukin 18 (interferon-gamma-inducing factor)
IL19	interleukin 19
IL1A	interleukin 1, alpha
IL1B	interleukin 1, beta
IL1R1	interleukin 1 receptor, type I
IL2	interleukin 2
IL24	interleukin 24
IL6	interleukin 6 (interferon, beta 2)
IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)
IL7	interleukin 7
IL7R	interleukin 7 receptor
IL8RB	chemokine (C-X-C motif) receptor 2
INCA	caspase recruitment domain family, member 17
ING3	inhibitor of growth family, member 3
IRAK2	interleukin-1 receptor-associated kinase 2
IRAK4	interleukin-1 receptor-associated kinase 4

IRF1	interferon regulatory factor 1
IRF3	interferon regulatory factor 3
ISG20L1	apoptosis enhancing nuclease
ITPR1	inositol 1,4,5-trisphosphate receptor, type 1
JAK1	Janus kinase 1
JAK2	Janus kinase 2
JMY	junction mediating and regulatory protein, p53 cofactor
JUN	jun proto-oncogene
KCNIP3	Kv channel interacting protein 3, calsenilin
KEAP1	kelch-like ECH-associated protein 1
KIAA0513	KIAA0513
KIF11	kinesin family member 11
KL	klotho
KNG1	kininogen 1
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LGALS1	lectin, galactoside-binding, soluble, 1
LGALS3	lectin, galactoside-binding, soluble, 3
LGALS8	lectin, galactoside-binding, soluble, 8
LITAF	lipopolysaccharide-induced TNF factor
LOC116143	LOC116143
LOC441931	LOC441931
LOC651771	LOC651771
LRDD	p53-induced death domain protein
LTA	lymphotoxin alpha (TNF superfamily, member 1)
LTBR	lymphotoxin beta receptor (TNFR superfamily, member 3)
MADD	MAP-kinase activating death domain
MAGEH1	melanoma antigen family H, 1
MAGI3	membrane associated guanylate kinase, WW and PDZ domain containing 3
MALT1	mucosa associated lymphoid tissue lymphoma translocation gene 1
MAP1S	microtubule-associated protein 1S
MAP2K1	mitogen-activated protein kinase kinase 1
MAP2K2	mitogen-activated protein kinase kinase 2
MAP2K3	mitogen-activated protein kinase kinase 3
MAP2K4	mitogen-activated protein kinase kinase 4
MAP3K1	mitogen-activated protein kinase kinase kinase 1
MAP3K10	mitogen-activated protein kinase kinase kinase 10
MAP3K5	mitogen-activated protein kinase kinase kinase 5
MAP3K7	mitogen-activated protein kinase kinase kinase 7
MAP3K7IP1	TGF-beta activated kinase 1/MAP3K7 binding protein 1
MAP3K7IP2	TGF-beta activated kinase 1/MAP3K7 binding protein 2
MAPK10	mitogen-activated protein kinase 10
MAPK11	mitogen-activated protein kinase 11
MAPK8	mitogen-activated protein kinase 8
MAPK8IP1	mitogen-activated protein kinase 8 interacting protein 1
MAPK9	mitogen-activated protein kinase 9
MBD4	methyl-CpG binding domain protein 4
MCL1	myeloid cell leukemia sequence 1 (BCL2-related)
MCTS1	malignant T cell amplified sequence 1
MDM4	Mdm4 p53 binding protein homolog (mouse)

MET	met proto-oncogene (hepatocyte growth factor receptor)
MIF	macrophage migration inhibitory factor (glycosylation-inhibiting factor)
MKI67	antigen identified by monoclonal antibody Ki-67
MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
MMP1	matrix metallopeptidase 1 (interstitial collagenase)
MMP3	matrix metallopeptidase 3 (stromelysin 1, progelatinase)
MMP7	matrix metallopeptidase 7 (matrilysin, uterine)
MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase
MOAP1	modulator of apoptosis 1
MRPL37	mitochondrial ribosomal protein L37
MSH2	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)
MSH6	mutS homolog 6 (E. coli)
MSX2	msh homeobox 2
MYB	v-myb myeloblastosis viral oncogene homolog (avian)
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)
MYCT1	myc target 1
MYD88	myeloid differentiation primary response gene (88)
NBN	nibrin
NCAM1	neural cell adhesion molecule 1
NDRG1	N-myc downstream regulated 1
NEU3	sialidase 3 (membrane sialidase)
NFATC2	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2
NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
NGFB	nerve growth factor (beta polypeptide)
NGFR	nerve growth factor receptor
NGFRAP1	nerve growth factor receptor (TNFRSF16) associated protein 1
NLRC4	NLR family, CARD domain containing 4
NLRC5	NLR family, CARD domain containing 5
NLRP10	NLR family, pyrin domain containing 10
NLRP11	NLR family, pyrin domain containing 11
NLRP13	NLR family, pyrin domain containing 13
NLRP2	NLR family, pyrin domain containing 2
NLRP3	NLR family, pyrin domain containing 3
NLRP4	NLR family, pyrin domain containing 4
NLRP5	NLR family, pyrin domain containing 5
NLRP6	NLR family, pyrin domain containing 6
NLRP7	NLR family, pyrin domain containing 7
NLRP8	NLR family, pyrin domain containing 8
NLRP9	NLR family, pyrin domain containing 9
NME1	non-metastatic cells 1, protein (NM23A) expressed in
NOD1	nucleotide-binding oligomerization domain containing 1
NOTCH2	notch 2
NOX4	NADPH oxidase 4
NR3C1	nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)
NSD1	nuclear receptor binding SET domain protein 1
NTN1	netrin 1
NUAK1	NUAK family, SNF1-like kinase, 1
NUAK2	NUAK family, SNF1-like kinase, 2
NUPR1	nuclear protein, transcriptional regulator, 1

OGG1	8-oxoguanine DNA glycosylase
OLR1	oxidized low density lipoprotein (lectin-like) receptor 1
OPA1	optic atrophy 1 (autosomal dominant)
OSGIN1	oxidative stress induced growth inhibitor 1
P2RX1	purinergic receptor P2X, ligand-gated ion channel, 1
P53AIP1	tumor protein p53 regulated apoptosis inducing protein 1
PADI4	peptidyl arginine deiminase, type IV
PAK1	p21 protein (Cdc42/Rac)-activated kinase 1
PAK6	p21 protein (Cdc42/Rac)-activated kinase 6
PARK2	parkinson protein 2, E3 ubiquitin protein ligase (parkin)
PARP1	poly (ADP-ribose) polymerase 1
PARVA	parvin, alpha
PAWR	PRKC, apoptosis, WT1, regulator
PCNA	proliferating cell nuclear antigen
PDCD11	programmed cell death 11
PDCD2	programmed cell death 2
PDCD2L	programmed cell death 2-like
PDCD4	programmed cell death 4 (neoplastic transformation inhibitor)
PDCD6	programmed cell death 6
PDCD6IP	programmed cell death 6 interacting protein
PDCL3	phosducin-like 3
PDE4B	phosphodiesterase 4B, cAMP-specific
PEBP1	phosphatidylethanolamine binding protein 1
PEBP4	phosphatidylethanolamine-binding protein 4
PEG10	paternally expressed 10
PEG3	paternally expressed 3
PERP	PERP, TP53 apoptosis effector
PGR	progesterone receptor
PHB	prohibitin
PHF17	PHD finger protein 17
PHLDA1	pleckstrin homology-like domain, family A, member 1
PHLPP	PH domain and leucine rich repeat protein phosphatase 1
PIAS1	protein inhibitor of activated STAT, 1
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit beta
PIK3CG	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit gamma
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
PIK3R3	phosphoinositide-3-kinase, regulatory subunit 3 (gamma)
PIK3R5	phosphoinositide-3-kinase, regulatory subunit 5
PINK1	PTEN induced putative kinase 1
PLA2G6	phospholipase A2, group VI (cytosolic, calcium-independent)
PLAGL1	pleiomorphic adenoma gene-like 1
PLAUR	plasminogen activator, urokinase receptor
PLEKHF1	pleckstrin homology domain containing, family F (with FYVE domain) member 1
PLEKHO1	pleckstrin homology domain containing, family O member 1
PLK1	polo-like kinase 1
PLK4	polo-like kinase 4
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1
PML	promyelocytic leukemia

PMS2	PMS2 postmeiotic segregation increased 2 (S. cerevisiae)
POU2F3	POU class 2 homeobox 3
PPARD	peroxisome proliferator-activated receptor delta
PPM1F	protein phosphatase, Mg2+/Mn2+ dependent, 1F
PPP1R13B	protein phosphatase 1, regulatory subunit 13B
PPP1R13L	protein phosphatase 1, regulatory subunit 13 like
PPP1R15A	protein phosphatase 1, regulatory subunit 15A
PPP2CA	protein phosphatase 2, catalytic subunit, alpha isozyme
PPP5C	protein phosphatase 5, catalytic subunit
PRDX1	peroxiredoxin 1
PRDX3	peroxiredoxin 3
PRDX5	peroxiredoxin 5
PRF1	perforin 1 (pore forming protein)
PRIMA1	proline rich membrane anchor 1
PRKAR1A	protein kinase, cAMP-dependent, regulatory, type I, alpha
PRKCB1	protein kinase C, beta
PRKCD	protein kinase C, delta
PRKCE	protein kinase C, epsilon
PRKD1	protein kinase D1
PRKRIR	protein-kinase, interferon-inducible double stranded RNA dependent inhibitor, repre
PRL	prolactin
PSEN2	presenilin 2 (Alzheimer disease 4)
PTAFR	platelet-activating factor receptor
PTEN	phosphatase and tensin homolog
PTGES	prostaglandin E synthase
PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxyge
PTMA	prothymosin, alpha
PTPN13	protein tyrosine phosphatase, non-receptor type 13 (APO-1/CD95 (Fas)-associated ph
PTPRC	protein tyrosine phosphatase, receptor type, C
PYCARD	PYD and CARD domain containing
QSOX2	quiescin Q6 sulfhydryl oxidase 2
RAB25	RAB25, member RAS oncogene family
RAC1	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac
RAD21	RAD21 homolog (S. pombe)
RAD51L1	RAD51 homolog B (S. cerevisiae)
RAD9A	RAD9 homolog A (S. pombe)
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1
RAPGEF1	Rap guanine nucleotide exchange factor (GEF) 1
RARG	retinoic acid receptor, gamma
RASA1	RAS p21 protein activator (GTPase activating protein) 1
RASSF1	Ras association (RalGDS/AF-6) domain family member 1
RB1	retinoblastoma 1
RBL2	retinoblastoma-like 2 (p130)
RBM5	RNA binding motif protein 5
RBMX	RNA binding motif protein, X-linked
RELA	v-rel reticuloendotheliosis viral oncogene homolog A (avian)
RELB	v-rel reticuloendotheliosis viral oncogene homolog B
RERE	arginine-glutamic acid dipeptide (RE) repeats
RIPK1	receptor (TNFRSF)-interacting serine-threonine kinase 1

RIPK2	receptor-interacting serine-threonine kinase 2
RIPK3	receptor-interacting serine-threonine kinase 3
RNASEL	ribonuclease L (2',5'-oligoisoadenylate synthetase-dependent)
RND3	Rho family GTPase 3
RNF41	ring finger protein 41
RNF7	ring finger protein 7
ROCK1	Rho-associated, coiled-coil containing protein kinase 1
RPL23	ribosomal protein L23
RPS13	ribosomal protein S13
RRAGA	Ras-related GTP binding A
RRAGC	Ras-related GTP binding C
RTKN	rhotekin
RUNX3	runt-related transcription factor 3
RXRA	retinoid X receptor, alpha
S100A11	S100 calcium binding protein A11
S100A8	S100 calcium binding protein A8
S100A9	S100 calcium binding protein A9
SAP30BP	SAP30 binding protein
SART1	squamous cell carcinoma antigen recognized by T cells
SCYE1	aminoacyl tRNA synthetase complex-interacting multifunctional protein 1
SDC2	syndecan 2
SELE	selectin E
SERPINB4	serpin peptidase inhibitor, clade B (ovalbumin), member 4
SERPINB9	serpin peptidase inhibitor, clade B (ovalbumin), member 9
SFRP2	secreted frizzled-related protein 2
SFRP4	secreted frizzled-related protein 4
SGPL1	sphingosine-1-phosphate lyase 1
SH3GLB1	SH3-domain GRB2-like endophilin B1
SHB	Src homology 2 domain containing adaptor protein B
SIAH1	seven in absentia homolog 1 (Drosophila)
SIVA1	SIVA1, apoptosis-inducing factor
SLAMF1	signaling lymphocytic activation molecule family member 1
SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), mem
SLK	STE20-like kinase
SMAD1	SMAD family member 1
SMAD2	SMAD family member 2
SMAD3	SMAD family member 3
SMAD7	SMAD family member 7
SMC6	structural maintenance of chromosomes 6
SMNDC1	survival motor neuron domain containing 1
SMPD1	sphingomyelin phosphodiesterase 1, acid lysosomal
SNCA	synuclein, alpha (non A4 component of amyloid precursor)
SON	SON DNA binding protein
SP1	Sp1 transcription factor
SPDYA	speedy homolog A (Xenopus laevis)
SPIN2A	spindlin family, member 2A
SPINK7	serine peptidase inhibitor, Kazal type 7 (putative)
SRGN	serglycin
SSTR3	somatostatin receptor 3

STAG1	stromal antigen 1
STAT1	signal transducer and activator of transcription 1, 91kDa
STK17A	serine/threonine kinase 17a
STK17B	serine/threonine kinase 17b
STK3	serine/threonine kinase 3
STK4	serine/threonine kinase 4
SULF1	sulfatase 1
SUMO3	SMT3 suppressor of mif two 3 homolog 3 (S. cerevisiae)
SUPV3L1	suppressor of var1, 3-like 1 (S. cerevisiae)
SYK	spleen tyrosine kinase
SYVN1	synovial apoptosis inhibitor 1, synoviolin
TADA3L	transcriptional adaptor 3
TAX1BP1	Tax1 (human T-cell leukemia virus type I) binding protein 1
TEGT	transmembrane BAX inhibitor motif containing 6
TEK	TEK tyrosine kinase, endothelial
TERT	telomerase reverse transcriptase
TFAP2A	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)
TGFB2	transforming growth factor, beta 2
TGFBI	transforming growth factor, beta-induced, 68kDa
TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-glutamyltransferase)
THAP2	THAP domain containing, apoptosis associated protein 2
THAP3	THAP domain containing, apoptosis associated protein 3
THYN1	thymocyte nuclear protein 1
TIAL1	TIA1 cytotoxic granule-associated RNA binding protein-like 1
TIMP1	TIMP metallopeptidase inhibitor 1
TIMP3	TIMP metallopeptidase inhibitor 3
TLR6	toll-like receptor 6
TMEM123	transmembrane protein 123
TNF	tumor necrosis factor
TNFAIP3	tumor necrosis factor, alpha-induced protein 3
TNFAIP8	tumor necrosis factor, alpha-induced protein 8
TNFRSF10A	tumor necrosis factor receptor superfamily, member 10a
TNFRSF10B	tumor necrosis factor receptor superfamily, member 10b
TNFRSF10C	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain
TNFRSF14	tumor necrosis factor receptor superfamily, member 14
TNFRSF19	tumor necrosis factor receptor superfamily, member 19
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A
TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B
TNFRSF21	tumor necrosis factor receptor superfamily, member 21
TNFRSF25	tumor necrosis factor receptor superfamily, member 25
TNFSF10	tumor necrosis factor (ligand) superfamily, member 10
TNFSF13B	tumor necrosis factor (ligand) superfamily, member 13b
TNFSF15	tumor necrosis factor (ligand) superfamily, member 15
TNFSF8	tumor necrosis factor (ligand) superfamily, member 8
TNFSF9	tumor necrosis factor (ligand) superfamily, member 9
TNS4	tensin 4
TOP1	topoisomerase (DNA) I
TOP2A	topoisomerase (DNA) II alpha 170kDa
TP53	tumor protein p53

TP53BP1	tumor protein p53 binding protein 1
TP53BP2	tumor protein p53 binding protein, 2
TP53I11	tumor protein p53 inducible protein 11
TP53I3	tumor protein p53 inducible protein 3
TP53INP1	tumor protein p53 inducible nuclear protein 1
TP73	tumor protein p73
TP73L	tumor protein p63
TPT1	tumor protein, translationally-controlled 1
TRADD	TNFRSF1A-associated via death domain
TRAF1	TNF receptor-associated factor 1
TRAF2	TNF receptor-associated factor 2
TRAF3	TNF receptor-associated factor 3
TRAF6	TNF receptor-associated factor 6
TRAF7	TNF receptor-associated factor 7
TRAIP	TRAF interacting protein
TRIAD3	ring finger protein 216
TRIAP1	TP53 regulated inhibitor of apoptosis 1
TRIM27	tripartite motif containing 27
TRIM32	tripartite motif containing 32
TRIM35	tripartite motif containing 35
TRIM69	tripartite motif containing 69
TRPS1	trichorhinophalangeal syndrome I
TXN	thioredoxin
TXN2	thioredoxin 2
TXNIP	thioredoxin interacting protein
UBE1L	ubiquitin-like modifier activating enzyme 7
UBE3A	ubiquitin protein ligase E3A
UGCG	UDP-glucose ceramide glucosyltransferase
UNC5B	unc-5 homolog B (C. elegans)
UNC5C	unc-5 homolog C (C. elegans)
UNC5D	unc-5 homolog D (C. elegans)
USP7	ubiquitin specific peptidase 7 (herpes virus-associated)
VAV1	vav 1 guanine nucleotide exchange factor
VCP	valosin containing protein
WDR35	WD repeat domain 35
WEE1	WEE1 homolog (S. pombe)
WISP1	WNT1 inducible signaling pathway protein 1
WT1	Wilms tumor 1
WWOX	WW domain containing oxidoreductase
XAF1	XIAP associated factor 1
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1
YAP1	Yes-associated protein 1
YWHAH	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta pol
ZAK	regulator of calcineurin 2
ZBTB16	zinc finger and BTB domain containing 16
ZBTB17	zinc finger and BTB domain containing 17
ZFAND5	zinc finger, AN1-type domain 5
ZFP36L2	zinc finger protein 36, C3H type-like 2
ZMYM4	zinc finger, MYM-type 4

ZNF148	zinc finger protein 148
ZNF346	zinc finger protein 346
ZNF443	zinc finger protein 443

AKNOWLEDEGMENTS

It has been a long journey and I am in debt with many people I met among the way.

First of all, I want to thank Prof. Neumaier, who gave me the opportunity to work in his laboratory. Since the very beginning he was thrilled by this project, and could easily pass on his enthusiasm to everyone involved in it. We spent several hours discussing all the aspects of the research, trying to evaluate all the possibilities and the implications of our discoveries. He taught me to look over the pure bioinformatics and genetic aspects of our research, and to continuously trying to have a general overview of all the biological implication of our work. I am grateful for his time, for his lessons and for his great family with whom we had the pleasure to spend some very good time during all these years.

I am grateful to Prof. Brors and Prof. Eils. Their help and knowledge have been of great support in the designing of the project and in the data analysis.

Prof. Brenner and Sabrina Hundt have been of great help providing the samples required for the project.

I am in debt with Prof. Binelli, for his comments and precious help in the statistical analysis.

I would like to thank Dr. Nittka (Steffi!) for your help, your advices, your patience and your time. You have been a very helpful figure during these years. For every problem we found on our way, you always had a "plan B" ready (and sometimes a plan C). Thank you!

Since the very first day, when they brought me to a dark-crazy-cave pub I realized I had found some amazing colleagues. I would like at first to thank Ingrid. She has helped me in all the basic and most important technical aspects of my life in Germany. With her help I could spot as apartment a castle isolated in the hills with a crazy landlord. With her help I was furthermore able from day 1 to ask, in German, for the Siemens' telephone number in Singapore. Thank you Ingrid for your invaluable help.

But she was not alone.

Miss T. helped me e.v.e.r.y time. With the time we got to know each other better and better and we spent some very good moments together. I am happy to say that she is one of my best friends. Thanks for everything.

Me, I would like to thank you. You mate, have been a precious friend in the last years. We spent some memorable time as that once at 3am under the star trying to take some pictures with fire (and beers). The force is strong with you, keep it like this! I'm glad I met you.

Diego and Flavia, my friends. I would like to thank you for your company and for your friendship (and those marvellous fried bananas).

Hani and Arash, thank you for the great Persian parties, dances and dinners we had in those years!

Teresa and Phillip. Thanks for all the time together, and for making me an (almost) excellent firefighter!

I will keep forever in my heart the Husarenstrasse party people. Arne and Daniela, the dirty-Dossenheimer Stefan with Mara cum Laude and loquace Lucas. I know right now everyone is going in different directions, but we will always find some time to party together. I miss all of you so much.

I would like to thank Francesco who spent all his last 3 summers with us in Heidelberg, M&M's and Gualtiero for the many unexpected visits, just for the weekend.

I am grateful to my parents, my brother and sister with Pietro and my parents in law. Every time they came to visit us, they brought happiness (and Italian food) in our family. Thank you.

As I said at the beginning this has been a long journey. I don't think I would have made it without my wife, Anna. She supports me, always. She comforted me in the moments of homesick, and joined me in the moments of joy. She is someone I can talk with, about anything, every day.
She is the spring of my happiness. I love you.